OECD Social, Employment and Migration Working Papers
No. 291

Not lost in translation: The implications of machine translation technologies for language professionals and for broader society

Francesca Borgonovi,

Justine Hervé,

Helke Seitz

https://dx.doi.org/10.1787/e1d1d170-en

Organisation for Economic Co-operation and Development

DELSA/ELSA/WD/SEM(2023)8

Unclassified **English text only**

27 March 2023

DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS
EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS COMMITTEE

Cancels & replaces the same document of 23 March 2023

Not lost in translation

The implications of machine translation technologies for language professionals and for broader society

OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS No.289

JEL classification: J21, J23, J28, Z13,

Authorised for publication by Stefano Scarpetta, Director, Directorate for Employment, Labour and Social Affairs.

This working paper was prepared in the framework of the OECD Skills Outlook 2023, supported by the European Commission through the Erasmus+ programme.

All Social, Employment and Migration Working Papers are now available through the OECD website at www.oecd.org/els/social-employment-and-migration-working-papers.htm

Francesca Borgonovi, Francesca.Borgonovi@oecd.org
Justine Hervé, jherve@stevens.edu
Helke Seitz, Helke.Seitz@oecd.org

JT03515213

OECD Social, Employment and Migration Working Papers

Acknowledgements

This working paper was prepared within the framework of the OECD Skills Outlook 2023, supported by the European Commission through the Erasmus+ programme. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the OECD member countries or the European Union.

The authors would like to thank OECD colleagues from the Centre for Skills, the Directorate for Science, Technology and Innovation, and the Directorate for Education and Skills, the Environment Directorate and the Development Centre for their review of the paper. In particular, they would like to thank Julie Lassébie, Fabio Manca, Luca Marcolin, El Iza Mohamedou, Mark Pearson, Quentin Vidal, Sarah Wildi, Ronald Bachmann and members of the Leibniz Institute for Economic Research in Essen for their valuable comments and feedback. The authors would also like to thank Mauro Pelucchi and Simone Perego for help with the analyses of data from European countries; Marika Boiron, Steve Dept and members of the OECD translation service for their willingness to share their views on their profession; Chiara Stramaccioni for help with the ESCO classification; Tatiana Bahous, Valentino Larcinese, Kavi Patel, and Dillon Smith from the London School of Economics for their work on the LSE MPA Capstone Report on 'Language skills in the age of Artificial Intelligence'; Jennifer Cannon and Duniya Dedeyn for their administrative and editorial work.

Abstract

The paper discusses the implications of recent advances in artificial intelligence for knowledge workers, focusing on possible complementarities and substitution between machine translation tools and language professionals. The emergence of machine translation tools could enhance social welfare through enhanced opportunities for inter-language communication but also create new threats because of persisting low levels of accuracy and quality in the translation output. The paper uses data on online job vacancies to map the evolution of the demand for language professionals between 2015 and 2019 in 10 countries and illustrates the set of skills that are considered important by employers seeking to hire language professionals through job vacancies posted on line.

Résumé

Ce document examine les implications des récentes avancées en matière d'intelligence artificielle pour un groupe spécifique de travailleurs, les « travailleurs du savoir », en se concentrant sur les possibilités de complémentarité et de substitution entre les outils de traduction automatique et les linguistes. L'apparition de ces outils pourrait être bénéfique puisqu'ils facilitent la communication entre individus de langues différentes, sans être pour autant exempte de risques, la traduction obtenue demeurant d'une précision et d'une qualité discutables. Nous nous servirons ici de données issues d'offres d'emploi publiées en ligne pour retracer l'évolution de la demande de linguistes, dans 10 pays, entre 2015 et 2019, et décrire l'ensemble des compétences considérées comme importantes par les employeurs qui cherchent à recruter des linguistes par le biais d'offres d'emploi publiées en ligne.

Table of contents

Tables

Figures

Boxes

1 Introduction

1. "*Sixty years ago, digital computers made information readable. Twenty years ago, the Internet made it reachable. Ten years ago, the first search engine crawlers made it a single database*" (Anderson, 2008[1]). Today, language translation is the new frontier of information exchange. Communication is key in today's interconnected world. Of the estimated 8 billion people who lived in the world at the end of 2022, about 5 billion were connected to the Internet. But whether on- or off-line, people's ability to access information and communicate with others depends on the language(s) they understand. Accessing information requires being able to comprehend the text and speech in which information is delivered, and information continues to be made available in a variety of languages. Therefore, making the most of the opportunities that are available due to global interconnectedness remains dependent on different economic and social agents being able to understand each other. Across the world, more than 7 000 different languages exist, although this number depends on how languages are classified – Max Weinreich famously remarked that the distinction between what is classified as a language and what is classified as a dialect is that a language is a dialect with an army – and many languages face imminent extinction because they are used by only a few, mostly elderly individuals (Bromham et al., 2021[2]; Joshi et al., 2021[3]; Potowski, 2013[4])

2. Around 88% of the world's population speak 200 different languages as their mother-tongue(s) or second language (Ethnologue - Languages of the World, n.d.[5]). Among first language speakers, Mandarin Chinese is the most widespread mother-tongue language: around 12% of the world population uses Mandarin Chinese as their first mother-tongue language. Mandarin Chinese is followed by Spanish (6%) and English (5%) (Eberhard, Simons and Fennig, 2022[6]). However, a large number of people speak English as a second language: estimates suggest that as many as 1.5 billion people worldwide could communicate in English in 2022, with the vast majority using English as a non-native language (Ethnologue - Languages of the World, n.d.[5]).

3. Exchanges of information between different agents whose primary communication medium is not the same language can take place if at least some of these agents are able to communicate in more than one language. Some individuals are able to communicate in more than one language in communities that are bi- or multi-lingual because of historical, geographical or demographic reasons or if they learnt one or more languages by participating in formal, non-formal or informal learning opportunities.

4. However, exchanges of information between agents whose primary communication medium is not the same language can also take place through mediators. The use of mediators allows individuals to communicate without the need to invest in language learning and to communicate across several language groups. Language professionals such as translators, interpreters and linguists have enabled communication between agents who otherwise could not have understood each other since the dawn of time. In the past decades, technological advances allowed individuals to use machine translation (MT) technologies as language mediators.

5. Over the past years, the landscape of machine translation tools has undergone a remarkable transformation thanks to advances in natural language processing. On one hand, the number of unique translatable language pairs increased from around 16 000 in 2019 to around 150 000 in 2022 (intento, 2022[7]). On the other hand, machine translation technologies today can translate some texts with a high level of accuracy – although the quality of translations is variable and depends on the algorithms being

used by the language models, the quantity, quality, and variety of the translations used to train the machine learning algorithms that are at the basis of artificial intelligence (AI) machine translation tools as well as the complexity of the text that is translated.

6. Technological advances shift the boundaries of what digital mediators can do and, by doing so, they change the opportunity cost for individuals to use digital rather than human mediators. The development of high-quality digital mediators entails high fixed costs initially, but after these initial investments are made, the marginal cost of developing translations of actual texts is close to zero (cost). Moreover, digital mediators offer near real-time translations (speed) of texts no matter their length (volume), and time between intention to translate and implemented translation are minimal (timeliness). By contrast, human mediators incur high initial fixed costs (because they have to develop a unique set of skills to be able to translate texts) as well as high marginal costs that accrue from engaging in each additional translation project. Human mediators also need time to be able to deliver translations, and the quality of the finished product depends on the amount of time individuals spend on the project and the experience they have accumulated over time. Table 1.1 provides an overview of core advantages and disadvantages of digital mediators and human mediators.

Table 1.1. (Dis)advantages of digital and human mediators

	Digital Mediators – AI machine translation tools	Human mediators – language professionals
Advantages	Speed, volume, cost, timeliness	Quality
Disadvantages	Quality	Speed, volume, cost, timeliness

7. The effects of the emergence of digital mediators on the language mediation market – conceived broadly in terms of the volume of information that is made available in multiple languages, as well as the employment opportunities and skills requirements of language professionals – depend on several factors. These include: 1) the quality of the language mediation delivered by digital mediators, 2) the expected returns from cross-language interactions, 3) the returns to the speed at which information exchanges occur but also on 4) the skillset of human mediators and 5) whether human mediators possess the skills needed to benefit from the enhanced productivity digital technology can deliver (quality). The emergence of digital mediators could lead to an overall expansion in the volume of information being translated into multiple languages and improve the welfare of individuals and societies requiring translations, especially if end users will have the capacity to select the right mediator provider– encompassing digital and human mediators - given their specific needs (for example prioritising speed or quality). It could also lead to the emergence of hybrid forms of language mediators, whereby human mediators work alongside digital mediators although such complementarity might lead to changes in the skills required of human mediators. On one hand, they may require the ability to understand how to make the most of technology. On the other hand, they may require skills to be able to engage in tasks that digital mediators cannot perform or perform poorly.

8. This paper first reviews the literature on the labour market implications of technological advances, with a particular emphasis on recent advances in artificial intelligence. It then discusses societal advantages that arise from the availability of information in multiple languages, but also some of the hazards that can arise when the accuracy of translations is poor either because of limitations in existing technology or because of malicious intents (Wang et al., 2021[8]) , potentially giving rise to misinformation (Lee and Qian, 2022[9]; Saghayan, Ebrahimi and Bahrani, 2021[10]). The paper does so by describing the capabilities but also limitations of existing AI machine translation tools. Next, it uses data on online job vacancies to map the evolution of the demand for language professionals between 2015 and 2019 in English-Speaking countries covering Australia, Canada, New Zealand, Singapore, the United Kingdom, and the United States, as well as, for a selected set of analyses, from EU countries covering France, Germany, Italy and Sweden. Results indicate that between 2015 and 2019 the number of vacancies posted

on line for language professionals remained relatively stable, a possible indication that given the state of machine translation technologies no large-scale substitution effect occurred. By leveraging information collected in the job ads on the skills that employers report seeking when hiring language professionals, the paper illustrates the relevance of AI-related skills for the work of language professionals as well as the continued relevance of content-specific knowledge skills and transversal skills. Language professionals have a crucial societal role as facilitators of communication across language boundaries and their work entails many of the non-routine tasks which may be exposed to automation due to AI’s unique capabilities and, as such, constitute an important case study to consider the labour market and skills implications of emerging AI technologies. The paper concludes by discussing lessons learnt for knowledge workers from the development and refinement of AI technologies but also the broader societal implications for the users of knowledge produced by digital and human actors.

2 Technological innovations and labour market effects for language professionals

9. Many of todays' political and social tensions arising in response to the automation of tasks previously carried out by humans on labour markets closely mirror those from 1779, when Ned Ludd supposedly broke knitting frames at the start of the industrial revolution in England (United Kingdom) lending the name to the Luddites movement in the process. Much of the debate back then and today revolves around the question of whether technologies substitute or complement workers, give rise to better or worse labour market conditions, and ultimately are associated with an increase or a decrease in labour market opportunities. Key to answering these questions is what tasks technologies are able to accomplish, what humans are able to do, and what skills humans need to be able to make the most of technological change (Violante, 2008[11]).

10. Empirical estimates over the impact of digital technologies on employment are mixed: whereas some empirical studies reveal that technological developments have led to a growth in employment opportunities (Dixon, Hong and Wu, 2021[12]; Koch, Manuylov and Smolka, 2021[13]) others suggest that technological developments led to lower employment possibilities for workers (Acemoglu and Restrepo, 2020[14]). Overall, to date, empirical evidence suggests that past waves of technological developments did not lead to overall lower employment opportunities and net job destruction in the long run (OECD, 2019[15]). In fact, over the 20th century the employment-to-population ratio rose and the unemployment rate did not change over the long-run (Autor, 2015[16]). The study by Georgieff and Milanez (2021[17]) found no evidence that recent technologies led to overall job destruction in the 21 countries under study at the broad country level, but that the risk of automation was a significant predictor of occupation-level employment growth: growth was lower in those jobs that were at the highest risk of automation. These results suggest that automation can lead to worse employment prospects for some. This is also in line with findings that job losses resulting from computer automation tend to be more pronounced for low-wage occupations (Bessen, 2016[18]), for occupations in the manufacturing sector (Mann and Püttmann, 2021[19]), and generally among workers conducting routine work (Gaggl and Wright, 2017[20]).

11. As a result of past waves of technological progress, today's workplaces demand people who can solve non-routine problems. Few workers, whether in manual or knowledge-based occupations, only use repetitive actions to perform their job tasks. In fact, evidence from the OECD Survey of Adult Skills, a product of the Programme for the International Assessment of Adult Competencies (PIAAC) indicated that in 2012 the majority of workers were confronted at least once a week in their job with simple problems requiring less than 30 minutes to find a solution and as many as one in ten workers were confronted every day with complex problems requiring at least 30 minutes to find a good solution (OECD, 2013[21]). As technologies that could perform rule-based tasks were introduced, the importance of people's ability to solve complex problems that could not be solved simply by applying pre-specified rules grew. Whereas computers gradually took over 'the expected', individuals increasingly had to deal with the 'unexpected

and the unfamiliar' often working alongside computers (Autor, Levy and Murnane, 2003[22]; Ikenaga and Kambayashi, 2016[23]; Spitz-Oener, 2006[24]).

12. Whereas in the past technological developments led to the creation of computers and robots that could only follow narrowly specified rules, machine learning algorithms allow automata to perform a considerably broader set of tasks that have no rule-based solutions. As a result, the set of tasks that can be performed by technologies is radically different. Whereas past trends from Germany, Japan and the United States indicated a decline in the demand for routine cognitive and manual skills and an increase in the demand for non-routine interactive and analytic skills (Autor, Levy and Murnane, 2003[22]; Ikenaga and Kambayashi, 2016[23]; Spitz-Oener, 2006[24]) it is possible that the advent of AI systems will radically change the demand for skills in the future as non-routine tasks are within the scope of what automata can perform reliably (Georgieff and Hyee, 2021[25]). On one hand, technology may obviate the need for humans to perform certain tasks. On the other hand, technologies may complement humans, requiring workers to learn to work effectively with new technologies (Arntz, Gregory and Zierahn, 2016[26]; Georgieff and Hyee, 2021[25]) as some tasks, but not all, will be affected by automation (Bessen, 2016[18]).

13. Estimates on the share of jobs rated at high risk of being automated depend on whether it is assumed that whole occupations can be automated or that task changes occur within occupations (Spitz-Oener, 2006[24]). With this difference in mind, the share of occupations at high risk of being automated [1] within the next two decades ranges from as low as 35% in Finland and 33% in Norway (Pajarinen, Rouvinen and Ekeland, 2015[27]), to 47% in the United States (Frey and Osborne, 2017[28]), while in Germany 59% of occupations are at risk of being automated (Brzeski and Burk, 2015[29]). Alternatively, a more fine-grained approach, that takes into account the automation of single tasks and skills provides lower estimates with 9%, 14% and 28% for jobs across OECD countries (Arntz, Gregory and Zierahn, 2016[26]; Lassébie and Quintini, 2022[30]; Nedelkoska and Quintini, 2018[31]).

14. Changes in the way workers carry out their jobs as a result of technological adoption might imply changes in the skills individuals need to master since the tasks they are engaged with will differ and, consequently, require workers to re-skill or up-skill (Lane and Saint-Martin, 2021[32]; Nedelkoska and Quintini, 2018[31]). Understanding the finer details underlying the transformative process of automation is crucial to identify changes in skills demands resulting from the adoption of technologies that automate certain tasks. According to the United States' O*NET (Occupational Information Network, n.d.[33]) database language professionals (officially classified as Interpreters and Translators code 27-3091.00) perform among others the following tasks that are key to maintaining translation quality and that so far cannot be substituted by AI-machine translation tools: Translate messages simultaneously or consecutively into specified languages, ***maintaining message content, context, and style as much as possible,*** Check translations of technical terms and terminology to ensure that they are ***accurate and remain consistent throughout translation revisions***; and Identify and resolve conflicts related to the ***meanings of words, concepts, practices, or behaviours***.[2]

[1] Frey, Carl Benedikt; Osborne, Michael A. (2017[28]) define high-risk occupations as occupations which have a probability of computerisation above 70%. This approach is adopted by Pajarinen, Mika; Rouvinen, Petri; Ekeland, Anders (2015[27]) for their estimations on Norway and Finland.

[2] According to the United States' O*NET (Occupational Information Network, n.d.[33]) database language professionals (officially classified as Interpreters and Translators code 27-3091.00) perform the following tasks: Follow ethical codes that protect the confidentiality of information; Translate messages simultaneously or consecutively into specified languages, orally or by using hand signs, maintaining message content, context, and style as much as possible; Listen to speakers' statements to determine meanings and to prepare translations, using electronic listening systems as necessary; Compile terminology and information to be used in translations, including technical terms such as those for legal or medical material; Check translations of technical terms and terminology to ensure that they are accurate and remain consistent throughout translation revisions; Identify and resolve conflicts related to the meanings of words,

15. Although few individuals worldwide work as language professionals (52 170 people were employed as Interpreters and Translators in the United States in 2021, corresponding to 0.037% of total employment, see Box 2.1), they have a crucial societal role as facilitators of communication across language boundaries and facilitate trade across economic agents working in different countries and language communities. Furthermore, the work of translators entails many of the non-routine tasks that previous waves of technological development did not expose to the threat of automation but which may be exposed due to AI's unique capabilities. It is therefore a key case study to consider the labour market and skills implications of emerging AI-machine translation technologies.

Box 2.1. Recent employment trends of language professionals in the United States

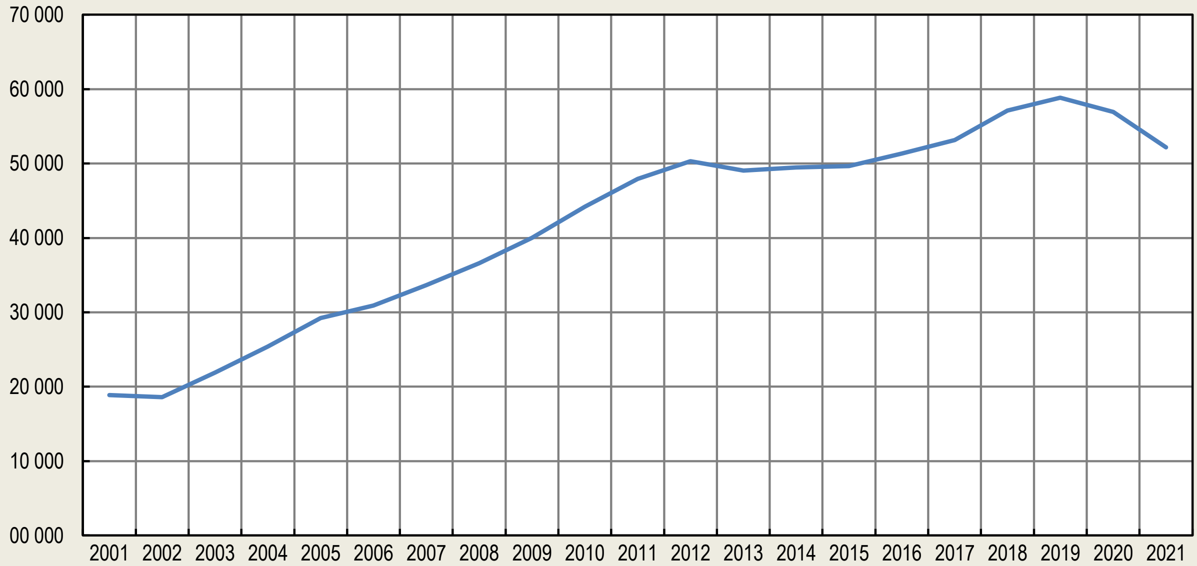
Over the past years, the volume of legal, political, informational content and documents translated by language professionals has increased. For example, the European Union requires legislation to be translated into 24 different language versions to conform with Council Regulation No 1 of 1985 (Rossi and Chevrot, 2019[34]).

Figure 2.1 shows the evolution of total employment for language professionals - interpreters and translators - for the United States, excluding self-employed individuals, suggesting an increase in the demand for language professionals. Over the past 20 years, employment of interpreters and translators has almost tripled. While employment in 2001 amounted to around 20 000, in 2021 more than 50 000 people were employed as interpreters or translators. Furthermore, the employment of interpreters and translators is expected to grow by around 20% between 2021-2031, an increase that is 15 percentage points above the average growth rate for all occupations (Bureau of Labor Statistics, U.S. Department of Labor, 2022[35]).

concepts, practices, or behaviours; Refer to reference materials, such as dictionaries, lexicons, encyclopaedias, and computerised terminology banks, as needed to ensure translation accuracy; Compile information on content and context of information to be translated and on intended audience; Adapt translations to students' cognitive and grade levels, collaborating with educational team members as necessary; Check original texts or confer with authors to ensure that translations retain the content, meaning, and feeling of the original material; Adapt software and accompanying technical documents to another language and culture; Educate students, parents, staff, and teachers about the roles and functions of educational interpreters; Proofread, edit, and revise translated materials; Train and supervise other translators or interpreters; Read written materials, such as legal documents, scientific works, or news reports, and rewrite material into specified languages; Travel with or guide tourists who speak another language; Discuss translation requirements with clients and determine any fees to be charged for services provided.

Figure 2.1. Employment trend of Interpreters and Translators, United States (2003-2021)

Total number of online job postings for language professionals, by year



Note: The figure shows the evolution of the estimated total employment rounded to the nearest 10 (excludes self-employed) for the occupation Interpreters and Translators.
Source: Bureau of Labor Statistics, U.S. Department of Labor (2022[36]), Occupational Outlook Handbook, www.bls.gov/ooh/.

3 Language, machine translation technology and language professionals

16. Cognitive scientists indicate that individuals' ability to share and communicate information through language is determined by their cognitive abilities (Johansson, 2021[37]). At the same time language is also what allowed humans to develop abstract thoughts, develop complex ideas about themselves and the world around them, and build these ideas across time and space. Moreover, the specific language individuals use to share and communicate information can determine their thoughts, preferences and behaviours including their sense of time and space (Boroditsky, 2001[38]; Evans, 2009[39]; Fuhrman and Boroditsky, 2010[40]; Miles et al., 2011[41]) their engagement in economically relevant behaviours such as saving decisions (Chen, 2013[42]; Sutter et al., 2018[43]), their educational and career choices (Rhodes et al., 2019[44]) and their adoption of gender stereotypes (Cimpian and Markman, 2011[45]; Lewis and Lupyan, 2020[46]). Languages, in other words, are not neutral to cognition but have the potential to shape how individuals perceive the world around them.

17. The use of language has complex implications for identity, social integration, education, and development. Therefore, whether languages are preserved or not has important geopolitical implications in multicultural and globally connected societies. Estimates suggest that every two weeks a language becomes extinct, often time leaving little written record of the cultural and intellectual heritage that was accumulated with its use and that around 40% of the languages that remain in use are endangered with only a few hundred languages being used in public education systems (UNESCO, 2010[47]). International efforts such as the celebration of the International Year of Indigenous Languages in 2019, the International Decade of Indigenous Languages (2022-2032) and the UNESCO Recommendation concerning the Promotion and Use of Multilingualism and Access to Cyberspace (see Box 3.1) represent important attempts to preserve, revitalise and promote languages that are at risk of going extinct.

Box 3.1. UNESCO Recommendation concerning the Promotion and Use of Multilingualism and Access to Cyberspace

The 2003 UNESCO Recommendation concerning the Promotion and Use of Multilingualism and Access to Cyberspace encourages international organisations, governments, civil society, academia and private sector organisations, including the IT industry, to collaborate in the development of multilingual content and systems, to facilitate access to networks and systems, to develop public domain content, and to seek equitable balance between the interests of rights-holders and the public interest (UNESCO, 2003[48]). Four rounds of reporting on the implementation of the Recommendation occurred so far, with the fifth round being planned for 2023. Only few countries reported on implementation, with decreasing reporting over time. In 2007, 32 member states reported, which decreased to 24 in 2011, 21 in 2015 and 17 in the last round conduced in 2019 (UNESCO, 2019[49]).

Concerning the development of multilingual content and systems, the 2003 Recommendation calls for the public and private sectors as well as the civil society at local, national, regional and international levels to work towards alleviating language barriers and promoting human interaction on the Internet by encouraging the creation and processing of, and access to, educational, cultural and scientific content in digital form, so as to ensure that all cultures can express themselves and have access to cyberspace in all languages, including indigenous ones.

The 2003 Recommendation urges member states and international organisations to encourage and support capacity-building for the production of local and indigenous content on the Internet. It encourages member states to formulate appropriate national policies on the crucial issue of language survival in cyberspace, designed to promote the teaching of languages, including mother tongues, in cyberspace. It highlights that international support and assistance to developing countries should be strengthened and extended to facilitate the development of freely accessible materials on language education in electronic form and to the enhancement of human capital skills in this area. It urges member states, international organisations and information and communication technology industries to promote collaborative participatory research and development on, and local adaptation of, operating systems, search engines and web browsers with extensive multilingual capabilities, online dictionaries and terminologies; to support international cooperative efforts with regard to automated translation services accessible to all, as well as intelligent linguistic systems such as those performing multilingual information retrieval, summarising/abstracting and speech understanding, while fully respecting the right of translation of authors. Finally, it identifies UNESCO, in co-operation with other international organisations, as the organisation that should establish a collaborative online observatory on existing policies, regulations, technical recommendations, and best practices relating to multilingualism and multilingual resources and applications, including innovations in language computerisation.

18. Given the wealth of evidence on the role languages have in shaping how individuals think about important economic and social phenomena, the ability of humans to tackle crucial questions of scientific and social significance depends on ensuring diversity in the languages individuals use to ponder about and solve problems, and in making the evidence arising of such efforts widely available across language communities. For example, it has been recently suggested that the prevalence of English as the language in which most scientific communication occurs may shape the very nature of scientific progress (Blasi et al., 2022[50]). Furthermore, the fact that most scientific publications are in English may lead to evidence produced in non-English languages being excluded from scientific repositories and search engines. Even when scientists whose mother tongue language is not English submit their findings for peer review in English-language outlets, reviewers may judge the quality of their research as worse because of lack of language proficiency making their insights less likely to be published and disseminated among the scientific community (Politzer-Ahles, Girolamo and Ghali, 2020[51]). These three sources of linguistic bias may ultimately lead to the exclusion of important insights useful for scientific research and understanding. In particular, indigenous communities are often especially hard hit by the consequences of human induced climate change and loss of biodiversity (Hunter, North and Slotow, 2021[52]). However, because of language barriers, they often fail to make the most of scientific evidence. Moreover, their knowledge remains largely excluded from research aimed at counteracting climate change and loss of biodiversity because such knowledge is not in English. Translation efforts are key to promote a knowledge base built on a wider set of languages. Box 3.2 describes the TranslatE project, which aims to reduce language barriers in the collection and use of scientific information in environmental science.

Box 3.2. The TranslatE project

Scientific evidence is crucial to address key environmental challenges, such as the loss of biodiversity. Although it is often assumed that the number of non-English language publications is decreasing as English becomes the language of science (Lynch et al., 2021[53]), in many fields the number of non-English-language publications is increasing at a similar rate to English-language publications (Chowdhury et al., 2022[54]). Linguistic diversity is essential for comprehensive evidence-based decision making (Nuñez et al., 2021[55]). The *Transcending language barriers to environmental sciences* (TranslatE) project aims to reduce language barriers in the collection and use of scientific information in environmental science, with the aim of ensuring that global biodiversity conservation efforts can rely on the full wealth of knowledge available rather than be determined by the language in which evidence is produced (Transcending language barriers to environmental sciences, n.d.[56]). It does so by considering the role of language in the global compilation and in the local application of scientific information in the domain of biodiversity conservation. The basic premise of TranslatE is that scientific studies that are not conducted and reported in English are considerably less likely to become widely known and used by the scientific community. At the same time, language barriers continue to prevent the use of valuable scientific evidence by local-level actors.

The TranslatE project considers the role of language in the global compilation of scientific information on biodiversity conservation by: 1) investigating the use of non-English-language scientific knowledge in environmental evidence syntheses and what consequences the inclusion or exclusion of non-English language scientific knowledge has and 2) assessing if there are systematic differences in the scientific knowledge that is published in different languages; and exploring how automated search systems could help identify important non-English-language scientific literature.

It considers the role of language in the local application of scientific information on biodiversity conservation by: investigating the use of scientific knowledge available in different languages in local decision-making and understanding its impact for local conservation efforts, and testing the effectiveness of machine/human translations for the uptake of scientific information in different communities.

Examples of work conducted as part of the TranslatE project include complementing evidence on amphibian heat tolerance produced in English with evidence in Portuguese, Spanish, simplified Chinese, or traditional Chinese (Pottier et al., 2022[57]); considering the role of language barriers in global bird conservation (Negret et al., 2022[58]); mapping (Chowdhury et al., 2022[54]) (Amano et al., 2021[59]; Chowdhury et al., 2022[54]) and proposing (Amano et al., 2021[60]; Khelifa, Amano and Nuñez, 2022[61]) effective solutions to redress lack of language diversity in environmental science. Studies estimate that up to 36% of conservation-related scientific documents may be published in non-English languages and that a majority of local-level decision makers, such as, for example, the directors of protected areas in Spain, reported language as a key barrier to the use of published scientific evidence in their conservation efforts (Amano, González-Varo and Sutherland, 2016[62]).

19. Ensuring full interdependence and connectivity across all possible language pairs in order to make knowledge universal is an unrealistic goal. However, it is possible to promote increased connectivity across language communities through a mixture of investments in language learning, the use of human and digital mediators, as well as in-depth assessments of the economics and the politics of knowledge production in different languages. Efforts to develop effective digital mediators of language are not new, but both the quality and the range of applications of the outputs of machine translation tools have significantly increased since the technology has moved from rule-based translation to statistical models to neural networks (Lopez, 2008[63]). Box 3.3 provides an overview of past and current developments of machine translation models.

Box 3.3. An overview of machine translation models

Machine translation is the process of automatically translating content from one language (the source) to another (the target). Translation was one of the first applications of computing power, starting in the 1930s. Rule-based machine translation (RBMT) was the earliest form of machine translation and was the dominant framework until the late 1980s. It was then followed by statistical machine translation (SMT) (Hutchins, 2006[64]). At the time of writing this working paper neural machine translation represents the most promising development in the field (Forcada, 2017[65]; Koponen, Salmi and Nikulin, 2019[66]; Stahlberg, 2020[67]) and is being used by software such as DeepL © (DeepL, n.d.[68]) and Google Translate © (Le and Schuster, n.d.[69]; Wu, 2016[70]). Supervised learning algorithms learn functions based on labelled training data. Labelled training data, with examples of input-output data pairs, are used to measure accuracy but also to enable the algorithm to learn iteratively. Unsupervised learning uses machine learning algorithms to analyse and cluster unlabelled data in order to discover hidden patterns (IBM, 2021[71]).

The following characterises the different main approaches to machine translation:

- *Rule-based Machine Translation* (RBMT) is based on dictionaries mapping words from the source to target language, as well as grammatical rules for source and target language developed by language experts and programmers. RBMT requires prominent human intervention to input linguistic rules and provide feedback on translation accuracy (Machine Translate, 2022[72]; Omniscien Technologies, 2022[73]).
- *Statistical Machine Translation* (SMT) generates translations based on statistical models. Large data sets of existing translations (aligned parallel corpora) are used as input. First, statistical probability calculations are performed to determine probable correspondences in the target language for elements of the source language. The most probable translations for the context are then selected using the language model of the target language. Finally, the most probable word order in the target language is generated. Whereas rule-based machine translation tools process written text word by word, statistical machine translation tools process text phrase by phrase. This is computationally intensive and requires a large amount of data on parallel corpora to achieve correct weight precision and deliver quality translation (Lopez, 2008[63]).
- *Neural Machine Translation (NMT)* uses artificial neural networks. In neural machine translation an artificial neural network (encoder) processes the source sentence word by word and transposes it into a mathematical representation in which each word is then represented in the context of the sentence as a multidimensional vector. From this abstract representation another artificial neural network (decoder) generates word by word the target-language sentence. Machine learning is used to train the system, using many corpora. Adaptive neural machine translation models can also be developed to incorporate information provided by professional translators during post-editing. The model therefore improves over time and can be tailored to the unique needs of specific technical fields or specific users, to ensure that terminology is aligned by conventions and requirements (Artetxe et al., n.d.[74]; Omniscien Technologies, 2022[75]).

Even though large improvements in the quality of machine translation have been observed, the number of languages for which machine translation systems have been built is around 100, and therefore only a small fraction of the more than 7 000 (un)official languages spoken around the world (Bapna et al., 2022[76]; Caswell and Bapna, 2022[77]). This results in machine translation systems being highly skewed towards European languages due to two main bottlenecks which are lack of digitised languages and lack of translated text.

20. A key problem of existing machine translation tools is that not all languages are represented in existing efforts and even when languages are represented, the quality of the translation varies markedly, for example because of differences in the amount of resources that are available to train AI systems (Fan et al., 2022[78]; Haddow et al., 2022[79]). Diversity in the quality of the translations produced by machine translation tools for high-resource and low-resource languages is likely to become more pronounced since neural networks models produce translations of a higher quality than rule-based and statistical models but only when they can rely on large training datasets, and compiling large and diverse datasets for low-resource languages is expensive, logistically challenging or both (Kuwanto et al., 2021[80]; Nekoto et al., 2020[81]; Orife et al., 2020[82]). At the moment, high-resource languages, i.e. languages for which large quantities of training data are available in digital form and for which machine translation tools work best, are the same languages dominating scientific and social exchanges. Unless this changes, for example because more digital examples of text become available, improvements in AI technologies could further exacerbate the concentration of prominence to a restricted set of languages rather than reduce existing disparities. Box 3.4 highlights challenges related to the development of machine translation models for low-resource languages as well as existing attempts to overcome such challenges.

Box 3.4. Solving the Low-Resource Language Problem: The Flores 101 and the No Language Left Behind Project-200 projects (MetaAI ©)

Multilingual systems, i.e. models that leverage information from multiple related languages (Arivazhagan et al., 2019[83]) (Fan et al., 2021[84]) (Zhang et al., 2020[85]) (Zoph et al., 2016[86]), have been used to tackle the low-resource language problem. However, translation quality remains a challenge (Tran et al., 2021[87]). Alternatively there have been attempts to develop effective machine translation tools for low-resource languages by collecting more data, including translations developed by human translators and monolingual data sources available on the web (Karakanta, Dehdari and van Genabith, 2017[88]) (Kreutzer et al., 2022[89]).

In July 2021 MetaAI © released the Flores-101 dataset, an open source many to many dataset covering 101 languages (Goyal et al., 2022[90]). This was followed in July 2022 by the No Languages Left Behind (NLLB-200) release, which covered 200 languages. Both Flores 101 and NLLB-200 are open-source projects developed by MetaAI © to facilitate direct translation of many low-resource languages. The goal of Flores 101 and NLLB-200 is to increase the opportunity for individuals to access and share content on the internet in a wider range of languages. In addition, the projects were designed to support translation of many languages that were previously not served by AI-machine translation and therefore could be used to broaden access and exposure to information. For example, high quality automatic translations could allow students and educators belonging to low-resource language communities to access digital resources compiled in different languages (Lee, 2019[91]). Similarly, improvements in translation quality could expand the accessibility of low-resource language knowledge to members of other language groups. Because language is intrinsically tied to culture, for many low-resource languages facing endangerment, the threat of losing one's language could also mean the erosion of one's heritage (Sallabank, 2013[92]). Finally, translation tools could challenge Western-centric modes of knowledge production and dissemination, and support knowledge production and exchange by indigenous communities (Bird, 2019[93]).

Attempts to extend machine translation to low-resource languages should weigh benefits alongside costs and risks. For example, the increase in digital participation and linguistic representation may compound and heighten discrimination and the surveillance of linguistic minorities (Treré, 2016[94]), and expose communities to misinformation, disinformation and hate content (Gereme et al., 2021[95]) (Hossain et al., 2020[96]). Finally, it is increasingly recognised that AI-systems, particularly systems that rely on extremely large data inputs, can have major negative environmental impacts. Estimates by

MetaAI © suggest that the overall impact of the NLLB-200 project, including data mining, backtranslation, modelling, developing final ablations, evaluation and training the NLLB-200 model was 104.31 tCO_2eq (Costa-jussà and al., 2022[97]). Ultimately, machine translation development and use have major technological, cultural, and societal impacts, and therefore requires an interdisciplinary and participatory approach and the involvement of multiple-stakeholders including the involvement of affected communities (Kusters et al., 2020[98]).

21. Rapid developments and improvements of machine translation technologies in recent years, have broadened their use thanks to the fact that many AI-powered translation tools are freely available on the Internet and allow to translate large amounts of text instantaneously and at no cost. In situations when accuracy is not critical, machine translation tools can help users understand the broad meaning of a text, a process which is called *gisting* (Forcada et al., 2018[99]). However, despite rapid market growth and technology improvements, the accuracy of machine translation technology continues to lag behind the quality of human translation. On top of the limitations detailed in Box 3.4 for low-resource languages, limitations that pertain to all language models include limitations related to *linguistics* and *human-coded biases*.

22. Linguistic limitations include obstacles related to untranslatable words and contextual meanings in cross-language translations, cultural context and cultural expectations, and ever-evolving nature of languages. In every language, words can be found that capture a very fine-grained meaning and for which no translation to another language exists. The *Dictionary of Untranslatables* by Cassin et al., (2014[100]) gathers linguistic subtleties by collecting and describing the complexity behind concepts embedded in single words with no equivalent translation. Additionally, machine translation technologies have difficulties evaluating or recognising metaphorical meanings, interpreting and translating hidden or subtle messages, or identifying contextual meanings that are not literal, such as humour, irony, or sarcasm (Ducar and Schocket, 2018[101]; Wallace and Kertz, 2014[102]).

23. Machine translation tools are also limited in their capability of adjusting translations to the cultural context or to meeting cultural expectations (Ducar and Schocket, 2018[101]). Languages and cultures differ in the phrases or gestures of conveying different messages. Such language subtleties for example include the way gratitude is expressed, what is the language appropriate way to say “thank you”, or the translation of the question “how are you?”. Therefore, nuanced understanding of the levels of formality, interpersonal relationships, and cultural expectations are required to communicate properly in that other language. Changes in slang, neologisms, and expressions over time impact the evolution of languages. The quality of machine translation technology requires a large amount of input data to be trained on and therefore involves an inherent gap between human and machine translation.

24. The development of machine translation technology is based on language models which require large amounts of data which are created by humans. Biases in data stemming from existing power structures are among the major issues of concern in artificial intelligence and machine learning models (Anaconda, 2020[103]). These biases in AI systems consequently result in incorrect output, and discriminatory output and predictions for certain populations (Smith and Ishita, 2020[104]). Although language professionals can also introduce biases in their translation outputs, those who commission translations to language professionals can require them to pay attention to cultural sensitivities. Moreover, the societal impact of the biases introduced by one language professional are quantitatively small since he or she can only develop few translations. By contrast, any biases introduced by AI systems have quantitatively large implications because the same bias is reflected in a very large number of translations.

25. Various sources of biases can potentially impact AI systems. One of the most prevalent biases is related to gender (Savoldi et al., 2021[105]). Machine translation technology performs less accurately when words or texts that are translated from a text that is rather gender-neutral are translated to a language that is not. For example, this could lead to machine translation technology only providing a single-gendered

translation or using masculine translation by default in male-dominated fields. AI systems are created by humans. Hence, they may integrate and mirror the perspective and knowledge of society of those who develop AI systems (Smith and Ishita, 2020[104]). Today, certain groups are underrepresented when it comes to AI development (e.g. no balanced share of women, Black or Hispanic employees in AI development) (Smith and Ishita, 2020[104]).

26. A final set of limitations pertains to ethical and legal implications. Issues related to ownership and privacy of the content inputted in machine translation systems data as well as the legal responsibility for the consequences associated with potential mistakes in documents being translated remain key challenges for the use of machine translations in high-stakes settings. Because of the limitations detailed, machine translation tools so far are not able to replicate the quality of human translations.

27. Due to these limitations and in order to achieve publishable translations of high-quality, even when machine translation models are used, human translators are often called in to work on the outputs of machine translators (Macken, Prou and Tezcan, 2020[106]). Despite these limitations, machine translation tools are used among language professionals, and even sometimes required by clients, as a computer aid in translation settings and it is claimed that familiarity with translation technology is a prerequisite for a successful career pathway (Cadwell, O'Brien and Teixeira, 2017[107]; Gaspari, Almaghout and Doherty, 2015[108]).

28. From an employer perspective, the deployment of machine translation technology is associated with beneficial effects. For instance, machine translation technology is productivity enhancing as higher volumes of content can be translated within a given time. Several studies provide evidence on speed gains following the use of machine translation software (Macken, Prou and Tezcan, 2020[106]; Plitt and Masselot, 2010[109]). However, such beneficial effects are not necessarily shared from a translator's perspective. Rather, the profession of translators is seen as being under pressure. Among language professionals reasons for the non-adoption of machine translation technology are related to the fear of being replaced by a machine, inducement to make particular errors, and the fear of losing language proficiency (Cadwell, O'Brien and Teixeira, 2017[107]; ELIS Research, 2022[110]; O'Brien, 2002[111]; Pym and Torres-Simón, 2021[112]). In addition to the threat of the profession through artificial intelligence, evidence is found that prices for translation developed by language professionals are decreasing (Vieira, 2018[113]; Moorkens, 2017[114]). On one hand, this reflects the increase in productivity that may arise from language professionals working alongside machine translation tools on certain output. On the other, it reflects a depreciation of the output of language professionals as alternatives (i.e. machine translations) become available to individuals and organisations commissioning translations. At the same time, if the translation market were to become highly specialised, with language professionals tackling only the most challenging projects that machine translators cannot tackle with sufficient quality, the price per word paid to language professionals to engage in such projects should become higher. Despite existing limitations of machine translation tools, the appeal of developing fast, timely and cheap translation may lead to their application in settings in which getting mis-translations can have very severe consequences. Box 3.5 details the importance of translations in the healthcare sector and what opportunities and challenges arise from the emergence of machine translation tools for patient care.

29. Upcoming innovations will likely further change the capabilities of machine translation. While neural machine translation has so far required the development of post-editing practices in the translation profession, the arrival of GPT4 on the market this year might mark a new step in the story of MT where AIs might acquire the ability to learn on their own. This would imply that machines might be able to start correcting themselves and not necessarily need the feedback of humans anymore, further reducing the scope of translators' roles. In addition, recent research suggests that new computer translation systems such as CUBBITT can already, in some circumstances, outperform human translation (Popel et al., 2020[115]).

Box 3.5. Linguistic barriers in healthcare: Opportunities and challenges for the use of machine-translation tools

Globalisation is increasing not only the number of people who live and work in a country in which the official language(s) or the most widely used language(s) differ from their mother tongue language(s), but it is also increasing the diversity of languages that are used. Linguistic superdiversity – conceived as the increase in the number of different languages that individuals residing in a country use to communicate as well as the share of such individuals in the population – poses unique challenges for communication (Vertovec, 2007[116]). Linguistic barriers can severely limit the well-being of individuals who have no or only limited ability to use the official language(s) used in a country. However, even when members of different language communities are proficient users of the official language(s) used in a country, they may face unique disadvantage in high stakes situations such as those occurring in health care sector and the criminal justice system (Vieira, O'Hagan and O'Sullivan, 2020[117]).

Interactions in these settings involve the use of technical terms and are marked by heightened stress and anxiety levels. Linguistic barriers may therefore be not only very consequential when people communicate, ether orally or in written form in these settings but may also be more prevalent, since even individuals who are proficient in the official language(s) used in a country may struggle with technical legal or medical terms. Moreover, many aspects of cognitive capacity tend to be lower when individuals experience acute and uncontrollable stressful situations (Sandi, 2013[118]). Therefore, the language skills of individuals whose primary language(s) is different from the official language(s) used in a country may struggle in stressful situations in medical and legal settings. The attempt for individuals to understand material that is available in a not mother-tongue language could also adversely influence individuals' outcomes by inducing ego depletion – defined as the transitory decline in self-regulatory capacity following an act of self-control (Baumeister, Muraven and Tice, 2000[119]; Baumeister, Vohs and Tice, 2007[120]).

Effective oral and written communication between patients and healthcare professionals is crucial for the provision of quality healthcare, especially as chronic conditions pose an increasingly large burden on individuals and societies. The management of chronic conditions or long-term illnesses often requires that individuals actively participate in their treatment regiments. Language barriers can lead to the emergence of disparities in the quality of care received by populations with limited proficiency in the language used in healthcare settings (Jacobs et al., 2006[121]) (Ohtani et al., 2015[122]) (Schwei et al., 2016[123]) (de Moissac and Bowen, 2017[124]; Jacobs et al., 2006[121]; Ohtani et al., 2015[122]; Schwei et al., 2016[123]) and interact with other sources of vulnerability, disadvantage and discrimination, such as having a disadvantaged socio-economic condition. The literature indicates that language barriers reduce access to health promotion/education resources (Brar et al., 2009[125]), preventive care (Cohen and Christakis, 2006[126]), cancer screening (McDonald and Kennedy, 2007[127]), mental health services (Schwei et al., 2016[123]), and referral to specialised services (Fryer et al., 2011[128]).

The legal foundations for guaranteeing access to language services in healthcare are grounded in laws on informed consent and civil rights laws (Teitelbaum, Cartwright-Smith and Rosenbaum, 2012[129]). Informed consent laws mandate that in healthcare situations individuals should have the ability to consent to the care they receive. The ability of patients to affirmatively give informed consent to treatment is a fundamental element of healthcare quality. Linguistic barriers prevent individuals' ability to consent to their care. In the United States Title VI of the 1964 Civil Rights Act provides the foundation for the right to language access upon which guidance documents are developed by the Department of Health and Human Services' Office for Civil Rights to prevent discrimination on the basis of linguistic proficiency (Chen, Youdelman and Brooks, 2007[130]). Despite the legal requirement to provide language services for individuals with limited English proficiency, healthcare providers may remain unaware of their legal responsibilities and/or may not prioritise the provision of language services at times of fiscal constraints with negative consequences for patient outcomes.

Guaranteeing the right to access healthcare for individuals with limited language proficiency can be achieved by investing in the cultural competence, skills and awareness of healthcare professionals, to ensure that they appreciate the consequences of linguistic barriers for health outcomes and patient care. It can also encompass establishing certification protocols to ensure quality standards of language professionals working in the healthcare sector, since language competency should be accompanied by specific technical knowledge and understanding of the healthcare sector to be most beneficial (Karliner et al., 2007[131]). Financial considerations remain a key barrier to the provision of high-quality language services. Ensuring that ad hoc funding is provided to guarantee these services can ensure take up by healthcare professionals and institutions. Finally, increasing the awareness of patients about their rights, and what opportunities are available to them to access trained medical language translators and interpreters is key (Chen, Youdelman and Brooks, 2007[130]).

In the absence of clearly defined guidelines and funding, machine translation software are increasingly being used to facilitate communication in healthcare settings. Studies of discharge instructions between English and commonly used languages in the United States such as Spanish, Chinese, Vietnamese, Tagalog, Korean, Armenian, and Farsi indicate that the quality of translated instructions varied markedly: whereas translated instructions from English into Spanish were generally high in terms of accuracy, accuracy was lower and more variable in the case of Farsi and Armenian (Taira et al., 2021[132]). In other words, machine translation software appear to perform poorly precisely in the instances in which they would be most valuable, such as aiding with the translation of lower- frequency languages or for which a smaller number of high quality certified language professional may be available.

Furthermore, the need for language professionals in healthcare is not only limited to translators working on written text, but plays a key role in facilitating oral communication between patients and healthcare providers serving the role of cultural brokers (Dohan and Levintova, 2007[133]; Hull, 2016[134]; Messias, McDowell and Estrada, 2009[135]), bridging the cultural divide that often exists between patients with limited language proficiency and clinicians. Because lack of language proficiency often coexist with individuals having an immigrant background or belonging to cultural minorities in a country, medical interpreters often play a role not only in delivering technically accurate information through translations but also using culturally specific and appropriate phrasing. Even when machine translation software are able to translate material in technically accurate ways, they perform very poorly on cultural adaptation dimensions and are unable to respond to perform the role of cultural brokers. Although some software are able to translate information delivered orally, the quality of such translations is often poor because mistakes occurring in the translation phase are magnified by mistakes occurring during the speech recognition phase.

Finally, medical interpreters are generally bound to the same confidentiality requirements as healthcare professionals. Given the highly sensitive nature of healthcare information, the use of machine translation tools may be especially sensitive and pose unique challenges with respect of data protection and ethical use of patient data (National Council on Interpreting in Health Care, n.d.[136]). To harness the potential of machine translation to improve the health outcomes of individuals facing linguistic barriers relies on identifying in which situations machine translation tools can complement or supplement medical translators and interpreters. It also requires continuously expanding and updating the knowledge base as technology improves on the risk arising from machine translation tools in various clinical scenarios, increasing the diversity in the training and evaluation of machine translation algorithms, comparing the performance variation among different machine translation algorithms, and expanding the outcomes used in the evaluation of machine translation for health practice (Khoong and Rodriguez, 2022[137]).

3.1. Changes in the skills set of language professionals

30. Previous sections illustrate that so far, machine translation technologies have the potential to enhance societal welfare but only if they complement rather than substitute the work conducted by language professionals in promoting inter-language communication. Individuals and institutions interested in engaging in inter-language communication by developing translations should carefully consider both the advantages but also the limitations of machine translation technologies given their unique needs (context, language, stakes associated with the translation). At the same time language professionals are bound to be required to react to the emergence of powerful technologies yielding rapidly improving outputs.

31. Some language professionals might fear losing their job and being replaced by machine translation technology (Pym and Torres-Simón, 2021[112]). Others may fear that their profession will lose intellectual profundity as they work alongside machine translation tools. Many language professionals consider translation tasks to be more than just a technical exercise: they aspire to develop texts that satisfy linguistic norms of a target culture taking into account the assumed knowledge of readers (Department of Computer and Information Science, Linköping University, 2017[138]). The task of translation texts might lead to an intellectual depreciation if translating is replaced by post-editing or clean-up of output translated by a machine (The Economist, 2017[139]). At the same time, the profession might gain in intellectual profundity if human translators tackle the most challenging, culturally driven aspects of translation, as well as post-editing the output produced by machines.

32. Translators require a broad range of skills. Skills related to the usage of machine translation technology among professional translators may result in changes in the skills requirements of language professionals. Machine translation technology may require language professionals to possess an extended skill set to make sure that on one hand they can make the most of machine translations to complement their work, and that on the other hand they can tackle the unique aspects of translation projects that cannot be effectively addressed by machines. For example, the competence framework developed by the European Master's in Translation (EMT) network which serves as a reference standard for translator training acknowledges the increased usage of machine translation technology. The competence framework defines five main competences areas. These are: 1) language and culture, 2) translation, 3) technology, 4) personal and interpersonal, and 5) service provision. As such, machine translation literacy and awareness, understanding of machine translation systems and the integration of machine translation into the workflow, and post-editing are becoming an integral part of the competences required of professional translators (European Commission, 2022[140]).

33. *Post-editing skills* are among the extended set of linguistic skills which are demanded when machine translation tools are integrated into professional translators' workflow (Rico and Torrejón, 2012[141]; Pym, 2014[142]; European Commission, 2022[140]; Koponen, Salmi and Nikulin, 2019[66]). *Post-editing* means that once machine translation output is generated, an editing, amending and correction process is conducted by human translators to achieve high quality translations (O'Brien et al., 2014[143]; Garcia, 2011[144]). Because post-editing needs to be done by someone proficient in the target language, post-editing is mostly carried out by professional translators (Vieira, Alonso and Bywood, 2019[145]). With the increased need for post-editing machine translation output, attempts have been made to harmonise and standardise this process. In 2017, the International Organization of Standards (ISO) developed a standard for *post-editing* (ISO 18587:2017), which provides the requirements for the process of full human post-editing and post-editors' competences and qualifications (International Organization of Standards, 2017[146]). Learning to post-edit machine-translated content is also finding its way into the education of professional translators, either as part of the formal education programme or non-formal training and education (Guerberof Arenas and Moorkens, 2019[147]) (see Box 3.6). Post-editing training for example includes knowledge on various kinds of machine translation systems or machine translation error analysis (Guerberof Arenas and Moorkens, 2019[147]). This is important since machine translation accuracy has,

despite strong advances over the past years, not been able to reach a human level of language proficiency, due to limitations related to linguistics, and biases.

34. Another set of skills comprises instrumental competences related to technology skills (Rico and Torrejón, 2012[141]). This involves for example raising awareness in the profession on the range of available technology that can be used in the translation process (Alcina, Soler and Granell, 2007[148]). Instrumental competences include understanding machine translation output and their integration in the workflow, knowledge about machine translation systems and their capabilities (Rico and Torrejón, 2012[141]) (European Commission, 2022[140]).

35. Finally, a key problem of existing machine translation tools is that they do not engage in communication. Although adaptive systems can incorporate 'feedback' on the quality of their output in new translations, such adaptation is reactive and initiated by the users rather than the system itself. Machine translation systems do not have the capacity to express doubt and feel self-doubt, to perceive uncertainty about their own translation predictions and ask for help. Because existing machine translation tools cannot understand if, when, and what they do not understand, they do not engage in a process of co-creation with the aim of satisfying the needs and intentions of those who seek their service. Delivering meaningful translations requires mediators to assign meaning to language and adjust translations in response to both verbal and nonverbal cues. When human mediators are involved, the language mediation process can become a collaborative effort between the mediator and the individual requesting the mediation. By proposing and weeding out alternatives and understanding context – whether technical and subject specific circumstances or cultural factors – human mediators can improve quality and generate meaningful content adapted to individual situations. Digital mediators, so far, do not engage in extensive communication with clients. Transversal skills such as the capacity to work with others, to communicate, to have cultural and situational awareness are therefore crucial aspects of the quality of human language mediators.

36. On the broader labor market effects of MT, the emergence of AI might lead to the emergence of a new category of language professionals, not exactly translators but rather "post-editors" who would be paid less than traditional translators (Lee and Qian, 2022[9]). Post-editors (PE) would devote less time and effort on the MT pre-translated text than to traditional translation projects, given that PE tasks tend to have relatively lower pay but identical, if not tighter, deadlines. Furthermore, if the jobs of translators were to become more digitised, the profession may face increasing expectations for remote working arrangements.

Box 3.6 Adaptations of education and training opportunities

Changing skills requirements and skills needs require the adaptation of education and training opportunities. This includes modifying the content and structure of educational programmes. Even though machine translation technology has made large advances in recent years, discussions about incorporating post-editing capabilities existed already in the early 2000s. Teaching post-editing was seen not only as a means to increase productivity and a way to embrace machine translation, but it was also acknowledged that post-editing of machine translated texts differ from post-editing of human translated texts in terms of cognitive processes (O'Brien, 2002[149]; Koponen, 2015[150]). Suggestions for additional training components comprised of skills and knowledge related to post-editing, machine translation technology and programming skills which are embedded in formal curricula or provided by language service providers (Koponen, 2015[150]).

The curricula of courses designed to train language translators have already started to adapt to the emergence of AI tools, including more content on computational linguistics and how best translators can make use of AI-machine translation tools. For example, in 2009, the Universitat Autònoma de Barcelona (UAB) introduced the first machine translation and post-editing modules as part of their Master's Degree for translators (Guerberof Arenas and Moorkens, 2019[147]). The university introduced modules in post-editing and machine translation. The latter covers the basic principles of machine translation technology, the types of engines that exist on the market, machine translation output as well as implementation of machine translation technology in the workflow. The post-editing module comprises basic knowledge of post-editing guidelines and diverse types of post-editing levels. Outside formal tertiary education of language professionals, training and education are additionally provided by a variety of language service providers.

Finland, a country where the usage of machine translation and post-editing skills were not as common as in other countries, introduced a course at the University of Helsinki to incorporate recent advances in machine translation technology into the curriculum of the Translation Studies programme. The aim was to introduce students to machine translation and post-editing but also to foster a positive attitude towards technology with the ability to critically assess and evaluate tools and processes. The course covered topics such as theory, history, practical use of machine translation and post-editing, as well as quality levels and guidelines of post-editing, and the evaluation of machine translation quality. The course did not cover lectures or practices on programming in general or building an MT system. Practical exercises included post-editing of machine translated texts (Koponen, 2015[150]).

Finally, outside of the training occurring in tertiary education institutions, the training of language professionals is provided by a variety of language service providers. For example, new private certification programmes such as the one run by RWS in the United Kingdom (RWS, 2022[151]), offer training to help linguists become post-editing specialists. In the United States and Australia, the ATA (American Translators Association, n.d.[152]) and NAATI (National Accreditation Authority for Translators and Interpreters, n.d.[153]) certifying bodies offer certification exams but no precise reference to MT proficiency is yet included in the exams' guidelines.

Box 3.7. What do language professionals think about the impact of AI on their work?

Data presented in this work identify the number of vacancies that were posted online advertising positions for language professionals as well as the skills demanded by prospective employers interested in hiring language professionals in a selected number of OECD countries. As such, they do not reflect the lived experience of language professionals, the skills they consider important to perform their job effectively given the current state of AI machine translation tools, and how they use such tools in their work. A questionnaire was shared with nine language professionals to gauge their perspectives and experiences. Respondents worked as professional translators, post-editors, and as members of a linguistic quality control agency. Given the nature of the questionnaire, the small number respondents and the non-representative nature of the selection of participants, information in this section can help to explore in detail the views of professionals and to explore how their attitudes are shaped by their daily work.

When they were asked about what skills they considered essential to be able to perform their jobs or facilitate their work, most respondents indicated that in their work they have to use a mix of knowledge skills, transversal skills, general digital skills, and AI-related skills. In particular, according to translators interviewed for this work, language professionals should possess an *'acute proficiency in the target language'*, '*good knowledge of the specialised areas covered in the translation assignment'*, '*cultural awareness*', '*problem-solving skills as well as collaborative problem-solving skills*', '*social perceptiveness*', '*time management*', '*proficiency in the use of translation technology such as CAT tools*', '*knowledge of AI-driven tools*', '*neural networks*' and '*computational linguistics*'. Depending on the specific role, professionals interviewed also indicated a wide range of ancillary skills necessary in specific functions. These ranged from 'people management and team building' to institutional knowledge', from 'psychometrics' to knowledge of speech recognition/speech-to-text and text-to-speech software'. Most indicated that digital skills played a crucial role in enabling them to perform well their work, with Computer Aided Translation (CAT) tools playing a particularly crucial role as digital aids to their work.

Most respondents agreed that AI-machine translation tools have changed the nature of their work or the work of language professionals more generally. At the same time, others indicated that AI-machine translation tools have not changed the nature of the job of language professionals but, rather, the processes involved in the delivery of translations and should be seen as a '*useful addition to the toolkit at the disposal of language professionals'.*

Some respondents indicated that because of the emergence of AI-machine translation tools they spent more time and effort dealing with '*technical issues at the expense of core tasks, i.e. understanding the source text and doing the required research to guarantee accuracy in the translation given the unique needs and context of the client*'. They felt that the AI-machine translation tools can hamper the creative and the learning process that translators acquire over time, reducing opportunities for skills development through experience. They also indicated that post-editing the output of AI-machine translation tools and proofreading text translated by colleagues entail very different processes and work, because the nature of the mistakes and problems in the two types of text is not the same. For example, whereas machines can now create convincing sentences, these are considerably more likely to contain major interpretation flaws compared to the translation output produced by professional translators.

Among those who did not consider AI-machine translation tools to have changed the nature of the work conducted by language professionals, AI-machine translation tools were viewed, so far, as aids that could be used to more efficiently and speedily perform tedious and repetitive tasks. By contrast, human input was considered to remain central to performing certain critical translation tasks, and as the driver of decisions over which tasks should be automated and which should continue to be predominantly or solely performed by humans. At the same time, professionals taking part in the survey did not consider AI-machine translation tools so far to outperform more traditional CAT tools, which many felt were more reliable than AI-machine translation tools.

Most respondents indicated that they use AI-machine translation tools in their work on a daily basis. At the same time such use appears to differ greatly across respondents. Some reported using AI-machine translation tools to post-edit content that was previously generated using AI-machine translation tools. Others reported using them to help them draft texts, while others indicated using AI-machine translation tools for '*inspiration*', i.e. '*feeding individual sentences or parts of sentences into the tool to get a few alternatives for how to translate a term or phrase*'. At the same time, some translators indicated that they were reluctant to use AI-machine translation tools because post-editing text generated by such tools requires attention to a different set of potential mistakes and post-editing assignments of work generated using AI-machine translation tools usually require very tight deadlines that leave little or no time to do the research work that would be required to ensure substantive accuracy.

Professionals interviewed expected AI-machine translation tools to have a significant impact on the demand for language professionals, their working conditions, the tasks they will be asked to perform and the skills they will need to possess to succeed in their work.

When respondents were asked to consider how the work of language professionals might change in the following five to ten years, many indicated that they expected the volume of the content that will be machine translated to increase dramatically, alongside a growing demand for rapid post-editing or full post-editing. Many respondents feared that such tasks will not be as well remunerated as the tasks currently performed by translators and therefore will attract less qualified professionals, despite the fact that high levels of skills are often needed to be able to post-edit text. In particular, to be able to rapidly post-edit text, translators engaging in post-editing need to possess significant knowledge of subject areas the text deals with as well as detailed knowledge of how a particular machine translation tool operates – i.e. what data it was trained on, what errors it typically generates, what systematic bias should be expected, what shortcomings to expect, to be able to identify and correct such biases as well as hallucinations. Some wondered about the extent to which acquiring high levels of skills in post-editing will be possible if junior translators will be deprived of the extensive and direct experience of engaging in original translation. In particular, one respondent indicated that '*today only senior translators seem to be able to most efficiently post-edit because their experience enables them to perceive the inaccuracies in a machine-translated text*'.

A recurring observation among respondents was the fact that individuals and institutions commissioning translations generally overestimate the quality of the output generated by AI-machine translation tools and underestimate the amount of time and the skills needed to ensure accuracy and quality through post-editing. To the extent that such perceptions will not change or the technology will dramatically improve, they expected that the working conditions of most language professionals will deteriorate. At the same time, many respondents reported that they expected the number of professionals working on original translations to decline sharply but that few, highly qualified, highly trained, and exceptionally talented professionals will continue to engage in transcreation of highly creative texts or to work in contexts where accuracy is crucial, such as the medical field or the criminal justice system.

When they were asked to consider the implications of development in AI for the education and training undertaken by aspiring language professionals, respondents indicated that they felt all new translators should receive high quality training in using AI tools as well as other digital resources such as CAT and terminology tools during their studies since ability to use such tools in their work is and will become an even more unavoidable part of the language industry. They also reported that, although investments in continuous professional development activities always played an important role, they will become essential and unavoidable for language professionals since technology will continue to evolve rapidly over time and engaging in lifelong learning will be the only way for these professionals to keep up with changing skills needs. They emphasised that it will be increasingly important for language professionals to acquire an understanding of how machine translation works and what errors and biases they are likely to encounter when facing using AI-machine translation technologies. They should also acquire the knowledge and experience needed to post-edit AI machine translated text efficiently and successfully. Other respondents emphasised the importance for education and training systems to continue equipping new professionals with very high levels of subject knowledge, since such knowledge is crucial in the post-editing phase to be able to understand the subtext and context and allows language professionals to detect inconsistencies and inaccuracies. Similarly, respondents indicated that education and training systems should continue to equip learners with the ability to structure their thinking and their work, to analyse business cases, to design workflows, test them, streamline them, and make them scalable. Many felt that communication and problem-solving skills will increasingly play an important role because of advances in AI technologies.

Most translators indicated that they used regularly AI-machine translation tools in their daily life outside of work, for *gisting* purposes in languages they do not know when accuracy does not matter and when they do not need to rely on the information contained in a text. Some indicated that they used such tools to produce text in their non-mother tongue languages, but they indicated that they used the text as the basis for further editing. Others indicated that outside of work they preferred to switch off all digital devices and take a deep breath in nature.

Source: Questionnaire distributed to a selected number of translators by the authors.

4 An empirical analysis of the demand for and the skills of language professionals

37. This section evaluates the demand for language professionals and the skills sought by employers in language professional job postings. Analyses are based on online vacancy data from Lightcast for the period 2015 and 2019 (see Box 4.1). The main analyses were conducted for online job postings available in English-speaking countries covering Australia, Canada, New Zealand, Singapore, the United Kingdom, and the United States. We complement this with evidence for four selected European Union countries – France, Germany, Italy and Sweden.

Box 4.1. Data description

Data collection

Lightcast (formerly known as Emsi Burning Glass) is a labour market analytics company that collects and standardises data on online job vacancies. Data on online vacancies collected by Lightcast are rich both in terms of size and granularity: they cover a wide range of vacancy sources and provide a variety of information for each job posting. Lightcast collects postings from over 80 000 online job sites to develop a comprehensive, real-time portrait of labour market demand including vacancies posted directly by employers and vacancies posted by agencies advertising temporary staffing needs thus covering a range of work opportunities available to freelance professionals. Lightcast identifies websites with employment-opportunity-related content using spider technology to search those sites for employment opportunities. Based on a retrieved list of job postings, job postings are deduplicated to avoid the same posting appearing multiple times. Data are extracted from job posting texts including company, industry, occupation, skills. Lightcast then uses natural language processing to identify what skills employers seek in their prospective employees based on information made available in the text of the posting. For each posting a number of skills demanded are identified. Comprehensive overviews of the data collection process, representativeness and limitations of Lightcast online job postings is provided in Brüning and Mangeol (2020[154]) and Samek, Squicciarini and Cammeraat (2021[155]). Although the data provide rich information on vacancies and the perspective of employers, they identify job openings (labour market flows) rather than employment levels (stocks). Furthermore, they reflect the expectations of prospective employers and no information from existing employees on the set of tasks and skills they perform. They only contain qualitative information on the range of skills employers seek rather than quantitative information on the level of proficiency employers expect employees to possess.

Analysis of vacancies of language professionals

Analyses conducted in this paper are based on online job vacancies for English-speaking (Australia, Canada, New Zealand,[3] Singapore, the United Kingdom, the United States) and selected EU countries (France, Germany, Italy, Sweden) for the time period 2015-2019. Due to the impact of the COVID-19 pandemic on labour markets, analyses are restricted to years prior to the outbreak of the pandemic. The analyses are further restricted to the occupation of *Professional Translators and Interpreters* (ISCO = 2643), also referred to as language professionals in the paper. Given the small overall number of vacancies for language professionals it is possible that for individual countries data may be volatile since changes in the requirements of a large company may lead to visible, sudden one-off increases in demand.

Whereas skills information from English-speaking countries was collected and stored in English, data for France, Germany, Italy and Sweden was collected in a range of languages (most texts were in the official language of the country being considered although other languages – including English were also used in some vacancies).

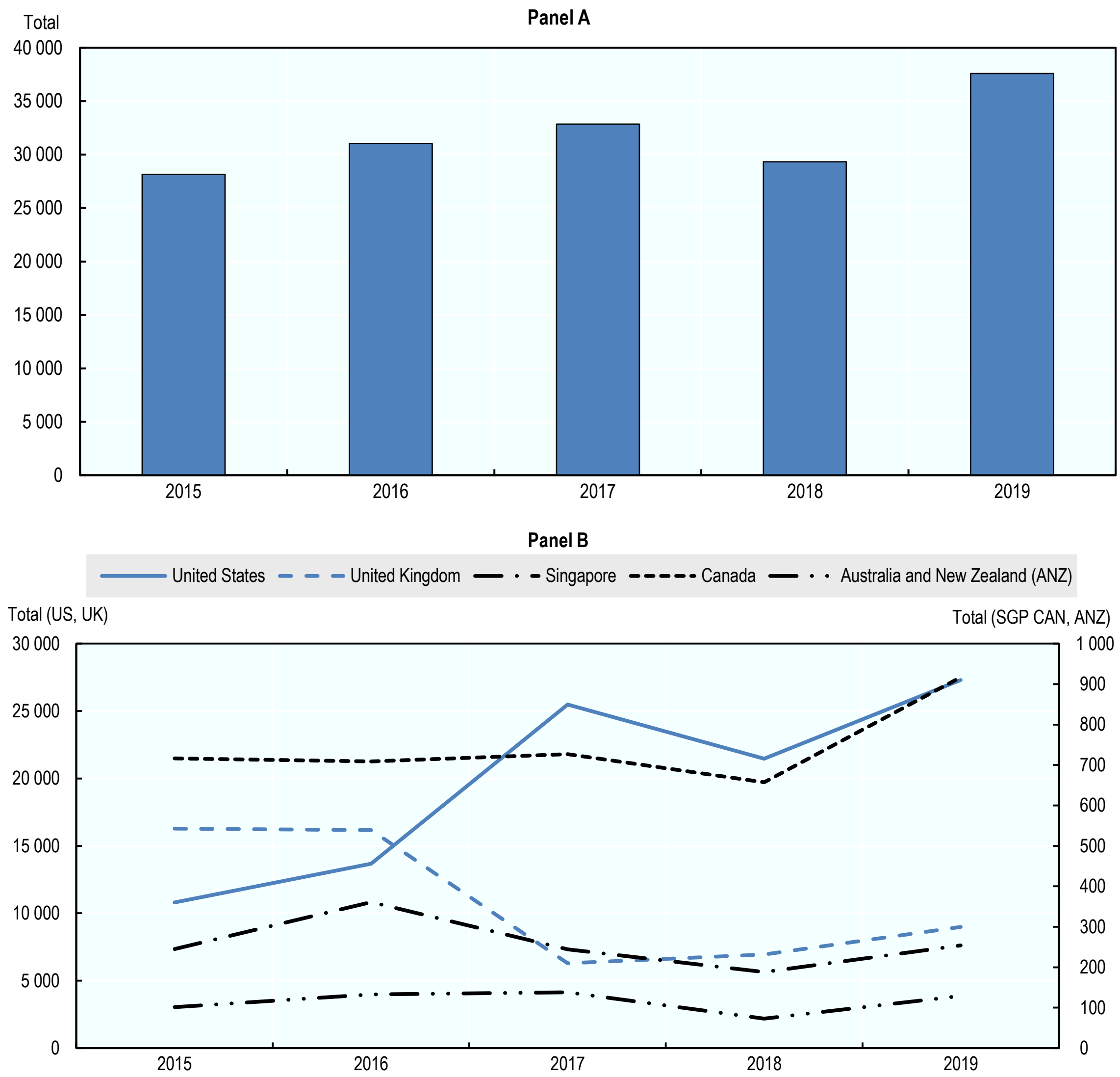
4.1. Evolution of online job postings for language professionals

38. Figure 4.1 illustrates the evolution of the total number of online job postings for language professionals between 2015 and 2019 for English-speaking countries Australia, Canada, New Zealand, Singapore, United Kingdom, and the United States. Panel A shows that total number of vacancies posted on-line for language professionals for all countries are rather stable over time. While around 28 000 online job postings were observed in 2015, four years later, in 2019, this number slightly increased by around 30%, totalling 37 000. Panel B shows the trend in the total number of vacancies for language professionals for the United Kingdom and the United States (left y-axis) and for Australia, Canada, New Zealand and Singapore (right y-axis). Increases in the total number of postings over time are observed for the United States, where online job postings more than doubled from 10 815 in 2015 to 27 310 in 2019. In Australia, New Zealand and Singapore, job postings have been more or less stable over time. For the United Kingdom, the total number of online job postings almost halved from 16 280 in 2015 to 8 994 in 2019. No explanations for this decline can be inferred from these data.

[3] Data for Australia and New Zealand are aggregated.

Figure 4.1. Trend in online job postings for language professionals, English-speaking countries (2015-2019)

Total number of online job postings for language professionals, by year



Note: Panel A shows the total number of online job postings for language professionals between 2015 and 2019 for Australia, Canada, New Zealand, Singapore, the United Kingdom, and the United States. Panel B shows the development of the total number of online job postings for language professionals by country, while numbers for the United States and the United Kingdom are shown on the left y-axis and for Singapore, Canada, Australia and New Zealand on the right y-axis.
Source: Authors' own compilation based on Lightcast™ (December 2022).

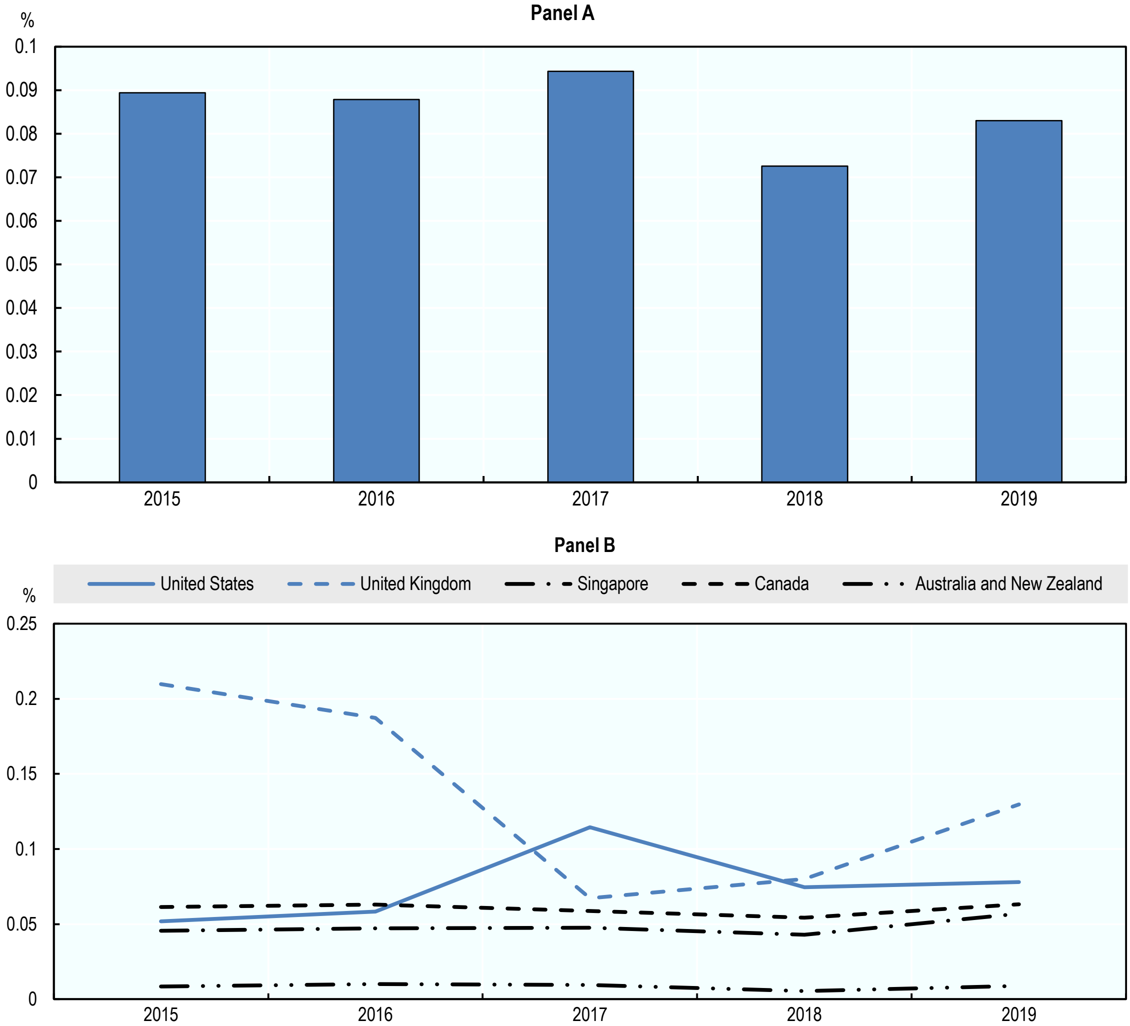
39. This relative stability in the total number of online job postings for language professionals over time does not necessarily reflect an increase in the demand and might reflect greater coverage of Lightcast of online labour markets or an increase over time in short-term contracts and turnover among language professionals. Overall vacancies published on-line in English-speaking countries increased from around 31 million to around 45 million between 2015 and 2019, an increase of more than 40%, reflecting the trend of more and more vacancies being published on line, being captured by Lightcast or an increase in short term contracts.

40. Figure 4.2 provides information on trends between 2015 and 2019 in the relative demand for language professionals relative to the overall demand expressed in terms of job vacancies by mapping trends in online job postings for language professionals as a percentage of overall online job postings. Panel A shows that the overall share of online job postings for English-speaking countries between 2015 and 2019, remained relatively stable. The share of language professionals did not change greatly varying between 0.07 and 0.09% of overall online job vacancies. Panel B provides the relative trend by country. Similarly to Panel A, results indicate that overall, the share of vacancies for language professionals remained stable between 2015 and 2019. In addition to English-speaking countries, absolute and relative trends in the demand for language professionals in selected EU countries are provided in Box 4.2.

41. The relative overall stability of online job postings for language professionals suggests that machine translation technologies did not replace human language professionals during the time period. Possible reasons as to why no descriptive evidence for a substitution effect on a larger scale can be observed is that machine translations accuracy still lags behind the accuracy of human professionals, that the quality of speech recognition hinders the quality of machine translations to replace effectively the work of human interpreters, and that machine translations cannot easily and readily communicate with clients to check on meaning and maintain quality. Instead of a substitution between humans and machines, a complementarity between these two might be at play, potentially leading to a specialisation of human translators into specific knowledge skills, post-editing skills and other forms of complementarity with machine translation tools. Essentially, the new translators' role might be to take care of translation parts that machines cannot yet tackle such as contextual meanings or cultural context.

Figure 4.2. Relative trend in online job postings for language professionals, English-speaking countries (2015-2019)

Percentage of online job postings for language professionals relative to online job postings, by year



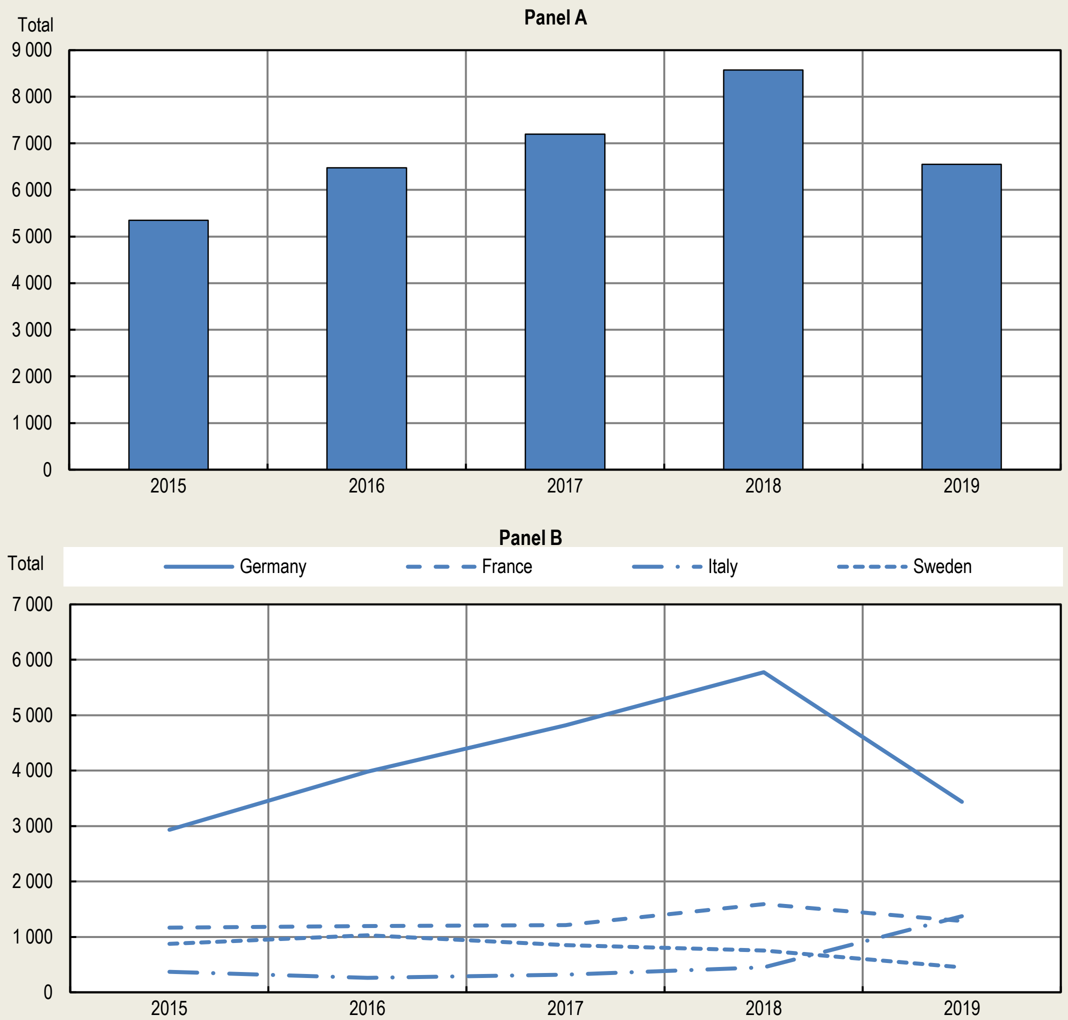
Note: Panel A shows the percentage of online job postings for language professionals relative to online job postings in a given year for Australia, Canada, New Zealand, Singapore, United Kingdom, and United States. Panel B shows the percentage of online job postings for language professionals relative to online job postings by country.
Source: Authors' own compilation based on Lightcast™ (December 2022)

Box 4.2. Evolution of language professionals in the EU

Panel A in Figure 4.3 shows the evolution of the total number of online job postings for language professionals between 2015 and 2019 for France, Germany, Italy and Sweden. It illustrates a slight increase in postings from 5 347 in 2015 to 6 546 in 2019. Overall, the slight increase in the total number of language professionals observed for selected EU countries is similar to the trend observed for English-speaking countries. Panel B in Figure 4.3 shows that in absolute numbers, job postings for language professionals were highest in Germany, followed by France. For all countries but Sweden, the total number of online job postings slightly increased between 2015 and 2019.

Figure 4.3. Evolution of absolute job postings, EU countries, (2015-2019)

Total number of online job postings for language professionals, by year

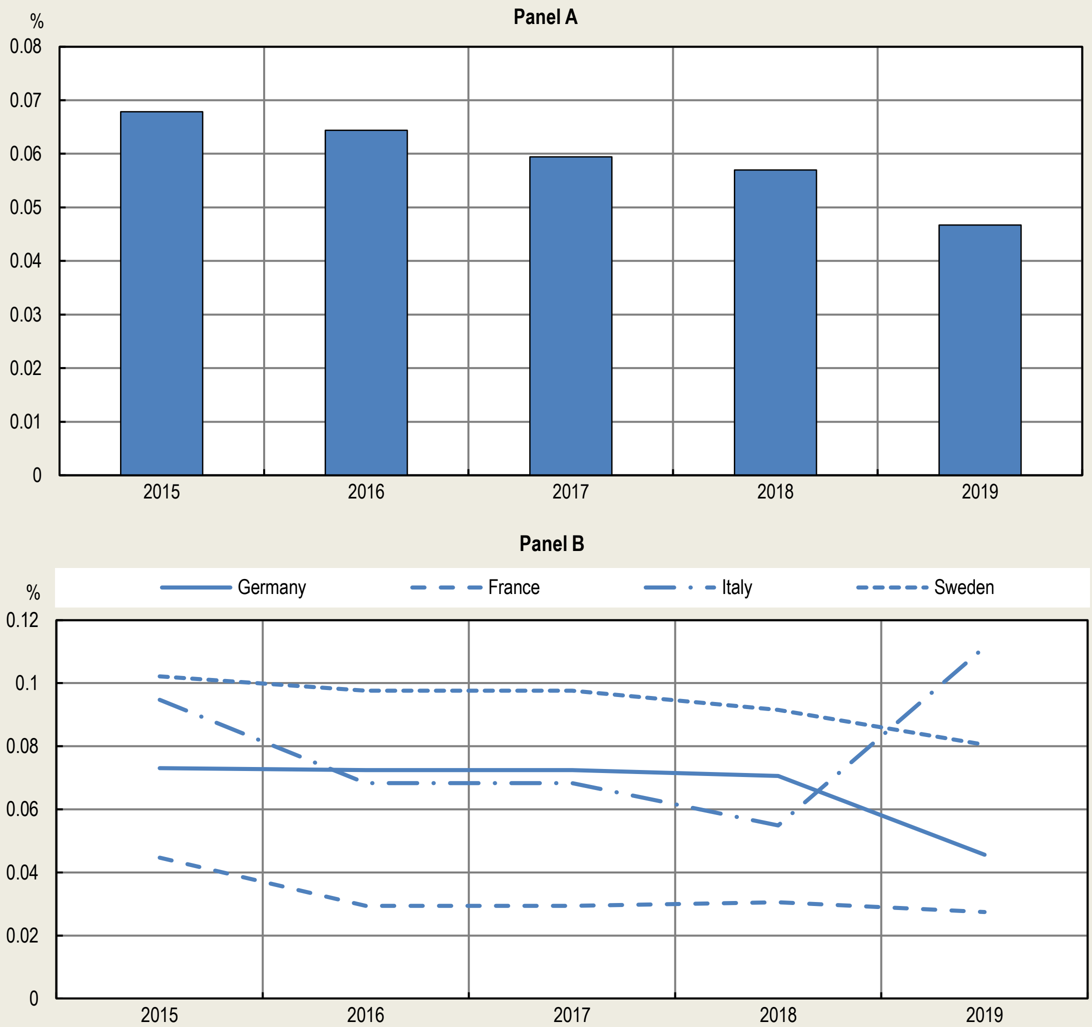


Note: Panel A shows the total number of online job postings for language professionals between 2015 and 2019 for selected European Countries: Germany, Italy, France and Sweden. Panel B shows the total number of online job postings for language professionals by country.
Source: Authors' own compilation based on Lightcast™ (December 2022).

Figure 4.4 shows the evolution of the relative share of job postings for all (Panel A) and by selected European Union countries (Panel B). Although the figures show a slight downward trend, the change is very small. The drop in the relative share between 2015 and 2019 is in fact just 0.021 percentage points (Panel A). A drop in the relative share of job postings does not necessarily reflect a drop in job postings of language professionals. Instead, it could reflect that other occupations are growing more in terms of shares, or the fact that certain contracts are not easily captured by online job posting data, such as self-employed contracts. Annex Figure A.1 illustrates the variation across selected EU countries in the percentage of online job postings available that are explicitly targeting freelancers, i.e. self-employed individuals, This slight downward trend is also reflected in Panel B in Figure 4.4, where in France, Germany and Sweden, the relative share of job postings is lower in 2019 than in 2015 and only for Italy, a small increase between 2018 and 2019 is observed.

Figure 4.4. Evolution of relative job postings, EU countries, (2015-2019)

Percentage of online job postings for language professionals relative to online job postings, by year



Note: Panel A shows the percentage of online job postings for language professionals relative to online job postings in a given year between 2015 and 2019 for selected European Union countries: France, Germany, Italy, and Sweden. Panel B shows the percentage of online job postings for language professionals relative to online job postings by country.
Source: Authors' own compilation based on Lightcast™ (December 2022).

4.2. The changing skills content of language professional occupations

42. Lightcast data detail, for each online job vacancy, skills keywords required by employers. On average in English-speaking countries each job posting mentioned 8 skills as being required or desired in 2015. This had grown to 12 in 2019. However, when considering all different skills labels identified in all job postings over 17 000 were mentioned and coded in the Lightcast dataset. In order to meaningfully analyse skills keywords, these need to be grouped into meaningful skills categories. Box 4.3 provides more information on skills keywords and grouping.

Box 4.3. From skills keywords to skills groups

Online vacancy data information by Lightcast provide a list of extracted skills keywords which are demanded in a specific job posting. The skills keywords extracted from all job postings amount to over 17 000 distinct skills keywords and a categorisation in broader categories is necessary in order to analyse such a large amount of information.

This paper builds on the taxonomy and mapping of over 17 000 distinct skills keywords into 61 skills categories that was developed by Lassébie et al. (2021[156]). The taxonomy is based on the Occupational Information Network (O*NET) database (Occupational Information Network, n.d.[33]) from the U.S. Department of Labor augmented by certain skills from the European Skills/Competences, Qualifications and Occupations (ESCO) classification. The mapping of skills keywords into each of the 61 skill categories is conducted using a supervised learning approach using the same data and English-speaking countries as in the present paper.

In the analyses presented the terms ‘skills’ and ‘skills keywords’ are used interchangeably whenever ungrouped skills extracted from job postings are analysed whereas the term ‘skills groups’ is used whenever skills are categorised or aggregated into broader skills groups.

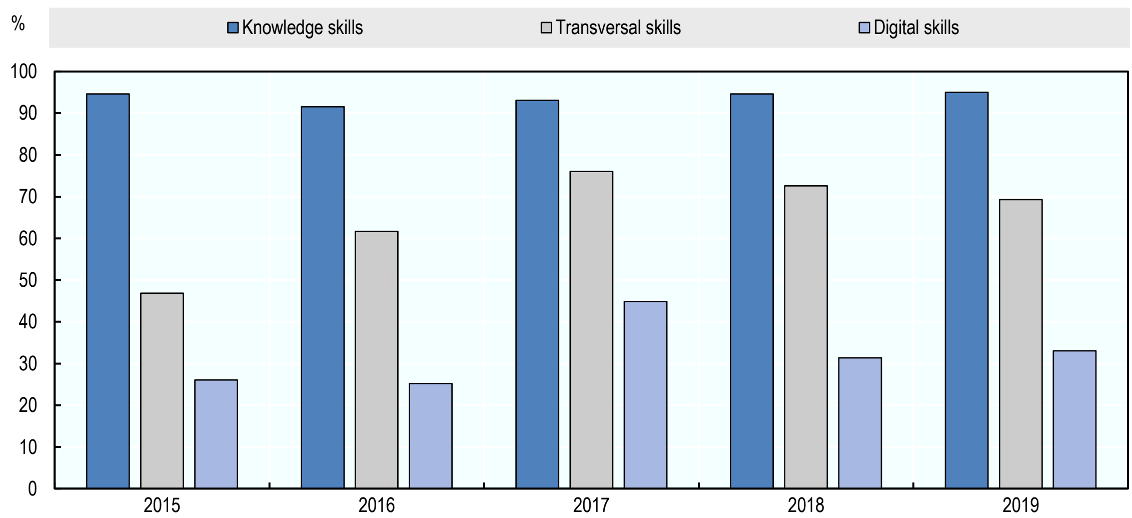
43. The following analyses focus on whether and how skills change. In order to do this, analyses consider broad aggregated skills groups up to more disaggregated skills groups. For the broadest level of aggregation, 61 skills categories were grouped into three groups: ***Knowledge skills***, ***transversal skills*** and ***digital skills*** (Table A.1 provides an overview on the categorisation of the different skills categories into skills groups). ***Knowledge skills*** represent skills categories that represent specific thematic knowledge (e.g. Biology, Medicine), ***transversal skills*** represent skills categories which are “seen as necessary or valuable for effective action in virtually any kind of work, learning or life activity” and “they are not exclusively related to any particular context” (e.g. *co-ordination*, *motivation*) (European Commission and CEDEFOP, 2021[157]). Lastly, ***digital skills*** represent digital skills categories such as *digital data processing or computer programming.*

44. Figure 4.5. illustrates the percentage of online job postings for language professionals mentioning any of the broader skill groups ***knowledge***, ***transversal*** and ***digital***. The ***knowledge group*** represents theoretical and factual knowledge acquired as the outcome of assimilating information through learning (European Commission, Directorate-General for Employment, Social Affairs and Inclusion, 2017[158]). Knowledge related skills keywords are mentioned most frequently in online job postings of language professionals. Between 2014 and 2019, the share of job postings mentioning **knowledge skills** is relatively stable: around 95% of job postings mentioned knowledge skills. ***Transversal skills***, relevant to a broad range of occupations and economic sectors (European Commission, Directorate-General for Employment, Social Affairs and Inclusion, 2017[158]), are the second most common mentioned skills group among language professionals. In 2015, 47% of postings mentioned ***transversal skills***, while this share increased to around 70% in 2019, an increase of more than 20 percentage points. Finally, the share of ***digital skills***,

remained relatively stable despite small fluctuations between 2015 and 2019. Between 2015 and 2019, the share of postings mentioning **digital skills** varied between 26% in 2015 and 33% in 2019.

45. These analyses show that while the demand for ***digital*** and ***knowledge skills*** remained stable over the study period, there was a slight increase in the demand for ***transversal skills***. This suggests that the increase in the demand for ***transversal skills*** did not come at the expense of other skills but, rather that the range of skills required of individual professionals broadened over time.

Figure 4.5. Percentage of online job postings for language professionals mentioning knowledge skills, transversal skills and digital skills, English-speaking countries (2015-2019)

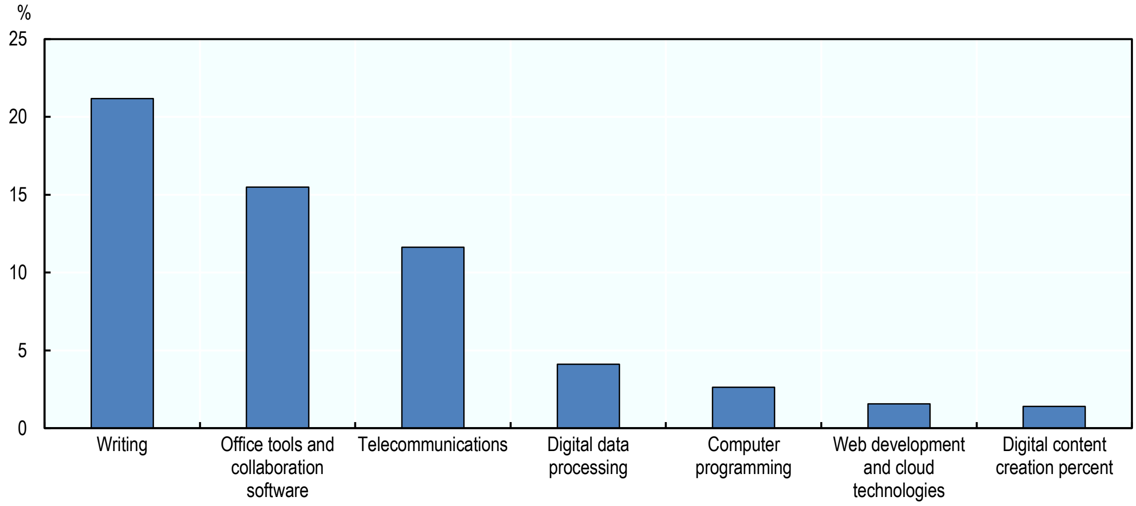


Note: The figure indicates the share of online job postings by language professionals mentioning knowledge, transversal and digital skills. Shares do not add up to 100% since several skills can be mentioned in one posting.
Source: Authors' own compilation based on Lightcast™ (December 2022).

46. Figure 4.6 provides a more granular picture of specific skills categories, in particular the ***digital skills*** group given this papers' focus on the impact of technology on skills. Even though Figure 4.5. did not show an increasing trend in the demand for ***digital skills***, changes could occur within skills subgroups. Figure 4.6 provides an overview of seven skills groups reflecting ***digital skills***: *Computer programming, digital data processing, digital content creation, office tools and collaboration software, telecommunications*, *web development,* and *writing*. ***Digital skills*** are requested in around one fifth of online vacancies advertising positions for language professionals. For example, on average, 21% of online job postings requested *writing skills*, 15% requested skills related to *office tools and collaboration software*, and 12% requested *telecommunication skills*. For the remaining skills, they were requested in less than 10% of online vacancies. Examples of highly demanded writing skills include: editing and post-editing, computational linguistics, machine translation (MT), computer-assisted translation (CAT), consecutive translation. Examples of highly demanded digital data processing skills include: machine learning, big data, metadata, ArcGIS, Data science. Examples of highly demanded computer programming skills include: artificial intelligence, C++, Linux, Java, Java Script.

Figure 4.6. Digital skill groups demanded in online job postings for language professionals, English-speaking countries (2015-2019)

Percentage of online job postings containing at least one skill that belongs to each of the categories



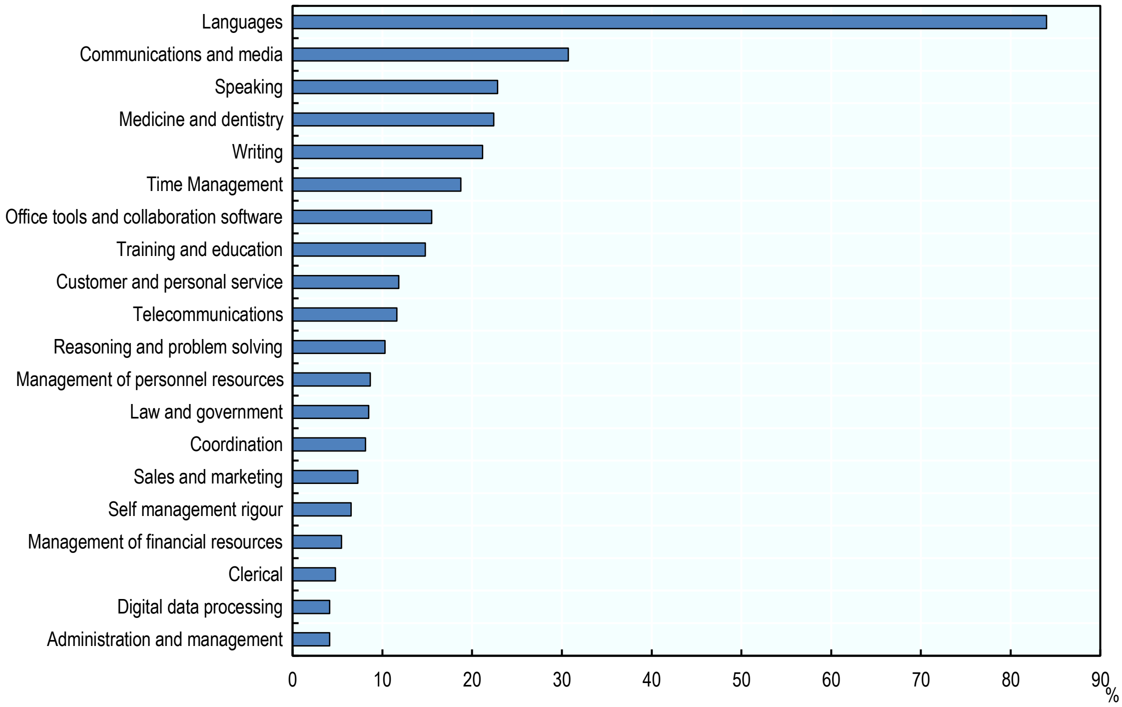
Note: The figure presents the average percentage of online job postings containing at least one skill that belongs to each of the categories. For each country – United States, United Kingdom, Australia, Canada and New Zealand, the average over the 2015-2019 period was calculated and then the average over the five countries was calculated.
Source: Authors' own compilation based on Lightcast™ (December 2022).

47. Figure 4.7. shows the top 20 out of 60[4] most frequently mentioned skills groups in online job postings of language professionals between 2015 and 2019. Among the top 20 skills groups, 45% represent ***transversal skills***, 40% represent ***knowledge skills*** and 15% represent ***digital skills***. Among the 20 most frequently mentioned skills groups in postings of language professionals, the top 5 skills groups mentioned in more than 1 out of 5 vacancies were: *language skills* (84%), *communications and media* (31%), *speaking* (23%), *medicine and dentistry* (22%) and *writing* (21%). *Writing* is grouped into the *digital skills* group because many of the detailed skills in that group have technology aspects. For instance, skills keywords such as *computational linguistics*, *machine translation* and *computer-assisted translation*, all belong to the *writing skill* category.

[4] The groups *Local Language* and *Foreign Language* from the taxonomy in Lassébie et al. (2021[156]) are aggregated into one category *Languages* in this working paper.

Figure 4.7. The top 20 skill groups of language professionals demanded by employers, English-speaking countries (2015-2019)

Percentage of online job postings mentioning one of the top 20 skill groups, by skills group



Note: The figure shows the percentage of online job postings mentioning one of the top 20 skills groups. The skill groups are ordered in descending order of the average percentage of online job postings mentioning that skill group between 2015 and 2019. The figure presents the average percentages averaging data from each country – United States, United Kingdom, Australia, Canada and New Zealand –over the 2015-2019 period.
Source: Authors' own compilation based on Lightcast™ (December 2022).

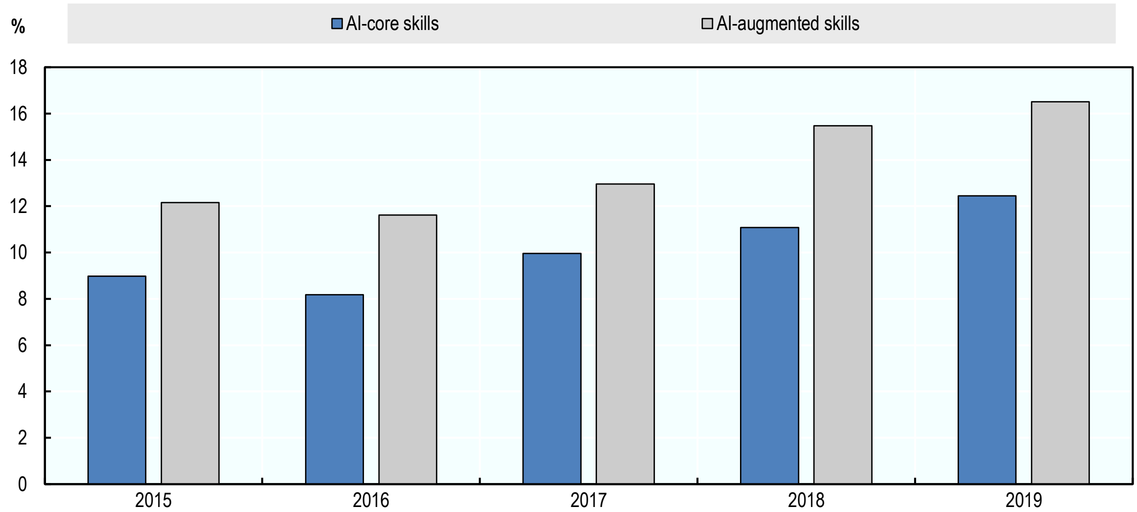
4.3. The evolution of skills related to the use of machine translation technology

48. Figure 4.8. illustrates the extent to which online job vacancies for language professionals demand prospective employees to possess skills related to artificial intelligence and machine learning. Two separate categorisations of artificial intelligence skills are used: the **AI-core skills** category borrows the AI-related skills categorisation developed by (Alekseeva et al., 2021[159]) using Lightcast data. The **AI-augmented skills** category builds on the core category but augments it with additional AI skills related to "computer programming" skills in the taxonomy developed by Lassébie et al., (2021[156]).

49. Figure 4.8. shows the evolution of job postings for language professionals requiring prospective applicants to have AI skills. Irrespective of the AI categories used, Figure 4.8. indicates that on average around one in ten postings for language professionals requested AI-core skills and around 14% requested AI-augmented skills. Furthermore, Figure 4.8. suggests an increase in the percentage of postings demanding AI skills from 2016 onward. While the share of postings requiring AI-core skills increased from 9% to 12% between, postings requesting AI-augmented skills increased from 12% to 17% between 2015 and 2019. The evolution of AI related skills in selected European Union countries is provided in Box 4.4.

Figure 4.8. Postings for language professionals mentioning AI skills, English-speaking countries (2015-2019)

Percentage of online job postings mentioning AI skills



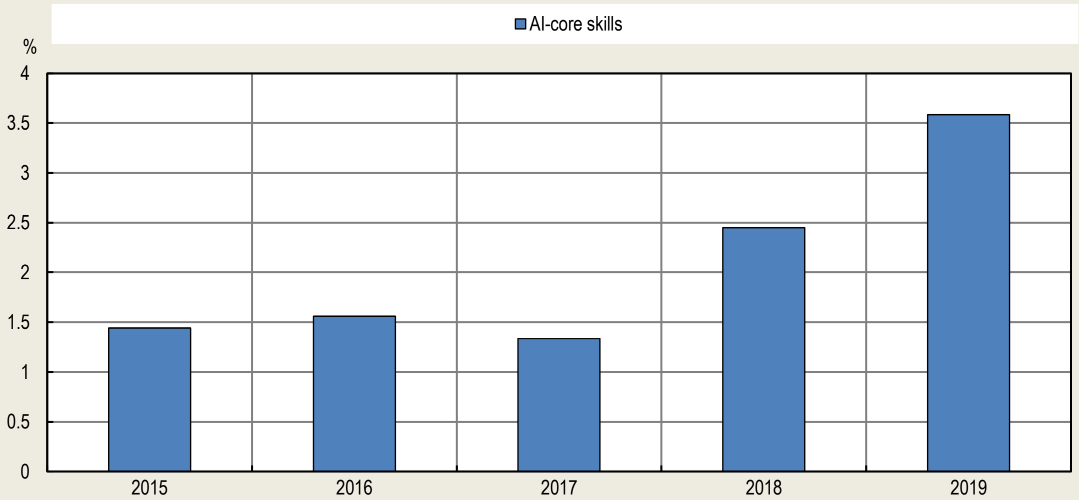
Note: Percentage of job postings mentioning AI skills is based on two different categorisations. The category AI-core skills uses a list of skills based on Alekseeva et al. (2021[159]), while the category AI-augmented skills complements this list with further AI skills related to "computer programming" skills in the taxonomy developed by Lassébie et al. (2021[156]).
Source: Authors' own compilation based on Lightcast™ (December 2022).

Box 4.4. Evolution of AI skills in the EU

Figure 4.9 shows AI-core skills[5] for EU countries which were extracted from Lightcast (an overview on these skills is provided in Table A.2.). In 2015 AI-core skills were mentioned in 1.4% of job postings and increased to 3.6% in 2019. The relevance of AI related skills is increasing as can be seen for countries in the European Union but also English-speaking countries as shown in Figure 4.8. . Data presented in this Figure 4.9 are not directly comparable to data presented in Figure 4.8. since the two utilise slightly different taxonomies given the different skill categorisation in English speaking countries and European Union countries developed by Lightcast.

Figure 4.9. Postings for language professionals mentioning AI skills, English-speaking countries (2015-2019)

Percentage of online job postings mentioning AI skills



Note: Percentage of job postings mentioning AI-core skills with skills keywords extracted from Lightcast for selected European Union countries: Germany, Italy, France and Sweden.
Source: Authors' own compilation based on Lightcast™ (December 2022).

50. While Figure 4.8. shows an overall increase in AI skills requested in English-speaking countries, the following sheds light on selected skills strongly related to machine translation technology. These are on the one hand skills related to the usage of *machine translation technology* and on the other hand skills related to the *post-editing* of machine translation output. In the following, we provide examples of these skills. In order to get a grasp of how frequently such skills appear in online job postings, Box 4.5 looks at examples related to machine translation technology, while Box 4.6 looks at skills related to post-editing.

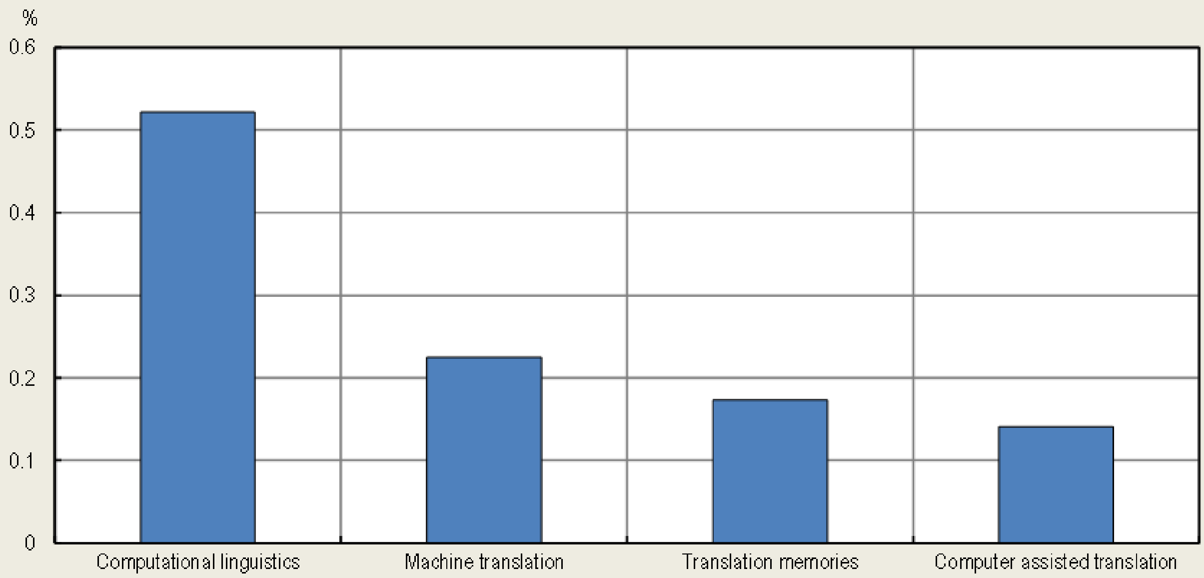
[5] AI-core skills for European countries differ from English-speaking countries in the type of AI keywords.

Box 4.5. Example of technical skills related to the use or development of machine translation technology

Figure 4.10 provides an overview of skills related to machine translation. Machine translation skills represent either skills related to the use or development machine translation technologies themselves. The share of postings mentioning a specific skills keyword related to AI skills is small since multiple keywords can refer to the broader skill requirement of AI skills. This is why, for example, in English-speaking countries in the sample around 10% of all postings mentioned AI-core skills but only around 0.5% of postings require the specific skills related to computational linguistics, 0.2% related to machine translation and less than 2% skills related to translation memories or assisted translation.

Figure 4.10. Skills groups related to machine translation technologies, English-speaking countries (2015-2019)

Percentage of online job postings mentioning machine translation related skills



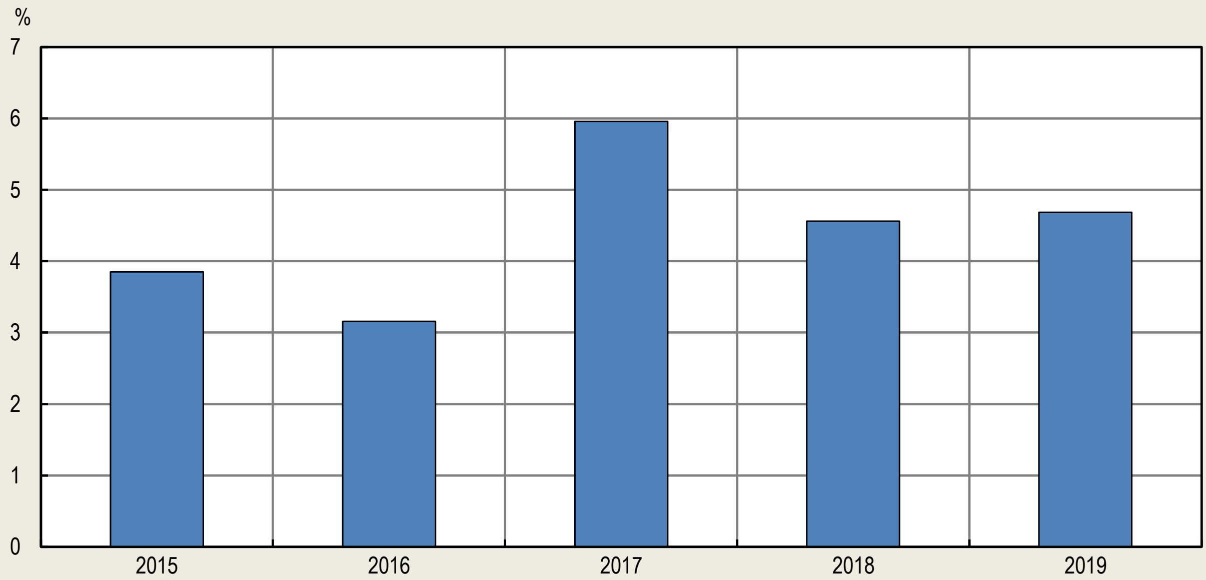
Note: The figure shows the average percentage of online job postings mentioning skills related to machine translation technology for English-speaking countries over the 2015-2019 period.
Source: Authors' own compilation based on Lightcast™ (December 2022).

Box 4.6. Example of skills groups related to editing and post-editing

Various skills emerge in online vacancy data that include editing and post-editing, meaning the processing of machine translated output. The growing importance of editing and post-editing skills could be expected considering the progressive digitisation of the profession and the quality gains associated with it (Garcia, 2011[144]). As Koehn (2009[160]) mentions, "If the goal of translations is to produce publishable text, machine translation might act as a first step, a human post-editor may then correct the output to a desired level of quality". Figure 4.11. shows the aggregated percentage of job postings requiring any skills related to editing and post-editing. This ranged from 3.9% in 2015 to 5.9% in 2017.

Figure 4.11. Skills groups related to editing and post-editing, English-speaking countries (2015-2019)

Percentage of online job postings mentioning editing and post-editing skills



Note: Percentage of online job postings mentioning skills related to editing and post-editing for English-speaking countries over the 2015-2019 period.
Source: Authors' own compilation based on Lightcast™ (December 2022).

5 Conclusions

51. "*Every time I fire a linguist, the performance of the speech recognizer goes up*" Frederick Jelinek[6] famously quipped when developing models for speech recognition and machine translation. In fact, over the past decades, advances in machine translation quality were achieved thanks to the move from theory driven linguistic models to statistical prediction models. However, it is equally true that over a decade after Jelinek's death the translation output developed using the most advanced machine translation models cannot match the quality of the output produced by professional human translators. In the inter-language communication market one could therefore claim what while Box's remark that '*all models are wrong but some are useful*'[7] continues to reflect reality. Although advances in machine translation technology have been driven by the development of theory-free, data driven prediction models (such as neural models) which have long outperformed theory driven linguistic models, the theory and practice guiding the work of professional linguists remains critical to the development of high-quality translations.

52. Machine translation technologies can play a key role to facilitate inter-language communication: machine translation in fact can deliver output cheaply, timely and quickly, and can tackle large volumes of data to be translated. However, the quality of the output remains lower than the quality of the output delivered by professional language professionals, especially for translations from/to low-resource languages and text loaded with cultural significance. Moreover, as the quality of machine translation tools depends on the quality of the data used to train the systems, such quality builds on prior work conducted by human translators, as well as on quality checks and input apported by human translators. The Mechanical Turk was an automatic chess-playing device in Vienna in 1770 by Wolfgang von Kempelen which appeared to be able to play chess autonomously whereas it was an elaborate shell in which a hidden human chess master directed the movements of the machine (Stephens, 2022[161]). Although machine translation tools are much more than empty shells, like the Mechanical Turk and other forms of artificial intelligence systems, they rely on the hidden work, past and present, of professional human translators to operate effectively.

53. By contrast, human translators can deliver high quality output and can engage in co-production with those requiring translations but they have clear capacity constraints: they require significant time to deliver translations, can only tackle small volumes of text at any one time and the marginal cost of each translation is high relative to the close to zero marginal cost of the output produced by machine translation tools.

54. Analyses based on data on online job vacancies for language professionals for a selected number of OECD countries between 2014 and 2019 indicate that the introduction of higher quality machine translation system did not lead to decreases in the demand for language professionals. This is in line with recent trends in employment and forecasts in employment in the United States which suggest increased

[6] Fred Jelinek was a pioneer in the field of automatic speech recognition and natural language processing. The exact origins are disputed and several alternative wordings have been proposed. In this work the quote is taken from the speech Hirschberg, Julia gave in 1998 at the 15th National Conference on Artificial Intelligence, Madison, Wisconsin titled "Every time I fire a linguist, my performance goes up", and other myths of the statistical natural language processing revolution.

[7] George Edward Pelham Box was a mathematician and professor of statistics.

opportunities for language professionals. This evidence suggests that, over the period, language professionals were not substituted by machine translation tools and, rather that machine translation tools complemented the work of language professionals.

55. It is not possible given the data at hand to detail with precision to what extent the introduction of increasingly accurate machine translation tools between 2014 and 2019 led to complementarity between machines and language professionals in the development of individual translation outputs or, rather, led to the specialisation of language professionals for the delivery of certain texts and machines being used for other types of texts. However, analyses of the set of skills demanded in vacancies for language professionals highlight: 1) the continued and growing relevance of *transversal skills* and *knowledge skills* – skills that allow language professionals to deliver high quality translations of text that machine translation tools cannot deliver with accuracy –2) the key role played by *digital skills* which have been an important component of the skill set required of language professionals at least since 2014 , and 3) skills that allow them to work alongside machine translation tools, such as, for example *post-editing skills*. In fact, as many as one in ten vacancies for language professionals required prospective applicants to have AI-core skills, i.e. skills that are directly associated with the knowledge of AI technology or the ability to use AI-related software, such as ”artificial intelligence”, ”machine vision”, ”deep learning” or ”speech recognition” (Alekseeva et al., 2021[159]), and the share of vacancies requiring AI-core skills increased, albeit slightly between 2014 and 2019.

56. The introduction of machine translation tools is reshaping the labour market opportunities of language professionals but also the opportunities businesses and individuals have to engage in inter-language communication. Today, language mediation systems are available at one's fingertips: one can request the translation of a text on a smartphone screen and receive immediate and free mediation. An expanded choice set can yield important societal benefits by broadening access to information produced in multiple languages. At the same time, unless potential users of language mediation systems will be educated about the opportunities and limitations of alternative language mediation providers, it is possible that digital mediators will create harm as well as enhance opportunities. In particular, as societies struggle to limit the spread and the deleterious effects of misinformation, disinformation and malinformation, the use of machine translation technologies could have the unintended effect of making such problem more acute (Caramancion, 2022[162]; Muda et al., 2021[163]). To the extent that un-checked translations of varying levels of accuracy and quality are made available to a large number of individuals and are accessed by individuals without a critical understanding of the nature of the translation process, machine translation tools could exacerbate existing problems associated with online content.

57. As Box 3.5 indicated, in high stakes settings machine translation tools remain inadequate, and the use of language professionals or even teams of language professionals should be privileged. One of the most notable examples of a commissioner recognising the critical role of translation was James Charles Stuart, also known as King James VI of Scotland (later named King James I of Great Britain and Northern Ireland at the death of Elizabeth I). At a time of deep religious fractures, in 1604 King James commissioned not a translator but a group of 47 scholars to translate the Bible books from Hebrew, Aramaic and Greek into English. In 1611 what became known as the 'Authorised version' of the Bible was published. The translated King James Bible established a canon for religious interpretation and practice and has, since then, been the reference for modern translations of the Bible and remains the English translation routinely used in Anglican and other English Protestant Churches over 400 years later. As the quest for timely, fast and cheap translations continues, the legacy over several centuries of major human translation efforts should be taken as a reminder of the potentially long-term benefits arising from investments in high-quality translations.

References

Acemoglu, D. and P. Restrepo (2020), “Robots and Jobs: Evidence from US Labor Markets”, *Journal of Political Economy*, Vol. 128/6, pp. 2188-2244, https://doi.org/10.1086/705716. [14]

Alcina, A., V. Soler and J. Granell (2007), “Translation Technology Skills Acquisition”, *Perspectives*, Vol. 15/4, pp. 230-244, https://doi.org/10.1080/13670050802280179. [148]

Alekseeva, L. et al. (2021), “The demand for AI skills in the labor market”, *Labour Economics*, Vol. 71, p. 102002, https://doi.org/10.1016/j.labeco.2021.102002. [159]

Amano, T., J. González-Varo and W. Sutherland (2016), “Languages Are Still a Major Barrier to Global Science”, *PLOS Biology*, Vol. 14/12, p. e2000933, https://doi.org/10.1371/journal.pbio.2000933. [62]

Amano, T. et al. (2021), “Ten tips for overcoming language barriers in science”, *Nature Human Behaviour*, Vol. 5/9, pp. 1119-1122, https://doi.org/10.1038/s41562-021-01137-1. [60]

American Translators Association (n.d.), *About the ATA Certification Exam*, https://www.atanet.org/certification/about-the-ata-certification-exam/ (accessed on 3 February 2023). [152]

Anaconda (2020), *2020 State of Data Science*, https://know.anaconda.com/rs/387-XNW-688/images/Anaconda-SODS-Report-2020-Final.pdf. [103]

Anderson, C. (2008), *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, https://www.wired.com/2008/06/pb-theory/. [1]

Arivazhagan, N. et al. (2019), “Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges”, https://arxiv.org/abs/1907.05019. [83]

Arntz, M., T. Gregory and U. Zierahn (2016), “The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis”, *OECD Social, Employment and Migration Working Papers*, No. 189, OECD Publishing, Paris, https://doi.org/10.1787/5jlz9h56dvq7-en. [26]

Artetxe, M. et al. (n.d.), “Unsupervised neutral machine translation”, *Published as a conference paper at ICLR 2018*, https://arxiv.org/pdf/1710.11041.pdf. [74]

Autor, D. (2015), “Why Are There Still So Many Jobs? The History and Future of Workplace Automation”, *Journal of Economic Perspectives*, Vol. 29/3, pp. 3-30, https://doi.org/10.1257/jep.29.3.3. [16]

Autor, D., F. Levy and R. Murnane (2003), "The Skill Content of Recent Technological Change: An Empirical Exploration", *The Quarterly Journal of Economics*, Vol. 118/4, pp. 1279-1333, https://doi.org/10.1162/003355303322552801. [22]

Bapna, A. et al. (2022), "Building machine translation systems for the next thousand languages", *Google Research*, https://arxiv.org/pdf/2205.03983.pdf. [76]

Baumeister, R., M. Muraven and D. Tice (2000), "Ego Depletion: A Resource Model of Volition, Self-Regulation, and Controlled Processing", *Social Cognition*, Vol. 18/2, pp. 130-150, https://doi.org/10.1521/soco.2000.18.2.130. [119]

Baumeister, R., K. Vohs and D. Tice (2007), "The Strength Model of Self-Control", *Current Directions in Psychological Science*, Vol. 16/6, pp. 351-355, https://doi.org/10.1111/j.1467-8721.2007.00534.x. [120]

Bessen, J. (2016), "How Computer Automation Affects Occupations: Technology, Jobs, and Skills", *Boston University School of Law, Law and Economics Research Paper* 15-49, https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1811&context=faculty_scholarship. [18]

Blasi, D. et al. (2022), "Over-reliance on English hinders cognitive science", *Trends in Cognitive Sciences*, Vol. 26/12, pp. 1153-1170, https://doi.org/10.1016/j.tics.2022.09.015. [50]

Boroditsky, L. (2001), "Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time", *Cognitive Psychology*, Vol. 43/1, pp. 1-22, https://doi.org/10.1006/cogp.2001.0748. [38]

Brar, S. et al. (2009), "Perinatal Care for South Asian Immigrant Women and Women Born in Canada: Telephone Survey of Users", *Journal of Obstetrics and Gynaecology Canada*, Vol. 31/8, pp. 708-716, https://doi.org/10.1016/s1701-2163(16)34274-8. [125]

Bromham, L. et al. (2021), "Global predictors of language endangerment and the future of linguistic diversity", *Nature Ecology & Evolution*, Vol. 6/2, pp. 163-173, https://doi.org/10.1038/s41559-021-01604-y. [2]

Brüning, N. and P. Mangeol (2020), "What skills do employers seek in graduates?: Using online job posting data to support policy and practice in higher education", *OECD Education Working Papers*, No. 231, OECD Publishing, Paris, https://doi.org/10.1787/bf533d35-en. [154]

Brzeski, C. and I. Burk (2015), *Die Roboter kommen. Folgen der Automatisierung für den deutschent [The Robots Come. Consequences of Automation for the German Labour Market]*, https://www.erc.de/wp-content/downloads/texte_tools/ING-DiBa_Economic-Research_Die-Roboter-kommen.pdf. [29]

Bureau of Labor Statistics, U.S. Department of Labor (2022), *Occupational Outlook Handbook*, Bureau of Labor Statistics, Department of Labor. [36]

Bureau of Labor Statistics, U.S. Department of Labor (2022), *Occupational Outlook Handbook, Interpreters and Translators*, https://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm (accessed on 21 September 2022). [35]

Cadwell, P., S. O'Brien and C. Teixeira (2017), "Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators", *Perspectives*, Vol. 26/3, pp. 301-321, https://doi.org/10.1080/0907676x.2017.1337210. [107]

Caramancion, K. (2022), "The Role of User's Native Language in Mis/Disinformation Detection: The Case of English", *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, https://doi.org/10.1109/ccwc54503.2022.9720812. [162]

Cassin, B. et al. (eds.) (2014), *Dictionary of Untranslatables*, Princeton University Press, https://doi.org/10.1515/9781400849918. [100]

Caswell, I. and A. Bapna (2022), *Unlocking Zero-Resource Machine Translation to Support New Languages in Google Translate*, Google Research Blog, https://ai.googleblog.com/2022/05/24-new-languages-google-translate.html. [77]

Chen, A., M. Youdelman and J. Brooks (2007), "The Legal Framework for Language Access in Healthcare Settings: Title VI and Beyond", *Journal of General Internal Medicine*, Vol. 22/S2, pp. 362-367, https://doi.org/10.1007/s11606-007-0366-2. [130]

Chen, M. (2013), "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets", *American Economic Review*, Vol. 103/2, pp. 690-731, https://doi.org/10.1257/aer.103.2.690. [42]

Chowdhury, S. et al. (2022), "Growth of non-English-language literature on biodiversity conservation", *Conservation Biology*, Vol. 36/4, https://doi.org/10.1111/cobi.13883. [54]

Cimpian, A. and E. Markman (2011), "The Generic/Nongeneric Distinction Influences How Children Interpret New Information About Social Others", *Child Development*, Vol. 82/2, pp. 471-492, https://doi.org/10.1111/j.1467-8624.2010.01525.x. [45]

Cohen, A. and D. Christakis (2006), "Primary language of parent is associated with disparities in pediatric preventive care", *The Journal of Pediatrics*, Vol. 148/2, pp. 254-258, https://doi.org/10.1016/j.jpeds.2005.10.046. [126]

Costa-jussà, M. and E. al. (2022), "No Language Left Behind: Scaling Human-Centered Machine Translation", https://arxiv.org/abs/2207.04672. [97]

de Moissac, D. and S. Bowen (2017), "Impact of language barriers on access to healthcare for official language minority Francophones in Canada", *Healthcare Management Forum*, Vol. 30/4, pp. 207-212, https://doi.org/10.1177/0840470417706378. [124]

DeepL (n.d.), "How does DeepL work?", https://www.deepl.com/en/blog/how-does-deepl-work (accessed on 8 January 2023). [68]

Department of Computer and Information Science, Linköping University (2017), "Comparing Machine Translation and Human Translation: A Case Study", *Proceedings of the Workshop on Human-Informed Translation and Interpreting Technology*, https://doi.org/10.26615/978-954-452-042-7_003. [138]

Dixon, J., B. Hong and L. Wu (2021), "The Robot Revolution: Managerial and Employment Consequences for Firms", *Management Science*, Vol. 67/9, pp. 5586-5605, https://doi.org/10.1287/mnsc.2020.3812. [12]

Dohan, D. and M. Levintova (2007), "Barriers Beyond Words: Cancer, Culture, and Translation in a Community of Russian Speakers", *Journal of General Internal Medicine*, Vol. 22/S2, https://doi.org/10.1007/s11606-007-0325-y. [133]

Ducar, C. and D. Schocket (2018), "Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate", *Foreign Language Annals*, Vol. 51/4, pp. 779-795, https://doi.org/10.1111/flan.12366. [101]

Eberhard, D., G. Simons and C. Fennig (2022), *Ethnologue: Languages of the World*, https://www.ethnologue.com/ (accessed on 8 January 2023). [6]

ELIS Research (2022), *European Language Industry survey*, https://ec.europa.eu/info/sites/default/files/about_the_european_commission/service_standards_and_principles/documents/elis2022-report.pdf (accessed on 8 January 2023). [110]

Ethnologue - Languages of the World (n.d.), *What are the top 200 most spoken languages?*, https://www.ethnologue.com/guides/ethnologue200 (accessed on 8 January 2023). [5]

European Commission (2022), *European Master's in Translation - Competence Framework 2022*, https://ec.europa.eu/info/sites/default/files/about_the_european_commission/service_standards_and_principles/documents/emt_competence_fwk_2022_en.pdf. [140]

European Commission and CEDEFOP (2021), *Towards a structured and consistent terminology on transversal skills and competences*, https://esco.ec.europa.eu/system/files/2022-05/MSWG%2014-04%20Report%20of%20the%20expert%20group%20on%20transversal%20skills%20and%20competences.pdf. [157]

European Commission, Directorate-General for Employment, Social Affairs and Inclusion (2017), *ESCO handbook: European Skills, Competences*, https://data.europa.eu/doi/10.2767/934956. [158]

Evans, V. (2009), *Language, Cognition and Space: The State of the Art and New Directions*, Advances in Cognitive Linguistics, Equinox. [39]

Fan, A. et al. (2022), "Beyond English-Centric Multilingual Machine Translation", *The Journal of Machine Learning Research*, Vol. 22, pp. 14839–4886, https://dl.acm.org/doi/abs/10.5555/3546258.3546365. [78]

Fan, A. et al. (2021), "Beyond English-Centric Multilingual Machine Translation", *Beyond English-Centric Multilingual Machine Translation*, Vol. 22, pp. 1-48, https://www.jmlr.org/papers/volume22/20-1307/20-1307.pdf. [84]

Forcada, M. (2017), "Making sense of neural machine translation", *Translation Spaces*, Vol. 6/2, pp. 291-309, https://doi.org/10.1075/ts.6.2.06for. [65]

Forcada, M. et al. (2018), "Exploring gap filling as a cheaper alternative to reading comprehension questionnaires when evaluating machine translation for gisting", *Proceedings of the Third Conference on Machine Translation: Research Papers*, https://doi.org/10.18653/v1/w18-6320. [99]

Frey, C. and M. Osborne (2017), "The future of employment: How susceptible are jobs to computerisation?", *Technological Forecasting and Social Change*, Vol. 114, pp. 254-280, https://doi.org/10.1016/j.techfore.2016.08.019. [28]

Fryer, C. et al. (2011), "The effect of limited English proficiency on falls risk and falls prevention after stroke", *Age and Ageing*, Vol. 41/1, pp. 104-107, https://doi.org/10.1093/ageing/afr127. [128]

Fuhrman, O. and L. Boroditsky (2010), “Cross-Cultural Differences in Mental Representations of Time: Evidence From an Implicit Nonlinguistic Task”, *Cognitive Science*, Vol. 34/8, pp. 1430-1451, https://doi.org/10.1111/j.1551-6709.2010.01105.x. [40]

Gaggl, P. and G. Wright (2017), “A Short-Run View of What Computers Do: Evidence from a UK Tax Incentive”, *American Economic Journal: Applied Economics*, Vol. 9/3, pp. 262-294, https://doi.org/10.1257/app.20150411. [20]

Garcia, I. (2011), “Translating by post-editing: is it the way forward?”, *Machine Translation*, Vol. 25/3, pp. 217-237, https://doi.org/10.1007/s10590-011-9115-8. [144]

Gaspari, F., H. Almaghout and S. Doherty (2015), “A survey of machine translation competences: Insights for translation technology educators and practitioners”, *Perspectives*, Vol. 23/3, pp. 333-358, https://doi.org/10.1080/0907676x.2014.979842. [108]

Georgieff, A. and R. Hyee (2021), “Artificial intelligence and employment : New cross-country evidence”*, OECD Social, Employment and Migration Working Papers*, No. 265, OECD Publishing, Paris, https://doi.org/10.1787/c2c1d276-en. [25]

Georgieff, A. and A. Milanez (2021), “What happened to jobs at high risk of automation?”*, OECD Social, Employment and Migration Working Papers*, No. 255, OECD Publishing, Paris, https://doi.org/10.1787/10bc97f4-en. [17]

Gereme, F. et al. (2021), “Combating Fake News in “Low-Resource” Languages: Amharic Fake News Detection Accompanied by Resource Crafting”, *Information*, Vol. 12/1, p. 20, https://doi.org/10.3390/info12010020. [95]

Gilles Adda, K. (ed.) (2019), *Designing for language revitalisation*, European Language Resources Association (ELRA), https://researchers.cdu.edu.au/en/publications/designing-for-language-revitalisation. [93]

Goyal, N. et al. (2022), “The Flores-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation”, *Transactions of the Association for Computational Linguistics*, Vol. 10, pp. 522-538, https://doi.org/10.1162/tacl_a_00474. [90]

Guerberof Arenas, A. and J. Moorkens (2019), “Machine translation and post-editing training as part of a master’s programme”, *The Journal of Specialised Translation* 31, https://www.jostrans.org/issue31/art_guerberof.pdf. [147]

Hachette UK, 2. (ed.) (2021), *The Dawn of Language: The story of how we came to talk*. [37]

Haddow, B. et al. (2022), “Survey of Low-Resource Machine Translation”, *Computational Linguistics*, Vol. 48/3, pp. 673-732, https://doi.org/10.1162/coli_a_00446. [79]

Hossain, M. et al. (2020), “BanFakeNews: A Dataset for Detecting Fake News in Bangla”, https://arxiv.org/abs/2004.08789. [96]

Hull, M. (2016), “Medical language proficiency: A discussion of interprofessional language competencies and potential for patient risk”, *International Journal of Nursing Studies*, Vol. 54, pp. 158-172, https://doi.org/10.1016/j.ijnurstu.2015.02.015. [134]

Hunter, N., M. North and R. Slotow (2021), “The marginalisation of voice in the fight against climate change: The case of Lusophone Africa”, *Environmental Science & Policy*, Vol. 120, pp. 213-221, https://doi.org/10.1016/j.envsci.2021.03.012. [52]

Hutchins, J. (2006), “Example-based machine translation: a review and commentary”, *Machine Translation*, Vol. 19/3-4, pp. 197-211, https://doi.org/10.1007/s10590-006-9003-9. [64]

IBM (2021), *Supervised vs. Unsupervised Learning: What’s the Difference?*, https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning (accessed on 13 December 2022). [71]

Ikenaga, T. and R. Kambayashi (2016), “Task Polarization in the Japanese Labor Market: Evidence of a Long-Term Trend”, *Industrial Relations: A Journal of Economy and Society*, Vol. 55/2, pp. 267-293, https://doi.org/10.1111/irel.12138. [23]

intento (2022), *The State of Machine Translation 2022*, https://inten.to/machine-translation-report-2022/ (accessed on 8 September 2022). [7]

International Organization of Standards (2017), *Translation services — Post-editing of machine translation output — Requirements*, https://www.iso.org/standard/62970.html (accessed on 25 October 2022). [146]

Jacobs, E. et al. (2006), “The Need for More Research on Language Barriers in Health Care: A Proposed Research Agenda”, *The Milbank Quarterly*, Vol. 84/1, pp. 111-133, https://doi.org/10.1111/j.1468-0009.2006.00440.x. [121]

Jennions, M. (ed.) (2021), “Tapping into non-English-language science for the conservation of global biodiversity”, *PLOS Biology*, Vol. 19/10, p. e3001296, https://doi.org/10.1371/journal.pbio.3001296. [59]

Joshi, P. et al. (2021), “The State and Fate of Linguistic Diversity and Inclusion in the NLP World”, *Computation and Language*, https://doi.org/10.48550/arXiv.2004.09095. [3]

Karakanta, A., J. Dehdari and J. van Genabith (2017), “Neural machine translation for low-resource languages without parallel corpora”, *Machine Translation*, Vol. 32/1-2, pp. 167-189, https://doi.org/10.1007/s10590-017-9203-5. [88]

Karliner, L. et al. (2007), “Do Professional Interpreters Improve Clinical Care for Patients with Limited English Proficiency? A Systematic Review of the Literature”, *Health Services Research*, Vol. 42/2, pp. 727-754, https://doi.org/10.1111/j.1475-6773.2006.00629.x. [131]

Khelifa, R., T. Amano and M. Nuñez (2022), “A solution for breaking the language barrier”, *Trends in Ecology & Evolution*, Vol. 37/2, pp. 109-112, https://doi.org/10.1016/j.tree.2021.11.003. [61]

Khoong, E. and J. Rodriguez (2022), “A Research Agenda for Using Machine Translation in Clinical Medicine”, *Journal of General Internal Medicine*, Vol. 37/5, pp. 1275-1277, https://doi.org/10.1007/s11606-021-07164-y. [137]

Koch, M., I. Manuylov and M. Smolka (2021), “Robots and Firms”, *The Economic Journal*, Vol. 131/638, pp. 2553-2584, https://doi.org/10.1093/ej/ueab009. [13]

Koehn, P. (2009), *Statistical Machine Translation*, Cambridge University Press, https://doi.org/10.1017/cbo9780511815829. [160]

Koponen, M. (2015), “How to teach machine translation post-editing? Experiences from a post-editing course”, *Conference: Proceedings of 4th Workshop on Post-Editing Technology and Practice (WPTP4)*. [150]

Koponen, M., L. Salmi and M. Nikulin (2019), "A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output", *Machine Translation*, Vol. 33/1-2, pp. 61-90, https://doi.org/10.1007/s10590-019-09228-7. [66]

Kreutzer, J. et al. (2022), "Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets", *Transactions of the Association for Computational Linguistics*, Vol. 10, pp. 50-72, https://doi.org/10.1162/tacl_a_00447. [89]

Kusters, R. et al. (2020), "Interdisciplinary Research in Artificial Intelligence: Challenges and Opportunities", *Frontiers in Big Data*, Vol. 3, https://doi.org/10.3389/fdata.2020.577974. [98]

Kuwanto, G. et al. (2021), "Low-Resource Machine Translation Training Curriculum Fit for Low-Resource Languages", https://arxiv.org/abs/2103.13272. [80]

Lane, M. and A. Saint-Martin (2021), "The impact of Artificial Intelligence on the labour market: What do we know so far?", *OECD Social, Employment and Migration Working Papers*, No. 256, OECD Publishing, Paris, https://doi.org/10.1787/7c895724-en. [32]

Lassébie, J. et al. (2021), "Speaking the same language: A machine learning approach to classify skills in Burning Glass Technologies data", *OECD Social, Employment and Migration Working Papers*, No. 263, OECD Publishing, Paris, https://doi.org/10.1787/adb03746-en. [156]

Lassébie, J. and G. Quintini (2022), "What skills and abilities can automation technologies replicate and what does it mean for workers?: New evidence", *OECD Social, Employment and Migration Working Papers*, No. 282, OECD Publishing, Paris, https://doi.org/10.1787/646aad77-en. [30]

Lee, K. and M. Qian (2022), "Misinformation in Machine Translation: Error Categories and Levels of Recognition Difficulty", in *Artificial Intelligence in HCI, Lecture Notes in Computer Science*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-031-05643-7_34. [9]

Lee, S. (2019), "The impact of using machine translation on EFL students' writing", *Computer Assisted Language Learning*, Vol. 33/3, pp. 157-175, https://doi.org/10.1080/09588221.2018.1553186. [91]

Le, Q. and M. Schuster (n.d.), "A Neural Network for Machine Translation, at Production Scale", *AI Google Blog*, https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html (accessed on 8 January 2023). [69]

Lewis, M. and G. Lupyan (2020), "Gender stereotypes are reflected in the distributional structure of 25 languages", *Nature Human Behaviour*, Vol. 4/10, pp. 1021-1028, https://doi.org/10.1038/s41562-020-0918-6. [46]

Lopez, A. (2008), "Statistical machine translation", *ACM Computing Surveys*, Vol. 40/3, pp. 1-49, https://doi.org/10.1145/1380584.1380586. [63]

Lynch, A. et al. (2021), "Culturally diverse expert teams have yet to bring comprehensive linguistic diversity to intergovernmental ecosystem assessments", *One Earth*, Vol. 4/2, pp. 269-278, https://doi.org/10.1016/j.oneear.2021.01.002. [53]

Machine Translate (2022), *Rule-based machine translation*, https://machinetranslate.org/rule-based-machine-translation (accessed on 14 December 2022). [72]

Macken, L., D. Prou and A. Tezcan (2020), "Quantifying the Effect of Machine Translation in a High-Quality Human Translation Production Process", *Informatics*, Vol. 7/2, p. 12, https://doi.org/10.3390/informatics7020012. [106]

Mann, K. and L. Püttmann (2021), "Benign Effects of Automation: New Evidence from Patent Texts", *The Review of Economics and Statistics*, pp. 1-45, https://doi.org/10.1162/rest_a_01083. [19]

McDonald, J. and S. Kennedy (2007), "Cervical Cancer Screening by Immigrant and Minority Women in Canada", *Journal of Immigrant and Minority Health*, Vol. 9/4, pp. 323-334, https://doi.org/10.1007/s10903-007-9046-x. [127]

Messias, D., L. McDowell and R. Estrada (2009), "Language Interpreting as Social Justice Work", *Advances in Nursing Science*, Vol. 32/2, pp. 128-143, https://doi.org/10.1097/ans.0b013e3181a3af97. [135]

Miles, L. et al. (2011), "Can a mind have two time lines? Exploring space–time mapping in Mandarin and English speakers", *Psychonomic Bulletin Review*, Vol. 18/3, pp. 598-604, https://doi.org/10.3758/s13423-011-0068-y. [41]

Moorkens, J. (2017), "Under pressure: translation in times of austerity", *Perspectives*, Vol. 25/3, pp. 464-477, https://doi.org/10.1080/0907676x.2017.1285331. [114]

Muda, R. et al. (2021), *People are worse at detecting fake news in their foreign language*, Center for Open Science, https://doi.org/10.31219/osf.io/p8su6. [163]

National Accreditation Authority for Translators and Interpreters (n.d.), *Become Certified*, https://www.naati.com.au/become-certified/ (accessed on 3 February 2023). [153]

National Council on Interpreting in Health Care (n.d.), *Ethics and standards of practice*, https://www.ncihc.org/ethics-and-standards-of-practice. [136]

Nedelkoska, L. and G. Quintini (2018), "Automation, skills use and training", *OECD Social, Employment and Migration Working Papers*, No. 202, OECD Publishing, Paris, https://doi.org/10.1787/2e2f4eea-en. [31]

Nekoto, W. et al. (2020), "Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages", *Findings of the Association for Computational Linguistics: EMNLP 2020*, https://doi.org/10.18653/v1/2020.findings-emnlp.195. [81]

Nuñez, M. et al. (2021), "Making ecology really global", *Trends in Ecology & Evolution*, Vol. 36/9, pp. 766-769, https://doi.org/10.1016/j.tree.2021.06.004. [55]

O'Brien, S. (2002), "Teaching Post-editing: A Proposal for Course Content", *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*, https://aclanthology.org/2002.eamt-1.11/. [149]

O'Brien, S. et al. (2014), *Post-Editing of Machine Translation: Processes and Applications*, Cambridge Scholars Publishing, New-castle upon Tyne. [143]

O'Brien, S. (2002), "Teaching post-editing: a proposal for course content", *European Association for Machine Translation Conferences/Workshops*, https://aclanthology.org/2002.eamt-1.11.pdf. [111]

Occupational Information Network (n.d.), , https://www.onetonline.org/. [33]

OECD (2019), *OECD Employment Outlook 2019: The Future of Work*, OECD Publishing, Paris, https://doi.org/10.1787/9ee00155-en. [15]

OECD (2013), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris, https://doi.org/10.1787/9789264204256-en. [21]

Ohtani, A. et al. (2015), "Language Barriers and Access to Psychiatric Care: A Systematic Review", *Psychiatric Services*, Vol. 66/8, pp. 798-805, https://doi.org/10.1176/appi.ps.201400351. [122]

Omniscien Technologies (2022), *What is Neural Machine Translation (NMT)?*, https://omniscien.com/faq/what-is-neural-machine-translation/ (accessed on 13 December 2022). [75]

Omniscien Technologies (2022), *What is Rules-Based Machine Translation (RBMT)?*, https://omniscien.com/faq/what-is-rules-based-machine-translation/ (accessed on 13 December 2022). [73]

Orife, I. et al. (2020), "Masakhane -- Machine Translation For Africa", https://arxiv.org/abs/2003.11529. [82]

Pajarinen, M., P. Rouvinen and A. Ekeland (2015), *Computerization Threatens One-Third of Finnish and Norwegian Employment*, https://www.etla.fi/wp-content/uploads/ETLA-Muistio-Brief-34.pdf. [27]

Plitt, M. and F. Masselot (2010), "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context", *The Prague Bulletin of Mathematical Linguistics*, Vol. 93/1, https://doi.org/10.2478/v10108-010-0010-x. [109]

Politzer-Ahles, S., T. Girolamo and S. Ghali (2020), "Preliminary evidence of linguistic bias in academic reviewing", *Journal of English for Academic Purposes*, Vol. 47, p. 100895, https://doi.org/10.1016/j.jeap.2020.100895. [51]

Popel, M. et al. (2020), "Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals", *Nature Communications*, Vol. 11/1, https://doi.org/10.1038/s41467-020-18073-9. [115]

Potowski, K. (2013), "Language Maintenance and Shift", in *The Oxford Handbook of Sociolinguistics*, Oxford University Press, https://doi.org/10.1093/oxfordhb/9780199744084.013.0016. [4]

Pottier, P. et al. (2022), "A comprehensive database of amphibian heat tolerance", *Scientific Data*, Vol. 9/1, https://doi.org/10.1038/s41597-022-01704-9. [57]

Pym, A. (2014), "Translation Skill-Sets in a Machine-Translation Age", *Meta*, Vol. 58/3, pp. 487-503, https://doi.org/10.7202/1025047ar. [142]

Pym, A. and E. Torres-Simón (2021), "Is automation changing the translation profession?", *International Journal of the Sociology of Language*, Vol. 2021/270, pp. 39-57, https://doi.org/10.1515/ijsl-2020-0015. [112]

Rhodes, M. et al. (2019), "Subtle Linguistic Cues Increase Girls' Engagement in Science", *Psychological Science*, Vol. 30/3, pp. 455-466, https://doi.org/10.1177/0956797618823670. [44]

Rico, C. and E. Torrejón (2012), "Skills and Profile of the New Role of the Translator as MT Post-editor", *https://redib.org/Record/oai_revista355-revista-tradum%C3%A0tica-tecnologies-de-la-traducci%C3%B3*, https://doi.org/10.5565/rev/tradumatica.18. [141]

Romanach, S. (ed.) (2022), "Language barriers in global bird conservation", *PLOS ONE*, Vol. 17/4, p. e0267151, https://doi.org/10.1371/journal.pone.0267151. [58]

Rossi, C. and J. Chevrot (2019), "Uses and perceptions of machine translation at the European Commission", *The Journal of Specialised Translation*, https://jostrans.org/issue31/art_rossi.pdf. [34]

RWS (2022), *New certification launches to help linguists and project managers become machine translation post-editing experts*, https://www.rws.com/about/product-news/2022/mt-certification/ (accessed on 3 February 2023). [151]

Saghayan, M., S. Ebrahimi and M. Bahrani (2021), "Exploring the Impact of Machine Translation on Fake News Detection: A Case Study on Persian Tweets about COVID-19", *2021 29th Iranian Conference on Electrical Engineering (ICEE)*, https://doi.org/10.1109/icee52715.2021.9544409. [10]

Sallabank, J. (2013), *Attitudes to Endangered Languages*, Cambridge University Press, https://doi.org/10.1017/cbo9781139344166. [92]

Samek, L., M. Squicciarini and E. Cammeraat (2021), "The human capital behind AI: Jobs and skills demand from online job postings", *OECD Science, Technology and Industry Policy Papers*, No. 120, OECD Publishing, Paris, https://doi.org/10.1787/2e278150-en. [155]

Sandi, C. (2013), "Stress and cognition", *WIREs Cognitive Science*, Vol. 4/3, pp. 245-261, https://doi.org/10.1002/wcs.1222. [118]

Savoldi, B. et al. (2021), "Gender Bias in Machine Translation", *Transactions of the Association for Computational Linguistics*, Vol. 9, pp. 845-874, https://doi.org/10.1162/tacl_a_00401. [105]

Schwei, R. et al. (2016), "Changes in research on language barriers in health care since 2003: A cross-sectional review study", *International Journal of Nursing Studies*, Vol. 54, pp. 36-44, https://doi.org/10.1016/j.ijnurstu.2015.03.001. [123]

Smith, G. and R. Ishita (2020), *Mitigating Bias*, https://haas.berkeley.edu/wp-content/uploads/UCB_Playbook_R10_V2_spreads2.pdf. [104]

Spitz-Oener, A. (2006), "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure", *Journal of Labor Economics*, Vol. 24/2, pp. 235-270, https://doi.org/10.1086/499972. [24]

Stahlberg, F. (2020), "Neural Machine Translation: A Review", *Journal of Artificial Intelligence Research*, Vol. 69, pp. 343-418, https://doi.org/10.1613/jair.1.12007. [67]

Stephens, E. (2022), "The mechanical Turk: a short history of 'artificial artificial intelligence'", *Cultural Studies*, pp. 1-23, https://doi.org/10.1080/09502386.2022.2042580. [161]

Sutter, M. et al. (2018), "Language group differences in time preferences: Evidence from primary school children in a bilingual city", *European Economic Review*, Vol. 106, pp. 21-34, https://doi.org/10.1016/j.euroecorev.2018.04.003. [43]

Taira, B. et al. (2021), "A Pragmatic Assessment of Google Translate for Emergency Department Instructions", *Journal of General Internal Medicine*, Vol. 36/11, pp. 3361-3365, https://doi.org/10.1007/s11606-021-06666-z. [132]

Teitelbaum, J., L. Cartwright-Smith and S. Rosenbaum (2012), "Translating Rights into Access: Language Access and the Affordable Care Act", *American Journal of Law & Medicine*, Vol. 38/2-3, pp. 348-373, https://doi.org/10.1177/009885881203800205. [129]

The Economist (2017), *Why translators have the blues*, https://www.economist.com/books-and-arts/2017/05/27/why-translators-have-the-blues. [139]

Tran, C. et al. (2021), "Facebook AI WMT21 News Translation Task Submission", https://arxiv.org/abs/2108.03265. [87]

Transcending language barriers to environmental sciences (n.d.), *https://translatesciences.com/*, https://translatesciences.com/ (accessed on 13 December 2022). [56]

Treré, E. (2016), "The Dark Side of Digital Politics: Understanding the Algorithmic Manufacturing of Consent and the Hindering of Online Dissidence", *IDS Bulletin*, Vol. 41/1, pp. 127-138, https://doi.org/10.19088/1968-2016.111. [94]

UNESCO (2019), *Fourth Consolidated Report on the Implementation by Member States of the 2003 Recommendation Concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*, https://unesdoc.unesco.org/ark:/48223/pf0000370315/PDF/370315eng.pdf.multi. [49]

UNESCO (2010), *Atlas of the World's Languages Languages in Danger*, https://unesdoc.unesco.org/ark:/48223/pf0000187026. [47]

UNESCO (2003), *Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*, https://www.unesco.org/en/legal-affairs/recommendation-concerning-promotion-and-use-multilingualism-and-universal-access-cyberspace?hub=66535. [48]

Vertovec, S. (2007), "Super-diversity and its implications", *Ethnic and Racial Studies*, Vol. 30/6, pp. 1024-1054, https://doi.org/10.1080/01419870701599465. [116]

Vieira, L. (2018), "Automation anxiety and translators", *Translation Studies*, Vol. 13/1, pp. 1-21, https://doi.org/10.1080/14781700.2018.1543613. [113]

Vieira, L., E. Alonso and L. Bywood (2019), "Post-Editing in Practice: Process, Product and Networks", *The Journal of Specialised Translation*, https://jostrans.org/issue31/art_introduction.pdf. [145]

Vieira, L., M. O'Hagan and C. O'Sullivan (2020), "Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases", *Information, Communication & Society*, pp. 1-18, https://doi.org/10.1080/1369118x.2020.1776370. [117]

Violante, G. (2008), "Skill-Biased Technical Change", in *The New Palgrave Dictionary of Economics*, Palgrave Macmillan UK, London, https://doi.org/10.1057/978-1-349-95121-5_2388-1. [11]

Wallace, B. and L. Kertz (2014), *Can Cognitive Scientists Help Computers Recognize Irony?*, https://cogsci.mindmodeling.org/2014/papers/005/paper005.pdf. [102]

Wang, J. et al. (2021), *Putting words into the system's mouth: A targeted attack on neural machine translation using monolingual data poisoning*, https://arxiv.org/pdf/2107.05243.pdf. [8]

Wu, Y. (2016), "Google's neural machine translation system: Bridging the gap between human and machine translation", *arXiv preprint arXiv*, https://arxiv.org/abs/1609.08144. [70]

Zhang, B. et al. (2020), "Improving Massively Multilingual Neural Machine Translation and Zero-Shot Translation", *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 1628–1639, https://aclanthology.org/2020.acl-main.148/. [85]

Zoph, B. et al. (2016), "Transfer Learning for Low-Resource Neural Machine Translation", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, https://doi.org/10.18653/v1/d16-1163. [86]

Annex A. Supplementary data

Table A.1. Categorisation of skills groups of language professionals

Categorisation of Skills Groups		
Knowledge Skills	Transversal skills	Digital Skills
Biology	Active Listening	Computer Programming
Building And Construction	Adaptability/Resilience	Digital Content Creation
Chemistry	Administration and Management	Digital Data Processing
Customer And Personal Service	Auditory and Speech Abilities	Ict Safety, Networks and Servers
Design	Clerical	Office Tools and Collaboration Software
Engineering, Mechanics and Technology	Communications and Media	Web Development and Cloud Technologies
Equipment Selection	Co-ordination	Writing
Fine Arts	Installation and Maintenance	
Food Production	Motivation/Commitment	
*Languages	Originality	
Geography	Persuasion and Negotiation	
History and Archaeology	Public Safety and Security	
Industry Knowledge	Quality Control Analysis	
Judgment and Decision Making	Reading Comprehension	
Law and Government	Reasoning and Problem-Solving	
Learning	Self-Management/Rigour	
Management Of Financial Resources	Social Perceptiveness	
Management Of Material Resources	Speaking	
Management Of Personnel Resources	Telecommunications	
Medicine and Dentistry	Time Management	
Philosophy and Theology	Visual Abilities	
Physical Abilities	Work Ethics	
Physics		
Production and Processing		
Psychology, Therapy, Counselling		
Psychomotor Abilities		
Quantitative Abilities		
Sales and Marketing		
Sociology and Anthropology		
Training and Education		
Transportation		

* While we rely on the categorisation developed by Lassébie et al., (2021[156]), we combine the categories "Local language" and "Foreign language" into the category "Languages".

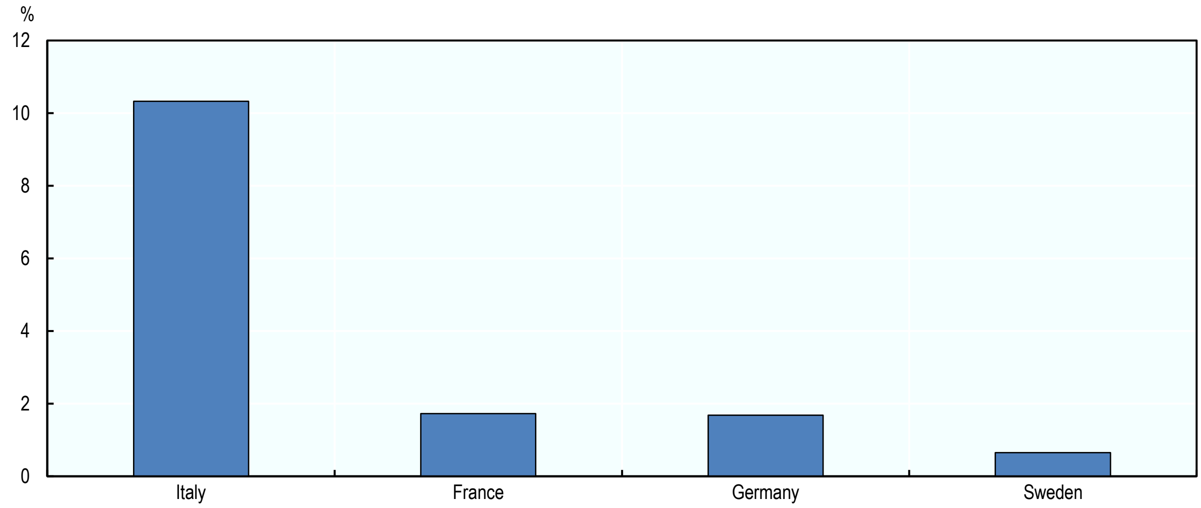
Table A.2. AI-core skills for selected EU countries

AI-core skills for selected EU countries		
analyse des sentiments	intelligent agent	robot operating system
apprentissage automatique	intelligenza artificiale	robot operating system (ros)
apprentissage supervisé	keras	robot programming
artificial intelligence	künstliche intelligenz	roboterprogrammierung
artificiell intelligens	lexical semantics	robotic systems
automatic speech recognition	lidar	scikit-learn
automatic speech recognition (asr)	linguistica computazionale	sémantique lexicale
autonome systeme	linguistique informatique	semi-supervised learning
autonomous systems	machine learning	sentiment analysis
beslutsstödsystem	machine translation	servoantriebe
bildbehandling	machine translation (mt)	servomoteur
bildverarbeitung	machine vision	simultaneous localization and mapping
chat bot	mahout	simultaneous localization and mapping (slam)
chatbot	maschinelle übersetzung	speech recognition
clustering algorithms	maschinelles lernen	spracherkennung
computational linguistics	maskininlärning	språkteknologi
computer vision	maskinöversättning	supervised learning
computerlinguistik	moses	système expert
convolutional neural network	mxnet	télédétection
datorlingvistik	natural language processing	telerilevamento
datorseende	natürliche sprachverarbeitung	tensorflow
decision trees	neural networks	text mining
deep learning	neurala nätverk	text to speech
digital agent	neuralt nätverk	text to speech (tts)
digital assistant	object recognition	tokenization
ESCOv1_12473	opencv	torch
expert system	opennlp	traduction automatique
fernerkundung	path planning	traduzione automatica
fjärranalys	pytorch	traitement automatique des langues
forêts aléatoires	random forest	traitement automatique du langage naturel
fouille de textes	recommender systems	unsupervised learning
gradient boosting	recurrent neural network	virtual agents
ibm watson	remote sensing	vision industrielle
image processing	réseau de neurones	vision par ordinateur
image recognition	réseaux de neurones	
intelligence artificielle	riconoscimento vocale	

Source: Skills extraction based on Lightcast ™.

Figure A.1. Share of postings explicitly available to freelance professionals (2015-2019)

Percentage of online job postings explicitly available to freelance professionals



Note: The figure shows the average percentage of online job postings explicitly available for freelance professionals for selected European Union countries: Germany, Italy, France and Sweden over the 2015-2019 period.
Source: Authors' own compilation based on Lightcast™ (December 2022).