

# ***Wikipedia as a source of monolingual and multilingual information about the heritage of Spain***

## ***A Wikipédia como fonte de informação monolíngue e multilíngue sobre o património construído de Espanha***

Dr. María-Dolores Olvera-Lobo<sup>1</sup>, Macarena Amo Valdivieso<sup>2</sup>, Juncal Gutiérrez-Artacho<sup>3</sup>,

<sup>1</sup> Tenured Professor of *Department of Information and Communication, University of Granada, Colegio Máximo de Cartuja, s/n, 18071, Granada (Spain) & CSIC, Unidad Asociada Grupo SClmago, Consejo Superior de Investigaciones Científicas, C/Albasanz, 26-28, 28037Madrid (Spain) [molvera@ugr.es](mailto:molvera@ugr.es)*

<sup>2</sup> *Postgraduate student of Faculty of Documentation and Communication, University of Granada, Colegio Máximo de Cartuja, s/n, 18071, Granada (Spain)*  
[mcrnamo@correo.ugr.es](mailto:mcrnamo@correo.ugr.es)

<sup>3</sup> *PhD Student ad Research Fellow of departament of Translation and Interpreting, University of Granada, C/ Buensuceso, 11, 18002, Granada (Spain)*  
[juncalgutierrez@ugr.es](mailto:juncalgutierrez@ugr.es)

### **ABSTRACT**

This study analyses the online encyclopaedia Wikipedia as a tool for the dissemination of monolingual and multilingual information about the heritage of Spain. . The sample is made up of catalogued buildings and sites of cultural interest in a region of Spain. For that purpose, the extent to which those buildings and sites are represented in Wikipedia by way of their own articles, that is to say, with information dedicated to each of them, has been examined. Three different types of search of the encyclopaedia (browsing, by way of the lists and through the internal search engine) were made during several months. Also, the extent to which Wikipedia includes multilingual information regarding the sample identified has been established, with identification of the number and type of languages in which the information is expressed, as well as the articles in languages other than Spanish which are the result of translation.

The results show that only 21.73% of the buildings and sites which constitute the final sample have a separate article in Wikipedia, of which only 20.85% are in several languages, with 50% being properly described as translations. This leads us to conclude that, given the collaborative nature of this tool, and also in view of the cultural and economic importance of historic and cultural heritage, institutions should take responsibility for the promotion the multilingual dissemination of this type of information.

**KEYWORDS:** Historic buildings and sites, Multilingual information, Heritage, Spain, Wikipedia.

## **RESUMO**

Este trabalho analisa a enciclopédia on-line Wikipédia como uma ferramenta de difusão de informação monolíngue e multilíngue sobre património histórico de Espanha. A amostra é composta por bens imóveis de interesse cultural catalogados de várias províncias espanholas. Para o efeito, estudou-se em que medida estes bens estão representados na Wikipédia através de um artigo próprio, ou seja, com informação independente dedicada a cada um deles. Procedeu-se à realização de três diferentes tipos de pesquisa na enciclopédia (navegação, por meio de anexos e através do motor de busca interno) durante vários meses. Por outro lado, determinou-se o grau em que a Wikipédia inclui informação multilíngue em relação à amostra, identificando o número e tipos de línguas nas quais se representa essa informação, bem como os artigos presentes noutras línguas diferentes do espanhol que são o resultado de uma tradução propriamente dita.

Os resultados mostram que apenas 21,73 % dos bens que constituem a amostra final possuem um artigo independente na Wikipédia, destes, somente 20,85 % se encontram em várias línguas, sendo 50% traduções propriamente ditas. Isto leva-nos a concluir que, dada a natureza colaborativa desta ferramenta e dada também a importância cultural e económica do património histórico, as próprias instituições deveriam responsabilizar-se em promover a difusão multilíngue deste tipo de informação.

**PALAVRAS-CHAVE:** edifícios e locais históricos, informação multilíngue, Património, Espanha, Wikipédia.

## INTRODUCTION

The Web is a well-established means of disseminating information. It is not subject to geographical boundaries and has an ever-more numerous audiences. Due to that nature, above all easy access to it from any part of the world, there is a need to offer multilingual information which is capable of meeting the needs of a wide range of users, thus overcoming language barriers. This is the context in which *Wikipedia* has come to be one of the main sources of information for Web users. According to its own definition, it is 'an encyclopaedia, understood to be a means for the compilation, storage and transmission of structured information. A *wiki* that can be edited by anyone, and which is free.<sup>1</sup> In effect, it is a free encyclopaedia, that is, there are no legal restrictions on use or distribution, and it can be modified. It is online (on the Web), collaborative (created by the online community), and multilingual (it aspires to include all the languages possible). It contains information in almost 300 different languages, of which English is the most representative with more than 4 million articles. A large part of the work on *Wikipedia* is done collaboratively, seeking to reach a degree of consensus. There are spaces for debate which facilitate co-ordinated participation, such as the Community Portal, the Village Pump and WikiProjects.

This study analyses the online encyclopaedia Wikipedia as a tool for the dissemination of monolingual and multilingual information about the heritage of Spain. The sample is made up of catalogued buildings and sites of cultural interest in a region of Spain. For that purpose, the extent to which those buildings and sites are represented in Wikipedia by way of their own articles, that is to say, with information dedicated to each of them, has been examined. Three different types of search of the encyclopaedia (browsing, lists and internal search engine) were made during several months. Further, there has been a review of the organisation of information in the encyclopaedia in categories and subcategories. Also, the extent to which Wikipedia includes multilingual information regarding the sample identified has been established, with identification of the number and type of languages in which the information is expressed, as well as the articles in languages other than Spanish which are the result of translation. The results show that only 21.73% of the buildings and sites which constitute the final sample have a separate article in Wikipedia, of which only 20.85% are in several languages, with 50% being properly described as translations. This

---

<sup>1</sup>A translation of the Spanish language version at:

[http://es.wikipedia.org/wiki/Wikipedia:La\\_enciclopedia\\_libre](http://es.wikipedia.org/wiki/Wikipedia:La_enciclopedia_libre) [Last accessed: 8 February 2015].

leads us to conclude that, given the collaborative nature of this tool, and also in view of the cultural and economic importance of heritage, Spanish institutions should take responsibility for the promotion the multilingual dissemination of this type of information.

## **WIKIPEDIA, SOURCE OF MULTILINGUAL INFORMATION**

The information on *Wikipedia* is organised, basically, into categories and articles. A category can include subcategories and links to articles. It can, in turn, belong to a parent category. In order to correctly categorise the entries or articles, a series of principles is followed, for example, such as not being redundant and including the article in all the categories to which by its nature it may belong. However, problems can arise when carrying out this organisation due to the application of Unicode standard — rules for the encoding of characters designed to facilitate the electronic processing of the information—.

For the creation of a standard article, *Wikipedia* offers a guide for the standardisation of the contents of the encyclopaedia. The guide recommends the inclusion of an introductory section or concise definition of the concept or subject of the article. The main body of the article must be structured into sections and subsections, and be accompanied by sections for notes and references, and for bibliographical references, as well as an additional bibliography and a section for external links. Finally, the categories in which the article is indexed must be stated, along with the interlanguage or *interwiki* links, that is to say, the links which connect articles on the same subject in different languages. They may be translations of the article or freshly written texts. Translations in *Wikipedia* are not generated by an integrated automatic translation tool but rather are done by contributors who, supposedly, have sufficient knowledge of the languages concerned. Nevertheless, although users are encouraged to provide correct translations in the language(s) in which they have language skills, this does not prevent the occasional instance when a contributor uses an automatic translator and effects a quick cut and paste, thereby incorporating low quality translations into the encyclopaedia. The pages or profiles of numerous contributors show signs, symbols and awards, amongst other things, which give information about language skills and other details of interest.

The rapid growth of *Wikipedia* in recent years has lead to numerous studies which make use of this resource in various settings. Responding to the academic interest

provoked by the access to multilingual information and the techniques of processing natural language involved in these processes (Peters, Braschler & Clough, 2004), *Wikipedia* has shown itself to be a resource for various multilingual applications (Sogra & Cimiano, 2012) such as translation of queries (Nguyen, Overwijk & Hauff, 2009), extraction from bilingual dictionaries (Ze *et al.*, 2012), automatic translations and the creation of comparable corpuses (Drexler *et al.*, 2014; Kiran, Santosh and Vasudeva, 2012), the multilingual search of answers (Olvera-Lobo & Gutiérrez-Artacho, 2011), amongst many other contexts.

In effect, its consideration as a source of multilingual information is a most noteworthy aspect of *Wikipedia*. Indeed, what is known as *Wikipedia* consists in truth not of one sole wiki but rather of a large number of wikis, each distinct and independent according to the language, thus forming a multilingual system. It follows that the contents, management and users of each Wikipedia are distinct and independent.

There are currently more than 280 Wikipedias according to the language in which the information is edited. The analysis presented herein focuses on *Wikipedia in Spanish*, which according to its number of content pages is ranked in tenth place amongst the different Wikipedias. As of 1st October 2014, it had 1,128,727 valid pages of content (a measurement taken in September 2013 revealed the existence of 1,050,076 pages of content, showing an increase of 78,651 articles, some 7% in 13 months).

The general objective of the research undertaken is to analyse the representation of the information on *Wikipedia* about buildings and sites of cultural interest in Spain, and the multilingual processing of that information. It must be borne in mind that heritage and the way it is managed undoubtedly influences activities such as tourism and economic development (Fernández Tabales & Santos, 1999; Fernando & López Guzmán, 2004). While there are subjects for whom ample multilingual treatment in the different Wikipedia is expected due to the global interest they generate, we consider that analysing the multilingual character of *Wikipedia* through local information makes it possible to determine whether this information source fulfils its aim of being a multilingual tool for Web users. When we refer to *local information*, we do so in the sense that it is linked to a specific geographical place, but is also of general interest as it forms part of the heritage, which is capable of generating attention and instilling in a wide range of *Wikipedia* and Web users a desire to become familiar with it and/or to visit it.

Other specific goals arise from the general objective set out in the section entitled *Research aims*, namely: (a) to know the extent to which Andalusian buildings and sites of cultural interest are represented in *Wikipedia* by way of their own articles, that is, with separate information dedicated to each; (b) to analyse how the information is organised in the encyclopaedia relating to the subject of study; (c) to determine the extent to which *Wikipedia* includes multilingual information in respect of the sample, identifying the number and types of languages in which that information is represented; and (d) to confirm which of the articles in languages other than Spanish are the result of translation.

## **METHODS**

The study sample comprises the buildings and sites, and activities of ethnological interest, of cultural interest of the provinces of Granada and Huelva contained in the *Catálogo General del Patrimonio Histórico Andaluz* (CGPHA) (General Catalogue of Andalusian Historic and Cultural Heritage), which is within the remit of the Education, Culture and Sports Department of the Regional Government of Andalusia. The said catalogue is a tool by which it is sought to safeguard the buildings and sites registered therein, as well as to facilitate enquires about and promotion of them. The two provinces chosen of the Autonomous Community of Andalusia have been selected because they have the largest and smallest number, respectively, of buildings and sites

of this type of heritage indexed in the catalogue. With the aim of establishing how the information about the sample is represented in *Wikipedia*, different types of search have been made in the encyclopaedia. Furthermore, the subject categories used by the encyclopaedia to organise the information have been analysed. Similarly, it has been determined how many of the buildings and sites in the sample have information in various languages, and which of them are translations. The procedure followed is set out below.

In order to know how many of these buildings and sites are described in a separate article in Wikipedia, dedicated exclusively to the building or site in question, three types of search were made during the period from one year (2013-2014):

a) browsing through the categories established in *Wikipedia* related to the sample, for example, 'Monuments and memorials of the province of...' and 'Defensive towers of...', amongst many others. In order to perform this process, a tour of the categories was made, starting with the most generic (Portal: Arts) until the most specific at the final level. Figure 1 provides an example of navigation through the categories of *Wikipedia*.

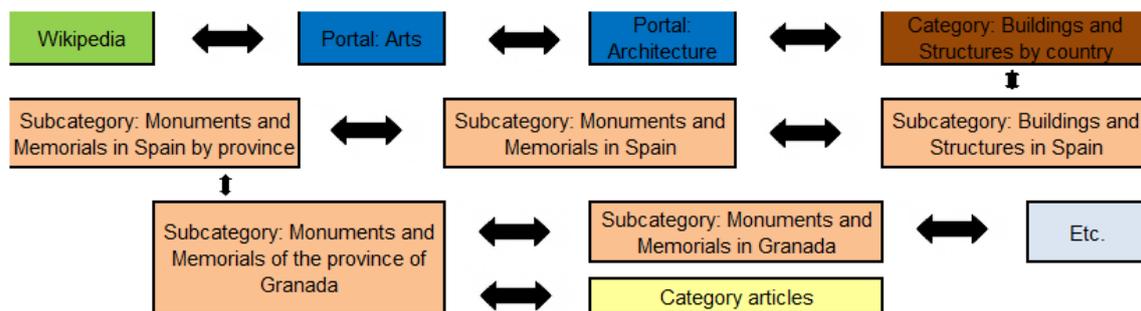


Figure 1. Example of navigation by browsing (2014)

b) analysis of the lists relating to the buildings and sites of cultural interest that make up the sample. These lists, included in the categories previously analysed by the browsing method, proved to be of interest given that, on the one hand, they helped to bring to light some buildings and sites that had gone unnoticed in the preceding procedure and, on the other hand, they revealed a number of mistakes on *Wikipedia* worth recording.

c) queries through the search engine integrated into *Wikipedia*, which made it possible to locate, by way of the name of the building or site in question, those which had not been found via the preceding two routes. This search engine only retrieves

information included in *Wikipedia in Spanish* since, as stated above, each *Wikipedia* is independent. This method allowed the recovery of some elements that were insufficiently categorised and which were not correctly included in the corresponding list.

On proceeding to the search for the various buildings and sites in the encyclopaedia, a number of problematic cases were detected. This was due to the manner in which they had been registered in *Wikipedia* or even in the Catalogue from where they had been obtained. Similarly, there were mistakes made by users when creating the encyclopaedia's content. Faced with such cases, the following criteria were adopted:

- a) Historic buildings included in more than one municipality – some buildings and sites appear listed more than once, given that they are included or pertain to more than one municipality. For the purposes of this analysis they have been taken into account only once.
- b) Buildings and sites with uncertain representation in the encyclopaedia – in some cases, the buildings and sites whose information is analysed do not have their own separate entry within *Wikipedia*, but it is seen, however, that in the article corresponding to the municipality to which it pertains, a certain amount of information about it is included. In order to register these cases, the category called 'uncertain representation' was created. However, these have not been included in the analyses made together with those buildings and sites which do have representation in *Wikipedia* in a separate article.
- c) Coincidences in some of the names of the buildings and sites – there are buildings and sites included in the *Catálogo General del Patrimonio Histórico Andaluz* which coincide with at least one of the alternative names that are assigned and listed in the section 'other name'. In that event, given that it does not really concern different buildings and sites, if in *Wikipedia* there is a separate article for that 'other name', that is the entry taken into account for the purposes of the analysis made herein.
- d) Errors of omission of buildings and sites – two types of omission have been detected: (i) catalogued buildings and sites of historic and cultural heritage which do not appear in *Wikipedia*, and (ii) buildings and sites present in *Wikipedia* which do not appear in the CGPHA. In neither of those cases have the buildings and sites been able to form part of the analysis.
- e) Unions of buildings and sites – in certain cases, what appears in the catalogue as one building or site, as in the case of 'The Alhambra and the Generalife,' appears in

*Wikipedia* as two separate articles, one “The Alhambra,’ the other ‘The Generalife’.  
In such cases, they have had to be considered as two distinct buildings or sites.

- f) Duplicated – sometimes there is more than one article in *Wikipedia* about the same building or site. This occurs because some buildings or sites have various alternative names in addition to the official name, giving rise to confusion. In that event, only one of the names has been taken into account, choosing the article which offers the most information.

After carrying out the searches on *Wikipedia* for the buildings and sites of historic and cultural interest which constituted the initial sample, and having applied the information selection criteria detailed above, the final sample constituted 971 buildings and sites, amongst which a wide variety of elements was found, ranging from buildings of international popularity to subaquatic spaces unknown to the public at large.

The search for buildings and sites of cultural interest in accordance with the various procedures described above revealed that of the 971:

- Only 211 (21.73%) of the buildings and sites of the sample were found to be represented in a separate *Wikipedia* article during the data compilation period.
- 719 (74.05%) were found to be not represented on *Wikipedia* during the data compilation period.
- 41 (4.22%), being the remainder, were deemed to be of uncertain representation. They cannot be included in the represented group as they did not have a separate article, but there was considerable information about them in the article within which they were included.

Therefore, the analysis centred on the 211 buildings and sites represented on *Wikipedia* by way of a separate article or entry dedicated thereto for which the following information was registered:

- Method of retrieval: it was recorded whether the information relating to the particular building or site was found through browsing, via the corresponding list or by using the internal search engine of *Wikipedia*.
- Position in the structure of *Wikipedia*: the categories and/or subcategories which nest the said buildings and sites are shown.
- Name of the building or site: in some cases more than one name has been included, when the name by which it appears in *Wikipedia* is one of the alternative names, or similar, instead of the official name by which it appears in the CGPHA.

- Municipality to which the building or site appears: this makes it possible to locate them geographically and to avoid the possibility of confusing it for buildings and sites in other places with the same or similar name.
- Date of creation: shows the age of the article and makes it possible to know, in the event that it is in more than one language, which was created first.
- Last date of revision: makes it possible to identify how up to date the information is.
- Languages in which it is available: interlanguage links which the article contains.
- Translations: where available in more than one language, it has been indicated which are translations of the Spanish article.

## RESULTS

### 1. METHOD OF RETRIEVAL

Of the 211 articles that finally made up the sample analysed, the vast majority (184 articles, 87%) had been located on *Wikipedia* through browsing given that it was the first method of information search applied with the aim of checking the structure of the subject categories and subcategories. Thus, as expected, a smaller number of articles was retrieved when searching through the list (15 articles, 7%) —the second method to be used— and, finally, through the search engine (12 entries, 6%), since there was no new search for the buildings and sites which had already been found through browsing. However, when applying those latter two retrieval methods, various articles were found which otherwise would have been missed. Indeed, if the categorisation of the articles were perfect and there were no human error, all the articles would have been retrieved through *browsing*.

### 2. CATEGORIES ON WIKIPEDIA

In the *Catálogo General del Patrimonio Histórico Andaluz*, from which the sample was obtained, the buildings and sites analysed were found to be organised around categories such as: activities of ethnological interest, set of historic buildings, historic gardens, place of ethnological interest, place of industrial interest, monuments and memorials, historic site, archaeological zone, heritage zone and others/not classified. *Wikipedia* has a much greater quantity and variety of categories and subcategories for

structuring the contents of the different subjects it contains. Historic and cultural heritage is no exception. As there is a greater diversity of categories in the encyclopaedia, each of them has a fewer amount of elements. Indeed, there are several *Wikipedia* categories where only one article associated with our sample has been found nested, as seen in table 1. The category 'buildings and sites of cultural interest' in the table includes all types of buildings and sites from our sample, which have been categorised in *Wikipedia* at that generic level and which did not have a more specific category linked to them. The second chart shows the categories with more than five elements in the encyclopaedia. The other items have been included in a single category called 'others'.

Table 1. Buildings and sites of the sample according to their category in *Wikipedia* (2014)

<b>Wikipedia Category</b>	<b>No. of buildings and sites</b>	<b>Percentage</b>
<b>Churches</b>	<b>40</b>	<b>18.96</b>
<b>Castles</b>	<b>32</b>	<b>15.17</b>
<b>Buildings and sites of cultural interest</b>	<b>29</b>	<b>13.74</b>
<b>Defensive towers</b>	<b>26</b>	<b>12.32</b>
<b>Monuments</b>	<b>14</b>	<b>6.64</b>
<b>Ritual dances of Huelva</b>	<b>14</b>	<b>6.64</b>
<b>Museums</b>	<b>8</b>	<b>3.79</b>
<b>Archaeological sites</b>	<b>8</b>	<b>3.79</b>
<b>Municipalities</b>	<b>6</b>	<b>2.84</b>
<b>Baths</b>	<b>5</b>	<b>2.37</b>
<b>Hermitages</b>	<b>4</b>	<b>1.90</b>
<b>Convents</b>	<b>3</b>	<b>1.42</b>
<b>Bridges</b>	<b>3</b>	<b>1.42</b>
<b>Chapels</b>	<b>2</b>	<b>0.95</b>
<b>Palaces</b>	<b>2</b>	<b>0.95</b>
<b>Monumental gates</b>	<b>2</b>	<b>0.95</b>
<b>Dolmens</b>	<b>2</b>	<b>0.95</b>
<b>Archives</b>	<b>1</b>	<b>0.47</b>
<b>Libraries</b>	<b>1</b>	<b>0.47</b>
<b>Cathedrals</b>	<b>1</b>	<b>0.47</b>
<b>Bastions</b>	<b>1</b>	<b>0.47</b>
<b>Parks and Gardens</b>	<b>1</b>	<b>0.47</b>
<b>Bullrings</b>	<b>1</b>	<b>0.47</b>

Historic sites	1	0.47
Royal Chancery	1	0.47
Botanical gardens	1	0.47
Paleontological sites	1	0.47
Andalusian coast	1	0.47
	211	

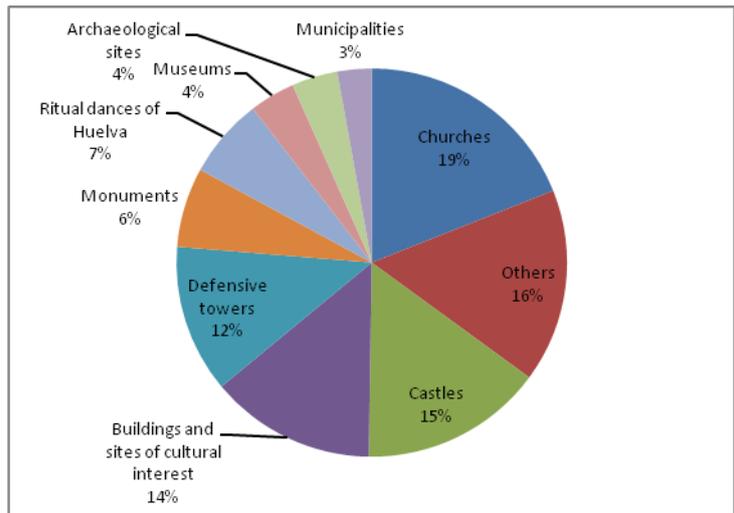


Figure 2. Principal buildings and sites of the sample according to their category in *Wikipedia* (2014)

The results demonstrate that, of the buildings and sites included in the study sample, churches and castles are the most represented in *Wikipedia*. Pursuant to the information in the CGPHA, both of those categories would be within the much more generic category ‘monuments and memorials’ which includes the larger part, more than 60%, of the buildings and sites of the sample.

### 3. MULTILINGUAL ARTICLES AND TRANSLATIONS IN WIKIPEDIA

The interlingual links of the 211 buildings and sites in the sample have been examined to check if there was an article about the building or site in a language other than Spanish. The amount of articles available in more than one language was very small (44 entries, 20.85%), with the large majority of articles only available in Spanish (167, 79.15%), although some specific entries have a large variety of interlingual links available, as is the case with the Alhambra, whose fame and worldwide recognition has

undoubtedly contributed to that fact. Nevertheless, it cannot be ruled out that in some cases it is possible that there were articles in other languages included in different Wikipedias which did not include their corresponding interlingual link.

As seen in table 2, of the 44 articles that are in more than one language, 22 (50%) are in two different languages, nine in more than five languages, five in three languages, five in four languages and three are in five different languages.

Table 2. Buildings and sites of the sample in more than one language in *Wikipedia* (2014)

		%	% of the total
<b>2 languages</b>	22	(50%)	(10.43%)
<b>3 languages</b>	5	(11.36%)	(2.37%)
<b>4 languages</b>	5	(11.36%)	(2.37%)
<b>5 languages</b>	3	(6.83%)	(1.42%)
<b>More than 5 languages</b>	9	(20.45%)	(4.26%)
<b>Total</b>	44	(100%)	(20.85%)

The language other than Spanish in which most of the multilingual articles of the sample are found is Portuguese (85.8% of the total of the entries), 36.4% are in English, 34.1% in French and 27.3% in Catalan. Chart 3 shows the most representative languages of the multilingual entries analysed.

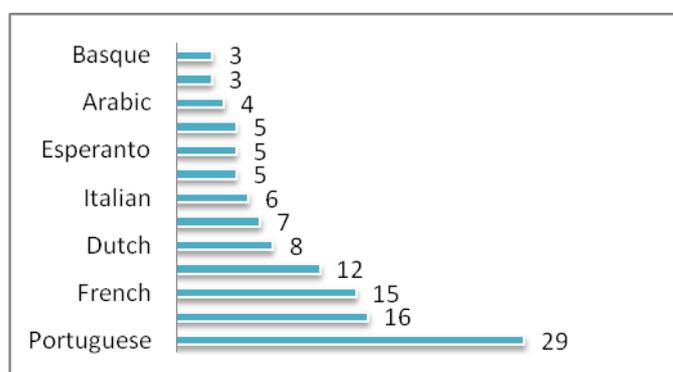


Figure 3. Most representative languages of the multilingual information of the sample (2014)

As shown, not all the interlingual links refer to translations, some refer to a newly prepared entry. It has been checked which of the multilingual entries are translations. Accordingly, it has been possible to detect that the language into which the original Spanish articles have most often been translated is Portuguese (16), followed by English (5), French (3), Catalan (1) and Galician (1). One reason for the greater

number of translations into Portuguese than other languages is due to the high productivity of some of the contributors who undertake translations of *Wikipedia* articles. A review of the multilingual entries of our sample (44 in total) shows that for those entries that are translations (22), the most productive author's first language is Portuguese and, according to the author, he or she also has a good command of English, French and Spanish. This user has carried out 14 of the 26 translations found.

## **CONCLUSIONS**

Given the inevitable limitations of the study sample and the dynamic nature of the information in *Wikipedia*, the results of this analysis are provisional. However, we consider that fact does not prevent us from reaching a number of useful conclusions. After completing the study, it can be confirmed that, in relation to our subject of interest, namely the subject domain of the historic and cultural heritage, and the sample used whose characteristics have been detailed, the coverage offered by *Wikipedia* is somewhat inadequate. The percentage of representation reaches only 21.73% of the buildings and sites of the sample and, further, many of the entries and articles included in the encyclopaedia contain only sparse information about the element with which they deal, and not only in other languages but also in Spanish. However, in many cases it is clear that the building or site concerned did not lend itself to the compilation of further information about it. A feature that undoubtedly influences representation is the fame or international recognition of the buildings and sites, with those that boast greater popularity being present or better represented in *Wikipedia*.

In addition, it has been verified that the encyclopaedia's complex framework of categories is occasionally counterproductive when it comes to retrieving the information. The majority of the mistakes found in the encyclopaedia are related to the linking of content, but these errors are not as numerous as might be thought and, further, this is offset by the rapid growth in content and speed of answer and updating.

It can also be confirmed that, using the interlingual links within an article, it is possible to access the corresponding entries in other languages pertaining to other Wikipedias, given that these links act as bridges assisting interaction between the different Wikipedias. Here, again we encounter the fact that only the entries for buildings and sites acknowledged to be international interest have a notable amount of interlingual links, as is the case of the Alhambra, the Capilla Real (Royal Chapel) of

Granada, and some historic centres of certain municipalities but, in general, the multilingual information about these aspects of Spain's historic and cultural heritage.

Finally, the examination of the *Wikipedia* articles with interlingual articles indicates that a number of them are translations from the original done by the users, although not all the articles in different languages which are about the same subject are necessarily translations. Some contributors prefer to create a new article in another language.

In summary, going beyond the polemic provoked by the quality of the information contained in *Wikipedia* (Saorín, 2012), it is beyond question that it is a remarkably useful and popular tool. However, much more could be obtained from it, above all from the perspective of the retrieval of multilingual information, that perspective being considered herein. The sparse coverage in the encyclopaedia regarding this type of information, of crucial importance from a cultural and economic point of view, allows us to conclude that, given that *Wikipedia* is a collaborative resource, it would be desirable for institutions to show a greater interest in promoting the dissemination of information about historic and cultural heritage.

### **Acknowledgements**

This research has been carried out within the framework of the project HAR2012-38562 Platform of Analysis and Access to multilingual web information about the historic and cultural building and sites of Spain, financed by the Ministry of Economy and Competitiveness.

### **REFERENCES**

- Drexler, J.; Rastogi, P.; Aguilar, J.; Durme, B.V.; Post, M. A Wikipedia-based Corpus for Contextualized Machine Translation, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, p.3593-3596, 2014.
- Fernández Tabales, A.; Santos, E. Turismo y Patrimonio Histórico en Andalucía: magnitudes generales y estrategias de planificación, *Boletín de la Asociación de Geógrafos Españoles*, v. 28, p.119–134, 1999.

- Fernando, L.; López Guzmán, T.J. El turismo como motor de desarrollo económico en ciudades patrimonio de la humanidad, *PASOS: Revista de Turismo y Patrimonio Cultural*, v. 2, n.2, p.243–256, 2004.
- Kiran, K.N.; Santosh, G.S.K.; Vasudeva, V. Multilingual Document Clustering using Wikipedia as External Knowledge, *Proceeding IRFC'11. Proceedings of the Second international conference on Multidisciplinary information retrieval facility*, p. 108-117, 2012.
- Nguyen, D.; Overwijk, A.; Hauff, C. WikiTranslate: Query Translation for Cross-Lingual Information Retrieval Using Only Wikipedia. *CLEF'08: Proceedings of the 9th Cross-language evaluation forum conference on Evaluating systems for multilingual and multimodal information access*, p. 58–65, 2009.
- Olvera-Lobo, M.D., Gutiérrez-Artacho, J. Language resources used in multi-lingual Question Answering Systems, *Online Information Review*, v.35, n.4, p.543–557, 2011.
- Peters, C.; Braschler, M.; Clough, P. *Multilingual Information Retrieval. From research to practice*, Springer, New York, 2004.
- Saorín, T. *Wikipedia de la A a la W*, UOC, Barcelona, 2012.
- Sogra, P.; Cimiano, P. Exploiting Wikipedia for cross-lingual and multilingual information retrieval, *Data & Knowledge Engineering*, v.74, p.26-45, 2012.
- Ye, Z.; Huang, J. X.; He, B.; Lin, H. Mining a multilingual association dictionary from Wikipedia for cross-language information retrieval. *J. Am. Soc. Inf. Sci.*, v.63, pp.2474–2487, 2012.