

An Interdisciplinary Approach to Human-Centered Machine Translation

Marine Carpuat¹, Omri Asscher², Kalika Bali³, Luisa Bentivogli⁴, Frédéric Blain⁵, Lynne Bowker⁶, Monojit Choudhury⁷, Hal Daumé III¹, Kevin Duh⁸, Ge Gao¹, Alvin Grissom II⁹, Marzena Karpinska³, Elaine C. Khoong¹⁰, William D. Lewis¹¹, André F. T. Martins¹², Mary Nurminen¹³, Douglas W. Oard¹, Maja Popovic¹⁴, Michel Simard¹⁵, François Yvon¹⁶

¹University of Maryland, ²Bar-Ilan University, ³Microsoft, ⁴Fondazione Bruno Kessler, ⁵Tilburg University, ⁶Université Laval,

⁷Mohamed bin Zayed University of Artificial Intelligence, ⁸Johns Hopkins University, ⁹Haverford College,

¹⁰University of California, San Francisco, ¹¹University of Washington, ¹²Instituto Superior Técnico, Universidade de Lisboa

¹³Tampere University, ¹⁴Dublin City University & IU University, ¹⁵National Research Council Canada ¹⁶Sorbonne-Université & CNRS

Correspondence: marine@umd.edu

Abstract

Machine Translation (MT) tools are widely used today, often in contexts where professional translators are not present. Despite progress in MT technology, a gap persists between system development and real-world usage, particularly for non-expert users who may struggle to assess translation reliability. This paper advocates for a human-centered approach to MT, emphasizing the alignment of system design with diverse communicative goals and contexts of use. We survey the literature in Translation Studies and Human-Computer Interaction to recontextualize MT evaluation and design to address the diverse real-world scenarios in which MT is used today.

1 Introduction

Machine Translation (MT) is one of the few NLP technologies that has been widely available online for decades. As both translation quality and internet access have improved (Gaspari and Hutchins, 2007), MT has gained a large and diverse user base. Millions of people use it to communicate across languages, including in settings where professional translators or interpreters are not realistically available (Nurminen and Papula, 2018; Kasperé et al., 2021; Vieira et al., 2022; Kenny et al., 2022).

As MT becomes increasingly embedded in everyday tools and tasks, the socio-technical gap between how the technology is developed and how it is used in real-world contexts is widening (Ackerman, 2000). Whereas initial MT systems were primarily used to support professional translators or narrow domains (Hutchins, 2001), today MT can be used by anyone with internet access in their daily life (Yvon, 2019; Kenny et al., 2022). However, MT does not yet fulfill its promise to enable communication across languages, particularly for users who may lack the language or domain expertise needed to make informed use of the translations (Liebling et al., 2020; Santy et al., 2021;

Valdez et al., 2023). This gap is further amplified by the rise of translation with general-purpose large language models (LLMs) (Vilar et al., 2023; Alves et al., 2024; Kocmi et al., 2024; Hendy et al., 2023). With such tools, translation can be integrated into broader workflows, where translation might be covert, making it even harder for users to assess its reliability. This can result in over-trust in MT (Martindale and Carpuat, 2018), which is particularly problematic in high-stakes scenarios where it can cause harm (Vieira et al., 2021), but also in under-use of MT tools in cases where they could be beneficial (O’Brien and Federici, 2019).

We argue that a human-centered approach to MT is needed: one that broadens what MT systems do to help users weigh risks and benefits and align system design with communicative goals. This approach echoes calls for human-centered AI (Capel and Brereton, 2023), which includes recognizing that people are at the heart of the development of any AI system (Vaughan and Wallach, 2021), emphasizing designing AI systems that augment rather than replace human capabilities, prioritizing human agency and system accountability (Shneiderman, 2022), and using human-centered design methods for AI systems (Chancellor, 2023).

To provide a foundation for human-centered MT, we argue that it is important to adopt an interdisciplinary approach that includes Translation Studies and Human-Computer Interaction (HCI). In this paper, we recontextualize MT research by surveying relevant literature in these fields. As Green et al. (2015) point out, the question of how to design effective human–MT interaction has been considered long before HCI, NLP, or AI were formalized disciplines. For example, Kay (1980/1997) introduced a cooperative interactive system as an alternative to fully automated translation to replace professional translators. As MT improved, these questions were revisited to design mixed-initiative post-editing interfaces (Green et al., 2013; Koehn

et al., 2014; Briva-Iglesias et al., 2023), highlighting the benefits of designing MT systems to augment, rather than replace, professional translators' abilities (O'Brien, 2024). As the MT user base has expanded from professional translators to professionals in other disciplines, as well as the general public (Savoldi et al., 2025), many relevant lessons can be drawn from theoretical and empirical work in Translation Studies and HCI. Accordingly, this survey results from discussions between co-authors across these fields. Translation studies and HCI experts identified key insights they wished to share with the MT researchers. These insights served as points of connection with the MT literature.

Considering MT's diverse uses (Section 2), we synthesize cross-disciplinary insights spanning MT literacy (Section 3), human-MT interaction (Section 4), and translation ethics (Section 5). We then outline research directions for human-centered MT evaluation (Section 6) and design (Section 7), illustrating interdisciplinary human-centered MT research with a healthcare case study (Section 8).

2 Understanding Contexts of Use

To develop human-centered MT, we must first understand how MT is used in the real world. While the body of research on users, contexts, and purposes has shown increased growth recently, the considerable size of the user population, estimated in 2021 at more than one billion (Nurminen, 2021a, p. 23), and growing variety of use contexts present a challenge for synthesizing that research into knowledge that can be used for designing systems that more directly serve user needs.

A classical framework distinguishes three use types (Hovy et al., 2002): assimilation, in which MT helps users get the gist of content in a foreign language (e.g., browsing news, triage) without requiring perfect quality; dissemination, in which MT content is shared with others, demanding higher quality (e.g., public announcements); and communication, in which MT supports live or interactive multilingual exchanges (e.g., chat, classrooms).

A wealth of MT research projects have considered different use cases over the years, but without much information sharing across settings: classroom speech translation (Lewis and Niehues, 2023), healthcare (Khoong et al., 2019; Valdez and Guerberof-Arenas, 2025), crisis response (Lewis et al., 2011; Escartín and Moniz, 2019), international patent processes (Nurminen, 2020), migra-

tion scenarios (Vollmer, 2020; Vieira, 2024; Pięta and Valdez, 2024), research and academic writing (Bowker and Ciro, 2019b; Ehrensberger-Dow et al., 2023; Bawden et al., 2024), customer support (Gonçalves et al., 2022), literary MT (Karpinska and Iyyer, 2023; Zhang et al., 2025a), and CAT/localization (Koehn et al., 2014; Lin et al., 2010), and intercultural collaboration platforms (Ishida, 2016). The examination of these contexts of use alone suggests some considerations that should impact MT design, beyond the general purpose of translation: risk management (error tolerance varies by domain), synchrony (real-time vs. delayed), urgency, shelf life, audience, interaction dynamics, modality/accessibility, and overtess of MT use (e.g., covert use of MT on a multilingual website or embedded in another application).

We also lack a deeper understanding of who uses online MT tools and how. Nurminen (2021a) estimates that 99.97% of MT users are not professional translators. "Machine Translation Stories" illustrate diverse uses by individuals from all walks of life, from music students translating old Italian arias to people using MT in their professional life (Nurminen, 2021b). A survey of 1,200 UK residents shows high satisfaction with MT for low-stakes uses but highlights a demand for better quality (Vieira et al., 2022). Another survey of 2,520 UK public service professionals reveals that 33% had used MT in their work, predominantly within health and social care sectors, but also across legal, emergency, and police services (Nunes Vieira, 2024). Formal training was uncommon, leading many professionals to rely on personal devices and publicly available tools like Google Translate and ChatGPT. But user needs are not met equally across socioeconomic and geographic contexts. For instance, interview studies showed that MT applications do not support effective cross-lingual communication for migrant workers in India and immigrant populations in the U.S., resulting in significant negative impacts on their daily lives (Liebling et al., 2020).

Human-centered MT should not just respond to user needs (Gasson, 2003), but consider more broadly how people are affected by MT, including the languages and perspectives of marginalized populations (Bender and Grissom, 2024), and considering both direct and indirect stakeholders (Friedman and Hendry, 2019, p. 39). These include recipients of translated content, institutions using MT at scale (Koponen and Nurminen, 2024),

writers of source texts (Taivalkoski-Shilov, 2019; Lacruz Mantecón, 2023), MT practitioners (Robertson et al., 2023), language learners, and broader language communities given evidence that language evolves through automation (Guo et al., 2024).

This complexity calls for further investigation of MT in context and for organizing use cases into a taxonomy that balances general-purpose development with contextual needs.

3 Machine Translation Literacy

Translation Studies research highlights a need for promoting machine translation literacy (Bowker and Ciro, 2019b) given the wide gap between how translation is approached by people within versus beyond the language professions. Professional translators have been trained in translation, which usually also involves acquiring a domain specialization (Scarpa, 2020), such as legal, medical or technical translation. As people, professional translators also have deep knowledge of the language pair in question, and the type of real-world knowledge and cultural knowledge that is necessary when translating between languages and cultures. Translators can bring all this information to bear on their understanding of the source text. They compensate for shortcomings in the source text (e.g., they can clarify the intended meaning of a sentence with poor punctuation or where a homophone has erroneously been used). Professional translators also operate within a sort of decision-making framework because they request (or even require) a translation brief from their client or employer (Munday et al., 2022). The translation brief is essentially a set of instructions and information that helps the translator to make sensible choices. For instance, the brief contains information about the intended purpose of the translation, where it will be published, who will read or use it, what the target reader's background (language variety, culture, education level) is. All of this information allows the translator to make informed decisions.

In contrast many MT users have no background in translation. They may not have the necessary linguistic knowledge, domain or cultural knowledge required to evaluate the adequacy of the translated text. They may have misconceptions about translation (Bowker, 2023), e.g., seeing it as an exact science or a task that can be done by any bilingual. They might not realize the importance of the translation brief. In short, they lack MT literacy, which

has been defined as “knowing how MT works, how the technology can be useful in a particular context, and what the implications are of using it for various purposes” (O’Brien and Ehrensberger-Dow, 2020).

This highlights the necessity of MT literacy and motivates a key direction in Human-Centered MT: designing tools that promote informed and responsible use, especially by lay users. Current tools lack this, but we will see that the existing literature provides a starting point.

4 Empirical Studies of MT Outside Professional Translation

Translation Studies and HCI offer extensive empirical research on human-MT interaction within various contexts, beyond professional translation. It reveals existing user strategies for using potentially imperfect MT, interventions that have already shown promise, and open research directions.

Post-editing The most studied human-MT interaction setting is probably post-editing, where people edit raw MT to improve it. It has received significant attention in the context of professional translation (Cadwell et al., 2016; Briva-Iglesias et al., 2023, among others), but it is also performed by other users, for instance when they translate their own source text as a writing aid in academic settings (Bowker, 2020a; Xu et al., 2024; O’Brien et al., 2018) or for scientific dissemination (Bawden et al., 2024). There is evidence that even monolingual users can interpret and revise MT output when provided with background knowledge or translation options (Hu et al., 2010; Koehn, 2010).

When users do not understand the target language, post-editing is not an option, but they still face a decision about whether to publish or share the raw MT outputs. Zouhar et al. (2021) studies the impact of augmenting raw MT with backtranslation, source paraphrasings and quality estimation feedback in such “outbound translation” settings, and show that backtranslation feedback increases user confidence in the produced translation, but not the actual quality of the text produced.

Augmented Outputs for Gisting Several studies show that augmenting MT outputs can improve comprehension and engagement, particularly when MT is used for understanding the gist of a text. Highlighting key words in source and target texts can improve people’s ability to understand difficult translations (Pan and Wang, 2014; Grissom et al.,

2024), and adding emotional and contextual cues promotes engagement with social media posts in a foreign language (Lim et al., 2018).

Research has also shown that users sometimes access outputs from multiple MT tools to better understand the errors associated with each individual output and, in doing so, enhance overall comprehension (Anazawa et al., 2013; Nurminen, 2019; Robertson et al., 2021). Other research has also indicated positive effects from exposure to outputs from multiple MT tools (Xu et al., 2014; Gao et al., 2015). Human-centered tools for MT gisting might therefore involve MT tools that embed a second MT tool directly into their user interface (Nurminen, 2020), or perhaps automatically show two outputs as a low-cost means of enhancing users' perceived transparency.

Source Understanding People use MT not only to gain access to texts across language boundaries, but also to augment and ensure their understanding of texts that are in languages they have limited competence in (Nurminen, 2021a). They might position a source text and its translation side-by-side and refer to both while reading, or they may look at both original and translated messages in an MT-mediated conversation (Nurminen, 2016). Recognizing this tendency, human-centered MT tools could make it easy to access original texts alongside their machine-translated versions, and provide affordances to compare them easily.

MT-mediated Communication HCI research has studied MT-mediated communication, and how the use of MT affects not only performance, but also interpersonal dynamics. Empirical evidence shows that people develop their own strategies to compensate for imperfect MT, such as adapting what they say (e.g., by employing redundant expressions and suppressing lexical variation in language use) (Yamashita and Ishida, 2006; Hara and Iqbal, 2015), using back-translation to assess outputs in a language they do not understand (Ito et al., 2023), or simply relying on their holistic understanding of the conversation to fill in gaps where the MT output does not make sense (Robertson and Díaz, 2022). Even when effective, these strategies come at a cost to communication: people communicate less naturally and authentically (Yamashita and Ishida, 2011) and might get misleading signals on translation quality (Tsai and Wang, 2015). Imperfect translations also affect interpersonal dynamics between interlocutors, increasing the risk

of participants misinterpreting their task partner's intent (Lim et al., 2022), misattributing communication breakdowns to human vs. MT-generated errors (Gao et al., 2014; Robertson and Díaz, 2022), and misassessing one another's contribution to the collaborative task (Xiao et al., 2024).

Trust Lay users' trust in MT is largely shaped by their perception of how MT-as-a-black-box functions, not just its intrinsic quality. Identical translations can be perceived differently when labeled machine vs. human-generated (Asscher and Glikson, 2021), and people might assign inconsistent ratings to the MT outputs before vs. after the label is disclosed (Bowker, 2009). Not all MT errors impact user trust equally: fluency or readability errors tend to lower trust more than adequacy errors, even though the latter can be more misleading when users rely on MT-generated meanings to inform their actions (Martindale and Carpuat, 2018; Popović, 2020). Factors like language proficiency, subject knowledge, and MT literacy influence how users perceive MT quality in gisting contexts (Nurminen, 2021a). MT literacy has also been shown to play a significant role in shaping translators' trust in MT (Scansani et al., 2019).

Taken together, this body of work highlights that building truly human-centered MT systems demands much more than generating fluent and adequate translations. It requires aligning system design with real-world communication practices, developing interaction strategies that empower users, and supporting their ability to assess risks in ideally independent and time-sensitive ways. Crucially, it also means empirically studying how these systems affect stakeholders, not just in terms of task performance, but also in how they shape interpersonal dynamics and shared understanding.

5 Ethicality of MT

The social implications of MT use extend beyond its immediate usefulness, bringing us into the realm of ethics. What does it mean for MT to be ethical? Surveying major frameworks of translation ethics (Koskinen and Pokorn, 2020) provides a foundation for addressing this question, highlighting the inherent multiplicity and conflicting perspectives in determining what is right or wrong in practice (Chesterman, 2001; Lambert, 2023).

For example, some approaches base translation ethics on strictly representing the original text's "precise" meaning and form, at all costs and un-

der all circumstances (Newmark, 1988). Other approaches emphasize a functional ethics of service, where ethical translation is defined by the translator's adherence to the requester's instructions, even if this means changing the source text or using it as mere inspiration (Holz-Mänttäri, 1984). Others prioritize alterity and social justice, viewing translation as a tool to challenge social and political inequalities by reframing communities' identity and values; in this case, ethical action might even involve refusing to translate the source text (Robinson, 2014). Several other translation ethics frameworks exist, each revolving around different priorities and values (Koskinen and Pokorn, 2020).

Today's influential ethical frameworks also imply that the translator's ethical response is necessarily situation- and text-dependent (Pym, 2012). By this we mean that for different texts, and in different situations, the ethical decision – whichever ethical framework one follows – may take different shapes. MT ethics, then, are no less situation-contingent than issues of MT usability or effectiveness.

Finally, a typology of the main approaches to translation ethics also reveals how some ethics are largely regional, or field-specific, inasmuch as they stem from the particular features of translation as a medium for intercultural communication (Pym, 2012, p. 57). In contrast, other approaches are more general in their concerns and values, and not intrinsic to the field of translation as such. Along these lines, it could be argued that a useful implementation of human-centered ethical evaluation in the case of MT should involve the compartmentalization of MT ethics from general AI ethics, and the preference for regional frameworks of ethics for MT (Asscher, 2025, p. 102–109). This implies a give-and-take between MT ethics oriented to the specificities of translation, on the one hand, and universal ethics, reminiscent of the general protocols of AI ethics proposed so abundantly in recent years, on the other hand (Floridi et al., 2018).

Relating these ethical insights to MT can apply to both the increasingly autonomous decision-making of the tool itself, and the social conditions that underpin its development and maintenance (Asscher, 2025, p. 98–101). The development and use of MT has already had vast consequences for many stakeholders. The ownership and distribution of anonymized translation data needed for the development of MT systems, and the re-use of this data to fine-tune MT, are some of the issues at stake, as there is currently no compensation for the origi-

nal human translators who created the data, and MT systems serve causes that are opaque to these translators and might contradict their values (Moorkens, 2022, p. 123–126). Issues of confidentiality and privacy are also pertinent, as personal translation data is utilized to train MT systems without regulation, rendering this data potentially identifiable (Nunes Vieira et al., 2022). The risks involved in high-stakes use of MT may strain the question of the moral and legal responsibility even further, for example in medical and legal situations, where translation errors may be particularly consequential (Vieira et al., 2021). Then, there are the sometimes problematic uses of MT in the professional translation workflow, and the broader issues of sustainability of the translation industry and environmental concerns (Bowker, 2020b; Skadiņa et al., 2023; Shterionov and Vanmassenhove, 2023). MT ethics also apply to the cultural and gender bias of contemporary LLMs (Gallegos et al., 2024), which may be manifested in translation, or the censorship recently enacted in some generative AI tools concerning certain charged historical occurrences, reinforcing unequal power relations across cultures (Wang et al., 2025; Bianchi et al., 2023).

Considering these points, human-centered MT research must pursue richer assessments of the moral consequences of its use in society. Studies of MT ethicality are valuable regardless of immediate implementability and can inform business and scientific leadership in governing the field and shaping MT agency and social implications.

6 Human-Centered MT Evaluation

MT evaluation has focused on benchmarking systems, or rating individual outputs, using automatic or human ratings of translation quality as ground truth (White and O'Connell, 1993; Koehn and Monz, 2006; Graham et al., 2013; Läubli et al., 2020; Freitag et al., 2021). Some recent proposals call for broadening its scope to measure social and environmental impact in addition to performance (Moorkens et al., 2024; Santy et al., 2021). A human-centered approach can draw from conceptualizations of the translation process and product quality from Translation Studies (Liu et al., 2024), and HCI methodology for evaluating systems in their socio-technical context (Liebling et al., 2022).

From Generic to Situated MT A key shift is from generic, context-independent evaluation toward situated assessments of fitness-for-purpose

and stakeholder impact. Holistic quality scores (Graham et al., 2013) are already complemented by fine-grained annotations such as MQM (Lommel et al., 2014). In contrast, Translation Studies work emphasizes evaluating translations based on their suitability for their intended purpose rather than adhering to a one-size-fits-all notion of quality (Bowker, 2009; Chesterman and Wagner, 2014; Colina, 2008). The impact of MT errors thus needs to be assessed in context (Agrawal et al., 2024), as general benchmarks may obscure rare but extreme errors (Shi et al., 2022). Expert knowledge might be required, for instance to determine whether an adequacy error poses a clinical risk (Khoong et al., 2019), or to assess social harms such as gender bias (Savoldi et al., 2021, 2024), name mistranslation (Sandoval et al., 2023), and lack of cultural awareness (Yao et al., 2024). Providing an “evaluation brief” (Liu et al., 2024) can describe the circumstances surrounding the translation creation, who it is for, and how it is intended to be used. Evaluation through question answering is another way to assess if translations preserve important information (Ki et al., 2025; Fernandes et al., 2025).

From Annotation to Human Studies Human studies that incorporate MT within the relevant end-user task can help us assess the impact of MT more comprehensively. Such tasks might align closely with the production and understanding of translations, such as post-editing MT (Castilho and O’Brien, 2016; Castilho and O’Brien, 2017; Bawden et al., 2024; Savoldi et al., 2024), reading comprehension (Jones et al., 2005; Scarton and Specia, 2016), gisting (Nurminen, 2021a) or triage tasks (Martindale and Carpuat, 2022). MT might be a tool in support of another task, such as collaborative information exchange in teams (Yamashita and Ishida, 2006), social media consumption (Lim et al., 2018), hiring and personnel decision making (Zhang et al., 2022) or housing information seeking (Xiao et al., 2025), and everyday conversations (Robertson and Díaz, 2022). As Santy et al. (2021) show, in such real-world cases, machine-aided translation systems can bring significant value to end-users. Nevertheless, this value is often contextualized within trade-offs among time, performance, and computational cost, especially given the limited technical accessibility and important occurrence of low-resource language settings.

From Static Benchmarks to Iterative Design Evaluations with human users do not only occur

at the end of a project; rather, they drive the iterative refinement cycle of the entire human-centered design process. This process typically begins with needs-finding studies to identify the social problem that technical solutions aim to resolve (Gao and Fussell, 2017; Gao et al., 2022; Xiao et al., 2024). It is often followed by co-design activities, where existing tools are used as technology probes to elicit inputs from targeted user groups on MT design. Subsequent phases include usability testing or clinical trials after each round of system development to determine the degree of success (Khoong and Rodriguez, 2022). There exist a wealth of frameworks to guide this process, including Human-centered design, Participatory Design, and Value Sensitive Design (Friedman and Hendry, 2019), all of which foreground the values of direct and indirect stakeholders. MT evaluation can also draw from frameworks for trustworthy AI, particularly methods for studying mental models (Bansal et al., 2019), trust calibration (Vereschak et al., 2021), and how a human-AI work system performs (Hoffman et al., 2023). These efforts aim to ensure that MT systems can account for the complex dynamics between system outputs, user interpretations, and downstream consequences, thereby requiring interdisciplinary collaborations and tailored study designs.

7 Human-Centered MT Design

This section outlines emerging techniques that can reframe MT as a contextual, potentially interactive process responsive to users’ needs, moving beyond traditional sequence transduction. It provides a richer toolbox to support MT literacy (Section 3) and build on past empirical studies of human-MT interaction (Section 4).

Richer Inputs, Many Outputs Human-Centered MT must adapt outputs to the audience and context. Research has already explored controlling formality (Sennrich et al., 2016; Rippeth et al., 2022), style (Niu et al., 2017; Agarwal et al., 2023), complexity (Agrawal and Carpuat, 2019; Oshika et al., 2024), and personalization (Mirkin and Meunier, 2015; Rabinovich et al., 2016). Adaptation may also require explaining content (Srikanth and Li, 2021; Han et al., 2023; Saha et al., 2025), or warning about cultural misunderstandings (Pituxcoosuvarn et al., 2020; Yao et al., 2024). However, it is still unclear how users and other stakeholders can guide these

systems in proactive and ecologically valid ways.

More contextual inputs are needed, similar to translator briefs (Castilho and Knowles, 2024). MT work has considered incorporating domain knowledge (Clark et al., 2012; Chu and Wang, 2018), style labels (Sennrich et al., 2016; Niu et al., 2017), example translations (Xu et al., 2023; Agrawal et al., 2023; Bouthors et al., 2024), and terminology (Alam et al., 2021; Michon et al., 2020). Some also address long-form (Karpinska and Iyyer, 2023; Peng et al., 2024) and conversational translation (Bawden et al., 2021; Pombal et al., 2024). However, these efforts usually consider one dimension of context at a time; we still need more holistic approaches that take a broad view of context (Castilho and Knowles, 2024) and incorporate knowledge and feedback needed for culturally appropriate outputs (Tenzer et al., 2024; Saha et al., 2025).

An Iterative Translation Process LLMs enable multi-stage translation workflows, including pre-editing, evaluation, and post-editing (Briakou et al., 2024; Alves et al., 2024). Pre-editing involves rewriting source texts to improve MT output (Bowker and Ciro, 2019a; Štajner and Popović, 2019; Ki and Carpuat, 2025), while post-editing—either human or automatic—is studied widely (Lin et al., 2022; Vidal et al., 2022; Ki and Carpuat, 2024). Yet, most work remains system-centric. Interactive approaches designed for professional translators (Green et al., 2013; Briva-Iglesias et al., 2023) suggest benefits from involving lay users with diverse goals and levels of proficiency.

Scale & Context How can we specialize models for specific contexts while reaping the benefits of scale (Team et al., 2022; Johnson et al., 2017; Vilar et al., 2023; Kocmi et al., 2024)? Work in this direction could build on efforts to structure resources for horizontal (across languages) and vertical (across domains) generalization (Ishida, 2006; Rehm, 2023), and techniques to support task (Ye et al., 2022; Alves et al., 2024), language (Blevins et al., 2024), and domain and terminology (Segonne et al., 2024) specialization in LLMs.

Decentering MT Centering people means recognizing that MT is often just one part of a broader workflow, where the MT output is not the end product. MT today often participate in content co-production with humans, rather than only for source-to-target conversion. This can be done via synchronized bilingual writing (Crego et al., 2023;

Xiao et al., 2024) or using translation as an aid for scientific writing (O’Brien et al., 2018; Steigerwald et al., 2022; Ito et al., 2023). In those settings, even when translating an abstract, the translation might be more of an adaptation than a literal translation (Bawden et al., 2024). Translation can be implicit or partial, when supporting simultaneous interpreters (Grissom et al., 2024), enabling natural translanguaging practices of bilinguals (Zhang et al., 2025b), or searching for texts written in a foreign language given a native language query (Galuščáková et al., 2022; Nair et al., 2022). In those settings, human-MT interface design is critical for lay users to remain aware of features of the targeted content and to develop strategies for navigating it (Petrelli et al., 2006). The need for intelligent interface design is particularly pronounced in LLM-powered multilingual communication and user interactions with conversational agents, where models must interpret and generate content for fluid language use while adapting to user goals, styles, and cultural norms. To support this, a prompt engineering playground with customized MT and user interfaces may enhance the accessibility of LLMs for a broader population (Mondshine et al., 2025).

Risk Management Reliable MT should help users weigh the benefits of MT against the risks it may pose. Quality estimation techniques designed for explainability have provided a good foundation toward this goal (Fomicheva et al., 2021; Guerreiro et al., 2023; Briakou et al., 2023; Specia et al., 2018). That said, growing evidence from user studies shows that more work is needed to identify and assess risks (Koponen and Nurminen, 2024), generate actionable feedback in user-specified contexts (Zouhar et al., 2021; Mehandru et al., 2023), determine when and how to disclose the use of MT (Simard, 2024; Xiao et al., 2024), provide useful descriptions of model properties (Mitchell et al., 2019), promote MT literacy among lay users (Bowker and Ciro, 2019a), and support the development of accurate user mental models (Bansal et al., 2019). Frameworks from human-centered explainable AI, such as seamless design (Ehsan et al., 2022), can help pinpoint gaps between system affordances and the needs of human stakeholders, fostering better alignment.

In sum, while existing work offers a rich toolbox for human-centered MT, more research is needed on designing interactions that preserve user agency and support effective, trustworthy use. This in-

cludes new interfaces that balance simplicity and flexibility, and foundational work on training models for controllability and context-awareness.

8 Case Study: Toward Reliable Translation for Clinical Care

Research on MT for clinical settings illustrates how human studies can drive the cycle of human-centered MT (Section 6) by understanding specific contexts of use (Section 2) to guide interface and model design decisions (Sections 4,7).

Understanding Needs Language barriers are a major source of healthcare disparities (Cano-Ibáñez et al., 2021), yet access to professional interpreters remains limited (Flores, 2005; Ortega et al., 2023). MT can potentially support clinical care, but reliability is a critical concern: MT errors can cause serious harm in, for example, discharge instructions from emergency departments (Khoong et al., 2019; Taira et al., 2021), pediatric care (Brewster et al., 2024) or urology (Rao et al., 2024), with disparate impact across languages. Yet, MT frequently mediates interactions between healthcare providers and patients in practice (Genovese et al., 2024). While dedicated MT tools have been developed for clinical settings (Starlander et al., 2005; Bouillon et al., 2005), generic apps such as Google Translate are still most commonly used (Nunes Vieira, 2024). In face of challenges such as time constraints, cultural barriers, and medical literacy gaps, clinicians develop their own workarounds when using MT, such as back-translation or relying on non-verbal cues to assess understanding (Mehandru et al., 2022).

Research Directions Generic MT tools thus often fall short in clinical care, and needs-findings studies motivate research into integrating pre-translated medical phrases, multimodal communication support, and interactive tools to assess mutual understanding. A human study evaluated feedback mechanisms to assist physicians in assessing the reliability of MT outputs in clinical settings, finding that quality estimation tools generally improve physicians' reliance on MT but fail to detect the most clinically severe errors (Mehandru et al., 2023). Complementary efforts focus on developing custom MT approaches that prioritize reliability and verifiability, by using vetted canonical phrases to scaffold the translation (Bouillon et al., 2017) or guide users in crafting better MT inputs (Robertson, 2023). While these works focus on text-based

MT, many healthcare use cases also warrant consideration of interaction using speech (Spechbach et al., 2019), sign language (Esselink et al., 2024) and pictographs (Gerlach et al., 2024). Cultural differences significantly impact the style and content of communication in healthcare (Kreuter and McClure, 2004; Brooks et al., 2019) and is another area where much research is needed. Khoong and Rodriguez (2022) further outline key domains for future research, including developing interactive tools for different types of communication; enhancing risk assessment, and assessing understanding and patient satisfaction on top of MT correctness.

9 Conclusion

Recontextualizing MT through Translation Studies and HCI highlights that truly supporting real-world needs demands understanding translation as a socio-technical process and designing user-centric tools. Each field offers important insights, and their synergy fuels new research.

Translation Studies provides theoretical and empirical frameworks for contextualizing assessments of translation quality, accounting for user diversity, and for framing translation as a process of situated decision-making that can inform our view of MT as it becomes part of increasingly diverse workflows. HCI complements this by focusing on real-world user experience with translation technologies, emphasizing needs, interface design, feedback, and collaboration in multilingual interactions. Both fields offer methods to evaluate stakeholder perceptions and behaviors, but mostly study the off-the-shelf MT and NLP tools which limits the space of interaction design. Conversely, MT/NLP offers a rich toolkit of generation, adaptation, and evaluation techniques, which are developed with less focus on user experience and context.

Interdisciplinary collaboration enables a shift towards genuinely human-centered systems, where users are active agents in a "machine in the loop" process. This approach poses key technical challenges for MT: how to personalize translation outputs, how to support interaction and control, how to model trust and adaptation, how to balance generalization and responsiveness to context, and how to sustain human agency in language use. Nevertheless, it promises greater real-world impact through more expansive conceptualizations of MT technology that support situated, embodied, and socially meaningful communication.

Limitations

Scope This survey is not exhaustive. While we aimed to highlight diverse perspectives, we cannot cover the breadth of the literature across Translation Studies, Human-Centered Interaction, Machine Translation and Natural Language Processing. To narrow down the scope, we facilitated discussions between experts in these disciplines to highlight connections and tensions across fields. We used the take-aways from these discussions to prioritize this survey. Furthermore, human-centered MT can draw upon insights and methodologies from many other disciplines, including linguistics and sociolinguistics, cognitive science and psychology, information science, communication studies, and education.

Multimodality Most work surveyed here focused on text translation, but human-centered MT must incorporate multiple modalities, such as speech, images, and gestures, reflecting the way people communicate. In addition to speech translation (Akiba et al., 2004; Agarwal et al., 2023) and its connection to simultaneous interpretation (Grissom et al., 2014; Wang et al., 2016), prior work has considered the role of vision in translating image captions and video-guided translation (Specia et al., 2016; Sulubacak et al., 2020). Pre-trained language models that encompass speech (Radford et al., 2022; Ambilduke et al., 2025) and vision (Radford et al., 2021; Chen et al., 2024) open new research directions. Further research with a human-centered perspective might include developing adaptive interfaces that detect errors (Han et al., 2024), seamlessly integrate multiple modalities and support repair (Sulubacak et al., 2020), thereby enabling more natural and effective human-computer interactions. However, a thorough treatment of multimodality in human-centered MT is beyond the scope of this paper.

Language Resource Disparities Unequal coverage and quality of MT techniques across languages remains a fundamental limitation which must be taken into account to develop human-centered MT. Many methods discussed, particularly in Section 7, are currently more feasible for high-resource languages. However, employing human-centered design methods and focusing on specific use cases can help develop strategies to mitigate disparities in translation quality across various languages, domains, and dialects (Santy et al., 2021).

Acknowledgements

This work was made possible by the NII Shonan Meeting on Human-Centered Machine Translation, organized by Marine Carpuat, Toru Ishida and Niloofar Salehi. We thank the National Institute of Informatics, Japan, for providing an excellent venue and support for productive discussions. We are also grateful to all participants for their contributions. We also thank Sharon O’Brien for earlier discussions.

References

- Mark S. Ackerman. 2000. *The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility*. *Human–Computer Interaction*, 15(2-3):179–203.
- Milind Agarwal, Sweta Agrawal, Antonios Anastasopoulos, Luisa Bentivogli, Ondřej Bojar, Claudia Borg, Marine Carpuat, Roldano Cattoni, Mauro Cettolo, Mingda Chen, William Chen, Khalid Choukri, Alexandra Chronopoulou, Anna Currey, Thierry Declerck, Qianqian Dong, Kevin Duh, Yannick Estève, Marcello Federico, and 43 others. 2023. **FINDINGS OF THE IWSLT 2023 EVALUATION CAMPAIGN**. In *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT 2023)*, pages 1–61, Toronto, Canada (in-person and online). Association for Computational Linguistics.
- Sweta Agrawal and Marine Carpuat. 2019. **Controlling Text Complexity in Neural Machine Translation**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1549–1564, Hong Kong, China. Association for Computational Linguistics.
- Sweta Agrawal, Amin Farajian, Patrick Fernandes, Ricardo Rei, and André F. T. Martins. 2024. **Assessing the role of context in chat translation evaluation: Is context helpful and under what conditions?** *Transactions of the Association for Computational Linguistics*, 12:1250–1267.
- Sweta Agrawal, Chunting Zhou, Mike Lewis, Luke Zettlemoyer, and Marjan Ghazvininejad. 2023. **In-context Examples Selection for Machine Translation**. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8857–8873, Toronto, Canada. Association for Computational Linguistics.
- Yasuhiro Akiba, Marcello Federico, Noriko Kando, Hiromi Nakaiwa, Michael Paul, and Jun’ichi Tsujii. 2004. Overview of the IWSLT evaluation campaign. In *Proceedings of the First International Workshop on Spoken Language Translation: Evaluation Campaign*, Kyoto, Japan.

- Md Mahfuz Ibn Alam, Ivana Kvapilíková, Antonios Anastasopoulos, Laurent Besacier, Georgiana Dinu, Marcello Federico, Matthias Gallé, Kweonwoo Jung, Philipp Koehn, and Vassilina Nikoulina. 2021. Findings of the WMT Shared Task on Machine Translation Using Terminologies. In *Proceedings of the Sixth Conference on Machine Translation*, pages 652–663, Online. Association for Computational Linguistics.
- Duarte M. Alves, José Pombal, Nuno M. Guerreiro, Pedro H. Martins, João Alves, Amin Farajian, Ben Peters, Ricardo Rei, Patrick Fernandes, Sweta Agrawal, Pierre Colombo, José G. C. de Souza, and André F. T. Martins. 2024. Tower: An Open Multilingual Large Language Model for Translation-Related Tasks. *Preprint*, arXiv:2402.17733.
- Kshitij Ambilduke, Ben Peters, Sonal Sannigrahi, Anil Keshwani, Tsz Kin Lam, Bruno Martins, Marcella Zanon Boito, and André F. T. Martins. 2025. From TOWER to SPIRE: Adding the Speech Modality to a Text-Only LLM. *Preprint*, arXiv:2503.10620.
- Ryoko Anazawa, Hirono Ishikawa, and Kiuchi Takahiro. 2013. Use of online machine translation for nursing literature: A questionnaire-based survey. *The Open Nursing Journal*, 7:22–28.
- Omri Asscher. 2025. *Machine Translation and Translation Theory*. Routledge.
- Omri Asscher and Ella Glikson. 2021. Human evaluations of machine translation in an ethically charged situation. *New Media & Society*, page 14614448211018833.
- Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7:2–11.
- Rachel Bawden, Eric Bilinski, Thomas Lavergne, and Sophie Rosset. 2021. DiaBLA: A corpus of bilingual spontaneous written dialogues for machine translation. *Language Resources and Evaluation*, 55(3):635–660.
- Rachel Bawden, Ziqian Peng, Maud Bénard, Éric Clergerie, Raphaël Esamotunu, Mathilde Huguin, Natalie Kübler, Alexandra Mestivier, Mona Michelot, Laurent Romary, Lichao Zhu, and François Yvon. 2024. Translate your Own: A Post-Editing Experiment in the NLP domain. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, pages 431–443, Sheffield, UK. European Association for Machine Translation (EAMT).
- Emily M Bender and Alvin Grissom, II. 2024. Power Shift: Toward Inclusive Natural Language Processing. *Inclusion in Linguistics*, page 199.
- Federico Bianchi, Tommaso Fornaciari, Dirk Hovy, and Debora Nozza. 2023. Gender and Age Bias in Commercial Machine Translation. In Helena Moniz and Carla Parra Escartín, editors, *Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation*, pages 159–184. Springer Verlag.
- Terra Blevins, Tomasz Limisiewicz, Suchin Gururangan, Margaret Li, Hila Gonen, Noah A. Smith, and Luke Zettlemoyer. 2024. Breaking the Curse of Multilinguality with Cross-lingual Expert Language Models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10822–10837, Miami, Florida, USA. Association for Computational Linguistics.
- Pierrette Bouillon, Johanna Gerlach, Hervé Spechbach, Nikolaos Tsourakis, and Ismahene Sonia Halimi Mallem. 2017. BabelDr vs Google Translate: A user study at Geneva University Hospitals (HUG). In *20th Annual Conference of the European Association for Machine Translation (EAMT)*.
- Pierrette Bouillon, Manny Rayner, Nikos Chatzichrisafis, Beth Ann Hockey, Marianne Santaholma, Marianne Starlander, Yukie Nakao, Kyoko Kanzaki, and Hitoshi Isahara. 2005. A generic multi-lingual open source platform for limited-domain medical speech translation. In *Proceedings of the 10th EAMT Conference: Practical Applications of Machine Translation*, Budapest, Hungary. European Association for Machine Translation.
- Maxime Bouthors, Josep Crego, and François Yvon. 2024. Retrieving Examples from Memory for Retrieval Augmented Neural Machine Translation: A Systematic Comparison. In *Findings of the Association for Computational Linguistics: NAACL 2024*, Findings of the Association for Computational Linguistics: NAACL 2024, pages 3022–3039, Mexico, Mexico. Association for Computational Linguistics.
- Lynne Bowker. 2009. Can Machine Translation meet the needs of official language minority communities in Canada? A recipient evaluation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 8.
- Lynne Bowker. 2020a. Chinese speakers’ use of machine translation as an aid for scholarly writing in English: A review of the literature and a report on a pilot workshop on machine translation literacy. *Asia Pacific Translation and Intercultural Studies*, 7(3):288–298.
- Lynne Bowker. 2020b. Translation technology and ethics. In *The Routledge Handbook of Translation and Ethics*. Routledge.
- Lynne Bowker. 2023. *De-Mystifying Translation: Introducing Translation to Non-translators*. Routledge, London.

- Lynne Bowker and Jairo Buitrago Ciro. 2019a. *Expanding the Reach of Knowledge Through Translation-Friendly Writing*. In *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, pages 55–78. Emerald Publishing Limited.
- Lynne Bowker and Jairo Buitrago Ciro. 2019b. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Emerald Publishing Limited.
- Ryan C. L. Brewster, Priscilla Gonzalez, Rohan Khazanchi, Alex Butler, Raquel Selcer, Derrick Chu, Barbara Pontes Aires, Marcella Luercio, and Jonathan D. Hron. 2024. *Performance of ChatGPT and Google Translate for Pediatric Discharge Instruction Translation*. *Pediatrics*, 154(1):e2023065573.
- Eleftheria Briakou, Navita Goyal, and Marine Carpuat. 2023. *Explaining with Contrastive Phrasal Highlighting: A Case Study in Assisting Humans to Detect Translation Differences*. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11220–11237, Singapore. Association for Computational Linguistics.
- Eleftheria Briakou, Jiaming Luo, Colin Cherry, and Markus Freitag. 2024. *Translating Step-by-Step: Decomposing the Translation Process for Improved Translation Quality of Long-Form Texts*. *Preprint*, arXiv:2409.06790.
- Vicent Briva-Iglesias, Sharon O’Brien, and Benjamin R. Cowan. 2023. *The impact of traditional and interactive post-editing on Machine Translation User Experience, quality, and productivity*. *Translation, Cognition & Behavior*, 6(1):60–86.
- Laura A. Brooks, Elizabeth Manias, and Melissa J. Bloomer. 2019. *Culturally sensitive communication in healthcare: A concept analysis*. *Collegian*, 26(3):383–391.
- Patrick Cadwell, Sheila Castilho, Sharon O’Brien, and Linda Mitchell. 2016. *Human factors in machine translation and post-editing among institutional translators*. *Translation Spaces*, 5(2):222–243.
- Naomi Cano-Ibáñez, Yasmin Zolfaghari, Carmen Amezua-Prieto, and Khalid Saeed Khan. 2021. *Physician-Patient Language Discordance and Poor Health Outcomes: A Systematic Scoping Review*. *Frontiers in Public Health*, 9:629041.
- Tara Capel and Margot Brereton. 2023. *What is Human-Centered about Human-Centered AI? A Map of the Research Landscape*. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI ’23, pages 1–23, New York, NY, USA. Association for Computing Machinery.
- Sheila Castilho and Rebecca Knowles. 2024. *A survey of context in neural machine translation and its evaluation*. *Natural Language Processing*, pages 1–31.
- Sheila Castilho and Sharon O’Brien. 2016. *Evaluating the impact of light post-editing on usability*. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 310–316, Portorož, Slovenia. European Language Resources Association (ELRA).
- Sheila Castilho and Sharon O’Brien. 2017. Acceptability of machine-translated content: A multi-language evaluation by translators and end-users. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 16.
- Stevie Chancellor. 2023. *Toward Practices for Human-Centered Machine Learning*. *Commun. ACM*, 66(3):78–85.
- Xi Chen, Josip Djolonga, Piotr Padlewski, Basil Mustafa, Soravit Changpinyo, Jialin Wu, Carlos Riquelme Ruiz, Sebastian Goodman, Xiao Wang, Yi Tay, Siamak Shakeri, Mostafa Dehghani, Daniel Salz, Mario Lucic, Michael Tschannen, Arsha Nagrani, Hexiang Hu, Mandar Joshi, Bo Pang, and 24 others. 2024. *On Scaling Up a Multilingual Vision and Language Model*. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14432–14444.
- Andrew Chesterman. 2001. *Proposal for a Hieronymic Oath*. *The Translator*, 7(2):139–154.
- Andrew Chesterman and Emma Wagner. 2014. *Can Theory Help Translators?: A Dialogue Between the Ivory Tower and the Wordface*. Routledge, London.
- Chenhui Chu and Rui Wang. 2018. *A Survey of Domain Adaptation for Neural Machine Translation*. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Jonathan H. Clark, Alon Lavie, and Chris Dyer. 2012. *One System, Many Domains: Open-Domain Statistical Machine Translation via Feature Augmentation*. In *Proceedings of the Conference of the Association for Machine Translation in the Americas*.
- Sonia Colina. 2008. *Translation Quality Evaluation: Empirical Evidence for a Functional Approach*. *The Translator*, 14(1):97–134.
- Josep Crego, Jitao Xu, and François Yvon. 2023. *BiSync: A Bilingual Editor for Synchronized Monolingual Texts*. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 369–376, Toronto, Canada. Association for Computational Linguistics.
- Maureen Ehrensberger-Dow, Delorme Benites , Alice, and Caroline and Lehr. 2023. *A new role for translators and trainers: MT literacy consultants*. *The Interpreter and Translator Trainer*, 17(3):393–411.

- Upol Ehsan, Q. Vera Liao, Samir Passi, Mark O. Riedl, and Hal Daume III. 2022. *Seamful XAI: Operationalizing Seamful Design in Explainable AI*. Preprint, arXiv:2211.06753.
- Carla Parra Escartín and Helena Moniz. 2019. Ethical considerations on the use of machine translation and crowdsourcing in cascading crises. In *Translation in Cascading Crises*. Routledge.
- Lyke Esselink, Floris Roelofsen, Jakub Dotlačil, Shani Mende-Gillings, Maartje de Meulder, Nienke Sijm, and Anika Smeijers. 2024. *Exploring automatic text-to-sign translation in a healthcare setting*. *Universal Access in the Information Society*, 23(1):35–57.
- Patrick Fernandes, Sweta Agrawal, Emmanouil Zarnis, André FT Martins, and Graham Neubig. 2025. Do LLMs understand your translations? evaluating paragraph-level MT with question answering. *arXiv preprint arXiv:2504.07583*.
- Glenn Flores. 2005. *The impact of medical interpreter services on the quality of health care: A systematic review*. *Medical care research and review: MCRR*, 62(3):255–299.
- Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. *Minds and Machines*, 28(4):689–707.
- Marina Fomicheva, Piyawat Lertvittayakumjorn, Wei Zhao, Steffen Eger, and Yang Gao. 2021. *The Eval4NLP Shared Task on Explainable Quality Estimation: Overview and Results*. Preprint, arXiv:2110.04392.
- Markus Freitag, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan, and Wolfgang Macherey. 2021. *Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation*. *Transactions of the Association for Computational Linguistics*, 9:1460–1474.
- Batya Friedman and David G. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press.
- Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. 2024. *Bias and Fairness in Large Language Models: A Survey*. *Computational Linguistics*, 50(3):1097–1179.
- Petra Galuščáková, Douglas W. Oard, and Suraj Nair. 2022. *Cross-language Information Retrieval*. Preprint, arXiv:2111.05988.
- Ge Gao and Susan R. Fussell. 2017. *A Kaleidoscope of Languages: When and How Non-Native English Speakers Shift between English and Their Native Language during Multilingual Teamwork*. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, pages 760–772, New York, NY, USA. Association for Computing Machinery.
- Ge Gao, Bin Xu, Dan Cosley, and Susan R. Fussell. 2014. *How beliefs about the presence of machine translation impact multilingual collaborations*. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW ’14, page 1549–1560, New York, NY, USA. Association for Computing Machinery.
- Ge Gao, Bin Xu, David C. Hau, Zheng Yao, Dan Cosley, and Susan R. Fussell. 2015. *Two is Better Than One: Improving Multilingual Collaboration by Giving Two Machine Translation Outputs*. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW ’15, pages 852–863, New York, NY, USA. Association for Computing Machinery.
- Ge Gao, Jian Zheng, Eun Kyoung Choe, and Naomi Yamashita. 2022. *Taking a Language Detour: How International Migrants Speaking a Minority Language Seek COVID-Related Information in Their Host Countries*. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):542:1–542:32.
- Federico Gaspari and John Hutchins. 2007. Online and free! Ten years of online machine translation: Origins, developments, current use and future prospects. In *Proceedings of Machine Translation Summit XI: Papers*, Copenhagen, Denmark.
- Susan Gasson. 2003. Human-centered vs. user-centered approaches to information system design. *Journal of Information Technology Theory and Application (JITTA)*, 5.
- Ariana Genovese, Sahar Borna, Cesar A. Gomez-Cabello, Syed Ali Haider, Srinivasagam Prabha, Antonio J. Forte, and Benjamin R. Veenstra. 2024. *Artificial intelligence in clinical settings: A systematic review of its role in language translation and interpretation*. *Annals of Translational Medicine*, 12(6):117.
- Johanna Gerlach, Pierrette Bouillon, Jonathan Mutual, and Hervé Speichbach. 2024. A Concept Based Approach for Translation of Medical Dialogues into Pictographs. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 233–242, Torino, Italia. ELRA and ICCL.
- Madalena Gonçalves, Marianna Buchicchio, Craig Stewart, Helena Moniz, and Alon Lavie. 2022. Agent and User-Generated Content and its Impact on Customer Support MT. In *Proceedings of the 23rd Annual Conference of the European Association for Ma-*

- chine Translation*, pages 201–210, Ghent, Belgium. European Association for Machine Translation.
- Yvette Graham, Timothy Baldwin, Alistair Moffat, and Justin Zobel. 2013. Continuous Measurement Scales in Human Evaluation of Machine Translation. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 33–41, Sofia, Bulgaria. Association for Computational Linguistics.
- Spence Green, Jeffrey Heer, and Christopher D. Manning. 2013. *The efficacy of human post-editing for language translation*. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’13*, pages 439–448, Paris, France. Association for Computing Machinery.
- Spence Green, Jeffrey Heer, and Christopher D. Manning. 2015. *Natural Language Translation at the Intersection of AI and HCI: Old questions being answered with both AI and HCI*. *Queue*, 13(6):30–42.
- Alvin Grissom, II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III. 2014. *Don’t Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation*. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1342–1352, Doha, Qatar. Association for Computational Linguistics.
- Alvin Grissom, II, Jo Shoemaker, Benjamin Goldman, Ruikang Shi, Craig Stewart, C. Anton Rytting, Leah Findlater, and Jordan Boyd-Graber. 2024. *Rapidly piloting real-time linguistic assistance for simultaneous interpreters with untrained bilingual surrogates*. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13548–13556, Torino, Italia. ELRA and ICCL.
- Nuno M. Guerreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André F. T. Martins. 2023. *xCOMET: Transparent Machine Translation Evaluation through Fine-grained Error Detection*. *Preprint*, arXiv:2310.10482.
- Yanzhu Guo, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel. 2024. *The Curious Decline of Linguistic Diversity: Training Language Models on Synthetic Text*. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3589–3604, Mexico City, Mexico. Association for Computational Linguistics.
- HyoJung Han, Jordan Boyd-Graber, and Marine Carpuat. 2023. *Bridging Background Knowledge Gaps in Translation with Automatic Explication*. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9718–9735, Singapore. Association for Computational Linguistics.
- HyoJung Han, Kevin Duh, and Marine Carpuat. 2024. *SpeechQE: Estimating the Quality of Direct Speech Translation*. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 21852–21867, Miami, Florida, USA. Association for Computational Linguistics.
- Kotaro Hara and Shamsi T. Iqbal. 2015. *Effect of Machine Translation in Interlingual Conversation: Lessons from a Formative Study*. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI ’15*, pages 3473–3482, New York, NY, USA. Association for Computing Machinery.
- Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. *How good are gpt models at machine translation? a comprehensive evaluation*. *Preprint*, arXiv:2302.09210.
- Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. 2023. *Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance*. *Frontiers in Computer Science*, 5.
- Justa Holz-Mänttäri. 1984. *Translatorisches handeln: Theorie und method*. Technical report, Suomalainen tiedeakatemia Helsinki.
- Eduard Hovy, Margaret King, and Andrei Popescu-Belis. 2002. *Principles of Context-Based Machine Translation Evaluation*. *Machine Translation*, 17(1):43–75.
- Chang Hu, Benjamin B. Bederson, and Philip Resnik. 2010. *Translation by iterative collaboration between monolingual users*. In *Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP ’10*, pages 54–55, New York, NY, USA. Association for Computing Machinery.
- W. John Hutchins. 2001. *Machine Translation over fifty years. Histoire, Epistémologie, Langage*, Tome XXII, fasc. 1:7–31.
- Toru Ishida. 2006. *Language Grid: An Infrastructure for Intercultural Collaboration*. In *Proceedings of the International Symposium on Applications on Internet, SAINT ’06*, pages 96–100, USA. IEEE Computer Society.
- Toru Ishida. 2016. *Intercultural Collaboration and Support Systems: A Brief History*. In *PRIMA 2016: Principles and Practice of Multi-Agent Systems*, pages 3–19, Cham. Springer International Publishing.
- Takumi Ito, Naomi Yamashita, Tatsuki Kurabayashi, Masatoshi Hidaka, Jun Suzuki, Ge Gao, Jack Jamieson, and Kentaro Inui. 2023. *Use of an AI-powered Rewriting Support Software in Context with Other Tools: A Study of Non-Native English Speakers*. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST ’23*, pages 1–13, New York, NY, USA. Association for Computing Machinery.

- Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. **Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation.** *Transactions of the Association for Computational Linguistics*, 5:339–351.
- D. Jones, E. Gibson, W. Shen, N. Granoien, M. Herzog, D. Reynolds, and C. Weinstein. 2005. **Measuring human readability of machine generated text: Three case studies in speech recognition and machine translation.** In *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 5, pages v/1009–v/1012 Vol. 5.
- Marzena Karpinska and Mohit Iyyer. 2023. **Large Language Models Effectively Leverage Document-level Context for Literary Translation, but Critical Errors Persist.** In *Proceedings of the Eighth Conference on Machine Translation*, pages 419–451, Singapore. Association for Computational Linguistics.
- Ramunė Kasperė, Jolita Horbačauskienė, Jurgita Motiejūnienė, Vilma Liubiniénė, Irena Patašienė, and Martynas Patašius. 2021. **Towards Sustainable Use of Machine Translation: Usability and Perceived Quality from the End-User Perspective.** *Sustainability*, 13(23):13430.
- Martin Kay. 1980/1997. **The Proper Place of Men and Machines in Language Translation.** *Machine Translation*, 12(1):3–23.
- Dorothy Kenny, Olga Torres-Hostench, Caroline Rossi, Alice Carré, Pilar Sánchez-Gijón, Sharon O’Brien, Joss Moorkens, Juan Antonio Pérez-Ortiz, Mikel L. Forcada, Felipe Sánchez-Martínez, and Gema Ramírez-Sánchez. 2022. **Machine Translation for Everyone.** Language Science Press.
- Elaine C. Khoong and Jorge A. Rodriguez. 2022. **A Research Agenda for Using Machine Translation in Clinical Medicine.** *Journal of General Internal Medicine*, 37(5):1275–1277.
- Elaine C. Khoong, Eric Steinbrook, Cortlyn Brown, and Alicia Fernandez. 2019. **Assessing the Use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions.** *JAMA Internal Medicine*, 179(4):580–582.
- Dayeon Ki and Marine Carpuat. 2024. **Guiding Large Language Models to Post-Edit Machine Translation with Error Annotations.** In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 4253–4273, Mexico City, Mexico. Association for Computational Linguistics.
- Dayeon Ki and Marine Carpuat. 2025. **Automatic Input Rewriting Improves Translation with Large Language Models.** *Preprint*, arXiv:2502.16682.
- Dayeon Ki, Kevin Duh, and Marine Carpuat. 2025. **ASKQE: Question answering as automatic evaluation for machine translation.** *arXiv preprint arXiv:2504.11582*.
- Tom Kocić, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Benjamin Marie, Christof Monz, Kenton Murray, Masaaki Nagata, Martin Popel, Maja Popović, and 3 others. 2024. **Findings of the WMT24 General Machine Translation Shared Task: The LLM Era Is Here but MT Is Not Solved Yet.** In *Proceedings of the Ninth Conference on Machine Translation*, pages 1–46, Miami, Florida, USA. Association for Computational Linguistics.
- Philipp Koehn. 2010. **Enabling monolingual translators: Post-editing vs. options.** In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT ’10*, pages 537–545, USA. Association for Computational Linguistics.
- Philipp Koehn, Michael Carl, Francisco Casacuberta, and Eva Marcos. 2014. **CASMACAT: Cognitive analysis and statistical methods for advanced computer aided translation.** In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, page 57, Dubrovnik, Croatia. European Association for Machine Translation.
- Philipp Koehn and Christof Monz. 2006. **Manual and Automatic Evaluation of Machine Translation between European Languages.** In *Proceedings on the Workshop on Statistical Machine Translation*, pages 102–121, New York City. Association for Computational Linguistics.
- Maarit Koponen and Mary Nurminen. 2024. **Risk management for content delivery via raw machine translation.** In *Translation, Interpreting and Technological Change : Innovations in Research, Practice and Training*, pages 111–135. Bloomsbury Publishing.
- Kaisa Koskinen and Nike K. Pokorn. 2020. **The Routledge Handbook of Translation and Ethics.** Routledge Handbooks in Translation and Interpreting Studies. Routledge.
- Matthew W. Kreuter and Stephanie M. McClure. 2004. **The role of culture in health communication.** *Annual Review of Public Health*, 25:439–455.
- Miguel L. Lacruz Mantecón. 2023. **Authorship and Rights Ownership in the Machine Translation Era.** In Helena Moniz and Carla Parra Escartín, editors, *Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation*, pages 71–92. Springer International Publishing, Cham.
- Jonathan Lambert. 2023. **Translation Ethics.** Routledge.

- Samuel Läubli, Sheila Castilho, Graham Neubig, Rico Sennrich, Qinlan Shen, and Antonio Toral. 2020. **A Set of Recommendations for Assessing Human-Machine Parity in Language Translation**. *Journal of Artificial Intelligence Research*, 67:653–672–653–672.
- William Lewis, Robert Munro, and Stephan Vogel. 2011. Crisis MT: Developing A Cookbook for MT in Crisis Situations. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 501–511, Edinburgh, Scotland. Association for Computational Linguistics.
- William D. Lewis and Jan Niehues. 2023. **Automatic speech translation in the classroom and lecture setting: Challenges, approaches, and future directions**. In Gloria Corpas Pastor and Bart Defrancq, editors, *IVITRA Research in Linguistics and Literature*, volume 37, pages 241–276. John Benjamins Publishing Company, Amsterdam.
- Daniel Liebling, Katherine Heller, Samantha Robertson, and Wesley Deng. 2022. **Opportunities for Human-centered Evaluation of Machine Translation Systems**. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 229–240, Seattle, United States. Association for Computational Linguistics.
- Daniel J. Liebling, Michal Lahav, Abigail Evans, Aaron Donsbach, Jess Holbrook, Boris Smus, and Lindsey Boran. 2020. **Unmet Needs and Opportunities for Mobile Translation AI**. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, pages 1–13, New York, NY, USA. Association for Computing Machinery.
- Hajin Lim, Dan Cosley, and Susan R. Fussell. 2018. **Beyond Translation: Design and Evaluation of an Emotional and Contextual Knowledge Interface for Foreign Language Social Media Posts**. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, pages 1–12, New York, NY, USA. Association for Computing Machinery.
- Hajin Lim, Dan Cosley, and Susan R. Fussell. 2022. **Understanding Cross-lingual Pragmatic Misunderstandings in Email Communication**. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1):129:1–129:32.
- Donghui Lin, Yoshiaki Murakami, Toru Ishida, Yohei Murakami, and Masahiro Tanaka. 2010. Composing Human and Machine Translation Services: Language Grid for Improving Localization Processes. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Jessy Lin, Geza Kovacs, Aditya Shastry, Joern Wuebker, and John DeNero. 2022. **Automatic Correction of Human Translations**. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 494–507, Seattle, United States. Association for Computational Linguistics.
- Ting Liu, Chi-ku Lo, Elizabeth Marshman, and Rebecca Knowles. 2024. Evaluation Briefs: Drawing on Translation Studies for Human Evaluation of MT. In *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 190–208, Chicago, USA. Association for Machine Translation in the Americas.
- Arle Lommel, Hans Uszkoreit, and Aljoscha Burchardt. 2014. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació*, (12):455–463.
- Marianna Martindale and Marine Carpuat. 2018. Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 13–25, Boston, MA. Association for Machine Translation in the Americas.
- Marianna J. Martindale and Marine Carpuat. 2022. A Proposed User Study on MT-Enabled Scanning. In *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 2: Users and Providers Track and Government Track)*, pages 377–393.
- Nikita Mehandru, Sweta Agrawal, Yimin Xiao, Ge Gao, Elaine Khoong, Marine Carpuat, and Niloufar Salehi. 2023. Physician Detection of Clinical Harm in Machine Translation: Quality Estimation Aids in Reliance and Backtranslation Identifies Critical Errors. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11633–11647, Singapore. Association for Computational Linguistics.
- Nikita Mehandru, Samantha Robertson, and Niloufar Salehi. 2022. Reliable and Safe Use of Machine Translation in Medical Settings. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’22, pages 2016–2025, New York, NY, USA. Association for Computing Machinery.
- Elise Michon, Josep Crego, and Jean Senellart. 2020. Integrating Domain Terminology into Neural Machine Translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3925–3937, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Shachar Mirkin and Jean-Luc Meunier. 2015. Personalized Machine Translation: Predicting Translational Preferences. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2019–2025, Lisbon, Portugal. Association for Computational Linguistics.

- Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. [Model Cards for Model Reporting](#). In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* ’19, pages 220–229, New York, NY, USA. Association for Computing Machinery.
- Itai Mondshine, Tzuf Paz-Argaman, and Reut Tsarfaty. 2025. Beyond English: The Impact of Prompt Translation Strategies across Languages and Tasks in Multilingual LLMs. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 1331–1354, Albuquerque, New Mexico. Association for Computational Linguistics.
- Joss Moorkens. 2022. Ethics and machine translation. In *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*, 2022, ISBN 978-3-98554-045-7, Págs. 121-140, pages 121–140. Language Science Press.
- Joss Moorkens, Sheila Castilho, Federico Gaspari, Antonio Toral, and Maja Popović. 2024. [Proposal for a Triple Bottom Line for Translation Automation and Sustainability: An Editorial Position Paper](#). *JoS-Trans: The Journal of Specialised Translation*, (41):2–25.
- Jeremy Munday, Sara Ramos Pinto, and Jacob Blakesley. 2022. [Introducing Translation Studies: Theories and Applications](#), 5 edition. Routledge, London.
- Suraj Nair, Eugene Yang, Dawn Lawrie, Kevin Duh, Paul McNamee, Kenton Murray, James Mayfield, and Douglas W. Oard. 2022. [Transfer Learning Approaches for Building Cross-Language Dense Retrieval Models](#). In Matthias Hagen, Suzan Verberne, Craig Macdonald, Christin Seifert, Krisztian Balog, Kjetil Nørvåg, and Vinay Setty, editors, *Advances in Information Retrieval*, volume 13185, pages 382–396. Springer International Publishing, Cham.
- Peter Newmark. 1988. *A Textbook of Translation*. New York Prentice Hall.
- Xing Niu, Marianna Martindale, and Marine Carpuat. 2017. [A study of style in machine translation: Controlling the formality of machine translation output](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2814–2819, Copenhagen, Denmark. Association for Computational Linguistics.
- Lucas Nunes Vieira. 2024. Uses of AI Translation in UK Public Service Contexts | CIOL (Chartered Institute of Linguists). <https://www.ciol.org.uk/ai-translation-uk-public-services>.
- Lucas Nunes Vieira, Carol O’Sullivan, Xiaochun Zhang, and Minako O’Hagan. 2022. [Privacy and everyday users of machine translation](#). *Translation Spaces*, 12(1):21–44.
- Mary Nurminen. 2016. Machine Translation-Mediated Interviewing as a Method for Gathering Data in Qualitative Research: A Pilot Project. In *New Horizons in Translation Research and Education* 4, pages 66–84. University of Eastern Finland.
- Mary Nurminen. 2019. Decision-making, Risk, and Gist Machine Translation in the Work of Patent Professionals. In *Proceedings of the 8th Workshop on Patent and Scientific Literature Translation*, pages 32–42, Dublin, Ireland. European Association for Machine Translation.
- Mary Nurminen. 2020. [Raw machine translation use by patent professionals: A case of distributed cognition](#). *Translation, Cognition & Behavior*, 3(1):100–121.
- Mary Nurminen. 2021a. *Investigating the Influence of Context in the Use and Reception of Raw Machine Translation*. Tampere University.
- Mary Nurminen. 2021b. Machine Translation Stories. <https://mt-stories.com/>.
- Mary Nurminen and Niko Papula. 2018. Gist MT Users: A Snapshot of the Use and Users of One Online MT Tool. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, pages 219–228, Alicante, Spain.
- Sharon O’Brien. 2024. [Human-Centered augmented translation: Against antagonistic dualisms](#). *Perspectives*, 32(3):391–406.
- Sharon O’Brien and Maureen Ehrensberger-Dow. 2020. [MT Literacy—A cognitive view](#). *Translation, Cognition & Behavior*, 3(2):145–164.
- Sharon O’Brien and Federico Marco Federici. 2019. [Crisis translation: Considering language needs in multilingual disaster settings](#). *Disaster Prevention and Management: An International Journal*, 29(2):129–143.
- Sharon O’Brien, Michel Simard, and Marie Josée Goulet. 2018. Machine Translation and Self-post-editing for Academic Writing Support: Quality Explorations. In *Translation Quality Assessment: From Principles to Practice*, 2018, ISBN 978-3-030-08206-2, Págs. 237-262, pages 237–262. Springer Suiza.
- Pilar Ortega, Natalie Felida, Santiago Avila, Sarah Conrad, and Michael Dill. 2023. [Language Profile of the US Physician Workforce: A Descriptive Study from a National Physician Survey](#). *Journal of General Internal Medicine*, 38(4):1098–1101.
- Masashi Oshika, Makoto Morishita, Tsutomu Hirao, Ryohei Sasano, and Koichi Takeda. 2024. [Simplifying Translations for Children: Iterative Simplification Considering Age of Acquisition with LLMs](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8567–8577, Bangkok, Thailand. Association for Computational Linguistics.

- Sharon O'Brien, Michel Simard, and Marie-Josée Goulet. 2018. Machine translation and self-post-editing for academic writing support: Quality explorations. *Translation quality assessment: From principles to practice*, pages 237–262.
- Mei-Hua Pan and Hao-Chuan Wang. 2014. Enhancing machine translation with crowdsourced keyword highlighting. In *Proceedings of the 5th ACM International Conference on Collaboration across Boundaries: Culture, Distance & Technology*, CABS ’14, pages 99–102, New York, NY, USA. Association for Computing Machinery.
- Ziqian Peng, Rachel Bawden, and François Yvon. 2024. À propos des difficultés de traduire automatiquement de longs documents. In *Actes de la 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1 : articles longs et prises de position*, pages 2–21, Toulouse, France. ATALA and AFPC.
- Daniela Petrelli, Stephen Levin, Micheline Beaulieu, and Mark Sanderson. 2006. Which user interaction for cross-language information retrieval? Design issues and reflections. *J. Am. Soc. Inf. Sci. Technol.*, 57(5):709–722.
- Hanna Pięta and Susana Valdez. 2024. Migration and translation technologies. In *The Routledge Handbook of Translation and Migration*. Routledge.
- Mondheera Pituxcoosuvarn, Yohei Murakami, Donghui Lin, and Toru Ishida. 2020. Effect of Cultural Misunderstanding Warning in MT-Mediated Communication. In *Collaboration Technologies and Social Computing : 26th International Conference, CollabTech 2020, Tartu, Estonia, September 8–11, 2020, Proceedings*, pages 112–127, Berlin, Heidelberg. Springer-Verlag.
- José Pombal, Sweta Agrawal, Patrick Fernandes, Emmanuel Zaranis, and André F. T. Martins. 2024. A Context-aware Framework for Translation-mediated Conversations. *Preprint*, arXiv:2412.04205.
- Maja Popović. 2020. Relations between comprehensibility and adequacy errors in machine translation output. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 256–264, Online. Association for Computational Linguistics.
- Anthony Pym. 2012. *On Translator Ethics: Principles for Mediation between Cultures*. John Benjamins.
- Ella Rabinovich, Shachar Mirkin, Raj Nath Patel, Lucia Specia, and Shuly Wintner. 2016. Personalized Machine Translation: Preserving Original Author Traits. *arXiv:1610.05461 [cs]*.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763. PMLR.
- Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust Speech Recognition via Large-Scale Weak Supervision. *Preprint*, arXiv:2212.04356.
- Pavithra Rao, Lauren M. McGee, and Casey A. Seidenman. 2024. A Comparative assessment of ChatGPT vs. Google Translate for the translation of patient instructions. *Journal of Medical Artificial Intelligence*, 7(0).
- Georg Rehm. 2023. European Language Grid: Introduction. In Georg Rehm, editor, *European Language Grid: A Language Technology Platform for Multilingual Europe*, pages 1–10. Springer International Publishing, Cham.
- Elijah Rippeth, Sweta Agrawal, and Marine Carpuat. 2022. Controlling Translation Formality Using Pre-trained Multilingual Language Models. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 327–340, Dublin, Ireland (in-person and online). Association for Computational Linguistics.
- Samantha Robertson. 2023. *Designing for Reliability in Algorithmic Systems*. Ph.D. thesis, University of California, Berkeley.
- Samantha Robertson, Wesley Deng, Timnit Gebru, Margaret Mitchell, Daniel Liebling, Michal Lahav, Katherine Heller, Mark Diaz, Samy Bengio, and Niloofar Salehi. 2021. Three Directions for the Design of Human-Centered Machine Translation. Technical report, Google Research.
- Samantha Robertson and Mark Díaz. 2022. Understanding and Being Understood: User Strategies for Identifying and Recovering From Mistranslations in Machine Translation-Mediated Chat. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’22, pages 2223–2238, New York, NY, USA. Association for Computing Machinery.
- Samantha Robertson, Zijie J. Wang, Dominik Moritz, Mary Beth Kery, and Fred Hohman. 2023. Angler: Helping Machine Translation Practitioners Prioritize Model Improvements. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI ’23, pages 1–20, New York, NY, USA. Association for Computing Machinery.
- Douglas Robinson. 2014. *Translation and Empire*. Routledge, London.
- Sougata Saha, Saurabh Kumar Pandey, Harshit Gupta, and Monojit Choudhury. 2025. Reading between the Lines: Can LLMs Identify Cross-Cultural Communication Gaps? In *Proceedings of the 2025 Conference*

- of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8043–8067, Albuquerque, New Mexico. Association for Computational Linguistics.
- Sandra Sandoval, Jieyu Zhao, Marine Carpuat, and Hal Daumé III. 2023. **A Rose by Any Other Name would not Smell as Sweet: Social Bias in Names Mistranslation**. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3933–3945, Singapore. Association for Computational Linguistics.
- Sebastin Santy, Kalika Bali, Monojit Choudhury, Sandipan Dandapat, Tanuja Ganu, Anurag Shukla, Jahanvi Shah, and Vivek Seshadri. 2021. **Language Translation as a Socio-Technical System: Case-Studies of Mixed-Initiative Interactions**. In *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS ’21*, pages 156–172, New York, NY, USA. Association for Computing Machinery.
- Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. **Gender Bias in Machine Translation**. *Transactions of the Association for Computational Linguistics*, 9:845–874.
- Beatrice Savoldi, Sara Papi, Matteo Negri, Ana Guerberof-Arenas, and Luisa Bentivogli. 2024. **What the Harm? Quantifying the Tangible Impact of Gender Bias in Machine Translation with a Human-centered Study**. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18048–18076, Miami, Florida, USA. Association for Computational Linguistics.
- Beatrice Savoldi, Alan Ramponi, Matteo Negri, and Luisa Bentivogli. 2025. **Translation in the Hands of Many: Centering Lay Users in Machine Translation Interactions**. *Preprint*, arXiv:2502.13780.
- Randy Scansani, Silvia Bernardini, Adriano Ferraresi, and Luisa Bentivogli. 2019. Do translator trainees trust machine translation? An experiment on post-editing and revision. In *Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks*, pages 73–79, Dublin, Ireland. European Association for Machine Translation.
- Federica Scarpa. 2020. **Research and Professional Practice in Specialised Translation**. Palgrave Macmillan UK, London.
- Carolina Scarton and Lucia Specia. 2016. A Reading Comprehension Corpus for Machine Translation Evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 3652–3658, Portorož, Slovenia. European Language Resources Association (ELRA).
- Vincent Segonne, Aidan Mannion, Laura Alonzo-Canul, Audibert Alexandre, Xingyu Liu, Cécile Macaire, Adrien Pupier, Yongxin Zhou, Mathilde Aguiar, Félix Herron, Magali Norré, Massih-Reza Amini, Pierrette Bouillon, Iris Eshkol Taravella, Emmanuelle Esparança-Rodier, Thomas François, Lorraine Goeuriot, Jérôme Goulian, Mathieu Lafourcade, and 7 others. 2024. Jargon : Une suite de modèles de langues et de référentiels d'évaluation pour les domaines spécialisés du français. In *Actes de la 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 2 : traductions d'articles publiés*, pages 9–10, Toulouse, France. ATALA and AFPC.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. **Controlling Politeness in Neural Machine Translation via Side Constraints**. pages 35–40. Association for Computational Linguistics.
- Ruikang Shi, Alvin Grissom, II, and Duc Minh Trinh. 2022. Rare but severe neural machine translation errors induced by minimal deletion: An empirical study on Chinese and English. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 5175–5180, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Ben Shneiderman. 2022. **Human-Centered AI**. Oxford University Press, Oxford, New York.
- Dimitar Shterionov and Eva Vanmassenhove. 2023. **The Ecological Footprint of Neural Machine Translation Systems**. In Helena Moniz and Carla Parra Escartín, editors, *Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation*, pages 185–213. Springer International Publishing, Cham.
- Michel Simard. 2024. Position Paper: Should Machine Translation be Labelled as AI-Generated Content? In *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 119–129, Chicago, USA. Association for Machine Translation in the Americas.
- Inguna Skadina, Andrejs Vasiljevs, Mārcis Pinnis, Aivars Bērziņš, Nora Aranberri, Joachim Van den Bogaert, Sally O’Connor, Mercedes García-Martínez, Iakes Goenaga, Jan Hajič, Manuel Herranz, Christian Lieske, Martin Popel, Maja Popović, Sheila Castilho, Federico Gaspari, Rudolf Rosa, Riccardo Superbo, and Andy Way. 2023. **Deep Dive Machine Translation**. In Georg Rehm and Andy Way, editors, *European Language Equality: A Strategic Agenda for Digital Language Equality*, pages 263–287. Springer International Publishing, Cham.
- Hervé Speechbach, Johanna Gerlach, Sanae Mzouri Karker, Nikos Tsourakis, Christophe Combescure, and Pierrette Bouillon. 2019. **A Speech-Enabled Fixed-Phrase Translator for Emergency Settings: Crossover Study**. *JMIR Medical Informatics*, 7(2):e13167.
- Lucia Specia, Stella Frank, Khalil Sima'an, and Desmond Elliott. 2016. **A Shared Task on Multi-modal Machine Translation and Crosslingual Image**

- Description.** In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 543–553, Berlin, Germany. Association for Computational Linguistics.
- Lucia Specia, Carolina Scarton, Gustavo Henrique Paetzold, and Graeme Hirst. 2018. *Quality Estimation for Machine Translation*. Morgan & Claypool Publishers.
- Neha Srikanth and Junyi Jessy Li. 2021. *Elaborative Simplification: Content Addition and Explanation Generation in Text Simplification*. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 5123–5137, Online. Association for Computational Linguistics.
- Sanja Štajner and Maja Popović. 2019. *Automated Text Simplification as a Preprocessing Step for Machine Translation into an Under-resourced Language*. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1141–1150, Varna, Bulgaria. INCOMA Ltd.
- Marianne Starlander, Pierrette Bouillon, Nikos Chatzichrisafis, Marianne Santaholma, Manny Rayner, Beth Ann Hockey, Hitoshi Isahara, Kyoko Kanzaki, and Yukie Nakao. 2005. Practicing Controlled Language through a Help System integrated into the Medical Speech Translation System (Med-SLT). In *Proceedings of Machine Translation Summit X: Papers*, pages 188–194, Phuket, Thailand.
- Emma Steigerwald, Valeria Ramírez-Castañeda, Débora Y C Brandt, András Báldi, Julie Teresa Shapiro, Lynne Bowker, and Rebecca D Tarvin. 2022. *Overcoming Language Barriers in Academia: Machine Translation Tools and a Vision for a Multilingual Future*. *BioScience*, 72(10):988–998.
- Umut Sulubacak, Ozan Caglayan, Stig-Arne Grönroos, Aku Rouhe, Desmond Elliott, Lucia Specia, and Jörg Tiedemann. 2020. *Multimodal machine translation through visuals and speech*. *Machine Translation*, 34(2):97–147.
- Breena R. Taira, Vanessa Kreger, Aristides Orue, and Lisa C. Diamond. 2021. *A Pragmatic Assessment of Google Translate for Emergency Department Instructions*. *Journal of General Internal Medicine*, 36(11):3361–3365.
- Kristiina Taivalkoski-Shilov. 2019. *Ethical issues regarding machine(-assisted) translation of literary texts*. *Perspectives*, 27(5):689–703.
- NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Hefernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, and 20 others. 2022. *No Language Left Behind: Scaling Human-Centered Machine Translation*. *Preprint*, arXiv:2207.04672.
- Helene Tenzer, Stefan Feuerriegel, and Rebecca Piekari. 2024. *AI machine translation tools must be taught cultural differences too*. *Nature*, 630(8018):820–820.
- Hsing-Lin Tsai and Hao-Chuan Wang. 2015. *Evaluating the Effects of Interface Feedback in MT-embedded Interactive Translation*. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA ’15*, pages 2247–2252, New York, NY, USA. Association for Computing Machinery.
- Susana Valdez, Ana Guerberof Arenas, and Kars Ligtenberg. 2023. Migrant communities living in the Netherlands and their use of MT in healthcare settings. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 325–334, Tampere, Finland. European Association for Machine Translation.
- Susana Valdez and Ana Guerberof-Arenas. 2025. “Google Translate is our best friend here”: A vignette-based interview study on machine translation use for health communication.
- Jennifer Wortman Vaughan and Hanna Wallach. 2021. *A Human-Centered Agenda for Intelligible Machine Learning*.
- Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. *How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies*. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):327:1–327:39.
- Blanca Vidal, Albert Llorens, and Juan Alonso. 2022. Automatic Post-Editing of MT Output Using Large Language Models. In *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 2: Users and Providers Track and Government Track)*, pages 84–106.
- Lucas Nunes Vieira. 2024. Machine translation and migration. In *The Routledge Handbook of Translation and Migration*. Routledge.
- Lucas Nunes Vieira, Minako O’Hagan, and Carol O’Sullivan. 2021. *Understanding the societal impacts of machine translation: A critical review of the literature on medical and legal use cases*. *Information, Communication & Society*, 24(11):1515–1532.
- Lucas Nunes Vieira, Carol O’Sullivan, Xiaochun Zhang, and Minako O’Hagan. 2022. *Machine translation in society: Insights from UK users*. *Language Resources and Evaluation*.
- David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2023. *Prompting PaLM for Translation: Assessing Strategies and Performance*. *Preprint*, arXiv:2211.09102.

- Stefan Mathias Vollmer. 2020. *The Digital Literacy Practices of Newly Arrived Syrian Refugees: A Spatio-Visual Linguistic Ethnography*. Ph.D. thesis, University of Leeds.
- Vivian WangVivian Wang, reporting from behind the Great Firewall, and had an intriguing conversation with DeepSeek’s chatbot. 2025. How Does DeepSeek’s A.I. Chatbot Navigate China’s Censors? Awkwardly. *The New York Times*.
- Xiaolin Wang, Andrew Finch, Masao Utiyama, and Eiichiro Sumita. 2016. A Prototype Automatic Simultaneous Interpretation System. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 30–34, Osaka, Japan. The COLING 2016 Organizing Committee.
- John S. White and Theresa A. O’Connell. 1993. Evaluation of Machine Translation. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.
- Yimin Xiao, Yuewen Chen, Naomi Yamashita, Yuexi Chen, Zhicheng Liu, and Ge Gao. 2024. (dis)placed contributions: Uncovering hidden hurdles to collaborative writing involving non-native speakers, native speakers, and ai-powered editing tools. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2).
- Yimin Xiao, Cartor Hancock, Sweta Agrawal, Nikita Mehandru, Niloufar Salehi, Marine Carpuat, and Ge Gao. 2025. Sustaining Human Agency, Attending to Its Cost: An Investigation into Generative AI Design for Non-Native Speakers’ Language Use. *Preprint*, arXiv:2503.07970.
- Bin Xu, Ge Gao, Susan R. Fussell, and Dan Cosley. 2014. Improving machine translation by showing two outputs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’14*, pages 3743–3746, New York, NY, USA. Association for Computing Machinery.
- Jitao Xu, Josep Crego, and François Yvon. 2023. Integrating Translation Memories into Non-Autoregressive Machine Translation. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1326–1338, Dubrovnik, Croatia. Association for Computational Linguistics.
- Wenbo Xu, Xiangping Chen, and Ping Li. 2024. A Survey on the Application of Online Machine Translation in Thesis Writing by English Majors. In *Proceedings of the 2024 9th International Conference on Distance Education and Learning, ICDEL ’24*, pages 365–370, New York, NY, USA. Association for Computing Machinery.
- Naomi Yamashita and Toru Ishida. 2006. Effects of machine translation on collaborative work. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, CSCW ’06*, pages 515–524, New York, NY, USA. Association for Computing Machinery.
- Naomi Yamashita and Toru Ishida. 2011. Conversational Grounding in Machine Translation Mediated Communication. In Toru Ishida, editor, *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*, pages 183–198. Springer, Berlin, Heidelberg.
- Binwei Yao, Ming Jiang, Tara Bobinac, Dify Yang, and Junjie Hu. 2024. Benchmarking Machine Translation with Cultural Awareness. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13078–13096, Miami, Florida, USA. Association for Computational Linguistics.
- Qinyuan Ye, Juan Zha, and Xiang Ren. 2022. Eliciting and Understanding Cross-Task Skills with Task-Level Mixture-of-Experts. *Preprint*, arXiv:2205.12701.
- François Yvon. 2019. The two paths of machine translation. *Hermès, La Revue*, 85(3):62–68.
- Ran Zhang, Wei Zhao, and Steffen Eger. 2025a. How Good Are LLMs for Literary Translation, Really? Literary Translation Evaluation with Humans and LLMs. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 10961–10988, Albuquerque, New Mexico. Association for Computational Linguistics.
- Yongle Zhang, Dennis Asamoah Owusu, Marine Carpuat, and Ge Gao. 2022. Facilitating Global Team Meetings Between Language-Based Subgroups: When and How Can Machine Translation Help? *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW1):90:1–90:26.
- Yongle Zhang, Phuong-Anh Nguyen-Le, Kriti Singh, and Ge Gao. 2025b. The news says, the bot says: How immigrants and locals differ in chatbot-facilitated news reading. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI ’25*, New York, NY, USA. Association for Computing Machinery.
- Vilém Zouhar, Michal Novák, Matúš Žilinec, Ondřej Bojar, Mateo Obregón, Robin L. Hill, Frédéric Blain, Marina Fomicheva, Lucia Specia, and Lisa Yankovskaya. 2021. Backtranslation Feedback Improves User Confidence in MT, Not Quality. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 151–161, Online. Association for Computational Linguistics.