Carlo Zoli, Silvia Randaccio

{carlo.zoli, silvia.randaccio}@smallcodes.com

Smallcodes and LinMiTech: two faces of the same new business model for the development of LRTs for LRLs

1 . LinMiTech Trentino

LinMiTech Trentino is a non-profit-making association that has as its goal the development and stimulation of the use of and training in, amongst its members and the speaking community, the linguistic, digital and new media technologies in the minority languages.

It was established on the initiative of the “Majon di Fascegn” Ladino Cultural Institute which, during the last twenty years, has made major efforts for the creation of technologies and IT resources. Together with the provincial bodies appointed to defend and promote minority languages, it has decided to share with them and to unify and harmonise their own projects with those that are in some way similar and that in the meantime have also been undertaken by them. All in a logic of reusing know-how and avoiding the duplication of expenses and resources. Completing the picture of the founder members is the company CELCT, an expression of the Bruno Kessler foundation (a major international mover in scientific research in the computational linguistics sector) and the TalenT Association, which embraces within it technological (software development) and professional (field researchers) partners as well as those from the world of university tuition.

2 . Smallcodes

It works in partnership with Smallcodes, a small software house based in Florence with a specific focus on the development of IT tools for minority languages. The aim of this collaboration is that of shortening the digital divide between majority languages and regional and minority languages. To achieve this goal, Smallcodes produces software systems for lexicography, spellchecking and neology/terminology planning for lesser-used languages, plus systems for toponymy cataloging and bibliographic archiving. These five modules are, according to the policy of Smallcodes, the first step towards a modern use of the language.

The interaction between LinMiTech and Smallcodes has given rise to a peculiar business model which starts from the awareness that that for less resourced languages language technologies are useless without adequate language resources. But, unlike big languages, LRs for minority languages cannot be found, either on the web or from classical sources, to the same extent as for major language. What one finds is inconsistent, due to graphical instability and to scarce presence in official environments. Therefore the parallel development of tools and resources cannot be separated at any level of the process of language rehabilitation. The code development without a correspondent increase in data retrieval may cause the production of unusable, or useless, tools.

Therefore, tools developers must be supported by customers and users who are also co-producers, co-developers and can supply their large amount of data to technology experts. From this awareness, the conclusion is drawn that it is crucial to have a federation of users who share commitment to providing intensive use of LTs and LRs. This federation of users should be non-profit in order to access funding for language policies or for scientific research world-wide. In this way, the software among this community is open-sourced but not in the totally voluntary type such as Libre Office, but instead it is guided by a leading industrial developer (such as Canonical and Open Office). LinMiTech, as the representative of the network of confederates has chosen Smallcodes as its technological partner because of SC’s precise focus on minority languages. In fact, a specific expertise in linguistics is very hard to find among developers.

Unlike classic open source development, the development of LTs depends both on the software codes and on the linguistic research. The traditional approach of open-source based on volunteers and donations is not applicable in this field because of the need for cooperation among linguistic experts and IT experts. In fact, the business model is not that the language experts or researchers adopt the system as users, basically using it “at their own risk” or contributing to the development, in a classical open-source fashion, without an integrated work with technology experts.

On the contrary, the LinMiTech-Smallcodes business model is that the software is centrally developed, and partnerships and funding opportunities are established every time a new language group enters the community. Every new language expert group adds new expertise, new funding, requests new features, but development is pursued in an industrial fashion, with attention to the latest web technologies, with highly resourced staff in an a “web 2.0 commercial way”. Then, the business itself is basically non-profit, but however this is different from software development done inside the linguistic academic world, which cannot have the structure and the attitude of a commercial software house.

Finally it is more common to find a commitment for sharing language resources (see for example OLAC, DoBeS), whereas Smallcodes focuses more on the sharing of software tools with language resources.

3. The linguistic network

The main consequence of this vision is that network among minorities is fundamental for the purpose of guaranteeing each language a systematic and constant presence in the written and in the IT world. But it must be clear that this presence may often have a symbolic importance, more than a functional one, which although shouldn't be underestimated. In fact, LTs for small languages are primarily designed to rise Ausbau (self-awareness) of a language and only on a second level to describe this language. An Italian-Sardinian automatic translator, in a world in which all the Sardinians are proficient in Italian, makes only sense in terms of recognition of Sardinian language. Such a translator has a totally different, if not opposed, aim of a Chinese-English translator. It is not developed to "describe" Sardinian language, but rather to "make" it.

This business model will improve its efficiency as many linguistic communities, bearer of LRs and lacking LTs, joins it as stake-holders of the organization.

4. Some examples of the model

We will give an example of the virtuous snowball effect that this cooperation among minorities has permitted to generate, in terms of

a) language resources for less described languages;
b) dramatic improvement of the tools;
c) significant cash flows for the industrial side of the model

The arrows below represent the history of events in the creation of our business model. Every single step represent a goal we achieved through the cooperation at the project of different institutions, associations, bodies, activists and so on.

Lexicographic tool for German-Badia Ladin → The Ladin side of the db re-used for Gardena Ladin-Italian → The Italian side of the db re-used for the Italian-Switzerland dialectological database → Great development of the tool that becomes a multi-dialectal lexicographic system (cfr. 5.1) → Multidialectal database for German varieties of Northern Italy (Mòcheno, Cimbrian, Sappada, Walser) → Etymological dictionary of Oto-Mangue language stock of Oaxaca, Mexico.

As it can be seen, the bootstrap phase was financed by small communities which are comparatively better resourced and funded than others. The first development and the first resources have allowed the group to join a much bigger working group such as Italian Switzerland which has in turn permitted to give very small communities, lacking of any kind of support and funding, to enter the community and use advanced technological tools.

Morphological analyzer for Sardinian language → Spellchecker for Sardinian → Spellchecker for Ladin → Spellchecker with dialectal background-driven mistakes for highly internally differentiated languages as Rrromani.

Learning applications for mobile devices for Rromani language → Same applications extended to Tamazight language.

5. A Brief description of the main tools

5.1 Lexicographic system

After having collected enough lexical material, it is possible to plan a lexicographic tool for the creation of dictionaries. In fact, lexical lists of various kinds are the necessary condition in order to set up the dictionary. They can be wordlists of local or global language (i.e. conforming to local varieties of the language or to the standardized spelling); they can also be imported form informal databases and being the result of an OCR or parsing of ancient dictionaries.

The figure 1 (see appendix) shows an example of 'standardizing' dictionary with registration of local varieties. Here is the extreme case of the entry otóbro ('October') which has around 150 different phonetic realizations ascribable to three consonantal macro-phenomena (1. maintenance of etymological t; 2. palatalization of t > c. 3. loss of b). As it can be seen, the standard forms have been chosen among those forms which are more "etymologically regular" (Lurà et al., 2009). Then (fig. 2), we have the same entry in a human-readable form (actually an XML + CSS which can be easily imported in a professional publishing tool as Adobe Indesign, see fig. 4); fig. 3 shows the XML of fig. 2 in the classic machine-readable form.

5.2 Spellchecking system

Another step is the creation of a fully integrated spell-checker for the minority language. The majority of spell-checking systems (e.g. HunSpell which is the base of LibreOffice, Firefox, Chrome, etc. proofing tools) are fed with wordlists which are not integrated and often not even exported from a coherent dictionary authoring system (Németh 2011); the same can be said for morphological engines or corpus analysis software, such as NOOJ (Ben Hamadou, Mesfar, Silberztein, 2010): they may provide powerful tools, but they are never integrated with a dictionary authoring and publishing system, and their use is normally confined to NLP specialists, and often well beyond the reach of traditional linguists not to say general public, school teachers or public administration staff. In fact, having an integrated system means that every change is reported automatically in both modules of the system and that the spell-checker is always up to date, and so is authoring, Web publication, Smartphone app generation, and even traditional paper publishing are all steps of a highly integrated procedure. This is especially useful in treating minority or lesser-used language, where the fieldwork is always active and new additions, changes, creation of neology and terminology, and even spell reforms are frequent events. As modern spell-checkers, our module works with a "best-guess" pattern of the rule, based on statistic algorithms, on Levenshtein distance (Levenshtein, 1966) and on double metaphone (Philips, 1990).

In addition, it includes dialectal-driven error patterns, which are fundamental for minority languages. In fact, every correction system sets up its guesses upon similarities of words. Our system adds to this method the awareness that, for semi- or recently standardized languages where the overwhelming majority of writers are de facto illiterate in their language, most errors can be caused by the knowledge of a word in one particular language variety that is not the standard form: in minority languages people do not only misspell: they simply can't write, even if they can perfectly speak (and write in the dominant language). The two word forms (standard and non-standard) may differ a lot sometimes: the non-standard word can be, for example, more similar to a word with a completely different meaning than to its standard equivalent; or it can also be so graphically far from the standard form that the system is not able to find the equivalence using the statistic algorithm or the standard pattern matching. The system must then know that there can be odd correspondences. We can offer a typical example from Sardinian language (the first language for which we developed the spell-checker): the word berbeghe (sheep) is pronounced /brebei/ in South Sardinia. If we analyze the differences among the two words, we can understand that a simple system would not be able to guess the standard form (berbeghe) starting from the non-standard one (brebei) (Corongiu, 2013). Conversely, our dialect-oriented spell-checker knows these odd correspondences and the rules that allow to guess them. Our system uses therefore two guess pattern, shown in the table below (fig. 5): the simple one detects "soundslike typical mistakes"; the advanced one detects "linguistic-background driven mistakes". See fig. 6 for MS Word and web interface of the "dialectal" spellchecker (Zoli, 2008). Smallcodes is currently addressing at Libre Office and Open Office to integrate a spellchecking system of Ladin language based on Smallcodes' algorythms in their system.

5.3 Terminology module

Another unmentioned tool is the terminology module, developed to be integrated in the dictionary. The creation of the terminology is a fundamental procedure if we want the language to be employed, for example, in school teaching (see for example fig. 7, which shows a collaborative webTool for neology, used by the authors of schoolbooks in Ladin Dolomitan), and administrative / official translation (see fig 8 & 9 for a tool of computer-aided technical translation for Sardinian languages, used by various public bodies). Languages which do not have a written tradition normally lack of technical lexicon. These new words need therefore to be created and the method for their creation already exists: the sources are the other international languages that have made this procedure before and the other minority languages that have already solved these issues. Another possibility is to re-use old words whose original meaning is losing importance in today's life and make these words express new meanings. A typical example is the vocabulary used for cars nowadays in Italian: this is nothing more than the recovered lexicon for horse carriages; similarly, the lexicon of Air Navigation is directly taken from Maritime Navigation vocabulary. English typically uses this strategy for neologisms, exploiting metaphors and meaning extensions of pre-existing words. Romance languages, on the other hand, favour the use of loan words, drawing inspiration from present or past prestigious languages.

5.4 Learning applications

A possible cross-development, as already mentioned, is an application in an e-book format, available off-line on iPad and iPhone. The application may include texts, images and image galleries, audio and video files. It is devoted to school teaching and contains therefore interactive exercises of different kinds: true / false questions, closed questions, links between text and image. In the application, it is possible to insert audio recordings of voice reading sections of text. For this particular activity, linguistic experts have to provide materials such as texts and exercises (in .doc format), audio files (recordings of texts reading), images for photo galleries and video.

The application can directly linked to a lexicon of the target language. This allows the simultaneous connection to any previously chosen entries in the dictionary, so that students have immediate information about selected terms in the text. The connection is made with a simple hover over a word (appropriately indicated by the graphics) that opens the card of the term in question. It

In parallel, it is possible to developed a web application, accessible online on every PC or tablet, which includes the same types of texts, exercises, images and image galleries, audio and video files of the e-book application. In this case, being the application always connected to the web, every entry in the lexicon is automatically updated in the application in case of any modification.

Optionally, the web application may include a "read-along" system. This is a system of rapid synchronization of text and reading, a reading-synchronized display of text segments ("karaoke" type), which is useful as an aid to lesser-used languages and / or languages whose speakers are more accustomed to orality. For the use of "read-along", Smallcodes has developed a unique system of rapid synchronization between audio and text, in which the operator can synchronize audio in a time equal to 1x the actual time of reading (i.e. to synchronize a one hour long audio it takes about a single hour).

6. Relationship with Academia

In many cases, language experts and linguistic resources providers are scholars coming from the academic world. So their focus in a “publish or perish” perspective is less concentrated on fallouts of their work on the community than it is on pure research. In this way, joining our community of users, they can have an operational arm to convert their research into usable tools and on our side, a strong partnership with scholars can lead to funding opportunities in the research field.

Bibliography

Ben Hamadou, Abdelmajid; Mesfar, Slim; Silberztein, Max. “Finite State Language Engineering: NooJ 2009”. International Conference and Workshop. Touzeur: Centre de Publication Universitaire, 2010.

Boukous, Ahmed. Phonologie de l’Amazighe. Rabat: Institut Royal de la Culture Amazighe, 2009.

Boukhris, Fatima. La Nouvelle Grammaire de l’Amazighe. Rabat: Institut Royal de la Culture Amazighe, 2008.

Corongiu, Giuseppe. Il sardo: una lingua normale. Cagliari: Condaghes, 2013.

Dell’Aquila, Vittorio; Iannàccaro, Gabriele. La pianificazione linguistica. Roma: Carocci Editore, 2011.

Kloss, Heinz “Abstandsprachen und Ausbausprachen”. In Göschel, Joachim; Nail, Norbert; Van der Els, Gaston. Zur Theorie des Dialekts: Aufsätze aus 100 Jahren Forschung. Zeitschrift fur Dialektologie and Linguistik, 1976.

Krauwer, Stevem. The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap Proceedings of SPECOM 2003, Moscow, 2013.

Fellbaum, Christiane, ed. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press, 1998.

Francopoulo, Gil et al. Lexical markup framework (LMK) Genoa: LREC, 2006.

Levenshtein, Vladimir I. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, 1966.

Lurà Franco et al. “Dalla carta al web: la versione informatica del lessico dialettale della Svizzera italiana”, In: Ruffino G., D'Agostino M. Storia della lingua italiana e dialettologia, atti del VIII Convegno Internazionale dell'Associazione per la Storia della Lingua Italiana, Palermo, 2009.

Németh, László http://hunspell.sourceforge.net/ (27/05/2013).

Philips Lawrence. Hanging on the Metaphone, In Computer Language, Vol. 7, No. 12 (December), 1990.

Scannel, Kevin. “New computational resources for indigenous and minority languages”, 17th annual NAACLT conference. Isle of Man, 2011.

Scannel, Kevin. “Semi-automated construction of semantic networks using web corpora”, Words, Texts and Dictionaries conference. University of Wales Centre for Advanced Welsh and Celtic Studies, Aberystwyth, 2008.

Vitali, Daniele. “Appello ai romagnoli per studiare la diversità dialettale” La Ludla XII, 2008.

Soria, Claudia, Zoli, Carlo. “New markets for Language Technology for minority languages”, Maaya Conference. Paris, 2012.

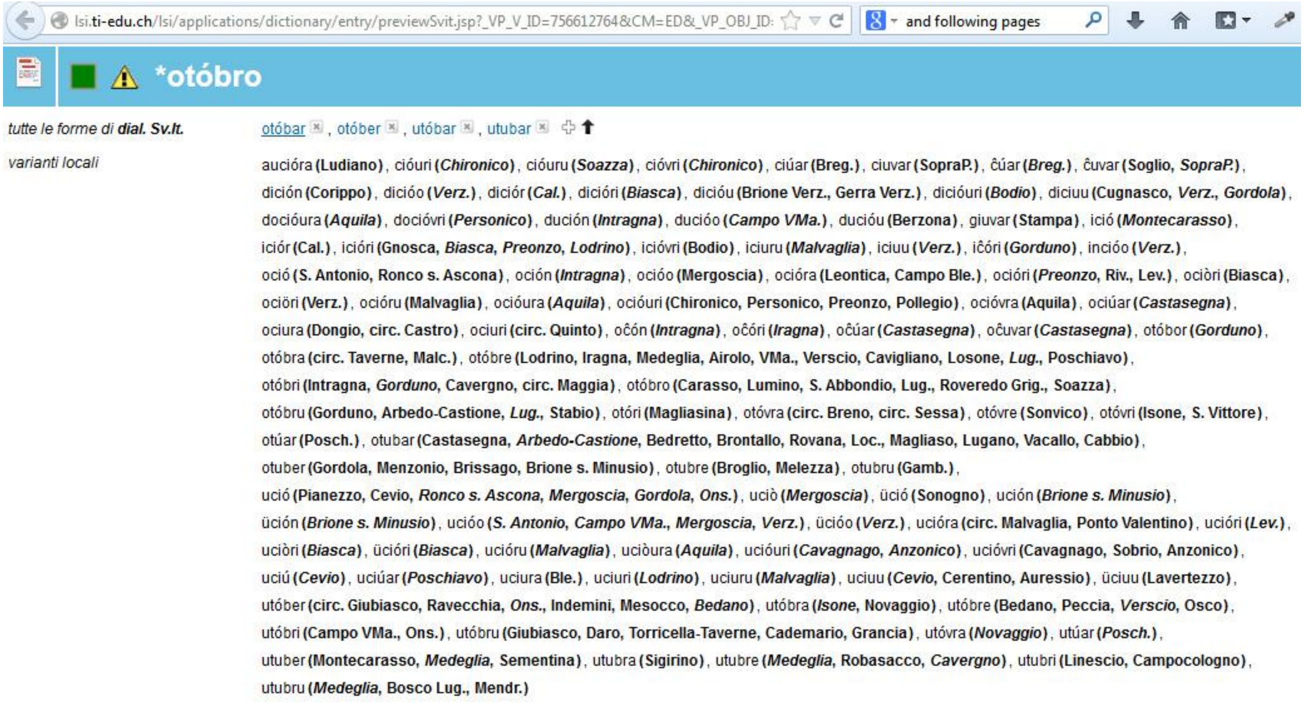
Videsott, Paul. Vocabolar dl Ladin Leterar / Wörterbuch des literarischen Ladinisch / Vocabolario del Ladino letterario (VLL). Projektbeschreibung, 2011.

Zoli, Carlo. “Encouraging the presence in the cyberspace of the lesser used languages through writing and proofing tools: the case of Sardinian language”, Maaya Conference. Paris, 2012.

Zoli, Carlo. “La scrittura standard del romagnolo: un’urgenza non rimandabile” La Ludla IX, 2012.

Zoli, Carlo. “Trattamento digitale delle lingue al servizio delle lingue meno usate”, Corongiu G., Romagnino C. Sa Diversidade de sas Limbas in Europa, Itàlia e Sardigna. Atos de sa cunferèntzia regionale de sa limba sarda, Macumere/Macomer, 2008.

Appendix


lsi.ti-edu.ch/lsi/applications/dictionary/entry/previewSvit.jsp?_VP_V_ID=756612764&CM=ED&_VP_OBJ_ID:
and following pages

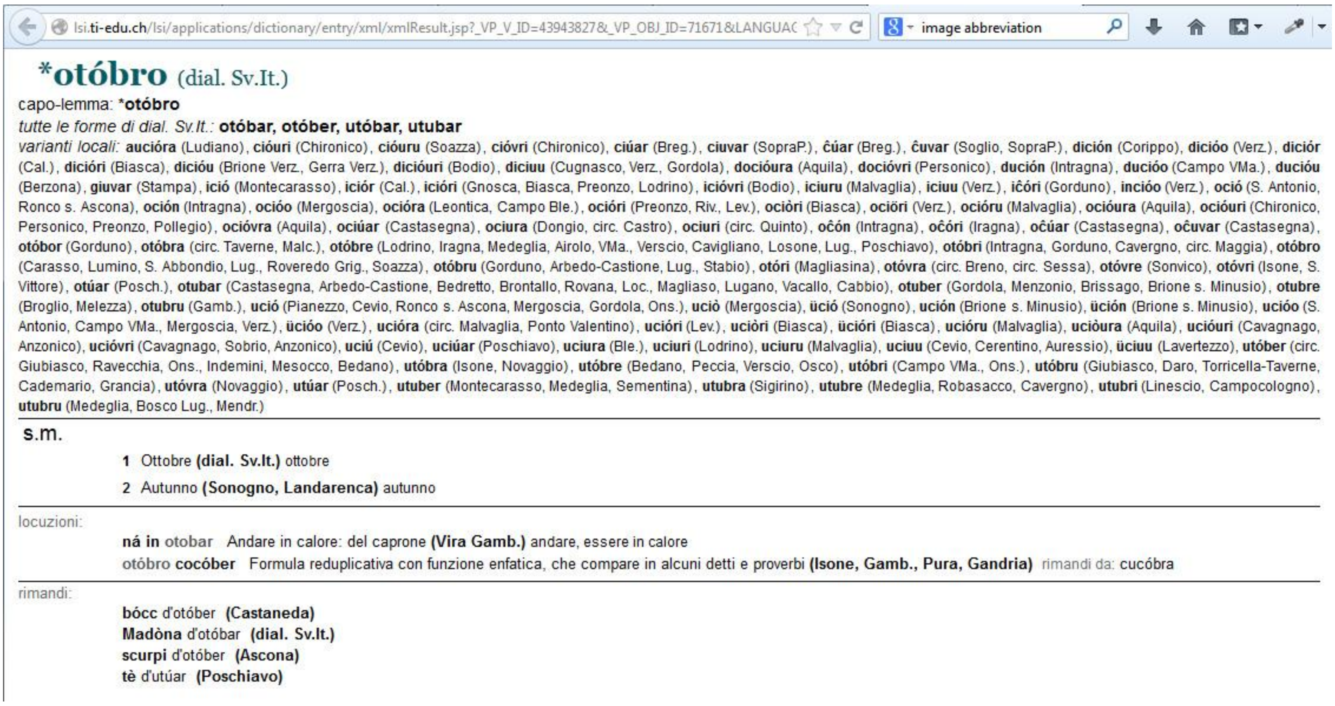
*otóbro

*tutte le forme di **dial. Sv.It.*** otóbar, otóber, utóbar, utubar

varianti locali

aucióra **(Ludiano)**, cióuri **(*Chironico*)**, cióuru **(*Soazza*)**, cióvri **(*Chironico*)**, ciúar **(Breg.)**, ciuvar **(SopraP.)**, ĉúar **(*Breg.*)**, ĉuvar **(Soglio, *SopraP.*)**,
dición **(Corippo)**, dicióo **(*Verz.*)**, diciór **(*Cal.*)**, dicióri **(*Biasca*)**, dicióu **(Brione Verz., Gerra Verz.)**, dicióuri **(*Bodio*)**, diciuu **(Cugnasco, *Verz.*, *Gordola*)**,
docióura **(*Aquila*)**, docióvri **(*Personico*)**, dución **(*Intragna*)**, ducióo **(*Campo VMa.*)**, ducióu **(Berzona)**, giuvar **(Stampa)**, ició **(*Montecarasso*)**,
iciór **(Cal.)**, icióri **(Gnosca, *Biasca*, *Preonzo*, *Lodrino*)**, icióvri **(Bodio)**, iciuru **(*Malvaglia*)**, iciuu **(*Verz.*)**, iĉóri **(*Gorduno*)**, incióo **(*Verz.*)**,
oció **(S. Antonio, Ronco s. Ascona)**, oción **(*Intragna*)**, ocióo **(Mergoscia)**, ocióra **(Leontica, Campo Ble.)**, ocióri **(*Preonzo*, Riv., Lev.)**, ociòri **(Biasca)**,
ociöri **(Verz.)**, ocióru **(Malvaglia)**, ocióura **(*Aquila*)**, ocióuri **(Chironico, Personico, Preonzo, Pollegio)**, ocióvra **(Aquila)**, ociúar **(*Castasegna*)**,
ociura **(Dongio, circ. Castro)**, ociuri **(circ. Quinto)**, oĉón **(*Intragna*)**, oĉóri **(*Iragna*)**, oĉúar **(*Castasegna*)**, oĉuvar **(*Castasegna*)**, otóbor **(*Gorduno*)**,
otóbra **(circ. Taverne, Malc.)**, otóbre **(Lodrino, Iragna, Medeglia, Airolo, VMa., Verscio, Cavigliano, Losone, *Lug.*, Poschiavo)**,
otóbri **(Intragna, *Gorduno*, Cavergno, circ. Maggia)**, otóbro **(Carasso, Lumino, S. Abbondio, Lug., Roveredo Grig., Soazza)**,
otóbru **(Gorduno, Arbedo-Castione, *Lug.*, Stabio)**, otóri **(Magliasina)**, otóvra **(circ. Breno, circ. Sessa)**, otóvre **(Sonvico)**, otóvri **(Isone, S. Vittore)**,
otúar **(Posch.)**, otubar **(Castasegna, *Arbedo-Castione*, Bedretto, Brontallo, Rovana, Loc., Magliaso, Lugano, Vacallo, Cabbio)**,
otuber **(Gordola, Menzonio, Brissago, Brione s. Minusio)**, otubre **(Broglio, Melezza)**, otubru **(Gamb.)**,
ució **(Pianezzo, Cevio, *Ronco s. Ascona*, *Mergoscia*, *Gordola*, *Ons.*)**, uciò **(*Mergoscia*)**, üció **(Sonogno)**, ución **(*Brione s. Minusio*)**,
üción **(*Brione s. Minusio*)**, ucióo **(*S. Antonio*, *Campo VMa.*, *Mergoscia*, *Verz.*)**, ücióo **(*Verz.*)**, ucióra **(circ. Malvaglia, Ponto Valentino)**, ucióri **(*Lev.*)**,
uciòri **(*Biasca*)**, ücióri **(*Biasca*)**, ucióru **(*Malvaglia*)**, uciòura **(*Aquila*)**, ucióuri **(*Cavagnago*, *Anzonico*)**, ucióvri **(Cavagnago, Sobrio, Anzonico)**,
uciú **(*Cevio*)**, uciúar **(*Poschiavo*)**, uciura **(Ble.)**, uciuri **(*Lodrino*)**, uciuru **(*Malvaglia*)**, uciuu **(*Cevio*, Cerentino, Auressio)**, üciuu **(Lavertezzo)**,
utóber **(circ. Giubiasco, Ravecchia, *Ons.*, Indemini, Mesocco, *Bedano*)**, utóbra **(*Isone*, Novaggio)**, utóbre **(Bedano, Peccia, *Verscio*, Osco)**,
utóbri **(Campo VMa., Ons.)**, utóbru **(Giubiasco, Daro, Torricella-Taverne, Cademario, Grancia)**, utóvra **(*Novaggio*)**, utúar **(*Posch.*)**,
utuber **(Montecarasso, *Medeglia*, Sementina)**, utubra **(Sigirino)**, utubre **(*Medeglia*, Robasacco, *Cavergno*)**, utubri **(Linescio, Campocologno)**,
utubru **(*Medeglia*, Bosco Lug., Mendr.)**

Fig. 1: An example of a standardizing dictionary with registration of local varieties.


lsi.ti-edu.ch/lsi/applications/dictionary/entry/xml/xmlResult.jsp?_VP_V_ID=43943827&_VP_OBJ_ID=71671&LANGUAC
image abbreviation

***otóbro** (dial. Sv.It.)

capo-lemma: ***otóbro**

tutte le forme di dial. Sv.It.: **otóbar, otóber, utóbar, utubar**

varianti locali: **aucióra** (Ludiano), **cióuri** (Chironico), **cióuru** (Soazza), **cióvri** (Chironico), **ciúar** (Breg.), **ciuvar** (SopraP.), **ĉúar** (Breg.), **ĉuvar** (Soglio, SopraP.), **dición** (Corippo), **dicióo** (Verz.), **diciór** (Cal.), **dicióri** (Biasca), **dicióu** (Brione Verz., Gerra Verz.), **dicióuri** (Bodio), **diciuu** (Cugnasco, Verz., Gordola), **docióura** (Aquila), **docióvri** (Personico), **dución** (Intragna), **ducióo** (Campo VMa.), **ducióu** (Berzona), **giuvar** (Stampa), **ició** (Montecarasso), **iciór** (Cal.), **icióri** (Gnosca, Biasca, Preonzo, Lodrino), **icióvri** (Bodio), **iciuru** (Malvaglia), **iciuu** (Verz.), **iĉóri** (Gorduno), **incióo** (Verz.), **oció** (S. Antonio, Ronco s. Ascona), **oción** (Intragna), **ocióo** (Mergoscia), **ocióra** (Leontica, Campo Ble.), **ocióri** (Preonzo, Riv., Lev.), **ociòri** (Biasca), **ociöri** (Verz.), **ocióru** (Malvaglia), **ocióura** (Aquila), **ocióuri** (Chironico, Personico, Preonzo, Pollegio), **ocióvra** (Aquila), **ociúar** (Castasegna), **ociura** (Dongio, circ. Castro), **ociuri** (circ. Quinto), **oĉón** (Intragna), **oĉóri** (Iragna), **oĉúar** (Castasegna), **oĉuvar** (Castasegna), **otóbor** (Gorduno), **otóbra** (circ. Taverne, Malc.), **otóbre** (Lodrino, Iragna, Medeglia, Airolo, VMa., Verscio, Cavigliano, Losone, Lug., Poschiavo), **otóbri** (Intragna, Gorduno, Cavergno, circ. Maggia), **otóbro** (Carasso, Lumino, S. Abbondio, Lug., Roveredo Grig., Soazza), **otóbru** (Gorduno, Arbedo-Castione, Lug., Stabio), **otóri** (Magliasina), **otóvra** (circ. Breno, circ. Sessa), **otóvre** (Sonvico), **otóvri** (Isone, S. Vittore), **otúar** (Posch.), **otubar** (Castasegna, Arbedo-Castione, Bedretto, Brontallo, Rovana, Loc., Magliaso, Lugano, Vacallo, Cabbio), **otuber** (Gordola, Menzonio, Brissago, Brione s. Minusio), **otubre** (Broglio, Melezza), **otubru** (Gamb.), **ució** (Pianezzo, Cevio, Ronco s. Ascona, Mergoscia, Gordola, Ons.), **uciò** (Mergoscia), **üció** (Sonogno), **ución** (Brione s. Minusio), **üción** (Brione s. Minusio), **ucióo** (S. Antonio, Campo VMa., Mergoscia, Verz.), **ücióo** (Verz.), **ucióra** (circ. Malvaglia, Ponto Valentino), **ucióri** (Lev.), **uciòri** (Biasca), **ücióri** (Biasca), **ucióru** (Malvaglia), **uciòura** (Aquila), **ucióuri** (Cavagnago, Anzonico), **ucióvri** (Cavagnago, Sobrio, Anzonico), **uciú** (Cevio), **uciúar** (Poschiavo), **uciura** (Ble.), **uciuri** (Lodrino), **uciuru** (Malvaglia), **uciuu** (Cevio, Cerentino, Auressio), **üciuu** (Lavertezzo), **utóber** (circ. Giubiasco, Ravecchia, Ons., Indemini, Mesocco, Bedano), **utóbra** (Isone, Novaggio), **utóbre** (Bedano, Peccia, Verscio, Osco), **utóbri** (Campo VMa., Ons.), **utóbru** (Giubiasco, Daro, Torricella-Taverne, Cademario, Grancia), **utóvra** (Novaggio), **utúar** (Posch.), **utuber** (Montecarasso, Medeglia, Sementina), **utubra** (Sigirino), **utubre** (Medeglia, Robasacco, Cavergno), **utubri** (Linescio, Campocologno), **utubru** (Medeglia, Bosco Lug., Mendr.)

s.m.

1 Ottobre **(dial. Sv.It.)** ottobre

2 Autunno **(Sonogno, Landarenca)** autunno

locuzioni:

ná in otobar Andare in calore: del caprone **(Vira Gamb.)** andare, essere in calore

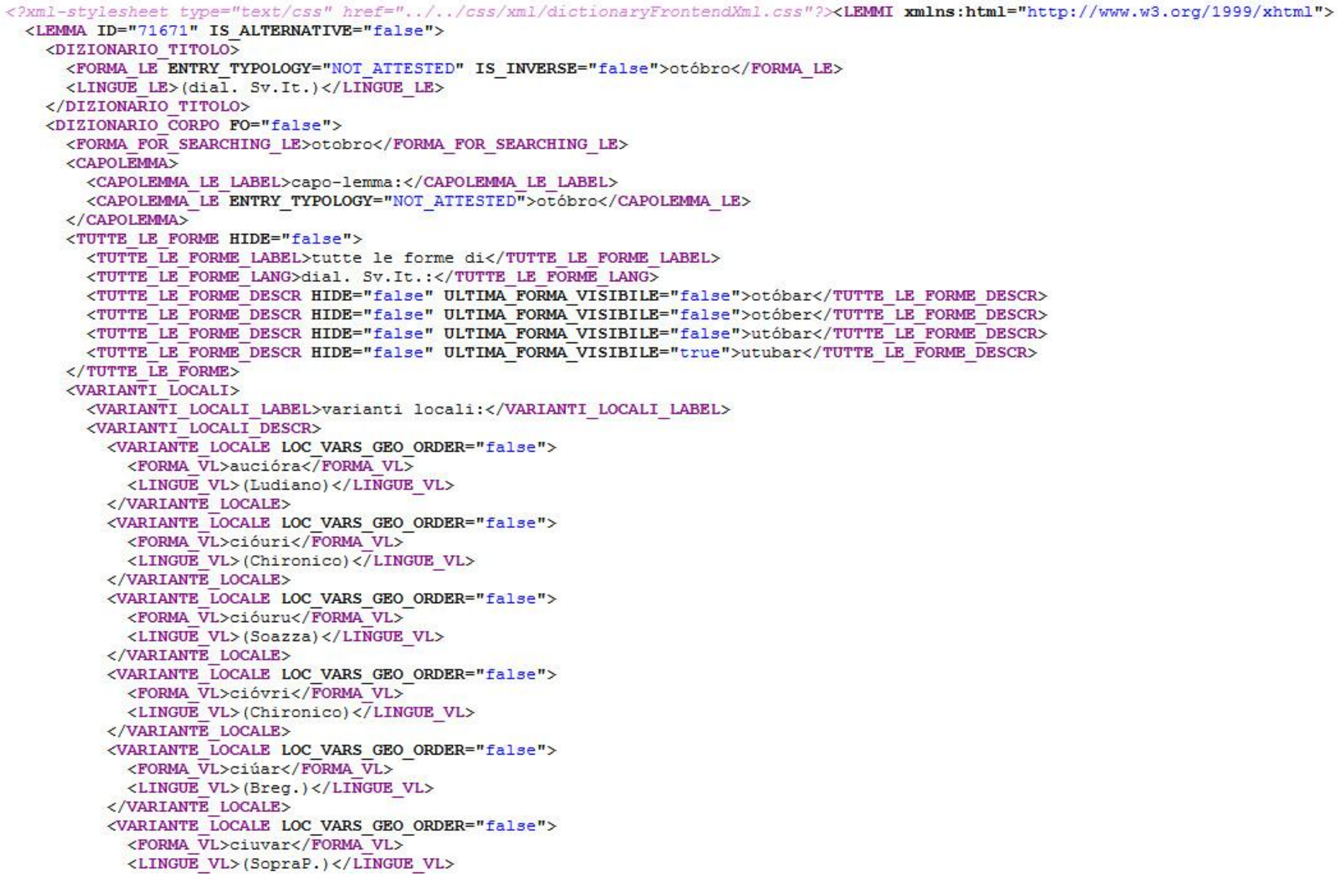
otóbro **cocóber** Formula reduplicativa con funzione enfatica, che compare in alcuni detti e proverbi **(Isone, Gamb., Pura, Gandria)** rimandi da: cucóbra

rimandi:

bócc d'otóber **(Castaneda)**
Madòna d'otóbar **(dial. Sv.It.)**
scurpi d'otóber **(Ascona)**
tè d'utúar **(Poschiavo)**

Fig. 2: The same entry in a human-readable form.[1]

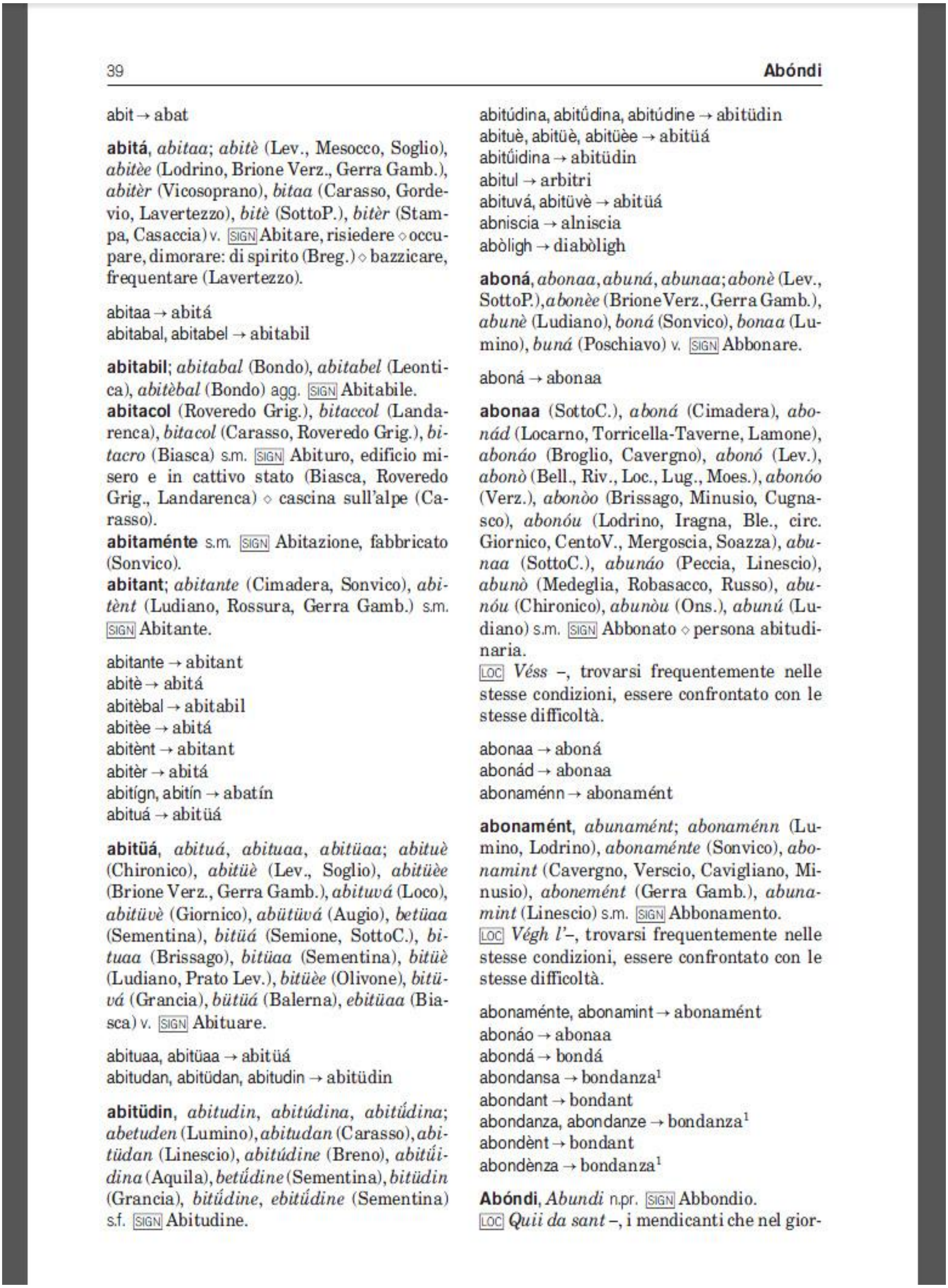
[1] Please note that the current tendency in normalization is to suggest a single graphic form but to allow free choices in local meanings and lexical types. The image shows the lexical type otóbro (‘October’) which in some places means ‘autumn, fall’.



```
<?xml-stylesheet type="text/css" href="../../css/xml/dictionaryFrontendXml.css"?><LEMMI xmlns:html="http://www.w3.org/1999/xhtml">
  <LEMMA ID="71671" IS_ALTERNATIVE="false">
    <DIZIONARIO_TITOLO>
      <FORMA_LE ENTRY_TYPOLOGY="NOT_ATTESTED" IS_INVERSE="false">otóbro</FORMA_LE>
      <LINGUE_LE>(dial. Sv.It.)</LINGUE_LE>
    </DIZIONARIO_TITOLO>
    <DIZIONARIO_CORPO FO="false">
      <FORMA_FOR_SEARCHING_LE>otobro</FORMA_FOR_SEARCHING_LE>
      <CAPOLEMMA>
        <CAPOLEMMA_LE_LABEL>capo-lemma:</CAPOLEMMA_LE_LABEL>
        <CAPOLEMMA_LE ENTRY_TYPOLOGY="NOT_ATTESTED">otóbro</CAPOLEMMA_LE>
      </CAPOLEMMA>
      <TUTTE_LE_FORME HIDE="false">
        <TUTTE_LE_FORME_LABEL>tutte le forme di</TUTTE_LE_FORME_LABEL>
        <TUTTE_LE_FORME_LANG>dial. Sv.It.:</TUTTE_LE_FORME_LANG>
        <TUTTE_LE_FORME_DESCR HIDE="false" ULTIMA_FORMA_VISIBILE="false">otóbar</TUTTE_LE_FORME_DESCR>
        <TUTTE_LE_FORME_DESCR HIDE="false" ULTIMA_FORMA_VISIBILE="false">otóber</TUTTE_LE_FORME_DESCR>
        <TUTTE_LE_FORME_DESCR HIDE="false" ULTIMA_FORMA_VISIBILE="false">utóbar</TUTTE_LE_FORME_DESCR>
        <TUTTE_LE_FORME_DESCR HIDE="false" ULTIMA_FORMA_VISIBILE="true">utubar</TUTTE_LE_FORME_DESCR>
      </TUTTE_LE_FORME>
      <VARIANTI_LOCALI>
        <VARIANTI_LOCALI_LABEL>varianti locali:</VARIANTI_LOCALI_LABEL>
        <VARIANTI_LOCALI_DESCR>
          <VARIANTE_LOCALE LOC_VARS_GEO_ORDER="false">
            <FORMA_VL>aucióra</FORMA_VL>
            <LINGUE_VL>(Ludiano)</LINGUE_VL>
          </VARIANTE_LOCALE>
          <VARIANTE_LOCALE LOC_VARS_GEO_ORDER="false">
            <FORMA_VL>cióuri</FORMA_VL>
            <LINGUE_VL>(Chironico)</LINGUE_VL>
          </VARIANTE_LOCALE>
          <VARIANTE_LOCALE LOC_VARS_GEO_ORDER="false">
            <FORMA_VL>cióuru</FORMA_VL>
            <LINGUE_VL>(Soazza)</LINGUE_VL>
          </VARIANTE_LOCALE>
          <VARIANTE_LOCALE LOC_VARS_GEO_ORDER="false">
            <FORMA_VL>cióvri</FORMA_VL>
            <LINGUE_VL>(Chironico)</LINGUE_VL>
          </VARIANTE_LOCALE>
          <VARIANTE_LOCALE LOC_VARS_GEO_ORDER="false">
            <FORMA_VL>ciúar</FORMA_VL>
            <LINGUE_VL>(Breg.)</LINGUE_VL>
          </VARIANTE_LOCALE>
          <VARIANTE_LOCALE LOC_VARS_GEO_ORDER="false">
            <FORMA_VL>ciuvar</FORMA_VL>
            <LINGUE_VL>(SopraP.)</LINGUE_VL>
```

Fig. 3: Machine-readable output of the same entry in a LMF (Francopoulo et al., 2006), compliant XML-schema.

Symmetrically, for the concept of "October", we could have many other lexical types, such as 'Month of St. Martin' or 'Month of chestnuts'.



abit → abat

abitá, *abitaa*; *abitè* (Lev., Mesocco, Soglio), *abitèe* (Lodrino, Brione Verz., Gerra Gamb.), *abitèr* (Vicosoprano), *bitaa* (Carasso, Gordevio, Lavertezzo), *bitè* (SottoP.), *bitèr* (Stampa, Casaccia) v. SIGN Abitare, risiedere ◇ occupare, dimorare: di spirito (Breg.) ◇ bazzicare, frequentare (Lavertezzo).

abitaa → abitá
abitabal, abitabel → abitabil

abitabil; *abitabal* (Bondo), *abitabel* (Leontica), *abitèbal* (Bondo) agg. SIGN Abitabile.
abitacol (Roveredo Grig.), *bitaccol* (Landarenca), *bitacol* (Carasso, Roveredo Grig.), *bitacro* (Biasca) s.m. SIGN Abituro, edificio misero e in cattivo stato (Biasca, Roveredo Grig., Landarenca) ◇ cascina sull'alpe (Carasso).
abitaménte s.m. SIGN Abitazione, fabbricato (Sonvico).
abitant; *abitante* (Cimadera, Sonvico), *abitènt* (Ludiano, Rossura, Gerra Gamb.) s.m. SIGN Abitante.

abitante → abitant
abitè → abitá
abitèbal → abitabil
abitèe → abitá
abitènt → abitant
abitèr → abitá
abitígn, abitín → abatín
abituá → abitüá

abitüá, *abituá*, *abituaa*, *abitüaa*; *abituè* (Chironico), *abitüè* (Lev., Soglio), *abitüèe* (Brione Verz., Gerra Gamb.), *abituvá* (Loco), *abitüvè* (Giornico), *abütüvá* (Augio), *betüaa* (Sementina), *bitüá* (Semione, SottoC.), *bituaa* (Brissago), *bitüaa* (Sementina), *bitüè* (Ludiano, Prato Lev.), *bitüèe* (Olivone), *bitüvá* (Grancia), *bütüá* (Balerna), *ebitüaa* (Biasca) v. SIGN Abituare.

abituaa, abitüaa → abitüá
abitudan, abitüdan, abitudin → abitüdin

abitüdin, *abitudin*, *abitúdina*, *abitǘdina*; *abetuden* (Lumino), *abitudan* (Carasso), *abitüdan* (Linescio), *abitúdine* (Breno), *abitǘidina* (Aquila), *betǘdine* (Sementina), *bitüdin* (Grancia), *bitǘdine*, *ebitǘdine* (Sementina) s.f. SIGN Abitudine.

abitúdina, abitǘdina, abitúdine → abitüdin
abituè, abitüè, abitüèe → abitüá
abitǘidina → abitüdin
abitul → arbitri
abituvá, abitüvè → abitüá
abniscia → alniscia
abòligh → diabòligh

aboná, *abonaa*, *abuná*, *abunaa*; *abonè* (Lev., SottoP.), *abonèe* (Brione Verz., Gerra Gamb.), *abunè* (Ludiano), *boná* (Sonvico), *bonaa* (Lumino), *buná* (Poschiavo) v. SIGN Abbonare.

aboná → abonaa

abonaa (SottoC.), *aboná* (Cimadera), *abonád* (Locarno, Torricella-Taverne, Lamone), *abonáo* (Broglio, Cavergno), *abonó* (Lev.), *abonò* (Bell., Riv., Loc., Lug., Moes.), *abonóo* (Verz.), *abonòo* (Brissago, Minusio, Cugnasco), *abonóu* (Lodrino, Iragna, Ble., circ. Giornico, CentoV., Mergoscia, Soazza), *abunaa* (SottoC.), *abunáo* (Peccia, Linescio), *abunò* (Medeglia, Robasacco, Russo), *abunóu* (Chironico), *abunòu* (Ons.), *abunú* (Ludiano) s.m. SIGN Abbonato ◇ persona abitudinaria.
LOC *Véss* –, trovarsi frequentemente nelle stesse condizioni, essere confrontato con le stesse difficoltà.

abonaa → aboná
abonád → abonaa
abonaménn → abonamént

abonamént, *abunamént*; *abonaménn* (Lumino, Lodrino), *abonaménte* (Sonvico), *abonamint* (Cavergno, Verscio, Cavigliano, Minusio), *abonemént* (Gerra Gamb.), *abunamint* (Linescio) s.m. SIGN Abbonamento.
LOC *Végh l'*–, trovarsi frequentemente nelle stesse condizioni, essere confrontato con le stesse difficoltà.

abonaménte, abonamint → abonamént
abonáo → abonaa
abondá → bondá
abondansa → bondanza^{1}
abondant → bondant
abondanza, abondanze → bondanza^{1}
abondènt → bondant
abondènza → bondanza^{1}

Abóndi, *Abundi* n.pr. SIGN Abbondio.
LOC *Quii da sant* –, i mendicanti che nel gior-

Fig. 4: XML above imported automatically into Adobe InDesign for automatic layout for printing.

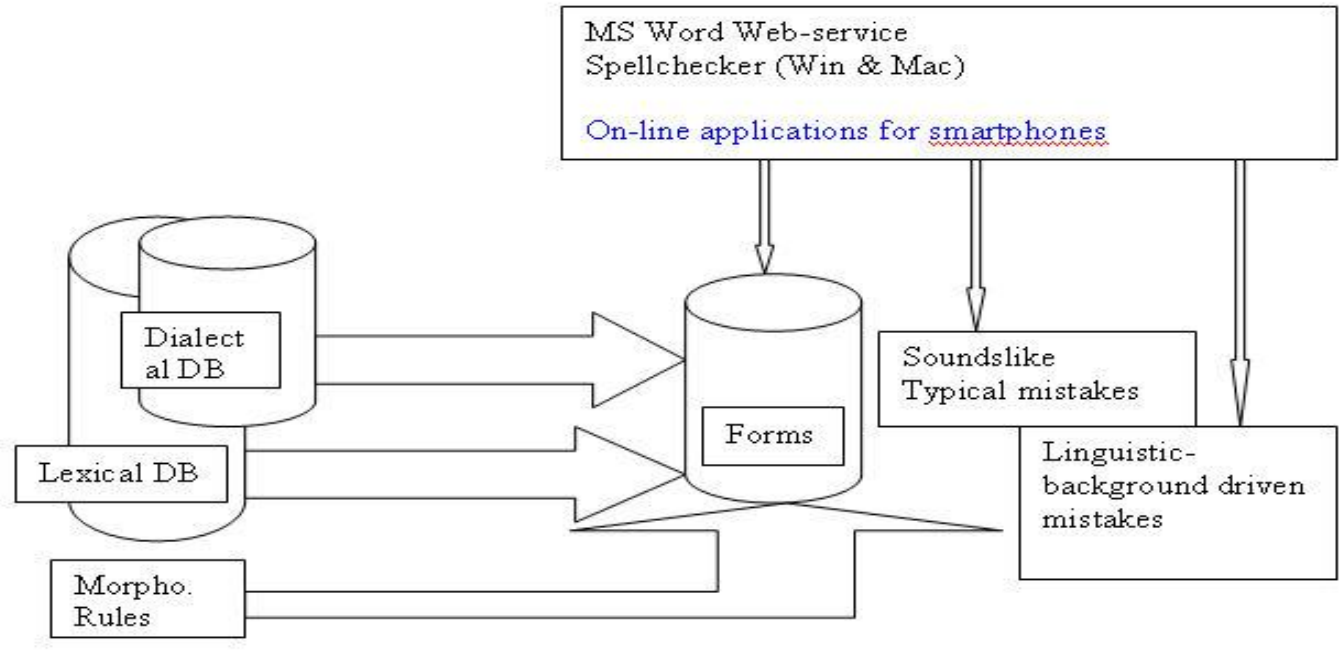
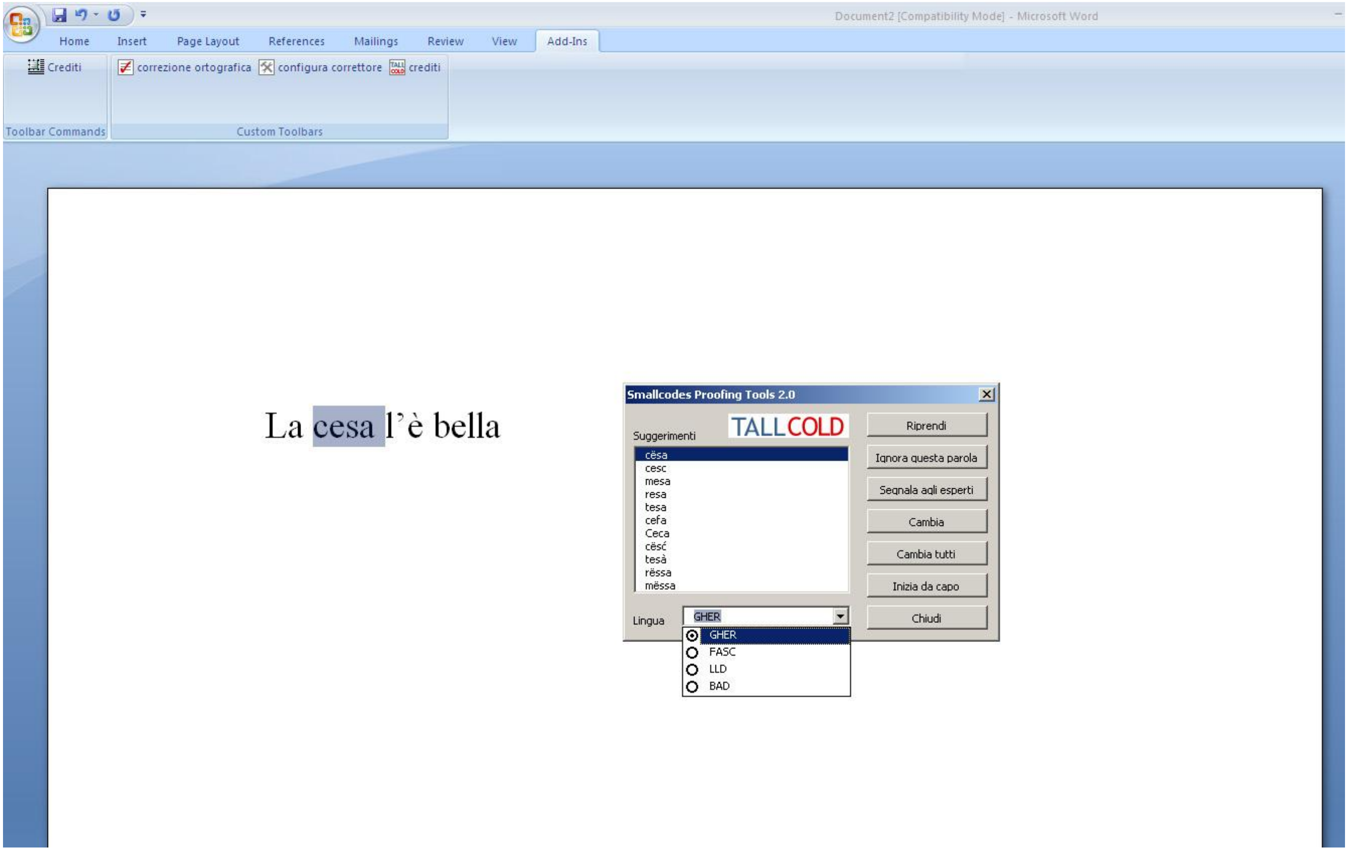


Fig. 5: functioning of an advanced spell-checking system



La cesa l'è bella

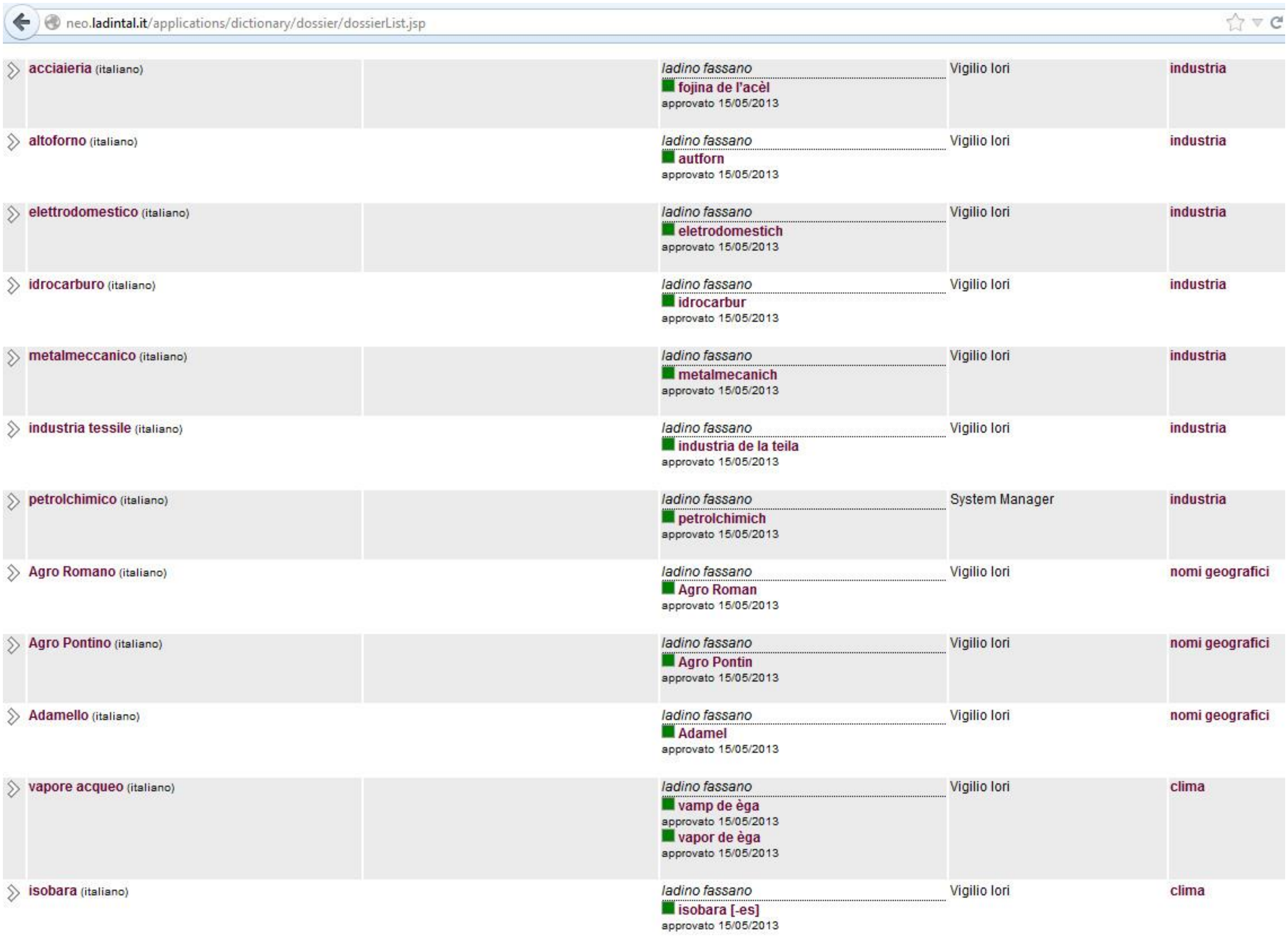
Suggerimenti

cësa
cesc
mesa
resa
tesa
cefa
Ceca
cësć
tesà
rëssa
mëssa

Lingua: GHER

● GHER
○ FASC
○ LLD
○ BAD

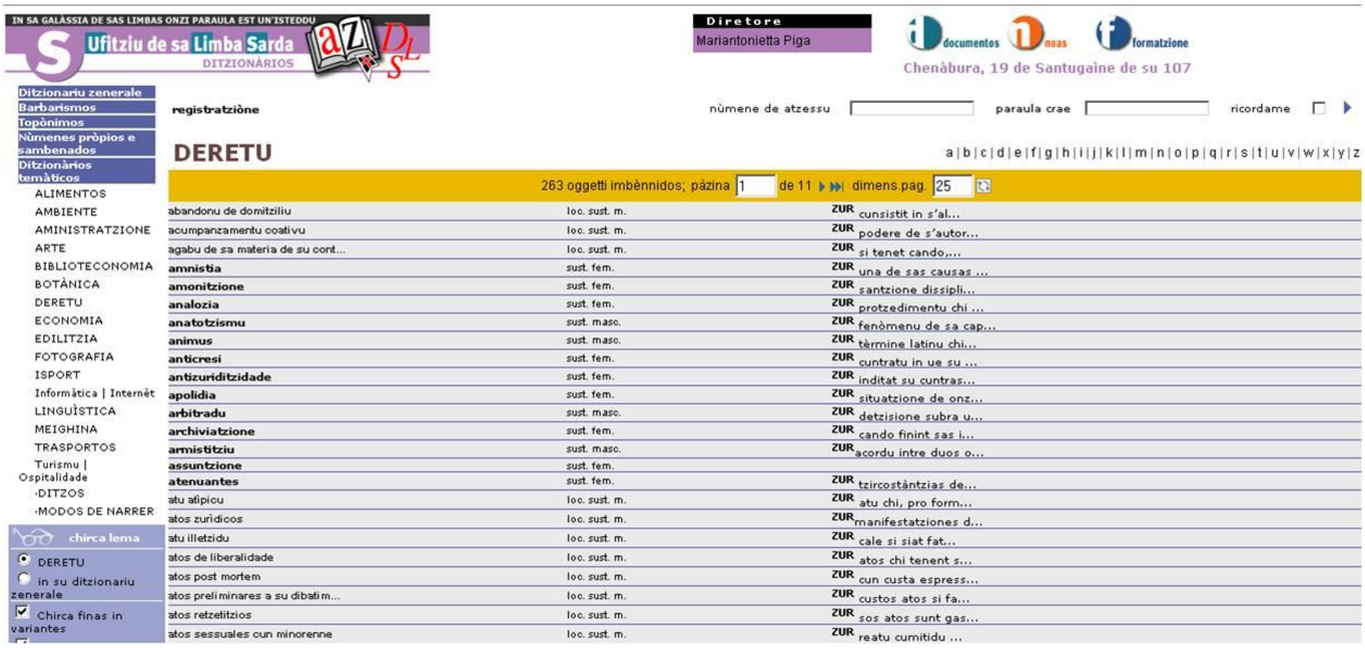
Fig. 6: Spell-checking of standard Ladin language with correction based on the typical errors caused by the three main dialectal backgrounds (corresponding to the three major oral dialects spoken in the respective alpine valleys: Gherdëina, Badiot, Fascian.



neo.ladintal.it/applications/dictionary/dossier/dossierList.jsp

acciaieria (italiano)		*ladino fassano* 🟩 **fojina de l'acèl** approvato 15/05/2013	Vigilio Iori	**industria**
altoforno (italiano)		*ladino fassano* 🟩 **autforn** approvato 15/05/2013	Vigilio Iori	**industria**
elettrodomestico (italiano)		*ladino fassano* 🟩 **eletrodomestich** approvato 15/05/2013	Vigilio Iori	**industria**
idrocarburo (italiano)		*ladino fassano* 🟩 **idrocarbur** approvato 15/05/2013	Vigilio Iori	**industria**
metalmeccanico (italiano)		*ladino fassano* 🟩 **metalmecanich** approvato 15/05/2013	Vigilio Iori	**industria**
industria tessile (italiano)		*ladino fassano* 🟩 **industria de la teila** approvato 15/05/2013	Vigilio Iori	**industria**
petrolchimico (italiano)		*ladino fassano* 🟩 **petrolchimich** approvato 15/05/2013	System Manager	**industria**
Agro Romano (italiano)		*ladino fassano* 🟩 **Agro Roman** approvato 15/05/2013	Vigilio Iori	**nomi geografici**
Agro Pontino (italiano)		*ladino fassano* 🟩 **Agro Pontin** approvato 15/05/2013	Vigilio Iori	**nomi geografici**
Adamello (italiano)		*ladino fassano* 🟩 **Adamel** approvato 15/05/2013	Vigilio Iori	**nomi geografici**
vapore acqueo (italiano)		*ladino fassano* 🟩 **vamp de èga** approvato 15/05/2013 🟩 **vapor de èga** approvato 15/05/2013	Vigilio Iori	**clima**
isobara (italiano)		*ladino fassano* 🟩 **isobara [-es]** approvato 15/05/2013	Vigilio Iori	**clima**

Fig. 7: An example of the work flow (with various status of approval) for the creation and consolidation of terminology in Ladin language: please note that the system is fully integrated with the dictionary module so that specific word-lists can be included or excluded from the general dictionary, exported for the Web or via Web-service for use within other applications.

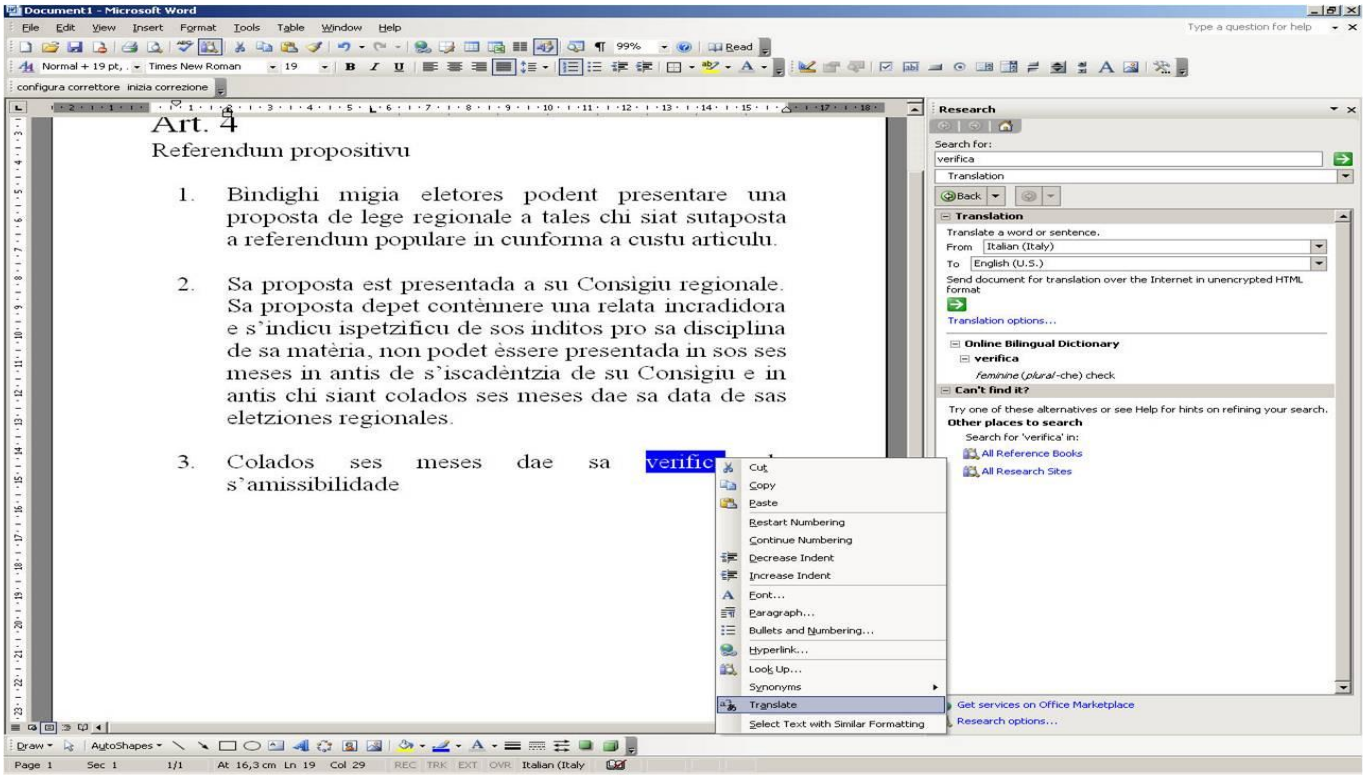


DERETU

263 oggetti imbènnidos; pàzina 1 de 11 dimens.pag. 25

abandonu de domitziliu	loc. sust. m.	ZUR	cunsistit in s'al...
acumpanzamentu coativu	loc. sust. m.	ZUR	podere de s'autor...
agabu de sa materia de su cont...	loc. sust. m.	ZUR	si tenet cando,...
amnistia	sust. fem.	ZUR	una de sas causas ...
amonitzione	sust. fem.	ZUR	santzione dissipli...
analozia	sust. fem.	ZUR	protzedimentu chi ...
anatotzismu	sust. masc.	ZUR	fenòmenu de sa cap...
animus	sust. masc.	ZUR	tèrmine latinu chi...
anticresi	sust. fem.	ZUR	cuntratu in ue su ...
antizuriditzidade	sust. fem.	ZUR	inditat su cuntras...
apolidia	sust. fem.	ZUR	situatzione de onz...
arbitradu	sust. masc.	ZUR	detzisione subra u...
archiviatzione	sust. fem.	ZUR	cando finint sas i...
armistitziu	sust. masc.	ZUR	acordu intre duos o...
assuntzione	sust. fem.		
atenuantes	sust. fem.	ZUR	tzircostàntzias de...
atu atìpicu	loc. sust. m.	ZUR	atu chi, pro form...
atos zurìdicos	loc. sust. m.	ZUR	manifestatziones d...
atu illetzidu	loc. sust. m.	ZUR	cale si siat fat...
atos de liberalidade	loc. sust. m.	ZUR	atos chi tenent s...
atos post mortem	loc. sust. m.	ZUR	cun custa espress...
atos preliminares a su dibatim...	loc. sust. m.	ZUR	custos atos si fa...
atos retzetitzios	loc. sust. m.	ZUR	sos atos sunt gas...
atos sessuales cun minorenne	loc. sust. m.	ZUR	reatu cumitidu ...

Fig. 8: Web page output of terminology module



Art. 4

Referendum propositivu

1. Bindighi migia eletores podent presentare una proposta de lege regionale a tales chi siat sutaposta a referendum populare in cunforma a custu articulu.

2. Sa proposta est presentada a su Consigiu regionale. Sa proposta depet contènnere una relata incradidora e s'indicu ispetzificu de sos inditos pro sa disciplina de sa matèria, non podet èssere presentada in sos ses meses in antis de s'iscadèntzia de su Consigiu e in antis chi siant colados ses meses dae sa data de sas eletziones regionales.

3. Colados ses meses dae sa verific
s'amissibilidade

Search for:
verifica
Translation

Translation

Translate a word or sentence.

From Italian (Italy)

To English (U.S.)

Send document for translation over the Internet in unencrypted HTML format

Translation options...

Online Bilingual Dictionary

verifica

feminine (*plural* -che) check

Can't find it?

Try one of these alternatives or see Help for hints on refining your search.

Other places to search

Search for 'verifica' in:

All Reference Books

All Research Sites

Fig. 9: Web-service output for word-to-word terminology translation.