

"Nuevos modelos para un aprovechamiento didáctico del periodismo digital: la consulta léxica en línea",

Próspero Morán

Profesor Asociado de la Universidad Carlos III de Madrid

Resumen.

En la búsqueda de nuevos modelos de escritura hipertextual en el ámbito del periodismo digital, profundizamos en la adopción de fórmulas informáticas que permitan la comprensión de los textos y los vocablos en ellos utilizados en contextos lingüísticos globales normalizados de forma diversa. Concretamente, se plantea el uso de modelos de hipertexto que permitan la lectura de periódicos digitales en lengua castellana, redactados en España, por parte de hablantes de español de Iberoamérica. El sistema se basa en licencias de distribución copyleft y aprovecha recursos disponibles en Internet.

Palabras clave: periodismo digital, software lingüístico, copyleft.

Introducción

Desde la irrupción de Internet son muchas las voces que se han alzado para considerar al nuevo Medio de medios “un peligro de primer orden para la humanidad”. No sólo desde una perspectiva política, social o económica, sino también informativa y desde luego cultural y lingüística.

La afirmación del Presidente de Francia, Jacques Chirac, respecto al impacto de Internet sobre la lengua, y especialmente sobre el francés es probablemente la más directa pero ha habido muchas otras de similar contundencia centradas en la amenaza del inglés sobre lenguas más o menos minoritarias en la sociedad del conocimiento.

Casi nadie gusta ya de recordarlo, pero apenas transcurrido cerca de medio siglo desde de la aparición de la Biblia de Gutenberg (1455) el Papa Alejandro VI ya había ampliado la censura a los libros seculares (1501). Cuatro siglos después, ocurría otro tanto tras la llegada del telégrafo, el teléfono y la tecnología de radiodifusión que iban a socavar los cimientos de la sociedad de entonces y a ser la correa transmisora de la subversión.

En todos los casos, las polémicas generadas se centraron en la libertad de prensa (tan importante para quien tenía una prensa o quien recibía una licencia de un gobierno para ocupar el espacio radioeléctrico a cambio siempre de cierta colaboración con el poder establecido). Pero casi nadie se ha detenido suficientemente en el aspecto más importante: el cultural y lingüístico. Lo que la imprenta permitió en primer lugar fue que las traducciones vernáculas de la Biblia llegasen a miles de personas iniciando una polémica, que persiste con el paso de los siglos, sobre el uso de las lenguas locales en lugares religiosos. Y no hace falta retrotraerse mucho para recordar cuando la creciente popularidad e influencia de la radio suscitó el establecimiento de normas para una pronunciación correcta y la conveniencia o no de permitir los dialectos y acentos locales, lo que, aún hoy en día, en pleno siglo xxi, sigue siendo fuente de controversia¹. La irrupción en los medios audiovisuales españoles de acentos periféricos, como el canario o el andaluz, en los informativos de la televisión pública pertenece a la historia más reciente, la de la transición democrática del último cuarto del pasado siglo².

A estas alturas, hasta los niños saben qué es Internet: Un conjunto de redes informáticas con reglas comunes que permite enviar mensajes desde cualquier ordenador (servidor o cliente) de una red a cualquier otro (servidor o cliente) de otra. Desde su creación y puesta en marcha en la década de los sesenta en los Estados Unidos ha crecido tan rápidamente que se encuentra ya extendida en uso eficaz por todo tipo de empresas, instituciones, gobiernos, asociaciones, organizaciones y, desde luego, particulares.

En este momento, con la Red de redes convertida en una realidad cotidiana, con la denominada aldea global asentada sobre la aldea territorial y cultural en que aún nos movemos, algunos estudiosos comienzan a preguntarse qué es un dialecto y si podemos o

¹ En realidad nos encontramos ante una máxima fundamental de la locución periodística: la importancia de la capacidad de reconocimiento de la voz del informador radiofónico y/o televisivo por el hecho de que la memoria auditiva sea superior a la visual. José María Carrascal lo recordaba cuando aseguraba que “no es lo que uno dice, no es la palabra o el pensamiento. Es la voz en su sentido más elemental, el tono, el timbre, lo que queda en el espectador”. Carrascal, José María (1992) “Al filo de la medianoche... y algo más”, Madrid, Espasa-Calpe.

²Ivan Tubau asegura que “puede decirse que los acentos regionales no afloraron en la radio y la televisión de España hasta bien entrada la democracia. La pionera más notoria fue la canaria Cristina García Ramos, que introdujo la fonética andaluza en los informativos nacionales. A partir de ahí comenzaron a parecer lícitos una serie de acentos que habían estado proscritos, incluidos el gallego y el catalán, hasta entonces frecuente objeto de chanza...” Tubau, Ivan (1993) “Periodismo oral”, Barcelona, Paidós.

no distinguir un dialecto unificado de una lengua o de una mezcla de ambas o de una lengua plagada de neologismos indefinibles. Esas posibles distinciones nos plantearían la existencia real de la aldea global y si Internet puede considerarse en la actualidad un medio lingüístico homogéneo. Pero no es esta, por más que resulte enormemente interesante y que convenga afrontarla no tardando mucho desde una perspectiva hispana y romanista, la línea de trabajo inicial que nos proponemos afrontar y de hecho estamos ya afrontando para caracterizar con modelos de valor añadido la prensa digital actual, que inexplicablemente está dejando de lado innumerables posibilidades lingüísticas y didácticas que la tecnología pone a su alcance³.

El concepto de variedad lingüística esta adquiriendo en la actualidad, dentro del ámbito del periodismo digital, una importancia desmedida alimentada sin duda por la premura en la redacción, la excesiva juventud e inexperiencia de los redactores y/o colaboradores infrapagados que pueblan los medios digitales y la falta de redactores-jefes expertos que al conocimiento del oficio periodístico sumen conocimientos tecnológicos imprescindibles para manejar con soltura hiperenlaces y sistemas de sindicación. Estos problemas y ausencias que caracterizan básicamente la información digital actual son caldo de cultivo idóneo para que se entremezcle en la redacción digital el habla y la escritura, los dialectos regionales y de clase, los géneros profesionales (como las jergas legales, deportivas, tecnológicas o científicas) y una enorme variedad de otros estilos de expresión.

David Crystal⁴ lo clarifica acudiendo a Naughton⁵ y a Tim Berners-Lee:

“Siendo el lenguaje un indicador tan sensible de los cambios sociales, resultaría verdaderamente muy sorprendente que un fenómeno tan radicalmente innovador no tuviera su correspondiente impacto en el modo de comunicarnos. Y así puede afirmarse. El lenguaje es el corazón de Internet, la actividad de la Red es interactividad. «La Red es, en realidad, un sistema que une un gran número de ordenadores y a la gente que los usa». Estas son las palabras de Naughton, y sus propias cursivas. Internet no es solamente un hecho tecnológico; es un hecho social, como ha destacado Berners-Lee, y su principal activo es el lenguaje”

³ Los diarios digitales aún mantienen la tendencia de reproducir casi literalmente la estructura y contenidos de la versión impresa de su cabecera. Tanto en España como en Iberoamérica se cuentan con los dedos de una mano las excepciones a esta norma. Aún no ha habido planteamientos innovadores en el ámbito de lo lingüístico, básicamente porque las líneas de trabajo aún se restringen a cuestiones estructurales. Recientemente, el diario argentino Clarin.com ha realizado una importante e innovadora renovación de diseño y estructura que resulta una excepción que confirma la regla del escaso riesgo que asumen los editores digitales en España e Iberoamérica.

⁴ Crystal, David (2003), “El lenguaje de Internet”,

⁵ Naughton, John (1999), *A brief history of the future: the origins of the Internet*, Londres, Weidenfeld and Nicolson.

Enormes posibilidades inexploradas

Los rasgos diferenciales de la variedad de una lengua son de distintos tipos reconociéndose habitualmente en muchos estudios estilísticos cinco clases para la lengua escrita que no están recibiendo demasiada atención por parte de los medios de comunicación digitales ni en lengua inglesa ni en lenguas hispánicas, que es el ámbito que deseamos acotar en esta aproximación a una realidad todavía por explotar. Esas cinco clases de lengua escrita se caracterizan por los rasgos que representan, comenzando en primer lugar por los rasgos gráficos.

1.- **Rasgos tipográficos:** En este sentido, la tipografía todavía sigue usándose en la Red con recursos propios a todas luces de la letra impresa o incluso sin aprovechar las posibilidades ya explotadas en los medios de comunicación impresos. Además de la presentación general de la página, el uso del espaciado o del interlineado e incluso del color, también se desdeñan en la actualidad en los medios digitales españoles en particular (y de todo el mundo en general)

2.- **Rasgos ortográficos:** Aún cuando puedan considerarse fronterizos con los rasgos tipográficos, los que caracterizan la ortografía de una lengua tienen una especial importancia en casos como el inglés donde se distingue al americano y al británico por la distinta grafía fonética de su pronunciación (como en “colour” frente a “color”). Este tipo de caracterización no tiene presencia en el ámbito de las lenguas hispánicas, salvo excepciones, pero podría irrumpir con fuerza en cualquier momento, sin que se haya previsto ningún tipo de método para sacar partido a posibilidades expresivas importantes.

3.- **Rasgos gramaticales:** Probablemente los que ofrecerán en el futuro mayores posibilidades, dadas las enormes posibilidades de la morfología y la sintaxis en la escritura tradicional que podrían en la escritura virtual general nuevas y ricas formas de expresión. Un caso significativo es el modelo del inglés religioso que permite el uso de pronombres arcaicos en la segunda persona del singular (thou, thee, thine...), pero hay otros y en una aldea global este tipo de usos específicos y no normalizados, que extrañarían a cualquier estudiante español del denominado inglés de la Reina, podrían explicarse en la misma

página web donde aparezcan sin hacerlo específicamente. Esto es, usando plataformas de software y modelos como el que aquí desarrollaremos y expondremos más adelante.

4.- **Rasgos léxicos:** el vocabulario de la lengua es el punto de partida de cualquier problema de lectura al que se puede enfrentar tanto un lector avezado como un aprendiz de esa lengua, bien por estar en edad de aprendizaje o por tener distinta lengua materna, o incluso por desconocer la jerga profesional que se esté empleando. Siguiendo a Crystal que nos ha servido de referente constante en este camino introductorio a nuestras propuestas, el caso del inglés jurídico es enormemente ejemplificador para el caso que nos ocupa: emplea algunas expresiones como *heretofore* (hasta ahora), *easement* (servidumbre) y *alleged* (presunto, alegado), así como algunas frases hechas como *signed sealed and delivered* (firmado, rubricado y aplicado) y expresiones latinas como *ex post facto* (de hechos posteriores) que difícilmente comprenderían lectores medios de la prensa digital británica o incluso algún que otro avezado y culto lector. Y esa importancia y la facilidad con la que se pueden aprovechar herramientas ya existentes con una pequeña ayuda de la programación informática y del software libre van a hacer que nos detengamos en estos rasgos para el modelo que propondremos en esta propuesta.

5.- **Rasgos discursivos:** Finalmente, la estructura de un texto, su coherencia, es otros factores en los que se pueden introducir modelos que aprovechen las capacidades informativas y didácticas que los medios digitales nos pueden ofrecer frente a todos los medios de masa anteriores.

Por supuesto, la prensa digital sigue siendo básicamente textual hasta el momento, pero no tardando mucho tendrá que dejar de serlo para penetrar por la senda del sonido y con ello por la senda del registro hablado que traerá seguramente diálogos interactivos y reconocimiento de voz, siguiendo los vaticinios de Negroponte y otros teóricos que nos han advertido de la aplastante lógica de que la comunicación con el ordenador camina ineludiblemente hacia la forma en que nos relacionamos más habitualmente y con mayor comodidad los humanos: la comunicación verbal.

En ese momento aparecerían otros dos rasgos que ofrecen igualmente múltiples posibilidades de apreciación y que podrían generar caracterizaciones y posibilidades de enorme interés en el ámbito periodístico. Se trata de los **rasgos fonéticos** (timbres de voz,

registros vocales y otras matizaciones que ofrecen posibilidades ingentes para una información más expresiva y matizada que en la televisión ya tienen referentes) y los **rasgos fonológicos** (el aprovechamiento de sonidos dialectales e incluso las posibilidades de la información plurilingüe de la que los sistemas de emisión dual en televisión son tan sólo la punta del Iceberg, un primer y tímido paso)

Modelos digitales de información léxica

Hemos apuntado más arriba que el vocabulario de la lengua es el punto de partida de cualquier problema de lectura al que se puede enfrentar tanto un lector avezado como un aprendiz de esa lengua y nuestros primeros esfuerzos se dirigen precisamente a ese ámbito convencidos de que en él pueden realizarse importantes avances sin grandes esfuerzos en fuerza laboral, dedicación de los profesionales en su labor redactora o consumo de ancho de banda.

Decíamos también que actualmente se detecta en los medios digitales una ausencia generalizada de profesionales formados y expertos que combinen conocimientos tradicionales del oficio periodístico con habilidades y conocimientos propios del mundo digital, pero estas carencias podrían estar siendo subsanadas con estrategias de aprovechamiento de las sinergias que permite la información digital, aunque para ello tendríamos que disponer al frente de esos medios de profesionales implicados en procesos innovadores y dispuestos a asumir riesgos, lo que en ámbitos de la empresa privada y en épocas sumidas en el miedo nacido tras la explosión de la burbuja tecnológica parece difícil exigir esas actitudes. Confiamos en que los medios públicos asuman ese carácter innovador y el riesgo necesario para innovar y abrir nuevos caminos.

Para imaginar siquiera la dificultad que entraña la problemática terminológica a la que se enfrentan el lector, de una parte, y el periodista diariamente en el ejercicio de su profesión, baste la transcripción de este mensaje de junio de 2003 de una lista de distribución referida a redes inalámbricas y que es un texto al que puede tener que enfrentarse cualquier periodista sin conocimiento alguno sobre redes y mucho menos sobre aquellas que se forman sin empleo de cableado alguno:

“Gracias Ed0 y S.J. . Ya va el puerto 80 cautivo con buen html que no se cachea; también he hecho cautivo el SMTP y así hago un relay con el postfix, mediante iptables (para que haga SASL). Si se os ocurre cómo hacer lo mismo con el POP, se admiten ideas (ojalá squid dejase POP!).

No pensaba que el OSPF fuera tan sencillo (reglas "a la vez" con el route clásico); es que con ese nombre...

En cuanto a autenticar dando un "ticket" con DHCP (una ip distinta) creo que se presta al spoofing. Voy a empezar con Radius (si alguien sugiere otra cosa, escucho), así que cualquier manual que me indiquéis será bien recibido.

El tema de un AP en vez de host_ap lo planteaba desde una postura práctica: subir un armatoste a la azotea con una prism2 de 90eur o un cacharrín (con un máximo de 16 usuarios) de 110eur. No sé que hacer.”

Esta es una conversación habitual en foros más o menos expertos que, como parece obvio, no está al alcance de muchos periodistas, ni tan siquiera entre los que se dedican a cubrir información tecnológica en medios generalistas, no digamos ya en los que tan sólo dedican su atención a la tecnología periódicamente y en medios que ni son de primera línea ni tienen influencia territorial que alcance a todo el Estado.

En análisis anteriores⁶ hemos puesto de manifiesto que enfrentarse a textos como el presente requeriría de técnicas y herramientas hipertextuales que, por desgracia siguen sin desarrollarse y que merecerían una mayor atención por parte de los investigadores y de los periodistas. Las herramientas lingüísticas en entornos hipertextuales de que disfruta la profesión periodística son aún mínimas por no decir inexistentes. Y hacia ese campo estamos enfocando nuestra atención algunos periodistas y profesores, en la convicción de que se hace necesario ofrecer soluciones concretas, en forma de modelos, que además empleen software libre o de código abierto y se distribuyan con licencias de “copyleft” o lo que es lo mismo exentas de derechos de autor. Con herramientas prácticas que abunden en la comprensión lingüística de los textos y los vocablos terminológicos escogidos tendremos posibilidades de avanzar en el ámbito de la normalización y la normativización lingüística del mundo digital.

⁶ Hemos abordado esta problemática en numerosas ocasiones, siendo la más reciente la ponencia "Problemas terminológicos de la redacción y la locución informativas en el periodismo tecnológico español", presentada en el II Encuentro Global de Ciberperiodismo, celebrado el 9 y 10 de marzo de 2004, en el Campus de La Berzosa de la Universidad Antonio de Nebrija de Madrid, cuyas actas están pendientes de publicación y en la ponencia "Del periodismo multimedia al periodismo digital: el perfil del metaperiodista" presentada en las Primeras Jornadas de Periodismo Digital de Murcia, celebradas en Mayo de 2004, coincidiendo con la Feria SICARM.

Pero sin llegar a los extremos de este tipo de textos confusos, plagados de jerga y acrónimos importados cuando no inventados, propios tan sólo de informaciones tecnológicas dirigidos a públicos y especializados, el lector se enfrenta actualmente a la necesidad de usar el diccionario en busca de significados léxicos no conocidos o simplemente no recordados. En la prensa escrita dispone probablemente de un medio también impreso para satisfacer esa necesidad con un procedimiento lento que exige una tarea de búsqueda y localización del término en cuestión.

Inexplicablemente, en el ámbito de la prensa digital seguimos teniendo que acudir a un procedimiento igualmente lento, de una celeridad propia del mundo real que no del virtual. Pese a que la tecnología puede impedirlo sin grandes esfuerzos, cada vez que encontramos un término en una información periodística cualquiera cuyo significado resulte dudoso para nosotros debemos acudir a un diccionario en línea⁷ y proceder a realizar una búsqueda del término en cuestión. Siempre de forma manual y a bastantes golpes de teclado y de ratón, cuando existen posibilidades de llegar a los mismos resultados, de forma integradora, con tan sólo un par de clics de ratón.

Lo que desmenuzamos aquí no es más que el punto de partida de un proceso que pretende facilitar el uso y aprovechamiento de los rasgos léxicos de la escritura digital en el ámbito de la información periodística, en un primer paso, extendiendo estos procedimientos al resto de rasgos propios de la lengua escrita y, en último término, de la lengua hablada. Y ese punto de partida es un modelo reproducible de herramienta que permita aplicar a cualquier hipertexto de forma sencilla sistemas de consulta interactiva sobre términos y vocablos contenidos en el texto que nos ocupe. Sistemas basados en simples atajos de teclado que ante la visualización de una página web nos permitan localizar en dos golpes de ratón/teclado un vocablo desconocido o del que deseemos conocer más profusamente significados, aclarar una dialogía o polisemia desconcertante, o simplemente traducir del anglosajón. Todo con la intervención automática de un diccionario asentado en una base de datos local o remota con el que también pueda interactuar el periodista en el momento inmediatamente posterior a la redacción de la noticia.

⁷ El principal diccionario de referencia es el que ofrece la Real Academia Española (RAE) en su página web, aunque los usuarios se inclinan por emplear portales que combinan la consulta al diccionario normativo del castellano con diccionarios de traducción entre lenguas, como ocurre con diccionarios.com.

Propuestas, en todo caso, conducentes a resolver los problemas terminológicos a que se enfrenta el periodista digital, pero también y sobre todo su lector/receptor.

Un modelo normalizado y normalizador

El modelo que proponemos pretende facilitar la búsqueda rápida de un término o un grupo de ellos y la consulta de su o sus significados sin necesitar anotar los vocablos dudosos, sirviendo al mismo tiempo para completar los comentarios periodísticos necesarios para aquellos términos que no aparezcan registrados en el diccionario normativo de la lengua en cuestión y que habitualmente explicaba a pie de página o entre paréntesis el periodista en cuestión.

En realidad lo que estamos proponiendo es un simple sistema que convierte cada palabra escrita en una página web⁸ en un hiperenlace a un diccionario que resuelva cualquier duda léxica.

En el modelo que mostramos aquí se usa una herramienta externa, en este caso el diccionario en línea de la Real Academia Española (RAE), pero podría sustituirse por cualquier diccionario instalado en un servidor interno del medio de que se trate o cualquier otro sistema que consulte una base de datos preparada al efecto.

Lo que hacemos es emplear un “script” programado con sencillez que consta de dos partes principales.

La primera es la más sencilla y consiste simplemente en responder a la pulsación de teclas del usuario y llamar al fichero que genera una lista de palabras (con su enlace a la página de la R.A.E.). De esta forma, en el texto de la página web en cuestión (que hemos colocado en la dirección www.prosperomorán.com/rae) bastará con pulsar en el teclado una letra (la “p” por ejemplo) para que todas las palabras que comiencen por esa letra en el texto aparezcan listadas en una ventana emergente (o pop-up) e hiperenlazadas de forma automática a una consulta contra la base de datos del diccionario de la RAE

⁸ Y por extensión también en una información periodística digital, que es el motivo central de aplicación de nuestro modelo, aunque no es el único factible.

La primera parte de nuestro modelo es un script que se ejecuta en el lado del cliente y que está programado en javascript:

```
<script language="javascript"><!--
/* Ejecutar la función “ handleKeyUp” en caso de que se produzca el evento “onkeyup” */
document.onkeyup = handleKeyUp;
function handleKeyUp(e)
{
// Obtener el evento (en caso de que no haya sido obtenido previamente)
if(!e)
    var e = window.event;
// key es el código de tecla
key = e.keyCode;
// si el código de tecla es >=65 (A) o <=90 (Z) o la tecla es la Ñ
if (((key>=65)&&(key<=90))||(key==192)) {
// pasar el código de tecla a la tecla pulsada propiamente
if (key==192) {
    key='ñ';
} else {
    key=String.fromCharCode(key);
}
}
/* abrir en una ventana aparte el fichero “palabras.php” pasándole como parámetro la tecla pulsada */
window.open("palabras.php?letra="+key+"","","width=400,height=300,scrollbars=yes");
}
}
//--></script>
```

A partir de los comentarios es fácil de entender lo que hace este código por parte de cualquier lector mínimamente formado en programación. Básicamente, el script permite que se ejecute una función para cada pulsación de tecla. Esa función comprueba que la tecla pulsada sea una letra y en ese caso abre una ventana donde llama al fichero PHP con el código que genera la lista de palabras.

El segundo script es un fichero al que hemos denominado “palabras.php” que devuelve la lista de palabras que comienzan por la tecla indicada y un enlace a la función de la R.A.E. mediante el cual obtenemos el significado de las mismas.

Ese fichero recibe un parámetro que es la letra pulsada, aunque ofrece múltiples posibilidades adicionales de fácil aplicación. Por ejemplo, se le podría pasar un segundo parámetro con el texto completo donde debe buscar. Pero, en vez de hacerlo así y como el texto lo estamos obteniendo de una base de datos, volvemos a buscar el texto antes de procesarlo.

Los pasos que se dan en este sencillo programa son los siguientes:

- Obtener el texto en el cual debemos realizar la búsqueda (mediante una consulta a una base de datos o haciendo una lectura del fichero).
- Eliminar etiquetas HTML que pueda haber en el mismo.
- Cambiar las entidades y sustituirlas por la letra equivalente. Debemos hacerlo, al menos, en todas las tildes (á é ...), en las eñes (ñ ...) y en las diéresis (ü).
- Pasar todo el texto a minúsculas (por comodidad y para que “Hola” no sea distinto a “hola”).
- Eliminar otros símbolos (números, signos de puntuación, caracteres extraños) y sustituirlos por un espacio.
- Separar las palabras por espacios y meterlas en una lista. Sólo hace falta incluir las que empiezan por el carácter buscado.
- Y, finalmente, ordenar la lista y poner el enlace al buscador de la R.A.E.

Este enlace se dirige hacia la dirección web <http://buscon.rae.es/draeI/SrvltGUIBusUsual> con dos parámetros: “TIPO_BUS” es que designa el tipo de búsqueda (coincidencia exacta de palabra, coincidencia parcial, etc.) Nos interesa una coincidencia exacta por lo cual ponemos “0”. El segundo parámetro es “LEMA” que recoge la palabra buscada.

Un ejemplo de enlace sería en este sentido:

http://buscon.rae.es/draeI/SrvltGUIBusUsual?TIPO_BUS=0&LEMA=palabra

Este segundo script puede visualizarse a continuación:

```
-----
<?php
// Obtener una lista de palabras (sin repetir) de un texto
/* Pasos:
- Eliminar HTML
- Cambiar entidades y sustituirlas por la letra
- Pasar a minúsculas
- Eliminar otros símbolos
- Separar por espacios y meterlos en una lista
*/
function entidades2tildes($str) {
    // minúsculas
    $str=str_replace("&aacute;","á",$str);
    $str=str_replace("&eacute;","é",$str);
    $str=str_replace("&iacute;","í",$str);
```

```

$str=str_replace("&oacute;","ó",$str);
$str=str_replace("&uacute;","ú",$str);
// mayúsculas
$str=str_replace("&Aacute;","Á",$str);
$str=str_replace("&Eacute;","É",$str);
$str=str_replace("&Iacute;","Í",$str);
$str=str_replace("&Oacute;","Ó",$str);
$str=str_replace("&Uacute;","Ú",$str);
// uml (en castellano sólo la u)
$str=str_replace("&uuml;","ü",$str);
$str=str_replace("&Uuml;","Ü",$str);
// eñes
$str=str_replace("&ntilde;","ñ",$str);
$str=str_replace("&Ntilde;","Ñ",$str);
// otros
$str=str_replace("&quot;","\"", $str);
$str=str_replace("&lt;","<",$str);
$str=str_replace("&gt;",">",$str);
return $str;
}
function eliminarnocaracteres($str) {
// para cada letra
for ($i=0;$i<strlen($str);$i++) {
// si no es un caracter (a-z ñ á é í ó ú ü) sustituirla por espacio
if
(!(((($str[$i]>='a')&&($str[$i]<='z'))||($str[$i]=='ñ')||($str[$i]=='ü')||((($str[$i]>='á')&&($str[$i]<='ú'))
)) {
    $str[$i]=" ";
}
}
return $str;
}

function norepetidas($words) {
    $temp=array();
    foreach ($words as $word) {
        $word=trim($word);
        if ((!(in_array($word,$temp))&&($word<>""))) {
            $temp[]=$word;
        }
    }
    // de paso las ordena :)
    sort($temp);
    return $temp;
}

function listadepalabras($str) {
// eliminar HTML
$str=ereg_replace("<[^>]*>"," ", $str);;
// eliminar "slashes" \
$str=stripslashes($str);

```

```

// cambiar entidades y sustituirlas por la letra
$str=entidades2tildes($str);
// pasar a minúsculas
$str=strtolower($str);
// eliminar símbolos no caracteres
$str=eliminarnocaracteres($str);
// construir la lista
$words=explode(" ",$str);
// que no estén repetidas
$words=norepetidas($words);
return $words;
}

// busca todas las palabras que empiecen por una letra en el array con todas las palabras
function busca($letra,$words) {
    $temp=array();
    foreach ($words as $word) {
        if ($word[0]==$letra) {
            $temp[]=$word;
        }
        // caso especial: tildes
        switch ($letra) {
            case 'a': if ($word[0]=='á') {
                $temp[]=$word;
            }
            break;
            case 'e': if ($word[0]=='é') {
                $temp[]=$word;
            }
            break;
            case 'i': if ($word[0]=='í') {
                $temp[]=$word;
            }
            break;
            case 'o': if ($word[0]=='ó') {
                $temp[]=$word;
            }
            break;
            case 'u': if ($word[0]=='ú') {
                $temp[]=$word;
            }
            if ($word[0]=='ü') {
                $temp[]=$word;
            }
            break;
        }
    }
    return $temp;
}

$letra=strtolower($letra);

```

```

// Obtener el mismo texto que "index.php" de un fichero
$texto=join("file('./contenido.html'));

// Obtener la lista de palabras que hay en el texto
$words=listadepalabras($texto);

// Buscar la letra
$words=busca($letra,$words);
?>
<html>
<head>
<title><? echo $letra; ?></title>
</head>
<body>
<?php

foreach ($words as $word) {
    echo
                                "<a
                                href=\"#"
onclick=\"window.open('http://buscon.rae.es/draeI/SrvltGUIBusUusual?TIPO_BUS=0&LEMA=".$
word." ','width=500,height=400,scrollbars=yes')\">".$word."</a><br>\n";
}

?>
</body>
</html>

```

El uso de este sistema de consulta de léxico en informaciones o textos digitales es el punto de partida de un sistema más complejo que en su materialización final pretende ofrecer al periodista bajo licencia “copyleft” y, sobre todo, al lector, un conjunto de herramientas lingüísticas que potencien la capacidad didáctica de la información, el uso pedagógico de los textos informativos y abran nuevas posibilidades al periodismo digital que le vayan confiriendo la personalidad que sin duda ya insinúa.

El sistema está siendo implementado en una publicación digital cultural en lengua asturiana que tiene presencia en la red desde 1995 y que tras un período de inanición ha irrumpido nuevamente con fuerza en abril de 2004: Asturnews. También está siendo mejorado y ampliado para su futuro uso, bajo licencia copyleft, por todos los medios de comunicación públicos y privados que lo deseen, tanto en lengua castellana, como en otras lenguas indoeuropeas.

BIBLIOGRAFÍA

- Adams, Lin, Lori Toomey y Elizabeth Churchill (1999), «Distributed research teams: meeting asynchronously in virtual space», en *Journal of Computer-Mediated Communication* 4 (4), . <http://www.ascusc.org/jcmc/vol14/issue4/adams.html>
- Baym, Nancy K. (1993), «Interpreting soap operas and creating community: inside a computer-mediated fan culture», en *Journal of Folklore Research* 30 (2/3), páginas 143-176.
- Baym, Nancy K. (1995), «The performance of humor in computer-mediated communication», en *Journal of Computer-Mediated Communication* 1(2).
<http://www.ascusc.org/jcmc/voll /issue2/bayni.html>
- Bechar-Israeli, Haya, (1996) «From <Bonehead> to <cLoNehEAd>: nicknames, play, and identity on Internet Relay Chat», en *Journal of Computer-Mediated Conununication* 1 (2).
<http://www.ascusc.org/jmc/voll /issue2/bechar.html>
- Berners-Lee, Tim, (1999) «*Weaving the Web*», Londres, Orion Business Books. (Trad. esp.: *Tejiendo la red*, Madrid, Siglo XXI, 2000)
- Crystal, David y Derek Davy, (1969) «*Investigating English styl«e*, Londres, Longman.
- Crystal, David y Derek Davy, (1976) «*Advanced conversational English«*, Londres, Longman, 1976.
- Crystal, David y Randolph Quirk (1964), «*Systems of prosodic and paralinguistic features in Englis«h*, La Haya, Mouton.
- Cumming, John D. (1995), «The Internet and the English language», en *English Today* 11 (1), págs. 3-8.
- Díaz Noci, Javier (2002). «*La escritura digital. Hipertexto y construcción del discurso en el periodismo electrónico«o*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Díaz Noci, Javier; Meso, KoI do (1999). «*Periodismo en Internet. Modelos de la prensa digita«l*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Díaz Noci, Javier; Meso, KoI do (con Armañanzas, Emiliana) (1996). «*El periodismo electrónico. Información y servicios multimedia en la era del ciberespacio«*. Barcelona: Ariel.
- García García, Francisco (1999). «Estructuras narrativas y nuevas tecnologías de la información». En. López, Xosé; Soengas, Xosé (eds.). *A comunicación audiovisual na era dixital*. Santiago de Compostela: Imprenta Universitaria, 90-92.

Gómez Y Méndez, José Manuel; Leal Gil, Eva (2001). «Delimitación del vocablo Cibernética y otras voces tecnológicas en periodismo». En: *Estudios sobre el mensaje periodístico*, 7, 95-107.

López, Xosé; Gago, Manuel; Pereira, Xosé (2000). «0 novo xornalismo electrónico». Santiago de Compostela: Edicións Lea.

López, Xosé; Otero, María Luisa (2001). «As factorías da Sociedade da Información. Produción de contidos para Internet». Santiago de Compostela: Edicións Lea.

Morán López, Próspero Manuel (2004) "Problemas terminológicos de la redacción y la locución informativas en el periodismo tecnológico español". II Encuentro Global de Ciberperiodismo, 9 y 10 de marzo de 2004. Universidad Antonio de Nebrija, Madrid.

Morán López, Próspero Manuel (2004) "Del periodismo multimedia al periodismo digital: el perfil del metaperiodista". Primeras Jornadas de Periodismo Digital, Mayo 2004, Jornadas SICARM, Murcia.

Morán López, Próspero Manuel (1996) «Llingua y Nueves Tecoloxías». Mieres, Ediciones del Norte.

Rodríguez de las Heras, Antonio (1991). «Navegar por la información». Madrid: Fundesco.