

Diseño de un etiquetario castellano para el traductor automático castellano–catalán interNOSTRUM

R. Canals-Marote, A. Esteve-Guillén, A. Iturraspe-Bellver y M.L. Forcada
interNOSTRUM

Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant, E-03071 Alacant
{raul,anna,amaia,mlf}@internostrum.com

Resumen La resolución de los homógrafos (formas superficiales para las que el analizador morfológico entrega más de una forma léxica) es fundamental en los programas de traducción automática. Por ejemplo, la palabra castellana *decepciones* puede ser sustantivo o verbo y sus traducciones al catalán son distintas en cada caso: *decepcions* o *decebis*. En este artículo describimos el conjunto de etiquetas (etiquetario morfosintáctico) usadas por el desambiguador léxico categorial de interNOSTRUM y los criterios lingüísticos y computacionales usados en su diseño.

1 Introducción

1.1 Etiquetarios

Un *etiquetario morfosintáctico* es un conjunto de etiquetas (categorías, *tags*) que podemos asignar a las palabras de un idioma de manera que las que reciben una determinada etiqueta tienen un comportamiento sintáctico similar, que puede estar relacionado con una morfología similar. Por ejemplo, la etiqueta “verbo léxico en pretérito imperfecto de indicativo” se puede aplicar a las palabras *cantábamos*, *bebían*, *veáis*, etc. También puede suceder que se pueda asignar más de una etiqueta a una palabra concreta: por ejemplo, la palabra *ahorro* puede recibir las etiquetas “sustantivo masculino singular” y “verbo léxico en presente de indicativo”, idealmente, un programa desambiguador debería poder usar el contexto —por ejemplo las etiquetas de las palabras vecinas— para elegir la etiqueta correcta.

Los etiquetarios pueden variar en grado de refinamiento: pueden ser muy gruesos (“sustantivo”, “verbo”, “pronombre”) o muy finos (“sustantivo masculino singular”, “verbo léxico en tercera persona del plural del pre-

sente de indicativo”, “pronombre átono dativo de tercera persona, masculino, singular”). En general, la elección depende de la aplicación concreta que se quiera dar al etiquetario. Una posible aplicación es la traducción automática.

1.2 El traductor automático interNOSTRUM

interNOSTRUM es un sistema de traducción automática (TA) del castellano al catalán (Canals et al. (2000), disponible en línea: www.internostrum.com). Como la mayoría de sistemas de TA, se trata de un sistema de traducción automática indirecta por *transferencia*; estos sistemas realizan la traducción en tres fases: el *análisis*, que convierte el texto origen en algún tipo de representación abstracta que destaca o extrae los detalles relevantes para la traducción; la *transferencia* propiamente dicha, que convierte la representación abstracta del texto origen en una representación abstracta de naturaleza similar pero referida al texto meta, y la *generación* o *síntesis* que genera el texto traducido a partir de esta representación. En la versión más idealizada del esquema de transferencia el módulo de análisis sólo depende de la lengua origen (y por tanto, sólo usa información sobre ésta), el módulo de generación sólo depende de la lengua meta (por tanto, sólo usa información sobre ésta) y sólo el módulo de transferencia incluye información bilingüe, tanto léxica como estructural.

La fase de análisis de interNOSTRUM se compone de tres subfases:

- el *análisis morfológico*, que entrega, para cada forma superficial (forma concreta de la palabra en el texto) una o más formas léxicas compuestas de *lema* o forma canónica (la forma que encontramos en el diccionario), categoría léxica, y información de flexión. Por ejemplo, el analizador convierte la forma superficial *can-*

tabas en la forma léxica compuesta por el lema *cantar*, la categoría *verbo léxico*, y la información de flexión *pretérito imperfecto, indicativo, 2ª persona, singular*; la salida del analizador morfológico de interNOSTRUM para este caso sería

```
cantabas/cantar<vblex><pii><2><sg>/
```

- si a una forma superficial le corresponde más de una forma léxica, el *desambiguador* analiza la categoría gramatical y la información de flexión de cada forma léxica y asigna una etiqueta a cada una de ellas; después, aplica criterios estadísticos a las etiquetas de las palabras vecinas y elige una de las etiquetas para la palabra homógrafa en cuestión. El etiquetador de interNOSTRUM no usa toda la información que le entrega el analizador; en el ejemplo de arriba, sólo necesita las dos primeras etiquetas, <vblex><pii>, para asignar a la forma léxica la etiqueta VLEXIPI que incluye el pretérito imperfecto y el indefinido de indicativo; la información de persona y número no se usa en la desambiguación categorial, se necesitará más tarde en las etapas de transferencia y generación.
- finalmente, un programa detecta (usando las etiquetas elegidas) aquellas construcciones de dos o más formas léxicas consecutivas (sintagmas rudimentarios) que requieran un tratamiento especial (por ejemplo, concordancia de género, de número, cambios de preposición, reordenamiento, etc.)

1.3 Estructura de la comunicación

El resto del artículo se organiza como sigue: en el apartado 2 se describen los factores que pueden influir en el diseño de un etiquetario y los criterios usados en interNOSTRUM para diseñarlo; el apartado 3 explica con mayor detalle cómo pueden usarse los problemas de traducción para guiar el diseño del etiquetario; el apartado 4 describe el etiquetario usado en interNOSTRUM, seguido de una breve conclusión.

2 Factores que influyen en el diseño del etiquetario

Cuando un equipo se plantea diseñar un etiquetario nuevo (o adaptar uno existente) para una aplicación concreta, debe tener en cuenta varios factores.

2.1 Motivación lingüística

El etiquetario tiene que estar motivado lingüísticamente, de manera que agrupe categorías de palabras con comportamientos sintácticos homogéneos y diferencie claramente las que tengan comportamientos desiguales: por ejemplo, la etiqueta *verbo* es demasiado gruesa. En cambio, si distinguimos gerundios, infinitivos, participios y verbos en forma personal, podemos decir que en castellano un artículo o una preposición no puede ir seguido de un verbo conjugado, pero sí de un infinitivo.

Pero si una palabra es homógrafa de tal modo que recibe dos etiquetas que son prácticamente imposibles de discernir usando información (reglas, estadística) basada en el contexto sintáctico inmediato, será muy difícil desambiguarla usando métodos tradicionales. Por ejemplo, el verbo *cantamos* puede estar en presente o en pretérito indefinido de indicativo; la traducción al catalán es diferente en cada caso (*cantem, vam cantar*) pero es difícil que un programa pueda usar el contexto sintáctico inmediato para decidir en qué caso se encuentra.

2.2 Universalidad *versus* adecuación

Por lo general, un etiquetario puede estar bien diseñado sin tener en cuenta una aplicación concreta, con la intención de que sirva para varias aplicaciones, o bien teniendo en cuenta la aplicación concreta. En interNOSTRUM, aunque esto haya supuesto una violación del principio (enunciado más arriba) que dicta que la fase de análisis de un sistema de traducción automática por transferencia tiene que estar diseñada de manera independiente de la lengua de llegada, hemos buscado siempre los etiquetarios que distinguían, en la medida de lo posible aquellas categorías que, en palabras homógrafas, darían lugar a traducciones diferentes si no se desambiguasen. Del mismo modo, hemos renunciado a hacer diferenciaciones que no dan lugar a problemas de traducción, o a cargar el etiquetario con información (como el género y número de las palabras) que se puede tratar adecuadamente en la subfase posterior (identificación de construcciones que requieren tratamiento especial). Finalmente, es obvia la conveniencia de usar un conjunto de etiquetas que se puedan inferir de manera sencilla (normalmente, eliminando parte de la información) a

partir de los indicadores categoriales y morfológicos que actualmente entrega el analizador morfológico de interNOSTRUM.

2.3 Tamaño del etiquetario

Un etiquetario extenso nos permitirá hacer distinciones más finas que un etiquetario pequeño. Esto sería una ventaja: por ejemplo, diferenciar singulares y plurales de los nombres comunes nos permitiría desambiguar palabras castellanas como *anàlisis* (catalán: *anàlisi* en singular, *anàlisis* en plural). Sin embargo, cualquier aumento en el tamaño del etiquetario tiene un efecto muy grande en la complejidad del programa desambiguador. Por ejemplo, si el etiquetador se basa, como el de interNOSTRUM, en un modelo oculto de Markov como el de Cutting et al. (1992) o el descrito en Manning y Schütze (1999, cap. 10) que consideran la estadística de aparición de bigramas (secuencias de dos etiquetas), duplicar el etiquetario cuadruplica el número de parámetros numéricos (probabilidades de transición) del programa etiquetador; si el etiquetador es de trigramas, el número de parámetros se octuplica. Si el desambiguador está basado en reglas, se habrán de formular más. Además de la complejidad intrínseca de un etiquetador tan grande, en el caso de los etiquetadores estadísticos existe el problema de la estimación de los parámetros numéricos a partir de los corpóra: si tenemos 40 etiquetas, hay 1600 bigramas y 64000 trigramas cuyas probabilidades deben estimarse. Se debe establecer, por tanto, un compromiso; el compromiso actual en interNOSTRUM se sitúa en 60 etiquetas.

2.4 Relación con otros etiquetarios

Existen varios etiquetarios para el castellano, la mayoría de ellos de propósito general: CRATER (Sánchez León y Nieto Serrano 1995), Projecte Corpus (<http://www.iula.upf.es/corpus/etiqaes.htm>), Xerox (<http://www.xrce.xerox.com/research/mltt/demos/doc/pos-spa-1.html>), y Parole (usado en combinación con el analizador MACO <http://www.lsi.upc.es/~nlp/parole-eng.html>). El diseño del etiquetario de interNOSTRUM no se ha hecho de espaldas a estos etiquetarios por al menos dos razones: la primera, porque así se ha podido aprovechar la experiencia contenida en su diseño, y la segunda, porque ello permite el uso de etiquetadores existen-

tes o textos etiquetados (manualmente o automáticamente) para evaluar nuestro etiquetador. La segunda razón es especialmente importante, porque podemos actuar de dos maneras diferentes:

- podemos usar un etiquetado (manual o automático) anterior $E_1(T)$ de un texto T de la siguiente manera: analizamos con nuestro analizador A_2 el texto T para generar un texto analizado $A_2(T)$. Donde $A_2(T)$ contiene homógrafos, consultamos $E_1(T)$ para desambiguarlos. Si $E_1(T)$ es un etiquetado suficientemente fino, será capaz de desambiguar completamente $A_2(T)$ y producirá $E_2(T)$, es decir, el texto T etiquetado con nuestras etiquetas.
- Si tenemos el programa etiquetador E_1 , podemos hacer lo mismo pero con cualquier texto.

Estas operaciones las podríamos denominar *etiquetado cruzado* (en inglés, *cross tagging*) y nos han servido para generar textos etiquetados con etiquetas de interNOSTRUM usando otros textos, como los ya etiquetados con CRATER que se pueden encontrar en ftp://ftp.lll1f.uam.es/pub/corpus/ITU_spanish_hand-corrected.tar.gz.

2.5 Las unidades multipalabra

Un caso especial lo presentan las llamadas *unidades multipalabra*: secuencias de más de una palabra que se decide tratar como si fuesen una única palabra. Por ejemplo, podemos tratar *a través de* como una preposición, *dirección general* como un sustantivo, *Sant Vicent del Raspeig* como nombre propio o *sacar de quicio* como un verbo. En interNOSTRUM hemos intentado que estas unidades multipalabra se correspondiesen en la medida de lo posible con partes de la oración claramente identificables que recibiesen las etiquetas correspondientes.

3 Diseño del etiquetario guiado por los problemas de traducción

¿Qué clase de problemas de traducción se pueden resolver con un etiquetario? Primero, vamos a distinguir qué problemas *no* se podrán resolver nunca con un etiquetador:

- La polisemia, es decir, el fenómeno por el cual formas superficiales (palabras polisémicas) que tienen una única forma

HOMÓGRAFO	TRADUCCIÓN
<i>ahorro</i> (subst./verbo)	<i>estalvi, estalvio</i>
<i>vino</i> (subst./verbo)	<i>vi, va venir</i>
<i>para</i> (prep./verbo)	<i>per (a), para/atura</i>
<i>como</i> (conj./verbo)	<i>com (a), menjo</i>

Tabla 1: Homógrafos castellanos categoriales con traducciones catalanas divergentes.

HOMÓGRAFO	TRADUCCIÓN
<i>cantamos</i> (pres./perf.)	<i>cantem, vam cantar</i>
<i>vivimos</i> (pres./perf.)	<i>vivim, vam viure</i>
<i>salen</i> (indic./subj.)	<i>surten, salin</i>
<i>venga</i> (indic./subj.)	<i>venja, vingui</i>

Tabla 2: Homógrafos castellanos no categoriales con traducciones catalanas divergentes.

léxica son ambiguas porque el lema puede tener más de una interpretación (p.ej. *estación, destino, escondite, librar*).

- Ciertos tipos de homografía dentro de la misma categoría. Hay homógrafos que no se pueden resolver porque la única cosa que cambia es el lema: *creo* (lemas *crear* y *creer*), *vendo* (lemas *vendar* y *vender*), etc.). Hay algunos casos dónde la resolución de estos homógrafos se podría abordar distinguiendo, por ejemplo, con categorías diferentes los verbos modales o auxiliares del resto de los verbos (léxicos), como *fui* (lemas *ir* y *ser*).

Es decir, podemos intentar resolver el resto de los homógrafos. Pero, como se ha dicho antes, en realidad sólo nos interesa resolver aquellos que afectan a la calidad de la traducción porque cada una de sus formas léxicas tiene una traducción diferente en alguna variedad del catalán; por ejemplo, no tiene sentido distinguir la 1ª y la 3ª persona de formas del imperfecto de indicativo o de subjuntivo o del presente de subjuntivo (*cantaba, cante, cantara* o *cantase*) porque las traducciones también son homógrafas en catalán (*cantava, canti, cantés*).

El caso más obvio son los homógrafos categoriales, como los que se muestran en la tabla 1. Sin embargo, existen otros casos dónde la homografía no afecta propiamente la categoría pero sí la traducción (tabla 2).

Es pertinente un comentario sobre la frecuencia de las palabras ambiguas. En cas-

Palabra	categorías	<i>f</i>
<i>la</i>	pronombre, artículo	0,051
<i>que</i>	conj., relativo	0,027
<i>los</i>	pronombre, artículo	0,027
<i>las</i>	pronombre, artículo	0,018
<i>para</i>	verbo, preposición	0,0091
<i>una</i>	verbo, artículo	0,0088
<i>como</i>	verbo, conjunción	0,0050

Tabla 3: Siete de los vocablos ambiguos más frecuentes en castellano y su frecuencia relativa observada en un corpus de textos periodísticos de un millón de palabras.

tellano, una de cada tres o cuatro palabras es homógrafa. Si el desambiguador comete errores en palabras muy poco frecuentes, esto no afecta demasiado a la calidad global de la traducción; el problema está cuándo comete errores en las palabras homógrafas más comunes. La tabla 3 indica siete de los homógrafos castellanos más frecuentes y la frecuencia relativa *f* con qué se observan en un corpus de aproximadamente un millón de palabras. Las palabras *para*, *una* y *como* son especialmente problemáticas porque sus traducciones son divergentes. Por poco que el desambiguador estadístico se equivoque en estas palabras tan frecuentes, se degrada notablemente la calidad de las traducciones. En el caso de interNOSTRUM la resolución incorrecta de homógrafos es una de las causas más frecuentes de errores de traducción.

Para diseñar el etiquetario, se construyó un catálogo completo de situaciones en las que la homografía afectaba la traducción: por un lado, estudiando la casuística existente, y por otro lado, evaluando las traducciones producidas por interNOSTRUM para detectar aquellos problemas que se pudiesen resolver mediante las distinciones oportunas. Cada distinción crea un etiquetario más extenso y, por tanto, más fino. Finalmente, se generaron todos los homógrafos que aceptaba el analizador morfológico de interNOSTRUM y que eran susceptibles de solución mediante un refinamiento del etiquetario y (usando los módulos restantes de interNOSTRUM) se discernieron aquellos que daban lugar a traducciones divergentes. Esta última estrategia podría parecer la más potente, pero tiene el inconveniente de que no da ninguna información sobre la frecuencia relativa de cada tipo de homografía (podemos hacernos una

idea mejor de las frecuencias estudiando textos reales), por ello es complementaria.

Hay un problema más: ¿qué divisiones de las etiquetas son las más adecuadas? Por ejemplo, si se quiere distinguir el presente de indicativo y el pretérito perfecto simple (*cantamos = cantem/vam cantar*), ¿se hacen dos etiquetas especiales, una para cada tiempo? ¿O se segrega sólo uno de los dos? ¿Cuál es el refinamiento (la partición) más adecuado? Una partición demasiado fina puede producir etiquetas tan poco frecuentes en los corpórea de entrenamiento que sean muy difíciles de distinguir estadísticamente.

Finalmente, conviene establecer un esquema claro de jerarquías de inclusión de las etiquetas más finas en etiquetas más gruesas (del estilo del que se hace en CRATER, Sánchez León y Nieto Serrano (1995)), para poder decidir de esta manera cuál es la granularidad adecuada.

4 Descripción del etiquetario de interNOSTRUM

Las tablas 4 y 5 describen el etiquetario que usa actualmente interNOSTRUM. Por lo general, la división en etiquetas ha sido inspirada por las indicaciones de la sección anterior, pero merecen explicación aparte algunas de ellas.

- Se han distinguido los antropónimos porque muchas veces aparecen en grupos de más de uno (*Juan López Sanchis*) o detrás de un sustantivo común (*El profesor López*).
- Se han distinguido los topónimos porque requieren la preposición *a* en las construcciones locativas en catalán (*en València* → *a València*).
- Se han distinguido los predeterminantes de otros vocablos con función de determinantes para evitar secuencias de más de un determinante. Así, podemos tratar bien secuencias como *una los dos cabos y...* donde la palabra *una* sólo puede ser verbo si se prohíben las secuencias de dos determinantes.
- Las formas conjugadas de los verbos se han distribuido en varias etiquetas para distinguir homógrafos castellanos que tenían traducciones divergentes al catalán.

- Se distinguen los verbos léxicos de otros verbos que actúan como auxiliares (*ser, haber*) o como modales *poder, deber*, por razón de las diferencias entre sus contextos sintácticos habituales y para poder desambiguar, al menos parcialmente, palabras como *podemos* (*podar* léxico, *poder* modal) o *fue* (*ser* auxiliar, *ir* léxico).

5 Conclusión

En este artículo hemos presentado el conjunto de etiquetas usadas por el desambiguador del traductor automático castellano-catalán interNOSTRUM. La elección de estas etiquetas ha estado motivada por criterios sintácticos, de manera que las categorías gramaticales que aparecen en los mismos contextos sintácticos han sido agrupadas en una única etiqueta, aunque también hemos considerado la distinción de las categorías que, en palabras homógrafas, dan lugar a traducciones diferentes. Además, hemos tenido en cuenta las frecuencias de aparición de los homógrafos para priorizar la resolución de aquellos que más degradan la calidad de las traducciones. En cuanto al número de etiquetas hemos buscado un compromiso entre la finura del etiquetario (su número de etiquetas) y la complejidad del programa desambiguador de manera que tenemos 60 etiquetas; ello supone para un modelo estadístico la estimación de 3600 bigramas o 216.000 trigramas, según el caso. A lo largo del diseño hemos tenido en cuenta el diseño de otros etiquetarios existentes para el castellano.

Agradecimientos: Este trabajo ha sido financiado por la Caja de Ahorros del Mediterráneo y por el Vicerrectorado de Nuevas Tecnologías de la Universidad de Alicante.

Referencias

- Canals, R., Esteve, A., Garrido, A., Guardiola, M. I., Iturraspe-Bellver, A., Montserrat, S., Pérez-Antón, P., andc Hermínia Pastor, S. O., y Forcada, M. L. (2000). interNOSTRUM: a Spanish-Catalan machine translation system. *Machine Translation Review (ISSN 1358-8346)*, (11). (also available at <http://www.bcs.org.uk/siggroup/nalatran/mtreview/>).
- Cutting, D., Kupiec, J., Pedersen, J., y Sibun, P. (1992). A practical part-of-speech tagger. En *Third Conference on Applied Natural Language Processing. Association*

ETIQUETA	DESCRIPCIÓN	EJEMPLOS
ACRONIM	Acrónimo	<i>CAM, ONU, sida</i>
NOMBRE	Sustantivo común	<i>casa, dígitos</i>
ANTROPONIM	Nombre propio de persona	<i>Javier, López</i>
TOPONIM	Nombre propio de lugar	<i>Valencia, Italia</i>
NPALTRES	Otros nombres propios	<i>Moulinex, Pentium</i>
ADJ	Adjetivo	<i>rojos, mío</i>
ADV	Adverbio	<i>más, ahora</i>
PREADV	Adverbio preadjetival o preadverbial	<i>muy, tan</i>
CNJCOORD	Conjunción coordinada	<i>y, e, pero</i>
CNJSUB	Conjunción subordinada	<i>si, aunque</i>
CNJADV	Conjunción adverbial	<i>aunque, sí</i>
DET	Determinantes	<i>aquel, el, su</i>
DETNEU	Determinante neutro	<i>lo</i>
INTNOM	Pronombre interrogativo	<i>cuál, quién</i>
INTADJ	Adjetivo interrogativo	<i>cuánto, qué</i>
INTADV	Adverbio interrogativo	<i>cómo, dónde</i>
INTERJ	Interjección	<i>adiós, ojalá</i>
NUM	Numeral cardinal	<i>diez, sesenta y cinco</i>
PREDET	Predeterminante	<i>toda</i>
PREDETNEU	Predeterminante neutro	<i>todo</i>
PREP	Preposición	<i>a, de</i>
PRNTON	Pronombre tónico	<i>ambos, nadie</i>
PRNTONNEU	Pronombre tónico neutro	<i>algo, eso</i>
PRNATNAC	Pronombre átono acusativo	<i>la, me</i>
PRNATNDAT	Pronombre átono dativo	<i>le</i>
PRNATNREF	Pronombre átono reflexivo	<i>se</i>
REL	Relativo	<i>cuyo, que, quien</i>
RELADV	Relativo adverbial	<i>adonde, cuando</i>
VLEXINF	Verbo léxico en infinitivo	<i>escuchar</i>
VLEXGER	Verbo léxico en gerundio	<i>escuchando</i>
VLEXPARTPI	Participio de verbo léxico	<i>escuchadas</i>
VLEXPFCI	Verbo léxico en presente, futuro o condicional de indicativo	<i>escucharíamos</i>
VLEXIPI	Verbo léxico en pretérito imperfecto o indefinido de indicativo	<i>escuchaba, escuché</i>
VLEXSUBJ	Verbo léxico en subjuntivo	<i>escuchara</i>
VLEXIMP	Verbo léxico en imperativo	<i>escuchad</i>

Tabla 4: El etiquetario de interNOSTRUM, primera parte.

for Computational Linguistics. Proceedings of the Conference. 31 March – 3 April 1992. Trento, Italy, pages 133–140.

Manning, C. D. y Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Sánchez León, F. y Nieto Serrano, A. F. (1995). Crater – Corpus Resources and Terminology Extraction deliverable report 22 – public domain pos tagger

for spanish. Technical report, Universidad Autónoma de Madrid. (disponible en http://www.l11f.uam.es/docs_en/final_report/drep22.html).

ETIQUETA	DESCRIPCIÓN	EJEMPLOS
VSERINF	Verbo <i>ser</i> en infinitivo	<i>ser</i>
VSEGER	Verbo <i>ser</i> en gerundio	<i>siendo</i>
VSEPARTPI	Verbo <i>ser</i> en participio	<i>sido</i>
VSEPF	Verbo <i>ser</i> en presente, futuro o condicional de indicativo	<i>soy, seríais</i>
VSEIPI	Verbo <i>ser</i> en pretérito imperfecto o indefinido de indicativo	<i>fue, era</i>
VSESUBJ	Verbo <i>ser</i> en subjuntivo	<i>fuese, sean</i>
VSEIMP	Verbo <i>ser</i> en imperativo	<i>sé</i>
VHBERINF	Verbo <i>haber</i> en infinitivo	<i>haber</i>
VHBERGER	Verbo <i>haber</i> en gerundio	<i>habiendo</i>
VHBERPARTPI	Verbo <i>haber</i> en participio	<i>habido</i>
VHBERPF	Verbo <i>haber</i> en presente, futuro o condicional de indicativo	<i>habríamos</i>
VHBERIPI	Verbo <i>haber</i> en pretérito imperfecto o indefinido de indicativo	<i>hubo</i>
VHBERSUBJ	Verbo <i>haber</i> en subjuntivo	<i>haya</i>
VMODALINF	Verbo modal en infinitivo	<i>deber, poder</i>
VMODALGER	Verbo modal en gerundio	<i>debiendo</i>
VMODALPARTPI	Verbo modal en participio	<i>debido</i>
VMODALPF	Verbo modal en presente, futuro o condicional de indicativo	<i>debemos</i>
VMODALIPI	Verbo modal en pretérito imperfecto o indefinido de indicativo	<i>debió</i>
VMODALSUBJ	Verbo modal en subjuntivo	<i>deba</i>
VMODALIMP	Verbo modal en imperativo	<i>debe</i>
LQUEST	Apertura de interrogación	<i>¿</i>
RPAR	Paréntesis o corchete derecho	<i>) ,]</i>
LPAR	Paréntesis o corchete izquierdo	<i>(, [</i>
CM	Coma	<i>,</i>
SENT	Signos de “fin de frase”	<i>. , ; , ? , !</i>

Tabla 5: El etiquetario de interNOSTRUM, segunda parte.