Computerisation of a corpus of personal correspondence spanning the 19th and 20th centuries: the study of linguistic variation

Maria-Pilar Perea

Universitat de Barcelona, Departament de Filologia Catalana,
Gran Via de les Corts Catalanes, 585, E-08007 Barcelona,
mpilar.perea@ub.edu

Članek predstavlja vsebino in tehnične vire CD-ROM-a *Epistolari d'Antoni M. Alcover (1880–1931)* (Korespondenca Antonija M. Alcoverja), objavljenega leta 2008, ki vsebuje celoten prepis osebne korespondence majorškega dialektologa iz obdobja od leta 1880 do leta 1931. Prikazuje, kako lahko računalniški program pomaga pri študiju jezikovnih variacij 3.529.159 besed iz korpusa korespondence. Med mnogimi drugimi lastnostmi posebno pozornost posvečamo narečnim značilnostim in jezikovni interferenci med katalonščino in kastiljščino na začetku 20. stoletja.

The article presents the contents and the technical resources of the CD-ROM "Epistolari d'Antoni M. Alcover (1880–1931)" [Antoni M. Alcover's correspondence], published in 2008, which contains the complete transcription of the Majorcan dialectologist's personal correspondence between 1880 and 1931. It shows how the computer program can help to study the linguistic variation throughout the 3,529,159 words of the corpus of correspondence. Among many other features, we pay special attention to the presence of dialect features and to the linguistic interference between Catalan and Castilian at the beginning of the 20th century.

Ključne besede: jezikovni korpus, računalniško podprto besedilno upravljanje, narečna variacija, jezikovna interferenca

Key words: Linguistic corpus, computer-based text management, dialect variation, linguistic interference

1 Introduction

The CD-ROM "Epistolari d'Antoni M. Alcover (1880–1931)" [Antoni M. Al-cover's correspondence], published in 2008, contains the complete transcription of the Majorcan dialectologist's personal correspondence between 1880 and 1931. Alcover kept 16,005 documents, including letters, postcards, visiting cards and other written materials, 638 of which were written by Alcover himself (drafts, copies or letters that he perhaps never intended to send). The computer program classifies the materials in two ways: according to the letters' authors, and according to the localities from where they were sent. The materials – written in Catalan, Castilian, German, French, Italian, Latin, and Esperanto – not only give a new vision of the life of the dialectologist, his environment, and the culture, politics and society of the end of the nineteenth century and the first third of the twentieth, but also provide rich information about the Catalan language of this period.

The presentation is organised as follows. Section 2 describes the correspondence and section 3 outlines the process of computerisation. The information obtained from this process is explained in section 4. Section 5 includes a small sample of letters in different languages, and section 6 shows the possible applications of the correspondence concerning the study of dialects. Finally, the concluding section provides a summary of the results and offers some suggestions for future research.

2 The correspondence

Before his death, the dialectologist Antoni M. Alcover (Manacor 1860 – Palma 1932)[1] stored his correspondence in 36 wooden boxes. In some cases, he himself wrote the descriptions on the envelopes or folders containing letters from the same author. Other envelopes were later annotated by others (cf. Perea 2009).

The boxes numbered from 1 to 17 include letters, in chronological order, dated between 1882 and 1928. However, some of them were not correctly classified and, in fact, the time interval spans the years 1877 to 1935. In general, these letters were written by members of Alcover's family or other local correspondents of little political, cultural or religious consequence. However, some letters from prominent figures have been found that should have appeared in other boxes. The relative disorder may have been produced by mistakes in the classification or by the consultations made over the years.

The second group of boxes (from 21 to 27) contains letters that are classified not chronologically but alphabetically, by the name of the correspondent. In general, with few variations, the description included in each box is "Cartes

---

[1] For information on Alcover's life and his contributions to dialectology, see Moll (1962, 1983), Massot (1985), Janer Manila (1996), and Perea (2001, 2005, 2006).

a Mn. Alcover que ofereixen especial importancia" [Letters to Father Alcover of special importance].

The third group of boxes (from 46 to 54) contain letters from correspondents considered "personages il·lustres" [illustrious figures] which are also ordered alphabetically.

Box 77 [A-22] includes letters from Alcover's youth, sent to his family and to his uncle, the priest Pere Josep Alcover. Most of these letters are written in the form of poems.[2] This box also contains documents dating from the end of the 19th century, which the dialectologist exchanged with Miquel Binimelis, attorney and alderman of the Town hall of Palma, on the subject of Carlism.[3]

The description of the content of box 132 is: "Collaboradors de l'Obra del Diccionari" [contributors to the Dictionary].[4] In fact, the box contains not only letters from the contributors, but texts from other correspondents whose letters also appear in other boxes [Bernhard Schädel, Leo Spitzer, Wilhelm Meyer-Lübke or even letters by Alcover himself written to different addressees].

Moreover, several boxes containing other materials have also been reviewed. The following ones include misplaced letters: 31 [70 letters], 33 [6], 61 [3], 63 [8], 76 [18], 77 [193], 79 [72], 80 [6], and 83 [3]. These documents provide complementary information.

3 The process of computerisation: methodological aspects

3.1 *The database*

The first stage in the computerisation process involved the scanning and transcription of the letters. Once transcribed, checked and compared with the original ones, the letters were entered into an *ad hoc* database for classification (Figure 1).

Figure 1 shows the screen used to classify the documents. The information required was related with the name of the author – including, whenever possible, a brief biographical description – the number of the box and the document, the language used, the type (letters, postcards, visiting cards, etc.), the description written on the envelope or folder, the size, the possible problems that the letter might present (i. e., when the date, the author or the addressee were unknown). Other information was also recorded, for instance the special characteristics of the document, the existence of a scanned reproduction when the letter contained drawings or was difficult to read (especially German letters written in Gothic script), or the indication of the edition of a document

---

[2] Janer Manila (1996) refers to these texts.

[3] Only a photocopy is preserved; the original letters are missing. The letters were published by Benàssar & Fullana (1993).

[4] That is to say: The *Diccionari català-valencià-balear* (DCVB), now on line: http://dcvb. iecat.net/. See Perea (2004).

when it had already been published. The transcription of the document was then included in the database (Figure 2).

Figure 1. The screen used to enter the letters into the database.

Figure 2. Screen showing a transcribed document. This figure corresponds to a letter from the Italian philologist Pier Enea Guarnerio.

Figure 2 shows the transcription of a letter. This document includes three types of marks in order to obtain a complete indexation: anthroponyms (8.451;[5] in fuchsia), toponyms (2.071; in blue), and key words (920; in green).

[5] The numbers given here indicate the number of entries of each index.

3.2 *Making searches*

The program included on the CD-ROM comprises a search screen. The letters are classified alphabetically by author (Figure 3) or by point of origin (Figure 4). The name of each author or place is followed by a figure in brackets which shows the number of letters related to these items.

Figure 3. Classification according to authors' name. The example corresponds to the chronology of letters sent by the German linguist Wilhelm Meyer-Lübke.

Figure 4. Classification according to point of origin.
The figure shows the place names beginning with the letter P appears.
In blue, the chronology of letters sent from Paris is shown.

In figure 3, each label in blue corresponds to a letter of the correspondence (14 documents) that Wilhelm Meyer-Lübke sent to Alcover. On the right there are two documents without dispatch date. The document may be accessed by clicking on each label. Figure 5 shows the transcription of the letter written on 20 November 1920. The left-hand column shows information related to the document and the author.

Figure 5. The transcription of a letter from Meyer-Lübke.

In the upper part of the screen of Figure 3, in orange, four labels classify the total amount of correspondence: 1. *Correspondents → A. M. Alcover* (the set of authors who sent letters to the dialectologist); 2. *A. M. Alcover → Correspondents* (the set of authors who received letters from Alcover; 638 letters have been preserved); 3. *Other correspondents* (the authors of 390 letters which were not addressed to Alcover but which were in his possession); 4. Other addressees (the 171 addressees who received letters from authors other than Alcover; i.e., the opposite of 3).

3.3 *Further searches*

More complex searches applying different criteria are also possible: by author, date, point of origin, subject, language, box, anthroponyms, toponyms, key words, etc. (Figure 6). Moreover, by searching for both free words and sequences a range of linguistic studies, such as lexis, morphology, syntax, etc. can be carried out (see § 5).

Figure 6. Searching according to different criteria – in this case,
according to key words.

4 Other information obtained from the correspondence

The complete computerisation of the correspondence has yielded a great deal of information regarding its contents. We know the names of the authors, their jobs, and the number of men and women or institutions that sent letters. We also know when the letters were received, the number of different types of documents, and the languages used. This information would have remained inaccessible without computerisation.

4.1 *Correspondents*

The 16,005 documents were written by 3,178 different correspondents. Moreover, 638 texts belonging to Alcover have been preserved. Some difficulties were encountered with the identification of the authors: there are 124 unsigned documents or with illegible signatures; in the case of 15 authors we know the first name, but not the surname; four letters are anonymous, and other documents are signed with pseudonyms or only with initials.

4.2 Correspondents' professions

Either by studying the authors' biographies, or through references in the letters, we have been able to identify the professions of the some of Alcover's correspondents. However, there are 563 whose professions remain unknown. Those

identified cover a total of 213 different jobs. The largest proportion are priests (905 members – 35% of the total); we also find writers, attorneys, secretaries, politicians, doctors, presidents and directors of various institutions, mayors, historians, philologists and linguists, journalists, managers, booksellers, teachers, military men, landowners, pharmacists, architects, librarians, publishers, printers, and traders.

4.3 *Sex*

Of the correspondents and addressees, 2,927 are men and 200 women. Fifty-one are groups, such as societies, institutions, companies, etc.

4.4 *Time span*

The letters cover a period of 50 years (from 1880 to 1931). However, four documents do not correspond to this time span: one written in 1877, and the remaining three were sent after Alcover's death, in 1932, 1934, and 1935.

Although we have been able to deduce the dates of some undated documents, 400 letters remain undated.

Figures 7–10 show the number of letters received each year. Figure 7 includes two tables (from 1880 to 1900 and from 1901 to 1931). The division of Figures 8–10 highlights, on the one hand, the increase in the correspondence from 1897 onwards and, on the other hand, the progressive decrease from 1926 to Alcover's death.

Years	Number of documents	Years	Number of documents	Years	Number of documents	Years	Number of documents
1880	4	1893	35	1906	1019	1919	542
1881	16	1894	33	1907	602	1920	503
1882	17	1895	43	1908	654	1921	392
1883	11	1896	71	1909	567	1922	399
1884	20	1897	110	1910	480	1923	290
1885	11	1898	210	1911	649	1924	185
1886	56	1899	272	1912	412	1925	179
1887	18	1900	425	1913	363	1926	230
1888	35	1901	392	1914	301	1927	160
1889	32	1902	508	1915	424	1928	188
1890	70	1903	521	1916	391	1929	194
1891	82	1904	504	1917	366	1930	162
1892	30	1905	727	1918	404	1931	92

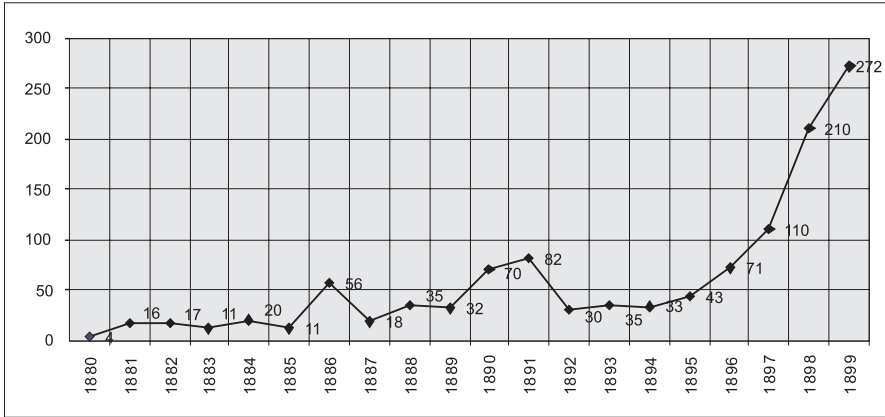Figure 7. Number of documents received each year by Alcover (1880–1931).

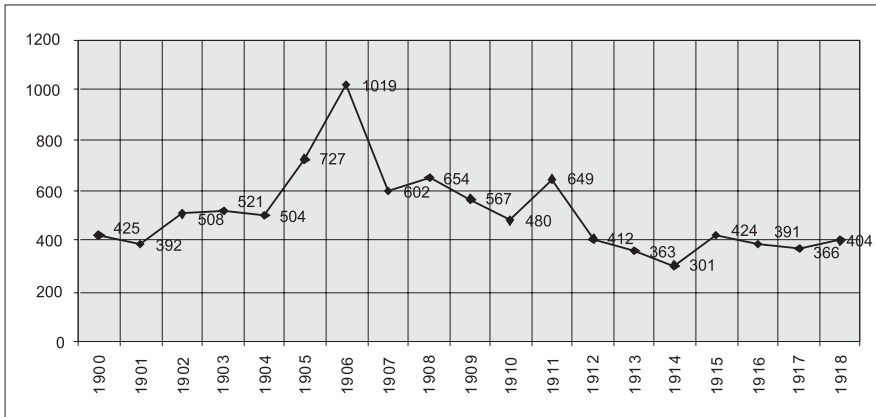Figure 8. Number of documents received between 1880 and 1899.

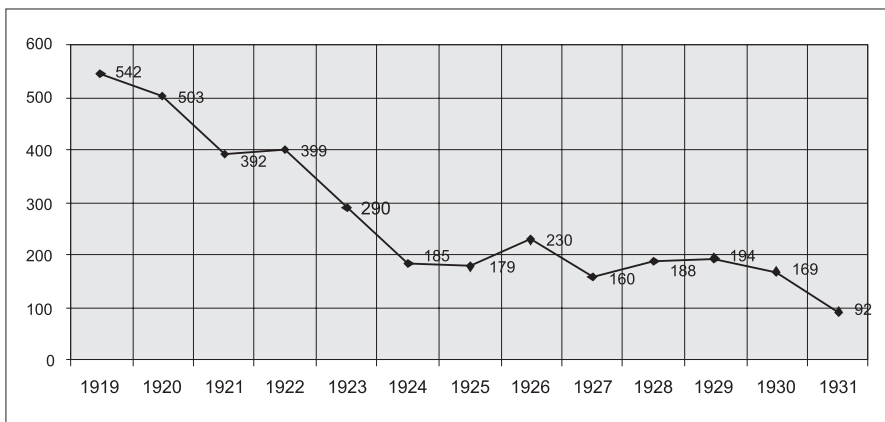Figure 9. Number of documents received between 1900 and 1918.

Figure 10. Number of documents received between 1919 and 1931.

Broadly speaking figures 8–10 each represent one of the three major periods in Alcover's life:

a) The years before his appointment as Majorca vicar general (1898). From 1880 to 1896 he received fewer than a hundred letters a year. The correspondents, during this period of his youth, were friends (Joan Guiraud, Josep Miralles, Francesc Casesnoves, Bartomeu Singala, Pere Orlandis, Miquel Costa, Mateu Rotger, Martí Llobera, and Andreu Pont Llodrà), seminary companions or teachers (Tomàs Forteza or Marià Aguiló). Alcover's interests in the cultural life of the period are reflected by his contacts with several Majorcan writers (Pere d'A. Peña, Miquel dels S. Oliver, Estanislau de K. Aguiló, and Antoni Frates). The first examples of reactionary and Carlist ideologies appear in his correspondence with Miquel Binimelis, the priest Fèlix Sardà i Salvany and the director of the Catholic journal *El Siglo Futuro.*

There is an increasing (though not systematic) progression in the number of letters received. The culminating moments are 1886 (when Alcover became a priest), 1890, 1891, and 1896. Many letters come from his Jesuit brother, Miquel, and, from 1886 onwards, from his farmer brother, Pere-Josep, who was Alcover's most assiduous correspondent over his lifetime (with a total of 659 letters). In 1897 the number of letters received increased when Alcover started the publication of his *Aplech de Rondaies* (Folk-tales) and was even higher in 1898.

b) Alcover's years in the public eye (1899–1920). From 1899 the number of letters received continues to increase, until the celebration of the First International Congress of the Catalan Language (1906). From 1900 to 1905, the average number of letters received is approximately 500 every year, with a peak in 1905 (727), because of the preparation of the congress. Throughout this period, letters from the contributors to the Dictionary (DCVB)[6] are the most numerous. There are also large numbers of letters from Alcover's family, especially during his frequent philological excursions, and there are letters of an ecclesiastical character, due to the post of vicar general which he held until 1916.

c) His last years, which can be divided into two periods: from 1921 to 1926 and from 1927 to 1931. From 1921, the letters received annually do not exceed the figure of 400. Thematically, letters of a political nature responding to Alcover's requests for funding in order to continue the Dictionary (DCVB) predominate. There are also a small number of letters from correspondents. This is the period of Alcover's correspondence with foreign dialectologists and philologists (Louis Gauchat, Jacob Jud, Wilhelm Meyer-Lübke, Leo Spitzer, H. Willy Paulyn, Gerhard Rohlfs, and Maurice Grammont) as a consequence of the publication of the first fascicle of the Dictionary.

As Alcover neared death, the number of letters he received decreased even further, falling to only 92 in 1931. Significantly it was his pupil Francesc de

---

[6] See footnote 4.

B. Moll who kept up most of the correspondence with contributors such as the linguists Alfons Par and Josep Calveras.

4.5 *Points of origin*

The 16,005 documents of the correspondence were sent from 721 localities. However, the point of origin is not specified in 360 letters, and cannot be deduced with certainty.

The places from which the highest number of documents were sent are: Palma de Majorca (3,382),[7] Barcelona (2,369), Manacor (1,650) [Majorca], Madrid (793), Son Crespí Vell (238) [Majorca], Valencia (238), Pollença (221) [Majorca], Halle (190) [Germany], Tortosa (190) [Catalonia], Vic (163) [Catalonia], Sóller (136) [Majorca], Ciutadella (129) [Menorca], Ibiza (129), la Torre Nova (129) [Catalonia], Perpiñán (127) [France], Tarragona (116) [Catalonia], Terrassa (102) [Catalonia], and Gerona (100) [Catalonia].

The distribution of the letters according to area is as follows, in descending order: Balearic Islands [7,406 documents from 134 localities]; Catalonia [4,605 doc. from 233 loc.]; Castile [865 doc. from 18 loc.]; Valencia [568 doc. from 45 loc.]; Germany [485 doc. from 44 loc.]; Italy [231 doc. from 29 loc.]; North Catalonia [176 doc. from 14 loc.]; France [163 doc. from 25 loc.]; Aragon [153 doc. from 13 loc.]; Andalusia [77 doc. from 13 loc.]; Switzerland [60 doc. from 10 loc.]; Argentina [53 doc. from 14 loc.]; Navarre and the Basque Country [33 doc. from 8 loc.]; Belgium [23 doc. from 7 loc.]; Cantabria [19 doc. from 6 loc.]; Cuba [19 doc. from 4 loc.]; Finland [16 doc. from 2 loc.]; Sardinia [14 doc. from 4 loc.]; Austria [13 doc. from 2 loc.]; England [11 doc. from 7 loc.]; Colombia [8 doc. from 1 loc.]; Galicia [6 doc. from 3 loc.]; United States [5 doc. from 2 loc.]; Czech Republic [5 doc. from 1 loc.]; The Philippines [5 doc. from 4 loc.],[8] etc.

4.6 *Typology of the documents*

The correspondence has been classified as follows: letters (11,151), postcards (2,246), visiting cards (993), B.L.M. cards ("Besa la mà") (259), telegrams (158), circular letters (121), notes (117), receipts (71), greeting cards (64), invitation cards ("Saluda") (63), announcements of birth, marriage or of change of residence (47), invoices (38), announcements of death (31), reports (4), and certificates (1).

---

[7] In the program, the name of several places has been transcribed in brackets; these symbols indicate that, though the point of origin does not appear in the document, it can be deduced.

[8] We do not include areas with fewer than 5 documents. This information appears in the CD-ROM.

Other documents have been classified according to the following rubrics: articles (23), index cards (5), speeches (5), word lists (2), and sermons (2). Materials that do not fit into the rubrics indicated are included under the generic name of "other documents" (87).

4.7 *Languages used in the correspondence*

The documents are written for the most part in Catalan (10,177) and in Spanish (4,363). German is the third language (399), followed by Italian (136) and English (48). A very small number of letters were written in Latin (26), in Occitan (14), in Esperanto (2), and in Galician (2).

Other letters are written using two languages: Spanish/Catalan (46), Catalan/Latin (14), German/Catalan (8), German/French (6), French/Catalan (4), German/Spanish (2), English/Spanish (2), Italian/Catalan (2), Catalan/Esperanto (1), Italian/German (1), Italian/French (1), and Latin/French (1).

4.8 *Topics of the letters*

When possible, the main topic (or topics) in the letters is indicated, with relation, in all cases, to Alcover's biography. When available, the description appearing on the envelope containing the letters on the same subject has been used; i.e., the letters received on the occasion of the death of Alcover's mother or the tribute paid to the Majorcan poet Joan Alcover, although documents from other boxes dealing with these topics were also added. In addition, 127 concepts have been defined to indicate the main topic(s) of the letters.

The most frequent topics refer to contributions to the Dictionary (1,599) and to subscriptions to the Dictionary's Bulletin (1,330).

Other numerically prominent[9] topics are: expressions of gratitude (735), philological excursions (623), the Congress of the Catalan Language (595), studies of the Catalan language (506), requests for information (494), requests for subsidies for the dictionary (417), recommendations (324), congratulations (284), illnesses (269), requests or offers (222), the attempt on Alcover's life (201), subscription to the dictionary (201), linguistic questions (198), folk-tales (195), conflicts with the Institut d'Estudis Catalans (195), political questions (174), Alcover's birthday (170), scholars (143), appointments (142), edition of Ramon Llull's works (134), administrative questions (125), requests for favours (107), opinions about the language in general (99),[10] etc.

---

[9] Detailed information concerning all described paragraphs is included in the CD-ROM.

[10] For example, contributors' letters that include linguistic information.

4.9 *Other information*

The computer program also includes a list of materials that were unrelated to the correspondence but were found in amongst the letters: notifications of death, programs of cultural activities, announcements, invitations, religious sheets, religious prints, and other documents.

The letters have been faithfully transcribed. In cases of incomprehensibility or ambiguity, punctuation has been added. Lexical solutions have been placed in brackets in cases of erroneous spelling ("ynemed delene [Magdalena]"). In a few cases, some abbreviations have been clarified. The documents have hardly been modified at all; only major writing errors or typing mistakes have been corrected.

5 Different samples of correspondence

As examples, we include transcriptions of three letters written in English, Italian, and French. The first was written by the philologist E. F. E. Suddard, who was working in the Centro de Estudios Históricos at that time.

[Suddard]
[Contestada 23 setembre 1924]

Centro de Estudios Históricos,
Almagro, 26,
Madrid, Sept. 3rd 1924

Dear Mossèn Alcover,

It is a very long time since I had the honour of hearing from you. I was for some time in Barcelona, and have now been in Madrid for several months. I do not expect to stay here much longer, so that if you care to write me, it would be safer to answer soon; especially in view of the distance and consequently slow communication between here and Palma.

If there is anything I can do for you in Madrid, I shall be most happy to serve you.

In Barcelona I saw Barnils, and several times went to his Institute at Vallvidrera, where he now has a nice laboratory.

I was told your subvention was renewed last year, and hope the same may be the case this year.

As for myself, I have no definite plans for the present, but except to remain in Spain, although not in Madrid, this winter.

I should be glad to know how the Dictionary is progressing, and to have new of yourself and you health.

Believe me; with kind remembrances,
Very respectfully yours,

E. F. E. Suddard

The second letter was written by Italian philologist Pier Enea Guarneiro. This letter is an example of the frequent discussions on linguistic questions that Alcover engaged in with foreign linguists.

[Postcard]
[Contestada 7 mars]

Milano, Donizetti 20
26-II-07

Illustre signore ed amico,

Un mio collega dell' Università desidererebbe sapere se nell' onomastica catalana, o solo majorchina, si abbia un nome di parentela, come "dalceri" o "dalcerto" o "dalcesto" o qualche cosa di simile. Si tratta di un nome che si trova in un' antica carta majorchina e non si riesce a decifrare.
Se Ella può sappermi dire qualche cosa, Le ne sarò molto grato.
Distratto da diverse cure, sia cui la morte del Maestro G. I. Ascoli, il gran glottologo, ho dovuto trascurare le cose catalane, ma presto le riprenderò. Saluto cordiale

P. E. Guarnerio

Al Sr. D. Antoni M. Alcover
Vicari general de Mallorca
Palma de Mallorca
Serra, 13
(Spagna)

Finally, the third letter was written by the dialectologist Jacob Jud and is a sample of the importance and influence abroad of Alcover's linguistic work.

[Postcard]
[Jud]
[Contestada 12 janer 1910]

Zurich 12 XII 10

Très honoré Monsieur,

Permettez-moi de vous remercier sincèrement de l'envoi de "Bolletí del Diccionari de la llengua catalana" que j'ai parcouru en partie déjà avec le plus vif intérêt. Tous ceux qui s'intéressent à l'oeuvre magnifique que vous avez entreprise avec tant de courage désirent ardemment le succès de la grand' oeuvre qui renouvellera la lexicographie catalane si imparfaitement connue aux romanistes. M. Griera qui poursuit ses études avec tant de désintéressement ici à Zurich, vous enverra le "montant" du prix du Bulletin que je lui remettrai l'un de ces jours.
Recevez, Monsieur, mes respects dévoués

J. Jud

Monsieur
Rev. Sr. Antoni Alcover
Serra 13
Palma de Mallorca
Espagne

---

6 The correspondence and the study of dialects

We stated at the beginning that the computer program makes it possible to conduct complex searches according to different criteria: author, date, point of origin, topic, language, or the use of anthroponyms, toponyms or keyword indexes. It is also possible to search for both free words and sequences inside the documents; this feature is particularly useful for carrying out linguistic studies of vocabulary, morphology, syntax, etc.[11]

6.1 *The study of verb morphology*

An example of the possibilities of searching for dialectal forms is found in the appearance of verb forms including a velar extension. Thus, the search for verb endings in "iga" refers us to the documents written by the priest from Ibiza Vicent Serra i Orvai or by the linguist Josep Calveras, whose letters include examples of this sort of ending: the subjunctives *càntiga* (Standard Catalan: *canti* 'I sing'), *estímiga* (St. Cat.: *estimi* 'I love'), *tàpiga* (St. Cat.: *tapi* 'I cover'), *adòbiga* (St. Cat.: *adobi* 'I mend'), *furàdiga* (St. Cat.: *foradi* 'I pierce'), *màtiga* (St. Cat.: *mati* 'I kill') [Figure 11].

Figure 11. The results of a linguistic search in one of the letters in the corpus.

---

[11] About other linguistic studies based on a corpus of correspondence, see Pallander-Collin (1999).

6.2 *Language contact*

Because most of the letters were written before the spelling reform promoted by the Catalan grammarian Pompeu Fabra at the beginning of the 20th century, several samples of natural language and non-normative writing are found. The linguistic interference between Catalan and Spanish in letters written in Catalan is very frequent. Here we have selected a number of words borrowed from Spanish, such as: *sust* or *susto* (Standard Catalan: *espant*; Spanish: *susto* 'fright'), *aconteixement* (St. Cat.: *esdeveniment*; Sp.: *acontecimiento* 'event'), *enfermetat* (St. Cat.: *malaltia*; Sp.: *enfermedad* 'illness'), *assunto* (St. Cat.: *assumpte*; Sp.: *asunto* 'matter'), *barato* (St. Cat.: *barat*; Span.: *barato* 'cheap'), *fallo* (St. Cat.: *fallada*; Sp.: *fallo* 'failure'), *enterro* (St. Cat.: *enterrament*; Sp.: *entierro* 'burial'), *recado* (St. Cat.: *encàrrec*; Sp.: *recado* 'message'), *plas* or *plaç* (St. Cat.: *termini*; Sp.: *plazo* 'term'), *lograr* (St. Cat.: *aconseguir*; Sp.: *lograr* 'to manage'), *inverosímil* (St. Cat.: *inversemblant*; Sp.: *inverosímil* 'improbable'), *despedida* (St. Cat.: *comiat*; Sp.: *despedida* 'farewell'), *tonteria* (St. Cat.: *beneiteria*; Sp.: *tontería* 'silly thing'). Their quantitative incidence can be checked in the documents.

In addition, if we search for the corresponding words in Standard Catalan in the documents, *espant*, *malaltia*, *barat*, *enterrament*, *encàrrec*, *termini*, *aconseguir*, and *comiat* appear, but *esdeveniment*, *fallada*, *inversemblant*, *assumpte* or *beneiteria* do not. This distinction shows the relatively recent appearance of new standard vocabulary with no tradition in the history of the Catalan language.

7 Conclusions

This paper describes the method used to computerise Alcover's correspondence, offering a small sample of its possible linguistic applications and demonstrating its intrinsic and extrinsic value in questions of history, politics, culture, literature, language, and biography. The computer program facilitates the preservation of the materials, and makes it possible to associate them with the letters that Alcover may have written, providing complete textual information and giving a full picture of the personality and works of the Majorcan dialectologist. The linguistic and dialectological study of the 16,005 letters is still to be undertaken. Finally, we are confident that the methodology used in the classification and analysis of the letters can be easily applied to documents of similar characteristics.

REFERENCES

Bartomeu BENNÀSSAR VICENS and Pere FULLANA, 1993: *Carlisme i integrisme a Mallorca (1887–1889): Epistolari entre Antoni Maria Alcover i Miquel Binimelis*, Mallorca: El Tall.

Gabriel JANER MANILA, 1996: *Com una rondalla. Els treballs i la vida de mossèn Alcover*, Palma: Editorial Moll.

Josep MASSOT I MUNTANER, 1985: *Antoni M. Alcover i la llengua catalana*, Barcelona: Publicacions de l'Abadia de Montserrat.

Francesc de B. MOLL, 1962: *Un home de combat (Mossèn Alcover)*, Palma de Mallorca: Moll.

– –, 1983: *Aspectes marginals d'un home de combat (Mossèn Antoni M. Alcover)*, Barcelona: Curial Edicions Catalanes / Publicacions de l'Abadia de Montserrat.

Minna PALLANDER-COLLIN, 1999: Male and female styles in 17th century correspondence: I think, *Language Variation and Change,* 11, 123–141.

Maria-Pilar PEREA, 2001: La metodologia de l'enquesta dialectal d'A. M. Alcover i F. de B. Moll aplicada a La flexió verbal en els dialectes catalans, *Zeitschrift für Katalanistik*, 14, 2001, 98–127.

– –, 2004: "The History of a Multi-dialectal Catalan Dictionary: The 'Diccionari català-Valencià-balear'", in: Julie Coleman & Anne McDermott (eds.), *Historical Dictionaires and Historical Dictionary Research*, Tübingen: Max Niemeyer, 109 –118.

– –, 2005: *Antoni M. Alcover dialectòleg, gramàtic, polemista*, Barcelona: Publicacions de l'Abadia de Montserrat.

– –, 2006: "La incidència de les variables socials en la dialectologia tradicional"; *Randa*, 56, 184–205.

– –, 2008: *Epistolari d'Antoni M. Alcover (1880–1931)*, Palma de Mallorca: Editorial Moll, CD-ROM edition.

– –, 2009: "La publicació de l'*Epistolari d'Antoni M. Alcover (1880–1931)*", *Estudis Romànics*, 31, 279–309.

DIGITALIZACIJA KORPUSA OSEBNE KORESPONDENCE IZ OBDOBJA 19. IN 20. STOLETJA: ŠTUDIJA JEZIKOVNIH VARIACIJ

CD-ROM *Epistolari d'Antoni M. Alcover (1880–1931)*, objavljen leta 2008, vsebuje celotno osebno korespondenco majorškega dialektologa Alcoverja iz obdobja od leta 1880 do leta 1931. Alcover je hranil 16.005 dokumentov, vključujoč pisma, razglednice, vizitke in druga pisna gradiva, od katerih jih je 638 napisal sam (osnutki, izvodi pisem, ki jih najbrž nikoli ni nameraval poslati). Računalniški program gradivo razvršča na dva načina: (1) glede na avtorja pisma in (2) glede na lokacijo, kamor so bila poslana. Gradiva, napisana v katalonščini, kastiljščini, nemščini, francoščini, italijanščini, latinščini

in esperantu, ne le, da nudijo nov pogled na dialektologovo življenje, njegovo okolje in kulturo, politiko ter družbo s konca devetnajstega in prve tretjine dvajsetega stoletja, temveč ponujajo tudi bogate informacije o katalonskem jeziku tistega časa. Članek predstavlja vsebino tehničnih virov CD-ROM-a in kaže, kako lahko računalniški program pomaga znanstveniku pri študiju lingvističnih variant 3.529.159 besed iz korpusa korespondence. Med drugim je posebna pozornost posvečena jezikovnim interferencam med katalonščino in kastiljščino z začetka 20. stoletja.