

# Researching contemporary English with corpora

Joseph Pentangelo

Ling 301, Fall 2020

College of Staten Island

CAMERAS ON PLEASE



(IF YOU DON'T MIND)

# Review: Blog Post #4

- Sentences should be clear. Don't feel any pressure to use lofty words or complex sentence structures.
- Introductions should be concise. No need to begin by talking about the murky origins of language itself. Begin with the topic that you are going to be discussing.
- Your introduction should describe what the rest of your paper will do. How will you explore your topic?
- You can use the introduction in Franco & Tagliamonte (2020) as a really solid model.



# Franco & Tagliamonte (2020)

Whereas most adverbs in English are formed with *-ly* or  $-\emptyset$  morphemes, a (albeit limited) set of adverbs ends in *-s*, predominantly adverbs ending in *-ward(s)*, like *forward(s)*, *afterward(s)*, and *inward(s)*, or *-way(s)*, such as *anyway(s)* or *halfway(s)*. An interesting attribute of these endings is that they are potentially two suffixes in one.<sup>1</sup> In what follows, we will refer to *-ward(s)/-way(s)* as suffixes, though we are noncommittal as to whether the suffix is *-s* or *-ward(s)/-way(s)*.

What does  $-\emptyset$  mean?

What does (s) mean?

This introductory paragraph throws us right into the action: most adverbs end in *-ly* or nothing, but some adverbs end in *-s*, and often in *-wards* and *-ways*. But these are interchangeable with *-ward* and *-way*, which is interesting.



CAMERAS ON PLEASE

Both have long been variable, *-ward(s)* since Old English (OED s.v. *-s, suffix1*) and *-way(s)* somewhat later. Attested variation (1, 2, 3) can be found from Early Modern English (1). Most of these forms can also be used with adjectival meaning (4), and some of them have gained additional functions (5, 6).

1. Variation in Early Modern English (examples from OED s.v. *anyway*):
  - a. Appease and ende all strifes, that *anie waie* touch the state. (1593, Thomas Bilson, *The perpetual gouvernement of Christes*)
  - b. If *any waies* I haue erred vnto you..it was but..by ignoraunce. (1592, Angell Day, *The English secretorie*).
2. So you're going around the circle so you go *forward*, and then you have to switch to *backwards* (F, 19, Toronto)
3. But *anyways*, he was dead there, *anyway* (M, 86, Swords)
4. It's *backwards* engineering, if you want to talk about that. (M, 33, Kapuskasing)
5. *toward(s)* as a preposition:  
Generally speaking, our orientation was out *toward* Clayton. We didn't come much to Perth. (M, 65, Almonte)
6. *anyway(s)* as a discourse-pragmatic (DP) marker (also see Ferrara, 1997):  
And ah my neighbourhood was ah suburban but there was still a fair amount of crime. Like I- I lived close to the- oh, I can't even remember the name of the mall. *Anyway*. But there was a- there was a fair amount of petty crime, and- and even some fairly violent crime. (M, 39, Kapuskasing)

Previous literature generally does not agree on the use of *-s* versus  $\emptyset$ . On the one hand, scholars like Bauer, Lieber, and Plag (2013:328) argued that functional differentiation plays a role, with  $\emptyset$  making an adjectival reading of the word available. On the other hand, other studies indicate that differences between varieties are more important (e.g., British versus American English, cf., Huddleston & Pullum [2002:615, 688, fn. 8]; Quirk, Greenbaum, Leech, & Svartvik [1985:667]), or even that the use of *-s* or  $\emptyset$  is "at the whim of the speaker" (Dixon, 2014:376). However, a systematic study of the use of these adverbs in spoken vernacular language is, to our knowledge, lacking. In this paper, we focus on data from contemporary Ontario English, mostly collected between 2003 and 2018, but with some historical data that were collected in the 1970s, 1980s, and 1990s, offering an apparent-time perspective for over 100 years. We will argue that the forms in *-ward(s)* and *-way(s)* have evolved differently in the last century, with linguistic and social factors playing contrasting roles depending on the suffix and on the word.

This paper is structured as follows. First, we provide an overview on adverbial *-s*. It will show that the literature is unclear about the variable use of this morpheme. Then, we describe the data and methods used in the analysis, followed by the results. Finally, we discuss and explain our findings in light of variationist sociolinguistic theory and principles of language change.

# Franco & Tagliamonte (2020)

- This paper has an excellent introduction.
- It is clear and concise (only a page and a half).
- Gives a review of the topic and sets out, in specific detail, how the research was carried out, and how this paper will be structured.



CAMERAS ON PLEASE

# Franco & Tagliamonte (2020)

- One section on historical forms, remainder of the paper on contemporary use in Ontario.
- How did they do their historical research?
- How did they do their contemporary research?
- What is a corpus?



# Corpus as a research tool

- A corpus is a compilation of texts that you can query. The plural of corpus is corpora.
- Franco & Tagliamonte (2020: 222):

The data come from the Ontario Dialects Project. [...] The data were collected in diverse locations in the province (Figure 1), from the largest urban center of the region, Toronto, to a multitude of smaller communities in the north and southeast of Ontario. The data were collected between 2003 and 2018 by means of sociolinguistic interviews with locals, although there are also diachronic materials available for some communities, offering an apparent-time perspective for over 100 years. In total, the extent of the archive at the time of writing comprised 18 communities, 1087 individuals, and 11,034,676 words (June 2019).

What does diachronic mean?



CAMERAS ON PLEASE



# Corpus as a research tool

- Franco & Tagliamonte queried the corpus for - $\emptyset$  and -s versions of -ward(s) and -way(s) words, and tallied the number of **tokens** for each **type**.

**type:** a given word

**token:** an instance of a given word

TABLE 3. *Distribution of -s and - $\emptyset$  per word (words listed by decreasing frequency of -s)*

word type	-s		- $\emptyset$		total <i>n</i>
	<i>n</i>	%	<i>n</i>	%	
<i>toward(s)</i>	527	91%	52	9%	579
<i>backward(s)</i>	95	88%	14	13%	109
<i>downward(s)</i>	4	80%	1	20%	5
<i>outward(s)</i>	3	75%	1	25%	4
<i>upward(s)</i>	4	67%	2	33%	6
<i>onward(s)</i>	2	50%	2	50%	4
<i>anyway(s)</i>	1481	23%	4904	77%	6385
<i>forward(s)</i>	11	13%	71	87%	82
<b>total</b>	<b>2127</b>	<b>30%</b>	<b>5074</b>	<b>70%</b>	<b>7174</b>

- How many tokens are there for the type *outwards*?
- How many tokens are there for the type *upward*?







# The Corpus of Contemporary American English

- We will be using the Corpus of Contemporary American English (COCA), available from [english-corpora.org](http://english-corpora.org)
- COCA has over 1 billion words with sources from 1990 to 2019, drawn from a range of written sources.
- Excellent for doing research on contemporary English.
- Unfortunately, its interface is very confusing.
- \* = wildcard (anything or nothing)
- ? = one blank letter



# Spill the tea

 **Corpus of Contemporary American English**           

SEARCHFREQUENCYCONTEXTOVERVIEW

**List** Chart Word Browse +

spill the tea [POS] ?

Find matching strings


Reset

☐ Sections Texts/Virtual Sort/Limit Options

 (HIDE HELP)

LOGGED IN

[Take a guided tour of English-Corpora.org \(Nov 2020\)](#)













 Download the corpus (and corpus-based frequency data) for offline use

The Corpus of Contemporary American English (COCA) is the only [large](#), genre-balanced corpus of American English. COCA is probably the [most widely-used corpus of English](#), and it is related to many other [corpora of English](#) that we have created, which offer unparalleled insight into [variation in English](#).






The corpus contains more than [one billion words](#) of text (25+ million words each year 1990-2019) from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, and (with the [update in March 2020](#)): TV and Movies subtitles, blogs, and other web pages.

Click on any of the links in the search form to the left for context-sensitive

# Spill the tea

 **Corpus of Contemporary American English**           

SEARCHFREQUENCYCONTEXTOVERVIEW

ON CLICK: [CONTEXT](#)  [TRANSLATE \(??\)](#)  [GOOGLE](#)  [IMAGE](#)  [PRON/VIDEO](#)  [BOOK](#) (HELP) WORD PROFILES: [SPILL](#) [TEA](#)

HELP	<input type="checkbox"/>	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	
1	<input type="checkbox"/>	SPILL THE TEA		3	<div></div>
					0.375 seconds

# Spill the tea

Corpus of Contemporary American English

SEARCH FREQUENCY CONTEXT OVERVIEW

(SHUFFLE)

CLICK FOR MORE CONTEXT ☐ [?] **SAVE LIST** CHOOSE LIST  CREATE NEW LIST  [?] SHOW DUPLICATES

1	2013	MAG	Atlantic	A	B	C	They are chocolate cakes, frosted doughnuts, cookies. It's very easy to <b>spill the tea</b> when you pour or take a sip, a feature added based on
2	2012	BLOG	straightfromthea.com	A	B	C	her with a black woman. Heard it might be that Porsche Stewart chick-Michelle, <b>spill the tea</b> , plz and ty! # You say that Kim's stepping out
3	2008	FIC	Bk:RedWolfConspiracy	A	B	C	He cleared his throat with a painful noise -- CHHRCK! -- nearly making Pazel <b>spill the tea</b> . The man had an appetite, too: four biscuits vanished into

- Note how important context is. Does “spill the tea” mean “gossip” in all of these citations?
- Also... what about “spill your tea” or “spill some tea”? Do we need to do separate searches for all of these variations?

# Spill \* tea

 **Corpus of Contemporary American English**           

SEARCHFREQUENCYCONTEXTOVERVIEW

ListChartWordBrowse +

[POS] ?

☐ Sections Texts/Virtual Sort/Limit Options

(HIDE HELP)

LOGGED IN

v 2020)

d frequency data) for offline

use

The Corpus of Contemporary American English (COCA) is the only **large**, genre-balanced corpus of American English. COCA is probably the **most widely-used corpus of English**, and it is related to many other **corpora of English** that we have created, which offer unparalleled insight into **variation in English**.

The corpus contains more than **one billion words** of text (25+ million words each year 1990-2019) from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, and (with the **update in March 2020**): TV and Movies subtitles, blogs, and other web pages.

Click on any of the links in the search form to the left for context-sensitive

# Spill \* tea

 **Corpus of Contemporary American English**           

SEARCHFREQUENCYCONTEXTOVERVIEW

ON CLICK: [CONTEXT](#)  TRANSLATE (??)  GOOGLE  IMAGE  PRON/VIDEO  BOOK (HELP) WORD PROFILES: [SPILL](#) [TEA](#)

HELP	<input type="checkbox"/>	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	
1	<input type="checkbox"/>		SPILL THE TEA	3	<div></div>
2	<input type="checkbox"/>		SPILL MY TEA	2	<div></div>
3	<input type="checkbox"/>		SPILL HIS TEA	2	<div></div>
4	<input type="checkbox"/>		SPILL YOUR TEA	1	<div></div>
5	<input type="checkbox"/>		SPILL HER TEA	1	<div></div>
6	<input type="checkbox"/>		SPILL ANY TEA	1	<div></div>
			TOTAL	10	

0.391 seconds

# Fun with Dated Slang

In groups, you will use COCA to explore dated slang. Each group will answer the following questions:

1. When is the first time your search term shows up?
  2. When is the most recent time your search term shows up?
  3. What year or years was it used the most?
  4. What does it seem to mean?
- Be aware that there are often several pages of results
  - If you get a “Become a premium member!” message, just refresh the page.

## Groups

1. bling(-bling)
2. booya(h)
3. wig out (include other tenses, e.g. wiggling out, wiggled out)
4. IRL (meaning ‘in real life’)
5. chillax (include other tenses, e.g. chillaxing, chillaxed)



# Fun with Dated Slang

Report back:

- bling(-bling)
- booya(h)
- wig out
- IRL
- chillax

1. When is the first time your search term shows up?
2. When is the most recent time your search term shows up?
3. What year or years was it used the most?
4. What does it seem to mean?

# Blog Post #5

- Due by 11:59pm on Wednesday, December 2<sup>nd</sup> (two weeks from now)
- 300 words minimum.
- Use COCA, COHA, or another corpus on [english-corpora.org](http://english-corpora.org), to examine 2 or more words or expressions relevant to your final project. You should use a corpus that is relevant to the timeframe and area you are examining. In the post, describe what corpus you used, and what search terms you used. Define the words or expressions (either from your own knowledge, the OED, [urbandictionary.com](http://urbandictionary.com), or whatever source is appropriate). Describe when they first appear in the corpus, and in what sort of contexts they appear (e.g. blog posts, television, everywhere). Explain how these words/expressions tie into your research.

# Word of the Year

- Free to attend this year!
- Link on our website.
- You can submit your own entries.

