

STATE-OF-THE-ART ON MONOLINGUAL LEXICOGRAPHY FOR THE BASQUE COUNTRY (BASQUE)

KLARA CEBERIO

The Elhuyar Foundation

ANTTON GURRUTXAGA

The Elhuyar Foundation

Ceberio, K., Gurrutxaga, A. (2019): State-of-the-art on monolingual lexicography for the Basque country (Basque). Slovenščina 2.0, 7 (1): 53–64.

DOI: <http://dx.doi.org/10.4312/slo2.0.2019.1.53-64>.

In this article, we give an overview of the evolution of Basque lexicography to the present, pointing out its main achievements and shortcomings, as well as its challenges for the future. Basque lexicography has a relatively short history, but a considerable amount of resources have been produced in the last 50 years, since the standardisation process began. After years of lexicographic work by different groups and publishers, a remarkable achievement is the *Dictionary of the Academy (Euskaltzaindiaren Hiztegia)*, a prescriptive updated dictionary recently published and based on historical and contemporary corpora. Although the number of monolingual products is noticeably increasing in the last years, Basque dictionary making has been specially productive for bilingual purposes, due probably to the sociolinguistic status of the language. On the other hand, specialized lexicography and terminology have been very active from the beginning of the standardisation process. Since the beginning of the XXI. century, use of corpora has known an increasing impulse. Many Basque dictionaries are freely available on the Internet.

Keywords: Basque, dictionary, corpus

1 BEGINNINGS

The history of Basque lexicography was not very rich until the language's process of standardisation began in 1968. The first attempt towards a general, unified dictionary was probably the *Diccionario Trilingüe del Castellano, Bascuence, y Latín* (Trilingual dictionary of Spanish, Basque and Latin) by Larramendi (1745). Larramendi tried to fill gaps in the language by creating neologisms for the missing technical and scientific words in Basque.

In his dialectal and historical dictionary (*Diccionario Vasco-Español-Francés; 1905-1906*), Azkue excluded recently borrowed terms and neologisms. Lhande published a dictionary (*Dictionnaire basque français; Dialectes Labourdin, Bas-Navarrais et Souletin, 1926*) which only covered the Eastern dialects. He used a broader criterion, in that he treated as Basque all words that were actually used.

One important point in this period is the absence of monolingual dictionaries. The lexicographic work focused mainly on bilingual dictionaries: compiling a common and dialectal lexicon using Spanish or French translations, or proposing Basque equivalents for Spanish or French words. For years, the lack of technical, scientific terminology and an official regulatory body which could take on the task of proposing the necessary neologisms, forced lexicographers to fill in these gaps on their own or to leave them out of their dictionaries (Azkarate, 1991).

When the process of standardisation began, the first task undertaken by Euskaltzaindia, the Academy of the Basque Language, was to unify the orthography, morphology and verbs; little attention was given to the lexicon itself, and it was limited to small lists of basic words. Thus, some lexicographers began to compile dictionaries with the aim of providing the user with resources to use Basque in “new” areas and registers. Of all the dictionaries produced from this endeavour, the *Hiztegia 80*, published by Kintana et al. in 1980, was the most successful. It was a clear attempt to make a modern, standard, medium-sized bilingual dictionary, using the unified orthography proposed in 1968 by the Academy.

2 SPECIALIZED DICTIONARIES

Since the mid-1970s, several organisations and working groups (mainly UEU, UZEI, and Elhuyar Foundation) undertook the task of expanding the use of Basque in specialist areas, especially science and technology, and providing the terminology and stylistic resources in Basque needed to do so. As a result, several technical, multilingual dictionaries were published. The task of defining words in Basque, a crucial step towards monolingual dictionaries, was first undertaken in technical areas. This field is still very active, and some of the most remarkable results are *Euskalterm*, the *Basque Public Term Bank* of the Basque Government, originally created by UZEI; and the *Zientzia eta Teknologiaren Hiztegi Entziklopedikoa* (*Encyclopaedic Dictionary of Science and Technology*) edited by Elhuyar. There are also some other interesting technical dictionaries (Environmental Sciences, Energy, Telecommunications, Nursery and Automotive, among others). In 2007, the Academy launched a project to collect and analyse the specialized lexicon used on technical texts, education books, popular science etc. The goal is to reflect the use of specialised lexicon and, in some cases, to give recommendations, but trying to minimize their amount, in order to generate the least impact on the real usage. Basic lexicons on Mathematics, Physics, Astronomy and Chemistry have been already published, and made available in PDF format (*Oinarrizko Lexiko Teknikoak*).

3 MONOLINGUAL DICTIONARIES

In 1984, the Academy approved a long-term plan for the development of dictionaries with the aim of laying the groundwork for developing a unified standard dictionary of Basque (Sagarna, 2010). On the one hand, the *General Basque Dictionary* should be a compilation of vocabulary used in the corpus of publications until 1970, and, on the other hand, the vocabulary used in the *Statistical Corpus of the 20th Century*. Both projects have already been completed. The *General Basque Dictionary* (1987-2005) has 16 volumes in paper format and each entry contains the following information: the lemma and the variants of the word as they appear in texts, information about the dialects corresponding to these variants, according to usage in the last hundred years, the meanings (by reference to Spanish equivalents), and the history and examples of usage

of the word, its compounds, expressions and etymology. Since 2009, the *General Basque Dictionary* has been available online.

Regarding monolingual general dictionaries, the *Euskararako Hiztegia* by the Adorez group was the first one to be published in 1986. Nevertheless, one of the most influential works was the *Hautalanerako Euskal Hiztegia* by Ibon Sarasola (1984–1995), and re-edited as *Euskal Hiztegia* (1996). It was compiled using the corpus of the *General Basque Dictionary*. For the first time, a Basque dictionary gave the date of the first occurrence of a word in the corpus. It also includes discussions and practical considerations about which word and in which form should be chosen for the standard lexicon. The three dictionaries mentioned above are only available in book format.

The nineties were very productive for Basque lexicography. Firstly, the Basque Academy created a commission to prepare the Unified Basque Dictionary (2000-2016). It is mostly an orthographic or spelling dictionary, a word list with some additional lexicographic data. In some cases, grammatical or semantic notes are added. In order to establish the standard form of a word, the use of that word both in a historical context and in modern times had been taken into account. Secondly, the first encyclopaedic dictionaries were published: *Elhuyar Hiztegi Entziklopedikoa* (1993), *Lur Entziklopedia* (1995), and *Harluxet Hiztegi Entziklopedikoa* (1998). The last two encyclopaedias can be browsed online. Thirdly, we should mention the Basque Wikipedia which was founded in 2001 and nowadays has about 284,000 articles. Finally, Elhuyar published a monolingual dictionary in 1994, the *Euskal Hiztegi Modernoa*.

In recent years, one of the most notorious works has been the Academy's monolingual dictionary, a work that due to its importance will be described in a separate section (see section 4).

4 BILINGUAL DICTIONARIES

In 1996, a new generation of Spanish-Basque bilingual dictionaries was released; for example, *Hiztegia 3000* (Adorez) and *Elhuyar Hiztegia*: standard size, detailed microstructure, run-on entries etc. Since then, the field of bilingual dictionaries has experienced unprecedented growth.

One of the most popular sites is *Elhuyar Hiztegiak*, which provides the option of searching three Basque bilingual dictionaries in a single click (from Basque to Spanish, French and English), together with the displaying of real examples from the *Elhuyar's Basque-Spanish web parallel corpus*. The dictionaries in this site are updated on a daily basis, mainly thanks to the contribution of users, who can propose new entries or senses, and changes in the content of the entries (equivalents, examples, etc.). Mobile apps for Android and iOS are also available and a version of the Spanish Basque option is available for Kindle e-readers.

An interesting project to be mentioned is *Nola Erran*, a French-Basque online dictionary which is not supported by a previous printed edition, edited by *Euskararen erakunde publikoa*.

5 SCHOOL DICTIONARIES

In the field of dictionaries for learners and students, Adorez published the first edition of his *Eskola Hiztegia* in 1995, Elhuyar published its first *Ikaslearen Hiztegia* (Dictionary of the Student) in 1997 and Vox published the *Euskara Ikaslearen Hiztegia* in 1999. All of them were published in book format. The last edition of Elhuyar's *Ikaslearen Hiztegia* (2008) can be consulted on the Internet using a code available with the purchase of the book.

We have no precise data about dictionary use at schools. Anyway, some shallow surveys made in education centres, and sales data we have access to, suggest that printed dictionaries are still being used at the classroom, maybe due to the shortage of monolingual school dictionaries available online.

6 CORPORA

Since 2005, Basque corpus building has been experiencing a remarkable growth, thanks to the *Corpus of Science and Technology*, the *Contemporary Reference Prose*, the *Lexicon Observatory*, *EHUskaratuak*, the *Elhuyar Web Corpus* and the *Corpus of Contemporary Basque*. The size of the last two corpora has reached the milestone of 100 million words (125 and 269, respectively).

A good example of corpus driven lexicography is the *Egungo Euskararen Hiztegia* (*Dictionary of Contemporary Basque*), a project designed to make available

online a dictionary of Basque as it is used today. The corpus behind is the already mentioned *Ereduzko Prosa Gaur (Contemporary Reference prose)*, a 25.1 million-word collection of books and newspaper articles written between 2000 and 2007 and selected for its linguistic quality. It is an ongoing work, which includes at present 58,034 entries and 435,201 examples.

According to the best of our knowledge, we believe that people in general do not know much about the existence of corpora. Even though the use of corpora has spread from the specialized linguists to other professionals who use Basque in their jobs (teachers, journalists, translators, editors...), or even to students who are specializing in the Basque language, our impression is that the potential of Basque corpora has not been fully exploited. Greater efforts should be made to extend the use of corpora, as they are a very valuable resource to access information about the real use of words in context.

7 DICTIONARY OF THE ACADEMY: A REFERENCE WORK

The *Euskaltzaindiaren Hiztegia* or the *Dictionary of the Academy* was first published in 2012 in book format. A second edition was published in 2016, and nowadays the *Dictionary* has more than 40,000 entries and 61,000 definitions. The previously mentioned *Unified Basque Dictionary* was taken as a starting point for the selection of list of words. The dictionary entries contain the following information: the lemma and the part of speech, information about the dialects, about the fields, registers, the scientific names of the living organisms and examples. Examples taken from corpora are a very valuable resource to clarify the right use of the word. The *Dictionary of the Academy* is available on the Internet since 2016, and a pdf version can also be downloaded. Currently, the Academy is using the aforementioned *Lexicon Observatory* corpus to enhance and update this dictionary. This project aims to create a monitor corpus, that is to say a corpus designed to show the changes that are occurring in the use of Basque. The corpus contains texts published after 2000, and it is being fed continuously. In 2017, the corpus size is 58.5 million words.

The *Dictionary of the Academy* constitutes an invaluable resource for Basque language users. It is a prescriptive updated dictionary based on historical and contemporary corpora.

8 USE OF INTERNET

As we have mentioned before, many Basque dictionaries are freely available on the Internet. It has to be taken into account that according to the latest study on the use of Internet by *Fundación Telefónica* (2016), 85% of Basque people between 16 and 74 usually use the Internet. So dictionary publishers are making a big effort to adapt their dictionaries to the new trends and make them available on the Internet.

9 HOW ARE THE DICTIONARIES FUNDED

Concerning the subject of how Basque dictionaries and other lexicographic works are funded, there are three types of agents mentioned in the present paper:

- institutions such as UEU, UZEI and Elhuyar: private organizations funded partly by public funds, at a very different extent;
- public institutions such as the Basque Academy and the University of the Basque Country: they are fully funded by public funds;
- dictionaries compiled and guided by an author such as *Morris Dictionary*: they are funded with private funds and receive public grants.

In terms of the perception of people about the quality of the lexicographic products according to who has compiled them, we do think that the way the work has been funded is not taken too much into account when users buy or use a product. They appreciate more being a Basque institution or company behind the compilation work, rather than the funds needed to compile it.

10 CROWDSOURCING

We do not have enough data regarding crowdsourcing in lexicography. However, Elhuyar has some short experience in the area of terminology. In reference to the *Zientzia eta Teknologiarean Hiztegi Entziklopedikoa* (*Encyclopaedic Dictionary of Science and Technology*), apart from being on the Internet, a step forward has been taken shortly. It is now a collaborative encyclopaedia, where anyone can participate and collaborate. The contents are being supervised by the experts in Elhuyar. We will evaluate the results of this experience in the future.

11 RESEARCHING ON DICTIONARY USE

As far as we know, almost nothing has been done in terms of researching dictionary use. Taking into account the short history of Basque lexicography, editors have had other priorities than researching dictionary use. Anyway, it is an issue that should be seriously considered in the near future.

12 CONCLUSIONS

Although Basque lexicography has a relatively short history, a considerable amount of resources have been produced in the last 50 years. As book sales in general and dictionaries in particular have come down in the Basque Country during the last ten-fifteen years, financial means have consequently been decreasing. On the other hand, although the number of monolingual products is noticeably increasing in the last years, we do believe that people in general use bilingual dictionaries (eu-es, eu-fr, eu-en) more than the monolingual ones. This fact can be attributable to two factors: the short history of Basque monolingual lexicography and the sociolinguistic status of the language.

REFERENCES

Articles:

- Azkarate, M. (1991). 'Basque Lexicography'. In *Wörterbücher. Dictionaries. Dictionnaires*, Berlin/New York: Walter de Gruyter. pp. 2371-2375.
- Sagarna, A. (2010). 'The Lexicographic Work of Euskaltzaindia - The Basque Language Academy 1984-2009'. In *Proceedings of EURALEX 2010*, Leeuwarden, The Netherlands.
- VV.AA. (2017): *La Sociedad de la Información en España 2016*. Madrid: Fundación Telefónica. Retrieved from https://www.fundaciontelefonica.com/artes_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/558/ (1 October 2017)

Dictionaries:

- Collins Elhuyar English Basque Dictionary euskara-ingelesa hiztegia*. (2016). E. Etxeberria, E. Pociello and K. Ceberio (eds.). Usurbil: Elhuyar Foundation.

- Diccionario Retana de Autoridades del Euskera.* (1976-1989). M. de la Sota Aburto (ed.). Bilbo: La Editorial Vizcaína.
- Diccionario Trilingüe Castellano, Bascuence y Latin.* (1745). M. Larramendi. Donostia-San Sebastián [Vol 1 436p., Vol 2 392 p] [Facsimile reprint 1984].
- Diccionario Vasco-Castellano.* (1981). P. Mugica. Bilbo: Mensajero.
- Diccionario vasco-español-francés.* (1905-06). R.M. Azkue. Bilbo: La Gran Enciclopedia Vasca.
- Dictionnaire basque-français et français-basque (Dialectes labourdin, bas-navarrais et souletin).* (1926). P. Lhande. Paris: Gabriel Beauchesne.
- Elhuyar Dictionary.* (2007). E. Etxeberria (ed.). Usurbil: Elhuyar Foundation.
- Elhuyar Hiztegi Entziklopedikoa.* (1993). I. Irazabalbeitia (ed.). Usurbil: Elhuyar Foundation.
- Elhuyar hiztegia: euskara-gaztelania / castellano-vasco.* (2006). A. Gurrutxaga and E. Etxeberria. Usurbil: Elhuyar Foundation.
- Eskola Hiztegia.* (1995). P. Uribarren, R. Badiola, L. Baraiazarra, J.L. Goikoetxea and G. Aurrekoetxea. Bilbo: Adorez Group.
- Europa Hiztegia - Eskola berrirakoa.* (1993). P. Uribarren, R. Badiola, L. Baraiazarra, J.L. Goikoetxea and G. Aurrekoetxea. Bilbo: Adorez Group.
- Euskal Hiztegi Modernoa.* (1994). A. Gurrutxaga (Ed.). Usurbil: Elhuyar Foundation.
- Euskal Hiztegia.* (1996). I. Sarasola. Donostia: Gipuzkoako Kutxa.
- Euskaltzaindiaren Hiztegia.* (2012). Bilbo: Euskaltzaindia.
- Euskara Ikaslearen Hiztegia.* (1999). I. Sarasola. Barcelona: VOX.
- Euskararako hiztegia.* (1986). P. Uribarren, R. Badiola, L. Baraiazarra, J.L. Goikoetxea and G. Aurrekoetxea. Bilbo: Adorez Group.
- Hautalanerako Euskal Hiztegia.* (1984-1995). I. Sarasola. Donostia: Gipuzkoako Kutxa.
- Hiztegia 3000.* (2002). P. Uribarren, R. Badiola, L. Baraiazarra, J.L. Goikoetxea and G. Aurrekoetxea. Bilbo: Adorez Group.
- Hiztegia 5000.* (2009). P. Uribarren, R. Badiola, L. Baraiazarra, J.L. Goikoetxea and G. Aurrekoetxea. Bilbo: Adorez Group.
- Hiztegia 80.* (1980): X. Kintana, J. Aurre, R. Badiola, et al. [Ekiten Group], Bilbo: Elkar.

- Ikaslearen Hiztegia*. (1997). E. Etxeberria (ed.). Usurbil: Elhuyar Foundation.
- Labayru Hiztegia* (2003). Bilbo: Ibaizabal Argitaletxea.
- Le Dictionnaire Elhuyar*. (2004). E. Etxeberria (ed.). Usurbil: Elhuyar Foundation.
- LUR hiztegi entziklopedikoa*. (1995). J. Zabaleta (ed.). Bilbao: Lur Argitaletxea.
- Morris Student Plus Hiztegia Euskara / Ingelesa - English / Basque*. (1998). M. Morris. Zarautz: Morris Academy.
- Orotariko Euskal Hiztegia*. (1987-2005). K. Mitxelena. Bilbo: Euskaltzaindia.
- Zehazki gaztelania-euskara hiztegia*. (2005). I. Sarasola. Irun: Alberdania.
- Zientzia eta teknologiaren hiztegi entziklopedikoa*, (2009). A. Gurrutxaga (ed.). Usurbil: Elhuyar Foundation.

Websites:

- Dictionary Morris Student Plus Hiztegia*. Retrieved from http://www1.euskadi.net/morris/indice_e.htm
- Elhuyar Hiztegiak*. Retrieved from <https://hiztegiak.elhuyar.eus>
- Euskalterm*. Retrieved from <http://www.euskara.euskadi.eus/r59-euskalte/eu/q91Eus-TermWar/kontsultaJSP/q91aAction.do>
- Euskaltzaindiaren Hiztegia*. Retrieved from http://www.euskaltzaindia.eus/index.php?option=com_hiztegiabilatatu&view=frontpage&Itemid=410&lang=eu
- Hiztegi Batu Oinarriduna*. Retrieved from http://www.euskaltzaindia.eus/index.php?option=com_hbo&view=frontpage&layout=aurreeratua&Itemid=411&lang=en
- Hiztegia 5000*. Retrieved from <http://www.bostakbat.org/azkue>
- Labayru Hiztegia*. Retrieved from <https://hiztegia.labayru.eus>
- Nola Erran*. Retrieved from <http://www.nolaerran.org>
- Orotariko Euskal Hiztegia*. Retrieved from <http://www.euskaltzaindia.net/oeh>
- Sinonimoen Hiztegia (Adorez)*. Retrieved from <http://www.bostakbat.org/azkue>
- Sinonimoen Hiztegia (UZEI)*. Retrieved from <http://www.uzei.eus/zerbitzuak-eta-produktuak/produktuen-katalogoa/sinonimoen-hiztegia/>
- Zehazki Hiztegia*. Retrieved from <http://ehu.es/ehg/zehazki/>
- Zientzia eta Teknologiaren Hiztegia*. Retrieved from <http://zthiztegia.elhuyar.eus>
- CorpEus*. Retrieved from <http://corpeus.elhuyar.eus/cgi-bin/kontsulta.py>

K. CEBERIO, A. GURRUTXAGA: *State-of-the-art...*

EPG – Contemporary Reference Prose. Retrieved from <http://www.ehu.es/euskara-orria/euskara/ereduzkoa/>

ItzulTerm. Retrieved from <http://itzulterm.elhuyar.eus/>

LBC –Lexicon Observatory. Retrieved from <http://lexikoarenbehatokia.euskaltzaindia.net/>

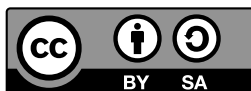
Statistical corpus of 20th century Basque. Retrieved from <http://xxmendea.euskaltzaindia.eus/Corpus/>

ZTC – Science and Technology Corpus. Retrieved from <http://www.ztcorpusa.eus>

STANJE ENOJEZIČNE LEKSIKOGRAFIJE: BASKIJA

V prispevku podajamo pregled razvoja baskovske leksikografije in izpostavimo glavne dosežke, spodrsaljaje in prihodnje izzive. Baskovska leksikografija ima sorazmerno kratko zgodovino, vendar pa je v zadnjega pol stoletja, odkar se je začel proces standardizacije, nastalo precejšnje število jezikovnih virov. Po letih leksikografskega dela različnih skupin in založnikov je nastal izjemni akademski preskriptivni slovar (*Euskaltzaindiaren Hiztegia*), ki je izšel nedavno in temelji na historičnih in sodobnih korpusih. Čeprav število enojezičnih del v zadnjih letih občutno narašča, pa je bila baskovska leksikografija še posebno produktivna pri izdelavi dvojezičnih slovarjev, kar gre najbrž pripisati socio-lingvističnemu statusu jezika. Po drugi strani pa sta bili področji specializirane leksikografije in terminologije izjemno aktivni že od samih začetkov procesa standardizacije. Od začetka 21. stoletja sicer močno narašča raba korpusov, številni slovarji baskovščine pa so prosto dostopni na spletu.

Ključne besede: baskovščina, slovar, korpus



To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna. / This work is licensed under the Creative Commons Attribution-Share-Alike 4.0 International.

<https://creativecommons.org/licenses/by-sa/4.0/>