

Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de ATI (Asociación de Técnicos de Informática). *Novática* edita también *Upgrade*, revista digital de CEPIS (Council of European Professional Informatics Societies), en lengua inglesa.

<<http://www.ati.es/novatica/>>
<<http://www.upgrade-cepis.org/>>

ATI es miembro de CEPIS (Council of European Professional Informatics Societies) y tiene un acuerdo de colaboración con ACM (Association for Computing Machinery). Tiene asimismo acuerdos de vinculación o colaboración con AdaSpain, AI2 y ASTIC

CONSEJO EDITORIAL

Antoni Carbonell Nogueras, Francisco López Crespo, Julián Marcelo Cocho, Celestino Martín Alonso, Josep Molins i Bertrán, Roberto Moya Quiles, César Pérez Chirinos, Mario Piattini Velthuis, Fernando Piera Gómez (Presidente del Consejo), Miquel Sarries Griño, Carmen Ugarte García, Asunción Yturbe Herranz

Coordinación Editorial
Rafael Fernández Calvo <rfoalvo@ati.es>

Composición y autoedición
Jorge Llácer

Traducciones
Grupo de Lengua e Informática de ATI
Coordinadas por José A. Accino (Univ. de Málaga) <jalfonso@ieev.uma.es>

Administración
Tomás Brunete, María José Fernández, Enric Camarero, Felicidad López

SECCIONES TÉCNICAS: COORDINADORES

Administración Pública Electrónica
Gumersindo García Aribas, Francisco López Crespo (MAP)
<gumersindo.garcia@map.es>, <flc@ati.es>

Arquitecturas
Jordi Tubella (DAC-UPC) <jortit@ac.upc.es>
Victor Viñals Yufera (Univ. de Zaragoza) <victor@unizar.es>

Auditoría SITIC
Marina Touriño, Manuel Palao (ASIA)
<marinatourino@marinatourino.com>, <manuel@palao.com>

Bases de Datos
Coral Calero Muñoz, Mario G. Piattini Velthuis
(Escuela Superior de Informática, UCLM)
<Coral.Calero@uclm.es>, <mpiattin@inf-cr.uclm.es>

Derecho y Tecnologías
Isabel Hernando Collazos (Fac. Derecho de Donostia, UPV)
<ihernando@legalek.net>
Isabel Davara Fernández de Marcos (Davara & Davara)
<idavara@davara.com>

Enseñanza Universitaria de la Informática
Joaquín Ezpeleta Mateo (CPS-LIZARR) <ezpeleta@posta.unizar.es>
Cristóbal Pareja Flores (DSIP-UCM) <cpareja@sisip.ucm.es>

Informática y Filosofía
Josep Corco (UIC) <jcorco@unica.edu>
Esperanza Marcos (ESSET-URJC) <euca@esset.urjc.es>

Informática Gráfica
Roberto Vivo (Eurographics, sección española) <rvivo@dsic.upv.es>

Ingeniería del Software
Javier Dolado Cosin (DLSI-UPV) <dolado@si.ehu.es>
Luis Fernández (PRIS-EL-UEM) <lufern@pris.esi.ueu.es>

Inteligencia Artificial
Federico Barber, Vicente Botti (DSIC-UPV)
<fvbotti@barber@dsic.upv.es>

Interacción Persona-Computador
Julio Abascal González (PI-UPV) <julio@si.ehu.es>
Jesús Lorés Vidal (Univ. de Lleida) <jesus@eup.udl.es>

Internet
Alonso Álvarez García (TID) <alonso@ati.es>
Lluc M. Pagés Casas (Indra) <lpages@ati.es>

Lengua e Informática
M. del Carmen Ugarte (IBM) <cugarte@ati.es>

Lenguajes Informáticos
Andrés María López (Univ. Carlos III) <amarin@it.uc3m.es>
J. Angel Velázquez (ESSET-URJC) <a.velazquez@esset.urjc.es>

Libertades e Informática
Alfonso Escolano (FIR-Univ. de La Laguna) <aescolan@ull.es>

Lingüística computacional
Xavier Gómez Guinovart (Univ. de Vigo) <xgg@uvigo.es>
Manuel Palomar (Univ. de Alicante) <mpalomar@dlsi.ua.es>

Mundo estudiantil
Adolfo Vázquez Rodríguez
(Rama de Estudios del IEEE-UCM) <a.vazquez@iee.org>

Profesión informática
Rafael Fernández Calvo (ATI) <rfoalvo@ati.es>
Miquel Sarries Griño (Ayto. de Barcelona) <msarries@ati.es>

Redes y servicios telemáticos
Luis Guíjarro Coloma (DCOM-UPV) <lguijar@dcom.upv.es>
Josep Solé Pareta (DAC-UPC) <pareta@ac.upc.es>

Seguridad
Javier Areitio (Redes y Sistemas, Bilbao) <jareitio@orion.deusto.es>
Composicion, Edición y Redacción ATI Valencia
Reino de Valencia 23, 46005 Valencia
Tlf./fax 963330392 <secreval@ati.es>

Software Libre
Jesús M. González Barahona, Pedro de las Heras Quirós
(GSYC-URJC) <jgb.pheras@gsyc.esset.urjc.es>

Tecnología de Objetos
Jesus Garcia Molina (DIS-UM) <jmolina@correo.um.es>
Gustavo Rossi
(LIFIA-UNLP, Argentina) <gustavo@sol.info.unlp.edu.ar>

Tecnologías para la Educación
Josep Sales Ruffi (ESPIRAL) <jsales@pie.xtec.es>

Tecnologías y Empresa
Pablo Hernández Medrano (Bluemat) <pablohm@bluemat.biz>

TIC y Turismo
Andrés Aguayo Maldonado, Antonio Guevara Plaza (Univ. de Málaga)
<laguayo.guevara@lcc.uma.es>

TIC para la Sanidad
Valentín Masero Vargas (DI-UNEX) <vmasero@unex.es>

Las opiniones expresadas por los autores son responsabilidad exclusiva de los mismos. *Novática* permite la reproducción de todos los artículos, salvo los marcados con © o *copyright*, debiéndose en todo caso citar su procedencia y enviar a *Novática* un ejemplar de la publicación.

Coordinación Editorial y Redacción Central (ATI Madrid)
Padilla 66, 3º, dcha., 28006 Madrid
Tlf./fax 914029391; fax.913093685 <novatica@ati.es>

Composicion, Edición y Redacción ATI Valencia
Reino de Valencia 23, 46005 Valencia
Tlf./fax 963330392 <secreval@ati.es>

Administración y Redacción ATI Cataluña
Via Laietana 41, 1º, 1º, 08003 Barcelona
Tlf./fax 934125235; fax.934127713 <secregen@ati.es>

Redacción ATI Andalucía
Isaac Newton, s/n, Ed. Sadiel, Isla Cartuja 41092 Sevilla
Tlf./fax 954460779 <secreand@ati.es>

Redacción ATI Aragón
Lagasca 9, 3-B, 50006 Zaragoza
Tlf./fax 976235181 <secreara@ati.es>

Redacción ATI Asturias-Cantabria <gp-astucant@ati.es>
Redacción ATI Castilla-La Mancha <gp-clmancha@ati.es>

Redacción ATI Galicia
Recinto Ferial s/n, 36540 Silleda (Pontevedra)
Tlf.986581413; fax 986580162 <secregal@ati.es>

Suscripción y Ventas: <<http://www.ati.es/novatica/interes.html>>, o en ATI Cataluña y ATI Madrid
Publicidad: Padilla 66, 3º, dcha., 28006 Madrid
Tlf.914029391; fax.913093685 <novatica.publicidad@ati.es>

Imprenta: 9-Impressió S.A., Juan de Austria 66, 08005 Barcelona.
Depósito Legal: B 15.154-1975
ISSN: 0211-2124; CODEN NOVAEC

Portada: Antonio Crespo Foix / © ATI 2003

SUMARIO

En resumen: El procomún del conocimiento <i>Rafael Fernández Calvo</i>	2
Monografía: Conocimiento abierto / Open Knowledge (En colaboración con Upgrade) Editores invitados: <i>Philippe Aigrain</i> y <i>Jesús M. González Barahona</i> Presentación. Propiedad y uso de la información y del conocimiento: ¿privatización o procomún? <i>Philippe Aigrain, Jesús M. González-Barahona</i> La Economía Política del procomún <i>Yochai Benkler</i> El redescubrimiento del procomún <i>David Bollier</i> La lengua en el medio digital: un reto político <i>José Antonio Millán</i> Nota sobre las patentes de software <i>Pierre Haren</i> Sobre la patentabilidad de las invenciones referentes a programas de ordenador <i>Alberto Bercovitz Rodríguez Cano</i> Eligiendo la herramienta legal correcta para proteger el software <i>Roberto Di Cosmo</i> Por favor, ¡pirateen mis canciones! <i>Ignacio Escobar</i> La normativa europea y norteamericana sobre propiedad intelectual en el 2003: protección legal antipiratero y derechos digitales <i>Gwen Hinz</i> 'Informática de confianza' y política sobre competencia: temas a debate para profesionales informáticos <i>Ross Anderson</i>	2 6 10 13 16 17 21 24 30
Secciones Técnicas	
Lengua e Informática	
El software libre y las lenguas minoritarias: una oportunidad impagable <i>Jordi Mas i Hernández</i>	36
Lenguajes informáticos	
Evaluación parcial de programas y sus aplicaciones <i>Pascual Julián Iranzo</i>	40
COMPAS: un compilador para un lenguaje imperativo con aserciones embebidas <i>Joaquín Ezpeleta Mateo, Pedro Gascón Campos, Natividad Porta Royo</i>	47
Seguridad	
Ocultación de imágenes mediante Esteganografía <i>David Atauri Mezquida, Luis Fernández Sanz, Matías Alcojor, Ignacio Acero</i> La confianza y la seguridad aspectos vitales para los servicios electrónicos <i>José A. Mañas Argemí</i>	52 58
Sistemas de Tiempo Real	
Sistemas Linux de tiempo real <i>Javier Miqueleiz Álamos</i>	63
Referencias autorizadas	69
Sociedad de la Información	
Personal y transferible	
Locos por los ordenadores (II): Ada Byron y Charles Babbage, o la bella y la bestia <i>Rafael Fernández Calvo</i>	75
Asuntos Interiores	
Coordinación editorial / Programación de <i>Novática</i>	76
Normas de publicación para autores / Socios Institucionales	79
Monografía del próximo número: «Ingeniería del Software: estado de un arte»	

Conocimiento abierto / *Open Knowledge*

José Antonio Millán
Lingüista y editor digital

<jam@jamillan.com>

La lengua en el medio digital: un reto político

Este artículo se publica bajo la Licencia 20 minutos

... y entre todos me la hicieron,
habla por habla,
soñando, sueña que sueña,
canta que canta.
Delante la tengo ahora,
toda tan ancha,
delante de mí ofrecida,
sin guardar nada,
onda tras onda rompiendo,
en mí -su playa-,
mar que llevo a todas partes,
mar castellana.

(Pedro Salinas, «Verbo» en «Todo más claro y otros poemas», 1949)

«La cuestión --dijo Humpty Dumpty-- es saber quién es el que manda..., eso es todo».

(Lewis Carroll, «A través del espejo»)

Resumen: el bien colectivo que constituye una lengua se convierte en una mercancía en cuanto pasamos al medio digital. Sin embargo, no hay ninguna razón por la que los hablantes de lenguas que no son el inglés, o los hablantes de lenguas minoritarias, o de variantes minoritarias de lenguas extendidas, no puedan beneficiarse de las muchas promesas de la lengua en el medio digital: interfaces orales, ayudas a la traducción, etc. Este artículo propone medios de política lingüística para que las investigaciones existentes (muchas de ellas financiadas con dinero público) puedan desembocar en programas que sirvan a las necesidades de la sociedad, y no aumenten su dependencia tecnológica.

Palabras clave: interfaces orales, lenguas minoritarias, licencia GPL, política lingüística, traducción.

En una sociedad ya muy mediada digitalmente, hay una tendencia creciente a incorporar el lenguaje natural para la comunicación automatizada entre sistemas y personas, y entre personas de distintas lenguas. Y ello por una razón muy clara: porque la lengua es un sistema de comunicación que la sociedad distribuye a todos, y que todos manejamos habitualmente. La lengua no sólo es la interfaz más común, sino que además es la más fina: no hay menú de elecciones o mapa clicable que pueda aportar todas las posibilidades que da una simple frase. Y si pensamos en la lengua hablada, incluso analfabetos o personas que dudarían ante un ratón o un teclado son capaces de contarle a un sistema automático (bien diseñado) lo que quieren.

Pero al llegar al terreno digital la lengua, un bien común de creación colectiva («entre todos me la hicieron»), gratuito («delante de mí ofrecida») y de uso ilimitado («sin guardar nada») se convierte en una mercancía. Para que las máquinas nos entiendan y nos hablen tienen que servirse de programas que tienen un desarrollo lento y caro, y que exigen conjuntos de datos estructurados (*corpus*, diccionarios) para su desarrollo. Incluso aunque existan estos programas y queramos (y podamos) pagarlos, es muy posible que no tengan en cuenta el conjunto de necesidades de nuestras sociedades. En este artículo voy a exponer los principios que deberían guiar futuras políticas públicas al respecto [1].

Voy a comenzar por una descripción impresionista de un horizonte de uso del software lingüístico, que formulé en otro contexto [2]: «¿Qué tipo de sistemas van a utilizar la lengua como interfaz? De todos los tipos: los de entrada de datos en general (de la agenda personal a sistemas profesionales), los de comercio electrónico (sistemas que busquen productos con determinadas características y vuelvan con descripciones y comparaciones), de ocio (localización de espectáculos, restaurantes, información turística...), educación y formación (sistemas de tutoría y evaluación automáticas), o investigación (localización de materiales, acceso 'inteligente' a bases de datos)».

Usaremos (a veces sin saberlo) estos programas cada vez más. Tendrán capacidades multilingües y serán capaces de emitir hipótesis sobre el interés que tendrá para nosotros un cierto material, traducirlo (con distintos grados de fiabilidad) y resumírnoslo. Serán nuestras herramientas de trabajo intelectual y profesional.

El entorno de estos sistemas configurará un sector económico de importancia. Pero en el caso del español el resultado será que aumentará la dependencia tecnológica de los países hispanohablantes, y se inclinará aún más su balanza de pagos [3]. Sí: los hablantes de muchas lenguas, incluso extendidas,

Autor

José Antonio Millán, lingüista y editor digital, dirigió la edición en CD-ROM del Diccionario de la Real Academia (1995) y creó el Centro Virtual Cervantes en Internet (1997). Ha realizado consultorías y desarrollos para numerosas instituciones, y ha trabajado en política lingüística en el marco de «Los tres espacios lingüísticos», 2001-2002. Es autor de los libros «Internet y el español» (2001) y «De redes y saberes. Cultura y educación en las nuevas tecnologías» (1998).

tendremos que pagar por usarlas en las redes, y los hablantes de lenguas o variantes minoritarias ni siquiera tendrán esa posibilidad: no existirán programas para ellos. Pagaremos por un diccionario informático de sinónimos del español de España o del francés de Francia (aunque sea como parte de un procesador de textos), pero uno del francés de Senegal o del español de Bolivia no lo conseguiremos ni pagando...

¿Por qué habría de ser así, existiendo, como existen para muchas lenguas, investigaciones y recursos --casi siempre financiados con dinero público-- que podrían ser la base del desarrollo de softwares lingüísticos a muy distintos niveles (generales y locales)? ¿Por qué este sector industrial, importante y estratégico, va a estar casi completamente colonizado para lenguas como el español, el francés o el portugués? ¿Por qué va a tardar en existir (si es que llega a hacerlo) para lenguas europeas con menos hablantes?

Para algunas lenguas, la razón puede ser que --por motivos de historia, y de recursos y dotación de sus universidades y otras instituciones-- no ha existido suficiente investigación de base. Pero para el español, para el portugués, para el francés o para el italiano, la verdadera razón es que sus respectivos gobiernos carecen, realmente, de una política lingüística digital. Este es un terreno especialmente resbaladizo, porque en él se juntan dos áreas en las que por lo general (y hablo del caso de España) los gobernantes carecen de conocimientos... y de ganas de solicitarlos: la política lingüística, y la política digital. Respecto a la primera, normalmente no saben ni que existe (salvo en las comunidades autónomas con lengua propia, donde se convierte en un instrumento político), y el alcance social de la cuestión digital tampoco acaba de ser comprendida (ahí está para España una legislación inadecuada, la vergonzosa gestión del dominio .es, la perpetuación de los monopolios de comunicaciones, etc.)...

¿Qué metas debería plantearse una política lingüística digital?:

- Garantizar que los recursos (como *corpus* y programas de desarrollo) y los conjuntos de datos estructurados que alimentan a los sistemas automáticos (como diccionarios) estén disponibles para nuevos desarrollos.
- Aumentar el número de agentes que desarrollan software lingüístico, para que crezca la calidad y la cantidad de opciones.
- Facilitar la incorporación al software lingüístico de las lenguas minoritarias o las variantes locales de lenguas extendidas.

En realidad estos tres puntos se resuelven en uno sólo: abrir al uso los recursos y los datos, mediante una licencia de utilización que garantice que los productos derivados sean igualmente *abiertos* y *reutilizables*. Estos son puntos básicos: en el caso del español (y quizás en otros), existen recursos en instituciones públicas o históricas que luego encuentran extraños problemas cuando se quieren usar para desarrollos. Algunos de recursos presumen de haberse «abierto» a Internet, pero eso sólo significa que les puede consultar: uno puede dar una palabra para que un *corpus* le devuelva sus ocurrencias, o para que se la analice un sistema morfológico. Pero este uso no es apto para desarrollos. *Abrir*, para que quede claro, es entregar un DVD o cualquier otro sistema de almacenamiento, con la totalidad del recurso, a quien lo pida. Al final de este artículo repasaré posibles objeciones a esta

forma de obrar. En cuanto al segundo punto: la *reutilización* se puede asegurar dando el recurso vía una licencia tipo GPL (*General Public License*) [4] o *Creative Commons* [5].

La situación actual (por lo menos en el caso del español) es que los recursos lingüísticos de los centros de investigación públicos no llegan de forma transparente a todas las empresas que podrían utilizarlos, sino a un reducidísimo número, que son los que desarrollan los programas de usuario final. Una política eficaz aspiraría a que los recursos de desarrollo de herramientas lingüísticas en poder de cualquier tipo de instituciones (tanto públicas como privadas) alcanzara a cualquiera que quisiera desarrollar software lingüístico. La visión dominante es que ésta es sólo una tarea de grandes empresas (y por añadidura, norteamericanas), pero la verdad es que tanto a nivel de datos, como de programas de desarrollo, o de pequeños softwares de usuario final, uno puede concebir todo tipo de desarrollos, muchos de ellos a la medida: por ejemplo, diccionarios de especialidad, orales y escritos, que complementarían los léxicos con los que ya contarán conversores habla/texto y sistemas de conversación [6].

Existen distintas vías que podrían lograr este fin. Una de ellas podría ser la creación de un Fondo de Recursos Lingüísticos, gratuito, abierto para cualquier entidad o individuo que quiera hacer desarrollos (bajo licencias como las mencionadas), para garantizar que los resultados del uso de estos recursos fueran igualmente abiertos y reutilizables.

Para constituir este Fondo, lo más realista sería comprar la licencia de uso de los recursos y datos lingüísticos a aquellas instituciones (universidades, o empresas) que las poseen. Puede parecer paradójico que frutos de la investigación pagada con dinero público deban ser comprados de nuevo para que alcancen al bien público, y lo es: pero parece la solución más práctica, comparada con otros caminos posibles... En paralelo, y dado que la titularidad del Fondo sería pública y su destino el bien común, se podría iniciar una campaña para que las instituciones cedieran gratuitamente sus recursos al Fondo (en vez de vendérselos).

¿Qué ámbito deberían tener este tipo de Fondos? Deberían constituirse por lenguas, más que por Estados. Muchas lenguas europeas están distribuidas por un gran número de países, en diferentes continentes (caso del francés, el portugués o el español), y sería absurdo limitarse a la variante europea: un Fondo de la Lengua Española, por ejemplo, debería reunir recursos de cuantas más variantes del español, mejor.

En una sociedad interconectada y multilingüe como la nuestra, podríamos intentar además ampliar el ámbito de los beneficios de una acción como esta, haciéndola extensible a diferentes lenguas. Para ello, cabría incentivar también recursos que saquen partido de la proximidad que exista entre ellas (por ejemplo, las lenguas descendientes del latín: español, francés, catalán, italiano, ...) para generar núcleos de morfología, sintaxis, lexicografía, etc. comunes a todas.

Los modelos de funcionamiento de estos Fondos podrían basarse en los que están vigentes en el *Linguistic Data Consortium* [7] o en la *European Language Resources Association* [8].

En otro lugar me he adelantado a las posibles objeciones que una acción así podría despertar: *«El copiar un corpus o un diccionario morfológico no los merma en nada, en ningún aspecto. Si todos los agentes que quieran hacer desarrollos lingüísticos consiguen libremente los resultados de esta imprescindible investigación de base lo más que puede pasar es que tengamos al poco tiempo una proliferación de programas que reconocen palabras, analizan frases, etc. Muchos de ellos no serán directamente utilizables por los usuarios finales, pero podrán formar parte de sistemas automáticos más elaborados, y el resultado final es que habrá más sistemas que usen nuestra lengua, de más tipos, y más baratos»*[9].

Una propuesta tan simple, barata y con unos beneficios tan claros para la sociedad; una solución que fomenta las capacidades de empresas y grupos de usuarios propios en detrimento sólo de los oligopolios; una propuesta que permitiría controlar un sector estratégico y de tránsito obligado para nuestras instituciones y ciudadanos, debería ser fácilmente asumida por las instancias gubernamentales que están para estas cosas.

La retórica vacua de alabanza de una lengua que nos hemos encontrado ya hecha y extendida podría dar paso a acciones que la defiendan realmente, en vez de perpetuar su situación de colonización tecnológica, con mal servicio a la sociedad que la usa.

Referencias

[1] Este artículo se beneficia de las discusiones que tuvieron lugar en los encuentros entre expertos lusófonos, francófonos e hispanohablantes en el marco de los Tres espacios lingüísticos (2001/2002), convocados por la 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/trespa.htm>). Quiero agradecer sus aportaciones a Daniel Pimienta e Isabel Trancoso y su constante apoyo a Daniel Prado.

[2] **José Antonio Millán**, «El español en la sociedad digital: una propuesta», intervención en el II Congreso Internacional de la Lengua Española, Valladolid, 16 a 19 de octubre del 2001 <http://cvc.cervantes.es/obref/congresos/valladolid/mesas_redondas/millan_j.htm>

[3] **José Antonio Millán**, «La lengua que era un tesoro», 28 de marzo del 2001 <<http://www.jamillan.com/tesoro.htm>> y su versión abreviada en inglés «How much is a language worth. A Quantification of the Digital Industry for the Spanish Language» <<http://www.jamillan.com/worth.htm>>.

[4] <<http://www.gnu.org/copyleft/gpl.html>>.

[5] <<http://creativecommons.org/>>.

[6] Este tipo de acciones, mediante las cuales, por ejemplo, la jerga de la neurobiología en portugués o el léxico de la ingeniería mexicana se incorporara a sistemas preexistentes exige, por una parte que estos sean abiertos y admitan ampliaciones, y en segundo lugar un trabajo colaborativo de creación de un corpus. Éste puede ser tan simple como llamadas telefónicas de voluntarios a un sistema automático que va almacenando muestras.

[7] <<http://www ldc.upenn.edu/>>.

[8] <<http://www.icp.grenet.fr/ELRA/home.html>>.

[9] «El español en la sociedad digital: una propuesta», cit.