



Conference on Bi- and multilingual universities – Challenges and future prospects

Saturday 3 September 2005

**Location: Metsätalo, Unioninkatu 40,
University of Helsinki**

SESSION 4: Language Policy and Minorities

Chair: Professor Pekka Sammallahti

Minority languages in Europe: The role of universities in creating a dynamic for the future

Miquel Strubell
Universitat Oberta de Catalunya
(Barcelona, Catalonia, Spain)
mstrubell@uoc.edu



Abstract: <http://www.palmenia.helsinki.fi/congress/bilingual2005/abstracts/Strubell.pdf>

A number of studies show that the demographic strength of many so-called "regional" or "minority" language groups is declining across Europe, despite a trend towards greater acceptance in mainstream discourse of the concept of linguistic diversity.

An example is given of a university that owes its very existence to the refusal to allow the use of a language other than the official language of the state in a state-run university, and to the response of a regional government committed to supporting its own language. This is the *Universitat Oberta de Catalunya*.

The role of a university in helping to develop a suitable dynamic favourable to the recovery of a so-called "minority language" is discussed in the framework of a model of sociolinguistic change that puts the individual and his/her perceptions at its centre.

...ooo000ooo...

Index

- [1. Introduction](#)
- [2. The Catalan case](#)
- [3. The Universitat Oberta de Catalunya \(Open University of Catalonia\)](#)
- [4. Research on multilingual tools at the Universitat Oberta de Catalunya](#)
- [5. Conclusion](#)
- [6. Some relevant bibliography](#)

1. Introduction

Since World War II there has been a growing movement among speakers of so-called "minority" languages (called "regional languages" in France; Nelde et al 1996; Williams et al 1997) calling for language rights. The initial aims of such movements have been varied, though it is probably true to say that in most cases the first claims have been for the language to be taught in the schools, including pre-schools, and/or for the language to be a (or the main) medium of instruction in such schools.

This we can record in place as far apart as **New Zealand** (on the left below, the photograph shows a Kōhanga reo – immersion – class, <http://www.nzhistory.net.nz/Gallery/tereo/history.htm>) and **Catalonia** (on the right below, the poster is from a 2003 campaign, <http://www.pelcatala.lateneu.org/>). Catalan-speakers and those wanting their children to learn the language at school had to campaign actively to see their language return to the schools in the Balearic Islands, the region of Valencia, and Catalonia proper.



In other cases, or at other times, claims have been for road signs (left, <http://www.llgc.org.uk/yngyrchu/Iaith/TyngedIaith/index-e.htm>; and/or the official use of the language, (right, <http://cymdeithas.com/deddfaith/>).



As you all know, official language planning agencies have been set up in a number of countries in order to redress the historic marginalisation, and even repression, of these languages.

The picture on the right (<http://www.gtj.org.uk/search/simpleSearch.php?srch=Welsh+not&lang=en>) shows the "Welsh Not", which used to be hanged around the neck of any school child caught speaking in Welsh; he or she would try and overhear a colleague doing the same, so as to pass on the punishment.



Despite these movements, the number of speakers of many such languages has been declining, often severely. It is reported that the number of Breton-speakers dying every year is now about 11,000; while little more than this number of schoolchildren of all ages are being educated bilingually. The INSEE 1998 survey (Clanché 2002; Héran, Filhon & Deprez 2002) for instance, revealed tragic declines in the speakers of all languages in France other than French, during the whole of the 20th century.

Curiously enough, (or perhaps simply because a three course meal does not begin with the dessert!) the demand for higher education through the medium of a "minority" language has not been at all widespread. The introduction of such languages into higher education probably started in two fields: the study of the languages themselves and their accompanying culture; and the training of school teachers.

This is fairly evident in Wales and Ireland, for instance, where few if any degrees in Engineering, Business Studies or the Sciences are offered through the medium of Welsh or Irish. (The picture below shows the [National University of Ireland, Galway](#))



[Return to the top](#)

2. The Catalan case.

Let us look at Catalan in more detail. In this case, at least in Catalonia proper, the (re)introduction of the local language as a medium of university instruction has not been limited in this way. For Biology degrees, for instance, nearly all lectures are in Catalan. Degrees in which the use of Catalan seems to be lowest are Law, Nursing and Veterinary Science, the first perhaps due to lack of confidence among lecturers, most of whom were of course trained monolingually in Spanish; the latter two because of the high proportion of non-Catalan students attending Catalan universities in this subject.

The issue of the use of Catalan gave rise to an outcry in late 1976 or 1977, when in a memorable *Paris Match* interview, the then-Prime Minister Adolfo Suárez was asked about the hypothetical use of Catalan at upper secondary level. In his dire ignorance (the recovery of Catalan in this field began even at university level towards the end of the Franco regime, in the early 1970s) he laughed it off: who, he asked, could even dream of nuclear chemistry being taught through the medium of Catalan?¹ Nuclear scientists and lecturers in many other

¹ "Se hará el bachillerato en vasco o en catalán?"

fields, largely scientific, protested vociferously: not only was such use possible, but it was actually frequent in all university faculties.

Thus it should be clear that for Catalans, reconquering higher education for the language was an important, and not just symbolic, objective.

It will be helpful at this point, I hope, to briefly describe the historical context of the Catalan language.

Curiously enough, the modernisation of Catalan, almost two centuries after having been banned for official use (except in Minorca, where for nearly all the 18th century it continued, under British rule, to be the official language), turned out not to be entrusted primarily to the university. This is not as surprising as it seems: at the same time that Catalan was banned by Philip V of Castile, all the Catalan universities were closed. Barcelona University (founded 1450) was razed to the ground, and was not re-established until over a century later, in 1837. So it might be argued that at the start of the 20th century, when the 1st International Conference of the Catalan language was held (in 1906), the Catalan authorities had no other choice than to set up a language academy outside the university. This was to be the "Secció Filològica" of the "Institut d'Estudis Catalans".

The driving force behind this modernisation was in fact a chemical engineer called Pompeu Fabra (1868—1948; photographs below), who held a chair in Bilbao (Basque country) from 1902 to 1911, yet who had already for years been a leading authority in linguistics, having studied in depth the grammars of other Romance languages.

The work of Fabra and his team culminated with the edition in 1932 of the [*Diccionari General de la llengua catalana*](#), a normative dictionary which coincided with the decentralisation of the control of Barcelona University, during the Spanish Second Republic (1931-1939). The board of directors of Barcelona University was re-elected, half being appointed by the Spanish Ministry of Education, and half by the (restored) Catalan government, or [*Generalitat de Catalunya*](#).

Given that the official status of Catalan was also restored in 1932, it was fitting that the institution became a formally bilingual university, each lecturer being free to choose whether to



Pompeu Fabra as a young man
<http://www.xtec.es/~malons22/personal/llengua%20al%20xx.htm>

Adolfo Suárez: "Su pregunta –perdóneme- es idiota. Encuéntreme en primer lugar profesores que puedan enseñar química nuclear en vascuence."

Adolfo Suárez, Spanish Prime Minister (1976)

Quoted from:

<http://www.behatokia.org/graf/..%5Ctxostenak%5CLINGUISTIC%20RIGHTS%20AND%20EDUCATION%20IN%20THE%20BASQUE%20COUNTRY.pdf>

The probable source is *Paris Match* no. 1466, July 1 1977.

lecture in Catalan or in Spanish, and all students being free to write their papers and examinations in either language. Unfortunately (or to be more precise, because of the Franco rebellion in 1936), this experience was short-lived. In 1939 there was a widespread purge of university teaching staff, many of whom chose to flee the country before the fascist troops reached Barcelona. All too many of these never returned. Those who became responsible for teaching the university students from then on had one thing perfectly clear in their minds: not to use Catalan in public.



© Arxiu Català - Josep Lluís Domènech - Barcelona - 2004

www.arxiucatala.com/pompeu.htm

At that time, the police also had a clearly defined role in regard to those pursuing higher education:



A widely distributed photograph taken by Manel Armengol (1977)

http://www2.ub.edu/comunicacions/revista_launiversitat/revista_25/Revista_25/universitat_3.htm

Please forgive this short digression on modern Catalan history. It was however necessary, I felt, to contextualise the next section of my paper.

[Return to the top](#)

3. The Universitat Oberta de Catalunya (Open University of Catalonia)



This section begins with the founding in 1973 of the [Universidad Nacional de Educación a Distancia, UNED](#), the equivalent of the British Open University. Despite growing demand, this university refuse point blank to offer courses through the medium of any language other than Spanish. In the early 1990s, the Catalan government negotiated with the Spanish Socialist government of the day the transfer to Catalan control of the UNED's operations in Catalonia. However, this proved to be legally impossible: according to its statutes

the UNED has to offer its courses throughout Spain. The Catalan government thus decided to set up a distance university of its own, with its first degree courses in Catalan. My university, the *Universitat Oberta de Catalunya*, thus owes its very existence to the vitality of Catalan, and to the stubbornness of those who refuse to accept the multilingual nature of Spain.

UOC had no particularly acute linguistic problems when it developed its University courses. It set up a language service



primarily to ensure the quality of the educational and operational texts, and made sure its full- and part-time staff were proficient in Catalan, so that it could function quite smoothly. The necessary lexicon was available for all the degree courses, which being distance courses do not include laboratory-based sciences, but do cover computer engineering and audiovisual communication, for example. Little research was needed in this sphere. So there were no language impediments to the growth of the university, which now has 21,300 undergraduates studying through the medium of Catalan, as well as about 2,400 graduates studying for postgraduate degrees, including masters and Ph.Ds... and a well-established reputation for quality and innovation.



Many other universities also have their language services. Students also coordinate activities to promote the use of the language at university level.

There was though one limitation: the size of the market to which the university could offer its courses. Catalan-speakers amount to a little over 7 million people, several million of whom, however, have Catalan as a second language. Furthermore, thanks to the memorable and enduring efforts of the Franco dictatorship, illiteracy in the language was generalised, and even today, thirty years on, a very high proportion of speakers over the age of say 40 feel extremely insecure when asked to use Catalan in writing.

The university coped with this by adopting the same principle as in nearly all other Catalan universities: students are free to write their papers and their exams in Spanish if they prefer. They can also correspond with the lecturer in either language. The number of Catalan-readers, as you can well imagine, is much higher than the number of Catalan-writers, so this practice broadened the potential market.

It was the Catalan government, whose annual grant provides the bulk of the University's income, that asked the university to try and extend into self-financing activities, and as a result, in association with a leading publisher, the university decided to start translating its course material into Spanish and to offer its degrees to Spanish-speakers across Spain. There are already more than 5,000 students enrolled on these courses. The name of the operation, "**Iberoamerica**", is somewhat misleading, for nearly all its undergraduates are

from Spain. Postgraduate students, on the other hand, do come from a wide range of countries: there are over 500 in Mexico alone, for instance.

[Return to the top](#)

4. Research on multilingual tools at the Universitat Oberta de Catalunya

It is in this context that one can start to describe UOC as a “bilingual” university, and I would like to devote most of my remaining time to outlining two interesting projects undertaken by colleagues of mine. One is on bilingual (Catalan-Spanish) discussion groups chosen from our campus (which, incidentally, is 100% internet-based) and how to cope with problems facing machine translations of the contributions to these groups; and the other is on the issue of the machine translation of educational and operational texts between Catalan and Spanish.

As regards the first ([Climent et al. 2003b](#)) the [Interlingua](#) project set out to help solve a recurrent problem in many linguistic contact situations: the tendency of speakers of the subordinate language to switch to the dominant language when they perceive that an interlocutor (even one in a whole group) is a dominant-language speaker, however bilingual that person may be. If machine translation can produce versions of sufficient quality, it might be invaluable in increasing the presence of a subordinate language, in this case Catalan. However, machine translation



http://www.comprendium.es/index_home_ca.html

systems rely on the use of standard forms of language and fail to translate spontaneous e-mails – which tend to contain spelling mistakes, typing errors, etc. - accurately. Thus before the translation can be undertaken, any written text must first be corrected, so that non-standard forms can be recognized.

The research project was based on a sample of 533 messages sent by 254 users of a computer science newsgroup on our virtual campus, in which students were free to exchange information, and opinions related to software, computers, etc. 76% of the messages were in Catalan, the rest in Spanish. Six evaluators (all bilingual full-time language consultants) analyzed machine translations of each of the contributions, in order to detect errors attributable either to the system or to the text written by the student.

Before being translated, each message was analysed by a special program in order to ascertain the language being used in each message. It is a refinement of a program, **TextCat**, developed by Gertjan van Noord (<http://odur.let.rug.nl/~vannoord/TextCat/>) and is based on the frequency of unigrams, bigrams and trigrams in a given text, and a comparison of the outcome with existing models for various languages.

Errors in the translations were classified into subdivisions of three basic categories:

- (a) those due to unintentional non-standard features (such as mistypes or deviations from the language norm),
- (b) those due to intentional non-standard features (such as code-switching or spelling innovations), and
- (c) those due to missing terminology.

The spell-checker was developed in both automatic and interactive modes. It corrects on the basis of similarities between a word not included in the dictionary and entries in the list and can select the most likely equivalent, based on criteria of similarity and of the relative frequency of various kinds of mistake. Here is an example given by the authors in another paper ([Climent et al. 2005](#)):

Aixo és un exemple de correccio de misatges de correu electronic

Aixo	això : aixe : ixo : així : baixo : pixo : aixà : aixol : mixo : aixa : fixo
correccio	correcció
misatges	missatges : visatges : miratges
electronic	electrònic

This gives the (correct!) version of the text:

Això és un exemple de correcció de missatges de correu electrònic

The check also contains a list of frequent mistakes that Catalans make, and their automatic corrections. Examples include "desde" (an interference from Spanish) which should be "des de", or "dongués" (a dialectal form) instead of "donés".

The authors comment that most errors seemed to be due to a poor command of Catalan, even though their sample consisted of highly-educated users. Their final recommendations were twofold:

- For **Catalan**, the main efforts should focus on automatic correction of unintentional deviations from language norms, mainly mistyping, orthography and mistakes caused by language contact. The results show that such unintentional deviations represent more than three times the amount of intentional non-standard deviations.
- The situation for **Spanish** was more balanced, which means that, in terms of frequency, efforts should be focused on feeding the MT system with terminology, as the incidence of errors due to missing terminology is slightly higher than that of unintentional and intentional non-standard deviations.

In conclusion, they underline the need for perfecting MT tools so that they can be fine-tuned to make them error-proof, yet also flexible enough to deal with the non-standard features typical of the e-mail register. Failure to do so would make it much more difficult for Catalan to survive in spaces such as forums in which speakers of other languages are present.

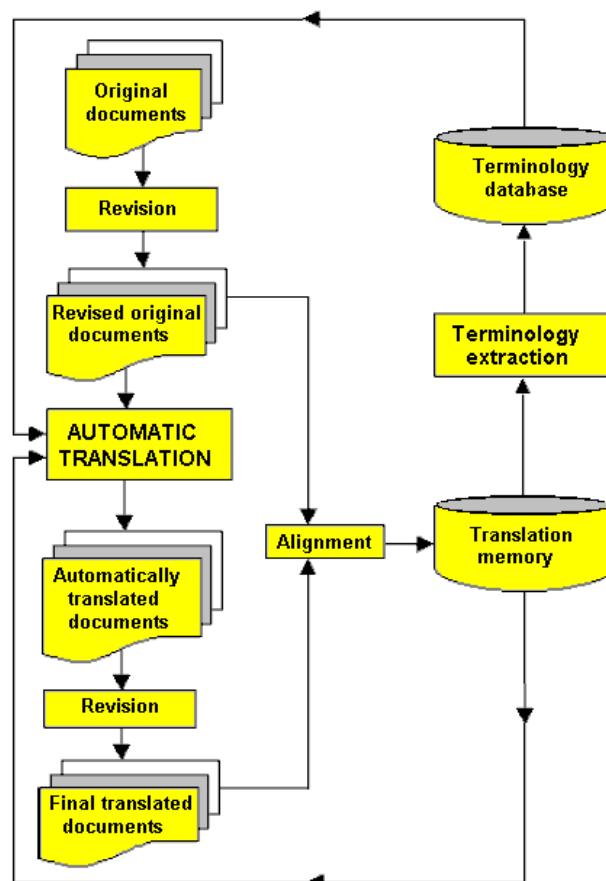
A second project that members of the same team worked on ([Climent et al. 2005](#)) involved much more formal texts closely related to the dynamic of individual academic subjects. Apart from the academic texts themselves, which form the core of the reading material for each subject course, there are a number of texts which change to varying degrees every semester, such as the course guidelines, the examination questions, and the continuous assessment exercises and assignments. As soon as the Spanish versions of degree courses began to be offered, steps had to be taken to ensure that apart from the difference in price (only the degrees in Catalan are subsidized by the Catalan government), there would be no differences in course content, requirements or standards. Moreover, huge amounts of such documents (in doc, rtf, txt and html formats) have to be translated into Spanish every semester in very short time-frames. The **TAA CV (Traducció Automatitzada a Campus**

Virtual, that is, Automated Translation on a Virtual Campus) project was designed with two aims in mind:

- a. to guarantee the linguistic quality of texts produced using an automated translation system, in order to drastically reduce the amount of human intervention needed to yield a printable outcome; and
- b. to provide the tools and training needed for the university staff involved in this work to be able to manage both the pre-edition and post-edition phases of the automated translation system.

One of the most important by-products of the project will be a bilingual terminology corpus, classified thematically, which can have a range of practical applications, including its possible expansion to include other pairs of languages. A program has been designed to automatically extract terminology from texts (n-grams) and terminological glossaries are built using a second program, *tond*, which allows translated equivalents to be found.

For the system to work, a range of supplementary computer programs is needed: a language detection program, correctors for both Catalan and Spanish, a terminology extractor, a translation equivalents search engine, and support programs for editing. The whole process is organized following a strict procedure:



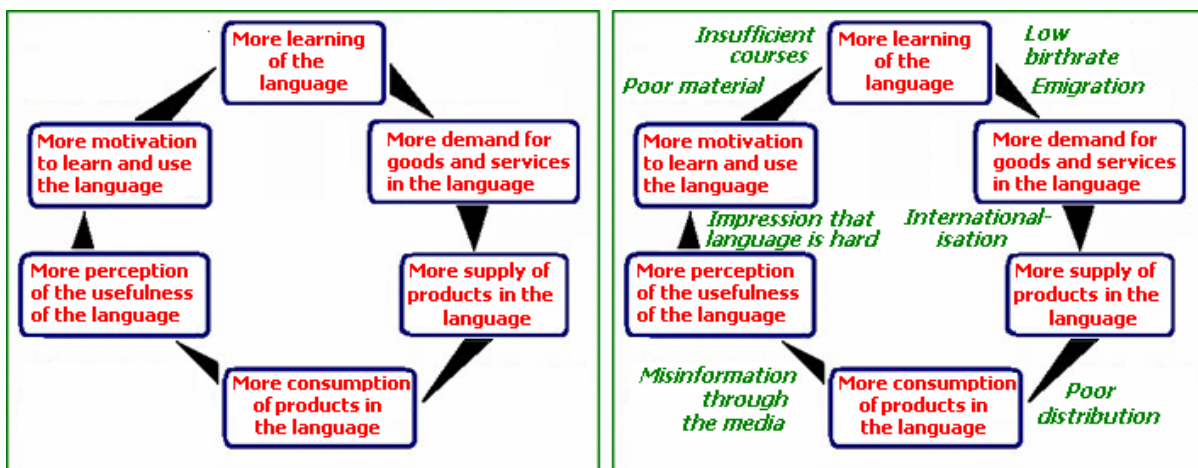
Among the aids for postediting, one is well worth mentioning, even though it has nothing to do with bilingualism: an automatic URL checker, which searches for the new URL when one proves to be inactive or incorrect. The authors say that it saves a great deal of work.

As should be clear from the procedure shown above, the corrected original text and the final version of the translation are incorporated into a translation memory which, following Moore (2002) can identify memorized sentences or segments in future tasks.

[Return to the top](#)

5. Conclusion

To the extent that the Open University of Catalonia has filled a gap in the availability of services in Catalan – in this case, distance university education – it is contributing to the prestige of Catalan, as a language which performs most if not all the high prestige functions of a modern society. Though it may seem contradictory, some of this prestige is growing out of the very fact that Catalan has increasingly, in our university, to co-exist with Spanish (and to a much lesser extent, with English) or to put it another way, there is an increasing need for tools which allow Catalan-speaking students to correspond with non-Catalan-speakers without the former having to abandon their language for this purpose.



Within the Catherine wheel model (Strubell 2001b; 2001c) such developments easily find their place. They can help to raise the profile of the language, making it more visibly useful and thus increasing motivation to learn and use it.

We hope to contact other universities at this conference who are having to face the same kinds of problems, with a view to establishing links and agreements and thus to advance together towards a world which comes to realize that its linguistic diversity is one of its greatest and most cherishable, yet also perishable, assets.

Kiitos paljo!

Tack!

Thank you!

Moltes gràcies!

[Return to the top](#)

6. Some relevant bibliography

- Clanché, François (2002). Langues régionales, langues étrangères: de l'héritage à la pratique. *INSEE Première*. No. 830. 4 pp. http://www.insee.fr/fr/ffc/docs_ffc/IP830.pdf
- Climent, Salvador; Miquel Strubell, Marta Torres & Glyn Williams (2002). "The Atlantis Observatory: Resources Available on the Internet to Serve Speakers and Learners of Minority Languages"; published in the *Proceedings of the Workshop Portability Issues in Human Language Technologies*, in *Language Resources and Evaluation Conference* (LREC, Las Palmas), <http://www.lreconf.org/lrec2002/>
- Climent, Salvador; Joaquim Moré & Antoni Oliver (2003a) *Building an environment for unsupervised automatic e-mail translation*. Working Paper Series WP03-002. Barcelona, IN3-Universitat Oberta de Catalunya. <http://www.uoc.edu/in3/dt/20244/index.html>
- Climent, Salvador; Joaquim Moré, Antoni Oliver, Míriam Salvatierra, Imma Sànchez, Mariona Taulé & Lluïsa Vallmanya (2003b). "Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation" *Journal of Computer-Mediated Communication*, 9 (1) November 2003. <http://jcmc.indiana.edu/vol9/issue1/climent.html>
- Climent, Salvador; J. Moré, A. Oliver, M. Salvatierra, I. Sànchez & M. Vázquez (2005). Technologies de la traducció per a la gestió de la doble oferta docent en català i castellà a la UOC. *Zeitschrift für Katalanistik/Revista d'Estudis Catalans* 18. p. 31-57 (ISSN 0932-2221). http://latina.phil2.uni-freiburg.de/pusch/zfk/18/04_Climent.pdf
- Héran, François; Alexandra Filhon & Christine Deprez (2002). La dynamique des langues en France au fil du XXe siècle, *Population et Sociétés*, No. 376: 1-4. http://www.ined.fr/publications/pop_et_soc/pes376.pdf.
- Moore, Robert (2002). "Fast and accurate sentence alignment of bilingual corpora", in *Machine Translation. From Research to Real Users (Proceedings of the 5th Conference of the Association for Machine Translation in the Americas, Tiburon, California)*. Heidelberg: Springer, pp. 135-244.
- Nelde, Peter; Strubell, Miquel & Williams, Glyn (1996). *Euromosaic. Production and reproduction of minority language groups in the EU*. European Commission, Luxembourg. See www.uoc.edu/euromosaic/
- Strubell, Miquel (1996). "Language planning and classroom practice in Catalonia", *Journal of Multilingual & Multicultural Development*, vol. 17, no. 24, pp. 262-275. <http://www.multilingualmatters.net/jmmd/017/0262/jmmd0170262.pdf>
- Strubell, Miquel (coord.)(2001a). *Lesser-Used Languages in States Applying for EU Membership (Cyprus, Czech Republic, Estonia, Hungary, Poland and Slovenia). Summary*. Education & Culture Series. EDUC 106A EN. Directorate-General for Research, European Parliament. Strasbourg. http://www.europarl.eu.int/workingpapers/educ/pdf/106a_en.pdf and http://cv.uoc.es/~grc0_003638_web/106a_en.pdf
- Strubell, Miquel (2001b). "Minorities and European Language Policies", a *Minorités et l'aménagement linguistique, Plurilingua XXII*: 45-58. http://www.geocities.com/m_strubell/viena.htm
- Strubell, Miquel (2001c). "Some aspects of a sociolinguistic perspective to language planning" in Kees De Bot, Shaak Kroon, Peter H. Nelde & Hans Van Der Velde (eds.) *Institutional Status and Use of National Languages in Europe. Plurilingua XXIII*, Asgard Verlag, Sy. Augustin: 91-106. ISBN: 3427642313.
- Strubell, Miquel (coord.)(2002). *The European Union and Minority Languages*. Education & Culture Series. EDUC 108 EN. Directorate-General for Research, European Parliament. Strasbourg.

Strubell, Miquel & Marta Torres Vilatarsana (2002). "El projecte Atlantis: Recursos digitals per a les llengües minoritzades de la UE, Recursos per a l'ensenyament del català". *Digit·Hum, Revista digital d'Humanitats*, vol. 5, 2003. (Comunicació presentada al *3r Simposi sobre l'ensenyament de català a no catalanoparlants*, Vic 46 de setembre de 2002.)
<http://www.uoc.edu/humfil/articles/cat/atlantis0303/atlantis0303.html>

Strubell, Miquel & Marta Torres Vilatarsana (2002). "L'Observatori Atlantis: recursos vinculats a les noves tecnologies per les llengües minoritzades de la UE". Comunicació presentada al *Congrés Mundial Sobre Polítiques Lingüístiques*, Barcelona, 16-20 d'abril de 2002.
<http://www.linguapax.org/congres/taller/taller5/Torres.html>

Strubell, Miquel (2004). "Small is beautiful: Bringing higher education closer to the people", in Caoilfhionn Nic Pháidín & Donla uí Bhraonáin (eds.) *University Education in Irish: Challenges and Perspectives*. Dublin, Fiontar, Dublin City University: 25-38.

Williams, G.; P. Nelde & M. Strubell (1997). "Del diagnòstic al remei: Euromosaic, llengua i desenvolupament econòmic", *Revista de llengua i dret*, 28: 145-186.

[Return to the top](#)