Computational Semantics and Pragmatics 2011

Raquel Fernández – ILLC, University of Amsterdam

Homework #2

Exercise 1: Distributional meaning

Have a look at this online tool: http://www.scottishcorpus.ac.uk/corpus/bnc/compair.php.

- 1.1 What do you think this tool does? Explain it in your own words.
- 1.2 The English adverbs '*utterly*' and '*absolutely*' are typically considered synonymous by most thesauri (http://thesaurus.com/browse/utterly). Compare them using this tool. What do you observe? What can you conclude about their meaning and usage?
- 1.3 Compare at least one more pair of words and comment on the results.

Exercise 2: Word usages across genres

The web interface of the Corpus of Contemporary American English (COCA) http://corpus.byu.edu/coca/ allows you to compare the *collocates* of a word (those words which appear frequently nearby) across different genres. This snap shot shows the results of a search for adjectival collocates of the word *'chair'* in the FICTION and ACADEMIC sections of the corpus:

410 MILLION WORDS, 1990	2010														
DISPLAY	SEE 0	CONTEXT: CLIC	CK ON WORD	[HELP]											
	SEC	1: 81,723,175	VORDS												
ARCH STRING		WORD/PHRA	SE TOP	ENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
	1	OVERSTUFFE	D	62	1	0.76	0.01	62.90	1	VICE	54	0	0.65	0.00	65.13
RD(S) chair	2	HARD		61	1	0.75	0.01	61.89	2	NATIONAL	38	0	0.46	0.00	45.83
ULLOCATES [j*] 4	3	NEARBY		38	0	0.46	0.00	46.50	3	PAST	65	2	0.78	0.02	32.03
4	4	STRAIGHT		45	1	0.55	0.01	45.66	4	INTERNATIONAL	28	1	0.34	0.01	27.60
POS LIST adiALL	5	5 SOFT		45	1	0.55	0.01	45.66	5	TECHNICAL	20	0	0.24	0.00	24.12
	6	HEAVY		37	0	0.45	0.00	45.27	6	ASSOCIATE	16	0	0.19	0.00	19.30
NDGM SEARCH RESET	8 7	ASLEEP		84	2	1.03	0.02	42.61	7	JOINT	14	0	0.17	0.00	16.88
	2 8	NEAREST		33	0	0.40	0.00	40.38	8	SOCIAL	13	0	0.16	0.00	15.68
IGNORE IGNORE	9	BROWN		32	0	0.39	0.00	39.16	9	SUCCESSFUL	12	0	0.14	0.00	14.47
SPOKEN SPOKEN FICTION MAGAZINE		KEYWORD IN CONTEXT DISPLAY Help / Information / contact SECTION: NO LIMITS												2	
NEWSPAPER ACADEMIC V ACADEMIC S		K FOR MORE C	-		[7] SAVE	цят)	снос	SE LIST	f [CREATE	NEW LIST			[?]	
RTING AND LIMITS	1 3	2009 ACAD E	nvironHealth	A B	C make t	heir so	hool o	listrict m	nore	sustainable. NEHA S	ustainability	Committee	Chair	and S	ustainat
DRTING RELEVANCE SEC1 : SEC2	2 2 2	2009 ACAD E	nvironHealth	A B	C for its	outstar	nding	efforts to	o mai	ke their school distri	ct more sust	ainable. NE	HA Sus	stainal	oility Co
MINIMUM FREQUENCY V 10 10 0	3 2	3 2007 ACAD MechanicalEng A B C for over 25 years as a technical committee member, a technical committee s										ittee se	creta	ry and g	
NIMUM FREQUENCY 💽 🗹 10 10		2007 ACAD M	lechanicalEng	AB	C Contro	l, inclu	ding 1	4 ASME	Sym	posiums and confer	ences. He se	erved as the	Tech	nical	Program
CK TO SEE OPTIONS	5 3	2007 ACAD M	lechanicalEng	AB	C NIH, O	NR and	the D	OOE. In 2	2003	he served as the Te	echnical Pro	ogram <u>Chai</u>	r for th	he AS	ME Inte
	6 3	2007 ACAD M	echanicalEng	AB	C Clark o	urrent	ly ser	ves as A	ssoci	ate Editor of the Jou	rnal of Vibra	ation and Ad	oustics	s, is <u>C</u>	hair of
	7 2	2005 ACAD M	lechanicalEng	AB	C of the	AIAA T	herm	ophysics	and	Electric Propulsion	echnical (ommittees,	and se	erved	as the f
	8 2	2005 ACAD M	lechanicalEng	AB	C Contro	l and t	he IEE	E Trans	actio	ns on Control Syster	n Technolog	y. He has s	erved a	as the	chair
	9 2	2005 ACAD M	lechanicalEng	AB	C particip	ant in	the a	ctivities	of the	ASME IGTI for ove	r 25 years,	serving as	chair (of tec	hnical
	5	ta la		1	-1 - 50		1						o. A.		>

- 2.1 Reproduce this search on the COCA website and explore the results: you can click on the collocate words to see the contexts where they appear, and change the value in the POS LIST field to look for collocates with Part-of-Speech other than adjective. Given these distributional results, what can you conclude about the meaning of the word 'chair'? Back up your claims with examples.
- 2.2 Compare the collocates of two other words across genres and explain your findings.

Exercise 3: Word usages from a diachronic perspective

Extracting the collocates of a word in different time periods can give us an indication of how the meaning of a word has changed over time. The web interface of the Corpus of Historical American English (COHA) http://corpus.byu.edu/coha/ allows you do this and much more. For instance, the following snap shot shows the nominal collocates of the word *'web'* in the 1950s and the 2000s.

CORPUS OF HISTORIC	LA		CICA	N EN	o Lito	9 M								R	AQUEL				
400 MILLION WORDS, 1810s	-200	0s																	
ISPLAY	SEE (CONTEXT	L: CLICK C	N WORD (A	[HELP]														
	SEC	1: 24,5	44,831 W	ORDS					SEC	2: 29,567,390 WORDS									
EARCH STRING								RATIO											
WORD(S) web	1	[SPIDE	R]	17	31	0.69	1.05	0.66	1	[SITE]	1142	0	38.62	0.00	3,862.36				
COLLOCATES [nn*] 4 + 4 +	2	[SOCI	ETY]	5	11	0.20	0.37	0.55	2	[PAGE]	90	0	3.04	0.00	304.39				
	3	[BELT]		4	4	0.16	0.14	1.20	3	[]	59	2	2.00	0.08	24.49				
	4	[CENT	ER]	4	5	0.16	0.17	0.96	4	[WORLD]	58	0	1.96	0.00	196.16				
ANDOM SEARCH RESET	5	[LIFE]		4	13	0.16	0.44	0.37	5	[SERVICE]	46	0	1.56	0.00	155.58				
	6	[TRAC	к]	3	1	0.12	0.03	3.61	6	[E-MAIL]	45	0	1.52	0.00	152.19				
1950 A 2 IGNORE A	7	[STRA	ND]	3	2	0.12	0.07	1.81	7	[INFORMATION]	38	0	1.29	0.00	128.52				
1940 2000																			
1930 1990 1920 1980		KEYWORD IN CONTEXT DISPLAY										Help / Information / contact							
1910 1900 1960	SECT	SECTION: NO LIMITS									PAGE: « < 1/4 » » SAMPLE: 100 200								
1890 🗸 1950 🗸	CLIC	K FOR M	ORE CONT	EXT			[?]	SAVE LI	ST	CHOOSE LIST	CREATE N	EW LIST			[7				
RTING AND LIMITS	1	1822 F	IC Odo	friedeOutca	st	A	зс	sumptuo	ous co	ouches once Ministered unto	their idle lu:	kury: The	e venor	nous s	pider ha				
ORTING FREQUENCY -	2	2 1827 FIC HopeLeslieVolume2				A	зс	faint sm	ile, "	ne lily	of the Ma								
	3	1830 F	IC Clar	enceATaleO	ur	A	зс	herself a	is she	e did so, " It is as fine as a <u>s</u>	pider's wet	o, a prett	y articl	e for ti	ne like of				
	4	4 1833 NF ByOrd			rderKing			no cause	for	alarm. " " Ships are like flies	in the spid	er's web	of the	sea. "	" Just no				
DE OPTIONS	5	1837 F	IC lion	ро		A	зс	But thou	, oh man! Whose hold on life is like the spider's web, Who hast thy footin										
SROUP BY	6	1838 F	IC Gen	tlemanLyon	s	A	3 C	de they	prese	ent Me now How's this?	the musty <u>s</u>	pider's v	veb He	re in m	iy mothe				
DISPLAY RAW FREQ	7	1838 M	AG USD	emRev		A	3 C	only ser	ves te	o render the former sensible	of their ent	angleme	nt in th	e polit	cal spid				
						100					who sits within as a spider in its web, you								

- 3.1 Using the COHA web interface, compare the collocates of another word or construction across different time periods and explain your findings.
- 3.2 Read this post on the Language Log http://languagelog.ldc.upenn.edu/nll/?p=2585. The snap shot below shows how to use the COHA interface to obtain the graph with the frequency of use of the word *'normalcy'* that appears in Mark Liberman's reply to one of the comments to the post.

					NG	Gardens														RA			
400 MILLION WORDS, 1810s-2000s COMPARE RESULTS: COCA COHA TIME BAC																							
ISPLAY	CLICK	CLICK ON BARS FOR CONTEXT											ICK ON COLUMN HEADINGS FOR FREQUENCY IN SUB-SECTION										
	SECTION	1810 0	1820	1830	1840 0	1850	1860	1870	1880	1890	1900	1910 0	1920 24	1930 27	1940 29	1950 25	1960 26	1970	1980 28	1990 34	2000		
EARCH STRING	PER MIL		0.00	0.00	0.00	0.00			0.00	0.00	0.00	0.00	0.94	1.10	1.19	1.02	1.08	1.13	1.11	1.22	45		
NORD(S) normalcy																							
																1. 1.		-	1.1				
COLLOCATES	SEE ALL YEARS																						
POS LIST	AT ONCE																						
ANDOM SEARCH RESET																							
			2				-	-	-												-		
1																							
1950 🗸 1950 🗸	CLICK F	OR MOR	E CONT	ТЕХТ			171	SAVE LIST	СНО	OSE LIS	ST	v	CRE	ATE NEV	N LIST	([7]						
ORTING AND LIMITS	1 192	0 NEWS	NYT-I	Let		No. Alter	State 1	im-, pea	hable.	if it ha	d, for e	xample	more	of the "	norma	lcv " of	Senato	or Hardi	na's utte	rance.	it miah		
SORTING FREQUENCY +		0 NEWS		10		AB	1	NNYBACH															
	3 192	0 NEWS	NYT-	Let		AB	C 19	201002 1	xt NO	TREATIE	S. NO	LEAGUE	. For Th	nev Con	flict with	h the D	octrine o	of Norn	alcy. T	o the E	ditor of		
THEODENCY I IS	4 192	0 NEWS	NYT-I	let				involver															
	AV - 0.000		NYT-	Bog	_	AB	C the	m. lie re	marke	d that th	ne Repu	blican o	andida	te was a	askina fi	or a ret	urn to "	norma	lcv " ar	nd asked	1:. " Do		
LICK TO SEE OPTIONS	5 192	0 NEWS																					
LICK TO SEE OPTIONS		0 NEWS	122200	100 100 P		A B	C his	hearers	" Well	, " he w	ent on	* a ret	urn to r	ormal	cy mea	ns a rei	turn to t	he good	l old tim	es of th	e long		
LICK TO SEE OPTIONS		0 NEWS	122200	Reg		A B		hearers															

Using the Chart display of the COHA web interface, find:

- (a) a word or construction whose frequency has increased to the present time;
- (b) a word or construction whose frequency has decreased since the 1800s.

Provide evidence for your claims and explain your findings.