New Media and the Humanities: Research and Applications

Edited by Domenico Fiormonte and Jonathan Usher

Proceedings of the first seminar
Computers, literature and philology
Edinburgh, 7-9 September 1998

2001
Published by the Humanities Computing Unit, University of Oxford
on behalf of the
Istituto Italiano di Cultura per la Scozia e l’Irlanda del nord
Centro Ricerche Informatica e Letteratura,
Università degli Studi di Roma "La Sapienza"

Francisco A. Marcos Marín

Where is electronic philology going?
The present and future of a discipline

I. Data

My students know very well how prone I am to reach conclusions only after the presentation and analysis of a concrete problem, usually linked to the activity of one of our research teams. In this case, my starting point will be Buenos Aires, where an Argentinian group of scholars, under my direction, is working on the cataloguing and studying of the Foulché-Delbosc collection in the National Library. The project is supported by the Bilateral Agreement between Spain and Argentina, and funded by a number of institutions1 in both countries.

Raymond Foulché-Delbosc was one of the major French hispanists of all time. He developed his activity mostly at the end of the 19th Century and the first third of the 20th. He founded the well known journal Revue Hispanique, in which he published, under different names, his editions of old texts, as well as substantial contributions to Spanish studies. Many of the books that came to light in those pages were part of his personal library. After his death this excellent collection went under the hammer in 1936. The catalogue of the auction was published and the content of the library was therefore known by everybody2. From that moment on, however, most scholars lost all trace of a substantial number of the books that were marked as ‘lost’ in the bibliographies, reference catalogues and databases devoted to Medieval and Classical Spanish. Nobody seemed to know who had bought more than one thousand two hundred of these books, and where they might be.

It would not be wise to reveal now, prematurely, what happened to these volumes without first taking advantage of the expectation created to explain what those Iberian bibliographical databases are, what their meaning from a new philological perspective is, and how we propose to present the contents of our libraries and archives to scholars in the digital era.

In 1990, the Spanish Agency for the Celebrations of the Quincentennial of the Discovery of America opened an Area of Language Industry. When the Spanish authorities invited me to take the reins of the Area, they had in mind the work that was being done in the Science Center IBM-UAM, at the Autonomous University of Madrid, whose main philological goal was the creation of UNITÉ, a package comprising several programs for critical editing of texts, including not only the collatio, but also a large part of the

1The Agencia Española de Cooperación Internacional, AECI, the Secretaría de Cultura of the Presidency of the Argentine Republic, the Secretaría de Estado of Universities, Ministerio de Educación y Cultura of Spain (PR1997-0019 002369550), the Universidad Autónoma of Madrid, and the Biblioteca Nacional de la República Argentina. It is only fair to acknowledge the debt to Esperanza Aguirre, Minister of Education and Culture of Spain, and Beatriz K. de Gutiérrez-Walker, Secretary of Culture of Argentina. Both gave courageous support to the project from the very beginning.
2The Catalogue was actually based on a previous publication, Catalogue de la Bibliothèque Hispanique de M. R. Foulché-Delbosc. Abbeville: Imprimerie F. Paillart, 1920. It was issued in Mayenne: Imprimerie Floch, 1936.
The first result of the UNITE project had been published in 1987, as the critical edition of an early 13th c. Spanish book, the Libro de Alexandre (Marcos Marín, 1985, and 1987). Among the different projects supported by the Area, one, ADMYTE, the Digital Archive of Spanish Manuscripts and Texts in CD-ROM, has been specially relevant for Spanish Philology. ADMYTE was the result of the collaboration of several institutions in different countries. Some of the projects included in the CDs existed before ADMYTE, and that made the Digital Archive feasible. In the field of bibliography, the name was BOOST, Bibliographia of Old Spanish Texts, which had by then reached its third edition on paper, published by the Hispanic Seminar of Medieval Studies in Madison, Wisconsin, and had become a basic tool for any library interested in Medieval books written in Castilian Spanish, Catalan, Galician, and Portuguese. Names such as those of professors Charles B. Faulhaber, Arthur L-F. Askins, Angel Gomez Moreno, Harvey Sharrer, Martha Shaffer or Gemma Avenzoa are associated with it.

The difference between BOOST and Philobillon, which is the name of the electronic version, illustrates clearly a first issue in the debate between old and modern Philology. BOOST had, as any printed book, a reduced number of indexes: titles, authors, subjects. Philobillon, which is implemented as a relational database in Advanced Revelation, allows more than 350 fields. It also permits the interactive handling of three different bibliographies: BETA, Bibliografía Española de Textos Antiguos (developed from the original BOOST, BITECA, Bibliografía de Textos Catalanes Antics, and BITAGAP, Bibliografía de Textos Antiguos Galegos e Portugueses). I will not dwell on the differences between indexing a book and browsing a database, because we are all well aware of them. It is actually more interesting to give a quick glance at the system, which is now available in two versions. The shortest and easiest version to use is implemented for the World Wide Web, its URL is http://sunsite.berkeley.edu/Philobillon/phhm.html. The largest version was distributed with ADMYTE–0, in 1993, and has been updated several times since then, with new releases of the program, which runs as a DOS program, or as a DOS window in a 32 bit operating system. It would be misleading also to limit ourselves to a consideration of the fact that the advantage of a database on an indexed book is limited to the possibility of including a larger number of fields. The main advantage is that it allows, and even makes, the scholar ask himself more questions, delve into many details surrounding his data, that may now be ordered and related to other aspects of his research.

The question of a better implementation of the program and the well-known limits of DOS should not concern us here: it is an issue for computer scientists to deal with. As philologists, we must realise the improvement implied by a system that offers the possibility of relating any copy of a manuscript to the rest of the copies, informing us not only of the basic details, such as author, date, copyist, location, catalogue description and identification, number of known copies and their location, titles; but also about the owners, persons related to the book or to any person related to it, bibliography, with new links to any names mentioned in the bibliography, and even more practical details, such as where the library is located, its address, phone, fax and e-mail numbers, what facilities it offers, opening hours, and even how to address the librarian, a courtesy sometimes not so irrelevant in the highly sophisticated Iberian world.

Philobillon, therefore, is a first model for descriptions and primary access to the information to be considered when a team finds a huge deposit of books, whose trace has been lost by the scholarly community. The scholars involved in that discovery, have, as their first goal, the need to make that information available as soon and as accurately as possible. This was actually the situation in Buenos Aires when, in 1996, Arthur Askins and Harvey Sharrer, in their search for manuscripts for the databases included in Philobillon, particularly manuscripts of Portuguese authors, visited the new building of the National Library in Agüero street, a name probably auspicious, as agüero, in Spanish, means 'omen'. When looking at the card files in the Sala del Tesoro, they noticed that several manuscripts carried the FD mark. Then they remembered the unknown destiny of the still missing manuscripts of the Foulech-Delbos collection. The quest for the treasury had started in the right place. Arthur and Harvey did not keep the discovery to themselves, and communicated it to other scholars, as Buenos Aires was currently hosting a medievalist conference. However, nobody paid special attention to the news, and, thanks to the collaboration established between Arthur Askins and our team since the days of ADMYTE, I took note of the information and started working in the Biblioteca Nacional with the assistance of some of my collaborators in Buenos Aires, with the aim of launching a joint research venture sponsored by the two governments, as is now the case. A catalogue is being written, and will be published not only on paper, but also across the web, in electronic format, which implies digital facsimiles of at least some chosen pages, and transcriptions of excerpts of each manuscript, until a CD-ROM with the whole versions can be produced.

Our initial idea was becoming more concrete. We had rediscovered, 60 years later, the whereabouts of a huge number of the books auctioned in Paris in 1936. It is only fair to add immediately that it is not entirely true that nobody knew that the books were in Buenos Aires. At least four notes had been published since the acquisition. They did not seem to reach the audience. The acquisition was duly reviewed in the Memoria of the Biblioteca Nacional corresponding to 1936, and in 1937, in the first issue of the Revista de la Biblioteca Nacional, ('La Biblioteca Nacional' 1937, 206) published again under the direction of Gustavo Martinez Zuviria, after a long silence following the publication of the last number of the Anales de la Biblioteca Nacional, which had been directed by Paul Groussac. In 1937 the copies entered the A sala de reservados, currently Sala del Tesoro, as Colección F-D.1 In 1992 a new reference to the collection (Salgado 1992) was

1 BETA is compiled by Charles B. Faulhaber, Angel Gomez Moreno, Angela Moll Decent, and Antonio Cortijo. BITAGAP is compiled by Arthur L-F. Askins, Harvey L. Sharrer, Martha E. Schaffer and Aida F. Dias. BITECA is the work of Vicente Beltrán, Gemma Avenzoa and the late Beatrice Cowieff. The disk 0 of ADMYTE, Archivo Digital de Manuscritos Españoles, was edited by Francisco Marcos Marin, Gerardo Meiro, Charles B. Faulhaber, John Nitti, Angel Gomez Moreno, and Aurora Martin de Santa Olalla, CD-ROM, Madrid: Miconet 1993. ADMYTE and UNITE are registered trademarks by Francisco Marcos Marín, John May designed Philobillon. Information about ADMYTE and related projects can be found on: http://www.BLLB.usm.es/~fmarcos/informes/admyte/admytexit.html.

2Whose pen name was Hugo Wast. He was appointed Director of the Biblioteca Nacional by the de facto government of General José Félix Uriburu, on the 10th of October 1931, and remained Director until 1955, with the exception of two very short periods in 1941 and 1943. During the first of them he served as Federal Interventor in the Catamarca province. During the second he was Minister of Justice and of Public Education in another de facto government, that of General Pedro Pablo Ramírez. (Cf. Sáez 1997). The person who actually purchased the books at the auction, Jorge Max Rohde, was also well known in the intellectual circles of Buenos Aires.

3 At that moment they were catalogued and included in the card files of the Reserved Room. Then they were almost lost, as I said. In any case, a reference was included by Milton A. Buchanan in his "Bibliographical Notes", Hispanic Review, 9, 1941, 228-230. Nobody could give us any information or assistance when Verónica Zumbraglia and I visited the Library in 1990, trying to prepare an ADMYTE CD-ROM with manuscripts and
Where is electronic philology going?

made, although it was included in a publication with limited diffusion. After moving to
the new location of the Library, Hugo Acevedo wrote the chapter on the history of the
National Library of Argentina published in the joint volume edited by the Association
of Libraries of Spain and Latin America, *ABINIA*, (Acevedo 1995) where he referred to
the acquisition in 1936 and put the accent, among other bibliographical treasures, on "varias
ediciones de La Celestina", with special mention of the volume of "Sevilla 1502", one of
the three extant copies of that printing (Olivetto 1998).

When the collection was rediscovered in 1996, and its full importance was duly
acknowledged, there existed only two sources of information about it: the old card files of
1937, whose information was very limited, and the more detailed, albeit also restricted,
information included in the catalogue of the auction. In the first case, the card files, only
the first of the several books often included inside the same binding was mentioned.
When a volume contains several books, there is no information of its whole content
on the card files. In the second source, the auction catalogue, there was no indication
of who had bought the books and, more concretely, which books had actually been
bought by the Argentinian government in 1936. There was, therefore, no possibility of
identifying and locating the books on the Library shelves. Georgina Olivetto undertook
the challenging task of preparing the lists of concordances between the card files and the
auction catalogue, with the valuable help of Librarian Hugo Acevedo, a person who has
a thorough knowledge of the deposits of the Biblioteca Nacional.

With these tables in our hands we might start preparing the way to give details of
the collection's contents to the scientific community. As I have already said, many of the
books were considered lost. The auction had taken place in 1936, times were not peaceful
after that date, neither in Spain nor in the rest of the world. Most people thought that
the books could have been destroyed. Even those who knew that there was a Foulché-Delbosc collection in Buenos Aires had no means of knowing how many books were
in the collection and what condition they were in, unless they visited the Library and
investigated the card files.

Although times are changing very quickly, editing the catalogue in book format is still
necessary. It is, however, insufficient. We have to establish a link between the old and the
new Philology, again, by preparing a format that can achieve three goals: publishing in
printed form, incorporation into the electronic catalogues of the Biblioteca Nacional that
uses Micro-Isis as its cataloguing database, and preparation for access through Internet.
Another requirement was the need for a system that includes text and pictures.

We were aware of the existence of a group of the Text Encoding Initiative that
has prepared an SGML proposal for the cataloguing of collections of ancient books,
which means that our system will have to be translated into that format in the future.

Philobiblon has evolved meanwhile from the DOS limited possibilities to its new web
presentation, Arthur Askins has announced the preparation of a CD-ROM with a new
version, and Charles Faulhaber, as director of the Bancroft Library, in Berkeley, has
kindly offered any collaboration needed for the development of our project, as we had
done already in the past. We opted for a text template (shown below in Appendix A
(Input template), as the most suitable form for further translations into the different
sources envisaged. An HTML version is built automatically from that source, the XML
version becoming the immediate goal.

The kind of information suitable for collection by a printed catalogue is not enough. It
has to be complemented in several ways. We cannot detail all of them now, but their main
spheres can be delineated. In this paper I will discuss the hypertext version, regardless
whether its concrete form is HTML, XML or SGML. All of them, from the linguistic
perspective, are dialects of the same basic markup language. After opening the hypertext,
the immediate options will lead us in the following directions: in the first place, the
interrogation of the database, to build our own indexes, according to the parameters we
can choose by interrelating the different fields of the template. Technically, it can be
done with a series of small programs or CGI, implementing different search options.
A second direction can take us to the text of the book itself, its transcription, preferably
in paleographic format. The third line will take us to the facsimile reproduction of the
book, usually in either compressed TIFF format, or JPEG.

For those who know our previous work, it will be clear that this implies an evolution
of the ADMYTE format. An example of the parallel columns of text and facsimile, which
allows easy checking and correction of foreseeable transcription mistakes is given in
Figure 1. A clear advantage is that the digital reproduction is far more accurate than

![Figure 1. Facsimile and transcription in ADMYTE-1](image-url)
human sight, which therefore means that we can store the information to a high degree of
definition, whilst for its presentation to the human eye a much reduced definition will
suffice. A rule of thumb is that it is sufficient to store the images at a resolution that
doubles that of its presentation. Storing at 600 dpi, for instance, will give clear printouts
at 300 dpi, a quality that improves clearly that of any photocopy. Digital reproduction
implies no harm to the books themselves, as opposed to the harm to both the bindings
and the actual pages, entailed in the forced position imposed by the photocopying, and
the heat and strong light necessarily involved.

2. Analysis

Once we have proceeded to the presentation of the raw data, the most important
objective for any philologist will be its analysis. Any computational analysis of texts is,
by definition, quantitative. The whole matter consists of knowing which elements of the
binary code are in which positions and comparing those elements or those positions,
or both, with other elements or positions. This is, of course, an extensional definition,
because it allows the definition of a set by its elements. Together with the extensional
definition, there are cases in which intensional definitions are possible, definition after
the properties that characterise the elements belonging to the same set.

In postulating a typology of Electronic Philology, we must take into account the data,
the procedures, and the results. The data can be either homogeneous or heterogeneous. In our
case, the different types of fields and the need to deal with texts, but also with images,
points to heterogeneity as a characteristic of the philological data. That characteristic
makes it necessary to develop standards that limit the presentation of the data, in such
a way that the effort made to collect them is not lost: data must be encoded in a way
that allows scholars to reuse them. The encoding of books in a catalogue, in our case,
involves structural description, and formal description as well, because each book is, at
least partially, treated as a unique copy.

The typology of procedures is more complex, and more interesting too, because the
researcher is not only dependent on the data, but also on his interests, the means and
devices at his disposal, and his limits.

By straight quantitative procedures we understand those aiming at obtaining a mass of
results justified simply by its dimension. Counting is a typical case of a quantitative
procedure. A few years ago, it was complicated to know the number of words, or vowels,
or particular marks, that constituted a text. Any word processor nowadays solves this
difficulty, adding much more information. We are so used to it, that we do not realize
how easy it is now to know that in Nebrija's dictionary Latinus-Hispanicus, included
in ADYETE-I, there are 240 words that appear more than 20 times (between 21 and
3,659, to be precise), or that the famous Spanish epic poem of the 12th century Cantar
de Mio Cid contains 439 instances of the word Cid with cedilla and only two of the same
word with a simple c. Quantitative procedures, however, are not limited to counting.
They can include any procedure that considers the data following strictly the extensional
characterization of a set; for instance, when comparing two files to know whether they
are or are not two instances of the same text. This activity must not be confused with the
collation, which involves more complex considerations.

Searches, contrasts, including the collatio, and all aspects involving an intensional
characterization, must be classified as qualitative procedures. These procedures are charac-
terised by a common feature in the definition of the compared or related elements; there-
fore, they take into account features that belong to the compared elements themselves.
ADYETE, for example, includes a formal glossary, a context-free glossary of lemmas,
and forms that permit an extremely productive association among variants of a lemma
that differ graphically and also morphologically. The system is limited by a rigidly
established hierarchy of associations between forms and their possible lemmata, but has
proved its usefulness as a general tool for the analysis of the highly polymorphic medieval
texts. Another significant example of a qualitative procedure is that of the collatio in
UNEITE, our system for critical editing of texts. The new UNITE release for 32 bits
systems collates simultaneously up to 50 versions of any ASCII text, in any language, in
either prose or verse (as can be seen in Appendix B below). It also provides the scholar
with a rich panoply of results, which will be considered shortly, because, as a part of the
recension, they are included in the last type of procedures.

The exploitation of huge textual resources has led to the development of complex
information retrieval systems, and rich query languages. These are characterised because
they act on a mass of data, which for the sake of the exposition we shall call a database,
extracting from that data, entities or references whose value is given in terms of quantity,
for instance, how many pages contain the word Cid. The entities or references defined
in these terms of quantity are called operands. Those operands can be related and
combined. An operator relates two operands and offers as the result of that combination
the combined value of the two related operands, which now may serve as a new operand
to be related to yet another operand. Any procedure that associates, for instance, a lemma
with its forms, operates by building lists of the elements, its operands, associated by the
operator = (association). Up to this point we are limited to the qualitative procedure, by
simply searching and locating.

However, we can go further by assigning a value to an operand, such as an integer n
that gives us the possibility of carrying out searches of words separated by n, or by using
the typically selective operators, the boolean operators (or, and, not). This third class, there-
fore, groups the so called selective procedures. The main application of selective procedures
for philologists goes far beyond the establishment of ranges or the application of boolean
operators. It resides, from our point of view, in the automatization of the recension from
the collatio. Once again, the advantage for the scholar consists not only of the amount of
data processed in a very short time, but also of the number of possibilities brought about
for selection by the application of quantitative and qualitative procedures.

UNEITE offers the possibility of producing either a short summary of the processes
involving in the collatio, or a much larger file with an extended version of the whole
development, both partially reproduced in Appendix C. The objective of a mostly
automatic recension is closer everyday.

We have considered the data and the procedures. It is also possible to categorise the
results in three types, separating that which is accumulative from that which is select and
that which is critical in the three types of primary, secondary, and tertiary results.

As primary results we consider those which are purely extractive or accumulative. Any
list of words, sorted in any order, can be a good example. Most word processors, and even
basic commands of operating systems, offer that possibility. More complex packages,
such as TACT, allow for changes in the distribution of the graphemes in different scales.
Greek, Arabic, Hebrew, Traditional Spanish. It is convenient to link those programs with routines that allow the cleaning and even the interpretation, of certain tags. An example could be the tag remark [RMK] in the texts coded following the HSMS system. Between the braces of that tag, words are included which express the opinions or doubts of the transcriber, and therefore do not belong to the original text. As observations written by the transcribers, they can even be in a language that differs from the language of the text that is being transcribed. This is often the case when American scholars transcribe Old Spanish texts; their remarks used to be written in English. If a list of words is produced automatically, disregarding that issue, the foreign words are also included. There is no necessity to give a concrete example of this failure here, but it could be produced, were it needed.

Primary results constitute a set from which secondary results can be obtained. This set can be interpreted as secondary can be selected from primary. An example is the list of selected words obtained from a whole list of words. The selected list constitutes the base to prepare basic, indexes, and concordances. There are many examples of commands, such as grep in UNIX, instructions or programs that help obtain secondary results: Word Cruncher, OCR, TACT are names that come easily to mind. However, it is remarkable how many colleagues do not feel at ease with those packages, and how much there is still to be done to familiarise scholars with these valuable tools.\footnote{As for the Spanish language, I gave examples in El Comentario Filológic con Apoyo Informático, Madrid, Síntesis, 1996.} In the field of editing, a typically select result is the result of the collatio, the file containing the so-called unified text, i.e., those elements extracted and grouped automatically by the program.

Tertiary or critical results are obtained from the selected ones, following an exact pattern. In this case, human interpretation is crucial. When we use a concordance or an index to build a dictionary, or the results of the collatio to prepare a critical edition, we are aiming at critical results. It goes without saying that different types of results can be combined in a final product, as when we prepare a critical edition, with an index and a list of words ordered by frequency.

The following table presents a scheme of the proposed quantitative analytic typology:

<table>
<thead>
<tr>
<th>Data</th>
<th>Homogeneous</th>
<th>Heterogeneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedures</td>
<td>quantitative</td>
<td>qualitative</td>
</tr>
<tr>
<td>Characteristics</td>
<td>extensional</td>
<td>intensional</td>
</tr>
<tr>
<td>Examples</td>
<td>inventories, comparisons</td>
<td>direct searches, collatio (as a procedure)</td>
</tr>
<tr>
<td>Results</td>
<td>primary</td>
<td>secondary</td>
</tr>
<tr>
<td>Characteristics</td>
<td>extractive, accumulative</td>
<td>differentiated</td>
</tr>
<tr>
<td>Examples</td>
<td>ordered word lists</td>
<td>indexes, concordances, collatio (as a result)</td>
</tr>
<tr>
<td>Criteria</td>
<td>accumulative</td>
<td>select</td>
</tr>
</tbody>
</table>

3. Back to the beginning

Once we have designed the broad perspectives of Electronic Philology, we can go back to the practical example with which we started our presentation. The Foulché-Delbosc collection constitutes an excellent test for many of the issues we have presented, in a somewhat simplified fashion.

The application of the lessons learned from Philobiblon or ADMYTE seems quite straightforward. The publication of the catalogue and the preparation of a CD-Rom with facsimile and transcription being starting points in this kind of task. However, I cannot finish this presentation without emphasising the core of our purpose. Electronic Philology is not only a way to make things faster and inventories more accurate. It is not just a tool. It is also, more deeply, a change in our attitude. For certain aspects we must understand that limiting ourselves to a well chosen set, a well designed corpus, for instance, can be suitable. There is also, however, the unlimited capability to deal with the whole: the whole Text, in a comprehensive corpus; the whole Work, in a comparison of all versions of the same work; the total number of copies of a manuscript; all variants of a construction, a formula, a sentence.

This is the reason why a typology is necessary; to adapt the means to the goals, the efforts to the needs. Philology has been, for centuries, the main humanist activity. Many people are still looking at computers as enemies of the intellectual activity that has characterized the approach to texts from the beginning. We are not convincing them by saying that we get more accurate accounts of words, or thousands of examples. What is different is not the quantity, it is the new insights, the new questions that we can ask, even more than the old questions that we can answer, a fact not to be forgotten. What is also different is the ease with which we can share our resources and our knowledge with scholars all over the world. The Foulché-Delbosc collection had been almost forgotten for sixty years, in spite of notices published in different printed sources. The situation, in less than two years, has changed dramatically: in a few months, more than one thousand texts, offering a richer view of Spanish Culture, will be ready for scholars to publish, study, and connect with other versions or with other aspects of human written production. That goal can be achieved only by means of the computer, and it is also the computer that is the ideal tool to make the texts themselves available to a large public, and also for research, without endangering them by direct unnecessary exposure. The ideal of preservation and usage is feasible thanks to the new techniques. Going beyond, discovering the new frontiers of Philology, does not depend on the computer. It depends on us, the scholars.
Appendix A. Input template

After [only for Medieval texts included in Philobiblon]
BIBST [only for texts of the Bibliography of Old Spanish Texts]

Author:
Title:
Other titles:
Language:

Notes to the book (bound with / occupies ff &&x2026 ;) [detailed description of volumes containing several works]
State of preservation:
MANID: [key in Philobiblon, when it exists]
TEKID: [key in Philobiblon, when it exists]

Version 1: qui__._.( 2 3 4 ) 1 la( 1 2 3 4 ) vida( 1 2 3 4 ) quisiere( ( 1 3 4 ) 2 ) de( 2 3 4 )

Version 2: /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ;

Version 3: /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ;

Version 4: /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ;

The process includes the results of the comparison for a whole section in the form of coincidences grouped for the maximum of versions as well as in pairs. The example shows the first stanza:

Estadísticas por grupos:
Versiones 1 2 3 4 : 28 coincidencias
Versiones 1 3 4 : 4 coincidencias

Estadísticas por parejas:
Versiones 3 4 : 32 coincidencias
Versiones 1 4 : 32 coincidencias
Versiones 1 3 : 32 coincidencias
Versiones 2 4 : 28 coincidencias
Versiones 2 3 : 28 coincidencias
Versiones 1 2 : 28 coincidencias

Appendix B. Qualitative procedure

A collatio of the first stanza of Gonzalo de Berceo's Vida de San Millán, 13th century (output limited to the first verse only)

Comienza el proceso de unificacion (July 20, 1996 7:13:43 pm)

Estrofa 1:

Estrofa original version 1:
Quien la Vida quiser de Sant Millan saber
Estrofa original version 2:
Qui la Vida quiere de San Millan saber
Estrofa original version 3:
Qui la Vida quiser de Sant Millan saber
Estrofa original version 4:
Qui la Vida quiser de Sant Millan saber

Version: 1A

Unificado: qui__._.la vida quiere de san__._.Millan saber

Version 1: qui__._.(2 3 4 ) 1 la( 1 2 3 4 ) vida( 1 2 3 4 ) quiser( ( 1 3 4 ) 2 ) de( 2 3 4 )

Version 2: /&.../&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ;

Version 3: /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ;

Version 4: /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&.../&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ; /&&x2026 ;
Appendix C. Selective procedure

These files list the procedures applied for the collatio and their results. The output has been limited to the first verse, for this example.

Short version:

Estrofa 1:
Versos: 1A, 2A, 3A, 4A
Unión de las palabras 'qui' y 'siete' en la versión 2
Letras comunes: Quien (1) Qui (2 3 4) ==> qui
Letras comunes: Sant (24) San (2 4) ==> san

Whole version:

Comienza el proceso de unificacion (July 20, 1998 7:13:43 pm)

Ficheros:
C:\UNITEX\mp1.duv(1)
C:\UNITEX\mp2.duv(2)
C:\UNITEX\mp3.duv(3)
C:\UNITEX\mp4.duv(4)
Estrofa 1:
Versos: 1A, 2A, 3A, 4A
Proceso de unificacion de posiciones (radio 6)
Se unifican la (1) la (2) la (3) la (4)
Se unifican Vida (1) vida (2) Vida (3) vida (4)
Se unifican quieser (1) quieser (3) quieser (4)
Se unifican de (1) de (2) de (3) de (4)
Se unifican Sant (1) Sant (3)
Se unifican Millan (1) Millan (2) Millan (3) Millan (4)
Se unifican saber (1) saber (2) saber (3) saber (4)
Se unifican Qui (2) Qui (3) Qui (4)
Se unifican San (2) San (4)
Proceso de union de palabras (radio union 4, radio comparacion 4)
Se unifican quieser (1 3 4) quieser (2)
Proceso de letras comunes (radio 4, minimos 1, 70%)
Se unifican Quien (1) Qui (2 3 4) ==> qui
Se unifican Sant (2 4) San (2 4) ==> san
Proceso de separacion de palabras (radio 4)