

Reutilización de datos lingüísticos para la creación de un sistema de traducción automática para un nuevo par de lenguas

Re-use of linguistic data to create a machine translation system for a new language pair

Carme Armentano-Oller
DLSI, Universitat d'Alacant
E-03071 Alacant
carmentano@dlsi.ua.es

Mikel L. Forcada
DLSI, Universitat d'Alacant
E-03071 Alacant
mlf@dlsi.ua.es

Resumen: Este trabajo estudia varias formas de reutilizar datos lingüísticos ya desarrollados para obtener rápidamente un sistema de traducción automática para un nuevo par de lenguas. En particular, se ha desarrollado un traductor entre el portugués y el catalán basado en la plataforma Apertium (www.apertium.org), a partir de los datos ya disponibles en esta plataforma para traducir entre portugués y español y entre español y catalán. Los resultados obtenidos indican que una simple composición de dos traductores completos es una buena primera opción, aunque también se muestran otros resultados muy interesantes obtenidos en poco tiempo usando las herramientas que proporciona esta plataforma.

Palabras clave: reutilización datos lingüísticos, traducción automática, Apertium

Abstract: This work examines various ways to re-use pre-existing linguistic data to quickly generate a machine translation system for a new language pair. In particular, a machine translation system between Portuguese and Catalan based on the Apertium platform (www.apertium.org) has been built from data existing in this platform for translating between Portuguese and Spanish and between Spanish and Catalan. The results obtained indicate that a simple composition of two complete translators is an adequate first option, but other very interesting results are shown which have been obtained in short time using the tools provided in the Apertium platform.

Keywords: re-use of linguistic data, machine translation, Apertium

1. Presentación

La traducción automática está ganando cada vez más importancia en nuestra sociedad; en las últimas décadas han ido apareciendo sistemas para un número creciente de pares de lenguas, que sin embargo, dejan aún muchos pares sin cubrir, especialmente cuando una de ellas es una lengua menor (Forcada, 2006). En este trabajo se han querido estudiar varias formas de reutilizar datos lingüísticos ya desarrollados en traductores automáticos para obtener rápidamente otro traductor para un nuevo par de lenguas. En particular, se han utilizado para traducir entre la lengua *A* y la lengua *C* los datos lingüísticos ya disponibles para traducir entre dos pares de lenguas con una lengua común *B* (el par *A-B* y el par *B-C*). Concretamente, se han utilizado los datos lingüísticos disponibles en la plataforma Apertium (www.apertium.org)

para los traductores español↔catalán (`apertium-es-ca`, versión 1.0) y español↔portugués (`apertium-es-pt`, versión 0.8) para construir un traductor portugués↔catalán (`apertium-pt-ca`).¹

A continuación se presenta la plataforma de traducción automática Apertium, se describe cómo funciona y se explica por qué se ha escogido para este trabajo. En la sección 3 se presentan los datos lingüísticos que se han utilizado y se explica cómo se han combinado en las diferentes configuraciones usadas para construir el traductor portugués↔catalán. Finalmente, en la sección 4 se presentan los resultados obtenidos, las conclusiones y se señalan algunas de las líneas que quedan abiertas.

¹En los paquetes Apertium se usan los códigos ISO 639-1 para designar los pares de lenguas. Así, se usa *ca* para catalán, *es* para español y *pt* para portugués.

2. *Apertium*

Apertium es una plataforma de traducción automática de código abierto. La plataforma incluye el motor de traducción, datos lingüísticos para varios de pares de lenguas y herramientas para desarrollar nuevos pares. Los programas, las herramientas y los datos lingüísticos se distribuyen con licencia GNU General Public License, versión 2.²

2.1. Cómo funciona Apertium

Apertium sigue un modelo clásico de traducción por transferencia basada en reglas (Hutchins y Somers, 1992, apart. 4.2), organizada en una serie de módulos dispuestos en cascada. Los módulos se comunican entre ellos mediante texto. Existen dos niveles del motor de traducción en Apertium: en el primer nivel la transferencia estructural se hace en un solo módulo, mientras que en el segundo nivel, para lenguas más distantes entre sí, la transferencia estructural se hace en tres etapas. Para desarrollar el traductor *apertium-pt-ca* se ha utilizado el motor de primer nivel, dado que la proximidad estructural entre las dos lenguas así lo permitía y también porque los dos traductores que se han utilizado como punto de partida también están desarrollados en el primer nivel.

Como se describe en detalle en (Armentano-Oller et al., 2006), los módulos lingüísticos que usa la primera versión de Apertium son los que se presentan a continuación (*A* es la lengua origen —en los ejemplos, español— y *B* es la lengua meta —en los ejemplos, catalán):

- **Analizador morfológico:** Analiza morfológicamente el texto a traducir a partir de la información del diccionario morfológico monolingüe que corresponda (*apertium-A-B.A.dix*). Para cada forma superficial del texto da todos los análisis o formas posibles. Así pues, ante la entrada “de la nube roja”, la salida de este módulo sería:

```

^de/de<pr>$
^la/el<det><def><f><sg>
  /lo<prn><pro><p3><f><sg>$
^nube/nube<n><f><sg>$
^roja/rojo<adj><f><sg>$

```

- **Desambiguador léxico categorial:** Basándose en modelos ocultos de Markov elige el análisis más probable para cada palabra ambigua, según su contexto. Siguiendo con el ejemplo, la salida del desambiguador sería:

```

^de<pr>$
^el<det><def><f><sg>$
^nube<n><f><sg>$
^rojo<adj><f><sg>$

```

- **Módulo de transferencia léxica:** A partir de la información del diccionario bilingüe *apertium-A-B.A-B.dix*, lee cada una de las formas léxicas y entrega su equivalente en la lengua meta. En el ejemplo, estas equivalencias serían:

```

de<pr> → de<pr>
el<det> → el<det>
nube<n><f> → núvol<n><m>
rojo<adj> → vermell<adj>

```

- **Módulo de transferencia estructural:** Trata los cambios estructurales entre las dos lenguas a traducir (cambios de género y número, reordenamientos, etc.). El fichero *apertium-A-B.trules-A-B.xml* contiene las reglas necesarias. En cada regla se detecta un patrón (una secuencia de formas léxicas) al que se le aplican los cambios y para el cual se genera una salida. Los patrones se suelen definir como secuencias de categorías gramaticales, pero se pueden detectar también lemas concretos. También en la salida se pueden generar lemas concretos, independientemente del diccionario bilingüe. En el caso del ejemplo, al traducir al catalán ha cambiado el género del sustantivo, por lo que hay que cambiar también el de los elementos que concuerdan con él, aplicando una regla con patrón *det-n-adj*. El resultado es:

```

^de<pr>$
^el<det><def><m><sg>$
^núvol<n><m><sg>$
^vermell<adj><m><sg>$

```

- **Generador morfológico:** A partir de las formas léxicas de la lengua meta genera las formas superficiales flexionadas

²<http://www.gnu.org/licenses/old-licenses/gpl-2.0.html>

correspondientes, utilizando el diccionario morfológico `apertium-A-B.B.dix`. Siguiendo con el mismo ejemplo, la salida de este módulo sería: `~de ~el núvol vermell`.³

- **Posgenerador:** Aplica algunas operaciones ortográficas de la lengua meta, tales como apostrofaciones y contracciones. Para cada lengua hay un fichero (`apertium-A-B.post-A.dix`) con las reglas que tiene que aplicar este módulo. En el caso del ejemplo hay que hacer la contracción entre la preposición *de* y el determinante *el*. Así, la salida de este módulo sería: `del núvol vermell`

2.2. Por qué Apertium?

Se ha decidido utilizar Apertium para estudiar la reutilización de datos porque la licencia que utiliza lo permite legalmente sin tener que pedir permisos expresamente, y porque el motor y los datos lingüísticos son independientes, lo que facilita el trabajo de desarrollo lingüístico.

3. Reutilización de datos

Se han combinado de distintas formas los datos lingüísticos de los dos traductores de partida, para poder estudiar qué ventajas e inconvenientes presentaba cada posible configuración (véase 3.3). También se han tenido que modificar datos lingüísticos y crear algunos nuevos, tanto de forma automática (véase 3.1) como manualmente.

La calidad de las traducciones se ha avaluado con textos de 10.000 palabras de noticias periodísticas publicadas en Internet,⁴ usando el programa `apertium-eval-translator`.⁵

3.1. El programa `apertium-dixtools`

El programa `apertium-dixtools`⁶ lleva a cabo distintas operaciones con los diccionarios de Apertium, una de las cuales es el *cruce* de diccionarios: a partir de los diccionarios de dos traductores `apertium-A-B` y

`apertium-B-C` genera los siguientes diccionarios:

- `apertium-A-C.A-C-crossed.dix`: Diccionario bilingüe *A-C* con los lemas comunes de los dos traductores de partida.
- `apertium-A-C.A-crossed.dix` y `apertium-A-C.C-crossed.dix`: Diccionarios monolingües de las lenguas *A* y *C* consistentes con el diccionario bilingüe generado, eso es, sólo con los lemas que se encuentran en el diccionario *A-C*.
- `consistent-bil-A-B.dix` y `consistent-bil-B-C.dix`: Diccionarios bilingües *A-B* y *B-C* sólo con los lemas comunes de los dos traductores de partida.

El programa `apertium-dixtools` es una herramienta muy útil para cruzar diccionarios, pero el cruce presenta aún algunos problemas:

- **Pérdida de cobertura:** Hay que considerar que los diccionarios generados contienen solo las entradas comunes (mismo lema con la misma categoría gramatical) de los dos traductores de partida, por lo que inevitablemente serán más reducidos. Además, a la hora de desarrollar los diccionarios de los traductores de partida no siempre se han seguido los mismos criterios lingüísticos, por lo que algunos lemas sí están en ambos traductores pero tienen asignadas distintas categorías gramaticales (por ejemplo, en un diccionario tienen la categoría *nombre* y en el otro, *adjetivo*): en estos casos los lemas no se cruzan.
- **Pérdida de unidades multipalabra:** En los diccionarios de Apertium hay también *unidades léxicas multipalabra* que no se han de traducir palabra por palabra, como *echar de menos*. Una parte de estas expresiones sí están en los diccionarios de uno de los traductores pero no en los del otro, sea porque no se han introducido o porque en el otro par de lenguas son innecesarias. En ambos casos, la expresión no se encontrará en los diccionarios cruzados generados por `apertium-dixtools`.

³El signo “~” marca palabras que pueden requerir cambios en la siguiente etapa (posgenerador).

⁴Textos accesibles en <http://xixona.dlsi.ua.es/~carmentano/avaluacions.html>

⁵<http://www.apertium.org/?id=apertium-eval-translator>

⁶Escrito en Java por Enrique Benimeli (ebenimeli@gmail.com), véase <http://xixona.dlsi.ua.es/wiki/Apertium-dixtools>

- **Errores de consistencia:** Para que el traductor funcione correctamente hace falta que los diccionarios sean *consistentes*, eso es, que todas las formas que se encuentran en cada uno de los diccionarios morfológicos tengan traducción en el diccionario bilingüe y su equivalente en el otro diccionario morfológico. En los diccionarios generados por `apertium-dixtools` hay dos tipos de errores de consistencia: los generados por el programa y los causados por los distintos criterios lingüísticos a la hora de desarrollar los diccionarios morfológicos de los traductores de partida (por ejemplo, los diccionarios del traductor `apertium-es-ca` tienen definidos los diminutivos de los adjetivos, cosa que no ocurre en los diccionarios del traductor `apertium-es-pt`).
- **Pérdida de secciones del diccionario:** En los diccionarios morfológicos los lemas se encuentran en *secciones* distintas, según cómo se deban segmentar sus formas superficiales en los textos. Actualmente, el programa `apertium-dixtools` no cruza las entradas que se encuentran en la sección denominada *postblank*, que contiene las entradas cuyas formas superficiales no van seguidas de un espacio en blanco (en el caso del catalán, las palabras apostrofas como *d'* o *s'*), pero que deben separarse de la siguiente con un blanco para su tratamiento posterior.

3.2. Datos utilizados

Los datos lingüísticos que se han utilizado son los que se describen a continuación:

- **Diccionarios:** Se han utilizado tres tipos de diccionarios:
 - diccionarios tomados de los traductores de partida
 - diccionarios `consistent-bil-pt-es.dix` y `consistent-bil-es-ca.dix`, generados automáticamente con `apertium-dixtools`
 - diccionarios `apertium-pt-ca.ca.dix`, `apertium-pt-ca.pt-ca.dix` y `apertium-pt-ca.pt.dix`, corregidos a mano a partir de

los diccionarios generados por `apertium-dixtools`

- **Probabilidades de desambiguación:** Se han copiado directamente de los traductores de partida.
- **Reglas de transferencia estructural:** Se han utilizado tres tipos de ficheros de definición de reglas de transferencia estructural:
 - ficheros copiados de los traductores de partida
 - ficheros `apertium-es-ca.trules-es-ca-m.xml`, `apertium-es-ca.trules-ca-es-m.xml`, `apertium-es-pt.trules-es-pt-m.xml`, y `apertium-es-pt.trules-pt-es-m.xml`, modificados a mano a partir de los de uno de los traductores de partida (véanse 3.3.3 y 3.3.4)
 - ficheros de reglas `apertium-pt-ca.trules-pt-ca.xml` y `apertium-pt-ca.trules-ca-pt.xml`, generados a mano (véase 3.3.5)
- **Diccionarios de postgeneración:** Se han copiado directamente de los traductores de partida.

3.3. Configuraciones estudiadas

Los datos lingüísticos de los traductores de partida se han combinado de las formas que se describen a continuación.

3.3.1. Composición de dos traductores completos

Consiste en traducir un texto del portugués al castellano con el traductor `apertium-pt-es`, y traducir el resultado al catalán con `apertium-es-ca`, o en sentido contrario para traducir del catalán al portugués.

3.3.2. Composición de los módulos de transferencia de dos traductores

También se ha probado a componer dos traductores componiendo sus módulos de transferencia. Así pues, para traducir del portugués al catalán se ha:

1. analizado el texto con el diccionario morfológico `apertium-pt-ca.pt.dix`

2. desambiguado el texto con las probabilidades del traductor `apertium-es-pt`
3. transferido al castellano con el módulo de transferencia `apertium-es-pt.trules-pt-es.xml` y el diccionario bilingüe `consistent-bil-pt-es.dix` y posteriormente transferido del castellano al catalán con el módulo de transferencia `apertium-es-ca.trules-es-ca.xml` y el diccionario bilingüe `consistent-bil-es-ca.dix`
4. generado en catalán con el diccionario morfológico `apertium-pt-ca.ca.dix`
5. aplicado las reglas de post-generación del diccionario `apertium-es-ca.post-ca.dix`

Para traducir del catalán al portugués se han aplicado los pasos equivalentes en sentido contrario.

3.3.3. Utilización del módulo de transferencia del primer traductor

Se han querido evaluar también los resultados de traducir con las reglas de transferencia estructural de solo uno de los dos traductores de partida. Primero se ha probado con las reglas utilizadas para traducir de la lengua de partida al español. Para ello, antes de aplicar las reglas ha hecho falta cambiar los lemas de salida de algunas de ellas, para que sean lemas de la nueva lengua meta. Los ficheros de reglas `*-m.xml` son el resultado de esta modificación.

A continuación, se ha traducido del portugués al catalán de la misma forma que en la configuración anterior, menos en el paso número 3: en este caso la transferencia se ha hecho directamente con las reglas de `apertium-es-pt.trules-pt-es-m.xml` y con el diccionario bilingüe `apertium-pt-ca.pt-ca.dix`.

Para traducir del catalán al portugués se han hecho las operaciones equivalentes en sentido contrario.

3.3.4. Utilización del módulo de transferencia del segundo traductor

También se ha querido probar el resultado de traducir aplicando reglas de transferencia estructural del español a la lengua meta. En este caso, ha hecho falta cambiar los lemas

de los patrones que hay que detectar en algunos casos, antes de aplicar las reglas, para que sean lemas de la nueva lengua de origen. Estos ficheros modificados también se llaman `*-m.xml`.

Así pues, para traducir del portugués al catalán se han aplicado los mismos pasos, con los mismos datos lingüísticos, que en la configuración anterior, pero cambiando el fichero de reglas de transferencia estructural por `apertium-es-ca.trules-es-ca-m.xml`. Para traducir del catalán al portugués se ha hecho el proceso equivalente en sentido contrario.

3.3.5. Utilización de un módulo de transferencia hecho a mano

Finalmente, se han creado a mano dos ficheros de reglas de transferencia estructural, uno para cada sentido de traducción, combinando las reglas de los módulos de los traductores de partida. Con ello, se ha creado un traductor `apertium-pt-ca` completo.⁷

Al combinar las reglas de transferencia estructural de dos traductores aparecen cinco situaciones distintas:

- **Reglas que son iguales en los ficheros de partida:** Algunas reglas son iguales en los dos traductores de partida. Cuando es así, la regla se puede copiar directamente en el nuevo fichero de reglas de transferencia.
- **Reglas que están sólo en uno de los dos ficheros:** En esta situación, la regla también se puede copiar, pero si está lexicalizada hay que cambiar los lemas. Este ha sido el caso de la regla que para traducir del catalán al castellano cambia la preposición *a* por *en* delante de un topónimo. En cambio, para traducir del castellano al portugués no hay ninguna regla que se ocupe de esta situación, ya que en portugués se utiliza *em*, equivalente a la preposición *en* del castellano.
- **Reglas que detectan el mismo patrón pero aplican acciones distintas:** En estos casos hay que crear una nueva regla que combine las acciones de las dos anteriores. Un ejemplo de ello sería el patrón *verbo-pronombre enclítico*: Por un lado, los pronombres *en* y *hi* del catalán no tienen traducción al español (ni al portugués), por lo que en

⁷Véase 2.1

`apertium-es-ca.trules-ca-es.xml`

se comprueba, antes de enviar el pronombre, si se puede traducir. Por otro lado, los pronombres del portugués pueden ir en posición proclítica, mesoclítica o enclítica, por lo que la regla con este patrón de `apertium-es-pt.trules-es-pt.xml` comprueba en qué posición tendrá que ir el pronombre. Así pues, la nueva regla combinada tiene que comprobar si el pronombre tiene traducción y en qué posición debe ir.

- **Reglas que detectan secuencias distintas de categorías léxicas que se solapan:** A menudo las reglas de los dos ficheros de partida detectan secuencias de categorías léxicas que se solapan o se incluyen. Al combinar las reglas para traducir del portugués al catalán se da el caso de todas las reglas que contengan verbos, ya que para traducir del español al catalán el pretérito perfecto se genera en su forma perifrástica (*canté* → *vaig cantar*).

3.3.6. Mejoras al traductor hechas a mano

A partir del primer modelo de traductor `apertium-pt-ca` completo se hicieron las siguientes mejoras:

- Cambios en los diccionarios:
 - Creación de las secciones *postblank* de los diccionarios (véase 3.1).
 - Introducción de las *expresiones multipalabra* de los diccionarios de partida.
 - Revisión de las equivalencias del diccionario bilingüe a partir de la observación de las traducciones hechas por el traductor: el programa `apertium-dixtools` cruza los diccionarios a partir de los lemas y su categoría gramatical. Eso puede crear errores en el caso de palabras con lemas homógrafos o palabras polisémicas, ya que un mismo lema puede estar en cada traductor con un significado distinto.
 - Introducción en los diccionarios de las 370 palabras desconocidas por el traductor que se dan con más frecuencia en un corpus de 10.000 palabras

Método	WER	% desc.
comp. 2 trad.	10,99 %	10,51 %
comp. 2 mod. transf.	26,55 %	13,05 %
1º mod. transf.	26,24 %	12,77 %
2º mod. transf.	28,33 %	12,60 %
mod. transf. nuevo	25,66 %	12,81 %
mejoras	14,10 %	10,19 %

Cuadro 1: Resultados obtenidos al traducir textos periodísticos del catalán al portugués (véase la sección 3)

- Cambios en los ficheros de reglas de transferencia estructural:
 - Cambio de la traducción de la estructura “*ir* + infinitivo” del portugués: Esta estructura se traduce al español por “*ir* + *a* + infinitivo” (por ejemplo, *vou cantar* → *voy a cantar*); en cambio, en el traductor español↔catalán, no se trata esta estructura. Se han hecho los cambios necesarios para que se traduzca por un verbo al futuro de indicativo (por ejemplo, *vou cantar* → *cantaré*).
 - Uso del determinante delante de porcentajes: En catalán y en español se usa un determinante delante de las expresiones de porcentaje (por ejemplo, *el aumento fue del 3 %*), mientras que en portugués, no (*o aumento foi de 3 %*). Se ha ampliado el número de reglas que tratan esta diferencia.

4. Resultados, conclusiones y líneas abiertas

En los cuadros 1 y 2 se pueden ver los resultados obtenidos en cada una de las configuraciones explicadas, expresados en términos de tasa de error por palabra o *word error rate* (WER) y en porcentaje de palabras desconocidas por el traductor.

4.1. Conclusiones

A partir de los resultados presentados podemos llegar a las siguientes conclusiones:

- Los resultados de la aplicación sucesiva de dos traductores completos son bastante buenos, lo que indica que esta es una buena configuración a utilizar cuando no se dispone de ningún traductor en-

Método	WER	% desc.
comp. 2 trad.	15,17 %	8,74 %
comp. 2 mod. transf.	19,28 %	10,37 %
1º mod. transf.	23,56 %	10,49 %
2º mod. transf.	19,33 %	10,28 %
mod. transf. nuevo	18,52 %	10,24 %
mejoras	16,67 %	9,06 %

Cuadro 2: Resultados obtenidos al traducir textos periodísticos del portugués a catalán (véase la sección 3)

tre las dos lenguas entre las que se quiere traducir.

- La función de cruce de diccionarios de *apertium-dixtools* es una herramienta muy útil para cruzar los diccionarios de dos traductores Apertium, pero, como se muestra (% de palabras desconocidas en los cuadros 1 y 2) se pierde cobertura léxica. Así pues, si se quiere utilizar *apertium-dixtools* para desarrollar un nuevo traductor hay que considerar que los diccionarios resultantes serán sólo una base a partir de la cual habrá que hacer revisiones y mejoras.
- Cruzar dos traductores Apertium permite desarrollar un nuevo traductor rápidamente y con poco esfuerzo, pero hay que asumir que en un principio se pierde calidad en la traducción. Así pues, saldrá a cuenta sólo si se hace con la intención de añadir después mejoras específicas para traducir entre el nuevo par de lenguas.
- Entre las mejoras que se tienen que aplicar al nuevo traductor hay que considerar el aumento de la cobertura léxica y la introducción de expresiones multipalabra.
- A la vista de los resultados de traducir utilizando uno de los módulos de transferencia estructural de los traductores de partida, vemos que, como las lenguas romances tienen pocas diferencias estructurales entre sí, se pueden adaptar fácilmente estos módulos para traducir entre otros pares de lenguas.

4.2. Líneas abiertas

Apertium dispone de herramientas y datos lingüísticos que se pueden reutilizar para desarrollar rápidamente y con poco esfuerzo los datos necesarios para traducir entre un

nuevo par de lenguas. Sin embargo, aún se pueden crear nuevas herramientas y mejorar las ya existentes para facilitar la reutilización de datos, como por ejemplo:

- Mejorar la función de cruce de diccionarios del programa *apertium-dixtools*.
- Creación de un sistema de cruce semiautomático de reglas de transferencia estructural.
- Adaptación de las herramientas para facilitar el intercambio de entradas entre los distintos diccionarios Apertium.

Además, se está trabajando en el desarrollo de programas para facilitar el intercambio de datos lingüísticos entre Apertium y otros sistemas de procesamiento automático del lenguaje: Freeling⁸ (Atserias et al., 2006), jspell⁹ (Almeida y Pinto, 1995), u OLIF (*Open Lexical Interchange Format*)¹⁰.

Bibliografía

- Almeida, J.J. y Ulisses Pinto. 1995. Jspell – um módulo para análise léxica genérica de linguagem natural. En *Actas do X Encontro da Associação Portuguesa de Linguística*, páginas 1–15, Évora 1994.
- Armentano-Oller, Carme, Rafael C. Carrasco, Antonio M. Corbí-Bellot, Mikel L. Forcada, Mireia Ginestí-Rosell, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, y Miriam A. Scalco. 2006. Open-source Portuguese-Spanish machine translation. En R. Vieira P. Quaresma M.d.G.V. Nunes N.J. Mamede C. Oliveira, y M.C. Dias, editores, *Computational Processing of the Portuguese Language, Proceedings of the 7th International Workshop on Computational Processing of Written and Spoken Portuguese, PRO-POR 2006*, volumen 3960 de *Lecture Notes in Computer Science*. Springer-Verlag, May, páginas 50–59.
- Atserias, Jordi, Bernardino Casas, Elisabet Comelles, Meritxell González, Lluís Padró, y Muntsa Padró. 2006. Freeling 1.3: Syntactic and semantic services

⁸<http://www.lsi.upc.edu/~nlp/freeling>

⁹<http://natura.di.uminho.pt/natura/natura?&topic=jspell>

¹⁰<http://www.olif.net/>

in an open-source NLP library. En *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)*, páginas 48–55. ELRA. Génova, Italia, mayo 2006.

Forcada, Mikel L. 2006. Open-source machine translation: an opportunity for minor languages. En *Strategies for developing machine translation for minority languages (5th SALT MIL workshop on minority languages)*. Organized in conjunction with LREC 2006.

Hutchins, W.J. y H.L. Somers. 1992. *An Introduction to Machine Translation*. Academic Press, London.