Monotrans: Human-Computer Collaborative Translation

Chang Hu[1,2], Benjamin B. Bederson[1,2] and Philip Resnik[1,3,4]
[1]Department of Computer Science, [2]Human-Computer Interaction Lab,
[3]Institute for Advanced Computer Studies, [4]Department of Linguistics,
University of Maryland
Contact: {changhu, bederson}@cs.umd.edu, resnik@umiacs.umd.edu
www.cs.umd.edu/hcil/monotrans

An enormous potential exists for solving certain classes of computational problems through rich collaboration between humans and computers. Take translation for example. Humans alone are expensive and can be surprisingly slow; despite significant recent advances, machine translation (MT) remains a crucial problem and fully automated high quality translation remains a distant dream for the vast majority of the world's language pairs. Usable translation quality can sometimes be obtained by statistical MT systems, but only for a minority of language pairs, and only in use cases where sufficient training text is available and the material being translated is reasonably similar to the material on which the system was trained.

Using the Web to reach non-professional human translators holds promise, and there has been some initial success with distributing translation over a crowd of bilingual users. However, compared to the total user population, the potential translator population is still small. For example, while Wikipedia currently has about 75,000 active contributors, there are fewer than 800 translators. With a much larger number of potential human helpers who speak only the source or target language, but not both, it seems natural to ask whether some combination of machine translation with volunteer *monolingual* speakers could result in high quality translation.

We propose a rethinking of the translation problem to bring together translation technology and human-computer interaction, producing a framework for translation that will exploit imperfect technology and limited human abilities in tandem to achieve capabilities neither can achieve alone.

The core of this framework is Monotrans, an iterative protocol in which the human participants work together to make sense of machine translations, and introduce redundant information to make their intended meanings clearer (Figure 1). This protocol makes it possible to detect and correct some translation errors, and to at least identify some passages that have errors even if they are not correctable given the available information. For example, "has cheeseburger" is a detectable error, even if it is not clear whether the intended meaning was "has cheeseburgers" or "have a cheeseburger". Back-translating a refinement and carrying along redundant information, e.g. a *picture* of multiple cheeseburgers, might help convey which of those alternatives the English speaker guessed, presenting the opportunity for confirmation or further correction.

Research Prototype

We built a research prototype as a multi-user web application (Figure 2). When a user logs in, the UI displays a book page in the user's language. Every sentence in the

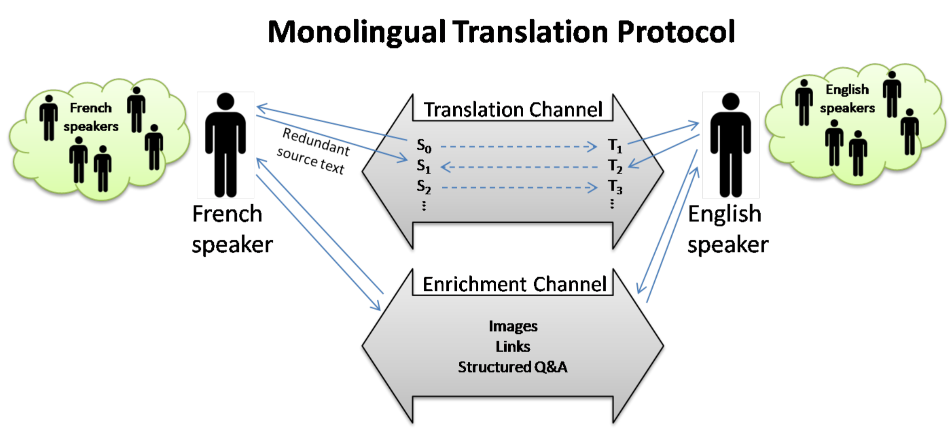

Figure 1 – Round-trip protocol. Dashed arrows show machine translation and solid arrows show human editing.

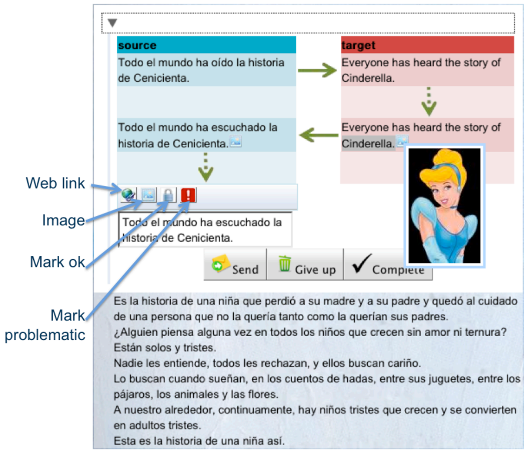

Figure 2 – In this prototype UI showing a page, a sentence is expanded, with an image attached to a word.

page is displayed with the most up-to-date translation hypothesis (or corresponding back-translation). The user can navigate through all available pages with navigation controls, or expand a sentence translation to edit it.

When a sentence is expanded, the UI shows the most up-to-date translation hypothesis with all previous translation hypotheses of this sentence and a rich editor where the sentence can be edited and annotated.

The rich editor currently includes the following elements for the enrichment channel, aimed at enhancing redundancy and communicating shared context.

- Image annotations
- Web link annotations
- Annotation of correct and incorrect parts of a sentence

Word level alignments necessary to perform annotation projection can be obtained from our own machine translation engines. Also, some machine translation services make word alignment information publically available to researchers along with the translation hypotheses. The Google Translate Research API (open to university research projects) is one such example.

Preliminary Results

We used MonoTrans to translate part of a children's book from Russian to Chinese. Chinese and Russian are commonly spoken languages in the world. However, they make good experimental candidates because they are very different from the perspective of linguistic typology.

In the experiments, two Russian speakers and four Chinese speakers formed four pairs to use the prototype. (One Russian speaker participated three times, with different content.) The participants were all native speakers of one language and had no knowledge of the other. They were all computer-literate and fluent speakers of English. While

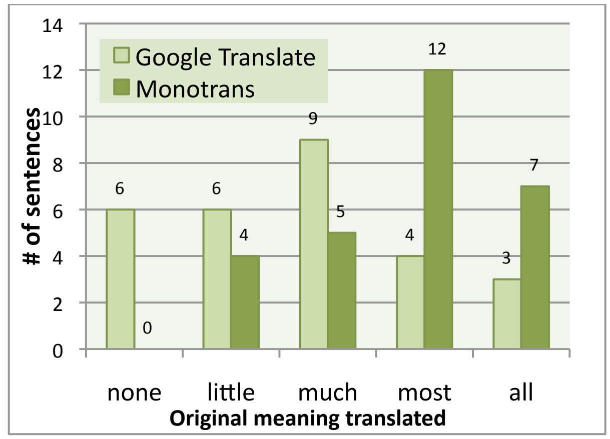

Figure 3 - Number of sentences in each adequacy category, rated by a professional translator.

most of the participants were computer science students and researchers, none of them work in the area of machine translation directly, and none of them were familiar with the details of this project. They were not linguists or linguistic students.

Participants worked on 6 pages (a total of 44 sentences) and finished translating 28 of them. This works out to approximately seven sentences per hour between any given pair of participants. It is about five times faster than the earlier "Wizard of Oz" experiment. With a standard rating procedure, sixteen of the 28 sentences translated with the prototype were rated as fully fluent and nineteen sentences of the 28 were rated as mostly or fully translated, by a professional translator not connected with the project.

The shift in adequacy is especially notable among these results. Completely inadequate MT outputs (none of the meaning preserved) dropped from 6 to 0. This means that the protocol helped the target language participants understand at least some of the meaning even when the original MT output quality was especially low and they had little to go on. In a coarse-grained way of thinking, if the adequacy rating could be categorized so that {none, little}=*bad* and {most, all}=*good*, then there would be a drop in *bad* (meaning) from 12 to 4 out of the 28, and there would be an increase in *good* from 7 to 19 of 28. That represents a factor of roughly 3 in each of the desired directions.

PAPERS

1. Chang Hu: Collaborative Translation by Monolingual Users. *CHI Extended Abstracts 2009*: 3105-3108
2. Chang Hu, Benjamin B. Bederson and Philip Resnik: Translation by Iterative Collaboration between Monolingual Users. *Graphics Interface 2010, Ottawa, Canada*