

**UPGRADE** is the European Journal for the Informatics Professional, published bimonthly at <<http://www.upgrade-cepis.org/>>

**Publisher**

UPGRADE is published on behalf of CEPIS (Council of European Professional Informatics Societies, <<http://www.cepis.org/>>) by NOVÁTICA <<http://www.ati.es/novatica/>>, journal of the Spanish CEPIS society ATI (Asociación de Técnicos de Informática <<http://www.ati.es/>>).

UPGRADE is also published in Spanish (full issue printed, some articles online) by NOVÁTICA, and in Italian (abstracts and some articles online) by the Italian CEPIS society ALSI <<http://www.alsi.it/>> and the Italian IT portal Tecnoteca <<http://www.tecnoteca.it/>>.

UPGRADE was created in October 2000 by CEPIS and was first published by NOVÁTICA and INFORMATIK/INFORMATIQUE, bimonthly journal of SVI/FSI (Swiss Federation of Professional Informatics Societies, <<http://www.svifsi.ch/>>).

**Editorial Team**

Chief Editor: Rafael Fernández Calvo, Spain <[rfoalvo@ati.es](mailto:rfoalvo@ati.es)>  
Assistant Editors:

- François Louis Nicolet, Switzerland, <[nicolet@acm.org](mailto:nicolet@acm.org)>
- Roberto Carniel, Italy, <[carniel@dgt.uniud.it](mailto:carniel@dgt.uniud.it)>

**Editorial Board**

Prof. Wolfgang Stucky, CEPIS President  
Fernando Piera Gómez and  
Rafael Fernández Calvo, ATI (Spain)  
François Louis Nicolet, SI (Switzerland)  
Roberto Carniel, ALSI – Tecnoteca (Italy)

**English Editors:** Mike Andersson, Richard Butchart, David Cash, Arthur Cook, Tracey Darch, Laura Davies, Nick Dunn, Rodney Fennemore, Hilary Green, Roger Harris, Michael Hird, Jim Holder, Alasdair MacLeod, Pat Moody, Adam David Moss, Phil Parkin, Brian Robson.

**Cover page and illustrations** designed by Antonio Crespo Foix, © ATI 2003

**Layout:** Pascale Schürmann

E-mail addresses for editorial correspondence: <[nicolet@acm.org](mailto:nicolet@acm.org)> and <[rfoalvo@ati.es](mailto:rfoalvo@ati.es)>

E-mail address for advertising correspondence: <[novatica@ati.es](mailto:novatica@ati.es)>

**Upgrade Newslist** available at <<http://www.upgrade-cepis.org/pages/editinfo.html#newslist>>

**Copyright**

© NOVÁTICA 2003. All rights reserved. Abstracting is permitted with credit to the source. For copying, reprint, or republication permission, write to the editors.

The opinions expressed by the authors are their exclusive responsibility.

ISSN 1684-5285

Coming issue:  
**“Software Engineering:  
The State of an Art”**

- 2 Editorial. Reassignment of Editorial Functions in Upgrade  
– Prof. Wolfgang Stucky (President of CEPIS)

## Open Knowledge

Guest Editors: Philippe Aigrain and Jesús M. González-Barahona

### Joint issue with NOVÁTICA

- 3 Presentation. Ownership and Terms of Use for Intangibles. Land Grab or Commons? – Philippe Aigrain and Jesús M. González-Barahona

*The guest editors present the issue, where they have offered the floor to a very diverse set of contributors, united by the effort to understand and promote information-based commons and convinced that a prosperous and more human economy can develop on its basis. They also provide a list of useful references for those interested in knowing more about this subject.*

- 6 The Political Economy of Commons – Yochai Benkler

*In this article the author defines the structure of the information commons, its sustainability, and its importance for democracy and for individual freedom.*

- 10 The Rediscovery of the Commons – David Bollier

*The author explains how a large part of the current information society is already a commons, which plays a vital role in the economical and cultural production.*

- 13 Language in the Digital Media: A Political Challenge – José-Antonio Millán

*The author analyses the situation of most languages in the digital domain, and how the promotion of a publicly available infrastructure of language-related software would serve to the societies using those languages.*

- 16 A Note on Software Patents – Pierre Haren

*A set of brief notes with his opinions on software patents is offered by the author.*

- 17 On the Patentability of Inventions Involving Computer Programmes – Alberto Bercovitz Rodríguez-Cano

*This article reproduces in substance the author’s address to the European Parliament during the hearing held on November 2002 on the Proposal for a European Directive on the Patentability of Computer Implemented Inventions.*

- 21 Legal Tools to Protect Software: Choosing the Right One – Roberto Di Cosmo

*This is an article in which the author analyses the different legal tools aimed at dealing with software protection.*

- 24 Petition to the European Parliament on the Proposal for a Directive on the Patentability of Computer-implemented Inventions – Several European computer scientists and engineers

*Petition written by several prestigious European computer scientists and engineers, related to the proposed Directive on software patents currently being discussed at the European Parliament.*

- 26 The Right to Read – Richard Stallman

*This is a short fiction story, 45 years ahead, in which the author, by extrapolating from some current trends, shows a future where access to information is tightly controlled, and the so called “trusted computing” is fully deployed.*

- 29 Please, Pirate My Songs! – Ignacio Escolar

*A musician describes the current situation of the music industry from his specific point of view.*

- 31 The EUCD and the DMCA in 2003: How Legal Protection for Technological Measures is shaping Consumers’ and Copyright Owners’ Digital Rights – Gwen Hinze

*This paper reviews the United States’ experience under the Digital Millennium Copyright Act and contends that Member States’ implementation legislation should include exceptions permitting circumvention for lawful uses and socially valuable activities. It also analyses a new technological protection regime contemplated by the 2003 draft of the European Union.*

- 35 ‘Trusted Computing’ and Competition Policy – Issues for Computing Professionals – Ross Anderson

*In this paper, the author gives an outline of Trusted Computing and sketches some of the possible effects on the computing business and the people who work in it.*

- 42 Software Patentability and CEPIS – Upgrade Editor’s Contribution

*The Editor of Upgrade introduces and publishes the positions on software patents put forward by two CEPIS member societies – GI, Germany, and ATI, Spain.*

# Language in the Digital Media: A Political Challenge

*José-Antonio Millán*

License: Licencia 20 min

“... and between them all they made it for me,  
with every utterance,  
dreaming, dream away,  
sing away.  
Now it is before me,  
in all its breadth,  
offered up before me,  
holding nothing back,  
wave upon wave breaking,  
on me –its beach-,  
a sea I carry with me everywhere,  
a Castilian sea.”

*(Pedro Salinas, “Verbo” (Word) in “Everything Clearer and other poems,” 1949)*

“The question is – said Humpty Dumpty – which is to be master, that’s all.”

*(Lewis Carroll, “Through the Looking Glass”)*

*The common good which is a language becomes a marketable commodity once we move into the digital media. However, there is no reason why speakers of languages other than English, or speakers of minority languages or of minority variants of widely used languages, should not be able to benefit from the many advantages the use of a language in the digital media can bring: spoken interfaces, translation aids, etc. This article proposes language policy measures which will enable existing research (much of which has been financed by public funds) to be developed into programmes which serve society’s needs without increasing its technological dependence.*

**Keywords:** GPL licences, language policy, minority languages, spoken interfaces, translation.

## 1 Introduction

In a society which is already highly digitally mediated, there is a growing tendency to incorporate natural language for automated communication between systems and humans, and between humans speaking different languages. This is happening for a very obvious reason: because language is a communication system which society disseminates to us all, and one which we all handle on a day to day basis. Language is not only the most common interface, but it is also the most refined: there is no option menu or clickable map capable of providing all the possibilities contained within a simple sentence. And as for spoken language, even people with literacy problems or those daunted by a mouse or a keyboard would be able to tell an automatic system (a well designed one) what they wanted.

But when we come to the digital world, language, a common good which is collectively created (“between them all they made it for me”), free (“offered before me”) and of unlimited

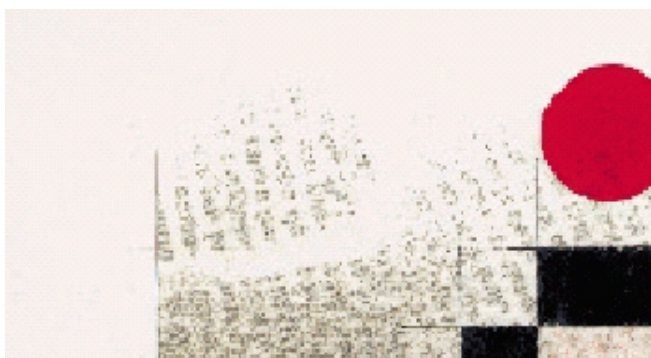
use (“holding nothing back”) becomes a marketable commodity. For machines to understand and talk to us they need to be equipped with programmes whose development is slow and expensive and require the existence of structured data sets (corpora, dictionaries). And even if such programmes were to

*José Antonio Millán*, linguist and pioneering digital editor in Spain, directed the CD-ROM edition of the Dictionary of the Spanish Royal Academy (1995), <<http://www.rae.es/>>, and created the Virtual Cervantes Centre (1997), <<http://cvc.cervantes.es/portada.htm>>. He has been a consultant and done development work for a great many institutions, and has worked on language policy in the framework of “*Los tres espacios lingüísticos*” (The three linguistic spaces), 2001-2002. He authored the books “*Internet y el español*” (Internet and the Spanish language) (2001) and “*De redes y saberes. Cultura y educación en las nuevas tecnologías*” (Of networks and knowledge. Culture and education in the new technologies) (1998). His personal website at <<http://jamillan.com>>, in Spanish, is worth a visit. <[jam@jamillan.com](mailto:jam@jamillan.com)>

exist and we were willing (and able) to pay for them, it is very likely that they would not take into account all the requirements of our societies. In this article I aim to explain the principles which should shape future public policies in this regard.

I'm going to begin with an impressionistic description of the future use of language software, which I put forward previously in another context [1]: "What kind of systems will language use as an interface? All kinds: data entry systems in general (from PDAs to professional systems), e-commerce systems (which search for products with certain characteristics and come up with descriptions and comparisons), leisure systems (location of shows, restaurants, tourist information...), education and training (automatic learning and evaluation systems), or research (search for material, 'intelligent' database access)."

We will be using these programmes more and more (sometimes without knowing we are doing so). They will have multilingual capabilities and will be able to give a hypothesis about how interesting a given piece of information may be for us, translate it (with varying degrees of reliability) and summarise it for us. They will be our intellectual and professional work tools.



## 2 The Linguistic Industry

These systems will give rise to an important economic sector. But in the case of Spanish, the result will be to increase the technological dependence of Spanish speaking countries and will tip the balance of payments against them even further [2]. The truth is that the speakers of many languages, even quite widely used ones, will have to pay to use them on networks, and the speakers of minority languages or variants will not even be able to do that: there will simply be no programmes for them. We may be able to pay for a computerized dictionary of synonyms in Spanish from Spain or in French from France (even as part of a word processor), but we won't be able to get one in Senegalese French or Bolivian Spanish however much we are prepared to pay...

Why should this be the case, given that for many languages there is the research and resources – almost invariably financed by public funds – which could provide the basis for the development of language software at all levels (both general and local)? Why is linguistic colonization going to almost completely close off this important and strategic industrial sector for languages like Spanish, French and Portuguese?

Why is it going to take such a long time before this sector can even be said to exist (if indeed it ever does) for the less widely spoken European languages?

For some languages the reason may be that – for historical reasons and due to universities' and other institutions' lack of resources and funding – there has never been enough basic research. But for Spanish, Portuguese, French or Italian the true reason is that their respective governments have never had any real digital language policy. This is an especially tricky issue, since it involves two problem areas which, generally speaking, (and here I speak for Spain) the powers that be lack both the knowledge and the political will to solve: language policy and digital policy. With regard to the former, they normally have no idea that such a policy even exists (except in the autonomous communities with their own language, in which it becomes a political instrument or bludgeon), and neither is the social importance of the digital issue fully understood (hence our current inadequate legislation here in Spain, the shameful way the .es domain is being managed, the perpetuation of communication monopolies, etc.)...

## 3 Which Goals for a Digital Language Policy?

What goals should be set for a digital language policy?:

- To guarantee that the resources (such as corpora and development programmes) and structured data sets that automatic systems (such as dictionaries) require are available for new projects.
- To increase the number of agents developing language software, so that the quality and quantity of options will also increase.
- To facilitate the incorporation of language software for minority languages or local variants of widely used languages.

These three points actually boil down to one thing: to open up the use of resources and data by means of a user licence which will ensure that derivative products are similarly *open* and *reusable*. These are fundamental issues: in the case of Spanish (and perhaps in others too) there are resources in public or historic institutions which encounter strange problems when they are required for use in projects. Some of these resources claim to have been "opened up" on the Internet, but this only means that they can be consulted: you can get the corpus to give you the number of hits for a word, or to analyse a morphological system. But this use is not enough for development projects. When we talk about an *open resource* we mean making the entire resource available on DVD or some other storage system to whoever asks for it. At the end of this article we will be taking a look at the possible objections to this way of working. With regard to the second point: reuse can be ensured by providing the resource under GPL (*General Public License*) [3] or *Creative Commons* [4] types of licences.

The current situation (at least as far as Spanish is concerned) is that the language resources of public research centres do not arrive transparently to all the enterprises that could use them, but only to a very small number which are those that are developing end user programmes. An effective policy would be one which aimed at putting language tool development resources currently in the hands of any kind of institutions (either public

or private) within reach of anyone wishing to develop language software. The dominant vision is that this is only a task for major companies (and US ones at that), but the truth of the matter is that both in terms of data and development programmes, or small end user programmes, there are all kinds of development projects that could be undertaken, many of them tailor made. For example, specialty dictionaries, oral and written, to complement the lexis which text/speech converters and conversation systems [5].

#### 4 Which Ways for a Digital Language Policy?

There are different ways this can be achieved. One way could be to create a Language Resource Bank, which would be free and open to any entity or individual who wanted to develop software (under the kind of licences mentioned before) to ensure that the results of the use of these resources were equally open and reusable.

In order to set up such a resource bank, the most realistic approach would be to buy the licence to use these language resources and data from those institutions (universities or enterprises) which hold them. It may seem paradoxical that the fruits of research paid for with public money should have to be bought again in order for them to benefit the common good, and it is: but it would seem to be the most practical solution compared to other possible ways. As a parallel action, given that the resource bank would be publicly owned and is intended to benefit the common good, a campaign could be initiated to encourage institutions to hand over their resources to the bank free of charge (rather than selling them to it).

What scope would these resource banks have? They should be set up by languages rather than by countries. Many European languages are spread over a large number of countries, in different continents (as is the case of French, Portuguese or Spanish), and it would be absurd to be restricted just to the European variant of the language. A Spanish Language Bank, for example, should ideally bring together resources from as many variants as possible.

In an interconnected and multilingual society such as ours, we might also try to widen the scope of benefits of an action such as this, by making it extendible to different languages. To do this it might be possible to promote resources which take advantage of the similarities that exist between many languages (for example, between the Latin-based languages: Spanish, French, Catalan, Italian, ...) to create kernels of morphology, syntax, lexicography, etc. common to all of them.

The functional model of these resource banks could be based on those of the Linguistic Data Consortium [6] or the European Language Resources Association [7].

Earlier I mentioned possible objections that such an action might arouse: *“Copying a corpus or a morphological dictionary does not in any way diminish it. If all agents wishing to work on linguistic projects can obtain the results of this indispensable basic research freely, the most that will happen is that in a short time we will have a proliferation of programmes for recognising words, analysing sentences, etc. Many of them will*

*not be directly usable by end users, but they may form part of more elaborate automatic systems, and the end result will be more systems, more kinds of systems and cheaper systems using our language”* [1].

Such a simple and cheap proposal, so clearly beneficial to society; a solution that promotes the capabilities of companies and user groups, and is detrimental only to the interests of corporate giants; a proposal which would give us control over a strategic sector which necessarily involves all our institutions and citizens, should be readily embraced by the relevant government bodies.

Rather than uttering empty rhetoric in praise of a language which was already created and widely used when we first encountered it, we should be taking real steps in its defence, instead of perpetuating its current situation of technological colonisation, a disservice to the society that uses it.

*Translation by Steve Turpin*

#### Acknowledgements

This article has benefited from discussions which took place among Portuguese, French and Spanish language experts during the *Tres Espacios Lingüísticos* (Three Linguistic Spaces) conferences (2001/2002), convened by the *Organisation Internationale de la Francophonie*, *Organización de Estados Iberoamericanos*, *Comunidades dos Países de Língua Portuguesa*, *Unión Latina y Secretaría de Cooperación Iberoamericana* <<http://www.jamillan.com/tresespa.htm>>. I would like to thank Daniel Pimienta and Isabel Trancoso for their contributions and Daniel Prado for his constant support.

#### References

- [1] José Antonio Millán, “El español en la sociedad digital: una propuesta” (The Spanish Language in the Digital Society: a Proposal), a contribution at the II Congreso Internacional de la Lengua Española, Valladolid, 16 to 19 of October 2001. <[http://cvc.cervantes.es/obref/congresos/valladolid/mesas\\_redondas/millan\\_j.htm](http://cvc.cervantes.es/obref/congresos/valladolid/mesas_redondas/millan_j.htm)>.
- [2] José Antonio Millán, “La lengua que era un tesoro” (The language that was a treasure), March 28 2001 <<http://www.jamillan.com/tesoro.htm>> and its abridged version in English “How much is a language worth? A Quantification of the Digital Industry for the Spanish Language”. <<http://www.jamillan.com/worth.htm>>.
- [3] <<http://www.gnu.org/copyleft/gpl.html>>.
- [4] <<http://creativecommons.org/>>.
- [5] This kind of actions, in which, for example, the jargon used in neurobiology in Portuguese or the lexis of Mexican engineering is incorporated into pre-existing systems require, firstly, for these systems to be open and allow additions to be made, and secondly, a collaborative work to create a corpus. This can be as easy as phone calls from volunteers or as sophisticated as an automatic system for storing samples.
- [6] <<http://www ldc.upenn.edu/>>.
- [7] <<http://www.icp.grenet.fr/ELRA/home.html>>.