

EL DESARROLLO TECNOLINGÜÍSTICO DEL EUSKERA

Araceli Díaz de Lezana
Viceconsejería de Política Lingüística
Gobierno Vasco

Barcelona, 21.05.04

1. Introducción.

En esta era de la Información y de la Comunicación han surgido y están surgiendo nuevas necesidades (en gran medida provocadas por Internet) que han impulsado a las universidades, centros de investigación y empresas a desarrollar nuevas tecnologías que integran un tratamiento especial de las lenguas. Es lo que se viene denominando "infoingeniería lingüística" o "tecnologías lingüísticas". Han aparecido nuevos negocios y nuevos mercados y, gracias a ese impulso dicho ámbito tecnológico avanza con gran rapidez. Estas tecnologías que también se conocen como "industrias de la lengua" son fundamentalmente éstas:

- El reconocimiento vocal (**S**peech-**T**o-**T**ext y **A**utomatic **S**peech **R**ecognition)
- La síntesis vocal (**T**ext-**T**o-**S**peech)
- El reconocimiento óptico de caracteres (**O**ptical **C**haracter **R**ecognition)
- La traducción automática
- La compresión de datos vocales
- El "data / audio mining" (búsquedas en grandes volúmenes de documentación)

Los gobiernos y las administraciones, conscientes del proceso de las transformaciones sociales, económicas y políticas que se están produciendo como consecuencia del desarrollo de las Tecnologías de la Información y las Comunicaciones, y la capacidad, prácticamente ilimitada, que su uso aporta para obtener, almacenar, procesar y compartir información, están elaborando políticas que permitan a la ciudadanía integrarse plenamente en esta Sociedad. Así, por ejemplo, la Unión Europea puso en marcha en 1999 la iniciativa "**eEuropa-Una Sociedad de la Información para todos**" que ha tenido continuación en diferentes planes de acción.

2. El euskera y las nuevas tecnologías de la información y de la comunicación (TICs).

El Gobierno Vasco ha venido promoviendo desde 1999 una serie de actuaciones en el marco de la "Iniciativa Euskadi 2000Tres" encaminadas a potenciar la incorporación plena de la sociedad vasca a la Sociedad de la Información poniendo el nuevo espacio digital al servicio de las personas.

Varios planes estratégicos en vigor sustentan la política del Gobierno Vasco en lo que concierne a las nuevas Tecnologías de la Información y de la Comunicación y el euskera:

√ 1999, El Plan General de Promoción del Uso del Euskera.

Este plan, elaborado por el Consejo Asesor del Euskera, aprobado por el Gobierno Vasco y ratificado por el Parlamento Vasco, establece los retos

principales del euskera para los próximos 10 años, diferenciados en tres niveles:

1. Objetivo principal: "decidir y promover las medidas de política lingüística necesarias para garantizar la posibilidad de vivir en euskera a quien así lo desee".
2. Tres objetivos estratégicos: garantizar la transmisión intergeneracional del euskera, incrementar el uso social del euskera mediante la oferta de productos y servicios en euskera; y mejorar la calidad de la lengua.
3. Y casi cien medidas u objetivos específicos en 17 sectores estratégicos de actuación, entre los que se incluye el apartado VI.3.2.c, relativo a las "*Nuevas tecnologías de la información y de la comunicación*".

En él se fijan las siguientes medidas:

Potenciar la industria de la ingeniería del euskera; priorizar la localización en euskera de productos de gran difusión en el mercado; impulsar la creación de una comunidad de usuarios avanzados; crear plataformas o redes de edición y traducción, así como diccionarios electrónicos que puedan estar al alcance de todos; desarrollar sistemas multimedia para el aprendizaje del euskera; y crear una infraestructura de herramientas básicas para el análisis del lenguaje.

✓ **2001-2004, El Plan de Ciencia, Tecnología e Innovación: el programa de Industrias de la Lengua e Ingeniería Lingüística.**

Este programa tiene los siguientes objetivos generales:

1. Fortalecer la capacidad científica y de innovación tecnológica de las empresas y centros de investigación asociados a la Industria de la Lengua.
2. Impulsar la penetración de las Industrias de la Lengua en el conjunto del tejido empresarial y social.
3. Aprovechar el conocimiento tecnológico emanado de los requerimientos planteados por la diversidad lingüística de la Comunidad Autónoma Vasca.

Los objetivos del Plan se concretan en seis líneas de acción:

- Desarrollo de técnicas de base de la Ingeniería del Lenguaje.
- Desarrollo de recursos lingüísticos.
- Fomento del uso de herramientas avanzadas y experimentación de plataformas capacitadoras.
- Almacenamiento, recuperación y búsqueda de información.
- Reconocimiento automático del habla, compresión digital de habla y conversión de texto a habla.
- Funcionalidades de soporte al tratamiento de texto en el contexto ofimático y conversión de texto a texto (traducción).

√ 2002-2005, Plan Euskadi en la Sociedad de la Información

El objetivo de este Plan es potenciar las tecnologías de la información en todos los niveles sociales: enseñanza, ciudadanía, empresa, cultura y Administración.

El Plan establece 10 líneas de actuación:

- respecto a los usuarios: Internet para todos, Empresa Digital y Administración on line,

- en relación a los Contenidos y Servicios: Euskadi en la Red, e-Formación, e-Salud y Sector de Contenidos, y
- en cuanto al Contexto e Infraestructuras: Tecnología y Normativa.

De este plan han surgido numerosas iniciativas, entre ellas, "Konekta zaitez", campaña que ha logrado que más del 34% de los hogares vascos cuente con un ordenador conectado a Internet, y "Empresa Digital", iniciativa dirigida a asesorar a las empresas en su incorporación a la nueva economía.

Otra iniciativa ha sido el "proyecto KZgunea", puesto en marcha en 2001, que consiste en la creación de una red de centros públicos gratuitos para la formación y el uso de las nuevas tecnologías de la información y de la comunicación, y pretende atraer a las nuevas tecnologías a sectores como amas de casa, jubilados, desempleados e inmigrantes. Gracias a acuerdos entre ayuntamientos, EUDEL (Asociación de Municipios Vascos) y Gobierno Vasco ya hay 225 centros creados en la CAV.

Además de estas iniciativas, en el Plan Euskadi en la Sociedad de la Información se recogen en el apartado "Euskadi en la Red" las iniciativas relacionadas con el euskera: "el desarrollo de recursos lingüísticos en la red; el desarrollo y promoción de contenidos en euskera en la red; la promoción de proyectos innovadores en materia de nuevas tecnologías; y las tecnologías de voz, traducción automática y tratamiento de textos aplicadas al euskera".

√ 2003-2005, Plan de Informática y Telecomunicaciones

Este plan integra los proyectos de Sistemas de Información surgidos de las iniciativas anteriores más otros proyectos, enmarcándose todos en las siguientes líneas estratégicas:

- Mejora de la gestión interna
- Administración electrónica
- Infraestructuras tecnológicas y Organización

3. El desarrollo tecnolingüístico del euskera.

El euskera ha avanzado en los últimos años en su integración en las nuevas tecnologías:

1. Publicación on-line de la prensa escrita en euskera (un diario escrito íntegramente en euskera, además de revistas de diferente índole).
2. Publicación de CDs para el aprendizaje del euskera y CDs con contenidos en euskera.
3. Páginas web multidisciplinares, buscadores en euskera (www.kaixo.com, www.aurki.com, www.jalgi.com) y boletines informativos electrónicos temáticos y foros de discusión (www.sustatu.com y www.erabili.com)
4. Integración en la red de corpus, diccionarios y demás herramientas útiles para los tecnohablantes de la lengua.
 - El corrector ortográfico del euskera XUXEN, realizado por el grupo IXA de la Facultad de Informática de la Universidad del País Vasco; se puede descargar gratuitamente desde [www.euskadi.net/euskara soft](http://www.euskadi.net/euskara_soft): versiones Office 97,

2000 y XP, OpenOffice.org para Windows y para Linux; y versión Word 98 y 2001 de Macintosh.

- EUSKALTERM, el Banco Terminológico Público Vasco (creado sobre la base del banco creado por la empresa UZEI en 1987). www.euskadi.net/euskalterm
- Diccionarios electrónicos:
 - Diccionario Elhuyar, español-euskera/euskera-español: www.euskadi.net/hizt_el
 - Diccionario 3000, español-euskera/euskera-español: www.euskadi.net/hizt_3000
 - Diccionario de sinónimos y antónimos: www.euskadi.net/hizt_sinon
 - Diccionario Morris inglés-euskera/euskera inglés: www.euskadi.net/morris
 - Diccionario enciclopédico Harluxet: www.euskadi.net/harluxet
 - Plug-in del Diccionario Elhuyar eu-es que se integra en Word 2000 y 2002 y en Internet Explorer: www.euskadi.net/euskara_soft
 - Diccionario Unificado (Hiztegi Batua): realizado por Euskaltzaindia-RALV, recopila la forma estándar de unas 20.000 palabras del léxico común del euskera; actualmente se está trabajando en la incorporación de 20.000 entradas más: www.euskaltzaindia.net
- Corpus estadístico del Euskera del Siglo XX, elaborado por la empresa UZEI por encargo de Euskaltzaindia-Real Academia de la Lengua Vasca: www.euskaracorpora.net
- Base de datos lexical del euskera (EDBL), elaborada por el grupo IXA de la Facultad de Informática de la Universidad del País Vasco: <http://ixa2.si.ehu.es/edbl/>
- Base de datos de toponimia de la CAV: www.euskadi.net/euskara_eaetoponimia

- Los nomenclátors de los nombres oficiales de las entidades de población de la Comunidad Autónoma Vasca, de los apellidos vascos (unos 10.000 recogidos hasta ahora), y de los nombres de persona: www.euskaltzaindia.net
- Boletín Oficial del País Vasco (BOPV), en euskera y español: www.euskadi.net/bopv

5. Traducción-localización a euskera de los productos de software de uso más generalizado:

- Windows 95, Windows 98; y el paquete de interfaz de Windows XP que se puede descargar desde www.microsoft.com/spain/windowsxp/euskera
- Word 6 y Office 2000 (Word, Excel y Outlook)
- Internet Explorer 4 y 5, y Outlook Express 5
- Sistema operativo Linux Mandrake, versiones 8.2 y 10.0, se puede descargar desde www.euskadi.net/euskara-soft
- Paquete de ofimática OpenOffice.org 1.0.2 y 1.1 para Windows y para Linux; está disponible en www.euskadi.net/euskara-soft
- Antivirus Panda Titanium
www.pandasoftware.es/titanium/euskara
- Antivirus Panda Platinum
www.pandasoftware.es/platinum
- Cursos Wit para autoaprendizaje (Word 2000 e Internet 2000)
- OCR para Omnipage versiones 11 y 12, y para Windows Word y Mac Word 2001; se puede descargar desde www.euskadi.net/euskara_soft

- Optenet: filtro de contenidos en Internet, www.optenet.com
- Prisma II: software de gestión para Pymes
- Cinco módulos de gestión de SAP; se pueden descargar desde www.euskadi.net/euskara_soft

A esto hay que añadir los programas de euskaldunización del software escolar (EIMA) del Departamento de Educación, Universidades e Investigación del Gobierno Vasco.

6. Construcción de escenarios virtuales:

- **HABENET**, para la enseñanza a distancia del euskera y para la formación del profesorado; proyecto realizado por el Instituto para la Euskaldunización y Alfabetización de Adultos cuyos objetivos son: mejorar las comunicaciones internas y externas de toda la red de HABE (HABE, euskaltegis, profesores, alumnos); facilitar el acceso al material didáctico, así como su homologación; desarrollar sistemas de autoaprendizaje y didáctica on-line; guiar mediante tutores la formación continua del profesorado; y mejorar y facilitar la gestión administrativa.
- **HIZPONET**, para optimizar la gestión de la planificación lingüística. Pretende mejorar las comunicaciones internas y externas de las infraestructuras sociales del euskera, públicas y privadas.
- **HIZNET**, curso de postgrado sobre planificación lingüística, organizado por la Fundación Asmoz de la Sociedad de Estudios Vascos (Eusko Ikaskuntza), la Universidad del País Vasco, la UEU (Universidad Vasca de Verano) y el Instituto de Sociolingüística SEI en colaboración con la Viceconsejería de

Política Lingüística. Este curso que se imparte por Internet está dirigido a técnicos de normalización lingüística, profesionales, investigadores y a todas aquellas personas que tengan responsabilidad en el ámbito de la normalización lingüística.

- **LINGUANET**, módulos de formación on line para técnicos de normalización lingüística que actúan en el ámbito socio-económico. Es una iniciativa de la Fundación Asmoz y de la Viceconsejería de Política Lingüística.
- La Viceconsejería de Política Lingüística y la Asociación de Municipios Vascos (EUDEL) en breve pondrán en marcha **módulos de formación on line** para **concejales de euskera**.

7. Almacenamiento, recuperación y búsqueda de información:

- Software Kapsula: primer software para la indexación y consulta, comercializado en euskera que utiliza técnicas de lematización (desarrollado como sistema de gestión documental). Proyecto desarrollado por las empresas Ametzagaina y Adur Software y financiado a través del Programa Intek del Departamento de Industria del Gobierno Vasco. www.kapsula.com

8. El reconocimiento automático del habla y la síntesis de voz

- Telefónica en colaboración con la Universidad del País Vasco ha desarrollado diversos proyectos de desarrollo de capacidades lingüísticas para la interpretación de la voz y la síntesis de mensajes sonoros en euskera.

- En 2003, el Gobierno Vasco firmó un convenio de colaboración con la empresa belga Scansoft Belgium para el desarrollo de los motores de síntesis y reconocimiento de voz, así como de gestión de diálogos, para el euskera. Agentes tecnológicos vascos (empresas y universidades) están elaborando los recursos lingüísticos necesarios para el proyecto: dos bases de datos acústicas, un léxico fonético y un gran corpus textual. El objetivo es que las empresas que desarrollan aplicaciones que utilizan tecnologías de voz dispongan de la posibilidad de integrar el euskera en sus productos.

9. El desarrollo de técnicas de base de la Ingeniería del Lenguaje, tales como herramientas para el procesamiento del lenguaje. Trabajos realizados por el grupo de investigación IXA de la Facultad de Informática de la Universidad del País Vasco www.ixa.si.ehu.es:

- Analizador morfológico MORFEUS
- Lematizador/etiquetador del euskera EUSLEM.
- Gain: buscador inteligente para Internet e intranet.
- Multimeteo-euskara: sistema de creación de pronósticos meteorológicos en euskera

También tenemos que mencionar el grupo de investigación DELi de la Universidad de Deusto que tiene una línea de investigación en las Tecnologías de la Información y de la Comunicación: www.deli.deusto.es

10. Funcionalidades de soporte al tratamiento de texto en el contexto ofimático y conversión de texto a texto (traducción).

- Memorias de traducción: el Servicio Oficial de Traducción del Instituto Vasco de Administración Pública sobre la base de la tecnología Trados está creando un corpus de memorias de traducciones jurídico-administrativas.
- Algunos prototipos de herramientas de ayuda a la traducción, desarrollados por el grupo IXA de la Universidad del País Vasco.
- Prueba de experimentación de la tecnología ATS aplicada al euskera, realizada en colaboración entre el Gobierno Vasco y la empresa catalana Automatic Trans.

11. Programas de ayuda a la investigación en ingeniería lingüística:

- Programa ETORTEK 2002-2005 del Departamento de Industria del Gobierno Vasco: agentes tecnológicos asociados de empresas y universidades vascas están desarrollando el proyecto HIZKING 21, la ingeniería lingüística para el siglo XXI.
- Programa INTEK del Departamento de Industria del Gobierno Vasco: dentro de este programa se desarrolló el proyecto ACOTE y sobre esa base se está desarrollando en la actualidad el proyecto AMETRA; los dos proyectos investigan herramientas de ayuda a la traducción.

12. Creación de empresas dedicadas a la infoingeniería lingüística: la empresa ELEKA; otras empresas que también están trabajando en este ámbito: Adur Software, Ametzagaina y CodeSyntax.

13. Integración de las tecnologías lingüísticas en la enseñanza: curso de postgrado HIZTEK, organizado por la Universidad Vasca de Verano.

14. Ayudas a la promoción de la integración del euskera en las Nuevas Tecnologías de la Información y de la Comunicación: la Dirección de Promoción del Euskera de la Viceconsejería de Política Lingüística del Gobierno Vasco publica una Orden anual. Así mismo, las Diputaciones Forales de Gipuzkoa y de Bizkaia también conceden subvenciones destinadas a este mismo fin. A esto hay que añadir también otras iniciativas locales.

4. Conclusiones y desafíos de futuro.

Es indudable que aunque la situación del euskera en lo que respecta a las TICs está lejos de alcanzar la situación de las lenguas mayoritarias, en los últimos años el euskera ha dado un gran salto hacia su integración en dichas tecnologías. En el futuro será necesario:

- Seguir ampliando la presencia del euskera en Internet, sobre todo ofreciendo contenidos y servicios para los niños y los jóvenes.
- Actualizar y alimentar las herramientas que ya están disponibles en Internet, como diccionarios, bases de datos, corpus, banco público de terminología, software en euskera, y habrá que crear nuevas como por ejemplo el corrector de gramática y de estilo (esperamos disponer de él en breve).
- Desarrollar las aplicaciones basadas en tecnologías de la voz que el mercado reclama: centros de llamada (operaciones bancarias, citas médicas, reservas, compras, información del tráfico, metereológica...), consulta de

documentos digitalizados, materiales para la enseñanza, juegos, etc.

- Desarrollar tecnologías que comprendan textos con rapidez: traducción automática de texto a texto y de voz a voz (speech-to-speech translation); una herramienta para resumir textos; y la herramienta que permita interrogar en grandes volúmenes de documentación y gestionar el conocimiento: indexar, clasificar, codificar y hacer extracciones o vaciados.

En consecuencia, las necesidades son claras, pero en el caso del euskera nos vemos obligados a establecer prioridades a la hora de emprender los proyectos puesto que no podemos desarrollar todos a la vez.

Tenemos que tener en cuenta que nuestro mercado es reducido pero, que poco a poco, tenemos que ir haciendo que la demanda aumente promoviendo el conocimiento y el consumo de los productos existentes. Es muy importante que ese mercado disponga de las aplicaciones en euskera que ya están disponibles en español, en francés o en inglés.

Las tecnologías lingüísticas no son perfectas pero sin duda son muy útiles y suponen una gran ayuda en nuestra sociedad actual.

La integración del euskera en las tecnologías lingüísticas conlleva la necesaria creación de recursos lingüísticos. Ese desarrollo tecnolingüístico es absolutamente necesario para la normalización de la lengua en dicho ámbito y para garantizar la presencia del euskera en el nuevo espacio social electrónico, telemático, digital y reticular, así como la supervivencia de nuestra lengua en la era de la globalización.