Silvia Randaccio, Carlo Zoli – Smallcodes; Claudia Soria – CNR-ILC

Standardised orthography - A shoe for a bare foot

Fourth Cambridge Conference on Language Endangerment, 2014, July 4th

Smallcodes srl (www.smallcodes.com) is a company based in Florence and made by a group of six people who believed that working with minority languages was economically sustainable and even profitable. We were wrong about that. But we also believed, and we still do, that the way to preserve these minority languages was the use of technology. Technology, by definition, is one of the most standardised and homogenizing field. It may sound bizarre that our way to fight cultural assimilation is the use of the most standardising thing in the world, namely technology, but that is what we firmly believe. That is why, together with the Institute of Computational Linguistics of National Research Council, we gained some experience in the use of standardised orthographies for the design of digital tools for minority languages.
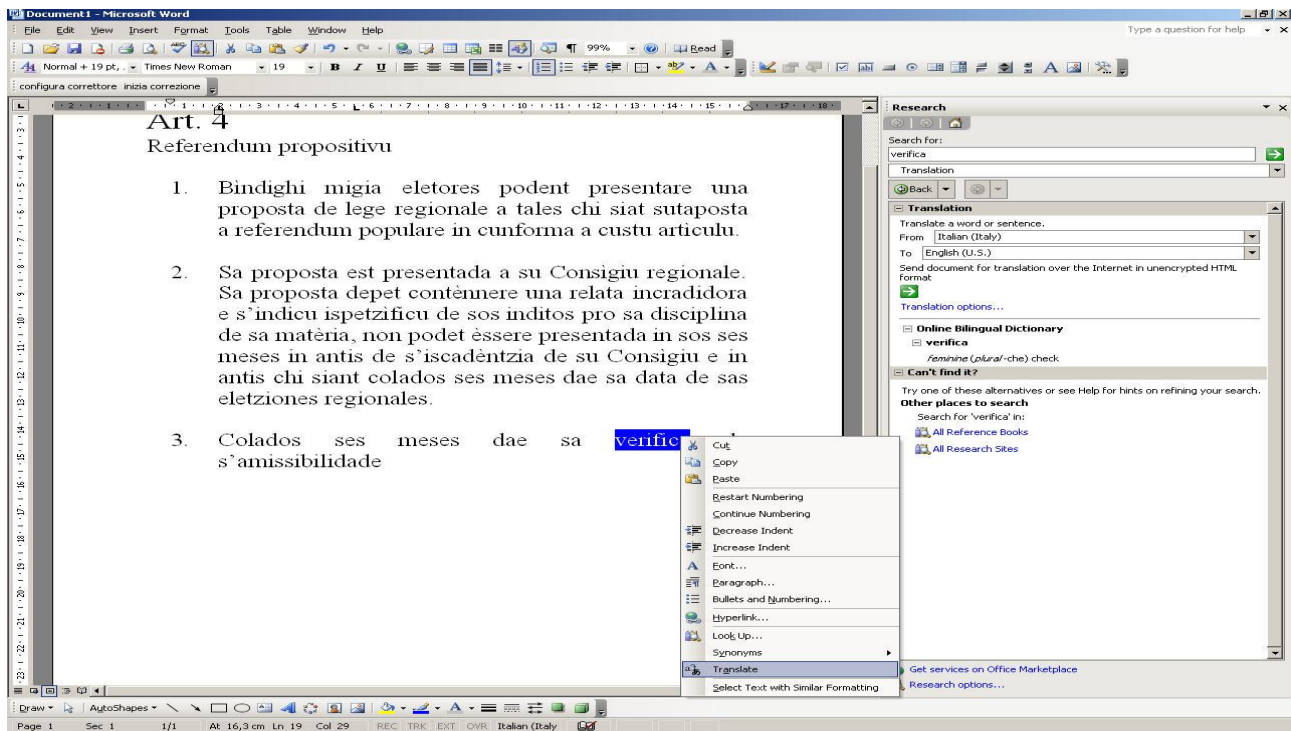
A few sentences of our *manifesto* summarize this concept:

- We believe that every serious program of cultural defence must start from language protection and that modernization must pass through a written – and possibly coherent and shared – form.

- We believe that digital technologies can largely contribute to the process of linguistic modernization and to the promotion and spread among youngest generations.

- We believe that, from a technological point of view, it is necessary to establish a high level of standardization in the creation of texts and databases. This is meant to assure a "long electronic life" to linguistic resources and to allow a real and easy exchange of information, resources and technologies.

How do we want to achieve these goals, and how? A few examples, both take from our experience with Sardinian language, can explain it.

The first one is a module we developed in order to create and support discussion groups for the elaboration of technical terminology. Normally, minority languages lack of technical words and, when they are raised to official languages (as it has been recently happening in Italy) they must necessarily have this kind of lexicon, which has to be created. In this way, we create discussion groups aimed at that, and at the end of the discussion one proposal is chosen and approved as official. This is a tool used by Sardinian Regional Linguistic office. The second picture is very clear: it shows a spellchecker we developed using morphology paradigms, which can be integrated into any writing software.

It is clear that these kind of resources, that we believe are very important for a small language such as Sardinian language, must necessarily admit only one standard form. A standard written form is necessary for every language to meet today's technological expectations: in fact only a standardised orthography can access modern tools such as Google Translator and the like. Moreover, writing is the only possible approach to the domains of administration, bureaucracy and education.

Therefore, language standardisation should not be regarded as an enemy of autochthonous languages but as a small price to be paid to support minority languages in the real world and in the cyberspace. If we want to save and preserve minority languages in the modern world, we must necessarily let them have access to the tools and resources of the same technological level as those of "bigger" languages.

The linguistic situation in Italy

The linguistic map of Italy, according to Unesco, shows the complexity of the linguistic situation in our country. As Tullio De Mauro says "In Europe, Italy has the highest rate of native linguistic diversity. This difference is not born recently or only a few centuries ago, but is rooted in the territory. A possible comparison is perhaps India" (2008). This quote tells the truth, even if in Italy this is not fully perceived.
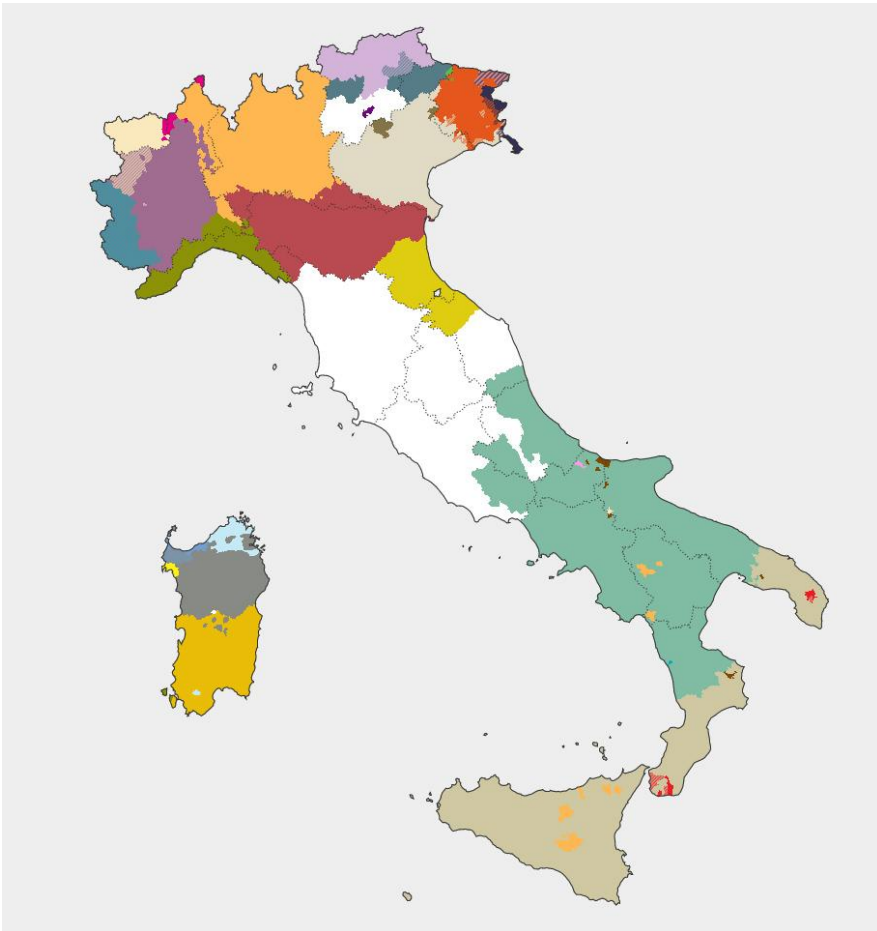
It sounds incredible, but Italian as a traditional language is spoken only in the blank space, and each of these colours has a different ISO code recognized by Unesco. Among these, for political and historical reasons, only twelve minorities have been officially recognized in 1999 by the Italian law. The others are still called, without any scientific basis, Italian dialects.

Smallcodes has worked or cooperated with most of these languages (especially with the twelve recognized languages for obvious reasons: they are those which mostly need technological support) as technologists, but also as consultants and together with the committees who developed standardized orthographies. We can therefore report about different approaches to orthography standardisation that were encountered over the years. About these, our engagement is direct and based on daily cooperation, with the implementation of didactic tools, of electronic dictionaries and more.

The approaches identified are: a top-down approach (as in the case of Fiulian and Dolomitan Ladin), a polynomic approach (as for Occitan), an "umbrella-like" approach (the case of Sardinian) and a mixed approach (as for Rromani). Of course there is a high degree of generalization in this categorisation, but these examples may be helpful to get into the problem.

Top-down approach

Using a top down approach means to apply the same method that was used with big national languages a few centuries ago. Nowadays, this is done ten time faster than before and in a completely different context where the spread of literacy has totally changed. With this approach, a prestigious variety is chosen among others to represent the whole language. The choice may be done according to the literary tradition or for geographical centrality, but there also cases in which an intermediate form is chosen among varieties.

In the case of Friulian, the common standard chosen was Central Friulian, spoken in the countryside around Udine (the city was instead too influenced by Venetian language). This standardisation now represents the common written language and, in some cases, even the spoken language of all Friuli. However, it has some features which are not general at all: one is the ending in feminine nouns in –e, not belonging to other varieties which have –a. And yet, this feature ended up becoming a hallmark of Friulian language – especially because it differs from the Italian form in –a - and it is commonly accepted.

Ladin Dolomitan was created taking into account the previous experience of Heinrich Schmid who standardised Romansh of Grisons in Switzerland. It follows the same principle: the creation of a variety which can be a landmark in written and also in spoken production. In this case, not an attested variety was chosen, but a kind of mixture of varieties.

Both examples can be used to point out another basic feature of a top-down approach in orthography: the standardisation of lexicon. This is a process which is still in progress even in big national languages. In Italian language, for example, both versions of the word for "chair" (*sedia* and *seggiola*) are theoretically accepted in standard language, but only one is increasingly emerging as the predominant. To see which of the two, one can simply have a look at an Ikea catalogue, where it is quite unlikely to find *seggiola*.

Polynomic approach

Occitan language has recently adopted a form of standardization based on literary tradition (Lamuela, 2008) but since a few years ago the common standard was the *Escolo dou Po* ("school of river Po"). This orthography was a very good example of polynomic orthography.

This approach was first introduced in French Corsica. Its aim is to reproduce all single local varieties, in terms of phonetics, lexicon, and even morphology, using a dialectological approach. It only tries to reduce the use of diacritics, accents, and an excessive number of graphic signs which are often used by local authors. So, this is not in all respects a form of standardization, but only a graphic indication for writing local varieties: the result is that the same word can be spelled differently according to the local use. In our opinion as technicians, this approach should be avoided, because modern technological tools need uniformity. Moreover, this approach can give a strong signal of inferiority to the language.

The new orthography of Occitan language is more unified and it is a real standard orthography with just a small degree of polynomy. The polymony is applied in this case only to the lexicon (and in this sense, this approach is the opposite of top-down) and sometimes to morphology when two forms are too different to be unified, but without exaggeration. Moreover, this orthography is more consistent with the literary tradition of the medieval Occitan troubadours.

"Umbrella-like" approach

A standard Sardinian language has started to be experimentally designed in the 80s. At that time, the dominant approach was top-down, with the normalisation of lexicon and reading. Time has shown that this approach cannot be used anymore because dialectal purism exists, even if we do not like it. This issue has been ironically called by some scholars a "Grandma's language syndrome": we all want to preserve the grandma's language, but it is always our own grandma and not someone else's one.

So the previous top-down approach has failed in Sardinia, and the LSU (*Limba Sarda Unificada* – "Unified Sardinian Language") has been changed into LSC (*Limba Sarda Comuna* – "Common Sardinian Language"). in 2006. The new LSC was designed with the aim of giving an "umbrella-like" approach to standardization.

The claim for oral standardisation was abandoned and this new orthography was studied to cover all local ways of speaking. This is a phonological and not phonetic spelling, and it is inevitably etymological. In fact, the "umbrella" which graphically covers all local forms such as *bennalzu*, *bennarzu*, *ghennariu*, *gennargiu*, ("January") cannot be found but in an hypothetic form of Proto-Sardinian *ghennàrgiu* which has ideally just left Latin to become something else.

With this kind of orthographies, the issue of morphology remains. In fact, morphology is always a problem for "umbrella-like" spelling systems because it is often impossible to standardise it. In Sardinian language, for example, the conditional tense is made with two different auxiliary verbs according to the geographical area. How can we find an umbrella that covers both? The question is still unanswered.

Mixed approach

Rromani language is not an Italian minority by definition but it can be said that it is the biggest European minority language and culture. It was not recognised by the law and it is still waiting for some form of recognition in Italy and elsewhere. Smallcodes' experience with Rromani language is due to the ongoing development of a teaching portal for the same scholar who created a new standard for Rromani language in the 90s (Courthiade, 1991). This was the first experiment of real unification of Rromani, after dozens of local orthographies spread all over the Rromani world, from Europe to Brazil.

This new approach has the aim of being an "umbrella-like" approach, but it has also a strong polynomic attitude, and that is why we can consider it a mixed approach. This polynomic attitude is necessarily strong because in the case of Rromani differences among varieties are not only caused by historical evolution, but also by geographical dispersion and by the tradition of gypsological studies which have failed in recognising the deep unity of Rromani language. Evidence has in fact shown that, if we do not consider the loanwords, original lexical differences cover some 20 or 30 words only.

Conclusion

As it has been pointed out, each linguistic situation facing the issue of standard orthography has its own peculiarities and the way to standardisation is hard. But standardisation is most probably the only possible way for the cultural redemption of minority languages. And in fact, even who rejects standardisation has to face the problem anyway. Such is the case of Italian Switzerland whose *Centro Dialettale della Svizzera Italiana*, whose approach is totally dialectological: during the construction of their dialectal dictionary, they also had to chose a main entry form for each word, and this was inevitably the intermediate form. Somehow, this act of choosing was a form of standardisation.

As a conclusion, it can be said that, despite all inconveniences and difficulties, standardization is fundamental because, as the title of our presentation says, "a standard orthography is like a shoe for a bare foot. At first it is a bit uncomfortable, but then the foot fits the shoe, the shoe fits the foot and, at that point, one can forge ahead in any terrain and in any weather".