



Tony Berber Sardinha (Sao Paulo Catholic University, Brazil)

Eliane Alambert

10/10/2009

<http://corpusbrasileiro.pucsp.br>



ACKNOWLEDGMENTS

- Fapesp, grant # 2008/00944-0

CORPUS BRASILEIRO PROJECT

- Description
 - Aims:
 - Compile a 1-billion word corpus of Brazilian Portuguese of the late 20th century, early 21st century
 - Make corpus available on the web, for free, for searching
 - Fill gap in resources for large, register-diversified, publicly available corpus of BP
 - Our intention is for the corpus to be a public reference for Brazilian Portuguese in terms of size and accessibility
 - Not a reference corpus, because it is not balanced, has little spoken language, does not span a large timeframe

CORPUS BRASILEIRO PROJECT

- Timeline
 - 2007: Idea came about
 - early 2008: Submitted grant application to Fapesp (grant making institution)
 - May 6, 2008: Received grant
 - April 30, 2010: Project will end
 - We consider project to be a long-term commitment
 - We consider this to be phase 1 of the project; phase 2: more improvements (more search features, further annotation, etc)
 - up to February 2009: Corpus was stored and running on a rental server at www.corpusbrasileiro.org; we had problems with that, closed it down
 - after that: We bought a server, installed at University, Corpus is being transferred to that server, at corpusbrasileiro.pucsp.br. Databases being rebuilt. No search interface there yet. No live demonstration...

CORPUS BRASILEIRO PROJECT







- Phase 1 goal is to produce a ‘basic’ version of the corpus, that is, with **a focus on concordancing and collocation**, which enables users to **quickly and simply**:
 - Get single word, 2 and 3-gram frequency lists, concordances, collocate tables for words
 - Get single word frequency lists for word + POS or word + lemma
 - for genre/register/whole corpus
 - Via a **clean, straightforward** interface, in order to make corpus **accessible** for a wide range of users, academics, and other people interested in the language (middle / high school teachers and students, for example, authors, journalists, etc.)

SQL ARCHITECTURE


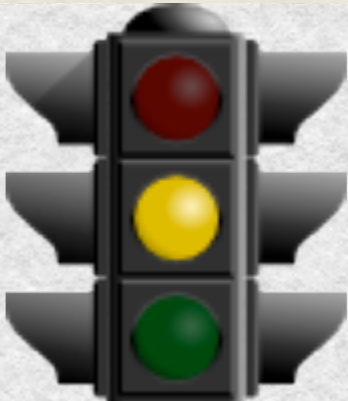

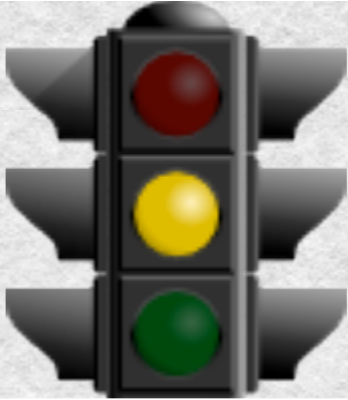

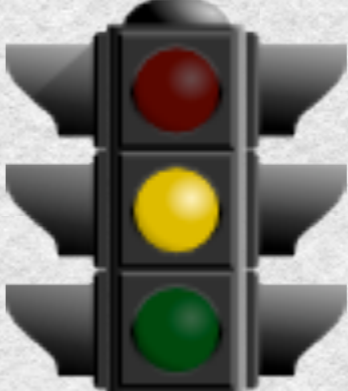
- Main challenge:
 - Enable fast searches, online
 - ‘One of the fundamental challenges in the creation of search engines for large corpora is the need to **balance size, annotation, and speed**. Some approaches may achieve two of the three goals quite well, but shortcomings with the third factor will severely diminish the usefulness of the corpus.’ (Davies, 2005 :307)
 - Decision: use relational database architecture, proven reliable and fast for BYU corpora, although these are about about 3 times smaller than ours.
 - SQL architecture is quite scalable, but we don’t know what impact a database this size will have on performance.

LARGE CORPORA

Very large,
corpora are
moving past the
1 billion word
mark

Bank of English		524 million		Collins
Collins Word Web		2.5 billion		Collins
Cosmas		2.2 billion		IDS

WACKY (WEB AS CORPUS KOOLOGY INITIATIVE)

deWaC		1.3 billion		WaCky site
itWaC		1.6 billion		WaCky site
ukWaC		1.9 billion		WaCky site

OFFLINE CORPORA OF PORTUGUESE

Large, 100+
million

Banco de
Português



750 million



PUCSP

AC/DC



352 million



Linguatca

Corpus de
Referência do
Português
Contemporâneo



335 million



Univ. Lisboa

ONLINE CORPORA

Large, but except for BYU, online corpora tend to be versions of offline counterparts

BNC Online		100 million	British Library
BNC-BYU		100 million	BYU
COCA		385+ million words	BYU
COSMAS		2.2 billion?	IDS
CREA		200 million	Real Academia
Time Magazine		100 million	BYU

ONLINE CORPORA OF PORTUGUESE

Small(ish), not
register diversified;
largest one is of a
single newspaper.

CETEM Público		180 million words	Linguatca
Corpus do Português		45 million	BYU
CETEM Folha		24 million	Linguatca
Lácio-Web		10 million	NILC
Banco de Português		1 million	PUCSP

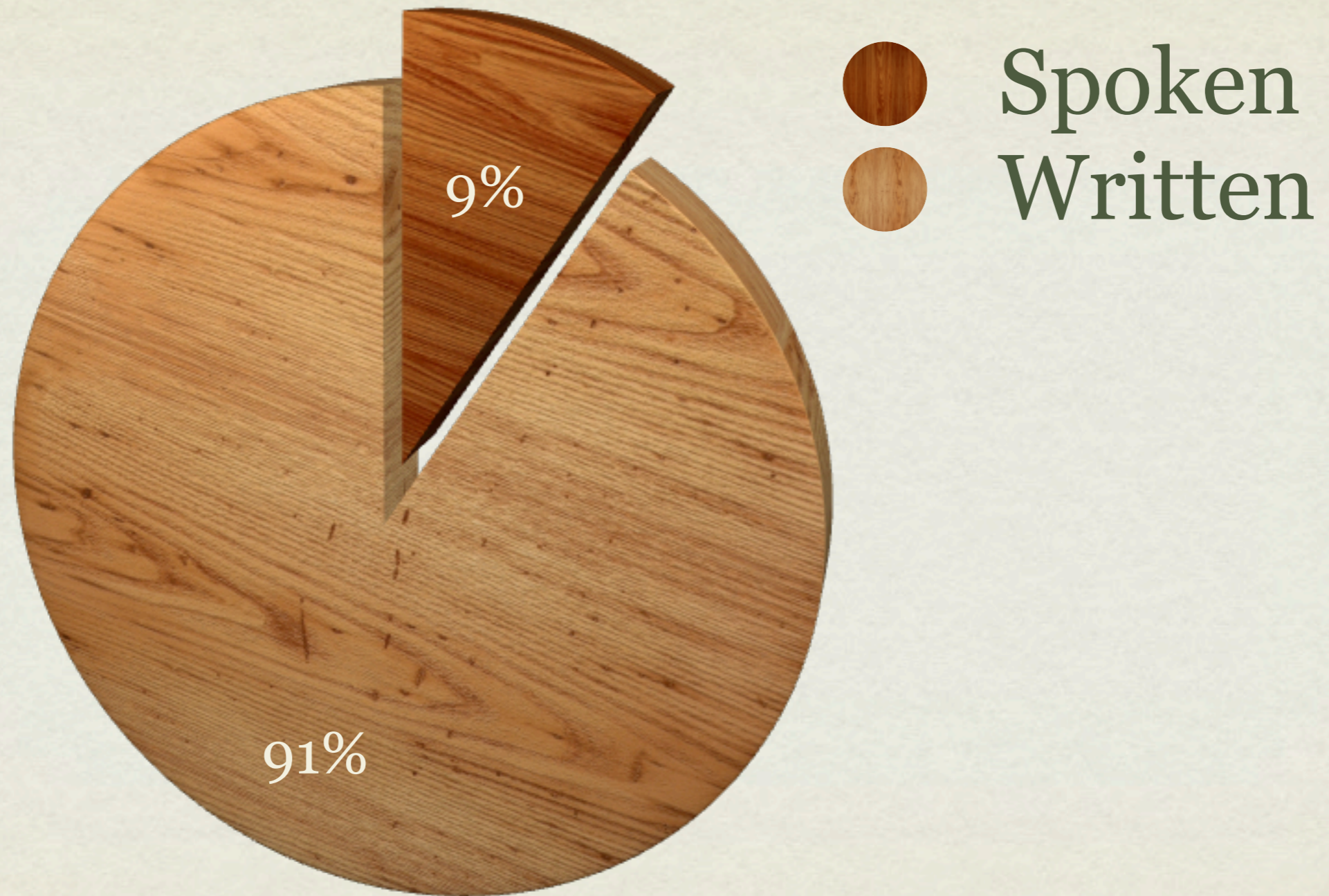
OBTAINING TEXTS

- We used our own previous collections:
 - Newspaper corpora
 - Transcriptions of talk
- We gathered offline sources as well
 - Sociolinguistic interviews
 - Meetings
- We searched the web for freely available online sources
 - Downloaded files via scripts
 - Converted, cleaned files

COLLECTION

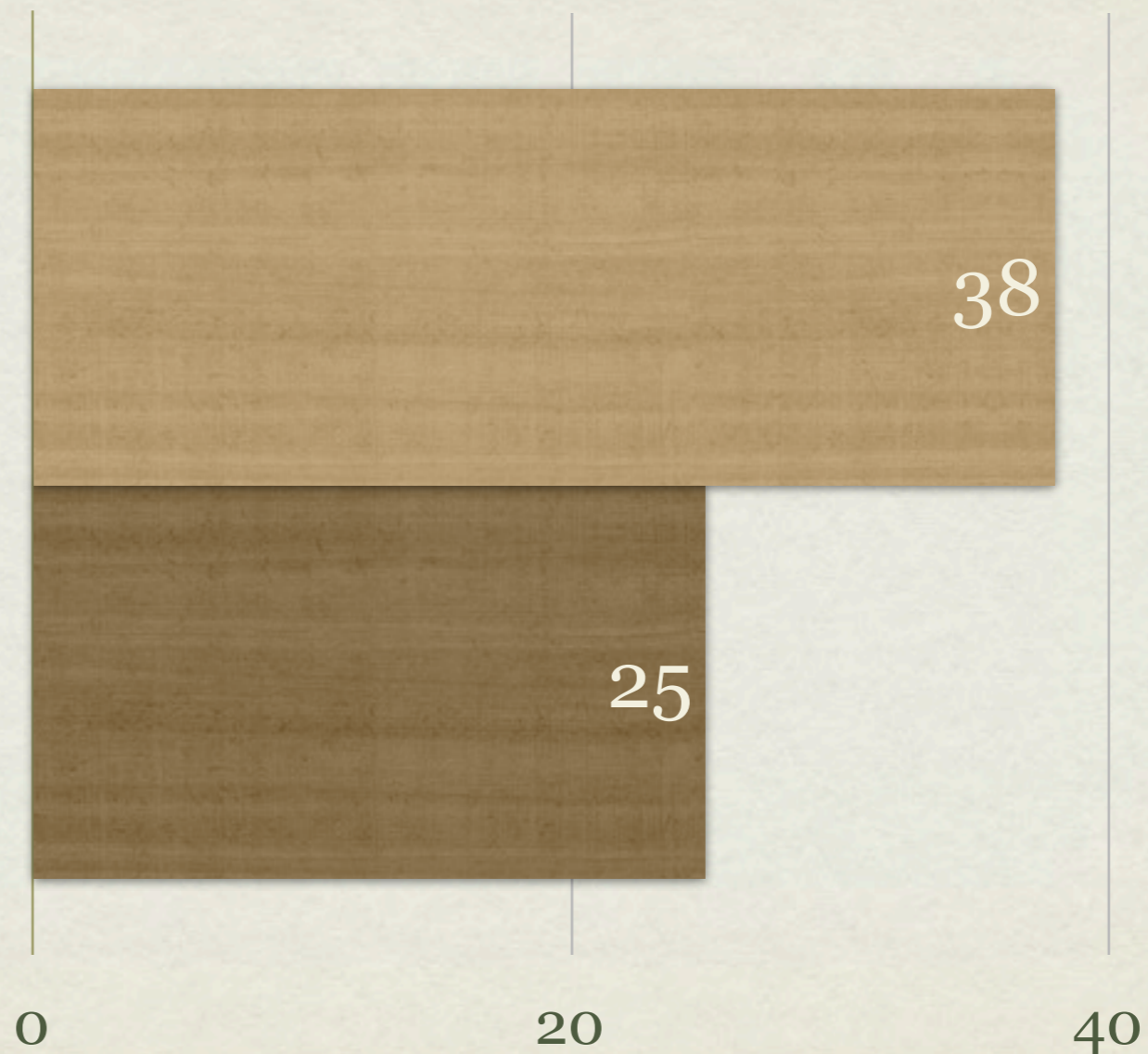


MODE



REGISTERS / GENRES

Registers
Genres



REGISTERS

Academic	508,185,548	53.1%
Various	227,831,149	23.8%
Education	96,438,254	10.1%
Politics	79,131,819	8.3%
Law	12,147,467	1.3%
Science	5,000,000	0.5%
Literature	3,300,761	0.3%
Health planning	3,152,869	0.3%
Economy	3,019,775	0.3%
Agronomy	2,260,561	0.2%
Computing	2,131,523	0.2%
Biology	1,909,844	0.2%
Political Science	1,613,627	0.2%
Health sciences	1,532,442	0.2%
Philosophy	1,488,153	0.2%
Communications	1,207,107	0.1%
Religion	822,196	0.1%

REGISTERS

Arts	631,319	0.1%
Business	526,373	0.1%
Cooking	436,971	0%
Medicine	408,068	0%
Maths	385,932	0%
Information science	328,225	0%
Psychology	280,945	0%
Chemistry	277,207	0%
Languages	265,467	0%
Geography	247,153	0%
Legal	246,437	0%
Environment	190,584	0%
Physics	186,424	0%
Weather	181,102	0%
Human rights	151,952	0%
Geoscience	145,690	0%
Pharmacy	117,117	0%
Vet Medicine	98,756	0%
Sports	74,604	0%
Architecture	40,945	0%
Conversation	21,430	0%

GENRES

Thesis	261,414,124	27.3%
Research article	246,771,424	25.8%
Newspaper various	222,697,136	23.3%
Technical document	107,781,080	11.3%
Congressional proceedings	77,330,504	8.1%
Curriculum standards	12,971,657	1.4%
Popular science reports	5,001,011	0.5%
Interview	4,518,276	0.5%
Various	4,087,172	0.4%
TV interview	3,371,725	0.4%
Magazine report	2,747,334	0.3%
Speech	1,779,712	0.2%
Short story	1,607,212	0.2%
Novels	1,549,601	0.2%

GENRES

Novels	1,549,601	0.2%
Handbook	874,087	0.1%
Magazine various	540,331	0.1%
Recipe	436,971	0%
Letter	275,817	0%
Law	246,437	0%
Medicine label	148,256	0%
Newspaper chronicle	143,948	0%
Soccer commentary	74,604	0%
TV debate	21,603	0%
Telephone	21,430	0%
Meeting	5,355	0%

FROM TEXT TO DATABASE

- Example:
 - No mínimo, 10 % do total de créditos exigidos para a graduação no ensino superior no país será reservado para a atuação dos alunos em ações extensionistas.
 - Translation: At least 10% of credits required for graduation in higher education in the country will be earmarked for students through extension actions.

Etiquetador Tree-Tagger

PUC/SP, LAEL, CEPRIL, DIRECT

Para que serve

Colocar etiquetas morfossintáticas em cada palavra de um corpus, por meio do etiquetador [Tree-Tagger](#).

Como usar

- [Fazer envio \(upload\)](#) do corpus que pretende usar.
- Escolher a língua.
- Depois de enviar o código do corpus enviado, digite no espaço correspondente do formulário abaixo e clique Fazer.

Língua ☒ Português

Código do corpus (seis dígitos):

Fazer

Limpar

FILES ARE TAGGED WITH
TREE TAGGER

FROM TEXT TO DATABASE: TAGGING AND LEMMATIZATION

Columns:
Word
POS
Lemma

no	S	em
mínimo	A	mínimo
,	2	,
10	C	@card@
%	V	<unknown>
do	S	de
total	J	total
de	N	de
créditos	J	crédito
exigidos	V	exigir
para	N	para
a	H	a
graduação	J	graduação

REGISTER TABLE: 7-GRAM STRUCTURE

word ID	register	Lemma	word 1	tag	...	word 6	tag
------------	----------	-------	--------	-----	-----	--------	-----

W₄: node
word

FROM TEXT TO DATABASE: 7-GRAMS (CONTINUED)

lemma
refers to
w4

id	g	l	w1	t1	w2	t2	w3	t3
149880843	ee	em	assegurar	V	que	E	,	2
149880844	ee	mínimo	que	E	,	2	no	S
149880845	ee	,	,	2	no	S	mínimo	A
149880846	ee	@card@	no	S	mínimo	A	,	2
149880847	ee	<unknown>	mínimo	A	,	2	10	C
149880848	ee	de	,	2	10	C	%	V
149880849	ee	total	10	C	%	V	do	S
149880850	ee	de	%	V	do	S	total	J
149880851	ee	crédito	do	S	total	J	de	N
149880852	ee	exigir	total	J	de	N	créditos	J
149880853	ee	para	de	N	créditos	J	exigidos	V

FROM TEXT TO DATABASE: 7-GRAMS

W₄: node
word

w4	t4	w5	t5	w6	t6	w7	t7
no	S	mínimo	A	,	2	10	C
mínimo	A	,	2	10	C	%	V
,	2	10	C	%	V	do	S
10	C	%	V	do	S	total	J
%	V	do	S	total	J	de	N
do	S	total	J	de	N	créditos	J
total	J	de	N	créditos	J	exigidos	V
de	N	créditos	J	exigidos	V	para	N
créditos	J	exigidos	V	para	N	a	H
exigidos	V	para	N	a	H	graduação	J
para	N	a	H	graduação	J	no	S

Server: localhost Database: cbr

Structure SQL Search Query Export Import Operations

	Table	Action						Records ¹	Type	Collation	Size	Overhead
<input type="checkbox"/>	ea							113,228	MyISAM	latin1_swedish_ci	11.3 MiB	-
<input type="checkbox"/>	eb							60,777	MyISAM	latin1_swedish_ci	5.6 MiB	-
<input type="checkbox"/>	ec							160,525	MyISAM	latin1_swedish_ci	14.8 MiB	-
<input type="checkbox"/>	ed							494,974	MyISAM	latin1_swedish_ci	46.5 MiB	-
<input type="checkbox"/>	ee01							10,000,000	MyISAM	latin1_swedish_ci	950.8 MiB	-
<input type="checkbox"/>	ee02							10,000,000	MyISAM	latin1_swedish_ci	948.0 MiB	-
<input type="checkbox"/>	ee03							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB	-
<input type="checkbox"/>	ee04							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB	-
<input type="checkbox"/>	ee05							10,000,000	MyISAM	latin1_swedish_ci	949.7 MiB	-
<input type="checkbox"/>	ee06							10,000,000	MyISAM	latin1_swedish_ci	951.8 MiB	-
<input type="checkbox"/>	ee07							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB	-
<input type="checkbox"/>	ee08							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB	-
<input type="checkbox"/>	ee09							6,366,377	MyISAM	latin1_swedish_ci	609.7 MiB	-
<input type="checkbox"/>	ef							3,032,012	MyISAM	latin1_swedish_ci	285.8 MiB	-

Registers

TABLES

<input type="checkbox"/>	eg01							10,000,000	MyISAM	latin1_swedish_ci	940.7 MiB
<input type="checkbox"/>	eg02							10,000,000	MyISAM	latin1_swedish_ci	941.3 MiB
<input type="checkbox"/>	eg03							10,000,000	MyISAM	latin1_swedish_ci	692.7 MiB
<input type="checkbox"/>	eg04							10,000,000	MyISAM	latin1_swedish_ci	692.1 MiB
<input type="checkbox"/>	eg05							10,000,000	MyISAM	latin1_swedish_ci	693.0 MiB
<input type="checkbox"/>	eg06							10,000,000	MyISAM	latin1_swedish_ci	692.4 MiB
<input type="checkbox"/>	eg07							10,000,000	MyISAM	latin1_swedish_ci	692.7 MiB
<input type="checkbox"/>	eg08							10,000,000	MyISAM	latin1_swedish_ci	691.7 MiB
<input type="checkbox"/>	eg09							10,000,000	MyISAM	latin1_swedish_ci	692.6 MiB
<input type="checkbox"/>	eg10							10,000,000	MyISAM	latin1_swedish_ci	691.6 MiB
<input type="checkbox"/>	eg11							10,000,000	MyISAM	latin1_swedish_ci	692.8 MiB
<input type="checkbox"/>	eg12							10,000,000	MyISAM	latin1_swedish_ci	693.3 MiB
<input type="checkbox"/>	eg13							10,000,000	MyISAM	latin1_swedish_ci	692.5 MiB

TABLES

<input type="checkbox"/>	eg14							10,000,000	MyISAM	latin1_swedish_ci	693.3 MiB
<input type="checkbox"/>	eg15							10,000,000	MyISAM	latin1_swedish_ci	692.5 MiB
<input type="checkbox"/>	eg16							10,000,000	MyISAM	latin1_swedish_ci	692.3 MiB
<input type="checkbox"/>	eg17							10,000,000	MyISAM	latin1_swedish_ci	692.3 MiB
<input type="checkbox"/>	eg18							10,000,000	MyISAM	latin1_swedish_ci	692.7 MiB
<input type="checkbox"/>	eg19							10,000,000	MyISAM	latin1_swedish_ci	692.6 MiB
<input type="checkbox"/>	eg20							10,000,000	MyISAM	latin1_swedish_ci	692.8 MiB
<input type="checkbox"/>	eg21							10,000,000	MyISAM	latin1_swedish_ci	693.0 MiB
<input type="checkbox"/>	eg22							10,000,000	MyISAM	latin1_swedish_ci	692.7 MiB
<input type="checkbox"/>	eg23							10,000,000	MyISAM	latin1_swedish_ci	692.0 MiB
<input type="checkbox"/>	eg24							10,000,000	MyISAM	latin1_swedish_ci	691.7 MiB
<input type="checkbox"/>	eg25							10,000,000	MyISAM	latin1_swedish_ci	692.4 MiB
<input type="checkbox"/>	eg26							3,732,527	MyISAM	latin1_swedish_ci	258.5 MiB

TABLES

<input type="checkbox"/>	eh							1,382,801	MyISAM	latin1_swedish_ci	128.1 MiB
<input type="checkbox"/>	ei							258,585,004	MyISAM	latin1_swedish_ci	18.3 GiB
<input type="checkbox"/>	ej							111,817,511	MyISAM	latin1_swedish_ci	7.7 GiB
<input type="checkbox"/>	ek							708,239	MyISAM	latin1_swedish_ci	68.4 MiB
<input type="checkbox"/>	el							914,786	MyISAM	latin1_swedish_ci	62.8 MiB
<input type="checkbox"/>	em							240,936	MyISAM	latin1_swedish_ci	17.3 MiB
<input type="checkbox"/>	en							289,390	MyISAM	latin1_swedish_ci	26.5 MiB
<input type="checkbox"/>	eo							4,319	MyISAM	latin1_swedish_ci	420.0 KiB
<input type="checkbox"/>	ex							199,154,880	MyISAM	latin1_swedish_ci	13.8 GiB
<input type="checkbox"/>	fa							86,323	MyISAM	latin1_swedish_ci	7.7 MiB
<input type="checkbox"/>	fb							22,033	MyISAM	latin1_swedish_ci	2.0 MiB
<input type="checkbox"/>	fc							1,803,404	MyISAM	latin1_swedish_ci	169.6 MiB

TABLES

<input type="checkbox"/>	fd01							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd02							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd03							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd04							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd05							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd06							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd07							10,000,000	MyISAM	latin1_swedish_ci	0.9 GiB
<input type="checkbox"/>	fd08							7,139,578	MyISAM	latin1_swedish_ci	681.7 MiB
<input type="checkbox"/>	x							166,247,219	MRG_MYISAM	latin1_swedish_ci	unknown
	61 table(s)	Sum						1,162,356,843	MyISAM	latin1_swedish_ci	73.5 GiB

TABLES

SEARCH

- Users can search for word, lemma or tag of node word plus word or tag anywhere within 3 words on each side of node, across the whole corpus or in selected registers.
- Query is first made to the database on a 1-gram table to display frequency of search term across all genres/registers, and on a 3-gram table, if search involves more than one term.
- Results are displayed to user (very fast, fraction of a second)
- Meanwhile, searches are made on all register tables, which are saved to temporary tables.
- User is given a link to see the concordances, which are presented in blocks of 50 to 1000 lines. Concordances are built with a LIMIT clause.

CONCORDANCES

- For each record selected by query, we also select records that are 7 rows behind and 7 rows ahead.
- These are then JOINed; tag columns are removed.
- This produces a concordance line with 10 words on either side of the node: 3 in the hit record plus 7 in the additional records.

CONCORDANCES

149880850	ee	de	%	V	do	S	total	J	de	N	créditos	J	exigidos	V	para	N
149880851	ee	crédito	do	S	total	J	de	N	créditos	J	exigidos	V	para	N	a	H
149880852	ee	exigir	total	J	de	N	créditos	J	exigidos	V	para	N	a	H	graduação	J
149880853	ee	para	de	N	créditos	J	exigidos	V	para	N	a	H	graduação	J	no	S
149880854	ee	a	créditos	J	exigidos	V	para	N	a	H	graduação	J	no	S	ensino	J
149880855	ee	graduação	exigidos	V	para	N	a	H	graduação	J	no	S	ensino	J	superior	A
149880856	ee	em	para	N	a	H	graduação	J	no	S	ensino	J	superior	A	no	S
149880857	ee	ensino	a	H	graduação	J	no	S	ensino	J	superior	A	no	S	País	J
149880858	ee	superior	graduação	J	no	S	ensino	J	superior	A	no	S	País	J	será	V
149880859	ee	em	no	S	ensino	J	superior	A	no	S	País	J	será	V	reservado	A
149880860	ee	país	ensino	J	superior	A	no	S	País	J	será	V	reservado	A	para	N
149880861	ee	ser	superior	A	no	S	País	J	será	V	reservado	A	para	N	a	H
149880862	ee	reservado	no	S	País	J	será	V	reservado	A	para	N	a	H	atuação	J
149880863	ee	para	País	J	será	V	reservado	A	para	N	a	H	atuação	J	dos	S
149880864	ee	a	será	V	reservado	A	para	N	a	H	atuação	J	dos	S	alunos	J

INTERFACE

Corpus Brasileiro

Palavra	Cat. Gram.
<input type="text" value="assim"/>	<input type="text" value="-----"/>
<input type="checkbox"/> KWIC <input type="checkbox"/> Ordenar: <input type="radio"/> ESQ <input checked="" type="radio"/> NOD <input type="radio"/> DIR	<input type="button" value="Pesquisar"/>
pesquisa avançada	
Contexto	
<input type="text" value="que"/>	<input type="text" value="0"/> <input type="text" value="1"/>
Registro/Gênero	Resultados
<input type="text" value="artigo (acadêmico)"/>	<input type="text" value="- 100 -"/>

INTERFACE

Resultados: 18 - aproximadamente 427306 ocorrências para assim (0.463 segundos)

N	Modo	Registro/gênero	Ocorrência
1.	escrito	artigos acadêmicos	164251
2.	escrito	textos de jornal	99898
3.	escrito	teses	99024
4.	escrito	documentos sobre educação	57142
5.	escrito	educação	2204
6.	escrito	literatura em geral	1320
7.	escrito	textos religiosos	1251
8.	falado	pronunciamento	941
9.	escrito	manual de informática	404
10.	escrito	revistas	309
11.	escrito	roteiros (cinema/tv)	205
12.	escrito	crônicas	101
13.	escrito	literatura - contos	87
14.	escrito	legislação	71
15.	escrito	bula de remédio	68
16.	falado	narração de futebol	14
17.	falado	debate político	14
18.	escrito	horóscopo	2
-	-	Todos	<u>427306</u>

527 occurrences,
0.397 seconds

Resultados: 527 (em 0.397 segundos) - Exibindo página 1 de 6

1. [02] nacional e internacional) apropriados às águas marinhas de forma clara e didática , permitindo **assim que** haja uma padronização e nivelamento metodológico entre os profissionais que trabalham nesta área .
2. [02] DSV Stena Marianos , no porto de Arraial do Cabo em janeiro de 1997 . **Assim que** chegamos , deparamo-nos com dois mergulhadores que acabavam de sair da câmara de saturação
3. [02] estoques do cliente através do EDI e providenciar o enchimento do tanque de produto , **assim que** o ponto de reposição for alcançado , até um nível predefinido ou conforme a
4. [02] que o essencial em tudo na vida é o amor vivido , partilhado . É **assim que** Deus ama todos os seres humanos . Educar assim não será um dos melhores
5. [02] inovações não poderiam fluir no velho esquema de centro planejador ? periferia executora . Tanto **assim que** os projetos , especialmente de São Paulo e Minas Gerais , foram amplamente discutidos
6. [02] superavam a sua malfadada , mas bem conhecida " pedagogia da repetência " . Foi **assim que** , em particular no Estado de São Paulo , chamados a participar nos anos
7. [02] conversava freqüentemente , teve sucesso quase imediato : não só teve sua ajuda amplamente aceita **assim que** chegou como também não levou mais do que cinco minutos para acertar toda a
8. [02] elevado (-3,01) , sugerindo que esta repetência está associada a outras anteriores.2 Confirma-se **assim que** a repetência apenas faz o aluno distanciar- se de seus colegas , não existindo
9. [02] Bernardo da Silva , presidente da UHC do Distrito Federal , nos anos 60 . **Assim que** Getulio Vargas fora eleito Presidente da República , estivemos com Sua Excelência expondo-lhe as
10. [02] que dar sorte / Na situação . (Alencar , 1980 : 169-70) Vemos **assim que** , nesse registro popular , o tema da mestiçagem é protagonizado pela imagem polissêmica
11. [02] de estudar a História da África , independente de qualquer outra motivação . Não é **assim que** fazemos com a Mesopotâmia , a Grécia , a Roma ou ainda a Reforma
12. [02] altos níveis competitivos e tecnológicos , especialmente nessa área de exportação de serviços . Tanto **assim que** , no caso dos projetos de consultoria , engenharia e construção , o Brasil
13. [02] acadêmica , política ou militância . Inserimos a militância nessa nossa abordagem pois Moura , **assim que** amadureceu , podemos dizer , intelectualmente , como ele mesmo nos revela , conseguiu
14. [02] terras da Nação " (terras baldias) , capítulo principal da lei . Foi **assim que** , por exemplo , o candidato à prefeitura do Movimiento de las Comunidades Negras
15. [02] Bernardo da Silva , presidente da UHC do Distrito Federal , nos anos 60 . **Assim que** Getulio Vargas fora eleito Presidente da República , estivemos com Sua Excelência expondo-lhe as
16. [02] que dar sorte / Na situação . (Alencar , 1980 : 169-70) Vemos **assim que** , nesse registro popular , o tema da mestiçagem é protagonizado pela imagem polissêmica
17. [02] de estudar a História da África , independente de qualquer outra motivação . Não é **assim que** fazemos com a Mesopotâmia , a Grécia , a Roma ou ainda a Reforma
18. [02] altos níveis competitivos e tecnológicos , especialmente nessa área de exportação de serviços . Tanto **assim que** , no caso dos projetos de consultoria , engenharia e construção , o Brasil
19. [02] acadêmica , política ou militância . Inserimos a militância nessa nossa abordagem pois Moura , **assim que** amadureceu , podemos dizer , intelectualmente , como ele mesmo nos revela , conseguiu
20. [02] vasoconstritor , e 1 mL de fentanil (50 µg) pelo cateter peridural , **assim que** a paciente referisse dor (a partir do momento que recobrava a consciência)
21. [02] o ópio , que era empregada com eficiência para a sedação de crianças 5. Era **assim que** a Deusa Ísis sedava seu filho Hórus 6. O ópio e , provavelmente ,
22. [02] Deméter , divindade grega , conhecia as propriedades sedativas e hipnóticas do ópio , tanto **assim que** , desesperada com o estupro de sua filha Perséfone , ingeriu essa substância para
23. [02] assim como deixou extraordinariamente abalado nosso amigo e colega Loïc . Contamos com seu artigo **assim que** ele possa voltar a trabalhar conosco . Referências Bibliográficas Imperialiste " . Actes de
24. [02] venda elevado que , de qualquer modo , liquidaria o livro em questão . É **assim que** certas decisões de pura comercialização editorial orientam a pesquisa e o ensino universitários no
25. [02] função das tradições históricas e constituem sempre um motivo de lutas na história . É **assim que** uma análise comparativa aparentemente rigorosa e generosa pode contribuir , sem que seus autores
26. [02] da qual não cabe , conforme mostra Hofbauer , a polarização branco/negro . Pode-se afirmar **assim que** , ao contrário da instrumentalização do repertório cultural afro-brasileiro e do Atlântico Negro para
27. [02] ser entregues aos comerciantes " brasileiros " lá estabelecidos (Rodrigues , 1988:98) . **Assim que** aportou na baía , após o malogro da viagem e dos negócios , vários
28. [02] ao caminho da salvação . (Zurara , 1973 , cap . XVI:86) E **assim que** onde antes viviam em perdição das almas e dos corpos , vinham de todo
29. [02] fazerem passar por livres . Mas ela o fez em área nitidamente rural , tanto **assim que** posicionou-se como " agregada " (Araújo , 2000) . Desse modo ,
30. [02] de rendas em lugar de como força de trabalho na própria casa ou explorando-os , **assim que** os ensinavam como artesãos e comerciantes e logo até os dotavam em ocasiões (
31. [02] lugar de nascimento dos indivíduos e de seus pais teve **que** ser desenvolvido . Foi **assim que** em 1980 apareceu a noção de " ancestry " (literalmente , ascendência)

94. [02] fatores de produção , como mão-de-obra barata , frouxa regulamentação e debilidade sindical . Foi **assim que** , entre 1960 e 1980 , a fatia rural do emprego fabril passou ,
95. [02] , que nos últimos anos têm crescido as reações internas a esse modelo . Tanto **assim que** , hoje , dificilmente , pode-se afirmar que a magistratura constitua um corpo homogêneo
96. [02] Zulaiê Cobra Ribeiro . 16 Na Argentina , o presidente Kirchner promoveu mudanças no Judiciário **assim que** assumiu o cargo . Dois magistrados renunciaram ao cargo de ministro da Corte Suprema
97. [02] cit. , p. 88) . O abalizado jurista conclui com asseverar : " É **assim que** o autoritarismo ascendente vem ganhando terreno no Direito Processual , à custa da subtração
98. [02] romances como nos contos , a da sua psychologia é identica a esta , mostrando **assim que** os seus processos litterarios , como proprios e pessoaes que são , derivão do
99. [02] chamá-la à esfera das idéias novas , das reformas , dos princípios dominantes . É **assim que** o teatro nascerá e viverá ; é assim que se há de construir um
100. [02] , dos princípios dominantes . É assim que o teatro nascerá e viverá ; é **assim que** se há de construir um edifício de proporções tão colossais e de um futuro

[Imprimir](#) - [Salvar como XLS](#) - [Salvar como TXT](#)

[Anterior](#) | [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [8](#) | [9](#) | [10](#) | [Próxima](#) - [Página inicial de busca](#)

Print, save as text or Excel



8th Brazilian Corpus Linguistics Meeting
3rd Brazilian Computational Linguistics School
Rio de Janeiro State University

November 11-14, 2009
elc2009@corpuslg.org

www.corpuslg.org/elc

Invited speakers: Mike Scott, Susan Hunston,
Alice Deignan, Steven Bird







THANK YOU!

- <http://corpusbrasileiro.pucsp.br>

José



Eli

