

# SPAIN: LEXICOGRAPHY IN IBERIAN LANGUAGES

Roser Saurí Colomer

Lab for Linguistics and Computation

Computer Science Department

Brandeis University

415 South Street, Waltham, MA 02454, USA

roses@cs.brandeis.edu

## Introduction

Spain has seven main autochthonous languages, currently recognized in different degrees: official (Spanish), co-official with Spanish in some of the areas where they are spoken (Catalan, Basque, Galician), locally recognized as co-official (Aranese), and not officially recognized (Asturian, Aragonese). Starting in medieval times, historical events consolidated Spanish as the language of culture in most of the regions forming present-day Spain. That fact gave rise to a diglossic situation in which Spanish was assumed as the high variety (used in official and written domains), while the other languages were relegated to low and spoken registers. Hence, whereas lexicography in Spanish shows a development very similar to other socially healthy Romance languages, the linguistic diglossia influenced the cultural development of the other languages mentioned above, and consequently also their lexicographic tradition. Except for Catalan, which managed to counteract the diglossic effect to a greater extent than the other languages, the lexicographic production of most of the minority languages in present-day Spain have been hindered for a long time by the social and political dominance of Spanish.

## Spanish

The earliest works on the Spanish lexicon were medieval glosses, such as the *Glosas Silenses* (circa 10th century), and Mozarabic and Latin-Spanish alphabetical glossaries, which date between the 12th and 15th centuries. The Renaissance was characterized by the proliferation of Latin-Spanish dictionaries (which laid the foundations for later monolingual dictionaries), bilingual dictionaries between Spanish and other modern languages, and specialized vocabularies in areas like medicine or agriculture. A noteworthy work of that period is *Universal vocabulario en latín y en romance* by Alfonso de Palencia, published in 1490 and considered a link between medieval and modern lexicography. The most remarkable works are, however, the two dictionaries authored by Elio Antonio de Nebrija, which initiated a new lexicographic tradition. They are entitled *Diccionario latino-español* and *Vocabulario español-latino*, and were compiled in 1492 and 1495, respectively.

The earliest monolingual general dictionaries were compiled in the 17th century, being *Tesoro de la lengua castellana* (1611), by Sebastián de Covarrubias, the most prominent. Beginning in the 18<sup>th</sup> century there was a general perception among intellectuals that Spanish was declining as a language of culture (the so-called Golden Century in Spanish literature had been two centuries earlier), and this feeling led to the creation of the Spanish Royal Academy (*Real Academia Española*, RAE) in 1714. RAE compiled two outstanding dictionaries. Between 1726 and 1739, it published *Diccionario de Autoridades*, remarkable in its time for its extensive use of literary and scientific sources and its coverage of both medieval and contemporary language. Forty years later, in 1780, RAE also published *Diccionario de la lengua castellana (DRAE)*, a dictionary that had been initially conceived as the short version of *Autoridades*, but that ended up

being used as the basis for several monolingual Spanish dictionaries. DRAE identified Castilian Spanish as the standard for Spanish until its 15<sup>th</sup> edition (1925), when regionalisms from other Spanish-speaking areas in Spain and America were also introduced.

Other dictionaries of that period are also worth noting, such as *Diccionario castellano con las voces de ciencias y artes*, compiled by Esteban de Terreros between 1786 and 1793, a dictionary that has been frequently neglected but which is of indisputable lexicographic value. Already in the 19<sup>th</sup> century, *Nuevo diccionario de la lengua castellana*, compiled in 1846 by Vicente Salvá, and *Diccionario nacional o gran diccionario clásico de la lengua española*, published by Ramón Joaquín Domínguez in 1847, are considered leading exponents of the best lexicography developed during that time.

The 20<sup>th</sup> century was a period of large and exceptional lexicographic productions. Among them we can mention the work entitled *Diccionario de construcción y régimen de la lengua castellana*, an 8-volume dictionary devoted to Spanish syntactic constructions, authored by Rufino José Cuervo and published throughout a period of more than 100 years (between 1886 and 1988). Equally singular are the two multi-volume etymological dictionaries *Diccionario crítico etimológico de la lengua castellana*, published by Joan Coromines between 1954 and 1957, and *Diccionario crítico etimológico castellano e hispánico*, compiled between the years 1980 and 1991 by Joan Coromines and José A. Pascual. Worth mentioning is also Samuel Gili Gaya *Tesoro lexicográfico*, appeared in 1947, which is a thesaurus of all Spanish dictionaries that had been published between 1492 and 1726. In addition to these remarkable works, Spanish lexicography in the 20<sup>th</sup> century also experienced a proliferation of general-user

dictionaries, the most worth noting of which are the appraised work of María Moliner *Diccionario de uso del español*, published in 1966, and the dictionary by Manuel Seco and his colleagues entitled *Diccionario del español actual*, which appeared in 1999. Function-specific dictionaries (synonyms, inverse, terminology or neology), and bi- or multi-lingual dictionaries were also compiled at that period. A number of dictionaries were published in electronic format, like DRAE 1992 edition or Moliner. Moreover, RAE started working on corpus linguistics-based projects for lexicography, and now its two corpora of historic and contemporary Spanish (CORDE and CREA) are available online, together with a database system (called BusCon) for dictionary look-up.

## **Catalan**

As in Spanish, medieval lexicography in Catalan is represented by glosses and Latin- or Hebrew-Catalan alphabetical glossaries (14th-15th centuries). The Catalan Renaissance also parallels the Spanish one in that it is characterized by the compilation of specialized glossaries, modern language bilingual vocabularies, and Latin-Catalan dictionaries. Within this period we find Joan Esteve *Liber elegantiarum*, appeared in 1489, which is the first Catalan vocabulary with lexicographic intention, and Gabriel Bosa *Diccionari llatí-català i català-llatí*, 1507, based on the lexicographic work in Spanish by Nebrija. Onofre Pou dictionary, entitled *Thesaurus puerilis* and published in 1575, is also worth mentioning of that period given its influence in subsequent production.

At the end of the 15th century, political developments brought Catalan into an acute diglossic situation that led to the so-called Decadence. Lexicographic work in Catalan during this period is represented by only three Latin-Catalan dictionaries, all of them compiled in the 17th century and which remained influential throughout the

subsequent century: *Fons verborum et phrasium*, published in 1637 by Antoni Font, *Thesaurus verborum ac phrasium*, 1640, by Pere Torra, and most remarkably *Gazophylacium catalano-latinum*, 1696, by Joan Lacavalleria. The Decadence lasted until the 18th century, when Catalan was banned from legal and educational spheres. The perception of decline resulted in dictionaries that aimed at preserving the language. At that same time, the claims of a Port Royal-inspired movement, led by Baldiri Rexach, in favor of using the student's own tongue for teaching, motivated a number of Catalan-Spanish dictionaries. Under the Enlightenment influence, there were attempts to emulate RAE work in several unsuccessful projects, but no monolingual dictionary of Catalan was ever published during that period.

The 19th century is known as the *Renaixença* (literally 'rebirth'). There was an enthusiasm for reviving Catalan as language of culture, which reflected in the proliferation of dictionaries, both dialectal and general. Note worthy works of that period are the 2-volume *Diccionario catalán-castellano-latino*, a collective work published between 1803 and 1805 by Esteve, Bellvitges, and Juglà, and *Diccionari de la llengua catalana*, 1939-1940, by Pere Labèrnia, which is considered the *Renaixença* dictionary. Both of them are trilingual Catalan-Spanish-Latin. Further, the positivist tendencies of those times led to compiling comprehensive dictionaries under historic-comparative methods. Good exponents of that are Marià Aguiló and Josep Balari posthumous published compilations. Such interest extended into the initial decades of the 20th century, and culminated with the impressive work of Josep Maria Alcover and Francesc de Borja Moll: a 10-volume dictionary entitled *Diccionari Català-Valencià-Balear* that was published between 1926 and 1962. *Alcover-Moll*, as it is popularly called, is a unique

work among all Romance languages. It covers all Catalan dialects from old and modern language, and provides etymological, patronymic and toponymic information.

A major concern among Catalan intellectuals at the beginning of the 20<sup>th</sup> century was to establish a prescriptive standard for Catalan that would help consolidate it as a modern language. This resulted in Pompeu Fabra *Diccionari general de la llengua catalana*, 1932, a dictionary assumed as normative by the Institute for Catalan Studies (*Institut d'Estudis Catalans*, IEC). Other major concerns in 20th-century Catalan lexicography were avoiding Spanish interference, communicating with other languages, and normalizing Catalan in all spheres, including technical and scientific fields. After Franco death, these led to a profusion of monolingual, bilingual, terminology, and neologism dictionaries, many of which are available in electronic format.

The main monolingual onomasiological dictionaries nowadays are *Enciclopèdia Catalana Gran diccionari de la llengua catalana*, 1998, and *Diccionari de la llengua catalana*, 1995, the normative dictionary compiled by IEC upon Fabra work. More function-specific products also exist (synonym, abbreviation, pronunciation, frequency, and inverse dictionaries), as well as unique works, like the 9-volume etymological dictionary entitled *Diccionari etimològic i complementari de la llengua catalana*, published between 1980 and 1991, or the 6-volume *Onomasticon Cataloniae*, appeared between 1989 and 1997, both of them authored by Joan Coromines. IEC is currently sponsoring the ambitious project of a corpus-based, comprehensive, and descriptive dictionary of contemporary Catalan.

## **Galician**

Galician early diglossia placed its culture in a weak situation already in Medieval Ages. Until the 18th century there was only one Galician vocabulary, written circa 1536

by Bachiller Olea. The 18th century featured the work of three unpublished lexicographers: Martín Sarmiento, who reached a level unsurpassed until the 20th century, Juan Sobreira, who conceived a dictionary based on DRAE, and Xosé Andrés Cornide. The first published dictionaries appeared in the second half of the 19th century, during a movement favorable to Galician regeneration. They were *Diccionario gallego-castellano*, published by Francisco Javier Rodríguez in 1863, *Diccionario gallego*, authored by Juan Cuveiro in 1876, and *Diccionario gallego-castellano*, published in 1884 by Marcial Valladares. These three dictionaries, which still took Spanish as a point of departure, were of inferior quality than the unpublished lexicographic work of the previous century.

Lexicographic production in the 20th century experienced a rebirth in the post-dictatorship 1980s. The first dictionaries completely in Galician appeared then. In 1982, Galician attained co-officiality status, and orthographic standards were created by the Royal Galician Academy (*Real Academia Galega*, RAG) and the Galician Language Institute (*Instituto da Lingua Galega*, ILG). In 1986, two remarkable general dictionaries were published by private companies (Estravis and Xerais). The first normative dictionary, *Diccionario da lingua galega*, was compiled by RAG and ILG in 1990. More specialized dictionaries (learner, bilingual, synonyms) appeared in the 1990s, together with terminological vocabularies. However, Galician lexicography of that period reflected the social controversy regarding the official orthographic norm, accused of being based too much on Spanish, instead of on a Portuguese or a more genuine historic-etymological view.

Currently, Galician lexicography is benefiting from corpus projects like CORGA and the Corpus of Modern Galician, which resulted in the *Tesouro Informatizado da Lingua Galega*, available on-line.

## **Basque**

Basque lexicography has been affected by the fragmentation between Eastern and Western dialects (spoken in French and Spanish domains, respectively), and by the diglossic situation of the language, which caused that Basque was not a written language, and therefore not used in official and scientific domains, until the 16<sup>th</sup> century. These facts made codifying Basque a controversial task, as reflected in its lexicographic tradition. The first dictionary was an unpublished Basque-French work by Silvain Pouvreau, dating from the 17th century, which covered only Eastern dialects. A century later, in 1745, Manuel Larramendi published *Diccionario trilingüe del castellano, bascuence y latín*, a Castilian-Basque-Latin dictionary which including all dialects, and that introduced several neologisms to fill onomasiological gaps in technical and scientific domains. Worth mentioning in the 19th century are José Francisco Aizkibel Basque-Spanish *Diccionario Basco Español titulado Euskeratik Erderara biurtzeko itz-tegia*, appeared in 1883, and the unpublished Basque-French dictionary by Maurice Harriet entitled *Dictionnaire Basque-Français*. Starting the 20<sup>th</sup> century, between 1905 and 1906, the 2-volume *Diccionario Vasco-Español-Francés* by Resurrección María Azkue was published, a substantial trilingual dictionary containing oral, written, dialectal and historical data, and which also introduced some new terminology.

A unified orthography for Basque was approved by the Basque Academy (*Euskaltzaindia*) at the end of the Franco era, in 1968, and the first monolingual dictionaries for Basque appeared in the 1980s. They are *Gaurko Euskara Idatziaren*

*Maiztasun Hiztegia*, a dictionary of frequencies edited by Ibon Sarasola in 1982, and *Hauta-lanerako Euskal Hiztegia*, also by Sarasola, a dictionary based on the literary tradition from 1750 to 1964 and published between 1984 and 1995. *Euskaltzaindia* has recently published the 8-volume *Diccionario general vasco. Orotariko euskal hiztegia*, 1987-2004, a descriptive and general dictionary of Basque, compiled from a diachronic corpus and available on CD-Rom, and it is currently working on a normative Unified Lexicon (*Hiztegi Batua*). Work on terminology and neologisms is also being carried out mainly by UZEI (Basque Centre for Terminology and Lexicography), its most important product being *Euskalterm*, a publicly available terminological database. Other more function- and audience-specific dictionaries are also produced, mostly by private companies.

### **Asturian**

The first known attempts to compile an Asturian dictionary were made in the 18<sup>th</sup> century by Gaspar Melchor de Xovellanos, Francisco Caveda y Solares, and Carlos González de Posada. Their work, mostly now lost, remained unfinished and unpublished. Language codification was also an important concern for 19<sup>th</sup>-century intellectuals in Asturias. Several of them, such as Juan Junquera Huergo and Julio Somoza, developed lexicographic work aiming at a standard dictionary, but this too remained unpublished. Only dialectal compilations were published in that century, alongside with the first Asturian-Galician dictionary, *Vocabulario de las palabras y frases bable*, by Apolinar Rato y Hevia de Argüelles (1891), which has been widely criticized for its inaccuracy.

The first half of the 20<sup>th</sup> century featured other unpublished dictionaries (by Pepín de Pría and Fernán Coronas, among others) and several dialectal glossaries, some of them of remarkable quality, such as *Vocabulario del bable de occidente*, by Bernardo

Acevedo y Huelves and Marcelino Fernández y Fernández, or *El bable de Cabranes*, by María Josefa Canellada. After the dictatorship, the production of dialectal glossaries increased and a considerable number of Asturian-Spanish dictionaries were published. In 1980, the Asturian Language Academy (*Academia de la Llingua Asturiana*, ALIA) was founded, and in 1981 it established the Asturian orthographic norms. Also significant is the fact that some publishing companies, such as Trabe, started developing work on Asturian lexicography in a consistent way. In 2000, ALIA published the first normative dictionary, *Diccionariu de la Llingua Asturiana*, a general dictionary that includes new terminology and neologisms. A large-scope etymological dictionary is currently being developed by Xosé Luis García Arias.

## **Aragonese**

*Glosas Emilianenes* (11<sup>th</sup> century), which has been considered written in Spanish for long but is in fact in Aragonese, constitutes the first lexicographic work (and written text) of a Romance language in the Iberian Peninsula. Despite such an early production, the declining of Aragonese power in the 15th century caused the cultural recession of that language and resulted, as a consequence, in an extremely sparse lexicographic production until the 19th century. Gerónimo de Blancas vocabulary, entitled *Índice donde se declaran algunos vocablos Aragoneses* and appeared in 1641, is the only work worthwhile mentioning for that period.

In the 19th century, two Aragonese-Castilian dictionaries were published: *Ensayo de un diccionario aragonés-castellano*, compiled in 1836 by Mariano Peralta, and *Diccionario de voces aragonesas*, 1859, by Jerónimo Borao. Other bilingual or dialectal vocabularies attest a renewed interest in Aragonese at the beginning of the 20th century. However, during Franco dictatorship, Aragonese lexicography was developed mainly in

the scope of academic research. Such is the case of Pedro Arnal vocabulary entitled *Vocabulario del alto-aragonés* and published in 1944, or Antoni Maria Badia i Margarit works, *Contribución al vocabulario aragonés moderno*, 1948, and *El habla del Valle de Bielsa*, 1950.

After Franco death, there was a gradual increase of lexicographic production. In 1987, Aragonese orthographic norms were approved. There were a number of dialectal vocabularies and terminological lexicons, either published independently or as part of wider works such as *Gran Enciclopedia Aragonesa* (Great Aragonese Encyclopedia). Several Aragonese-Castilian dictionaries appeared at that time, including *Diccionario aragonés. Aragonés-castellano y castellano-aragonés* published in 1992 by Rafael Andolz Canela, or the *Vocabulario básico bilingüe aragonés-castellano y castellano-aragonés*, 1997, by Antonio Martínez Ruiz. Nevertheless, no general monolingual dictionary exists yet. The project *Tresoro d'a Luenga Aragonesa*, currently ongoing, aims at creating a public domain database containing all Aragonese words published between the 17<sup>th</sup> and 20<sup>th</sup> centuries. Its initial product is *Endize de bocables*, an index of 90,000 inflected forms, published in 1999 by the Institute for High Aragon Studies (*Instituto de Estudios Altoaragoneses*), available also in CD-Rom.

## **Aranese**

Lexicographic work in Aranese started late and is scarce. As a matter of fact, Aranese was not a written language until the end of the 19<sup>th</sup> century. At that time, some Occitan and Gascon dictionaries had been compiled already, but none on the Aranese dialect. The first works on Aranese lexicon appeared in the 20<sup>th</sup> century: Jusèp Condò Aranese vocabulary, published in 1914 in *Butlletí de Dialectologia Catalana* (Bulletin of Catalan Dialectology), and Joan Coromines dissertation entitled *Vocabulario aranés*,

published in 1931. Aranese orthography was created in 1983 based on the norms of the Institute for Occitan Studies (*Institut d'Estudis Occitans*). In the 1990s, the *Conselh Generau d'Aran* (Aranese Government) published several dictionaries of different nature, including the Aranese-Spanish-Catalan-French general dictionary by Frederic Vergés, entitled *Petit diccionari*, and other works of terminological nature, most of which are at least trilingual Aranese-Catalan-Spanish given Aran Valley political attribution.

## **Bibliography**

- Alvar Ezquerro, Manuel (1995) 'Los diccionarios del español en su historia'. *International Journal of Lexicography* 8: 173-201.
- Arias Cabal, Á. (1996) 'La lexicografía asturiana. Cronología de doscientos años d'intentos'. *Lletres Asturianas* 60: 41-63.
- Azkarate, M. (1991) 'Basque Lexicography'. In Hausmann et al. (eds.): 2371-2375.
- Bajo Pérez, E. (2000) *Los diccionarios. Introducción a la lexicografía del español*. Oviedo: Trea.
- Haensch, G. (1991) 'Katalanische Lexikographie'. In Hausmann et al. (eds.): 1770-1788.
- Haensch, G. (1991) 'Spanische Lexikographie'. In Hausmann et al. (eds.): 1738-1767.
- Haensch, G. (1997) *Los diccionarios del español en el umbral del siglo XXI*. Salamanca, Universidad de Salamanca.
- Hausmann, F.J., Reichmann, O., Wiegand, H.E. & Zgusta, L. (eds.) *Wörterbücher. Ein Internationales Handbuch zur Lexikographie* (3 vols.). Berlin, New York: Walter de Gruyter.

- Instituto de Estudios Altoaragoneses (1998) *Fuens lexicograficas de l'aragonés. Catalogo de repertorios lexicograficos aragoneses dende o sieglo XVII dica 1998*. Uesca: IEA.
- Pensado, J.L. (1982) 'La lexicología gallega en el siglo XVIII'. In Kremer, D. & Lorenzo, R. (eds.) *Tradición, actualidade e futuro do galego*. Santiago de Compostela: Universidade de Santiago de Compostela, Xerais: 85-98.
- Pensado, J.L. (1988) 'La lexicografía gallega decimonónica'. In Lorenzo, R. (ed.), *Coloquio de Lexicografía*. Santiago de Compostela: Universidad de Santiago de Compostela, Xunta de Galicia: 49-56.
- Pensado, J.L. (1991) 'Galician Lexicographie'. In Hausmann et al. (eds.): 1736-1738.
- Perea, M.P. (1998) 'De re lexicogràfica'. *Els Marges* 60, 96-120.
- Perea, M.P. (2000) 'Compleció d'un buit lexicografic: la lexicografia catalana (1989-1994)'. *Els Marges* 66, 77-107.
- Soberanas, A.J. & Colon, G. (1986) *Panorama de la lexicografia catalana. De les glosses medievals a Pompeu Fabra*. Barcelona: Enciclopèdia Catalana.

## **Abstract**

Spain has seven main autochthonous languages: Aragonese, Aranese (a dialect of Gascon), Asturian, Basque, Catalan, Galician, and Spanish. For all of them it currently exists a body of lexicographic work, although they present different levels of achievement. Starting in medieval times, the empowerment of Spanish as the language of culture in most of present-day Spain led the other languages to a diglossic situation that strongly affected their cultural development. Among other areas, this is reflected in their lexicographic tradition, which had a less active production than Spanish and other socially healthy languages.

**Keywords:** Aragonese, Aranese, Asturian, Basque, Catalan, Galician, Gascon, Spanish, lexicography, diglossy.