The advantages of using relational databases for large historical corpora

Mark Davies Brigham Young University http://davies-linguistics.byu.edu/ mark_davies@byu.edu

Outline

- Introductory comments
- VIEW/BNC (100m words)
- Corpus of Historical English (OED, 37m)
- Corpus do Português (July 2006, 45m)
- Examples of annotation via relational databases
- Proposed 200m word corpus of historical English



 Function of intended use
 Product of biases of corpus creators
 CED, CEEC, etc. perfectly fine for historical pragmatics
 Bottom line: not make size a limitation

Accessibility

Via Web
Cost (free?? to end-users)
User interface (may be a much larger and diverse audience than we had first intended)

Searchability

Spelling, morphology: wildcards
Syntax: POS tagging
Semantics: Collocates, comparing collocates, synonyms, user-defined lists
[Pragmatics]
Q: Is what we study a function of the

corpus, or what we really want to study?

Size: Corpus del Español (<u>www.corpus</u>delespanol.org) ; 100m words 1200s-1900s

	DAVIES/NEH] Corpus del español - Micr	osoft Interne	t Explorer										<u>_ 8 ×</u>	
Ē	e <u>E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp)												
) Back 🔹 💮 🖌 🔀 💋 🎾	🔵 Search 🦻 🚽	Favorites	<i>C</i>	- 🎍 🧕	- 🗾 🏾	🗉 🧐 🌏) 🛍 🕅	1 🔏					
Ado	ress 🙆 http://www.corpusdelespanol.org/												🗕 📑 Go	
G	oogle -	Search 👻	S 🔊 🖇	Ӯ 3999 blocke	ed 🛛 🐴 Cheo	:k 👻 🗐 Au	itoFill 🕒 📧	Options 🥖	9					
1	Mark Davies					DAÑO		_ ·			Usaresp	añol		
	BYU/ISU/NEH		00	RPUS	DEL ES	PANOL	_			I	ntroduction	1/ Tour		
<u>SEA</u>	RCH SORT B	<u>v</u>	LIMITS	<u> </u>	INDEX	<u>+</u>	GROUP BY	<u>SHOW</u>	<u>l</u>	VERIFY	RES	ET		
hor	nbre *.adj * 1900s	-				100	FORMS 🚽	HITS	-	NO 🚽	SL	JBMIT		
ACC	aerouun													
There	are three options to select words/phrases	for the keywor	d in context d	isplay in the l	ower window [more] (<u>Pr</u>	oblems seeing	examples bel	<u>ow?</u>)					
SEE CONTEXT HELP HISTORICAL 1900s														
WORDS / PHRASES V 12005 V 13005 V 14005 V 15005 V 16005 V 17005 V 18005 V Total V Oral V Lit V Text														
1	HOMBRE JOVEN						<u>4</u>		<u>39</u>	<u>85</u>	<u>13</u>	<u>66</u>	<u>6</u>	
2	HOMBRE MODERNO								Z	<u>82</u>	12	<u>6</u>	<u>64</u>	
3	HOMBRE HERIDO						5	1	<u>4</u>	47	<u>44</u>	3		
4	HOMBRE ALTO				1	7	4	2	20	40	3	35	2	
5	HOMBRE FUERTE	V	1	1	8	16	1	7	16	31	2	15	14	
6	HOMBRE BUENO	V	93	1	4	23	8	21	37	30	8	18	4	
7	HOMBRE NUEVO			-	-	11	3	7	21	26	2	12	12	
	HOMBRE MADURO	2					2	4	11	26	8	17	1	
							-	-		20	2	<u></u>	÷ _	
< <		a ah a la A rra - 1	aduma Ta	and the	ale de tra						KEYW	ORDS IN	CONTEXT	
rore 1/1	SIGLO TEXT	n the lettmost	column, To re	r-sort, click of RE-SO	r the desired (RT BY: <u>L2</u>	L1	C	<u>R1 R2</u>						
1	19_ORAL Entrevista (ABC)		nosotros ten	emos que trat	arla como ta	. El hon	nbre moderno	de hoy, ap	parte de perde	r respeto a r	nuchas otr	as		
2	19_ORAL Habla Culta: Bogot 19_ORAL Entropyists (ABC)	s	i, es interesar	ntisimo; él va	a hablar sobr roto dol alma	eel hon del bon	nbre moderno	en pugna	con con el a	mbiente; mu	y interesai coiritual d	nte •		
4	19 ORAL Habla Culta: San J.,	est	e momento, e:	stán siendo e:	kaminadas po	rel hon	nbre moderno	para dete	rminar si, en r	realidad está	n sirviendo	-), si		
5	19_ORAL La Pecadora, Haban	president	es americano	s. ENRIQUE:	El nuestro es	un hon	nbre moderno	y abierto.	UGARTE: Pe	ro demasiad	o convenci	do de qu	e	
6 19_ORAL Entrevista (ABC) iglesias laicas que nos harían mucha falta. El hombre moderno ha querido prescindir de las creencias														
7	19_ORAL Entrevista (ABC) 19_ORAL Habla Cultar San 1		es la perdida :	de seguridade	s metatisicas	El hon	nbre moderno	es como s	i perdiera el r	espeto al oro de revolució:	len de la	tiono		
9	19 ORAL Entrevista (ABC)	D	ueda echar su	s raíces la so	mbra errante	del hom	nbre moderno,	Francia de	ese proceso i esdeñó muv or	onto a Branc	usi, ques, no:	ebe su		
10	19_ORAL Habla Culta: La Pa	yo ha	go, čno? ENC	Entonces, t	ú has elegido	un hon	nbre moderno.	INF.ECl	aro. INF.AH	oy en día to	los son			
<u>11</u>	19_ORAL Entrevista (Chiapas)	ima	ginación y el	pensamiento ·	y la inversión	del hom	bre moderno;	en este m	omento de la	plática me d	esconecté	y ya no		
12	19_ORAL Tres esqueletos y	de la co	rrupción? CIE	RRA LA TOM	A. ¿El <mark>origen</mark>	del hom	bre moderno?	Se dice qu	je proviene de	Africa. O m	ás			
120														

Features: VIEW: view.byu.edu (British National Corpus; 100m words)

🚰 [DAVIES] Variation in English Words and P	hrases (VIEW) -	Microsoft Internet Explorer					_ 8 ×					
Eile Edit View Favorites Tools Help							2					
🛛 😋 Back 🔹 🕥 🖌 💌 😰 🏠 🔎	Search	vorites 🧭 🔗 - 嫨 💿 - 🗌	, 💽 🧶	I 🥏 🇯	1 🕸 %							
Address 🕘 http://view.byu.edu/							💌 🔁 Go					
Google -	Search 🝷 🍣 🛛	🚿 嶜 3999 blocked 🛛 🖑 Check 👻	🗒 AutoFill 🧧	🔰 巀 Optic	ons 🖉							
SEARCH STRING (HELF		W:VARIATION IN E	NGLIS	SH W	ORDS AND	PHRASES						
WORD/PHRASE break * [n*]												
(INSERT TAG) -select-	Mark	Davies / Brigham Y	oung l	Jniver	sity							
SEARCH CUSTOMIZED LISTS		RD TO SEE IN CONTEXT				DON'T SHOW PART						
DISPLAY (HELF) [DISTRIB]	FOR FREQUENCY IN ALL 70 REGISTE	RS		[HE	LP IN INTERPRETING	THIS VIEW]					
© TABLE DISTRIB WORD/PHRASE TOKENS PER MIL IN REG1 PER MIL IN REG2 REG2 O CHART C CHART DISTRIB WORD/PHRASE TOKENS TOKENS PER MIL IN REG1 PER MIL IN REG2 REG2												
CHART	1	BREAK THE NEWS	32		1.98		76.22					
C SURROUNDING (HELP	2	BREAK THE SILENCE	22		1.36		52.40					
-select- 💌 5 💌 5 💌	3	BREAK HER HEART	8		0.49		38.11					
SORT BY (HELF	2 4	BREAK THE SPELL	10		0.62		23.82					
FREQUENCY	5	BREAK HIS NECK	8		0.49		19.05					
O PERCENT	6	BREAK MY HEART	8		0.49		19.05					
REGISTER 1 (HELF	7	BREAK A PROMISE	4		0.25		19.05					
SPOKEN	8	BREAK FOR LUNCH	4		0.25		19.05 💌					
FICTION 5	KEYY	ORDS IN CONTEXT				More information	•					
NON-FICTION MISC	LIMIT BY	PART OF SPEECH: NO	LIMIT P	Y GENRE:	FICTION [SEE ALL]		(HELP)					
	1 CAM	through. "Party, party," she yelled (out to break	the spel	l and Hosanna ran from	n where he had been	sitting watchin					
REGISTER 2	2 FP1	It was a grey-haired British sergear	nt who was <u>t</u>	t he first to	- break the spell of in	nmobility and shock, b	by moving forv					
IGNORE MIN	3 GOL	give us a song?" Hope welcomed <u>th</u>	<u>ie</u> chance to	<u>break</u> th	<u>e spell</u> he had cast, s	eeing now that his bes	t course lay					
FICTION	4 G1S	him into her. He lay still, not wantin	g to <u>break</u>	<u>the spell,</u>	attempting to work ou	it what was different. I	Before					
NEWS LIMIT?	5 G1S	release her, and she had no will wit	h which to <u>b</u>	reak the	<u>spell</u> . Even so, <u>the</u> re	was a developing dua	lity in her					
	6 HA7	took Johnny into <u>the</u> cottage, might	t this in som	e way <u>bre</u>	ak the spell and spoi	l <u>the</u> magic? Would Jo	hnny simply d					
	7 HGT	will until, steeled and streng <u>the</u> ned	by her desp	erate nee	d to <u>break the</u> spell,	Merrill wrenched hers	elf away to op					
OPTIONS (HELF	8 HGY	listening, enjoying <u>the</u> peace. Roby	n waited for	him to <u>br</u>	eak the spell of this (wonderful, strangely c	ompanionable					
# HITS 100	9 HTM	as his because talking seemed an ir	ntrusion, as	though it n	night <u>break the</u> <u>spell</u>	of enchantment she f	elt wrapped in					
SEE POS TAGS NO 🔽	10 HTM	the fire burning in Bird Spirit Land.	Your flame	will <u>break</u>	the spell." Fierce Eye	es scowled. "Get this c	ver					
SEARCH RESET	T						Þ					

BNC / VIEW features

Word document

Corpus of Historical English: view.byu.edu/che

🚰 Corpus of Historical English [I	DAVIES] -	Microsof	t Inte	rnet Explorer										_ 5	×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites	<u>T</u> ools <u>H</u> e	elp													7
🛛 🚱 Back 👻 🕥 🖌 😫 💈	ا 🏠 🛿	Searc	:h 🤞	🏷 Favorites 🛛 🔗 🗸	è 🖸 🗸		9	11 🖉	20	8					
Address 🙆 http://view.byu.edu/d	che/x.asp													- 🔁 🛛	io
Google	•	G Searc	h +	🕥 🛛 🚿 🔄 3999 blocked	🍄 Check	• 🗐 Auto	oFill 🕒	🔁 Optior	ns 🖉						
Based on the Oxford English Dictionary (37 million words, Old English - Modern English) SEARCH Click on the number in a column to see [KEYWORD IN CONTEXT] display															
SEARCH		Click	on t	he number in a column to se	e [KEYWOR	ND IN CO	NTEXT] d	lisplay						[HELP]	
WORD(S)/PHRASE(S)	HELP			CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s	
light		1 +		LIGHT				137	149	812	1988	1328	4473	4078	
[CUSTOMIZED LISTS] [+OE]		2 +		DELIGHT					1	174	418	241	435	218	
• FREQUENCY (TABLE)	HELP	3 +		FLIGHT				13	19	97	273	182	399	904	
•• FREQUENCY (TABLE) 3* FEIGHT 13 19 97 273 182 399 904 •• FREQUENCY (CHART) 4* LIGHTS 6 43 218 138 432 658															
C RELEVANCE				SLIGHT				20	5	30	189	155	761	544	
C SURROUNDING WORDS		5		DELICHTS				20	J	27	144	77	07	64	
		6		DELIGHTS					-	37	144		97	64	
		7 +		LIGHTNING				1	2	51	132	112	242	180	
	HELP	8 +		LIGHTLY				25	42	83	108	51	233	221	
				DELICUTED .						- 20	- 00		474	100	믁
FREQUENCY IN CENTURY	HELP		[KEY	WORD IN CONTEXT									ormation		
		1	1509	BARCLAY: Shyp of Folys	Because t	he <u>lightr</u>	ning or th	nunder vie	olent s	uffreth th	ee and th	ny house	to be vnt	rent.	
DISPLAY		2	1555	T. PHAER: neid	And from	the skye:	s the <u>ligh</u>	<mark>itning</mark> fy	ers do fla	ashe wtigi	risly stea	uen.			
SORT BY CENTURY	HELP	3	1556	W. TOWERSON: Voy.	The 4.day called Ter	we had 1 nados.	terrible th	nunder an	d <u>lightn</u>	inq , with	exceedir	ng great (gusts of r	aine,	
	HELD	4	1558	PHAR: neid	Me the fat	her of Go	ods Beb	plasted wi	th his <u>lia</u>	i <mark>htning</mark> M	vynd.				
RAW # -		5	1561	Burn. Paules Ch.	On Wedne	sdayw	vas seene	e a marue	ilous gre	at fyrie <u>li</u>	ightning	l.			
# UTTC		6	1561	T. NORTON: Calvin's Inst.	To set the	aire on f	ier with <u>I</u>	ightning	flames.						
# HIIS	HELP	7	1563	W. FULKE: Meteors	If <u>lightni</u>	n q or any	/ other in	flamation	be in th	e upper p	art of the	ese cloud	s.		
100		8	1563	W. FULKE: Meteors	The most	dangerou	us, violen	t and hur	tfull kind	of <u>lightn</u>	ling is ca	alled Fulm	ien.		
SUBMIT RESET		9	1563	W. FULKE: Meteors	Fulgur is t	hat kinde	of <u>light</u> .	ning whic	ch followe	eth thund	er.	6			
		10	1263	W. FULKE: Meteors	heard.	ine <u>liqnti</u>	ning app	eare unto	ius, a go	od pretty	While be	rore the	thunderc	ар ре	
		11	1563	FULKE: Meteors	The Sea-C to be cove	Calfe is no ered with	ever hurt their skir	with <u>ligh</u> nnes.	itning: v	vherefore	the Emp	erours te	nts were	woont	•
(e) Done												📃 🙆 I	internet		_

Corpus of Historical English: Size and distribution

Century	# words
1000s	207,594
1100s	53,359
1200s	313,717
1300s	1,077,330
1400s	1,388,448
1500s	3,182,053
1600s	5,127,445
1700s	3,688,076
1800s	10,422,785
1900s	11,063,571
TOTAL	36,524,378

Simple frequency of word or phrase over time: turn on

🚰 Corpus of Historical English [DAVIES] - M	licrosoft Internet E	kplorer						S (10)	🏙 💌 📓	0 0	<u> </u>			
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp											1			
🖙 Back 👻 🤿 🗸 🔯 🖓 Search	😹 Favorites 🛛 🛞 M	edia 🏼 🌀	B- 🥔 🛛	- 🗐 📖	,									
Address 🕘 http://view.byu.edu/che/x.asp										-	∂Go Links '			
Based on 37 million words of text from Old English to Present-Day English		COR	PUS Creat	OF edb	HIS y Ma	TOR Irk D	ICAL avie	. EN s / E	GLIS BYU	H				
SEARCH	Click on a bar to	o see [KEY] er below ti	WORD IN C	DNTEXT] dis for table di	splay						[HELP]			
WORD(S)/PHRASE(S) HELP	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s			
C FREQUENCY (TABLE) HELP FREQUENCY (CHART) RELEVANCE SURROUNDING WORDS														
LIMITS	PER MILLION	0.0	0.0	0.0	0.0	0.0	0.0	0.2	1.4	2.3	5.0			
SURROUNDING WORDS HELP	FREQUENCY	0	0	0	0	0	0	1	5	24	55			
FREQUENCY IN CENTURY HELP	[KEYWORD	IN CONT	EXT]						Mo	ore informatio	n 💌			
DISPLAY SORT BY CENTURY 1900s	1 1900 The I to ge 2 1901 Pleas 3 1902a White 4 1904 Such	Newlands g et considera se, Nanna, e she was g objections	round is the able <u>turn or</u> don't <u>turn (</u> gone the ma to the mea	most diffic the ball. <u>In</u> the dark n from the ningfulness	ult to make gas compar of the mec	runs on in ny called to hanical ima	the whole c <u>turn on</u> th ge as <u>turn</u>	f South Afr e meter. <u>on</u> the diffi	ica, the bov culties of de	vlers always efining mase	s being able – s, length,			
RAW #	5 1909 I dor	, and their n't know the	combination e moment h	s, cannot b e delivers t	e discussed he ball whic	in this con h way it wil	nection. I <u>turn on</u> p	itching.						
# HITS HELP	6 1917 If so 7 1919 She'l	meone will I screen we	kindly re- ell, and she'	carbon the s one of the	arc for us a e few that ca	and <u>turn or</u> an <u>turn on</u>	<u>n</u> the electri the tears w	c current w hen she wa	e shall be n ants to.	eady to star	rt showing.			
	9 1934 Cove 10 1942 Imm	 8 1926 Most fat cows and heavy heifers lost around 25c and in instances the down-<u>turn on</u> better grades was even greater. 9 1934 Cover One with a Gun (v. phr.): to <u>turn on</u> the heat. 10 1942 Immediately you are grounded, <u>turn on</u> your back and unlock the quick release mechanism. 												
	11 1946 Thei 12 1949 Thei 12 1950 Thei	mmediate majority	action is to ; were Zionis	turn on the ts who at	e reciprocal the eve of	course. the war had	l been waiti	ng for their	turn on th	e immigrati	on quota.			
(A)	13 1953 The 1 14 1953 Philo	rimmer wit sopher's ai	rguments wi	t urn on his hich <u>turn o</u>	n these terr	s nand ove ns are apt,	r nand as th sooner or l	ater, to stai	gs over. rt to rotate i	idly. There i	is nothing			

Spelling changes: vn* (chart)

🚰 Corpus of Historical English [DAVIES	5] - Microsoft Internet	Explorer						- E	🎇 💌 📓	00	0 -	8 ×		
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>F</u>	<u>H</u> elp													
⇔Back • ⇒ • 🙆 😰 🚮 🥘 Sea	arch 📓 Favorites 🍘	Media 🌀 🛙	3- 🎒 🖸	- 🗉 📖										
Address 🙆 http://view.byu.edu/che/x.asp	p									•	ểGo Li	nks »		
Based on 37 million words of text from Old English to Present-Day English	Based on 37 million words of text from Old English to Present-Day English CORPUS OF HISTORICAL ENGLISH Created by Mark Davies / BYU SEARCH Click on any bar or number below for a list of matching strings													
SEARCH	Click on any b	ar or number	below for a	a list of mat	ching string	ļs					[HELP	1		
WORD(S)/PHRASE(S) HELP	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s			
C FREQUENCY (TABLE) HELP FREQUENCY (CHART) C RELEVANCE C SURROUNDING WORDS	PER MILLION	38.5	187.4	1.711.7	3,710.1	3.917.3	4,554.9	1.783.1	4.9	0.8	0.6	-		
LIMITS	FREQUENCY	8	10	537	3997	5439	14494	9143	18	8	7			
SURROUNDING WORDS HELP	2													
FREQUENCY IN CENTURY HELP	2 [KEYWOR 1 1908 Well, 1	D IN CONTE	хт] Carthlick,	but I haven	't <u>followe</u> a	l it up muc	:h. To tell γε	e the truth,	Mo I class 'em	re informatio all alike (n oriests,	-		
SORT BY CENTURY HELP	2 1951 The St follow 3 1955 Beaur 4 1956 We follow	ns, salvos, an ate Legislatu y ed it up with nont College lowed it, up	nd all the lo ire at Alban th a narcotion, who won t to the sprin	t of 'em. ly ordered a c clinic of its their juniors g hole on th	an inquiry, v s own. at Molesey ne edge of t	which met la a week ag he flat land	ast week. A o, <u>followec</u> I, a no-good	nd this wee <u>1 it up</u> by w bit of bog	k the Ameri vinning the j hole on the	can Legion unior-senio edge of ara	has or eights. able land.			
# HITS HELP	5 1961 Whys many 6 1965 I <u>fo</u>	he should ha guns. llowed it up	ve <u>followe</u> with a ri	e d it up bea	its me. You ith the scho	didn't ask l ool cap.	her? Of co	ourse not! A	is I told you	, she carrie	es too			
										Internet				

Spelling changes (table): rank-frequency listing by 1500s

曫 Corpus of Historical English [DAVIES] - I	Microsoft In	ternet Explorer						B 🛃	8 💌	20	00	3			
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u>	ools <u>H</u> elp	ı														
🗢 Back 🔹 🤿 🗸 🙆	Q Search	🚡 Favorite	es 🎯 Media 🧭 🛃 🎒) • E	11											
Address 🕘 http://view.byu.edu/c	he/x.asp		· · · · · · · · · · · · · · · · · · ·									• @	'Go Lin	ıks '		
Google -	- 0	🖁 Search 🕞	💽 🛛 🕸 1431 blocked 🛛 🗳 Che	eck 👻 😽	AutoLink	. +										
Based on 37 million words of text from Old English to Present-Day English Created by Mark Davies / BYU Click on the number in a column to see [KEYWORD IN CONTEXT] display																
SEARCH		Click or	the number in a column to se	e [KEYW	ORD IN	солтехт] display						[HELP	1		
WORD(S)/PHRASE(S)	HELP		CENTURY 1000s 1100s 1200s 1300s 1400s 1500s 1600s 1700s 1800s 1900s + VNTO													
vn*		1 +	1 ⁺ VNTO 253 979 3459 1573													
[CUSTOMIZED LISTS] [+OE]		2 +	VNDER			77	429	374	1208	736	2					
FREQUENCY (TABLE)	HELP	3 +	VNTILL				8	10	177	156						
C FREQUENCY (CHART)		4 +	VNLESSE					1	118	124						
C RELEVANCE		5 *	VNDERSTAND				10	11	112	82	1					
C SURROUNDING WORDS		6 +	VNDERSTANDING				6	1	92	95						
LIMITS		7 +	VN			4	85	97	90	55	2	1	3			
SURROUNDING WORDS	HELP	8 +	VNIUERSALL					1	70	36						
5 - LEFT/RIGHT 5 -			COMP.				-	4.4		- 20						
FREQUENCY IN CENTURY	HELP	INT	RODUCTION								More info	rmation		-		
FREQ DISPLAY FREQ DISPLAY RAW # # HITS 100 SUBMIT	HELP HELP	This o use t years Pleas Th Ex morp years	corpus of historical English is to o the corpus to find the freque s. e feel free to take a look at eir ree minute tour , an overvie tended Sample Searches , o hology (word forms), syntax (s.	aken fror ncy and ther or th w of the concrete grammal	m 37 mi use of w me follow main fea example r), lexis (llion wor ords, phr ing pages tures of t s of how (vocabula	rds of tex ases, and s, which pr the corpus the corpu ary), and s	t from Old constructio rovides sar (with exar s can be us remantics (English to ons in each nple searc mples of a sed to exa word mea	Present- n century hes of th ctual que mine cha ning) dur	Day Eng during t e corpus ries) inges in s ring the p	lish. You he past 1 : spelling, past 1000	can .000			

Chart display (by century): to * [davies:up] [CUSTOMIZED LISTS]



🕝 Internet

Table display (1500s): to * [davies:up] [CUSTOMIZED LISTS]

🚰 Corpus of Historical English [[DAVIES] -	Microsoft	Internet Explorer										_6	<u>ı ×</u>	
<u> File E</u> dit <u>V</u> iew F <u>a</u> vorites	<u>T</u> ools <u>H</u> e	elp												•	
🛛 🌏 Back 👻 🕘 👻 😰) 🏠 [Search	n 🧙 Favorites 🥝 🔗	- 👌 💿	-	E 🗐		1	8						
Address 🙆 http://view.byu.edu/	/che/x.asp	-												<u>5</u> 0	
Google -	-	G Search	🔹 🗣 🚿 🖓 3999 blocke	ed 🛛 🗳 Check	- 🗐 AU	utoFill 🕒	💫 Optic	ins 🏼 🖉							
Based on the Oxford English Dictionary (37 million words, Old English - Modern English) CORPUS OF HISTORICAL ENGLISH Created by Mark Davies / BYU SEARCH Click on the number in a column to see [KEYWORD IN CONTEXT] display															
SEARCH		Click	on the number in a column	to see [KEYW	ORD IN C	ONTEXT]	display						[HELP]		
WORD(S)/PHRASE(S)	HELP		CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s		
to * [davies:up]		1	TO TAKE VP 3 14 6 TO SET VP 13 11 11												
[CUSTOMIZED LISTS] [+OE]		2	Z TO SET VP 13 11 3 TO MAKE VP 6 10 9												
C FREQUENCY (TABLE)	HELP	3	3 TO MAKE VP 6 10 9 4 TO COME VP 1 7 3												
FREQUENCY (CHART)		5	4 TO COME VP 1 7 3 5 TO LIFT VP 7 1 7 1												
C RELEVANCE		6	TO BEARE VP						7	4					
C SURROUNDING WORDS		7	TO TAKE UP					1	7	38	46	70	70		
LIMITS		8	TO SET UP						7	43	36	54	117		
	HELD	9	TO FILL VP						6	7					
5 - LEFT/RIGHT 5		10	TO MAKE UP						6	45	40	93	85		
FREQUENCY IN CENTURY	HELP	C	USTOMIZED LISTS: MODI	FY LIST							More inf	ormation		-	
			modify a list		LISTS YO	U HAVE (CREATED		M	ODIFY L	IST				
			iniouny a list.		UP M	D	NOW	M D	¬ _						
	HELP	Ac	ld or delete words from you	r list and					up			WORD	DLIST		
		cli	ck on [Submit]										E OF		
PAW #	HELP								lur	2	-	- WOF	RDS		
		R	eturn to main menu							, 					
# HITS	HELP	0	verview of wordlists						vr	ope					
SUBMIT RESET			luppe 🔽												
									s	ubmit f	Reset				

🥝 Internet

ど Done

Morphology: *ly (+1900s -1800s -1700s)

🚰 Corpus of Historical English [[DAVIES] - M	icrosoft In	ternet Explorer							🐝 💌	20	000	3 _ 8	Ι×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> o	ools <u>H</u> elp													
😓 Back 🔹 🤿 🗸 🔯 🖉	🕄 Search	😹 Favorite	es 🛞 Media 🧭 🛃 🎒 🧕] - 🗐 📖										
Address 🙆 http://view.byu.edu/ch	ne/x.asp											•	≻Go Link	s ×
Based on 37 million w of text from Old Eng to Present-Day Engl	ords lish ish		CORPUS Creat	OF edb	HIS ум	STO ark	RIC Da	C <mark>AL</mark> vies	E N ; /	GL] BYU	SH			
SEARCH		Click on	the number in a column to see	EKEYWORI		TEXT] di	splay						[HELP]	-
WORD(S)/PHRASE(S)	HELP		CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s	
*ly		1 *	BASICALLY										220	
[CUSTOMIZED LISTS] [+OE]		2 +	ROUTINELY										53	
FREQUENCY (TABLE)	HELP	3 +	ELECTRONICALLY										47	
C FREQUENCY (CHART)		4 +	REPORTEDLY										39	
C RELEVANCE		5 +	ENVIRONMENTALLY										36	
SURROUNDING WORDS		6 +	UNDERSTANDABLY										28	
LIMITS		7 +	RECURSIVELY										22	
SURROUNDING WORDS	HELP	8 +	GOOGLY										19	
		9 +	HILLBILLY										17	
FREQUENCY IN CENTURY	HELP	10 +	PHONEMICALLY										16	
+19005 -18005 -17005 -180		11 +	MINIMALLY										16	
DISPLAY		EXT	ENDED SAMPLE SEARCHES								More info	rmation		•
SORT BY CENTURY	HELP		, and and a second products of the second pr				, ,	- 						
FREQ DISPLAY	HELP	down, MORE	/downe/doun/doune/dovn/dovi	<u>ne</u> . This will	show the	changes	; in spelli	ng for th	is one wo	ord over 1	the past :	1000 yea	ars.	
# HITS 100	HELP	1. You 2. You 3. Ea	u can easily examine the overa u can see which <u>words ending i</u> irly Modern English (e.g. 1500s Present Day English, which off	ill frequency <u>n -/v</u> (mosti -1600s) use	/ of prefix y adverb ed the cor	kes (e.g. s) occur mparative + + AD1 (<u>pre-</u>) or for the fi e (-er) ar	suffixes rst time nd super <i>Ds delie</i> s	(e.g. <u>-tio</u> in the 19(lative (-e	<u>n</u>) over ti DOs. st) suffixi	ime. es more f	frequent	ly oop	
SUBMIT RESET		easily unlike	risence bay English, which out find cases of long adjectives (aly in the 1800s or 1900s	with the sup	erlative s	uffix in th	he 1500s	(<u>group1</u>	, <u>qroup2</u> ,	, <u>group 3</u>), which v	would be	can	
		SYNT Looki part o	TAX / GRAMMAR ng for changes in grammar is s of speech. Yet by using word al	omewhat c ternates (e.	hallengin g. <i>can/wi</i>	g at the p ill/should	oresent ti , etc for 1	me, sinc the class	e the cor of moda	pus is no Is) or <u>cus</u>	t (yet) ar stomized	notated lists of w	for <u>vords</u> ,	

🔮 Internet

ē

Morphology/lexicon: word roots: *light* (+1900s -1600s)

🚰 Corpus of Historical English [[DAVIES] - M	1icrosoft In	nternet Explorer						E .] 🏙 💆	22	00	3	<u>א צ</u>		
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u>	ools <u>H</u> elp															
🗢 Back 🔹 🤿 🔹 🔯	🔇 Search	🚡 Favorite	ies 🛞 Media 🧭 🛃 🎒	0 - E	13.											
Address 🙆 http://view.byu.edu/ch	ne/x.asp											• 6	≷Go Lin	iks "		
Based on 37 million w of text from Old Eng to Present-Day Engl	ords Iish Iish		CORPU: Crea	SOF ted	Н] Бу	[ST Mar	DRI k Da	CAL vie	- EN \$ /	IGL В Ү І	ISH '					
SEARCH		Click or	n the number in a column to :	see [KEYW	ORD IN C	CONTEXT]	display						[HELP	j 🔺		
WORD(S)/PHRASE(S)	HELP		CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s			
light		1 +	SUNLIGHT								1	104	163			
[CUSTOMIZED LISTS] [+OE]		2 +	LIGHTWEIGHT										111			
FREQUENCY (TABLE)	HELP	3 +	HIGHLIGHT										49			
FREQUENCY (CHART)		4 +	LIGHT-WEIGHT									8	46			
C RELEVANCE		5 +	HIGHLIGHTS										41			
SURROUNDING WORDS		6 +	6 * ENLIGHTENMENT 1 15 41													
LIMITS		7 +	SPOTLIGHT										34			
SURROUNDING WORDS	HELP	8 +	HEADLIGHTS										34			
			LIMELICUT									-				
FREQUENCY IN CENTURY	HELP	TH	REE MINUTE TOUR								More inf	ormation		_		
		The	Corpus of Historical English w	ill allow yo	u to see:									-		
	HELD	1	The frequency of words ()	wane, clene	ed, or fed	l^er/fethe	r/feather) nhrasi	es (thou	hast or ti	urn on), o	r narts o	of			
1900s -			words (e.g. <u>heart*</u> (sorted	by freque	ncy in th	e <u>1500s</u> o	r the <u>180</u>	<u>()</u> 0s), <u>*shi</u> t	o (by <u>140</u>	<u>Os</u> or <u>170</u>	<u>10s), *frie</u>	<u>nd*</u> (by				
FREO DISPLAY	HELP		<u>1700s</u>), or the substring <u>al</u>	<u>??m*</u> (by <u>1</u>	<u>.800s</u>). (N	lote: <u>char</u>	<u>ts vs. list</u>	<u>s)</u>								
RAW #		2.	1400s/1500s to 1600s [ent	ering / lea	ving] or *	*light* 16	00s to 19	ge from ()Os [ente	ring / lea	ry to and ving])	mer, (e.g	, "ment				
# HITS	HELP	3.	Which words occur mos	t often wi	th other	s (often r	elated to	the mear	ning of th	e word),	e.g.:					
100			 words immediately after words within five words 	hard (<u>150</u> of breaks (<u>0s, 1900:</u> 1700s, 1	<u>s)</u> [sorted 900s) or <i>f</i>	by uniqu See (160)	eness: <u>15</u> 15. 1900s	5 <u>00s</u> , <u>190</u> :)	<u>0s]</u>						
SUBMIT RESET		4.	Changes in which words	occur m	ost ofter	n with ot	hers (sor	netimes a	a clue to :	changes i	n meanin	g), e.g.				
			words after hard [<u>+1500</u>	<u>ls/-1900s</u> ,	-1500s/+	1900s] or	r before <i>n</i>	neat [<u>+15</u>	500s/-190	<u>10s, -150</u>	<u>)s/+1900</u>	<u>s]</u>				
			words near <i>market</i> [<u>+16</u>	0057-1900	s, <u>-1600s</u>	<u>7+19008]</u>	or match	[<u>+1600</u> 9	<u>7-1900s</u> ,	-1600s/4	19005]					
		Note that in each case, the link runs an actual query against the corpus and returns the results (i.e. these are not "canned"														
	searches). Feel free to see how the search form has been filled out, and then modify this to create your own queries.															
		sear	ches). Feel free to see how th	ie search fi	orm has l	been filled	l out, and	then mo	dify this t	o create	your own	queries.	mea			

Lexicon: New words in 1900s: * (+1900s -1800s)

🚰 Corpus of Historical English [D	AVIES] - Mi	icrosoft Int	ernet Explorer							🛯 🏙 💌		0	0	<u> I</u> <u>×</u>		
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> o	ols <u>H</u> elp															
🕁 Back 🔹 🤿 🗸 🙆 🖄	🔇 Search	😹 Favorites	: 🎯 Media 🧭 🛃 🎒 💽	- 🗏 👔	1											
Address 🕘 http://view.byu.edu/ch	e/x.asp												∂Go Lin	iks "		
Based on 37 million wo of text from Old Engl to Present-Day Engli	ords ish ish		CORPUS Creat	OF edl	HI by M	STC 1 a r k	RI Da	CAI vie	. EN \$ /	IGL BYU	ISH					
SEARCH			CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s			
WORD(S)/PHRASE(S)	HELP	1 *	RADIO					2					1827			
*		2 +	TELEVISION										1300			
[CUSTOMIZED LISTS] [+OE]		3 *	SOVIET										779			
FREQUENCY (TABLE)	• FREQUENCY (TABLE) ILLEF 4 * JA22 7/4 • FREQUENCY (CHART) 5 * ELECTRONIC 721															
FREQUENCY (CHART)		5 *	ELECTRONIC										721			
		6 +	* TECHNIQUES 721 * TV 591													
SORROONDING WORDS		7 *	TV TV 591 8+ DNA 581 9+ GENE 7 6 486													
		8*														
0 - LEFT/RIGHT 0 -	HELP	9 +														
FREQUENCY IN CENTURY	HELP	THR	THREE MINUTE TOUR													
+1900s -1800s -1700s																
DISPLAY		The fo	llowing searches provide a quic	k overvi	ew of the	types of	queries	that you	can carr	y out. Als	o, feel fi	ree to loo	k at			
SORT BY CENTURY	HELP	the pa to exa	ge " <u>Extended Sample Searches</u> mine changes in spelling, morp	: from th hology (•	<u>e Corpus</u> word forr	", which ns), synt	provides ax, lexis	concrete (vocabul	example ary), and	es of how I semanti	the corp ics (word	ous can b meaning	e used))			
FREQ DISPLAY	HELP	during	the past 1000 years.													
RAW # 🔽		The Co	orpus of Historical English will a	llow you	to see:											
# HITS 100 SUBMIT RESET	HELP	 The Corpus of Historical English will allow you to see: 1. The frequency of words (wane, cleped, or fed^er/fether/feather), phrases (thou hast or turn on), or parts of words (e.g. heart* (sorted by frequency in the 1500s or the 1800s), *ship (by 1400s or 1700s), *friend* (by 1700s), or the substring ql??m* (by 1800s). (Note: charts vs. lists) 2. Which words or phrases have entered into or left the language from one century to another, (e.g. *ment 1400s/1500s to 1600s [entering / leaving] or *light* 1600s to 1900s [entering / leaving]) 3. Which words occur most often with others (often related to the meaning of the word), e.g.: words immediately after hard (1500s, 1900s) [sorted by uniqueness: 1500s, 1900s] words immediately after hard (1500s, 1900s) or fere (1600s 1 1900s) 														
		4.	words within five words of b Changes in which words or	reaks (<u>1</u> cur mos	<u>700s, 190</u> st often	<u>JOs</u>) or fa with oth	ice (<u>1600</u> ers (son	<u>is, 1900s</u> netimes -) a clue to	changes	in meani	na), e.a.				
ê							212 (2011					Internet				

Relevancy / Z-score like listing: * seat

🚰 Corpus of Historical English [D/	AVIES] - N	Microsof	t Internet Explorer						2 <mark>8</mark> 8	1		2 5 0		₽l×	
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> oo	ols <u>H</u> elp														
🗢 Back 🔹 🤿 🖉 🙆 🚳	🗟 Search	😹 Fav	orites 🍘 Media 🎯 🗟 🗸 🎒	J - 1	I 📖										
Address 🙆 http://view.byu.edu/che	e/x.asp												• ĉ	≻Go	
Google -	- G	Search	🝷 💽 🛛 🏧 1752 blocked 🛛 👋	Check 🔫	🔨 AutoLi	nk 👻									
Based on the Oxford En Dictionary (37 million wo Old English - Modern Eng	glish ords, glish)		CORPU Crea	SO ated	F H	IS Ma	ר ס R rk נ	IC/ Davi	ALE ies	і NG / В Ү	LIS U	H			
SEARCH		Click	on the number in a column to	see [KE)	WORD I		XT] displ	ау					[HELP	1 🔺	
WORD(S)/PHRASE(S) * seat	D(S)/PHRASE(S) at HELP CENTURY 1000s 1100s 1200s 1300s 1400s 1500s 1600s 1700s 1800s 19 at														
FREQUENCY (TABLE)	HELP	2	2 PILOT'S SEAT												
FREQUENCY (CHART)		3	RUMBLE SEAT										5.88% 4		
RELEVANCE SURROUNDING WORDS		4	MERCY SEAT							0.80% 3		0.76% 2	3.97% 5		
LIMITS		5	PASSENGER SEAT										2.20% 7		
SURROUNDING WORDS	HELP	6	REAR SEAT										2.17% 12	•	
FREQUENCY IN CENTURY 1900s>3	HELP		INTRODUCTION								More	information	1 000/ h	•	
DISPLAY SORT BY CENTURY 1900s FREQ DISPLAY RAW # HITS 100 SUBMIT RESET		TT W pt Pl 	INTRODUCTION More information This corpus of historical English is based on the material in the Oxford English Dictionary (OED). It contains 37 million words of texts from Old English to Present-Day English. You can use to the corpus to find the frequency and use of words phrases, and constructions in each century during the past 1000 years. Please feel free to take a look at either or the following pages, which provides sample searches of the corpus: Three minute tour, an overview of the main features of the corpus (with examples of actual queries) Extended Sample Searches, concrete examples of how the corpus can be used to examine changes in spelling, morphology (word forms), syntax (grammar), lexis (vocabulary), and semantics (word meaning) during the past 1000 years. [Copyright / licensing issues]												

Lexical bundles: * * (+1900s -1800s)

🕙 Corpus of Historical English [[DAVIES] - N	1icrosol	ft Inter	net Explorer											3 ×
<u> </u>	<u>T</u> ools <u>H</u> elj	p													.
🛛 🌍 Back 🔹 🕥 👻 🙎	1 🏠 🖌	🔘 Sear	ch 🤸	🎖 Favorites 🛛 🔗 - 🍃		- 📃 🏾	. 🧐	I 🥏 🌋	1 🕸	28					
Address 🕘 http://view.byu.edu/	che/x.asp														Go
Google -	•	G Searc	:h 👻 🍯	📡 🚿 🔄 3999 blocked '	🍄 Check	• 🗐 AL	utoFill 🕒	🕽 💽 Opti	ons 🌛						
Based on the Oxford E Dictionary (37 million v Old English - Modern E	inglish words, nglish)			CORPUS Creat	OF e d	HI by I	ST(Mar	ORI k Da	CA avie	LEN s/	IGL BYI	ISH			
SEARCH		24	AIM W/ 2											105	
WORD(S)/PHRASE(S)	HELP	25	AN	AEROPLANE										103	
* *	11221	27	TH	E OVERALL										100	
[CUSTOMIZED LISTS] [+OE]		28	тн	E INTERNET										99	
• FREQUENCY (TABLE)	HELP	29	OF	APPROXIMATELY										93	
		30	TH	E MOVIE										93	
		31	OF	TELEVISION										90	
		32	TH	E ELECTRONIC										90	
SURROUNDING WORDS		33	TE	CHNIQUES OF										90	
LIMITS		34	ST	AINLESS STEEL										89	
SURROUNDING WORDS	HELP	35	TH	E TV						-				87	
0 V LEFT/RIGHT 0 V		36	EI	AL						2				87	
FREQUENCY IN CENTURY	HELP		[KEYW	ORD IN CONTEXT]								More inf	formation		-
DISPLAY		1	1925	A. W. JUDGE: Automobile Engines	Althou <u>overa</u> (maxin	gh the ba <u>II</u> length	alance an and the car use.	id torque cylinder o	of the op limensior	posed two ns limit it t	o-cylinder :o power	r engine a units of a	ire excelle bout 12 to	ent, <u>the</u> 0 15 H.P	.] 🗎
SORT BY CENTURY	HELP	2	1936	Jrnl. R. Aeronaut. Soc.	A com propul	bination (sion.	of ship ar	nd hydrov	ane can	only redu	ce <u>the o</u>	verall re	sistance t	0	
FREQ DISPLAY	HELP	3	1951	G. LEWIS: Mod. Florist	The fo Interfle	ur large (ora, set u	groups Jp in 194	are close 6 to contr	ly conne ol and d	cted and o evelop the	compose intercha	<u>the ove</u> nge of ov	r all organ erseas or	isation ders.	
# HITS 100	HELP	4	1951	K. L. LITTLE: Mende of Sierra Leone	<u>The o</u> of com politica nowad	verall pi ponent v al entity v ays, the	icture i villages. T which, in section o	s one of s This comb the older of a chiefd	mall tow ination o sense, c lom.	ns around f town an orrespond	l each of d villages ls to what	which is s constitut t is officia	pread a r es a socia lly termed	number il and 1,	
SUBMIT RESET		5	1953	Colour Terms Sci. & Industry	Colour images	balance s in a col	, the rela our repro	ationship l oduction.	between j	the over	all intens	ities of th	e differen	t colour	
		6	1953	J. S. HUXLEY: Evolution in Action	With a becam	hundred e visible.	-thousan	id-fold spi	ed-up [d	of film] <u>t</u>	the over	all proce	sses of ev	olution	
		7	1954	Electronic Engin.	The sh	ort line v	vith which	h we are	concerne	d in this a	rticle is d	lefined as	one for v	vhich	-

🥝 Internet 👘

Semantics (collocates): hard * (+1900s -1500s)

🚰 Corpus of Historical English [[DAVIES]-N	Microsoft	Internet Explorer						- B] 🏙 🗾	22	0	35	×			
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> i	ools <u>H</u> elp												1				
🗢 Back 🔹 🤿 🗸 🙆	Q Search	😹 Favor	rites 🛞 Media 🎯 🖏 - 🤅	🗐 • 🖸 🛃	13												
Address 🙆 http://view.byu.edu/ch	ne/x.asp											•	∂Go Links	; »			
Based on 37 million w of text from Old Eng to Present-Day Engl	ords lish lish		CORP Cre	US OF	HI by I	ST(Marl	DRI k Da	CAL Ivie	. EN \$ /	IGL : вүц	ISH '						
SEARCH		Click	on the number in a column	to see [KEYW(DRD IN C	ONTEXT]	display						[HELP]	-			
WORD(S)/PHRASE(S) hard *	HELP	1	CENTURY HARD DISK	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s 29				
	HELP	2	2 HARD TIMES 2 1 11 18 3 HARD CORE 1 17 4 HARD WAY 16 16 5 HARD DRIVE 13														
C FREQUENCY (CHART)		4															
C RELEVANCE C SURROUNDING WORDS		6 7	HARD TIME HARD COPY									2	13 11				
LIMITS		7 HARD COPY 8 HARD STUFF 9 HARD KNOCKS															
SURROUNDING WORDS	HELP	9	8 HARD STUFF 2 2 11 9 HARD KNOCKS 3 10 10 HARD UP 1 19 10														
FREQUENCY IN CENTURY	HELP	[]	KEYWORD IN CONTEXT]								More inf	ormation		•			
DISPLAY		1 19	917 This Division had alrea 936 The fellows who have .	dy had a very . small double-	hard tim dip stand	ie From Is are h	the mon aving a <u>h</u>	nent of do ard time	e-busing <u>e</u> .	its life wa	s made v	ery diffic	ult for it.	-			
SORT BY CENTURY	HELP	3 19	956 Old-time sterno drinke 961 Fresh from the land of	rs made the sis-boom-bah	flophouse , the Ar	e forenooi mericans	n murky (had a ha	with their rd time	hard-tim at first le	e breath. arning to	applaud (good play	/ by				
FREQ DISPLAY	HELP	5 19	either team. 967 It has taken twenty year the speary reputation of	ars to get rid o	f the ama	ateurs, an	d the pro	fessional	s are h	aving a <u>h</u>	ard time	<u>e</u> breakin	g down				
# HITS	HELP	6 19	970 Ann-Margret is giving h Flab is reality.	nim a <u>hard tim</u>	1e on the	home fro	int, too, t	ossing ou	t little zin	igers abo	ut his adv	ancing a	ge like				
SUBMIT RESET		7 19 8 19	970 Smith told Petty Officer 974 Even people who've be lack pep.	r David Lewis, en out awhile	We are g with hepa	oing to ha ititis or m	ave a sit-i ononucleo	n and giv osis have	e the Jim a <u>hard t</u>	imy a <u>ha</u> time mak	<u>rd time</u> . king their	way bac	k. They				
		9 1	978 I was having a <u>hard ti</u> right.	<u>me</u> just breath	ning. But	they've g	ot me pa	dded up p	oretty goo	od right n	ow and I	think I'll	be all				
		10 19	990 Critics have always have him to art-historical a	d a <u>hard time</u> antecedents.	pegging	Zush s	ome rega	ard him a	s an outs	ider artist	t Others	s prefer t	o link				
Ø		11 1	991 Tve wondered whether	I could ever s	et this qu	ad before	e you in s	ome fairr	ness to it	and its o	aalings, ti	nings I've	e had a	-			
e Done											J 🛛 🛛 🖓 🖓	Internet					

Semantics (collocates): * meat (+1900s -1500s)

🚰 Corpus of Historical English [I	DAVIES] - N	Microsoft Internet Explorer					- E] 讔 💆	× 2	0	3	Ð×				
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u>	ools <u>H</u> elp											-				
🗢 Back 🔹 🤿 🖌 🙆 🚮	🔇 Search	💽 Favorites 🛞 Media 🎯 🖏 - 🎒 🧿	- 🗐 📖													
Address 🙆 http://view.byu.edu/ch	ne/x.asp									- 6	≷Go Li	nks »				
Based on 37 million w	ords	COPPIIS	OF HI	IST/	DT		E N		тен							
of text from Old Eng	lish Jich	Creat		Mari		via	- LN	RVI								
to Presencebay Eng	IISH	Creat	eaby	riai		vie	<u> </u>		<u> </u>							
SEARCH		Click on the number in a column to see	EKEYWORD IN	CONTEXT]	display						[HELF	ղ 🔺				
WORD(S)/PHRASE(S)	HELP	CENTURY	1000s 1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s					
* meat		1 MINCED MEAT						7	4	5	16					
[CUSTOMIZED LISTS] [+OE]		2 RED MEAT 2 11 3 FRESH MEAT 6 6 3 3 12 10 4 IN MEAT 6 6 1 4 5 9 5 COLD MEAT 6 6 1 4 19 8 6 LEAN MEAT 6 7 1 6 7														
	HELP															
SURROUNDING WORDS		COOKED MEAT								1	6					
LIMITS		7 COORED MEAT 7 <th7< th=""> 7 <th7< th=""> <th7< th=""> 7 <th7< <="" td=""></th7<></th7<></th7<></th7<>														
SURROUNDING WORDS	HELP															
	HELP								More inf	formation						
+1900s -1500s	<u></u>	The Corpus of Historical English will	allow you to see:						<u></u>			_				
			anon ,oa to 500.	·								-				
DISPLAT		1. The frequency of words (war	ne, <u>cleped</u> , or <u>fea</u>	d^er/fethe	<u>r/feather</u>), phrase	es (<u>thou l</u>	hast or tu	<u>irn on</u>), o	r parts d	of					
1900c	HELP	words (e.g. <u>heart*</u> (sorted b	y frequency in th	ie <u>1500s</u> o	r the <u>180</u>	<u>0s), *ship</u>	o (by <u>140</u>	<u>0s</u> or <u>170</u>	<u>10s</u>), <u>*frie</u>	<u>end*</u> (by						
119003		<u>1700s</u>), or the substring <u>al??n</u>	<u>n*</u> (by <u>1800s</u>). (f	Vote: <u>char</u>	<u>ts vs. list</u>	5) 				*						
FREQ DISPLAY	HELP	1400s/1500s to 1600s [enteri	na / leavina] or 3	*liaht* 160	onguai 10s to 190	je from d 10s fente	ring / lea	ry to ano vinol)	ther, (e.ç	j. ≃ment						
		3. Which words occur most o	ften with other	rs (often r	elated to	the mear	ning of the	e word),	e.g.:							
# HITS	HELP	words immediately after ha	rd (<u>1500s</u> , <u>1900</u>	<u>s</u>) [sorted	by uniqu	eness: <u>15</u>	00s, <u>190</u>	<u>0s]</u>								
100		words within five words of a	breaks (<u>1700s</u> , <u>1</u>	<u>.900s</u>) or <i>f</i>	ace (<u>1600</u>	<u>)s, 1900s</u>)		:							
SUBMIT RESET		 4. Unanges in which words of words after bard [+1500s/- 	-1900s1500s/4	n with oti -1900sl or	t ers (sor thefore <i>a</i>	netimes a beat [+15	1 Clue to (100 -190</td <td>manges i Os1500</td> <td>n meanin 1<!--+1900</td--><td>.g), e.g. ⊲l</td><td></td><td></td></td>	manges i Os1500	n meanin 1 +1900</td <td>.g), e.g. ⊲l</td> <td></td> <td></td>	.g), e.g. ⊲l						
		words near market [+1600	s/-1900s, -1600s	s/+1900s]	or match	[+1600s	/-1900s,	-1600s/+	1900s]							
		Note that in each case, the link runs	an actual query	against th	e corpus	and retur	ns the re	sults (i.e.	these ar	e not "ca	nned"					
		searches). Feel free to see how the :	search form has	been filled	l out, and	then mo	dify this t	o create	your own	queries.						
												-				

Semantics (wide-range collocates): market [5L/5R] (+1900s -1600s)

Corpus of Historical English [DAVIES] - M	1icrosoft Inte	rnet Explorer		🗳 🛃 👪 💇	× 2 0 0 0 _	. 8 ×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u>	[ools <u>H</u> elp						1
⇐ Back • ⇒ • 🖄 🐼	Q Search	😹 Favorites	@Media 33 및 - 43 이 - 13 11				1
Address 🕘 http://view.byu.edu/c	he/x.asp					▼ ∂ 60	Links »
Based on 37 million w of text from Old Eng to Present-Day Eng	vords glish glish		CORPUS OF HI Created by I	STORICA Mark Davi	LENGLI es/byu	(SH	
SEARCH		Click on a	word to see [KEYWORD IN CONTEXT] display			[HEL	LP] 🔺
WORD(S)/PHRASE(S)	HELD		WORD	1900s	1600s	PERC	
market		1	STOCK	85	0	100.00%	
[CUSTOMIZED LISTS] [+OE]		2	HAS	81	0	100.00%	
	HELD	3	COMMON	78	0	100.00%	
FREQUENCY (TABLE)	TILLE	4	NEW	64	0	100.00%	
C FREQUENCY (CHART)		5	BEEN	53	0	100.00%	
C RELEVANCE		6	BLACK	31	0	100.00%	
SURROUNDING WORDS		7	PRICES	31	0	100.00%	
LIMITS		8	MORE	28	0	100.00%	
	LUCI D	9	FREE	26	0	100.00%	
5 - LEST/DICHT 5 -	HELP	10	SHARE	26	0	100.00%	
		11	CAN	25	0	100.00%	
FREQUENCY IN CENTURY	HELP	12	WORLD	23	0	100.00%	
+1900s -1600s		13	EUROPEAN	22	0	100.00%	
DISPLAY				~	<u>^</u>	100,000	
SORT BY CENTURY	HELP	THRE	E MINUTE TOUR			More information	•
19005 -			<u>1700s</u>), or the substring <u>qi??m"</u> (by <u>1800s</u>). (N	ote: <u>cnarts vs. lists</u>)			-
	LICIO	2.	Which words or phrases have entered into or	• left the language from Settate 1000s to 1000s For	n one century to anot	her, (e.g. *ment	
RAW #	HELP	3	Which words occur most often with others	ight" 1600s to 1900s [<u>er</u> s (often related to the mi	<u>itering</u> / <u>leaving</u>)) eaping of the word) le		
			words immediately after <i>bard</i> (1500s, 1900s)) [sorted by uniqueness:	1500s. 1900s]	.g.,	
# HITS	HELP		words within five words of breaks (1700s, 19	00s) or face (<u>1600s</u> , <u>190</u>	<u>10s</u>)		
100		4. (Changes in which words occur most often	with others (sometime	s a clue to changes in	i meaning), e.g.	
SUBMIT RESET			words after <i>hard</i> [<u>+1500s/-1900s</u> , <u>-1500s/+1</u>	<u>1900s]</u> or before <i>meat</i> [<u>+</u>	-1500s/-1900s, -1500	<u>s/+1900s]</u>	
		· ·	words near <i>market</i> [<u>+1600s/-1900s</u> , <u>-1600s/</u>	<u>'+1900s]</u> or match [<u>+160</u>	<u>)0s/-1900s</u> , <u>-1600s/+</u> ;	<u>1900s]</u>	
		Note the	at in each case, the link runs an actual query a	gainst the corpus and re	turns the results (i.e.	these are not "canned"	
		searche	es), reel free to see now the search form has b	een niled out, and then r	nouny this to create y	our own queries.	
							-
e						🔰 🔡 Internet	

More information....

Corpus of Historical English [I	DAVIES] - M	icrosoft Interr	net Explorer		🏜 🖭 👪 💓	× 2 0 0 0 _	B×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> o	ools <u>H</u> elp						-
← Back • ⇒ • 🖾 🙆 🖓	C Search	Revorites	🖗 Media 🚳 🛃 - 🚑 🖸 - 🗎 🛍				
							inter »
	ne/x.asp						links
Based on 37 million w of text from Old Eng to Present-Day Engl	ords lish lish		CORPUS OF HIS Created by M	STORICA lark Davi	es / Byu	ISH	
SEARCH		Click on a v	vord to see [KEYWORD IN CONTEXT] display			[HEL	.P] 📥
WORD(S)/PHRASE(S)	HELP		WORD	1900s	16005	PERC	
market		1	STOCK	85	0	100.00%	
[CUSTOMIZED LISTS] [+OE]		2	HAS	81	0	100.00%	
	HELP	3	COMMON	78	0	100.00%	
C FREQUENCY (TABLE)		4	NEW	64	0	100.00%	
FREQUENCY (CHART)		5	BEEN	53	0	100.00%	
C RELEVANCE		6	BLACK	31	0	100.00%	
SURROUNDING WORDS		7	PRICES	31	0	100.00%	
LIMITS		8	MORE	28	0	100.00%	
	HELD	9	FREE	26	0	100.00%	
	<u>nele</u>	10	SHARE	26	0	100.00%	
		11	CAN	25	0	100.00%	
FREQUENCY IN CENTURY	HELP	12	WORLD	23	0	100.00%	
+1900s -1600s		13	EUROPEAN	22	0	100.00%	_
DISPLAY		L	00000	~	2	100.000	
SORT BY CENTURY	HELP	EXTEN	DED SAMPLE SEARCHES			More information	
19005 -		CDELLTN	IC:			Introduction	-
FREQ DISPLAY RAW #	HELP	1. You ca and see	(a) an easily compare alternate spellings in a wide that this spelling has increased greatly since th that its use decreased at the same time that -I	range of words. For ex e 1500s. You can then NG was on the rise.	ample, you search fo search for an earlier	Three minute tour Extended sample searches Background info / commen	its
# HITS	HELP	2. You ca	an easily look at the <u>decline of the spelling of [</u>	<u>u] as [v]</u> , by looking at	the overall frequency	Search by words and phras Search for surrounding wor	;es rds
100		[vn-].				Limiting by frequency	-
		3. You se	earch for the exact word form <u>down</u> and see th	at its frequency has inc	creased quite a bit sin	Customized word lists	-
SUBMIT		Suspecti <u>down/do</u>	ng that it may have been spelled differently be wne/doun/doune/dovn/dovne. This will show th	fore the 1600s, you ca he changes in spelling f	n use wildcards to sea or this one word over	r ich for alternate forms: the past 1000 years.	
		MORPHO	LOGY / WORD FORMS				
		1. You ca	an easily examine the overall frequency of pref	ixes (e.g. <u>pre-</u>) or suff	ixes (e.g. <u>-tion</u>) over	time.	
		2. You ca	an see which <u>words ending in -//</u> (mostly adver	bs) occur for the first t	ime in the 1900s.		-
I		 B. Early 	modern Endlish (e.d. 1500s-1600s) used the co	nmnarative (-er) and si	uneriative (-est) suffic	es more trequentiv	

Corpus do Português: 45m words, 1200s-1900s: www.corpusdoportugues.org

🚰 [Davies/Ferreira/NEH] O Corpus do Portuguê	s (45m	1 words, 1200s-1900s) - Mi	croso	ft Inte	rnet E	xplore	r										_ 8	×
Eile Edit View Favorites Tools Help																	_	7
🛛 😋 Back 🝷 💮 🖌 😰 🐔 🔎 Se	arch 🤇	📩 Favorites 🛛 🔗	-	0	-] 🦉	10	<u>(</u>	\$1 🔏	3							
Address 🕘 http://view.byu.edu/cdp/																-	🔁 Go	,
Google - C Sea	arch 🝷	😂 🛛 🚿 嶜 3999 blocked		🎸 Cheo	k + ∖	🗐 Auto	oFill 🧯	3 🔁 0	ptions	ß								
SEARCH STRING (HELP) WORD/PHRASE *temp*	0	CORPUS DO	РО	RTL	JGU	ÊS	(45	MIL	LION	W	ORD)s, :	120	0s-1	.900)s)		
(INSERT TAG) -select-	М	ark Davies (Brigh	nam	Υοι	ing	Univ) an	id Mic	hael	J. F	errei	ira (Geoi	rget	own	Uni	V)	
SEARCH CUSTOMIZED LISTS	SEE	CONTEXT: CLICK ON WO	RD (ALL SE	стіо	NS) OF	R NUM	BER (ON	E SECT	ION)					٢M	10RE I	HELP1	
DISPLAY (HELP) WORD 13 14 15 16 17 18 PtAc PtN PtFc PtOr BrAc BrN BrFc BrOr TOTAL															Γ			
© TABLE 1 TEMPO 542 252 6182 4150 2908 12706 1744 2615 3761 1239 2492 2048 3709 1381 48002																		
C CHART 1 1 1 1 1 1 1 2 2 2 5 1 2 2 1 3 1 2 1 3 1 3 1 2 2 1 3 1 2 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 </td <td></td>																		
2 TEMPOS 44 298 497 526 433 1620 344 590 630 126 347 299 567 103 6424 C SURROUNDING (HELP) 3 TEMPLO 9 98 295 259 508 612 162 31 87 6 114 31 72 26 2310																		
SURROUNDING O TEMPLO 9 9 9 255 508 612 162 31 87 6 114 31 72 26 2310 -select- • • • • 1 1 89 650 44 30 4 660 94 52 14 1639																		
SORT BY (HELP)	5	TEMPESTADE	9	28	60	106	65	392	51	45	76	6	30	18	66	14	966	
	6	TEMPORAL	6	65	183	88	59	125	47	34	42	10	122	18	36	3	838	
OPERCENT	7	TEMPLOS		14	164	73	89	144	95	12	27	1	119	5	26	7	776	
SECTION SOFT (HELP)	8	TEMPORADA					1	10	30	139	8	26	24	366	65	24	693	
	9	TEMPERAMENTO		7	1	25	37	289	28	8	78	7	23	18	81	16	618	
	10	TEMPERATURAS						7	237	25	9	3	227	20	4		532	-
1900s-ALL 1900s-ALL 1900s-BRAZ		KEYWORDS IN CONTEX	т										More	informa	ation			-
1900s-PORT 1900s-PORT	LIM	IT BY PART OF SPEECH: N	0			L	ІМІТ В	Y SECTI	ON: 17	00S [SEE AI	LL]					(HELP)	-
ORAL 🔽 ORAL 💌	1	17:Rosario:Frutas	de t	odas a	as alm	as, qu	e estã	o debai>	o do se	eu gov	erno,8	& juris	dição <u>t</u>	tempo	oral, 8	. espiri	itual: qu	L
MIN FREQ MIN FREQ	2	17:Coutinho:Economica	imp	erio de	e may	or valo	or que	o mund	o todo;	no co	rpo ter	n ajuri	isdiçaĉ	i <u>tem</u> j	<u>ooral</u> ,	& na a	alma a e	5
	3	17:NavioConselho	mei	o das	brenh	as por	oficio	e por in	terêsse	, fazer	ndo a s	sua feli	icidade	e <u>tem</u>	poral	e eteri	na, e da	3
OPTIONS (HELP)	4	17:Coutto:Brasil	sob	re tant	os qu	e os m	liseráv	eis tinha	am pad	ecido e	e pade	ciam, s	se leva	antou	um <u>ter</u>	mpora	<u>al</u> tao fo)
# HITS 100	5	17:Coutto:Brasil	Gra	ça divi	ina, de	evern a	as Mis:	soens do	Maran	hao o	seu es	piritua	l, e <u>te</u>	mpor	<u>al</u> aug	mento)* motiv	1
DISPLAY RAW FREQ 💽	6	17:Pita:America	a Ju	irisdiça	io de t	trez. N	o cive	l a Relaç	;ao da f	3ahia,	no <u>Te</u>	mpora	al ao (Gover	no do M	/larant	1áo, e n	1
GROUP BY WORDS -	7	17:Aires:Vaidade	a ju	irisdiça '	o de t	rës; n	o espi	ritual ao	bispad	o de P	ernam •	buco, I	no <u>ter</u>	npora	il ao g	overno	o do Ma	I
SAVE LISTS NO 🔽	8	17: Aires: Vaidade	aca	DOU CO	matu	igida,		m a mor	te, toi a	pena	temp	<u>oral</u> , e	e por o	consec	tercia	i pena	curta, p	F
SEARCH RESET	10	17:Barros:Vieira	per	petua, nortor	enqua		es se tieber	perpetua muisto	im den sebreur	ro da	sua m	esma (estera	, mas 200 1/	<u>temp</u>	<u>orai</u> , e	s exting a dectei	
	10	17.Darros.vieira		4	, en c	≊.	:.	J	:							L	n uesiel	-
																	•	
E Done														0	Interne	et		

Frequency tables and chart displays (*cujo* 'whose')

🚈 [Davies/Ferreira/NEH] O Corpus do Portuguê	(45m words, 1200s-1900s) - Microsoft Internet Explorer	💾 🛃 🕷 💌 🗶 🖉 🖸 🚫 💽 🖃 2
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp		18
🗢 Back 🔹 🔿 👻 😰 🚮 🔯 Search 🛛 🐼 F	avorites 🐠 Media 🧭 🛃 🗸 ᢖ 💽 🛪 📃 🎎	
Address 🕘 http://www.corpusdoportugues.org/		▼ ∂°60
Google - C Sear	h 👻 🔄 1752 blocked 🛛 🏘 Check 👻 AutoLink 👻	
SEARCH STRING HELP WORD/PHRASE cujo (INSERT TAG) -SELECT- v	O CORPU Click this button to always allow popups on www.corpusdoportugues.org Mark Davies To let an individual popup through, press the 'Ctrl' key while clicking on a link.	1200s-1900s)
SEARCH CUSTOMIZED LISTS	CLICK ON ANY BAR TO SEE THE EXAMPLES FROM THAT CENTURY, DIALECT, OR RE	EGISTER
DISPLAY (HELP)		
O TABLE	DECADE 1300 1400 1500 1600 1700 1800 1900 PtAc PtN PtFc PtOr BrAc BrN	BrFc BrOr PT BR AC NEW FIC OR
CHART SURROUNDING WORDS (HELP) -SELECT- SORT BY (HELP) FREQUENCY PERCENT	# HITS 37 227 610 522 372 1422 2103 437 350 272 20 474 243 # (MIL) 1.4 3.1 4.9 4.0 2.6 11.9 24.0 3.4 3.7 3.6 1.3 3.4 3.8	274 33 1079 1024 911 593 546 53 3.6 1.3 12.0 12.0 6.8 7.4 7.2 2.6
SECTION SORT . (HELP)	PER MIL 25.8 72.2 125 131 141 120 87.7 128 95.6 75.7 15.4 141 64.5	75.8 26.0 90.1 85.2 134 79.9 75.8 20.7
1 ALL 2 ALL 1900s-ALL 1900s-ALL 1900s-ALL 1900s-BR 1900s-BR 1900s-BR 1900s-PT ORAL ORAL MIN FREQ MIN FREQ 5	OVERVIEW: INTRODUCTION NOTE: The corpus and corpus interface are still being finalized. The corpus w that time, the corpus will be unavailable at times, and your searches may pro and improvements by July 2006)	Introduction
OPTIONS (HELP) # HITS 100 DISPLAY RAW FREQ GROUP BY WORDS SAVE LISTS NO	This website allows you to quickly and easily search more than 45 million w texts from the 1200s to the 1900s. The interface allows you to search for e lemmas, part of speech , or any combinations of these. You can als (collocates) within a ten-word window (e.g. all nouns somewhere near <i>cade</i> nouns near <i>girar</i>). The corpus also allows you to easily compare the frequency of and distributio constructions across texts, in at least three ways:	vords in more than 50,000 Portuguese xact words or phrases, wildcards, o search for surrounding words eva, all adjectives near <i>mulher</i> , or all on of words, phrases, and grammatical
a		Internet

Frequency tables and chart displays (*difícil.* * *de* [vr*] 'hard to VVI')

🚰 [Davies/Ferreira/NEH] O Corpus do Portugué	s (45m words, 1200s-1900s) - Microsoft Internet Explorer 🛛 👻 🛃	8 💌 🛛 🖉 🖸 🖸 🔽 🖃 🖄													
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp		100 A													
🖙 Back 🔹 🔿 🔹 🙆 👔 🖓 🚱 Search 🔬	iavorites 🛞 Media 🧭 🛃 🖌 🛃 💽 🗸 🧮 🛄														
Address 🕘 http://www.corpusdoportugues.org/		▼ @Go													
Google -	ch 👻 👰 1752 blocked 🛛 🏘 Check 👻 🦄 AutoLink 👻														
SEARCH STRING HELP	O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s	-1900s)													
WORD/PHRASE difficil.* de [vr*]															
(INSERT TAG) -SELECT	Mark Davies (Brigham Young O) / Michael J. Ferreira (Georgetown O)														
SEARCH CUSTOMIZED LISTS	CLICK ON ANY BAR TO SEE THE EXAMPLES FROM THAT CENTURY, DIALECT, OR REGISTER														
DISPLAY (HELP)															
© TABLE															
C CHART															
SURROUNDING WORDS (HELP)															
-SELECT- 💽 5 💽 5															
SORT BY (HELP)															
• FREQUENCY	# HITS 0 0 5 2 8 95 370 51 95 40 31 25 39 43 46	217 153 76 134 83 77													
	# (MIL) 1.4 3.1 4.9 4.0 2.6 11.9 24.0 3.4 3.7 3.6 1.3 3.4 3.8 3.6 1.3	12.0 12.0 6.8 7.4 7.2 2.6													
Section SORT (HELP)	PER MIL 0.0 0.0 1.0 0.5 3.0 8.0 15.4 14.9 26.0 11.1 23.9 7.4 10.4 11.9 36.3	18.1 12.7 11.2 18.0 11.5 30.0													
1 ALL															
1900s-ALL 1900s-ALL 1900s-BR	OVERVIEW: INTRODUCTION	Introduction													
1900s-PT 1900s-PT		_													
ORAL VORAL	NOTE: The corpus and corpus interface are still being finalized. The corpus will be onlin	ne in late June 2006. Before													
MIN FREQ MIN FREQ	and improvements by July 2006)	urate results. (List of fixes													
# HITS 100															
DISPLAY RAW FREQ V	This website allows you to quickly and easily search more than 45 million words in m	ore than 50,000 Portuguese													
GROUP BY WORDS -	texts from the 1200s to the 1900s. The interface allows you to search for exact word	ds or phrases, wildcards,													
SAVE LISTS NO 🗸	(collocates) within a ten-word window (e.g. all nouns somewhere near <i>cadeia</i> , all ad	jectives near <i>mulher</i> , or all													
SEARCH RESET	nouns near girar).	· · · · · ·													
	The corpus also allows you to easily compare the frequency of and distribution of word.	s obrases and grammatical													
	constructions across texts, in at least three ways.	s, philosos, and grammatical													
(a) iavascript:x(12)		i Internet													

Frequency tables and chart displays (*difícil.** *de* [vr*] 'hard to VVI') [1900s-Pt]

🚰 [Davies/Ferreira/NEH] O Corpus do Portuguê	is (451	n words, 1200s-1900s) - Microsoft	Inter	net Ex	plorer						1	🖭 🗱	W			0 8		۶Ľ
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp																		
🗢 Back 🔹 🔿 🗸 🙆 👘 🖓 🖓	avoriti	es 🛞 Media 🧭 🖏 - 🚑 💽 -		<u>11</u>														
Address Address Address Address			_														• @	'Go
Google - G Sear	ch 👻	💽 🛛 🔯 1752 blocked 🛛 🖑 Check	- 88	AutoL	ink 👻													
SEARCH STRING HELP			пĉ	е <i>(</i> л	5 M I		10	NI 317	0 P	ne	1.9	0.0 c .	. 1.0	0.0.6	\ \			
WORD/PHRASE difficil.* de [ur*]				5 (4	0 1411		.10	- 1N - 1V		08,	12	005	19	005	<u> </u>			
(INSERT TAG)	Ma	rk Davies (Brigham Young U)	/ Mie	chael	J. Ferr	eira	9 (G	eorge	towr	ηU)								
	CEE.	CONTENT: CLICK ON WORD (ALL	05.07	TONE	OD NU	MDE	-n //		CTI	NAL)					r	MODE		1 🔺
	SEE	WORD	12	14 1			ER (U	DNE SE		DN)	D+Or	Bréo	BPN	BrEo	BrOr	SUB	TOTA	-1 =
	1DIFÍCIL DE EXPLICAR5 $\underline{1}$ $\underline{5}$ $\underline{2}$ 2 3 1 8 19																	
	2 DIFÍCIL DE CONTROLAR 4 2 4 2 6 6																	
	2 DIFICIL DE CONTROLAR 1 3 2 1 6 8																	
SURROUNDING WORDS (HELP)	3 DIFICIL DE ENCONTRAR 1 3 2 1 1 6 8 4 DIFÍCIL DE FAZER 1 1 2 1 1 2 2 5 9																	
-SELECT- J 5 J	4 DIFÍCIL DE FAZER 1 1 2 1 1 2 2 5 9 5 DIFÍCIL DE ENTENDER 1 1 2 1 1 2 1 1 4 10																	
SORT BY (HELP)	5 DIFÍCIL DE ENTENDER 1 1 2 1 1 2 1 1 4 10 6 DIFÍCIL DE ESQUECER 1 1 3 1 4 5																	
FREQUENCY	7	DIFÍCIL DE DIZER				1	2		1		<u>3</u>					4	6	
C PERCENT	8	DIFÍCIL DE COMPREENDER				1	2		<u>3</u>		1				1	4	7	
SECTION SORT . (HELP)	9	DIFICÍLIMOS DE DESFAZER								4						4	4	
1 ALL 🛋 2 ALL 🛋	10	DIFÍCIL DE ATINGIR							2		<u>1</u>				1	3	4	
1900s-ALL	<u> </u>									-	_		Teday		_		-	<u> </u>
1900s-BR 1900s-BR		OVERVIEW: INTRODUCTION											Intro	oductio	n			_
19005-P1																		-
		NOTE: The corpus and corpus int	erfac 	e are la stri	still bei	ng f	finali	ized. T	he co	orpus	will be	e onlin in secu	e in la moto i	ate Jur	ne 200	16. Befr	ore	
MIN FREQ MIN FREQ		and improvements by July 20	(aliao 06)	ie at t	mes, a	na y	our	search	ies m	iay pi	oduce	inaccu	irate r	esuits	. (LISI	. от пя	es	
			Ċ.															
OPTIONS (HELP)																		
# HITS 100																		
CROUR BY		This website allows you to quick texts from the 1200s to the 1900	ly ar Os. T	id easi he inte	ly sear rface a	ch n dlow	nore Is vi	e than « nu to s	45 m eard	ullion h for	words	in mo word	re tha s or i	an 50, abras	000 Pi es. w	ortugue ildcar	ese ds.	
SAVELISTS NO -		lemmas, part of speech, or	any	comb	ination	s of	the	ese. Y	íou c	an a	lso se	arch	for	surro	undin	g wor	ds .	
SEARCH RESET		(collocates) within a ten-word wi	indov	/ (e.g.	all no	JNS	som	newher	e ne	ar ca	deia, i	all adje	ective	s near	r mult	<i>ier,</i> or	all	
		nouns near <i>girar</i>).																
		The corpus also allows you to ea	silv o	ompa	e the f	real	Jeno	v of ar	nd dis	stribu	tion of	words	. ohra	ses, a	nd ara	ammat	cal	
		constructions across texts in at le	ast t	hree M	avs			,						, .				-
bttp://www.corpusdoportugues.org/v3.asp2s-p8ww	4-difíc	il8:c4-8:w5-da8:c5-8:w6-antendar8:c6-												2	1 Inter	net		

Lemmatized: **fazer.*** (fazer = to make, do): +1300s-1600s, -1900s

🚰 [Davies/Ferreira/NEH] O Corpus do I	Português (45n	n words, 1200s-1900	is) - Microsoft In	ternet Explore				_	- 8 ×
<u> </u>	Help								-
🛛 😋 Back 🝷 🕥 🖌 😫 🐔	Search	🛧 Favorites 🛛 🚱	🔗 - 🖕 🛛	J • 🗾 🛃	S 🖉	🗶 🕸 🔏			
Address 🕘 http://view.byu.edu/cdp/									🗲 Go
Google -	🛛 🖸 Search 🕞	• 🍛 🚿 🙅 3999	blocked 🛛 🗳 CH	ieck 👻 🗐 Autol	Fill 🕒 🔁 Oj	otions 🖉			
SEARCH STRING	(HELP)			UGUÊS	(45 MTU		DS 1200s	-1900s)	
WORD/PHRASE fazer.*							.03, 12003	190037	
(INSERT TAG) -select-	<u>-</u> N	1ark Davies (B	B <mark>r</mark> igham Yo	oung Univ) and Mic	hael J. Fe <mark>r</mark> r	eira (Georg	etown Univ)	
SEARCH CUSTOMIZED LISTS									
DISPLAY	(HELP)	WORD(S)		SEC 1	SEC 2	P/MIL 1	P/MIL 2	PERC 1/2	_
TABLE		1 FEYTO		1251	0	93.04	0.00	1.00	
OCHART		2 FEZESSE		1053	0	78.31	0.00	1.00	
C CHINKI	()	3 FAZË		1040	0	77.34	0.00	1.00	
	(HELP)	4 FECTO		824	0	61.28	0.00	1.00	
-select- 💌 0 💌 0	•	5 FEZERA		804	0	59.79	0.00	1.00	
SORT BY	(HELP)	6 FIZERÃO		739	0	54.96	0.00	1.00	
FREQUENCY		7 FFAZER		727	0	54.07	0.00	1.00	
C PERCENT		8 FEZEROM		658	0	48.94	0.00	1.00	
SECTION SORT .	(HELP)	9 FEZO		553	0	41.13	0.00	1.00	
		10 FEZERŐ		538	0	40.01	0.00	1.00	
1800s		11 FAZIÃO		502	0	37.33	0.00	1.00	-
1700s 1900s-BR 1600s 1900s-PC	AZ	KEYWORDS IN CO	DNTEXT				More info	rmation	•
1500s	LIM	IT BY PART OF SPE	ECH: YES		LIMIT BY SEC	TION: YYY [SEE	: ALL]	(HF	ELP) 📥
1300s FICT	1	15:Dinis:Braganca	a que lha de=	ra por rëda sa	ibuda co fforo	que me ende de	uiã a <u>ffazer</u> ass?	como era cõtëudo	ë hu
MIN FREQ MIN FREQ	Q 2	15:Dinis:Braganca	a . E outrossi d	os dictos morac	lores do dicto	logar de Sesulffi	deuë a <u>ffazer</u> se	ruiço A nosso senh	or el
5 🔽 🛛 🗆	3	15:Dinis:Braganca	a dinheiros E	outrossi os dict	os moradores	do dicto logar de	e sesulffe deuë a <u>f</u>	i fazer seruiço a no	osso
OPTIONS	(HELP) 4	15:Dinis:Braganca	a - Iha dera poi	r Renda, sabud	a & a fforo qu	e mende au?ã de	: <u>ffazer</u> assi como) era cõteudo ë hû:	a car
# HITS 100	5	15:Dinis:Braganca	a alffandega /	A alde?a de Sar	nbadí por sent	ëça cõ ssas perti	ėėças pera Aiuda (de <u>ffazer</u> o meu ca	asteli
DISPLAY RAW FREQ	• 6	15:Dinis:Braganca	a eu que uola	desse por uos:	sa Alde?a cõ s	sas pertëëças pe	ra aiuda de <u>ffaze</u>	<u>:r</u> o meu Castello d	lessa
GROUP BY WORDS 🚽	7	15:Dinis:Braganca	a dicta Aldeya	a de sanbadj cõ	sas pertëëça	s por uossa Alde	a pera Aiuda de <u>f</u>	fazer esse meu C	astel
SAVE LISTS NO 🔽	8	15:Dinis:Braganca	a & que Retijr	iha pera ssi as	Rendas que o	s de ssanbadi au	?ä de <u>ffazer</u> & os	padröados das eiç	greia:
SEARCH RESET	9	15:Dinis:Braganca	a & ao Sabad	o ou?r os preito	osna Vila E ë - 0 l	na vila se <u>ttazer</u>	todau?a iustica. E	:sta postura deue∖ Ia Caa ka	/aler
	10	15:Dinis:Braganca	a peraeumj	oprigar por ele	s & polos que	an de uijr A <u>ffaz</u> '- F else dewi ' :	<u>er</u> a dicta çerca d Ganan annia	e Cen braças de m	nuro 💌
e								🥝 Internet	

Tagged for part of speech: mulheres [aj*] (ADJ women) (sorted by 1800s)

🚰 [Davies/Ferreira/NEH] O Corpus do Portugué	≜ s (45m	1 words, 1200s-1900s) - Mic	rosoft Int	ernet I	Explore	er										_ 8	×
Eile Edit Yiew Favorites Tools Help																4	7
🔀 Back 🔹 💮 🖌 🔀 🛃 🔎 Se	earch S	🛃 Favorites 🥝 🔗	🎍 🖸] • [, .	3	通る	2 🛍	*	**							
Address () http://view.byu.edu/cdp/			- UR-				~ ~								-] 🔁 G	0
Google -	arch 👻	🕥 🛷 🖓 3999 blocked	ABC Che	eck 👻	🗐 Aut	oFill	B	Option:	s 🖉								
SEARCH STRING (HELP)					IÊS	(4	5 M 1		N NC			20)e - 1				
WORD/PHRASE mulheres [aj*]	ĬĬĬ					1	5 111				, , ,	201			,5,		
(INSERT TAG) -select-	M	ark Davies (Brigh	am Yo	ung	Univ) a	nd №	licha	el J.	Ferre	ira ((Geoi	get	own	Uni	v) –	
SEARCH CUSTOMIZED LISTS																	
DISPLAY (HELP)	SEE	CONTEXT: CLICK ON WOR	RD (ALL S	ECTIO	NS) O	r Nu	MBER (ONE SE	CTION)				[М	10RE I	HELP	1
TABLE	WORD 13 14 15 16 17 18 PtAc PtN PtFc PtOr BrAc BrN BrFc BrOr SUB TOTA 1 MULHERES BONITAS 15 6 5 4 1 15 31														TOTAL		
O CHART	1	MULHERES BONITAS					15			6		5	4	1	15	31	
C CONNECT	2	MULHERES PERDIDAS					<u>12</u>			1			2		12	15	
O SURROUNDING (HELP)	3	MULHERES HONESTAS				1 1	- 2			1			1		9	13	
-select- 🔽 0 🔽 0 🔽	4	MULHERES BELAS					Z			5					7	12	
SORT BY (HELP)	5	MULHERES CASADAS				5	<u>6</u>	1		4 1	4	3	3		6	27	
FREQUENCY	6	MULHERES LIVRES					4			1	1		1		4	7	
O PERCENT	7	MULHERES FEIAS					4			1					4	5	
SECTION SORT (HELP)	8	MULHERES AJOELHADAS					4								4	4	
	9	MULHERES DESSE					<u>3</u>			1					3	4	
PORT-FICT	10	MULHERES DESTE				1	3								3	4	-
PORT-ACAD 1900s-ALL 1900s-BRAZ		KEYWORDS IN CONTEXT	Г									More	inform	ation			-
1800s 1900s-PORT	LIM	IT BY PART OF SPEECH: N	0		L	IMIT	BY SE	CTION:	1800S	[SEE A	LL]					(HELP)	
1600s 🔽 ORAL 工	1	18:Almeida:Gatos5	- Vivo	lá. Há i	muitos	anos	s - diss	e-me el	le Te	rra de <u>n</u>	nulher	es bo	<u>nitas</u>	, segur	ndo m	e const	ē
MIN FREQ MIN FREQ	2	18:Dinis:Fidalgos	fantást	ico, fei	rido no	seu	orgulh	o cereb	ral em	eminên	cia de c	retino	- ade	us <u>mu</u>	Ihere	s boni	t
5 🗌 0 🗖	3	18:Dinis:Pupilas	- Mauri	ício ten	n essa	habi	lidade	de ser v	visto to	dos os c	lias por	as <u>m</u>	ulher	es bo	nitas	da terr	а
OPTIONS (HELP)	4	18:Caminha:Bom-crioulo	reconh	eceu q	ue era	obs	ervada	, se é q	ue cert	o instint	o, pecu	liar da	as <u>mu</u>	Iheres	s boni	itas , lh	с
# HITS 100	5	18:Alencar:Lucíola	decerto	o, quer	ia um	bem	louco	ão pequ	ieno, p	referia-c) a toda	s as <u>r</u>	nulhe	res be	onitas	s do mu	п
DISPLAY RAW FREQ	6	18:Lopes:Viúva	que de	vora c	om um	ha for	me can	ina, cor	no qua	ndo con	templay	/a um	a muli	tidão d	e <u>mul</u>	lheres	1
GROUP BY WORDS	7	18:Alencar:Correr	de son	s, aque	ela onc	la de	perfur	nes, de	toillete	es, e de	mulhe	res b	onita	s que s	se alas	strava p	n
SAVE LISTS NO -	8	18:Alencar:Correr	faremo	os, se f	or nec	essái	rio, cor	no dizia	Afons	o Karr a	propós	ito da	s <u>mul</u>	heres	boni	<u>tas;</u> fai 	'n
SEARCH RESET	9	18:Machado:Datas	Verdes	, na qu	Jal são	com	nendad	ores do	númer	o os nar	morado	s que	despr	ezam .	as <u>mu</u>	lheres	_
	10	18:Macedo:Mulheres	de ser	admira	ada; m	ias di	esculpo	u-a diz	endo q	ue era u	im dese	jo nat	ural à	s <u>mul</u>	heres	; bonit	² ▼
														•		▶	

Collocates: mulheres // [aj:fs] (ADJ women) (sorted by 1900s)

🚰 [Davies/Ferreira/NEH] O Corpus do Portuguê	is (45	m words, 1200s-1900s)	- Mi	croso	oft I	nteri	net I	xplore	ar 👘				8			7 🛛 🖉	2 3	0 💽	_ 8	×
Eile Edit View Favorites Tools Help														-						
← Back → → → 🙆 🖗 🖓 🥘 Search 📓	Favorit	es 🛞 Media 🌀 🖏	- 8	0	-		<u>11</u>													
Address Address Address						_													- <i>@</i>	50
Google - G Sear	rch 👻	👰 🛛 👰 1752 blocked	ABC	Che	ck •	- 38	Aut	oLink	•										_	
SEARCH STRING HELP	0	CORPUS DO P	O R	е т П	GI	IÊ S	s (45	итгі	ION	JWO) R D	S . 1	200	s - 1	9005	0			
WORD/PHRASE mulher																	· /			
(INSERT TAG) -SELECT-	Ma	ark Davies (Brigham '	You	ng l	J) 7	Mic	chae	el J. Fi	erreira	(Ge	orget	own U	り							
SEARCH CUSTOMIZED LISTS	RESI			NCY	(HF	ELD)											r	MORE	HELP	1 🔺
	ELP) SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION)															Έ				
OTABLE WORD 13 14 15 16 17 18 PTAC PTN PTFC PTO BrAC BrN BrFc BrOr SUB TOT 9														%	-					
C CHART	1	GRÁVIDA						3	7	1		1		3	9	2	27	30	10.54	
	2	ADÚLTERA						4			- 4						4	8	7.69	
SURROUNDING WORDS (HELP)	3	NUA						2			25		3		13	2	43	45	5.85	
adj.FSG 🔽 0 🔽 4 💌	4	IDOSA						6		2	3				1		6	12	5.71	
SORT BY (HELP)	5	GORDA					1	13		1					17		25	39	4.16	
C FREQUENCY	6	DEITADA						1			6			1			11	12	3.61	
• PERCENT	7	MAGRA						1		2	2				6	2	12	13	2.89	
SECTION DISPLAY (HELP)	8	PRENHE			4	2		1			4				1		5	15	2.56	
	9	CASADA						40	2	2		1	2	4	11	1	28	68	2.55	
	10	BONITA					2	72			10		3	4	31	3	51	125	2.54	
1900s-ALL 1900s-BR 1900s-BR		KEYWORDS IN CONTE	EX⊤												I	ntroducti	on			-
1900s-PT 1900s-PT	LIM	IT BY PART OF SPEECH	: NO								LII	MIT BY	SECT	ION:	NO					-
ORAL 🔽 ORAL 💌	1	18:Machado:Jucanda			Q	uano	do e	ntrou r	ia sala i	de es	pera,	viu um	na <u>mul</u>	l her d	e pé,	<u>magra</u> ,	, amare	lada, e	nvolvid	ē
MIN FREQ MIN FREQ	2	19N:Pt:Expr			du	Jran	te o	ano ai	nterior.	# Es	ta jove	em <u>mu</u>	<u>ilher r</u>	nagra	a, deli	cada e o	de voz t	ranquil	a é con	is
4 🔽 🛛 🗆	3	19N:Pt:Público			рa	ara a	a pro	tegere	em ver	caixa	Rosa	é uma	nulh <u>mulh</u>	er m	agra	e frágil,	a quem	o cab	elo clar	0
OPTIONS (HELP)	4	19:Fic:Pt:Tavares:Insul	omis	sos	. N	lo lu	gar	frontei	ro ão d	ele ia	aquel	a <u>mul</u>	<u>her m</u>	agra	, nova	, mas d	e idade	quase	indefini	í٨
# HITS 100	5	19:Fic:Pt:Amorim:Masc	ara		. V	'eio	abri	r a por	ta a Rit	a, <u>m</u>	<u>ulher</u>	do Lor	nguinha	os, <u>ma</u>	<u>aqra</u> ,	vestida	de farra	apos si	ujos.	
DISPLAY RAW FREQ 🗸	6	19:Fic:Br:Carvalho:Beb	ado	5	, п	o ca	aso,	da mo	rta, um	a <u>mu</u>	<u>lher</u> r	nuito t	pranca	e <u>ma</u>	<u>qra</u> , o	com os a	cabelos	escorri	idos sol	c
GROUP BY WORDS -	7	19:Fic:Br:Holanda:Burr	0		a	alm	a do	distrit	o. Mafra	a, a <u>r</u>	nulhe	r <u>mag</u>	ira , lev	vantou	J-se, p	ostou-s	e atrás	do poe	ta e co	n
SAVE LISTS NO 🔽	8	19:Fic:Br:Louzeiro:Dev	otos		ol	hand	do o	táxi q	ue se di	istan:	ciava.	Uma <u>n</u>	nulher	rjove 	m e <u>m</u>	<u>aqra</u> , p	bés des:	calços,	atrave	s
SEARCH RESET	9	19:Fic:Br:Louzeiro:Pixo	te		Er	a tu	ido c	jue sat	ola. Lem	nbra-	se vag	ament	te da <u>n</u>	nulhe	<u>r, ma</u>	qra ep	alıda, a	VOZ SL	imida, l	J
	10	19;FIC:Br:Kego:Pedra	-		a	vida	па	casa p	aroquia Io do fo	n. Era Hiller	a uma	mulhe	er <u>ma</u>	<u>qra</u> e	aita, l	Je vozis Janta	seca e a	utorită	riā,	
		19;FIC:Dr:Neto;Turblina	0		, е	sma	iyav	a u D0	io ue re	ijau;	a <u>mu</u>	<u>ner, r</u>	nayra	, uiste	s, con	na ienta	mente,			ŕ
								_		-		_						a		- E

Collocates: mulheres // [aj:fs] (ADJ women) (1900s vs. 1800s)

🚰 [Davies/Ferreira/NEH] O Corpus do Portugi	uês (45m words,	1200s-1900s) - M	licrosoft	Internet Exp	lorer				200	S5	×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp											
⇐ Back	Favorites 🛞 M	edia 🎯 🖏 👍	3 🖸 •	· 🗐 📖							
Address 🕘 http://www.corpusdoportugues.org/										• ?	Go
Google - C Se	arch 👻 💽	1752 blocked 🛛 🐴	🌾 Check	👻 🔆 AutoLir	ik 🕶						
SEARCH STRING HELP	O CORI	PUS DO PO	RTUG	UÊS (4	5 MILL	ION	WORDS, 1200)s-1900	s)		
WORD/PHRASE mulher	Mark Davi	oc (Brigham Vo	una LIX	/ Michael 1	Forroira	(Coord	atown U)				
(INSERT TAG) -SELECT-		es (brighain ro	ung oj		, renena	Georg	jetown O)				
SEARCH CUSTOMIZED LISTS	RESULTS OR	DERED BY FREQU	ENCY (I	HELP)					[M(ORE HELP] 🔺
DISPLAY (HELP)	SEE CONTEX	T: CLICK ON WOR	RD								
C TABLE	WOR	D #۱	REG 1	# REG 2	% REG 1		WORD	# REG 2	# REG 1	% REG 2	
C CHART	1 INFA	ME	6	0	1.00	1	BRASILEIRA	14	0	1.00	
SURROUNDING WORDS (HELP)	2 INS	ENSÍVEL	6	0	1.00	2	SOZINHA	14	0	1.00	
adj.FSG 🔹 0 💌 4 💌	3 ADC	RADA	5	0	1.00	3	DIFERENTE	8	0	1.00	
SORT BY (HELP)	4 DES	MAIADA	5	0	1.00	4	DESEJADA	8	0	1.00	
C FREQUENCY	5 INFE	RNAL	5	0	1.00	5	VERDE	6	0	1.00	
• PERCENT	6 TAN	ТА	5	0	1.00	6	CLARA	5	0	1.00	
SECTION DISPLAY - (HELP)	7 SEG	UINTE	4	0	1.00	7	IMPORTANTE	5	0	1.00	
	8 PER	VERSA	4	0	1.00	8	EXCEPCIONAL	5	0	1.00	-
1900s-ALL 1800s	KEYWO	RDS IN CONTEXT	Г					Introduct	tion		Ŧ
1700s 1900s-PT		RT OF SPEECH: N	0				LIMIT BY SECTION:	NO			
1500s 🔽 ORAL 🔽	1 19N:Pt:B	eira	há se	eis anos. Ago	ra, ela era	uma <u>m</u>	<u>ulher sozinha</u> , firme	e, de olhar ve	rde e cabe	lo apanhad	。
MIN FREQ MIN FREQ	2 19N:Pt:J	ornal	como	péquevais	er agora? A	۹ <u>mulhe</u>	e r não fica <u>sozinha</u> . A	Apesar de ser	mos humil	des, estamo	25
4 🗸	3 19:Fic:Pt	:Melo:Autópsia	patró	ões depois de	comer, e	mandarı	n a <u>mulher</u> trabalhar	<u>sozinha</u> em	cima deles	s. Anica não	1
OPTIONS (HELP)	4 19:Fic:Pt	:Melo:Autópsia	de lit	pertação. Ass	im, que po	odia faze	r <u>mulher</u> <u>sozinha</u> , c	om marido a	usente, filh	io pequeno	P
# HITS 100	5 19:Fic:Pt	:Melo:Autopsia	, eue,	, pois claro, t é raro que o	: mesmo <u>n</u> narceiro al	<u>hulher</u> <u>s</u>	sozinha tem e o med a mulher sozinha	o a espera na	a solidao ar	ntiga a decisão in	
CROUR BY	7 19:Fic:Bi	:Amaral:Amigos	pare	e raro que o s cuio perfil (era muito s	emelhar	a <u>mumer</u> <u>soznina</u> , nte: uma mulher soz	yue tennyue zinha e um h	omossexu/	al que não t	0
SAVE LISTS NO V	8 19:Fic:Bi	:Dantas:Cartilha	cond	enada ão infi	erno? É do	ditado:	mulher sozinha é a	ndorinha. Ah,	, Virgem de	a Guadalupe	9. 1
SEARCH RESET	9 19:Fic:B	:Dantas:Cartilha	porq	ue não tem e	stofo próp	rio, fibra	i de <u>mulher</u> que sabe	e viver <u>sozin</u>	<u>ha</u> ? Embor	ra, muitas v	e
	10 19:Fic:B	Queirós:Dora	o gu	arda me segi	irou o puls	o: - <u>Mul</u>	her <u>sozinha</u> na rua,	tarde da noit	e, a ordem:	n	
	11 19:Fic:B	:Teixeira:Rua	pros	peridade cad	a vez maio	ir. Mas d	eixei minha <u>mulher</u> ;	sozinha no h	iotel. Com	licença. Nis	ti 💌
	•									•	1

Collocates: cadeia // [nn*] (ADJ string) (FICT vs. ACAD)

ē

	عتد الكله
<u>File Edit View Favorites Tools H</u> elp	1
4 Back $\bullet \Rightarrow - 2$ 2 2 2 2 Carch 🙀 Favorites (Media 3 2 $\mathbb{R}_{2} \bullet \mathbb{R}_{3} \bullet \mathbb{R}_{3}$	
Address 🕘 http://www.corpusdoportugues.org/	∂ Go
Google - 🔽 💽 Search - 😥 🔤 1752 blocked 🛛 🍄 Check - 👯 AutoLink -	
SEARCH STRING HELP O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)	
WORD/PHRASE cadeia Mark Davies (Brigham Young LL) (Michael 1, Eerreira (Coorgetewn LL)	
(INSERT TAG) -SELECT-	
SEARCH CUSTOMIZED LISTS RESULTS ORDERED BY FREQUENCY (HELP) [MORE HEL	·] 🔺
DISPLAY (HELP) SEE CONTEXT: CLICK ON WORD	
C TABLE WORD # REG 1 # REG 2 % REG 1 WORD # REG 2 # REG 1 % REG	2
C CHART 1 PAI 7 0 1.00 1 SUPRIMENTOS 54 0 1.00	
SURROUNDING WORDS (HELP) 2 PORTA 7 0 1.00 2 REACÇÃO 21 0 1.00	
NOUN - 5 - 5 - 3 DIAS 6 0 1.00 3 REAÇÃO 17 0 1.00	
SORT BY (HELP) 4 GENTE 6 0 1.00 4 REACÇÕES 13 0 1.00	
• FREQUENCY 5 PRESOS 6 0 1.00 5 CONCEITO 10 0 1.00	
© PERCENT 6 ANOS 5 0 1.00 6 GESTÃO 10 0 1.00	
SECTION SORT (HELP) 7 FILHO 5 0 1.00 7 MONTANHAS 10 1 0.91	
1 1900s-8R A 8 HOMENS 5 0 1.00 8 CARACTERES 9 0 1.00	-
1900s-BR Introduction	-
CRAC FICT FICT	-
FICT ACAD register. For example, find the frequency and distribution of cujo, noun + caro, or forms of ter+ que + infinitive.	
MIN FREQ MIN FREQ You can also search for collocates (words nearby a given word), which often provides insight into the meaning of a	
0 0 given word. For example, you can search for the most common nouns near <i>cavalo</i> , verbs near <i>papel</i> , adjectives	
OPTIONS (HELP) the collocates of the same word in different sections (e.g. nouns near <i>cadeia</i> in FICTION and ACADEMIC, or in the	
# HITS 100 1900s vs the 1800s).	
GROUP BY WORDS V	
SAVE LISTS NO V	
SEARCH RESET VARIATION (HISTORICAL / REGISTER / DIALECT)	
You can easily limit the searches and sort the results by register (genre) . This allows you to see how words and	
phrases vary across speech and many different types of written texts. For example, find -mento nouns in academic,	-

Word comparisons: [nn*] {agudo/aguçado/afilado} 'sharp N'

🚰 [Davies/Ferreira/NEH] O Corpus do Portugu	ês (45m words	, 1200s-1900s)	- Micr	osoft Inl	ternet	: Explorer				B] 🏙 🖬 🔟 🖉		0 🖸	_ 8 ×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp														1
두 Back 🔹 🔿 🗸 🙆 🖓 🖓 Search 😹	Favorites 🛞 🕅	1edia 🥶 民	- 🎒	0 - E	I 11									
Address Address Address						·								- ∂Go
Google - G Sea	arch 👻 😫	a 1752 blocked	ABC C	iheck 👻	👯 Al	utoLink 👻								
SEARCH STRING HELP	O COR	PUS DO P	ORI	rugu	ÊS	(45 MILLIO	N W C	D R I	DS,	120	0s-1900s)		
WORD/PHRASE do/agudo/aguçado}	Marily David			- 112 7 1										
(INSERT TAG) -SELECT-	Mark Dav	ies (Brignam	Youn	g 0) / r	MICHA	iei J. Ferreira (G	eorgeti	own	0)					
SEARCH CUSTOMIZED LISTS	RESULTS OR	DERED BY FRE	QUEN	CY (HEL	.P)							[]	MORE	HELP] 🔺
DISPLAY (HELP)	AFIL	DO +	-	%		AGUDO	+	-	%		AGUCADO	+	-	%
C TABLE	1 NARIZ	INHO 1		1.00	1	ANGULO	8		1.00	1	ESPÍRITO	1		1.00
C CHART	2 84842	21TO 1		1 00	2	GRITO	6		1 00	2		1		1.00
SURROUNDING WORDS (HELP)	3 DEL AT	0 1		1.00		STLVO	5		1.00	2	RUMOR	1		1.00
-SELECT- V 0 V 0 V				1.00	3	ENEADTE	3		1.00	3	CENCO	1		1.00
SORT BY (HELP)	4 KOST			1.00	4	ENFARTE	4		1.00	4	SENSO	1		1.00
	5 NARIZ	. 12	+ 2	0.88	5	OLHAR	4		1.00	5	DENTE	1	1	0.50
• PERCENT	6 FOCIN	HO 3	3	0.50	6	SOM	3		1.00	6	INSTINTO	1	1	0.50
SECTION SORT - (HELP)					7	ACENTO	2		1.00	7	OLHO	1	1	0.50
					8	ÁPICE	2		1.00	8	FOCINHO	1	5	0.17
					9	ASSOBIO	2		1.00					-
1900s-ALL 1900s-BR 1900s-BR	KEYWO	RDS IN CONT	EXT								Introduction	on		•
1900s-PT 1900s-PT	LIMIT BY PA	RT OF SPEECH	I: NO				LII	MIT	BY SEC	TION	: NO			
	1 19Ac:Pt:E	inc	, d	esigna-s	e tam	bém deste modo	a função	de	um âng	julo <u>a</u>	gudo num triân	gulo rea	tângu	ılo, definida
MIN FREQ MIN FREQ	2 19Ac:Pt:E	inc	se	mpre m	enos	do que 180º ou ra	ds (ângu	ulo <u>a</u>	<u>qudo</u> (ou obt	:uso) e um arco	maior t	erá se	mpre
	3 19:Fic:Pt:	Regio:Principe	en	itre os lo	oiros c	abelos anelados, f	formand	o un	n ângul	o <u>aqu</u>	ido cujo vértice	mais o	u men	os cairia na
OPTIONS (HELP)	4 19:Fic:Pt:	Ribeiro:Planici	e pa	ipel e o i	relógi	o de madeira rem	atado en	n ân	gulo <u>ac</u>	<u>udo</u> .	No meio da cas	a, ladri	lhada	a lambaz
# HITS 100	5 19Ac:Br:I	.ac:Thes	qu	iina latei	ral de	sua margem, forr	nando u	m âr	ngulo <u>a</u>	<u>qudo</u>	(FIG. 2.5), às v	ezes		
DISPLAY RAW FREQ 💽	6 19Ac:Br:I	.ac:Thes	m	argem, (em vis	sta dorsal, forman	do um â	ngul	o <u>aqud</u>	<u>lo</u> . La	bro cerca de me	eia vez	mais I	ongo que
GROUP BY WORDS -	7 19Ac:Br:I	.ac:Thes	m	argem, (em vis	sta dorsal, forman	do um â	ngul	o <u>aqud</u>	<u>lo</u> . La	bro cerca de un	na vez e	e meia	mais
SAVE LISTS NO 🔽	8 19:Fic:Br	Holanda:Burro),q	ue tinha	o per	fil dentro de um â	ngulo <u>ac</u>	<u>juda</u>	ı, volto	u a fr	eqüentar o escr	itório qu	Jando	saía às
SEARCH RESET														

Synonyms: [=gritar]: synonyms of 'to shout'

🚰 [Davies/Ferreira/NEH] O Corpus do Portugué	ès (45	m words, 1200s-1900s) -	Mici	osofi	t Inter	net Ex	plore	r				-	1 👪	W				×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp																		1
수 Back 🔹 🔿 🗸 🙆 🖓 🔞 Search 家 I	Favorit	es @Media 🚳 🛃+	4	3	- 🗐	63												
Address bttp://www.corpusdoportugues.org/			-														.	Go.
		🚳 🗌 🐼 1750 klaskad	ABC .	-Ll.	- 464	A sub-st	i i - I.	_									(·	
Google +	i i		<u> </u>	CHECK	- T []	Auto		•										
SEARCH STRING HELP	0	CORPUS DO P	O R	тис	G U Ê S	S (4	15 M	ИІСС	ION	woi	RDS,	120) O s - :	1900)s)			
WORD/PHRASE [=gritar]	M	ark Davies (Bricham)	(oun	a U)	Z Mir	haol	1 E	orroir:	(Ceor	notow	n LD							
(INSERT TAG) -SELECT-		ark bavies (brighain i	oun	ig v)				enene		Jerow								
SEARCH CUSTOMIZED LISTS	SEE	CONTEXT: CLICK ON W	ORD	(ALL	. SECT	TIONS) OR	NUMBE	R (ONE	SECT	(ON)					[MORI	E HELP	.] 🔺
DISPLAY (HELP)		WORD	13	14	15	16	17	18	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	TOTAL	
• TABLE	1	GRITAR [S]			13	38	18	222	2	23	209	23		20	163	9	740	
CCHART	2	RECLAMAR [S]		1	1	3	3	78	10	33	20	5	26	39	48	17	284	
	3	PROTESTAR [S]			2	3	6	84	8	34	30	6	5	37	18	2	235	
O SURROUNDING WORDS (HELP)	4	BRADAR [S]		18	48	27	3	73		2	22			1	4	2	200	
-SELECT 5 - 5 -	5	QUEIXAR-SE [S]		1	2	7	13	77	1	7	31	1		3	10		153	
SORT BY (HELP)	6	APREGOAR [S]	36	20	34	19	7	17		4	7		1	3	3		151	
FREQUENCY	7	BERRAR [S]			9	16		46		1	34	10			26	1	143	
C PERCENT	8	LADRAR [S]	2	2	9	16	13	25	1		43			1	7		119	
SECTION SORT 💽 (HELP)	9	RALHAR [S]		1		1	1	81			8	3			12		107	
1 ALL 🔼 2 ALL 👗	10	CHIAR [S]			4	18	8	27		2	16	1		1	7		84	
	<u></u>						_				_			-	_			
1900s-BR 1900s-BR		KEYWORDS IN CONTE	хт										1	Introdu	ction			
1900s-PT 1900s-PT	31	18:Azevedo:Cortiço		de	saias.	Jerôr	nimo	perdeu	a paciê	ncia e	ia pro t	testar	brutaln	nente	contra s	emelha	nte inva	si 🔺
ORAL 🔽 ORAL 💌	32	18:Azevedo:Cortiço		cor	n vont	tade o	le afa	istar-se	e, mas s	em an	imo de	prote	e <mark>star</mark> , p	or aca	nhame	nto, ten	tou reat	ar
MIN FREQ MIN FREQ	33	18:Azevedo:Mulato		vo	ntade	até de	e que	ixar se	ão pai;	de <u>pr</u>	otesta	<u>r</u> contr	a aquel	as con	traried	ades qu	e a fazia	ī
5 🗖 5	34	18:Macedo:Luneta		Há	só um	na voz	que	pode e	e há de j	orotes	<mark>star</mark> , é	a minh	na, a vo	z susp	eita,			
OPTIONS (HELP)	35	18:Pompéia:Ateneu		Est	a cara	angue	jola, (enorme	e e pesa	da, qu	e pareo	cia pro	testar	a cad	a solav	anco, co	ontra o d	:a
# HITS 100	36	18:Taunay:Inocência		alg	umas	pauta	das d	le incit	amento,	parec	eu que	rer o c	argueir	o <u>prot</u>	estar (contra o	tratame	en:
DISPLAY RAW FREQ 🗸	37	18:Taunay:Inocência		me	smo.	Oh! S	r. Ga	rcia! qu	uis prot e	<u>estar</u>	Pereira	Nac	da; dig	o-lhe	isto do	coração		
GROUP BY WORDS	38	18:Alencar:Correr		no	ator q	lue ca	íra er	n falta	, nem de	prot	estar (contra	o ato do	os dire	tores p	or uma	semelha	nt
SAVE LISTS NO -	39	18:Cunha:Artigos		de	Loure	nço d	e Méc	licis. Pi	recisamo	os, por	ém, <u>pr</u>	otesta	<u>ar</u> tamb	ém co	ntra a i	ntroduç	ão deste	F
SEARCH RESET	40	18:Alencar:Garatuja		na	quela .	capita	nia, c	jue vin	ham tod	os uni	dos em	corpo	protes	tar co	ntra a	violênci	a inaudit	a
	41	18:Alencar:Garatuja		er	evolta	do ne	leos	entime	ento do l	pelo, ia	prote	star, (quando	parec	eu-lhe d	que de r	novo ent	re
ARCH RESET For Michael Control a Violencia Inductia 41 18:Alencar:Garatuja e revoltado nele o sentimento do belo, ia protestar, quando pareceu-lhe que de novo entre 42 18:Machado:Textos dessa confissão, xê-se que o noeta queria principalmente protestar contra o caminho que l																		

Customized lists: querer.* [davies@gritar-syn]: want to "shout"

🚰 [Davies/Ferreira/NEH] O Corpus do Portugué	≜ <mark>s (</mark> 45	m words, 1200s-1900s) - Mici	rosoft II	ntern	net Ex	plore	r					-] 🏙	7 🛛	20		3 _	. 8 ×
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp																		-
수 Back 🔹 -> 👻 🙆 🚮 🔞 Search 📷	Favori	tes 🛞 Media 🎯 🛃 🎝	3 -		13,													
Address 🗃 http://www.corpusdoportugues.org/																	•	∂Go
Google - C Sea	rch 🔫	👰 🛛 🕾 1752 blocked 🛛 🗳	Check ,	*	AutoL	ink 🖣	•											
SEARCH STRING HELP	0	CORPUS DO POR	тисі	JÊS	; (4	5 N	41L	LIC	N W	ORI	os,	120	0 s - 1	900	s)			
WORD/PHRASE querer.* [davies@gi				8 A) - I				- 17										
(INSERT TAG) -SELECT-	IM	ark Davies (Brigham Your	ig () /	MIC	naei	Ј. Не	erreir	a (G	eorge	town	0)							
SEARCH CUSTOMIZED LISTS	SEE	CONTEXT: CLICK ON WORD) (ALL S	ECTI	IONS'		NUMP	ER (ONE SE	CTIO	N)					[MORE		P. 1 🔺
DISPLAY (HELP)		WORD	1	3 14	4 15	16	17	18	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	Тот	AL
(TABLE	1	OUIS GRITAR						8			1				5		14	4
O CHART	2	OUIS PROTESTAR		_				3		1	2				1		7	
	3	OUER GRITAR		_				4									4	
SURROUNDING WORDS (HELP)	4	OUERIA PROTESTAR		_				1			1						2	
-SELECT- 🔽 5 🔽 5	5	4 QUERIA PROTESTAR 1 1 1													2			
SORT BY (HELP)	6	OUEREMOS RECLAMAR								1						1	2	
• FREQUENCY	7	OUERO GRITAR			_						1				1		2	
C PERCENT	8	OUERO CLAMAR		_										1			1	
SECTION SORT (HELP)	9	OUIZ GRITAR									1						1	
	10	OUISESSE GRITAR													1		1	
	<u></u>																	
1900s-ALL 1900s-ALL 1900s-BR 1900s-BR		KEYWORDS IN CONTEXT											I	ntroduc	tion			-
1900s-PT 1900s-PT	31	18:Azevedo:Cortiço	de sa	ias	Jerôn	imo p	berde	uap	aciênci	a e ia	prot	estar l	orutalm	ente c	ontra s	emelha	nte in	ivas 📥
ORAL SCAL	32	18:Azevedo:Cortiço	com	vonta	ade d	e afa	star-s	se, m	ias sem	n anim	no de	protes	star, po	or acar	nhamer	nto, teni	tou re	eatar
MIN FREQ MIN FREQ	33	18:Azevedo:Mulato	vonta	nde a	até de	quei	xar s	e ão	pai; de	prot	estar	contra	aquela	is cont	rarieda	des que	e a fa	ziam
5 🗖 5	34	18:Macedo:Luneta	Há số	i uma	a voz	que	pode	e há	de <u>pro</u>	otesta	<mark>ar</mark> ,éa	a minha	a, a voz	suspe	ita,			
OPTIONS (HELP)	35	18:Pompéia:Ateneu	Esta	carar	nguej	ola, e	enorm	nee	pesada,	, que	pareci	a <u>prot</u>	<u>estar,</u>	a cada	solava	anco, co	ontra	o ca
# HITS 100	36	18:Taunay:Inocência	algur	nas p	pauta	das d	le inci	itame	ento, pa	areceu	ı quer	er o ca	rgueiro	prote	e star c	ontra o	trata	men [.]
DISPLAY RAW FREQ 💌	37	18:Taunay:Inocência	mesr	no.C	Dh! Sr	r. Gar	rcia! d	quis j	protest	t <mark>ar</mark> Pe	ereira.	Nada	a; digo	o-lhe is	to do o	oração		
GROUP BY WORDS	38	18:Alencar:Correr	no at	or qu	Je caí	íra en	n falt	a, ne	m de <u>p</u>	rotes	star o	ontra o	ato do	s diret	ores po	r uma s	semel	lhant
SAVE LISTS NO 🔽	39	18:Cunha:Artigos	de Lo	uren	nço de	e Méd	icis. P	Precis	samos,	porér	n, <u>pro</u>	otesta	<u>r</u> també	im con	tra a ir	ntroduçã	šo de:	ste p
SEARCH RESET	40	18:Alencar:Garatuja	naqu	ela c	apitar	nia, q	ue vi	nhan	n todos	unido	s em	corpo j	protest	tar cor	ntra a v	violência	a inau	Idita
	41	18:Alencar:Garatuja	e rev	oltad	do nel	eos	entim	nento	do bel	o, ia 🛾	orotes	star, q	uando p	barece	u-lhe q	ue de n	ovo e	entre
	42	18:Machado:Textos	dessa	a con	nfissão	o, vê-	-se qu	Je o	poeta q	jueria	princi	palmer	nte pro t	testar I	contra	i o cami	inho a	tne ▲
	•																	

Customized lists: estou + emotions (I am + emotions): 1900s

🚰 [Davies/Ferreira/NEH] O Corpus do Português	; (45m words, 1200s-1900s) - Microsoft	Internet Explorer									
Eile Edit View Favorites Tools Help											
🛛 🌍 Back 🔹 🕥 🖌 🗾 👔 🏠 🔎 Sea	arch 🧙 Favorites 🕢 🔗 - 嫨	o • 🔜 🗄 💐 🤣 🗱 🕉									
Address 🕘 http://view.byu.edu/cdp/			💌 🏓 Go								
Google -	rch 🕞 🕤 🕺 S999 blocked 🦓	Check 👻 🖓 AutoFill 🕒 🍢 Options 🌛									
SEARCH STRING (HELP)	O CORPUS DO POR	T UGUÊS (4 5 MILLION WO	RDS, 1200s-1900s)								
(INSERT TAG) -select-	Mark Davies (Brigham \	Young Univ) and Michael J. Fer	reira (Georgetown Univ)								
SEARCH CUSTOMIZED LISTS	SEE CONTEXT: CLICK ON WORD (AL	SECTIONS) OR NUMBER (ONE SECTION)									
DISPLAY (HELP)	WORD 13	14 15 16 17 18 PtAc PtN PtFc Pt0	Or BrAc BrN BrFc BrOr SUB TOTAL								
• TABLE	1 ESTOU CONTENTE	2 24 <u>7</u> <u>4</u> 2	<u>2 9 1</u> 25 51								
O CHART	2 ESTOU CANSADO	1 1 1 29 4 3 2	<u>4 4</u> <u>4</u> 21 53								
C SURROUNDING (HELP)	3 ESTOU SATISFEITO	2 1 17 4 2 3	<u>1</u> <u>3</u> <u>3</u> 16 36								
-select- 💌 0 💌 0 💌	4 ESTOU ARREPENDIDO	<u>3</u> <u>2</u> 16 38									
SORT BY (HELP)	5 ESTOU FELIZ	1 <u>3 1</u>	<u>4 1 6</u> 15 16								
• FREQUENCY	6 ESTOU CONFIANTE	3 2	<u>3</u> <u>2</u> 1010								
O PERCENT	7 ESTOU ENGANADO	1 4 2 1	<u>4</u> <u>3</u> 1015								
SECTION SORT - (HELP)	8 ESTOU CONVICTO	5 2	<u>1</u> <u>1</u> 99								
	9 ESTOU SEGURO	2 3 2 4 <u>1</u> 2	<u>3 2 1</u> 9 20								
	10 ESTOU ANSIOSO	9 2 1 2	1 2 8 17 🔽								
1900s-ALL 1900s-BRAZ 1900s-BRAZ	CUSTOMIZED LISTS: MODIFY	LIST	More information								
1900s-PORT	To modify a list:	LISTS CREATED [DAVIES]	MODIFY LIST								
MIN FREQ S OPTIONS (HELP)	AL IN THOMANY A list. N FREQ Add or delete words from your list and click on [Submit] Add or delete words from your list MULHER_AJ M D CORES M D MULHERES M D ROUPA M D MULHERES M D ROUPA M D										
# HITS 100 DISPLAY RAW FREQ GROUP BY WORDS SAVE LISTS NO											
			feliz								
iavascript:x("[davies@mulheres]")			Submit Reset								

Customized lists: User-created lists (can correct our "errors"): sair 'to leave'

🚰 [Davies/Ferreira/NEH] O Corpus do Portugué	ês (45m w	ords, 120)0s-1900s) - Microsoft	Intern	et Explorer		* *	🖭 🗱 👿 🛛	2005 _	BX
<u>File E</u> dit <u>V</u> iew F <u>a</u> vorites <u>T</u> ools <u>H</u> elp										1
🗢 Back 🔹 🤿 🗸 🔯 🖓 Search 👔	Favorites	Media	🎯 🖪 - 🎒 💽 -	· E i	Ŭ,					
Address 💩 http://www.corpusdoportugues.org/										∂~Go
Google -	rch 👻 💽	👰 17	52 blocked 🛛 🍄 Check	- *	AutoLink 👻					
SEARCH STRING HELP	ос	ORPU	S DO PORTUG	UÊS	(45 MI	LLION W	ORDS, 12	00s-1900s	;)	
WORD/PHRASE sair.*	– Mark	Davies	(Brigham Voung LI)	7 Mieł	hael 1 Ferre	eira (George	town II)			
(INSERT TAG) -SELECT-	Mark		(brightam round o)	7 101101		sina (deorge				
SEARCH CUSTOMIZED LISTS	SAVE	.IST	NAME davies	LI	IST sair			Submit Reset	HELP	
DISPLAY (HELP)										
• TABLE	SEE CO	NTEXT: C	CLICK ON WORD						[MORE HELP]
			WORD(S)		SEC 1	SEC 2	P/MIL 1	P/MIL 2	PERC 1/2	
SURROUNDING WORDS (HELP)	1		SAYU		145	0	15.32	0.00	153.16	
-SELECT 5 - 5 -	2		SAIROM		92	0	9.72	0.00	97.18	
CIEREQUENCY	3		SAYROM		94	2	9.93	0.11	91.88	
PREQUENCI PERCENT	4		SAYRAM		46	1	4.86	0.05	89.93	
SECTION SORT . (HELP)	5		SAIDAS		110	3	11.62	0.16	71.68	_
1 1 2 1	6		SAHYD				3.40	0.05	64 51	_
1800s 1800s 1800s 1700s 1700s	cu	STOMIZ		IST	55	L	5,49	Introducti	04.31	
1600s 1600s 1500s 1500s		5101121	ED LISTS, MODIFY L	.191						
1400s 1400s 1300s I 1300s I	To mo	dify a lis	t:	LIST	S CREATED	DAVIES		MODIFY LIST		Ē
MIN FREQ MIN FREQ	Add or	delete w	ords from your list	SAI	R M	FAZ_PHRA	SES M	r>		
5 🗆 5	and cli	ck on [Su	ubmit]	GRI	TAR-SYN M	DEVER	M	ORDLIST NAME		
OPTIONS (HELP)				FAZ:	1 M	FAZER	M	IST OF WORDS		-
# HITS 100	Retur	<mark>n to ma</mark> i	in menu	ЕМО		CORES	M	sayu	-	
GROUP BY WORDS -	Overv	<u>view of v</u>	<u>vordlists</u>	MUU		BOURA	M s	sairom savrom		
SAVE LISTS YES 🗸				MOL	D		D	ayram		
SEARCH RESET				MULI	HERES D		5	saidas sahvr	-	
							1	Submit Recet		
							-	Submit Reset		-
é)									🥝 Internet	

Advantages of relational databases

 Size, speed, and annotation
 Unlimited, modular tables (synonyms, customized lists, etc)
 Helps in annotating the corpus

POS tagging

- Bootstrap from existing tagged corpus
- Probabilistic: citie, then emptie (update tens of millions of tokens at a time)
- Rule-based: to emptie, the emptie CONJ
- Even words that are not in the existing corpus: DET XX PREP/CONJ

Lemmatization

English: not much of a problem (work, flye, returnes, but saw, ground, wound, etc)

Much worse for other languages (Sp trabajo 'work/trabajo, Iwork/trabajar)

Sp hubiese = 56/72 possible forms

Lemmatization (historical)

- > Apply modern lemmas (finish)
- > Apply spelling rules: *tie: emptie / empty
- Use frequency information from database to find overly frequent lemmas (goo / go, had lane / had lain)
- Use frequency information to target highly frequent unknown forms first (students / see context / provide modern form / words acquire modern features)
- > a pease of: compare to PDE: a piece / peas / peace of
- Can also use contextual words (pease)

Interface for annotating Old Portuguese (blocks of text)

Lie Fait Aiew Lavolites Tools Helb	
年 Back 🔹 🔿 🗸 🔞 🖓 🖓 Search 📾 Favorites 🛞 Media 🧭 🖏 🖬 🖕 🚍 🖸 🗸 🚍 🚺	
Address 🕼 http://view.byu.edu/neh/byID.asp	- @60
Google - C Search - 😥 🔊 1752 blocked 🦓 Check - 👯 AutoLink -	
BY TEXT / PASSAGE BY WORD HAS LEMMA NO LEMMA	
15_braga_Text-job 💌 WORD LEM POS FREQ 50 -10000 500 POS	1
OFFSET 200 OPT > 4 x MPT	`
SUBMIT RESET GO	
	-
38036659 , , pu 3/13/11 >>	
38036660 o o atd:ms 1282528 >> CONTEXT WORD V// WORD	-
38036661 qual qual pn 41410 >>	-
38036662 se se pnp 418287 >> WURD offereceo	
38036663 offereceo oferecer VJ:3s 122 >> LEMMA oferecer	
38036664 ao a prp+ 60968 >>	
38036665 capitao capitao nn:ms 5370 >> - POS	
38036666 de de prp 1651256 >>	
Submit Reset	
38036669 pos	
38036670 trazer VP: 3079 >>	
38036671 a 0 atd:fs 1562557 >>	
MORE	
1 37465452 oferecer VJ:3s vulgares, todos os empregos, que vos offereceo o tempo, mais dignos de amor.	
2 37467220 oferecer VJ:3s aventejado papel, com o qual subitamëte se offereceo a el Rey, e Valido Castelhano ;	
3 37470344 oferecer VJ:3s Com zelo digno de hum varão piadoso, <u>offereceo</u> Diogo Soares, o primeiro motivo aos olhos	
4 38036663 oferecer VJ:3s hum mercador gentio China, o qual se offereceo ao capitão de Malaca de nos trazer a	
5 38157478 oferecer VJ:3s presente hûa vez, de cousa que se offereceo a caso, querer elle dali por diãte	
6 38184588 oferecer VJ:3s hû cauallo que logo veo, & lho offereceo dizëdo Tomai senhor este que creo que estes	
7 38200379 oferecer VJ:3s porem contentouse o pobre homë com que lhe offereceo hû homem da terra que fosse trabalhar a	
8 39996071 oferecer VJ:3s de ouelhas ; & que de seus gados offereceo a Deos os primogenitos : & Caim filho	
9 40407090 oferecer VJ:3s justa remuneração devida a quem com tanta gloria offereceo a vida pela Fè, & honra da	

Cone 🕑

Interface for annotating Old Portuguese (by word, POS, lemma)

🎒 Ne	w Page 1 - Mi	crosoft II	nternet E	xplorer						8	8 🚾	× 2 0 0 s .	. 8 ×
Eile	<u>E</u> dit <u>V</u> iew	Favorites	<u>T</u> ools	<u>H</u> elp									-
4⊐ В	ack 👻 🔿 👻 (🗵 🕼 🖞		Search 🛛 🗽 Fay	vorites 🛞 Media 🏼 🏹 🖏 🕯	- 👍 💽	- [i i i i i i i i i i i i i i i i i i i					-
Addr	ess 🙆 http://v	view.bvu.e	du/neh/bv	ID.asp								▼	<i>i</i> ∂ Go
Go			•	C Search	🗸 🚱 🛛 🔊 1752 blocked	ABC Cher	-k -	👯 AutoLink 🕞					
	0		_			V							
B	TEXT / PAS	SAGE			BY WORD					HAS LEMMA		NO LEMMA	
1	5_braga_Text	-job		-	WORD LE	м Сонт	AR	POS		FREQ 50 -100	00	500 pos	
OF	FSET 200				CORPUS OLD 🖵 # 20	0				OPT > 4 × MF	т		
	SUBMIT RE	SET			SUBMIT RESET					GO		GO	
													-
10	317	conte			VS:1s/VS:3s	>>		[_
11	316	contav	am		VI:3p	>>		CONTEXT		WORD 💌 /	'/ 📃	WORD 💌	
12	282	contei			VJ:1s	>>		WORD		cõta			-
13	235	cõta			VP:3s	>>							_
14	216	contar	am		VJ:3p/VM:3p	>>				contar			
15	195	cõtar			VR:	>>		POS		VP:3s			
16	194	contas	se 		VT:1s/VT:3s	>>							
17	174	contou	-lhe		VJ:3s+	>>		Submit Res	set				
18	173	contar	a		VM:1s/VM:3s	>>							
19	162	contou	-me		VJ:3s+	>>							
20	158	contar	ei 2		VF:15	>>							
21	141	contar	a 		VF:35	>>	Ţ						
22	137	contan	105		VP:10	>>		J					
M	DRE												-
1	37062574	contar	VP:3s	entendam	& declarem & Julguë na Re	partiçam	n & <u>ci</u>	<u>õta</u> dellas assi como	o nos titullos	seguintes do pam			
2	37063631	contar	VP:3s	se paguara	a huû Real. E pella dicta <u>c</u>	<u>õta</u> & Re	parti	çã se paguarão as c	cousas deste	e foral			
3	37497309	contar	VP:3s	inquietàrão) dessa bãda. Dãdolhes a s	i mesmo	<u>cőt</u>	<u>a</u> dos q se të reduzi	do, ou redu:	zirë			
4	37497455	contar	VP:3s	o tëpo q es	tiver alojada, seja por <u>cõt</u>	<u>ta</u> dos cu	Ipad	os, se fará conta de	tudo				
5	37538951	contar	VP:3s	que nunca	o virão, do que se <u>cõta</u> pe	ellos exp	erimë	étados. Tres ondas,	que parece				
6	37548281	contar	VP:3s	a outra nag	ção do mundo, o darlhe <u>cõ</u>	i <u>ta</u> desta	gene	erosa paixão, a quer	m sómente	nòs			
7	37613772	contar	VP:3s	vitoria. Por	èm porq inteiramëte não f	icasse á	<u>cõta</u>	de interiores movin	mentos, que	o desejo muytas			
8	37631296	contar	VP:3s	pode comp	rehender alguma escondid	la senter	ıça. P	'or esta <u>cõta</u> se poe	em aquella f	amosa interpretaçã	o[57]		
9	37710441	contar	VP:3s	sëpre uiuë	e ë esta parte tã pouca <u>cô</u>	<u>íta</u> deue	s de t	fazer delle, como de	e ser				
, (e) D	one			N		<i>w</i> . ,	^					🙆 Internet	

Interface for annotating Old Portuguese (by lemma frequency I)

🍯 New	Page 1 - Mi	crosoft Inte	rnet Explorer							🖻 🖭 🏙 I	w 🛛 🖉 🖸 🖸 🖸	_ 8 ×
<u>F</u> ile <u>E</u>	dit <u>V</u> iew	F <u>a</u> vorites	<u>T</u> ools <u>H</u> elp									1
🗘 Back	$\tau \Rightarrow - ($	2 🕅 🖓	🛛 🔞 Search 🛛 👔 Fav	/orites 🌒	Media 🧭 🖏 🎒 🕟	-		談				
Address	🕘 http://v	iew.byu.edu/	neh/byID.asp				_				•	ୖ୶୕ଡ଼
Goog	e - surrecte	d "brigham vo	oung" 🔽 🔀 Search	- 😨 🛛	🔯 1752 blocked 🍄 Che	ck ·	- 8	🕻 AutoLink 👻 👸 animals	👸 re	surrected 👩 brigham you		
							_	<u> </u>	<u> </u>		-	_
ВҮ Т	EXT / PAS	SAGE		BY WOR	D					HAS LEMMA	NO LEMMA	
12_0	coimbra_te:	xt-msc12	-	WORD		AR		POS		FREQ 50 - 10000	500 pos	
OFFS	ET 200			CORPUS	OLD 💽 # 200					OPT > 2 × MPT	0.01	
SU	BMIT RE	SET		SUBMIT	TRESET					GO	GO	
34	2.87	4366	VIDE		CHANGE ALL			CHANGE ALL LEMMAS:				
35	2.84	4090	CASTELA		CHANGE ALL							
36	2.77	4064	NAU		CHANGE ALL			DANO	то		Submit	
37	2.45	4041	SENHORIA		CHANGE ALL							
38	2.42	4034	FEITA		CHANGE ALL							
39	2.48	3944	DANO		CHANGE ALL							
40	2.11	3793	BISPO		CHANGE ALL							
41	2.02	3718	NAVIO		CHANGE ALL							
42	2.47	3709	FORTALEZA		CHANGE ALL							
43	2.15	3676	ORDENAR		CHANGE ALL							
44	2.20	3533	N		CHANGE ALL							
45	2.96	3422	ARMADA		CHANGE ALL	-						

🥭 Nei	v Page 1 - Mic	rosoft I	nternet Exp	olorer							🖴 🖭 👸	8 W	x / C 🖸 🔽 🗉	. 8 ×
Eile	<u>E</u> dit <u>V</u> iew	F <u>a</u> vorites	s <u>T</u> ools <u>I</u>	<u>H</u> elp										-
<⇒ Ba	ck → → ~ @	2) 🕅 i	상 (@) Sea	arch 🗔 Favo	orites @Media	61 B A 1	₹ 🗸		1 組					
Addre:	s 🙆 http://vi	ew.bvu.e	:duínehíhvID	.asp	~		_	_					•	∂ Go
Goo		d "brigber		C Search	🗸 🚳 🖾 175	2 blocked ABC Ch	eck ·	- 3	🗧 AutoLipk 👻 🐻 apima	als 👸	resurrected 🔊 brigham	vouna		x
	- Isquecce	a brighan		G boardin								young		
BY	TEXT / PAS	SAGE			BY WORD						HAS LEMMA		NO LEMMA	7 -
12	_coimbra_tex	t-msc12	2	-	WORD	LEM CON	TAR		POS		FREQ 50 -10000	5	500 POS	
OFF	SET 200				CORPUS OLD	- # 200					OPT > 2 × MPT			
s	UBMIT RE	SET			SUBMIT	RESET					GO		GO	
														-
	_						_ _							_
KW.	.C		4252			CHANG	E	Н.			WORD 🗾 //		WORD 💌	
	2345		dano danos		nn:ms	>>		Н.	WORD		dãno			
3	251		dänos		nn:mp	>>		Н.			dano			-
4	228	0	dapno		nn:ms	>>		Н.						_
5	224	0	damno		nn:ms	>>		Н.	POS		nn:ms			
6	217	0	dampno		nn:ms	>>		1						
7	106	0	danno		nn:ms	>>			Submit Reset					
8	100	c.	dapnos		nn:mp	>>								
9	87	0	damnos		nn:mp	>>								
10	62	c c	dãpno		nn:ms	>>								
11	45	0	dannos		nn:mp	>>	-							
														
	RE													
1	38153155	dano	nn:ms	a elle : & a	assi vendolhe ei	ueste <u>dano</u> me (conte	nto	, & quero que me arri	Inquem				_ _
2	38157207	dano	nn:ms		ie dar esmolia, i	& o <u>dano</u> que so anda da dãna av	cceae	e au fai	os ingratos, irata de i	num				_
	38214492	dano		que o post	Deos que pagas	enua uu <u>uanu</u> yu së o grëde mel &	dãn	те. о о	z eni sua casa, uanuu we fizeră à Painha cu	, a vrmä	8,			_
5	39344895	dano		ate o nass	arem para se e	se o grade mark suitar o dâno gu	e com	uy vis	so se fazi Háio	a yrrina.	, α			_
6	39348315	dano		sem nagar	rem coimas : co	mtal que não fac	inos e	ănr	n nas nouidades : nor	oue faz	endoo nagarão os dão	05		-
7	39348350	dano	nn:ms	trazer os o	ditos gados. E qu	uando fizerem d á	ino n	as	nouidades, farão S 2	penhor	a Quarta			
8	39348369	dano	nn:ms	em tanto o	gado que baste	para pagar o <u>dãi</u>	<u>10</u> &	coi	mas. E não prenderão	os pas	tores			
9	39361752	dano	nn:ms	dinheiro :	de maneira que	não soomente n	ecebe	m	dâno no preço em qu	ue as co	mprão fiadas,			
		i.	i						N . I .				Takarat	_
e													internet	

Composition of proposed 200 million word historical corpus

	SP	OKEN/INFORMAL	FI	CTION		NEWS	ACA	ADEMIC	OED	TOTAL
	SIZE	SOURCES	SIZE	SOURCES	SIZE	SOURCES	SIZE	SOURCES		
1500s	5	EEBO-Dr CED	10	EEBO-Fic	1	??	13	EEBO-NF	1	30
1600s	5	EEBO-Dr BAILEY CED	10	EEBO-Fic	1	NEWSBOOKS (NEWDIGATE)	13	EEBO-NF	1	30
1700s	5	LION-Dr BAILEY CED LATE 18C	10	LION-Fic	1	LAMPETER ZEN	13	LION-NF	1	30
1800s	9	LION-Dr PG-Dr BAILEY	16	LION-Fic PG-Fic	7	MOA-Jrnl SCAN-News (LON TIMES) (NY TIMES)	16	LION-NF PG-NF	2	50
British	4		7		3		8			23
American	4		7		3		8			23
Misc	1		2		1		0			4
1900s	9	MOVIES RADIO ORALHIST BNC, LLC, SEC, (LDC) PG-Dr SCAN-Dr	21	PG-Fic SCAN-Fic BNC-Fic	14	SCAN-Per (LON TIMES) (NY TIMES) BNC-News NDNP	14	PG-NF SCAN-NF BNC-NF	2	60
British	4		9		6		7			27
American	4		9		6		7			27
Misc	1		3		2		ο			6
TOTAL	33		67		24		69		7	200
PERCENT	16.5		33.5		12.0		34.5		3.5	100%

Conclusion

 \succ Relational databases: size, speed, and annotation Very wide range of searches Information from database can be used as an integral part of the tagging process Already working in several online corpora, more in planning stages