

The Challenges of Digitizing Spanish Grammars for Italian Speakers Across Centuries

Silvia Muzzupappa^{1,*}, Giorgio Maria Di Nunzio^{2,†}

¹Department of Literary and Linguistic Studies, University of Padova, Via E. Vendramini, 13, 35137 Padova, Italy

²Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy

Abstract

In this paper, we will discuss the design of a database for the diachronic study of the production of grammars (*grammarography*) for the teaching of Spanish to an Italian-speaking audience between the sixteenth and nineteenth centuries. We will focus on the grammarography of the Spanish language as second language (the so-called *Gramáticas de Español para Italianos*, henceforth GEI). The aim of this database is to provide a common ground for the researchers in this field to establish the definition of the grammaticographical textual genre of GEI. The database will allow a more in-depth analysis of each work, facilitating through a graphical interface the recognition of their common and uncommon features.

Keywords

Grammaticography, Spanish grammar, Italian grammar, Linguistic database

1. Introduction

In this paper, we want to discuss the design and implementation of a digital resource for the study of Spanish grammar for an Italian-speaking audience published in the sixteenth and nineteenth centuries (the so-called *Gramáticas de Español para Italianos*, henceforth GEIs). This study falls within the research area named 'grammarography' (or grammaticography [1], as the art and craft of writing grammars). Our proposal addresses the parallelism and divergences with respect to both Spanish and Italian grammar tradition. Like all second language grammars, GEIs make descriptive and theoretical choices in order to adapt, privilege, or omit some topics with respect to first language grammars. However, the history of Spanish [2] [3] [4] [5] [6] or Italian grammars [7] [8] [9] has been focused on a single language so far and not on both in comparison. So, the analysis of texts teaching foreign languages has been subordinated to a secondary role. The only interest shown so far is on their derivation with respect to the grammatical tradition of each language individually considered.

Instead, our project aims to research both the historical shape of second language grammars and the choices made by grammarians. At the moment there are few studies on the history of

19th IRCDL (*The Conference on Information and Research science Connecting to Digital and Library science*), February 23–24, 2023, Bari, Italy

*Corresponding author.

†These authors contributed equally.

✉ silvia.muzzupappa@phd.unipd.it (S. Muzzupappa); dinunzio@dei.unipd.it (G. M. D. Nunzio)

🌐 <https://www.dei.unipd.it/~dinunzio/> (G. M. D. Nunzio)

🆔 0000-0001-7116-9338 (G. M. D. Nunzio)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

GEIs, their features and the relationship with Spanish and Italian grammatical tradition [10] [11] [12] [13] [14]. This project is part of the line of research of *Epigrama* project, developed by the University of Padua [15]. It studies a corpus of critical and digital editions of some GEIs and intends to enrich it through computer technologies both for the codification of texts and quantitative and qualitative data analysis.

For this reason, we are designing a database to provide general as well as specific information on the external structure of the works and quantitative and qualitative analysis of the extracted data.

The objectives of this proposal are:

- the description and analysis of the grammarisation process of the parts of speech in the GEI corpus to have a systematic framework of the classification methods of the typical elements of this genre, such as definitions, terminology, examples, etc.;
- the identification of the distinctive features of a didactic grammar as an autonomous grammaticographical textual genre in terms of external (the layout) and internal structure (the content) of each text;
- the design and implementation of a database containing the mentioned features to define the genre of this kind of grammar by quantitative and qualitative data analysis.

The paper is organized as follows:

- Section 1: Introduction, where we introduce the situation at the moment;
- Section 2: The GEI corpus, where we introduce the documents collected so far and we analyze their structure and content;
- Section 3: Design of the database, where we imagine how the database should be designed;
- Section 4: Conclusion and future work, where we describe our expectations.

2. The GEI corpus

In 2016, a corpus of relevant texts has begun to be built within *Epigrama*¹ project by the University of Padua. To quote the official website of the project "the corpus includes grammars or grammatical treatises, introductions to translations and lexicographic works, as well as any type of text in which metalinguistic observations useful for teaching Spanish to Italians are found". For our project, the corpus published on *Epigrama* website has been extended to nine grammars, for a total of nine authors, which in detail are:

1. Mario Alessandri d'Urbino, *Il paragone della lingua toscana et castigliana*, Napoli, Mattia Cancer, 1560 [16] [17];
2. Mattia Chirchmair, *Gramatica spagnuola e italiana*, Firenze, Giuseppe Manni, 1709 [18];
3. Antonio Fabro, *Grammatica per imparare le lingue italiana, francese e spagnola*. Venezia, Gio. Dini, 1627 [19];
4. Lorenzo Franciosini, *Gramatica spagnola e italiana*, Venezia, Sarzina, 1624 [20] [21];

¹<http://www.epigrama.eu/>

5. José Martínez de Valdepeñas, Grammatica della lingua spagnola. Genova, Stamperia di Giovanni Franchelli, 1786;
6. Giovanni Miranda, Osservazioni della lingua castigliana, Vinegia, Gabriel Giolito de Ferrari, 1566 [22] [23];
7. Josef Faustino Perles y Campos, Gramatica Española, o' modo de entender, leier, y escrivir Español, Napoli, Parrino & Mutii, 1689;
8. Gennaro Sisti, Traduzione dal francese del nuovo metodo di Porto Reale. Con cui agevolmente s'insegna la lingua spagnola, Napoli, Serafino Porsile Regio Stampatore, 1742 [24];
9. Massimo Troiano, Compendio del signor Massimo Troiano con le annotazioni del signor Argisto Giuffredi, Firenze, 1601 [25].

This selection of documents was carried out according to the following criteria:

- their first edition is always between the sixteenth and nineteenth centuries;
- Italian is used as a metalanguage and presents the equivalents of categories and linguistic forms for comparative or contrastive analysis;
- they are bilingual or multilingual texts;
- their content is written both in Spanish and Italian (in one case, in French too).

As mentioned above, in order to identify the distinctive features of GEI as a grammaticographical textual genre, we decided to work on two parallel fields. These fields cannot be separated from each other: the macro-structure (which identifies the layout of the grammar) and the micro-structure (which identifies the content of the grammar).

2.1. Macro-Structure: The Layout

The macro-structure is the identification of the topics covered in each grammar and their subdivision into chapters, sections and paragraphs. More simply, it concerns the document layout that can be seen by the reader. An initial starting point was taken from San Vicente and Lombardini's work [14], appropriately revisited in order to build the database (still in the design phase) to collect, catalog and process information and data for the definition of the grammatical textual genre of GEIs. One of the most difficult challenges faced so far has been the definition of the structure of the elements composing a grammar and, consequently, the identification of the related entities and properties of the database. We have started to identify the main common themes for most of the grammars, which are:

- Parts of speech of both languages (we can find them in all grammars);
- Dialogues;
- Lexical tables;
- Proverbs;
- Letter patterns;
- Pronunciation and spelling rules.

These themes for all grammars coincide with the nine parts of speech (hereinafter PoS) of traditional grammar, which are noun, adjective, article, pronoun, verb, adverb, preposition, conjunction and interjection.

A first result of this cross-sectional analysis shows that, although the parts of speech constitute the main topics shared by all the texts, these are presented by each author with different denominations and classifications. Moreover, the order and specificity with which the same topics are dealt with vary from one grammar to another. In addition, although the books are divided into chapters, sections and paragraphs, sometimes there are no titles associated with them, while in other cases the authors presented generic titles for different topics. It was the reader's task to understand to which part of speech the author was referring to. So, to make them identifiable in our database, we need to do the same without altering the original structure of the text.

A second important point is dialogues. They are usually presented in a graphic form divided into two columns: the first in Spanish (or the original language), the second contains the Italian translation. We assume that these dialogues already exist in other works aimed to teach Spanish language to foreigners, adapted and translated into a didactic perspective for an Italian-speaking audience.

Regarding the lexicon, it can be noted that some books present to the reader lists of Spanish words accompanied by the Italian translation, grouped together by specific areas. For example, there are bilingual glossaries on food, animals, degrees of kinship, materials, parts of the house, etc.

Similarly proverbs and letters are often presented in the texts with Italian translation in order to provide the reader with expressions suitable for a specific written or oral register, formal or informal, within native Spanish speakers context.

It can be deduced that, while the pure treatment of the parts of speech may fall within a theoretical knowledge, these last three elements would rather be linked to a practical use of Spanish as a foreign language.

Finally, the glosses deserve a separate study. They are annotations in a smaller editorial format than the body of the text, placed at the side of it, normally at the beginning of the paragraph, which provides an indicative title or summarize the theme explained in the paragraph they are referring to. In the definition of the macro-structure, we decided to consider glosses as part of the chapter (or section) they accompany and not to transcribe them in a separate parallel document, since it would have been more difficult to maintain the structural relationship between the chapter and the gloss.

In Table 1, we show a summary of the macro-structure elements for each grammar in the corpus.

2.2. Micro-Structure: The content

The micro-structure includes the content of the grammars, which we can synthesize in four areas (when present):

- The definitions and the terms with which the PoS and their related topics are presented by the authors;

Table 1

Summary of macro-structure elements in the corpus.

Author	Parts of speech	Dialogs	Lexicon	Proverbs	Letters	Glosses
Mario Alessandri d'Urbino (1560)	x					
Mattia Chirchmair (1709)	x	x	x	x		
Antonio Fabro (1627)	x	x	x	x		
Lorenzo Franciosini (1624)	x					x
José Martínez de Valdepeñas (1786)	x	x			x	
Giovanni Miranda (1566)	x					x
Josef Faustino Perles y Campos (1689)	x		x	x	x	
Gennaro Sisti (1742)	x	x	x			
Massimo Troiano (1601)	x				x	x

- Paradigms of verbs or declensions of nouns, adjectives, pronouns and articles;
- The examples, with reference to the statements, the length, the linguistic register and, above all, the language in which they were written to show the reader practical use cases;
- The terminology used by each author to identify grammatical objects and concepts.

We need to identify and catalog these four areas in each text to compare them with the nomenclature of today's grammars, established by Enciclopedia Treccani ² for Italian and by Real Academia Española for Spanish, to have a clear overview of the diachronic evolution and development of the GEI. This will allow us to finally define the grammaticographical textual genre as an autonomous genre, not subordinated to first language grammars. Starting from the four criteria listed above, the content of each grammar can be summarized as follows: in all books we find definitions, examples and paradigms (both declensions and conjugations), except, in the last case, in Sisti's work.

With regard to the examples, an important distinction can be made: as mentioned in the paragraph concerning the micro-structure, each section that deals with the Spanish language from a practical point of view can be a valid example for the reader in order to speak Spanish fluently. This is the reason why, Spanish dialogues accompanied by Italian translations are a useful tool to identify register, context, terms and lexical collocations within a broad, organic and elaborated speech. This is true for the oral plan of the Spanish language too. A similar consideration can also be made on the written plan by reading the models of letters proposed by Perles y Campos and Martínez de Valdepeñas. Finally, we cannot exclude short essential sentences proposed by the author to complete his theoretical explanation in the considered chapter.

2.3. The challenges of each book

2.3.1. Massimo Troiano's *Compendio* (1601): more unique than rare

Massimo Troiano's *Compendio* [25] is the only text among those examined presented entirely in the form of two dialogues, without any real division into chapters based on the parts of the speech or on any other logical classification. In this case, the glosses are a fundamental help

²https://www.treccani.it/enciclopedia/elenco-opere/La_grammatica_italiana

for the reader, which helps "splitting" the dialogue based on the topics covered in each "split". Another peculiarity is that dialogues written by Massimo Troiano alternate with annotations introduced by Argisto Giuffredi. In some cases, these annotations are real corrections of the concepts expressed in the dialogue; in other cases, they include Giuffredi's opinions or reasoning on the subject; sometimes they offer a useful hint to the reader to improve its level of knowledge of both the Italian and Spanish languages.

Furthermore, concerning the micro-structure, it is important to emphasize that in the Massimo Troiano's *Compendio* paradigms are included within the part of dialogue affected by the theme and for this reason they are presented as a simple statement. In addition, it is not easy to identify them quickly by scrolling through the pages of the text if not, once again, with the help of the glosses that act as "bookmarks". This text is interesting because it is placed in a peripheral area of the prototype of grammatical textual genre compared to the other books of the corpus, on both macro-structure and micro-structure.

2.3.2. Sisti's *Traduzione* (1742): a grammar dealing with poetry too

In terms of macro-structure, Sisti includes in his *Traduzione* [24] an entire chapter dedicated to poetry in which, he introduces the concept of rhyme, verse and other concepts related to poetic production. He also proposes different types of compositions of Spanish tradition, some of which are composed following the Italian tradition.

In the micro-structure, however, we can note that paradigms (both declensions and conjugations) do not appear in his grammar. Conversely, grammatical rules are illustrated in incomplete lists without all the entries or by reporting only the endings.

2.3.3. Fabro's *Grammatica* (1627): a perfect trilingual balance

It should be noted that Fabro's *Grammatica* [19] is the only trilingual grammar (Italian, Spanish and French) of the corpus, without the prevalence of one language over the others. This means that every content written in one language has a symmetrical and specular image with respect to the contents of other languages. Even in terms of the graphic layout, the contents of the entire book maintain this balanced distinction with the use of three separate columns: one for French, one for Italian and one for Spanish.

Furthermore, the lists of linguistic equivalences (Spanish-Italian) introduced within the dialogues, merging in a single chapter two areas (lexicon and dialogues), that are typically separated in GEI (at least in relation to what it has emerged so far).

3. Design of the database

The database is being designed to manage the large amount of data that will be extracted from the mentioned books, preserving the existing relations, common and uncommon elements. The database will ensure that the standards of accessibility, sharing and reliability will be respected to assure that this work can be extended or used in different contexts.

The database intends to standardize the common features that we normally look for in the corpus of grammars built for this task, both at structural and content level. This will allow

us to preserve the fidelity of textual data contained in the books, while defining a structural order that will better illustrate the evolution of the texts and the textual genre of the GEIs. Therefore, the database has specific qualitative features (that match the definition of macro and micro-structure) which can be extended, linked and related, once again, to multiple contexts and choices (for example, other bilingual teaching grammar of different languages, not only Spanish). We think that a correct implementation of such a database will make the definition of GEIs easier and more analytical, but it will also allow us to study new textual genres or link defined ones with deeper and more standardized relations, opening space for new lines of research.

Of course, the first implementation will be the most important one and therefore we are still tackling some very critical aspects on both design and technical point of view. At the time of this writing no demo is still available.

3.1. Additional Information

Some extra-textual data go beyond the strict reproduction both from the graphic point of view (for example, the digital editions of the works) and from the point of view of the information of a single book. In our case these additional data can be summarized as:

- Name of the author (s);
- Grammar title;
- Year of publication;
- Edition number (if there are more);
- Name of the publisher (s);
- Metalanguage, the language used to describe the language of the teaching grammar;
- Object language, the language described by the metalanguage.

As stated in the paragraph about micro-structure, we are referring to today's nomenclature of Enciclopedia Treccani to identify, define and possibly associate (where possible) current concepts and terminology with those used by authors in their books.

Starting from this premise, a series of common attributes have been identified among the books of the corpus (closely related to the common themes that are the parts of speech) that can be summarized as follows:

- The attributes of gender, number and case can be applied to the variable parts of the speech that include the name, pronoun, article and adjective;
- The verb will have as attributes the mode, the tense, the conjugation, the person and the number;
- Classification, on the other hand, is a transversal attribute to all parts of the speech since it allows them to be classified and divided into different types.

3.2. Entity-Relationship Schema

The design of the conceptual schema of the database is shown in Figure 1.

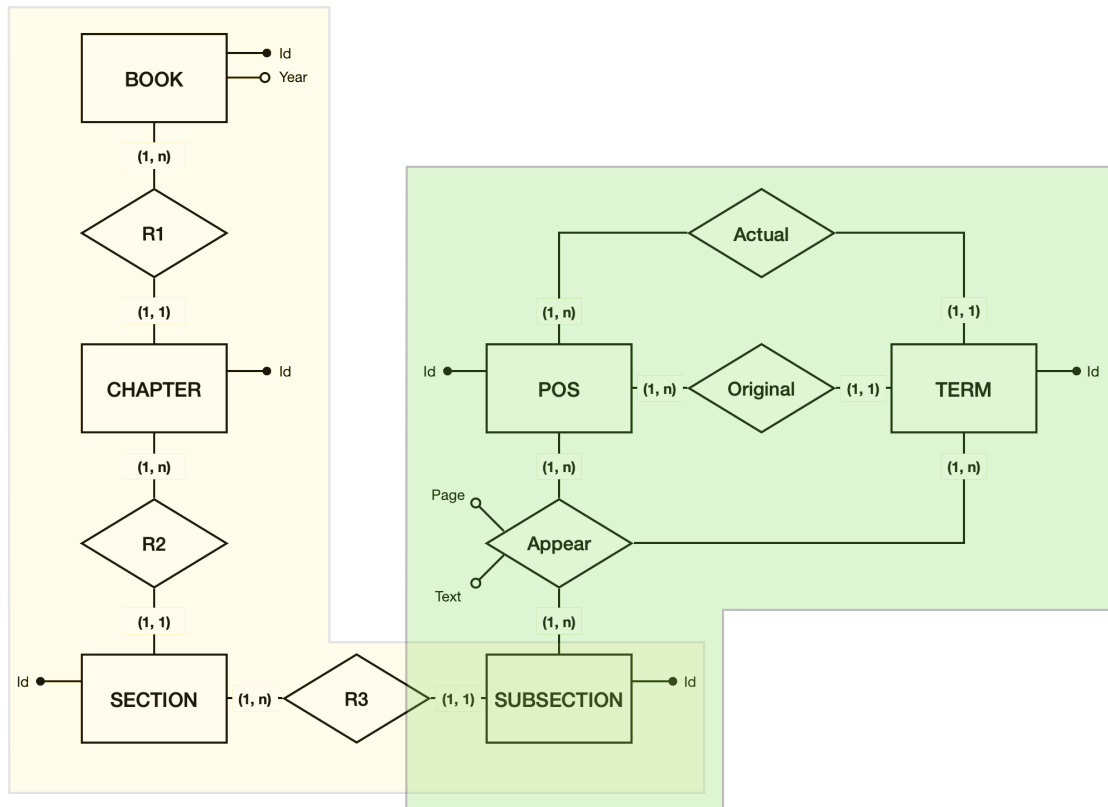


Figure 1: The Entity-Relationship Schema. Only Identifiers (Ids) and some properties are shown.

The macro-structure of the grammars is highlighted with a yellowish area. This part is designed to capture the structure of the grammar: each book (grammar) is composed of chapters, each chapter is subsequently subdivided into sections and subsections. In reality, not shown in the ER schema, there is also the possibility that a section in a chapter exists, but it is not explicitly mentioned by the author (i.e., there is a clear change in the content but the section is not mentioned).

The micro-structure of the grammars is highlighted with a greenish area. This part allows the researcher to document in detail each occurrence of each part of speech in a specific page. The association "Appear" stores these occurrences and it also gives the possibility to save the actual orthography (how it appeared in the text) to the related concept (i.e., the two orthographic variants *feminino* and *femminino* will refer to the same term *femminile* in the entity **TERM**). The two associations 'Original' and 'Actual' record what POS was linked to that particular term in that grammar and what PoS is today for the same term, respectively.

3.3. Example of Queries

The ER schema will give the possibility to the researchers in grammarography to answer, as an example, the following questions:

- In all the grammars, find the PoS "noun" and show the distribution of occurrences in the pages across years;
- Find all the conjugations relative to the typology "irregular verb" and show the context in which they appear;
- Find all the chapters that discuss about nouns and adjectives in the same subsections and in different subsections;
- Find all the orthographic variants of the PoS pronouns.

4. Conclusion and Future Work

The examination present in this paper is still an ongoing work in its early design phase. Nevertheless, we already believe that the definition of the grammaticographical genre of GEIs through the use of a specific-crafted database is one more step into the direction of a deeper integration between humanistic studies and technologically supported data analysis. We expect to be able to better define this genre, supporting the classic processes for such a definition with complex, richer and better-visualized analytical data.

But we also expect to define a new general approach to these studies, using wider and wider databases to build comparisons that as for today are impossible (or very expensive anyhow). For this reason we are carefully designing our implementation to keep the door open to future expansions, based on Spanish or Italian works but also on other languages. This means that the elements that define micro and macro-structure of books and grammars must be encapsulated in data structures that will allow us to faithfully represent the original text, but also define a common schema to support comparison between very different structures. Some of these problems have been presented in this work: the definition of chapters, sections and topics even when the original author does not use them, or use them in so many different ways that each grammar is apparently a whole different format. But as we can easily imagine, this is not completely true: some topics may vary, but others are just presented in different forms. While this macro-observation is easy to do with an analogical approach, others (like different translations, different approaches to some grammatical explanations or different use of verbs or adjectives) can be very costly without a machine-assisted approach.

Our efforts are now directed towards the building of a first demo of our database with a subset of all the data we need to analyze. This means that we have defined the data we want to extract from each grammar and how to link it (as shown in our ER schema) and we have already parsed a wide portion of our corpus. To meet our goal we need to keep working in both these directions: parsing the remaining texts on one hand; convert our schemes and designs into a proper-functional database on the other.

References

- [1] U. Mosel, *Grammaticography: The art and craft of writing grammars*, *Catching Language: The Standing Challenge of Grammar Writing* (2008). doi:10.1515/9783110197693.41.
- [2] M. L. Calero Vaquera, *Historia de la gramática española: 1847-1920 de A. Bello a R. Lenz*, *Biblioteca románica hispánica. 2, Estudios y ensayos*, Gredos, Madrid, 1986.

- [3] A. Ramajo Caño, *Las gramaticas de la lengua castellana desde Nebrija a Correas*, Universidad de Salamanca, Salamanca, 1987.
- [4] J. J. Gómez Asencio, *El castellano y su codificación gramatical. De 1492 (A. de Nebrija) a 1611 (John Sanford)*, volume 1 of *Beltenebros*, Fundacion Instituto Castellano y Leones de la Lengua, Burgos, 2006.
- [5] J. J. Gómez Asencio, *El castellano y su codificación gramatical. De (B. Jiménez Patón) 1614 a 1697 F. Sobrino*, volume 2 of *Beltenebros*, Fundacion Instituto Castellano y Leones de la Lengua, Burgos, 2006.
- [6] J. J. Gómez Asencio, *El castellano y su codificación gramatical. De 1700 a 1835*, volume 3 of *Beltenebros*, Fundacion Instituto Castellano y Leones de la Lengua, Burgos, 2011.
- [7] S. Fornara, *Breve storia della grammatica italiana, nuova ed ed.*, Carocci, Roma, 2019.
- [8] S. Fornara, *La trasformazione della tradizione nelle prime grammatiche italiane (1440-1555)*, Aracne, Roma, 2013.
- [9] C. Trabalza, *Storia della grammatica italiana*, A. Forni, Bologna, 1963.
- [10] L. Kukenheim, *Contributions a l'histoire de la grammaire italienne, espagnole et francaise a l'epoque de la Renaissance*, Noord-Hollandsche Utgevers-Maatschappij, Amsterdam, 1932.
- [11] A. M. Gallina, *Prime grammatiche spagnole ad uso degli Italiani (sec. 16.)*, Tip. commerciale, Venezia, 1975.
- [12] E. Pietrobon, A. Polo (Eds.), *Apprendere una lingua tra uso e canone letterario: gli esempi nella riflessione linguistica in Europa (secoli XVI-XVIII)*, Ledizioni, Milano, 2022.
- [13] F. San Vicente (Ed.), *Grammatica e insegnamento linguistico: approccio storiografico: autori, modelli, espansioni*, number 2019, 12 in *Quaderni del CIRSIL, CLUEB*, Bologna, 2019.
- [14] H. E. Lombardini, F. San Vicente, *Gramáticas de español para itálofonos (siglos 16.-18.)*, Studium Sprachwissenschaft. Beiheft, Nodus, Münster, 2015.
- [15] C. Castillo Peña, *Epigrama: Un portal para la edición digital de textos gramaticales (2020)*.
- [16] G. M. Alessandri, *Il paragone della lingua toscana et castigliana di M. Gio. Mario Alessandri d'Urbino*, prima editione ed., appresso Mattia Cancer, In Napoli, 1560.
- [17] G. M. Alessandri, *La tradición gramatical del español en Italia: Il paragone della lingua toscana et castigliana di Giovanni Mario Alessandri d'Urbino: estudio y edición crítica*, CLEUP, Padova, 2017.
- [18] M. Chirchmair, *Gramatica spagnuola, e italiana, composta da Mattia Chirchmair maestro di lingue nell'Accademia de' Nobili di Firenze, nella stamperia di Giuseppe Manni nella stamperia di Francesco Ricciardo*, In Firenze E di nuovo in Napoli, 1709.
- [19] A. Fabre, *Grammatica per imparare le lingue italiana, francese, e spagnola, ad istanza di Domenico Sforzini alla Vittoria in Piazza Nauona, per Francesco Corbelletti*, In Roma, 1626.
- [20] L. Franciosini, *La Grammatica della lingua Spagnuola per uso degli Italiani*, presso Giacomo Sarzina, In Venetia, 1624.
- [21] L. Franciosini, *La tradición gramatical del español en Italia. La Gramatica spagnola e italiana de Lorenzo Franciosini. Estudio y edición crítica*, Cleup, Padova, 2016.
- [22] G. Miranda, *Osseruationi della lingua castigliana di M. Giouanni Miranda diuise in quatro libri: ne' quali s'insegna con gran facilità la perfetta lingua spagnuola. Con vna tauola delle*

cose principali, che nella presente opera si contiene, appresso Gabriel Giolito de' Ferrari, In Vinegia, 1566.

- [23] G. Miranda, *La tradición gramatical del español en Italia. Las Observaciones della lingua castigliana de Giovanni Miranda. Estudio y edición crítica*, Epigrama, Cleup, Padova, 2018.
- [24] G. Sisti, *Traduzione dal francese in italiano del nuovo metodo di Porto Reale. Con cui agevolmente s'insegna la lingua spagnola. Con l'aggiunzione di due dialoghi ed un copioso nomenclatore in fine fatta da D. Gennaro Sisti ...e consecrata [al] ... Cardinal Belluga, per Serafino Porsile regio stampatore*, In Napoli, 1742.
- [25] G. Miranda, M. Troiano, A. Giuffredi, *Il Compendio del signor Massimo Troiano tratta dall'Osservazioni della Lingua Castigliana del signor Giovanni Miranda. Con le Annotazioni del signor Argisto Giuffredi*, appresso Bartolommeo Sermartelli il giouane, In Firenze, 1601.