



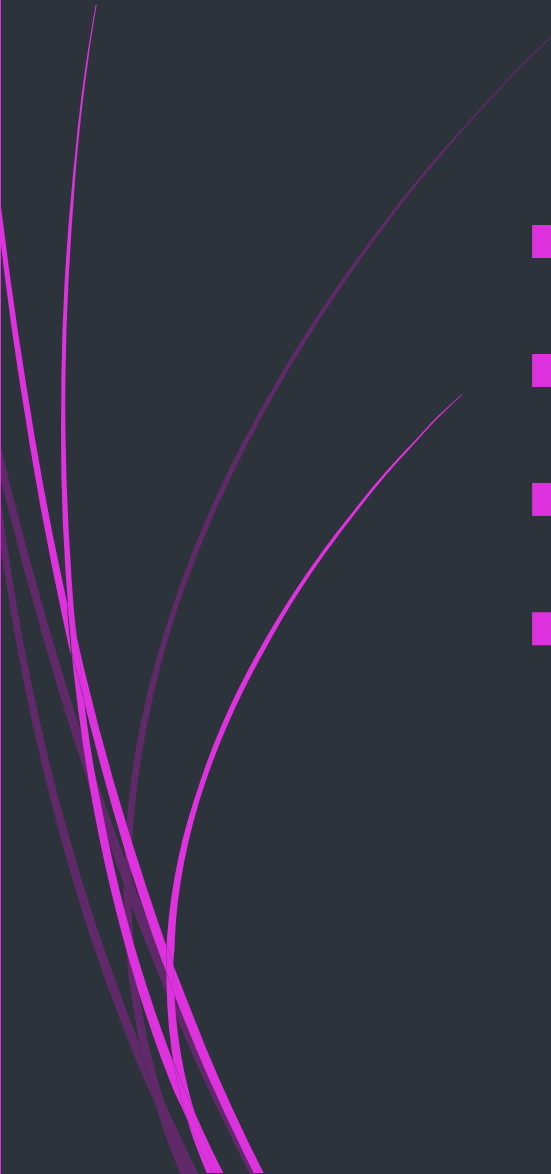
Using Corpora to Create Academic Writing Assignments

Sarah Deutchman

Waseda University



Agenda

- Overview of terminology
 - Creating vocabulary lists
 - Formulaic language
 - Argument Frames
- 



Corpora

- ▶ Corpora can be defined as "a collection of naturally occurring, computer-readable texts, often comprising many millions of words, which is considered more or less representative of a particular domain of language use" (Hyland, 2013, p. 248).



Formulaic Sequences

- It is a broad term that covers various lexical units including idioms, proverbs, collocations, lexical bundles, and conventionalized expressions (Schmitt & Carter, 2004).
- Colligations can be defined as the environment in which a word would be used, what words are connected to it, and what the grammatical patterns are (Sinclair 1996, 1999, 2004; Hoey 1997a, 1997b, as cited in Hoey & O'Donnell, 2008).
- Collocations can be defined as what words co-occur together or word associations (Hoey, 2014).



Connection of corpora to academic writing

- Understanding of multiword expressions define the writer's linguistic fluency (Hyland, 2012).
- Being able to choose certain phrases over others helps to control academic discourse (Hyland, 2012).
- More experienced writers use formulaic language.
- Learner corpus can be used to check student's errors.



The Assignment

- ▶ With computers tracking the Web sites we visit and surveillance cameras filming us in many public places, many people feel that their right to privacy has been violated. Do you think that this kind of monitoring is justifiable? Support your position with reasons and examples from your own experience, observations, or reading.
- ▶ What would a student need to be able to complete the assignment?
- ▶ Vocabulary, Grammatical Patterns, Argumentative Essay Organization, Help with arguments



Vocabulary

- What key terms would likely be used in this essay?
- Surveillance, surveillance camera, tracking cookie, privacy, monitoring

Creating A Vocabulary List

- ▶ It is possible to create corpora based on a certain assignment and then create a searchable list based on parts of speech

The screenshot shows a web interface for creating a vocabulary list. At the top, there are navigation tabs: "List" (highlighted with a blue box), "Chart", "Word", and "Browse +". Below the tabs is a search input field containing the word "monitor" and a "[POS]?" label. To the left of the input field is a small blue square icon. Below the input field is a "Reset" button. Underneath, there are three tabs: "Sections" (with an unchecked checkbox), "Texts/Virtual" (highlighted with a grey background), and "Sort/Limit Options". On the left side, there is a scrollable list with the following items: "FIND TEXTS", "MY CORPORA", a dashed line separator, "climate change", "global", and "leadership". On the right side, there are four buttons: "Create corpus" (highlighted with a yellow background), "Edit corpora", "Find keywords", and "Refresh list".

Adverbs associated with monitor

MONITORVERB [9,029,987 WORDS, 5000 TEXTS] (0.9% OF TOTAL) NOUN VERB ADJ **ADV** N+N ADJ+N [\[ALL CORPORA\]](#) [SAVE LIS](#)

HELP	WORD (CLICK FOR CONTEXT)	FREQ	# TEXTS	SPECIFIC		ENTIRE CORPUS	EXPECTED
				FREQ	TEXTS		
1	SHARPLY	381	333	90	100	11,293	102.7
2	AGGRESSIVELY	220	202			6,796	61.8
3	UNUSUALLY	152	137			6,109	55.6
4	ABROAD	440	319			18,255	166.0
5	PRIVATELY	240	211			10,189	92.7
6	EARLY	1968	1467			84,149	765.3
7	WIDELY	724	632			32,222	293.0
8	FORMALLY	192	173			8,558	77.8

List Chart Word Browse +

surveillance [POS]?

Find matching strings Reset

Sections Texts/Virtual Sort/Limit Options

1

- IGNORE
-
- TV/MOVIES
- BLOG
- WEB-GENL
- SPOKEN
- FICTION
- MAGAZINE
- NEWSPAPER
- ACADEMIC

2

- IGNORE
-
- TV/MOVIES
- BLOG
- WEB-GENL
- SPOKEN
- FICTION
- MAGAZINE
- NEWSPAPER
- ACADEMIC

ANALYZE



Select entries below | Then click button to analyze text; see word definitions; find related phrases

NEW



SAVE



TRANSLATE



ANALYZE

1 2010 ACAD 8 am; Treatment of Children to notify bus riders of unacceptable sound levels (i.e. signal) visual surveillance or community outreach #Target behaviors Target behav

Can see which words are usually associated with surveillance based on frequency

EDIT TEXT	SAVE TEXT	<input checked="" type="radio"/> WORD	<input type="radio"/> PHRASE
FREQ RANGE	1-500	501-3000	> 3000
341 WORDS	30 %	15 %	24 %

CLICK ON ANY WORD BELOW FOR A FULL WORD SKETCH

#1: to **notify bus riders** of **unacceptable** sound levels (i .e., **signals**), **visual surveillance**, or community **outreach**. # Target **behaviors**. **Target behavior codes** referred to the **measurement**

#2:; out of about 250 ,000 **soldiers**, only 175 ,000 can be used in **patrol**, **surveillance**, and **battle missions**. **Apparently**, even this **diminished** figure **compared** to the **initial**

#3: from a **variety** of public **sources**, including the Bureau of the Census and the Surveillance, Epidemiology, and End Results Program (SEER). **Rural-urban continuum codes** are

#4: in **superficial hemorrhagic gastritis** which may **progress** during the first 24-48 **h** and requires **specific surveillance** 38. The **quantity** of **ingested corrosive agent** is **related** to the **accidental/voluntary pattern** of

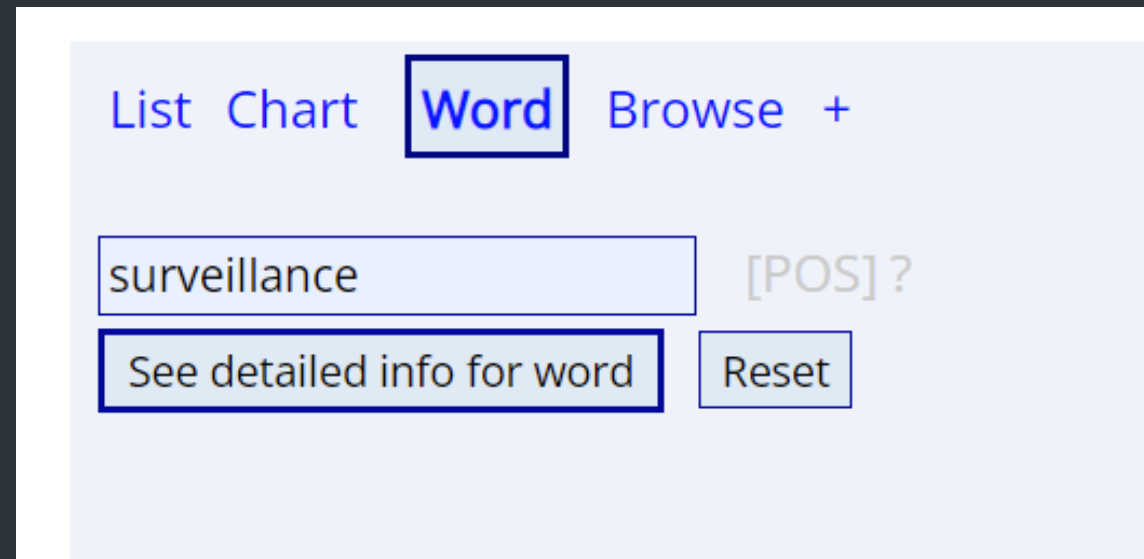
#5: **presentation**, **contained esophageal disruption**, and **minimal contamination** of

(CLICK ANY WORD FOR FULL WORD SKETCH)

LOW FREQ	MID FREQ	HIGH FREQ
<p>9: surveillance</p> <p>2: epub, infections, resistance</p> <p>1: accidental/voluntary, clonal, contamination, continuum, corrosive, criteria, diminished, disruption, dissemination, esophageal, gastritis, genome, genomic, hemorrhagic, ingested, invasive, isolates, labs, measurement, microbiology, minimal, neonatal, notify, outreach, pathogens, patrol, population-based, presentation, provision, quantity, riders, rural-urban, specialized, stable, stream, streptococci, superficial, unacceptable, visual</p>	<p>2: al, available, codes, et</p> <p>1: acknowledged, agent, apparently, battle, behavior, behaviors, blood, bridging, bus, compared, conducted, contained, data, developed, h, hidden, highly, hospital, hospitals, initial, measures, missions, pattern, period, population, prepare, progress, reduce, referred, related, signals, soldiers, sources, spaces, specific, surrounding, target, variety</p>	<p>16: the</p> <p>11: and, of</p> <p>7: to</p> <p>6: in</p> <p>5: for</p> <p>3: a, is</p> <p>2: are, from, public</p> <p>1: about, all, as, be, better, by, can, causing, centers, community, control, development, doi, during, even, figure, first, health, help, i, if, including, levels, may, model, one, only, or, out, over, part, program, project, requires, sound, their, this, turn, understand,</p>

Word Function

- ▶ Can see where and when the word is used. What words are associated with it.
- ▶ Links to translation sites.
 - ▶ Linguee (kanji corpora), Reverso(translated words in context), WordRef (English Japanese Dictionary)
- ▶ Links to video corpora.
 - ▶ Yarn(entertainment oriented), Youghish, Playphrase



surveillance (NOUN) #4115

BLOG WEB TV/M SPOK FIC MAG NEWS ACAD

1. close observation of a person or group

DMOCG[E]

YouGlish PlayPhrase Yarn

JA: Google WordRef Reverso Linguee

SYNONYMS NEW: DEFIN +SPEC +GENL

shadowing investigation, observation, reconnaissance, scrutiny, stakeout, surveillance, tailing

TOPICS (more)

intelligence, monitoring, privacy, outbreak, oversight, prevention, disease, infection, transmit, warrant, terrorist, enforcement, epidemic, terrorist, monitor, infect, laboratory, routinely, acute, drone

COLLOCATES (more)

NOUN camera, video, system, program, intelligence, tape, footage, data

VERB conduct, capture, monitor, release, install, obtain, review, drone

ADJ electronic, foreign, constant, domestic, active, increased, behavioral, secret

ADV under, allegedly, overhead, eg, improperly, prior, statewide, mechanically



Transitions and Connectors

- ▶ What transitions and connectors are most commonly used in an argumentative essay?
- ▶ Despite, Some people might say, on the other hand, while, whereas
- ▶ Additionally, Moreover, For example
- ▶ All of these can be looked up in COCA and analyzed with a KWIC

Results of 'in spite of' search

- In spite of + determiner+ adjective+ noun

Text	Phrase	Phrase
immigrant Argentine was irreversible ,	in spite of	a brief period during
faith in the cartridges they fire . This	in spite of	a cloud that once
up , just like their predecessors and "	in spite of	a developmental tendency
d decade of fire war , young girls . #	In spite of	a growing body of
gorical perception than NH children ,	in spite of	a hearing deprivation period .
scape from it . Rather , we show that	in spite of	a Malthusian constraint
ell density or neural crest migration ,	in spite of	a marked reduction in
ntation modeling . The latter is useful	in spite of	a small sample size
Americans agitated for political rights	in spite of	a state law that
ort to ensure a decentralized politics	in spite of	a stronger federal government ,
ground assumes a positive outcome .	In spite of	a terrible background ,
num , " of course , is that somehow ,	in spite of	all his precautions
m and those opposing it is not over .	In spite of	all regime attempts
ce-mixup . " " Race mixing " occurred	in spite of	all this , of course .

Academic Formula List(AFL)

- ▶ Patterns that reoccur in academic writing which is separated into written and spoken discourse (Simpson-Vlach & Ellis, 2010).
 - ▶ On the other hand, Despite the fact, It should be noted
- ▶ A list can be found on the AEP Foundation Website
 - ▶ <https://www.eapfoundation.com/vocab/academic/afl/>
- ▶ Can be used with COCA or SKELL
 - ▶ <https://skell.sketchengine.eu/#result?lang=en&query=despite%20the%20fact&f=concordance>



Academic Phrase Bank

- The academic phrase bank provides phrases to help with dissertation writing and other academic writing. The phrases are broken into separate headings based on purpose (Morley, 2014).
- For this essay ‘being cautious’ would be the most useful heading. The section ‘being cautious when giving explanations’, and ‘devices that distance the author from a proposition’ are useful.
- Highlighting the sections to look at could prevent students from being overwhelmed or these items could be added to the assignment sheet.
- <https://www.phrasebank.manchester.ac.uk/using-cautious-language/>



Modals

- Argumentative essays need to use cautious language
- Can be a source of confusion for students
- What is the difference between would and could?
- Possibility
- Sketch Engine for Language Learning can be used to show the words in context.



Brainstorm Arguments For/ Against

- What are some arguments for surveillance?
Examples?
- What are some arguments against surveillance?
Examples?

Key Arguments

For	Against
Control of crime (Does surveillance make it easier to catch criminals?)	Invasion of privacy (How can your personal information be used against you? Do you want companies to be able to read your text messages? Can people find where you live? (GPS))
Improved customer service (Amazon calculating user preferences 'You might like...') * See an ad on Facebook for something you bought on Amazon. (data profiles)	Costly (How expensive is a surveillance system?)
Ability to monitor and respond to national issues (e.g., the spread of coronavirus, disasters)- Track where confirmed cases have been and their contacts. (Make sure you can prove this though. Underreporting is a huge problem)	Issues of data safety (How might other people access your information? Leak your passwords?)



Other Things to Look Into

- Edward Snowden
- Watch lists
- Mass surveillance
- Right to privacy
- Metadata
- Breach of privacy
- Terrorism
- PRISM
- Phone spy (stalkerware)
- Tracking cookies
- Predictive data mining
- Cambridge Analytica
- Blackmail
- Deepfakes
- Facial recognition software
- Social Credit System (China)
- Big Brother (1984)
- Pegasus (Israeli spyware)



Argument Essay Frame

- ▶ Introduction
 - ▶ Define key terms (surveillance, privacy)
 - ▶ Explain why the reader should care about the issue/ Why the issue is debated
 - ▶ Thesis: Transition+ other side's position, your position + two predictors
- ▶ Body Paragraph 1
 - ▶ First supporting point
 - ▶ Give more details about the point
 - ▶ Example
 - ▶ How does this point prove your position?



Argument Essay Frame

- ▶ Body Paragraph 2
 - ▶ Second supporting point
 - ▶ Give more details about the point
 - ▶ Example
 - ▶ How does this point prove your position?
- ▶ Body Paragraph 3
 - ▶ Counter argument (one point from the other side)
 - ▶ Explain why that argument is not important or is not strong
 - ▶ Give example to prove your points
 - ▶ Restate why the counter argument wasn't strong or ways to address the counter argument



Argument Essay Frame

➤ Conclusion

- Restate the thesis
- Restate main points
- Explain why it was important to talk about the issue
- Add a prediction, recommendation, or final comment (no new information)



Conclusion

- ▶ It is important to understand vocabulary load when choosing essay topics
- ▶ If the essay topic cannot be chosen, then it is important to break down the topic in a way that helps students succeed.
- ▶ Planning step by step and providing sample essays can help students.
- ▶ Teaching how to search corpora can help students on future assignments.

References

- ▶ Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>
- ▶ Hoey, M. (2014). Words and Their Neighbours. In J.R. Taylor (Ed.), *The Oxford Handbook of the Word*. <https://doi.org/10.1093/oxfordhb/9780199641604.001.0001>
- ▶ Hoey, M., & O'Donnell, M. B. (2008). Lexicography, grammar, and textual position. *International Journal of Lexicography*, 21(3), 293–309. <https://doi.org/10.1093/ijl/ecn025>
- ▶ Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150- 169.
- ▶ Hyland, K. (2013). Corpora and innovation in English language education. In Hyland, K & Wong, L. (Eds.) *Innovation and change in language education* (pp 248-262). Routledge.
- ▶ Morley, J. (2014). *Academic phrasebank: A compendium of commonly used phrasal elements in academic English in PDF format*. The University of Manchester.
- ▶ Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1–22). John Benjamins. <https://doi.org/10.1075/llt.9>.
- ▶ Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.