THE UNIVERSITY OF MELBOURNE

Faculty of Arts

Research Unit for Multilingualism and Cross-cultural Communication (RUMACCC)

Understanding and Improving Machine Translations for Emergency Communications

Understanding and Improving Machine Translations for Emergency Communications

A report to the Victorian Government

June 2024

Authors:

Professor John Hajek
Professor Anthony Pym
Dr Yu Hao
Dr Maria Karidakis
Ms Ambrin Hasnain
Ms Anila Hasnain
Dr Juerong Qiu
Dr Ke Hu
Ms Rachel Macreadie

This report was produced by the Research Unit for Multilingualism and Cross-Cultural Communication (RUMACCC), School of Languages and Linguistics, Faculty of Arts, The University of Melbourne, for the Victorian Department of Families, Fairness and Housing (DFFH).

Acknowledgments

The University of Melbourne acknowledges the support of the Victorian Government for the funding received for this project. The project was supported by the Victorian Government through the Department of Families, Fairness and Housing. We also wish to thank the members of the project steering committee for their advice and feedback, as well as the participants in this study for sharing their insights with us.

Ethics approval of this research project was received from the Human Research Ethics Committees, Faculty of Arts, University of Melbourne (Ethics Reference Number 2024-27953-49410-3).

Table of Contents

1. Executive summary 8

1.1. Project background and purpose 8

1.2. Key findings 10

1.3. Recommendations 11

1.4. A simple guide to writing for machine translation in multilingual emergency communication (pre-editing of source texts) 11

Spell out who has to take an action. 12
Spell out what the action is on. 12
Indicate where things are. 13
Avoid culture-specific terms. 13
Avoid two-part verbs. 14
Avoid elements with multiple grammatical functions. 14
Where possible, avoid terms that can be gratuitously gendered. 14
Use the second person only for actions to be taken. 14
Remove ambiguities that are due to the order of elements in a sentence. 15

2. Overview of the report 17

3. Project background 18

3.1. Policies and guidelines on the use of machine translation in Victoria and beyond. 18

3.2. Disadvantages of machine translation 20

3.3. Advantages of machine translation 21

3.4. Disadvantages and advantages of generative AI 21

3.5. The integration of machine translation into different workflows 22

4. Research design 24

4.1. Initial experiments with training machine-translation systems 24

4.2. Selection of variables 26

4.2.1. Machine translation system 26
4.2.2. Target languages 26
4.2.3. Sample texts 27
4.2.4. Evaluation strategy 28

4.3. Research procedure 28

4.3.1. Phase 1: Post-editing of machine translation errors 28
4.3.2. Phase 2: Pre-editing of the original English-language text 28
4.3.3. Phase 3: Reception of emergency messages 29

5. Results 32

5.1. Actionability of raw machine translations (questionnaires) 32

5.2. Machine-translation correction (post-editing) 35

5.3. The issue of language specificity 38

5.3.1. Machine translation problems specific to Chinese 38
5.3.2. Machine translation problems specific to Greek 39
5.3.3. Machine translation problems specific to Spanish 39
5.3.4. Machine translation problems specific to Dari 40
5.4. English source-text correction (pre-editing) 41
5.5. Evaluations by speakers of the languages 43
5.5.1. Chinese 43
5.5.2. Greek 45
5.5.3. Spanish 46
5.5.4. Dari 47
6. Recommendations 49
6.1. Should raw machine translation be used? 49
6.2. How should emergency messages be written? 50
6.3. When should human translation be used? 50
6.4. How can machine translation be combined with human translation? 50
6.5. What workflows can integrate machine translation? 50
6.6. How should machine translations be labelled? 52
6.7. How should we *not* evaluate machine translation? 52
6.8. Will translation automation improve? 53
Appendices 54
Appendix A: Original and pre-edited source texts 54
Appendix B: Machine translation problems as indicated by post-editing 58
Appendix C: Questions testing the actionability and comprehension of raw machine translations 60
Appendix D: Example of a template for emergency texts 62
References 63

List of Tables

Table 1. Percentages of readers' answers to questions about the raw machine translations (with number of participants in brackets). 34
Table 2. Total time and processing time per word in post-editing machine translations. 36
Table 3. Comparison between original English source text message and a pre-edited version of the same message. 42

List of Figures

Figure 1. Workspace in the translation memory suite Phrase, integrating machine translation, a pre-established glossary, and previous translations (from the translation memory). 26
Figure 2. Translation strategies and steps, depending on urgency. 52

Glossary

Actionability: A text's ability to have the reader take the desired action.

AI (artificial intelligence): Computer technology with the ability to perform and augment tasks that humans can perform. While AI is at work in all electronic translation technologies, the term has nevertheless come to be associated with the use of LLMs (Large Language Models) to perform language tasks based on prompts.

Controlled authoring: The writing of a text following a special set of rules. The rules may be for easy-access reading, the application of standardised terminology and syntax, or to avoid issues known to be problematic for machine translation ('pre-editing').

Generative AI: Artificial intelligence with the ability to perform operations based on in-putted instructions ('prompts') and a very large database; a commonly used term for LLMs (Large Language Models).

Glossary: A list of terms with their field-specific definitions. Translators compile and use bilingual glossaries.

GPT (generative pre-trained transformers): A type of Large Language Model (LLM) system used, for example, in ChatGPT.

Language pair: The source language and target language on which a particular translation or editing process operates.

LLMs (Large Language Models): Sets of algorithms that process very large databases (theoretically as big as 'the Internet') to carry out language tasks like those that humans perform.

Machine translation: The automated rendition of text or speech from one human language to another.

NAATI: The Australian National Accreditation Authority for Translators and Interpreters.

Neural machine translation: A machine-translation system that incorporates deep learning. These systems have been used since about 2016, enhancing accuracy and context-sensitivity (for example, in Google Translate).

Non-translation: The non-provision of a translation in a particular target language.

Original text: Here, the English-language text in its current form, without pre-editing. We thus refer to the 'original source text' and the 'pre-edited source text' to distinguish between the two kinds of texts that a translator can work from.

Output: Here, the text produced by a machine-translation system; also called the 'machine translation'.

Post-editing: The correction of raw machine-translated text.

Pre-editing: The editing of a source text to reduce the number of errors in the passage through machine translation; this is a type of 'controlled authoring'.

Prompt: User's instruction or command given to a generative AI system.

Raw machine translation: A machine translation that has not been corrected or edited; also called 'unedited machine translation'.

Re-narration: Conveying or re-telling a message in one's own words, which may be in a different language.

Segment: In a translation-memory system, the unit of language stored in the system and paired with a previous translation of it. Segments are usually sentences but can also be sets of sentences.

Source language: The language of the original text, which the translator translates from (i.e., English in this report).

Source text: The text that is translated into another language, also referred to as the 'original text' in cases where there is no pre-editing. When there is pre-editing, we refer to the 'pre-edited source text'.

Target language: The language a text is translated into.

Template: A monolingual document that comprises all the elements (usually 'chunks' at the level of short paragraphs) that might be used in a future message. To produce a particular emergency message, the author selects and combines the appropriate chunks (see Appendix D).

Translation memory: A database comprising matching segments in different languages. The database allows users to store and potentially reuse their previous translations in their new translation projects.

Translation memory suite: A set of electronic tools that recycle previous translations, these days with numerous added features such as machine-translation feeds, term bases, revision tools, and quality controls. Also called computer-aided translation tools ('CAT tools').

Unedited machine translation: A machine translation that has not been corrected or edited; also called 'raw machine translation'.

1. Executive summary

1.1. Project background and purpose

A major challenge faced by any multilingual society in emergencies or other critical situations that require either attention or a person to take urgent action, is to ensure that multicultural communities have timely and appropriate access to accurate information in their languages. This challenge is being met with relative success in Victoria and elsewhere in Australia, partly due to new government communication policies and practices that focus on community-based translations and re-narrations of official messages, as well as the government policy of relying on professionally certified translators.

Meanwhile, the use of machine translation in everyday communication is increasing. As developments in machine translation and related artificial intelligence (AI) technologies are ongoing, there are increasing expectations about their ability to supplement or replace human translation.

Progress in these areas has a knock-on effect on another part of government policy: the traditional avoidance of translation technologies, which are officially deemed unreliable. That said, several countries around the world have reported use of machine translation in emergency situations and disaster contexts for reasons of speed, accessibility and cost efficiency. However, while there have been recent technological improvements, serious problems have also been identified, in particular concerning poorly written source texts, inaccurate or ambiguous translations, and the lack of sufficient electronic resources for some languages to train machine translation systems on.

Against this background, we have investigated the use of machine translation in urgent 'behaviour-change communication' in an effort to improve emergency messaging to multicultural communities. This innovative study is the first to evaluate raw machine translations, as well as two editing processes—machine translation results that have been corrected by a human (post-edited), and machine translation of source texts which have been corrected by a human (pre-edited). In doing so, we provide data on problems that may arise when using machine translation in emergency communications and insights into what may be required to address them. As a result, we provide practical suggestions for how to reduce or avoid machine translation problems and risks, to create more efficient and effective, as well as culturally appropriate emergency communication strategies.

Accordingly, we evaluated three different approaches to machine translation:

1. Raw machine translation (machine-translated text that has not been corrected or edited; also referred to as unedited machine translation);
2. Post-edited machine translation (raw machine-translated text that has subsequently been corrected by a human); and
3. Pre-edited English-language source text followed by machine translation (the original text is amended to avoid problems in the passage through machine translation).

With English as our source language, we analysed four different target languages: Chinese, Dari, Greek, and Spanish—all of which are important community languages in Victoria. From linguistic and communicative perspectives, each has a different historical relationship to English. They also each present specific challenges, such as differences in writing systems, in the extent and quality of existing machine translation-related resources, and in potential issues regarding linguistic variation.

This project had three phases:

- Phase 1: NAATI-accredited human translators evaluated and corrected the raw machine translations (post-editing).
- Phase 2: Human translators pre-edited the English-language source texts before submitting them to machine translation.
- Phase 3: Interviews were carried out with fluent speakers of Spanish (n=5), Greek (n=4), Chinese (n=6), and Dari (n=3) from Victoria. We first conducted a questionnaire survey concerning the raw machine translations. This was to identify how the readers made sense of some of the problems in machine translation: whether they would adopt the desired behaviour change (actionability) and whether they understood basic information (comprehension). We then asked each participant to assess: (a) the raw machine translations, (b) the post-edited machine translations, and (c) the raw machine translations of pre-edited texts. They were also invited to comment on problematic translations that might cause misunderstandings or a lack of trust.

To capture rich data in understanding the effectiveness and efficiency of each workflow, in Phases 1 and 2 we also recorded and analysed:

- Differences in accuracy between the raw, post-edited, and pre-edit-based machine translation texts.
- The time each translator took to correct the raw machine translations in Phase 1, which allowed us to evaluate the variable qualities of the raw machine translations in the four languages.
- Each translator's screen activities for the text corrections required as part of Phase 1 and Phase 2 workflows, which allowed us to map the 'human touch'

onto specific parts of the texts and to see how individual translation problems were solved.

1.2. Key findings

1. In the first instance, problems in the raw machine translation texts were identified in all target languages under evaluation. In some cases, these were shared across all languages; in other cases, they were language-specific. Crucially, some of the errors could put readers at life-threatening risk due to mistranslations. This risk can nevertheless be reduced in various ways.
2. Risks can be reduced by developing translation memories and glossaries. Rather than extract such resources from previous translations, in this case it is advantageous, prior to any urgent situation, to produce human translations of the templates currently available for emergency messaging. Those human translations can be used to override machine translations in situations of urgency.
3. Risks can be further reduced by post-editing (correcting machine translations) and pre-editing (rewriting English-language source texts in a clear way). Both are found to be effective ways to enhance the actionability and understandability of behaviour-change messages.
4. Risks can be mitigated by human post-editing of machine-translated texts but are more efficiently reduced through human pre-editing to make the original English-language texts clearer, less ambiguous and less prone to errors in the passage through machine translation.
5. Post-editing can also enhance the perceived status and trustworthiness of the emitting institution and indicates respect for the receiving language and culture.
6. Pre-editing standardises the English texts to improve the quality of automated translation. It generally involves clarifying syntax and eliminating ambiguities. It may also require awareness of grammatical differences across various languages.
7. There is a slight preference by receivers for the results of pre-editing rather than post-editing. None of our participants preferred the raw machine translations, given the lower levels of accuracy and comprehension.
8. The time required for human post-editing will vary according to language—and will be greater for less-resourced languages such as Dari and other languages of the many newly emerging multicultural communities in Victoria.
9. If a text is to be translated into more than two languages or so (depending on factors such as text quality and the availability of electronic language resources), then pre-editing tends to be more cost-effective than post-editing.

1.3. Recommendations

We have asked whether, when and how machine translation can be used for high-stakes texts in situations of urgency. Our research supports the following recommendations:

1. Raw machine translation should *not* be used for emergency texts *as English-language source texts are currently written*. It might nevertheless be used as a last resort when adequate preparations have been made, especially the writing of clear source texts, the prior human translation of templates, the use of human post-editing when possible, and the identification of appropriately selected bilingual community-based contact persons.
2. All emergency messaging should apply the basic principles of clear, explicit writing in the English-language source text. This can vastly improve the quality of machine translations.
3. Certified translators should be employed whenever possible, particularly for translations of English-language templates.
4. Professionals can also post-edit machine translation output, when there is time, and pre-edit original texts and templates, independently of time constraints.
5. Raw machine translation can be combined with other communication solutions, especially the clear writing of source texts, the human translation of templates, and the identification of bilingual community-based contact persons who can explain the translations. The various elements in the workflows depend at each stage on the time available.
6. Raw machine translations should always be clearly labelled as such, so that consumers, service providers and translators are immediately aware of potential communication issues.
7. For emergency messaging, machine translations should not be compared directly with human translations. Such comparisons overlook the need for urgency. On the ground, the comparison is more often with the absence of any translation at all.
8. Machine translation and generative AI are both improving in quality, but not to the extent that we can be sure that no high-stakes errors are made.

1.4. A simple guide to writing for machine translation in multilingual emergency communication (pre-editing of source texts)

Most problems with machine translation can be solved by applying just one rule: *make everything as explicit as possible*. Anything that is implicit, in the sense that it relies on knowledge that is not expressed in the text, can create problems for machine translation.

The original English-language texts we have been working on successfully apply the basic guidelines for behaviour-change communication: the sentences are short, the words are from everyday language, and the second person (*you*) is used. However, these principles do not always make the texts friendly to machine translation.

The following general recommendations are based on what we have discovered when pre-editing the three documents to make them friendly to machine translation. Although the guidelines are very similar to the general rules for clear technical writing in English, there are a few extra considerations that are due to the varying nature of different languages. The recommendations go from the most important to the most optional:

1. Spell out who has to take an action.
2. Spell out what the action is on.
3. Indicate where things are.
4. Avoid culture-specific terms.
5. Avoid two-part verbs.
6. Avoid elements with multiple grammatical functions.
7. Where possible, avoid terms that can be gratuitously gendered.
8. Use the second person only for actions to be taken.
9. Remove ambiguities that are due to the order of elements in a sentence.

Spell out who has to take an action.

Sometimes a human can tell who has to act by applying knowledge about what happens in the world, but machine translation must sometimes guess. Some languages make a guess obligatory, which can lead to mistakes:

> *Original: It is too late to leave.*
> *Back-translated machine translation: It is too late for us to leave.*
> *Re-write: It is too late for you to leave.*

Spell out what the action is on.

Make sure it is clear what is being done to what. Again, humans can figure out the relations, but machines are not so good at it. We would not throw away fish organs and then eat them, but a raw machine translation might tell us to do precisely that:

> *Original: internal organs must be removed from the fish and discarded before eating*
> *Back-translated machine translation: internal organs must be removed from the fish and discarded before eating them*
> *Re-write: internal organs must be removed from the fish and discarded before you eat the fish*

In the example below, the verb *position* does not have a corresponding term in Chinese, which forces the machine translation to guess rather unhappily:

> *Original: Try to position the car towards the approaching fire.*
> *Back-translation: Try to park your car near the fire.*
> *Re-write: Try to park the car with the front towards the fire that is coming towards you.*

Indicate where things are.

The following is the one serious error in all the machine translations of the 'Too late to leave' message concerning bushfires:

> *Original: If you are caught in fire in your car*
> *Back-translated machine translation: If your car is on fire*
> *Re-write: If you are in your car and you cannot escape the fire*

The English reader instinctively groups the words as follows: 'If you are caught in fire [in your car]'; the machine translation sees it as 'If you are caught in [fire in your car]'.

This can also concern cases where an idea seems very clear in English, but the relations become difficult in other languages:

> *Original: Do not swim in affected areas or use water for cooking, drinking, washing or showering*
> *Back-translated machine translation: Do not swim in affected areas or use any water for cooking, drinking, washing or showering*
> *Re-write: Do not swim in affected areas or use water from those areas for cooking, drinking, washing or showering.*

In the following example, the idea of *indoors* does not have a specific meaning in some languages:

> *Original: Go indoors*
> *Back-translated machine translation: Go to the interior [of the country?]*
> *Re-write: Go inside a house or a shed*

Avoid culture-specific terms.

Often a term is perfectly clear within Australia culture, but other languages do not have a corresponding term or expression, for instance, examples such as *yabbie*, *football oval* and *u-turn*. The easiest solution is often to avoid the Australian term:

> *Original: mussels, crayfish, yabbies*
> *Back-translated machine translation: mussels, lobster, lobsters*
> *Re-write: shellfish*

> *Original: Make a u-turn and travel to safety.*

Back-translated machine translation: Do a turn in U and travel safe.
Re-write: Go back and travel to a safe location.

Original: football oval
Back-translated machine translation: oval [shaped] football
Re-write: football field

Avoid two-part verbs.

Verbs that have two parts, as in *look up*, can be interpreted in different ways: to look up a word is not the same as to look up to the sky. They are better avoided:

Original: slow down and turn on your headlights
Back-translated machine translation: calm yourself and turn on your headlights
Re-write: reduce speed and turn on your headlights

Avoid elements with multiple grammatical functions.

In the following examples, *shelter* could be a noun or a verb, and *boiling* could be an adjective or a gerund (a verb used as a noun). The difference may not matter much for actionability, but the ambiguities make the reading difficult:

Original: shelter inside a house
Back-translated machine translation: a shelter inside a house
Re-write: find shelter inside a house

Original: Boiling water bursts the algae cells…
Back-translated machine translation: To boil water bursts the algae cells…
Re-write: When water boils, it bursts the algae cells….

Where possible, avoid terms that can be gratuitously gendered.

Since machine translation selects the most probable options in the databases, doctors tend to become men and nurses tend to become women. The sexist bias can be avoided by using a more neutral term like *healthcare professional*:

Original: seek medical advice from your local doctor or Nurse-on-Call on 1300 60 60 24.
Back-translated machine translation: seek medical advice from your local [male] doctor or [female] nurse at 1300 60 60 24.
Re-write: seek medical advice from your local healthcare professional or by calling the professionals always available at 1300 60 60 24.

Use the second person only for actions to be taken.

Although use of the second person (*you*) is generally recommended for behaviour-change messaging, the pronoun becomes problematic in languages that distinguish between different kinds of second person, mostly between formal and informal

variants. This was the case, for instance, in all our translations into Spanish, Chinese and Greek, where the raw machine translations shifted between the formal and informal pronouns.

These shifts do not affect actionability, but they can compromise the perceived trustworthiness of the text. In our focus group for the machine translation into Spanish, for example, the problem was held to be major by the language teachers but less so by the other readers.

There is a trade-off to be calculated here. On the one hand, the second person makes the text more direct and the actions more explicit. On the other, the varying translations may decrease trustworthiness. We estimate that it is worth keeping the second person when there is no easy alternative, thus accepting the risks of visible (but not critical) mistakes in some languages.

To mitigate the risk, however, one might want to remove the second person in cases where the sentence does not concern a direct action, although this is a very optional consideration:

> *Original: What you should do*
> *Back-translated machine translation: What you [formal / informal] should do*
> *Re-write: What to do*
>
> *Original: The extreme heat is likely to kill you*.
> *Back-translated machine translation: The extreme heat is likely to kill you [formal / informal]*.
> *Re-write: The extreme heat is likely to be deadly.*

Note that the second-person problem can also be solved efficiently by using AI solutions with a specific prompt (e.g. 'Translate into Chinese in a formal register/ using the formal second person').

Remove ambiguities that are due to the order of elements in a sentence.

Sometimes ambiguities exist both in English and in the other languages. Avoiding them is just part of good writing, and the problems do not usually affect actionability.

The following sentence could be saying that plants are being grown for food while they are being processed and packed. It makes more sense to imitate the order in which actions really occur:

> *Original: Irrigation water should not come in contact with plants being grown for food during processing and packing.*
> *Back-translated machine translation: [same]*

Re-write: Irrigation water should not come into contact with plants that are being processed and packed to be consumed as food.

In the following example, we guess one should have a friend for *all* the activities, but the sentence does not make that clear:

Original: Always swim, dive or surf with a friend.
Back-translated machine translation: [same].
Re-write: Always go with a friend when you swim, dive or surf.

A more common problem is when a description can be interpreted in several ways. Does the following mean large schools of fish *and* large schools of other animals *and* large schools of wildlife? In English, the specific term *school* is applicable to fish only, so the ambiguity is resolved. But this specific term does not exist in all other languages:

Original: Keep away from large schools of fish, seals or other wildlife as these can attract sharks.
Back-translated machine translation: [same]
Re-write: Keep away from areas where there are seals, numerous fish or other wildlife. These can attract sharks.

2. Overview of the report

In the preceding executive summary, we presented our key findings (1.2), our recommendations (1.3), and a practical guide to writing English-language source texts for machine translation (1.4).

The following report is structured as follows.

In Section 3, we provide details on the project background, including a literature review of local and global machine translation policies, the advantages and disadvantages of machine translation and generative AI, and different machine translation workflows.

In Section 4, we present the research design, explain our selection of variables and elaborate on the research procedure.

In Section 5, we present our results on the actionability and comprehension of (1) raw machine translations, (2) the post-editing process (including language-specific problems), and (3) the pre-editing process. We present the ways community members received the three kinds of translation outputs.

In Section 6, we provide recommendations on when and how to incorporate machine translation in emergency communications.

In Appendix A, we present the original and pre-edited source texts side by side.

In Appendix B, we detail problems encountered and time spent on correcting errors by translators during the post-editing process.

In Appendix C, we include the questionnaire given to community members to test the actionability and comprehension of raw machine translations.

In Appendix D, we give an example from the document templates currently in use.

This research was conducted in Melbourne from August 2023 to May 2024. It was guided in part by discussions with the designated Machine Translation Project Steering Committee established by the Victorian Department of Families, Fairness and Housing, which has an active role in supporting and improving the delivery of emergency information to all communities in Victoria. Progress and interim findings were discussed over the course of the project with the steering committee which also provided valuable feedback.

3. Project background

The recent experience of the global COVID-19 pandemic in Victoria and elsewhere in the world has demonstrated the importance of and difficulty in communicating critical information designed to protect communities quickly and effectively. This is particularly the case when targeting multicultural communities, due to language and cultural barriers (Hajek et al., 2022; Karidakis et al., 2022; Sengupta et al., 2024).

Victoria is a national and international leader in efforts (including targeted research) to improve multicultural communication outcomes (e.g., the 'Better practice guide for multicultural communications' (Victorian Department of Families, Fairness and Housing, 2023)). This is driven in part by need: Victoria is among the most culturally and linguistically diverse States and Territories in Australia, with more than 270 different language communities living within its borders. At the same time, Victoria continues to show significant long-term population expansion as a result of large-scale immigration from overseas—with a resultant ongoing increase in the number and size of multicultural communities within its borders. This evolving diversity also increases the need for effective communication in a rising number of languages.

The use of machine translation *outside* of emergency communication is rapidly evolving around the world. It brings with it the possibility of extreme time savings, which are of clear interest for emergency messaging. However, one must be wary of unrealistic expectations regarding the accuracy of machine translations.

Most previous research on the use of machine translation for emergencies has concerned policy issues, the advantages and disadvantages of machine translation with respect to unaided human translation, and the various ways in which technology can be integrated into translation workflows. There has been very little attention given to the actual consequences of machine translation and how the various problems can be resolved and managed by appropriate human intervention. That is a specific focus of this report.

3.1. Policies and guidelines on the use of machine translation in Victoria and beyond

Current Victorian Government policy 'advises against the use of automated interpreting and translating tools, which cannot at present be guaranteed to be accurate' (Department of Premier and Cabinet, 2017, p. 10). The policy further notes that where machine translation is used, its output should always be checked for accuracy by a translator with a formal credential from the National Accreditation Authority for Translators and Interpreters (NAATI).

At the Federal level, the *Australian Government Language Services Guidelines* (2019) similarly recommends that 'Australian Government agencies should engage NAATI-credentialed translators to post-edit [i.e., correct] machine translation output' (Department of Home Affairs, 2019), which is broadly in line with International Standard ISO 18587:2017 on post-editing. The guidelines also suggest that machine translation should be run through an automatic post-editor and then checked by a NAATI-credentialed translator. The policy goes on to list further actions that need to be considered when using machine translation: risk management, the use of pre-editing, the compiling of glossaries and ongoing quality control by NAATI (Department of Home Affairs, 2019, p. 45). The document also recognises that 'machine-translated output may be less reliable (or not viable) for minor languages' (Department of Home Affairs, 2019, p. 46).

Australian policy is currently broadly in line with developments around the world (for details, see Pym et al., 2024). For example, the European Commission's in-house neural machine translation system, known as eTranslation, comes with the caveat: 'Use it to get the gist of a text or as the starting point for a human-quality translation. If you need a perfectly accurate, high-quality translation, the text must still be revised by a skilled professional translator' (European Commission, 2023). Singapore similarly has its own neural machine translation system: SG Translate Together. Since 2021, the Singaporean government has sought to improve the system by inviting registered users, called 'citizen translators', to post-edit its output (Government of Singapore, 2024). The United States (US) digital.gov platform features an 'Introduction to Translation Technology' (United States Government, 2023) which, similar to the Australian Government guidelines, discourages the use of raw machine translation. Despite acknowledging that artificial intelligence has the 'potential of making translations more accurate as it learns', the guidelines also state that 'agencies should work with competent human translators for all translations, including translations supported by translation technology'. However, the United States Department of Health and Human Services 'seek[s] comment on the use of machine translation in health programs and activities generally, other possible approaches to address this issue, and whether there should be an exception to this provision to allow for the limited use of machine translation in exigent circumstances' (United States Government, 2022, p.47862). This indicates openness to further explore potential uses of machine translation tools.

Several countries have reported on the use of machine translation in emergency situations. For instance, following the Haiti earthquake in 2010, within days a machine translation system was set up, providing emergency communication translations for Kreyòl, a resource-poor language (Lewis, 2010).

There are also some guidelines currently available concerning the labelling of machine translations, in the same way as food products have labels providing information for

consumers. For example, the International Organization for Standardization has developed ISO 11669:2024, which recommends the label 'unedited machine translation' (UEMT) as a warning to users. ASTM International, which also develops international standards across a wide range of materials and activities, has prepared ASTM F2575-23 (2023), a guideline, which recommends distinguishing between raw unedited machine translation and bilingually reviewed translation. The latter category includes unaided human translation, human translation with translation tools, and post-edited machine translation (see Melby & Lester, 2024).

3.2. Disadvantages of machine translation

Most of the existing scholarly work on machine translation emphasises a range of disadvantages. Pym et al. (2022), for example, identified serious problems in the use of machine translation for healthcare texts. Specifically, they found issues such as untranslated text embedded in images, omissions, as well as ambiguous syntax (i.e. sentence structure) and terms that may have context-specific meaning. They further found poorly written source text to result in nonsensical translations, which at times give the opposite meaning to what was intended. Those issues appeared to occur more frequently between languages that are not historically related to each other, languages that lacked sufficient translated language training data, and in contexts where the source text differs from the data the translation system was trained on (Bowker & Buitrago Ciro, 2019; Hu et al., 2010). However, to our knowledge, none of these studies investigates cases of actual harm ensuing from misunderstandings of machine translations, although they give warnings of negative possibilities. Here, on the other hand, we do investigate ways in which inaccuracies in machine translations can affect receivers' decisions about actions to take.

When machine translations are compared with human translations, the quality of the latter is generally perceived as being superior in terms of accuracy and nuance (Bowker, 2009; Hale & Liddicoat, 2015; Pym et al., 2022). One crucial differentiator is assumed to be humans' use of real-world knowledge to interpret the meaning of a text in a specific context. This enables more accurate resolutions of lexical and structural ambiguities, including cultural meanings, allusions, implied meanings, irony and other stylistic techniques (Bowker & Buitrago Ciro, 2019; Hale & Liddicoat, 2015; O'Mara & Carey, 2019). Liddicoat (in Hale & Liddicoat, 2015) argues that machine translation systems like Google Translate 'can create an illusion of comprehension without ensuring the reality of that comprehension' (pp. 22-23). Human translators are also seen as being more capable to translate culturally or context specific terms and expressions (O'Mara & Carey, 2019).

3.3. Advantages of machine translation

Despite these drawbacks, a range of benefits have been identified. The major advantages are speed, accessibility and cost efficiency, especially when compared to human translators (Bowker, 2009; O'Mara & Carey, 2019). Even with basic connectivity and understanding, machine translations can be extremely beneficial, for instance, to residents of or visitors to rural areas, who may not have easy access to other translation tools (O'Mara & Carey, 2019). Embedding machine translation in emergency SMS communications can also increase immediate access to vital information (Lewis, 2010). Further, if no translators and interpreters are available, the use of machine translation system ensures that a translation of some kind can still be provided (Bowker, 2009).

The problem remains as to whether a possibly inaccurate machine translation presents more or fewer risks than no translation at all (Bowker, 2009). However, what we do know is that multicultural communities can perceive inaccurately machine-translated material, such as signage, as being disrespectful and offensive, as it can result in nonsensical and impolite language. Thus, attempts to be inclusive can sometimes have the opposite effect (Angermeyer, 2017; Hajek et al., 2022).

The use of machine translations has also been shown to give users a sense of independence and confidentiality (Pym et al., 2022). For example, refugees, asylum seekers and young migrants have been found to prefer machine translations rather than depend on official government interpreters, who were not trusted (Pokorn & Čibej, 2018).

3.4. Disadvantages and advantages of generative AI

A recent innovation in text translation is the increasing availability and use of Large Language Models (also known as LMMs) as used in systems such as ChatGPT and Google Gemini, collectively known as 'generative AI'. Compared to more established neural machine translation systems such as Google Translate and DeepL, which draw on databases of past human translations, Large Language Models draw on vast language resources in order to write texts that are similar to those that humans would write. In evaluations based on automatic metrics, translations done with generative AI systems sometimes score higher than those done with neural machine translation (Hu et al., 2024; Na et al., 2024), although it depends very much on the language pair (Jiao et al., 2023). ChatGPT translations, for example, have been found to be inferior to neural machine translations for Irish (Castilho et al., 2023), Icelandic and Hausa (Hendy et al., 2023). They generally do not perform well for low-resource and non-cognate languages, due to the limited availability, quality and accessibility of appropriate training data for AI systems to learn from, as well as differences in cultural and social nuances (Ghosh & Caliskan, 2023). For these reasons, there has been no quantum leap in translation quality across the board. This means that, for many of the languages that are spoken in

Victoria, machine translation may still be a better option than generative AI translation, at least until sufficient high quality electronic language resources have been generated as training data for the latter.

Interestingly, generative AI systems perform significantly better when judged at paragraph level rather than sentence level (Karpinska & Iyyer, 2023), indicating the extent to which they can adjust texts to suit contexts. A very viable solution for many language pairs is thus to **use machine translation and then have the result revised by generative AI with respect to specific purposes or readership profiles**. Since emergency messaging often needs to be tailored to specific target groups, generative AI systems can be used for that tailoring. An example would be the distinction between a formal and an informal *you* in many languages, which is very problematic for machine translation from English but can be solved, as previously noted, by specifying a formal or informal readership when writing a prompt in generative AI.

3.5. The integration of machine translation into different workflows

The translation of a text from one language to another requires some kind of process, often referred to as workflow. Studies have been carried out on several workflows in which machine translation can be used together with other communication strategies.

Post-editing: Machine translation is mostly used with human post-editing (Bowker, 2009; Bowker & Buitrago Ciro, 2019; Uekusa & Lee, 2023), sometimes after a quality evaluation has been made to identify the text requiring most attention. Various degrees of post-editing are possible. Bowker (2009) trialled four translation workflows in linguistic minority communities in Canada: (a) raw machine translations, (b) unaided human translations, (c) light post-editing, where only content errors were corrected, and (d) heavy post-editing, where all errors were tackled. Both forms of post-editing gave acceptable results, although the French speakers involved in testing preferred heavily post-edited translations, citing cultural preservation as the single most important reason for wanting an accurate translation.

Pre-editing: Editing an original text in a way that reduces potential machine translation errors is called 'pre-editing'. Pym et al. (2022) examined raw machine-translated COVID-19 information on a Catalan government website. They then pre-edited the Catalan source texts to eliminate the ambiguities that gave rise to machine translation errors. The simplified texts yielded more accurate and acceptable results in both English and French, demonstrating that pre-editing may prove more efficient than post-editing if a text is to be translated into multiple target languages.

Human mediators: The use of non-professional, bi- or multilingual community contact persons can help ensure that community members are aware of translated information,

particularly during emergency situations, and can provide supplementary explanations when end-users request them to do so. Similar mediation practices have been reported in studies on communication practices during the COVID-19 pandemic (Hajek et al., 2022; Karidakis et al., 2022). In a localised emergency context, community members with local languages can be the people who receive machine-translated messages, explain them, and re-narrate them, often changing from written to spoken mode. As they have relevant linguistic and cultural knowledge, they are often in a position to compensate for problems in machine translations, without duplicating or replacing the functions of a professional translator or interpreter. These workflows have received scant attention in the literature, although an interview study by Cadwell (2020) shows how foreigners relied on a range of friends, contacts and non-official sources to receive translated information in the 2011 Great East Japan Earthquake. In many cases, these non-official persons were turned to because they were highly trusted.

Pre-translated templates: Given the repetitive and predictive nature of emergency messaging, templates can be prepared that include the main text units or 'chunks' from which a new message can be compiled. Such templates exist for the main emergency messaging in English in Victoria and are used extensively and successfully, e.g. through the VicEmergency website (see Appendix D). Since there is no time pressure for the elaboration and updating of the templates, professional translators should be employed for the main community languages in areas at risk. Those human-translated templates could also be used to train dedicated machine translation systems or could be fed into commercially available translation-memory software like Trados as databases (paired translations and/or dedicated glossaries). A translation memory is a database of bilingual segments. It allows users to store and re-use their previous translations in future work between the same two language in similar domains (see 4.1 below). It works more or less like human memory—the more memories you have, the more you can draw on in a new scenario. These possibilities have yet to be explored in the literature.

Translation across media: These different workflows should also account for the various media that are preferred by different cultural groups. Following the 2010 earthquake in Haiti, crowdsourced volunteers were responsible for geotagging, translating and classifying the SMS messages, and to route them to appropriate aid agencies. As noted by scholars reporting on the project (Hester et al., 2010; Lewis, 2010), this process could have been simplified by integrating a machine-translation engine—which Lewis (2010) and colleagues promptly developed in response to the crisis—for example to provide preliminary translations that volunteer translators could then correct instead of having to translate the SMS themselves.

4. Research design

This study examines the production and reception of machine-translated emergency messages into several target languages. We set out to compare three ways of using machine translation: (a) raw (unedited), (b) post-edited, and (c) with a pre-edited source text. Our evaluations are based on two main sources of information: (a) from Phase 1, what professional translators did when correcting (post-editing) the machine translations, and (b) from Phase 3, what community-language readers said about the results of the three ways of using machine translations.

4.1. Initial experiments with training machine-translation systems

Initial experiments were conducted to determine if it was feasible to develop language resources for a dedicated machine-translation system, which could be trained to perform well in the specific domain of emergency services. This would be similar to the government machine-translation system developed in Singapore, where Singapore-specific terms and expressions can override the translations proposed by a general system like Google Translate. We therefore took a sample of previous emergency translations in our project languages, aligned them into databases of paired sentences, and automatically extracted term bases (glossaries) from them using ChatGPT. The general hope was that those language-specific databases would then improve the accuracy and acceptability of machine translations. Alternatively, it was thought that they could be fed into a translation-memory suite like Trados, Phase or MemoQ, where they would override the solutions provided by input from a public machine-translation system or AI feed.

Our experiments nevertheless indicated that this general approach was not worth pursuing, for the following reasons:

1. To recycle pre-existing translations, one needs to extract paired sentences in two languages from previous translations. This extraction process was consistently perturbed by the need to convert PDF files and the use of text in images from which text extraction was not possible.
2. The automatic extraction of terminology required significant editing by a human expert in each language concerned, largely because of variation in the terms used in the human translations over the years.
3. Relatively few terms and sentences were repeated across different text genres. For example, a text on how to prepare for bushfires certainly shares the same *topic* as a ‘Too late to leave’ message (they are both about bushfires), but the two texts have almost no sentences in common. They actually give very different kinds of information and express it in very different ways.

Together, these factors meant that the possible gains from recycling previous translations of sentences were less than could be justified by the effort involved.

We then decided to focus instead on time-sensitive emergency messaging, where communication must be quick and effective. That urgency factor can potentially justify taking certain risks with machine translation, given that the comparison is not between an automatic and a human translation, but between an automatic translation and no translation at all (non-translation).

We also became aware of the excellent Word templates used for this kind of messaging by VicEmergency in Victoria (see Appendix D). Since almost all emergencies are like previous emergencies, these templates include almost all the text elements that an author, in a particular case, can then select and combine to produce an emergency text. As a result, instead of extracting language resources from previous translations or contemplating a dedicated machine-translation system based on those resources, one should produce human translations of the templates, which have been developed and tested specifically in the context of emergency communications in Victoria.

We therefore saw that it made more sense to test the viability of a model with the following steps:

1. Ensure that the English-language templates are written in a way that is friendly to machine translation (pre-edited).
2. Have human translators render the templates into the main languages at risk.
3. Convert those translations into translation memories (stored translations) that can be used to override proposals made by public machine-translation systems (Pym & Hao, 2024).
4. For problematic terms, establish glossaries that can similarly be used to override the machine translation suggestions.

It is important to stress here that current technologies allow machine translation to be easily combined with several other ways of solving translation problems.

Figure 1 shows how various kinds of translation suggestions can be integrated into the one workspace, in this case the translation-memory suite Phrase. In this interface, the original text is segmented into individual sentences and presented on the left-hand side; next to them are the raw machine translations. Human translators can then make corrections to those machine translation suggestions by drawing on the translation memories (records of previous translations, especially of templates), on glossaries, or on additional machine translation suggestions, all listed on the right-hand side. For example, to translate the term *Emergency Warning*, the raw machine translation into Spanish suggests *Advertencia*. However, the glossary and the translation memory give *Aviso* as the preferred translation, so the post-editor can see those proposals and may decide to adopt them in the post-editing space.

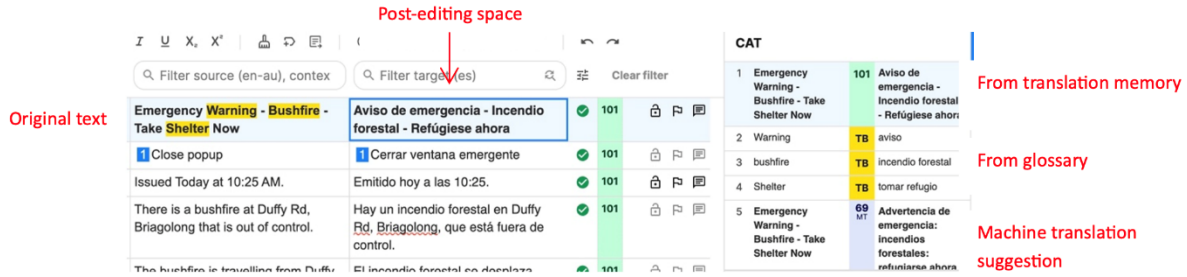

Figure 1. Workspace in the translation memory suite Phrase, integrating machine translation, a pre-established glossary, and previous translations (from the translation memory).

We are not suggesting, however, that translations must be done with this kind of technology in urgent situations—since interaction with the technology takes time. The technology is more useful for *preparing* for machine translation by developing translation memories and glossaries and ensuring that they help solve the problems of machine translation.

Within that general model, our main research activity has thus focused on testing the quality of machine translation with and without post-editing, as well as the ability of pre-editing to solve problems in a range of different languages.

4.2. Selection of variables

The main variables were selected as follows.

4.2.1. *Machine translation system*

We selected Google Translate as the main machine translation system, primarily because it is the system that can most easily be used with websites and Android. It is also the system most widely available around the world in addition to having the broadest range of languages: some 133 languages in 2023, which is still much less than the 270 or so languages spoken at homes in Victoria. That said, Google Translate does not specifically include Dari. We therefore used Google Translate for Persian, of which Dari is a variety, to test the viability of machine translation between very closely related and/or overlapping languages.

The original source material in English was fed into Google Translate in December 2023 to provide us with the initial raw translations for evaluation and modification as necessary.

4.2.2. *Target languages*

Four target languages were chosen for our main experiment: Chinese, Spanish, Greek, and Dari/Persian. Since a key variable for machine translation is how many electronic resources are available for a given target language, we work with high-resource cognate

(i.e. related to English) languages (Spanish and Greek), a high-resource non-cognate language (Chinese), plus a right-to-left low-resource language (Persian), where we were interested in how accessible the output was for speakers of Dari. These four languages are also used by large communities in Victoria, each with a different profile, and specific challenges.

Spanish is the official language of 21 nations around the world, with large communities from many of these countries residing in Victoria. While Spanish is related to English, using the same Roman alphabet, and sharing a significant amount of technical language (due to English borrowing from French and Latin, which are closely related to Spanish), there is also the potential for lexical and other dialect differences across the many Spanish-speaking nations.

Greek is more distantly related to English (which however has borrowed significant technical lexicon from Greek) and uses a different writing system. Older members of the Greek community in Victoria are often only literate in Greek.

The Chinese-speaking community is expanding rapidly in Victoria and their language uses a character-based writing system. All three target languages (Spanish, Greek, Chinese) are well-established in machine translation, and are hence known as high-resource languages, i.e. they have large vocabularies and textual sources to train machine translation systems with.

Dari is of particular interest for a number of different reasons. It is very distantly related to English and is written with a right-to-left Perso-Arabic script. It is the first or second language of a large part of the emerging Afghani community in Victoria. However, there is also considerable variation in the use of language names amongst members of this community—who may variously use the terms ‘Dari’ and ‘Persian’ to describe the same language that they use. Also closely related is Hazaragi, which is preferred by many Hazaras as the name for their language variety. It also remains true that across broader society ‘Persian’ is most linked to Iran. Given the much larger number of Iranian Persian speakers in the world, any machine translation system will likely be trained on Iranian Persian and therefore show Iranian dialect features.

4.2.3. *Sample texts*

Since the use of machine translation is mainly justified by extreme time savings, we decided against using texts that had a long shelf life (guidelines, instruction manuals, etc.), for which professional translators should be properly employed. We therefore selected three texts issued for different types of emergencies on the VicEmergency website, which deal with bushfire (481 words in the original source text), water quality (395 words) and sharks (179 words).

4.2.4. *Evaluation strategy*

Machine translation systems are usually evaluated by automated metrics that compare the output to a human-produced reference translation (Pym & Hu, 2024). This is the case for BLEU (Bilingual Evaluation Understudy), METEOR (Metric for Evaluation of Translation with Explicit ORdering), and TER (Translation Edit Rate), which quantifies the number of edits required to change a machine translation into the reference translation. Since these metrics are automatic, they are applied to long stretches of text for which means are calculated.

For the present study, we decided against using these merely quantitative metrics. In an emergency text, one error can be trivial, without consequence, while another with the same formal properties can be high stakes, a question of life or death. We needed an evaluation system that takes account of these qualitative differences.

We therefore decided to test a small number of machine-translated texts in two main ways: first, by looking at the actual changes (post-edits) that translators made to the texts to improve their accuracy and comprehension, and second by collecting feedback from speakers of the languages on three different translations: (a) the raw machine translation from Google Translate, (b) a *post-edited* version of that translation, and (c) the raw machine translation of a *pre-edited* version of the original English-language text.

4.3. Research procedure

The research consisted of three consecutive phases.

4.3.1. *Phase 1: Post-editing of machine translation errors*

The procedure for the post-editing analysis was as follows:

1. The texts were fed through *Google Translate*.
2. That raw machine-translation output was then corrected (post-edited) by a NAATI-certified translator for actionability (see below) and major stylistic errors.
3. The post-editing process for each text was screen-recorded. The data show the breakdown of time spent on tackling complex translation problems across the chosen languages (see Appendix B). This indicates not only where the machine-translation errors were but also how important the post-editors considered them to be.
4. Brief notes were written up for each text and each language (3 x 4) listing the main translation errors and possible ways in which they could be solved.

4.3.2. *Phase 2: Pre-editing of the original English-language text*

The notes generated in Phase 1 were then used to rewrite (pre-edit) the English texts in such a way that they would go through machine translation with only minor errors. In

theory, the clearer the English text, the fewer the errors in the machine translations. This principle is particularly important when the one text is to be translated into many languages, as a change in the original text can potentially avoid errors in all the translations.

The pre-editing process followed these steps:

1. Working on the original English-language texts alongside their machine translations into Chinese and Spanish, we changed the English text until the translation problems were solved. This mostly meant removing colloquial expressions, clarifying ambiguous syntax (sentence structure), and spelling things out where necessary (Appendix A).
2. On the basis of those pre-edits, a short set of guidelines was compiled explaining how to write in English so that the text is optimally machine-translation-friendly (Appendix D).

4.3.3. *Phase 3: Reception of emergency messages*

In the last phase, we examined how community members understood and evaluated the three kinds of translations we had produced: (a) raw machine translations, (b) post-edited machine translations, and (c) raw machine translations of pre-edited English texts.

First, we conducted pilot studies with the raw machine translations into Spanish and Chinese. This was done as part of a translation teaching and learning activity in a translation class of 29 Spanish speakers in Spain and a similar class of 44 Chinese-speakers in Melbourne.

Once we had identified the main inaccuracies in the raw machine translations in Spanish and Chinese, ten main problems were selected, covering a range of difficulties (particularly complex syntax and unclear terms). Multiple-choice questions were then devised to test how well readers of our four languages could solve those problems (Appendix C).

Some of those questions concerned *actionability*, by which we mean the text's ability to have the reader take the desired action. For example, in the bushfire warning, we noted that the expression *u-turn* had been machine-translated literally into languages where it was not the normal term. We therefore asked:

> *If you are driving towards the warning area, what should you do?*
> *1. Turn around and go in the other direction.*
> *2. Find a church to pray in.*
> *3. Park in a safe place.*
> *4. Enter the warning area.*
> *5. Not sure.*

To select the right answer (number 1), the reader would have to interpret correctly the translation used for *u-turn* and/or apply some common sense. If they selected the right answer, we judged the translation to be actionable.

Other questions concerned *comprehension*, by which we mean the reader's ability to resolve ambiguity. For example, the text on water quality refers to *new outbreaks or changes that might indicate improved water quality*, where the phrase *new outbreaks* can erroneously be interpreted in machine translation as improving water quality. Here the question was:

Why is regular testing being conducted?

1. *To identify new outbreaks that improve water quality.*
2. *To check if people see the warning signs.*
3. *To see changes that improve water quality.*
4. *To test the internal organs of the fish.*
5. *Not sure.*

The correct answer was 3.

We recruited community members from each language community, including four Greek, six Chinese, five Spanish, and three Dari participants. The Chinese community members also included international students, as they represent a large cohort in Victoria who may not have relevant information on emergency situations. While most participants were fluent in English and were thus not necessarily users of translations themselves, they all felt confident to comment from the perspective of translation users in their community.

As noted, three sets of materials in Chinese, Spanish, Greek, and Dari for each type of emergency (bushfire, water quality, and sharks) were used:

1. Texts A: Raw machine translations of the three emergency texts.
2. Texts B: Post-edited machine translations (generated in Phase 1).
3. Texts C: Raw machine translations of three pre-edited English-language source texts (generated in Phase 2).

Focus groups and individual interviews with participants were conducted from April 3 to April 21, 2024, with the following procedures:

1. The participant read and signed the ethics consent form.
2. The participant read Text A (a raw machine translation of the bushfire text) and answered the corresponding multiple-choice questions. Their responses were recorded using the online survey platform Qualtrics.
3. The participant then read Texts B and C and discussed their responses to the survey questions with the researcher.

4. The participant identified errors in the texts, determined which of the three texts addressed their points of confusion, decided which text they trusted most and why, and whether they would follow those recommended actions during a bushfire.
5. Steps 1 to 4 were repeated for the shark and water quality texts.

5. Results

We have tested the viability of machine translations for emergency messaging in three ways, each of which gives a different kind of result.

As part of Phase 3, we asked members of each language community to interpret translations using a questionnaire. Our initial multiple-choice questionnaire tested how readers made sense of ten inaccuracies in the machine translations. Some of the questions tested *actionability* (which of four actions would the reader take); others tested *comprehension* (which of four interpretations would the reader use to disambiguate a sentence).

Regarding raw machine translations, the results of the questionnaire suggest that, even in our small sample of texts, there were at least two instances where a translation blocked sense-making and could have led to a wrong action being taken. One of them concerned all languages, the other only concerned Chinese. That is, **there is a real risk in the use of raw machine translation**, even though most readers were able, given time, to work around stylistic problems and bracket out items that did not make sense. We are interested in how that risk can be reduced.

Our analysis of how translators corrected the raw machine translations (post-editing as part of Phase 1) gives a different view, showing how much work can be required to make machine translations linguistically acceptable in each target language. However, we also found that this corrective **post-editing enhances the perceived status and trustworthiness** of the emitting institution (such as a government authority) and indicates respect for the target language and culture.

In Phase 2, the analysis of rewriting of how the English-language texts can be rewritten (pre-editing) shows that **most machine translations problems can be solved by writing very explicit source text**. When our focus groups were shown three different machine translation outputs (raw, post-edited, pre-edited) as part of Phase 3, **they most often said they preferred the pre-edited machine translations, followed by the post-edited machine translations. None of our readers preferred the raw machine translations.**

5.1. Actionability of raw machine translations (questionnaires)

In the pilot study on Spanish, the possible answers to the questions did not include a ‘not sure’ option, since we wanted to force the reader to decide on an action. However, after studying the machine translations into all the languages, we added a ‘not sure’ option to accommodate translations that reproduced or created ambiguities. We thus allowed for interpretations in which the reader would not adopt the desired behaviour

change but would not take a negative action either—they would have to seek further information. For all analyses involving community feedback, we therefore give the percentage of correct answers followed by the percentage of ‘not sure’ and wrong answers in Table 1.

Table 1. Percentages of readers' answers to questions about the raw machine translations (with number of participants in brackets).

		Spanish (5)			Chinese (6)			Greek (4)			Dari (3)		
	Question	Correct (%)	Not sure (%)	Wrong (%)	Correct (%)	Not sure (%)	Wrong (%)	Correct (%)	Not sure (%)	Wrong (%)	Correct (%)	Not sure (%)	Wrong (%)
Bushfire	1. Take refuge where?	100	0	0	100	0	0	75	25	25	100	0	0
	2. Why long sleeves?	100	0	0	100	0	0	100	0	0	66	0	33
	3. What is a u-turn?	100	0	0	66	0	33	75	25	0	100	0	0
	4. Car on or in fire?	33	33	33	17	66	17	0	100	0	33	33	33
Sharks	1. Where is the warning for?	33	33	33	83	0	17	75	25	0	33	33	33
	2. What do you have to do?	66	0	33	66	17	17	75	25	0	66	33	0
	3. What brings the sharks?	0	66	33	66	17	17	50	50	0	0	66	33
Water	1. Why the analyses?	0	66	33	83	17	0	100	0	0	0	66	33
	2. What should you do?	33	66	0	100	0	0	50	25	25	33	66	0
	3. What can you eat?	60	0	40	83	0	17	25	25	25	33	66	0

Most participants across language groups answered the questions concerning the bushfire text correctly, except for question 4: 'position your car towards the approaching fire' (see Appendix C), in which we observed points of confusion (for Chinese and Greek participants in particular) and misunderstanding. With respect to texts issued for increased shark activity and the blue-green algae bloom (water quality), Chinese and Greek participants generally demonstrated greater comprehension of the behaviour-change messages than their Spanish and Dari counterparts. The main takeaway here is that **raw (unedited) machine translation should not be used in emergency communication scenarios unless other measures are also taken (such as pre-editing, post-editing and the human translation of templates). Without such measures, machine translation may fail to convey the messages clearly (signalled by confusion) and could thus put multicultural community members at risk (signalled by incorrect answers to the questions)**.

5.2. Machine-translation correction (post-editing)

With respect to Phase 2 of our research, Table 2 shows the total time spent on correcting non-optimal machine translations in Greek, Spanish, Chinese, and Dari, as well as the average processing time for each text. It includes the amount of time translators needed to closely read the machine output, check against the original English texts, and revise the raw machine translation to avoid stylistic infelicities and semantic errors that may affect actionability.

A methodological problem here was that some translators corrected more enthusiastically than others, despite being instructed only to correct what was strictly necessary. One translator into Spanish was particularly perfectionist, so the task was given to two other translators as well and the results for the three were averaged. A certain tendency to perfectionism can be found not only among translators (who have a professional interest in finding machine translation to be inadequate) but also in discussions among some receivers, especially language teachers (who find their personal preferences to be the most correct).

On all three texts, the post-editors for Greek and Spanish tended to work faster than those working on Chinese and Dari, as indicated by the processing time per word. This can be attributed to the principle that machine translation works better between languages in the same family (English, Spanish, Greek) than with non-cognate, i.e. unrelated languages (English and Chinese), thus requiring fewer corrections. Machine translation also works better with languages that have extensive electronic resources, which is the case for Chinese, but not for Dari. That would explain why Dari required more time, even though it is technically closer to English than Chinese. As an added complication, the translator working on the Dari text often needed to address differences between (Afghan) Dari and (Iranian) Persian, which are two varieties of the

same language. Extra time was spent ensuring that any Iranian vocabulary would not cause difficulties for Dari speakers in understanding the messages.

Table 2. Total time and processing time per word in post-editing machine translations.

	Bushfire		Sharks		Water quality	
	Total time in minutes	**Milliseconds per word**	**Total time in minutes**	**Milliseconds per word**	**Total time in minutes**	**Milliseconds per word**
Greek	23:09m	2890ms	07:37m	2350ms	13:37m	1970ms
Spanish	24:16m	3210ms	06:06m	2040ms	08:23m	1270ms
Chinese	24:44m	2980ms	11:36m	3690ms	14:41m	2130ms
Dari	52:09m	6460ms	23:36m	7710ms	66:02m	10030ms

Appendix B outlines the places where the machine translation did not perform optimally, along with the post-editing times required to fix them. Some items were problematic across all four languages. For instance, a literal machine translation of the English structure *due to parts of a whale carcass* does not sound idiomatic in Greek, Spanish, Chinese, or Dari. This required each translator to spend extra time searching for corresponding structures and rewriting the entire sentence (Chinese: 01:59m; Greek: 01:08m; Dari: 02:55m). Other issues were problematic to different degrees in different languages, as indicated by the variation in the time taken to resolve them. The expression *make a u-turn*, for example, does not appear as such in the other languages but may be understandable: it was machine-translated literally into Dari and Spanish, where the post-editors replaced it with the normal expressions in those languages. The machine translations of the same expression into Greek and Chinese required no changes.

Aside from the stylistic problems, we observed translators spending time correcting a few problematic translations that may affect actionability. The most serious case is the clause *if you are caught in fire in your car*, which machine translation can process in two different ways:

(a) If you are caught in [fire in your car] = there is fire *in* your car
(b) If you are caught in fire [in your car] = there is fire *around* your car

In context, interpretation (b) makes sense because of the subsequent sentences: *Park off the road behind a solid structure. Try to position the car towards the approaching fire*. However, the machine translations in Greek, Spanish, Chinese and Dari all opted for interpretation (a) 'if your car catches fire', which is simpler and thus statistically more probable. So did the AI alternatives ChatGPT3.5, Google Gemini and Microsoft Copilot when we tested this text in December 2023. This is a clear error that caused readers serious confusion about the subsequent instructions (which were all translated correctly): *close all windows*, *turn off the air con*, and *get down as low as possible*

[inside your car]. If the car is on fire, none of those instructions will do much good. Each post-editor spent around two minutes tackling this problematic sentence, accounting for some 7% of the total task time for that text in Chinese and Greek. We also observed screen activities occurring outside of Word. For example, the Chinese translator checked the meaning of the phrase *caught in fire* using an online dictionary. The post-editors across the board chose to allocate a fair amount of time to this problem, which clearly concerns actionability. Our Spanish post-editors all opted for literal versions that retained the syntactic ambiguity of the English sentence: they spent time on the problem but did not actually resolve it one way or the other.

Interestingly, in May 2024, Google Translate, DeepL, GPT4.0, Google Gemini and Copilot no longer opt for interpretation 'there is fire in your car' when working into Spanish. They all now reproduce the ambiguity of the English sentence by translating word for word: 'If you are in fire in your car', showing that even after additional potential translation training over time, the new translated outcomes may *not* improve clarity or comprehension in the target language. Another translation problem concerns which exact location is being referred to in one of the messages. The ambiguity of the English phrase *Issued for Lake Bunga, Lake Tyers Beach* in the water quality source document comes from the way the comma is used in English. It could be interpreted in three ways:

(a) Lake Bunga and Lake Tyers Beach
(b) Lake Bunga, on the beach of Lake Tyers
(c) Lake Bunga Beach and Lake Tyers Beach

This ambiguity is retained in the machine translation into Chinese: it took the translator almost two minutes (around 16% of the total time on task for this particular document) to search for these places and the map of coastal regions of Victoria. The map indicates these are two neighbouring places and there is only one ocean beach next to Lake Tyers. Thus, the translator chose the first interpretation based on extralinguistic information.

Additionally, when translating into Chinese and Dari, the underlined subordinate clause in *Identify new outbreaks or changes <u>that might indicate improved water quality</u>* creates headaches for machine translation. The English structure introduced by *that* modifies *changes*, whereas the Chinese translation mistakenly used it to also modify *new outbreaks*, thus giving conflicting information, i.e., 'the new outbreaks of algae are indicators of improved water quality'. The Spanish and Greek translations were fine in this case.

Chinese and Dari also tend not to differentiate between *crayfish* and *yabbies* in everyday language use: both were translated as 'lobsters' (the same translation for both words). We found out that there are no equivalent terms for these shellfish in (Afghan) Dari, perhaps because it is a landlocked country. The machine translation into Spanish used the Anglicism *yabbies*.

In machine translations into Chinese, Greek and Spanish, the informal and formal second-person pronouns were used inconsistently throughout the texts (i.e. always *you* in English, but informal *tú* and formal *usted* in Spanish, informal *εσύ* and formal *εσείς* in Greek, and informal *你* and formal *您* in Chinese. This stylistic infelicity does not cause misunderstandings. However, it may affect the trustworthiness of the emergency messaging. It took up much of the Spanish post-editors' time, who were especially aware of the problem.

5.3. The issue of language specificity

As can be seen from the examples above, some problems in machine translation affect all languages. They can usually be solved by writing the original text more clearly. Other problems, however, appear in just one or several languages and may require post-editing. Here we give examples where such problems concern actionability and comprehension, bearing in mind that these are the kinds of problems that our interview participants had to solve when reading the raw machine translations.

5.3.1. *Machine translation problems specific to Chinese*

As is indicated by the time differences in Table 2 and Appendix B, Chinese involved some translation problems that were more significant than in the other languages:

> *EN: Try to position the car towards the approaching fire.*
> *ZH MT: 尝试将汽车停在靠近火场的位置*
> *Back translation: Try to park your car close to a fire.*

The English here instructs people who take shelter inside a car to park their car facing the direction of the approaching fire, based on the logic that the windshield is typically more resistant to heat than the rear window and the front of the car has fewer flammable components compared to the rear, where the petrol tank is located. However, the Chinese machine translation distorted the original message, suggesting people park their car *close* to a place that is on fire. The grammatical differences between Chinese and English partly explain this translation error. In English, *position* can be either a noun (a position) or a verb (to position a car), whereas in Chinese, the closest word *位置* is always a noun. The machine translation thus must search for an alternative verb.

> *EN: Cooling systems*
> *ZH MT: 冷却系统 [literal translation]*

Here the Chinese translation is a literalism from English, which refers to the cooling system of a fridge rather than a home cooling system (*制冷系统*) like fans or air conditioners.

5.3.2. *Machine translation problems specific to Greek*

The problems in Greek that may affect actionability are as follows:

> *EN: Get down as low as possible below window level and cover up with a pure woollen blanket.*
> *GR MT: Κατεβείτε όσο το δυνατόν χαμηλότερα κάτω από το επίπεδο του παραθύρου και καλύψτε το με μια καθαρή μάλλινη κουβέρτα.*
> *Back translation: Get as low as possible below the window level and cover it with a clean woollen blanket*

The English says you should cover yourself with a pure woollen blanket when taking shelter inside a car, whereas the Greek says the blanket is to protect the car windows. This is a clear case of confusion, creating significant risk.

> *EN: Slow down and turn on your headlights.*
> *GR MT: Χαλαρώστε και ανάψτε τους προβολείς σας*
> *Back translation: Calm down and turn on your headlights.*

The English says that if there is a bushfire while traveling, you should slow down and turn on your headlights. The implied message is that smoke from the fire can significantly reduce visibility, making it difficult to see the road and other cars. Slowing down allows for better reaction time and turning on headlights helps improve your visibility to others. The Greek translation omits one of the key instructions: the driver might calm down but not slow down.

5.3.3. *Machine translation problems specific to Spanish*

Spanish required some changes to be made to disambiguate texts and provide clearer instructions:

> *EN: internal organs removed from the fish and discarded before eating.*
> *ES MT: desecharlos antes de comerlos*
> *Back-translation: …removed from the fish and discarded before you eat them.*

The Spanish translation says one has to throw away the internal organs before eating them. This error is likely to affect actionability concerning the fish organs, but one would hope that common sense would resolve it.

> *EN: shelter in a room*
> *ES MT: refugio en una habitación*
> *Back-translation: [a] refuge in a room*

The English word *shelter* can be a verb or a noun. Here in the original it is a verb, but it has been incorrectly translated as a noun. As a result, in the absence of a verb in the Spanish translation, there is no clear instruction to take action to mitigate risk.

EN: move indoors
ES MT: mudarte al interior
Back-translation: move yourself to the interior

The Spanish 'interior' (machine translation of *indoors*) is general and could imply inside a house, inside a shed, or the interior of the country. There is also an inappropriate use of the Spanish verb *mudarte*, which is more commonly used for permanent relocations, as in moving house.

5.3.4. *Machine translation problems specific to Dari*

In addition to lower accuracy and reduced quality stemming from Dari being a low-resource and non-cognate language, the main challenge here was dealing with the differences between (Afghan) Dari and (Iranian) Persian dialect variations. For example, Dari and Persian have different words for *lake* and *river*, and the Persian word for *car* is used in Dari to mean 'car engine', while the Dari word is closer to the Persian word for 'motor'. Further, Dari uses many loan words from English, while loan words in Persian are mostly derived from French. For example, Dari uses 'shower', while Persian uses 'douche'.

EN: 12:27 AM
DA MT: ظ.ق 12:27 [transliteration of a.m.]

EN: 10:25 AM
DA MT: صبح 10:25 (in the morning)

The time locator *am/pm* is not used in Dari, which leads to some confusing translations. Interestingly, *am/pm* was treated differently in all three texts. While *AM* was either rendered as a transliteration of 'a.m.', or as 'in the morning', *3:14 PM* lost its reference to *pm* in the third text; it was translated as '15:14'.

EN: algae bloom
DA MT: شکوفایی جلبک
Back-translation: algae blossom

Instead of translating *algae bloom* to indicate an increase in algae, *bloom* is translated in the sense of a blooming flower. This has an impact on the clarity and actionability of the warning.

EN: It is too late to leave
DA MT: برای رفتن خیلی دیر است
Back-translation: It is too late to go.

The English text instructs the reader that it is too late to leave their property due to a bushfire. The Dari translation translates *to leave* as 'to go' which changes the meaning and clarity of the instructions.

> *EN: solid structure*
> *DA MT: سازه جامد*
> *Back-translation: solid [state of matter] structure*

In English, *solid* is used to describe a state of matter but is also used for objects that have a solid quality, i.e. are firm and stable. In Dari, different adjectives are used to describe 'solid' when referring to the state of matter or the quality of an object. The machine translation of *solid structure* was thus nonsensical in Dari.

5.4. English source-text correction (pre-editing)

Once we had used post-editing to identify the main problems in the machine translations for Chinese and Spanish, in Phase 2 we then went back to the original English texts and asked how the English could be re-written to avoid the problems. This is called 'pre-editing'.

Since Chinese and Spanish are very different languages, our assumption was that when a change in the English text solved a machine-translation problem in both languages, then it would probably do so for many other languages as well. This assumption would later be tested when we asked how many readers would prefer the raw machine translations of the pre-edited texts (Texts C in the reception tests).

To do the pre-editing in this case, the research team worked in Google Translate with live interfaces for English-Chinese and English-Spanish. This way, when we changed the English text, we could immediately see how the change affected the machine translations. We then used trial-and-error until we had solutions that worked for both languages.

The comparison below shows how we dealt with the main problem in the raw machine translations. The original *If you are caught in fire in your car* became *If you are in your car and you cannot escape the fire*. That is, we made the relations very explicit (Table 3).

Table 3. Comparison between original English source text message and a pre-edited version of the same message.

Original English source text	Pre-edited English source text
If you are travelling, do not enter the warning area. Make a U-turn and travel back to safety. If you are currently driving, slow down and turn on your headlights. Smoke will make it difficult to see. **If you are caught in fire in your car**: Park off the road behind a solid structure to block the fire's heat or pull over to cleared area. Try to position the car towards the approaching fire.	If you are travelling, do not enter the warning area. **Go back** and travel to **a safe location**. If you are currently driving, **reduce speed** and turn on your headlights. Smoke will **lower visibility**. **If you are in your car and you cannot escape the fire**: Park off the road behind a solid structure to block the fire's heat or pull over to cleared area. Try to **park the car with the front towards the fire that is coming towards you.**

The comparison also shows that changes were necessary in sentences that seem very clear to the English reader. The expression *u-turn* did not translate well in Spanish, so it became 'go back'; the structure *will make it difficult to see* created problems because the algorithms struggle to identify what *it* refers to, so it became 'lower visibility'; the two-word verb *position towards*, which machine translation turned into 'park near the fire' in Chinese, was similarly made very explicit in the pre-edited version.

The purpose of the pre-editing is to remove idiomatic English terms (e.g., *u-turn*) and context-dependent expressions (e.g., *back to safety*) which create problems for machine translation. In theory, it can thus lead to reasonably correct machine translations in multiple languages. However, pre-editing also requires awareness of grammatical differences across various languages. For instance, as noted, *position* in English can be either a noun (a position) or a verb (to position a car), while in Chinese the closest word 位置 is always a noun. Therefore, pre-editing involves extra work for the translator to find an alternative expression. This also means that stylistic problems such as repetition (as in 'towards… towards' in the last sentence in the above comparison) are not considered negatively—clarity is the primary aim.

As noted above, an interesting problem was the second person *you*, where the raw machine translations mix the formal and informal in Spanish, Greek and Chinese. In Spanish, for example, the problem is not just a mixing of the formal *usted* and the informal *tú*: the second person is marked in all conjugated verbs, so the problem concerns all instructions like 'Park…' or 'Try to…', even when the word *you* does not appear in the English text. It is possible to rewrite the English so that the second person is not used at all (as in 'One must…' or 'It is advisable to…'), which would solve the translation problem. However, one of the consistent recommendations for emergency texts is that the second person should be used, given that it addresses the reader directly and thus has an impact (Sengupta, et al., 2024). We therefore retained the second-person structures in cases where the receiver was directly instructed to take an action. We judged that the resulting problems in the Spanish, Greek and Chinese

machine translations would concern stylistics, not comprehension or actionability. As it happened, once we made the text more explicit and thus more formal in structure, many of the machine translations in Spanish shifted from the informal *tú* to the formal *usted*, even when we had not planned for that to happen. A similar improvement was found in Chinese, with a shift from the informal 你 to the formal 您.

Since this pre-editing was a trial-and-error process involving three languages, it took much more time than the post-editing of the machine translations. However, it ideally only has to be done once for translations into many languages. As a general rule (and bearing in mind that it depends on the text and the languages), **if a text is to be translated into two languages or more, then pre-editing will be more time-efficient than post-editing**. Further, when pre-editing and post-editing are combined, the time required for fixing erroneous machine translations can be reduced considerably.

Once we had completed the pre-editing, we organised our changes around a set of general principles, outlined under 1.4. When those principles are applied in the writing of the English text, especially templates, then the process is technically called 'controlled authoring'. Its benefits are not only in time savings when translating, which can be considerable, but also in improved comprehension and actionability for the texts when read in English.

5.5. Evaluations by speakers of the languages

From the above, we had three translations (raw, post-edited and pre-edited) for each of the three original texts (bushfire, sharks and water quality). For each language, those nine texts were then read and assessed by speakers of the languages, and we then discussed the translations in interviews. Each interview lasted for an average of just over an hour. We were particularly interested in the risks presented by the raw machine translations (Texts A) and the degrees to which those risks were mitigated by the post-edited machine translations (Texts B) and machine translations of pre-edited originals (Texts C).

5.5.1. *Chinese*

The Chinese translations of the three texts were tested in a postgraduate translation class of 44 students (the pilot study), three Chinese international students with mixed academic backgrounds (two women and one man, all in their late 20s), and three community members who had lived in Australia for over a decade (two women in their 30s and 50s respectively, one man in his 30s). All participants were native speakers of Chinese. The six interviews were carried out separately in Melbourne.

In the pilot study, the students were invited to read Texts A (raw machine translations) and then answer the corresponding questions. Many questions seemed challenging for them. The most problematic was the question about when you should ‘position your car

towards the approaching fire', to which only 9 out of 44 (18%) gave the right answer. This low percentage can be explained by the fact that machine translation introduced a serious distortion of the original meaning: 'position your car towards the fire' was mistakenly rendered as 'park your car next to a fire' (see the analysis of this example in Section 5.3.1). Almost two-thirds of the students (n=25, 57%) chose 'not sure', which is not surprising since they could not find any relevant information in Texts A. Around 25% of the students (n=11) were misled by the machine translation and chose the most counterintuitive option 'when parking your car next to a place that is on fire'.

Other translation problems reported above also seemed to hinder the participants' understanding of the emergency messages. For instance, while most students (n=33, 70%) understood that the aim of regular testing is to identify changes that indicate improved water quality, some students (n=3, 7%) were not sure, and six students (14%) chose what the machine translation had told them: 'the new round of outbreaks of blue-green algae indicating improved water quality', which is again counterintuitive.

Six participants were invited to read the three versions of the three texts. When they were asked to vote (3 texts x 6 participants = 18 votes), Texts C (based on pre-edited texts) received half of the votes (9 out of 18) because readers found the message clear and the texts easy to read. This was presumably because the *English-language* texts had been rewritten in a clear way and thus gave concise sentence structures in the Chinese translations. Texts B received 7 votes, favoured for their natural flow resembling human translation, indicating the effectiveness of post-editing. Surprisingly, Texts A (raw machine translations) of the shark text obtained two votes. Readers perceived it as adhering more closely to the English text than the other two translations, even though the participants agreed that a Text A 'reads most like a machine-generated version'. This preference stemmed from the expectation that translated official notices should adhere closely to the source text. The Chinese readers' tolerance of stylistic problems suggests that, for this community, translated emergency messages should follow the English texts as closely as possible.

Most of the Chinese interviewees gave correct answers to these questions, except for the problematic one about 'position your car towards'. Only one out of six participants chose the right answer, while two-thirds were confused even after spending quite a lot of time going over the raw machine translation. All participants managed to resolve the points of confusion after reading Text B (post-edited texts) and Text C (machine translations based on pre-edited texts).

In all three kinds of texts, although the Chinese participants did not pick up on typos or grammatical errors, they occasionally commented on unidiomatic expressions: 'I think I get it, but it does not sound right'. They did not find the switch between the formal 您

and informal 你 to be a major issue, although several participants asked during the interviews whether Texts A were machine translations.

At the same time, we observed that the interviewees demonstrated some preferences and expectations with respect to the kind of language that should be used in emergency texts. For instance, some did not like seeing key information (such as directives and instructions) buried in a large chunk of text. They preferred sentences that placed the action verb at the very beginning (‘Give me the verb!'), followed by contextual and explanatory information, as is the norm in Chinese. Similarly, in emergency communication, the participants expected to see texts in a formal register using a firm tone of voice, which they believed to be an indication of an official message from an authoritative source. However, one Chinese participant noted:

> *I want to see texts in a high register with an assertive tone, because it sounds like the message talks about a serious matter. But I understand some people from our community may prefer a message that is more straightforward and down to earth*.

In addition, most Chinese participants tended to show zero tolerance for ambiguity in emergency messages, especially regarding food security and life-or-death matters such as bushfires.

5.5.2. *Greek*

The focus group for the translations into Greek comprised four bilingual women who are all tertiary-educated and are involved in community organisations.

They gave no wrong answers to any of the comprehension questions based on the raw machine translation, except for the question based on the mistranslation of ‘your car is on fire’, where they all answered ‘Not sure’. They said they answered many of the questions on the basis of conjecture and common sense.

They all found that the raw machine translations (Texts A) had numerous linguistic problems, commenting that they seemed to have been translated ‘word for word’. Several participants questioned the formal register used in Texts A and sometimes Texts B and pointed out that choice of specific words was too academic, too sophisticated, and sometimes archaic. This was the case for the Greek equivalents of *structure* and *difficult visibility,* where more informal alternatives were suggested. Such cases were considered to be a source of difficulty or confusion for the lay Greek person reading the text. Texts C were perceived to have a simpler sentence structure, which added comprehension:

> *The third one [Text C] is always like logical, sequential in in understanding the immediate threat. It also spells it out with a little bit more clarity.*

When asked to choose between the three versions of the three translations (3 texts x 4 readers = 12 votes), most of the preferences were for Texts C (the results of pre-editing) because the shorter sentences gave greater clarity and the pre-editing resulted in words that were more in common usage. Texts B (post-edited translations) nevertheless gained three votes, indicating that the post-editing achieved acceptable results.

5.5.3. *Spanish*

As mentioned, the translation of the bushfire text into Spanish was tested by a university class of 29 Spanish speakers in Spain (the pilot study of the bushfire text), two accredited Spanish-English translators in Spain (both women in their 30s), and three Spanish teachers from Latin America in Melbourne (two men and one woman, all in their 30s).

In the pilot study, the Spanish students answered all questions correctly except for the one about 'If you are caught in fire in your car', where only 12 of the 29 (41%) gave the right answer. Since the questionnaire was administered electronically, we could see that the students took about eight times longer to answer this question than any other question, indicating that the mistranslation had seriously blocked sense-making. To allow such situations of extended doubt to be detected in our data, where the translation consumed time but would be unlikely to lead to a wrong action, we decided to add a 'Not sure' option to all questions for future interviews. In the ensuing general discussion, it took the class some two minutes to note the abovementioned problem with the mixes of the formal and informal second person.

This general pattern held for the other participants, who gave very few wrong answers but opted for the 'Not sure' option for several of the questions when they were looking at the raw machine translation (Texts A), especially the mistranslation of 'the car on fire'. In many cases, the right answers were given based on common sense. One of the Spanish translators commented: 'It's all wrong, but we understand it perfectly—we are used to machine translations'. All the Spanish translators and teachers found numerous lexical and syntactic problems in the raw machine translation, even though they could generally understand it and act upon it. Regional varieties tended not to be a problem. Only one word was picked up by a Latin American participant as being used in Spain but not in the rest of the Spanish-speaking world: *jersey* as the term for a woolen jumper. Otherwise, there were no problems. For example, a car is a *coche* in Spain, but a *carro* or *auto* in Latin America. Google Translate gave the formal *automóvil*, which would not be anyone's most natural option but was understood by all.

The two male language teachers were very concerned about the switches between the formal and informal second persons, both in the raw machine translation and in the result of the pre-editing (the problem did not appear in the post-edited texts B). They argued that this stylistic problem was important because it created 'cognitive

dissonance' and 'distracted from the information'. For this reason, they preferred Texts B, the post-edited versions, which had a consistent second person. The translators nevertheless preferred Texts C, the results of pre-editing, because of the greater clarity.

5.5.4. *Dari*

Three speakers of Dari were interviewed separately: a woman in her 30s, and two men in their early 20s and late 40s respectively. As indicated in the times taken for the post-editing (Appendix B), the machine translations were much more problematic than those into the other languages. This was partly because the electronic resources for Dari are not as developed as they are for the other languages we investigated, but also because the machine translation was into Persian, which has dialect differences with respect to Dari. We were therefore particularly interested in how the readers negotiated these differences.

We asked whether it was unusual for a Dari speaker to confront an emergency text in Persian. Two participants said they had never received emergency communication in either language variety, while the third had only seen one such communication—in Persian—on television during floods in Shepparton.

One participant said that he personally had no problem with the Persian, while a second estimated that Dari speakers would understand about 80-85% of the translations. The third said that 'if we use Dari, then we have to keep and stay in Dari. Um, so we should not actually use Persian because, it will be really hard, especially when it comes to emergency'. This participant added that the messaging should be spoken, not written:

> *First, they cannot read in the first language, so there's no way that they will read it. But the ones who can read it, as soon as they see a few mistakes, they kind of like just. Because you're an emergency, it will be good if they do it quickly audio message instead of just reading it.*

The participants' answers to the comprehension questions indicated significantly greater difficulties than was the case in the other languages.

For the bushfire questions, answers based on Text A were mostly correct except for the question about the car being on fire, where one reader was correct, another was incorrect but gave the correct answer after seeing Text C, and the third was not sure even after seeing all texts.

For the translations about sharks, there was at least one incorrect answer for each question, and for the question 'What brings the sharks?', there was only one correct answer, which came after seeing Text C. The results were similar for the translations about water quality, where there were incorrect answers for the question 'Why the analyses?', which were not corrected after seeing Text B and Text C.

As for the preferred translations, all three preferred Text B for the bushfire document and Text C for the water quality one, while for the shark activity document there were two votes for Text C and one vote for Text B. In all, Text C gained just one more vote than Text B.

These answers suggest that, in this case, the raw machine translations cannot be considered better than no translation at all, and the improvements brought about by post-editing and pre-editing do not mitigate all the risks. Given more general literacy concerns for this community, communication here would be better in spoken format, perhaps with the aid of a machine translation plus the original English, where doubts can be addressed and explained.

6. Recommendations

Our questions were if, when and how machine translation can be used for high-stakes texts in situations of urgency. Our initial survey of emergency texts suggested that, given the existence of templates for these texts, it was not efficient to extract translation memories or glossaries from previous translations, and we should therefore *not* envisage developing a dedicated machine translation system. Instead, we decided to focus mainly on ways of improving public machine translations through post-editing and pre-editing. On that basis, our study allows us to propose some answers to the basic questions.

6.1. Should raw machine translation be used?

Raw machine translations should *not* be used for emergency *texts as English-language source texts are currently written*. However, machine translations might be used as a last resort whenever adequate preparations have been made, especially the writing of clear source texts, the prior translation of templates, the use of human post-editing when possible, and the identification of appropriately selected bilingual community-based contact persons.

Working on raw machine-translated emergency texts, we found several errors that could have led the reader to take erroneous actions. We estimate that the consequences of those errors could be life-threatening. This means that, even though raw machine translations can provide substantial time savings, the associated risk of incorrect information can outweigh the benefits of those savings. Whenever possible, it is preferable to take the time to have the raw machine translations post-edited by a professional.

Given the current original source texts, we also found that the problems of machine translation are far greater for low-resource languages and cases where there are competing varieties of a language. It would be a mistake to believe that some of the excellent results achieved in translation between major languages can be generalised to all the languages spoken at homes in Victoria.

Raw machine translations might nevertheless be used as a last resort whenever the source texts are very clear, model templates have been translated by professionals, glossaries and translation memories have been developed and applied, and bilingual community-based contact persons have been identified to explain any comprehension problems.

6.2. How should emergency messages be written?

All emergency messaging should apply the basic principles of clear, explicit writing in the English source text.

Application of these principles can vastly improve the quality of machine translation, sometimes eliminating errors entirely. The clearer you write, the better the machine translations. The principles can also ensure that the messages are better understood in English.

6.3. When should human translation be used?

Certified translators should be employed whenever possible.

In particular, professionals should be employed to ensure high-quality translations of the templates that have been developed for emergency messages.

The selection of target languages for professionals to translate into should be based on the census data on the languages used in the areas at risk. This can also mean selecting the languages spoken by communities known to have difficulties in accessing information in English (on the principles of language triage, see Pym, 2023).

6.4. How can machine translation be combined with human translation?

Human translators can post-edit machine translation output, when there is time, and pre-edit source texts and templates, independently of time constraints.

Certified translators should be employed to correct (post-edit) machine translations whenever there is enough time to do so, with a pragmatic focus on effective (actionable) messaging within time constraints, rather than on complete accuracy.

Post-edited machine translations can also become the basis for professionals to pre-edit English-language source texts. This process should improve future machine translations for a wide range of target languages.

Another way to combine human and machine translation is to use translated templates as input for translation memories that can be used to override a public machine translation system (see 4.1 above and Appendix D below).

6.5. What workflows can integrate machine translation?

Raw machine translation should not be used in isolation from other communication solutions, especially the clear writing of original English-language texts, post-editing, and the identification of appropriately selected bilingual community contact persons.

In this study, we found that it was not cost-effective to extract translation memories from previous human translations of texts on the same topic. It was more efficient to work from the document templates that are currently being used (see Appendix D).

The ideal workflow for the integration of machine translation involves pre-editing of the original text (or the use of controlled authoring principles in templates), post-editing of the machine translation (when time is available), and community-based bilingual contact persons who can disseminate and explain the messaging in appropriate media. The contact persons should not be regarded as unpaid translators and should not be used to replace professional translators. That said, people who have academic qualifications in the language concerned, including NAATI qualifications, should be considered low-risk candidates for inclusion in lists of contact persons. There are many more people with language and translation qualifications than are actually employed as professional translators, creating a large overlap between the professional and non-professional groups. These contact persons could be viewed in the same way as Country Fire Authority volunteers organise lists of contact persons with telephone numbers, each of whom can contact further people. It would then be a matter of extending that approach to identify contact persons with sufficient language skills to alert end-users to the existence of the translated information and to clear up misunderstandings.

Of these factors, the most important are pre-editing and the identification of bilingual contact persons. These are tasks that are not affected by urgency.

The various steps that can be taken thus depend on the relative urgency of each situation. Figure 2 provides an overview of this relation.

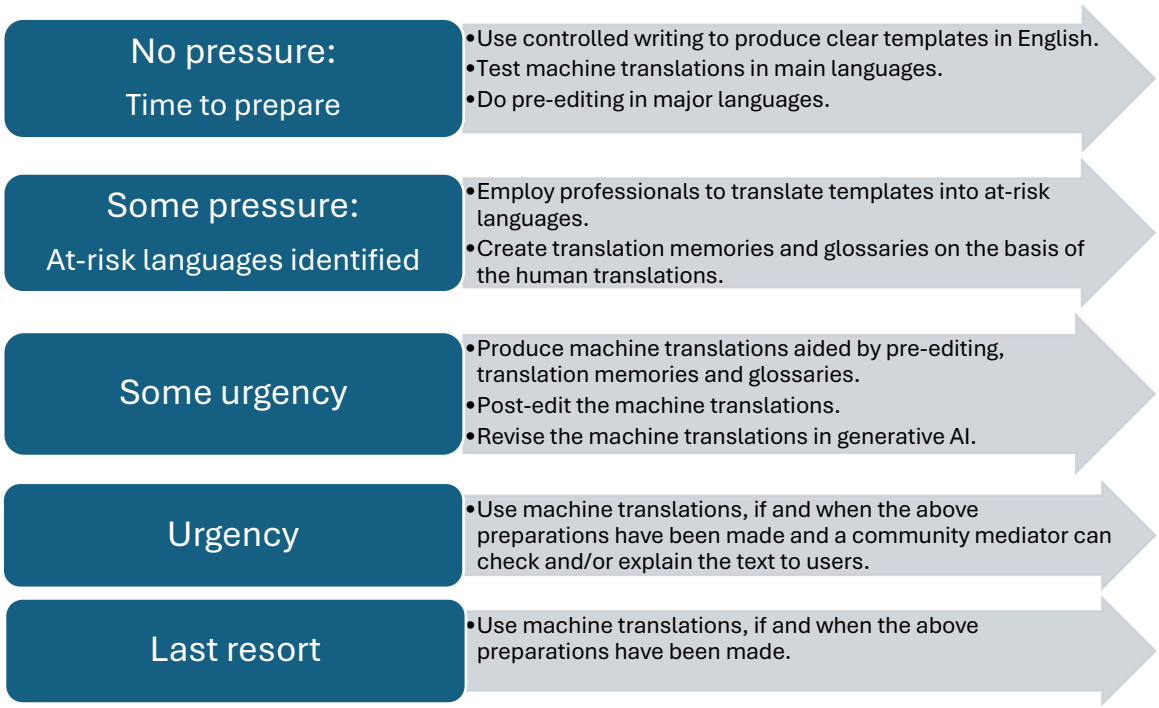

Figure 2. Translation strategies and steps, depending on urgency.

6.6. How should machine translations be labelled?

Raw machine translations should always be clearly labelled as such, so that consumers, service providers and translators are immediately aware of potential communication issues.

Machine translations that are based on pre-edited source texts should also be labelled with an appropriate warning. Even though they will have fewer errors than raw machine translations when the source text was not pre-edited, there is always the possibility that errors in low-resource target languages will compromise the text's actionability.

That said, we do not consider it necessary to label post-edited machine translations when a certified translator has done the post-editing.

6.7. How should we *not* evaluate machine translation?

For emergency messaging, a direct comparison of machine translations with human translations is inadequate, as it overlooks urgency issues.

Our comparisons here have thus been between raw machine translation, human post-edited translation, and raw machine translation based on pre-edited English texts. This more complex mode of evaluation is necessary in order to identify appropriate workflows.

6.8. Will translation automation improve?

Machine translation and generative AI are both improving in quality, but not to the extent that we can be sure that no high-stakes errors are made.

We noticed improvements in Google Translate in the period of our research, and we encountered problems (notably the second person *you* in Spanish, Greek and Chinese) that are handled better by generative AI. One can expect those improvements and advantages to increase. There is no guarantee, however, that the use of general public systems will entirely remove mistranslations, especially in the numerous languages that have few electronic resources.

We do not recommend waiting for perfect machine translations. It is more important to adopt measures that can improve the current use of machine translations.

Appendices

Appendix A: Original and pre-edited source texts

The following are the original English-language texts that we used in our tests, along with the pre-edited versions done on the basis of problems in the Chinese and Spanish machine translations. The main pre-edits are indicated in red. The pre-edited versions were then fed into Google Translate and the raw outputs were labelled Texts C in our interviews.

Original bushfire warning	Pre-edited bushfire warning
Emergency Warning—Bushfire—Take Shelter Now Issued Today at 10:25 AM. • There is a bushfire at Duffy Rd, Briagolong that is out of control. • The bushfire is travelling from Duffy Road in a south-easterly direction. • This fire is threatening homes and lives. • It is too late to leave the area safely so you must take shelter now. **You are in danger and need to act immediately to survive.** **The safest option is to take shelter indoors immediately. It is too late to leave.** **Leaving now would be deadly.** **What you should do:** **You should move indoors:** • Protect yourself by wearing long sleeves and trousers, made from pure cotton or wool. Wear leather boots. • Bring your pets inside. • Close all exterior doors, windows and vents and turn off cooling systems. • You must take shelter before the fire arrives. The extreme heat is likely to kill you well before the flames reach you. • Shelter in a room that has two exits, such as a door or window including one directly to the outside. It is important to be able to see outside so you know what is happening with the fire. **If your home catches on fire:** • Move away from the rooms that are on fire, closing doors behind you. • As soon as the bushfire has passed the house, or conditions inside become unbearable, you need to get out and go to an area that has already been burnt. Staying inside a burning building will almost certainly end in death. • It may still be too hot to remain outside, so you will need to seek shelter in another structure or last resort option.	**Emergency Warning—Bushfire—Take Shelter Now** Issued Today at 10:25 AM. • There is a bushfire at Duffy Rd, Briagolong, that is out of control. • The bushfire is travelling from Duffy Road in a south-easterly direction. • This fire is threatening homes and lives. • It is too late to leave the area safely so you must take shelter now. **You are in danger and need to act immediately to survive.** **The safest option is to find shelter inside a house or a shed immediately. It is too late for you to leave.** **Leaving now would be deadly.** **What to do:** **You should go inside a house or a shed:** • Protect yourself by wearing trousers and a shirt or jumper with long sleeves. The clothes should be made of pure cotton or wool. Wear leather boots. • Bring your pets inside. • Close all exterior doors, windows and vents and turn off all fans and air conditioning. • You must take shelter before the fire arrives. The extreme heat is likely to be deadly before the flames reach you. • Find shelter in a room that has two exits, such as a door or window, including one directly to the outside. It is important to be able to see outside so you know what is happening with the fire. **If your home catches on fire:** • Move away from the rooms that are on fire, closing doors behind you. • As soon as the bushfire has passed the house, or conditions inside become unbearable, you need to get out and go to an area that has been burnt. Staying inside a burning building will almost certainly end in death. • It may still be too hot to remain outside, so you will need to seek shelter in another structure.

If you cannot get indoors, last resort options include: • Shelter in the middle of a large open area like a ploughed paddock, football oval or sporting reserve. • Get into a large body of water like a dam, lake, river, the ocean or inground pool. • Try to protect yourself from the fire's heat. **If you are travelling:** • If you are travelling, do not enter the warning area. Make a u-turn and travel to safety. • If you are currently driving slow down and turn on your headlights. Smoke will make it difficult to see. **If you are caught in fire in your car:** Park off the road behind a solid structure to block the fire's heat or pull over to cleared area. • Try to position the car **towards** the approaching fire. • Turn on your hazard lights and headlights. • Close all windows. • Turn off the air-conditioning and shut all the air vents. • Turn your car engine off. • Get down as low as possible below window level and cover up with a pure woollen blanket. **Drinking water:** • Smoke and ash may impact your tank water. • To avoid contamination, block off water tanks until the roof and gutters have been flushed by rain. This message was issued by Forest Fire Management Victoria.	**If you cannot go inside a house or shed, the last available options include:** • Go to the middle of a large open area like a ploughed paddock or an area used for sports. • Go into a large body of water like a dam, lake, river, the ocean or inground pool. • Try to protect yourself from the fire's heat. **If you are travelling:** • If you are travelling, do not enter the warning area. Go back and travel to a safe location. • If you are currently driving, reduce speed and turn on your headlights. Smoke will lower the visibility. **If you are in your car and you cannot escape the fire:** Park off the road behind a solid structure to block the fire's heat or pull over to cleared area. • Try to park the car with the front towards the fire that is coming towards you. • Turn on your hazard lights and headlights. • Close all windows. • Turn off the air-conditioning and shut all the air vents. • Turn your car engine off. • Get down as low as possible below window level and put a pure woollen blanket over yourself. **Drinking water:** • Smoke and ash may impact your tank water. • To avoid contamination, block off water tanks until the roof and gutters have been flushed by rain. This message was issued by Forest Fire Management Victoria.

Original shark warning	Pre-edited shark warning
Issued Today at 12:27 AM. This **Advice** message is being issued for Lake Bunga, Lake Tyers Beach. • Due to parts of a whale carcass, there could be an increase in shark activity. • A whale carcass can attract sharks to the area and mean they are closer to the shore than normal. • While it is not uncommon for sharks to be present off the Victorian coast, you should exercise additional caution in the area. • This Advice replaces the Advice issued at 9:05 AM on Thursday 14 December 2023. **Avoid the area. Stay informed and do not enter the water at closed beaches.** **What you should do:** • If you see sharks, report the sighting by calling Triple Zero (000) or notify lifesavers immediately if you are at a patrolled beach. **Staying safe at nearby beaches:**	Issued Today at 12:27 AM. This **Advice is** being issued for Lake Bunga and Lake Tyers Beach. • Parts of a whale carcass are in the water. This could increase shark activity. • A whale carcass can attract sharks to the area and mean they are closer to the shore than normal. • While it is not uncommon for sharks to be present off the Victorian coast, you should exercise additional caution in the area. • This Advice replaces the Advice issued at 9:05 AM on Thursday 14 December 2023. **Avoid the area. Stay informed and do not enter the water at closed beaches.** **What to do:** • If you see sharks, report the sighting by calling Triple Zero (000) or notify lifesavers immediately if you are at a patrolled beach. **To stay safe at nearby beaches:**

• Always swim, dive or surf with a friend. • Swim between the red and yellow flags on patrolled beaches. • Don't swim in places where human or animal waste enters the water. • Keep away from large schools of fish, seals or other wildlife as these can attract sharks.	• Always go with a friend when you swim, dive or surf. • You must swim between the red and yellow flags on patrolled beaches. • If human or animal waste enters the water, do not swim in those places. • Large schools of fish, seals and other animals can attract sharks. Stay away.

Original water-quality warning	Pre-edited water-quality warning
Issued Last Tuesday at 3:14 PM. This message is being issued for water quality at Lake Bolac. • High levels of blue-green algae have been detected in the lake. • The algal bloom is expected to remain until there is significant rainfall to flush the lake or cooler conditions slow the algae. • Regular testing is being conducted to identify new outbreaks or changes that might indicate improved water quality. • This notification will be reviewed weekly and updated as the situation changes. **Warning signs have been installed at the lake.** **You should avoid direct contact with affected water in Lake Bolac.** **Do not swim in affected areas or use water for cooking, drinking, washing or showering. Boiling the water will not make it safe.** **What you should do:** • Visit Parks Victoria website (https://www.parks.vic.gov.au/places-to-see/parks/lake-bolac-highway-park) for more information. • Observe information signs posted at the lake. **Protect your health:** • Direct contact with water affected by blue green algae can cause skin irritation, sore eyes, ears and nose. If swallowed it can cause cramps, nausea and vomiting. • If you do come into contact with affected water, wash your skin immediately in clean cool water. • Boiling affected water does not make it safe to drink. Boiling water bursts the blue green algae cells and releases toxins into the water making it more likely that you will experience symptoms. • If you are experiencing any health issues after contact with affected water seek medical advice from your local doctor or Nurse-On-Call on 1300 60 60 24. **Protect your pets:** • Pet owners should prevent pets from drinking or having direct contact with contaminated water.	Issued Last Tuesday at 3:14 PM This message is being issued with reference to water quality at Lake Bolac. • High levels of blue-green algae have been detected in the lake. • The algal bloom will probably remain until there is significant rainfall to flush the lake or cooler temperatures to slow the growth of the algae. • Regular testing is being conducted to detect new outbreaks or identify changes that might indicate improved water quality. • This notification will be reviewed weekly and updated as the situation changes. **Warning notices have been installed at the lake.** **You should avoid direct contact with affected water in Lake Bolac.** **Do not swim in affected areas or use water from those areas for cooking, drinking, washing or showering. Boiling the water will not make it safe. Boiling the water will not make it safe to use.** **What you should do:** • Visit Parks Victoria website (https://www.parks.vic.gov.au/places-to-see/parks/lake-bolac-highway-park) for more information. • Observe information notices posted at the lake. **Protect your health:** • Direct contact with water affected by blue green algae can cause skin irritation, and pain in your eyes, ears and nose. If swallowed, it can cause cramps, nausea and vomiting. • If you do come into contact with affected water, wash your skin immediately in clean cool water. • Boiling affected water does not make it safe to drink. When water boils, it bursts the blue green algae cells and releases toxins into the water, which makes it more like that people will have symptoms. • If you are experiencing any health issues after contact with affected water, seek advice from your local medical professional or call Nurse-On-Call on 1300 60 60 24. **Protect your pets:** • Pet owners should prevent pets from drinking or having direct contact with contaminated water.

Agriculture information: • Irrigation water contaminated with blue-green algae should not be sprayed on vegetables and fruit, or come in contact with plants being grown for food during processing and packing. • Fruit and vegetables grown in the affected areas are safe to consume if farmers follow recommendations regarding blue-green algae **Impacts in your area:** • The water is green as a result of the algae bloom. **Boating and fishing:** • Boating and fishing is still allowed in these areas, however care needs to be taken. • Fish caught from affected areas should be rinsed and cleaned thoroughly in fresh water, and internal organs removed from the fish and discarded before eating. • Do not eat mussels, crayfish or yabbies caught from the affected area. This message was issued by Department of Energy, Environment and Climate Action.	**Agriculture information:** • Irrigation water contaminated with blue-green algae should not be sprayed on vegetables and fruit, or come into contact with plants that are being processed and packed to be consumed as food. • Fruit and vegetables grown in the affected areas are safe to consume if farmers follow recommendations regarding blue-green algae. **Impacts in your area:** • The water is green as a result of the algae bloom. **Boating and fishing:** • Boating and fishing is still allowed in these areas; however, care needs to be taken. • Fish caught from affected areas should be rinsed and cleaned thoroughly in fresh water and internal organs removed and discarded before the fish is eaten. • You should not eat shellfish caught from the affected area. This message was issued by Department of Energy, Environment and Climate Action.

Appendix B: Machine translation problems as indicated by post-editing

The following table indicates the time taken by post-editors to correct the main problems (in bold) in the raw machine translations. The times for Spanish are the means for three post-editors.

	Example in English	Time on task in minutes			
		Chinese	Spanish	Greek	Dari
Bushfire	'from **Duffy Road** in a south-easterly direction'	00:17	00:04	00:23	01:13
	'homes and **lives**'	--	00:06	--	01:09
	'It is **too late to leave** the area safely so you must **take shelter now**. '	00:13	00:21	00:41	04:28
	'Leaving now **would be deadly**'	--	00:05	00:39	02:10
	'**long sleeves** and trousers, **made from** pure cotton or wool'	--	00:06	--	01:47
	'**Cooling** system'	00:10	00:08	--	--
	'You must **take shelter before the fire arrives**'	00:23	--	00:17	--
	'An area that **has already been burnt**'	00:47	00:43	00:13	00:55
	'a **burning building**'	00:10	--	--	00:36
	'will **almost certainly** end in death'	00:13	00:14	00:42	00:17
	'Shelter in another structure or **last resort option**'	03:14	00:40	02:40	00:22
	'a ploughed **paddock**, football **oval** or **sporting reserve**'	01:01	00:39	00:57	00:47
	'Get into a large body of water like a dam, **lake**, **river**, the ocean or **inground pool**.'	--	00:15	--	01:44
	'Make a **u-turn**'	--	00:15	--	00:34
	'If you are **caught in fire in your car**.'	01:47	00:54	01:40	01:53
	'Park off the road behind a **solid structure** to block the fire's heat or pull over to **cleared area**.'	01:53	00:38	00:58	01:02
	'**Position** the car towards'	00:38	00:16	00:26	01:29
	'Turn your **car engine** off.'	--	--	--	00:29
	'**Get down as low as possible below** window level and **cover up** with a pure woollen blanket.'	00:23	01:43	00:20	03:47
	'**block off** water tanks **until** the roof and gutters have been flushed by rain.'	--	--	--	01:24
Sharks	'Issued for **Lake Bunga, Lake Tyers Beach**'	01:51	--	--	--
	'**Due to parts of** a whale carcass'	01:59	00:20	01:08	02:55
	'You should **exercise additional caution** in the area.'	--	--	--	01:57
	'Stay informed and do not enter the water at **closed beaches**.'	00:10	--	--	02:23
	'If you see sharks, report the sighting by calling **Triple Zero (000)** or notify **lifesavers** immediately if you are at a **patrolled beach**.'	00:09	00:24	--	04:07
	'Always swim, dive or surf **with a friend**'	--	--	00:27	00:40
	'**Large schools of** fish, seals, or other wild animals'	00:10	--	00:29	00:46
Water	'Identify new outbreaks or changes **that** might	01:05	--	--	03:12

quality	indicate improved water quality.'				
	'Do not swim in the affected areas or use **water** for cooking, drinking, washing, or **showering**'	00:50	00:17	--	03:02
	'Observe **information signs posted at the lake**'	00:09	--	--	03:50
	'or come in contact with **plants being grown for food during processing and packing**'	03:33	--	02:56	06:09
	'**Boating** and fishing'	--	--	00:20	--
	'Internal organs removed from the fish and discarded **before eating**'	00:19	--	00:16	03:35
	'Do not eat **mussels, crayfish or yabbies** caught from the affected area.'	01:01	00:20	00:24	00:15

Appendix C: Questions testing the actionability and comprehension of raw machine translations

The following are the questions we used in our interviews with readers to test the viability of the raw machine translations. Each question tests how a translation error was construed. Tests of actionability are marked with an asterisk (*). The correct answers are in bold.

Bushfire

1. Where should you take refuge immediately?*
    a. **Inside a house or shed**
    b. In a car to escape
    c. In a wardrobe
    d. In the interior of the country
    e. Not sure
2. What should you do with long sleeves?*
    a. Take them off because they could burn.
    b. **Put on a shirt with long sleeves.**
    c. Use the sleeves to put out the fire.
    d. Use the sleeves to remove embers.
    e. Not sure.
3. If you are travelling into a warning zone, what should you do?*
    a. **Turn around and go in the opposite direction.**
    b. Find a church to pray in.
    c. Park somewhere.
    d. Enter the warning zone.
    e. Not sure.
4. When should you position your car to face the fire?*
    a. When you are as low as possible below the level of the windows.
    b. **When the fire is very close to the car and there is no time to escape by driving the car.**
    c. When you are in a cleared area.
    d. When you have parked your car next to the fire.
    e. Not sure.

Sharks

1. Where has there been an increase in shark activity?
    a. **Lake Bunga and Lake Tyers Beach**

b. Lake Bunga, which is part of the wider Lake Tyers Beach area
c. All patrolled beaches in Victoria
d. Areas along the coast between Lake Bunga and Lake Tyers Beach
e. Not sure

2. What should you do to stay safe in the affected areas when engaging in water activities?*
    a. Surf with a friend, but swim or dive alone.
    b. **Go with a friend for all water activities.**
    c. Report any friend who swims, dives, or surfs alone by calling 000.
    d. Swim in places where human or animal waste enters the water.
    e. Not sure.
3. Which of the following is ***not*** a cause of increased shark activity?
    a. A whale carcass in the water
    b. Large schools of fish and seals
    c. The shouting of a swimmer in trouble
    d. **Lifesavers on the beach**
    e. Not sure

Water quality

1. Why is regular testing being conducted?
    a. To identify new outbreaks that improve water quality.
    b. To check if people see the warning signs.
    c. **To see changes that improve water quality.**
    d. To test the internal organs of the fish.
    e. Not sure.
2. What should you do after the outbreak of blue-green algae?*
    a. Throw away the internal organs of the fish before you eat them.
    b. Boil the water from the affected areas before using it.
    c. **Keep your pet away from drinking or entering the water.**
    d. Clean the shellfish caught from the affected area thoroughly before you eat them.
    e. Not sure.
3. Which of the following food items from the affected areas are safe to consume?*
    a. Mussels, crayfish, or yabbies cleaned in fresh water
    b. Fruit and vegetables
    c. Processed and packed edible plants
    d. **Gutted fish cleaned in fresh water.**
    e. Not sure

Appendix D: Example of a template for emergency texts

The following is an example page from the templates currently used by Emergency Management Victoria.

EMCOP Publishing
Community Notification Template

Critical Details	Drag & Drop
This **Emergency Warning** is being issued for [warning_locality]. There is a bushfire at [warning_incident_location] that is [status]. The bushfire is travelling from [location] in a [direction] direction towards [location]. The fire started [on/near] [road/intersection] in [location]. This fire is threatening homes and lives. It is too late to leave the area safely so you must take shelter now.	This message is for people at [location]. This bushfire could impact [location] any time between [time] and [time]. This Emergency Warning replaces the [warning level] issued at [time]. [Weather conditions have changed /Unexpected conditions have occurred]. The fire has now crossed road/ landmark and is moving towards road/landmark. A wind change is expected around (time), which will cause the fire to change directions towards (location/landmark). Conditions can become very dangerous and unpredictable. Firefighters have been unable to stop the fire and it has now crossed road/ landmark and is moving towards road/landmark. The wind has now changed direction and the fire is headed towards [location]. If you have not already left, the time to safely evacuate has now passed. Take shelter indoors immediately. It is now too dangerous to leave.

References

Angermeyer, P. S. (2017). Controlling Roma refugees with 'Google-Hungarian': Indexing deviance, contempt, and belonging in Toronto's linguistic landscape. *Language in Society*, *46*(2), 159–183. https://doi.org/10.1017/S0047404516001020

ASTM. (2023). F2575. *Standard practice for language translation*. https://www.astm.org/f2575-23.html

Bowker, L. (2009). Can machine translation meet the needs of official language minority communities in Canada? A recipient evaluation. *Linguistica Antverpiensia*, *8*, 123–155. https://doi.org/10.52034/lanstts.v8i.248

Bowker, L., & Buitrago Ciro, J. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing. https://doi.org/10.1108/9781787567214

Cadwell, P. (2020). Trust, distrust and translation in a disaster. *Disaster Prevention and Management*, *29*(2), 157-174. https://doi.org/10.1108/DPM-11-2018-0374

Castilho, S., Mallon, C., Meister, R., & Yue, S. (2023). Do online machine translation systems care for context? What about a GPT model? *24th Annual Conference of the European Association for Machine Translation (EAMT 2023)*, European Association for Machine Translation.

Department of Home Affairs. (2019). *Australian Government language services guidelines: Supporting access and equity for people with limited English*. Commonwealth of Australia. https://immi.homeaffairs.gov.au/settlement-services-subsite/files/language-services-guidelines.pdf

Department of Premier and Cabinet. (2017). *Effective translations: Victorian Government guidelines on policy and procedures*. Victorian State Government. https://www.vic.gov.au/using-translation-services

European Commission. (2023). *eTranslation: The European Commission's machine translation system*. European Commission Directorate-General for Communication. https://commission.europa.eu/resources-partners/etranslation_en

Ghosh, S., & Caliskan, A. (2023). ChatGPT perpetuates gender bias in machine translation and ignores non-gendered pronouns: Findings across Bengali and five other low-resource languages. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society,* 901-912. https://doi.org/10.1145/3600211.3604672

Government of Singapore. (2024, May 16). *SG Translate Together*. https://www.sgtranslatetogether.gov.sg/

Hajek, J., Karidakis, M., Amorati, R., Hao, Y, Sengupta, M., Pym, A., & Woodward-Kron, R. (2022). *Understanding the experiences and communication needs of culturally and linguistically diverse communities during the COVID-19 pandemic*. The University of Melbourne, report prepared for the Victorian Department of Families, Fairness and Housing.

Hale, S., & Liddicoat, A. (2015). The meaning of accuracy and culture, and the rise of the machine in interpreting and translation: A conversation between Sandra Hale and Anthony Liddicoat. *Cultus: The Journal of Intercultural Mediation and Communication*, *8*, 14–26.

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Young, J. K., Afify, M., & Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation (arXiv:2302.09210*)*. *arXiv*. https://doi.org/10.48550/arXiv.2302.09210

Hester, V., Shaw, A., & Biewald, L. (2010). Scalable crisis relief: Crowdsourced SMS translation and categorization with Mission 4636. *Proceedings of the First ACM Symposium on Computing for Development*, 1–7. https://doi.org/10.1145/1926180.1926199

Hu, C., Bederson, B. B., & Resnik, P. (2010). Translation by iterative collaboration between monolingual users. *Proceedings of Graphics Interface 2010*, 39–46. https://doi.org/10.1145/1837885.1837902

Hu, Y., Chen, C., Yang, C.-H. H., Li, R., Zhang, D., Chen, Z., & Chng, E. S. (2024). GenTranslate: Large language models are generative multilingual speech and machine translators (arXiv:2402.06894). arXiv. https://doi.org/10.48550/arXiv.2402.06894

International Organization for Standardization (ISO). (2024). *ISO 11669:2024: Translation projects. General guidance*. https://www.iso.org/standard/79089.html

Jiao, W., Wang, W., Huang, J., Wang, X, Shi, S., & Tu, Z. (2023) Is ChatGPT a good translator? Yes, with GPT-4 as the engine (arXiv:2301.08745). arXiv. https://doi.org/10.48550/arXiv.2301.08745

Karidakis, M., Woodward-Kron, R., Amorati, R., Hu, B., Pym, A., & Hajek, J. (2022). Enhancing COVID-19 public health communication for culturally and linguistically diverse communities: An Australian interview study with community representatives. *Qualitative Health Communication*, *1*, 61-83. https://doi.org/10.7146/qhc.v1i1.127258

Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist (arXiv:2304.03245). *ArXiv*. https://doi.org/10.48550/arXiv.2304.03245

Lewis, W. D. (2010). Haitian Creole: How to build and ship an MT engine from scratch in 4 days, 17h, & 30 min. *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, 27–28 May 2010.

Melby, A. & Lester, G. (2024). *Consumer protection labels for translations*. https://www.tranquality.info/wp-content/uploads/2024/04/ConsumerProtectionLabels-for-Translations.pdf

Na, H., Wang, Z., Maimaiti, M., Chen, T., Wang, W., Shen, T., & Chen, L. (2024). Rethinking human-like translation strategy: Integrating Drift-Diffusion model with large language models for machine translation (arXiv:2402.10699). *arXiv*. http://arxiv.org/abs/2402.10699

O'Mara, B., & Carey, G. (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. https://doi.org/10.1111/1467-8500.12388

Pokorn, N. K., & Čibej, J. (2018). 'It's so vital to learn Slovene': Mediation choices by asylum seekers in Slovenia. *Language Problems and Language Planning*, *42*(3), 288–307. https://doi.org/10.1075/lplp.00024.pok

Pym, A. (2023). Triage and technology in healthcare translation. In G. Palumbo, K. Peruzzo & G. Pontrandolfo (Eds.) *What's Special about Specialised Translation?* (pp. 247-268). Peter Lang.

Pym, A., Ayvazyan, N., & Prioleau, J. (2022). Should raw machine translation be used for public-health information? Suggestions for a multilingual communication policy in Catalonia. *Just. Journal of Language Rights & Minorities, Revista de Drets Lingüístics i Minories*, *1*(1), 71–99. https://doi.org/10.7203/Just.1.24880

Pym, A., Ayvazyan, N., González Núñez, G., Macreadie, R., Takeda, K., & Zhou, S. (2024). *Language policy on machine translation and generative AI in government communications*. Humanities Commons. https://hcommons.org/deposits/item/hc:64111/

Pym, A., and Hao, Y. (2024). *How to Augment Language Skills. Generative AI and Machine Translation into Language Learning and Translator Training*. Routledge.

Pym, A. & Hu, K. (2024). Machine translation. An Overview. In H. Nesi & P. Milin (Eds.) *International Encyclopedia of Language and Linguistics*. 3rd Edition. Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00014-4

Sengupta, M., Pym, A., Hao, Y., Hajek, J., Karidakis, M., Woodward-Kron, R., & Amorati, R. (2024). On the transcreation, format and actionability of healthcare translations. *The*

International Journal for Translating and Interpreting Research, 16(1), 121-141. https://doi.org/10.12807/ti.116201.2024.a07

Victorian Department of Families, Fairness and Housing (2023). *Better practice guide for multicultural communications*. https://content.vic.gov.au/sites/default/files/2023-08/Better-practice-guide-for-multicultural-communications.pdf

Uekusa, S., & Lee, S. (2023). Exploring the role of language ideology in disaster contexts: Case study of the 2011 Great East Japan Earthquake and Tsunami. *International Journal of the Sociology of Language*, *2023*(284), 37–57. https://doi.org/10.1515/ijsl-2022-0061

United States Government. (2022, August 4). Nondiscrimination in Health Programs and Activities. 87 FR 47824 (to be codified at 42 CFR 438, 440, 457, 460, 80, 84, 86, 91, 92, 147, 155, & 156) (pp. 47824-47920). https://www.federalregister.gov/documents/2022/08/04/2022-16217/nondiscrimination-in-health-programs-and-activities

United States Government. (2023, May 11). *Introduction to translation technology*. Digital.Gov. https://digital.gov/resources/introduction-to-translation-technology/