

LANGUAGE AND COMMUNICATION TECHNOLOGY COLLOQUIA

Rovereto, November 9th 2006

THE ALTR EXPERIENCE: SOME QUESTIONS AND PROPOSALS OF AN ELECTRONIC VERSION OF DIALECTAL DICTIONARIES

Patrizia Cordin

Facoltà di Lettere e Filosofia. Via S.Croce, 65 - 38100 Trento
patrizia.cordin@lett.unitn.it

ABSTRACT

In this paper I will present the main criteria that have been followed in the planning and in the partial realization of the *ALTR – Archivio lessicale dei dialetti trentini*, a lexical database (now published in CD-Rom version), based on the different linguistic varieties (Lombard, Venetian, Ladin) spoken in Trentino. The whole corpus selected for the *ALTR* consists in more than thirty dictionaries and atlases referring to thirteen linguistic areas of the province. A standard form for each entry is proposed, to illustrate how the electronic hyper-dictionary (integrated dictionary) can maintain the original data of the sources and, at the same time, can give uniformity to their presentation. In the transfer from print dialectal dictionaries to electronic ones some problems arise. Here I will briefly discuss three topics in particular: a) phonetical transcriptions, b) Italian translations of dialectal headwords, c) lexical field markers. In the conclusion, I will present some possible paths for searching in the *ALTR* dictionary.

1. The project

The project of the *ALTR*, which is now partially realized with about 35.000 entries and published in a CD-Rom, has started six years ago¹. The goal of the research was the proposal of an electronic “dictionary of dictionaries” for the Trentino area. In fact, many dialectal dictionaries have been published, above all in the last thirty years²: some of them present the lexicon of a single village (or town), others the lexicon spoken in one or two valleys. The dictionaries selected as sources for our database are about thirty. The input of the CD-Rom is constituted by only five dictionaries, that cover a relevant part of the province; they are: Aneggi 1984, Groff 1955, Prati 1960, Quaresima 1964, Tissot 1976. Prati and Quaresima dictionaries, are quite complex, in particular for the great amount of sending from one entry back to another one. Therefore the input from these dictionaries is only partial (from A to C, plus many entries which are sent back).

The richness of the production implies a great variety of lexicographic criteria: each author focuses different aspects and chooses a different form for the presentation of the data. As a result, it is very difficult to compare the various works. Therefore, in projecting our “archive” it was necessary to think how to fill some gaps of the original works, and above all, how to give uniformity to such different texts, in order to allow for a flexible way of searching, which is impossible with the print versions of the dictionaries.

2. The parts of the lexical record.

¹ The *ALTR* collaborators are: Patrizia Cordin, coordinator, Ilaria Adami, Serenella Baggio, Stefano Bernardini, Giovanni Bonfadini, Lidia Floess, Tiziana Gatti, Corrado Grassi, Gabriele Iannaccaro, Antonella Mott, Daniele Rando, Lotte Zoerner.


² The first dialectal vocabulary in Trentino was published in 1856 by Giovan Battista Azzolini. It presents the lexicon of Rovereto variety.

The uniformity of the presentation and the possibility of comparing the data have been reached thanks to a common lexical record, where each entry is organized in different parts; it represents the base unity of the *ALTR*.

ARCHIVIO LESSICALE DEI DIALETTI TARENTINI

Nuova Duplica Trova tutti Trova standard Indice Indice ascii Importa immagine Tabella Lista Stampa Note Elimina Copia Esol

Lemma
grumjàl

N° Fonte	Trascrizione	Trascrizione corretta
1 di 1	grumjàl	
Area	Sottoarea	Località
Area 13 Primiero	13A Val Cismon	non disponibile per sottoarea
Fonte	1976 Tissot	
Definizione	Termine italiano	Grammatica
grembiule che copre la parte anteriore della gonna, annodato	grembiule	s.m., pl.
Varianti in uso	Varianti antiche	Derivati
Uso	Osservazioni	Immagine
vestiario e calzature	il fabbro (<i>fàaro</i>) portava, lavorando, un	
N° Esempio	Esempi	Italiano
1 di 2	el grumjàl de la festa	nell'antico costume era bello, a fiorellini
2 di 2	prima de dir mal, vòrdete so 'l grumjàl	prov.: prima di parlare male di altri, vedi tu

Software by Centro Tecnologie Multimediali Università degli Studi di Trento

Different fields are proposed in the record. The first one is the dialectal **headword** in its original transcription (the one that each author has chosen); on the same line, we find the phonetical **transcription** with the symbols that have been chosen for the *ALTR*; when required, the correction of transcription mistakes found in the original source (**trascrizione corretta**) is given.

On the second line we find the information regarding the source (**author, area, subarea and village**).

In the central part of the record the **definition** is given, followed by a field named **Italian term**, where one or more translations are proposed for the dialectal word; the **grammatical category** follows (where the traditional names of the parts of the speech are used: nouns; transitive, intransitive, reflexive verbs; pronouns; adjectives; conjunctions,...). **Variants** (alternative forms) and **derivates** follow. We find then **lexical field markers (uso)** where the main domain of use for the word are suggested; the **notes** of the author about the word follow (when present), and the **pictures** given in the original dictionaries.

Finally, **examples** and collocations appear, with their **italian** translation (when present) and with an indication of the **lexical field** (where possible).

Two colours distinguish the different parts of the entry: white and yellow. What is given in the white fields reproduces the original source faithfully; in the yellow fields, on the other hand, new information is provided by the *ALTR* staff.

Sometimes new information can be added in the white fields in order to for an easier and richer search in the electronic dictionary. In this case, however, what is proposed by the *ALTR* staff is put inside brackets. In fact, the accuracy of the electronic version to the original sources is a very important criterion that we have tried to keep in mind in each step of the planning of the new dictionary.

3. Main topics

The “yellow sections” (which add new information to the original sources) have raised the most interesting problems in the planning the *ALTR*. The solutions that have been adopted can still be discussed. I will present here the reasons of the choices that have been made regarding a) phonetical transcriptions, b) italian translations of the headword, c) and lexical field markers (uso).

3.1. Transcription

In the section TRASCRIZIONE, the writing of the headword according to a simplified phonetical alphabet is given. It is necessary in fact to give uniformity to the great variety of alphabets adopted by the different lexicographers (cfr. Sanga 1979-80). The *ALTR* does not introduce symbols that are different from the usual alphabet; it only adds precise information about the position of accent, it distinguishes open and close vowels, and fricatives +/- voiced (see p.p. slides).

IPA has not been used, because of two main reasons: first of all, the *ALTR* is based on a corpus of written texts, and not of oral materials.³ Secondly, the electronic archive wants to be an easy tool of research for non specialist readers too, who could be discouraged by IPA.

3.2. *Italian translation*

There is a field in the archive named *TERMINE ITALIANO*: here one or more Italian words are given as a translation of the dialectal headword⁴. In the original sources the Italian translation often corresponds to the definition. However, definitions are not always transparent: each author uses a different style and a different language for defining the words, and it is not unusual to find here old and very specialistic terms, that are not easily understandable.⁵

Let's see a simple example. The word *rampon* (crampon) in Trentino dialects has two different meanings. In our Archive only two dictionaries register the word and we find three records: two present the same meaning, through different definitions, which are proposed by two authors (Aneggi and Tissot); the third record concerning the variety of Cembra, introduces a second meaning of the word (*riavolo*) which is very specialized.

In this case, the field "Italian term" has two evident advantages: a) in any search it binds together two records that have the same meaning; b) it renders the specialistic definition given by Aneggi for the second meaning of the word easier to be understood thanks to a paraphrase (instrument with a long handle used for cleaning casks). More generally, the field transforms monolingual dialectal dictionaries in a bidirectional dictionary where the reader can go from dialect to Italian, but also from Italian to dialect.

3.3. *Thematic field markers (uso)*

Another important problem that the *ALTR* staff has discussed concerns the field named *USO* (better to say: thematic or lexical field markers).

What we intend to give here is the possibility of a research, based on semantic markers, that puts together words used within the same domain. The result could be useful not only for people interested

³ On this topic, see Bonfadini - Iannaccaro (2005).

⁴ Per l'indicazione del termine italiano si è fatto riferimento in particolare al *DISC* (1997) e a De Mauro (2000).

⁵ "Opaque" definitions are given in particular for words referred to plants, animals and instruments.

to linguistic aspects of the lexicon, but also for people interested to ethnography, or in general to elements of “material, popular culture”.

However, the introduction in the dictionary of semantic markers raises some delicate problems of classification (see Jackson and Marengo). The *ALTR* semantic markers list -as it is now proposed- is not definitive: it consists of 58 markers, that have been recognized on the basis of what the authors underlined about the specific employ of the word in the definitions, in the notes and in the examples. The 58 markers are rather dishomogeneous: in fact, we find all together very broad categories (*credenze/ beliefs*), nomenclature (lists of names: for plants, animals, places, person names, activities...), and specific semantic areas (*slitte/ sledges*). A better classification should be articulated in more levels, with general themes subarticulated in one or more specific ones: however, this classification can be done only with a wide input (that is when the elaboration of the sources will be in a more advanced phase), in order to recognize the main themes treated and to organize them in a correct semantic hierarchy.

4. Ways of searching

A main purpose of the *ALTR* is to be a flexible tool that allows for searching both expert and non expert users. Therefore, three different ways of research are offered: a) geographical; b) selection from list; c) free research.

In each research one or more fields can be inquired. Let's see some examples.

TERMINE ITALIANO **EX** ubriaco, stupido (drunk; stupid)

TRASCRIZIONE: **EX** sequence *ns 34 records*. Notice: in the transcription field the *ALTR* associates automatically the selected symbol with the same + diachritics

USO: **EX** parti del corpo (parts of the body).

We can also combine more than one research: for instance we can ask to find all the proverbs (ESEMPI italiano) that contain a name of an animal (USO). **EX** 34 records

In another research, combining three fields, we can ask to find all those verbs (GRAMMAR) that in the examples (ESEMPI) can appear with a specific adverb –or particle- (phrasal verbs): **EX** *for a* in a particular area (TISSOT).

Finally, the *ALTR* allows for printing the results of the search. Six different types of reports are possible: general report, headwords report, grammar report, alternative forms report, derivatives report, examples report.

References

A. Dictionaries

- ANEGGI, A. (1984) *Dizionario cembrano (triangolo Sover-Montesover-Piscine): parole e cose- frasi - modo di dire - proverbi del dialetto della valle di Cembra* (revisione linguistica a cura di Piera Rizzolatti), S. Michele all'Adige (Trento), Museo degli usi e costumi della gente trentina.
- AZZOLINI, G. (1856); edizione completa 1976, *Vocabolario vernacolo-italiano pei distretti roveretano e trentino del professore Giambattista Azzolini*, Trento, Provincia Autonoma, Assessorato alle Attività Culturali.
- DE MAURO, T. (2000) *Il dizionario della lingua italiana per il terzo millennio*, Torino, Paravia.
- DISC (1997) *Dizionario Italiano Sabatini Coletti*, a cura di Francesco Sabatini e Vittorio Coletti, Firenze, Giunti.
- GROFF, L. (1955) *Dizionario trentino-italiano: oltre 3000 voci, frasi, detti e proverbi trentini, note sulla struttura e diffusione del dialetto. Florilegio di poesie e prose dialettali; con biografie di 15 autori del Trentino*, Trento, Monauni.
- PRATI, A. (1960) *Dizionario valsuganotto*, Venezia, Roma, Istituto per la collaborazione culturale.
- QUARESIMA, E. (1964) *Vocabolario anaunico e solandro raffrontato col trentino*, Venezia, Roma, Istituto per la collaborazione culturale.
- TISSOT, L. (1976) *Dizionario primierotto: parole, frasi, modi di dire, proverbi del dialetto di Primiero*, Trento, Provincia Autonoma di Trento.

B. Other references

- BONFADINI G., IANACCARO G. (2005) 'La questione della grafia', in Cordin P.(2005), 59-74.
- CORDIN, P. (2002) *L'Archivio Lessicale dei Dialetti Trentini (ALTR). Problemi metodologici e proposte per l'organizzazione di una banca dati*, in "La Dialettologia oltre il 2001", Atti del Convegno, Sappada/Plodn (Belluno), 1-5 luglio 2001, a cura di G. Marcato, Padova, Unipress, 363-372.
- CORDIN, P. (a cura di) (2005) *L'Archivio lessicale dei dialetti trentini*, Trento, Dipartimento di Scienze filologiche e storiche.
- CORDIN, P., GATTI, T. (2006) 'Dai dizionari dialettali su carta ai dizionari in rete. Aspetti metodologici e questioni aperte', in Bruni F. e Marcato C. (a cura di), *Lessicografia dialettale*. Atti del Convegno, Venezia 9-11 dicembre 2004, Venezia, Centro interuniversitario di studi veneti.

- CORDIN P. (2006), 'L'Archivio lessicale dei dialetti trentini', in CORINO E., MARELLO C., ONESTI C. (eds.), *XII Euralex International Congress Proceedings*, Torino, 6th-9th September, 2006, Torino, Dell'Orso, vol. II, 655-660.
- SANGA, G. (1979-80) 'La grafia dei dialetti', *Rivista italiana di dialettologia*, 4, 236-238.