Tema 5: Corpus electrónico

TECNOLOGÍAS DE LA TRADUCCIÓN

¿Qué es un corpus?

- Conjunto de textos en formato electrónico que han sido seleccionados y clasificados mediante determinados criterios lingüísticos y que se utiliza como muestra de una lengua o una variante lingüística.

Tipos de corpus

- Según la modalidad de la lengua:
  - corpus escrito
  - corpus oral
    - transcripciones de grabaciones/grabaciones y transcripciones fonéticas
- Según el número de lenguas:
  - Monolingüe
  - Bilingüe/multilingüe

Tipos de corpus II

- Según la relación entre las lenguas:
  - Comparable
  - Paralelo
    - Alineados
    - No alineados
- Según la especificidad de los textos:
  - General
    - Corpus general de referencia
  - Especializado

Tipos de corpus III

- Según aspectos cronológicos:
  - Sincrónico
  - Diacrónico
- Según el proceso de creación de corpus:
  - Abierto
  - Cerrado

Tipos de corpus IV

- Según la información que se añada al corpus:
  - Corpus simple o en bruto
  - Corpus anotado
    - Información extratextual (referencia bibliográfica, fecha de inclusión, autor, etc.)
    - Información textual (títulos, capítulos, intervenciones de un hablante, etc.)
    - Etiquetado semántico, morfológico y sintáctico

Etiquetado de corpus

1	El	el	det>2	@PREMOD DET MSC SG
2	mineral	mineral	subj>8	@NH N MSC SG
3	de	de	pm>4	@POSTMOD PREP
4	arcilla	arcilla	mod>2	@NH N FEM SG
5	más	mucho	ad>6	@PREMOD ADV CMP
6	puro	puro	ads>2	@POSTMOD A MSC SG
7	,	,		
8	es	ser	main>0	@MAIN V IND PRES SG P3
9	decir	decir	comp>8	@MAIN V INF
10	,	,		
11	caolinita	caolinita		@NH Heur N FEM SG
12	no	no	ad>13	@PREMOD ADV
13	desordenada	desordenado	ads>11	@POSTMOD A FEM SG
14	tal	tal	ads>13	@POSTMOD A UTR SG
15	como	como	pm>17	@PREMARK CS
16	la	la	obj>17	@NH PRON Pers FEM SG ACC
17	presente	presentar	man>9	@MAIN V SUB PRES SG P3
18	en	en	pm>20	@PREMARK PREP
19	la	la	det>20	@PREMOD DET FEM SG
20	arcilla	arcilla	loc>17	@NH N FEM SG
21	china	chino	ads>20	@POSTMOD A FEM SG

Corpus para la traducción

- Obtener ejemplos de uso real de la lengua.
- Adquirir conocimiento conceptual sobre un campo de especialidad.
- Adquirir conocimiento lingüístico: terminología, colocaciones, equivalentes, etc.
- Obtener información sobre frecuencia de uso, distribución, etc.

Corpus en línea

- Español
  - CREA y CORDE (RAE)
  - Corpus del Español (Mark Davies)
  - http://www.corpusdelespanol.org/
  - Bwananet (IULA, UPF) (español, catalán e inglés)
- Catalán
  - Corpus Textual Informatitzat de la Llengua Catalana
  - http://ctilc.iec.cat/
- Inglés
  - British National Corpus
  - http://www.natcorp.ox.ac.uk/
  - http://corpus.byu.edu/bnc

¿Qué podemos obtener de los corpus?

- Listas de concordancias monolingües

Cómo citar el CORPUS **Concordancias.**

Pantalla: 1 de 3. Siguiente 1 2 3 Ver párrafos

N°	CONCORDANCIA			AÑO
1	os dividen en tres: depresión tropical, cuando el	ciclón	alcanza vientos de hasta 63 kilómetros por hor **	2003
2	r o los lugares que azotará y la intensidad de un	ciclón	tropical. Hace siglos los hombres se defendían **	2003
3	ndas arrasadas, así como puentes y carreteras. El	ciclón	que golpeó el puerto de Tampico, México, en 19 **	2003
4	dieron la vida en la India a causa del paso de un	ciclón	. Dos años después, en 1979, el devastador hura **	2003
5	uadas. Las razones que provocan que el paso de un	ciclón	por una nación determinada deje un balance neg **	2003
6	d 1997 1997 10 107 P Reuter Un	ciclón	deja cientos de muertos en Bangladesh El devas **	1997
7	ja cientos de muertos en Bangladesh El devastador	ciclón	que barrió las costas de Bangladesh el lunes d **	1997
8	pescadores e isleños desaparecidos tras el último	ciclón	, que se ha cobrado hasta ahora 700 víctimas mo **	1988
9	mo tropical como consecuencia de dos sistemas: el	ciclón	Ágata, que se encuentra frente a la costa atlá **	1986
10	"Hemos tenido casos en que se nos pedía dar a un	ciclón	el nombre de la suegra", recuerda la meteorólo **	1997
11	con el Centro Nacional de Huracanes de Miami, el	ciclón	denominado Joyce tiene, en estos momentos, la **	2000
12	cifras emanadas del organismo estadounidense). El	ciclón	se traslada en dirección Oeste, a 36 kilómetro **	2000
13	o estará al Norte de Venezuela, pero el radio del	ciclón	es de 95 kilómetros, por lo cual serán las ráf **	2000
14	r debido a las fuertes lluvias ocasionadas por el	ciclón	, mientras que miles han sido evacuadas de la p **	2000
15	hay otra opción para prevenir la formación de un	ciclón	. La única alternativa es hacer predicciones co **	1996
16	ca el especialista. En el Atlántico, "se formó un	ciclón	tropical que alcanzó nivel de tormenta tropica **	2002
17	rsas condiciones que favorecen la formación de un	ciclón	tropical. Una de ellas es que la temperatura d **	2003
18	una zona de baja presión en superficie. Cuando un	ciclón	se genera, el estado del tiempo sufre un cambi **	2003
19	n velocidades de hasta 62 kilómetros por hora. Un	ciclón	se convierte en una tormenta tropical cuando a **	2003
20	P Ciclones tropicales ORFILIO PELÁEZ ¿Qué es un	ciclón	tropical? El ciclón tropical se define como un **	2002
21	les ORFILIO PELÁEZ ¿Qué es un ciclón tropical? El	ciclón	tropical se define como un centro de bajas pre **	2002
22	so, llegar a mil. ¿Cómo se clasifican? El término	ciclón	tropical es genérico y se les comienza a llama **	2002
23	m/h. Solo al huracán podemos considerarlo como un	ciclón	tropical plenamente desarrollado. Para el caso **	2002
24	s factores favorables, entonces puede formarse un	ciclón	tropical. Al requerir de aguas cálidas, dichos **	2002

¿Qué podemos obtener de los corpus?

- Listas de concordancias bilingües
  - play >
    (COMPARA)

«[…] and not being able to **play** the piano.»	«[…] e à incapacidade de **tocar** piano.»
Joe wanted to switch partners and **play** the best of three sets, […]	Joe queria trocar de parceiros e **jogar** de novo, uma melhor de três, […]
[…] he likes to **play** the father in our relationship.	[…] gosta de **fazer a** figura paterna no nosso relacionamento.

¿Qué podemos obtener de los corpus?

- Listas de palabras
  - Lemas > =, *rate*, *market*, *price*, *good*, *capital*, *investment*, etc. (BwanaNet English Economy corpus)
- Colocaciones
  - Sustantivos que son colocaciones de la forma *television* > radio, news, show, cable, network, station, series, etc. (BYU-Corpus of Contemporary American English)
- Listas de agrupaciones
  - Agrupaciones de 3 palabras que incluyen *mesita*> *mesita de noche*, *mesita de madera*, *mesita de luz*, *mesita del teléfono*, etc. (CREA)

¿Cómo buscamos en un corpus?

- Por lema
  - do > do, did, does, doing (BYU-British National Corpus)
- Por forma exacta
  - houses> houses
- Por forma truncada
  - hous* > house, housewife, housekeeper, house-doctor, houses, etc.
- Por categoría gramatical
  - adjective+noun > commercial legislation, fiscal protection, Social Fund (BwanaNet English Law corpus)

¿Cómo buscamos en un corpus?

- Presencia de todos los elementos
  - ciclón AND tropical

CONCORDANCIA

os de los ciclones los dividen en tres: depresión tropical, cuando el ciclón alcanza vientos de hasta 6
os dividen en tres: depresión tropical, cuando el ciclón alcanza vientos de hasta 63 kilómetros por hor
vientos de hasta 63 kilómetros por hora; tormenta tropical, si sus vientos adquieren una velocidad de e
r o los lugares que azotará y la intensidad de un ciclón tropical. Hace siglos los hombres se defendían
lugares que azotará y la intensidad de un ciclón tropical. Hace siglos los hombres se defendían de los

- Presencia de alguno de los elementos
  - severo OR grave

asos atendidos se dio una situación especialmente grave. "Se les ha prestado la atención habitual en es
1995 1995 10 107 P Escocia, en grave situación. Frío ártico en el norte de Europa y
n convertido la zona en un lugar peligroso por el grave riesgo de aludes. El Instituto Nacional de Mete
ección Civil mantiene la alerta L, la segunda más grave, para prevenir los efectos de las lluvias fuert
e gases contaminantes al ritmo actual supondrá un severo cambio del clima en España en los próximos año

- Ausencia de elementos
  - cresta Y NO ola

CONCORDANCIA

ácil, no tienen ahora la sensación de estar en la cresta a pesar de que "tanto los premios como el hech
onada fiesta para solaz de sus fans. Entre alguna cresta y varios pelajes rubio platino, no bajaron del
os, del V a.C. hasta el V d. C son señalados como cresta y decadencia del teatro occidental. Pero es qu
ubango y Mozámedes, sólo para bajar al mar por la cresta de una montaña hicieron una carretera de curv

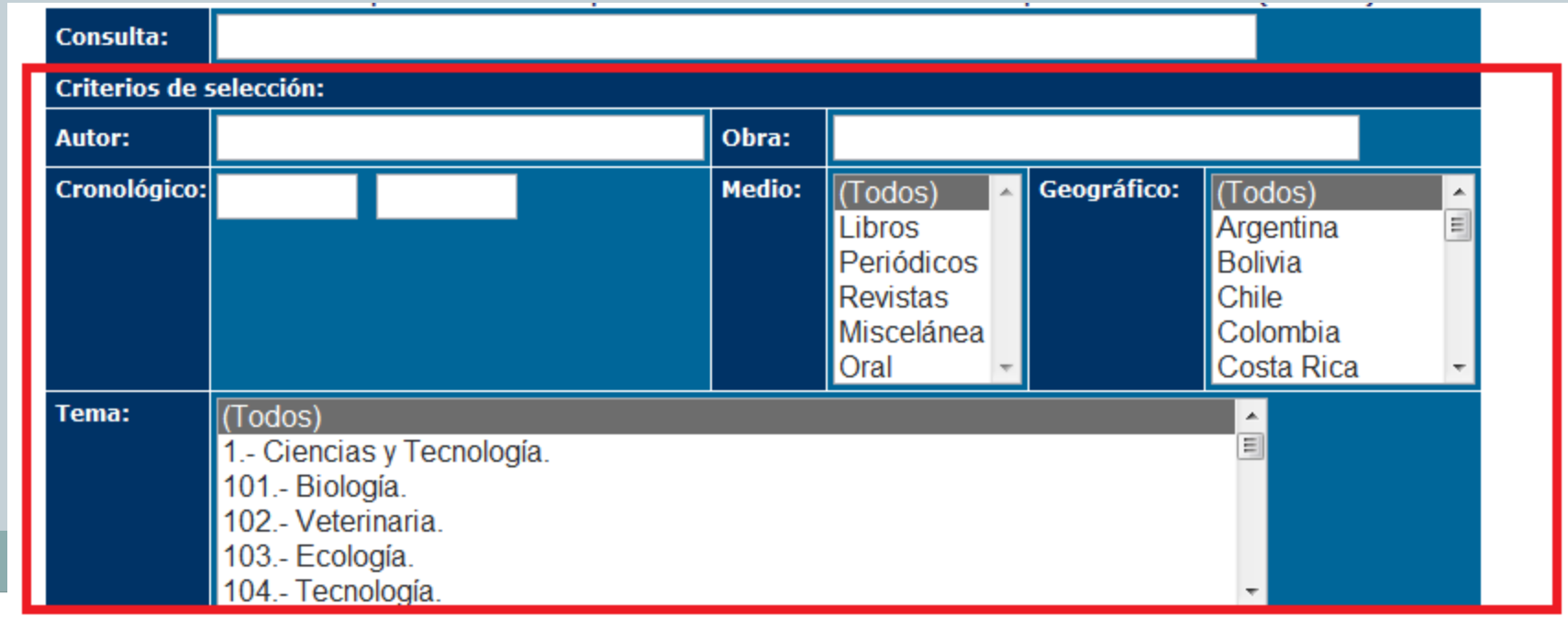
¿Cómo buscamos en un corpus?

- Combinación continua
  - llover a * > llover a mares, llovía a cántaros, llueve a torrentes, lloviendo a raudales, etc. (BYU-Corpus del español)
  - "raining cats and dogs" > It was **raining cats and dogs** and the teachers were running in and out helping us get our stuff in and just couldn't do enough for us. (BNC)
- Combinación discontinua
  - arrow [within 9] noun > A gaily-painted **quiver**, full of **arrows** // He could draw an **arrow** from his **quiver** [...] (BYU-Oxford English Dictionary corpus)

¿Cómo buscamos en un corpus?

- Filtros

categoría gramatical
área temática
Idioma
tipo de texto
área geográfica
autor
fecha
lugar del texto



CREA

- Corpus de referencia actual de la lengua española.
- Textos de países de habla hispana, desde 1975 hasta 2004.
- Permite:
  - Filtrar por autor, obra, fecha, tema, medio y procedencia geográfica.
  - Datos estadísticos
  - Concordancias
  - Agrupaciones

Concordancias (RAE)

Consulta:	***chasco*, en todos los medios, en *CREA***
Resultado:	**134 casos en 111 documentos.**

OBTENCIÓN DE EJEMPLOS

Recuperar | Concordancias. | Normal. | **Clasificación:** | **Agrupación:** | **Marcas:**

Cómo citar el CORPUS **Concordancias.**

Pantalla: 1 de 6. Siguiente 1 2 3 4 5 6 Ver párrafos

N°	CONCORDANCIA			AÑO	AUTOR
1	eflexión y el supuesto -que eso no se sabe nunca-	chasco	de encontrarse una demanda de separación al en **	1995	PRENSA
2	uipo a comienzos de la temporada, pero también el	chasco	monumental de Atenas y la desorganización que **	1994	PRENSA
3	neral del 27-E fue un "éxito" ("se han llevado un	chasco	quienes daban por finiquitados a los sindicato **	1994	PRENSA
4	suprema a "Bamboleo" para ocultar alegremente el	chasco	que le infringe la inspiración de Simón Díaz. **	1999	PRENSA
5	que crea que sabe mucho inglés se llevará flor de	chasco	si escucha al llegar a Londres la palabra dogg **	2004	PRENSA
6	tuada en la calle 3ª N° 20-40 se llevaron un gran	chasco	cuando la supuesta "casa de brujas" fue encont **	1975	PRENSA
7	l en mayo, y la segunda en junio. Defensores del '	chasco	' El miércoles a las 8 de la noche, la Selecci **	1997	PRENSA
8	otras, fácil de conquistar, pero, se llevaron un	chasco	, porque cuando la pretendían en amores, ella l **	1997	PRENSA
9	hasta ayer invicto AS Roma a manos del Inter, el	chasco	del Milán en Parma y el desliz juventino permi **	2000	PRENSA
10	n lo correcto". Pero reconoció que se llevaron un	chasco	con el precio que le asignaron, pues obviament **	1997	PRENSA
11	és y en República Dominicana, los autobuses. Otro	chasco	parecido vivió su señora cuando una vez a su h **	1997	PRENSA
12	os. Ambiente de miedo que olía a linchamiento. El	chasco	resultó tan vivo y tan chocante que hubo despu **	1996	PRENSA
13	las apagadas (de susto: ni jugó) como Insúa. Vaya	chasco	. **	2003	PRENSA
14	eo absoluto. ¿Tanto tiempo currándotelo para este	chasco	? Le das un abrazo y dices que ya lo arreglaréi **	2003	PRENSA
15	película sí fue premiada, pero yo no. Me llevé un	chasco	". Una portavoz de la Seminci lo considera un m **	2003	PRENSA
16	a. Fracasó en su intento de fichar a Romario y el	chasco	se repitió después con Jordi Cruyff. No ha sid **	1997	PRENSA
17	ar un poco grande. Unos y otros se han llevado un	chasco	. José María Escondrillas no será seguramente u **	1986	PRENSA
18	ada de Estados Unidos calificó las armas de "gran	chasco	". Las ametralladoras Gau-19 Gatling, que la ad **	2000	PRENSA
19	de generación eléctrica. La afirmación resultó un	chasco	. Parece un castigo. La naturaleza le está cobr **	1998	PRENSA
20	racia sin demasiada conciencia de ello. Su primer	chasco	fue no poder votar la Constitución y ahí empez **	1995	PRENSA

OBTENCIÓN DE EJEMPLOS

Recuperar | Agrupaciones. | Normal. | Clasificación: | Agrupación: | Marcas:

Cómo citar el CORPUS

Agrupaciones.

De 2 palabras	%	Casos	De 3 palabras	%	Casos	De 5 palabras	%	Casos
profesar la	12.22	11	*profesar la religión*	5.55	5	*profesar y practicar su propia*	2.22	2
profesar como	6.66	6	*profesar como monja*	3.33	3	*profesar tomó el nombre de*	1.11	1
profesar el	6.66	6	*profesar su fe*	3.33	3	*profesar una religión o una*	1.11	1
profesar y	6.66	6	*profesar y practicar*	2.22	2	*profesar la antropología. Confiesa sus*	1.11	1
profesar en	6.66	6	*profesar una religión*	2.22	2	*profesar muchos sistemas políticos, aunque*	1.11	1
profesar una	5.55	5	*profesar la fe*	2.22	2	*profesar" en el contexto de*	1.11	1
profesar su	4.44	4	*profesar en la*	2.22	2	*profesar el intelectualismo o el*	1.11	1
profesar por	3.33	3	*profesar cualquier religión*	2.22	2	*profesar con ellos y no*	1.11	1
profesar un	3.33	3	*profesar en éste*	1.11	1	*profesar la tesis de que*	1.11	1

Resultados con estadísticas (RAE)

Consulta:	***chasco*, en todos los medios, en *CREA***
Resultado:	**135 casos en 111 documentos.**

Filtros: Casos

Ratio: 10

Mantener documentos (Sólo para filtro sobre casos).

Filtrar

OBTENCIÓN DE EJEMPLOS

Recuperar

Concordancias

Normal.

Clasificación:

Agrupación:

Marcas:

Cómo citar el CORPUS

Estadísticas

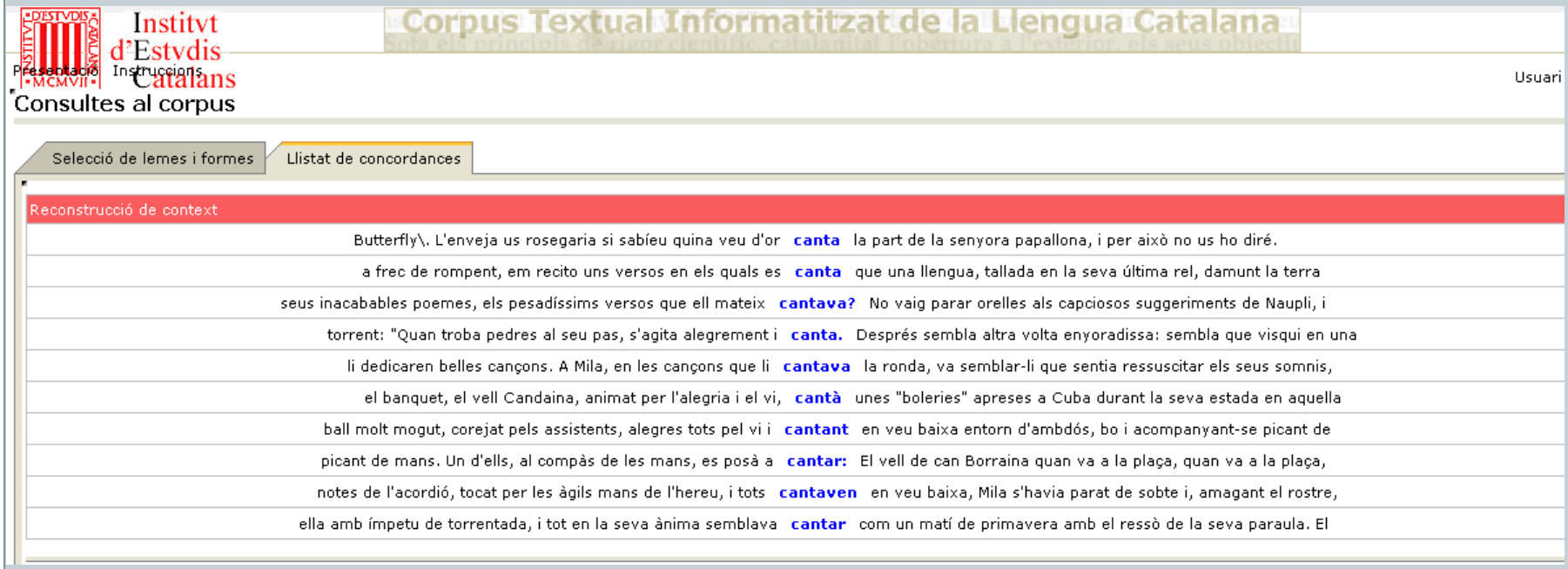
Año	%	Casos
1986	10.00	12
1995	9.16	11
2002	8.33	10
1997	7.50	9
1988	6.66	8
2003	6.66	8
1991	5.83	7

País	%	Casos
ESPAÑA	58.77	77
ARGENTINA	5.34	7
CUBA	4.58	6
MÉXICO	4.58	6
PERÚ	4.58	6
COLOMBIA	3.81	5
VENEZUELA	3.81	5

Tema	%	Casos
7.- Ficción.	53.33	72
3.- Política, economía, comercio y finanzas.	12.59	17
2.- Ciencias sociales, creencias y pensamiento.	11.85	16
5.- Ocio, vida cotidiana.	11.85	16
4.- Artes.	6.66	9
9.- Oral.	2.22	3
1.- Ciencia y Tecnología.	1.48	2

CTILC

- Institut d'Estudis Catalans
- Textos literarios y no literarios, fechados entre 1833 y 1988.
- Corpus lematizado
- Permite buscar concordancias por lema y por forma.


Institvt d'Estvdis Catalans
Corpus Textual Informatitzat de la Llengua Catalana
Presentació Instruccions
Usuari
Consultes al corpus
Selecció de lemes i formes
Llistat de concordances
Reconstrucció de context
Butterfly\. L'enveja us rosegaria si sabíeu quina veu d'or canta la part de la senyora papallona, i per això no us ho diré.
a frec de rompent, em recito uns versos en els quals es canta que una llengua, tallada en la seva última rel, damunt la terra
seus inacabables poemes, els pesadíssims versos que ell mateix cantava? No vaig parar orelles als capciosos suggeriments de Naupli, i
torrent: "Quan troba pedres al seu pas, s'agita alegrement i canta. Després sembla altra volta enyoradissa: sembla que visqui en una
li dedicaren belles cançons. A Mila, en les cançons que li cantava la ronda, va semblar-li que sentia ressuscitar els seus somnis,
el banquet, el vell Candaina, animat per l'alegria i el vi, cantà unes "boleries" apreses a Cuba durant la seva estada en aquella
ball molt mogut, corejat pels assistents, alegres tots pel vi i cantant en veu baixa entorn d'ambdós, bo i acompanyant-se picant de
picant de mans. Un d'ells, al compàs de les mans, es posà a cantar: El vell de can Borraina quan va a la plaça, quan va a la plaça,
notes de l'acordió, tocat per les àgils mans de l'hereu, i tots cantaven en veu baixa, Mila s'havia parat de sobte i, amagant el rostre,
ella amb ímpetu de torrentada, i tot en la seva ànima semblava cantar com un matí de primavera amb el ressò de la seva paraula. El

British National Corpus (BNC)

- Corpus monolingüe del inglés actual, oral y escrito, de más de 100 millones de palabras.
- Diferentes interfaces para realizar búsquedas.
- Búsquedas complejas en http://bncweb.lancs.ac.uk/
- Permite:
  - Búsqueda por distancia contextual
  - Por lema y forma
  - Por categoría gramatical
  - Con comodines
  - Datos estadísticos: distribución geográfica, edad, sexo, etc.
  - Colocaciones
  - Etc.

Collocation parameters:			
Information:	collocations	Statistics:	Log-likelihood
Collocation window span:	3 Left - 3 Right	Basis:	whole BNC
Freq(node, collocate) at least:	5	Freq(collocate) at least:	5
Filter results by:	Specific collocate:	and/or tag: any verb	Submit changed parameters Go!

There are 4766 different types in your collocation database for "[word="awareness"%c]". (Your query "awareness" returned 3531 hits in 1118 different texts)

No.	Word	Total No. in whole BNC	Expected collocate frequency	Observed collocate frequency	In No. of texts	Log-likelihood value
1	raise	6,043	1.072	123	92	925.9591
2	increase	7,239	1.285	66	60	391.3238
3	raising	2,443	0.434	50	46	376.7774
4	heighten	116	0.021	19	18	224.7968
5	promote	3,142	0.558	30	30	180.5525
6	develop	8,533	1.514	36	31	159.3617
7	encourage	5,070	0.900	21	16	92.205
8	heightening	43	0.008	7	7	82.7219
9	mining	528	0.094	9	1	64.5094
10	developing	3,093	0.549	13	13	57.4394
11	promoting	1,498	0.266	10	9	53.1478

Corpus BwanaNet

- Programa de explotación de corpus de diferentes especialidades.
- Desarrollado por el Institut Universitari de Lingüística Aplicada (IULA), de la Universitat Pompeu Fabra.
- Incluye textos originales y paralelos en inglés, catalán y español de las áreas de informática, medio ambiente, derecho, medicina, genoma y economía.

Corpus BwanaNet

- El programa de corpus permite:
  - Seleccionar ámbito temático
  - Consultar por lema o por forma
  - Consultar por distancia contextual
  - Consultar por categoría gramatical
  - Concordancias multilingües
  - Listas de palabras
  - Etc.

5. Resultados: concordancias

Atención! el número de concordancias que se muestran esta limitado a 50

Selección realizada:

Lengua de los documentos: Castellano
Documentos paralelos: Inglés
Ámbitos temáticos seleccionados: Informática
Número de palabras: 291901
Cantidad de documentos: 22

a:[lemma="memoria"] :: ((a.doc_area="i")) within s;

Número de concordancias: 50

1<i00018><trad><oo> <s>El router seleccionado debe estar normalmente activado y tener suficiente ***memoria*** para procesar información de encaminamiento NLSP y generar el seudo LSP para su LAN .</s>

<s>The router you choose should be typically up and should have enough memory to process NLSP routing information and generate the pseudonode LSP for its LAN .</s>

2<i00018><trad><oo> <s>Para determinar el tiempo de procesamiento y el uso de ***memoria*** de l router , emplee la utilidad MONITOR .</s>

<s>To determine the processing time and memory usage of a router , you use the MONITOR utility .</s>

3<i00018><trad><oo> <s>Si la asignación de la ***memoria*** entre el motor MS y el motor de E/S no tiene establecido el ajuste adec INETCFG no se ejecutará .</s>

<s>If the allocation of memory between the MS Engine and the IO Engine is not set to an adequate value , INETCFG does not run .</s>

Actividad

	Oral /escrito	Monolingüe/ multilingüe	Comparable /paralelo	General /especializado	Sincrónico /diacrónico	Bruto/anotado Tipo de anotación
CREA						
CORDE						
CTILC						
BNC						
BwanaNet						

Corpus ad hoc

Crearemos nuestro propio corpus cuando necesitemos:

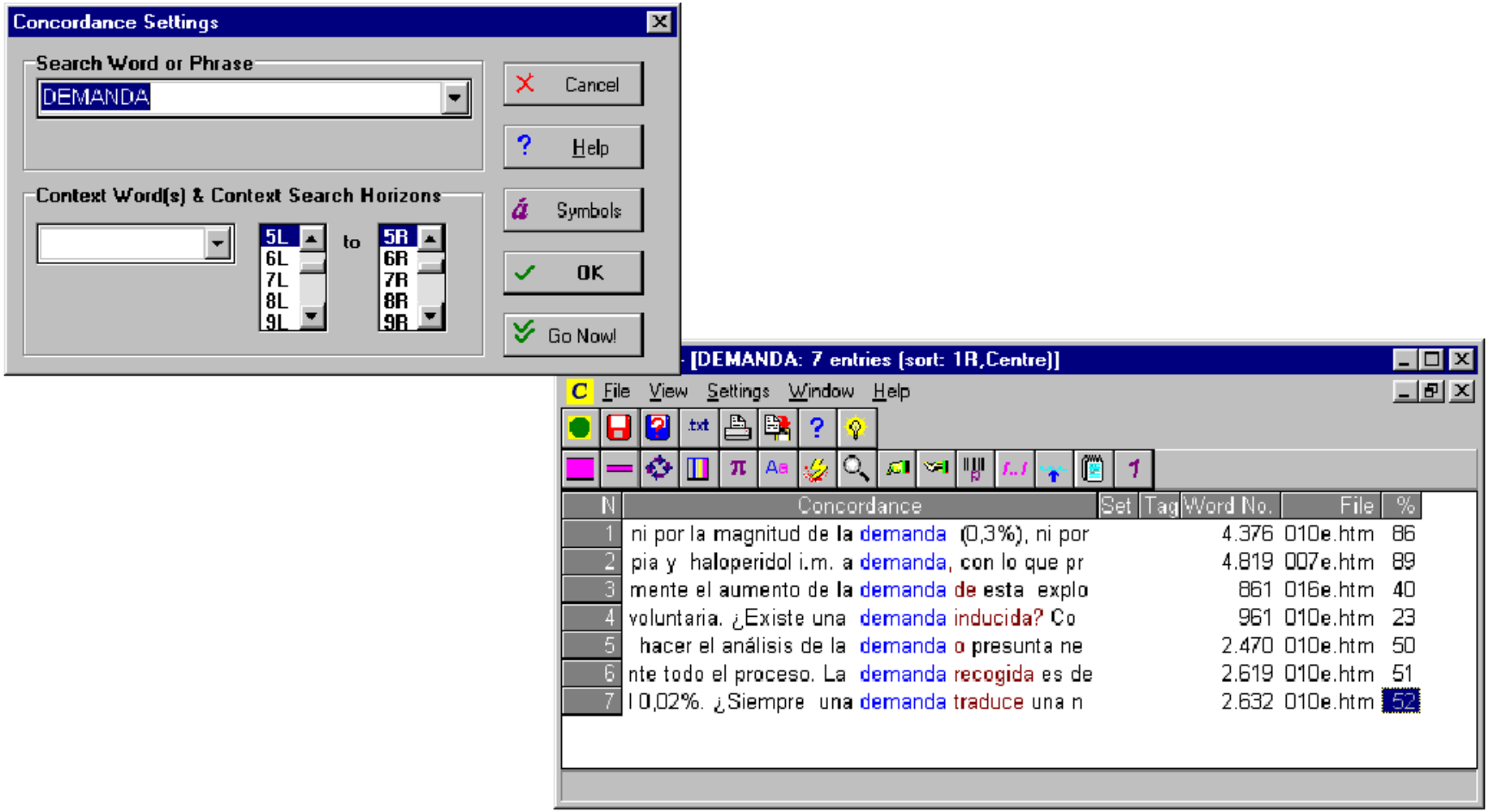
- Obtener información sobre el tema de la traducción
- Obtener información sobre el género
- Obtener información lingüística y/o terminológica
- Alimentar memoria de traducción

Creación de un corpus ad hoc

- Recopilación de textos en Internet
  - Criterios de calidad: autoría, actualidad, corrección, destinatarios, etc.
- Recopilación de traducciones propias anteriores
  - En formato adecuado y de manera organizada

Explotación del corpus

- Programas
  - Wordsmith Tools
  - AntConc
  - MonoConc Pro
  - Otros
- Funciones
  - Concordancias
  - Listas de palabras
  - Colocaciones
  - Agrupaciones
  - Estadísticas


Concordance Settings
Search Word or Phrase
DEMANDA
Cancel
Help
Symbols
OK
Go Now!
Context Word(s) & Context Search Horizons
5L
6L
7L
8L
9L
to
5R
6R
7R
8R
9R
[DEMANDA: 7 entries (sort: 1R,Centre)]
File View Settings Window Help
N Concordance Set Tag Word No. File %
1 ni por la magnitud de la demanda (0,3%), ni por 4.376 010e.htm 86
2 pia y haloperidol i.m. a demanda, con lo que pr 4.819 007e.htm 89
3 mente el aumento de la demanda de esta explo 861 016e.htm 40
4 voluntaria. ¿Existe una demanda inducida? Co 961 010e.htm 23
5 hacer el análisis de la demanda o presunta ne 2.470 010e.htm 50
6 nte todo el proceso. La demanda recogida es de 2.619 010e.htm 51
7 l 0,02%. ¿Siempre una demanda traduce una n 2.632 010e.htm 52

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	%	mmas
1	DE	8.104	8,59	111	100,00	
2	LA	3.032	3,22	109	98,20	
3	Y	2.570	2,73	108	97,30	
4	#	2.332	2,47	101	90,99	
5	EN	2.248	2,38	110	99,10	
6	EL	2.196	2,33	107	96,40	
7	QUE	2.029	2,15	111	100,00	
8	LOS	1.633	1,73	108	97,30	
9	A	1.630	1,73	106	95,50	
10	PARA	1.332	1,41	107	96,40	
11	UN	1.170	1,24	106	95,50	
12	UNA	1.022	1,08	105	94,59	
13	CON	985	1,04	104	93,69	
14	LAS	978	1,04	103	92,79	
15	SE	844	0,90	99	89,19	
16	DEL	721	0,76	102	91,89	
17	ES	624	0,66	103	92,79	
18	POR	563	0,60	100	90,09	
19	MÁS	497	0,53	92	82,88	
20	COMO	454	0,48	92	82,88	
21	O	433	0,46	81	72,97	
22	SU	423	0,45	90	81,08	
23	NO	352	0,37	73	65,77	
24	AL	338	0,36	90	81,08	
25	DATOS	330	0,35	56	50,45	
26	SUS	266	0,28	77	69,37	
27	SISTEMA	260	0,28	56	50,45	

Bibliografía

Bowker, L. (2000). “Towards a Methodology for Exploiting Specialized Target Language Corpora as Translation Resources”, International Journal of Corpus Linguistics 5(1): 17-52.

Pastor, V. y A. Alcina (2009): “Search techniques in corpora for the training of translators”, en *International Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning*. Bulgaria.

Sánchez Gijón, P. (2003). Els documents digitals especialitzats: utilització de la lingüística de corpus com a front de recursos per a la traducció. Tesis disponible en <http://www.tdx.cbuc.es/ >. Barcelona, Universidad Autónoma de Barcelona