



The Gateway Wing

Data-driven learning to improve student writing

Heather Smyser, PhD

Defense Language Institute English Language Center

Beyond Basics: LEARN Workshop

Rockville, MD

26 Sept 2017

heathersmyser@us.af.mil

We Train the Future



Data-driven learning



The Gateway Wing

No explicit teaching of rules

Direct access to data

Corpus as a tool

We Train the Future



Benefits of Corpus Use



The Gateway Wing

Useful for chronic, learner-specific problems
discrete grammar

Acar, Geluso, & Shiki, 2011; Boulton, 2009;
2010; Gaskell, & Cobb, 2004; Yoon, 2008

Lexical Phrases

Li & Schmitt, 2009

Fluency

Luo, 2016; O'Sullivan & Chambers, 2006

Perception of writing

Charles, 2014



The Gateway Wing

Little research on how effective DDL is for L2 writing

We Train the Future



Use of Corpora in the Classroom: Learners



The Gateway Wing

- International ESL students
- In 1st-year composition

FULFILLMENT OF ENGLISH PROFICIENCY	MINIMUM FOR MOST MAJORS	MINIMUM FOR ENGINEERING, BUSINESS, NURSING, AND SBS LAW ONLY
TOEFL (Institution Code: 4832)	70	79
IELTS	6.0	6.5
SAT Critical Reading (Institution Code: 4832)	580 or higher	580 or higher
ACT English (Institution Code: 0096)	21	21
Pearson Test of English (PTE) Academic	53	53
IB (English A – Higher Level)	5 or higher	5 or higher (Engineering, Nursing, and SBS Law ONLY)
Cambridge Certificate of Proficiency in English (ESOL) CPE	B or higher	Not an eligible option
Cambridge GCSE, O Level or higher (English)	C or higher	Not an eligible option
Completion of 4 years of regular English classes at an accredited American or IB high school (domestic or abroad)	C or higher	C or higher (Engineering ONLY)
Endorsement from the University of Arizona's Center for English as a Second Language (CESL)	Yes	Yes
Completion of ENGL 101 OR 102 OR (107 AND 108) or Equivalent Course. For course description information click here .	C or higher	C or higher (Engineering & SBS Law ONLY)



Use of Corpora in the Classroom: Context



The Gateway Wing

- Freshman English Composition Course
- Homework assignment: use Wordandphrase.info

WORD AND PHRASE . INFO DAVIES | BYU | COCA

FREQUENCY LISTS - [ANALYZE TEXTS](#) | [ALL GENRES](#) - ACADEMIC LOG IN [HELP](#)

ENTER TEXT BELOW [MY TEXTS](#)

[HELP](#) WORD PHRASE

Select individual words in the text to see "word sketches"

You can enter any text that you would like in the form at the left -- for example, a paper that you've written, or a newspaper article that you've copied from another website. After inputting the text, you can then see useful information about words and phrases in that text, based on data from the [Corpus of Contemporary American English \(COCA\)](#).

First, it will **highlight** all of the medium and lower-frequency words in your text and create lists of these words that you can use offline. This **frequency data** can help language learners focus on new words, and it can allow you to see "what the text is about" (i.e. text-specific words). You can also have it show you the **"academic" words** in your text.

Second, you can click through the **words** in the text to see a detailed "word sketch" of any of the words -- showing their definition, and detailed information for the word from COCA -- collocates (which provide meaning into the meaning and usage of the word), re-sortable concordance lines, and the frequency of the word (overall, and by genre).

Finally, you can do powerful searches on selected **phrases** in your text, to show related phrases in COCA. In this way, this resource is like a **"collocational thesaurus"** to see what related phrases are most likely in different styles of English. For example, if you click on the words *potent argument* in the text that you enter, it will suggest alternate ways to express this (e.g. *powerful* or *convincing argument*), and it will show you the frequency of those phrases in COCA -- overall, and by genre. This will help you use "just the right phrase", based on a huge collection of native English texts.

MATCHING PHRASES FREQUENCY BY GENRE (CHART) | DEFINITIONS | COLLOCATES | RE-SORTABLE CONCORDANCE LINES



Using Corpus: Assignment



The Gateway Wing

1. Enter the website <http://www.wordandphrase.info> into your browser.
2. Click on the Input/analyze texts box.
3. Take one paragraph from your draft and copy and paste it into the search box.
4. Make sure the word button at the bottom right of the input box has a box around it. Click search. (Be patient. It might take a while to load your text.)
5. What are the values for each chart at the top? Copy and paste those.
6. How much of your paper is composed of academic English?
7. Click on one word in blue or green. Hit search. What do you see? In which domain(s) do you most often see this word?
8. Clear the search. (You might have to hit another word in your text to get back to the search boxes.) It takes a few minutes to load.
9. Search for another word in your text. What do you see? In which domain(s) do you most often see this word?
10. Click on the percentage in red at the right side of the search box. Wait for it to load. What do you see?
11. Search for a couple of words under this setting. What differences do you notice between this and the words in blue/green?
12. Rewrite your paragraph, sentence by sentence. Aim for 20% Academic English in each sentence.



SEE LISTS

FREQ RANGE

1-500

501-3000

> 3000

ACAD

HELP

264 WORDS

70 %

17 %

13 %

7 %

WINE *n* (RANK 1453, FREQ 30088)

	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
CLICK BAR TO LIMIT					
STORED	12	53	57	72	8
MORE	1942	7073	8066	11781	1226



Activity Results



The Gateway Wing

Pre-WordandPhrase.info		
Word	Frequency	Abstract?
the	90	0
of	64	0
to	58	0
and	48	0
in	42	0
a	35	0
is	33	0
i	22	0
have	17	0
s	17	0
for	15	0
more	15	0
students	15	0
university	14	0
by	13	0
it	13	0
that	13	0
this	13	0
about	12	0
are	12	0
from	12	0
people	12	0
one	11	0
as	10	0
study	10	0

Post-WordandPhrase.info		
Word	Frequency	Abstract?
the	189	0
of	132	0
to	121	0
and	113	0
in	88	0
a	75	0
is	69	0
i	45	0
have	35	0
s	34	0
for	33	0
students	33	0
more	32	0
this	29	0
university	29	0
that	28	0
about	27	0
are	27	0
by	26	0
it	26	0
people	23	0
study	23	0
from	22	0
one	22	0
with	22	0



Activity Results, cont.



The Gateway Wing

who	10	0
their	9	0
which	9	0
with	9	0
because	8	0
my	8	0
or	8	0
be	7	0
countries	7	0
different	7	0
has	7	0
other	7	0
there	7	0
we	7	0
an	6	0
can	6	0
church	6	0
like	6	0
on	6	0
so	6	0
student	6	0
was	6	0
his	5	0

as	21	0
which	21	0
who	21	0
has	20	0
their	19	0
different	18	0
my	17	0
or	17	0
because	16	0
other	16	0
there	15	0
be	14	0
countries	14	0
like	14	0
can	13	0
church	13	0
on	13	0
was	13	0
an	12	0
so	12	0
systems	12	1
they	12	0
we	12	0



Student sample



The Gateway Wing

Pre-Word and Phrase

Arizona is **now moving** more into becoming a bigger player in wine, **which provide** more and more opportunity for people to enter the wine industry. Since Arizona wine producing region actually is some of the oldest in the U.S., **which have been first settled by Spanish settlers in the 16th century. The variety of grapes brought in from Europe was well suited for the hot climes and produced large yields.** We have a lot potential to work on the winery in Arizona. However, when it comes to wine, people usually think of California. Even like University of Arizona such a large scale and high-ranking university do not have relative majors of viticulture and enology society. “I want to try to foster more interest with students. We are a huge university. Why can’t we start opening a new major or maybe have a wine library or laboratory like the UCs school?” mentioned by the vice president, Marita Hill, who had studying wine and having internship in a vineyard in Italy before. Because the process of making wine is a lot to do with chemistry, the club needs people who have some background with chemistry as well. Wine is more like a kind of art, it really needs to put a lot effort in every aspect during processing, which makes this club none-exclusive, they involves business, agriculture, chemistry environment-related knowledge and so on. Arizona is worth developing the viticulture and winery industry more, and it is time to be aware more of the enology society, which is the thing that this club provides.

Post-Word and Phrase

Arizona is **now proceeding** more into becoming a bigger player in wine, **providing** more **occupancy** in the wine industry. **According to Jim Seder, a wine inquirer, mentioned** “Since Arizona wine producing region actually is some of the oldest in the U.S., which have been first settled by Spanish settlers in the 16th century. The variety of grapes brought in from Europe was well suited for the hot climes and produced large yields.” In fact, wine grapes do not appreciate the sweet life. Vineyards planted in a rocky, inhospitable landscape produce more concentrated juice with a more complex flavor, which exactly is the geological structure that Arizona has contained. **As Arizona Experience Organization declared, “Vineyard growers generally look for warm days and cool nights with a certain ratio of sunlight to shade.”** And some of the Arizona’s higher elevation zones like Sonoita in Santa Cruz County, Willcox in Cochise County, and the Verde Valley in Yavapai County, completely fit this bill. The hot days and naturally chilly nights of the high altitude regions balance acidity and sweetness. Winegrowers are constantly experimenting with temperature manipulation like canopy management to provide shade on hot days. Nowadays, Arizona has 942 acres under vine and 83 bonded wineries, which has been developing for years and has a great potential to work more on the winery.

We Train the Future



Discussion



The Gateway Wing

- Consistent with other corpus-informed studies that showed no improvement in writing process (See Gaskell & Cobb, 2004)
- No improvement in level of text
- No change in number of abstract terms appearing in most frequently used words of corpus
- Increase in length (consistent with Luo, 2016)



Discussion



The Gateway Wing

- Not appropriate for one-time fix
- Input could be too complex
- Writing topics too varied to be measured by corpus



Questions?



The Gateway Wing

- heather.smyser@us.af.mil



Thank you!



The Gateway Wing

We Train the Future



References



The Gateway Wing

- Acar, A., Geluso, J., & Shiki, T. (2011). How can search engines improve your writing? *CALL-EJ*, 12(1), 1-10.
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37-51.
- Boulton, A. (2010). Data-Driven Learning: Taking the computer out of the equation. *Language Learning: A Journal of Research in Language Studies*, 60(3), 534-572. doi: 10.1111/j.1467-9922.2010.00566.x
- Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, 35, 30-40. doi: 10.1016/j.esp.2013.11.004
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(4), 301-319. doi:10.1016/j.system.2004.04.001
- Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology*, 14(1), 28-44.
- Li, J. & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85-102. doi: 10.1016/j.jslw.2009.02/001
- Luo, Q. (2016). The effects of data-driven learning activities on EFL learner's writing development. *SpringerPlus*, 5, 1-13. doi: 10.1186/s40064-016-2935-5
- O'Sullivan, I., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49-68. doi: 10.1016/j.jslw.2006.01.002
- Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10(3), 130-139. doi:10.1016/j.jeap.2011.03.003
- Yoon, H. (2008). More than a linguistic experience: The influence of corpus technology on L2 academic writing. *Language Learning and Technology*, 12(2), 31-48.