



**DIGITAL
SCHOLARSHIP**

www.digital.cahss.ed.ac.uk

Workshop: Using Corpus-based Methods in
the Social Sciences and Humanities

Using Corpus-Based Methods in the Social Sciences and Humanities

Dr Laura A. Cariola

**School of Health in Social Science &
Institute for Advanced Studies in the
Humanities**

What is a Corpus?

- “Corpus” comes from the Latin word “body” – plural “corpora”
- A corpus is a collection of texts
 - Not random
 - Purposeful
 - Sampled to be representative of the subject investigated
 - Machine-readable for computer analysis
 - Spoken or written data

What is a Corpus?

“A corpus is a collection of pieces of language text in electronic form, selected according to external criteria* to represent, as far as possible, a language or language variety as a source of data for linguistic research”.
(Sinclair, 2005)

*External criteria refers to language variety, text type, text location etc.

Why use Corpus Methods?

- Quick computerised processing of medium to big data
- Provides an objective and balanced picture
- Identifies quantitatively typical and atypical patterns
- Uses reliable statistics
- Provides contextual qualitative examples
- Tests existing hypotheses or enables to develop new hypotheses
- Produces objective evidence vs. intuition

What can Corpus Analysis do?

- **Quantitative analysis:**
 - Counts lexical items
 - Language changes over time
 - Language use of speakers/characters
 - Semantic field
 - Semantic prosody (coarse grain)
 - Comparison of text — e.g., chapters, sections
 - Grammatical features
- **Qualitative analysis**
 - Contextual differences
 - Semantic prosody (fine grain)

What can Corpus Analysis do?

- **Key words** – highly occurring words
- **Parts-of-speech** (grammatical properties)
- **Semantic prosody**
 - Words that occur frequently with other words
 - Create positive and negative meaning
 - Produce unconscious associations
 - E.g. battling cancer (war metaphor - aggressive) vs. going through cancer (journey metaphor - softer)

Different Types of Corpora

- **Reference corpus** — often large, includes spoken and written texts, provides comprehensive information about a language/ standard comparison to another corpus
- **Monitor corpus** — grows in size over time (e.g., Bank of English)
- **Balanced corpus** — balanced according to type, genre, situation (e.g., FLOB)
- **Sample corpus** — small size of carefully selected text
- **Annotated corpus** — includes linguistics/grammatical information (e.g., BNC, Brown, LOB)
- **Unannotated corpus** — plain text without any linguistic information

Different Types of Corpora

- **General corpora** — Large language variety (e.g., BNC)
- **Specialised corpora** — specific to a genre (e.g., Shakespeare)
- **Monolingual or multilingual corpora** — (e.g., COMENEGO: Multilingual Corpus of Business and Economics)
- **Parallel corpus** — translated texts, bilingual or multilingual (e.g., Bible)
- **Comparable corpus** — collection of similar texts in different language/ regional differences (e.g., Brown and LOB)
- **Spoken and written corpora** — (e.g., BNC Spoken 2014)
- **Native speaker and learner corpora** — (e.g., The Leiden Learner Corpus)

Corpus in the Social Sciences and Humanities: Law

Focus on use of legal language & forensic linguistics:

- Interpreting and analysing legal texts
- Comparing different legal texts
- Discursive construction of actors
- Authorship analysis (e.g., forgery, text messages)
- In 2013 BYU Law School started the first class on law and corpus linguistics

Corpus in the Social Sciences and Humanities: Law

- The Corpus of Historical English Law Reports 1535–1999 (CHELAR): A resource for analysing the development of English legal discourse (Fanego et al., 2017)
- Comparative legal linguistics (Matilla 2016)
- 12 Calculating legal meanings? Drawbacks and opportunities of corpus-assisted legal linguistics to make the law (more) explicit (Vogel, 2017)
- Patterns in language and law (Solan, 2017)

Corpus in the Social Sciences and Humanities: Politics

- Much research on political presentations
- Often paired with Critical Discourse Analysis
- Analysis of political discourses, e.g., newspaper discourses, manifestos, social media discussions
- Political ideologies
- Presentations of “us” vs. “others”
- Cognitive fallacies to justify actions

Corpus in the Social Sciences and Humanities: Politics

- Priming Language Political Issues as Issues of State Security: A Corpus-Assisted Discourse Analysis of Language Ideological Debates in Estonian Media Before and After the Ukrainian Crisis (Siiner et al., 2017)
- The Cultural Political Economy of Austerity in Germany: A Corpus-Assisted Critical-Realist Discourse Analysis (Griebel, 2017)
- Evaluation in political discourse addressed to women: Appraisal analysis of Cosmopolitan's online coverage of the 2014 US midterm elections (Mayo & Taboada, 2017)

Corpus in the Social Sciences and Humanities: Literature

- Mostly associated with “corpus stylistic”
- Analysis of literary text with focus on:
- Distribution of words to identify textual features that characterise an author or particular text
- Focus on grammar and stylistic
- Translations
- Authorship attribution and literary style
- Focus on particular style of author

Corpus in the Social Sciences and Humanities: Literature

- Explorations in an English Poetry Corpus: A Neurocognitive Poetics Perspective (Jacobs, 2018)
- Corpus Methodologies in Literary Translation Studies: An Analysis of Speech Verbs in Four Spanish Translations of *Hard Times* (Ruano, 2017)
- Woman's World Portrayed in Literary Works of Jane Austen (Praminatih & Nafiah, 2018)

Corpus in the Social Sciences and Humanities: Pedagogy

- Error analysis
- Language assessments
- Competencies at different ages, education
- Academic literacy
- Literacy and bilingualism

Corpus in the Social Sciences and Humanities: Pedagogy

- What does corpus linguistics have to offer to language assessment? (Xi, 2017)
- Academic Phraseology: A Key Ingredient in Successful L2 Academic Literacy (Granger, 2017)
- Historical literacy in bilingual settings: Cognitive academic language in CLIL history narratives (Lorenzo, 2017)

Building your own Corpus

- **1) Formulate a research question**
- **2) Identify sources to collect your data:**
 - Existing corpus
 - Data archives
 - Books
 - Transcripts of spoken language
 - Blogs
 - Newspapers
 - Social media
- **3) Check copyright**

Building your own Corpus

- **4) After you have collected your data:**
 - Clean up the corpus — pictures, mark-up etc.
 - Keep notes of changes to the original corpus
 - Convert the text into an acceptable format (e.g., .txt)
- **5) Choose an appropriate corpus software, such as WordSmith**

Corpus Size

- There aren't any set parameters
- Dependent on the purpose of the research
- Studies exploring collocations require bigger corpus size
- Specialized corpora can be much smaller

Basics Terms in Corpus Research

- **Type/token**
 - We want to compile a corpus to do research relevant to our topic
 - 13tokens (words)
 - 11 types (different words) (we, want, to, compile, a, corpus, carry, do, relevant, our, topic)
- **Word-form / lemma**
 - Writ|e, writ|es, writ|ing, wr|ote (form of lemma write)

Exploring a Corpus

- **Word Lists**
 - List that displays words (alphabetically or frequency order) of a corpus
- **Keyword Lists**
 - To identify high frequency occurring words in a corpus relatively to another corpus (reference corpus)
 - High frequency = high key vs. low key
- **Concordances**
 - Examples of a given word or phrase in a text
 - KWIC Key Word in Context
 - Node word is in central position
- **Collocations**
 - Co-occurrence of two or more words/lexical items

Word Lists

WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	% person	.emmas	Set
1	THE	1,148	5.21	1	00.00	0.96	
2	TO	753	3.41	1	00.00	0.96	
3	AND	642	2.91	1	00.00	0.96	
4	HE	577	2.62	1	00.00	0.91	
5	HIS	550	2.49	1	00.00	0.92	
6	OF	429	1.95	1	00.00	0.94	
7	WAS	409	1.85	1	00.00	0.93	
8	HAD	352	1.60	1	00.00	0.88	
9	IT	352	1.60	1	00.00	0.93	
10	IN	348	1.58	1	00.00	0.94	
11	THAT	333	1.51	1	00.00	0.91	
12	A	285	1.29	1	00.00	0.91	
13	AS	242	1.10	1	00.00	0.94	
14	GREGOR	199	0.90	1	00.00	0.89	
15	WITH	199	0.90	1	00.00	0.90	
16	SHE	196	0.89	1	00.00	0.75	
17	HIM	188	0.85	1	00.00	0.90	
18	HER	187	0.85	1	00.00	0.76	
19	WOULD	187	0.85	1	00.00	0.79	
20	NOT	176	0.80	1	00.00	0.92	
21	BUT	171	0.78	1	00.00	0.96	
22	AT	169	0.77	1	00.00	0.91	
23	FOR	166	0.75	1	00.00	0.91	
24	THEY	156	0.71	1	00.00	0.77	

Keyword Lists

Kafka's Metarphosis vs. The Trial

N	Key word	Freq.	%	Texts	RC	Freq.	RC	%	Log_L	Log_R	P	emmas	Set
1	GREGOR	199	0.90	1	0				626.07	142.21	0.0000000000		
2	GREGOR'S	99	0.45	1	0				311.46	141.20	0.0000000000		
3	SISTER	96	0.44	1	0				302.02	141.16	0.0000000000		
4	FATHER	96	0.44	1	5				264.54	6.20	0.0000000000		
5	MOTHER	82	0.37	1	0				257.98	140.93	0.0000000000		
6	SAMSA	34	0.15	1	0				106.97	139.66	0.0000000000		
7	PARENTS	26	0.12	1	0				81.80	139.27	0.0000000000		
8	FAMILY	31	0.14	1	6				67.52	4.30	0.0000000000		
9	GRETE	22	0.10	1	1				61.45	6.39	0.0000000000		
10	CHIEF	37	0.17	1	18	0.02			55.23	2.97	0.0000000000		
11	CLERK	34	0.15	1	19	0.02			46.63	2.77	0.0000000000		
12	FOOD	14	0.06	1	0				44.05	138.38	0.0000000000		
13	COUCH	17	0.08	1	2				41.63	5.02	0.0000000000		
14	BODY	24	0.11	1	9	0.01			41.02	3.35	0.0000000000		
15	VIOLIN	12	0.05	1	0				37.75	138.16	0.0000000000		
16	FURNITURE	15	0.07	1	3				32.37	4.26	0.0000000098		
17	FLAT	20	0.09	1	10	0.01			29.38	2.93	0.0000000566		
18	BOSS	9	0.04	1	0				28.31	137.74	0.0000001002		
19	CLEANER	9	0.04	1	0				28.31	137.74	0.0000001002		
20	CHARWOMAN	9	0.04	1	0				28.31	137.74	0.0000001002		
21	HERE	19	0.09	1	204	0.24	-24.70	-1.49	0.0000006672				
22	IVE	4	0.02	1	100	0.12	-25.17	-2.71	0.0000005231				
23	HAS	4	0.02	1	103	0.12	-26.33	-2.75	0.0000002851				
24	YOU'RE	6	0.03	1	120	0.14	-26.42	-2.39	0.0000002716				
25	THAT'S	11	0.05	1	160	0.19	-27.35	-1.93	0.0000001667				
26	ASKED	16	0.07	1	199	0.24	-28.94	-1.70	0.0000000718				
27	CAN	15	0.07	1	194	0.23	-29.46	-1.76	0.0000000542				
28	DON'T	12	0.05	1	183	0.22	-32.67	-2.00	0.0000000080				
29	IT'S	16	0.07	1	218	0.26	-34.96	-1.83	0.0000000004				
30	MAN	7	0.03	1	180	0.21	-45.98	-2.75	0.0000000000				
31	ARE	6	0.03	1	227	0.27	-68.65	-3.31	0.0000000000				
32	IS	26	0.12	1	425	0.50	-80.54	-2.10	0.0000000000				
33	ME	17	0.08	1	370	0.44	-86.00	-2.51	0.0000000000				
34	SAID	51	0.23	1	770	0.91	-136.24	-1.98	0.0000000000				
35	I	58	0.26	1	863	1.02	-150.68	-1.96	0.0000000000				
36	YOU	51	0.23	1	1,039	1.23	-231.58	-2.41	0.0000000000				

Kafka's The Trial vs. Metarphosis

N	Key word	Freq.	%	Texts	RC	Freq.	RC	%	Log_L	Log_R	P	emmas	Set
1	K	1,176	1.40	1	0				546.73	142.84	0.0000000000		
2	YOU	1,039	1.23	1	51	0.23	231.58	2.41	0.0000000000				
3	I	863	1.02	1	58	0.26	150.68	1.96	0.0000000000				
4	SAID	770	0.91	1	51	0.23	136.24	1.98	0.0000000000				
5	LAWYER	193	0.23	1	0		89.73	140.23	0.0000000000				
6	ME	370	0.44	1	17	0.08	86.00	2.51	0.0000000000				
7	IS	425	0.50	1	26	0.12	80.54	2.10	0.0000000000				
8	COURT	158	0.19	1	0		73.45	139.94	0.0000000000				
9	ARE	227	0.27	1	6	0.03	68.65	3.31	0.0000000000				
10	PAINTER	128	0.15	1	0		59.51	139.64	0.0000000000				
11	S	125	0.15	1	0		58.11	139.60	0.0000000000				
12	TRIAL	99	0.12	1	0		46.03	139.27	0.0000000000				
13	MAN	180	0.21	1	7	0.03	45.98	2.75	0.0000000000				
14	MISS	107	0.13	1	1		41.54	4.81	0.0000000000				
15	LENI	88	0.10	1	0		40.91	139.10	0.0000000000				
16	UNCLE	87	0.10	1	0		40.45	139.08	0.0000000000				
17	IT'S	218	0.26	1	16	0.07	34.96	1.83	0.0000000004				
18	B	75	0.09	1	0		34.87	138.87	0.0000000006				
19	DIRECTOR	74	0.09	1	0		34.40	138.85	0.0000000016				
20	DON'T	183	0.22	1	12	0.05	32.67	2.00	0.0000000080				
21	RSTNER	67	0.08	1	0		31.15	138.70	0.0000000210				
22	JUDGE	92	0.11	1	2		29.71	3.59	0.0000000474				
23	CAN	194	0.23	1	15	0.07	29.46	1.76	0.0000000542				
24	PRIEST	63	0.07	1	0		29.29	138.61	0.0000000594				
25	BUSINESSMAN	79	0.09	1	1		29.12	4.37	0.0000000650				
26	ASKED	199	0.24	1	16	0.07	28.94	1.70	0.0000000718				
27	DOORKEEPER	61	0.07	1	0		28.36	138.57	0.0000000978				
28	THAT'S	160	0.19	1	11	0.05	27.35	1.93	0.0000001667				
29	CASE	74	0.09	1	1		26.93	4.28	0.0000002083				
30	YOU'RE	120	0.14	1	6	0.03	26.42	2.39	0.0000002716				
31	HAS	103	0.12	1	4	0.02	26.33	2.75	0.0000002851				
32	WOMAN	70	0.08	1	1		25.18	4.20	0.0000005198				
33	IVE	100	0.12	1	4	0.02	25.17	2.71	0.0000005231				
34	BACK	149	0.18	1	82	0.37	-26.73	-1.07	0.0000002309				
35	WOULD	441	0.52	1	187	0.85	-28.48	-0.70	0.0000000916				
36	FLAT	10	0.01	1	20	0.09	-29.38	-2.93	0.0000000566				
37	FURNITURE	3		1	15	0.07	-32.37	-4.26	0.0000000098				
38	BODY	9	0.01	1	24	0.11	-41.02	-3.35	0.0000000000				
39	CLERK	19	0.02	1	34	0.15	-46.63	-2.77	0.0000000000				

Concordances

Concord

File Edit View Compute Settings Windows Help

N	Concordance	Set	Tag	Word #		Sent		Para		Sect		File	Date	%
				#	Pos	#	Pos	#	Pos	#	Pos			
1	everything and stayed there immobile. "Gregor!" shouted his sister, glowering at			3,278	78	2	0	77		0	77	morphosis.txt	May/08 00:00	60%
2	, and before she even realised it was Gregor that she saw screamed. "Oh God			3,251	76	34	0	50		0	50	morphosis.txt	May/08 00:00	60%
3	go back in the living room for a while?" Gregor could see what Grete had in mind			3,176	72	2	0	75		0	75	morphosis.txt	May/08 00:00	59%
4	: "Mother's fainted, but she's better now. Gregor got out." "Just as I expected",			3,628	88	2	0	27		0	27	morphosis.txt	May/08 00:00	61%
5	the door shut with her foot. So now Gregor was shut out from his mother,			3,448	81	4	0	47		0	47	morphosis.txt	May/08 00:00	60%
6	salts to bring her mother out of her faint; Gregor wanted to help too - he could			3,324	80	23	0	23		0	23	morphosis.txt	May/08 00:00	60%
7	of Gregor, he might have made her ill, so Gregor hurried backwards to the far end			2,695	56	19	0	94		0	94	morphosis.txt	May/08 00:00	57%
8	. His mother was not used to the sight of Gregor, he might have made her ill, so			2,687	56	11	0	86		0	86	morphosis.txt	May/08 00:00	57%
9	drawers, groaning, out of the room than Gregor poked his head out from under			2,610	54	18	0	09		0	09	morphosis.txt	May/08 00:00	57%
10	looked around. Her eyes met those of Gregor on the wall. Perhaps only			3,127	70	7	0	26		0	26	morphosis.txt	May/08 00:00	59%
11	picture at least, now totally covered by Gregor, would certainly be taken away by			3,054	66	10	0	53		0	53	morphosis.txt	May/08 00:00	59%
12	, and then went back out to Grete. Gregor kept trying to assure himself that			2,747	60	2	0	46		0	46	morphosis.txt	May/08 00:00	57%
13	of a chase as everything went so slowly. Gregor remained all this time on the floor,			4,251	04	2	0	50		0	50	morphosis.txt	May/08 00:00	64%
14	nonetheless lifted his feet unusually high. Gregor was amazed at the enormous			4,157	01	2	0	56		0	56	morphosis.txt	May/08 00:00	64%
15	look of determination, walked towards Gregor. He probably did not even know			4,137	99	50	0	36		0	36	morphosis.txt	May/08 00:00	64%
16	it, hit squarely and lodged in his back; Gregor wanted to drag himself away, as			4,525	12	15	0	24		0	24	morphosis.txt	May/08 00:00	65%
17	another one immediately flew at him; Gregor froze in shock; there was no			4,428	08	13	0	27		0	27	morphosis.txt	May/08 00:00	65%
18	onto the wall or ceiling. Whatever he did, Gregor had to admit that he certainly			4,282	05	5	0	81		0	81	morphosis.txt	May/08 00:00	64%
19	the hall, could see straight away that Gregor had the best intentions and would			3,743	92	32	0	42		0	42	morphosis.txt	May/08 00:00	62%
20	for some act of violence. That meant Gregor would now have to try to calm his			3,685	91	4	0	84		0	84	morphosis.txt	May/08 00:00	62%
21	wouldn't listen, would you." It was clear to Gregor that Grete had not said enough			3,654	90	6	0	53		0	53	morphosis.txt	May/08 00:00	61%
22	up tightly in his overcoat between Gregor and his mother, would always			3,978	97	91	0	77		0	77	morphosis.txt	May/08 00:00	63%
23	laying there entombed in his bed when Gregor came back from his business			3,904	97	17	0	03		0	03	morphosis.txt	May/08 00:00	62%
24	both angry and glad at the same time. Gregor drew his head back from the door			3,813	94	2	0	12		0	12	morphosis.txt	May/08 00:00	62%
25	chest of drawers was something that Gregor could do without if he had to, but			2,579	53	9	0	78		0	78	morphosis.txt	May/08 00:00	57%
26	exactly how everything looked, what Gregor had eaten, how he had behaved			1,257	18	47	0	56		0	56	morphosis.txt	May/08 00:00	51%
27	. She left the sheet where it was. Gregor even thought he glimpsed a look			1,127	15	2	0	26		0	26	morphosis.txt	May/08 00:00	50%
28	clear enough that it was no pleasure for Gregor to cut himself off so completely,			1,113	13	33	0	12		0	12	morphosis.txt	May/08 00:00	50%
29	made her call out: "Let me go and see Gregor, he is my unfortunate son! Can't			1,328	21	22	0	27		0	27	morphosis.txt	May/08 00:00	51%
30	sister at first persuaded her against it. Gregor listened very closely to all this,			1,298	20	2	0	97		0	97	morphosis.txt	May/08 00:00	51%
31	His mother also wanted to go in and visit Gregor relatively soon but his father and			1,284	19	11	0	83		0	83	morphosis.txt	May/08 00:00	51%
32	was able to do so, but as time went by Gregor was also able to see through it all			0,700	03	39	0	99		0	99	morphosis.txt	May/08 00:00	48%
33	pane of the window open from then on. If Gregor had only been able to speak to			0,623	02	3	0	22		0	22	morphosis.txt	May/08 00:00	48%
34	began to talk of the need to earn money, Gregor would always first let go of the			0,375	97	13	0	74		0	74	morphosis.txt	May/08 00:00	47%
35	had threatened her and tried to bite her. Gregor went straight to hide himself			0,962	10	2	0	61		0	61	morphosis.txt	May/08 00:00	49%
36	room would have been no surprise for Gregor as it would have been difficult for			0,911	09	17	0	10		0	10	morphosis.txt	May/08 00:00	49%
37	deeply for a little while. She would alarm Gregor twice a day with this running			0,793	07	5	0	92		0	92	morphosis.txt	May/08 00:00	49%
38	the way it was before so that when Gregor comes back to us again he'll find			2,142	39	20	0	41		0	41	morphosis.txt	May/08 00:00	55%
39	, quietly, almost whispering as if wanting Gregor (whose whereabouts she did not			2,067	38	9	0	66		0	66	morphosis.txt	May/08 00:00	54%

concordance collocates plot patterns clusters timeline filenames source text notes

Dispersion Plots

Concordance Plot list (unsaved)

File Edit View Compute Settings Windows Help						
N	File	Words	Hits per 1,000 person		Plot	
1	Gregor Metamorphosis	2,099	199	9.00	0.902	

KeyWords Plot list (unsaved)

File Edit View Compute Settings Windows Help						
N	Key word	arsion	yness	Link tokens	Hits	Plot
1	Overall			0	885	
2	GREGOR	.902	6.07	7	199	
3	GREGOR'S	.777	11.46	6	99	
4	SISTER	.816	2.02	5	96	
5	FATHER	.812	4.54	6	96	
6	MOTHER	.730	7.98	10	82	
7	SAMSA	.226	6.97	2	34	
8	PARENTS	.723	11.80	0	26	
9	FAMILY	.740	7.52	1	31	
10	GRETE	.456	11.45	3	22	
11	CHIEF	.429	5.23	1	37	
12	CLERK	.440	6.63	2	34	
13	FOOD	.699	4.05	1	14	
14	COUCH	.569	11.63	1	17	
15	BODY	.778	11.02	1	24	
16	VIOLIN	.091	7.75	0	12	
17	FURNITURE	.223	2.37	2	15	
18	FLAT	.749	9.38	1	20	
19	BOSS	.574	8.31	0	9	
20	CLEANER	.000	8.31	1	9	

Collocates

Collocate List (unsaved)

File Edit View Compute Settings Windows Help

N	Word Set	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	GREGOR	1	201	1	1					1	199	1				
2	TO	1	72	25	47	5	5	2	8	5		6	10	14	11	6
3	THE	1	65	26	39	7	6	7	6			2	9	6	10	12
4	HIS	1	42	13	29	2	1	6	4				13	8	2	6
5	HAD	1	42	7	35	2	5					22	4	4	2	3
6	AND	1	38	19	19	5	3	2	3	6		6	3	2	3	5
7	THAT	1	38	23	15	1	1	2	4	15		4	4	2	3	2
8	WAS	1	37	14	23	3	1	7	2	1		13	1	3	4	2
9	OF	1	25	17	8	5	2	3	2	5					4	4
10	HE	1	24	2	22		1		1			4	5	6	4	3
11	IN	1	24	14	10	3	4	5	2			2	4		2	2
12	IT	1	20	11	9	1	4	3	3				4	3		2
13	AS	1	19	9	10	2	1	3		3		6	1	1	1	1
14	WOULD	1	17	4	13	1		1	1	1		10	1	1	1	
15	BUT	1	15	9	6	1	2	1		5			2	3	1	
16	A	1	15	5	10		2	3					4	1	2	3
17	HAVE	1	14	5	9	4			1			2	2	4	1	
18	FROM	1	13	2	11		2						1	3	2	5
19	HIMSELF	1	13	0	13								5	3	2	3
20	AT	1	13	7	6	1	1		2	3			1	2	2	1
21	NOT	1	12	5	7	1	2		2				4	1	1	1
22	SO	1	12	7	5	1		1	3	2		1	1		2	1
23	SHE	1	11	7	4	2	3	2					2	1		1
24	SAID	1	11	7	4					7		3				1
25	OUT	1	11	2	9	1	1					1	3	2	1	2
26	WITH	1	11	4	7	1	3					2	1		2	2
27	NO	1	10	5	5	2		2	1					3		2
28	HER	1	10	6	4	2	1	2	1						1	3
29	HIM	1	10	6	4	1	1	1	2	1				1	1	2
30	BACK	1	10	3	7				2	1		2	2	1	2	
31	ALL	1	9	3	6		1	1	1				2	1	1	2
32	BEEN	1	9	4	5		2	2					1	2		2
33	THIS	1	9	5	4	1	1	1	1	1				2		2
34	SEE	1	9	5	4	1	1	1		2			2			2
35	FOR	1	9	7	2		1	1	1	4		1			1	
36	BY	1	9	6	3			1		5					1	2
37	WAY	1	8	3	5		1	1		1				1	2	2
38	MOTHER	1	8	6	2	2	1	1	2					1	1	
39	DID	1	8	3	5	1				2		3	1		1	

concordance collocates plot patterns clusters timeline filenames source text notes

Semantic Patterns

Concordance Pattern List (unsaved)

File Edit View Compute Settings Windows Help

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5	
1		THE	TO	THE	THE	THAT	GREGOR	HAD	HIS	TO	TO	THE
2		TO	THE	HIS	TO	SAID		WAS	THE	HIS	THE	HIS
3		OF	IN	WAS	HIS	AND		WOULD	TO	THE	HE	TO
4		AND	AND	IN	IT	WHEN		TO	IT	HE	WAS	HE
5		HAD	HAD	TO	THAT	TO		AND	HE	HAD	THAT	AND
6		AS	WITH	AND	AND	OF		AS	HIMSELF	HAVE	OF	FROM
7		HIS	BUT	HER	OF	BY		COULD	HAD	HIMSELF	ROOM	HAD
8		IN	IT	OF	SO	BUT		HE	THAT	WAS	AND	DRWARD
9		HAVE	OF	IT	METHING	WHILE		THAT	IN	NO	TIME	OF
10		SHE	SHE	THAT	MOTHER	IT		SAID	NOT	IT	ONE	HIMSELF
11		BUT	FROM	FOR	HER	FOR		IN	ONLY	BUT	ON	HER
12		OR	ME	AS	IN	NOW		DID	OUT	FROM	SO	WAS
13		WAS	BE	SHE	THIS	DOOR		WENT	INTO	AT	WAY	BY
14		AT	EVEN	WAY	FERENT	AS		WANTED	NEVER	THERE	WITH	ON
15		BEEN	NOT	WANTED	BACK	AT		SLOWLY	AND	OVER	THERE	THAT
16		NO	ROOM	SEEMED	WHY	HIM		WITH	STILL	THAT	THEY	OUT
17		HER	HOUGHT	FAMILY	WAS	DID		THOUGH	SEE	THIS	NOW	ONLY
18		USED	USED	DOWN	AT	TIME		THE	SHE	UP	DO	SEE
19		WORDS	THEN	BEEN	SOON	THOUGH		EALISED	ABOUT	AGAIN	FROM	WITH
20		WITH	SISTER	HIM	NOT	WAS		GREGOR	YOU	WHEN	GET	WITHOUT
21		LET	THAT	NO	HIM	VOICE		HAVE	ALL	AND	BEING	WORK
22		MOTHER	BEEN	LOOKED	MOMENT	DWARDS		EVEN	HOUGHT	HEAD	AS	WAY
23		HIM	DOOR	LOOK	EVENING	HER		BACK	MORE	DOOR	AT	THEY
24		DOOR	EVER		DRWARD	IF		CAME	BUT	FIRST	BACK	THINK
25			ALL		ON	DRIVE		HEARD	HAVE	FAMILY	IN	THIS
26			AT			GREGOR		ON	CALLED	HOW	MOVED	DAY
27			BACK			SO		ONLY	AT	OTHER	NOT	DRWAY
28			FIRST			THERE		MADE	PUSHED	OUT	HIS	HEARD
29			HER			MADE		EARNED	BE	BEEN	HAD	BEEN
30			HIS			SEE		ISTENED	BACK	BE	HEAR	ABOUT
31			HOW								HIMSELF	ALL
32			FOR									AWAY
33			GO									IT
34			HE									MUCH
35												NO
36												INTO
37												HIM
38												I'M
39												IN

concordance collocates plot patterns clusters timeline filenames source text notes

Clusters or n-grams

Concordance Cluster List (unsaved)

File Edit View Compute Settings Windows Help

N	Cluster	Freq.	Set	Length	Related
1	HE HAD BEEN	16		3	
2	HE DID NOT	14		3	
3	HE WOULD HAVE	12		3	
4	THAT HE HAD	12		3	
5	THAT HE COULD	10		3	
6	AS IF HE	10		3	
7	SO THAT HE	9		3	
8	HE HAD TO	9		3	
9	THAT HE WAS	9		3	
10	HE WANTED TO	8		3	
11	WHERE HE WAS	8		3	
12	HE COULD NOT	8		3	
13	HE WAS NOT	8		3	
14	HE HAD NOT	7		3	
15	AND HE WAS	7		3	
16	HE BEGAN TO	6		3	
17	AND IF HE	6		3	
18	AND HE WOULD	6		3	
19	HE WAS IN	6		3	
20	THAT HE WOULD	6		3	
21	WHEN HE HAD	6		3	
22	WHAT HE WAS	5		3	
23	WHEN HE WAS	5		3	
24	HE FELT THAT	5		3	
25	HAD REACHED THE	5		3	
26	BUT HE HAD	5		3	
27	HE WOULD NOT	5		3	
28	HE SAID TO	5		3	
29	HE LAY THERE	5		3	

Metaphors

- Corpus analysis of metaphors, when language is not literal but figurative
- Involves in-depth manual assessment of concordances
- Metaphors are figures of speech in which a word or phrase is applied to an object or action to which it is not literally applicable.
- Metaphor involve talking, and potentially thinking, about one thing in terms of another.

Conceptual Metaphors

- **“Metaphors we Live by”** (Lakoff & Johnson, 1980)
- Human thought is largely metaphorical / image schematic:
 - **TIME IS MONEY** “your are wasting your time”
 - **ARGUMENT IS WAR** “I demolished his argument”
 - **HAPPY IS UP, SAD IS DOWN** “I am feeling up”, “I am feeling low”
 - **IDEAS ARE PLANTS** “an idea come to fruition”
 - **LIFE IS A JOURNEY** “He had a head start in life”
 - **LOVE IS A JOURNEY** “The relationship has hit a dead-end street

LOVE IS A JOURNEY

Source domain JOURNEY —> Target domain LOVE

Travelers —> lovers

Destination —> goals

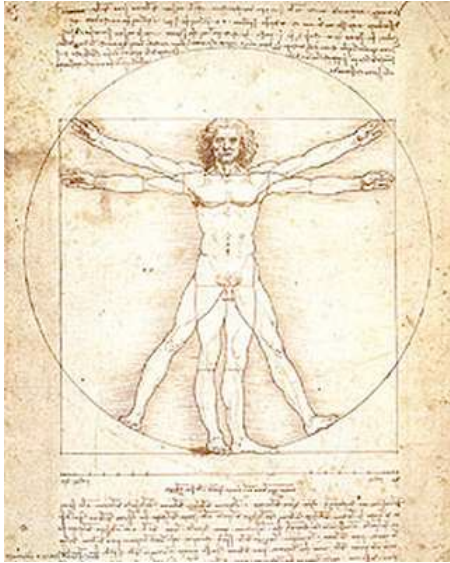
Vehicle —> relationship

Metaphor Analysis

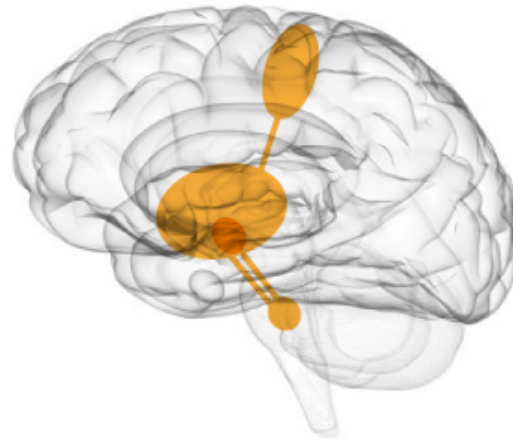
- Metaphor Identification Procedure (MIP) (Pragglejazz Group, 2007)
- Procedural steps to identify metaphors:
 - 1. Read the entire text–discourse to establish a general understanding of the meaning.
 - 2. Determine the lexical units in the text–discourse.

Metaphor Analysis

- 3. (a) For each lexical unit in the text, establish its meaning in context.
- 3. (b) For each lexical unit, determine if it has a more basic contemporary meaning.
- 3. (c) If the lexical unit has a more basic current–contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it.
- 4. If yes, mark the lexical unit as metaphorical.



EMBODIMENT CIRCUITRY



Summary of Analytic Steps

Analysis of one text (single corpus approach)

1. Conduct a word list to explore word frequencies in your corpus
2. Explore the concordances of 'interesting' words to identify their function and meaning in context
3. Clusters provide further analytic insight

Summary of Analytic Steps

Analysis of one corpus compared to a reference corpus:

1. Conduct a word list to explore word frequencies in your corpus
2. Compare key words that identifies salient words and their frequencies in your corpus compared to the reference corpus
3. Explore the concordances of 'interesting' words to identify their function and meaning in context

Software Packages

- Wordsmith (£60) standalone software
<http://www.lexically.net/wordsmith/>
- SketchEngine (free trial/ £8.51 per month) web-based
<https://www.sketchengine.co.uk/>
- AntConc (free) web-based
<http://www.laurenceanthony.net/software.html>
- Wmatrix (£50 per year) web-based
<http://ucrel.lancs.ac.uk/wmatrix/>

Books for Further Reading

- Baker, P., Hardie, A. & McEnery, A.(2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Baker, P. (2006) *Using corpora in discourse analysis*. London: Continuum.
- McEnery, A. & Wilson, A. (1999).*Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, A. & Hardie, A. (2012) *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2003). *Reading concordances: an introduction*. London: Pearson Education.

Further Resources

Journals:

- Corpora

<http://www.euppublishing.com/loi/cor>

- ICAME <http://clu.uni.no/icame/journal.html>

- International Journal of Corpus Linguistics

<https://benjamins.com/#catalog/journals/ijcl/main>

Google Scholar, Library Alerts and specialised subject journals