Exploring translators’ expectations of Wikipedia: A qualitative review

Elisa Alonsona, Bryan J. Robinsonb,*

*Corresponding author. Tel.: +034954978289; fax: +034954349817.
E-mail address: elialonso@upo.es

Abstract

We approach Wikipedia as a technology integrated in a social network surrounding the translation process. In that network, interactions take place between translators and Wikipedia. This paper's goal is to explore potential uses that translators could expect from Wikipedia. We have based our interpretative and qualitative work on a review of the state-of-the-art literature dealing with Wikipedia in several disciplines. We have related Wikipedia’s epistemology to the tools categorization proposed by Austermühl (2001) according to the translation process described by Holmes (1988). We have concluded that translators might use Wikipedia expecting to find linguistic, semantic, terminological, lexicographic and cultural information.

Keywords: Wikipedia; translation; translators; translation tools.

1. Introduction

One decade after its birth, Wikipedia has attained monumental proportions and this extraordinary project has clearly marked society at large and collectives, like translators, in particular. The conceptual framework where this work stands is based on the idea of “technology-mediated translation” or “trans-humanisation of translation” (Alonso, & Calvo, 2015), an approach where technologies, whether specific to translation or not, are not considered isolated tools, but instruments with the potential to develop further actions and interactions. Corpas (2004, p. 257) affirms that...
teletranslation practices, time pressure, and the levels of quality translators must achieve, have made the Internet an unrivalled resource, whereas Pinto (2004/2011) states that in the current information and knowledge society, individuals must be able to manage information and evaluate its quality in order to succeed in any activity. But this is not always easy, especially because new collaborative patterns and new revision and distribution models dismantle the modern model of authorship. One of the signs of our time is this never-ending process of creation. As Castells says (1996/2010, p. 31), unlike the earlier agricultural and industrial societies, our current informational society is defined by a cumulative feedback loop between innovation and the uses of innovation.

In our opinion, Wikipedia is a perfect example of how innovation loops can surpass any foreseen implementation of new work models and technologies. In the current technological paradigm, translators might be both consumers and producers (Cronin 2010) of Wikipedia articles, and in the present study we propose to focus on how they could use the encyclopaedia to solve translation problems.

A survey conducted by Torres (2012) shows that Wikipedia is a tool used by many translators but, in order to contribute to a better picture of the translator-Wikipedia binomial, a number of challenges remain. Firstly, we will try to summarise what Wikipedia is, what technology it relies on, and what conventions have allowed its foundation and development.

In order to explore the translators' potential expectations of Wikipedia, we will hypothesise about how Wikipedia could be present in the translation processes. We will explain how the qualitative review produced in this paper has been used to design a survey about the use and perceptions of Wikipedia among professional translators (Alonso, 2015 a).

This paper is exploratory and interpretative and seeks to shed some light on the way the translator may use Wikipedia in the translation process, bringing up additional considerations and pointing out the challenges that remain unclear.

2. Wikipedia: Features and conventions

The following paragraphs will be devoted to reviewing the state of the art literature dealing with Wikipedia as an open encyclopedic project, its epistemological features and its main conventions.

Over the last decade, Wikipedia and its founders, Wales and Sanger have attracted the attention of society and the mass media for a number of reasons; among others, Wikipedia's editorial policy (Waters 2006, & Rajan, 2009), its open access policy (de Vrieze, 2012), its neutrality (Bernstein, 2011) or its reliability (McHenry, 2004).

From academia, a number of disciplines, especially artificial intelligence (Artificial Intelligence, 2013), education and information sciences (Head, & Eisenberg, 2010; Aibar, & Fuster, 2012; Thornton-Verma, 2012) have contributed to understanding Wikipedia, its role in society and in education, and to developing a series of technologies that exploit its possibilities.

The historian Burke (2012) has devoted some attention to the emergence of Wikipedia as a popular source of knowledge. In his proposal, where he follows Foucault's conception of knowledge and social policy, Burke analyses the mechanisms of information retrieval and affirms that (2012, p. 109): “The most obvious and widespread form of information retrieval is human memory, but its limits and fallibility have provoked a long series of attempts to supplement it with artificial aids”.

According to Alonso (2015 b), Burke’s statement links, on the one hand, to the consideration of Wikipedia as an external aid, in the form of an electronic memory of humanity, and, on the other hand, to the postulates about “collective intelligence” proposed by Lévy (1997) or “participatory culture” set out by Jenkins et al. (2009, p. xi).

Wikipedia defines itself (Wikipedia 2001/2016) as an “Internet encyclopedia, supported by the non-profit Wikimedia Foundation”. As Ayers et al. (2008, p. 37) have pointed out, encyclopaedias “have traditionally been published as comprehensive guides to some defined area of knowledge”. However, what made Wikipedia exceptional was the successful implementation of a new editorial, authoring, and distribution model (Ayers et al., 2008, p. 37).

In the same vein, Benkler (2006, p. 70–71) thinks that the most remarkable features of Wikipedia are its use of wiki technology (the collaborative authorship tool), the GNU Free Documentation License, (inherited from free software) and its neutral point of view. The following paragraphs aim at exploring how these features set up Wikipedia's epistemology.
2.1. Wikipedia conventions

The successful implementation of the new authorship and distribution model of Wikipedia relies on a number of conventions that have been designed and incorporated by its founders and its community. In this sense, Stvilia et al. (2005, p. 444) have pointed out that "as an online encyclopedia, Wikipedia draws heavily on the well-established genre of printed encyclopedia by importing its form conventions and use norms". However, Wikipedia incorporates conventions that are not present in traditional encyclopedias. For example, it encourages the use of references at the end of every article, and the list of authors who have contributed to an article is visible in the “View history” tab of its text editor.

Though it is easy to make relatively small contributions to Wikipedia, the encyclopedia is in fact a very complex project (Saorín, 2012, Ayers et al., 2008). Furthermore, the fact that it provides the technical resources for any contributor to write content, does not mean that any content is acceptable.

Any author who wishes to publish in Wikipedia and become part of its community must devote many hours to reading its policies, guidelines, technical manuals, and style manuals (Ayers et al., 2008, p. 11–12). The community has been creating these papers by consensus since Wikipedia was founded and it strongly recommends users become familiar with all this very useful information, despite it being somewhat overwhelming for new users.

Although the Wikipedia community has done much to pursue the implementation of these policies, the encyclopedia has been involved in many public debates and scandals that have questioned its accuracy, neutrality and quality (Ayers et al., 2008, p. 51–58, and Benkler, 2006, p. 70–71).

In order to enhance its prestige, in addition to the use of banners warning about the deficiencies of some articles, the Wikipedia community has agreed a series of desirable characteristics of quality articles and those that meet these criteria can be proposed by the community, and approved by editors to become “featured articles”. Featured articles are considered the best articles Wikipedia has to offer, in the opinion of the Wikipedia editors (Wikipedia 2004/2016).

In addition, there are some visual signs that indicate the quality of an article in Wikipedia. Thus, the gold star in a featured article is a visible sign that certifies its quality from the Wikipedia editors’ point of view. In addition, the Wikipedia Version 1.0 Editorial Team (Wikipedia, 2006/2016) has designed a set of quality labels to indicate how close an article is to distribution-quality level. From highest to lowest, these are: FA (Featured Article), A (A-Class), GA (Good Article), Bplus, B, C, Start, Stub (very basic description of the topic), etc.

3. Wikipedia for translators

The approach we will pursue in the following lines is not focused on the individuals that translate Wikipedia articles, though, as a reference, the survey conducted by McDonough Dolmaya (2012), that follows this path, reveals that 68% of her survey’s respondents, who translated Wikipedia, did not have any formal training in translation, and this same percentage had never worked as professional translators. These figures suggest that professional translators do not participate in a great extent in translating Wikipedia. As a result, according to a recent survey by the same author (McDonough Dolmaya, 2016) – with a few exceptions – most translations contained in her Wikipedia corpus contained both transfer and language or style problems, and that these problems (particularly omissions) remained incorrect for months or years.

Indeed, there is an interesting research line focused on translation in crowdsourced environments, but our approach in this paper points to another field. This section is aimed at determining how Wikipedia could be present as an aid in the translation process. Our starting point is the idea of the “trans-humanization of translation” proposed by Alonso & Calvo (2015), that conceives translation as an extended cognitive, anthropological and social system or network which integrates human translators and technologies, whether specific to translation or not. Their approach is technology-mediated and envisages technologies in action and interaction with the human, fostering a plethora of instrumental developments, not only as isolated fragmentary tools utterly dominated by the human.

More specifically, in order to explore the possibilities of Wikipedia within the idea of a trans-human translation process, we will review the categorization of processes, proposed by Holmes (1988), and of tools, proposed by Außermühl (2001), as well as the state of the art research about Wikipedia, languages and translation.

For translators there is no longer any question of whether or not to use computers and information networks. The use of information and communication technology (ICT) is essential in the lives of today’s language professionals
In 2001, Austermühl analyzed the different types of translation tools proposed by Melby (1982) for the translation model described by Holmes (1988), offering a more dynamic approach. Surpassing the static classification of tools, Austermühl (2001, p. 9–15) focused on translation as a process, and on the tools that assist the translator in each of the three main stages: reception, transfer, and formulation. In Table 1, we show how the processes and sub-processes described by Austermühl relate to the tools to be used in each of them.

Table 1. Translation tools (Austermühl 2001) and the corresponding translation processes and sub-processes (Holmes 1988).

<table>
<thead>
<tr>
<th>Process</th>
<th>Sub-processes (Holmes)</th>
<th>Tools (Austermühl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reception</td>
<td>Text Understanding</td>
<td>Terminology Database</td>
</tr>
<tr>
<td></td>
<td>Linguistic Decoding</td>
<td>Reception Dictionaries</td>
</tr>
<tr>
<td></td>
<td>Information Retrieval</td>
<td>Electronic Encyclopedia</td>
</tr>
<tr>
<td></td>
<td>Problem Analysis</td>
<td>Knowledge Database</td>
</tr>
<tr>
<td></td>
<td>Text Analysis</td>
<td>Term Extraction</td>
</tr>
<tr>
<td>Transfer</td>
<td>Cultural Adaptation</td>
<td>Integrated Hypermedia Knowledge Bases</td>
</tr>
<tr>
<td>Formulation</td>
<td>Creative/Technical Writing</td>
<td>Terminology Production Dictionaries</td>
</tr>
<tr>
<td></td>
<td>Linguistic Encoding</td>
<td>Electronic Archives</td>
</tr>
<tr>
<td></td>
<td>Text Verification</td>
<td>Digital Corpora</td>
</tr>
<tr>
<td></td>
<td>Text Archiving</td>
<td>Document Management</td>
</tr>
</tbody>
</table>

Wikipedia is a comprehensive electronic encyclopedia. Since it was born on the Internet and has been developed through wiki technology it can also be considered an integrated hypermedia knowledge base. Although it is not intended to be considered as a dictionary, the truth is that it contains terminological and lexicographic material in many languages. It can also be considered a multilingual digital corpus, as suggested by recent works that try to exploit Wikipedia to retrieve multilingual information (Adafre, & de Rijke, 2006; Potthast et al.; 2008; Jones et al., 2008; Schönholen et al., 2008; Adar, Skinner, & Weld 2009; Nguyen et al. 2009; Ye et al., 2012). From the field of NLP (Natural Language Processing), as stressed by Aguado de Cea (2013), DBpedia – the ontology behind Wikipedia – can provide tools to edit interlanguage mappings and help build terminological resources. Particularly interesting is the research conducted by Oliver and Climent (2012) and the tool they have developed, WordNet, which seeks to process Wikipedia as a major semantic knowledge corpus for natural language processing tasks. These works suggest, that the potential of Wikipedia as a source of linguistic, semantic, terminological and lexicographic information has just started to be exploited, and that translators might find it useful to use Wikipedia for these purposes in all the main translation processes, as well as in most of the sub-processes.

General concern about the quality and accuracy of Wikipedia articles should not necessarily discourage translators from using Wikipedia. As Désilets et al. (2009, p. 7) observed in an ethnographic study about how translators use tools and resources to resolve translation problems, translators do not shy away from this kind of resource despite the apparent lack of control.

In the same vein, Lagoudaki (2006) asked a group of professional translators about their use of Internet in relation to their work. She found (2006, p. 10) that they made heavy use of the Internet as a source of knowledge and linguistic information. As stated above, Torres (2012) conducted a survey about the use of translation technologies by translators. When asked about the terminological and lexicographical resources that they used, more than 70% of the respondents reported using Wikipedia. In the same vein, according to Olvera & Gutierrez (2011), traditionally popular multilingual resources such as machine translators, dictionaries and corpora were increasingly giving way to ontologies such as Wikipedia.

Indeed, as stated by Saorín (2012, p. 11), Wikipedia cannot be considered a single encyclopedia, but a network of knowledge bases or cultural systems, one per language. In her epistemological analysis, Alonso (2015b) has suggested the possibility of considering Wikipedia as a cultural polysystem, applying the postulates of the theory elaborated by Even-Zohar (1990) and Lotman (1999), as well as from the social imaginary enounced by Vázquez-Medel (2008/2009). In more empirical works, different communities of practices have been reported by Hara, Schachaf, & Hew (2010) to have typical behaviors on Wikipedia discussion places across different languages. Following this same path, the research conducted by Rinser, Lange, & Naumann (2013) tries to find an effective approach to the
identification of groups of Wikipedia articles in different languages describing the same real-world entities. Applying visualization techniques, Biuk-Aghai, Pang & Si (2013) have represented maps of the most collaborative categories or topics across the Wikipedias in English, German, Chinese, Swedish and Danish, and have found interesting differences in the distribution of co-authoring over the Wikipedias studied. Finally, van der Velden (2013) explores the possibilities of decentering design in Wikipedia in order to adapt it for the production of indigenous knowledge.

The above lines depict a landscape where translators are comfortably installed in a highly globalized, technified environment in which their favorite documentation resource is the Internet, and where Wikipedia is present. They appear to have adapted to the new media and are able to quickly scan the results obtained from Internet search engines and select the appropriate information. They do not hesitate to use generalist, loosely controlled resources. From the traditional, “purist” perspective of translation, they can be considered to take “risky” decisions. However, they seem confident in their own abilities. Regarding the review of existing literature, we can observe a real wave of new research focused on Wikipedia as a linguistic and semantic resource that also contains a great deal of cultural information.

4. Use and perceptions of Wikipedia among translators

As explained in the introduction, the qualitative review pursued in this paper has recently been used to design an online survey (Alonso, 2015 a) aimed at exploring, from a technological and sociological perspective, how translators from the sample (a total of 412 professionals) conducted their work, the needs they experienced, and the tools and resources (human or human-driven) they resorted to when translating. More specifically, that interpretative and descriptive work looked at how participants used Wikipedia and analyzed their perceptions of this tool.

The method of that proposal was based on the work produced by Head and Eisenberg “How today's college students use Wikipedia for course related research” (2010). In order to adapt Head and Eisenberg’s model from the higher education arena to the translation context, the qualitative review conducted in this paper -together with previous surveys about translation technologies, and focus group sessions- were of paramount importance, since they enabled us to draft three separated lists addressing key topics that were incorporated in the online survey about:

a) The needs that professionals could have when translating (Alonso, 2015 a, p. 98): To find information about a topic; To find the meaning of a term; To find equivalent terms in two or more languages; To find reliable sources for a topic; To find previous translations of a term or sentence; To find the use of a term or sentence in context; To use translation memories; To create your own glossaries; To create your own corpus; To view the images associated with a term or idea; To search library resources; To use mobile devices (ex. smartphones, tablets); To negotiate your translation decisions with clients, agencies or colleagues; To consult with experts.

b) The tools and resources that professionals could use when translating (Alonso, 2015 a, p. 99): Wikipedia; Other online encyclopedias; Printed encyclopedias; Google; Search engines other than Google (ex. Yahoo!, Ask.com); Translation memories; Monolingual or aligned online corpora (e.g., Linguee, EUR-Lex, CREA corpus); Bilingual or monolingual printed corpora; Bilingual or monolingual online dictionaries; Bilingual or monolingual printed dictionaries; Terminology databases; Machine translation; Translators’ forums and mailing lists; Translation blogs; Blogs about other topics; Image search engines (ex. Google Images); Resources available in libraries; Mobile devices (ex. smartphones, tablets); Subject experts; Translator colleagues.

c) The purposes related to professionals’ use of Wikipedia (Alonso, 2015 a, p. 101): To find information about a topic; To find the meaning of cultural references; To find the meaning of a term; To find equivalent terms in two or more languages; To find reliable sources for a topic; To find previous translations of a term or sentence; To find the use of a term or sentence in context; To view the images associated with a term or idea; To create or feed your translation memories; To create your own glossaries; To create parallel texts or reference corpora; To justify your translation decisions before clients, agencies or colleagues.

The survey results (Alonso, 2015) suggested that most respondents made extensive use of all sorts of technologies when translating, amongst which TM and MT/post-editing were not the most popular. They also resorted to human (or human-driven) resources (translator colleagues, experts, social networks, blogs, etc.) to meet their needs (general documentation, terminological/lexicographical, visual).

According to the results, respondents had a good overall opinion of Wikipedia (usefulness, reliability and ease of use) and most of them reported using it when translating. However, some results suggested the existence of some kind of controversy or censorship with regard to the use of Wikipedia in professional contexts.
Interestingly, the hypothesis proposed in this paper: that Wikipedia could be present in the three main phases of translation (Reception, Transfer and Formulation) identified by Holmes (1988) seems to be confirmed by the results of Alonso’s survey. However, there is a clear emphasis on the purposes related to the Reception phase (‘Find information about a topic’ ranked 1, ‘Find the meaning of cultural references’ ranked 2, ‘Find the meaning of a term’ ranked 3, ‘View the images associated with a term or idea’ ranked 5, ‘Find the use of a term or sentence in context’ ranked 7, ‘Find reliable sources for a topic’ ranked 8); and, to a lesser extent, to the Transfer process (‘Find equivalent terms in two or more languages’ ranked 4, ‘Find previous translations of a term or sentence’ ranked 6).

5. Conclusions

As stated above, our approach, aimed at better understanding the binomial translators-Wikipedia, relies on the consideration of Wikipedia as a technology integrated in an extended cognitive, anthropological and social system or network surrounding the translation process, that has been called “trans-humanization of translation” or “technology-mediated translation” by Alonso, & Calvo (2015). In that network, a number of interactions take place between human translators and technologies and, as a consequence, it is worth exploring the expectations translators could have towards tools. In the same vein, we consider that Wikipedia is a great example of how new production models based on information technologies create a feedback loop between innovation and the uses of innovation in the way suggested by Castells (1996/2010).

We have noted that, despite its short history, Wikipedia has already attracted the attention of different disciplines, such as Artificial Intelligence, Information Sciences, Education and Translation Studies. In our epistemological analysis, we have pointed out that there is a relatively broad consensus around the idea of considering Wikipedia as an online encyclopedia, based on the principles of free software movement, and as an example of collaborative patterns. We have observed that, despite the fact that Wikipedia shares some common features with the genre of encyclopedias, it shows a number of unique elements, a set of conventions or policies – such as Verifiability (V), Non Original Research (NOR), and Neutral Point of View (NPOV)- that have been promoted and developed by its founders and its supporting community.

Almost any debate about Wikipedia inevitably has to deal with the fact that the encyclopedia has been involved in many public debates and scandals that have questioned its accuracy, neutrality, and quality. In this sense, we have pointed out that a number of mechanisms have been developed and implemented in order to promote the encyclopedia’s self-criticism and reflexivity, under the form of warnings that are exhibited to inform about the status or the quality of articles, as well as through the discussions behind decisions or changes in articles or policies. Another element that is aimed at improving the quality of Wikipedia is the existence of guidelines and a selection of featured articles, as well as some additional external signs, such as the gold star or a system based on quality labels that are shown in some article headings.

There is some evidence, such as the survey conducted by Torres (2012), which indicates that translators make heavy use of Wikipedia as a terminological and lexicographical resource, and this suggests that these could be, at least in part, some of their expectations. Traditional approaches to quality stress the importance of considering customers’ expectations which, in our case, has led us to hypothesize about the expectations translators could have towards Wikipedia. In our explorative and interpretative approach, we have tried to determine whether Wikipedia could be present in the translation processes and sub-processes described by Holmes (1988), according to the tools categorization proposed by Austermühl (2001). In our review of existing literature, we have pointed out a real wave of new research focused on Wikipedia as a multilingual, semantic and cultural knowledge base. These works suggest, that the potentiality of Wikipedia as a source of linguistic, semantic, terminological and lexicographic information has just started to be exploited, and that translators might find it useful to exploit Wikipedia for these purpose in all the main translation processes (Reception, Transfer, Formulation), as well as in most of the sub-processes, since it could fit almost under every categorization tool proposed by Austermühl (database, dictionary, knowledge base, digital corpora, etc.). As explained above, Alonso’s (2015 a) survey results seem to confirm these hypotheses and reveal that Wikipedia is particularly useful during the Reception and Transfer phases of translation.

In brief, Wikipedia has many strong points (size, hypermedia design, growth rate, diversity of topics, peer/crowd review, traceability, open concept, free distribution, plurality, ubiquity, multilingual, neutral point of view, consensus, resilience, unlimited, etc.) and many weaknesses (uneven quality, risky, non-professionally produced, subversive, loosely controlled, etc.). The origin of its most outstanding features probably has much in common with that of its weaknesses: its collaborative nature. For translators, translator trainers and the translation industry, Wikipedia is a
revolutionary tool. The key to success will depend on how we manage it in order to take advantage of its power, while controlling its limitations and risks.

Finally, we think that the arguments and details raised in this contribution could help inform further quantitative surveys or ethnographic works focused on the translator’s use, expectations and perception of Wikipedia, or contribute to the design of training programs for translators.

References


Aron, J. (2013). A tale of information ethics and encyclopaedias; or, is Wikipedia just another Internet scam? Conference on Web Search and Data Mining (pp. 94–103). New York (USA).


