

## TRABAJAR EN NEOLOGÍA CON UN ENTORNO INTEGRADO EN LÍNEA: LA ESTACIÓN DE TRABAJO OBNEO

M. TERESA CABRÉ CASTELLVÍ Y ROSA ESTOPÀ BAGOT  
UNIVERSIDAD POMPEU FABRA (BARCELONA)

**Resumen:** La detección, almacenamiento y análisis de neologismos han sido durante siglos una tarea manual, individual y enormemente ardua, que, a menudo, se ha basado en criterios intuitivos. Pero el panorama actual de la investigación en este ámbito ha cambiado significativamente y el principal motor de este cambio ha sido el uso de las tecnologías aplicadas al trabajo neológico. En este artículo nos proponemos tres objetivos: a) en primer lugar, presentar la metodología del Observatorio de Neología de Barcelona, explicando cómo se pasó de un proceso de trabajo manual a un proceso de trabajo automatizado, y cómo trabajar usando herramientas aisladas y disgregadas se substituyó por trabajar con un entorno integrado (Estación de trabajo OBNEO); b) en segundo lugar, describir las características de la Estación de trabajo OBNEO y de las herramientas que lo componen: la Plataforma de trabajo (Plataforma OBNEO), el detector de neología (BUSCANE0), el gestor de diccionarios (GESTODIC OBNEO) y el consultor de neologismos (Buscador BOBNEO); c) y, en tercer lugar, analizar el cambio metodo-

**Abstract:** For centuries, the extraction, storage, and analysis of neologisms have been manual, individual, and extremely arduous tasks, which have often been based on intuitive criteria. The last few decades, however, have seen significant changes in research in this field. These changes are due mainly to the use of technology applied to work in neology. This article has three main goals. The first is to introduce the methodology used at the Observatori de Neologia in Barcelona and explain the transition from manual to computerized work as well as the shift from working with separate and non-integrated tools to using an integrated environment, namely the OBNEO Workstation. The second goal is to describe the features of this workstation and its integrated tools: the work platform (OBNEO Platform), the neologism extraction tool (BUSCANE0), the dictionary management system (GESTODIC OBNEO), and the neologism search engine (OBNEO Searcher). Finally, the last goal is to examine how using an integrated online work environment such as the one at the Observatori has

lógico que ha supuesto para el investigador en neología utilizar una estación de trabajo integrado en línea como la que se utiliza en el Observatorio.

**Palabras clave:** neología, tecnología, tratamiento automático de la información, extracción de neología.

led to changes in methodology for researchers in neology.

**Keywords:** neology, technology, automatic information processing, neologism extraction.

## 1. INTRODUCCIÓN

La neología es un hecho natural y habitual de todas las lenguas: se crean día a día decenas de palabras nuevas. Si ello es así, el estudio de la neología debería ser un punto clave en la investigación sobre el funcionamiento de una lengua. Pero, a pesar de la magnitud y la importancia de este fenómeno lingüístico, los neologismos, como objeto de estudio de la lingüística teórica, han recibido escasa atención por considerarlos un aspecto marginal. Sablayrolles (2000) muestra a través de un extenso estudio el insignificante lugar que la neología ha ocupado en los principales modelos lingüísticos contemporáneos. Esta falta de interés por parte de los lingüistas es, según Vallès (2004), una consecuencia lógica de la contraposición teórica entre palabras posibles y palabras reales, y entre competencia y actuación; pues muchos lingüistas se han dedicado más a la competencia léxica que a la actuación. En nuestra opinión, el estudio de los neologismos es imprescindible en un modelo gramatical basado en el uso real de la lengua, que se proponga explicar el sistema y el uso léxicos de los hablantes de una comunidad lingüística.

En contraposición con este interés desigual por parte de los lingüistas teóricos, la neología ha sido a menudo foco de atención para los lingüistas con finalidades aplicadas. Así, lexicógrafos, terminógrafos, técnicos lingüistas, traductores e, incluso, logopedas o analistas del discurso, han tenido en cuenta la neología en su práctica profesional cotidiana. En estos contextos profesionales, las palabras nuevas y, más concretamente, la falta de palabras (los vacíos denominativos), suelen ser «un problema» que hay que resolver. Para ello, precisan abordar una serie de decisiones teórico-metodológicas sobre la neología: qué se entiende por neología, qué se considera «palabra nueva», cómo se detecta y cómo

se procesa una palabra nueva, y cómo se almacena la información. Decisiones todas ellas difíciles de tomar y no exentas de polémica.

Durante muchos años, para recolectar nuevas palabras, los investigadores han tomado como referencia el criterio psicolingüístico (Rey, 1976): la intuición que tiene todo hablante sobre las palabras de su lengua –lo que Sablayrolles denomina *sentimiento de neologicidad*. Sin embargo, el trabajo manual, asistemático e intuitivo no es adecuado para gestionar grandes cantidades de palabras, ni cuando se quiere estudiar un fenómeno de manera global, y menos aun cuando se trabaja en equipo en un mismo proyecto. Así pues, por razones de eficiencia, sistematicidad y coordinación, los investigadores en neología, han ido abandonando, poco a poco, el criterio psicolingüístico y han pasado a adoptar un criterio más objetivo: el criterio lexicográfico, beneficiándose al mismo tiempo de las ventajas que les ofrecía la microinformática (bases de datos, diccionarios en línea, acceso a texto digital, exploradores de corpus, etc.). Así, el desarrollo de programas de constitución y gestión de bases de datos, por ejemplo, han permitido eliminar las «fichas en papel». Han sido, sin embargo, el uso de herramientas –aisladas o integradas en plataformas– como los buscadores de neología, la elaboración de diccionarios electrónicos desplegados morfológicamente, el acceso a diccionarios digitales, y la posibilidad de acceder a grandes cantidades de textos digitales, los factores que han cambiado realmente el trabajo sistemático en neología, en concreto el trabajo de los observatorios de neología.

## 2. EL OBSERVATORIO DE NEOLOGÍA DE BARCELONA (OBNEO): METODOLOGÍA DE TRABAJO

En el marco de las lenguas neolatinas han surgido, en los últimos cuarenta años, observatorios de neología<sup>1</sup> con el objetivo de recoger las palabras nuevas

---

1 Los observatorios que estudian la neología de una lengua se gestan en los años setenta del siglo pasado en el ámbito francófono: en Francia, son pioneros los trabajos de Louis Guilbert y su equipo, y los observatorios de neología creados por Bernard Quemada y coordinados por Gabrielle Quemada; en Québec, la Oficina quebequesa de la lengua francesa (OQLF) abrió en 1973 un observatorio de préstamos del inglés.

Algunos ejemplos de observatorios nacidos a partir de estas primeras experiencias son: Centre de Terminologie et Néologie de la Université de l'exINALF (1987), Observatorio de neología de Barcelona (1988), Observatorio de Neología de la Universidad de Vigo (1996), Osservatorio de

que nacen y se usan en una lengua. Los datos almacenados por los observatorios muestran cómo evoluciona el léxico de una lengua, su comportamiento en contextos determinados o para variedades lingüísticas concretas.

El Observatorio de neología de Barcelona (OBNEO) nace como un grupo de investigación en 1988 en la Universidad de Barcelona, y desde 1994 se incorpora como proyecto de investigación en el Instituto Universitario de Lingüística Aplicada de la Universidad Pompeu Fabra.

Desde sus orígenes, el objetivo principal de OBNEO es analizar el fenómeno de la aparición de palabras nuevas o neologismos en el uso, tanto para el catalán como para el castellano. La especificidad de la investigación en neología que realiza OBNEO desde 1988 es el vaciado de textos publicados en medios de comunicación de gran difusión. En sus orígenes y hasta el año 2000, el objetivo de OBNEO fue exclusivamente la neología de la prensa diaria editada en Barcelona. A partir del 2000 se diversifican las fuentes de vaciado y se abren dos nuevas líneas de investigación: la neología oral (procedente principalmente de textos radiofónicos) y la neología procedente de textos de prensa no diaria, lingüísticamente poco controlados.

## 2.1. Concepto de neologismo

La noción de neologismo es ciertamente compleja. Por el carácter relacional del concepto nuevo, la novedad no existe en sí misma, sino solo por referencia a algo (Rey 1976; Cabré, et al. 2003); es necesario, por lo tanto, definir qué es un neologismo o cuando una unidad es un neologismo, hay que establecer también para quién una lexía es nueva y con relación a qué corpus.

El criterio de base con el que trabaja OBNEO para establecer la neologicidad de una palabra es lexicográfico: se considera neologismo cualquier palabra que no aparece en un corpus lexicográfico de exclusión (un conjunto de diccionarios –que pueden variar según las finalidades de la recopilación<sup>2</sup>– utilizados para determinar si las unidades son neológicas).

---

Neologia della Lingua Italiana (1998), Observatório de Neologismos Científicos e Técnicos de la Universidad de Sao Paulo (1998), Grupo de neología de la Universidad de Málaga (2003), el Observatoire de la néologie du français contemporain de Paris<sup>7</sup> (2004), etc.

2 El establecimiento del corpus de exclusión determina los neologismos que resultan del vaciado de textos; si los diccionarios seleccionados son obras normativas, la cantidad de neologis-

Es necesario precisar que cuando el vaciado se hace totalmente manual, antes de la aplicación del criterio lexicográfico, prevalece, obviamente, la intuición neológica de quien vacía un texto, porque, por un lado, el neólogo nunca tiene en mente un diccionario; y por otro, hay unidades nuevas (especialmente las formadas por sufijación o prefijación) que le pasan desapercibidas a causa de su alto grado de predictibilidad formal o semántica.

A pesar de la eficacia del criterio lexicográfico, teniendo en cuenta las características y los objetivos del proyecto –básicamente: la necesidad de sistematización y objetividad cuando se trabaja en equipo y la utilidad de los resultados para diversas aplicaciones lexicográficas–, es evidente que utilizar exclusivamente este criterio tiene unos costes, como ya señaló hace unos años Corbeil (1971), pues se recogen como neologismos unidades que representan poca novedad o ninguna, porque están perfectamente instaladas en la lengua, pero que por razones diversas no aparecen en los diccionarios de referencia.

## 2.2. Detección y almacenamiento de los neologismos

En OBNEO el vaciado de los neologismos de prensa hasta el 1994 se realizó de manera totalmente manual: se leían los textos y, si se *sospechaba* que una palabra podía ser nueva, se consultaban en los diccionarios para confirmar la neologicidad o no de la unidad. Los neologismos seleccionados manualmente se registraban en una base de datos.

En los años noventa se empezó a extender la microinformática, surgieron las ediciones digitales de los periódicos y, con ello, los investigadores pudieron disponer corpus textuales informatizados y explorarlos. No es de extrañar que en este contexto nuevo de trabajo, en el marco de OBNEO, se creara en 1995 un detector semiautomático de neologismos SEXTAN (Vivaldi, 2000). SEXTAN es un programa que permite detectar en un texto digitalizado palabras nuevas, exclusivamente las formadas por recursos formales y los préstamos, con relación a un diccionario de exclusión. SEXTAN funciona con un diccionario desplegado morfológicamente: compara cada una de las ocurrencias del texto –previamente

---

mos que resultará del vaciado será muy superior que si se consideran obras descriptivas, generales y especializadas, con una nomenclatura más amplia.

etiquetado y desambiguado morfológicamente— con el diccionario y si detecta que una unidad del texto no está en su diccionario, la propone como neológica. El programa SEXTAN ofrece tantas fichas como ocurrencias de una misma unidad léxica para que el usuario seleccione la que le parezca más adecuada. SEXTAN recupera automáticamente del texto, para cada unidad seleccionada, su categoría gramatical, contexto, fecha, fuente, página y sección.

Aunque actualmente ya no se usa, OBNEO trató con SEXTAN la mayor parte de los textos de prensa desde 1995 hasta 2006, lo que permitió extraer de manera sistemática y exhaustiva miles de neologismos formales de los textos analizados. Para paliar sus insuficiencias por cuanto con este sistema quedaban sin detectar la totalidad de los neologismos semánticos, los sintagmáticos, los sintácticos y los que son el resultado de una conversión o de una lexicalización, OBNEO combinó el vaciado manual con el vaciado semiautomático con SEXTAN.

A pesar de su utilidad, el uso de SEXTAN suponía una serie de limitaciones, entre las que destacamos las siguientes:

- Se debía aplicar en un ordenador local y, por lo tanto, no se podía trabajar simultáneamente en equipo ni en línea.
- Funcionaba sólo con texto pretratado morfológicamente y, por consiguiente, para utilizarlo el equipo de trabajo necesitaba disponer de un analizador y un desambiguador morfológicos.
- Dependía excesivamente de la configuración formal de un periódico, de manera que cuando los periódicos cambiaban los formatos —que era muy a menudo— el programa debía adaptarse específicamente.
- Era dependiente de lengua.

Teniendo en cuenta estas limitaciones y con el fin de potenciar el trabajo autónomo, colectivo y en línea, en 2005 el equipo de OBNEO diseñó una estación de trabajo integrada en línea (Estación OBNEO). Esta estación de trabajo neológico permite la detección, selección, almacenamiento y explotación de la neología de manera simultánea a través d'Internet.

### 3. LA ESTACIÓN OBNEO<sup>3</sup>

Este entorno de trabajo integrado en línea ha supuesto para el neólogo, y para OBNEO, un paso adelante en la investigación en neología, pues ha permitido reunir en un mismo entorno digital diferentes herramientas básicas de la cadena de trabajo, facilitando enormemente la recopilación, el almacenamiento, el control y la explotación de los datos y ha favorecido además el trabajo en red entre investigadores y entre grupos de investigación que se encuentran físicamente separados.

Las ventajas de trabajar con un estación en red se pueden resumir en los siguientes puntos:

- Potencia el trabajo en equipo con una metodología compartida
- Hace posible el intercambio de datos en línea de manera rápida
- Permite utilizar fuentes de exclusión distintas en cada momento, según las necesidades de cada investigación
- Permite extraer listados de neologismos con búsquedas filtradas cruzando todos los campos de información
- Facilita las investigaciones con datos contrastivos
- Hace posible el trabajo intergrupar
- Aumenta la autonomía del investigador.

La Estación OBNEO está concebida como un entorno modular que incluye las siguientes herramientas:

- PLATAFORMA OBNEO: herramienta principal del entorno que gestiona y configura todas las herramientas y las fichas de neologismos (<http://obneo.iula.upf.edu/pobneo/index.php>).

---

3 Este proyecto ha sido financiado por tres proyectos de investigación dirigidos por M. Teresa Cabré: *Cooperació en Neologia Catalana: desenvolupament d'una plataforma de treball en xarxa via Internet* (IEC: 2005-2007), *Desenvolupament d'una plataforma de treball en xarxa via Internet en neologia catalana i millora de les estratègies de detecció automàtica de neologismes* (Acció complementari del MEC: JHUM2004-20159-E, 2005-2006), y *Desenvolupament d'una plataforma de treball en xarxa via Internet en neologia catalana i millora de les estratègies de detecció automàtica de neologismes* (DURSI, 2005-2006).

- BUSCANEEO: detector y extractor de neologismos (<http://obneo.iula.upf.edu/buscaneo/>).
- GESTODIC OBNEO: gestor que permite incorporar y gestionar diversos diccionarios que pueden activarse y desactivarse como corpus de exclusión (<http://obneo.iula.upf.edu/gesdic/>).
- BUSCADOR BOBNEO: interfaz de acceso libre para todos los usuarios de Internet que permite realizar búsquedas sobre las fichas de los neologismos validados, es decir sobre el banco de neologismos del OBNEO (BOBNEO) (<http://obneo.iula.upf.edu/bobneo/index.php>).

Estas herramientas permiten al neólogo:

- Configurar sus corpus digitales a partir de Internet
- Buscar automáticamente candidatos a neologismos formales
- Gestionar las fichas de neologismos
- Crear distintos corpus de exclusión
- Gestionar diccionarios de exclusión
- Vaciar textos con diferentes corpus de exclusión
- Crear nuevos diccionarios de exclusión ad hoc
- Autoconfigurar los campos y los valores de las fichas
- Consultar diccionarios digitales en línea
- Extraer estadísticas sobre los datos neológicos
- Consultar los datos almacenados
- Imprimir, importar y exportar fichas de neologismos en formatos diversos
- Llevar un control de los usuarios
- Preservar los datos de cada equipo
- Ayudar al neólogo en la toma de decisiones a través de menús de ayuda.

En realidad la Estación OBNEO está formada por dos grandes tipos de aplicaciones diferentes:



- Aplicaciones de gestión de datos: aplicaciones encargadas de gestionar toda la información del entorno. En este grupo se incluyen las Plataformas OBNEO, el BUSCANEEO y el GESTODIC OBNEO.
- Aplicaciones de consulta de fichas: aplicaciones encargadas de ofrecer una interfaz de consulta de fichas. Pertenecen a este grupo los Buscadores ONBNEO de acceso abierto<sup>4</sup>.

Las herramientas que integran la Estación OBNEO trabajan sobre una misma base de datos, es decir, se comportan como una serie de interfaces que dan acceso a los datos de maneras diversas. Así, dada la gran cantidad de puntos en común que tienen las aplicaciones de la Estación OBNEO, lo más adecuado para su desarrollo fue una arquitectura en capas que facilita su creación, mantenimiento y ampliación<sup>5</sup>, pues separa la implementación de los diferentes niveles de funcionalidad, de manera que un cambio en una capa tiene el mínimo efecto en el resto. Además, el tipo de desarrollo seleccionado facilita la creación de nuevas aplicaciones que trabajen sobre los mismos datos, pues solo es necesario crear una interfaz adecuada y ocasionalmente añadir nuevas funcionalidades en la capa de dominio.

Para implementar la Estación OBNEO se utilizaron cuatro tecnologías básicas: HTML, DTML (Javascript + CSS), PHP i Perl<sup>6</sup>; aunque en su interfaz, las herramientas funcionan como aplicaciones web basadas en tecnología HTML.

### **3.1. La Plataforma OBNEO**

La Plataforma OBNEO es, como hemos dicho, la herramienta central de la estación y la que permite:

---

4 Y también los Buzones neológicos que permiten a un usuario externo enviar al Observatorio neologismos. Buzón neológico: <http://obneo.iula.upf.edu/bustia/index.php> i Buzón neológico escolar: <http://obneo.iula.upf.edu/bustiaInfantil/index.php>

5 Laura López Oros es la ingeniera responsable de la construcción y mantenimiento del Entorno OBNEO.

6 Para una descripción técnica del proyecto puede consultar Cabré, Estopà, López WP (en prensa).

- Configurar los usuarios, los grupos de trabajo, los datos de soporte y los idiomas de la interfaz.
- Gestionar los datos de los neologismos.
- Almacenar las fichas de neologismos (es decir el banco de datos de neologismos).

Con relación a la gestión de usuarios, la Plataforma OBNEO prevé diferentes tipos de perfil de usuario para realizar las tareas de gestión de la neología:

- Usuarios externos no registrados
- Usuarios externos registrados
- Colaboradores
- Correctores
- Validadores
- Gestores
- Gestor-administrador.

La Plataforma OBNEO gestiona tanto usuarios individuales de la Estación, como grupos de usuarios. Así mismo, los usuarios gestores pueden gestionar los datos básicos de todos los campos de las fichas; de manera que pueden ampliar, modificar, suprimir las características de cada campo. Por ejemplo, pueden añadir categorías gramaticales, fuentes de vaciado, regiones, notas, etc. Además cada usuario conectado puede gestionar sus datos personales, modificar su contraseña y el idioma que por defecto usará la plataforma. Actualmente, la interfaz de usuario de la Plataforma se ha desarrollado en catalán y castellano, sin menos cabo que pueda desarrollarse en otras lenguas.

En cuanto a la gestión de datos neológicos, la Plataforma OBNEO ofrece actualmente cuatro tipos de ficha según el texto del cual procede el neologismo:

- Neologismos procedentes de textos de prensa diaria
- Neologismos procedentes de textos de prensa no diaria
- Neologismos procedentes de textos orales
- Neologismos procedentes de textos escolares.

El sistema prevé para cada ficha cuatro situaciones, que aseguran el control, a distintos niveles, sobre la información: incompleta, completa, corregida, validada.

En cuanto a las funcionalidades de la gestión de las fichas neológicas, la Plataforma contempla las siguientes:

- Edición
- Consulta
- Impresión
- Importación
- Exportación
- Extracción de estadísticas.

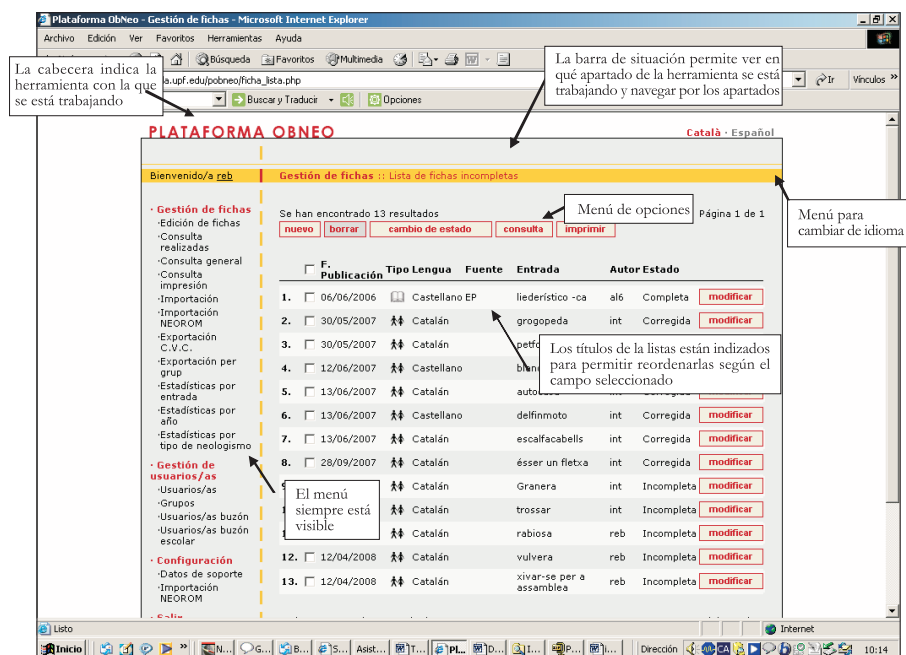


Figura 1. Captura de pantalla de la página principal de la Plataforma OBNEO

### 3.1.1. Edición de fichas

El módulo de edición de fichas permite realizar las siguientes tareas:

- Dar de alta una nueva ficha
- Modificar una o un conjunto de fichas
- Borrar una o un conjunto de fichas
- Cambiar de estado una o un conjunto de fichas
- Consultar una o un conjunto de fichas.

Este módulo incluye:

- a) Mecanismos de control para asegurar la sistematicidad y calidad en forma de:
  - Menús cerrados (por ejemplo, junto al campo de código de nota se encuentra un botón que abre una ventana con la lista de todos los códigos disponibles que se pueden seleccionar y que, una vez seleccionados, se añaden automáticamente en el campo de nota y de código de nota),
  - Campos obligatorios que no permiten avanzar si previamente no se han rellenado
- b) Remisiones a la metodología de trabajo, que facilitan el trabajo al neólogo, a través de menús de ayuda (por ejemplo, junto al campo tipo de neologismo se ofrece la opción de abrir una ventana donde se incluye una descripción detallada de cada tipo de neologismo existente con ejemplos concretos)
- c) Acceso a recursos que ayudan a tomar decisiones en la selección y compleción de los datos, como:
  - Diccionarios digitales (al lado del campo entrada se puede acceder al corpus de exclusión lexicográfico)
  - Extracción de estadísticas (junto al campo de tipo de neologismo se ofrece el acceso a cómo ha sido clasificado un neologismo en el Banco de neologismo del Observatorio).

### 3.1.2. Consulta e impresión de fichas

A través del módulo de consulta e impresión, la Plataforma OBNEO permite acceder a todas las fichas validadas del banco de datos (consulta general) o a las fichas que cada usuario ha creado o modificado. La búsqueda de información es muy flexible: se pueden realizar búsquedas sobre todos los campos de la ficha neológica y usar condiciones cruzadas en todos los campos.

Los resultados se pueden mostrar por orden alfabético o de frecuencia y se pueden imprimir en formatos estándares preestablecidos o diseñar un formato a través de un menú que permite seleccionar los campos que se quieren imprimir de cada neologismo, o exportar directamente los datos a un fichero .txt.

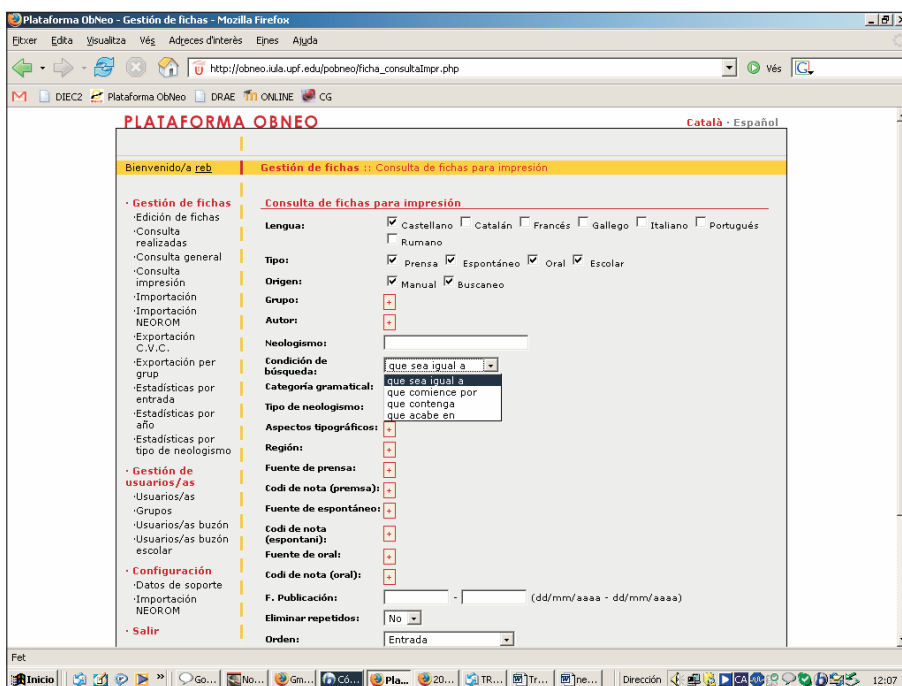


Figura 2. Captura de pantalla del módulo de consulta de fichas para imprimir

### *3.1.3. Importación y exportación de fichas*

La importación de fichas permite descargar un fichero en el formato actual de trabajo, así, pues, se pueden recuperar fichas de neologismos de otros formatos en los que antiguamente trabajaba OBNEO o en los que pueden trabajar otros grupos de investigación.

Mediante la exportación de fichas, se puede reunir un determinado grupo de registros en un fichero zip que el usuario puede guardar en un disco.

### *3.1.4. Extracción de estadísticas*

Actualmente la Plataforma OBNEO realiza distintos cálculos estadísticos sobre los neologismos:

- Cálculo por entrada: el sistema muestra el número de veces que se ha encontrado un neologismo, cómo ha sido clasificado y su relación porcentual con el resto de entradas.
- Cálculo por año: el sistema muestra una lista, ordenada alfabéticamente, de neologismos de las fichas validadas que cumplan los criterios de búsqueda indicados correspondientes a un período. Para cada neologismo, muestra la categoría gramatical, el número de fichas encontradas y el porcentaje respecto al total de fichas encontradas.
- Cálculo por tipo: el sistema genera una lista de los tipos de neologismos con el número y el porcentaje de fichas encontradas de cada tipo respecto al total de fichas encontradas en un período indicado previamente.

## **3.2. BUSCANEO**

BUSCANEO es una herramienta web de detección de neologismos formada por dos módulos:

- Gestor de diccionarios
- Buscador de neologismos

BUSCANEO, a diferencia de SEXTAN, trabaja con textos no marcados lingüísticamente. El texto de entrada se puede obtener a partir de un fichero local, pero, sobre todo, está pensado para trabajar directamente con textos de páginas web, especialmente con los procedentes de medios de comunicación escritos.

Este buscador de neologismos permite analizar sitios web y extraer de sus contenidos candidatos a neologismo. La herramienta posee tres funcionalidades:

- Gestión de sitios web
- Busca de neologismo
- Gestión de candidatos.

### *3.2.1. Gestión de sitios web*

Este módulo permite establecer:

- La periodicidad de búsqueda de un sitio web
- La lista de diccionarios que se usará para saber si las palabras encontradas son o no candidatas a neologismos
- La lista de direcciones prohibidas
- Los filtros para analizar o no una determinada página web.

### *3.2.2. Búsqueda de neologismos*

La búsqueda de neologismos es la funcionalidad central de la herramienta. Se encarga de revisar automáticamente los sitios web programados y para cada uno de ellos analiza la página indicada como página inicial y después analiza sucesivamente las páginas según su configuración. Para cada página se extrae la información siguiente:

- La lista de enlaces referenciados en la página<sup>7</sup>.

---

<sup>7</sup> Es importante notar que se considera que una dirección corresponde a una página que forma parte del sitio web si: a) la dirección es relativa, b) si la dirección es absoluta pero cumple al menos una de las dos condiciones siguientes: 1) su path coincide o pertenece al path

- La lista de candidatos a neologismo encontrados dentro del texto de la página<sup>8</sup>, que se guardará en la base de datos, almacenando para cada uno de ellos el sitio web donde se ha encontrado y la lista de contextos en los que aparece.

### 3.3.3. Gestión de candidatos

Este módulo permite al usuario gestionar los candidatos procedentes de las búsquedas en los sitios web. Para ello, dispone de un buscador de candidatos por fecha y sitio web que le ofrece una lista paginada y ordenada alfabéticamente de los candidatos. Cada candidato se presenta acompañado de la lengua, el nombre del sitio web y la fecha de publicación, y un enlace con los contextos. A partir de esta información el usuario tiene dos posibilidades: rechazar el candidato o aceptarlo. Si lo rechaza, se borra de la base de datos de candidatos. Si, por el contrario, lo acepta, el sistema pasa a una página en la que se muestra el formulario para dar de alta el candidato como neologismo en la Plataforma OBNEO. Una vez se han completado las informaciones con relación al neologismo (categoría gramatical, tipo de recurso de formación, marcas pragmáticas, etc.), éste queda incorporado en la Plataforma OBNEO y se elimina de BUSCANEO.

A continuación mostramos una de las pantallas del BUSCANEO, la que permite aceptar o rechazar un candidato a neologismo:

---

de la dirección principal del sitio web o de sus direcciones alternativas; 2) o su path no coincide ni pertenece a un subdirectorio del path de ninguna de las direcciones prohibidas del sitio web. Así, una dirección solo se analizará si pasa todos los filtros definidos en la configuración. BUSCANEO está configurado para que durante el proceso de análisis del sitio web analice como máximo 100 páginas. Esta cifra, pero, puede modificarse si se desea.

8 Para buscar los candidatos el BUSCANEO extrae todas las palabras contenidas en el texto de la página separadas por signos de puntuación o espacios. Se descartan automáticamente las direcciones de páginas web y de correo electrónico. Del resto de palabras candidatas a neologismo, se descartan las que cumplen alguna de las siguientes condiciones: contiene alguna mayúscula, empieza por un dígito, tiene una longitud inferior a 2, contiene más de tres letras iguales seguidas, esta recogida en uno de los diccionarios indicados en la configuración del sitio web, está dentro de un fragmento en otro idioma.



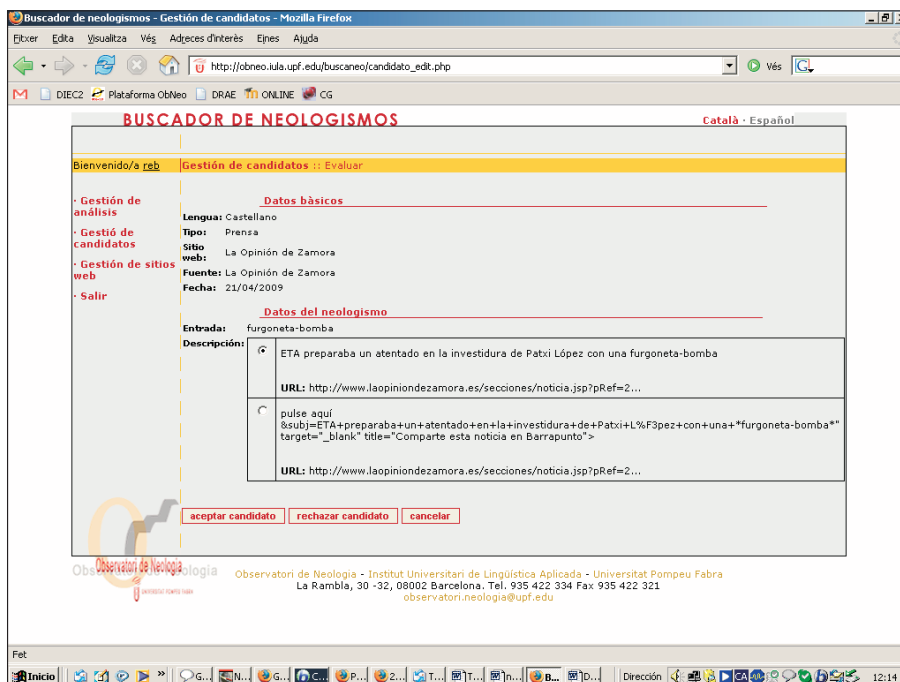


Figura 3. Captura de pantalla de BUSCANEO

Parece obvio decir que BUSCANEO supone un avance en relación con SEXTAN. Los hitos que la herramienta ha alcanzado son los siguientes:

- Permite trabajar con textos no etiquetados morfológicamente y, por tanto, es una herramienta independiente de lengua.
- Opera a través de Internet, por lo que se pueden procesar los textos que se difunden directamente a través de la red.
- Es independiente del formato de presentación –muy mutante– de las páginas web.
- Permite el uso de distintos diccionarios, con lo que se constata en todo momento que una unidad es neológica en relación a una(s) fuente(s) lexicográfica(s) concreta(s).

- Ofrece un 100% de cobertura en la detección de neología formal y de préstamos.
- Integra los resultados directamente en la Plataforma de OBNEO.

A pesar de estos avances en relación sobre todo a la autonomía del neólogo y a la integración de los datos, BUSCANEEO, hoy por hoy, aun no permite detectar neologismos semánticos, neologismos procedentes de cambios de categoría flexiva ni neologismos sintagmáticos, que hay que seguir detectando manualmente.

### **3.3. GESTODIC OBNEO**

El Gestor de diccionarios de OBNEO está formado por distintos diccionarios desplegados morfológicamente. Este módulo es abierto y flexible. Abierto, porque permite introducir nuevos diccionarios; flexible, porque, antes de iniciar una sesión de detección de neologismos, ofrece la opción de seleccionar los distintos diccionarios que se pueden utilizar.

La herramienta incluye tres módulos:

- Gestión de diccionarios, desde donde se administran los diccionarios
- Gestión de tipos flexivos, que permite asignar a cada lema un tipo de flexión. Las reglas de flexión que se utilizan en esta aplicación se basan en las reglas definidas en el proyecto PAROLE
- Gestión de lemas, que lista los lemas existentes en un diccionario.

### **3.4. Buscador BOBNEO**

El Buscador BOBNEO, de acceso totalmente libre, permite a cualquier usuario conectado a Internet realizar consultas sobre las fichas validadas almacenadas en el banco de neologismos del Observatorio de Neología (BOBNEO). El usuario después de seleccionar el idioma de los neologismos sobre los que quiere hacer la consulta, debe expresar una condición de búsqueda (simple o avanzada). Con las opciones avanzadas, puede restringir o precisar más su búsqueda y acceder a muchas más informaciones a través de una selección por el corpus de vaciado, la categoría gramatical, el tipo de neologismo, los aspectos

tipográficos, etc. Como resultado de búsqueda, se obtiene una lista paginada de las fichas validadas que cumplen los criterios de búsqueda seleccionados, como se puede observar en la siguiente figura sobre la búsqueda de los neologismos que empiezan por la forma *freak*:

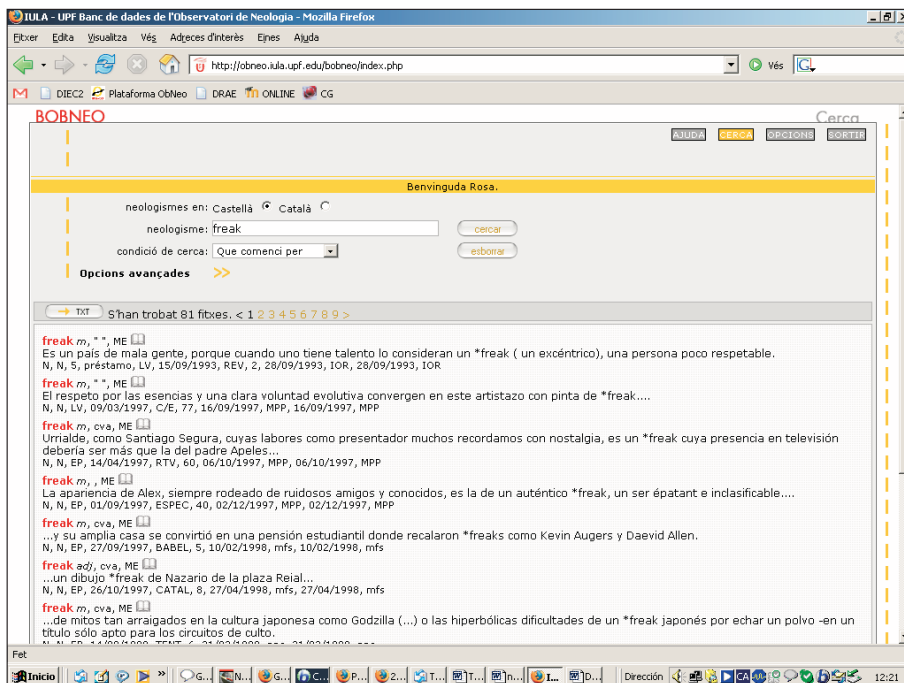


Figura 4. Captura de pantalla de la búsqueda del neologismo *freak* en BOBNEO

#### 4. VALORACIÓN DEL ENTORNO OBNEO Y NUEVOS RETOS PARA EL FUTURO

No cabe duda de que el trabajo neológico ha avanzado con la integración de las herramientas en una estación de trabajo en línea. El primer paso es la posibilidad de consultar a través de Internet el Banco de neologismos del Observatorio BOBNEO. El segundo hito alcanzado es la búsqueda automática

de candidatos a neologismos formales y la integración de los neologismos a la Plataforma; BUSCANEO ha supuesto para el neólogo ganar en autonomía, agilidad, en tiempo y en sistematicidad. Finalmente, la Plataforma OBNEO ha permitido el paso más cualitativo con relación al trabajo del investigador pues permite trabajar modularmente, virtualmente y en equipo. La Plataforma OBNEO además ofrece al neólogo múltiples recursos y ayudas que le facilitan el trabajo de vaciado, análisis y estudio de los neologismos.

Con este proceso integrado de trabajo el OBNEO ha mejorado la situación y posibilidades de trabajo en los siguientes aspectos:

- Autonomía
- Agilidad
- Copoperación en los aspectos metodológicos
- Intercambio de datos
- Teletrabajo
- Diversidad de corpus de exclusión
- Sistematicidad
- Ahorro de tiempo
- Seguridad
- Flexibilidad

Aún así hay que seguir trabajando en el ámbito tecnológico, principalmente para alcanzar las siguientes operaciones:

- Incorporación de nuevos cálculos estadísticos
- Utilización de diferentes corpus de exclusión lexicográficos (especializados) y no lexicográficos
- Detección de neologismos formados por resemantización o sintagmación
- Tratamiento automático de vaciado neológico de textos orales.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- CORBEIL, Jean-Claude (1971): «Aspects du problème néologique». *La banque des mots*, 2, 123-136.
- CABRÉ, M. Teresa *et al.* (2000): *La neologia en el tombant de segle*. Barcelona: Observatori de Neologia, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. (Sèrie Activitats, 5).
- CABRÉ, M. Teresa *et al.* (2002): «La Lexicografia i la identificació automatitzada de neologia lèxica» en Paz Battaner; Janet de Cesaris (ed.) *De lexicografia. Actes del I Symposium Internacional de Lexicografia (Barcelona, 16-18 de maig de 2002)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 287-294.
- CABRÉ, M. Teresa *et al.* (2003): «L'observatoire de néologie: conception, méthodologie, résultats et nouveaux travaux» en Sablayrolles, J. (ed.) *L'innovation lexicale*. Paris: Honoré Champion Éditeur.
- ESTOPÀ, Rosa (2008): «L'observation de mots nouveaux en espagnol et en catalan au cours de deux décennies : l'Observatori de Neologia de l'IULA». *Neologica*, 2, 169-184.
- OBSERVATORI DE NEOLOGIA (2004a): *Metodología del trabajo en neología: criterios, materiales y procesos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. [Versión digital: <http://www.iula.upf.edu/repositori/04mon009.pdf>].
- OBSERVATORI DE NEOLOGIA (2004b): *Llengua catalana i neologia*. Barcelona: Meteora.
- REY, Allain (1976): «Néologisme : un pseudo-concept ?». *Cahiers de lexicologie*, 28, 3-17.
- SABLAYROLLES, Jean-François (2000): *La néologie en français contemporain. Examen du concept et analyse de productions néologiques récentes*. Paris: Honoré Champion.
- VALLÈS, Teresa (2004): *La creativitat lèxica en un model basat en l'ús: una aproximació cognitiva a la neologia i a la productivitat*. Barcelona: Publicacions de L'Abadia de Montserrat.
- VIVALDI, Jorge (2000): «SEXTAN: prototip d'un sistema d'extracció de neologismes» en M. Teresa Cabré *et al.* (ed.) *La neologia en el tombant de segle*.

Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 165-173.