Assessing User Needs for a Wiki Translation Resource

Alain Désilets
National Research Council of Canada
alain.desilets@nrc-cnrc.gc.ca

Caroline Barrière
National Research Council of Canada
caroline.barriere@nrc-cnrc.gc.ca

Jean Quirion
Université du Québec en Outaouais
jean.quirion@uqo.ca

Professional translators are among the heaviest users of dictionaries, terminological resources, encyclopedias and other reference material. The resources they employ (e.g. Termium[1], Grand Dictionnaire Terminologique[2], IATE [3], the Robert-Collins bilingual dictionary) tend to be proprietary and have very tightly controlled editing policies. Typically, their content can only be edited by highly qualified terminologists working for the organization that produced them.

The advent of collaborative wiki resources like Wikipedia[4] and Wiktionary[5], challenge this assumption of tight edit control [3]. Indeed, using an open editing policy, Wikipedia was able to achieve, in the short span of three years, a level of coverage and accuracy comparable to that of Encyclopedia Brittanica [4]. Given this, it is natural to ask whether translators could benefit from a more open and non proprietary resource. One whose content would be created collaboratively by a wider community which might include translators or even the public at large. In the rest of this paper, we will refer to this type of resource as a *wiki translation resource* (or simply *wiki resource* when the context makes it clear that we are talking about a resource for translators).

This position paper presents a *short summary* of research we have done in this area. For more details , see [2].

Our research questions were:

- *Question 1:* What might be the advantages of a wiki resource compared to more conventional closed resources currently used by translators?

- *Question 2:* To what extent do existing wiki resources like Wikipedia, Wiktionary and OmegaWiki already constitute a satisfactory translation resource?

- *Question 3:* If existing wiki resources are not useful to translators in their current state, how might they be improved and transformed into a satisfactory translation resource?

We investigated these questions using field data collected through Contextual Inquiry[6] and interviews of translators doing their normal day-to-day work in a variety of environments (home-based freelancers, small translation agencies, large government translation departments, and distributed communities of volunteer translators).

Regarding the first question, we found that a wiki resource might help translators by providing them with a single resource with broad coverage of all types of translation problems (terminology, phraseology, general language), in all content domains (medical, legal, technical, etc.). This would improve over the current situation where translators spend a lot of time carrying out the exact same searches in multiple resources, because none of them offers comprehensive coverage, and sometimes still do not find what they are looking for. While building such a large unified resource is clearly outside of the means and expertise of any one publisher or organization, it may be that a worldwide translation community could create it through a massively collaborative process like wiki. This is not far fetched, considering that Wikipedia achieved in three years, a size, coverage and quality comparable to that of Encyclopedia Britannica [4].

Regarding the second question, we found that while translators stand to benefit from a wiki resource, none of the existing ones fill that need in its current state. This is evidenced by the fact that in the course of our contextual interviews, none of our subjects explicitly consulted a wiki resource to resolve a translation problem, even though many of them knew of their existence. A first issue with wiki resources as it pertains to translation, is that they do not have sufficient coverage of typical problems encountered by translators. Indeed, when we evaluated three wiki resources (Wikipedia, Wiktionary, OmegaWiki) for the set of translation problems experienced by subjects during our interviews, we found that the best of the three

---

[1] TERMIUM: http://www.termium.gc.ca/

[2] Grand Dictonnaire Terminologique (GDT): http://www.granddictionnaire.com/

[3] Inter-Active Terminology for Europe (IATE): https://iate.cdt.eu.int/iatenew/login.jsp

[4] Wikipedia: http://www.wikipedia.org/

[5] Wiktionary: http://www.wiktionary.org/

[6] Contextual Inquiry: http://en.wikipedia.org/w/index.php?title=Contextual_inquiry&oldid=190730351

(Wikipedia) had only 31%. coverage In contrast, TERMIUM, a conventional closed terminology database, offered a much larger coverage of 76%. A second issue with existing wiki resources is that that their user interface does not make it easy to carry out translation related tasks such as: finding an appropriate translation for a problem, adding a new translation for a problem, and assessing the trustworthiness of a particular translation for a problem. For example, we found that on Wikipedia and Wiktionary, adding the simplest possible form of a bilingual entry (i.e. a term in one language, connected to its equivalent in an other language, without any additional detail such as definition, domain, etc) required no less than 14 user actions. This is way too much, considering that our interviews indicate that translators work under tight time pressure, and that they do not feel it is worth spending too much time creating terminological entries. A third issue with existing wiki resources is that they lack credibility with translators, who seem very concerned about reliability of their sources. Note however that this may not constitute a showstopper, because translators also seem to have no qualms about looking in less-than-reputable sources when they do not find what they need in trustworthy sources. Nevertheless, our interviews indicate that it might be important for wiki resources to at least address translator's perception of poor reliability.

Regarding the third question, we found that while it would be easy enough to design a user interface that is better suited to the needs of translators, it would be hard to retrofit it onto existing resources like Wikipedia and Wiktionary, without making them harder to use by the bulk of their non-translator constituency. Instead, we propose the creation of a separate, specialized resource for translators, possibly based on the OmegaWiki infrastructure. We show how such a resource could be bootstrapped by automatically mining cross-lingual data from Wikipedia, Wiktionary and OmegaWiki, and that this would result in a respectable initial coverage of 50% for typical translation difficulties. We also show how research on automatic quality assessment of Wikipedia entries [1] could help address translator's perception of poor reliability of open wiki resources.

References

[1] De Alfaro and L., Adler, T, 2007. A Content-Driven Reputation System for Wikipedia. Proceedings of WikiMania 2007, Taipei, Taiwan, August 3-5, 2007.

[2] Désilets, A., Barrière, C., Quirion, J. Making WikiMedia Resources more Useful to Translators. Proc. Wikimania 2007, Taipei, Taiwan, Aug 3-5, 2007.

[3] Désilets, A. Translation Wikified: How will Massive Online Collaboration Impact the World of Translation? Proc. Translating and the Computer (29). November 29-30, 2007. London, United Kingdom. Publication Date: February 2008.

[4] Giles, Jim, 2005. Internet encyclopaedias go head to head, Nature 438, pp. 900-901.