

NEOLOGÍA INFORMÁTICA DEL GALLEGO Y EQUIVALENCIAS TERMINOLÓGICAS EN EL CORPUS XIGA/SLI

XAVIER GÓMEZ GUINOVART, ANXO M. LORENZO SUÁREZ
Universidade de Vigo

RESUMEN. *Este trabajo examina las diversas modalidades de coexistencia de términos equivalentes a través del estudio del uso real de la lengua documentado en el corpus Xiga/SLI (compilado en el Seminario de Lingüística Informática de la Universidade de Vigo) de textos sobre informática en gallego, agrupando los ejemplos analizados en función de las características de su distribución libre, complementaria o mixta entre las diferentes categorías textuales del corpus analizado.*

PALABRAS CLAVE: *Corpus técnicos, lenguaje de la informática, términos equivalentes, neología.*

ABSTRACT. *This paper analyses the various ways in which equivalent terms can occur in real texts through the study of a Galician technical corpus on computing (Xiga/SLI Corpus) built at the Vigo University Computational Linguistics Group. We propose and describe a classification of the terminological distribution observed in the corpus into three categories: free distribution, complementary distribution, and mixed distribution.*

KEYWORDS: *Technical corpora, language of computing, equivalent terms, neology.*

1. INTRODUCCIÓN

El corpus Xiga/SLI, compilado en el Seminario de Lingüística Informática¹ (SLI) de la Universidade de Vigo, es un corpus textual de gallego que recoge material escrito de diversos ámbitos y registros en el campo de la informática, incluyendo artículos periodísticos, manuales y textos de ayuda, menús y mensajes de programas, discusiones y noticias (*news*) en grupos de Internet, y textos académicos, escolares e informativos. La compilación de este corpus por parte del SLI tiene como objetivo principal su utilización como material de referencia y estudio en diversas actividades académicas de investigación que se están llevando a cabo en estos momentos (Gómez Guinovart y Lorenzo Suárez 1994, Lorenzo Suárez y Gómez Guinovart 1996, Gómez Guinovart 1999), y dentro de las cuales se enmarca la elaboración de un diccionario actualizado de informática en gallego. Dentro de este proyecto terminológico, el corpus Xiga/SLI está destinado a constituirse en una herramienta de consulta imprescindible para conocer el uso real del gallego escrito en la informática y, a partir de esta constatación, poder realizar propuestas terminológicas normativas ajustadas a la realidad lingüística plasmada en los textos. Por otra parte, el corpus Xiga/SLI, al contener el conjunto de términos que hoy están en circulación en el ámbito de la informática, va a permitir realizar investigaciones lingüísticas fundamentadas sobre el uso social de las

propuestas terminológicas, permitiendo observar de manera empírica la dirección de la selección léxica, es decir, las vacilaciones y preferencias actuales en la selección de términos informáticos en la lengua gallega escrita contemporánea.

2. COMPOSICIÓN DEL CORPUS

La versión del Xiga/SLI que empleamos en este trabajo está formada por 5 categorías de textos que conjuntamente totalizan más de 400.000 palabras. A continuación presentamos un sumario del material analizado en cada categoría y, en las Tablas 1-6, la tabulación del cómputo de líneas, palabras y caracteres proporcionado por la utilidad *wc* (*word count*) de Linux, en orden de mayor a menor número de palabras en el corpus de estudio, primero desglosado por categorías (Tablas 1-5) y al final en conjunto (Tabla 6 y Gráfico 1). Véanse en el Apéndice las referencias bibliográficas correspondientes.

X – Prensa escrita

x1	12919	151987	993006
x2	6241	40288	277913
	19160	192275	1270919

Tabla 1

M – Manuales y textos de ayuda

m1-t	6609	24074	159402
m2-t	3353	15773	105463
m3-f	1883	12029	85418
m4-x	2727	9274	78650
m5	581	1303	15020
m6	547	6021	39724
m7-x	925	2301	22357
m8-x	212	2668	17540
m9-x	1134	3151	34747
	17971	76594	558321

Tabla 2

G – Discusiones y noticias (*news*) en Internet

g1	8676	69119	441588
g2	1099	6599	45345
	9775	75718	486933

Tabla 3

T – Textos académicos, didácticos e informativos

t1	1091	9045	63922
t2	258	1774	13198
t3	275	6289	42865
t4	1703	6750	52755
t5-f	595	5050	33936
	3922	28908	206676

Tabla 4

P – Menús y mensajes de programas

p1-t	3808	16202	114435
p2-x	1151	4602	34736
p3-x	401	1407	9728
p4-x	395	1298	9090
p5-x	341	826	6578
p6-x	602	1766	12850
p7-x	188	425	3191
p8-x	171	528	3991
	7057	27054	194599

Tabla 5

	<i>L</i>	<i>P</i>	<i>C</i>
X	19160	192275	1270919
M	17971	76594	558321
G	9775	75718	486933
T	3922	28908	206676
P	7057	27054	194599
Total	57885	400549	2717448

Tabla 6

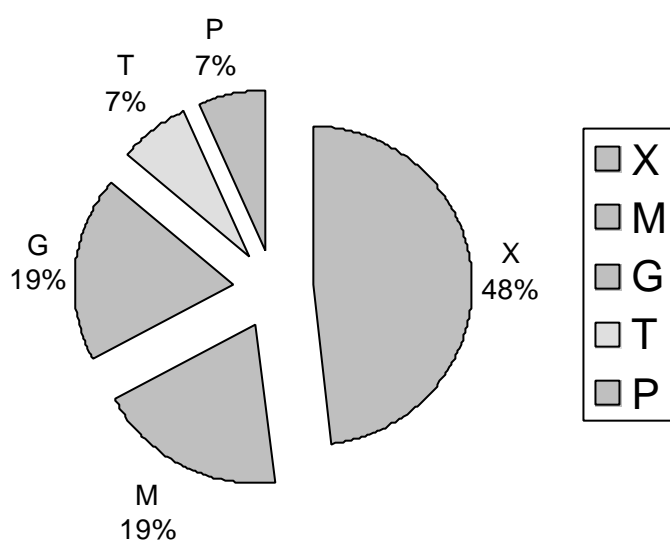


Gráfico 1

3. ANÁLISIS DE LA EQUIVALENCIA TERMINOLÓGICA EN EL CORPUS

En este trabajo analizaremos la plasmación en el corpus Xiga/SLI de un conjunto de ejemplos significativos de diversas posibilidades de selección léxica de términos equivalentes por parte de la comunidad lingüística usuaria. A efectos expositivos, agruparemos los ejemplos analizados en función de las características de su distribución entre las diferentes categorías textuales del corpus examinado, y emplearemos para el análisis el programa TACT de análisis textual (Pérez Guerra 1998).

3.1. Distribución complementaria

Dado un conjunto de términos sinónimos o equivalentes, diremos que sus miembros están en distribución complementaria si se cumple que cada miembro de este conjunto es utilizado por un mismo grupo de textos, y se cumple también que cada grupo de textos utiliza un único término de este conjunto. Aunque esta distribución complementaria no se puede observar en el corpus Xiga/SLI de manera estricta utilizando la división en categorías temáticas del corpus aquí presentada, sí que es posible observar algunos ejemplos muy característicos de este tipo de uso, como ilustramos a continuación con los equivalentes *fiestra/xanela/ventá* y *icona/icono*.

3.1.1. *fiestra/xanela/ventá*

La frecuencia constatada en el corpus de estos tres términos informáticos equivalentes es de 310 menciones de *fiestra* (194 veces en singular y 116 en plural), 190 de *xanela* (112+88) y 56 de *ventá* (48+8). Los cálculos crudos, sin embargo, no reflejan una distribución altamente excluyente (ver Figuras 1-3, donde se señalan con un asterisco los casos de distribución no complementaria) que parece indicar una elección consciente entre los términos contendientes. Así, 3 subcorpus sólo emplean *fiestra*, mientras que 6 sólo utilizan *xanela* y 4 se limitan a *ventá*². La alternancia efectiva entre estos términos sólo se manifiesta claramente en [M4-X] (12 usos de *fiestra* frente a 21 de *xanela*) y, con mucha menos fuerza, en [P2-X] (1 aparición de *fiestra* frente a 6 de *ventá*).

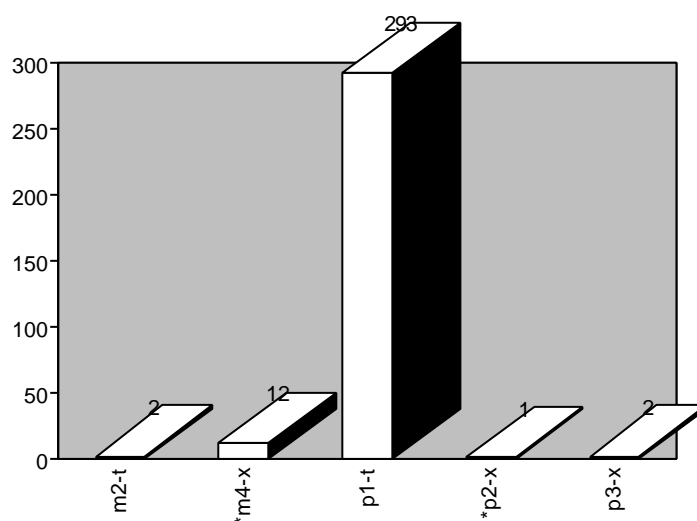


Figura 1. Distribución de *fiestra*.

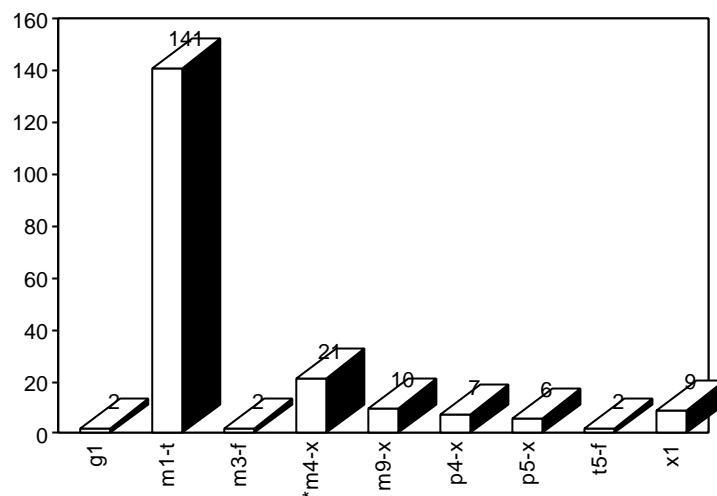


Figura 2. Distribución de *xanela*.

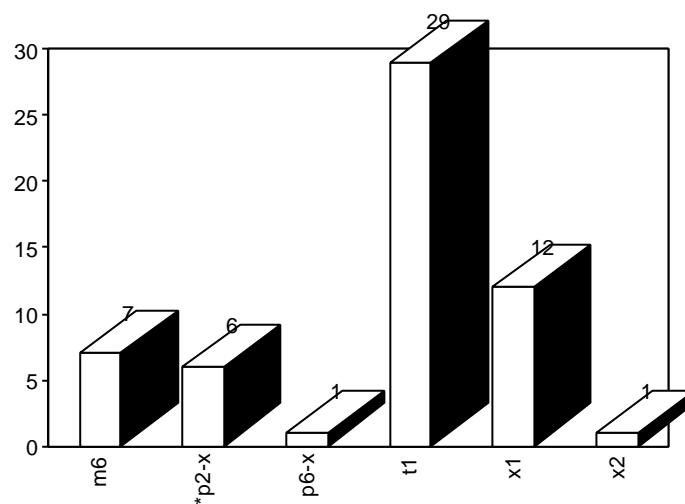


Figura 3. Distribución de *ventá*

3.1.2. *icona/icono*

De esta pareja de términos equivalentes, el empleado más veces y en más fuentes es el masculino castellanizante *icono*, con 110 (62+48) menciones en el corpus, a pesar de ser el femenino *icona*, con sólo 54 (33+21) menciones, la forma normativa establecida por la Real Academia Galega (1997). Su distribución en el corpus (ver Figuras 4-5) sería totalmente

complementaria de no darse la alternancia de [P1-T], posiblemente debida a que los diferentes programas que componen este subcorpus están traducidos por diversas personas con distintos hábitos de selección terminológica, y la falsa apariencia de alternancia que se observa en [M6], donde los ejemplos de *icono* en masculino aparecen precisamente como ilustraciones de los errores ortográficos que es capaz de subsanar el programa corrector.

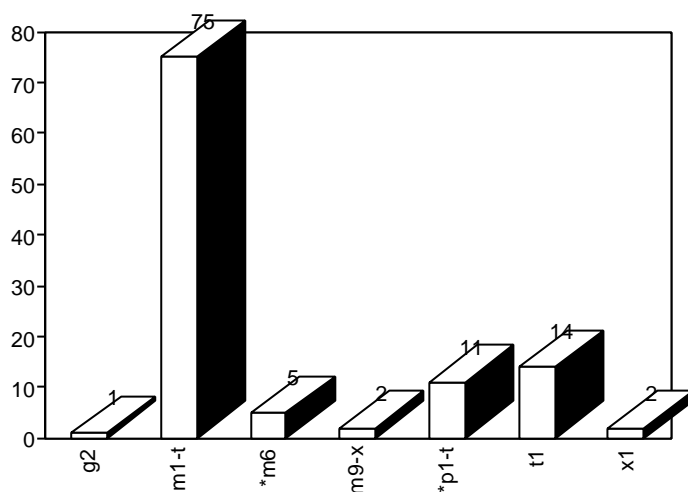


Figura 4. Distribución de *icono*

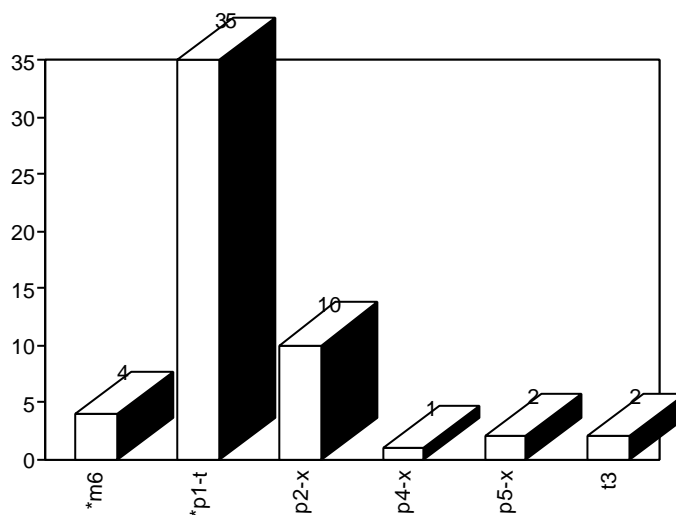


Figura 5. Distribución de *icona*

Al lado de esta pareja, se encuentran en el corpus otros dos términos equivalentes mucho más minoritarios en frecuencia y dispersión: el lusismo *icone*, con 12 (11+1)

menciones localizadas todas en [M3-F], y la forma idiosincrásica con desplazamiento acentual *ícono*, citada 6 veces en [M4-X].

3.2. *Distribución libre y mixta*

El concepto de ‘distribución libre’ alude al uso concurrente de los diferentes miembros de un conjunto de términos equivalentes por parte de todos los textos del corpus. Como sucede con la distribución complementaria, tampoco es posible observar la distribución libre en el corpus Xiga/SLI de manera estricta. Sin embargo, se pueden observar en el corpus algunos casos que ilustran adecuadamente este uso, como ejemplificaremos con los equivalentes *ficheiro/arquivo*; y un gran número de casos de distribución irregular o mixta que comparten características de la distribución complementaria y de la libre, como ilustraremos mediante los equivalentes *endereço/dirección*.

3.2.1. *ficheiro/arquivo*

El término de mayor frecuencia de esta pareja es *ficheiro*, empleado 1.006 (516+490) veces en el corpus, mientras que *arquivo* es utilizado sólo 551 (309+242). Como contraste, la dispersión de *arquivo* (23 subcorpus) es mayor que la de *ficheiro* (17). En cuanto a su distribución, constituye un ejemplo bastante ilustrativo de distribución libre, ya que la selección léxica es variable en 15 subcorpus, lo que contrasta con el uso exclusivo de *ficheiro* en [P1-T] y [T3], y el de *arquivo* en 8 subcorpus (ver Figuras 6-7, donde se destacan con un asterisco los casos de distribución complementaria).

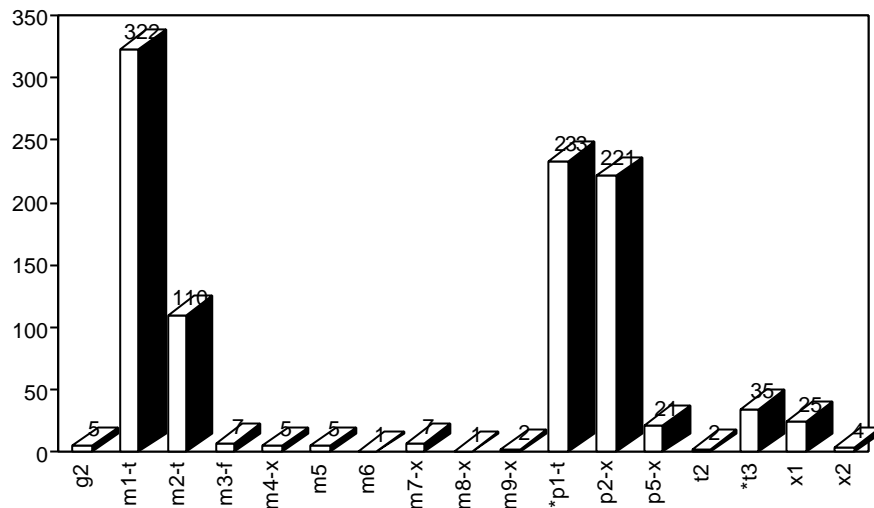


Figura 6. Distribución de *ficheiro*

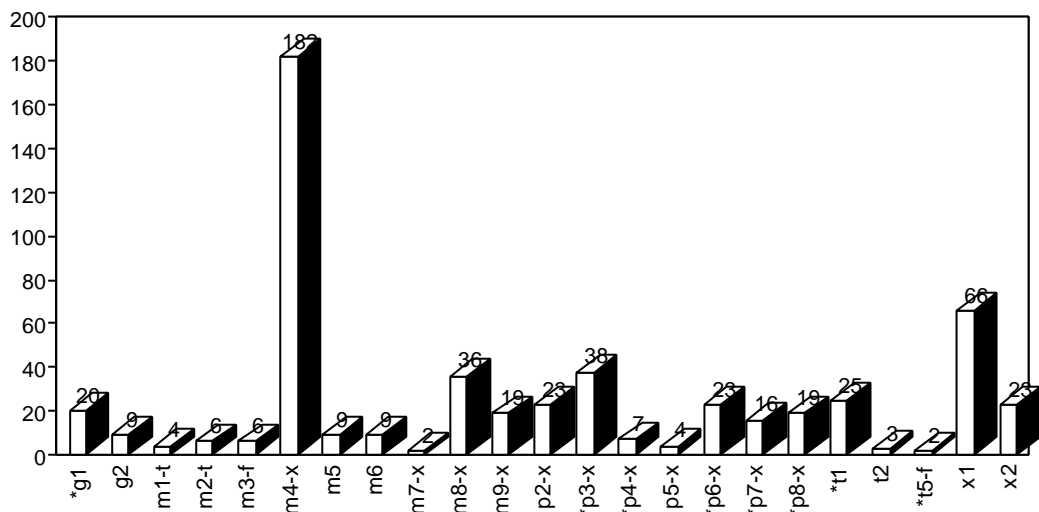


Figura 7. Distribución de *arquivo*

3.2.2. *endereço/dirección*

El término *endereço* es más frecuente en el corpus que su equivalente *dirección*, contando el primero con 345 (250+95) menciones frente a las 118 (86+32) del segundo (entre las que incluimos la propuesta ortográfica alternativa *dirección*, mencionada una vez en [T4]). Además, se constata que el primero posee un uso más específico que el segundo. Casi todas las menciones de *endereço* en los diversos subcorpus (con la única excepción de [X1], el menos técnico) representan el uso de este término en informática: “endereço electrónico”,

“enderezado de páginas”, etc. En cambio, el uso de *dirección* en el corpus parece manifestar una mayor tendencia a aparecer en acepciones ajenas a la terminología informática: así, “dirección de empresa” o “dirección oposta” frente a “picar na dirección” o “direcciones de correo”.

Enderezo y *dirección* se encuentran repartidos en un número semejante de subcorpus (respectivamente, 14 y 15). En cuanto a su distribución, se observa en 8 subcorpus la distribución libre, mientras que en 13 es complementaria favorable a *enderezo* (en 6) o a *dirección* (en 7) (ver Figuras 8-9).

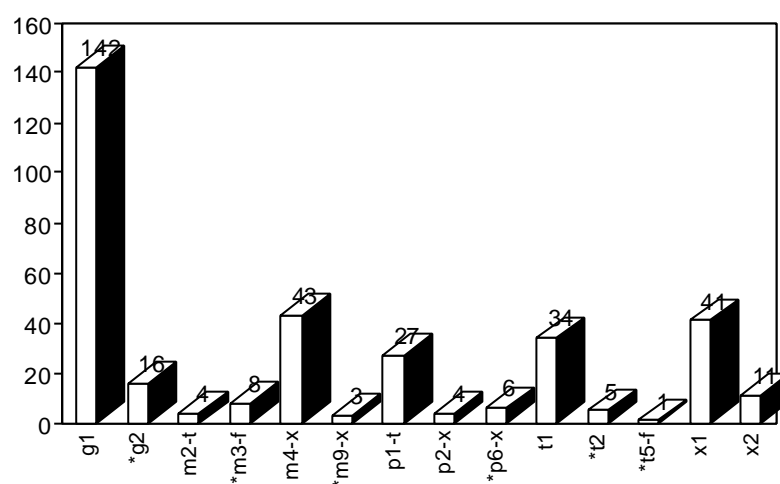


Figura 8. Distribución de *enderezo*

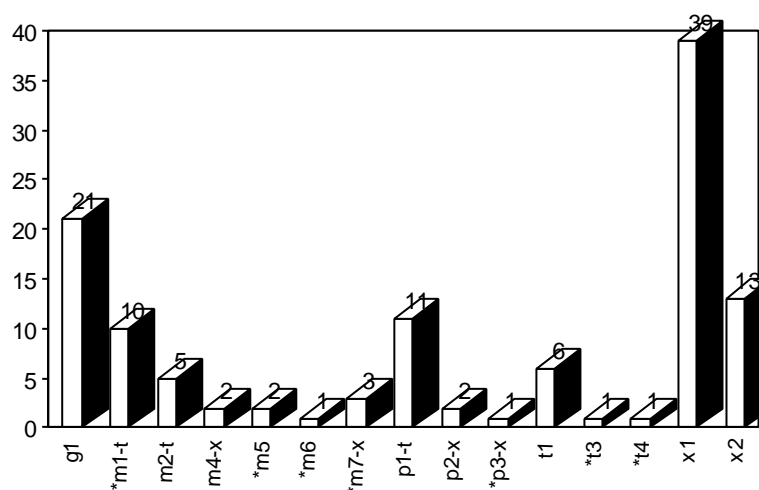


Figura 9. Distribución de *dirección*

4. CONCLUSIONES

En el análisis que presentamos en este trabajo, examinamos las diversas modalidades de coexistencia de términos equivalentes, a través del estudio del uso real de la lengua documentado en el corpus Xiga/SLI de textos sobre informática en gallego. De la observación minuciosa del corpus (de la que sólo hemos puesto algunos ejemplos, debido a las limitaciones de espacio propias de este tipo de trabajos), tratamos de inferir la dirección de la selección léxica de equivalentes emprendida por el gallego y así fundamentar en este uso social nuestra labor terminológica en el campo de la informática.

Un examen del contenido lingüístico de este corpus nos lleva a afirmar, con todas las reservas y prudencia necesarias, que la inestabilidad y la fluctuación son dos rasgos característicos que definen, en el momento actual, el estado de la terminología informática en gallego. La falta de tradición en la utilización de la lengua gallega en este ámbito del conocimiento facilita la extensión de formas alternantes para expresar los mismos objetos o procesos, un fenómeno al que también contribuye la utilización de opciones gráficas alternativas a la normativa oficial.

APÉNDICE: TEXTOS ANALIZADOS

[X1] Texto íntegro de 40 ejemplares de *Océano Internet*, suplemento semanal del diario *O Correo Galego*, publicados entre el 22 de enero y el 25 de noviembre de 2000.

[X2] Texto íntegro de 8 ejemplares de *Díxitos*, revista trimestral del Centro de Supercomputación de Galicia, publicados entre octubre de 1998 y octubre de 2000.

[G1] 385 mensajes datados entre el 22 de mayo y el 27 de noviembre de 2000 producidas en el seno del grupo de discusión de la Ciberirmandade da Fala en el servidor Egroups.

[G2] 18 mensajes sobre temas informáticos producidos en el seno del grupo de noticias de soc.culture.galiza.

[M1-T] Manual del escritorio de trabajo GNOME, traducido del inglés dentro del Proyecto Trasno³ de traducción del sistema operativo Linux al gallego.

[M2-T] 6 documentos de ayuda, del tipo denominado 'mini-HOWTO' en el Proyecto de Documentación de Linux o LDP (Proyecto Trasno).

[M3-F] Documento de ayuda de Miguel Conde titulado “Todo o que sempre quixo saber e nunca ousou preguntar sobre as novas”, versión 1.7 (14/01/98), traducida al gallego por Ramón Flores⁴.

[M4-X] Manual del gestor de transferencias de ficheros Net Vampire, distribuido a través del Proxecto Xis⁵ de traducción de programas para Windows al gallego.

[M5] 4 textos de ayuda sobre Linux extraídos de la revista electrónica *Galinux*⁶.

[M6] Textos de ayuda de 4 aplicaciones informáticas para Windows desarrolladas por la empresa Imaxin⁷.

[M7-X] Manual de GIIRC, utilidad para el programa de diálogo en línea mIRC, realizada por Rami Rodríguez Suárez y distribuida a través del Proxecto Xis.

[M8-X] Manual de RegCleaner, programa administrador del registro de Windows, traducido por Galego21 y distribuida a través del Proxecto Xis.

[M9-X] Textos de información en las páginas web del Proxecto Xis sobre las características y procedimientos de instalación de los programas distribuidos.

[P1-T] Traducción de los menús y mensajes de programas de 14 programas para Linux (Proxecto Trasno).

[P2-X] Traducción del programa Windows Commander (Proxecto Xis).

[P3-X] Traducción del programa CD-DA Easy CD (Proxecto Xis).

[P4-X] Traducción del programa Radiator (Proxecto Xis).

[P5-X] Traducción del programa Context (Proxecto Xis).

[P6-X] Traducción del programa WebZip (Proxecto Xis).

[P7-X] Traducción del programa Install Maker (Proxecto Xis).

[P8-X] Traducción del programa RegCleaner (Proxecto Xis).

[T1] Ramos, J. M. 2000. “Internet básico”. [Documento de Internet disponible en <http://members.es.tripod.de/joseramos7/rede.htm>]

[T2] Andión, N. 1998. “Introducción a Internet”. [Documento de Internet disponible en <http://www.uvigo.es/webs/sli/virtual/introga.htm>]

[T3] Magán, F. 1996. “Estándares de representación da información textual e multimedia”. *Lingüística e informática*. Eds. X. Gómez Guinovart y A. M. Lorenzo Suárez. Compostela: Tórculo. 153-186.

[T4] Plataforma polo Galego na Informática. 1995. “Informe: estado actual do galego na informática”. [Documento de Internet disponible en http://www.geocities.com/CollegePark/Library/4280/informe_1995.html]

[T5-F] Flores, R. 2000. “Computadores galegófonos”. [Documento de Internet disponible en <http://mx4.xoom.com/ramonflores.1/PARADISO/Computadores-galegofonos.html>]

NOTAS

1. <http://www.uvigo.es/webs/sli/>
2. También se documentan en el Xiga/SLI 9 menciones de *xanela* en [X1] y 2 en [G1] que no se corresponden con el término aquí analizado, sino que aluden a la denominación de un catálogo de recursos de Portugal y norte de Galicia disponible en Internet (referido como “www.xanela.com”).
3. <http://trasno.gpul.org/>.
4. <http://www.simil.com/scg/newsfaq.htm>.
5. <http://www.galego21.org/xis/>.
6. <http://www.geocities.com/galinux2000/>.
7. <http://www.imaxin.com>.

BIBLIOGRAFÍA

- Gómez Guinovart, X. 1999. “A neoloxía terminolóxica informática en castelán, catalán e galego”. *Viceversa. Revista Galega de Traducción* 5: 75-83.
- Gómez Guinovart, X. y A. M. Lorenzo Suárez. 1994. *Vocabulario de informática (galego, inglés, castelán)*. Vigo: Servicio de Normalización Lingüística da Universidade de Vigo.
- Lorenzo Suárez, A. M. y X. Gómez Guinovart. 1996. “Terminoloxía, informática e lingua galega”. *Cadernos de Lingua* 13: 5-33.
- Pérez Guerra, J. 1998. *Análisis computarizado de textos: una introducción a TACT*. Vigo: Universidade de Vigo (Servicio de Publicacións).
- Real Academia Galega. 1997. *Diccionario da Real Academia Galega*. A Coruña: Real Academia Galega.
- Rodríguez Río, X. 1996. “Os traballos terminolóxicos en lingua galega: unha aproximación á súa situación e ás súas necesidades”. *Cadernos de Lingua* 13: 35-74.