

Quality in multilingual crisis communication: a reception-oriented perspective

YU HAO

University of Melbourne, Australia

Multilingual crisis communication relies on timely and reliable translation, yet what constitutes a “good” translation remains conceptually underexplored in this high-stakes, culturally loaded field. This position paper reviews previous studies and argues “quality” should be treated not only as a text’s intrinsic property but as real-world communicative success. More specifically, this position paper conceptualizes quality as a series of effect-related parameters: the degree to which intended audiences can access messages in time (“timeliness”), make sense of them (“understandability”), trust them as credible and relevant (“trustworthiness”) and act on them appropriately (“actionability”).

Keywords: crisis communication; behavioral-change emergency messaging; quality assessment; reception-oriented perspective

Introduction

Crisis translation is often regarded as a matter of language justice and equity (e.g., O’Brien 2022). Unequal and delayed access to reliable translated information can exacerbate existing vulnerabilities and create new risks for public safety during health crises and other emergencies including earthquakes, bushfires and floods. Thus, a major challenge faced by multicultural societies in crisis and emergency situations is the need to provide reliable translated messages so that all communities have equitable access to critical information in their own languages. This challenge is particularly acute given that crisis communication constitutes a high-stakes genre whose purpose is not only to inform but also to elicit appropriate behavioral responses in a timely manner; not to mention the linguistic and cultural nuances involved when engaging with divergent communities that experience barriers in accessing official messaging, which is often initially prepared in the society’s dominant language.

Over the years, studies have examined translation in these high-stakes communicative scenarios (e.g., studies reviewed in Federici 2020; Todorova, 2024). To mitigate the challenge of multicultural crisis translation, some studies focus on the training of non-professional citizen translators (e.g., Federici and Cadwell 2018), which is important in situations where professional translation is either unavailable (in the cases of low-resource languages) or overwhelmed due to limited human resources. Bilingual/bicultural members of a community can thus be trained as citizen translators to create locally available and well-adapted services. This challenge is also being met with advances in machine translation and generative AI systems that automate or augment communication across various domains. In this context, recent studies in this field have explored the use of fully automated translation (e.g., Hajek et al. 2024; do Carmo 2025) and semi-autonomous workflows involving both human and non-human agents (e.g., Hao and Pym 2026). Others examine “trust” in multilingual communication, which they view as a layered and socially mediated construct. On this view, linguistic clarity seems to enable initial comprehension while culturally grounded relationships and credible mediators sustain the desirable behavioral changes (e.g., Macreadie et al. 2025).

A central question merits careful analysis: how do we know whether a translated crisis-related message is good enough? Although “quality” is addressed either directly or indirectly in many studies on crisis communication, there remains limited conceptual clarity with respect to what constitutes “good” translation in public emergencies. This becomes increasingly relevant, as translation continues to shift from static text production to dynamic and increasingly iterative human-AI interactions. This position paper, therefore, reviews and examines how “translation quality” has been constructed in crisis messaging, with special attention to the shift from a purely textual evaluation towards a reception-oriented framework.

Translation quality redefined: Why reader response matters?

This section presents a conceptual discussion of the theories over the years, elaborating on what grounds we should judge the “success” of translation and why textual comparison alone could be reductive.

Although the equivalence paradigm generally privileges source-text resemblance, Nida (1964) contends that translation should also aim to produce “the same effect on the reader as the original text did” (Nida 1964: 159). While this dynamic or functional equivalence perspective has been criticized as reductive by those with a *Skopos*-oriented mindset (e.g., Vermeer 1989), since translations may aim to achieve effects different from those of the source text, it nonetheless highlights the need to factor in how target readers might understand and respond to a translation.

As shown in the *Skopos* paradigm (Vermeer 1989), the evaluative center of gravity has shifted away from retrospective textual comparison between translation and its start text, as anchored in the equivalence paradigm (see more detailed discussion in Pym 2023). As *Skopos* theory allows that a single text can be legitimately translated in different ways, each yielding a valid translation (Reiss and Vermeer 1984; Vermeer 1989), “quality” is inherently purpose- and readership-dependent. Thus, quality assessment models that prioritize formal proximity to the original could risk mistaking surface-level mimicry for high quality, even when the intended communicative purpose might not be achieved. Also according to Vermeer’s *Skopos* rules (Reiss and Vermeer 1984), the purposes that could govern all subsequent translation decisions might be defined by various actors who expect value from translation, including the commissioner’s rationale (Nord 2005), the reader’s intentions (Vermeer 1989) and/or what the trained translator judges the purpose to be (Vermeer 1989; Holz-Mänttari 1984). In these views, readers and commissioners are made epistemically important: if they have a say in what translation should be made *for*, then empirical evidence should be gathered on whether translation functions effectively for the end users in the intended context.

Further, norm-based theories, as proposed by Toury (1995/2012) based on sociological concepts developed by Itamar Even-Zohar and earlier sociologists, help explain why readers’ sense of “normal” matters in quality assessment. In descriptive translation studies (DTS), translation is characterized as norm-governed behavior shaped by the conventions and expectations of the target culture (Toury 1995/2012). That is, a translation’s uptake and legitimacy are mediated by shared norms constructed by communities of readers, critics, publishers and other gatekeepers at particular historical moments and within specific cultural contexts. In this sense, any account of “quality” must remain grounded in the cultural and historical conditions under which translations are produced, circulated and interpreted. Moreover, Toury’s polarized distinction between adequate and acceptable translation (Toury 1980) is pertinent here: whereas an adequate translation adheres to source-text norms, an acceptable translation aligns with target-side norms to achieve naturalness and credibility. So, what counts as “good” depends only partly on its relation to the source text and perhaps more predominantly on how it positions itself *vis-à-vis* the recipient culture’s expectations of well-formed and credible texts.

Chesterman (1998) further expands the effects-orientation in Nida’s (1964) position on dynamic equivalence into a conceptual framework which differentiates “proximate (cognitive and affective) effects”, “behavioral effects” and “tertiary effects.” He argues that translation quality should be assessed through empirical investigation of its effects, including its impact on readers’ mental states, subsequent behaviors and macro-level cultural change, such as shifts in norms, genres and ideological repertoires. All of those cannot

be captured through sentence-level comparison alone. Chesterman's three "laws" further emphasize the dynamic status of translation effects: translations may produce different effects on different readers ("law of heterogeneous effect"); effects may vary over time even for the same reader ("law of changing effect"); and translations often generate multiple effects simultaneously ("law of multiple effects") (Chesterman, 1998). If such effects are heterogeneous and temporally unstable, this raises an important question for translation quality assessment: for whom, under what conditions can a translation be considered "good"? Beyond textual analysis, it therefore requires us to examine empirically how people actually understand and appreciate the translated text, as well as the communicative conditions under which translations are received, interpreted and acted upon.

Recent studies on quality in multilingual crisis communication

This section then reviews how translation quality has been conceptualized and assessed in existing literature on multilingual communication, particularly in healthcare and emergency settings. Previous literature shows divergent ontological positions with respect to what constitutes "quality," which in turn give rise to two broad approaches to its empirical investigation: (1) an output-focused approach, grounded in the textual and linguistic analysis of translations and (2) a reception-oriented approach, which prioritizes how recipients understand, interpret, respond to and act upon the translation. Within the first approach, quality is treated as an intrinsic property of the translated text and is often evaluated through manual error analysis of the translated text or automatic evaluation metrics (e.g., BLEU, METEOR). While the reception-oriented approach conceptualizes quality in terms of communicative outcomes and assesses quality through human-centered protocols in addition to textual analysis. This study hastens to add that these two approaches are not mutually exclusive but intertwined. Textual analysis could nevertheless be conducted by the researchers in preparation for eliciting any empirical evidence from the authentic recipients (e.g., Hajek et al. 2024; Hao and Pym 2026). Some reception studies may also invite bilingual (non-)professionals to compare the translation against its original (e.g., O'Brien et al. 2018) or in relation to the target language conventions.

Output-focused quality: textual analysis and error taxonomies

Within the output-focused paradigm, translation quality is conceptualized as a measurable attribute of the translated artefact itself. This approach has been applied to the evaluation of outputs produced from scratch, generated fully autonomously, or developed through hybrid workflows, through the analysis of linguistic features with reference to start-text fidelity and target-language

norms of fluency and naturalness. These analyses of quality seem predominantly negative and reductive, grounded in an error-based, deficit-oriented mindset (e.g., Staiano et al. 2025; Guo 2026). That is, quality is inferred from deviations from an assumed ideal: the presence, severity and distribution of errors (as in Multidimensional Quality Metrics, MQM framework) serve as indicators of lower quality: quality scores decrease as the number and gravity of errors increase. Textual analysis could either be deductively based on existing error taxonomies or proceed inductively through observing recurring error patterns in the texts at hand.

Over the years, this output-focused approach has been adopted to examine the feasibility of machine-assisted crisis translation. For instance, O'Brien et al. (2019) compared three translation modes: from-scratch translation, raw machine translation and post-edited translation. In this study, professional evaluators and bilingual participants were asked to assess these outputs in terms of adequacy (adherence to source meaning) and fluency (linguistic well-formedness), often supplemented by error annotation procedures. The aim was to identify systematic errors in machine translation in crisis contexts, particularly for low-resource language pairs.

Similarly, Pym et al. (2022) conducted a detailed textual analysis of raw machine translation in public health communication settings. Their study highlights the risks of using raw machine output in high-stakes scenarios, which identified recurrent error patterns such as untranslated text in images, omissions and ambiguity in context-dependent terms and syntactic structures. On top of these studies, this output-focused mindset also seems central to the training of bilingual mediators in crisis settings. For example, Federici et al. (2019) highlight the need for both professional and citizen mediators to maintain terminological consistency, domain-specific accuracy and stylistic appropriateness in their output, particularly when dealing with risk-sensitive content.

However, the output-focused perspective can potentially be problematic, as it tends to treat semantic accuracy and naturalness alone as the primary indicators of quality. In this sense, this reductive view seems to equate quality with communicative success in crisis messaging, assuming that textual fidelity and naturalness will always lead to clear understanding and desirable behavioral changes among its recipients. Researchers in this area have increasingly recognized the limitations of relying solely on this approach for quality assessment. As O'Brien and Federici (2020) argue, crisis communication often involves multimodal, time-sensitive and culturally embedded messages, in which linguistic correctness alone may not be sufficient to guarantee effective communication.

Reception-oriented quality: Understandability, actionability, trustworthiness and timeliness

A growing body of research has advocated for a reception-oriented perspective to translation quality (e.g., O'Brien et al. 2018; Hajek et al. 2022). Within this paradigm, quality is no longer merely treated as an inherent attribute of the translated text, but also as an emergent outcome from the interactions between the text, its prospective readers and the broader communicative context.

Recent studies reflect a broader epistemological shift towards evaluating quality as a context-dependent phenomenon, grounded in user experience and communicative outcomes. Beyond textual analysis, these reception studies all empirically examine how culturally, socially and age-heterogeneous groups understand and respond to emergency messages in various settings, although each has paid particular attention to certain aspects. For instance, O'Brien and Cadwell (2017) conducted a series of comprehension tests in a case study in Kenya, to assess the extent to which original and translated messages are understood by prospective readers. The translated public health messaging in Kiswahili was found more likely to elicit desirable sense-making than the English original.

In parallel, many have argued that in crisis contexts, where timely and accurate action can be lifesaving, the translation's ability to elicit desirable behavioral changes should also be prioritized. For instance, Ogie and Perez (2020) compared community feedback on intended behavioral responses to texts translated by professional translators and citizen mediators. Their findings indicate that professionally translated texts may fail when the source text itself is ambiguous or overly technical. Responses to translated emergency messages depend largely on clarity, simplicity and cultural resonance, areas in which citizen translation seems to outperform its counterpart.

Recent studies have also factored in time-based evaluation, measured in seconds per word and task completion time. These measures are particularly pertinent in crisis and public health contexts, where timely messaging could profoundly influence communicative impact. Working with automated translation (such as post-editing machine or AI output) has traditionally been associated with speed gains compared with from-scratch translation (e.g., O'Brien 2011; Plitt and Masselot 2010). However, previous studies (e.g., Pym et al. 2022) also showed that raw automated output should not be used in crisis translation despite the faster turnaround, given that even if the output is 99% accurate, the remaining 1% of errors could potentially lead to fatal consequences during emergencies. In other words, speed should not be traded off against reliable, trustworthy messages and overall communicative effectiveness.

Lastly, this shift in focus also emphasizes trustworthiness as an evaluative dimension, particularly in high-stakes contexts where credibility can potentially shape how recipients respond to the message. For instance, Pym et al. (2022) focused on this dimension in a COVID-19 public messaging context, using qualitative and observational methods. Trust is inferred indirectly from cases of distrust, such as translation errors, delayed or inaccessible messaging and from communities' reliance on alternative mediators or information sources other than the official government translation. Macreadie et al. (2025) further argue that translation is not only about conveying an accurate message but about building trust between institutions and audiences in the context of cross-cultural vaccination messaging. The study distinguishes between "thin trust" (based on surface accuracy and institutional authority) and "thick trust" (grounded in cultural relevance and social relationships) and reports that most institutional translations achieve only the former. That is, even linguistically accurate translations can sometimes fail in public health messaging, especially if audiences do not perceive the text as culturally trustworthy. More recent studies in this field on human-centered AI also introduce evaluation criteria such as trustworthiness, safety and user control (e.g., Briva-Iglesias and O'Brien 2026).

Drawing on the literature, this position paper argues that, beyond linguistic correctness, translation quality in crisis settings should be evaluated in terms of its "effects" along four dimensions. Each dimension should be empirically tested with authentic message recipients in context.

- Understandability: the degree to which receivers can make sense of the text (e.g., in O'Brien and Cadwell 2017; Rossetti 2019; Alexander and Pescaroli 2020; Ogie and Perez 2020);
- Actionability: the extent to which receivers are attentive to information and make desirable behavioral changes (e.g., Sengupta et al. 2024)
- Trustworthiness: whether receivers perceive the message as credible (e.g., O'Brien 2022; Pym et al. 2022; Seale et al. 2024; Macreadie et al. 2025) and lastly
- Timeliness: whether receivers can access up-to-date information through convenient channels in a timely manner (O'Mara and Carey 2019; Hajek et al. 2024).

These four dimensions are not mutually exclusive and can interact with one another in real-life reception. That said, any causal relationships among them should not be assumed. For instance, highly readable texts may not necessarily generate greater perceptions of trustworthiness (e.g., Macreadie et al. 2025); while recipients may still change their behavior during emergencies even if they do not fully trust a translation, particularly when no alternative sources of information are available (Rajkhowa et al. 2025). Further empirical evidence is needed to understand how these dimensions interact dynamically

with one another, which would also help conceptualize what constitutes a high-quality translation in crisis settings. Although previous studies have focused on one or more of these dimensions, the present model does not assume that any single dimension should be prioritized over the others without empirical justification. In addition, the recipients' prior knowledge about emergencies, their stylistic expectations of warnings and safety instructions and their trust in the authoritative source of messages could also shape how the translated message is received.

A range of human-centered measures have been used to assess quality in relation to one or more of the four dimensions, including

- Relative preference evaluation, in which bilingual recipients compare and rank different versions of translation (of the same start texts). This method can elicit intuitive user responses to translation quality, often prompted via questions like “Which text do you find most satisfactory to read?” and “Which text would you prefer to recommend to your friends and family?” This method can be used to compare how effective different translation workflows are and the quality of different versions of automated crisis translations.
- Direct assessment, in which authentic users exercise a scalar judgement of the translation, on a scale commonly from 0 to 10. The participants may not have access to the start text or alternative translations. This method could be used to measure all four quality parameters as described above.
- Task-specific comprehension questionnaires provide a measure to explicitly test users' understanding of crisis-related information. This could be particularly useful to flag cases in which “official language” fails to meet the linguistic and cultural norms of the affected populations. Similarly, questions can be designed also to focus on actionability, such as “does it clearly specify what to do or avoid?”
- In the same vein, task-based gap-filling tasks also seek to examine the actual use of translation in terms of successful information uptake, but perhaps within a more controlled setting. It requires participants to restore missing information in a reference text based on the available translation.

Alongside these methods, qualitative approaches, including individual surveys and interviews, focus-group discussions and community quality validation workshops remain powerful ways to yield rich insights into user perceptions of trustworthiness and cultural appropriateness.

Empirical studies conducted by Melbourne research team

This section reviews a series of research initiatives undertaken in Melbourne, Australia, between 2020 (during the COVID-19 pandemic) and 2024 (when Anthony Pym started to take on a more honorary role at the University of Melbourne). These projects focused on multilingual communication aimed at engaging Culturally and Linguistically Diverse (CALD) communities in the context of COVID-19 and other emergencies, such as bushfires and floods. Anthony was the “brain” behind all these initiatives. The focus here is not on seeking advice from experienced translators or accreditation authorities on translation quality, but on hearing from the authentic recipients of the translated public message (i.e., members across various CALD communities) about the potential effects of translated material on them. I was first involved in the COVID-19 project during my PhD candidature to assist with data collection and analysis; and from this project, the empirical data have revealed diversified preferences across various communities, offering more nuanced and contextualized insights than accuracy and grammatical correctness on the text level. Later, I was further involved more closely in the research design and research output preparation in the two subsequent emergency communication projects. I share with the team the ethos that a text is not complete until it has been interpreted. In the field of crisis translation, it is particularly important to examine how authentic readers interpret and respond to the message alongside textual analysis.

These initiatives provide a longitudinal empirical account of how translation quality has been reconceptualized in practice. Some shared features of these projects include: 1) “quality” is situated within a broader socio-cultural context as a communicative success, i.e., the prospective audience have timely access to the translated message that is comprehensible, trustworthy and actionable; 2) non-experts are often recognized as epistemic authorities across the projects, i.e., CALD community organizations, leaders and members; 3) the most advanced technologies available at each point, from neural machine translation to ChatGPT-4omini, are tested in the context of multilingual public messaging.

Early pandemic insights (2020-2022, Hajek et al. 2022; Sengupta et al. 2024)

During the COVID-19 pandemic, the Melbourne research team investigated how CALD communities accessed and evaluated the government-issued translation of public health materials (Hajek et al. 2022). This work situates “quality” within the lived experiences of CALD communities, providing some of the first evidence of community-led and machine-assisted translation practices in crisis settings.

More specifically, interviews were conducted with 58 community members and leaders from African, Pasifika, Afghan, Myanmar, Indian

subcontinent, Arabic and Chinese communities that were resident in Victoria, Australia. Our participants, although being heterogeneous in terms of knowledge expertise, all held a clearly defined role in the community, e.g. represent a community organization, advocate for the community and/or being widely perceived as important people with high standing in their community. Through these interviews, we explore how participants and their communities have followed and received information about COVID-safe behavior, testing and vaccination; we also asked about their level of satisfaction and to what extent they trust the government-issued materials. These questions were then followed by a series of quality evaluation tasks, in which community feedback was collected on authentic materials, e.g., official pamphlets and posters on “How to Handwash” and “After Vaccination” available in public spaces and online.

Here, quality is explicitly defined in terms of whether multilingual healthcare messaging enables users to access, understand, trust and act upon behavior-change messaging in a timely manner (Hajek et al. 2022; Sengupta et al. 2024). It is evaluated by community participants in terms of ease of access, frequency of updates, message completeness and relevance. Quality is signaled by whether translated messages enable recipients to engage effectively with information, especially messages related to vaccination.

For context, as we know, Melbourne was under a nearly nine-month lockdown, making it one of the cities that experienced the longest COVID-19 lockdowns worldwide. In the later stages of the pandemic, the government introduced policies that associated vaccination rates with the easing of lockdown restrictions. Part of the reason lies in the various conspiracy theories and misinformation related to vaccination. In addition to concerns about potential health impacts, vaccination is viewed negatively in some communities. These beliefs include claims that vaccines contain microchips, vaccines alter DNA, and lockdowns and vaccination campaigns were simply implemented as mechanisms of government control. Consequently, the government made a significant effort in seeking effective ways to engage (non-)English-speaking communities for vaccination. During the interviews, most participants whose community relied on translation for information access reported that official translations were often delayed, linguistically inconsistent (mixing two languages), or culturally misaligned (Hajek et al. 2022). For instance, many community participants flagged the official translation of medical information about vaccines and antibodies as overly technical and thus difficult for them and for their people to follow, as these communities were mainly refugees and had relatively lower educational and literacy levels. In other cases, participants complained that the steps described in the official translation for booking vaccination appointments, receiving the vaccine and/or staying safe afterwards were so confusing that they did not know what to do. Some images and colors on the official pamphlets were also perceived as culturally inappropriate or offensive. For instance, several

Pacific communities flagged the issue that using real people's photos in the vaccination jab pamphlet may scare their communities and that cartoons should be used instead; red and black were reported as inappropriate by a few African communities, as these colors may cause negative psychological impacts. As a result, bilingual community leaders and organizations began producing timely, culturally appropriate materials for local dissemination via social media platforms such as Twitter and WhatsApp groups; Google Translate was also reported to be used in some cases.

At the same time, "quality" is also found to be intertwined with trust, which depends not only on textual parameters, i.e., grammatical correctness and idiomatic expression, but also on the perceived credibility of sources (in our case, Victorian Government) and mediators, e.g., NAATI-accredited translators or bilingual community members. Interestingly, we also observed several trust-based strategies adopted by the government in the later stages to boost vaccination rates. These strategies re-positioned mediators and sources of information as originating from within the community itself, including efforts to engage bicultural healthcare workers to produce tailored vaccination promotion videos for their own LOTE (languages other than English)-community and to mobilize community or religious leaders to disseminate vaccine-related information and encourage uptake, via community engagement and outreach initiatives to promote vaccination awareness and public-health messaging.

Our study also argues for space for legitimate "transcreation" (Sengupta et al. 2024), understood as the adaptation of messages to specific audiences, including additions, omissions, reformatting and changes in layout and visual design, all aimed at enhancing the effects of the translated message on its authentic recipients. That said, whether transcreation or literal translation works better should still be tested empirically with authentic community participants. Quality here emerges as relational and context-dependent and in many cases should be co-constructed with its authentic recipient communities.

Machine translation for emergency messaging (2023-2024, Hajek et al. 2024)

The urgency embedded in emergency communication that requires timely responses motivates us to explore the potential of neural machine translation in this domain. Quality is evaluated across a range of machine-assisted workflows and is explicitly conceptualized in relation to "actionability" in our emergency messaging project, i.e., the ability of translated messages to prompt appropriate behavioral responses among recipients (Hajek et al. 2024). Although having "actionability" underlined, urgency and "timeliness" are nevertheless embedded in certain types of emergency communication that requires rapid response, i.e., emergency alerts and crisis broadcasts. Similarly, "understandability" and "trustworthiness" were also tested to understand whether community participants comprehended the intended instructions and

whether they would recommend the translated message to their family, friends and community.

Adopting a reception-oriented perspective, the study examines three translation conditions: raw machine translation, machine translation post-edited by a human (post-edited) and machine translation of pre-edited source texts (pre-edited). The output of each of these workflows was empirically evaluated with members of Spanish-, Greek-, Chinese- and Dari-speaking communities in Melbourne through individual semi-focused interviews. Participants were also invited to complete comprehension tasks after reading the raw machine translation and to indicate their preferences after reading all three versions.

The results of comprehension tests show that most participants across four language groups were misled by ambiguities and distortions in automated translation. Many participants also flagged counter intuitive machine translations that back translate into English as “park your car next to the fire” or “if your car is on fire, park off the road behind a solid structure” (Hajek et al. 2024: 36), which creates risks to public safety and may have serious, even fatal, consequences. Similarly, when providing preference evaluations, none of the participants chose raw machine translation as the preferred option. Taken together, despite clear speed gains, raw machine translation without “human touch” should not be used for emergency alerts and safety instructions. On the other hand, post-edited and pre-edited versions were consistently viewed as more accessible and easier to understand; participants reported that they were more likely to follow the instructions in these versions and share them with family and community members.

This study further forges links between linguistic features and actionability: clearer, more explicit and syntactically simpler translations are more likely to be perceived as easy-to-understand and thus elicit perceived behavioral change. Pre-editing plays an important role in this regard, as it standardizes texts, resolves ambiguities and simplifies syntactic structures. This also explains why the results showed a slight preference among our participants for pre-edited outputs over post-edited ones. At the same time, post-editing processes that used cultural adaptation seem to enhance the perceived legitimacy and trustworthiness of the message source.

Taken together with time-based evaluation, the findings indicate that raw machine translation is faster than pre-editing (particularly when translating into multiple languages), followed by post-editing. Across the four languages, the Dari translators spent the longest time in post-editing machine translation of the same text. The project formed the basis for the following recommendations: (Hajek et al. 2024): 1) the urgency inherent in emergencies justifies the inclusion of automated translation, however, gains in speed associated with fully automated translation should not be traded off against reliability and overall communicative effectiveness (see examples of counterintuitive automated translation); 2) pre-editing that fixes problems in

the ST may offer a more cost-effective solution than post-editing when texts are to be translated into multiple languages; 3) the time required for post-editing tends to increase for less-electronic-resourced languages, such as Dari in our case. Finally, preparatory work conducted prior to emergencies, such as the development of translation memories, glossaries and pre-translated templates, can be used to override raw automated output in time-sensitive situations.

Cultivating a reception-oriented mindset about quality in the GenAI era (2024-2025, Hao and Pym 2026)

Our recent studies on GenAI-translated emergency messaging introduce an explicit pedagogical move toward cultivating a reception-oriented mindset of quality among our translation students (Hao and Pym 2026; Hao and Wang unpublished). A series of studies invited Chinese trainee and novice translators in Melbourne to evaluate emergency-related translations generated by ChatGPT, from the perspectives of their LOTE-speaking (Languages Other than English) communities. When revising ChatGPT initial translations and prompt-based revisions, students were asked: “Is this message easy to understand?” and “Does it allow someone from your community to act on it immediately?”

Guided by these questions, the studies place students in evaluative situations where they must simulate the situational constraints of prospective end users. Likewise, in think-aloud protocol (TAP) tasks, participants were asked whether they would recommend a ChatGPT translation “to a non-English speaking member of your community,” thereby directly embedding reception into the evaluation task. Our findings suggest that participants consistently prioritized Chinese texts (translated from English) with clear subjects and imperative structures, while rejected syntactically complex or ambiguous constructions on the grounds that they would be difficult to process “when people are under stress” during emergencies. One participant explicitly articulated this reception-oriented stance through their complaints about clumsy structures copied literally from the English text that sought to explain the context and rationale behind the evacuation instruction: “The goal is to stay safe... Just tell me what to do!” (Hao and Wang, unpublished).

Interestingly, this perspective also led students to accept non-literal and trans-creative strategies as suggested by ChatGPT, such as generalization (e.g., rendering “do not light a match” as “do not light fire”) and legitimate omissions or additions when these better aligned with the intended behavioral outcome.

At the same time, we argue that reception-oriented evaluation must be paired with a “vigilant, low-trust” attitude (Hao and Pym 2026) towards AI outputs. The classroom experiment shows that over-reliance on GenAI leads students to accept fluent but deeply flawed translations, whereas a vigilant

stance enables them to detect problems and further iterate automated revisions (Hao and Pym 2026). Similarly, participants in the TAP study consistently rejected ChatGPT revisions that introduced omissions which could lead to potentially inappropriate life-saving actions, even when these revisions were stylistically concise and polished (Hao and Wang unpublished). The results of our classroom experiment could possibly have broader implications in (non-)professional crisis translation. Given that this vigilance becomes more than ever required of human in the loop of iterative GenAI workflows, where translation is no longer linear but recursive, involving multiple rounds of prompting, evaluation and/or revision. Translators must therefore continuously decide when to trust, refine, or override AI outputs.

In doing so, students begin to internalize a key principle underpinning our projects: quality is not equivalent to fidelity, but to communicative effectiveness under real-world conditions. In this sense, quality in the GenAI era should no longer be associated merely with a static textual property but with an ongoing, situated and human-centered judgement.

Concluding remarks

This position paper sought to discuss what constitutes a “good” translation in multilingual communication during health crises and disasters. It reviews a longitudinal set of empirical studies conducted in Melbourne, in which “quality” is treated less as an intrinsic property of a translated text and more as related to its effects on prospective recipients under real-world constraints, especially in high-stakes and culturally-loaded contexts. That is, rather than having experienced professionals and accreditation authorities examine the quality of translation against the ST and target-language conventions, quality is ontologically perceived as the degree to which intended audiences can access messages in time (“timeliness”), make sense of them (“understandability”), trust them as credible and relevant (“trustworthiness”) and act on them appropriately (“actionability”). This paper then advocates that, in addition to mono- or bilingual analysis on the text level, quality should also be grounded in user experience and communicative outcomes during crises and emergencies.

A good translation is more than just a correct one, in the sense of accurately reproducing the original meaning and being grammatically error-free and stylistically appropriate. The COVID-19 project demonstrates that community-perceived cultural appropriateness and trust in information sources can substantially influence responses to translated messages. As we have seen, delayed access to official translations that are culturally misaligned and difficult to read may undermine comprehension and behavioral changes, whereas locally mediated translations can enhance community relevance and perceived trustworthiness.

At the same time, machine translation and generative AI systems are reported to have great potential in crisis communication (i.e., speed gains), however, fully automated output that introduces ambiguities and distortions should be avoided as even a small probability of unsafe interpretation may have serious consequences. These increasingly iterative workflows, involving multiple rounds of human-AI interaction, will inevitably require human translators to remain engaged in the loop to exercise situated judgement. More specifically, in crisis settings, humans in the loop could be essential to enhance accountability, recipients' willingness to act upon and share messages within their communities. Achieving this requires a vigilant and low-trust stance towards AI-generated outputs, alongside a deep understanding of the preferences and lived experiences of prospective recipients, including how they process and act upon translated instructions under stress during emergencies.

By repositioning "quality" as relational and situated, this paper calls for quality assessment research that combines textual analysis with user-centered, empirically grounded protocols. Future studies, that said, should test the proposed four dimensions of translation effects across different languages, modalities and institutional settings, to connect reception measures more directly to observable behavioral outcomes.

Declaration of AI use

In preparing this manuscript, the author used ChatGPT5.5 for grammar check. All AI-assisted content was reviewed, edited and verified by the author, who takes full responsibility for the accuracy and integrity of the final text.

References

- Alexander, David E., and Gianluca Pescaroli. 2020. "The role of translators and interpreters in cascading crises and disasters: Towards a framework for confronting the challenges". *Disaster Prevention and Management: An International Journal* 29(2): 144–156.
- Briva-Iglesias, Vincent, and Sharon O'Brien. 2026. "Human-Centered AI Language Technology (HCAILT): An empathetic design framework for reliable, safe and trustworthy multilingual communication". *International Journal of Human-Computer Interaction*: 1–15.
- Cadwell, Patrick, Sharon O'Brien, and Erin DeLuca. 2019. "More than tweets: A critical reflection on developing and testing crisis machine translation technology". *Translation Spaces* 8(2): 300–333.
- Chesterman, Andrew. 1998. "Causes, translations, effects". *Target: International Journal of Translation Studies* 10(2): 201–230.

- do Carmo, Félix. 2025. *Evidence Review Report on the Use of AI for Multilingual Communication in Public Services: With a Specific Focus on the NHS*. Guildford: Centre for Translation Studies, University of Surrey. <https://doi.org/10.15126/901630>.
- Federici, Federico, Minako O'Hagan, Sharon O'Brien, and Patrick Cadwell. 2019. "Crisis translation training challenges arising from new contexts of translation". *Cultus* 12: 246–279.
- Federici, Federico. 2020. "Translation in contexts of crisis". In E. Bielsa and D. Kapsaskis (eds) *The Routledge handbook of translation and globalization*. London and New York: Routledge. 176-189.
- Federici, Federico. 2022. "Translating hazards: Multilingual concerns in risk and emergency communication". *The Translator* 28(4): 375–398.
- Federici, Federico, and Patrick Cadwell. 2018. "Training citizen translators: Design and delivery of bespoke training on the fundamentals of translation for New Zealand Red Cross". *Translation Spaces* 7(1): 20–43.
- Guo, Jing. 2026. *An Empirical Study on Foreign Residents' Expectations of Crisis Translation in China during COVID-19: User-centred Translation Quality as a Multidimensional Quality Metrics Error Typology*. Doctoral thesis. Auckland: The University of Auckland.
- Hajek, John, Maria Karidakis, Riccardo Amorati, Yu Hao, Medha Sengupta, Anthony Pym, and Rita Woodward-Kron. 2022. *Understanding the Experiences and Communication Needs of Culturally and Linguistically Diverse Communities during the COVID-19 Pandemic*. Melbourne: The University of Melbourne. Available at: https://arts.unimelb.edu.au/__data/assets/pdf_file/0020/4071170/Report_CALD_COVID.pdf
- Hajek, John, Anthony Pym, Yu Hao, Maria Karidakis, Ambrin Hasnain, Anila Hasnain, Juerong Qiu, Ke Hu, and Rachel Macreadie. 2024. *Understanding and Improving Machine Translations for Emergency Communications*. Melbourne: The University of Melbourne. DOI: 10.17613/jthe-m639.
- Hao, Yu, and Anthony Pym. forthcoming in 2026. "Can generative AI help students revise translations? Yes, with low, vigilant trust". In I. Feinauer and A. Marais (eds) *The Reality of Revision*. Amsterdam: John Benjamins.
- Hao, Yu, and Yizhou Wang. Unpublished manuscript. "Evaluating GenAI translation for emergency communication: A multimodal think-aloud protocol study".
- Holz-Mänttari, Justa. 1984. *Translatorisches Handeln: Theorie und Methode*. Helsinki: Academia Scientiarum Fennica.
- Macreadie, Rachel, Kadija Bouyzourn, Anthony Pym, and Reine Meylaerts. 2025. "Thick and thin trust in translated cross-cultural vaccination messaging". *Translation Studies* 18(2): 265–282.

- Nida, Eugene. 1964. *Toward a Science of Translating, with Special Reference to Principles and Procedures involved in Bible Translating*. Leiden: E. J. Brill.
- Nord, Christiane. 2005. *Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-oriented Text Analysis*. Amsterdam and Atlanta: Rodopi.
- O'Brien, Sharon. 2011. "Towards predicting post-editing productivity". *Machine Translation* 25(3): 197–215.
- O'Brien, Sharon. 2022. "Crisis translation: A snapshot in time". *InContext* 2(1): 84–108.
- O'Brien, Sharon, and Patrick Cadwell. 2017. "Translation facilitates comprehension of health-related crisis information: Kenya as an example". *Journal of Specialised Translation* 28: 23–51.
- O'Brien, Sharon, and Federico Federici. 2020. "Crisis translation: Considering language needs in multilingual disaster settings". *Disaster Prevention and Management: An International Journal* 29(2): 129–143.
- O'Brien, Sharon, Federico M. Federici, Patrick Cadwell, Jim Marlowe, and Benjamin Gerber. 2018. "Language translation during disaster: A comparative analysis of five national approaches". *International Journal of Disaster Risk Reduction* 31: 627–636.
- O'Brien, Sharon, Michel Simard, and Marie-Josée Goulet. 2018. "Machine translation and self-post-editing for academic writing support: Quality explorations". In J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty (eds) *Translation Quality Assessment: From Principles to Practice*. Cham: Springer International Publishing. 237–262.
- O'Mara, Brigid, and Gemma Carey. 2019. "Do multilingual androids dream of a better life in Australia? Effectiveness of information technology for government translation to support refugees and migrants in Australia". *Australian Journal of Public Administration* 78(3): 449–471.
- Ogie, Rodney I., and Pascal J. Perez. 2020. "Collaborative translation of emergency messages (Co-TEM): An Australian case study". *International Journal of Disaster Risk Reduction* 50: 101920.
- Plitt, Mirko, and François Masselot. 2010. "A productivity test of statistical machine translation post-editing in a typical localisation context". *Prague Bulletin of Mathematical Linguistics* 93(1): 7–16.
- Pym, Anthony. 2023. *Exploring Translation Theories*. London and New York: Routledge.
- Pym, Anthony, Nune Ayvazyan, and Jennifer M. Prioleau. 2022. "Should raw machine translation be used for public-health information? Suggestions for a multilingual communication policy in Catalonia". *Journal of Language Rights & Minorities* 1(1–2): 71–99.
- Reiss, Katharina, and Hans J. Vermeer. 1984. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer. Trans. Christiane Nord as

- Towards a General Theory of Translational Action: Skopos Theory Explained*. Manchester: St. Jerome, 2013.
- Rossetti, Alessandra. 2019. "Intralingual translation and cascading crises: Evaluating the impact of semi-automation on the readability and comprehensibility of health content". In F. Federici and S. O'Brien (eds) *Translation in Cascading Crises*. London: Routledge. 219–240.
- Seale, Holly, Ben Harris-Roxas, Anthony E. Heywood, Ibrahim Abdi, Abdulmajeed Mahimbo, Lisa Woodland, and Elizabeth Waller. 2024. "'It's no use saying it in English': A qualitative study exploring community leaders' perceptions of the challenges and opportunities with translating and interpreting COVID-19 related public health messaging to reach ethnic minorities in Australia". *PLOS One* 19(2): e0284000.
- Sengupta, Medha, Anthony Pym, Yu Hao, John Hajek, Maria Karidakis, Rita Woodward-Kron, and Riccardo Amorati. 2024. "On the transcreation, format and actionability of healthcare translations". *Translation & Interpreting* 16(1): 121–141.
- Staiano, Maria Carmen, Lifeng Han, Johanna Monti, and Francesca Chiusaroli. 2025. "ITALERT: Assessing the quality of LLMs and NMT in translating Italian emergency response text". In *Proceedings of Machine Translation Summit XX: Volume 1*. 566-577.
- Todorova, Miglena. 2024. "Translation and emergency". In C. V. Angelelli (ed) *The Routledge Handbook of Translation and Sociology*. London and New York: Routledge. 383-399.
- Toury, Gideon. 1980. *In Search of a Theory of Translation*. Tel Aviv: Porter Institute for Poetics and Semiotics.
- Toury, Gideon. 1995/2012. *Descriptive Translation Studies – and Beyond*. Revised edition. Amsterdam and Philadelphia: John Benjamins.
- Vermeer, Hans J. 1989. *Skopos und Translationsauftrag*. Heidelberg: Institut für Übersetzen und Dolmetschen.