Filling the Gap: A Three-Language Philological Dictionary Based on Contexts from Authoritative Sources

Laura Cignoni
Consiglio Nazionale delle Ricerche

This paper describes the methodology adopted for the creation of a multilingual-English-Italian-French-philological dictionary, designed to meet as far as possible the requirements of users in the field of philology, who need to use specific terms in a language other than their own. The project is addressed to graduate and post-graduate students, tutors and scholars, translators and interpreters for whom a glossary of specialised terms relative to a given universe of discourse is essential. The dictionary defines a variety of terms associated with philology and extends to other closely connected disciplines such as textual criticism, codicology, palaeography, epigraphy, papyrology, genetic criticism, etc. The three-language glossary is arranged in conventional form-each lexical entry listed alphabetically-and the English terms are followed by their equivalents in Italian and French. The Greek or Latin words from which many philological words have derived and frequently used to refer to a particular concept or phenomenon are also included. The project involves recording not just the term but also a brief contextualized definition in each language, accurately quoting the source, certified and scientifically reliable, from which the information was drawn. Alongside these definitions, a number of other contextualizations appear, also derived from authoritative sources, and different types of illustrations-e.g. manuscripts, stems, images of people and places-relative to the terminology are provided. The textual data and images will be included into an application (Alpha version) of the PINAKES project, released in March 2007, which is able to deal with different types of information-text, scientific objects, tables or graphics. This ongoing dictionary project, at present covering a total of around 1,000 words, is constantly enriched with new entries, definitions and contextualizations in the different languages.

1. Introduction

The starting point of this project was the awareness of the existence of many monolingual and bilingual dictionaries in printed form and on-line dedicated to specific sectorial activities (ranging from the more traditional disciplinary fields of history, philosophy, economics, computer science, medicine, to less academic fields such as cinema, cooking, gardening, etc.), while there seemed to be no specialized multilingual dictionary treating the philological terms appropriately. It was thus decided to start a multilingual project, aimed at the creation of a dictionary responding to the growing needs of expert users in the field of philology, who have to use the same terminology of their native language in another language. It is a well-known fact that the knowledge of one's own language is not always sufficient when the specialized terms of this language are involved. Therefore, the availability of a handbook of this type can be extremely useful to facilitate the learning and acquisition of this particular terminology.

Like all specialised dictionaries, this enterprise omits the words we use to communicate in everyday situations and which are found in general dictionaries, and limits itself to the field of philology in its broadest sense. This work does not intend to replace but rather to act as a complementary and integrative support to other important works mentioned below, avoiding duplication of efforts and waste of money.

This paper is organized as follows: the introduction starts with an account of the need for a specialized multilingual philological dictionary and presents the target audience, the disciplines and languages involved in the project. Section 2 describes the methodology that we used in the project, while Section 3 is dedicated to the treatment of definitions with some examples from

authoritative sources. Finally, after the presentation of the software in Section 4, we outline the possibility of exploiting the software as a tool for didactic purposes and conclude by emphasizing on our plans for future development and research in this field.

1.1. *Target audience*

This work is aimed at the world of users of philology who are dealing with a particular domain of knowledge, using and sharing a highly specialized jargon different from general language. The potential users of the undertaking are students who have started to dedicate themselves to research and need to learn more about the subject; University professors or scholars who have a high level of expertise but must render in a second language the terms relative to that particular discipline; translators of a work or interpreters who have a low level of expertise but need to dispose of specialised terms in their native language and are to be provided with all the documentation relative to a specific topic, for example before a congress.

The languages involved in the project are English, Italian and French. For a potential Italian user wanting to know the translation of the term *manoscritto* in English and in French the translations are *manuscript* and *manuscrit* respectively. The Italian word *pergamena* is *parchment* in English and *parchemin* in French, the Italian *pagina contigua* is rendered with *facing page* in English and *contre-page* or *page opposée* in French, and these equivalents are not so predictable. More difficult and problematic for an Italian is the translation of a term like *lezione* (intended as "the text of a work, as it has been handed down by the manuscripts") to be rendered with *reading* (not *lesson*!) in English and *leçon* in French.

1.2. *Disciplines involved*

The project moving from philology extends to other related disciplines, for example textual criticism (a branch of philology aimed at identifying the text which is as close as possible to the lost original, after comparison of the survival copies), codicology (historical discipline that studies the manuscript codex), palaeography (science that deals with the ancient Scriptures), epigraphy (a sub-branch of palaeography, which studies ancient inscriptions, generally found on hard surfaces such as stone), papyrology (science that studies ancient literary texts as preserved in manuscripts written on papyrus), genetic criticism (which emphasizes on the different stages of production of a manuscript starting from the study of the autographs).

We decided to begin with the more general texts explaining the basic facts relative to the different subjects and containing a high density of terms likely to contain statements that could be used as definitions, and then to concentrate on one specific discipline at a time.

1.3. *Languages*

The languages involved in the project are English, Italian and French. The Greek or Latin entry from which a word was derived and useful to those wishing to have a fuller and clearer understanding of the meaning of a word is also reported. See for example the term *acephalous*, which derives from the Greek *aképhalos*, "without head", formed by the privative alpha, expressing negation, and *kephalé* "head", indicating a manuscript in which the very beginning or the first pages of a manuscript are missing.

Some concepts have no equivalents in the three languages treated, and are known by their Greek or Latin term, as is the case of *lacuna*, *hápax légomenon*, *homoiotéleutos*, *constitutio textus*, *stemma codicum*, etc.), and these are the terms generally used in all three languages. An example is provided by the Latin term *incunabula*, plural of *incunabulum*:

> **Incunabula**, GREETHAM D.C., *Textual Scholarship. An Introduction*. Library of Congress. Cataloguing in-publishing-data, 1994, p. 22.
> (literally "swaddling clothes") refers to those books printed on or before 31 December 1500, that is, in the "infancy" of printing.

2. Methodology

Starting from a small sample of monographic texts and other publications in printed or electronic form dealing with the above-mentioned disciplines and in the three languages considered, we extracted a list of specialized terms and included them into a rudimentary database in the form of a glossary. We used the spreadsheet Microsoft Excel to store the corresponding words in four vertical columns, each assigned to a different language, the first for the Greek or Latin word (when available), the second, third, and fourth to English, Italian and French respectively. The following is a sample list of terms contained in the Excel application of Word (Figure 1).

	quill	calamo	calame
	quire	quaderno	cahier
	rag	straccio	chiffe - panne - peille
Lat. lectio	reading	lezione	leçon
Lat. recensio	recension	recensione	recension
	recension, *open*	recensione aperta	recension ouverte
	recension, *closed*	recensione chiusa	
	reference text	testimone	témoin
	replica	copia imitativa - copia iconica	copie figurée
	reportatio	reportatio	reportatio
	restore (to)	restituire	restituer
	roll	rotolo, rollo	rotulea, rôle
Lat. scholia	scholia	scolio	scolie
	scribe - copyist	copista, amanuense, scriba	copiste - écrivain - scribe
script	script - display	scrittura distintiva	lettres capitulaires
Lat. scriptorium, *pl*. scriptoria	scriptorium	compositoio - scrittorio	scriptorium
selectio	selection	selectio	selectio
	sheet	foglio	feuille
	siglum	sigla	sigle

Figure 1

2.1. *Organisation of the entries*

All the headwords of the glossary appear in bold character, and are listed in alphabetical order. A headword composed by more than one word, for the sake of alphabetical ordering, is treated as a single word. In the headwords formed by two terms we evidenced the most relevant, often inverting the two words, taking into account only the first part of the expression, the one preceding the comma. Therefore, an entry like *synchronic variant* in English is listed as *variant, synchronic* and will appear before *variant, substantial* and after *variant, immediate*:

> variant
> variant, alternative
> variant, apparatus
> variant, author
> variant, correction
> variant, evolutive
> variant, genetic
> variant, immediate

variant, synchronic
variant, substantial
variant, late

This disposition, necessary for the preliminary phase, that of inclusion of the lemmas in the glossary, will not be relevant when the data are inserted in the Pinakes software, as we shall see later. The dictionary does not contain information like grammatical category of the entry, distinction between transitive or intransitive verbs, countables and uncountables, parts of speech, phonetic description, etc., since the scope of this dictionary is to clarify the meaning of a word or concept rather than usage. The original idea of labelling each lexical entry with an abbreviation for each specific category or field of research was also abandonned owing to the fact that some words could be included in more than one category. In the early stage of the project we did not want to go into too much detail but we have kept track of these specifications and we intend to insert them in the future as they can be extremely helpful for the user.

Extensive work is needed to cover all the disciplines related to philology and to include languages other than the ones we have started with.

3. Definitions

For each lexical entry, as well as the translation in the three languages, English, Italian and French, we extracted one or more definitions, as well as contextualizations drawn from authoritative sources, aimed at exemplifying the meaning of the headword.

The definitions contained in important dictionaries and encyclopedias have not been included, unless the name of the author who produced such definition is specified, as occurs in the Italian dictionary edited by Beccaria where the definitions of the single words in different areas (linguistics, philology, metrics, etc.) are provided by well-known scholars (1994).

The definitions contained, for example, in the on-line encyclopaedic dictionary Wikipedia are clear, simple, well-formulated and exhaustive, but they are not drawn from authentic, scientifically reliable and acknowledged sources. The same can be said about the Vocabulaire Codicologique by V.D. Muzerelle (1985), an excellent work, which provides the user with the equivalents in four languages (French, English, Italian and Spanish) and definitions in French but, again, there is no mention of the author who wrote it out, the person who actually "signed" the definitions.

The definitions of the lexical entries in the three languages which carry the signature of the relative authors make this enterprise different from other lexicographical works. The purpose is to identify the terms of a domain and to retrieve contextual fragments that are significant for the user. While it was relatively easy to track down the definitions for the general and more frequent terms such as *code*, *princeps*, *witness*, *tree*, more detailed terms as for example *uncial* and *half-uncial capitals*, which pertain to the field of palaeography, may require the support of more specialized texts.

The contextualized definitions identified within the different works have been memorized in separate files, labelled with a name formed by a lexical entry, author's name, source from which the context was drawn, page number in the case of a paper work, electronic address in the case of material coming from the web. In this way the file is unique and can be identified easily.

The files containing the contextualized definitions in the three languages are stored in separate folders (one for each language), with sub-folders relative to the different sources. The following are some definitions of words in the three languages, each followed by complete bibliographical reference.

> **Recension,** HALL F. W. (1913). *A companion to classical texts*. University of Chicago Press, 1913, p. 108.
>
> By recension is meant the selection of the most trustworthy documentary evidence as a basis on which to found the text.
>
> **Palinsesto → Abrasione**, DEL POPOLO C., in G. L. BECCARIA, *Dizionario di linguistica e di filologia metrica, retorica*, Giulio Einaudi editore, Torino, 1994, p. 5.

(lat. *abrasio* 'raschiatura'). Anche *esarazione, rasura*. Termine con cui si indica l'operazione di raschiatura del material scrittorio (in genere la pergamena), operata per vari motivi.

Archétype, COLLOMP P., *La critique des textes*, Les Belles-Lettres, Paris, 1931, p. 20.

L'archétype est le plus récent ancêtre commun des manuscrits existants ... Lorsque les manuscrits divergent, ou bien les leçons sont ... l'une est évidemment bonne et les autre évidemment fausses, et alors, si l'archétype n'avait pas la bonne leçon, sa postérité ne peut l'avoir retrouvée que par contamination ou correction, ... s'il avait la bonne leçon il était sur ce point pareil à l'original.

In the case of *archétype* above, the definition, clear and well-formulated, was originally too long, and therefore it was decided to perform some minor cuts to the text reducing it to the present form.

Those parts of a text that did not produce a definition—either because it was too long and/or difficult to extrapolate or required the knowledge of background information of a subject—have been exploited as contextualizations. As we know very well, no definition is ever completely exhaustive, but the consecutive reading of a number of quotations at different levels of analysis and difficulty can help give a fuller and clearer understanding of what a word means.

Some passages with contextualized definitions are sometimes semantically very rich, and provide additional useful information which can be exploited as cross-reference to other lemmas.

4. Software

The data collected in the files and folders will be included in the system Pinakes, which is one of the main research projects of the "Digital Renaissance Foundation" (New Technologies for Cultural Heritage) of the Istituto e Museo della Storia della Scienza in Florence *http://www.rinascimento-digitale.it/index.php?SEZ=31.* The aim of the project is to group a relevant number of existing research projects of critical editions to share software, technologies, know-how and provide open and free downloadable applications to the large community of humanities studies and research.

The information is stored in different modules, correlated by hypertextual links, one for each language and independent one of the other. The system creates no limits as to size of page and alphabetical ordering; in fact, the availability of space offered by the computer makes it possible to include an amount of data that could not be contained in the paper dictionary.

The application of the Pinakes system, released in March 2007, is able to handle information of various types—text, tables or graphs, structured texts in HTML, XML, etc.—as well as digital objects like images, tri-dimensional objects, video, etc.

The user who registers with Pinakes is assigned a password which allows him/her to share the data with the others availing themselves of the system. Figure 2 shows the names of some of the ongoing Pinakes projects.
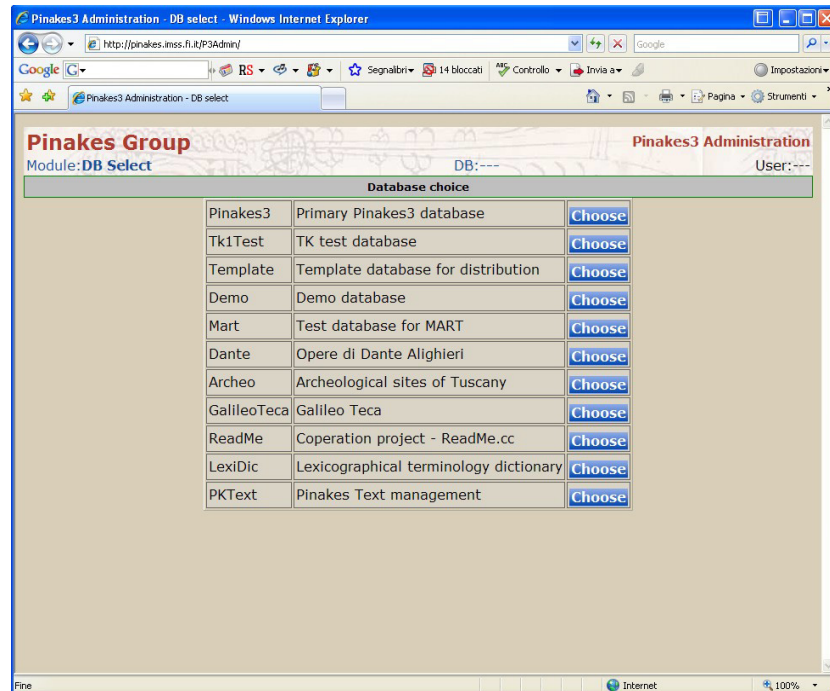
Figure 2

Sometimes a term, or a contextualized definition is found for one language but not for another. In this case, the space designed to include for example a term or a description of that particular term will remain empty until the missing information is found.

Figure 3 shows the term *acephalous* in English and *acefalo* in Italian, while the French term with its corresponding definition is missing. As we can see, the Pinakes system envisions the possibility of introducing languages other than the ones so far included in the project.
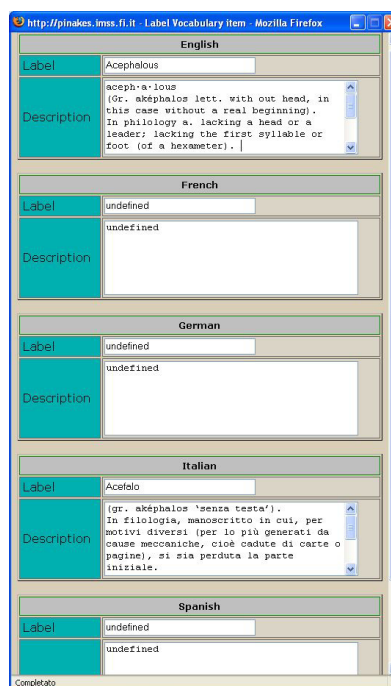
Figure 3

Terms like *capitalis quadrata* (the square capitals used for the large, angular inscriptions of Roman monumental writing), or *ostrakon* (piece of pottery, or stone, containing scratched-in words or other forms of writing which may give clues as to the time when the piece was in use) may require the support of an image which helps make out the meaning of a word: for example, Figure 4 shows images of *ostraka* from Medinet Madi in the Fayyum Region of Egypt (Bresciani and Menchetti 2004).

Figure 4

The absence of an illustration can render a definition difficult if not impossible to understand; the illustrations are a supplement to the textual explanations as they clarify the definitions and tell us more about the word in question. They contribute to answering the questions of a certain type of reader, and can help provide additional information on some concepts, for example the image of a manuscript, of a stem, of a *scriptorium*, etc.

When inserting an illustration, the system asks the user to specify the details necessary to describe the object, providing a number of details such as characteristics, provenance, current location, etc. Figure 5 shows a screenshot from Pinakes, with the detailed data that it is necessary to insert for the illustration of the term *manuscript*.
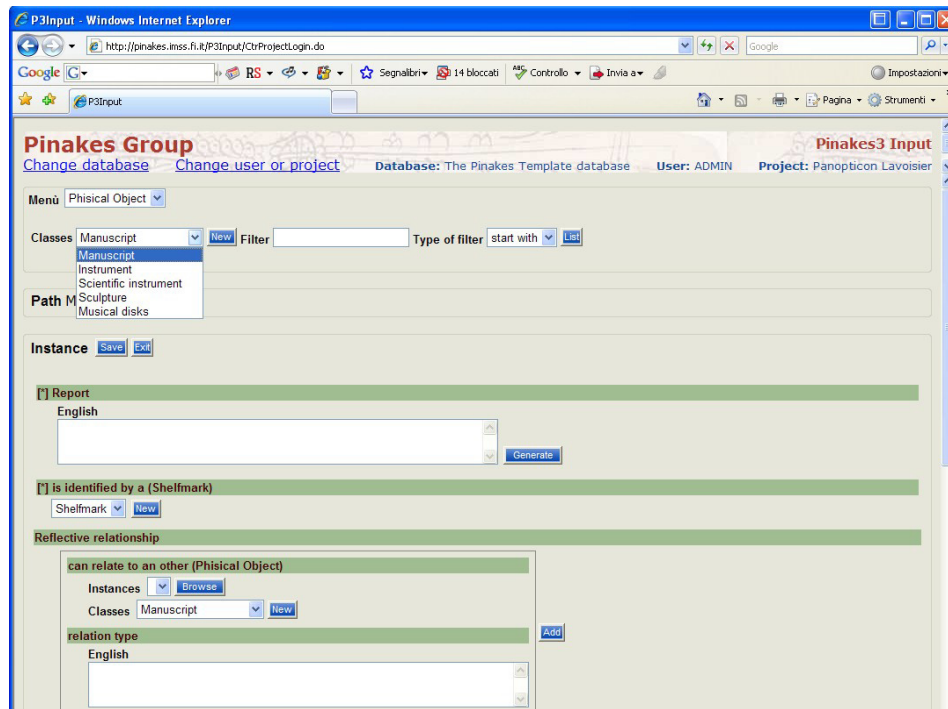

Figure 5

The electronic medium makes it possible to click directly on the word that has been chosen, thus replacing the traditional vertical layout of the text and lookup of the words in alphabetical order.

The definitions and contextualizations relative to each lexical entry can include links cross-referencing to the pages of other sites, as for example those of *Wikipedia* and of the *Vocabulaire codicologique* of Denis Muzurelle mentioned above.

5. Final remarks

This paper is an account of the design, progress and achievements of the project on philological terminology which was started at the beginning of 2007. The issues and problems that need to be addressed when compiling a multilingual contextualized dictionary from a number of different sources have been discussed.

This three-language lexicon can be used in institutional or academic settings for educational purposes, for the teaching of philological terminology, providing information that is currently not so easy to reach on the part of the student.

A University student or foreign language learner can read the short definition of a term in a language other than his own and then check the corresponding term supplied with one or more definitions and/or contextualizations in his own language. Let us suppose that a user looks up the Italian and French translations of the word *draft*. Once he has found the Italian lemma, he can learn the translation in the other two languages (*brutta copia* in Italian, and *brouillon* in French, and then read the relative definitions and contextualizations.

The hypertextual functions allow the user to consult a large number of definitions at different levels of depth, moving from one environment to another, visualizing the translation of the same term in the different languages, with the relative definitions and contextualized examples, each of which is well documented, with specification of the author and precise bibliographical reference.

The reading of various definitions which can be reached through appropriate references within the same passage can contribute to the comprehension of a term, help the user understand better, have a clearer and more exhaustive picture of a particular word, concept, phenomenon.

The software also allows the user to create cognitive learning routes, as reference-point and/or consultation for independent construction of knowledge, with an in-depth study of the subject, fostering integrated and efficient learning. When reading the definition given by an eminent author, the user might be encouraged to go and read that work or others by the same or by other authors.

This material will be handled by a system which allows data navigation in the web using the most common browsers as well as data publication in electronic (eg.: CD, DVD) or printed form.

Acknowledgements

References

Avalle, D'Arco S. (1978). *Principi di critica testuale.* Padova: Antenore.

Beccaria, G. L. (ed.) (1994). *Dizionario di linguistica e di filologia, metrica, retorica.* Torino: Giulio Einaudi Editore.

Bédier, J. (1929). La Tradition manuscrite du lai de l'Ombre, réflexions sur l'art d'éditer les Anciens Textes. Paris: Librairie Ancienne Honore Champion.

Bresciani, E.; Menchetti, A. "L'application des technologies numériques aux texts non alphabétiques de l'ancienne Egypte". *Linguistica computazionale* 20-21.

Collomp, P. (1931). *La critique des textes*. Paris: Les Belles-Lettres.

Dain, A. (1964). *Les manuscrits*. Paris: Les Belles Lettres.

Deppman, J.; Ferrer, D.; Groden, M. (eds.) (2004). *Genetic criticism.* Philadelphia: University of Pennsylvania Press.

Fiormonte, D. (2003). *Scrittura e filologia nell'era digitale.* Torino: Bollati Boringhieri.

Greetham, D. C. (1994). *Textual Scholarship*. *An Introduction*. Library of Congress cataloguing in-publishing-data.

Hall, F. W. (1913). *A companion to classical texts*. Chicago: University of Chicago Press.

Kenney, E. J. (1974). *The Classical Text. Aspects of Editing in the Age of the Printed Book*. Berkeley: University of California Press.

Maas, P. (1975). *Critica del testo*. Firenze: Le Monnier.

Macé, C. et al. (eds.) (2006). "The evolution of texts: confronting stemmatological and genetical methods". *Linguistica computazionale* 24-25. 255-283.

McGann, J. J. (1983). *A critique of modern textual criticism*. Chicago: University of Chicago Press.

Muzerelle, D. (2002-2003). *Vocabulaire codicologique*. Edition hypertextuelle, Version 1.1.

Numerico, T.; Vespignani, A. (eds.) (2003). *Informatica per le scienze umanistiche*. Bologna: Il Mulino.

Pasquali, G. (1974). Storia della tradizione e critica del testo. Mondadori: Milano.

Pearson, J. (1998). Terms in context. Studies in Corpus Linguistics. Amsterdam: John Benjamins.

Stussi, A. (2006). *Fondamenti di critica testuale*. Bologna: Il Mulino.

Timpanaro, S. (1981). *La genesi del metodo del Lachmann*. Padova: Liviana Editrice.