ALT Explored:
Integrating an Online Dialectometric Tool and an Online Dialect Atlas

Martijn Wieling[+], Eva Sassolini[*], Sebastiana Cucurullo[*], Simonetta Montemagni[*]

[+]University of Groningen, CLCG – Netherlands

[*]Istituto di Linguistica Computazionale "Antonio Zampolli" – CNR – Italy

E-mail: m.b.wieling@rug.nl, {eva.sassolini,nella.cucurullo,simonetta.montemagni@ilc.cnr.it}

Updated (corrected links after Gabmap update): 2021-08-18

Abstract

In this paper, we illustrate the integration of an online dialectometric tool, *Gabmap*, together with an online dialect atlas, the *Atlante Lessicale Toscano* (ALT-Web). By using a newly created url-based interface to *Gabmap*, ALT-Web is able to take advantage of the sophisticated dialect visualization and exploration options incorporated in *Gabmap*. For example, distribution maps showing the distribution in the Tuscan dialect area of a specific dialectal form (selected via the ALT-Web website) are easily obtainable. Furthermore, the complete ALT-Web dataset as well as subsets of the data (selected via the ALT-Web website) can be automatically uploaded and explored in *Gabmap*. By combining these two online applications, macro- and micro-analyses of dialectal data (respectively offered by *Gabmap* and ALT-Web) are effectively and dynamically combined.

Keywords: dialectometry, digital dialectal resources, interoperability of tools and resources

1. Introduction

Computational techniques in the humanities have a long history, dating back to Father Roberto Busa in the late 1940s. Since then, research in the area, initially dubbed as "Humanities computing", was focused on the development of formalisms suitable for rich text representation (see e.g., the Text Encoding Initiative, or TEI) on the one hand, and the design and implementation of software tools for a wide range of different kinds of textual analysis on the other hand. These two trends were combined in such a way that textual data augmented with annotations were fully searchable by developed tools. Although computational analysis techniques allowed to explore data from different perspectives, under this approach the focus of analysis typically remained on individual data (typically words). Adopting Moretti's (2013) terminology, we will refer to this type of analysis as "close reading", which represents a central practice in humanities studies, based on the study of individual features contained in a text as well as their variation and history.

Nowadays, the field of "Humanities computing" is increasingly known as "Digital Humanities". This terminology change is associated with a shift of emphasis from the analysis of individual data to data mining techniques enabling the identification of patterns in large (text) collections. The growth in size and accessibility of digital databases combined with advances in data mining have opened up new ways of creating meaning through what in Moretti's words is called "distant reading". Distant reading explicitly ignores the individual features that close reading concentrates on, in favour of detecting large-scale trends, patterns and relationships from wide collections of data. These patterns are not discernable from a single text or detailed analysis. As Burdick et al. (2012) claim, distant reading is therefore not just a "digitization" or "quickener" of classic humanities methodologies. It is, rather, a new way of doing research wherein computational methods allow for novel sets of questions to be posed and answered.

Whereas close reading software tools are within reach of traditional scholars of the arts and humanities, this is not typically the case with distant reading tools. Reasons underlying this state of affairs range from the quite usual scepticism of researchers in humanities towards quantitative and empirical analysis techniques to the difficulties inherent in the preparation and the computational processing of data.

Recently, in the Digital Humanities literature, distant reading is no longer contrasted with close reading, but they are seen as complementary perspectives combining the top-down global view obtained through distant reading with the more traditional local view of individual features characteristic of close reading. In this paper, we demonstrate how a combined view of close and distant reading can open new research perspectives and answer new questions. This claim will be exemplified with the analysis of dialectal data. In particular, the paper presents the integration of an online dialect atlas together with an online application for dialectometric analyses. By integrating these two systems, the user is not only able to conduct a qualitative fine-grained analysis of the data via the online dialectal resource, but is also able to effortlessly obtain a quantitative aggregate view of (subsets of) the data.

The dialect atlas is the *Atlante Lessicale Toscano* (ALT; Giacomelli et al., 2000), a specially designed linguistic atlas focusing on dialectal variation within Tuscany, a region which has a special status in the complex puzzle of linguistic variation in Italy. A few years after its construction, ALT was made available as an online resource, called ALT-Web,[1] providing close reading

[1] http://serverdbt.ilc.cnr.it/altweb/

functionalities to inspect the data at the level of individual question items and/or individual locations (Cucurullo et al., 2006). No facilities to carry out aggregate analyses of the data were included in either of the two versions.
The distant reading functionalities are provided by the online application *Gabmap* (Nerbonne et al., 2011; Leinonen et al., forthcoming), which has been developed by the University of Groningen in the framework of CLARIN-NL as an accessible opensource web application to analyze language variation data. *Gabmap* allows both aggregate dialectometric analyses and data inspection at the level of the individual items.
In the following sections, we illustrate how we have integrated ALT-Web and *Gabmap* and highlight the functionality of the resulting system, dubbed *ALT Explored*. With *ALT Explored*, linguists and dialectologists are able to explore, quantitatively and qualitatively analyze and visualize the ALT data in order to support their research. From a more general perspective, this study can also be seen as an example of integration of resources and services within the pan-European CLARIN ERIC infrastructure. Whereas *Gabmap* is already part of CLARIN, ALT-Web will be integrated soon, thanks to the recent participation of Italy in the project.
The paper is organized as follows. Section 2 introduces the ALT atlas and its online counterpart in detail, and discusses the functionality of *Gabmap*. Section 3 explains the steps necessary to pre-process the ALT data for use in *Gabmap*, while Section 4 discusses the functionality of the integrative system. Finally, Section 5 ends with some conclusions.

2. Background

2.1 Atlante Lessicale Toscano

The *Atlante Lessicale Toscano* (ALT) is a specially designed linguistic atlas in which dialectal data have both a diatopic and diastratic characterization. The adjectives qualifying this linguistic atlas in its name are “lexical” and “Tuscan”. ALT is lexical in the sense that its main focus is on lexical variation. ALT is Tuscan in the sense that it is a regional atlas focusing on dialectal variation within Tuscany, a region where both Tuscan and non-Tuscan dialects are spoken. Specifically, Gallo-Italian dialects are spoken in Lunigiana and in small areas of the Apennines.
The interviews of the *Atlante Lessicale Toscano* (ALT; Giacomelli et al., 2000) were conducted between 1974 and 1986 in 224 localities in Tuscany, with 2,193 informants selected with respect to various parameters, including age, socio-economic status, education and culture. Tuscan dialects are spoken in 213 out of the 224 investigated locations (with a total of 2,060 informants). The interviews were carried out by a group of trained fieldworkers who employed a questionnaire of 745 target items, designed to elicit variation mainly in vocabulary, semantics and phonetics. A dialectal corpus with these features lends itself to investigations concerning geographic or horizontal (diatopic) variation as well as social or vertical (diastratic) variation (see Wieling et al., 2014 for an example). *ALT Explored* focuses on both dimensions of lexical variation.
During ALT interviews, informants were asked two main types of questions: onomasiological questions starting from concepts and looking for the lexical items designating them (a typical onomasiological question is “How is this concept designated or named?”), and semasiological questions starting from word forms and asking for their meanings (a typical semasiological question is “Which meanings does this word have?”). Each informant responded to a total of 460 onomasiological (eliciting a lexical item for a concept) and 285 semasiological (eliciting a meaning for a word form) questions. In what follows, we will focus on the subcorpus of answers to onomasiological questions.

2.2 ALT-Web

ALT-Web is the online version of ALT: it contains the digitized responses to all 745 onomasiological and semasiological questions. In ALT-Web, all dialectal responses are assigned different levels of representation, with a first level rendering the original phonetic transcription and other levels containing normalized representations of the original form encoded in standard Italian orthography.
In this multi-level representation model, dialectal data are encoded in layers of progressively decreasing detail going from phonetic transcription to different levels of normalized representations abstracting away from details of speakers’ pronunciation. For the phonetic transcription, a geographically specialized version of the “Carta dei Dialetti Italiani” (CDI) transcription system (Grassi et al. 1997) was used, henceforth referred to as CDI-ALT. In what follows, we will focus on the phonetic and normalized representation levels. Concerning the latter, the most abstract normalized representation was selected, i.e. the representation level abstracting away from productive phonetic variation within Tuscany and mainly reflecting lexical and (limited) morphological variation.
ALT-Web provides flexible and dynamic query facilities which permit the user to interactively access the corpus of dialectal data and to navigate through it on the basis of his or her research interests. Information can be accessed and retrieved on the basis of a wide range of parameters which can be flexibly combined. For example, lexical data can be queried on the basis of the question, the location, and/or the socio-economic features of the informant(s). Furthermore, questions can be filtered on the basis of their relevance with respect to a given semantic field or linguistic register. ALT-Web also contains a basic geographical visualization facility providing for each answer presence/absence maps: frequency information is not taken into account in this visualization. For more details about ALT-Web, the interested reader is referred to Cucurullo et al. (2006).

2.3 Gabmap

Gabmap (Nerbonne et al., 2011; Leinonen et al., forthcoming) is an online web application for conducting dialectometric analyses. Besides enabling a researcher to conduct aggregate analyses (such as cluster analysis, or visualizing the data using multidimensional scaling) of dialect data, *Gabmap* also offers facilities for visualizing and inspecting the original data. All visualizations can be downloaded as vector or bitmap graphics. *Gabmap* has been positively evaluated by Snoek (2014: 206): "Gabmap is excellent software that permits the mapping and comparison of linguistic data in a fast and generally painless manner".

To use *Gabmap*, a free account needs to be created. After logging in, language variation data can be uploaded and the results of the dialectometric analysis can be inspected. The data file should contain the lexical variants provided as answers to the questionnaire items (possibly in phonetic transcription) in each location. In addition, a mapping file (i.e. a "kml" file) with the geographical coordinates of each location as well as the outline of the dialect region is necessary to adequately visualize the results on a map.

The aggregate analyses in *Gabmap* are based on an underlying distance matrix between all *N* pairs of locations (i.e. an *N* x *N* matrix). These distances can be calculated on the basis of binary distinctions (same or different), but are more frequently based on the Levenshtein distance (i.e. the edit distance; Levenshtein, 1965) which is equal to the minimum number of insertions, deletions and substitutions to transform one phonetically transcribed pronunciation to another. Dialect distances between each pair of locations are obtained by averaging the distances for each individual linguistic item.

Gabmap offers several analyses on the basis of this distance matrix. A first analysis offered by *Gabmap* is cluster analysis. After a user has selected a specific clustering algorithm and determined the number of desired clusters, *Gabmap* shows the resulting cluster map, with locations in the same cluster having the same color. As cluster analysis is inherently instable and subjective (i.e. how many clusters to choose), *Gabmap* also allows the user to obtain a probabilistic (or fuzzy) clustering, where each location has a certain probability of being in a cluster with other locations.

A useful alternative to clustering offered by *Gabmap* is multidimensional scaling, which takes advantage of the structure in the original distance matrix, to reduce the distance matrix to an *N* x 3 matrix. The values in the three columns can then be mapped to the colors red, green and blue. In this way each location can be assigned a specific color. Locations (i.e. dialects) which are linguistically similar will have a similar color.

Besides these two types of analyses, there are various other options implemented in *Gabmap*, including the inspection of the data via distribution maps (showing the relative frequency of a form in a location), the visualization of difference maps (i.e. lines connect each pair of locations, the thickness of the line indicates the similarity between the locations), and the visualization of reference point maps (visualizing the differences from the perspective of a pre-selected location) à la Goebl (2010). Some additional information about many of these options is discussed by Wieling (2013), Snoek (2014) and Leinonen et al. (forthcoming).

3. Data Pre-processing

At the time when the digital editions of ALT were designed and developed, the use of the CDI-ALT phonetic representation was the only option, in line with the Italian tradition of dialectological studies which preferred the CDI transcription system with respect to the International Phonetic Alphabet (IPA). Nowadays, this choice needs to be revised for both usability and technical reasons. First, to make the ALT corpus usable by the wider international community of dialectologists and linguists who may not be familiar with the CDI notation. Secondly – and primarily for the specific concerns of this study – to guarantee the interoperability of ALT-Web and *Gabmap* resources and tools. The CDI-ALT phonetic representation was thus converted to the International Phonetic Alphabet (IPA) representation. This conversion was done automatically on the basis of 158 ordered conversion rules encoded as PERL regular expressions.

In ALT, phonetically transcribed data were represented through a hybrid encoding schema including both compositional and atomic representations which, depending on the processing task, were automatically converted into each other (Montemagni and Paoli, 1989-90; Cucurullo et al., 2006). Compositional representations (third column in the Appendix table) encode each phonetic symbol with a basic sign which can be further specified through diacritics (conveying information e.g., about stress or nasality of vowels). Atomic representations (second column in the Appendix table) show a 1:1 correspondence between CDI-ALT phonetic symbols and computer codes. The CDI-ALT compositional representation was taken as the starting point for the conversion to IPA. In the Appendix, the correspondence table between the CDI-ALT and IPA notations is reported. It is clear that in most cases a one-to-one correspondence was found, with the following exceptions:

- a CDI-ALT phonetic segment combined with one diacritic corresponds to an individual IPA phonetic segment (e.g. [À] > [ɛ]);
- an individual CDI-ALT phonetic segment is converted into a IPA phonetic segment combined with a diacritic (e.g. [Ø] > [tʲ]).

Interestingly enough, there are three different cases (highlighted in grey in the correspondence table) in which two different CDI-ALT segments are assigned the same IPA representation: this is the case of the weakened realization of palatal affricates, e.g. [â] and [®], whose IPA representation coincides with the representation of the [Ç] and [Ð] CDI-ALT segments, i.e. the voiceless and voiced postalveolar fricatives [ʃ] and [ʒ].

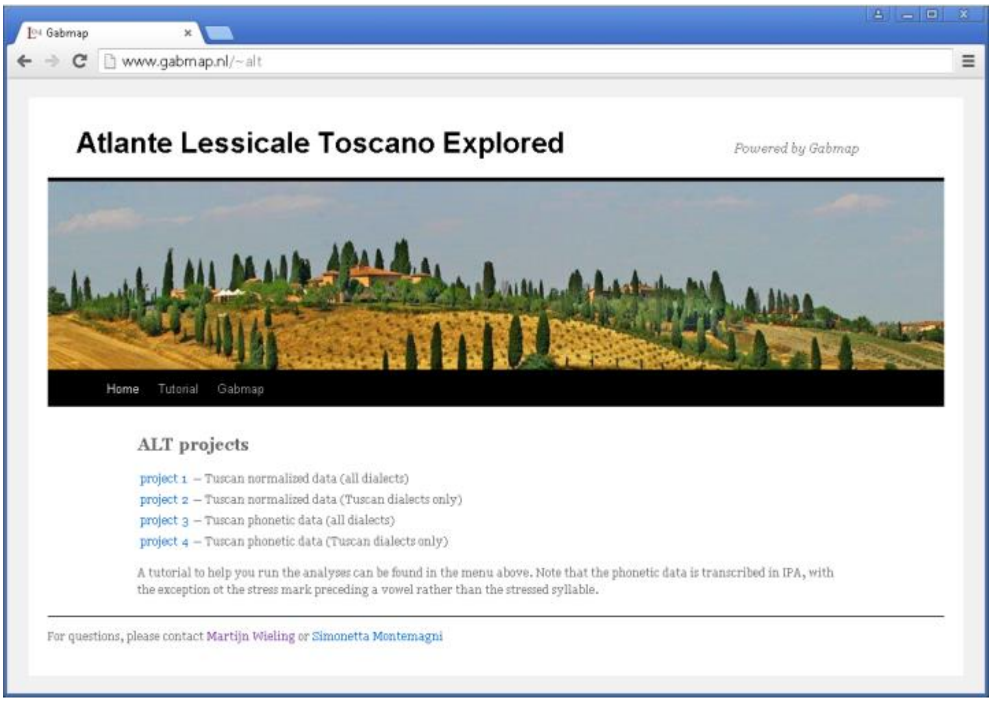

Figure 1: Four custom *Gabmap* projects in *ALT Explored*

Our final remark with respect to the conversion involves the strategy we employed to mark word stress. In IPA, word stress is marked by inserting a vertical line (ˈ) before the stressed syllable in the phonetic transcription of the word, while in the CDI-ALT notation the stress is associated with the vowel in the stressed syllable. Due to the difficulty of automatically segmenting the word in syllables, we decided to follow the CDI-ALT stress notation, i.e. to insert the stress mark (i.e. the vertical line) before the vowel in the stressed syllable, also in the IPA notation.

4. Design and functionality of ALT Explored

After careful analysis of the ALT data, and the way they are organized and indexed within ALT-Web, we identified two integration possibilities with *Gabmap*. On the one hand, a new access functionality to the ALT dialectal corpus was designed and implemented, making it possible to carry out dialectometric analyses of the whole dataset, or user-defined subsets based on linguistic and extra-linguistic filters. On the other hand, the ALT-Web visualization of distributional data was extended with a new map type accounting for relative frequency of items within investigated locations. In this case, the integration with *Gabmap* is carried out at the level of individual queries.

As *Gabmap* is available as opensource software, we have customized the original source code to allow for the integration with ALT-Web along the lines depicted above. The changes can be classified in three categories, which are illustrated in the following subsections.

4.1 Custom Gabmap projects

The first change involved the creation of four custom Gabmap projects on the basis of the dataset containing all answers to the onomasiological ALT questions. We created four projects to represent both the phonetically transcribed data (converted to IPA) and the normalized data, both for all locations in the ALT dataset (224) and for the subset of 213 locations where a Tuscan dialect is spoken. For this purpose, the ALT-Web data was converted to the tabular form required for *Gabmap* and the appropriate mapping files including the locations were created. In each location, the phonetic transcription (or normalized form) of each individual speaker was retained. Note that a relatively similar customization of *Gabmap* has been made for another dataset focusing on lexical variation in contemporary English dialects, the *BBC Voices* data (i.e. BBC Voices explored; Wieling, 2013).

The four individual projects can be accessed directly via the links at https://gabmap.nl/alt (see Figure 1) or from the ALT-Web main navigation page, where this is proposed as a third access functionality aimed at providing the results of aggregate analyses of the ALT corpus: this new access functionality is meant to integrate and extend the basic and advanced access close reading functionalities already present in ALT-Web.

4.2 Inspection of custom-generated data

The second change involved creating another direct url-based interface to *Gabmap*. The purpose of this interface was to allow a user to inspect language variation data in (a temporary account in) *Gabmap*, without the need for creating an account or manually uploading data

(but instead by simply clicking a link). Clearly, this interface facilitates the integration of *Gabmap* in online repositories of dialect atlas data, such as ALT-Web. For example, ALT-Web is currently in the process of being adapted to allow users to analyze custom subsets of ALT data in *Gabmap*. Each subset is created on the basis of setting specific filters. The current filters allow the user to specify the age range, the province(s) in which the locations should be located, and the semantic field(s) of the questions included: in the near future filters could be extended to include gender, job type and education level of the speaker. After setting one or more filters, the ALT-Web application creates the appropriate *Gabmap* datafile (where phonetic transcriptions are converted to IPA), stores it on the ALT-Webserver and generates a direct link to *Gabmap*. When the user clicks on this link, the data is uploaded to *Gabmap*, and the resulting analyses can be directly inspected in a temporary *Gabmap* account. As an example, consider the following url (from ALT-Web) to display the data with respect to the semantic field of 'Agriculture':
https://gabmap.nl/bin/accountALT?action=ALT&username=_tmpALT&description=Agriculture&kmlurl=https://gabmap.nl/resources/Tuscany-all.kml&dataurl=http://dbtvm1.ilc.cnr.it/proveALT/MAT/Matrix_fon_agr_ALL.txt.

4.3 Integration of Gabmap distribution maps

The third change involved the extension of the *Gabmap* application by allowing direct (url) access to the distribution maps. Consequently, when the user supplies both the question item and the variant as url parameters, the corresponding distribution map is shown visualizing the relative frequency of use of the variant throughout Tuscany. This functionality has been integrated in ALT-Web as part of the basic access functionality. In accessing the data, the user can choose one of two search paths, corresponding to the typical access keys to the data of a linguistic atlas: the questionnaire item through which the dialectal word was elicited, or the locality in which it was witnessed. Regardless of the type of selection, the final result is a list of entries satisfying the user request. For each dialectal variant in the list, it is possible either to look up the corresponding entries in the ALT-Web database or to project the result onto a map. If the *Gabmap* distribution map is selected, the result is visualized within *Gabmap*. Figure 2 shows an example of this type of map. The darkness of the color in the map reflects the relative frequency of the selected variant in each location. In contrast, the original ALT-Web visualization only distinguished presence and absence of a given answer, regardless of the number of informants who attested it.
As the ALT-Web application uses a different transcription system than *Gabmap*, the CDI-ALT to IPA conversion script (see Section 3 above) is used here to convert the CDI-ALT transcription of individual items (transmitted via the url) to the IPA transcription used in *Gabmap*. For example, the CDI-ALT compositional representation [suSi8na] (a dialectal variant of *susina*, 'plum') would be converted to [suzˈina] in IPA. Figure 2 visualizes the distribution map of this variant.[2]

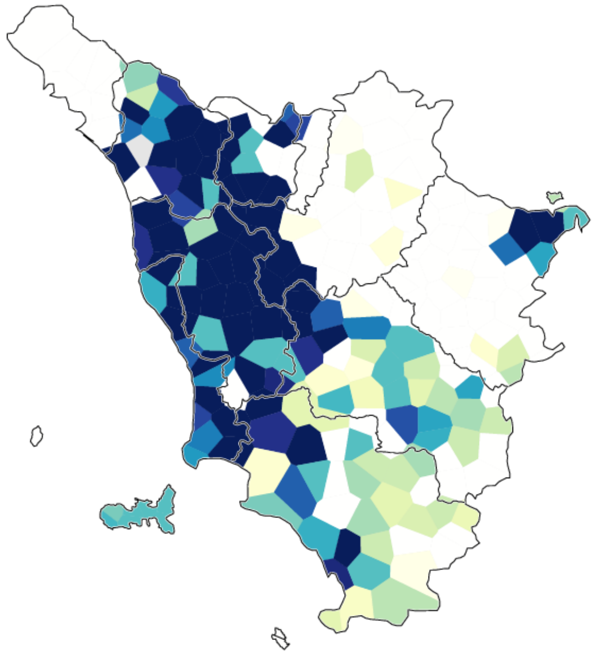

Figure 2: Distribution map of [suzˈina] for the concept *susina*, 'plum'. Darker colours indicate a greater relative frequency of use

5. Conclusion

The main contribution of this work can be summarized as follows. From the perspective of online digital dialectal resources such as ALT-Web, the integration with *Gabmap* in *ALT Explored* provides the opportunity to conduct quantitative analyses which were not possible before. From the perspective of *Gabmap*, the flexible extensions represent a unique opportunity to become the visualization and exploration tool for an entire linguistic atlas.
While the two online applications are both useful in their own right, their integration becomes more than the simple sum of the two parts. Within *ALT Explored*, the dynamic combination of macro- and micro-analysis, or to put it in Moretti's words of "close and distant reading", allows the dialectologist to alternate between separate views of the data, zooming in and out, searching for large-scale patterns and then focusing on fine-grained analyses.
The combination of close and distant reading of textual data is becoming more and more important in the field of digital humanities. To our knowledge, our contribution represents the first attempt in this direction in the sub-area of digital dialectology.

[2] This distribution map can be viewed online by using the following url:
https://gabmap.nl/bin/accountALT?username=altdemo&action=ALTDISTMAP&project=3&var=suSi8na&convert=1&item=101-susina#fig.

Acknowledgements

The research reported in this article was carried out in the framework of the Short Term Mobility program of international exchanges funded by the National Research Council (CNR, Italy). We are grateful to Çağrı Çöltekin, Marc Kemps-Snijders, Jan Pieter Kunst, and Juan Bontje for their help with configuring *Gabmap*. Thanks are also due to Elisabetta Carpitelli, for valuable advice on issues related to the CDI-ALT to IPA conversion.

References

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., Schnapp, J. (2012). *Digital Humanities*. MIT Press.

Cucurullo, N., Montemagni, S., Paoli, M., Picchi, E., Sassolini, E. (2006). Dialectal resources on-line: The ALT-Web experience. *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genova, 1846–51.

Giacomelli, G., Agostiniani, L., Bellucci, P., Giannelli, L., Montemagni, S., Nesi A., Paoli M., Picchi E., Poggi Salani T. (2000). *Atlante lessicale Toscano*. Roma: Lexis Progretti Editoriali.

Goebl, H. (2010). Dialectometry: Theoretical pre-requisites, practical problems and concrete applications (mainly with examples from the Atlas Linguistique de la France 1902-1910). *Dialectologia*. Special issue, I, Geolinguistics around the World, 63-77, http://www.publicacions.ub.edu/revistes/dialectologia SP2010/.

Grassi, C., Sobrero, A., Telmon, T. (1997). *Fondamenti di Dialettologia Italiana*, Roma-Bari, Laterza.

Leinonen, T., Çöltekin, Ç., & Nerbonne, J. (forthcoming). Using Gabmap. *Lingua*.

Levenshtein, V. (1965). Binary codes capable of correcting deletions, insertions and reversals. Doklady *Akademii Nauk SSSR*, 163:845–848. In Russian.

Moretti, F. (2013). *Distant Reading*. London, Verso.

Montemagni, S., Paoli, M. (1989/1990). *Dalla parola al bit (e ritorno): percorsi dall'inchiesta sul campo alla banca dati dell'ALT*. *Quaderni dell'Atlante Lessicale Toscano*, VI/VIII, pp. 7-52.

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., Leinonen, T. (2011). Gabmap – a web application for dialectology. *Dialectologia: revista electrònica*, SI II, 65-89.

Snoek, C. 2014. Review of Gabmap: Doing Dialect Analysis on the Web. *Language Documentation & Conservation*, 8, 192-208.

Wieling, M. (2013). Voices dialectometry at the University of Groningen. In: Clive Upton and Bethan Davies (eds.) *Analysing 21st-century British English: Conceptual and methodological aspects of the BBC 'Voices' project*. London: Routledge, pp. 208-218.

Wieling, M., Montemagni, S., Nerbonne, J., Baayen, R.H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language*, 90(3), 669-692.

Appendix 1

CDI-ALT to IPA Conversion Table

	CDI-ALT	CDI-ALT composit. represent.	IPA	Notes
1.	ə	@	ə	
2.	ə́	@8	ˈə	
3.	a	a	a	
4.	ä	a3	æ	
5.	ä̃	a37	æ	
6.	ä̃́	a378	ˈæ	
7.	ä́	a38	ˈæ	
8.	ã	a7	a	
9.	ã́	a78	ˈa	
10.	á	a8	ˈa	
11.	b	b	b	
12.	ƀ	b5	β	
13.	č	C	ʧ	
14.	ḉ	C5	ʃ	See n.106
15.	d	d	d	
16.	ď	D	dʲ	
17.	đ	d5	ð̆	
18.	e	e	e	
19.	e̤	e0	e	
20.	ẽ̤	e07	e	
21.	ẽ̤́	e078	ˈe	
22.	é̤	e08	ˈe	
23.	ẹ	e1	e	

	CDI-ALT	CDI-ALT composit. represent.	IPA	Notes
24.	ẹ̃	e17	e	
25.	ẹ̃́	e178	ˈe	
26.	ẹ́	e18	ˈe	
27.	ȩ	e2	ɛ	
28.	ȩ̃	e27	ɛ	
29.	ȩ̃́	e278	ˈɛ	
30.	ȩ́	e28	ˈɛ	
31.	e̯	e4	e̯	
32.	é	e8	ˈe	
33.	f	f	f	
34.	ğ	G	ʤ	
35.	g	g	g	
36.	ǥ	g5	ɣ	
37.	ǥ́	G5	ʒ	See n.107
38.	ġ	g6	gˠ	
39.	h	h	h	
40.	i	i	i	
41.	i̧	i2	i	
42.	ĩ̧	i27	i	
43.	ĩ̧́	i278	ˈi	
44.	í̧	i28	ˈi	
45.	i̥	i4	j	
46.	ĩ	i7	i	
47.	ĩ́	i78	ˈi	
48.	í	i8	ˈi	
49.	ğ̃	J	gʲ	
50.	č̃	j	kʲ	
51.	k	k	k	

	CDI-ALT	CDI-ALT composit. represent.	IPA	Notes
52.	kʼ	k5	x	
53.	l	l	l	
54.	lʼ	L	ʎ	
55.	ł	l6	ʟ	
56.	m	m	m	
57.	n	n	n	
58.	ń	N	ɲ	
59.	ṇ	n1	ɳ	
60.	ṅ	n6	ŋ	
61.	o	o	o	
62.	o̤	o0	o	
63.	õ̤	o07	o	
64.	ṍ̤	o078	ˈo	
65.	ó̤	o08	ˈo	
66.	ọ	o1	o	
67.	ọ̃	o17	o	
68.	ọ̃́	o178	ˈo	
69.	ọ́	o18	ˈo	
70.	o̧	o2	ɔ	
71.	õ̧	o27	ɔ	
72.	ṍ̧	o278	ˈɔ	
73.	ó̧	o28	ˈɔ	
74.	ö	o3	ø	
75.	ö̃	o37	ø	
76.	ö̃́	o378	ˈø	
77.	ö́	o38	ˈø	
78.	ó	o8	ˈo	
79.	o̧̧	o9	ɔ	

	CDI-ALT	CDI-ALT composit. represent.	IPA	Notes
80.	ǫ̨́	o98	ˈɔ	
81.	p	p	p	
82.	ᵽ	p5	ɸ	
83.	ḍ	R	ɖ	
84.	r	r	r	
85.	ṛ	r1	ɽ	
86.	s	s	s	
87.	S	S	z	
88.	t	t	t	
89.	tʹ	T	tʲ	
90.	ŧ	t5	θ	
91.	u	u	u	
92.	u̧	u2	u	
93.	ü̧	u27	y	
94.	ǘ̧	u278	ˈu	
95.	ú̧	u28	ˈu	
96.	ü	u3	y	
97.	ü̃	u37	y	
98.	ü̃́	u378	ˈy	
99.	ǘ	u38	ˈy	
100.	u̯	u4	w	
101.	ũ	u7	ũ	
102.	ṹ	u78	ˈũ	
103.	ú	u8	u	
104.	v	v	v	
105.	ž	w	ʑ	See n.109
106.	š	x	ʃ	See n. 14
107.	ʃ̆	X	ʒ	See n. 37
108.	ś	x5	ɕ	
109.	´ʃ	X5	ʑ	See n.105
110.	∂	Y	ɟ	
111.	ẓ	Z	dz	
112.	z	z	ts	