

Localisation in the Basque Country

Luistxo Fernandez, Code & Syntax - LFernandez@codesyntax.com - <http://www.codesyntax.com>

The Basque Country, located in the border between Spain and France, is a small industrialised region of around three million inhabitants, out of whom one million or less speak Basque. This is the original native language of the region, an isolated pre-indoeuropean language, which has the undeserved reputation of being a too difficult language to learn. But this is a wrong perception. It happens to be very different from neighbouring languages in its syntax, morphology (with its distinctive ergative case mark), and the lexicon. Any Italian speaker reading a Catalan text may be able to roughly grasp its meaning. That would not happen with a Basque text.

Basque is a minority language even in its own territory, in terms of number of speakers and social functions. It is an endangered language that suffered a deep recession during the 20th century. After the dictatorship of Franco in Spain, a strong social and cultural movement emerged in support of the language. This movement was partly promoted by the newly elected local authorities, with the Basque Government as the most salient institution, and as a result of it over the past 20 years the language has made substantial progresses in areas such as education, the media, and the administration. Statistics about the command and usage by the population are also improving. But despite all these advances, as we enter the 21st century, it is becoming clear that new technologies must be applied to the language to guarantee its survival. At the frontline of the technological challenge is localisation.

Probably, most of us first heard of localisation from Microsoft around 1995. The Windows 95 and 98 operating systems, as well as other MS Office packages have since been localised into Basque. In the Basque Country several research groups and companies have contributed to this. But this effort has been totally subsidized by the Basque local authorities, and Microsoft has been the main beneficiary of it.

The localisation process applied the tools and procedures set up by Microsoft, and the participation of Basque experts was largely restricted to the translation task. Consequently, the know-how transferred to the local participants was more related to terminological issues than to real localisation procedures. One of the problems was that the Basque localisation effort was undertaken by the Spanish division of Microsoft, which had more concern towards general marketing matters regarding Spain than with localisation proper. Basque localisers detected some deficiencies that they could not mend, such as, for example, the syllabic partition of the MS Word processor, which did not meet with Basque standards.

Microsoft, who is reputed for its well engineered i18n and l10n projects, did not give much priority to the issue. There was no real technical problem, nor was the Basque locale too complex. The real matter was one of organisation, different priorities, and little concern for real localisation, besides getting the check from the Basque administration.

On top of this, the constantly-upgrading pace of Microsoft software made the life and profit of the localised products extremely short (at the expense of Basque taxpayers, of course). Apart from that, marketing and distribution were awful.

Local authorities seem to have switched channels recently: Now, it is time to localise free software! There is already a Mandrake Linux localised version being distributed, and the Open Office suite will come next. This new effort seems fine, at least Mr. Gates will get zero royalties. However, the new official push goes in the same direction, the localisation of 'big' desktop infrastructure. In my view, this 'big-desktop' strategy leads nowhere. I would say that we are localising just menus. It is okay, but in the end, what is what we really need? I, personally, need no Basque menu. What I really need is Basque food! A word processor with Basque menus is fine, but if it cannot cut Basque syllables correctly, what kind of Basque text am I going to write with it? In my opinion, the real prospects for Basque localisation should focus in another direction, and not towards the 'big-desktop' effort.

My hopes for the future are on research teams specialised in natural language processing (NLP), and on some small companies working around that field. Multidisciplinary teams of linguists and computer scientists have allied in Basque universities and have produced interesting results in areas such as speech recognition and synthesis, morphological and syntactic analysis, spell checking...

These advances are not what we would call localisation, properly, because they pertain to another area of language engineering. Yet, they provide the best basis for true Basque localisation. Translating the huge amount of message-strings that any operating system contains is a very expensive task; however the benefits are dubious. Adapting a web application server to handle dates under the Basque style, or being able to lematise the root of a term in a search facility (a very important aspect for any Basque locale due to the postpositional nature of its morphology) is a much more interesting accomplishment. It is nice to be able to develop websites or Internet services that can handle such Basque little quirks. It is in that manner that language engineering can foster localisation. Fortunately, some small companies

have joined NLP research teams to develop this kind of practical localisation software for Basque; and it is in the internet where it is becoming most visible in the form of Basque and multilingual sites.

This trend brings hopeful prospects to the Basque localisation community. There is a big concern for multilingualism from the part of the administration and other public institutions, such as universities, museums, social services and the like, since they need to communicate at least in Basque and Spanish, both internally and with the society at large. It is in the realm of I-net applications that these needs will have to be solved, in the first place.

Localising something like Gnome is a formidable task, with results visible almost to nobody. Localising a reporting tool for an intranet or a public website is a much easier undertaking, and besides has immediate and much more visible results. The desktop application must be looked for, acquired, installed, only to find out in the end that a new upgraded version has just been released. Conversely, the localised facility on a web site can be very easily accessed and put at work through a simple click on the *Language Change* button of the interface.

There is a clear need for localisation, and it is up to Basque localisation companies to join NLP research teams to meet this demand. This has, in turn, some important strategic implications. From the point of view of the skills, the morphological properties of Basque and Spanish are different enough as to make the task of developing bilingual products an intellectually challenging activity. It is almost as making a localisation test. Basque morphology very often creates length-problems, similar to those faced in German localisation, and there are also some tricky syntactic aspects, such as inverse word-order, post-positional case marking (similar to that of Finnish), etc. The expertise we may have will help very little with encoding issues (as those posed by Asian languages), but will give us a good basis for localising on a western-European scale, at least.

In my opinion, developing a Basque localisation expertise will not just provide solutions for this marginal language that we still speak, but will be an opportunity for our companies to open to wider markets, ranging from local exporter companies (the local industry is quite internationalised), to the participation in European ventures and projects.

This in itself is quite an important challenge.

Basque alone does not offer a big enough market to sustain a company. The 'big-desktop' strategy, the localisation of either proprietary or open source fat systems, is only sustainable through public subsidy. The ever-upgrading nature of such systems shows no prospect of taking us beyond this subsidy-policy, and offers, if any, very far-off benefits, which will anyway fall into the hands of MS royalty collectors, or the free-software community. It will hardly pay back to the local users, who will see almost no change on their computer screens. It is by no means a sustainable strategy.

A sustainable policy should be based on real market opportunities, not on subsidies. The globalized web is that true market. Multilingual solutions for that market is where Basque localisers should focus on. The conjunction of our local need for multilingualism together with our locally attained skills will plot the route to develop solutions for others. And it is the outside market that will, in turn, sustain our own local effort.