



23 August 2017

Via ECFS

Ms. Marlene H. Dortch
Secretary
Federal Communications Commission
445 Twelfth Street, S.W.
Washington, DC 20554

Re: Notice of *Ex Parte* Communication
Wireless Emergency Alerts, PS Docket No. 15-91

Dear Ms. Dortch:

On 21 August 2017, the undersigned along with Chris Wendt and Paula Boyd, all from Microsoft Corporation, spoke by telephone with Megan Henry and Linda Nagel from the Commission's Public Safety and Homeland Security Bureau (PSHSB) to discuss matters pertaining to auto-translation and improvements to the wireless emergency alert (WEA) system under consideration in the above-referenced docket. Please find below a summary of that conversation.

In an emergency context, auto-translation should be conducted by the alert originator rather than the device. When the WEA system is used during an emergency, having alert originators perform the translation, including using auto-translation, is preferred to a device-based auto-translation system. Device-based translation that employs auto-translation sites or apps would require the use of an internet connection and could generate both network congestion and congestion of auto-translation sites during intense, geographically-concentrated use. In addition, the translation capability would depend upon the availability of a device's data connection. By contrast, an alert originator's auto-translation would not impose the same heavy demand on the data network or auto-translation service servers. The variety of translations could be sent to devices and the message could be displayed according to the user's language preference (assuming the device supported the font needed for the chosen language). A device must have downloaded the requisite font in order to display alert translation into a language that does not use Latin characters.

Different language pairs result in different auto-translation quality levels. When using auto-translation, there are significant quality differences between language pairs which are influenced by the similarity of the source language and the target language. In addition, given that machine translation is an application of machine learning, the amount and quality of

available previously translated material that the system has consumed will affect the quality of the translation. Another factor contributing to quality levels is the suitability of the algorithms being employed for the particular language pair.

- English-Spanish is the most easily machine translatable language pair because there is a large amount of quality training material available and the two languages are fairly similar in grammar.
- Regardless of the particular language pair, suitability of the original text is an important factor in auto-translation. A degradation in translation quality can arise from typographic errors, missing punctuation, and the like. The suitability of the original text could be reduced *e.g.*, it might be more error-prone, when it is being produced in a high-pressure, time-sensitive situation such as an emergency which, in turn, could degrade translation accuracy or understandability.
- Nothing can guarantee 100 percent accuracy, and there is the potential for error in translating mission-critical instructions (*e.g.*, “do not drive west” could be mistranslated into “drive west” or “do not drive east”). A slight error that is otherwise acceptable in non-emergency circumstances could have severe consequences for life-saving instructions during an emergency. A wireless emergency alert system should seek to minimize content errors, particularly for those alerts that could affect personal safety.

Human judging is used to test the understandability of auto-translation. Human-judged tests constitute the baseline for determining general understandability of auto-translation, which considers semantic accuracy, fluency, and lexicon. Microsoft uses a 1-4 scoring system where 1 means not understandable and 4 means perfect. Microsoft’s scoring system does not distinguish between an understandable but wrong translation, on the one hand, and a translation that is not understandable, on the other; both would receive a score of 1.

- For a prediction of auto-translation behavior for specific language pairs in emergency alerts, Microsoft suggests performing a test with a representative sampling of emergency alert messages, using a minimum of 200 messages and, although expensive, human judges.
- Although human-judged tests set the baseline for accuracy determinations, Microsoft uses the algorithmic bilingual evaluation understudy (BLEU) system for its daily testing and improvement. The BLEU system generates a score between 0 and 100 (100 meaning that every word in the auto-translation was exactly the same *and* in the same place as the reference translation) for the relative difference between machine translation output and one or more human translations. This system is not a measure of understandability, but allows Microsoft to determine whether or not improvements result from changes made to the system.

Specialized auto-translation. Auto-translation systems can be created that specialize in terminology relevant to a particular domain, such as weather. These systems statistically elevate

Ms. Marlene H. Dortch

23 August 2017

Page 3 of 3

the terminology relevant to that domain so that when the computer makes a probabilistic determination between choices of different words and phrases, it will prefer those more closely related to the identified domain (in this example, words and phrases having to do with weather). However, this approach on its own will not necessarily result in translations with sufficiently reliable accuracy to be utilized for mission-critical messages and instructions during an emergency.

Respectfully submitted,



Gunnar Halley
Senior Attorney, CELA - Regulatory Affairs
Microsoft Corporation

cc (via e-mail): Megan Henry
Linda Nagel