

# **Aplicacións da lexicografía bilingüe baseada en cónpora na elaboración do Dicionario CLUVI inglés-galego**

Xavier Gómez Guinovart, Eva Díaz Rodríguez, Alberto Álvarez Lugrís  
Universidade de Vigo

## **Resumo**

Neste artigo presentamos a segunda edición (2008) do Dicionario CLUVI Inglés-Galego, un dicionario bilingüe de traducións contextualizadas que se está a desenvolver a partir dos datos do Corpus CLUVI (Corpus Lingüístico da Universidade de Vigo).

## **Abstract**

This paper presents the second edition (2008) of the English-Galician CLUVI Dictionary, a corpus-based dictionary of contextual translations which is being elaborated from data in the CLUVI Corpus (Linguistic Corpus of the University of Vigo).

## **1. Introducción**

O *Dicionario CLUVI Inglés-Galego* (Gómez Guinovart et al., 2008) é un dicionario baseado na colección de textos ingleses traducidos ao galego que forma parte do *Corpus CLUVI* (Gómez Guinovart, 2003) e constitúe, ao noso entender, o primeiro dicionario baseado en cónpora da lexicografía galega. Todas as palabras inglesas que aparecen nas súas entradas están documentadas nos textos en inglés traducidos ao galego recompilados no corpus paralelo CLUVI. Alén diso, todas as traducións galegas recollidas no dicionario para esas palabras son traducións reais identificadas nas versións galegas dos textos ingleses do corpus. Finalmente, para cada tradución identificada, o dicionario fornece un exemplo real do seu uso tal como está documentado no corpus. Desde o ano 2005, está dispoñible na web, no enderezo <http://sli.uvigo.es/dicionario>. A primeira edición do dicionario, na súa versión 1.5 (2006) alcanzaba as 6.677 entradas e 10.807 traducións. A segunda edición, dispoñible desde setembro de 2008, incorpora ao dicionario un maior número de entradas e equivalencias tiradas do Corpus CLUVI

(20.000 entradas con 30.000 traducións e 60.000 exemplos), ao tempo que amplía os datos lexicográficos contidos nos artigos con información sobre americanismos e variantes ortográficas e con notas de interese gramatical, tradutolóxico e normativo. O obxectivo destes engadidos é que a ferramenta resultante poida ser realmente útil tanto na docencia do inglés como na tradución inglés-galego. Aínda que as entradas desta obra están redactadas só na dirección de tradución inglés-galego, o sistema de busca implementado permite recuperar tamén as entradas a partir das súas traducións ao galego, converténdose así tamén nun dicionario galego-inglés. A editorial Reverso ten previsto publicar a versión impresa desta segunda edición do dicionario a comezos de 2009.

Neste artigo imos presentar as principais características desta nova edición do *Dicionario CLUVI inglés-galego*, así como as técnicas empregadas para a súa elaboración a partir do corpus de traducións CLUVI.

## **2. Antecedentes**

A orientación á investigación aplicada baseada en córpora textuais está cada vez máis consolidada como unha metodoloxía frutífera para a descrición e análise dos fenómenos lingüísticos en practicamente todos os seus aspectos. O estudo da lingua a través dos córpora permite achegarse dunha maneira empírica ao uso real da linguaxe no seu contexto. A análise das unidades léxicas dun corpus textual permite observar dun modo moi realista a súa potencialidade semántica, a súa frecuencia de uso, a súa combinatoria e, nos córpora paralelos, as súas posibles traducións.

O primeiro caso de éxito na introdución do uso de córpora textuais informatizados na práctica lexicográfica débese á Universidade de Birmingham e á editorial Collins, promotora do dicionario *Cobuild* (Sinclair, 1987), cuxa primeira edición data de 1987. Naquela altura, o proxecto Cobuild supuxo unha grande innovación, xa que por vez primeira utilizaba un corpus representativo de textos para a elaboración dun dicionario dunha lingua, neste caso, do inglés contemporáneo.

Tralo éxito do Cobuild, a metodoloxía de traballo da lexicografía baseada en córpora foi seguida por outras grandes editoriais, como Oxford University Press, Longman e Larousse (que colaboraron na elaboración do British National Corpus) e Cambridge University Press.

No caso do español, podemos ver exemplos recentes da aplicación desta metodoloxía no dicionario publicado pola editorial SGEL a partir do corpus *Cumbre* (Sánchez, 2001) ou no dicionario de colocacións *Redes* (Bosque, 2004) baseado nun corpus xornalístico de 250 millóns de palabras da editorial SM. A metodoloxía de traballo da lexicografía baseada en córpora estase a utilizar tamén para o catalán na elaboración por parte do IEC do *Diccionari descriptiu de la llengua catalana* baseado no Corpus Textual Informatizat de la Llengua Catalana (Rafel, 1997). En Galicia, o corpus de referencia do galego denominado *Tesouro Informatizado da Lingua Galega* (Santamarina, 2003) constitúe a base do dicionario de uso da lingua galega dirixido polo profesor Antón Santamarina, en fase de preparación; e o corpus paralelo CLUVI, elaborado no Seminario de Lingüística Informática da Universidade de Vigo, é a fonte no que se alicerza o *Diccionario CLUVI inglés-galego*, obxecto deste artigo.

### **3. Do corpus ao dicionario**

O Corpus CLUVI está formado por un conxunto de córpora textuais paralelos de diversos rexistros especializados da lingua galega contemporánea que suman uns 22 millóns de palabras. As traducións na combinación inglés-galego que forman parte do CLUVI e que serviron de base para o dicionario pertencen ao Corpus UNESCO de divulgación científica inglés-galego-francés-español (3.724.620 palabras), ao Corpus LOGALIZA de localización de software inglés-galego (2.375.157 palabras), ao Corpus TECTRA de textos literarios inglés-galego (1.476.020 palabras) e ao Corpus VEIGA de subtitulación de cinema inglés-galego (78.032 palabras), un subconxunto textual do CLUVI que ascende a uns 8 millóns de palabras.

A primeira fase no traballo de elaboración do dicionario foi precisamente a preparación desta fonte textual, a súa codificación informática e a anotación das equivalencias de tradución entre as oracións en inglés e galego dos textos paralelos recompilados. Adoptouse a oración como unidade de tradución de referencia porque a súa identificación e delimitación se poden automatizar case por completo. Neste sentido, empregamos unha definición de oración de base puramente formal que podería resumirse como "o fragmento de texto situado entre dous puntos ortográficos ou equivalentes (signos de puntuación que supoñan unha pausa maior, como as admiracións e as interrogacións)". Esta definición ten, por suposto, as súas excepcións e deficiencias, como veremos máis abaixo, pero permítenos desenvolver un traballo máis rápido e cómodo cos textos.

Despois de identificar e etiquetar as oracións dos orixinais e das traducións, establecemos as correspondencias entre as unidades de tradución de ambos os dous grupos de textos, dando lugar así ao que se denomina tecnicamente un corpus paralelo aliñado a nivel de oración. Neste corpus, de forma xeral, a cada unidade do texto orixinal correspóndelle outra unidade do texto traducido. Hai, non obstante, excepcións: as chamadas asimetrías da tradución, que constitúen a principal dificultade para o aliñamento e, en consecuencia, tamén para o tratamento automático da extracción léxica bilingüe a partir do corpus paralelo. As asimetrías de tradución pódense definir como correspondencias non biunívocas e alteracións na orde das unidades de tradución. Se o habitual é que as correspondencias entre orixinal e tradución sexan biunívocas (1:1), abundan os casos nos que esta correspondencia pode ser 1:0, 1:1/2, 1:2, 0:1, etc. Do mesmo xeito, podemos atopar nas traducións fragmentos desprazados, mesmo oracións enteiras que cambian de lugar. A existencia destes fenómenos motivou o uso dunha serie de etiquetas especiais para identificalos e facilitar o traballo de extracción automática de equivalencias léxicas bilingües candidatas a seren lemas do dicionario. A este efecto, unha vez identificados e marcados os casos de omisión, adición e reordenamento, xerouse unha nova copia do corpus no que estas asimetrías se eliminaron, do mesmo xeito que algúns signos de puntuación,

os díxitos e as palabras gramaticais con maior índice de frecuencia. Todos estes elementos complican o proceso de extracción léxica ben por seren repetitivos, ben por supoñeren casos de dislocación entre o texto inglés e mais o galego.

Para conseguir que un corpus paralelo etiquetado quede aliñado no nivel da palabra (isto é, para conseguir que se establezan no corpus as equivalencias léxicas de tradución) existen diversos algoritmos estatísticos que traballan con índices de atracción ou asociación léxica e cos datos da coaparición de elementos léxicos nas unidades de tradución bilingües aliñadas. No noso caso, empregamos a ferramenta informática para o aliñamento léxico *NATools* (Simões, 2008). Atopáranse máis detalles sobre proceso de extracción léxica bilingüe automática en Gomez Guinovart e Sacau Fontenla (2005). O resultado obtido da extracción automática é un dicionario bilingüe probabilístico ao que lle cómpre unha revisión manual para mellorar a súa precisión; trátase fundamentalmente de eliminar equivalencias erradas e de engadir outras que, malia estaren documentadas no Corpus CLUVI, non se recollen no dicionario probabilístico por distintas razóns de índole cuantitativo. Para identificar estas ausencias na listaxe inicial de equivalencias, comprobamos a presenza no dicionario de todas as formas léxicas presentes tres ou máis veces nos textos en inglés do corpus, e consultamos os leuarios e as equivalencias ofrecidas por outros repertorios léxicos bilingües de orientación elemental e escolar, como o dicionario bilingüe inglés-catalán da colección *Tutor* de Collins (Saeton e McCann, 2004) ou o dicionario inglés-galego de Xerais (Salgado, 1999).

Tras esta revisión, o leuario da segunda edición do *Dicionario CLUVI Inglés-Galego* recolle todos os lemas que aparecen un mínimo de tres veces no conxunto dos córpora dos que se partiu, ao carón dun conxunto de palabras que consideramos de interese malia posuír unha presenza menor nos textos. En total, esta segunda edición do dicionario contén 20.000 entradas documentadas no corpus, acompañadas de 30.000 traducións e 60.000 exemplos.

#### 4. Microestrutura do dicionario

Unha vez rematada a fase de selección dos lemas no corpus e de construción da macroestrutura, o proceso de edición do dicionario continuou pola súa microestrutura, é dicir, pola estrutura interna das entradas ou artigos lexicográficos. Deste xeito, engadíronse as categorías gramaticais ás que se pode adscribir cada lema; notas gramaticais e de uso tanto dos lemas ingleses como das traducións galegas, cando era o caso; un exemplo bilingüe para cada unha das traducións fornecidas con cadansúa referencia ao Corpus CLUVI; e, de ser relevante, información sobre os usos fraseolóxicos do lema, un apartado presente en 1 de cada 4 entradas. A información fraseolóxica, que acrecenta o valor deste dicionario, recolle as combinacións lexicalizadas do lema con outra ou outras palabras, acompañadas das súas posibles traducións exemplificadas documentadas no corpus.

O *Dicionario CLUVI Inglés-Galego* editouse integramente no formato XML, respectando unhas normas de codificación abondo sinxelas definidas na súa correspondente DTD (definición do tipo de documento):

```
<!ELEMENT dicionario (entrada+)>
<!ELEMENT entrada (lema, super_cat+)>
<!ELEMENT lema (#PCDATA)>
<!ELEMENT super_cat (categoria, acepcion+)>
<!ELEMENT categoria (#PCDATA)>
<!ELEMENT acepcion (plurilex?, traduccion, exemplo)>
<!ELEMENT traduccion (#PCDATA)>
<!ELEMENT plurilex (#PCDATA)>
<!ELEMENT exemplo (en, gl, fonte)>
<!ELEMENT en (#PCDATA)>
<!ELEMENT gl (#PCDATA)>
<!ELEMENT fonte (#PCDATA)>
```

Esta DTD lese do seguinte xeito: un documento de tipo *dicionario* está formado por un conxunto de entradas; cada entrada inclúe ademais do lema un conxunto

de informacións traditolóxicas agrupadas en función das posibles categorías gramaticais do lema. Cada un destes conxuntos (denominados *super\_cat* na DTD) pode conter unha ou máis acepcións, dependendo da polisemia de cada lema en cada categoría gramatical. A información agrupada en cada acepción inclúe a tradución en galego, un exemplo de uso documentado no corpus e, opcionalmente, a expresión plurilexématica (fraseolóxica) da que forma parte o lema na acepción. Por último, cada exemplo consta dun fragmento textual do Corpus CLUVI en inglés, a súa tradución ao galego, e a referencia da obra na que se documenta o exemplo. Deste xeito, cada entrada do dicionario pode incluír unha ou máis categorías gramaticais con unha ou máis traducións, sendo codificada internamente como se ilustra a seguir mediante un exemplo:

```
<entrada>
<lema>own</lema>
<super_cat><categoria>transitive verb</categoria>
<acepcion><traducion>ser dono</traducion>
<exemplo><en>One relic hast thou in thy treasury, handed down from the Moslems who
once @owned# Toledo --a box of sandal-wood containing a silken carpet: </en>
<gl>Tes unha reliquia no teu tesouro, herdada dos musulmáns que nalgún tempo @foron
donos# de Toledo... unha arqueta de pao de sándalo que contén unha alcatifa de
seda.</gl><fonte>ALH (778)</fonte></exemplo></acepcion>
<acepcion><traducion>posuír</traducion>
<exemplo><en>Kino and Juana came slowly down to the beach and to Kino's canoe,
which was the one thing of value he @owned# in the world. </en>
<gl>Kino e Juana baixaran de vagar ata a praia e achegáranse á canoa de Kino, que era
o único de valor que el @posuía# no mundo.</gl><fonte>PER
(239)</fonte></exemplo></acepcion>
<acepcion><plurilex>own up</plurilex>
<traducion>confesar</traducion>
<exemplo><en>But she did not mind @owning up# to it in the least; one must admit
that.</en>
<gl>Pero alomenos non lle importaba @confesalo#; iso había que recoñecelo.
</gl><fonte>CAR (1083)</fonte></exemplo></acepcion></super_cat>
<super_cat><categoria>adjective</categoria>
```

<acepcion><traducion>propio</traducion>

<exemplo><en>It was about this little kid that wouldn't let anybody look at his goldfish because he'd bought it with his @own# money. </en>

<gl>Falaba dun neno pequeno que non deixaba ve-lo seu peixe a ninguén porque o mercara co seu @propio# diñeiro.</gl><fonte>VIX (20)</fonte></exemplo></acepcion></super\_cat>

<super\_cat><categoria>pronoun</categoria>

<acepcion><plurilex>of one's own</plurilex><traducion>de seu</traducion>

<exemplo><en>He thought it very possible that Master Randolph's sister was a coquette; he was sure she had a spirit @of her own#; but in her bright, sweet, superficial little visage there was no mockery, no irony. </en>

<gl>Considerou que sería moi posible que a irmá de Master Randolph fose unha coqueta. Estaba seguro de que tiña un espírito @de seu#. Mais no seu pequeno, elegante, doce e superficial rostro non había farsa nin ironía ningunha.</gl><fonte>DAI (124)</fonte></exemplo></acepcion></super\_cat>

</entrada>

As entradas, codificadas en XML, son convertidas ao formato de impresión do dicionario mediante unha folla de estilo XSL. Así, o exemplo anterior da entrada de *own*, correspóndese coa seguinte presentación impresa:

**OWN** ▪ *transitive verb*

ser dono ► *One relic hast thou in thy treasury, handed down from the Moslems who once owned Toledo –a box of sandal-wood containing a silken carpet.* Tes unha reliquia no teu tesouro, herdada dos musulmáns que nalgún tempo foron donos de Toledo... unha arqueta de pao de sándalo que contén unha alcatifa de seda. [ALH]

posuír ► *Kino and Juana came slowly down to the beach and to Kino's canoe, which was the one thing of value he owned in the world.* Kino e Juana baixaran de vagar ata a praia e achegáranse á canoa de Kino, que era o único de valor que el posuía no mundo. [PER]

♦ **own up** confesar ► *But she did not mind owning up to it in the least; one must admit that.* Pero alomenos non lle importaba confesalo; iso había que recoñecelo. [CAR]

▪ *adjective*

propio ► *It was about this little kid that wouldn't let anybody look at his goldfish because he'd bought it with his own money.* Falaba dun neno pequeno que non deixaba ve-lo seu peixe a ninguén porque o

mercara co seu propio diñeiro. [VIX]

▪ *pronoun*

♦ **of one's own** de seu ► *He thought it very possible that Master Randolph's sister was a coquette; he was sure she had a spirit of her own; but in her bright, sweet, superficial little visage there was no mockery, no irony.* Considerou que sería moi posible que a irmá de Master Randolph fose unha coqueta. Estaba seguro de que tiña un espírito de seu. Mais no seu pequeno, elegante, doce e superficial rostro non había farsa nin ironía ningunha. [DAI]

Como se pode ver, a microestrutura do *Diccionario CLUVI Inglés-Galego* é semellante á doutros dicionarios bilingües, pero dásele máis importancia (e máis espazo) ás traducións dos lemas e aos exemplos de contextualización. Se analizamos detalladamente a estrutura interna dos artigos, podemos reparar en que os seus elementos constituíntes amosan certa variabilidade, reflexo da heteroxeneidade e das irregularidades da lingua inglesa. Así por exemplo, destácanse no diccionario os casos de lemas con flexión irregular, como no caso de *louse*:

**louse** ▪ *noun* ♪ Pl: *lice*

piollo ► *A louse crawled over the nape of his neck and, putting his thumb and forefinger deftly beneath his loose collar, he caught it.* Un piollo andaba a arrastrarse pola súa caluga, e metendo destramente o polgar e o furabolos no colar frouxo apañouno. [RET]

ou as diferenzas entre as formas tipicamente británicas e estadounidenses:

**plead** ▪ *intransitive verb* ♪ Este verbo é irregular en inglés americano, adquirindo a forma *pled* para o pasado e o participio. En inglés británico, simplemente engade *-ed*.

argumentar ► *He had long been patient, pleading with the sinful soul, giving it time to repent, sparing it yet awhile.* Durante moito tempo fora paciente, argumentando coa alma pecadora, dándolle ocasión de arrepentirse, autorizándolle un prazo máis aínda. [RET]

alegar ► *Martins felt he had got all he could from Crabbin, so he pleaded tiredness, a long day, promised to ring up in the morning, accepted ten pounds' worth of bafs for immediate expenses, and went to his room.* Martins pensou que xa obtivera de Crabbin todo o que podía, de maneira que se disculpou alegando que estaba canso, que fora un día moi longo e prometeu chamalo ao día seguinte; aceptou dez libras de vales para primeiros gastos e foise á habitación. [TER]

suplicar ► *Charles and Hal begged her to get off and walk, pleaded with her, entreated, the while she wept and importuned Heaven with a recital of their brutality.* Charles e Hal pregáronlle que descendese e que seguise a pé; suplicáronllo de xeonllos mentres ela, desfacéndose en bágoas, clamaba ós ceos coa descrición dos tormentos que tiña que aturar. [CHA]

### **jumper** (🇺🇸 sweater) ▪ *noun*

xerseí ► *A kind-faced woman in a hand-knitted jumper said wistfully, "Don't you agree, Mr. Dexter, that no one, no one has written about feelings so poetically as Virginia Woolf?"* Unha muller de rostro infantil que levaba un xerseí de calceta dixo ansiosa: – Señor Dexter, ¿non está de acordo comigo en que ninguén, repito, ninguén describiu os sentimentos dun modo tan poético coma Virginia Woolf? [TER]

saltador ► *He told Stephen that his name was Athy and that his father kept a lot of racehorses that were spiffing jumpers.* Explicoulle a Stephen que se chamaba Athy e que o seu pai tiña unha chea de cabalos de carreiras, excelentes saltadores. [RET]

### **humour** (🇺🇸 humor) ▪ *noun*

humor ► *And the funny thing was, I sort of missed him after we moved, because he had a helluva good sense of humour and we had a lot of fun sometimes.* E o caso foi que despois tiña morriña del, pois era un tío con sentido do humor e, ás veces, pasabámolo ben. [VIX]

### **lorry** (🇺🇸 truck) ▪ *noun*

camión ► *Loaded into trucks and vans, they are exported from one country to the next, even making it to the display windows of genuine bookshops –perhaps innocently, perhaps not.* Cargados en camións ou furgonetas, expórtanse dun país a outro e, o que é peor, chegan a invadi-los escaparates das librerías legais. ¿Inxenuidade ou complicidade dos libreiros? [C07]

Nestes e noutros casos nos que o lema amosa variantes formais, o dicionario conta con xogos de referencias cruzadas, de xeito que a busca por calquera das variantes vai obter sempre como resultado todas as formas posibles de cada lema.

Poden acompañar aos lemas, ademais, diversas notas sobre os seus usos en inglés, que serven para delimitar máis correctamente o sentido que se lle dá á palabra en cada caso:

**must** ▪ *modal verb* 💡 A diferenza entre *must* e *have to* radica en que o modal designa unha obriga imposta polo suxeito, unha obriga interna, mentres que *have to* expresa obrigas externas. Ex.: *I must finish my work today* > *Debo rematar o traballo hoxe* (é un deber que me propoño); *You have to drive on the right* > *hai que*

conducir pola dereita (obríganme as leis nacionais).

deber ► *Further, there is the belief that every one during life or after must go on pilgrimage to San Andrés de Teixido.* Aínda máis, crese que toda persoa, en vida ou despois dela, debe ir en romaría a San Andrés de Teixido. [GAL]

deber de ► *I shone my torch on him: he hadn't got a gun; he must have dropped it when my bullet hit him.* Enfoqueino coa lanterna: non levaba pistola, debeu de deixala caer cando lle disparei. [TER]

ter que ► *Allowing that my theory of the crime was the correct one, and I believe that it must be the correct one, then obviously the Wagon Lit conductor himself must be privy to the plot.* Supoñendo que a miña teoría sobre o crime sexa correcta, e eu coído que ten que ser correcta, é evidente que o revisor de Wagon Lit ten que ter parte no complot. [ASA]

haber que ► *Athy grinned and turned up the sleeves of his jacket, saying: It can't be helped; It must be done.* Athy fixo un aceno e arregazou as mangas da chaqueta dicindo: Non se pode evitar; Hai que facelo. [RET]

▪ *noun*

necesidade ► *"Decentralization is a must," said Dr Indra Djati Sidi, director-general of primary and secondary education, adding that educational programmes devolving more authority to districts for teacher training and construction of schools had actually started two years ago.* "Descentralizaré unha necesidade", declara Indra Djati Sidi, director xeral de ensino primario e secundario, precisando que a delegación nos distritos dunha maior autoridade en materia de formación de mestres e de construción de escolas se iniciara hai dous anos. [C12]

**gallon** ▪ *noun*

galón (4,55 litros) ► *A tanker pulls up to fill a 100-gallon water tank for \$6 to the owner who then sells a bucketful for five cents.* É tamén un camión cisterna que, logo do pagamento de 6 dólares, enche o tanque de 400 litros dun propietario que revende a auga a cinco cents o caldeiro. [C29]

O *Diccionario CLUVI Inglés-Galego* ofrece tamén información ortográfica, morfolóxica e de uso sobre as traducións galegas dos lemas. Infórmase, por exemplo, das mudanzas ortográficas da última reforma normativa do ano 2003. Téñase en conta que moitos dos textos do Corpus CLUVI dos que se tiran exemplos son anteriores a esta reforma, polo que as advertencias son sempre oportunas:

### **hassle** ▪ *noun*

conflito ✍ A forma ortográfica correcta deste termo é *conflito*. Na normativa anterior a 2003, a forma ortográfica correcta era *conflicto*, que é a que se recolle no exemplo. ► *By starting a hotline, some 50 Slovenian teenagers have become pros at listening, conversing and gently settling everyday hassles*. É o nome dunha liña aberta iniciada e animada por uns cincuenta adolescentes eslovenos. ¿Cal é a súa función? escoitar, dialogar e resolver con calma os conflictos de tódolos días. [C06]

### **service** ▪ *noun*

servizo ✍ A forma ortográfica correcta deste termo é *servizo*. Na normativa anterior a 2003, a forma ortográfica correcta era *servicio*, que é a que se recolle no exemplo. ► *He first elicited Michel's name and address, his length of service, and the length of time he had been on this particular route*. Solicítoulle en primeiro lugar que lle dixese o seu nome completo e enderezo, anos de servicio na compañía e cánto tempo levaba traballando naquela liña en concreto. [ASA]

misa ► *The sudden bursts of laughter and then one voice (Minta's) speaking alone, reminded her of men and boys crying out the Latin words of a service in some Roman Catholic cathedral*. As súbitas gargalladas e logo unha voz (a de Minta) falando soa lembráronlle os homes e nenos que entoan as palabras en latín dunha misa nunha catedral católica romana. [CAR]

Recóllense, do mesmo xeito, as indicacións da reforma normativa de 2003 sobre as variantes ortográficas admitidas:

### **impregnable** ▪ *adjective*

inexpugnable ~ inexpugnábel ► *Its heavenly walls totally impregnable, impervious but to the pious, who enter it submissively and with humility*. Os seus muros celestes son totalmente inexpugnables, agás para os piadosos, que entran con mansedume e humildade. [C28]

### **piracy** ▪ *noun*

pirataría ~ piratería ► *Furthermore, computer product manufacturers have developed "technical measures" to fight "piracy" and prevent copying (of software, databases and so on), which keeps users from enjoying their full rights under the principle of fair use*. De por parte, os fabricantes de produtos informáticos instauraron medidas técnicas para loitar contra da piratería e facer imposible a copia (de programas, bases de datos, etc.). Tales disposicións impídenlles ós usuarios exercer-lo seu lexítimo dereito ó uso xusto. [C25]

De grande axuda cremos que poden ser as advertencias sobre erros de lingua habituais, como algúns castelanismos moi estendidos:

## **fraud** ▪ *fraude*

fraude ✎ En galego este termo ten xénero feminino. ► *Excessive liberalization leads to private monopolies, capital flight, unpalatable levels of inequality and lack of trust in the market due to financial fraud.* A liberalización radical ten como resultado a aparición de monopolios privados, provoca fugas de capitais, desigualdades intolerables e perda de confianza no mercado por mor das fraudes financeiras. [C11]

Canto ao caudal léxico recollido neste dicionario, procurouse incluír, ademais do que poderíamos chamar léxico estándar ou común da lingua inglesa (e galega), outros grupos de palabras como a) palabras de creación recente ou con novos significados, procedentes sobre todo dos campos das tecnoloxías da información e algunhas disciplinas científicas:

## **hacker** ▪ *noun*

hacker ► *Other firms are offering anonymous or "pseudonymous" browsing tools, packages to control cookies –the strings of code that are planted on the user's computer by websites– and services to block hacker intrusion.* Outras empresas ofrecen instrumentos de navegación anónimos ou "pseudónimos", programas de control de cookies (delatores electrónicos) e servicios de protección contra os hackers ou intrusos informáticos. [C07]

pirata informático ► *Throughout the 1980s hackers took the symbolic role of the militant opposition.* Nos anos oitenta, os piratas informáticos simbolizaron a oposición militante. [C15]

b) nomes propios de persoa, país, cidade, ríos, mares, etc.; é dicir, artigo de contido máis enciclopédico ca lexicográfico:

## **Euclid** ▪ *proper noun*

Euclides ► *Greek thinkers such as Euclid, Ptolemy and Dionysius Thrax, respectively the "inventors" of geometry, map-making and grammar, worked there.* Nela traballaron sabios gregos coma Euclides, Claudio Ptolomeo e Dionisio Tracio, considerados como os "pais" respectivos da xeometría, da cartografía e da gramática. [C14]

## **Mauritius** ▪ *proper noun*

illa Mauricio ► *Researchers say abnormally warm conditions persisted in sea water for more than five months in 1998, causing extensive damage to corals around island nations including Seychelles, Mauritius, Maldives and Sri Lanka.* Os especialistas afirman que polo menos durante cinco meses dese

ano os mares e os océanos rexistraron temperaturas máis altas do normal, o que danou dun modo moi considerable os arrecifes que rodean as Seicheles, a illa Mauricio, as Maldivas e Sri Lanka. [C11]

### **Hague** ▪ *proper noun*

a Haia ► *But it was not until 1998 that 120 countries (out of 160 participants) meeting in Rome adopted a statute for a permanent International Criminal Court (ICC) to sit in The Hague (Netherlands).* Houbo que agardar ata xullo de 1988 para que 120 dos 160 países representados se reunisen en Roma para aproba-los estatutos dunha Corte Penal Internacional (CPI) que se establecerá na cidade holandesa da Haia. [C23]

### **Ganges** ▪ *proper noun*

Ganxes ► *In December 1996, recently elected governments in both India and Bangladesh decided to resolve decades of acrimony over the sharing of the waters from the Ganges, one of the most culturally and economically significant rivers on earth.* En decembro de 1996, os gobernos recentemente elixidos da India e Bangladesh decidiron superar anos de animosidade pola distribución das augas do Ganxes, un dos ríos con maior significación cultural e económica da Terra. [C28]

c) termos especializados que chegan á lingua común a través dos textos de divulgación científica:

### **mitochondrial** ▪ *adjective*

da mitocondria ► *One possible application of therapeutic cloning could be the treatment of women whose mitochondrial DNA –the genetic material that provides energy to the cell– is damaged, and therefore risks passing this defect on to their children.* Podería servir, por exemplo, para trata-las mulleres nas que o adn das mitocondrias (o material xenético que proporciona enerxía á célula) está danado, con risco de transmitir esta deficiencia ós seus fillos. [C18]

### **haemoglobin** ▪ *noun*

hemoglobina ► *Sickle-cell anaemia (an abnormal form of the red pigment of the blood, haemoglobin) affords some protection against the deadly form of the malaria parasite, Plasmodium falciparum.* Así mesmo, a anemia falciforme ou drepanocitose (unha enfermidade da hemoglobina do sangue) procura un certo grao de protección contra a malaria falciparum –forma mortal de paludismo. [C18]

Polo que respecta ao contido das entradas, cabe destacar, por último, a sección de fraseoloxía, que nalgúns casos, como no dos *phrasal verbs*, chega a ser moi

amplio. A entrada do verbo *go*, por exemplo, contén doce unidades deste tipo cun total de dezasete traducións:

**go** ▪ *transitive verb*

♦ **to go through** atravesar ► *She went through the line of brush when the moon was covered, and when it looked through she saw the glimmer of the great pearl in the path behind the rock. Atravesou a sebe cando a lúa se ocultou, e cando volveu a aparecer viu o brillo da gran perla no camiño detrás da pedra.* [PER]

pasar por ► *She did not like it that Jasper should shoot birds; but it was only a stage; they all went through stages. Non lle agradaba que Jasper matase paxaros; pero estaba no tempo, todos pasan por esas etapas.* [CAR]

revisar ► *And, even without any view to a formal study of the transcendental philosophy, great enlargement would be given to the understanding, by going through a Kantian dictionary, well explained, and well illustrated. E, mesmo sen vistas a un estudio formal da filosofía transcendental, daríase unha grande ampliación ó entendemento revisando un diccionario kantiano, ben explicado e ben ilustrado.* [LET]

♦ **to go past** pasar por diante de ► *Lily went past Mr Carmichael holding her brush to the edge of the lawn. Lily pasou por diante de señor Carmichael levando o pincel na man e foi ata o lindeiro do prado.* [CAR]

▪ *intransitive verb*

ir ► *He would go straight to Mrs Ramsay, because he felt somehow that she was the person who had made him do it. Iría dereito xunta a señora Ramsay porque dalgún xeito intuíu que fora ela quen llo fixera levar a cabo.* [CAR]

marchar ► *I wouldn't go away from here without having seen that old castle. Non marcharía de aquí sen ve-lo tal castelo.* [DAI]

chegar ► *My plan was to go as far as possible that night, and then, building a fire, to sleep in the protection of its glare. O meu plano era chegar tan lonxe como me fose posible durante aquela noite e, entón, prendería unha fogueira para durmir ó abeiro da súa luz.* [TEM]

pasar ► *They had nothing to say, but something seemed, nevertheless, to go from him to her. Non tiñan nada que dicir, pero, así e todo, algo semellaba pasar del a ela.* [CAR]

quedar ► *I should think she'd rather go indoors. Pensaba que prefería quedar na casa.* [DAI]

♦ **go across** cruzar ► *Presently a granite bridle-path goes across towards Lage through a splendid country of moor and pines, with a dozen tors of poised blocks of granite, purple and sheer, looking twice their real height against the clear turquoise sky. Un pouco máis alá un camiño de ferradura vai cruzando cara a Laxe a través dunha espléndida paisaxe de breixos, piñeiros e unha ducia de cotos con elegantes*

penedos ergueitos contra o ceo azul turquesa que semellan te-lo dobre da súa altura. [GAL]

♦ **go away** liscar ► *And I was outside the shelter by myself in the dark and the twisty things had gone away.* E estaba soño fóra das cabanas e esas cousas retortas xa liscaran. [SEN]

♦ **go back** volver ► *It would be best now, I think, if you were to go back to your post.* E agora, creo que o mellor sería que volvese ó seu posto. [ASA]

♦ **go by** pasar ► *Horned toads watched the family go by and turned their little pivoting dragon heads.* Os lagartos cornudos ollaban pasa-la familia e retorcían as súas pequenas cabezas xiratorias de dragón. [PER]

♦ **go down** baixar ► *So she must go down and begin dinner and wait.* Daquela debería baixar, deixar que principiase a cea e agardar. [CAR]

♦ **go in** entrar ► *I don't wish to go in, and I don't intend to.* Non desexo entrar e non teño ningunha intención de facelo. [ESP]

♦ **go off** sair ► *And Minta walked on ahead, and presumably Paul met her and she went off with Paul in the garden.* E Minta seguía camiñando en cabeza, presumiblemente Paul se ía atopar con ela, e logo sairían ó xardín. [CAR]

marchar ► *Without saying a word, the only token of her errand a basket on her arm, she went off to the town, to the poor, to sit in some stuffy little bedroom.* Sen dicir unha palabra, sendo un queipo que levaba no brazo o único sinal da súa ocupación, marchaba á vila, onda os pobres, a sentar nun cuarto pequeno e mal ventilado. [CAR]

♦ **go on** seguir ► *Anyhow, said Lily, tossing off her little insincerity, she would always go on painting, because it interested her.* De calquera xeito, dixo Lily, desfacéndose da súa lixeira insinceridade, ela quería seguir pintando porque lle interesaba. [CAR]

continuar ► *They both felt uncomfortable, as if they did not know whether to go on or go back.* Ámbolos dous se sentiron incómodos, coma se non soubesen se continuar ou se deixalo. [CAR]

♦ **go out** sair ► *The girl came back, making signs to him to be quick and go out quietly by the back.* A rapaza volveu acenándolle para que se dese présa e saíse pola parte de atrás caladamente. [RET]

apagarse ► *Then the light burned my fingers and fell out of my hand, going out as it dropped, and when I had lit another the little monster had disappeared.* Entón, o misto queimoume os dedos e caeume da man, apagándose mentres caía; e, cando acendín outro, o pequeno monstro desaparecera. [TEM]

♦ **go up** subir ► *"When they go up," he said, "we will slip away, down to the lowlands again."* -Cando eles suban -dixo-, nós baixaremos outra vez á chaira. [PER]

## 5. Interface de consulta

Por último, para realizar as buscas no dicionario, a interface web de consulta en Internet, no enderezo <http://sli.uvigo.es/diccionario>, ofrece unha caixa onde

podemos escribir o termo ou a cadea de caracteres (o fragmento de palabra) que queremos buscar, xunto cunha serie de botóns para indicar a dirección da busca (inglés-galego ou galego-inglés) e se desexamos buscar palabras completas ou todas as palabras que comece polo cadea indicada:

haemo	<input checked="" type="radio"/> En inglés	<input checked="" type="radio"/> Comeza con
Buscar	Borrar	<input type="radio"/> En galego <input type="radio"/> Palabra exacta

Como primeiro resultado da consulta, obteremos unha listaxe de todos os lemas que comezan coa cadea introducida na caixa de texto; a partir desta lista, podemos ir aos artigos completos correspondentes:

Resultado da procura:

- + **haemoglobin**, hemoglobina ... >>>
- + **haemophilia**, hemofilia ... >>>
- + **haemophiliac**, hemofílico ... >>>
- + **haemorrhagic**, hemorráxico ... >>>

Establecendo a dirección de consulta galego-inglés, obteremos a listaxe das entradas que conteñen entre as traducións a palabra galega indicada. Por exemplo, a busca de "neno", daríanos como resultado:

Resultado da procura:

- + **babe**, meniño ... >>>
- + **baby**, neno ... >>>
- + **boy**, rapaz ... >>>
- + **child**, fillo ... >>>
- + **infant**, bebé ... >>>
- + **kid**, neno ... >>>
- + **preppy**, neno ben ... >>>
- + **stillborn**, neno nacido morto ... >>>
- + **whizzkid**, neno prodixio ... >>>

Deste xeito, aínda que as entradas do dicionario están redactadas só na dirección de tradución inglés-galego, a súa consulta informatizada permite tamén o seu uso a partir do galego, converténdose así tamén nun dicionario galego-inglés apto para

a tradución inversa.

## 6. Conclusións

Neste artigo presentamos o proceso que seguimos para a creación da segunda edición (de 2008) do *Diccionario CLUVI Inglés-Galego*, baseándonos nos datos fornecidos polo Corpus CLUVI de traducións. Desde o Seminario de Lingüística Informática da Universidade de Vigo agardamos que sexa unha ferramenta útil tanto para estudantes e usuarios de inglés de todos os niveis, como para os profesionais da tradución e da interpretación, que ata agora tiñan que recorrer a “dicionarios ponte” inglés-portugués ou inglés-castelán. Agardamos tamén todas as críticas e suxestións que os usuarios nos fagan chegar e que sen dúbida contribuirán a mellorar a súa calidade en futuras edicións.

## 7. Bibliografía

Bosque, Ignacio (2004). *Diccionario Redes: Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM.

Gómez Guinovart, Xavier (coord.), Alberto Álvarez Lugrís e Eva Díaz Rodríguez (2008). *Diccionario CLUVI Inglés-Galego*, 2ª edición. <<http://sli.uvigo.es/diccionario/>>.

Gómez Guinovart, Xavier (dir.) (2003-). *Corpus CLUVI (Corpus Lingüístico da Universidade de Vigo)*. Vigo: Universidade de Vigo. <<http://sli.uvigo.es/CLUVI/>>.

Gómez Guinovart, Xavier e Elena Sacau Fontenla (2005): "Técnicas para o desenvolvemento de dicionarios de tradución a partir de córpora aplicadas na xeración do Diccionario CLUVI Inglés-Galego". *Viceversa: Revista Galega de Traducción*, 11, pp. 159-171.

Rafel, Joaquim (dir.) (1997). *Corpus Textual Informatizat de la Llengua Catalana*. Barcelona: Institut d'Estudis Catalans. <<http://ctilc.iec.cat/>>.

Saeton, Maggie e Carol McCann (coord.) (2004). *Collins Tutor English-Catalan Català-Anglès. Diccionari bilingüe per al nivell inicial*. Glasgow: HarperCollins Publishers.

Salgado, Benigno (1999). *Diccionario elemental inglés-galego galego-inglés*. Vigo:

Edicións Xerais de Galicia.

Sánchez, Aquilino (dir.) (2001). *Gran diccionario de uso del español basado en el Corpus lingüístico CUMBRE*. Madrid: Sociedad General Española de Librería.

Santamarina Fernández, Antón (dir.) (2003-). *Tesouro informatizado da lingua galega (TILG)*. Santiago de Compostela: Universidade de Santiago de Compostela. <<http://www.ti.usc.es/TILG/>>.

Simões, Alberto Manuel Brandão (2008). *Extracção de recursos de tradução com base em dicionários probabilísticos de tradução*. Tese de doutoramento. Braga: Universidade do Minho.

Sinclair, John (ed.) (1987). *Collins Cobuild English Language Dictionary*. 1ª edición. Londres: Collins.