

# Scientific Writing in English: Techniques and Tools

Using Corpus of Contemporary American English  
(COCA)

Ethel Schuster, Ph.D.  
Northern Essex Community College, USA  
[eschuster@necc.mass.edu](mailto:eschuster@necc.mass.edu)

Ethel Schuster, Ph.D.

# Questions

- What makes a phrase or text sound natural?
- How to recognize a natural vs. an unnatural phrase/text?
- How to write natural/fluent text?
- What computational tools does Corpus Linguistics offers us?
  - Write better?

# Language

- Made up of prefabricated "chunks"
  - Fixed or semi-fixed
    - combinations of words
  - Master them
  - language sounds
    - natural
    - fluent

# Types of “Chunks”

- Colligations
- Collocations

# Colligations

- **Combinations of words**
  - in grammatical categories
  - e. g., verb, noun, preposition

# Examples

## 1. LOOK + preposition

- He was **looking at** component
- He was **looking for** the answer

## 2. BELIEVE + that clause

- He **believed that** he would survive

## 3. TRY + to infinitive

- He **tried to** run the marathon

# Collocations

- Words “strung” together as units
  - Will sounds “right” to native speaker
  - E. g., “dream + about/of”  
(instead of “dream + with”)

# Types of Collocations

- Verbal
- Nominal
- Adjectival
- Adverbial



# Example: Verbal

└─ Nouns can be subject or object of verb

1. N + V

“The **team identified** the problem...”

2. V + N

“Our group has done the research..”

# Example: Nominal

## └ Nouns as compounds

### 1. N + N

“There is little **research data** related to obesity and THR”

### 2. N + of + N

“From the patient's **point of view**, they are likely to have little expectation that you can solve their problem”

# Example: Adjectival

└ Adjectives combine with certain nouns

1. Adj + N

“They conducted a **longitudinal study** of 54 children from first through fourth grades”

“The **present/current study** employed both elicited and spontaneous language measures”

# Example: Adverbial

- Verbs and adjectives combine with certain adverbs

- Order is usually fixed

“He had a **fairly extensive** knowledge of the colonies and their operation”

“Although **widely acknowledged** among traditional cultures, sleep paralysis is one of the most ...”

# Exception: Order NOT Fixed

- Areas with overlapping high values could be **carefully studied** through risk assessment.
- The team **carefully studied** every part of the virus.
- I'm **carefully studying** my setup and becoming very familiar with it.
- This material should be approached with an open mind, **studied carefully** and critically considered.
- The question would have to be **studied carefully** first.

# Binomials

- Two words belonging to same grammatical category
- For example, verbs, nouns, adjectives, linked by conjunction or preposition

# Order?

- Reversible = order can be inverted  
*social or political or  
political or social*
- Irreversible = order is fixed  
*come and go or  
better or worse*

# Corpus Linguistics

- Based on assumption that language is probabilistic
  - certain combinations of words have higher probability of occurring than others, although these other ones may be perfectly grammatical



# Examples

- Why should *increase or decrease* be more recurrent than *decrease or increase*?
- Why is *results suggest* much more frequent than *results suggested* in academic language?
- Why is *fill the gap* preferred to *bridge the gap* in academic discourse although latter is recurrent in everyday language?


# Corpus of Contemporary American English (COCA)

- Corpus with over 460 million words, with texts ranging from 1990 to the current year (they are updated annually)
- Covered genres: from spoken language and fiction to journalistic and academic
- [www.americancorpus.org](http://www.americancorpus.org)

# Using COCA

- Main tool available on both sites is a **concordancer**
  - Brings you ALL occurrences in corpus of the word you are studying
  - Called KWIC (Keyword In Context) concordance, because word is shown in context

# Let's use it!

- 
- **Sample words**
  - **Suggestions**
  - **See tutorials**

# COCA

- Main tool available is **concordancer**
  - Shows ALL occurrences of word(s)
  - Called a KWIC (Keyword In Context) concordance, i.e., word shown in context
  - Larger portion of context obtained by clicking on concordance line

# Search: study

1. Enter word

2. Results

3. Frequency

4. Concordance lines

The screenshot shows the 'WORDS OF CONTEMPORARY AMERICAN ENGLISH' search interface. The search term 'study' is entered in the search box. The results table shows a frequency of 144875 for the word 'STUDY'. Below the search interface, a concordance table is displayed, showing the word 'study' in context across various sources and dates.

LINE	YEAR	ACAD	COMMUNICATIONS	A	B	C	CONTEXT	FREQ
1							STUDY	144875

LINE	YEAR	ACAD	COMMUNICATIONS	A	B	C	CONTEXT
1	2012	ACAD	Communications	A	B	C	buffering accumulated over the course of the past decade is finally sufficient to motivate systematic <b>study</b> . A 2007 study of nearly 2,000 hosts connect
2	2012	ACAD	Communications	A	B	C	course of the past decade is finally sufficient to motivate systematic <b>study</b> . A 2007 <b>study</b> of nearly 2,000 hosts connected through cable and DSL comp
3	2012	ACAD	Communications	A	B	C	or access link, has been key to the exposure of bufferbloat. A 2010 <b>study</b> of 130,000 measurement sessions revealed widespread, severe overbufferin
4	2012	ACAD	Communications	A	B	C	situation is clearly worse than shown. Focusing only on cable customers, the same <b>study</b> showed the equipment had two dominant buffer sizes: 128KB
5	2012	ACAD	Communications	A	B	C	easily copied. Dhamija et al. also found that even experienced participants in the <b>study</b> had trouble with picture-in-picture attacks showing screenshots
6	2012	ACAD	Communications	A	B	C	show whatever they want. Also in 2006, Downs et al.9 conducted a complementary <b>study</b> examining phishing email messages. As in Dhamija et al., 8
7	2012	ACAD	Communications	A	B	C	it would be safe to give it again. Sheng et al.30 conducted a followup <b>study</b> involving a large-scale survey examining demographics and phishing succe
8	2012	ACAD	Communications	A	B	C	the site's brand name (see Figure 3). However, a 2007 <b>study</b> by Jackson et al.18 found EV certificates ineffective in protecting people from phishing att
9	2012	ACAD	Communications	A	B	C	. Many Web sites offer advice on how to identify phishing sites. A 2010 <b>study</b> by Kumaraguru et al.24 found this kind of information useful in helping pe
10	2012	ACAD	Communications	A	B	C	fake Web sites but only if they actually read the material. In an earlier <b>study</b> , Kumaraguru et al.23 found that simply emailing anti-phishing material w
11	2012	ACAD	Communications	A	B	C	an intervention that teaches them about phishing and how to protect themselves. In a <b>study</b> with more than 500 participants, Kumaraguru et al. found
12	2012	ACAD	Communications	A	B	C	Cranor, I.f., and hong, j.i. you've been warned: an empirical <b>study</b> of the effectiveness of web browser phishing warnings. In Proceedings of the CHI C
13	2012	ACAD	Communications	A	B	C	interest, this feature is simply not visible. Thus, it is important to <b>study</b> how a visually impaired individual, who can not use feedback from the camera s
14	2012	ACAD	Communications	A	B	C	, s. and bagherinia, h. blind guidance using mobile computer vision: a usability <b>study</b> . In Proc. ACM SIGACCESS Conference on Computers and Access
15	2012	ACAD	Communications	A	B	C	study in which artists made line drawings intended to convey specific 3D shapes. The <b>study</b> was designed so that drawings could be registered with ren
16	2012	ACAD	Communications	A	B	C	the underlying differential properties of the 3D surface. Lines drawn by artists in this <b>study</b> largely overlapped one another, particularly along the ocul

# Search: study “co-occurs” with ?

1. Enter word
2. Enter category, e.g., any verb

Can be done with COLLOCATES

**CORPUS OF CONTEMPORARY AMERICAN ENGLISH**  
450 MILLION WORDS, 1990-2012

SEARCH STRING: study  
COLLOCATES: [+v]  
POS LIST: vb, V, N, P, R, S, A, C, J, F, O, Q, X, Y, Z, ., !, @, #, \$, %, ^, &, \*  
MINIMUM: SEARCH ASSET

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]

	CONTEXT	FREQ
1	IS	554
2	DID	414
3	CAN	354
4	CONDUCTED	348
5	WAS	337
6	WILL	307
7	SAYS	288
8	DO	252
9	USED	239
10	COULD	219
11	BEGAN	204
12	HAVE	195
13	WANT	191

KEYWORD IN CONTEXT DISPLAY

SECTION: NO LIMITS

CLICK FOR MORE CONTEXT	[?]	[SAVE LIST]	CHOOSE LIST	.....	CREATE NEW LIST	[?]
1	2012	MAG	OrganicGarden	A B C	massive disinvestment by grocery stores in this part of Chicago. A friend <b>conducted a study</b> in which he asked local children what color a banana was; based on wh...	
2	2012	NEWS	CSMonitor	A B C	a geophysicist at the University of London who led the British-Italian team that <b>conducted the study</b> , in a prepared statement. # Using boulders as benchmarks for a	
3	2012	ACAD	Communications	A B C	compared viewers' perceptions of hand-drawn versus computer-generated pen-and-ink illustrations. Phillips et al.19 <b>conducted a study</b> similar to ours, in which artist	
4	2012	ACAD	GeneralMusicToday	A B C	p. 235). A decade later, Thomas Goolbsy (1994b) <b>conducted a study</b> in which he found music perception to be necessary for accurate and "musical "	
5	2012	ACAD	LanguageSpeech	A B C	adolescence and into early adulthood. For example, McChes-Bilback (1992) <b>conducted a study</b> in which 120 students ages 10, 14, and 18 years were asked to	
6	2012	ACAD	LanguageSpeech	A B C	explore the understanding of ABNs, Nippold et al. (1999) <b>conducted a study</b> that involved four groups of students (n = 240) whose mean ages were	
7	2012	ACAD	Education	A B C	the individual spent his childhood, social mind and education. Yousef was <b>conducted a study</b> (1998) in which he reported that the decision style may be affected by	
8	2011	SPOK	Fox_Back	A B C	something that would make them move as a herd. So, he <b>conducted a study</b> of Americans eating habits and then he managed to find a doctor who concluded that	
9	2011	MAG	America	A B C	. What Should We Ask? Back in 1971, John N. Kotre <b>conducted a study</b> of 100 young Catholic adults. Fifty of these, by their own definition,	
10	2011	MAG	Cosmopolitan	A B C	not intentionally, says cognitive psychologist Mara Mather, PhD. She recently <b>conducted a study</b> at the University of Southern California that found that when men ar	
11	2011	MAG	Cosmopolitan	A B C	for Media and the Public Agenda at the University of Maryland, which <b>conducted the study</b> . Make regular gadget-free plans. For example, leave your phone in the ca	
12	2011	MAG	Prevention	A B C	, associate professor of crop and soil sciences at Penn State, who <b>conducted the study</b> . (Some producers of both organic and conventional eggs enrich their hens' fe	
13	2011	NEWS	Houston	A B C	to home, the University of Houston's Hobby Center for Public Policy <b>conducted a study</b> that clearly showed the positive impact historic designation has on property v	
14	2011	ACAD	NaturalHist	A B C	to biologist Joe Walston of WCS and his colleagues. The team recently <b>conducted a study</b> of forty-two sites where tigers are still found (so-called source sites); the	
15	2011	ACAD	TeachLibrar	A B C	) at all? In 2009, the United States Department of Education <b>conducted a study</b> that determined that online learning is more effective than F2F, but that hybrid instru	
16	2011	ACAD	CollegeStud	A B C	second year than non-first generation students. Phinney and Haas (2003) <b>conducted a study</b> on both perceived social support and the perceived lack of social suppo	

# Results?

- Show list of verbs that occur in specified position(s)
- Can identify *conducted* as frequent collocate, which would give us verbal collocation:

*conduct a study.*



# Results?

- Go down in the list, find
  - *Conduct* (position 20) with 117 occurrences
  - *conducted* shows 348 occurrences

# What we learn?

- This also tells us that *conducted a study* is twice as common as *conduct a study*

In other words, the collocation occurs more frequently in its **past** form

# Verbs + “study”

- See first 20 verbs that occur 1 or 2 positions to the right of *study*
  - *conducted* is among them

	CONTEXT	FREQ
1	<u>WAS</u>	<u>7175</u>
2	<u>IS</u>	<u>3464</u>
3	<u>WERE</u>	<u>2530</u>
4	<u>FOUND</u>	<u>2398</u>
5	<u>CONDUCTED</u>	<u>1487</u>
6	<u>HAS</u>	<u>1162</u>
7	<u>ARE</u>	<u>1109</u>
8	<u>SHOWED</u>	<u>927</u>
9	<u>PUBLISHED</u>	<u>882</u>
10	<u>BE</u>	<u>812</u>
11	<u>EXAMINED</u>	<u>737</u>
12	<u>REPORTED</u>	<u>714</u>
13	<u>HAD</u>	<u>674</u>
14	<u>DID</u>	<u>663</u>
15	<u>SHOWS</u>	<u>559</u>
16	<u>HAVE</u>	<u>549</u>
17	<u>WILL</u>	<u>538</u>
18	<u>SUGGESTS</u>	<u>530</u>
19	<u>SUGGEST</u>	<u>488</u>
20	<u>INCLUDED</u>	<u>466</u>

# Other verbs?

- Related to what the *study* investigated
  - what the *study* ...  
*showed/shows/ suggests*

>> Examine concordance lines for each verb to establish exact context in which collocation is customarily used

# NOTE

- Unexpected results occur
- E.g., adjectives + “investigation”?  
Yields “further investigation”

# Look at Concordances

- This collocation occurs frequently with the verb **WARRANT**, as in
  - ✓ These results are consistent with previous findings and warrant further investigation...
  - ✓ ... this observation warrants further investigation.
  - ✓ Because further investigation of these treatments is warranted, the committee is...
  - ✓ ... we believe that further investigation on this topic is warranted.
  - ✓ This hypothesis warrants further investigation.
  - ✓ This possibility warrants further investigation.

# Longer conventional expression? Search...

**CORPUS OF CONTEMPORARY AMERICAN ENGLISH**

0 MILLION WORDS, 1990-2012

START [v] ?

history | lists | profile |

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]

COMPARE [v] ? SIDE BY SIDE [v] ?

		CONTEXT	FREQ
1		WARRANTS	33
2		WARRANT	29
3		WARRANTED	22
4		WARRANTING	2
5		WARRANT	2
6		WARRANTS	1
TOTAL			89

1,514 se

KEYWORD IN CONTEXT DISPLAY

Help / information / contact

SECTION: NO LIMITS

CLICK FOR MORE CONTEXT [?] [SAVE LIST] CHOOSE LIST [v] CREATE NEW LIST [v] [?]

1	2012	NEWS	WashPost	A	B	C	risk factor for the development of skin cancer; however, this observation <b>warrants further investigation.</b> " They suggested that doctors pay particular attention
2	2012	ACAD	EducationTreatment	A	B	C	variability in final evaluation results when applied to the same body of literature <b>warrants further investigation</b> to reduce sources of discrepancies and refine cu
3	2012	ACAD	ExceptionalChildren	A	B	C	collected, rather than expect constant growth throughout the year. This hypothesis <b>warrants further investigation.</b> # Our findings further suggest that growth of
4	2012	ACAD	ExceptionalChildren	A	B	C	persistent low achievement from those who profit from instruction). This possibility <b>warrants further investigation.</b> # Despite the promise of these findings, cert
5	2011	ACAD	Bioscience	A	B	C	et al. 2010). Although this presents an attractive hypothesis that <b>warrants further investigation,</b> the mechanisms behind the patterns in natural populations are
6	2011	ACAD	CommCollegeR	A	B	C	faculty members is related to a program being transfer oriented or occupationally oriented <b>warrants further investigation.</b> Conclusions and Implications for Pra
7	2011	ACAD	CommCollegeR	A	B	C	possibility of differential unmet need between occupational-technical courses and transfer-oriented courses remains and <b>warrants further investigation.</b> Authors
8	2011	ACAD	InstrPsych	A	B	C	of coeducational classes. Examining the referral rates of boys in single-gender classrooms <b>warrants further investigation.</b> # Although not conclusive without fur
9	2010	ACAD	EnvironHealth	A	B	C	in humans (She et al., 2007). This worrisome trend <b>warrants further investigation.</b> # Maternal exposure to PCBs is predominantly through the food chain. Infan
10	2010	ACAD	SchoolCounsel	A	B	C	constraints in order to maintain some level of school-community collaboration. This topic <b>warrants further investigation.</b> # Finally, partnership-related training d
11	2010	ACAD	SchoolCounsel	A	B	C	be motivated by additional factors not captured in this study. This relationship <b>warrants further investigation.</b> # Further, it is important to note that partnership
12	2009	ACAD	SportBehavior	A	B	C	Yet, self-criticism has been ignored in past related sport psychology research and <b>warrants further investigation.</b> The rationale for this study is based primarily
13	2008	NEWS	AssocPress	A	B	C	' testimony that he " never used anabolic steroids or human growth hormone <b>warrants further investigation.</b> " // The seven-time Cy Young Award winner gave
14	2008	ACAD	CollegeStud	A	B	C	as headache or pain syndromes like temporomandibular disorder in susceptible individuals. This <b>warrants further investigation</b> to better understand the relation
15	2008	ACAD	EarNoseThroat	A	B	C	may broaden once longer-term follow-up data become available. Therefore, this approach <b>warrants further investigation.</b> # L: References -- (n1.) McClelland S
16	2007	ACAD	Archaeology	A	B	C	" Some of the government archaeologists will even leak information to me that <b>warrants further investigation.</b> " he says. # Spangler is also a savvy operator -

Ethel Schuster, Ph.D.



# Findings?

- Forms in which WARRANT occurs:
  - 89 occurrences.
- Simple present tense form (62) is more frequent than the past tense (22)

# Examine concordance lines...

- Most recurrent structure is
- *Det + WARRANT + further investigation*
  
- See...

# Findings...

<b>Determiner</b>	<b>Noun Phrase</b>	<b>WARRANT</b>	
This	hypothesis	warrants	<u>further</u> investigation.
This	worrisome trend	warrants	<u>further</u> investigation.
This	<u>X.X</u>	warrants	<u>further</u> investigation.
These	findings	warrant	<u>further</u> investigation.
That	<u>X.X</u>	warranted	<u>further</u> investigation.

# Concordance for *warranted*

- └• Verb in passive construction, e.g.,
  - ✓ ... saying no further investigation is warranted.
  - ✓ ... reasonable grounds to believe that further investigation is warranted.
  - ✓ Further investigation is warranted to test the thesis that ...
  - ✓ ... and found that no further investigation was warranted.

# Longer conventional expression? Search...

- We find

*“reasonable grounds to believe that further investigation is warranted”*

- In 5 of the 22 concordance lines

# Examples found

- ✓ ... that the Department of Justice find that no reasonable grounds to believe that further investigation is warranted.
- ✓ ... saying there are no reasonable grounds to believe that further investigation is warranted.
- ✓ ... if she determines that there are reasonable grounds to believe that further investigation is warranted.
- ✓ ... the Attorney General concludes that "there are no reasonable grounds to believe that further investigation is warranted..."
- ✓ the Attorney General determines that reasonable grounds do exist to believe that further investigation may be warranted.

# Closing Remarks

- Sometimes field of work is too specific
- Corpora do not have enough material in domain to answer all your questions

>> Should build customized corpus