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

Size

- 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

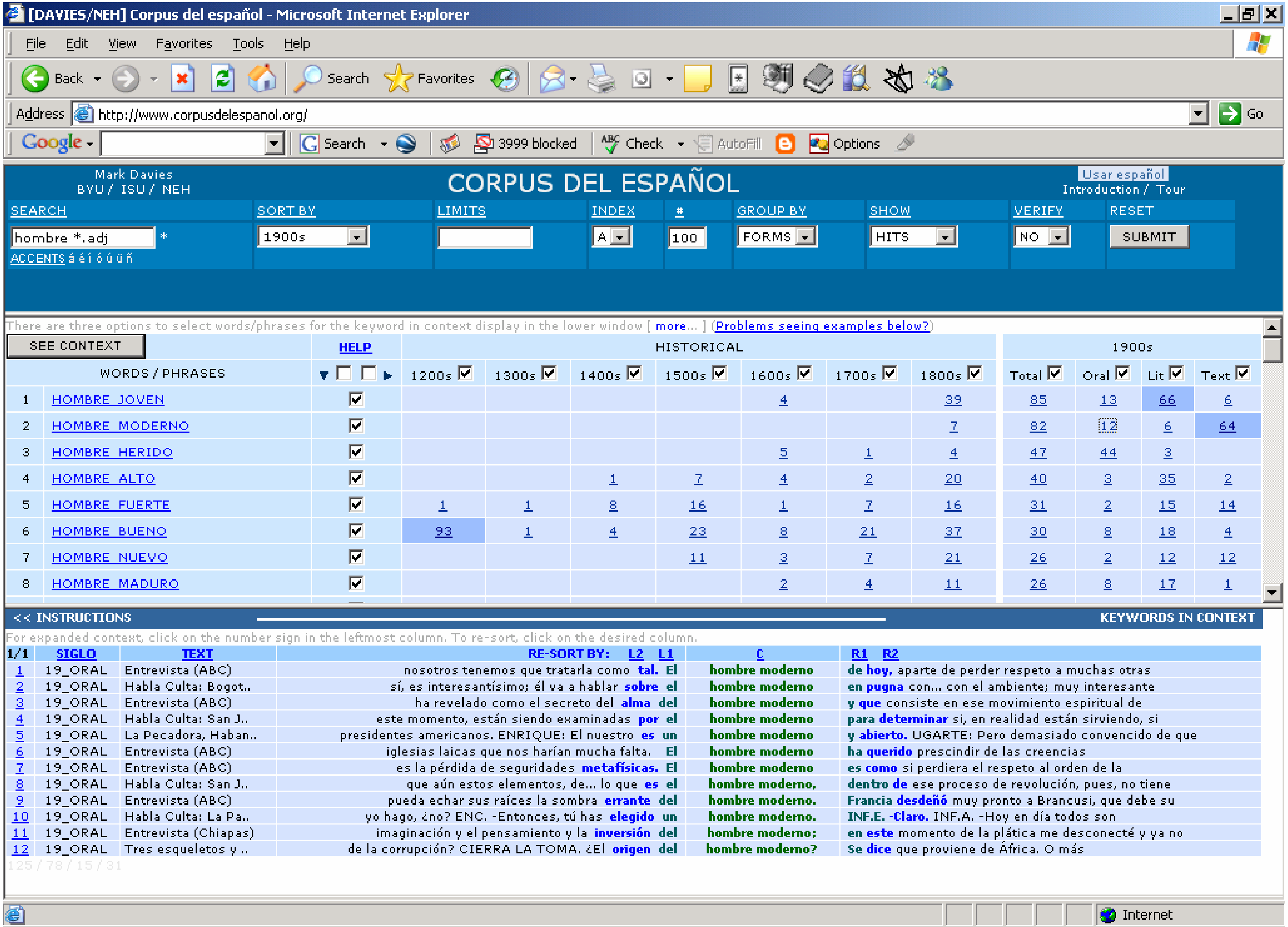
Accessibility

- 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 (www.corpusdelespanol.org) ; 100m words 1200s-1900s


[DAVIES/NEH] Corpus del español - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.corpusdelespanol.org/ Go
Google Search 3999 blocked Check AutoFill Options
Internet

Mark Davies
BYU / ISU / NEH

CORPUS DEL ESPAÑOL

Usar español
Introduction / Tour

SEARCH	SORT BY	LIMITS	INDEX	#	GROUP BY	SHOW	VERIFY	RESET
hombre *.adj *	1900s		A	100	FORMS	HITS	NO	SUBMIT
ACCENTS á é í ó ú ü ñ								

There are three options to select words/phrases for the keyword in context display in the lower window [more...] (Problems seeing examples below?)

SEE CONTEXT		HELP	HISTORICAL							1900s			
	WORDS / PHRASES	▼ □ □ ▶	1200s ☑	1300s ☑	1400s ☑	1500s ☑	1600s ☑	1700s ☑	1800s ☑	Total ☑	Oral ☑	Lit ☑	Text ☑
1	HOMBRE JOVEN	☑					4		39	85	13	66	6
2	HOMBRE MODERNO	☑							7	82	12	6	64
3	HOMBRE HERIDO	☑					5	1	4	47	44	3	
4	HOMBRE ALTO	☑			1	7	4	2	20	40	3	35	2
5	HOMBRE FUERTE	☑	1	1	8	16	1	7	16	31	2	15	14
6	HOMBRE BUENO	☑	93	1	4	23	8	21	37	30	8	18	4
7	HOMBRE NUEVO	☑				11	3	7	21	26	2	12	12
8	HOMBRE MADURO	☑					2	4	11	26	8	17	1

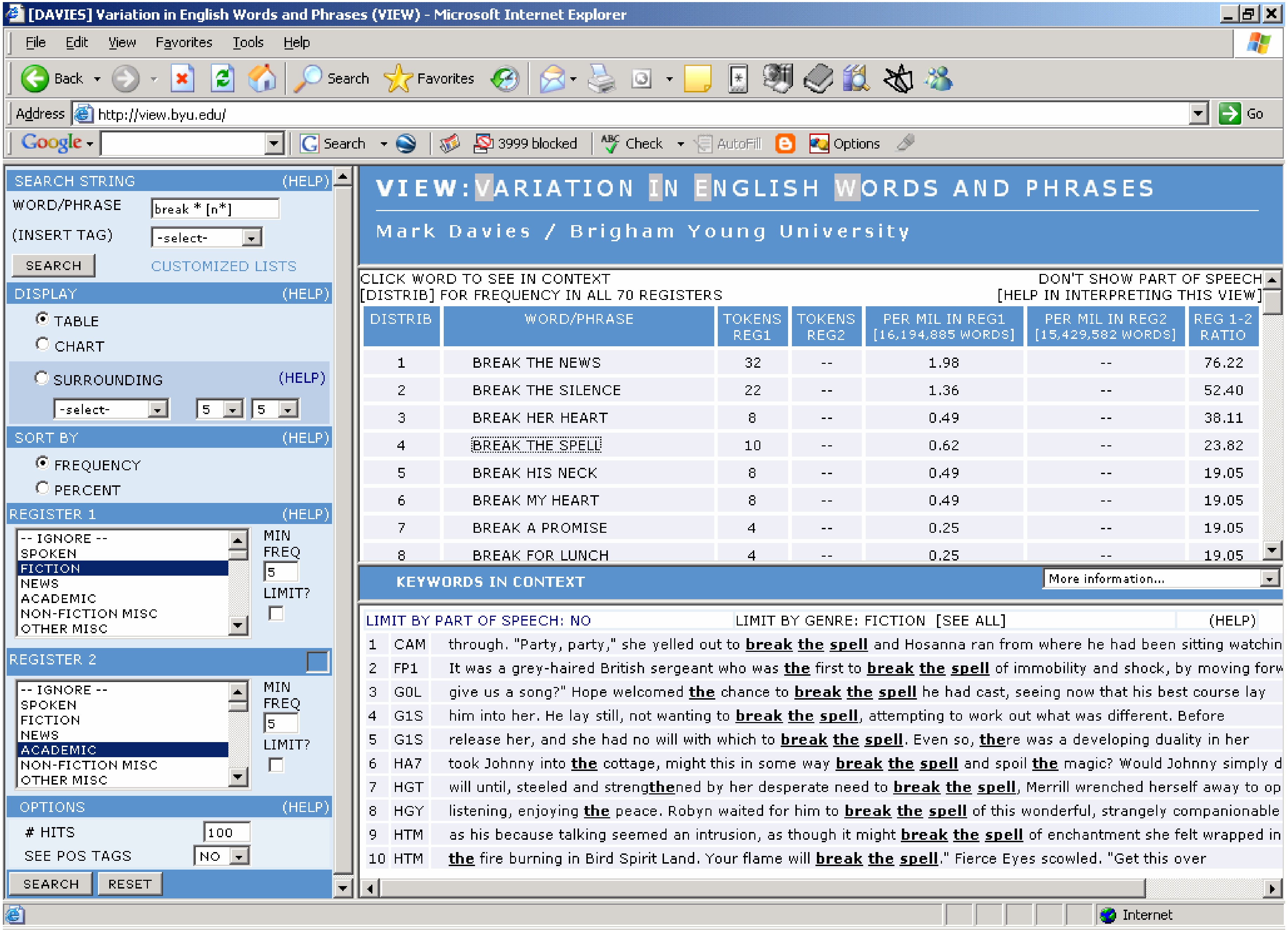
<< INSTRUCTIONS KEYWORDS IN CONTEXT

For expanded context, click on the number sign in the leftmost column. To re-sort, click on the desired column.

1/1	SIGLO	TEXT	RE-SORT BY: L2 L1	C	R1 R2
1	19_ORAL	Entrevista (ABC)	nosotros tenemos que tratarla como **tal. El**	**hombre moderno**	**de hoy,** aparte de perder respeto a muchas otras
2	19_ORAL	Habla Culta: Bogot..	sí, es interesantísimo; él va a hablar **sobre el**	**hombre moderno**	**en pugna** con... con el ambiente; muy interesante
3	19_ORAL	Entrevista (ABC)	ha revelado como el secreto del **alma del**	**hombre moderno**	**y que** consiste en ese movimiento espiritual de
4	19_ORAL	Habla Culta: San J..	este momento, están siendo examinadas **por el**	**hombre moderno**	**para determinar** si, en realidad están sirviendo, si
5	19_ORAL	La Pecadora, Haban..	presidentes americanos. ENRIQUE: El nuestro **es un**	**hombre moderno**	**y abierto.** UGARTE: Pero demasiado convencido de que
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)	es la pérdida de seguridades **metafísicas. El**	**hombre moderno**	**es como** si perdiera el respeto al orden de la
8	19_ORAL	Habla Culta: San J..	que aún estos elementos, de... lo que **es el**	**hombre moderno,**	**dentro de** ese proceso de revolución, pues, no tiene
9	19_ORAL	Entrevista (ABC)	pueda echar sus raíces la sombra **errante del**	**hombre moderno.**	**Francia desdeñó** muy pronto a Brancusi, que debe su
10	19_ORAL	Habla Culta: La Pa..	yo hago, ¿no? ENC. -Entonces, tú has **elegido un**	**hombre moderno.**	**INF.E. -Claro.** INF.A. -Hoy en día todos son
11	19_ORAL	Entrevista (Chiapas)	imaginación y el pensamiento y la **inversión del**	**hombre moderno;**	**en este** momento de la plática me desconecté y ya no
12	19_ORAL	Tres esqueletos y ..	de la corrupción? CIERRA LA TOMA. ¿El **origen del**	**hombre moderno?**	**Se dice** que proviene de África. O más

125 / 78 / 15 / 31

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



Address: http://view.byu.edu/

SEARCH STRING (HELP)

WORD/PHRASE: break * [n*]

(INSERT TAG): -select-

SEARCH   CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE
○ CHART
○ SURROUNDING (HELP)   -select-   5   5

SORT BY (HELP)

● FREQUENCY
○ PERCENT

REGISTER 1 (HELP)

-- IGNORE --
SPOKEN
FICTION
NEWS
ACADEMIC
NON-FICTION MISC
OTHER MISC

MIN FREQ: 5
LIMIT? □

REGISTER 2 □

-- IGNORE --
SPOKEN
FICTION
NEWS
ACADEMIC
NON-FICTION MISC
OTHER MISC

MIN FREQ: 5
LIMIT? □

OPTIONS (HELP)

HITS: 100
SEE POS TAGS: NO

SEARCH   RESET

VIEW: VARIATION IN ENGLISH WORDS AND PHRASES

Mark Davies / Brigham Young University

CLICK WORD TO SEE IN CONTEXT
[DISTRIB] FOR FREQUENCY IN ALL 70 REGISTERS

DON'T SHOW PART OF SPEECH
[HELP IN INTERPRETING THIS VIEW]

DISTRIB	WORD/PHRASE	TOKENS REG1	TOKENS REG2	PER MIL IN REG1 [16,194,885 WORDS]	PER MIL IN REG2 [15,429,582 WORDS]	REG 1-2 RATIO
1	BREAK THE NEWS	32	--	1.98	--	76.22
2	BREAK THE SILENCE	22	--	1.36	--	52.40
3	BREAK HER HEART	8	--	0.49	--	38.11
4	BREAK THE SPELL	10	--	0.62	--	23.82
5	BREAK HIS NECK	8	--	0.49	--	19.05
6	BREAK MY HEART	8	--	0.49	--	19.05
7	BREAK A PROMISE	4	--	0.25	--	19.05
8	BREAK FOR LUNCH	4	--	0.25	--	19.05

KEYWORDS IN CONTEXT   More information...

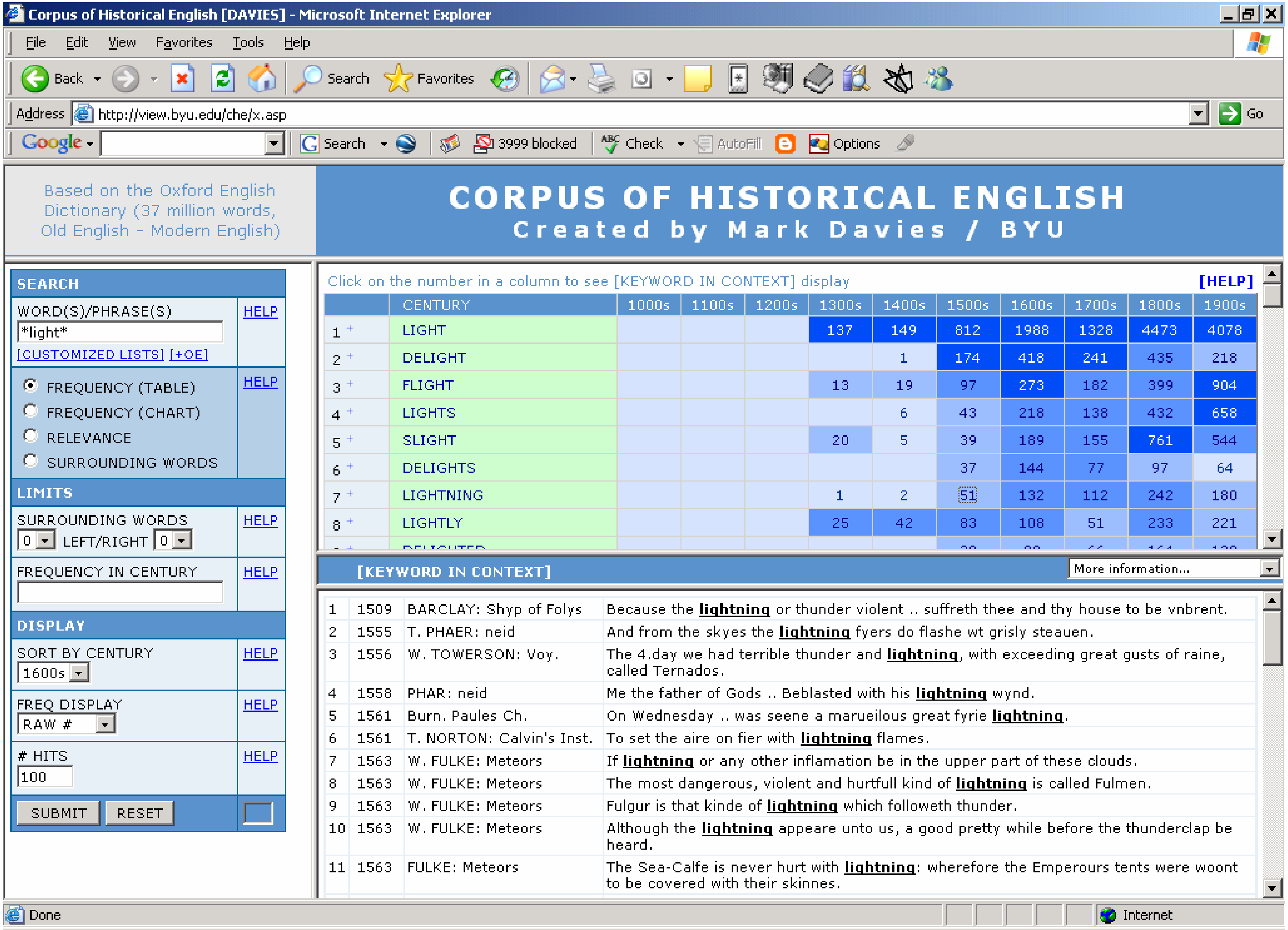
LIMIT BY PART OF SPEECH: NO   LIMIT BY GENRE: FICTION [SEE ALL]   (HELP)

1	CAM	through. "Party, party," she yelled out to **break the spell** and Hosanna ran from where he had been sitting watchin
2	FP1	It was a grey-haired British sergeant who was **the** first to **break the spell** of immobility and shock, by moving forw
3	G0L	give us a song?" Hope welcomed **the** chance to **break the spell** he had cast, seeing now that his best course lay
4	G1S	him into her. He lay still, not wanting to **break the spell**, attempting to work out what was different. Before
5	G1S	release her, and she had no will with which to **break the spell**. Even so, **the**re was a developing duality in her
6	HA7	took Johnny into **the** cottage, might this in some way **break the spell** and spoil **the** magic? Would Johnny simply d
7	HGT	will until, steeled and streng**the**ned by her desperate need to **break the spell**, Merrill wrenched herself away to op
8	HGY	listening, enjoying **the** peace. Robyn waited for him to **break the spell** of this wonderful, strangely companionable
9	HTM	as his because talking seemed an intrusion, as though it might **break the spell** of enchantment she felt wrapped in
10	HTM	**the** fire burning in Bird Spirit Land. Your flame will **break the spell**." Fierce Eyes scowled. "Get this over

BNC / VIEW features

Word document

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


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://view.byu.edu/che/x.asp Go
Google Search 3999 blocked Check AutoFill Options
Done
Internet

Based on the Oxford English Dictionary (37 million words, Old English – Modern English)

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S) HELP
light
[CUSTOMIZED LISTS] [+OE]

HELP
● FREQUENCY (TABLE)
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS HELP
0 LEFT/RIGHT 0

FREQUENCY IN CENTURY HELP

DISPLAY

SORT BY CENTURY HELP
1600s

FREQ DISPLAY HELP
RAW #

HITS HELP
100

SUBMIT RESET

Click on the number in a column to see [KEYWORD IN CONTEXT] display [HELP]

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1 +	LIGHT				137	149	812	1988	1328	4473	4078
2 +	DELIGHT					1	174	418	241	435	218
3 +	FLIGHT				13	19	97	273	182	399	904
4 +	LIGHTS					6	43	218	138	432	658
5 +	SLIGHT				20	5	39	189	155	761	544
6 +	DELIGHTS						37	144	77	97	64
7 +	LIGHTNING				1	2	51	132	112	242	180
8 +	LIGHTLY				25	42	83	108	51	233	221

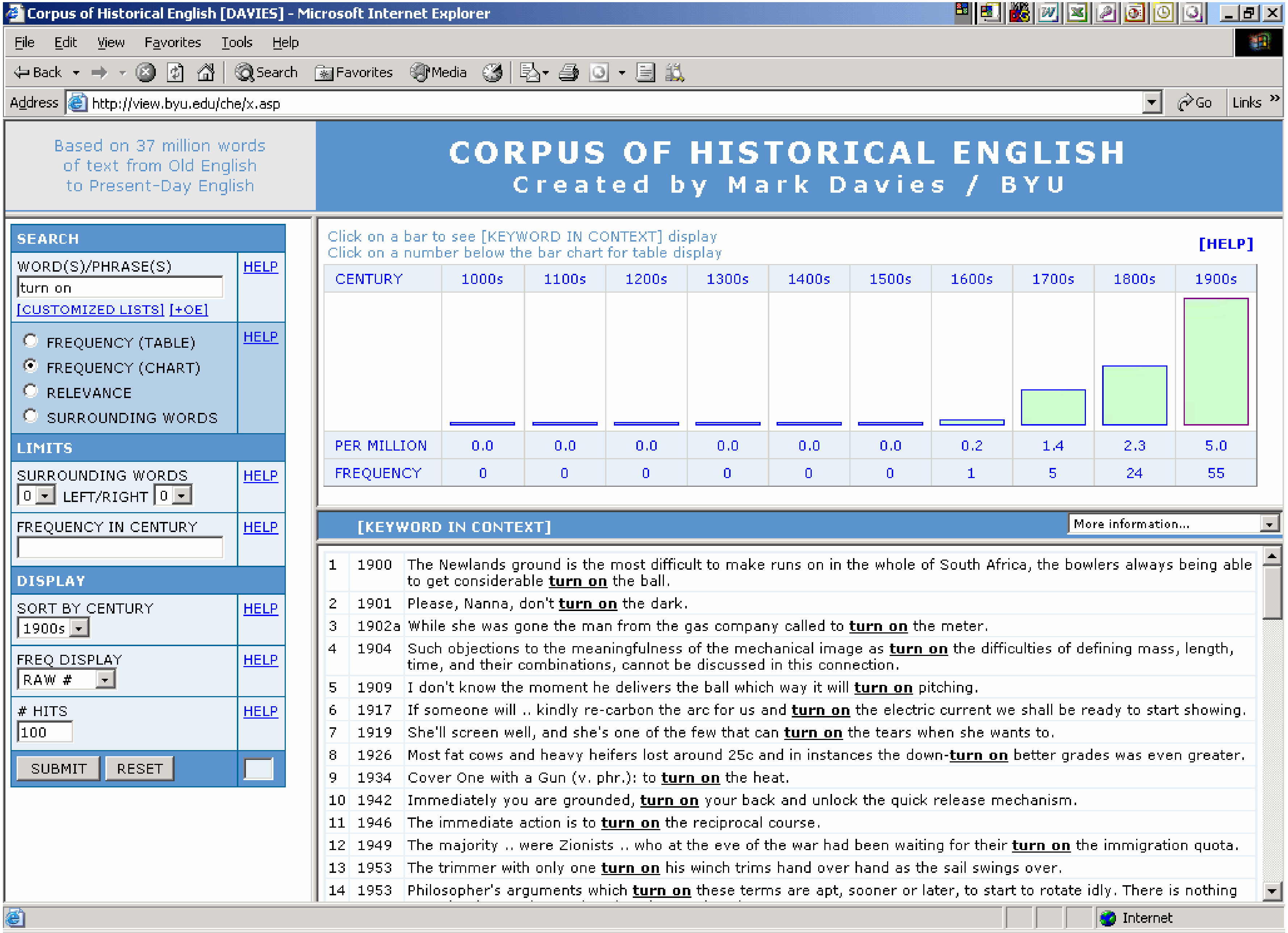
[KEYWORD IN CONTEXT] More information...

1	1509	BARCLAY: Shyp of Folys	Because the **lightning** or thunder violent .. suffreth thee and thy house to be vnbrent.
2	1555	T. PHAER: neid	And from the skyes the **lightning** fyers do flashe wt grisly steauen.
3	1556	W. TOWERSON: Voy.	The 4.day we had terrible thunder and **lightning**, with exceeding great gusts of raine, called Ternados.
4	1558	PHAR: neid	Me the father of Gods .. Beblasted with his **lightning** wynd.
5	1561	Burn. Paules Ch.	On Wednesday .. was seene a marueilous great fyrie **lightning**.
6	1561	T. NORTON: Calvin's Inst.	To set the aire on fier with **lightning** flames.
7	1563	W. FULKE: Meteors	If **lightning** or any other inflamation be in the upper part of these clouds.
8	1563	W. FULKE: Meteors	The most dangerous, violent and hurtfull kind of **lightning** is called Fulmen.
9	1563	W. FULKE: Meteors	Fulgur is that kinde of **lightning** which followeth thunder.
10	1563	W. FULKE: Meteors	Although the **lightning** appeare unto us, a good pretty while before the thunderclap be heard.
11	1563	FULKE: Meteors	The Sea-Calfe is never hurt with **lightning**: wherefore the Emperours tents were woont to be covered with their skinnes.

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] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): turn on HELP

[CUSTOMIZED LISTS] [+OE]

○ FREQUENCY (TABLE)
● FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS
HELP

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0 HELP

FREQUENCY IN CENTURY: HELP

DISPLAY

SORT BY CENTURY: 1900s HELP

FREQ DISPLAY: RAW # HELP

HITS: 100 HELP

SUBMIT RESET

Click on a bar to see [KEYWORD IN CONTEXT] display
Click on a number below the bar chart for table display

[HELP]

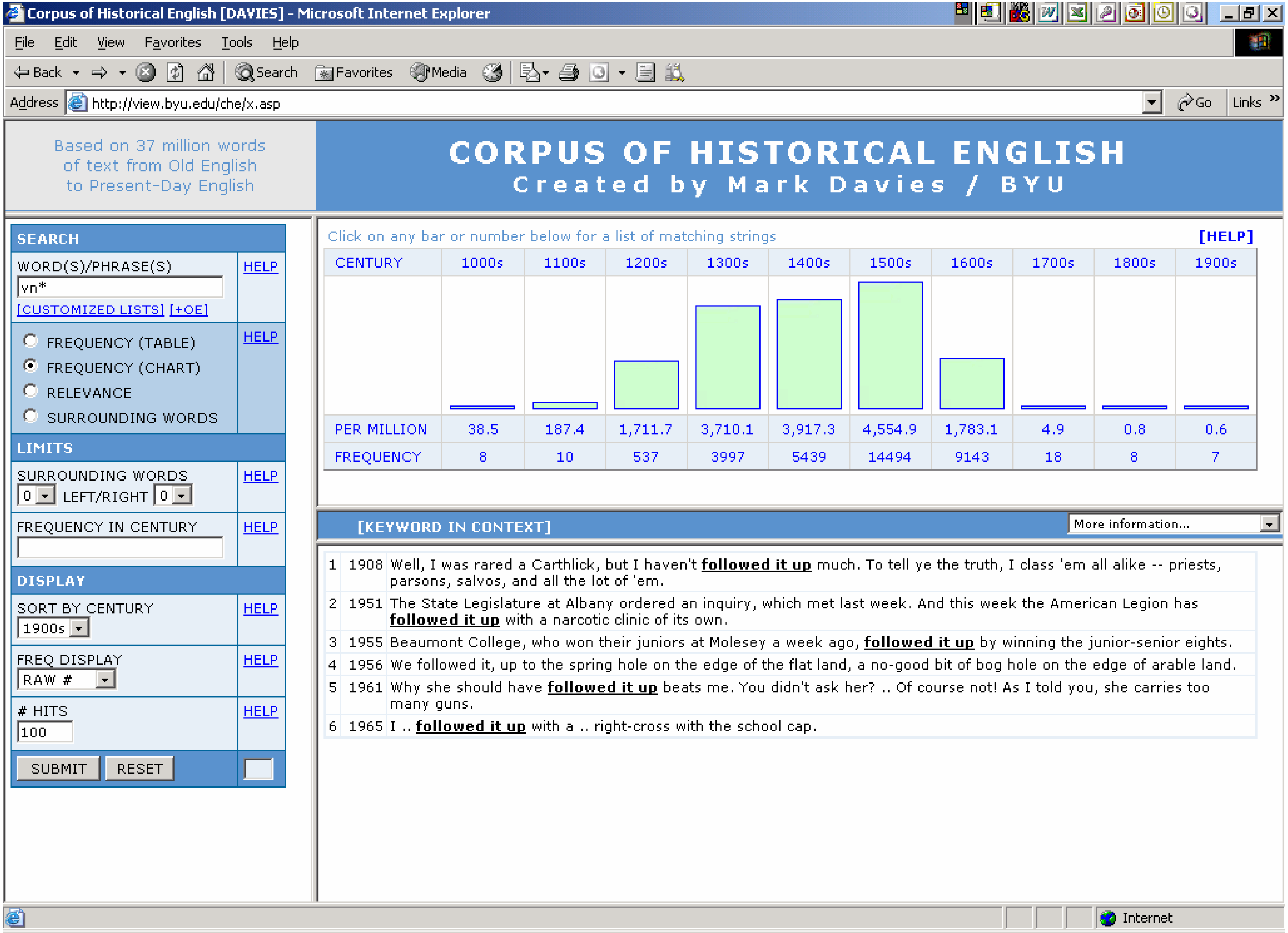
CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
PER MILLION	0.0	0.0	0.0	0.0	0.0	0.0	0.2	1.4	2.3	5.0
FREQUENCY	0	0	0	0	0	0	1	5	24	55

[KEYWORD IN CONTEXT]

More information...

1	1900	The Newlands ground is the most difficult to make runs on in the whole of South Africa, the bowlers always being able to get considerable **turn on** the ball.
2	1901	Please, Nanna, don't **turn on** the dark.
3	1902a	While she was gone the man from the gas company called to **turn on** the meter.
4	1904	Such objections to the meaningfulness of the mechanical image as **turn on** the difficulties of defining mass, length, time, and their combinations, cannot be discussed in this connection.
5	1909	I don't know the moment he delivers the ball which way it will **turn on** pitching.
6	1917	If someone will .. kindly re-carbon the arc for us and **turn on** the electric current we shall be ready to start showing.
7	1919	She'll screen well, and she's one of the few that can **turn on** the tears when she wants to.
8	1926	Most fat cows and heavy heifers lost around 25c and in instances the down-**turn on** better grades was even greater.
9	1934	Cover One with a Gun (v. phr.): to **turn on** the heat.
10	1942	Immediately you are grounded, **turn on** your back and unlock the quick release mechanism.
11	1946	The immediate action is to **turn on** the reciprocal course.
12	1949	The majority .. were Zionists .. who at the eve of the war had been waiting for their **turn on** the immigration quota.
13	1953	The trimmer with only one **turn on** his winch trims hand over hand as the sail swings over.
14	1953	Philosopher's arguments which **turn on** these terms are apt, sooner or later, to start to rotate idly. There is nothing

Spelling changes: **vn*** (chart)



http://view.byu.edu/che/x.asp

Based on 37 million words
of text from Old English
to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S) HELP

vn*

[CUSTOMIZED LISTS] [+OE]

HELP

○ FREQUENCY (TABLE)
● FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS HELP
0 LEFT/RIGHT 0

FREQUENCY IN CENTURY HELP

DISPLAY

SORT BY CENTURY HELP
1900s

FREQ DISPLAY HELP
RAW #

HITS HELP
100

SUBMIT RESET

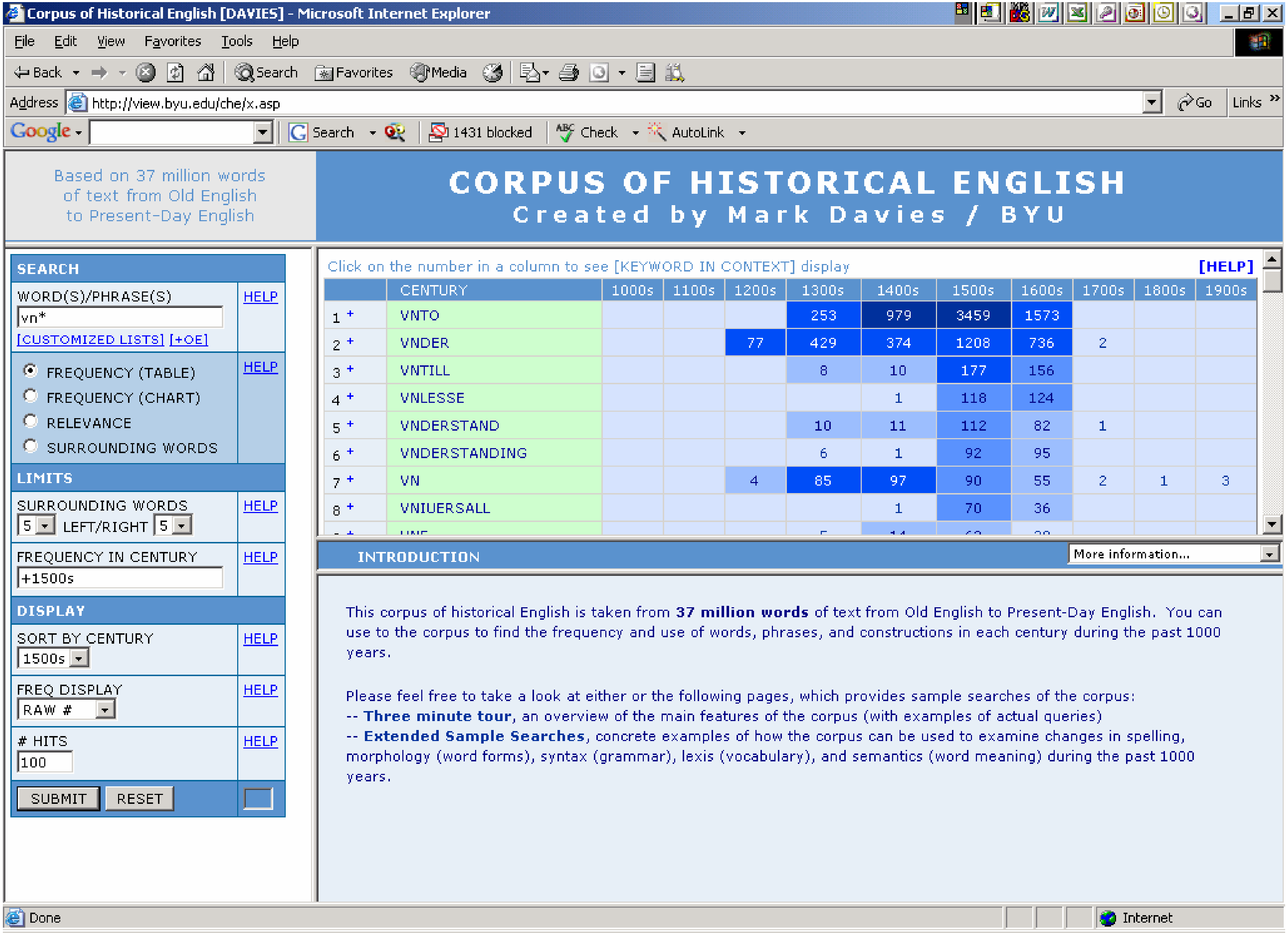
Click on any bar or number below for a list of matching strings [HELP]

CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
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
FREQUENCY	8	10	537	3997	5439	14494	9143	18	8	7

[KEYWORD IN CONTEXT] More information...

1	1908	Well, I was rared a Carthlick, but I haven't **followed it up** much. To tell ye the truth, I class 'em all alike -- priests, parsons, salvos, and all the lot of 'em.
2	1951	The State Legislature at Albany ordered an inquiry, which met last week. And this week the American Legion has **followed it up** with a narcotic clinic of its own.
3	1955	Beaumont College, who won their juniors at Molesey a week ago, **followed it up** by winning the junior-senior eights.
4	1956	We followed it, up to the spring hole on the edge of the flat land, a no-good bit of bog hole on the edge of arable land.
5	1961	Why she should have **followed it up** beats me. You didn't ask her? .. Of course not! As I told you, she carries too many guns.
6	1965	I .. **followed it up** with a .. right-cross with the school cap.

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


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Google Search 1431 blocked Check AutoLink

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): vn*

HELP
[CUSTOMIZED LISTS] [+OE]

● FREQUENCY (TABLE)
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS: 5 LEFT/RIGHT 5

FREQUENCY IN CENTURY: +1500s

DISPLAY

SORT BY CENTURY: 1500s

FREQ DISPLAY: RAW #

HITS: 100

Click on the number in a column to see [KEYWORD IN CONTEXT] display [HELP]

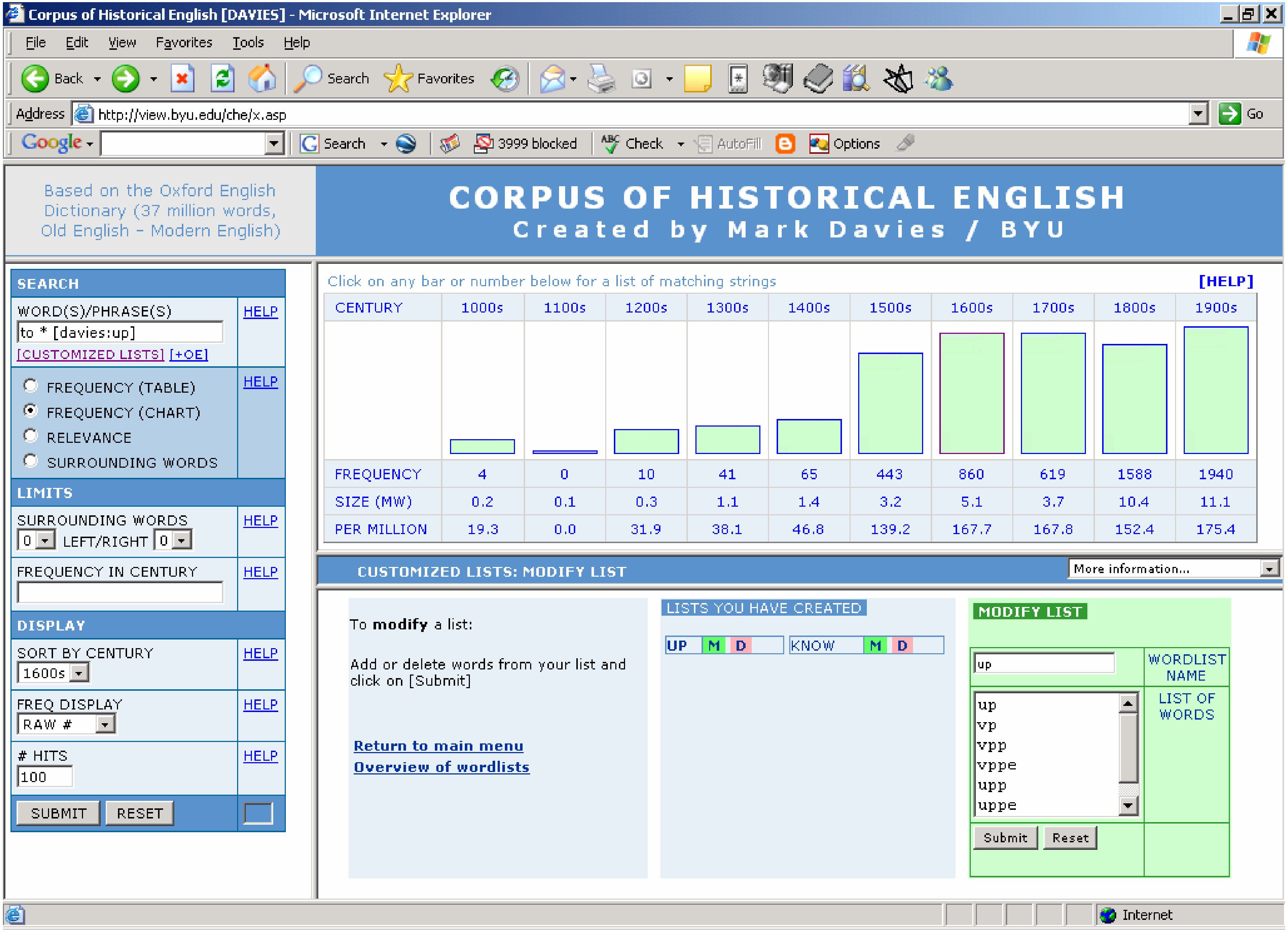
	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1 +	VNTO				253	979	3459	1573			
2 +	VNDER			77	429	374	1208	736	2		
3 +	VNTILL				8	10	177	156			
4 +	VNLESSE					1	118	124			
5 +	VNDERSTAND				10	11	112	82	1		
6 +	VNDERSTANDING				6	1	92	95			
7 +	VN			4	85	97	90	55	2	1	3
8 +	VNIUERSALL					1	70	36			
[illegible] +	[illegible]				[illegible]	[illegible]	[illegible]	[illegible]			

INTRODUCTION

This corpus of historical English is taken from **37 million words** of text 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.

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



Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Address http://view.byu.edu/che/x.asp
Go
Google
Search
3999 blocked
Check
AutoFill
Options
Internet

Based on the Oxford English Dictionary (37 million words, Old English - Modern English)

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): to * [davies:up]

[CUSTOMIZED LISTS] [+OE]

○ FREQUENCY (TABLE)
● FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0

FREQUENCY IN CENTURY:

DISPLAY

SORT BY CENTURY: 1600s

FREQ DISPLAY: RAW #

HITS: 100

Click on any bar or number below for a list of matching strings [HELP]

CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
FREQUENCY	4	0	10	41	65	443	860	619	1588	1940
SIZE (MW)	0.2	0.1	0.3	1.1	1.4	3.2	5.1	3.7	10.4	11.1
PER MILLION	19.3	0.0	31.9	38.1	46.8	139.2	167.7	167.8	152.4	175.4

CUSTOMIZED LISTS: MODIFY LIST

To **modify** a list:

Add or delete words from your list and click on [Submit]

Return to main menu
Overview of wordlists

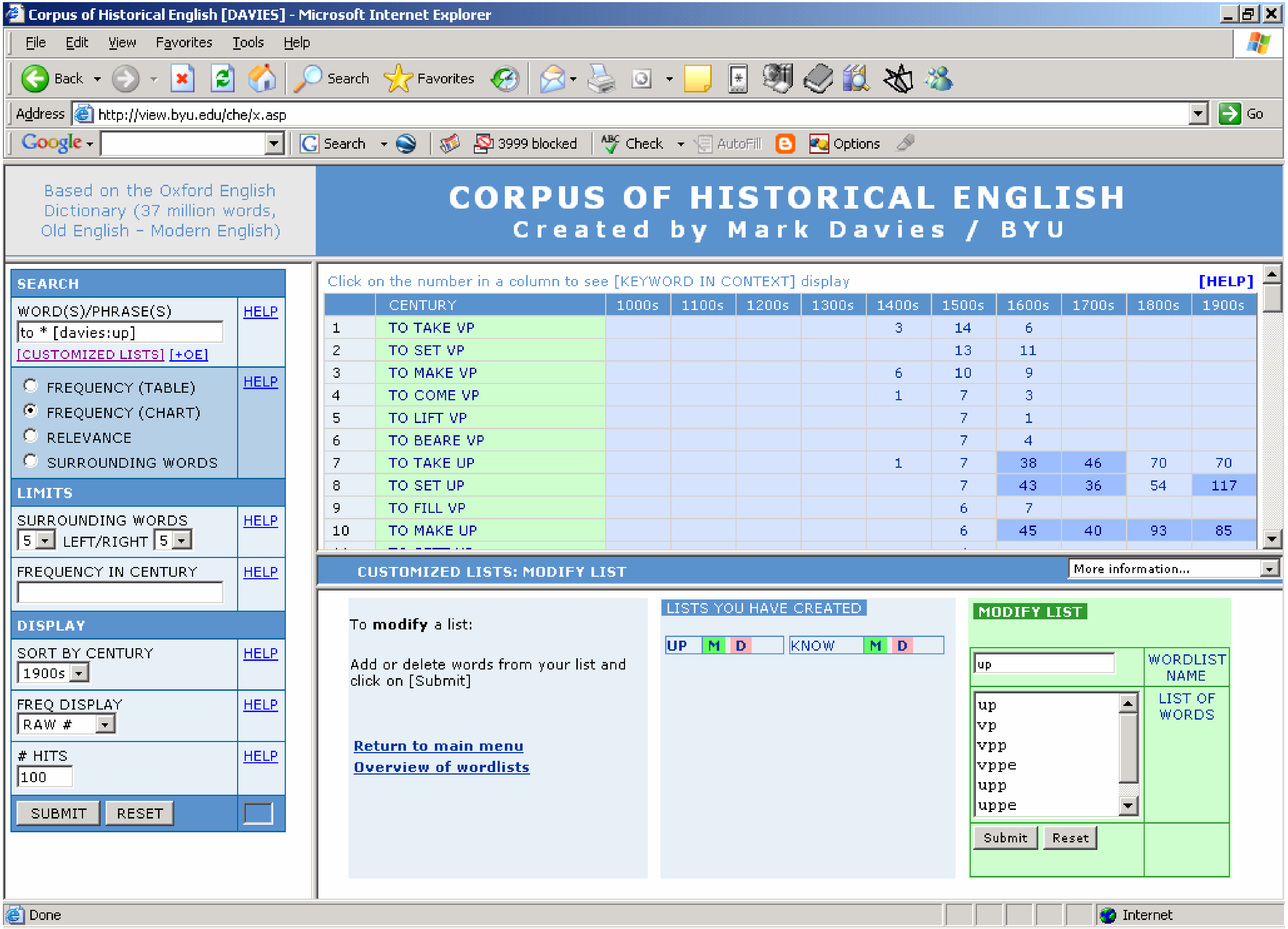
LISTS YOU HAVE CREATED

UP	M	D	KNOW	M	D

MODIFY LIST

up	WORDLIST NAME
up vp vpp vppe upp uppe	LIST OF WORDS

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



Based on the Oxford English Dictionary (37 million words, Old English – Modern English)

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S) HELP
to * [davies:up]
[CUSTOMIZED LISTS] [+OE]

HELP
○ FREQUENCY (TABLE)
● FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS HELP
5 LEFT/RIGHT 5

FREQUENCY IN CENTURY HELP

DISPLAY

SORT BY CENTURY HELP
1900s

FREQ DISPLAY HELP
RAW #

HITS HELP
100

SUBMIT RESET

Click on the number in a column to see [KEYWORD IN CONTEXT] display **[HELP]**

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1	TO TAKE VP					3	14	6			
2	TO SET VP						13	11			
3	TO MAKE VP					6	10	9			
4	TO COME VP					1	7	3			
5	TO LIFT VP						7	1			
6	TO BEARE VP						7	4			
7	TO TAKE UP					1	7	38	46	70	70
8	TO SET UP						7	43	36	54	117
9	TO FILL VP						6	7			
10	TO MAKE UP						6	45	40	93	85

CUSTOMIZED LISTS: MODIFY LIST

More information...

To **modify** a list:

Add or delete words from your list and click on [Submit]

Return to main menu
Overview of wordlists

LISTS YOU HAVE CREATED

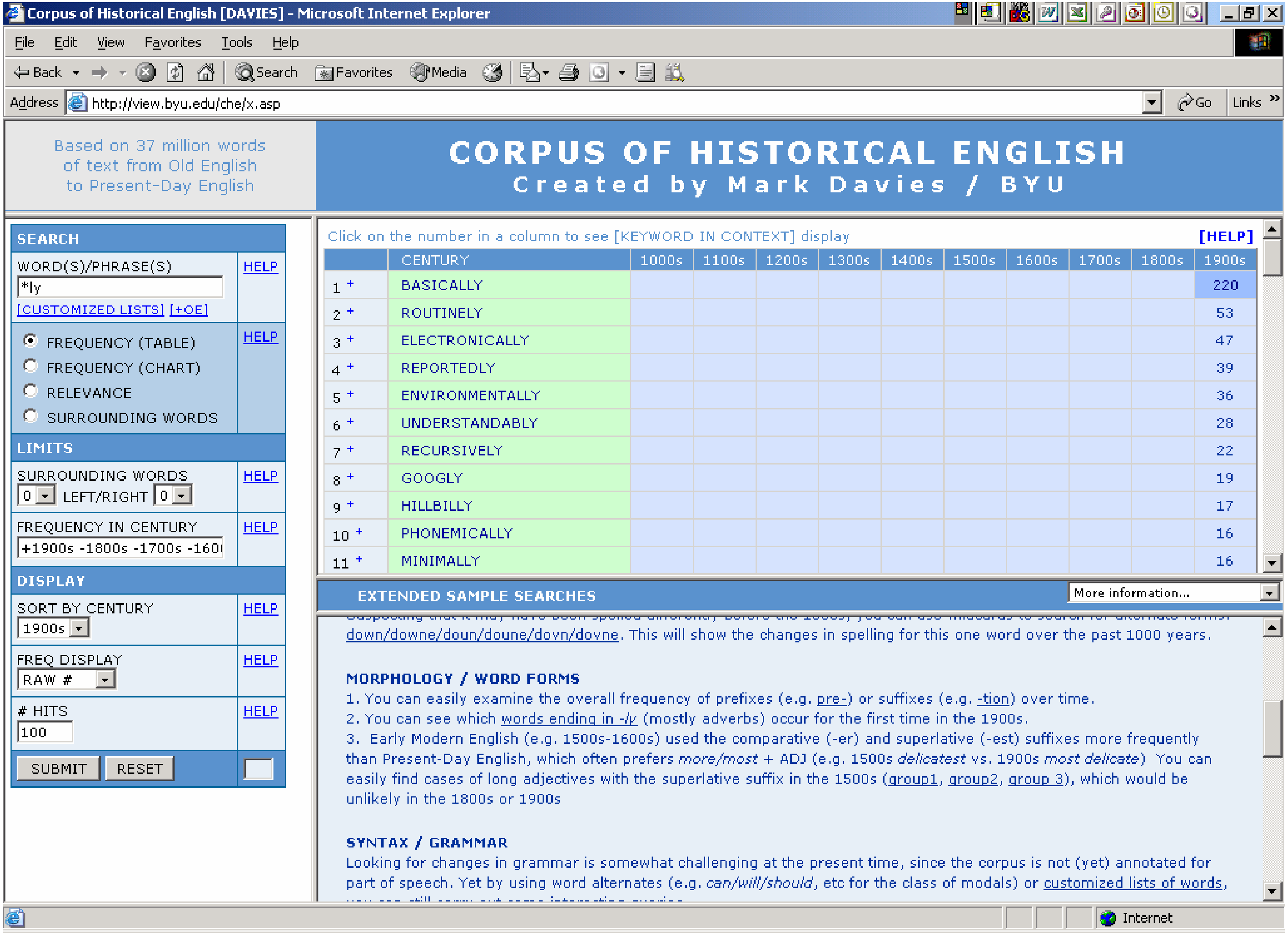
UP	M	D	KNOW	M	D

MODIFY LIST

up	WORDLIST NAME
up vp vpp vppe upp uppe	LIST OF WORDS

Submit Reset

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


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): *ly HELP
[CUSTOMIZED LISTS] [+OE]

● FREQUENCY (TABLE) HELP
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0 HELP

FREQUENCY IN CENTURY: +1900s -1800s -1700s -160 HELP

DISPLAY

SORT BY CENTURY: 1900s HELP

FREQ DISPLAY: RAW # HELP

HITS: 100 HELP

SUBMIT RESET

Click on the number in a column to see [KEYWORD IN CONTEXT] display [HELP]

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1 +	BASICALLY										220
2 +	ROUTINELY										53
3 +	ELECTRONICALLY										47
4 +	REPORTEDLY										39
5 +	ENVIRONMENTALLY										36
6 +	UNDERSTANDABLY										28
7 +	RECURSIVELY										22
8 +	GOOGLY										19
9 +	HILLBILLY										17
10 +	PHONEMICALLY										16
11 +	MINIMALLY										16

EXTENDED SAMPLE SEARCHES

More information...

down/downe/doun/doune/dovn/dovne. This will show the changes in spelling for this one word over the past 1000 years.

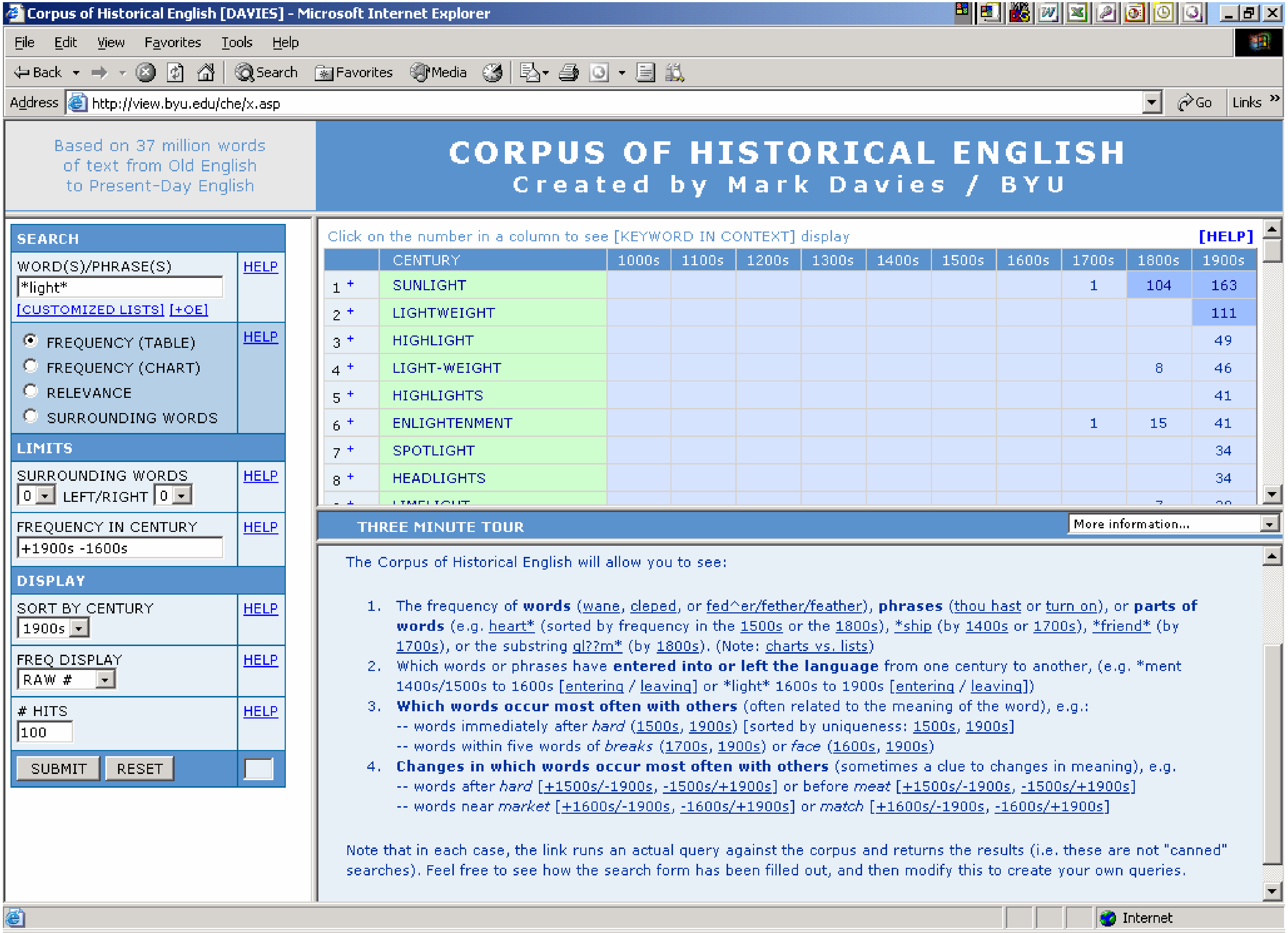
MORPHOLOGY / WORD FORMS

1. You can easily examine the overall frequency of prefixes (e.g. pre-) or suffixes (e.g. -tion) over time.
2. You can see which words ending in -ly (mostly adverbs) occur for the first time in the 1900s.
3. Early Modern English (e.g. 1500s-1600s) used the comparative (-er) and superlative (-est) suffixes more frequently than Present-Day English, which often prefers *more/most* + ADJ (e.g. 1500s *delicatest* vs. 1900s *most delicate*) You can easily find cases of long adjectives with the superlative suffix in the 1500s (group1, group2, group 3), which would be unlikely in the 1800s or 1900s

SYNTAX / GRAMMAR

Looking for changes in grammar is somewhat challenging at the present time, since the corpus is not (yet) annotated for part of speech. Yet by using word alternates (e.g. *can/will/should*, etc for the class of modals) or customized lists of words,

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


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
http://view.byu.edu/che/x.asp
Go
Links
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): *light* HELP
[CUSTOMIZED LISTS] [+OE]

● FREQUENCY (TABLE)
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS
HELP

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0 HELP

FREQUENCY IN CENTURY: +1900s -1600s HELP

DISPLAY

SORT BY CENTURY: 1900s HELP

FREQ DISPLAY: RAW # HELP

HITS: 100 HELP

SUBMIT RESET

Click on the number in a column to see [KEYWORD IN CONTEXT] display **[HELP]**

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1 +	SUNLIGHT								1	104	163
2 +	LIGHTWEIGHT										111
3 +	HIGHLIGHT										49
4 +	LIGHT-WEIGHT									8	46
5 +	HIGHLIGHTS										41
6 +	ENLIGHTENMENT								1	15	41
7 +	SPOTLIGHT										34
8 +	HEADLIGHTS										34

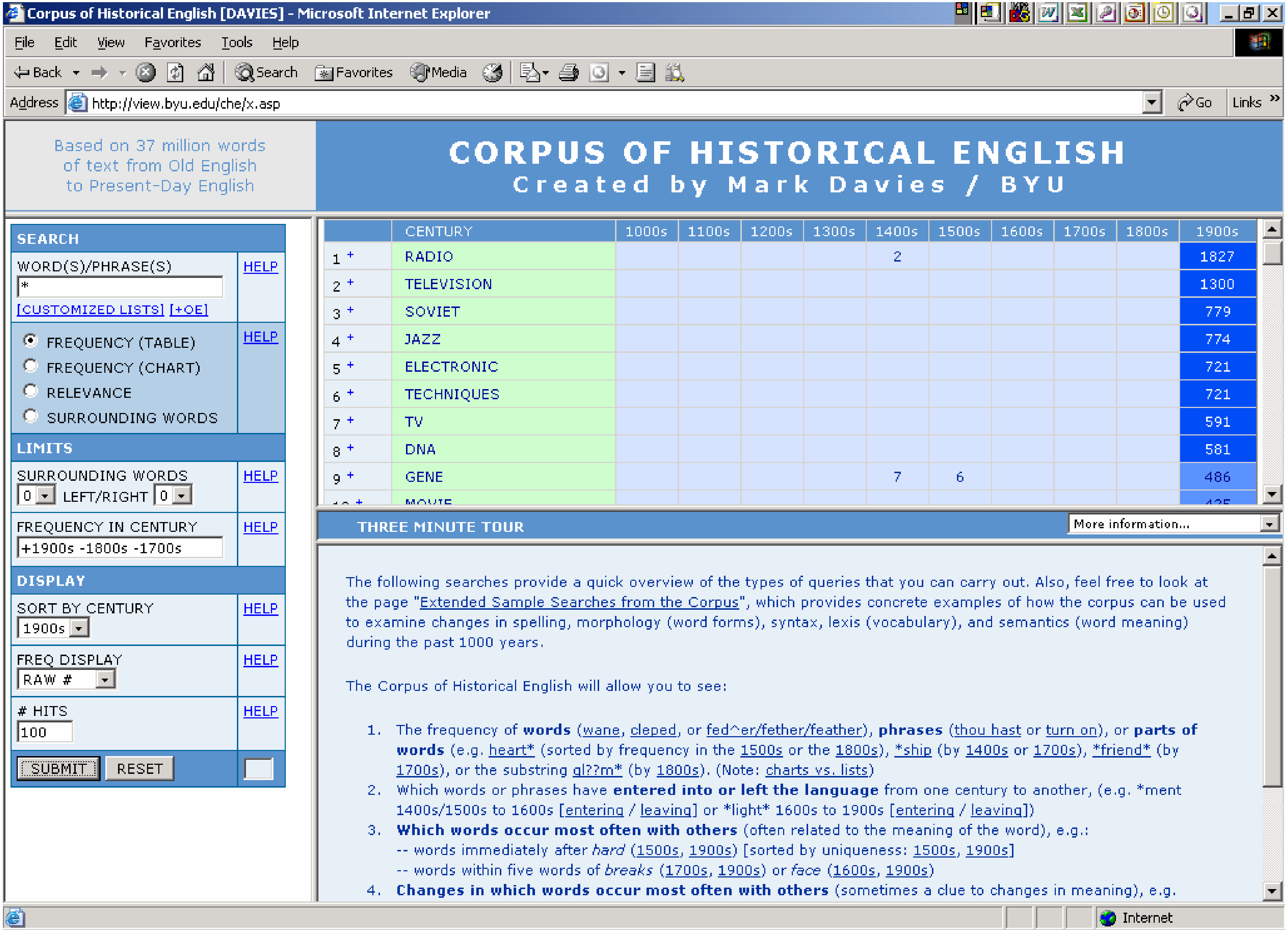
THREE MINUTE TOUR More information...

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 gl??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 within five words of *breaks* (1700s, 1900s) or *face* (1600s, 1900s)
4. **Changes in which words occur most often with others** (sometimes a clue to changes in meaning), e.g.
   -- words after *hard* [+1500s/-1900s, -1500s/+1900s] or before *meat* [+1500s/-1900s, -1500s/+1900s]
   -- words near *market* [+1600s/-1900s, -1600s/+1900s] or *match* [+1600s/-1900s, -1600s/+1900s]

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.

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


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH

Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): * HELP

[CUSTOMIZED LISTS] [+OE]

● FREQUENCY (TABLE) HELP
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0 HELP

FREQUENCY IN CENTURY: +1900s -1800s -1700s HELP

DISPLAY

SORT BY CENTURY: 1900s HELP

FREQ DISPLAY: RAW # HELP

HITS: 100 HELP

SUBMIT RESET

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1 +	RADIO					2					1827
2 +	TELEVISION										1300
3 +	SOVIET										779
4 +	JAZZ										774
5 +	ELECTRONIC										721
6 +	TECHNIQUES										721
7 +	TV										591
8 +	DNA										581
9 +	GENE					7	6				486
10 +	MOVIE										425

THREE MINUTE TOUR

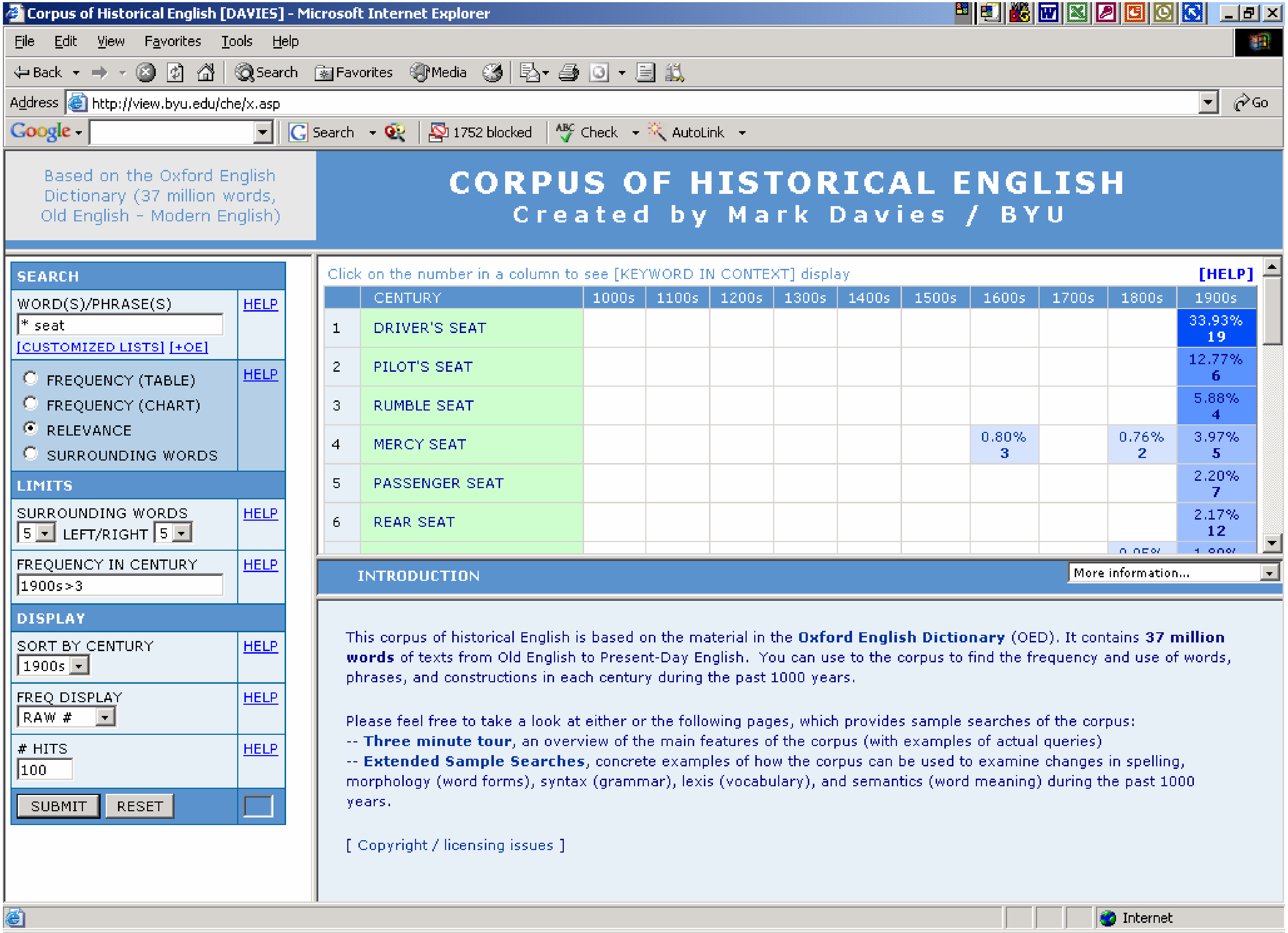
More information...

The following searches provide a quick overview of the types of queries that you can carry out. Also, feel free to look at the page "Extended Sample Searches from the Corpus", which provides concrete examples of how the corpus can be used to examine changes in spelling, morphology (word forms), syntax, lexis (vocabulary), and semantics (word meaning) during the past 1000 years.

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 gl??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 within five words of *breaks* (1700s, 1900s) or *face* (1600s, 1900s)
4. **Changes in which words occur most often with others** (sometimes a clue to changes in meaning), e.g.

Relevancy / Z-score like listing: * **seat**


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go
Google Search 1752 blocked Check AutoLink
Internet

Based on the Oxford English Dictionary (37 million words, Old English - Modern English)

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S) HELP
* seat
[CUSTOMIZED LISTS] [+OE]

HELP
○ FREQUENCY (TABLE)
○ FREQUENCY (CHART)
● RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS HELP
5 LEFT/RIGHT 5

FREQUENCY IN CENTURY HELP
1900s>3

DISPLAY

SORT BY CENTURY HELP
1900s

FREQ DISPLAY HELP
RAW #

HITS HELP
100

SUBMIT RESET

Click on the number in a column to see [KEYWORD IN CONTEXT] display **[HELP]**

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1	DRIVER'S SEAT										33.93% **19**
2	PILOT'S SEAT										12.77% **6**
3	RUMBLE SEAT										5.88% **4**
4	MERCY SEAT							0.80% **3**		0.76% **2**	3.97% **5**
5	PASSENGER SEAT										2.20% **7**
6	REAR SEAT										2.17% **12**

INTRODUCTION

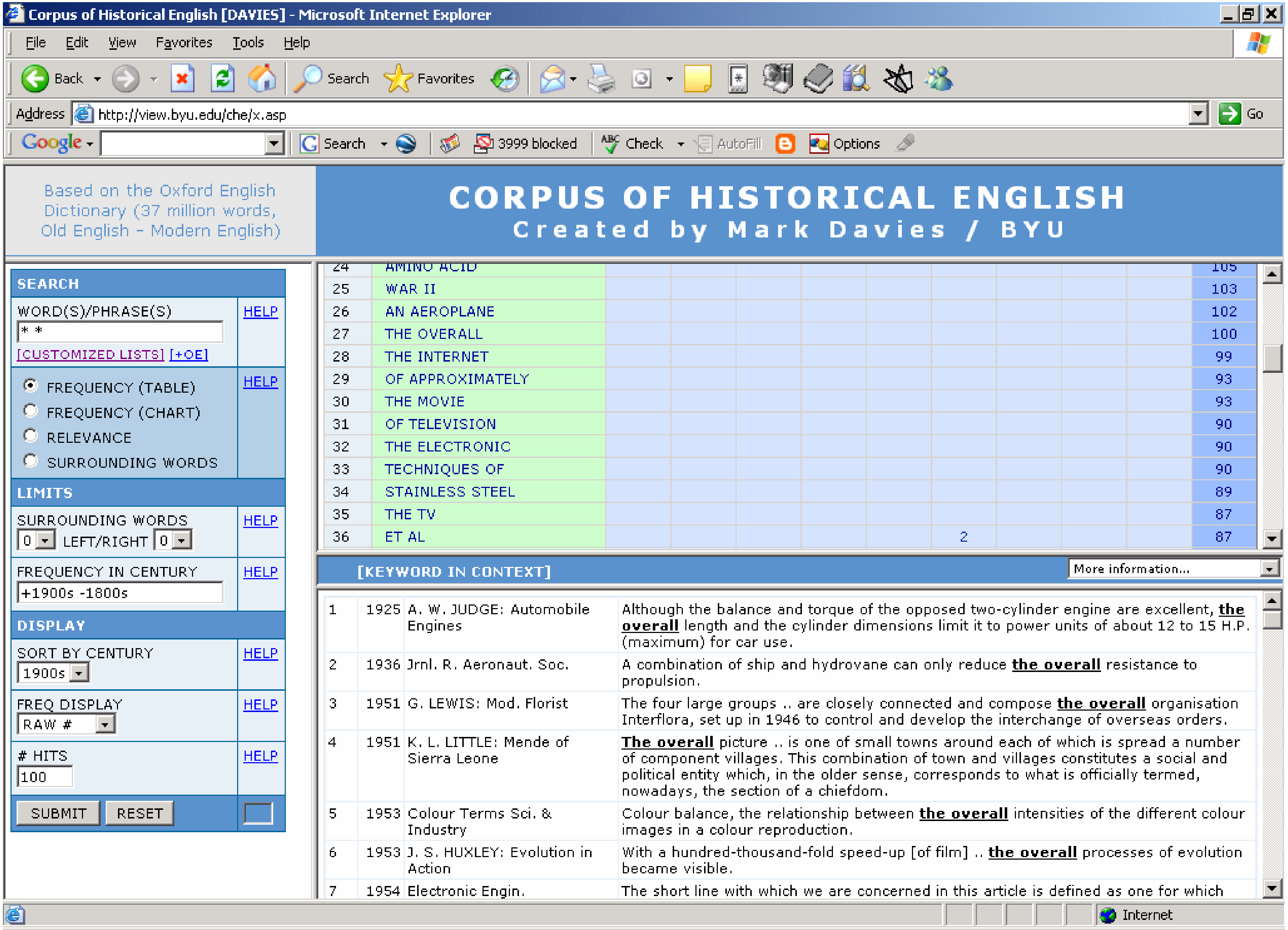
More information...

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] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://view.byu.edu/che/x.asp Go
Google Search 3999 blocked Check AutoFill Options
Internet

Based on the Oxford English Dictionary (37 million words, Old English – Modern English)

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): * * HELP
[CUSTOMIZED LISTS] [+OE]

- ● FREQUENCY (TABLE)
- ○ FREQUENCY (CHART)
- ○ RELEVANCE
- ○ SURROUNDING WORDS

HELP

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0 HELP

FREQUENCY IN CENTURY: +1900s -1800s HELP

DISPLAY

SORT BY CENTURY: 1900s HELP

FREQ DISPLAY: RAW # HELP

HITS: 100 HELP

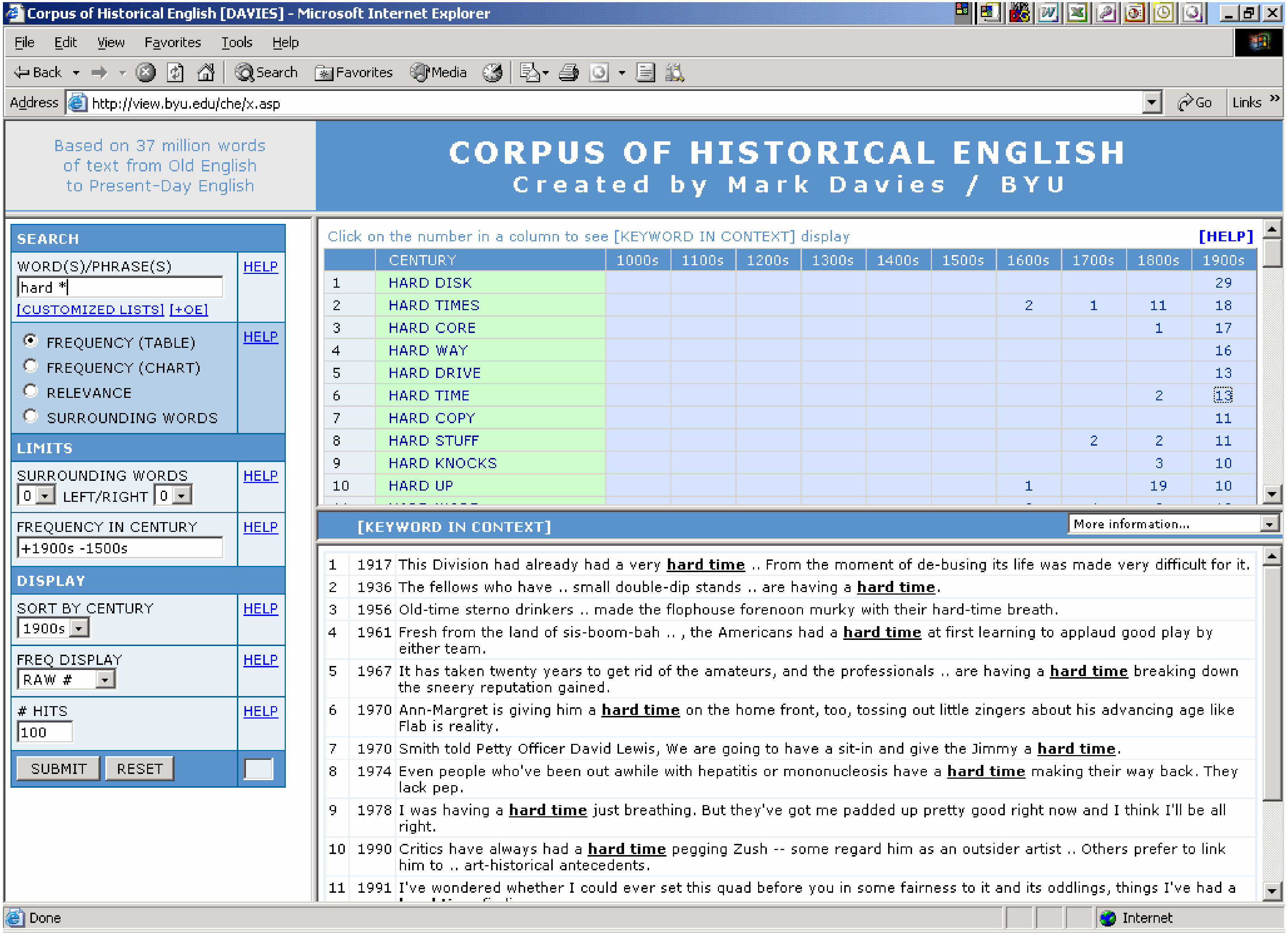
SUBMIT RESET

24	AMINO ACID										105
25	WAR II										103
26	AN AEROPLANE										102
27	THE OVERALL										100
28	THE INTERNET										99
29	OF APPROXIMATELY										93
30	THE MOVIE										93
31	OF TELEVISION										90
32	THE ELECTRONIC										90
33	TECHNIQUES OF										90
34	STAINLESS STEEL										89
35	THE TV										87
36	ET AL						2				87

[KEYWORD IN CONTEXT] More information...

1	1925	A. W. JUDGE: Automobile Engines	Although the balance and torque of the opposed two-cylinder engine are excellent, **the overall** length and the cylinder dimensions limit it to power units of about 12 to 15 H.P. (maximum) for car use.
2	1936	Jrnl. R. Aeronaut. Soc.	A combination of ship and hydrovane can only reduce **the overall** resistance to propulsion.
3	1951	G. LEWIS: Mod. Florist	The four large groups .. are closely connected and compose **the overall** organisation Interflora, set up in 1946 to control and develop the interchange of overseas orders.
4	1951	K. L. LITTLE: Mende of Sierra Leone	**The overall** picture .. is one of small towns around each of which is spread a number of component villages. This combination of town and villages constitutes a social and political entity which, in the older sense, corresponds to what is officially termed, nowadays, the section of a chiefdom.
5	1953	Colour Terms Sci. & Industry	Colour balance, the relationship between **the overall** intensities of the different colour images in a colour reproduction.
6	1953	J. S. HUXLEY: Evolution in Action	With a hundred-thousand-fold speed-up [of film] .. **the overall** processes of evolution became visible.
7	1954	Electronic Engin.	The short line with which we are concerned in this article is defined as one for which

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


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Done
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): hard *
HELP
[CUSTOMIZED LISTS] [+OE]

● FREQUENCY (TABLE)
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS
HELP

LIMITS

SURROUNDING WORDS: 0 LEFT/RIGHT 0
HELP

FREQUENCY IN CENTURY: +1900s -1500s
HELP

DISPLAY

SORT BY CENTURY: 1900s
HELP

FREQ DISPLAY: RAW #
HELP

HITS: 100
HELP

SUBMIT RESET

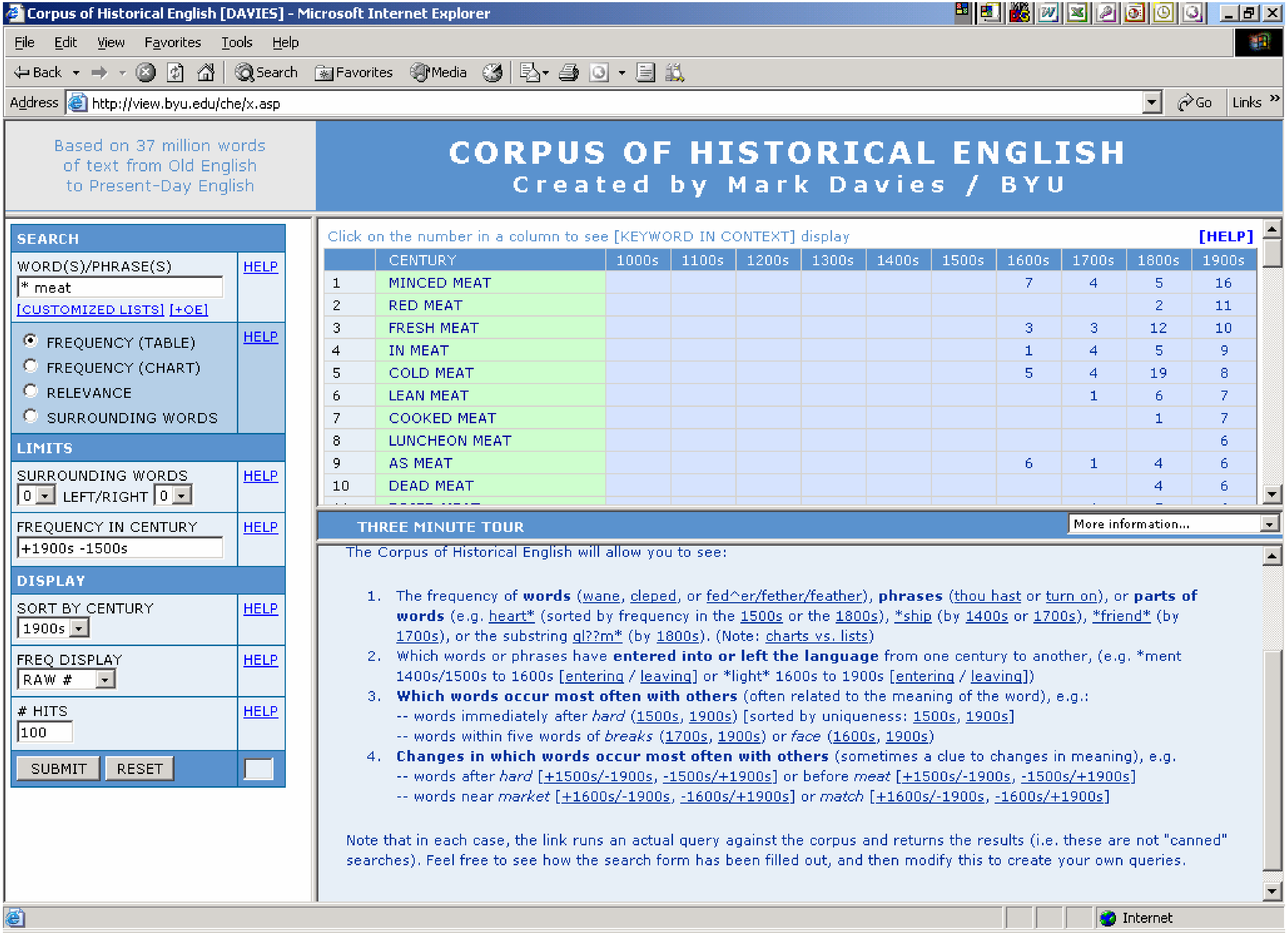
Click on the number in a column to see [KEYWORD IN CONTEXT] display [HELP]

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1	HARD DISK										29
2	HARD TIMES							2	1	11	18
3	HARD CORE									1	17
4	HARD WAY										16
5	HARD DRIVE										13
6	HARD TIME									2	13
7	HARD COPY										11
8	HARD STUFF								2	2	11
9	HARD KNOCKS									3	10
10	HARD UP							1		19	10

[KEYWORD IN CONTEXT] More information...

1	1917	This Division had already had a very **hard time** .. From the moment of de-busing its life was made very difficult for it.
2	1936	The fellows who have .. small double-dip stands .. are having a **hard time**.
3	1956	Old-time sterno drinkers .. made the flophouse forenoon murky with their hard-time breath.
4	1961	Fresh from the land of sis-boom-bah .. , the Americans had a **hard time** at first learning to applaud good play by either team.
5	1967	It has taken twenty years to get rid of the amateurs, and the professionals .. are having a **hard time** breaking down the sneery reputation gained.
6	1970	Ann-Margret is giving him a **hard time** on the home front, too, tossing out little zingers about his advancing age like Flab is reality.
7	1970	Smith told Petty Officer David Lewis, We are going to have a sit-in and give the Jimmy a **hard time**.
8	1974	Even people who've been out awhile with hepatitis or mononucleosis have a **hard time** making their way back. They lack pep.
9	1978	I was having a **hard time** just breathing. But they've got me padded up pretty good right now and I think I'll be all right.
10	1990	Critics have always had a **hard time** pegging Zush -- some regard him as an outsider artist .. Others prefer to link him to .. art-historical antecedents.
11	1991	I've wondered whether I could ever set this quad before you in some fairness to it and its oddlings, things I've had a

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



Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S) HELP: * meat
[CUSTOMIZED LISTS] [+OE]

● FREQUENCY (TABLE) HELP
○ FREQUENCY (CHART)
○ RELEVANCE
○ SURROUNDING WORDS

LIMITS

SURROUNDING WORDS HELP: 0 LEFT/RIGHT 0

FREQUENCY IN CENTURY HELP: +1900s -1500s

DISPLAY

SORT BY CENTURY HELP: 1900s

FREQ DISPLAY HELP: RAW #

HITS HELP: 100

SUBMIT RESET

Click on the number in a column to see [KEYWORD IN CONTEXT] display [HELP]

	CENTURY	1000s	1100s	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
1	MINCED MEAT							7	4	5	16
2	RED MEAT									2	11
3	FRESH MEAT							3	3	12	10
4	IN MEAT							1	4	5	9
5	COLD MEAT							5	4	19	8
6	LEAN MEAT								1	6	7
7	COOKED MEAT									1	7
8	LUNCHEON MEAT										6
9	AS MEAT							6	1	4	6
10	DEAD MEAT									4	6

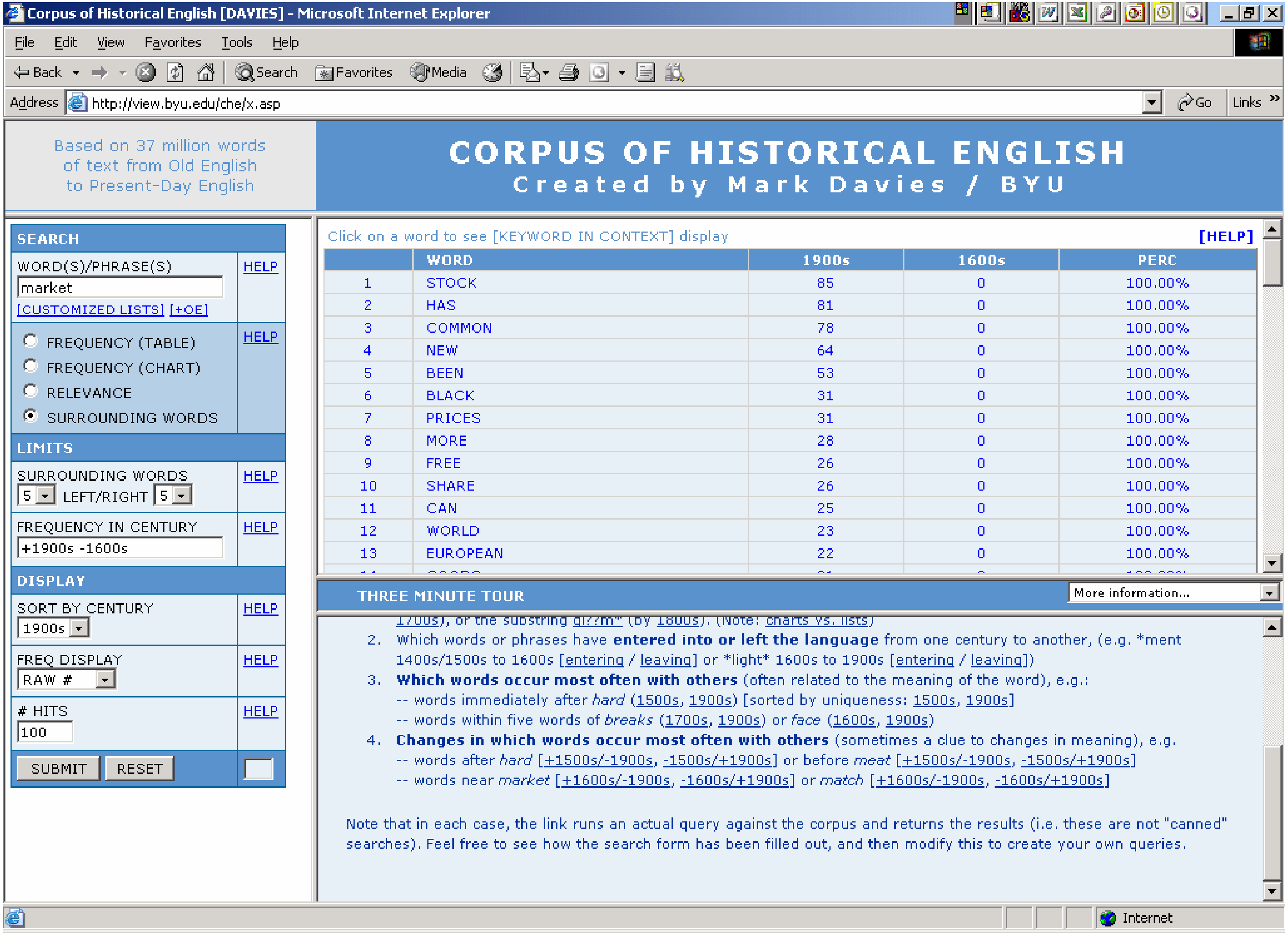
THREE MINUTE TOUR

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 gl??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 within five words of *breaks* (1700s, 1900s) or *face* (1600s, 1900s)
4. **Changes in which words occur most often with others** (sometimes a clue to changes in meaning), e.g.
   -- words after *hard* [+1500s/-1900s, -1500s/+1900s] or before *meat* [+1500s/-1900s, -1500s/+1900s]
   -- words near *market* [+1600s/-1900s, -1600s/+1900s] or *match* [+1600s/-1900s, -1600s/+1900s]

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.

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



Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH

Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S): market

HELP
[CUSTOMIZED LISTS] [+OE]

○ FREQUENCY (TABLE)
○ FREQUENCY (CHART)
○ RELEVANCE
● SURROUNDING WORDS

LIMITS

SURROUNDING WORDS: 5 LEFT/RIGHT 5

FREQUENCY IN CENTURY: +1900s -1600s

DISPLAY

SORT BY CENTURY: 1900s

FREQ DISPLAY: RAW #

HITS: 100

Click on a word to see [KEYWORD IN CONTEXT] display [HELP]

	WORD	1900s	1600s	PERC
1	STOCK	85	0	100.00%
2	HAS	81	0	100.00%
3	COMMON	78	0	100.00%
4	NEW	64	0	100.00%
5	BEEN	53	0	100.00%
6	BLACK	31	0	100.00%
7	PRICES	31	0	100.00%
8	MORE	28	0	100.00%
9	FREE	26	0	100.00%
10	SHARE	26	0	100.00%
11	CAN	25	0	100.00%
12	WORLD	23	0	100.00%
13	EUROPEAN	22	0	100.00%

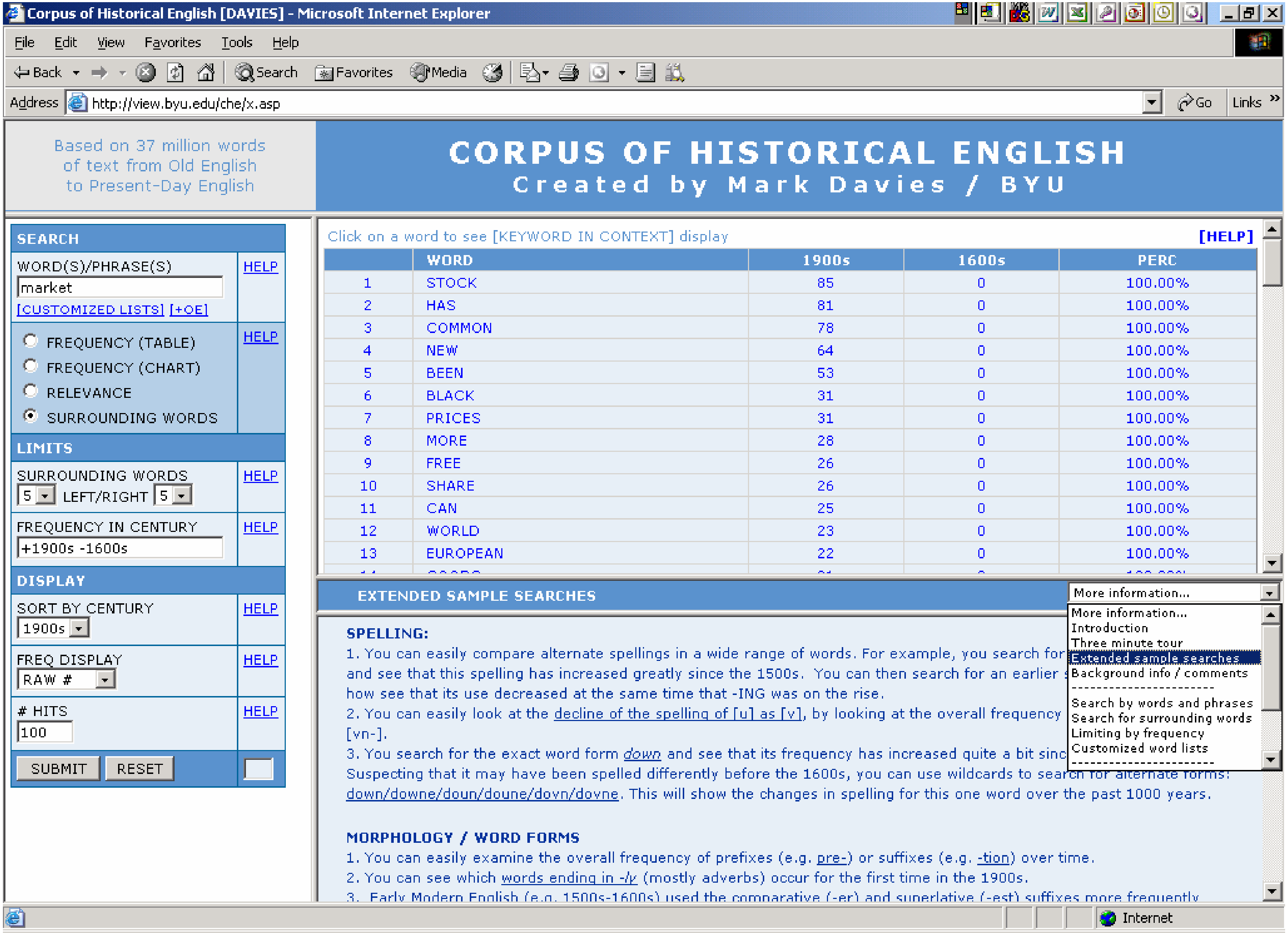
THREE MINUTE TOUR

1700s), or the substring qi??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 within five words of *breaks* (1700s, 1900s) or *face* (1600s, 1900s)
4. **Changes in which words occur most often with others** (sometimes a clue to changes in meaning), e.g.
   -- words after *hard* [+1500s/-1900s, -1500s/+1900s] or before *meat* [+1500s/-1900s, -1500s/+1900s]
   -- words near *market* [+1600s/-1900s, -1600s/+1900s] or *match* [+1600s/-1900s, -1600s/+1900s]

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.

More information….


Corpus of Historical English [DAVIES] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/che/x.asp
Go Links
Internet

Based on 37 million words of text from Old English to Present-Day English

CORPUS OF HISTORICAL ENGLISH
Created by Mark Davies / BYU

SEARCH

WORD(S)/PHRASE(S) HELP
market
[CUSTOMIZED LISTS] [+OE]

○ FREQUENCY (TABLE) HELP
○ FREQUENCY (CHART)
○ RELEVANCE
● SURROUNDING WORDS

LIMITS

SURROUNDING WORDS HELP
5 LEFT/RIGHT 5

FREQUENCY IN CENTURY HELP
+1900s -1600s

DISPLAY

SORT BY CENTURY HELP
1900s

FREQ DISPLAY HELP
RAW #

HITS HELP
100

SUBMIT RESET

Click on a word to see [KEYWORD IN CONTEXT] display [HELP]

	WORD	1900s	1600s	PERC
1	STOCK	85	0	100.00%
2	HAS	81	0	100.00%
3	COMMON	78	0	100.00%
4	NEW	64	0	100.00%
5	BEEN	53	0	100.00%
6	BLACK	31	0	100.00%
7	PRICES	31	0	100.00%
8	MORE	28	0	100.00%
9	FREE	26	0	100.00%
10	SHARE	26	0	100.00%
11	CAN	25	0	100.00%
12	WORLD	23	0	100.00%
13	EUROPEAN	22	0	100.00%

EXTENDED SAMPLE SEARCHES

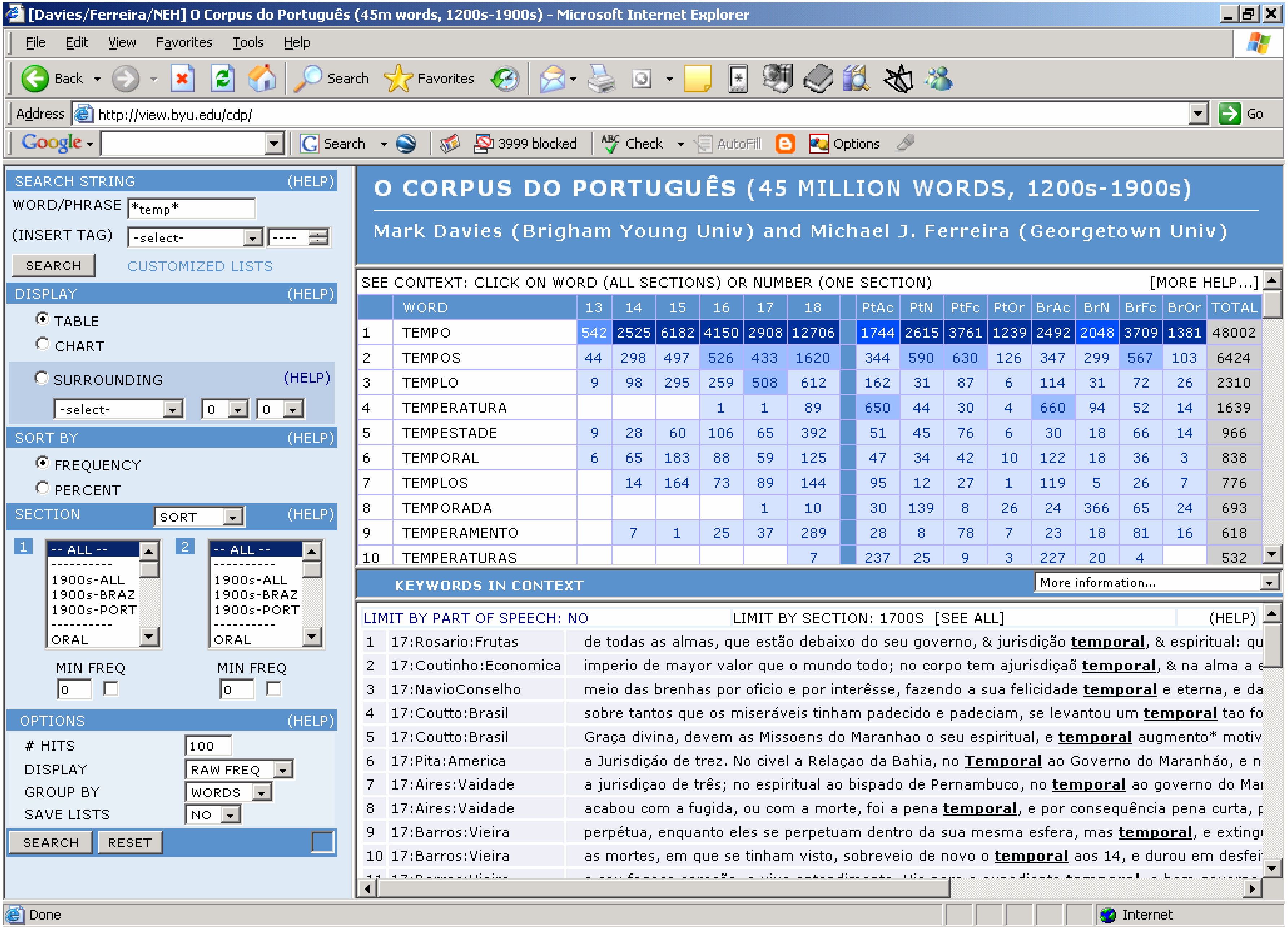
SPELLING:

1. You can easily compare alternate spellings in a wide range of words. For example, you search for and see that this spelling has increased greatly since the 1500s. You can then search for an earlier how see that its use decreased at the same time that -ING was on the rise.
2. You can easily look at the decline of the spelling of [u] as [v], by looking at the overall frequency [vn-].
3. You search for the exact word form *down* and see that its frequency has increased quite a bit sinc Suspecting that it may have been spelled differently before the 1600s, you can use wildcards to search for alternate forms: down/downe/doun/doune/dovn/dovne. This will show the changes in spelling for this one word over the past 1000 years.

MORPHOLOGY / WORD FORMS

1. You can easily examine the overall frequency of prefixes (e.g. pre-) or suffixes (e.g. -tion) over time.
2. You can see which words ending in *-ly* (mostly adverbs) occur for the first time in the 1900s.
3. Early Modern English (e.g. 1500s-1600s) used the comparative (-er) and superlative (-est) suffixes more frequently

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



[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer

Address: http://view.byu.edu/cdp/

SEARCH STRING (HELP)

WORD/PHRASE: *temp*

(INSERT TAG): -select- | ----

SEARCH   CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE
○ CHART

○ SURROUNDING (HELP)
-select- | 0 | 0

SORT BY (HELP)

● FREQUENCY
○ PERCENT

SECTION: SORT (HELP)

1: -- ALL --, ----------, 1900s-ALL, 1900s-BRAZ, 1900s-PORT, ----------, ORAL

2: -- ALL --, ----------, 1900s-ALL, 1900s-BRAZ, 1900s-PORT, ----------, ORAL

MIN FREQ: 0 □   MIN FREQ: 0 □

OPTIONS (HELP)

# HITS	100
DISPLAY	RAW FREQ
GROUP BY	WORDS
SAVE LISTS	NO

SEARCH   RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young Univ) and Michael J. Ferreira (Georgetown Univ)

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION) [MORE HELP...]

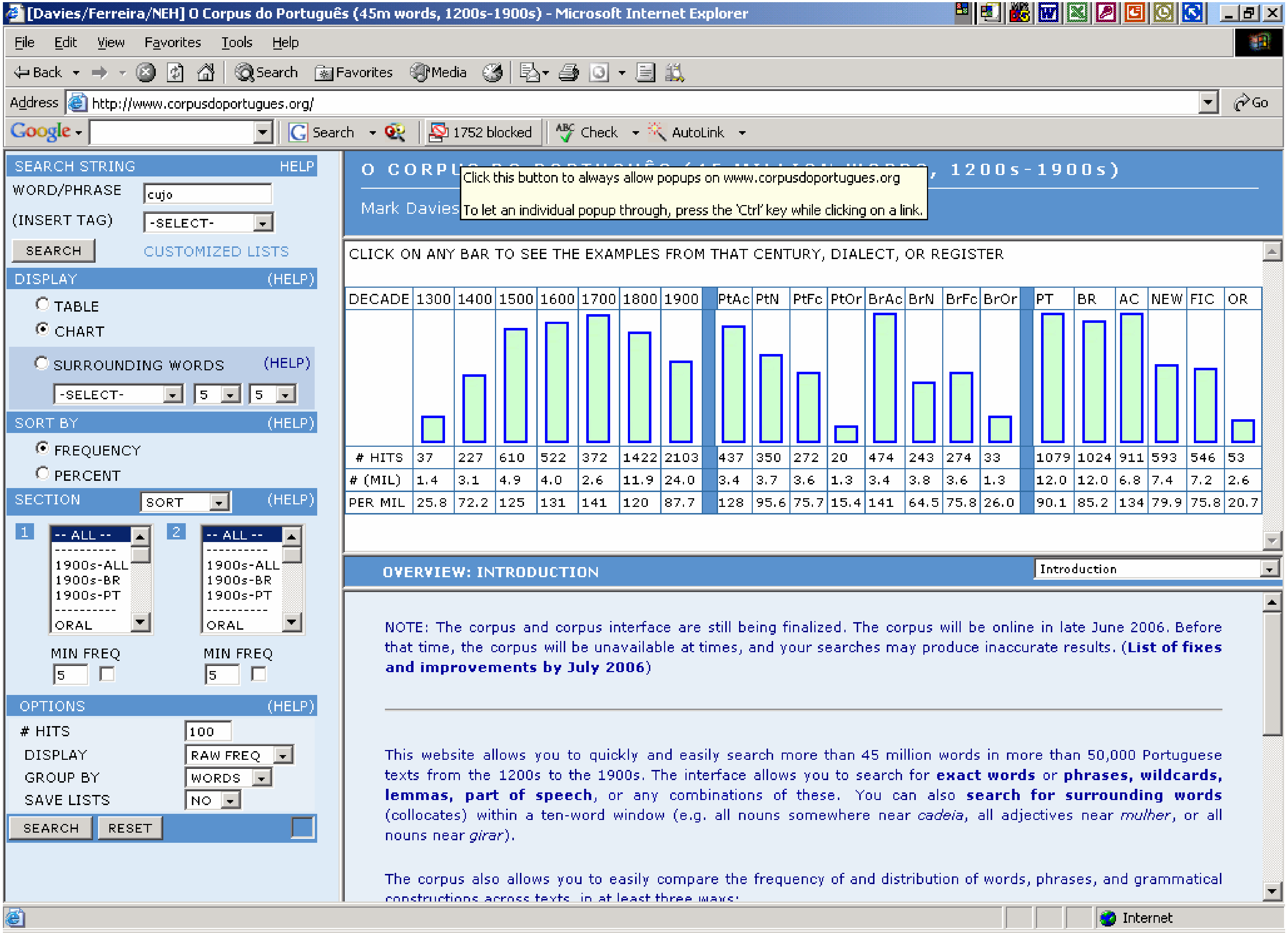
	WORD	13	14	15	16	17	18	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	TOTAL
1	TEMPO	542	2525	6182	4150	2908	12706	1744	2615	3761	1239	2492	2048	3709	1381	48002
2	TEMPOS	44	298	497	526	433	1620	344	590	630	126	347	299	567	103	6424
3	TEMPLO	9	98	295	259	508	612	162	31	87	6	114	31	72	26	2310
4	TEMPERATURA				1	1	89	650	44	30	4	660	94	52	14	1639
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
7	TEMPLOS		14	164	73	89	144	95	12	27	1	119	5	26	7	776
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

KEYWORDS IN CONTEXT   More information...

LIMIT BY PART OF SPEECH: NO   LIMIT BY SECTION: 1700S [SEE ALL]   (HELP)

1	17:Rosario:Frutas	de todas as almas, que estão debaixo do seu governo, & jurisdição **temporal**, & espiritual: qu
2	17:Coutinho:Economica	imperio de mayor valor que o mundo todo; no corpo tem ajurisdiçaõ **temporal**, & na alma a e
3	17:NavioConselho	meio das brenhas por ofício e por interêsse, fazendo a sua felicidade **temporal** e eterna, e da
4	17:Coutto:Brasil	sobre tantos que os miseráveis tinham padecido e padeciam, se levantou um **temporal** tao fo
5	17:Coutto:Brasil	Graça divina, devem as Missoens do Maranhao o seu espiritual, e **temporal** augmento* motiv
6	17:Pita:America	a Jurisdição de trez. No civel a Relaçao da Bahia, no **Temporal** ao Governo do Maranhão, e n
7	17:Aires:Vaidade	a jurisdiçao de três; no espiritual ao bispado de Pernambuco, no **temporal** ao governo do Mar
8	17:Aires:Vaidade	acabou com a fugida, ou com a morte, foi a pena **temporal**, e por consequência pena curta, p
9	17:Barros:Vieira	perpétua, enquanto eles se perpetuam dentro da sua mesma esfera, mas **temporal**, e exting
10	17:Barros:Vieira	as mortes, em que se tinham visto, sobreveio de novo o **temporal** aos 14, e durou em desfei

Frequency tables and chart displays (*cujo* ‘whose’)



WORD/PHRASE: cujo

(INSERT TAG): -SELECT-

○ TABLE

● CHART

○ SURROUNDING WORDS (HELP) -SELECT- 5 5

● FREQUENCY

○ PERCENT

HITS: 100

DISPLAY: RAW FREQ

GROUP BY: WORDS

SAVE LISTS: NO

Click this button to always allow popups on www.corpusdoportugues.org

To let an individual popup through, press the 'Ctrl' key while clicking on a link.

CLICK ON ANY BAR TO SEE THE EXAMPLES FROM THAT CENTURY, DIALECT, OR REGISTER

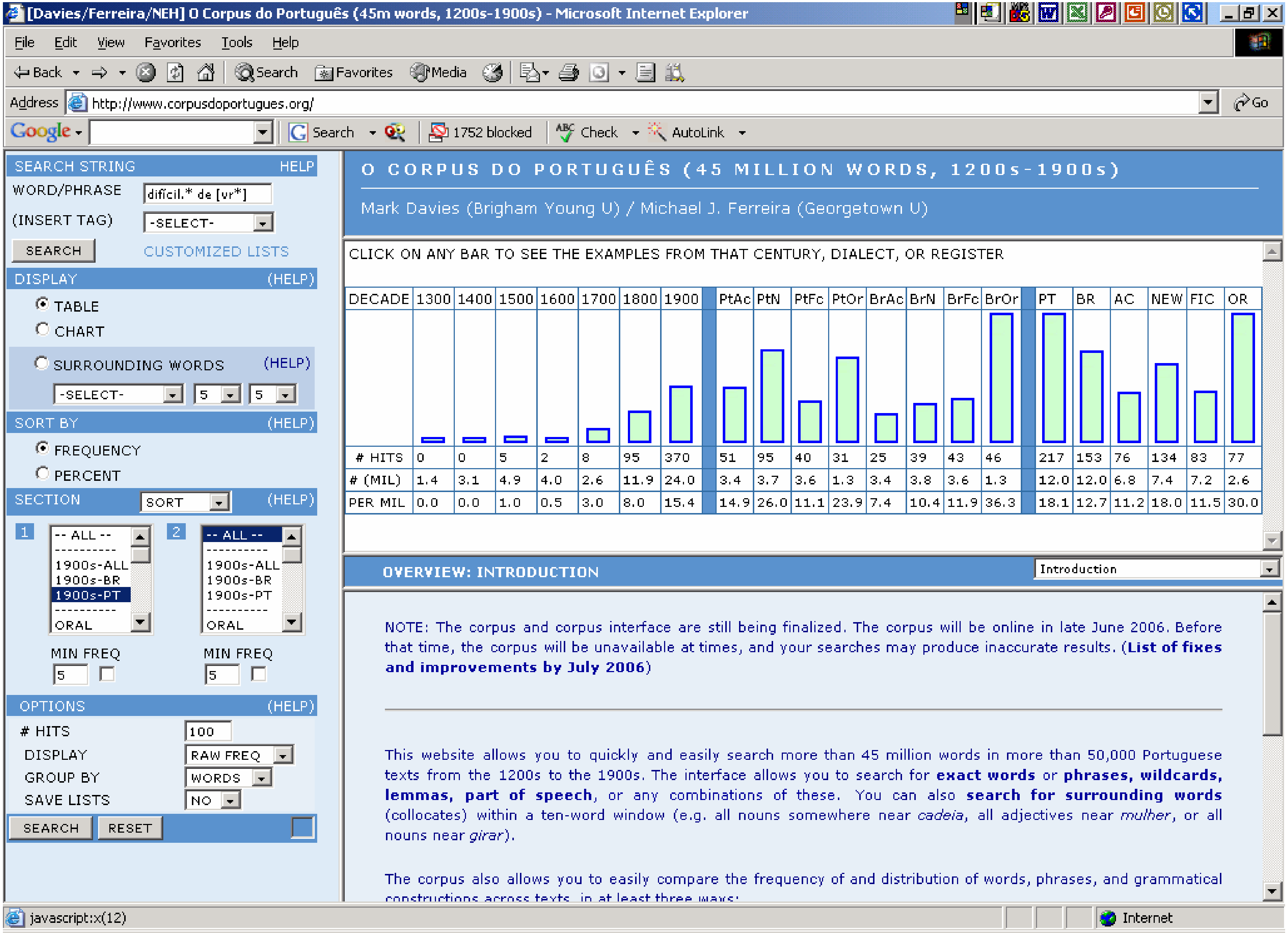
DECADE	1300	1400	1500	1600	1700	1800	1900	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	PT	BR	AC	NEW	FIC	OR
# HITS	37	227	610	522	372	1422	2103	437	350	272	20	474	243	274	33	1079	1024	911	593	546	53
# (MIL)	1.4	3.1	4.9	4.0	2.6	11.9	24.0	3.4	3.7	3.6	1.3	3.4	3.8	3.6	1.3	12.0	12.0	6.8	7.4	7.2	2.6
PER MIL	25.8	72.2	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

NOTE: The corpus and corpus interface are still being finalized. The corpus will be online in late June 2006. Before that time, the corpus will be unavailable at times, and your searches may produce inaccurate results. (**List of fixes and improvements by July 2006**)

This website allows you to quickly and easily search more than 45 million words in more than 50,000 Portuguese texts from the 1200s to the 1900s. The interface allows you to search for **exact words** or **phrases, wildcards, lemmas, part of speech**, or any combinations of these. You can also **search for surrounding words** (collocates) within a ten-word window (e.g. all nouns somewhere near *cadeia*, all adjectives near *mulher*, or all nouns near *girar*).

The corpus also allows you to easily compare the frequency of and distribution of words, phrases, and grammatical constructions across texts, in at least three ways:

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


[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.corpusdoportugues.org/ Go
Google Search 1752 blocked Check AutoLink
SEARCH STRING HELP
SEARCH
CUSTOMIZED LISTS
DISPLAY (HELP)
SORT BY (HELP)
SECTION (HELP)
OPTIONS (HELP)
SEARCH RESET
OVERVIEW: INTRODUCTION
Introduction
javascript:x(12)
Internet

WORD/PHRASE: dificil.* de [vr*]

(INSERT TAG): -SELECT-

DISPLAY: ● TABLE ○ CHART ○ SURROUNDING WORDS (HELP) -SELECT- 5 5

SORT BY: ● FREQUENCY ○ PERCENT

SECTION: SORT

1: -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL — MIN FREQ 5 □

2: -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL — MIN FREQ 5 □

HITS: 100

DISPLAY: RAW FREQ

GROUP BY: WORDS

SAVE LISTS: NO

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

CLICK ON ANY BAR TO SEE THE EXAMPLES FROM THAT CENTURY, DIALECT, OR REGISTER

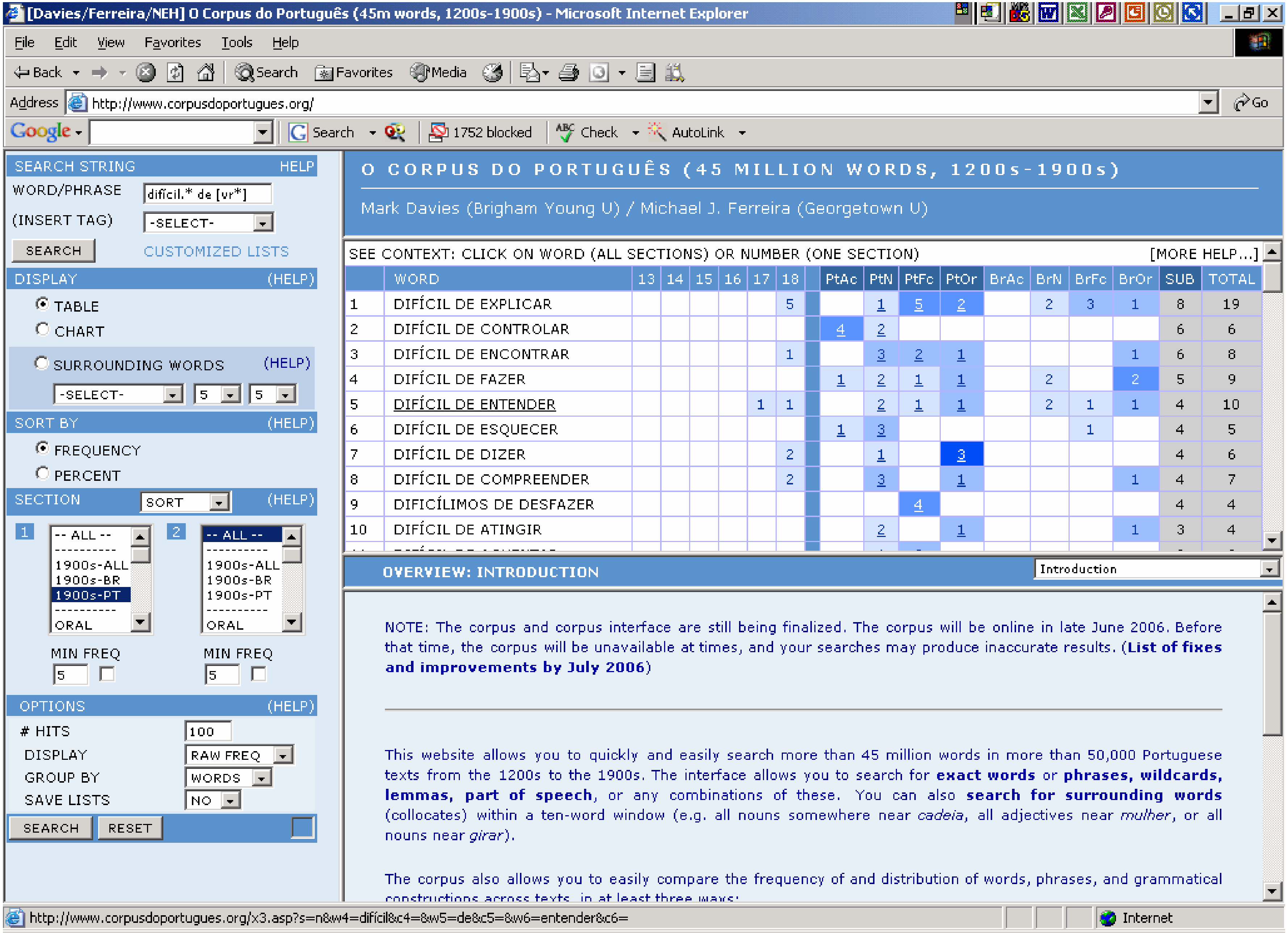
DECADE	1300	1400	1500	1600	1700	1800	1900	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	PT	BR	AC	NEW	FIC	OR
# HITS	0	0	5	2	8	95	370	51	95	40	31	25	39	43	46	217	153	76	134	83	77
# (MIL)	1.4	3.1	4.9	4.0	2.6	11.9	24.0	3.4	3.7	3.6	1.3	3.4	3.8	3.6	1.3	12.0	12.0	6.8	7.4	7.2	2.6
PER MIL	0.0	0.0	1.0	0.5	3.0	8.0	15.4	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

NOTE: The corpus and corpus interface are still being finalized. The corpus will be online in late June 2006. Before that time, the corpus will be unavailable at times, and your searches may produce inaccurate results. (**List of fixes and improvements by July 2006**)

This website allows you to quickly and easily search more than 45 million words in more than 50,000 Portuguese texts from the 1200s to the 1900s. The interface allows you to search for **exact words** or **phrases, wildcards, lemmas, part of speech**, or any combinations of these. You can also **search for surrounding words** (collocates) within a ten-word window (e.g. all nouns somewhere near *cadeia*, all adjectives near *mulher*, or all nouns near *girar*).

The corpus also allows you to easily compare the frequency of and distribution of words, phrases, and grammatical constructions across texts, in at least three ways:

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



[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
http://www.corpusdoportugues.org/
Go
Google
Search
1752 blocked
Check
AutoLink

SEARCH STRING HELP

WORD/PHRASE: dificil.* de [vr*]

(INSERT TAG): -SELECT-

SEARCH · CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE
○ CHART

○ SURROUNDING WORDS (HELP) -SELECT- 5 5

SORT BY (HELP)

● FREQUENCY
○ PERCENT

SECTION SORT (HELP)

1: -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / **1900s-PT** / ---------- / ORAL

2: **-- ALL --** / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL

MIN FREQ 5 □ MIN FREQ 5 □

OPTIONS (HELP)

HITS: 100
DISPLAY: RAW FREQ
GROUP BY: WORDS
SAVE LISTS: NO

SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION) [MORE HELP...]

	WORD	13	14	15	16	17	18	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	SUB	TOTAL
1	DIFÍCIL DE EXPLICAR						5		1	5	2		2	3	1	8	19
2	DIFÍCIL DE CONTROLAR							4	2							6	6
3	DIFÍCIL DE ENCONTRAR						1		3	2	1				1	6	8
4	DIFÍCIL DE FAZER							1	2	1	1		2		2	5	9
5	DIFÍCIL DE ENTENDER					1	1		2	1	1		2	1	1	4	10
6	DIFÍCIL DE ESQUECER							1	3					1		4	5
7	DIFÍCIL DE DIZER						2		1		3					4	6
8	DIFÍCIL DE COMPREENDER						2		3		1				1	4	7
9	DIFICÍLIMOS DE DESFAZER									4						4	4
10	DIFÍCIL DE ATINGIR								2		1				1	3	4

OVERVIEW: INTRODUCTION · Introduction

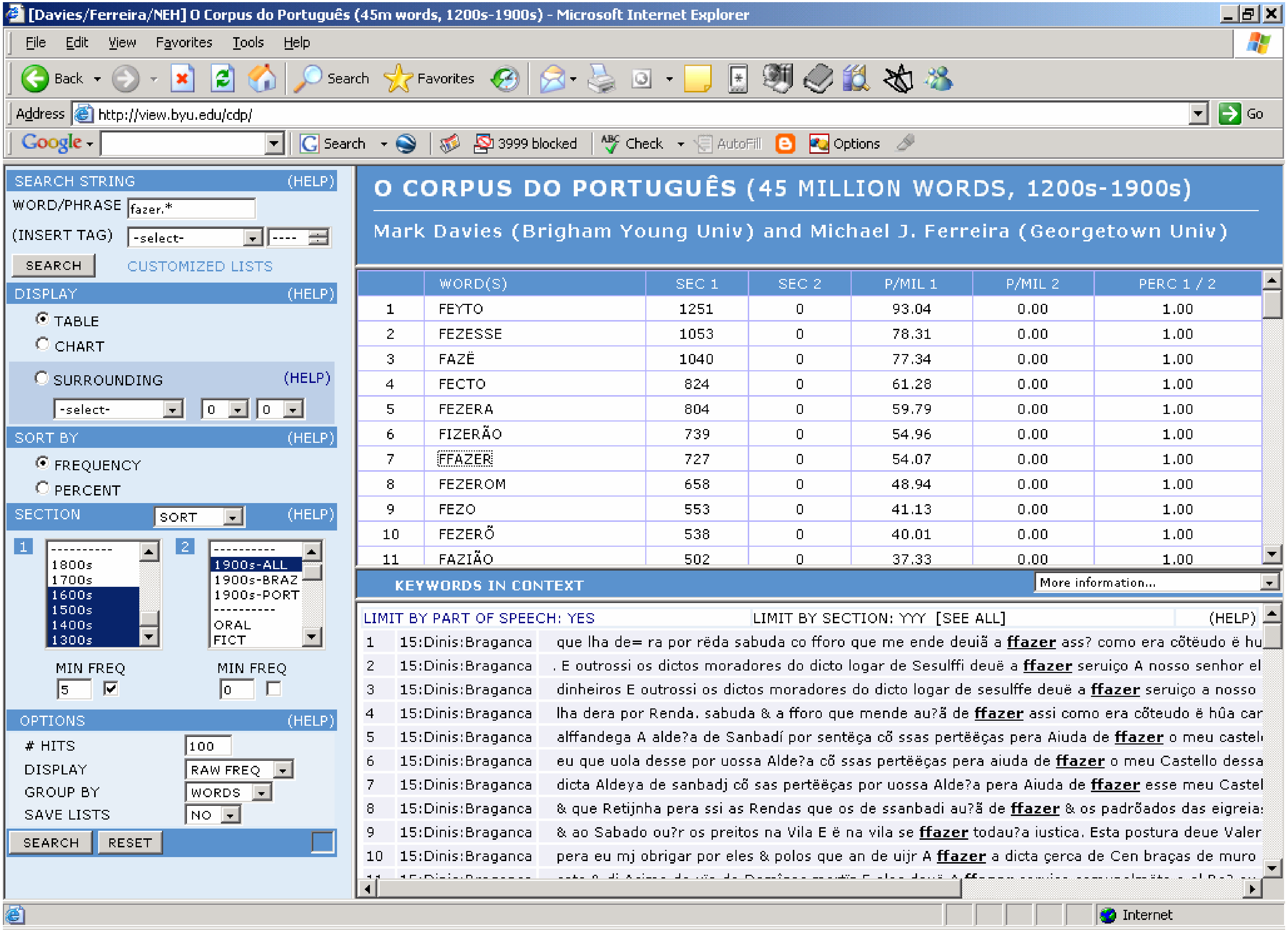
NOTE: The corpus and corpus interface are still being finalized. The corpus will be online in late June 2006. Before that time, the corpus will be unavailable at times, and your searches may produce inaccurate results. (**List of fixes and improvements by July 2006**)

This website allows you to quickly and easily search more than 45 million words in more than 50,000 Portuguese texts from the 1200s to the 1900s. The interface allows you to search for **exact words** or **phrases, wildcards, lemmas, part of speech**, or any combinations of these. You can also **search for surrounding words** (collocates) within a ten-word window (e.g. all nouns somewhere near *cadeia*, all adjectives near *mulher*, or all nouns near *girar*).

The corpus also allows you to easily compare the frequency of and distribution of words, phrases, and grammatical constructions across texts, in at least three ways:

http://www.corpusdoportugues.org/x3.asp?s=n&w4=dificil&c4=&w5=de&c5=&w6=entender&c6=

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


[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://view.byu.edu/cdp/ Go
Google Search 3999 blocked Check AutoFill Options
Internet

SEARCH STRING (HELP)

WORD/PHRASE: fazer.*

(INSERT TAG): -select- ----

SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE
○ CHART
○ SURROUNDING (HELP)
-select- 0 0

SORT BY (HELP)

● FREQUENCY
○ PERCENT

SECTION SORT (HELP)

1: ---------- 1800s 1700s **1600s 1500s 1400s 1300s**

2: ---------- **1900s-ALL** 1900s-BRAZ 1900s-PORT ---------- ORAL FICT

MIN FREQ 5 ☑ MIN FREQ 0 □

OPTIONS (HELP)

# HITS	100
DISPLAY	RAW FREQ
GROUP BY	WORDS
SAVE LISTS	NO

SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young Univ) and Michael J. Ferreira (Georgetown Univ)

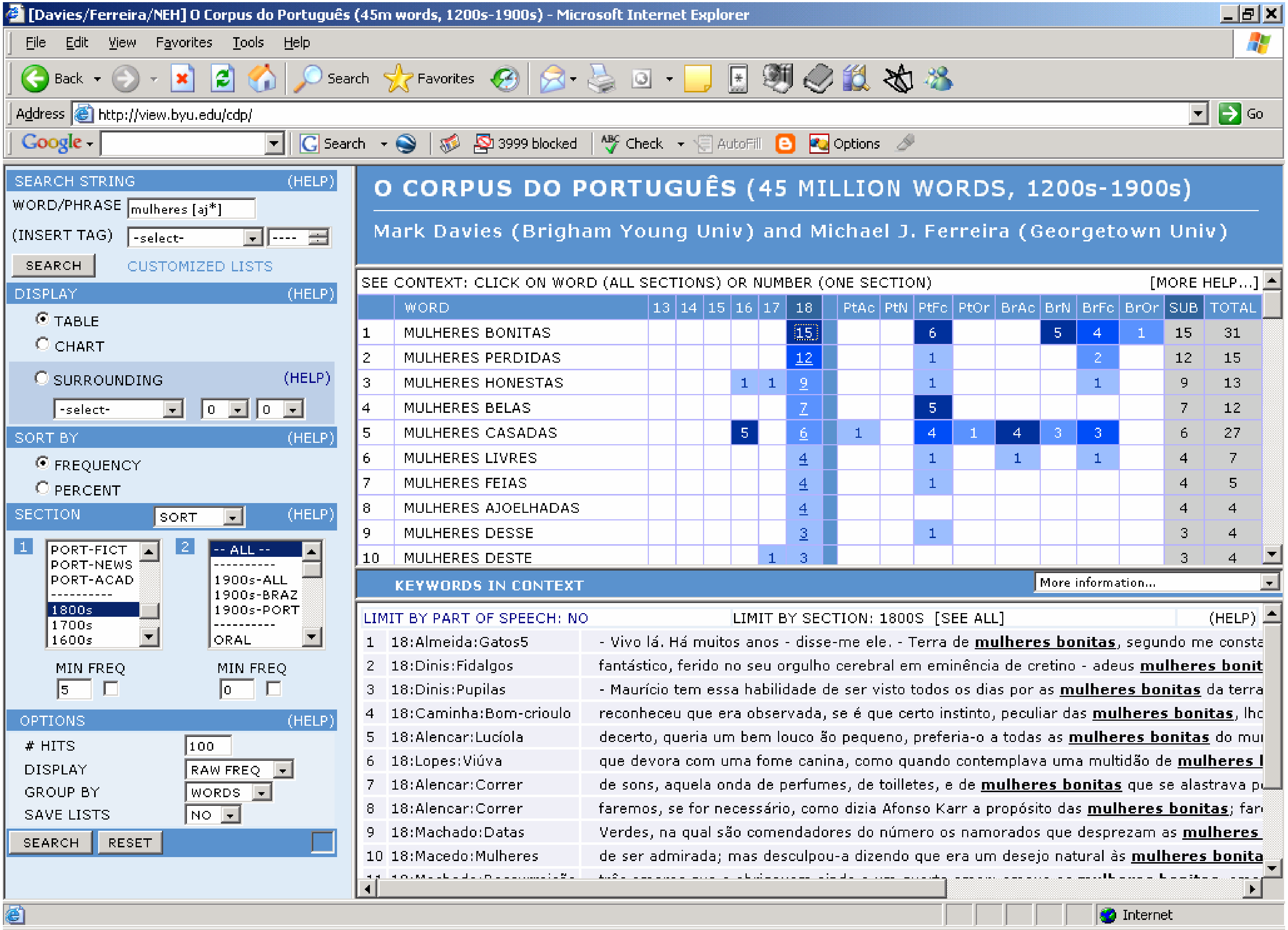
	WORD(S)	SEC 1	SEC 2	P/MIL 1	P/MIL 2	PERC 1 / 2
1	FEYTO	1251	0	93.04	0.00	1.00
2	FEZESSE	1053	0	78.31	0.00	1.00
3	FAZË	1040	0	77.34	0.00	1.00
4	FECTO	824	0	61.28	0.00	1.00
5	FEZERA	804	0	59.79	0.00	1.00
6	FIZERÃO	739	0	54.96	0.00	1.00
7	FFAZER	727	0	54.07	0.00	1.00
8	FEZEROM	658	0	48.94	0.00	1.00
9	FEZO	553	0	41.13	0.00	1.00
10	FEZERÕ	538	0	40.01	0.00	1.00
11	FAZIÃO	502	0	37.33	0.00	1.00

KEYWORDS IN CONTEXT More information...

LIMIT BY PART OF SPEECH: YES LIMIT BY SECTION: YYY [SEE ALL] (HELP)

1	15:Dinis:Braganca	que lha de= ra por rëda sabuda co fforo que me ende deuiã a **ffazer** ass? como era cõtëudo ë hu
2	15:Dinis:Braganca	. E outrossi os dictos moradores do dicto logar de Sesulffi deuë a **ffazer** seruiço A nosso senhor el
3	15:Dinis:Braganca	dinheiros E outrossi os dictos moradores do dicto logar de sesulffe deuë a **ffazer** seruiço a nosso
4	15:Dinis:Braganca	lha dera por Renda. sabuda & a fforo que mende au?ã de **ffazer** assi como era cõteudo ë hũa car
5	15:Dinis:Braganca	alffandega A alde?a de Sanbadí por sentëça cõ ssas pertëëças pera Aiuda de **ffazer** o meu castel
6	15:Dinis:Braganca	eu que uola desse por uossa Alde?a cõ ssas pertëëças pera aiuda de **ffazer** o meu Castello dessa
7	15:Dinis:Braganca	dicta Aldeya de sanbadj cõ sas pertëëças por uossa Alde?a pera Aiuda de **ffazer** esse meu Castel
8	15:Dinis:Braganca	& que Retijnha pera ssi as Rendas que os de ssanbadi au?ã de **ffazer** & os padrõados das eigreia
9	15:Dinis:Braganca	& ao Sabado ou?r os preitos na Vila E ë na vila se **ffazer** todau?a iustica. Esta postura deue Valer
10	15:Dinis:Braganca	pera eu mj obrigar por eles & polos que an de uijr A **ffazer** a dicta çerca de Cen braças de muro

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



SEARCH STRING (HELP)
WORD/PHRASE mulheres [aj*]
(INSERT TAG) -select- ----
SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)
● TABLE
○ CHART
○ SURROUNDING (HELP)
-select- 0 0

SORT BY (HELP)
● FREQUENCY
○ PERCENT

SECTION SORT (HELP)

1 PORT-FICT, PORT-NEWS, PORT-ACAD, ----------, 1800s, 1700s, 1600s
2 -- ALL --, ----------, 1900s-ALL, 1900s-BRAZ, 1900s-PORT, ----------, ORAL

MIN FREQ 5 □
MIN FREQ 0 □

OPTIONS (HELP)
HITS 100
DISPLAY RAW FREQ
GROUP BY WORDS
SAVE LISTS NO
SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young Univ) and Michael J. Ferreira (Georgetown Univ)

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION) [MORE HELP...]

	WORD	13	14	15	16	17	18		PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	SUB	TOTAL
1	MULHERES BONITAS						15				6			5	4	1	15	31
2	MULHERES PERDIDAS						12				1				2		12	15
3	MULHERES HONESTAS				1	1	9				1				1		9	13
4	MULHERES BELAS						7				5						7	12
5	MULHERES CASADAS				5		6		1		4	1	4	3	3		6	27
6	MULHERES LIVRES						4				1		1		1		4	7
7	MULHERES FEIAS						4				1						4	5
8	MULHERES AJOELHADAS						4										4	4
9	MULHERES DESSE						3				1						3	4
10	MULHERES DESTE					1	3										3	4

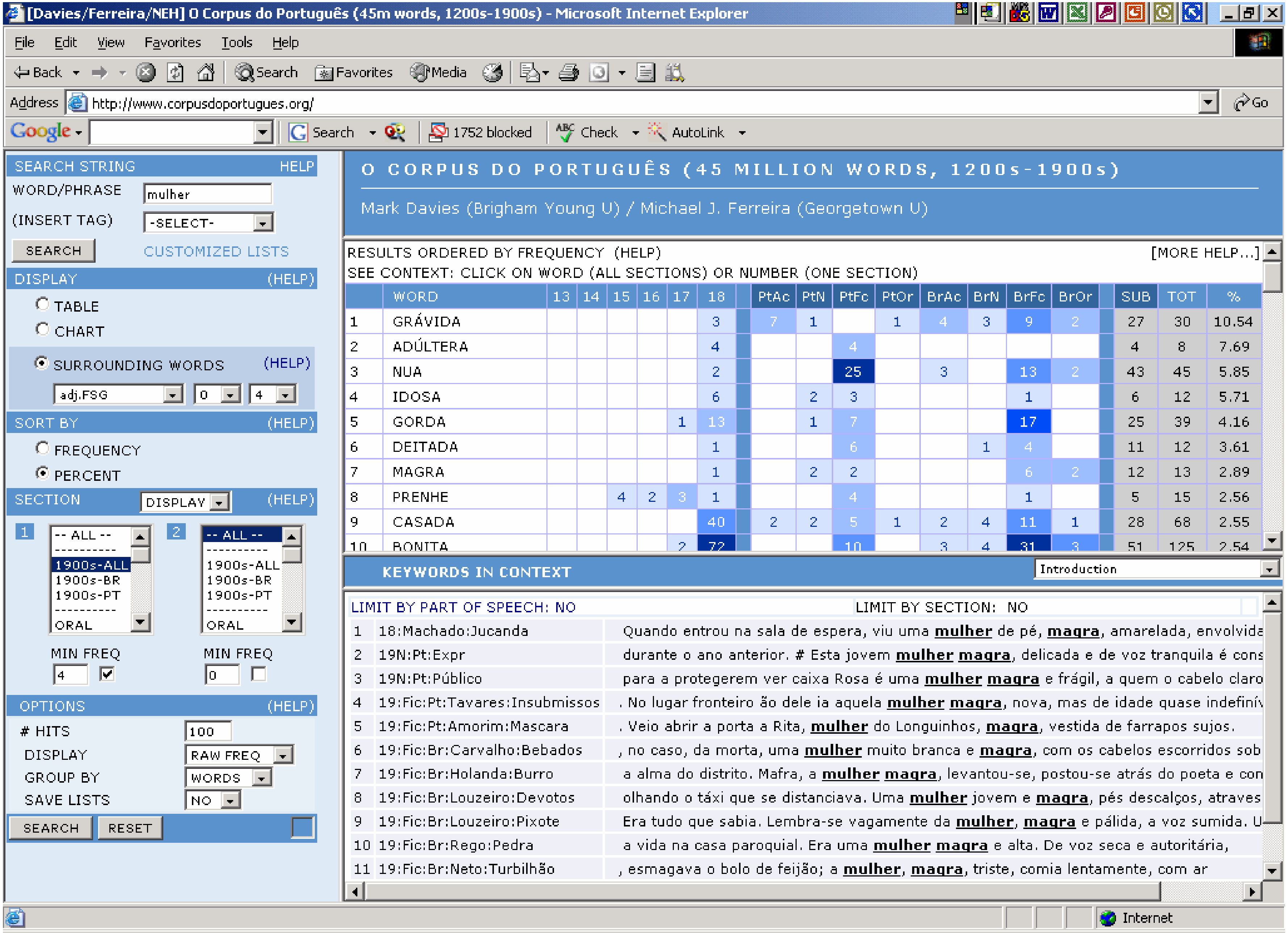
KEYWORDS IN CONTEXT

More information...

LIMIT BY PART OF SPEECH: NO LIMIT BY SECTION: 1800S [SEE ALL] (HELP)

1	18:Almeida:Gatos5	- Vivo lá. Há muitos anos - disse-me ele. - Terra de **mulheres bonitas**, segundo me consta
2	18:Dinis:Fidalgos	fantástico, ferido no seu orgulho cerebral em eminência de cretino - adeus **mulheres bonit**
3	18:Dinis:Pupilas	- Maurício tem essa habilidade de ser visto todos os dias por as **mulheres bonitas** da terra
4	18:Caminha:Bom-crioulo	reconheceu que era observada, se é que certo instinto, peculiar das **mulheres bonitas**, lhe
5	18:Alencar:Lucíola	decerto, queria um bem louco ão pequeno, preferia-o a todas as **mulheres bonitas** do mu
6	18:Lopes:Viúva	que devora com uma fome canina, como quando contemplava uma multidão de **mulheres**
7	18:Alencar:Correr	de sons, aquela onda de perfumes, de toilletes, e de **mulheres bonitas** que se alastrava p
8	18:Alencar:Correr	faremos, se for necessário, como dizia Afonso Karr a propósito das **mulheres bonitas**; far
9	18:Machado:Datas	Verdes, na qual são comendadores do número os namorados que desprezam as **mulheres**
10	18:Macedo:Mulheres	de ser admirada; mas desculpou-a dizendo que era um desejo natural às **mulheres bonita**

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


[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.corpusdoportugues.org/
Go
Google Search 1752 blocked Check AutoLink
Internet

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

SEARCH STRING HELP

WORD/PHRASE: mulher

(INSERT TAG): -SELECT-

SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)

○ TABLE
○ CHART
● SURROUNDING WORDS (HELP)

adj.FSG 0 4

SORT BY (HELP)

○ FREQUENCY
● PERCENT

SECTION DISPLAY (HELP)

1: -- ALL -- / ---------- / **1900s-ALL** / 1900s-BR / 1900s-PT / ---------- / ORAL

2: **-- ALL --** / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL

MIN FREQ 4 ☑ MIN FREQ 0 □

OPTIONS (HELP)

HITS 100
DISPLAY RAW FREQ
GROUP BY WORDS
SAVE LISTS NO

SEARCH RESET

RESULTS ORDERED BY FREQUENCY (HELP) [MORE HELP...]
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION)

	WORD	13	14	15	16	17	18	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	SUB	TOT	%
1	GRÁVIDA						3	7	1		1	4	3	9	2	27	30	10.54
2	ADÚLTERA						4			4						4	8	7.69
3	NUA						2			25		3		13	2	43	45	5.85
4	IDOSA						6		2	3				1		6	12	5.71
5	GORDA					1	13		1	7				17		25	39	4.16
6	DEITADA						1			6			1	4		11	12	3.61
7	MAGRA						1		2	2				6	2	12	13	2.89
8	PRENHE			4	2	3	1			4				1		5	15	2.56
9	CASADA						40	2	2	5	1	2	4	11	1	28	68	2.55
10	BONITA					2	72			10		3	4	31	3	51	125	2.54

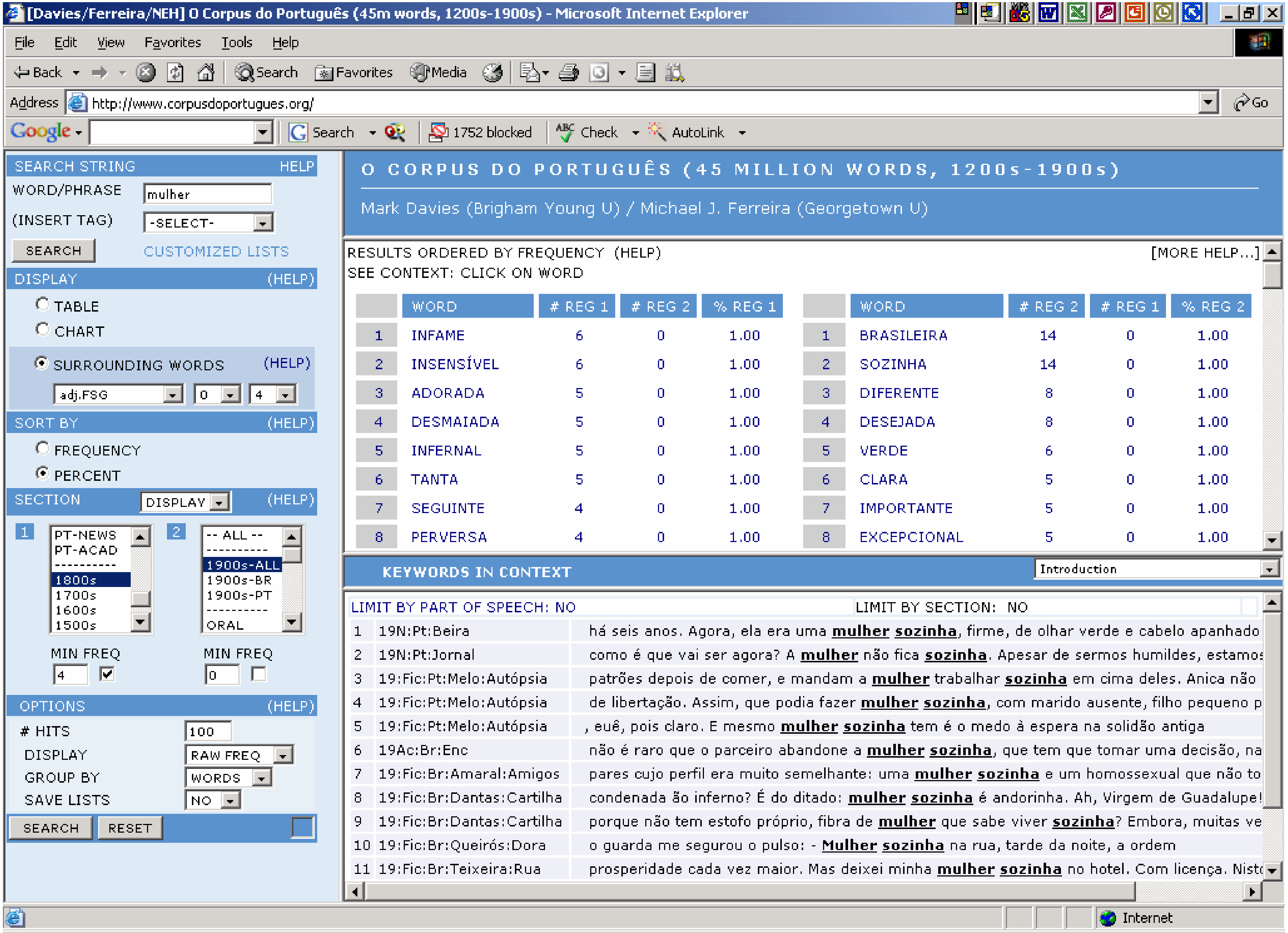
KEYWORDS IN CONTEXT

Introduction

LIMIT BY PART OF SPEECH: NO   LIMIT BY SECTION: NO

1	18:Machado:Jucanda	Quando entrou na sala de espera, viu uma **mulher** de pé, **magra**, amarelada, envolvida
2	19N:Pt:Expr	durante o ano anterior. # Esta jovem **mulher** **magra**, delicada e de voz tranquila é cons
3	19N:Pt:Público	para a protegerem ver caixa Rosa é uma **mulher** **magra** e frágil, a quem o cabelo claro
4	19:Fic:Pt:Tavares:Insubmissos	. No lugar fronteiro ão dele ia aquela **mulher** **magra**, nova, mas de idade quase indefiní
5	19:Fic:Pt:Amorim:Mascara	. Veio abrir a porta a Rita, **mulher** do Longuinhos, **magra**, vestida de farrapos sujos.
6	19:Fic:Br:Carvalho:Bebados	, no caso, da morta, uma **mulher** muito branca e **magra**, com os cabelos escorridos sob
7	19:Fic:Br:Holanda:Burro	a alma do distrito. Mafra, a **mulher** **magra**, levantou-se, postou-se atrás do poeta e con
8	19:Fic:Br:Louzeiro:Devotos	olhando o táxi que se distanciava. Uma **mulher** jovem e **magra**, pés descalços, atraves
9	19:Fic:Br:Louzeiro:Pixote	Era tudo que sabia. Lembra-se vagamente da **mulher**, **magra** e pálida, a voz sumida. U
10	19:Fic:Br:Rego:Pedra	a vida na casa paroquial. Era uma **mulher** **magra** e alta. De voz seca e autoritária,
11	19:Fic:Br:Neto:Turbilhão	, esmagava o bolo de feijão; a **mulher**, **magra**, triste, comia lentamente, com ar

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


[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
http://www.corpusdoportugues.org/
Go
Google
Search
1752 blocked
Check
AutoLink
Internet

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

SEARCH STRING — HELP

WORD/PHRASE: mulher

(INSERT TAG): -SELECT-

SEARCH — CUSTOMIZED LISTS

DISPLAY (HELP)

○ TABLE
○ CHART
● SURROUNDING WORDS (HELP)

adj.FSG — 0 — 4

SORT BY (HELP)

○ FREQUENCY
● PERCENT

SECTION: DISPLAY (HELP)

1: PT-NEWS, PT-ACAD, ----------, **1800s**, 1700s, 1600s, 1500s

2: -- ALL --, ----------, **1900s-ALL**, 1900s-BR, 1900s-PT, ----------, ORAL

MIN FREQ: 4 ☑ — MIN FREQ: 0 □

OPTIONS (HELP)

HITS: 100

DISPLAY: RAW FREQ

GROUP BY: WORDS

SAVE LISTS: NO

SEARCH — RESET

RESULTS ORDERED BY FREQUENCY (HELP) [MORE HELP...]

SEE CONTEXT: CLICK ON WORD

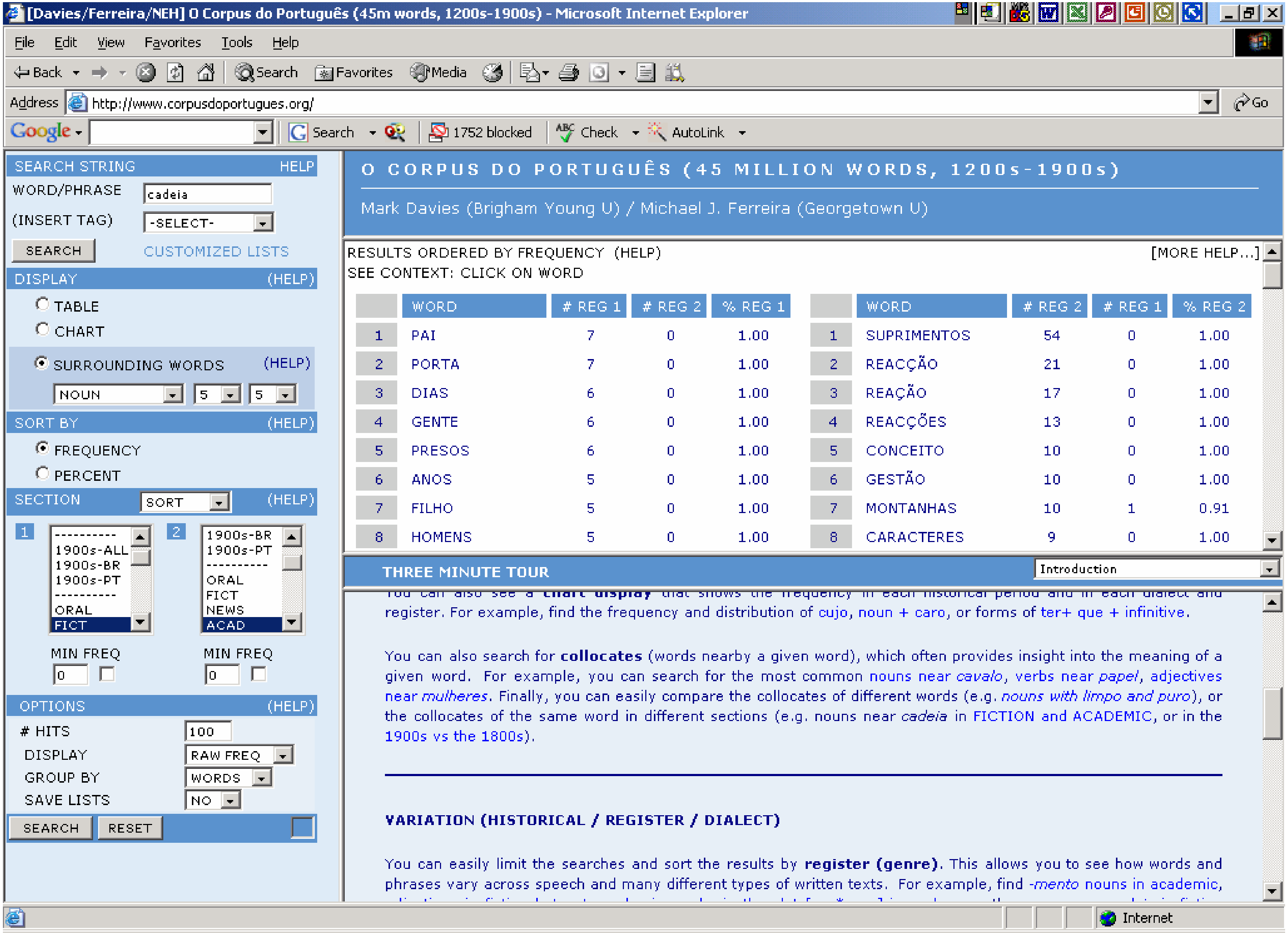
	WORD	# REG 1	# REG 2	% REG 1		WORD	# REG 2	# REG 1	% REG 2
1	INFAME	6	0	1.00	1	BRASILEIRA	14	0	1.00
2	INSENSÍVEL	6	0	1.00	2	SOZINHA	14	0	1.00
3	ADORADA	5	0	1.00	3	DIFERENTE	8	0	1.00
4	DESMAIADA	5	0	1.00	4	DESEJADA	8	0	1.00
5	INFERNAL	5	0	1.00	5	VERDE	6	0	1.00
6	TANTA	5	0	1.00	6	CLARA	5	0	1.00
7	SEGUINTE	4	0	1.00	7	IMPORTANTE	5	0	1.00
8	PERVERSA	4	0	1.00	8	EXCEPCIONAL	5	0	1.00

KEYWORDS IN CONTEXT — Introduction

LIMIT BY PART OF SPEECH: NO — LIMIT BY SECTION: NO

1	19N:Pt:Beira	há seis anos. Agora, ela era uma **mulher** **sozinha**, firme, de olhar verde e cabelo apanhado
2	19N:Pt:Jornal	como é que vai ser agora? A **mulher** não fica **sozinha**. Apesar de sermos humildes, estamos
3	19:Fic:Pt:Melo:Autópsia	patrões depois de comer, e mandam a **mulher** trabalhar **sozinha** em cima deles. Anica não
4	19:Fic:Pt:Melo:Autópsia	de libertação. Assim, que podia fazer **mulher** **sozinha**, com marido ausente, filho pequeno p
5	19:Fic:Pt:Melo:Autópsia	, euê, pois claro. E mesmo **mulher** **sozinha** tem é o medo à espera na solidão antiga
6	19Ac:Br:Enc	não é raro que o parceiro abandone a **mulher** **sozinha**, que tem que tomar uma decisão, na
7	19:Fic:Br:Amaral:Amigos	pares cujo perfil era muito semelhante: uma **mulher** **sozinha** e um homossexual que não to
8	19:Fic:Br:Dantas:Cartilha	condenada ão inferno? É do ditado: **mulher** **sozinha** é andorinha. Ah, Virgem de Guadalupe!
9	19:Fic:Br:Dantas:Cartilha	porque não tem estofo próprio, fibra de **mulher** que sabe viver **sozinha**? Embora, muitas ve
10	19:Fic:Br:Queirós:Dora	o guarda me segurou o pulso: - **Mulher** **sozinha** na rua, tarde da noite, a ordem
11	19:Fic:Br:Teixeira:Rua	prosperidade cada vez maior. Mas deixei minha **mulher** **sozinha** no hotel. Com licença. Nistc

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



[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.corpusdoportugues.org/
Go
Google Search 1752 blocked Check AutoLink
Internet

SEARCH STRING HELP

WORD/PHRASE: cadeia

(INSERT TAG): -SELECT-

SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)

○ TABLE

○ CHART

● SURROUNDING WORDS (HELP)

NOUN 5 5

SORT BY (HELP)

● FREQUENCY

○ PERCENT

SECTION SORT (HELP)

1: ----------, 1900s-ALL, 1900s-BR, 1900s-PT, ----------, ORAL, **FICT** (selected)

2: 1900s-BR, 1900s-PT, ----------, ORAL, FICT, NEWS, **ACAD** (selected)

MIN FREQ 0 □ MIN FREQ 0 □

OPTIONS (HELP)

# HITS	100
DISPLAY	RAW FREQ
GROUP BY	WORDS
SAVE LISTS	NO

SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

RESULTS ORDERED BY FREQUENCY (HELP) [MORE HELP...]

SEE CONTEXT: CLICK ON WORD

	WORD	# REG 1	# REG 2	% REG 1		WORD	# REG 2	# REG 1	% REG 2
1	PAI	7	0	1.00	1	SUPRIMENTOS	54	0	1.00
2	PORTA	7	0	1.00	2	REACÇÃO	21	0	1.00
3	DIAS	6	0	1.00	3	REAÇÃO	17	0	1.00
4	GENTE	6	0	1.00	4	REACÇÕES	13	0	1.00
5	PRESOS	6	0	1.00	5	CONCEITO	10	0	1.00
6	ANOS	5	0	1.00	6	GESTÃO	10	0	1.00
7	FILHO	5	0	1.00	7	MONTANHAS	10	1	0.91
8	HOMENS	5	0	1.00	8	CARACTERES	9	0	1.00

THREE MINUTE TOUR Introduction

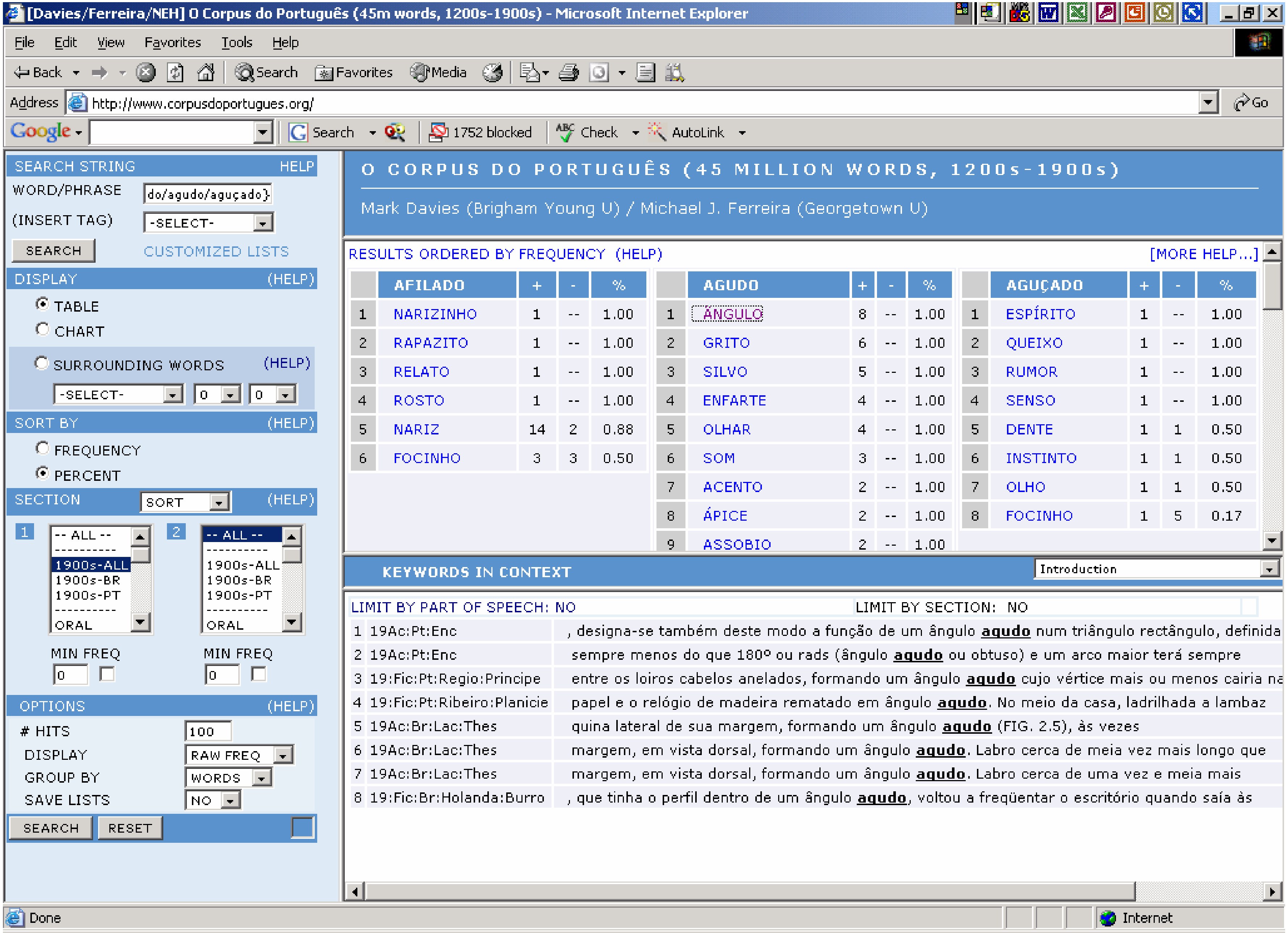
You can also see a **chart display** that shows the frequency in each historical period and in each dialect and register. For example, find the frequency and distribution of cujo, noun + caro, or forms of ter+ que + infinitive.

You can also search for **collocates** (words nearby a given word), which often provides insight into the meaning of a given word. For example, you can search for the most common nouns near *cavalo*, verbs near *papel*, adjectives near *mulheres*. Finally, you can easily compare the collocates of different words (e.g. *nouns with limpo and puro*), or the collocates of the same word in different sections (e.g. nouns near *cadeia* in FICTION and ACADEMIC, or in the 1900s vs the 1800s).

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 Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
http://www.corpusdoportugues.org/
Go
Google
Search
1752 blocked
Check
AutoLink
Done
Internet

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

SEARCH STRING — HELP

WORD/PHRASE: do/agudo/aguçado}

(INSERT TAG): -SELECT-

SEARCH — CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE

○ CHART

○ SURROUNDING WORDS (HELP)

-SELECT- 0 0

SORT BY (HELP)

○ FREQUENCY

● PERCENT

SECTION: SORT (HELP)

1: -- ALL --, ----------, 1900s-ALL, 1900s-BR, 1900s-PT, ----------, ORAL

2: -- ALL --, ----------, 1900s-ALL, 1900s-BR, 1900s-PT, ----------, ORAL

MIN FREQ: 0 □ MIN FREQ: 0 □

OPTIONS (HELP)

HITS: 100

DISPLAY: RAW FREQ

GROUP BY: WORDS

SAVE LISTS: NO

SEARCH RESET

RESULTS ORDERED BY FREQUENCY (HELP) [MORE HELP...]

	AFILADO	+	-	%
1	NARIZINHO	1	--	1.00
2	RAPAZITO	1	--	1.00
3	RELATO	1	--	1.00
4	ROSTO	1	--	1.00
5	NARIZ	14	2	0.88
6	FOCINHO	3	3	0.50

	AGUDO	+	-	%
1	ÂNGULO	8	--	1.00
2	GRITO	6	--	1.00
3	SILVO	5	--	1.00
4	ENFARTE	4	--	1.00
5	OLHAR	4	--	1.00
6	SOM	3	--	1.00
7	ACENTO	2	--	1.00
8	ÁPICE	2	--	1.00
9	ASSOBIO	2	--	1.00

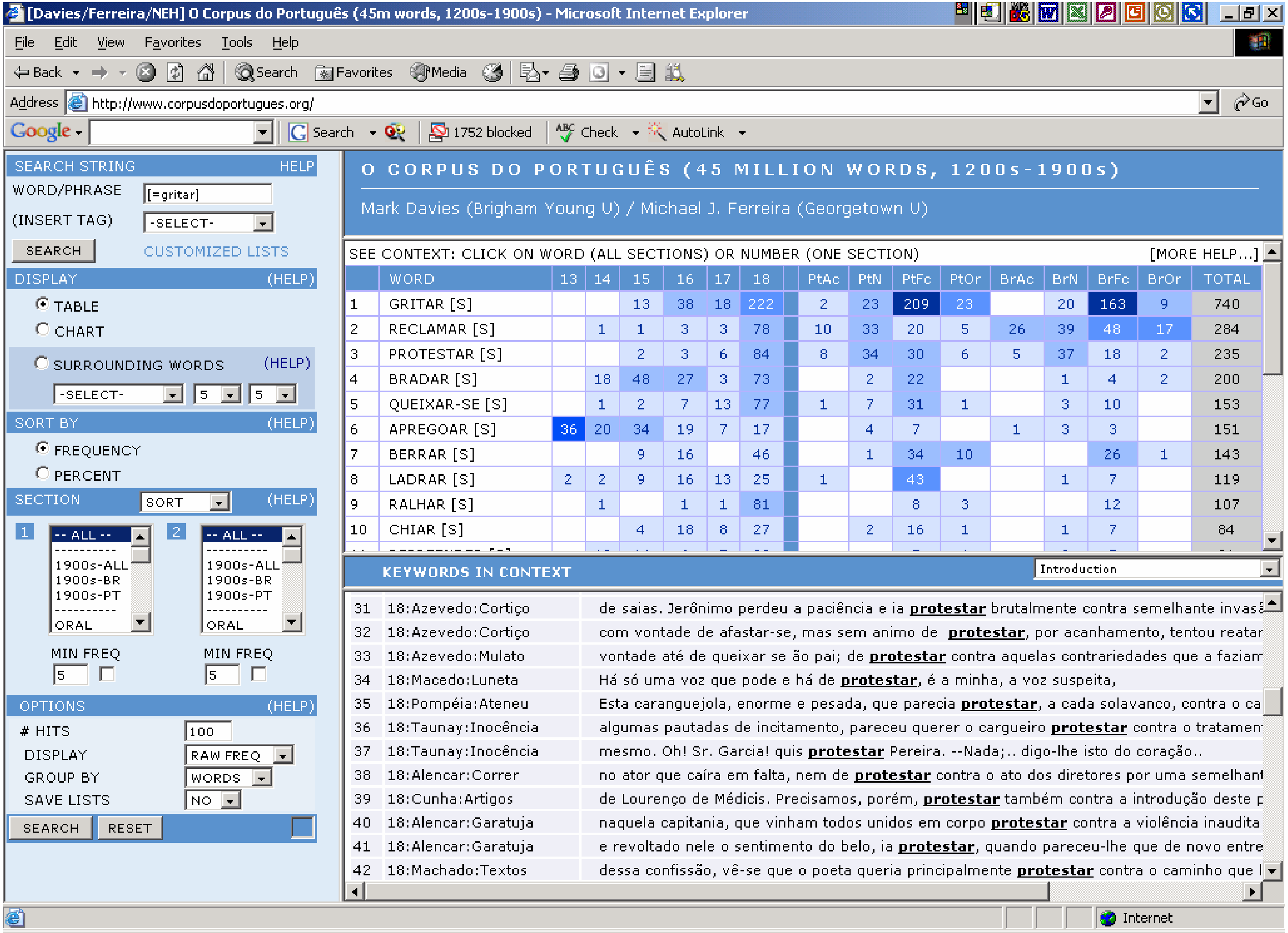
	AGUÇADO	+	-	%
1	ESPÍRITO	1	--	1.00
2	QUEIXO	1	--	1.00
3	RUMOR	1	--	1.00
4	SENSO	1	--	1.00
5	DENTE	1	1	0.50
6	INSTINTO	1	1	0.50
7	OLHO	1	1	0.50
8	FOCINHO	1	5	0.17

KEYWORDS IN CONTEXT — Introduction

LIMIT BY PART OF SPEECH: NO LIMIT BY SECTION: NO

1	19Ac:Pt:Enc	, designa-se também deste modo a função de um ângulo **agudo** num triângulo rectângulo, definida
2	19Ac:Pt:Enc	sempre menos do que 180º ou rads (ângulo **agudo** ou obtuso) e um arco maior terá sempre
3	19:Fic:Pt:Regio:Principe	entre os loiros cabelos anelados, formando um ângulo **agudo** cujo vértice mais ou menos cairia na
4	19:Fic:Pt:Ribeiro:Planicie	papel e o relógio de madeira rematado em ângulo **agudo**. No meio da casa, ladrilhada a lambaz
5	19Ac:Br:Lac:Thes	quina lateral de sua margem, formando um ângulo **agudo** (FIG. 2.5), às vezes
6	19Ac:Br:Lac:Thes	margem, em vista dorsal, formando um ângulo **agudo**. Labro cerca de meia vez mais longo que
7	19Ac:Br:Lac:Thes	margem, em vista dorsal, formando um ângulo **agudo**. Labro cerca de uma vez e meia mais
8	19:Fic:Br:Holanda:Burro	, que tinha o perfil dentro de um ângulo **agudo**, voltou a freqüentar o escritório quando saía às

Synonyms: **[=gritar]**: synonyms of ‘to shout’


[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
http://www.corpusdoportugues.org/
Go
Google
Search
1752 blocked
Check
AutoLink
Internet

SEARCH STRING HELP

WORD/PHRASE [=gritar]

(INSERT TAG) -SELECT-

SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE

○ CHART

○ SURROUNDING WORDS (HELP)

-SELECT- 5 5

SORT BY (HELP)

● FREQUENCY

○ PERCENT

SECTION SORT (HELP)

1 -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL

2 -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL

MIN FREQ 5 □ MIN FREQ 5 □

OPTIONS (HELP)

# HITS 100

DISPLAY RAW FREQ

GROUP BY WORDS

SAVE LISTS NO

SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

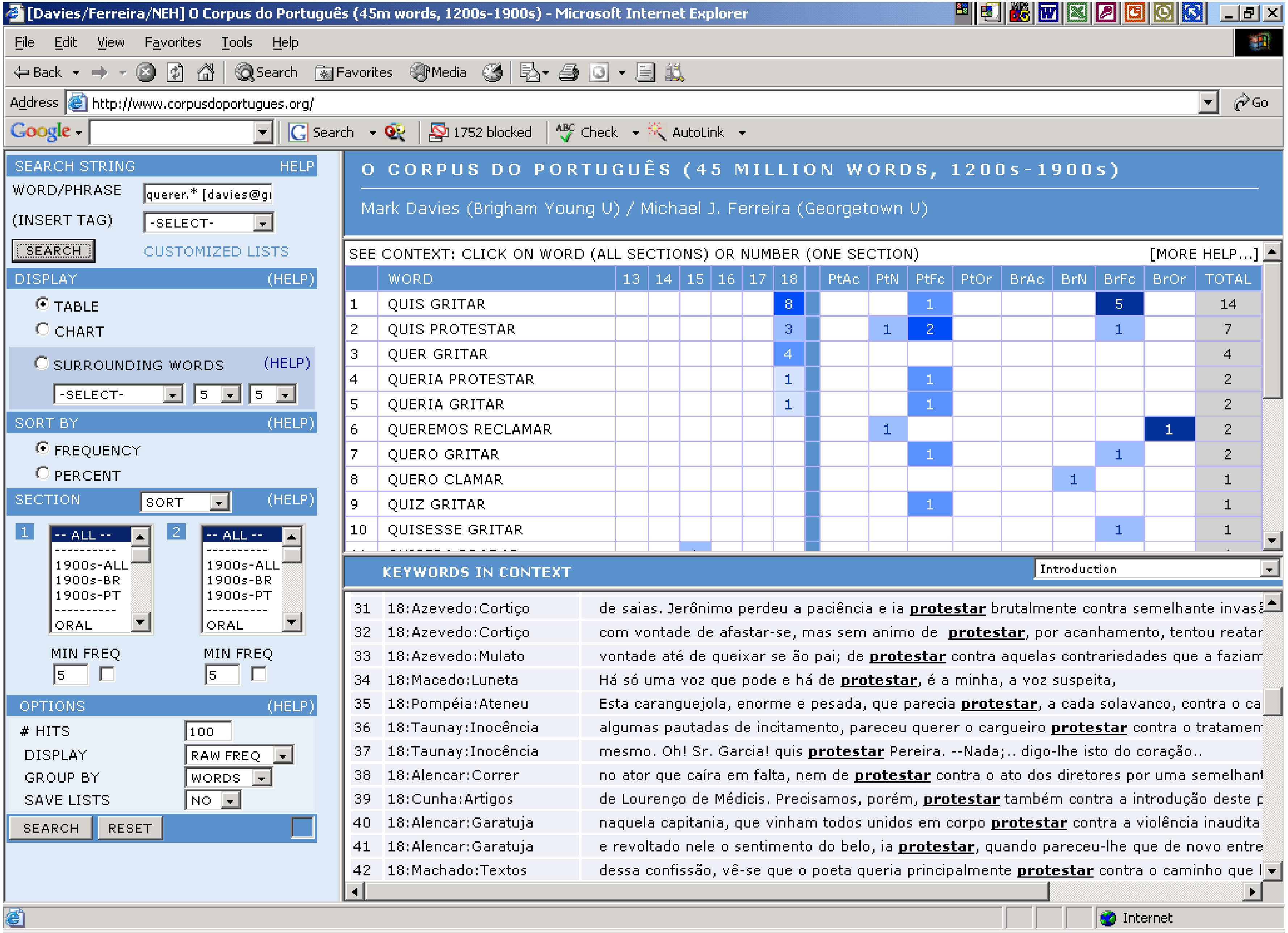
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION) [MORE HELP...]

	WORD	13	14	15	16	17	18		PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	TOTAL
1	GRITAR [S]			13	38	18	222		2	23	209	23		20	163	9	740
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
4	BRADAR [S]		18	48	27	3	73			2	22			1	4	2	200
5	QUEIXAR-SE [S]		1	2	7	13	77		1	7	31	1		3	10		153
6	APREGOAR [S]	36	20	34	19	7	17			4	7		1	3	3		151
7	BERRAR [S]			9	16		46			1	34	10			26	1	143
8	LADRAR [S]	2	2	9	16	13	25		1		43			1	7		119
9	RALHAR [S]		1		1	1	81				8	3			12		107
10	CHIAR [S]			4	18	8	27			2	16	1		1	7		84

KEYWORDS IN CONTEXT Introduction

31	18:Azevedo:Cortiço	de saias. Jerônimo perdeu a paciência e ia **protestar** brutalmente contra semelhante invasã
32	18:Azevedo:Cortiço	com vontade de afastar-se, mas sem animo de **protestar**, por acanhamento, tentou reatar
33	18:Azevedo:Mulato	vontade até de queixar se ão pai; de **protestar** contra aquelas contrariedades que a faziam
34	18:Macedo:Luneta	Há só uma voz que pode e há de **protestar**, é a minha, a voz suspeita,
35	18:Pompéia:Ateneu	Esta caranguejola, enorme e pesada, que parecia **protestar**, a cada solavanco, contra o ca
36	18:Taunay:Inocência	algumas pautadas de incitamento, pareceu querer o cargueiro **protestar** contra o tratamen
37	18:Taunay:Inocência	mesmo. Oh! Sr. Garcia! quis **protestar** Pereira. --Nada;.. digo-lhe isto do coração..
38	18:Alencar:Correr	no ator que caíra em falta, nem de **protestar** contra o ato dos diretores por uma semelhan
39	18:Cunha:Artigos	de Lourenço de Médicis. Precisamos, porém, **protestar** também contra a introdução deste p
40	18:Alencar:Garatuja	naquela capitania, que vinham todos unidos em corpo **protestar** contra a violência inaudita
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, vê-se que o poeta queria principalmente **protestar** contra o caminho que l

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



[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.corpusdoportugues.org/ Go
Google Search 1752 blocked Check AutoLink
Internet

SEARCH STRING HELP
WORD/PHRASE querer.* [davies@gı
(INSERT TAG) -SELECT-
SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)
● TABLE
○ CHART
○ SURROUNDING WORDS (HELP)
-SELECT- 5 5

SORT BY (HELP)
● FREQUENCY
○ PERCENT

SECTION SORT (HELP)
1 -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL
2 -- ALL -- / ---------- / 1900s-ALL / 1900s-BR / 1900s-PT / ---------- / ORAL
MIN FREQ 5 □ MIN FREQ 5 □

OPTIONS (HELP)
HITS 100
DISPLAY RAW FREQ
GROUP BY WORDS
SAVE LISTS NO
SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

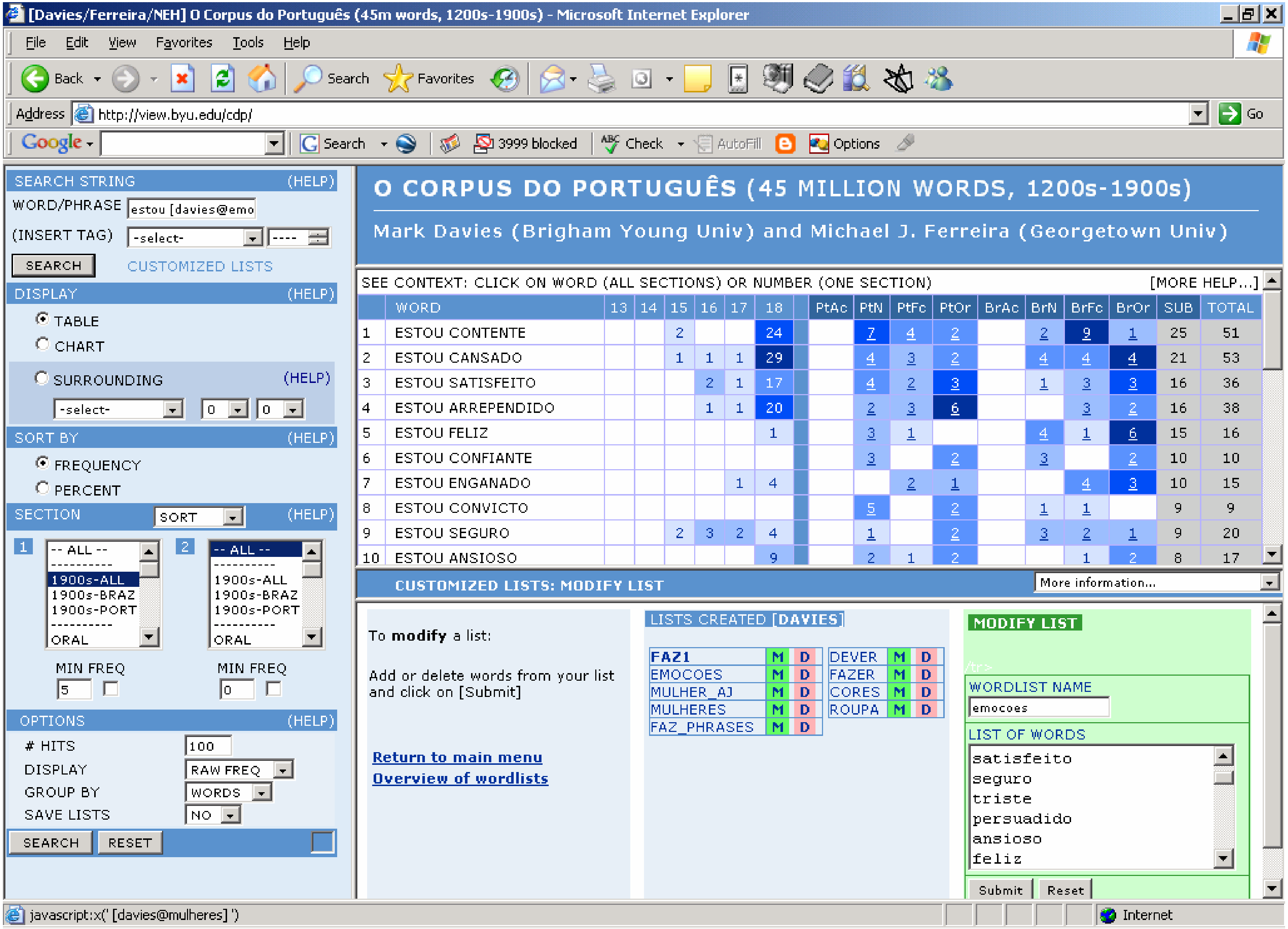
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION) [MORE HELP...]

	WORD	13	14	15	16	17	18		PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	TOTAL
1	QUIS GRITAR						8				1				5		14
2	QUIS PROTESTAR						3			1	2				1		7
3	QUER GRITAR						4										4
4	QUERIA PROTESTAR						1				1						2
5	QUERIA GRITAR						1				1						2
6	QUEREMOS RECLAMAR									1						1	2
7	QUERO GRITAR										1				1		2
8	QUERO CLAMAR													1			1
9	QUIZ GRITAR										1						1
10	QUISESSE GRITAR														1		1

KEYWORDS IN CONTEXT Introduction

31	18:Azevedo:Cortiço	de saias. Jerônimo perdeu a paciência e ia **protestar** brutalmente contra semelhante invasã
32	18:Azevedo:Cortiço	com vontade de afastar-se, mas sem animo de **protestar**, por acanhamento, tentou reatar
33	18:Azevedo:Mulato	vontade até de queixar se ão pai; de **protestar** contra aquelas contrariedades que a faziam
34	18:Macedo:Luneta	Há só uma voz que pode e há de **protestar**, é a minha, a voz suspeita,
35	18:Pompéia:Ateneu	Esta caranguejola, enorme e pesada, que parecia **protestar**, a cada solavanco, contra o ca
36	18:Taunay:Inocência	algumas pautadas de incitamento, pareceu querer o cargueiro **protestar** contra o tratamen
37	18:Taunay:Inocência	mesmo. Oh! Sr. Garcia! quis **protestar** Pereira. --Nada;.. digo-lhe isto do coração..
38	18:Alencar:Correr	no ator que caíra em falta, nem de **protestar** contra o ato dos diretores por uma semelhant
39	18:Cunha:Artigos	de Lourenço de Médicis. Precisamos, porém, **protestar** também contra a introdução deste p
40	18:Alencar:Garatuja	naquela capitania, que vinham todos unidos em corpo **protestar** contra a violência inaudita
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, vê-se que o poeta queria principalmente **protestar** contra o caminho que l

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



[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer

Address http://view.byu.edu/cdp/

SEARCH STRING (HELP)

WORD/PHRASE: estou [davies@emo

(INSERT TAG): -select- | ----

SEARCH | CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE
○ CHART
○ SURROUNDING (HELP)

-select- | 0 | 0

SORT BY (HELP)

● FREQUENCY
○ PERCENT

SECTION: SORT (HELP)

1: -- ALL -- / ---------- / 1900s-ALL / 1900s-BRAZ / 1900s-PORT / ---------- / ORAL

2: -- ALL -- / ---------- / 1900s-ALL / 1900s-BRAZ / 1900s-PORT / ---------- / ORAL

MIN FREQ: 5 □ | MIN FREQ: 0 □

OPTIONS (HELP)

- # HITS: 100
- DISPLAY: RAW FREQ
- GROUP BY: WORDS
- SAVE LISTS: NO

SEARCH | RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young Univ) and Michael J. Ferreira (Georgetown Univ)

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (ONE SECTION) [MORE HELP...]

	WORD	13	14	15	16	17	18	PtAc	PtN	PtFc	PtOr	BrAc	BrN	BrFc	BrOr	SUB	TOTAL
1	ESTOU CONTENTE			2			24		7	4	2		2	9	1	25	51
2	ESTOU CANSADO			1	1	1	29		4	3	2		4	4	4	21	53
3	ESTOU SATISFEITO				2	1	17		4	2	3		1	3	3	16	36
4	ESTOU ARREPENDIDO				1	1	20		2	3	6			3	2	16	38
5	ESTOU FELIZ						1		3	1			4	1	6	15	16
6	ESTOU CONFIANTE								3		2		3		2	10	10
7	ESTOU ENGANADO					1	4			2	1			4	3	10	15
8	ESTOU CONVICTO								5		2		1	1		9	9
9	ESTOU SEGURO			2	3	2	4		1		2		3	2	1	9	20
10	ESTOU ANSIOSO						9		2	1	2			1	2	8	17

CUSTOMIZED LISTS: MODIFY LIST | More information...

To **modify** a list:

Add or delete words from your list and click on [Submit]

Return to main menu
Overview of wordlists

LISTS CREATED **[DAVIES]**

FAZ1	M	D	DEVER	M	D
EMOCOES	M	D	FAZER	M	D
MULHER_AJ	M	D	CORES	M	D
MULHERES	M	D	ROUPA	M	D
FAZ_PHRASES	M	D			

MODIFY LIST

</tr>

WORDLIST NAME: emocoes

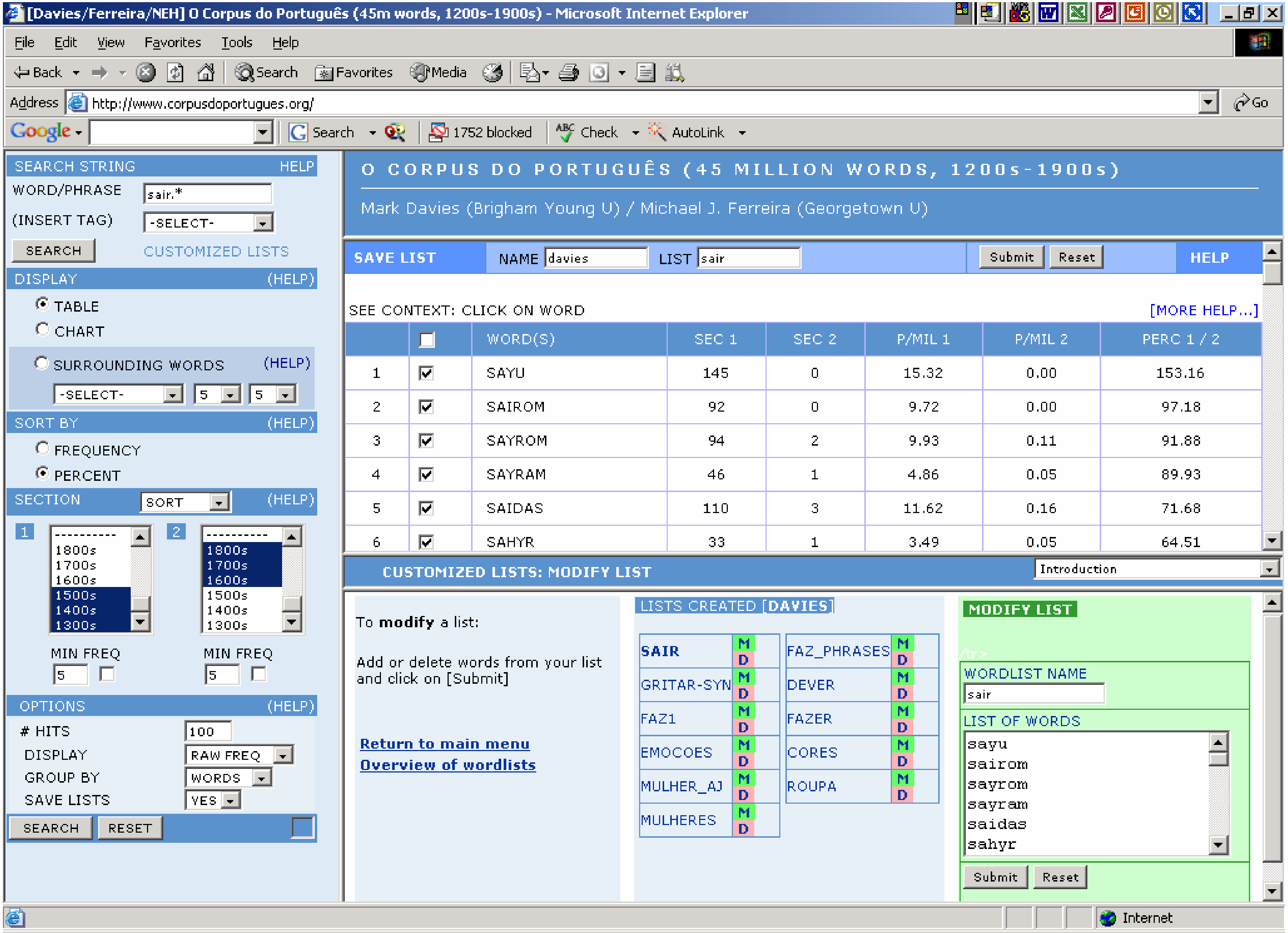
LIST OF WORDS:

satisfeito
seguro
triste
persuadido
ansioso
feliz

Submit | Reset

javascript:x('[davies@mulheres]') Internet

Customized lists: User-created lists (can correct our “errors”): **sair ‘to leave’**


[Davies/Ferreira/NEH] O Corpus do Português (45m words, 1200s-1900s) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.corpusdoportugues.org/
Go
Google
Search 1752 blocked Check AutoLink
Internet

SEARCH STRING HELP

WORD/PHRASE: sair.*

(INSERT TAG): -SELECT-

SEARCH CUSTOMIZED LISTS

DISPLAY (HELP)

● TABLE
○ CHART
○ SURROUNDING WORDS (HELP)

-SELECT- 5 5

SORT BY (HELP)

○ FREQUENCY
● PERCENT

SECTION: SORT (HELP)

1: ----------, 1800s, 1700s, 1600s, 1500s, 1400s, 1300s

2: ----------, 1800s, 1700s, 1600s, 1500s, 1400s, 1300s

MIN FREQ 5 □ MIN FREQ 5 □

OPTIONS (HELP)

HITS: 100
DISPLAY: RAW FREQ
GROUP BY: WORDS
SAVE LISTS: YES

SEARCH RESET

O CORPUS DO PORTUGUÊS (45 MILLION WORDS, 1200s-1900s)

Mark Davies (Brigham Young U) / Michael J. Ferreira (Georgetown U)

SAVE LIST NAME davies LIST sair Submit Reset HELP

SEE CONTEXT: CLICK ON WORD [MORE HELP...]

	□	WORD(S)	SEC 1	SEC 2	P/MIL 1	P/MIL 2	PERC 1 / 2
1	☑	SAYU	145	0	15.32	0.00	153.16
2	☑	SAIROM	92	0	9.72	0.00	97.18
3	☑	SAYROM	94	2	9.93	0.11	91.88
4	☑	SAYRAM	46	1	4.86	0.05	89.93
5	☑	SAIDAS	110	3	11.62	0.16	71.68
6	☑	SAHYR	33	1	3.49	0.05	64.51

CUSTOMIZED LISTS: MODIFY LIST Introduction

To **modify** a list:

Add or delete words from your list and click on [Submit]

Return to main menu
Overview of wordlists

LISTS CREATED [**DAVIES**]

SAIR	M D	FAZ_PHRASES	M D
GRITAR-SYN	M D	DEVER	M D
FAZ1	M D	FAZER	M D
EMOCOES	M D	CORES	M D
MULHER_AJ	M D	ROUPA	M D
MULHERES	M D		

MODIFY LIST

/tr>

WORDLIST NAME: sair

LIST OF WORDS:
sayu
sairom
sayrom
sayram
saidas
sahyr

Submit Reset

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

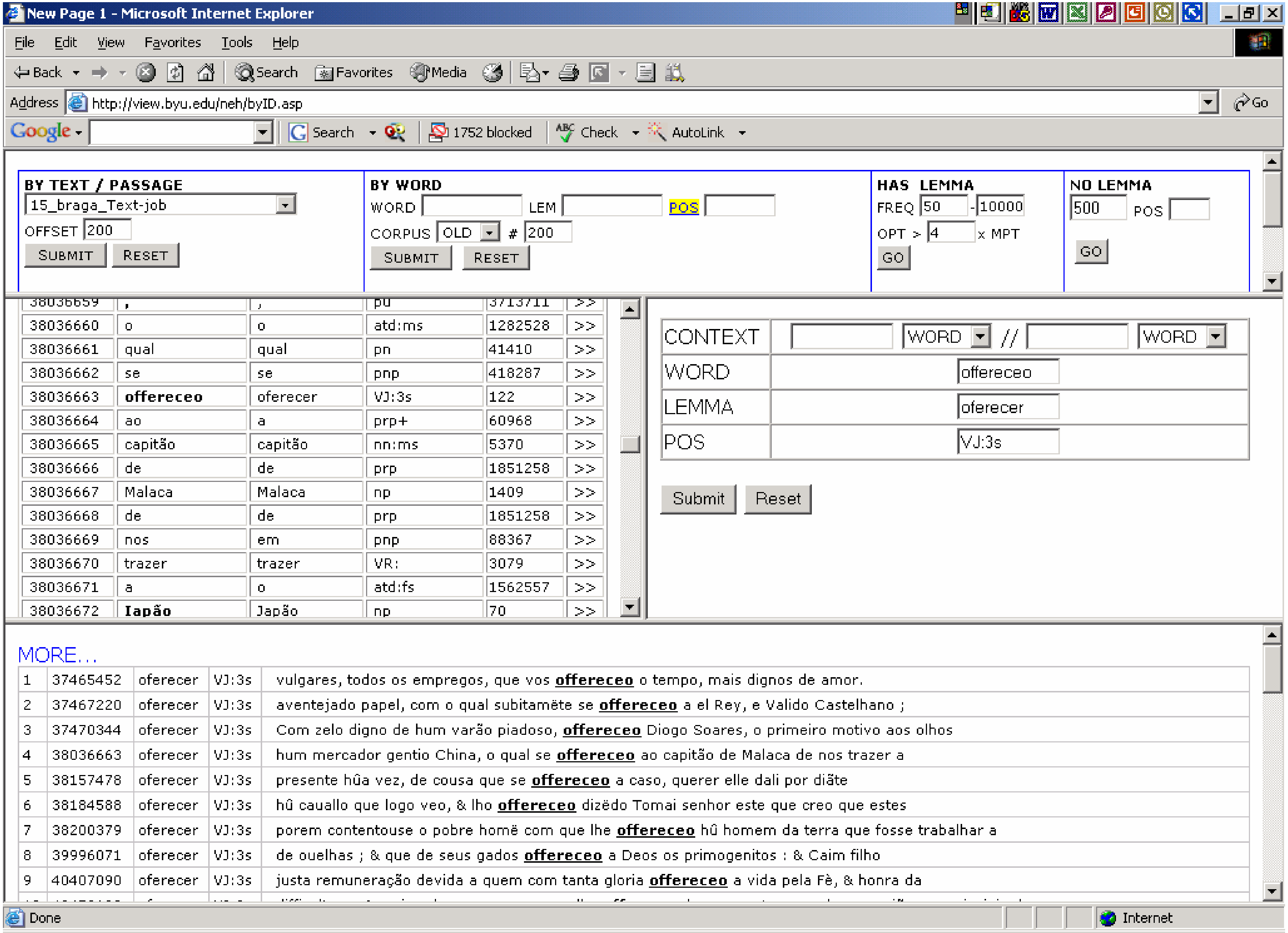
Lemmatization

- English: not much of a problem (*work*, *flye*, *returnes*, but *saw*, *ground*, *wound*, *etc*)
- Much worse for other languages (Sp trabajo ‘work/trabajo, I-work/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)



New Page 1 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Go
Search 1752 blocked Check AutoLink
SUBMIT RESET
GO
Submit Reset
Done
Internet

Address http://view.byu.edu/neh/byID.asp

BY TEXT / PASSAGE
15_braga_Text-job
OFFSET 200

BY WORD
WORD
LEM
POS
CORPUS OLD
# 200

HAS LEMMA
FREQ 50 - 10000
OPT > 4 x MPT

NO LEMMA
500
POS

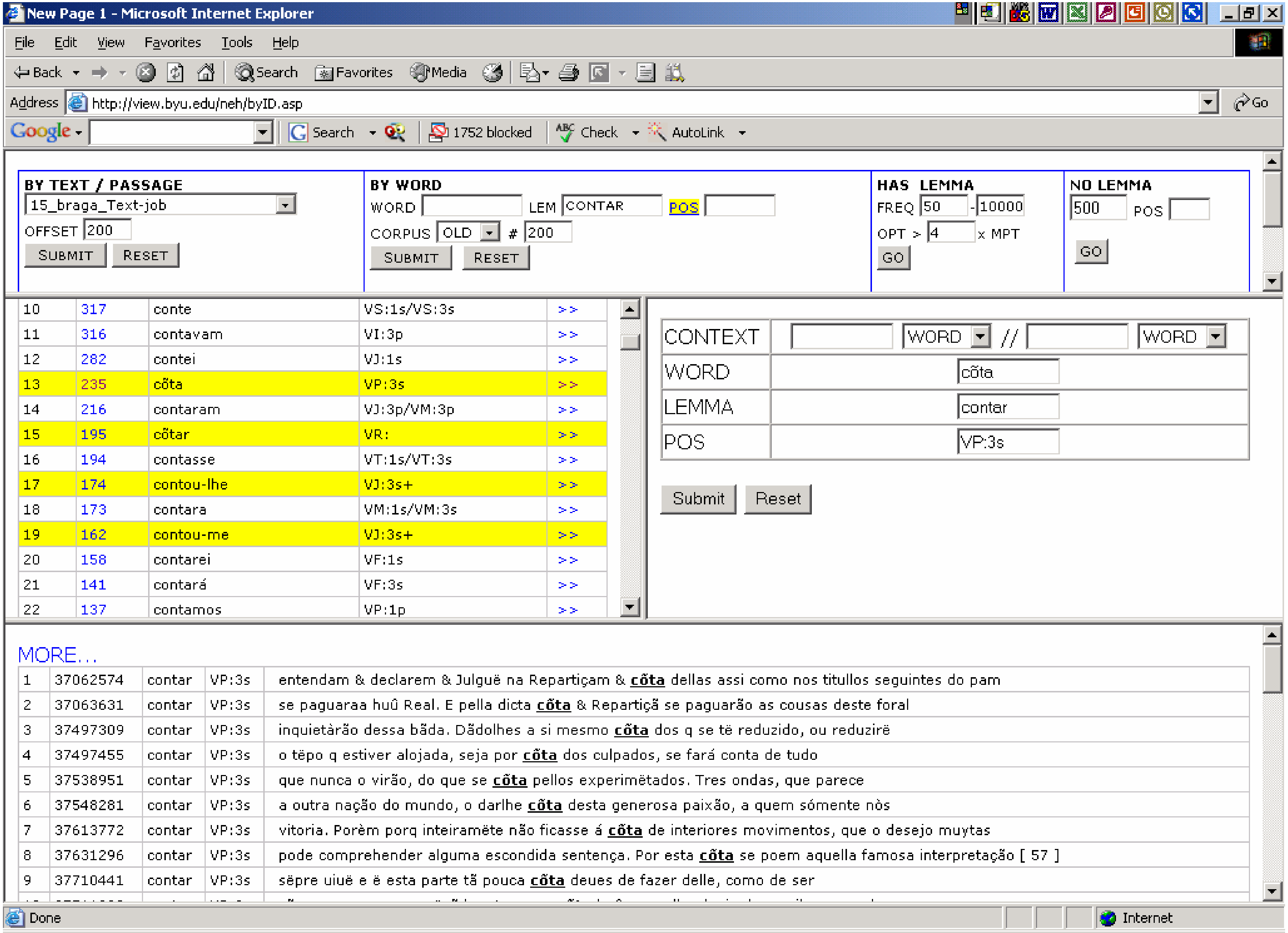
38036659	,	,	pu	3713711	>>
38036660	o	o	atd:ms	1282528	>>
38036661	qual	qual	pn	41410	>>
38036662	se	se	pnp	418287	>>
38036663	**offereceo**	oferecer	VJ:3s	122	>>
38036664	ao	a	prp+	60968	>>
38036665	capitão	capitão	nn:ms	5370	>>
38036666	de	de	prp	1851258	>>
38036667	Malaca	Malaca	np	1409	>>
38036668	de	de	prp	1851258	>>
38036669	nos	em	pnp	88367	>>
38036670	trazer	trazer	VR:	3079	>>
38036671	a	o	atd:fs	1562557	>>
38036672	**Iapão**	Japão	np	70	>>

CONTEXT		WORD //	WORD
WORD	offereceo		
LEMMA	oferecer		
POS	VJ:3s		

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 subitamẽte se **offereceo** a el Rey, e Valido Castelhano ;
3	37470344	oferecer	VJ:3s	Com zelo digno de hum varão piadoso, **offereceo** 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 hũa vez, de cousa que se **offereceo** a caso, querer elle dali por diãte
6	38184588	oferecer	VJ:3s	hũ cauallo que logo veo, & lho **offereceo** dizẽdo Tomai senhor este que creo que estes
7	38200379	oferecer	VJ:3s	porem contentouse o pobre homẽ com que lhe **offereceo** hũ 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

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


New Page 1 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://view.byu.edu/neh/byID.asp
Google Search 1752 blocked Check AutoLink
Done
Internet

BY TEXT / PASSAGE
15_braga_Text-job
OFFSET 200
SUBMIT RESET

BY WORD
WORD LEM CONTAR POS
CORPUS OLD # 200
SUBMIT RESET

HAS LEMMA
FREQ 50 - 10000
OPT > 4 x MPT
GO

NO LEMMA
500 POS
GO

10	317	conte	VS:1s/VS:3s	>>
11	316	contavam	VI:3p	>>
12	282	contei	VJ:1s	>>
13	235	cõta	VP:3s	>>
14	216	contaram	VJ:3p/VM:3p	>>
15	195	cõtar	VR:	>>
16	194	contasse	VT:1s/VT:3s	>>
17	174	contou-lhe	VJ:3s+	>>
18	173	contara	VM:1s/VM:3s	>>
19	162	contou-me	VJ:3s+	>>
20	158	contarei	VF:1s	>>
21	141	contará	VF:3s	>>
22	137	contamos	VP:1p	>>

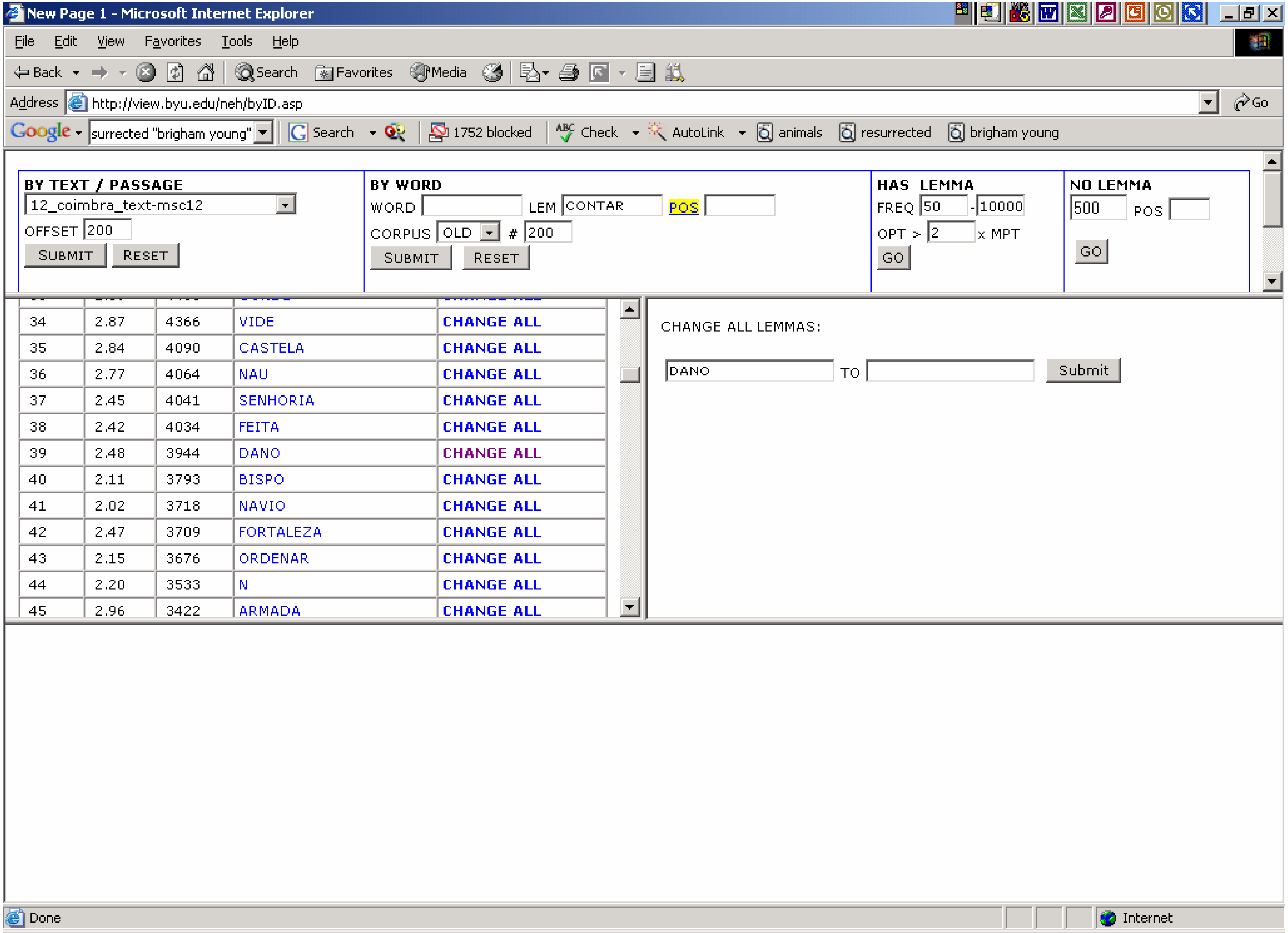
CONTEXT	WORD // WORD
WORD	cõta
LEMMA	contar
POS	VP:3s

Submit Reset

MORE...

1	37062574	contar	VP:3s	entendam & declarem & Julguë na Repartiçam & **cõta** dellas assi como nos titullos seguintes do pam
2	37063631	contar	VP:3s	se paguaraa huû Real. E pella dicta **cõta** & Repartiçã se paguarão as cousas deste foral
3	37497309	contar	VP:3s	inquietàrão dessa bãda. Dãdolhes a si mesmo **cõta** dos q se të reduzido, ou reduzirë
4	37497455	contar	VP:3s	o tëpo q estiver alojada, seja por **cõta** dos culpados, se fará conta de tudo
5	37538951	contar	VP:3s	que nunca o virão, do que se **cõta** pellos experimëtados. Tres ondas, que parece
6	37548281	contar	VP:3s	a outra nação do mundo, o darlhe **cõta** desta generosa paixão, a quem sómente nòs
7	37613772	contar	VP:3s	vitoria. Porèm porq inteiramëte não ficasse á **cõta** de interiores movimentos, que o desejo muytas
8	37631296	contar	VP:3s	pode comprehender alguma escondida sentença. Por esta **cõta** se poem aquella famosa interpretação [57]
9	37710441	contar	VP:3s	sëpre uiuë e ë esta parte tã pouca **cõta** deues de fazer delle, como de ser

Interface for annotating Old Portuguese (by lemma frequency I)



New Page 1 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Go
Search
1752 blocked
Check
AutoLink
animals
resurrected
brigham young
Done
Internet

Address http://view.byu.edu/neh/byID.asp

Google surrected "brigham young"

BY TEXT / PASSAGE

12_coimbra_text-msc12

OFFSET 200

SUBMIT RESET

BY WORD

WORD

LEM CONTAR

POS

CORPUS OLD

# 200

SUBMIT RESET

HAS LEMMA

FREQ 50 - 10000

OPT > 2 x MPT

GO

NO LEMMA

500

POS

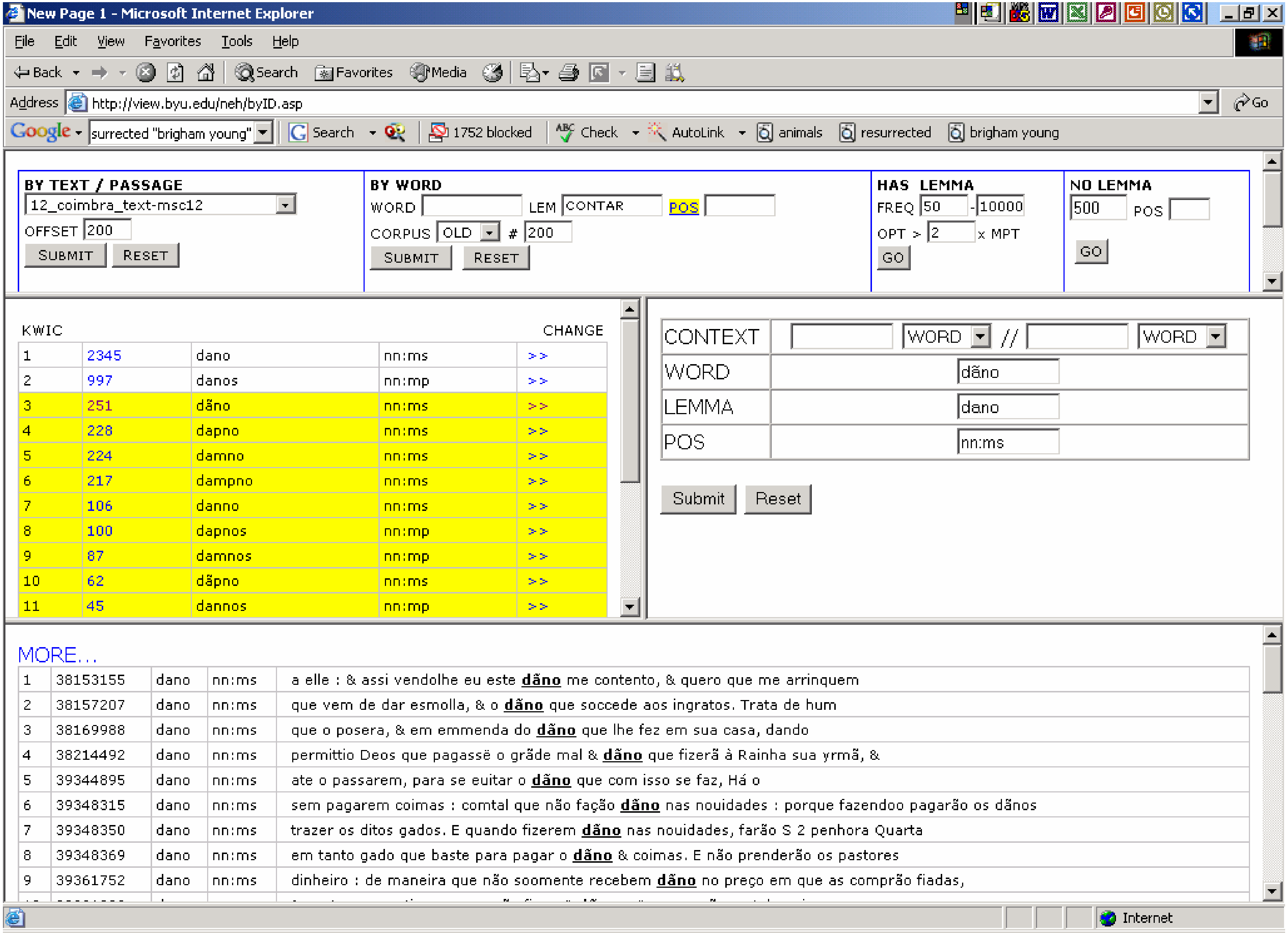
GO

34	2.87	4366	VIDE	CHANGE ALL
35	2.84	4090	CASTELA	CHANGE ALL
36	2.77	4064	NAU	CHANGE ALL
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

CHANGE ALL LEMMAS:

DANO TO

Submit


New Page 1 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address http://view.byu.edu/neh/byID.asp
Go
Google surrected "brigham young"
Search
1752 blocked
Check
AutoLink
animals
resurrected
brigham young
Internet

BY TEXT / PASSAGE
12_coimbra_text-msc12
OFFSET 200
SUBMIT RESET

BY WORD
WORD   LEM CONTAR   POS
CORPUS OLD   # 200
SUBMIT RESET

HAS LEMMA
FREQ 50 - 10000
OPT > 2 x MPT
GO

NO LEMMA
500 POS
GO

KWIC   CHANGE

1	2345	dano	nn:ms	>>
2	997	danos	nn:mp	>>
3	251	dãno	nn:ms	>>
4	228	dapno	nn:ms	>>
5	224	damno	nn:ms	>>
6	217	dampno	nn:ms	>>
7	106	danno	nn:ms	>>
8	100	dapnos	nn:mp	>>
9	87	damnos	nn:mp	>>
10	62	dãpno	nn:ms	>>
11	45	dannos	nn:mp	>>

CONTEXT	WORD // WORD	
WORD	dãno	
LEMMA	dano	
POS	nn:ms	

Submit Reset

MORE...

1	38153155	dano	nn:ms	a elle : & assi vendolhe eu este **dãno** me contento, & quero que me arrinquem
2	38157207	dano	nn:ms	que vem de dar esmolla, & o **dãno** que soccede aos ingratos. Trata de hum
3	38169988	dano	nn:ms	que o posera, & em emmenda do **dãno** que lhe fez em sua casa, dando
4	38214492	dano	nn:ms	permittio Deos que pagassë o grãde mal & **dãno** que fizerã à Rainha sua yrmã, &
5	39344895	dano	nn:ms	ate o passarem, para se euitar o **dãno** que com isso se faz, Há o
6	39348315	dano	nn:ms	sem pagarem coimas : comtal que não fação **dãno** nas nouidades : porque fazendoo pagarão os dãnos
7	39348350	dano	nn:ms	trazer os ditos gados. E quando fizerem **dãno** nas nouidades, farão S 2 penhora Quarta
8	39348369	dano	nn:ms	em tanto gado que baste para pagar o **dãno** & coimas. E não prenderão os pastores
9	39361752	dano	nn:ms	dinheiro : de maneira que não soomente recebem **dãno** no preço em que as comprão fiadas,

Composition of proposed 200 million word historical corpus

	SPOKEN / INFORMAL		FICTION		NEWS		ACADEMIC		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**		**0**			**6**
TOTAL	**33**		**67**		**24**		**69**		**7**	**200**
PERCENT	**16.5**		**33.5**		**12.0**		**34.5**		**3.5**	**100%**

Conclusion

- 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