

### NRC Publications Archive Archives des publications du CNRC

### Making WikiMedia resources more useful for translators

Désilets, Alain; Barrière, Caroline; Quirion, J.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

### Publisher's version / Version de l'éditeur:

Proceedings of Wikimania 2007, The International Wikimedia Conference, 2007

NRC Publications Archive Record / Notice des Archives des publications du CNRC : https://nrc-publications.canada.ca/eng/view/object/?id=18e82f47-be13-420f-9110-527e4b64b9bf https://publications-cnrc.canada.ca/fra/voir/objet/?id=18e82f47-be13-420f-9110-527e4b64b9bf

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at <a href="https://nrc-publications.canada.ca/eng/copyright">https://nrc-publications.canada.ca/eng/copyright</a> READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site <u>https://publications-cnrc.canada.ca/fra/droits</u> LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.









National Research Council Canada Conseil national de recherches Canada

Institute for Information Technology

Institut de technologie de l'information



### Making Wikimedia Resources More Useful for Translators \*

Désilets, A., Barrière, C., Quirion, J. August 2007

\* published in the Proceedings of the Wikimania 2007, The International Wikimedia Conference. Taipei, Taiwan. August 3-5, 2007. NRC 50383.

Copyright 2007 by National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.



Copyright, National Research of Canada and Jean Quirion, Université du Québec en Outaouais, 2007.

This work is distributed under a dual license, namely under the GNU Free Documentation License version 1.2 (terms of the license described at http://www.gnu.org/licenses/fdl-1.2.txt) and the Creative Commons Attribution License (terms of the license described at http://creativecommons.org/licenses/by/3.0/).

### Making WikiMedia resources more useful for translators

Alain Désilets Caroline Barrière Jean Quirion National Research Council of Canada National Research Council of Canada Université du Québec en Outaouais alain.desilets@nrc-cnrc.gc.ca caroline.barriere@nrc- jean.quirion@uqo.ca cnrc.gc.ca

#### Abstract

In this paper, we explore the concept of a wiki translation resource, that is, a free, open, massively collaborative wiki based resource that translators could use to find translations of problematic words, terms or expressions. Using field data collected through contextual observation and interviews of translators in their workplace, we specifically investigate three research questions. First, what might be the advantages of a wiki resource, compared to more conventional closed resources currently used by translators? Second, to what extent do existing wiki resources like Wikipedia, Wiktionary and OmegaWiki already constitute a satisfactory translation resource? Third, 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? Regarding the first question, we show how a wiki resource might improve on the myriad of online resources currently used by translators (bilingual dictionaries, generic or domain specific terminology databases) by providing a single free tool with a wider coverage of all types of translation problems and topic domains. Also, it could achieve economies of scale for freelance translators, by allowing them to share expertise and data within a worldwide community of practice. Regarding the second question, we show that in their current state, existing wiki resources are not very useful to translators because they lack sufficient coverage of typical translation problems. Also, their user interfaces do not make it easy to carry out key 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. Regarding the third question, we describe what research and development would have to be done to turn each of those existing wiki resources into a satisfactory translation resource. Based on this, we conclude that OmegaWiki is the most promising platform, and that it can indeed be evolved into a resource that translators could use in their daily work.

### **1. Introduction**

Professional translators are among the heaviest users of dictionaries, terminological resources, encyclopedias and other reference material (De Saint-Robert, 1991; Bowker and Pearson, 2002). The resources they use (e.g. Termium<sup>1</sup>, Grand Dictionnaire Terminologique<sup>2</sup>, IATE <sup>3</sup>, 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<sup>4</sup> and Wiktionary<sup>5</sup>, challenge this assumption of tight edit control. 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 (Giles, 2005). Given that, 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.

Our work looks at this very possibility, that is, creating a free, open, Wikipedia like resource that *anybody* can edit (even non translators), with the purpose of helping translators – and eventually the general public - deal with the translation of problematic words, terms and expressions. 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).

In this paper, we specifically investigate three specific questions.

- *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 investigate these questions using field data collected through contextual observation and interviews of translators at their workplace.

- <sup>2</sup> Grand Dictonnaire Terminologique (GDT): http://www.granddictionnaire.com/
- <sup>3</sup> Inter-Active Terminology for Europe (IATE): https://iate.cdt.eu.int/iatenew/login.jsp
- <sup>4</sup> Wikipedia: http://www.wikipedia.org/
- <sup>5</sup> Wiktionary: http://www.wiktionary.org/

<sup>&</sup>lt;sup>1</sup> TERMIUM: http://www.termium.gc.ca/

To the best of our knowledge, this paper constitutes the first investigation of the usefulness of wiki resources for translators. While there have been studies and books written on the use by translators of more conventional resources such as online dictionaries and terminological databases (L'Homme, 2000; Bowker 2002; Bowker and Pearson, 2002; Varantola, 2006), none of them looked specifically at the use of wiki resources.

The remainder of the paper is organized as follows. Section 2 describes the methodology followed to collect the field data used in our investigation of the three research questions. In turn, Section 3, 4 and 5 then analyze that data to shed light on the three research questions. Finally, Section 6 provides conclusions and directions for future research.

### 2. Research Methodology

Having introduced the concept of a wiki translation resource, we now describe the methodology used to gather field data for our investigation.

Our research is based on user field data collected in the context of a wider on-going project called OPLT (*Observation des Pratiques Languagières Technologisées* which might translate to "*Observing Uses of Language Technology*"). This joint project by the National Research Council of Canada and the Université du Québec en Outaouais aims at better understanding the technological practices and needs of professional translators, by observing and interviewing them in their normal workplace.

For this study, we used a specific interviewing technique called *Contextual Inquiry* (Beyer and Hotzblatt, 1998). Contrarily to other techniques, Contextual Inquiry calls for interviewing the user in his normal work environment, and to focus the interview on an actual task. The user is simply asked to carry out a particular task in the way she usually does, and to think aloud through the process, that is, explain what she is doing every step of the way. The interviewer does not remain passive and will frequently interrupt with probing questions to clarify what the subject said, or ask her to explain behavior that she has not verbalized explicitly.

So far in our OPLT work, we have interviewed five professional translators recruited from three types of work environments: home based freelance, medium size translation company and academia. Subjects, all female, were translating from French to English and were based in Canada.

Each subject was interviewed in the context of carrying out two translation tasks: a *natural* and a *controlled* task (50 minutes each). In the *natural* task, we asked the subject to work on whatever document was in her in-tray at the time. The purpose of this task was to maximize the ecological validity of the data by making sure that we observed the subject working on a document which is representative of what she usually translates. In the *controlled* task, we asked all subjects to translate a same short document (a nontechnical newspaper article). The purpose of this task was to provide a common point of reference across all subjects.

During the interviews, the translator and interviewer's voices were recorded and the translator's screen was captured as video. The audio was later transcribed to text. Copies of the source documents being translated as well as the translations produced by the subjects were also collected. The audio, video and text documents collected during our contextual inquiries provide us with a detailed account of what the subjects did and why. This constitutes valuable *qualitative* information about how translators use dictionaries and terminology databases, and what they need from such resources. The data also provides information that we can use in *quantitative* analysis.

For example, we compiled a list of 59 cases of *translation problems* encountered by our subjects. By translation problem, we mean a word or expression which a subject had difficulty translating, and for which she had to consult various resources. In the remainder of this paper, we will also refer to a an appropriate translation of this problematic word, term or expression, as a *solution* to the translation problem.

For each problem, we collected:

- The English word, term or expression which posed a difficulty.
- The various resources that the subject consulted in trying to resolve that problem.
- The French word, term or expression that the translator eventually used to translate the problematic expression.

Note that our analysis focuses mainly on a subset of 42 out of these original 59 cases. More specifically, we put aside problems that pertained to English expressions which one could not reasonably expect to find in a dictionary or terminological resource. Such expressions are therefore not useful in evaluating the adequacy of a particular dictionary or terminological resource. For example, one subject had difficulty finding a French equivalent for the expression "Go Huskies!", which refers to the rally cry of a college's football team. Obviously, no dictionary or terminology database should be expected to help with that sort of problem.

The 42 relevant cases were identified by keeping any case for which either:

- One of our subjects consulted a dictionary or terminological database when trying to resolve that problem.
- After the fact, researchers on our team were themselves able to resolve the problem using one of the following resources: the bilingual Robert-Collins dictionary, the TERMIUM database or the GDT database. These were the three resources consulted most often by our subjects overall.
- In two cases ("head salesman" and "education gap"), the terminology expert on our research team (Quirion) deemed that the problem should be covered by a wiki translation resource, even though no subject had searched in a dictionary or terminological resource, and the term could not be found in the three resources listed above.

It should be noted that our data may not be representative of the translator population at large since our sample so far only includes professional Canadian translators working from English to French. Nevertheless, the transcripts of our interviews show much commonality in the practices and tools used by these subjects. Therefore, we are reasonably confident that our data is representative of how this particular sub population translates. But

the picture might be different for translators who are from a different population. For example, translators working in a non Canadian context, or who are translating between different language pairs (or for that matter, are translating between English and French, but in the opposite direction). Also, our data may not be representative of the work practices of "hobbyist" translators who translate content on open sites like Wikipedia and Global Voices<sup>6</sup>.

# 3. What might be the advantages of a wiki translation resource?

Having described our research methodology and the field data we collected, we now analyze that data to validate the need for a wiki translation resource and identify the advantages it might have over existing resources currently used by translators.

# **3.1** What kinds of problems do translators encounter and what resources do they consult to resolve them?

If we are to understand the advantages of a wiki translation resource, we must first understand the needs of translators when they consult a resource to resolve a translation problem. Table 3.1 gives a list of the different types of translation problems that our subjects encountered.

Type of problem	Examples		
Terminology	subsidiary, fuel-oil		
Phraseology	on short notice, for more than a decade		
General language	grave, fiery, step		
Cultural or Country- Specific Realities	Go Huskies!, liberal Indian Affairs critic		
World Knowlege	Sun (name of a computer company), former Rep. Joseph Kennedy		

#### Table 3.1: Types of translation problems

*Terminology* problems are terms (either a single word or a multi word expression) that map to a well defined concept in a particular field. For example, in the context of the oil industry, the term "fuel oil" refers to a very specific kind of oil derivative used for the production of power or heat. For this type of problem, our subjects typically consulted terminology databases like TERMIUM and the GDT, as well as corporate or personal homegrown lexicons and databases. However, for terms that have become part of common day vocabulary (e.g. "email"), they also sometimes consulted general dictionaries like Robert-Collins.

*Phraseology* problems are commonly used general language expressions which have a specific meaning and connotation, and are often difficult to render in a different language (e.g. "on short notice"). In general, phraseology problems are not well covered

by dictionaries and terminological resources. Therefore, our subjects usually ended up finding a solution by relying on their own memory and experience, or searching in parallel bilingual texts from their translation archives (bitexts). It is worth noting that even though terminological database lookup was rarely successful for phraseology problem, our subjects often first looked there in the odd chance that there might be an entry for the problem (and occasionally there would be). Our subjects seemed to view database lookup as a quick low cost strategy that could save them a more involved search in bitexts, or consultation with colleagues. This tells us that if a wiki resource was able to cover phraseology problems, it would be an attractive and unique feature of that resource, compared to more conventional terminology databases.

*General language* problems are words which, while not specialized, are nevertheless difficult to translate exactly in a particular context. For example, the word "fiery" used in the sense of "hot tempered", could translate to many French words, some of them stronger and more negatively connotated than others. In a situation like that, our subjects usually turned to a bilingual or unilingual dictionary, or a thesaurus, to find a French equivalent that has the same intensity and connotation as "fiery" does in the original English text.

Cultural or Country Specific Realities are expressions that refer to a particular concept in a particular culture, which is difficult to translate or has no equivalent in another culture. For example, the colloquial expression "Go Huskies!" which refers to the rally cry one might use for a football team, does not have a straightforward equivalent in French. In a situation like this, our subjects never turned to dictionaries and terminology databases. Instead, they would consult colleagues or search through corpuses. This tells us that expressions like this clearly do not belong in a wiki translation resource, and we did not consider them in our list of 42 relevant cases. Another example of culture or country specific problem is the term "Liberal Indian affairs critic" which refers to a political position that only exists in Canada. Some of our subjects looked in terminology databases for this sort of thing. Therefore, it is reasonable and realistic to expect a wiki translation resource to help with cases like this one (and therefore we considered that particular example, and others like it, in our list of 42 relevant cases).

Finally, *World Knowledge* problems are expressions or words which the translator does not understand because he lacks specific world knowledge An example is the word "Sun" used to refer to the computer company of the same name. A translator who does not know that such a company exists might experience problem understanding what it refers to, and how to translate it. Such problems are typically not covered by dictionaries and terminological databases, and our subjects usually tried to resolve them by doing Internet searches with Google. However, such world knowledge is often covered by resources like Wikipedia, and therefore it is not unreasonable to think that a wiki translation resource could cover this kind of problems.

## **3.2** How often might translators consult a wiki translation resource?

Using our user field data, we can get an idea of how useful a wiki translation resource could be to translators. In other words, how often do translators encounter a problem which one might expect to resolve using such a wiki resource?

<sup>&</sup>lt;sup>6</sup> Global Voices: <u>http://www.globalvoicesonline.org/</u>.

The answer is that our subjects could have used a wiki translation resource for 2% of the words they translated. Assuming a conservatively small ten words per sentence, we found that this was equivalent to one word in every five sentences.

Also, we observed that in all but one case, all the resources consulted by our subjects were online as opposed to paper based.

Therefore, we can say that a specialized wiki resource, if it was properly designed for the needs of translators, might be consulted very frequently.

# **3.3** Why might a translator use a wiki translation resource instead of a conventional one?

It should be clear by now that translators currently need to use a wide range of resources when trying to resolve translation problems. There are two reasons for this. Firstly, different resources specialize in different types of translation problems (see section 3.1). Secondly, many of them (even multi domain terminology databases like GDT and TERMIUM) tend to have better coverage of certain domains than others (e.g. education, administration, science), due to the particular focus of the organization that produces them.

Our interview transcripts indicate that for translators, having to look in so many resources is a real inconvenience. For example, they constantly needed to resort to fairly complex (but largely tacit) heuristics to decide which resource to consult in what circumstances. We observed that these heuristics were not fail safe and that often, the first resource consulted by the subject turned out to be a bad choice. By that, we mean that it did not have an entry for the particular translation problem while their second or third choice did. Some translators also commented on the fact that they did not like having so many resources opened at once, because they needed to constantly navigate between them. Some of them went through great length to arrange those various resources in their task bar and their web browser, so as to facilitate navigation between them. Finally, there were always cases where the translator did not find any relevant solution, in spite of having consulted several resources.

Therefore, one possible advantage of a wiki resource is that it might grow into a single resource with a broader coverage of all problem types and all content domains. This might allow translators to resolve most problems by consulting a single resource, thus avoiding inconvenient navigation between multiple resources. While building such a large unified resource is clearly outside of the means and expertise of any one publisher or organization, it may very well 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 (Giles, 2005).



the GNU Free Documentation License (see Project:Copyrights for details). If you

Illustration 3.1.: Wiki markup for a Wiktionary page

A wiki resource could also achieve economies of scale for freelance translators, by allowing them to share expertise and data within a worldwide community of practice. Our freelance subjects sometimes commented on the importance of belonging to a network of translators that they could consult and share information with. The Wikipedia experience shows that wikis can not only lead to high quality content, but that they are also a good tool for fostering the emergence of active communities.

# 4. Do existing wiki resources already constitute a satisfactory translation resource?

Having established that a wiki translation resource would indeed be useful to translators, we now ask ourselves whether existing wiki resources already constitute such a satisfactory translation resource.

### **4.1 Resources evaluated**

In our study, we look at three existing wiki resources, namely, Wikipedia, Wiktionary and OmegaWiki. As a point for comparison, we also use TERMIUM, a conventional terminology database. We now describe each of those resources in turn.

Wikipedia and Wiktionary, both started by the WikiMedia foundation, are respectively a wiki based encyclopedia and dictionary that anyone in the world can consult and edit. In spite of this open editing policy, these resources have comprehensive content whose accuracy compares favorably to that of more conventional resources (Giles, 2005). Although both are available in multiple languages, versions in different languages are largely grown independently from one another and by different linguistic communities. However, a page in one language often (but not always) has links to pages that describe the same concept in some (but usually not all) other languages. Also, neither of those resources is, properly speaking, a terminology database. Although each page on those sites is about one particular word, term or concept, information about them is captured not in the form database records, but in the form of markup tags that specify the content of the page and how it is to be displayed. For example, Illustration 4.1 displays the wiki markup for the Wiktionary page for English word "subsidiary". This markup is created manually by authors of pages, and therefore, it does not always follow a uniform standard (although the editing tools try to nudge authors towards uniformity through non compulsory templates for pages). For example, in Illustration 4.1 we see that links to translations are encoded using different tags for different languages (e.g. "\* *French:* [[filiale]] (1)" for the French translation and "[zh:subsidiary]" for the Chinese translation). As we shall see later, this tag based data model makes it difficult to tailor these resources to the needs of translators.

OmegaWiki is a wiki resource that aims at providing information on all words and terms of all languages. It was designed as a generic terminological resource, that is, it was not built specifically for the needs of translators. Contrarily to Wikipedia and Wiktionary, OmegaWiki is a proper multilingual terminology database. In other words, information about a word, term, expression and its translations in different languages is encoded explicitly and directly in a relational database (not as tags). This makes the data much easier to manipulate programmatically, and as we will see, this resource can therefore be more easily tailored to the needs of translators. As of this writing, OmegaWiki is still in beta stage, and editing of its content is restricted to a very small number of core contributors. But the development team plans to open it up for editing as soon as stability of underlying database is ensured. Also, the content of the database is still very sparse, because the project is still in its infancy.

TERMIUM is the terminology database of the Translation Bureau of the Government of Canada. It is one of the most comprehensive and most widely used terminology databases for the English-French language pair<sup>7</sup>. While *editing* of its content is restricted to Bureau personnel, it can be *consulted* by external paying customers. Like OmegaWiki, it is a proper terminology database. Of the four resources we looked at, it is the only one that was designed and grown specifically with the needs of translators in mind.

#### 4.2 Do translators use existing wiki resources?

When evaluating the usefulness of existing wikis as translation resources, a first question to ask is whether translators currently use them for that purpose. Somewhat surprisingly, our field data indicates that the answer is a clear no, at least for the translators we observed and interviewed.

Indeed, at no point in the course of our interviews did we observe translators going explicitly to a wiki resource while trying to resolve a translation problem. However, in one case (and one case only) a subject did a Google search for the word "*burough*", then

	Wikipedia	Wiktionary	OmegaWiki (June 07)	OmegaWiki (Aug 2, 07)	TERMIUM
Has english entry	66.7%	50.0%	0%	28.6%	81.0%
Has English entry in right sense	54.8%	45.2%	0%	23.8%	76.2%
Has French equiv.	31.0%	33.3%	0%	14.3%	76.2%
Has French equiv. in correct sense	31.0%	26.2%	0%	11.9%	76.2%

Table 4.1: Trying to resolve translation problems with different resources.

<ol> <li>Subject Field(s)</li> <li>Commercial and Other Bodies (Law)</li> </ol>	Subject Field(s) - Organismes commerciaux ou no
subsidiary corporation Source	corporation filiale Source
Subsidiary Source CORRECT	filiale Source CORRECT, FEM
Subsidiary company Source	société filiale Source CORRECT, FEM
division Source SEE OBS	
<b>DEF</b> - A corporation in which another corporation owns directly, or indirectly through one or more intermediary corporations, a majority of the shares carrying the right to elect at least a majority of the members of the board of directors. Source	DEF – Société juridiquement inc société mère, généralement du indirectement, une participation membres du conseil d'administ
OBS - In current business vocabulary, the term "division" is used as a synonym for "subsidiary". Source	
OBS - Compare with "parent (company)". Source	
KEY TERM(S) – legal subsidiary	
2004-03-16	
<ul> <li>Subject Field(s)</li> <li>General Vocabulary</li> </ul>	Subject Field(s) - Vocabulaire général
Source ADJ	ADJ

**Illustration 4.1: A TERMIUM entry** 

found the Wikipedia page for that word, and went there to consult it.

The fact that our subjects almost never used a highly publicized wiki resource like Wikipedia is somewhat surprising. One can venture two possible explanations. On one hand, it could be that translators do not know about those resources, or have not thought of them as a possible place to look for solutions to translation problems. On the other hand, it could be that existing wiki resources, in their current state, do not fit their needs.

Because our interviews were not aimed specifically at understanding translators' use of wiki resources (they were done as part of the wider OPLT project that looks at their use of technology in general), we did not have an opportunity to ask them specifically why they did not use them. However as we will see, further analysis clearly shows that in their current state, existing wiki resources are indeed not very well suited to the needs of translators. One can therefore conjecture that this may play a major role in their lack of use by that community.

# 4.3 How many of the problems could have been resolved using existing wiki resources?

Having established that translators do not currently use wikis as translation resources, we next ask ourselves what might happen if they did.

We took each of the 42 relevant translation problems and tried to resolve them by going to the four resources mentioned above (three wiki based resources and TERMIUM). In the case of OmegaWiki, because it was going through such rapid growth at the time of this study we decided to perform this analysis twice: once at the beginning of June 2007, and once on August 2, 2007.



Illustration 4.2: An English Wikipedia page

a a w	article discussion modifier h
WIKIPÉDIA L'encyclopédie libre	Filiale
Navigation	Example territe est une ébauche à compléte
<ul> <li>Accueil</li> <li>Portails thématiques</li> <li>Index alphabétique</li> <li>Une page au hasard</li> <li>Contacter Wikipédia</li> </ul>	Une <b>filiale</b> est une entreprise dont le capital : direction, l'administration et le contrôle par l' maison mère sur le plan juridique.
Contribuer	Voir aussi [modifier]
<ul> <li>Aide</li> <li>Communauté</li> <li>Modifications récentes</li> </ul>	- Succursale
<ul> <li>Accueil des nouveaux arrivants</li> <li>Faire un don</li> </ul>	Portail de l'économie
Rechercher	Catégories : Wikipédia:ébauche droit • Dro

#### **Illustration 4.3: Corresponding French page**

For the other three resources, analysis was done at the beginning of June 2007 only. Table 4.1 summarizes the results of our searches through these four resources. Looking at the last row, we can see that existing wiki resources have very little coverage of translation problems, compared to a more conventional terminology database like TERMIUM. Indeed, the percentage of problems that could be resolved with the best wiki resource (Wikipedia), is only 31.0 p. cent, compared with 76.2 p. cent for TERMIUM. Wikitonary follows behind with 26.2 p. cent. As for OmegaWiki, we could not solve even a single of the observed problems using that resources at the end of June 2007. However, as of August 2, this resource had greatly improved, and we found we could solve 11.9% of the observed problems.

Table 4.1 also provides a break down of the reasons why each resource fails at helping with translation problems. The first row displays the percentage of the cases where we could find an entry for the English word or expression. We can see that Wikipedia and Wiktionary starts reasonably high (66.7 p. cent and 50.0 p. cent respectively), but still short of TERMIUM (81.0 p. cent). For OmegaWiki, the coverage is already zero at that point for the June analysis, and is still very low (28.6 p. cent) as of the August analysis.

The second row shows the percentage of cases for which we not only found an entry for the English term, but where the entry also contained a definition matching the sense used in the source text being translated. For example, in one case Wikipedia had an entry for the word "fiery", but that entry did not include a definition for the "hot tempered" sense used in the source text. Here, we see that Wikipedia coverage experience a sharp drops of 11.9 percentage points. In contrast, the decrease in coverage is smaller for Wiktionary (4.8 points), TERMIUM (3.8 points) and OmegaWiki (4.8 points for the August analysis). OmegaWiki's coverage for the June analysis stays of course at zero for the rest of the table.

The third row shows the percentage of cases where we not only found an English entry with the correct sense, but that entry also had a link to a French entry. Here, Wikipedia, Wiktionary and OmegaWiki (August analysis) all incur a sharp drop in coverage while TERMIUM stays the same. The later is understandable since in TERMIUM is a resource designed specifically for translation purposes, where entries are always created for the sole purpose of providing an equivalent in another language.

Finally, the fourth row shows the percentage of cases where at least one of the French entries associated with the English entry was for the sense used in the source text being translated. Here, we see that Wikipedia and TERMIUM stay the same, while Wiktionary experiences a significant drop (7.1 points) and OmegaWiki (August analysis) stays about the same (2.4 points drop).

# 4.4 How easy is it to consult wiki resources for translation purposes?

Given this lack of coverage of the kinds of problems typically encountered by translators, it is not surprising that they seldom use those resources. However, this does not seem to be the only issue with wikis for translation use. Indeed, when searching in the various resources to produce Table 4.1, we found that the wiki resources were much harder and less efficient to use than TERMIUM, even for cases where relevant material was found in them. While we have not formally measured this usability gap, we can illustrate some of the issues through a Cognitive Walkthrough (Wharton, C. et al., 1994). This is a well known usability inspection method whereby a group of developers or usability experts go through the steps that an end user would have to follow to complete a specific task. Any usability issue that arises in the process is written down and reported as areas needing improvement.

In this section, we present a walkthrough for TERMIUM and the three wiki resources, for a simple task where the end user needs to find a French translation for the English word "subsidiary". While





### Illustration 4.6: An OmegaWiki English page with French translations

appropriate one. This may at worse involve scrolling up and down a few times. This high efficiency can be attributed to various good design choices in TERMIUM. The screen is compact and only contains information that is relevant to a translator. In particular, definitions of terms and their equivalents are short, but complete enough to allow the user to make sure that the sense being presented is the same as the sense used in the source text being translated. Also, the entry only displays information about the language pair (English-French) that the user is working in. Both English and French are presented on a single screen and in parallel. All senses of the English term are listed on a same screen, along with their respective translations. Finally, although we cannot see this from looking at a single entry, all TERMIUM entries present the information in the exact same format and layout, which means the user can quickly get used to scanning them rapidly to find specific information.

In contrast, finding an appropriate French translation for "subsidiary" on Wikipedia is much more complicated and slow. Illustration 4.3 shows what the user sees after searching for that word on the English Wikipedia. We can see that the presentation of the information is not as well suited to the needs of translators. While the page contains a lot of interesting information about a subsidiary, most of it is irrelevant from a translation perspective. In fact, in order to see an actual translation for "subsidiary", the user must locate the "Français" link that appears on the left, and click on it. In the case shown in Illustration 4.3, this is not such a big problem because translations for that particular term are only shown for eight languages. But suppose the density of inter language links on Wikipedia was increased so that a page like this one contains links to translations in say, 50 languages or more, we can see that the user would incur a significant visual scanning and scrolling overhead when trying to locate the Francais link.

Once the user has located and clicked on the *Français* link, he is presented with the screen displayed in Illustration 4.4. While this page does display a French translation ("filiale"), the user has now lost the definition of the English term, and may have to toggle between the two pages in his decision process. Moreover, the

Illustration 4.4: An English Wiktionary page

19 19 E	article discussion modifier historique
子维ビ	subsidiary
Wiktionnaire Le dictomatre libre navigation • Page d'accueil • Portails • Une page au hasard	Sommaire (masquer) 1 Anglais 1.1 Étymologie 1.2 Adjectif 1.3 Nom commun
contribuer	
Communauté     Pages de discussion     Modifications récentes	Anglais
aide	🍣 Étymologie
<ul> <li>Aide</li> <li>Modèles</li> <li>Conventions</li> <li>Faire un don</li> </ul>	→ Étymologie è compléter. (Ajouter)
rechercher	
Consulter Rechercher	subsidiary 1. Supplémentaire. 2. Inférieur. 3. Auxiliaire
Pages liées	J. Auxiliane.
<ul> <li>Suivi des liens</li> </ul>	🔯 Nom commun
<ul> <li>Importer une image ou un</li> </ul>	subsidiany
<ul> <li>Pages spéciales</li> <li>Version imprimable</li> </ul>	1. Filiale, succursale.

### Illustration 4.5: Following the "Français" link from an English Wiktionary page

this is based on a single example, it is nevertheless representative of the many issues we encountered when searching for the 42 relevant translation problems.

Illustration 4.1 displays what the end user sees after searching for the English word "subsidiary" in TERMIUM. This single screen provides all the information that the user needs to quickly resolve the problem. A quick scan through the various translations of the various senses of the English word allows him to choose the most



### Illustration 4.7: Adding link to French translation on the English Wikipedia page.



### Illustration 4.8: Adding reciprocal link to the English translation on the French Wikipedia page

page only lists one possible translation of "subsidiary". While it is possible in Wikipedia to associate an English page with more than one French page (this would be rendered as multiple *Français* links in Illustration 4.3) the user can never see more than one such translation at a time. The user could also get alternate translations by following links in the *Voir aussi* section (i.e. "*See also*"), but again, he can only see one translation at a time.

When we add these things together, we get a user experience which is far from optimal for translators.

We now turn our attention to Wiktionary. In Illustration 4.5, we see that presentation of the information for the English term is somewhat better suited to the needs of translators. The information on the page is much more compact than its Wikipedia counterpart and most of it is relevant to translators. But it still falls short of TERMIUM in terms of the needs of translators. Just like the case for Wikipedia, the user needs to locate the *Français* 



### Illustration 4.9: Adding a new French translation in an OmegaWiki entry

link amongst a potentially very long list of languages, all but one of them completely irrelevant to his current task of translating to French. Also, presentation of the information is not consistent. In some cases, the translation information is accessed through a *Français* link on the left and sometimes through a link in the *Translations* section. Sometimes, these two presentation styles are used on a same page (as in the case of Illustration 4.5), possibly forcing the user to scan not one, but two lists of languages.

In cases where translations are listed in the Translations section, Wiktionary does show the English word and its multiple translations in the target language on a single page. But again, imagine an entry for a word that has four different senses all of which must are translated in 50 different languages. With such an entry, it would be difficult for the user to quickly scan it to locate the translation in the desired language. On the other hand, in cases where the translation is accessed through a Français link, the user cannot immediately see the English and French words on a same page. However, when the use clicks on that Français link (Illustration 4.6), he is presented with a screen that is well suited to his needs as a translator. This screen shows English and French words on a same page, displays multiple choices of equivalents, and only shows information relevant for the English-French pair. The one shortfall is that it does not directly display the definitions of the English term and its various translations, which means the user may have to go back and forth between English and French pages in his decision making.

Turning now to OmegaWiki (Illustration 4.7), we see that its presentation of the information is always well suited to the needs of translators. Like for TERMIUM, the user can quickly make a decision by looking at that single screen. The entry is compact and only shows information that is relevant to translators (in

particular, it only shows information for the language pair that the translator is interested in). Moreover, the entry is complete and shows most of the relevant information on a single page: English term, list of possible translations, each with a short definition. From a presentation point of view, the only piece missing is a short definition of the English term (the user has to click on it to see its definition).

Summarizing this section, we have shown that among the three wiki resources, OmegaWiki is the only one whose user interface consistently allows fast and easy consultation by translators. It might seem that these kinds of user interface issues are just minor irritants and that they do not really affect usefulness of a resource. But remember that, according to our data, a translator might consult a wiki translation resource once every five sentence. Also, our interviews with translators clearly indicate that they are very conscious of the passing of time because they typically work to tight deadlines, and are paid by the word. They often spoke about running out of time and needing to take shortcuts to get the translation done on time. Given those observations, translators need tools that can help them make decisions rapidly and efficiently. Each access to a wiki translation resource should thus provide the necessary information at a glance and not through lots of navigation in different pages. Failing that, translators are not likely to use the resource at all.

# 4.5 How easy is it to edit wiki resources for translation purposes?

In the previous section, we looked at how easy it is for users to consult the various resources when looking for solutions to translation problems. But what about editing their content? The short answer as we shall see, is that none of the three wiki resource evaluated allow the user to carry out that task easily and efficiently.

Editing is an important issue for a wiki translation resource, since a central concept behind wiki is collaborative editing. A wiki translation resource can only be successful if it can effectively support a community of users who can grow the resource and keep it up to date.

Continuing on our previous example of "subsidiary", we present a walkthrough for a simple scenario where the user searches for a French equivalent for that term, and only finds an English entry with no French equivalent. The user then decides to create a French translation "filiale", and associate it with the English word "subsidiary" in the wiki resource.

To complete this simple task with Wikipedia, the user would have to do the following:

• Edit "subsidiary page": Click on the *Edit* link on the "subsidiary" page. This displays the content of the "subsidiary" page in wiki markup (Illustration 4.8).

= 1 click

• Add link to "filiale" page: Scroll down the edit page to locate where the list of alphabetically sorted language links is specified. Scan that list of links to locate the exact place where the *Français* link should go (in alphabetical order). Click after the entry that should precede the *Français* link, and type the following markup tag: "[[fr:filiale]]". Click Save. This displays

the "subsidiary" page with the newly added *Français* link on the left (as per previous Illustration 4.3).

= 1 scroll, 1 alphabetical scan, 2 clicks, 1 keyboard operation.

- Create French "filiale" page and open it for editing: Select the title of the English page, and do Ctrl-c to copy it to the clipboard (Note: actually this step would only be needed for long titles where one might make a mistake or misremember it when creating a back link to the English page in the next step). Scroll down to the list of language links on the left, locate the *Français* link and click on it.
  - = 1 selection, 1 keyboard operation, 1 scroll, 1 click
- Add back link to "subsidiary" page: You are now ready to enter content into the "subsidiary" page (Illustration 4.9). Create a back link to the English "subsidiary" page, by typing the following markup: "*[[en:"*, then do Ctrl-v to paste the title of the English page copied in the previous step, then type markup "*]]*". Click on "Save".
  - = 1 keyboard operation, 1 click
- Go back to English Wikipedia: You are now seeing seeing a page like the on shown in previous Illustration 4.4. Scroll down to the list of language links, click on the *English* link that now appears on the left. This is done to go back to the English Wikipedia where the user can do his next search for an English term to translate.

= 1 scroll, 1 click

**Total:** 6 clicks, 3 scrolls, 3 keyboard operations, 1 selection, 1 alphabetical scans = 14 user actions.

Of course, none of this is difficult to do, but it sufficiently interrupts the flow of the translator that he may not bother adding the entry altogether. However, as pointed out in Désilets et al. 2006, most of those tedious steps could be easily automated so that the translator can create a French equivalent with just a few actions.

The procedure for adding a French equivalent in Wiktionary is essentially the same as in Wikipedia, therefore we do not present it.

On OmegaWiki, the same scenario would play out as follows:

- Open "subsidy" page for editing: Click on the Modifier (Modify) link on the page for "subsidiary".
   = 1 click
- Specify French as the target language: You now see the page on Illustration 4.10. Scroll down to the *Synonymes et traductions* (i.e. *Synonyms and translations*) section of the page and click on the languages pick list (at the right of the plus sign icon). A list of languages whose name start with "a" are listed. To see the French language, you must type "fr" in the text box (left of "Clear"). You now see *français* in the list of the target language and can click on it.

= 1 scroll, 2 clicks, 1 keyboard operation



Illustration 4.10: WeBiText user interface



Illustration 4.11: TerminoWeb user interface

- Add "fililale" equivalent: In the text box beside the languages pick list, type the French equivalent "filiale".
  - = 1 click, 1 keyboard operation
- Save the page: Scroll down to the bottom of the page and click on the *Save* button.
  - = 1 scroll, 1 click

```
Total: 5 clicks, 2 scrolls, 2 keyboard = 9 user actions
```

OmegaWiki therefore requires significantly less user actions (5 to be exact) than Wikipedia or Wiktionary, but it is still much more than seems necessary.

Note that we have not presented walkthroughs for editing content in TERMIUM, because we did not have access to those features for our study. Therefore, we are not in position to compare the usability of editing in wiki resources to that of TERMIUM. However we do not think a comparison is necessary to see that editing translation content in the wiki resources is too inefficient and that those features are not likely to be used by translators who are hard pressed for time and are not paid specifically to create terminology entries in a wiki resource.

# 4.6 How easy is it to evaluate the trustworthiness of a translation?

Given that anyone in the world (including people who are neither terminologists nor translators) can modify a wiki resource, it is worth asking whether translators would be inclined to trust it, and if not, how the resource could be modified to promote trust.

Our interviews clearly indicate that translators do not blindly trust any dictionary or terminological database, even high quality ones like TERMIUM. Our subjects often talked spontaneously about their reasons for trusting such and such resource only for such and such domain or only in such and such circumstances. Even when looking at a trusted resource, they would not trust all entries equally. For example, if a particular translation was marked as having been "normalized" (i.e. the organization producing the resource took an official position that this particular translation was the one to be favored), some translators would have more confidence in it. When consulting an in house terminology database, some translators would also pay attention to the name of the people who created or edited the entry. For example, a translation that was entered by a junior translator might be given less confidence than a translation that was entered by a senior translator or vetted by a terminologist.

While being able to assess the trustworthiness of a particular translation seemed important to our subjects, it is not clear to what extent they needed special visual and textual cues in the resource to help them do this. Often, it seemed that our subjects relied mostly on their own knowledge and experience to decide if a particular translation was appropriate. For example, if they saw a French term that was a word for word translation from the English, or if the proposed translation was one that they had never seen or heard before, our subjects tended to doubt its correctness. In situations like these, our subjects often did a Google search to verify that the French translation was actually being used on trusted sites (e.g. sites from the Government of Canada). Translators seem to have developed a kind of 6<sup>th</sup> sense which allows them to very rapidly assess the appropriateness of a suggested equivalent in a particular context, and they use this extensively to perform a kind of "downstream" quality control.

### The following definitions and translations were found on WikiPedia and Wiktionary.

Subsidiary

- A subidiary, in <u>business</u>, is an entity that is controlled by another entity (c.f. <u>WikiPedia</u>)
- auxiliary or supplemental (c.f. Wiktionary)
- secondary or subordinate (c.f. Wiktionary)
- of, or relating to a <u>subsidy</u> (c.f. <u>Wiktionary</u>)
- A company owned by the parent company or holding company (c.f. Wiktionary)
- (music) a subordinate theme (c.f. Wiktionary)

#### French Translations

Filiale (c.f. <u>WikiPedia</u>)

Une filiale est une <u>entreprise</u> dont le capital a été formé pour plus de sa moitié du capital par des apports d'une autre société dite <u>société mère</u>, qui en assure généralement la direction, l'administration et le contrôle par l'intermédiaire d'une ou plusieurs personnes, administrateurs ou gérants qu'elle a désignés. <u>Examples of use</u>

#### Succursale (c.f. WikiPedia)

Une succursale est l'établissement commercial autonome et durable d'une société Examples of use

Supplémentaire (c.f. <u>Wiktionary</u>) Examples of use

Inférieur (c.f. <u>Wiktionary</u>) Examples of use

Auxilliaire (c.f. <u>Wiktionary</u>) Examples of use

Examples of

#### Illustration 4.12: Mockup of a specialized translator view of Wikipedia and Wiktionary data for a term

As it turns out, none of the existing wiki resources provide any explicit cues to help translators evaluate the trustworthiness of solutions to a translation problem. While our user field data indicates that such cues would be useful to translators, we cannot conclude that they are absolutely necessary, even in the context of an open wiki resource. It could be that translators would be able to judge based on their own knowledge and experience, whether a particular translation taken from the wiki resource is appropriate or not for their particular needs. It could also be that the absence of this kind of cue is more of an issue for junior translators than for experienced ones. More focused interviewing needs to be done with translators in order to clarify these specific issues.

# 5. Towards a wiki translation resource: How to improve existing wiki resources?

Summarizing our discussion so far, we have shown that while translators could benefit from a wiki translation resource, none of the existing wikis (Wikipedia, Wiktionary, OmegaWiki) fulfill that role well.

While somewhat disappointing, these findings are not surprising given that none of those resources were designed and grown specifically for the needs of translators. Even OmegaWiki was started as a general terminological resource, and is not targeted specifically at translators. In a sense, it is unfair to evaluate these resources for a task that they weren't designed for in the first place. But the fact remains that, unless they can deal with the issues we raise, translators are not likely to use them. We see this as a lost opportunity for wiki resources, one that deserves to be dealt with.

Fortunately, we believe that it is possible to transform existing wiki resources into something that is truly suited to the needs of translators. In this section, we look at the research and development work needed to turn each of the three wikis into a satisfactory translation resource. Based on this, we then try to determine which of the three resources seems like the best platform to build a wiki translation resource.

We look in turn at each of the issues uncovered in the previous section, namely: lack of coverage of translation problems, and lack of efficient support in the user interface for consulting, editing and assessing the trustworthiness of solutions to such problems.

# 5.1 Improving coverage of translation problems

As pointed out earlier, existing wiki resources currently do not have sufficient coverage of translation problems to be truly useful to translators. In this section, we look at how that issue could be resolved for three existing wiki resources.

One way to improve coverage would be to actively involve translators and terminologists in the expansion of translation related content. But for this to happen, we need to provide those communities with tools they can use to easily populate the resource with data.

A first step in that direction would be to provide import tools that translators can use to easily pour their own lexicons and terminology databases into a wiki resource. Our interviews indicate that many freelance translators keep a record of translation problems they encounter frequently, as well as the solutions they chose in different contexts. Such homegrown resources can grow to become quite large, and could benefit other translators. Larger translation companies also keep similar records, but it is not clear that they would be willing to share them with a larger community (since those in house terminology database provide them with a competitive edge). In any case, those records of previously encountered problems are sometimes stored in terminology databases that can export their data into standard formats like Term Base Exchange (TBX). If a wiki resource was able to read such a format, we would have an easy way to pool terminology and lexical resources from thousands of users. The OmegaWiki team is currently developing such tools.

Note however that our interviews also indicate some translators do not use terminology databases for their records. Some freelance translators simply use spreadsheets or even free form word processor documents for that purpose. Such tools obviously cannot export their data to a standard format like TBX. However, it might be possible to implement programs which could read such documents and, with possible assistance from the user, figure out their structure so that information can be extracted from it.

The simple tools described above would be useful to allow translators to contribute existing records to a wiki resource. But what about translation problems for which neither the wiki resource nor the user have a solution? Below, we describe two tools under development at the National Research Council of Canada: WeBiText and TerminoWeb. Both of those assist a user in doing a terminological searches on the web, and could be extended so that the results of those searches can be easily contributed to a wiki resource.

WeBiText<sup>8</sup> is essentially a bitext system which uses the whole web as a corpus. Given an English term like "subsidiary", it searches the web for parallel English-French sentences that employ that term. A typical search results screen is shown in Illustration 5.1. One can see how this system could act synergistically with a wiki translation resource. For example, if a user searched the wiki for "subsidiary" and did not find an appropriate translation, there could be a button that allows him to launch a WeBiText search for that term. Conversely, if the user then finds an appropriate translation for "subsidiary" in WeBiText, there could be a single button that allows him to pour that hit back into the wiki.

In a way, it may seem that a wiki translation resource is redundant with a WeBiText. But our Contextual Inquiry study clearly shows that translators make heavy use of both terminological databases and corpuses (including bitexts), and that they tend to look in terminological databases first. The reasons for this are not clear yet, but it seems to be due to the fact that in terminological databases, the information is preanalyzed and immediately accessible. Indeed, during our interviews we noticed that subjects tended to take much longer to evaluate and analyze search results from Google or a bitext system than for dictionary or terminological database entries. It seemed they needed to read many of the hits before they could form a picture of the various translations found, as well as their respective contexts of use and quality.

TerminoWeb (Barrière and Agbago 2006) is another system being developed at the NRC, which might offer a different and complementary approach to populating a wiki translation resource

<sup>&</sup>lt;sup>8</sup> www.webitext.org

with content. It was designed to assists terminologists in doing thematic searches. The system allows the user to retrieve web content on a particular topic domain, and to automatically extract important terms that relate to that topic. Illustration 5.2 shows a sample screen, where the system has suggested terms for the "paragliding" domain. The user can then select suggested terms for inclusion in the thematic search results. The system has many other useful features which are described in detail in Barrière and Agbago 2006.

While TerminoWeb was developed for unilingual thematic searches, beta users have requested features to help them find translations of the various terms. One way to provide this would be to combine TerminoWeb's features with those of WeBiText. In addition, one can imagine features which allow users to, at the push of a single button, contribute the results of a multilingual thematic search to a wiki resource. Such a tool might allow terminologists and translators to efficiently participate in the effort to extend coverage of that wiki resource.

An altogether different approach to increasing the coverage of wiki resources would be to implement tools that can automatically extract translation related content from one wiki resource, and pour it into another. This would be particularly useful for OmegaWiki, since its coverage of translation problems is currently much lower than that of Wikipedia and Wiktionary (see Table 4.1).

Such an automatic extraction could be implemented easily using simple heuristics. For example, given the Wikipedia entry shown in Illustration 4.3 for term "subsidiary", the system could extract relevant French translations as follows:

- Extract a short definition of "subsidiary" by taking the first sentence of that page.
- Follow the *Français* link on the "subsidiary" page to get the page on Illustration 4.4. Parse the title of that page to get the French equivalent "filiale", and grab the first sentence as its short definition.
- Acquire additional French translations by following links in the *Voir aussi* (i.e. *See also*) section of the "filiale" page on Illustration 4.4. In this case, this would result in the addition of translation "*succursale*", for which the system could again, grab the first sentence as short definition.
- Finally, the system would pour all of this information into OmegaWiki.

Similar heuristics could easily be implemented for Wiktionary as well.

If implemented, such a system would achieve a 50.0 p. cent coverage of our observed translation, which is significantly better than the coverage of any individual wiki resources (which lie at 31.0 p. cent, 26.2 p. cent and 11.9 p. cent respectively).

#### **5.2 Improving the consultation user interface**

As pointed out earlier, the user interface of two of the existing wiki resources (Wikipedia and Wiktionary) does not support effective consultation to resolve translation problems. In this section, we look at what could be done to address that issue.

Unfortunately, changing the user interface of Wikipedia and Wiktionary to support more efficient and streamlined consultation

by translators, may in turn make it harder to consult for the general user population. For example, "fixing" the user interface of Wikipedia for translators would call for the removal of most of the text displayed for "subsidiary" on Illustration 4.3, or at least, moving that text to another page which the user could not see without doing an extra click on a link (e.g. a *Details* link). This is clearly suboptimal for the majority of Wikipedia users who are not translators.

A better approach would be to implement a specialized interface to Wikipedia and Wiktionary which would automatically extract the information that is relevant to the translator (using the approach described above in Section 5.1), and then show it in a way that is most appropriate for them (e.g. something along the lines of Illustration 5.3).

While implementing these heuristics is not difficult, one possible problem is that extracting all this information on the fly may be significantly increase the load on an extremely busy server like Wikipedia. This in turn may make the system slow to respond to user actions.

#### **5.3 Improving the editing user interface**

As pointed out earlier, the user interface of all three existing wiki resources does not support easy and efficient editing of translation related content. In this section, we look at what would be needed to address that issue in each of those resources.

Improving the editing user interface of OmegaWiki is a simple matter of streamlining the editing process in its existing interface.

In the case of Wikipedia and Wiktionary however, we are faced with a non trivial technical challenge. As pointed out in the previous section, we cannot change the consultation user interface of those resources without negatively affecting the general user population, and we must therefore superimpose a specialized interface for translators on top of the existing content.

While superimposing a read-only specialized interface to the original data seems relatively straightforward, we run into difficulties when we want to allow translators to use that same interface to edit the original content. Because this specialized interface gathers data from several pages and presents them in a unified single page view, it would need to keep track of where each piece of data came from. The reason for this is that if the user changes a particular piece of data on the unified view, the system needs to reproduce that change in the original source of the data. For example, in the case of "subsidiary", if the user changes the short definition of "succursale", the system needs to know that this information actually came from the first sentence of the first Voir aussi (i.e. See also) link on the French page pointed to by the Français link on the English page for "subsidiary". Only if it knows that, can the system then go to the "filiale" page and change its first sentence.

This problem illustrates the advantages of the OmegaWiki relational data model which captures relationships between words and translations directly in the database as opposed to indirectly through markup tags typed manually (and often non uniformly) in a collection of interrelated pages. With OmegaWiki, it is trivial to reliably extract translation information from the database, present it to the user for modification, then put the modified information back into the database. While far from trivial, such tracking of disparate translation information on Wikipedia and Wiktionary could still be implemented. However, there is also a danger that translators might make a change which, while appropriate for the needs of translators, is not appropriate for the needs of the general user population. For example, a translator might change the short definition of "filiale", not realizing that he is in fact changing the first sentence of the Wikipedia page for that term. And while the new short definition might be appropriate when displayed in the context of the specialized translator interface, it might not be appropriate as the first sentence for the Wikipedia page of that term.

# 5.4 Allowing translators to assess trustworthiness of entries

As pointed out earlier, none of the existing wiki resources provide easy ways for translators to assess the trustworthiness of a particular solution to a translation problem. Given the fact that those resources can be modified by anyone, this might be a serious limitation. In this section, we look at what would need to be done to introduce such features in those wiki resources.

An easy to implement approach would be to add fields to the data, where contributors could provide examples of use of a term and its translation, with references to the actual source where the example was taken from. This would allow contributors to substantiate their entries with examples taken from well reputed sources (e.g. the Government of Canada, the European Parliament). While adding such fields in OmegaWiki is trivial (because OmegaWiki is in fact based on a relational data model), in the case of Wikipedia and Wiktionary, we run again into the problem that this data would have to be encoded in the form of tags distributed across several pages. This in turn makes it harder for the system to allow the user to modify this data through a streamlined user interface specialized for translators.

Another approach would be to, again, combine a wiki translation resource with the WeBiText system. For each pair of original term-translation pair, there would be a link *Examples of use* (as shown in Illustration 5.3) which when clicked, would spawn a WeBiText search. This search could even be restricted to web sites which are deemed reputable by this wiki translation community. Examples that the user deems particularly useful and relevant could then be poured back into the wiki at the click of a single button on the WeBiText search results page.

A third approach would be to provide an automatically generated quality score for an entry, based on a reputation system like the one described by (De Alfaro and Adler, 2007). This system calculates the quality of a Wikipedia entry by looking at the "reputation" of the people who have created and edited it. The reputation of a contributor is in turn computed by looking at the history of his contributions, and evaluating whether these contributions tend to "stick" or on the contrary, be quickly overridden by other contributors. These automatic reputation scores have been shown to strongly correlate with the Wikipedia community's actual evaluation of quality and reputation.

Another approach would be to allow users to explicitly rate a particular solution. Similar voting schemes have been used successfully on sites like E-bay, to evaluate the trustworthiness of buyers and sellers.

# 5.5 What existing wiki resource is the best platform for a wiki translation resource?

Looking at the previous discussion, we can see that of the three existing resources, OmegaWiki is the one that could most easily be transformed into a satisfactory wiki translation resource. Its user interface is already well suited to consultation by translators. Also, because it uses a relational data model to capture information about words and their relationships (as opposed to the wiki markup model used in Wikipedia and Wiktionary), we can more easily improve its user interface for editing and assessing the trustworthiness of translation related information. This relational data model also makes it easier to integrate tools for expanding its content (e.g. WeBiText, TerminoWeb or a Wikipedia/Wiktionary robot for extracting translation relevant content).

The only disadvantage of OmegaWiki compared to Wikipedia and Wiktionary, is that it is not as well known. But we believe that a properly modified OmegaWiki would catch on rapidly in this community.

### 6. Conclusions and future work

In conclusion, we have introduced the new concept of a wiki translation resource, which we define as an open, free wiki based resource that translators can use to find translations of problematic words, terms or expressions. Using user field data we have shown that, while translators stand to benefit from such a wiki resource, none of the existing ones fill that need in its current state. Firstly, these resources do not have sufficient coverage of typical problems encountered by translators. Secondly, their user interface does not make it easy to carry out key 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.

We have outlined a research and development agenda that could lead to such a wiki translation resource by improving and extending the existing OmegaWiki site.

In the future, we plan to work on parts of this R&D agenda, namely:

- Collaborate with the OmegaWiki team to build, test and evaluate a robot capable of extracting translation relevant information from Wikipedia and Wiktionary, and pour it into OmegaWiki.
- Collaborate with the OmegaWiki development team to improve their user interface, so that translators and terminologists can more easily and efficiently contribute or edit its content.
- Continue collecting user data from a wider range of translators, in order to identify special problems for those categories of users. In particular, translators who work on language pairs other than English and French, and non professional "hobbyist" translators who contribute to sites like Wikipedia and Global Voices.
- Continue development of the WeBiText bilingual web search system, and look for ways that it could work synergistic ally with OmegaWiki.
- Continue development of the TerminoWeb terminology extraction system, and look for ways that it could work synergistic ally with WeBiText and OmegaWiki.

Other interesting issues which we do not plan to investigate ourselves, are the development and testing of an automatic rating system for translations contained in a wiki translation resource, as per De Alfaro and L., Adler, or manual voting schemes such as the ones used on e-bay to rate sellers and buyers.

### References

Barrière and C., Agbago, A., 2006. TerminoWeb: A Software Environment for Term Study in Rich Contexts (TerminoWeb : Un environnement logiciel pour l'étude de termes en présence de contextes riches), International Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006). Beijing, Chine. Les 25 et 26 août 2006. NRC 48765.

Beyer H. and Holtzblatt K, 1998. Context Design: A Customer-Centered Approach to Systems Designs (Morgan Kauffman).

Bowker and L. And Pearson, J. (2002). Working with Specialized Languages. A practical guide to using corpora, Londres/New York, Routledge, xiv, 242 p.

Bowker, L., 2002. <u>Computer-aided translation technology : a</u> practical introduction. Ottawa : University of Ottawa Press, c2002.

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

De Saint-Robert, M.-J., 1991. Terminologie et documentation : le point de vue d'un terminologue aux Nations Unies, Pierre Pelou (dir.), La documentation internationale, coll. Systèmes d'Information et Nouvelles Technologies, Paris, ESF, p. 128-149.

Désilets, A., Gonzalez, L., Paquet, S., Stojanovic, M., 2006. Translation the Wiki Way. Proc. WikiSym 2006. Odense, Denmark. August 21-23, 2006. Giles, Jim, 2005. Internet encyclopaedias go head to head, Nature 438, pp. 900-901.

L'Homme, Marie-Claude, 2000. Initiation à la traductique, Brossard (Québec), Linguatech éditeur inc., 395 p.

Miller, G. A., 1995 WordNet: a lexical database for English Full text. Communications of the ACM, Vol 38, Issue 11 (November 195), pp 39-41, ISSN: 0001-0782.

Varantola, Krista, 2006. The Contextual Turn in Learning to Translate, Lynne Bowker (dir.). Lexicography, Terminology, and translation. Text-Based Studies in Honour of Ingrid Meyer, coll. Regards sur la traduction, Ottawa, Presses de l'Université d'Ottawa, p.215-226.

Wharton, C., et al., 1994. The Cognitive Walkthrough Method: A Practictioner's Guide. In Noobsen, Jakob, and Mack, R. (eds). Usability Inspection Methods. Chchester: John Wiley and Sons. ISBN 0-471-01877-5.

### Acknowledgements

This paper could not have been written without the hard work of the OPLT team (Louise Brunette, Christiane Melançon and Geneviève Patenaude) who spent countless hours producing, transcribing and encoding the user data that served as the basis for our analysis. The authors are also indebted to Gerard Meijssen and Kim Bruning of the Open Progress Foundation (<u>http://openprogress.org/</u>) who helped us better understand and evaluate OmegaWiki.