Below are the presentation slides from Shelley Staples' presentation. If you are a member of Michigan State University and would like the access codes to the CROW or MACAWS corpora that were provided during the presentation, please e-mail Adam Pfau (pfauadam@msu.edu).

Using Corpora for Pedagogy and Research

Shelley Staples

Overview of Today's Talk

- Crow (Corpus and Repository of Writing):
  - Corpus and computational projects
    - Reporting verbs
    - Citations
    - Genre Classification
  - Corpus-based language teaching + classroom-based research

- MACAWS (Multilingual Corpus of Assignments—Writing and Speech)
  - Corpus analysis (Picoral, Novikov, Sommer-Farias)
  - Corpus-based language teaching

Our Lab Team

https://writecrow.org/team/










Our Institutions

Acknowledgments


CROW
Corpus & Repository of Writing

Crow Corpus

Institution	Course	Number of texts	Number of words	Average word count
Northern Arizona University	ENGL 105	1,174	1,561,604	1,330.16
Purdue University	ENGL 106i	7,362	6,503,644	883.77
University of Arizona	ENGL 106	1,334	1,142,210	856.23
University of Arizona	ENGL 107	398	349,997	879.39
University of Arizona	ENGL 108	671	634,949	946.27
Total		**10,939**	**10,192,404**	**931.75**

Crow Corpus - Demographics

Institution	Countries	Years in School	Majors
Purdue	China—5,236 (71%) India – 537 (7%) South Korea – 530 (7%) Malaysia – 334 (4%) Indonesia – 52 (< 1%) Turkey – 47 (< 1%) Thailand – 42 (< 1%) Ecuador – 41 (< 1%)	1 – 6,162 (84%) 2 – 835 (11%) 3 – 235 (3%) 4 – 128 (2%)	Engineering – 1,479 (20%) Sciences – 1,441 (20%) Liberal Arts – 1,216 (17%) Management – 933 (13%)
University of Arizona	China – 1,070 (45%) Saudi Arabia – 275 (12%) Malaysia – 175 (7%) Mexico – 98 (4%) Kuwait – 62 (3%) South Korea – 34 (1%) Thailand – 31 (1%)	1 – 1,597 (67%) 2 – 455 (19%) 3 – 255 (11%) 4 – 70 (3%)	Sciences – 1,008 (42%) Management – 360 (15%) Social & Behavioral – 349 (15%) Engineering – 195 (8%) Arts & Sciences – 161 (7%)

Crow Repository

Institution	Course	Number of texts
Purdue University	ENGL 106i	68
University of Arizona	ENGL 101	3
University of Arizona	ENGL 102	37
University of Arizona	ENGL 106	147
University of Arizona	ENGL 107	88
University of Arizona	ENGL 108	37
Total		380

Corpus and Repository of Writing (Crow)

- Access at: http://crow.corporaproject.org
- Interface intended to be user-friendly for teachers
- UX studies ongoing
- Please contact us for access at https://crow.corporaproject.org/authorize
- Offline corpus is also under development for public use

Corpus Research Projects with Crow Reporting Verbs

- Descriptive study (Kwon, Staples, & Partridge, 2018)
  - Investigated semantic classes and rhetorical functions of reporting verbs based on Charles, 2006 and Friginal, 2013
  - Compared L2 FYW literature review with Friginal 2013, undergrads in a forestry class
  - Found key differences in the ways L2 FYW used reporting verbs
    - Less lexical variety
    - Less register awareness
    - More self-reference and uncited generalization; less attribution to outside sources

Corpus Research Projects with Crow Reporting Verbs

- Intervention study (Shin, Velázquez, Swatek, Staples, Partridge, 2018)
  - Developed materials based on Kwon et al. (2018)
  - Implemented in 45-minute workshop
  - Compared pre-post workshop groups with control group (corpus texts from the same semester without intervention)
  - Intervention group improved variety and register awareness; less progress on rhetorical functions

L2 Journal
An electronic refereed journal
for foreign and second language educators

Corpus Research Projects with Crow Citations

- Descriptive study (Gao, Picoral, Staples, & MacDonald, under review)
  - Investigated citation use in terms of form (integral, nonintegral, hybrid)
  - Investigated citation use in terms of rhetorical function (based on Petrić, 2007)
  - Investigated pedagogical materials in relation to citation use
  - Differences across two assignments (lit review and researched argument)
  - Found new form (hybrid) and limited use of rhetorical functions (mostly attribution); also found relationship between model papers provided by instructors and students' texts

Corpus Research Projects with Crow Citations

- Citation Classifier (in progress, with Picoral, Novikov, Sanchez, and Gao)
  - Picoral used human coding to create two deep neural network classifiers
  - Tool predicts form and function for new texts based on these key determinants
    - Overall accuracy of 65% for form (5 labels) and 74% (5 labels) for function
  - Testing accuracy of identification on new data
    - More varied assignments; different institutions
    - Next up: writers from different backgrounds (Spanish Heritage Language writers)

Corpus Research Projects with Crow Genre Classification

- Pilot analysis (SSLW 2019, with Tardy):
  - Applied two frameworks for genre classification to 10 UArizona assignments (Carter, 2007; Nesi & Gardner, 2012)
    - Used corpus texts and repository materials
    - Identified one metagenre based on Carter (2007)(research from sources) and two new ones (analysis of artifact, rewriting a text)
    - Identified five genre families based on Nesi & Gardner
    - Many assignments were hybrids between the genre families
    - Many student papers did not fall into the same genre families (there was a lot of variation in the explication of the assignments)

Corpus Research Projects with Crow Genre Classification

- Genre Classifier (in progress with Picoral, Velázquez, Novikov, Goulart, Shin, and Reppen)
  - Logistic regression to identify key linguistic features of four assignments (Literacy Narrative, Research Proposal, Synthesis, Researched Argument)
  - Machine learning model to label sentences as one of the four assignments based on linguistic features
  - Literacy Narrative the most precise (.81 precision)
  - Researched Argument the most difficult to classify, with low precision (.53) and virtually no recall (.01)
  - Next steps: dig into the linguistic features to examine qualitatively in texts; combine sentence labelling to create a text-based label

Corpus Pedagogy: Crow

- Genre and register-based materials development (with Conrad, Dang, & Wang) based on:
  - Teacher needs analysis
  - Word lists
  - Examination of concordance lines
  - Evaluation: Surveys, focus groups, and examination of texts
- Materials take the form of:
  - Full texts (model texts)
  - Larger excerpts from texts
  - Crowcordance lines
  - iDDL concordance lines
  - Word lists

Larger Text Excerpts

Part II: Identifying Showing vs. Telling

Now that you have observed common features of "showing" and "telling", you can practice identifying these two ways of presenting information. As you read the following excerpts from two students' Literacy Narratives, change the text color in which the authors are **telling** about something to purple and change the text color that are **showing** something to blue. The first question below asks for your copy/pasted text of Excerpt 2. The excerpt has been coded as a model.

Excerpt 1

I remember first coming to the United States when I was 15 years old as a foreign exchange student. Even though there are not too many significant differences between Russian, European, and American cultures I experienced cultural shock. On my first day of school I saw people being so open, smiling all the time, some girls were wearing pajamas to school and didn't brush their hair (I found out that it was called messy bun), I saw boys that dressed like girls, a lot of kids had tattoos, all kinds of people expressed themselves in different ways. High school students seemed to be so grown up for their age and so different compared to teenagers in Russia. My first mistake was comparing both cultures and judging American culture.

Excerpt 2

Maybe Full Metal Jacket is not as popular as 2001: A Space Odyssey, it is still a Kubrick's film. It will be sold out quickly. For this reason, I was supposed to buy the ticket in person as early as possible, at any cost.

It was a morning in November. When I arrived at the film archive, it was 7 am. The pale sky applied a white filter covered on the ground. There were several people waiting outside the gate of the archive. I walked to the end of the line and read a book to kill time. When I was reading, I heard a boy passing by the waiting line asked his father: "Dad, What are those big brothers waiting for?"

"Watching a movie." The father said.

Crowcordance activity

Activity 3, Part II: Using “this + summary word” as a transition

In addition to transition phrases like “for example” and “such as” another method of transitioning between sentences is to use “this” alone or with a summary word that captures the meaning of the ideas presented in the previous sentence.

Take a look at the examples below. Then answer the questions that follow.

Crow | Genre Analysis

1. Listeners of spoken word won't get the chance to refer to the text, they will only take away the essence of what was being conveyed and maybe a few iconic phrases. This makes it imperative for the poet to make their poems easy to understand and interesting.

4. The color choice surrounds the schools team colors of Red White and Blue. This choice also relates it to the athletic family as a whole because all of the sports posters use this color scheme.

5. An informal letter is written to a person in a personal fashion. This can be written in first, second or third person.

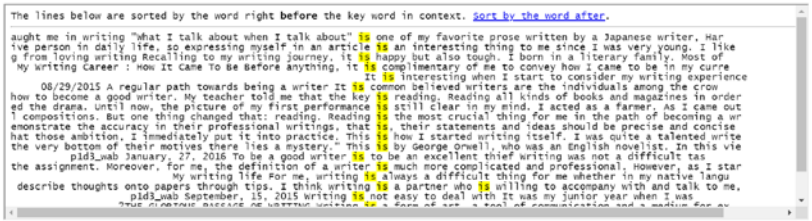
6. As it is ending of the letter it should be short and good. This part will leave final impression of the writer on the receiver.

1. Identify what “this” refers to in each of the sentences.
2. Compare the sentences that have this + summary word and the sentences that use this alone. Was it easier to identify what “this” meant with the summary word?
3. What other reasons might you want to use a summary word with “this”?
4. Now, look at the sentences that only use “this” instead of “this + summary word”. What summary word could you add after “this”?

Concordance Line Activity

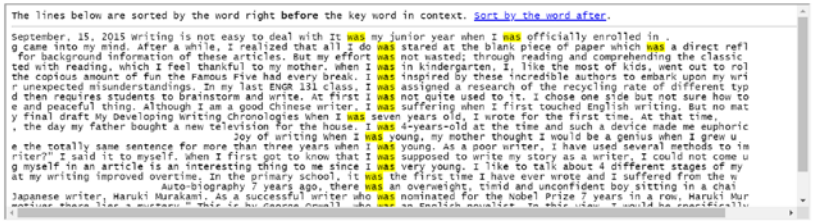
Activity 4, Part II: *Is* vs. *Was*

Now, examine the use of "is" and "was" (the most common verbs in Literacy Narratives) in the excerpts below.



The lines below are sorted by the word right **before** the key word in context. Sort by the word after.

aught me in writing "What I talk about when I talk about" **is** one of my favorite prose written by a Japanese writer, Har
ive person in daily life, so expressing myself in an article **is** an interesting thing to me since I was very young. I like
g from loving writing Recalling to my writing journey, it **is** happy but also tough. I born in a literary family. Most of
My writing **Career** : How It Came To Be Before anything, it **is** complimentary of me to convey how I came to be in my curre
It **is** interesting when I start to consider my writing experience
08/29/2015 A regular path towards being a writer It **is** common believed writers are the individuals among the crow
how to become a good writer. My teacher told me that the key **is** reading. Reading all kinds of books and magazines in order
ed the drama. Until now, the picture of my first performance **is** still clear in my mind. I acted as a farmer. As I came out
l compositions. But one thing changed that: reading. Reading **is** the most crucial thing for me in the path of becoming a wr
emonstrate the accuracy in their professional writings, that **is**, their statements and ideas should be precise and concise
hat those ambition. I immediately put it into practice. This **is** how I started writing itself. I was quite a talented write
the very bottom of their motives there lies a mystery." This **is** by George Orwell, who was an English novelist. In this vie
pld3_wab January. 27. 2016 To be a good writer **is** to be an excellent thief writing was not a difficult tas
the assignment. Moreover, for me, the definition of a writer **is** much more complicated and professional. However, as I star
My writing life For me, writing **is** always a difficult thing for me whether in my native langu
describe thoughts onto papers through tips. I think writing **is** a partner who **is** willing to accompany with and talk to me,
pld3_wab September, 15, 2015 Writing **is** not easy to deal with It was my junior year when I was
THE GLORIOUS PASSAGE OF WRITING writing **is** a form of art, a tool of communication and a medium for ex



The lines below are sorted by the word right **before** the key word in context. Sort by the word after.

September, 15, 2015 Writing is not easy to deal with It **was** my junior year when I **was** officially enrolled in .
g came into my mind. After a while, I realized that all I do **was** stared at the blank piece of paper which **was** a direct refl
for background information of these articles. But my effort **was** not wasted; through reading and comprehending the classic
ted with reading, which I feel thankful to my mother. When I **was** in kindergarten, I, like the most of kids, went out to rol
the copious amount of fun the Famous Five had every break. I **was** inspired by these incredible authors to embark upon my wri
r unexpected misunderstandings. In my last ENGR 131 class, I **was** assigned a research of the recycling rate of different typ
d then requires students to brainstorm and write. At first I **was** not quite used to it. I chose one side but not sure how to
e and peaceful thing. Although I am a good Chinese writer, I **was** suffering when I first touched English writing. But no mat
y final draft My Developing Writing Chronologies When I **was** seven years old, I wrote for the first time. At that time,
, the day my father bought a new television for the house. I **was** 4-years-old at the time and such a device made me euphoric
Joy of writing When I **was** young, my mother thought I would be a genius when I grew u
e the totally same sentence for more than three years when I **was** young. As a poor writer, I have used several methods to im
riter?" I said it to myself. When I first got to know that I **was** supposed to write my story as a writer, I could not come u
g myself in an article is an interesting thing to me since I **was** very young. I like to talk about 4 different stages of my
at my writing improved overtime. In the primary school, it **was** the first time I have ever wrote and I suffered from the w
Auto-biography 7 years ago, there **was** an overweight, timid and unconfident boy sitting in a chai
Japanese writer, Haruki Murakami. As a successful writer who **was** nominated for the Nobel Prize 7 years in a row, Haruki Mur

Questions for Activity 4 Part II:

1. First, take a look at the language surrounding the two verbs. How does it differ?
2. Why is present tense being used? Past tense? In other words, what different functions are used with the different tenses?

Word list activities

Part I: Frequent Words in Academic Texts

One way to understand register is through the different vocabulary used in a register. Below, we provide the most frequent words from the academic texts you are reading in the first unit of English 106 (Table 1). First, examine the frequent words from the academic texts. Then, answer the questions that follow. Please note that we omitted words that are common in all registers, such as "the", "and" and "of".

Table 1. Frequent Words in Academic Texts

Rank	Word	Frequency per 10,000 words
1	speech	197.2
2	community	181.6
3	language	148.6
4	they	136.5
5	may	111.2
6	social	96.9
7	use	84.8
8	communities	79.3

Part II: Frequent Words in the Rewrite

Now, look at the frequent words in the Rewrite in Table 2 below. Then answer the questions that follow.

Table 2. Frequent Words in the Rewrite

Rank	Word	Frequency per 10,000 words
1	language	318.9
2	you	180.3
3	I	162.3
4	we	144.2
5	people	140.9
6	use	112.7
7	English	103.7
8	can	92.4
9	some	91.3
10	they	91.3

Classroom-based Research with Crow

Course #1 (ENGL 107)

Fall 2019-Spring 2020: Four focus groups with five instructors over three months

Spring 2020:

- Surveys from four instructors and 54 students
- Eight observations of four instructors **face-to-face** classes
- 250 student papers

Summer 2020:

- One focus group and survey with one instructor
- Observation of **asynchronous** course

Fall 2020:

- Two focus groups with two instructors
- Surveys from two instructors and 17 students
- Two observations of two instructors' **synchronous and asynchronous** online classes
- ~180 student papers

Results: Effectiveness (Survey, N = 7)

Unit	Effective for activity	Effective for SLW
Literacy Narrative (6)	Effective-4 Very effective -2	Effective-3 Very effective-2 **Somewhat effective - 1**
Genre Analysis (4)	Effective-3 Very Effective - 1	Effective-3 Very effective – 1

- Students (some, not all) were really engaged/motivated by seeing examples from international students like themselves
- Led to improvement in writing (use of transitions, showing vs. telling)
- Led to creative language use (student decided to use 3rd person instead of 1st person in the literacy narrative)
- Appreciated having the materials created rather than "exploring/using the tool in a more open-ended class format"; connected to learning outcomes

Results: Ease of Implementation (Survey, N = 7)

Unit	Preparation	Ease of implementation
Literacy Narrative (6)	Somewhat prepared-2 Prepared -2 Very prepared-2	**Very difficult - 1** Somewhat easy-1 Easy-1 Very easy-3
Genre Analysis (4)	Somewhat prepared-1 Prepared-1 Very prepared - 2	Easy-2 Very easy-2

- Still required time to scaffold materials for students
- More training/support desired
- "Easy to fit into my classroom" (with a bit of a bump on the transition to asynchronous instruction)
- Appreciated that neither instructor nor student needed to be expert user of "this tool". Could "copy/paste" what I wanted to use

Outreach

- Workshops at national and international conferences (Symposium on Second Language Writing, Teaching and Language Corpora, CALICO)
- TESOL affiliate conferences such as AZTESOL and CATESOL
- High schools in Tucson
- Writing Center tutors at community college
- EFL teachers in Sonora, Mexico
- **Crow for Teachers – corpus-based materials for writing and language teaching**
- **CIABATTA – tools for building corpora**
- **Virtual workshop series – coding, grants, etc.**

MACAWS

Multilingual Academic Corpus of Assignments - Writing and Speech

MACAWS Team

Dr. Bruna Sommer-Farias
MSU

Dr. Adriana Picoral

Aleksey Novikov

Mariana Bertho

Valentina Vinokurova

MACAWS Corpus

L2	Mode	# of students	# of texts	# of words	Average word count
Russian	Writing	100	765	104,813	137.00
Russian	Speech	72	269	19,241	71.50
Portuguese	Writing	255	2,075	536,168	258.40
Total		427	**3,109**	**660,222**	212.36

MACAWS Corpus Demographics

L2	L1s	Other L2s
Russian	English (only) 68% Spanish (only) 2% Russian (only) 2% English and Russian 2% Other/unidentified L1s 26%	Spanish 64% Other Romance 25% German 15% ASL 6% Chinese 5%
Portuguese	English (only) 32% Spanish (only) 45% English and Spanish 45% Portuguese 1% Other L1s 6%	Spanish 32% Other Romance 17% German 5% ASL 2% Chinese 2%

Multilingual Corpus of Assignments – Writing and Speech

- Access at: http://macaws.corporaproject.org
- Please contact us for access at https://macaws.corporaproject.org/authorize
- Offline corpus is also under development for public use

Corpus Research Projects: MACAWS

- Picoral (2020)
  - Comparisons between MACAWS (four semesters of L2 Portuguese) and L1 Portuguese, Spanish, and English corpora
  - Learners compared across L1 English L2 Spanish, L1 Spanish L2 English, and L1 English L1 Spanish groups for copula choice (ser, estar) using logistic regression and word embeddings
  - Evidence of L1/L2 influence by copular construction (not wholesale); e.g., prepositional predicates do seem to be influenced by Spanish but not adjectival predicates
- Novikov (2021)
  - Complexity across MACAWS (four semesters of L2 Russian)
  - **Morphological** and syntactic complexity
  - Some patterns mirror those for English (noun-noun sequences increasingly used; decrease in “Because” clauses)
  - Findings mediated by task and mode
- Sommer-Farias & Picoral (in progress)
  - Lexical bundle use across genres in L2 Portuguese

Corpus Pedagogy: MACAWS

2018-2019 Paper-based Materials Development + Asynchronous Webinar

2019-2020 Professional Development Workshops—Local and International

- Goal: Independent activity design
- Reality: Even though teachers interested in developing activities, needed more scaffolding

2019-2020 iDDL Tool Creation

Spring/Summer 2020 Additional Web-based Materials Development

Summer-Fall 2020 Implementation in PORT and RSSS classrooms

Fall 2020 Focus Groups and Surveys (instructors and students) for feedback on materials

iDDL Materials

MACAWS Concordance lines box 1.

ощадь. Мы ели в ГУМ. мой друг, который не знает как говорить по-русски, они сказали "GYM" и не гум. Я много смеялся. Мы ходили по
Университете в Аризоне, в Тусоне. Чут чут понимаю и говорю по-русски, а неплохо пишу и читаю. Мои дедушка и папа русские, так
орю хорошая по-английском и по-французски. я неплохо говорю по-русски. я говорю по-английском в доме.
о-английски, я неплохо говорю по-испански, и я плохо говорю по-русски. Я хорошо понимаю по-английски, я неплохо понимаю по- ис
се. я учусь в Университете Аризоне потому что ч хачу говорю по-русски. я изучаю математику и русский язык. я люблю оба. я учусь
потому что они любет русский язык, хотят читать литературу по-русски. Я хочу быть переводчик и люблю русский язык. Русски такж
я изучаю русский язык в университет, я тоже читаю медленно по-русски, немного понимаю и плохо говорю. Дома моя фамилия говорим
язык. Моя подруга Была из Москвы. Она заставила меня по-русски культура. Я люблю Америку но Россия красивый. Очень реки и
-французски, тоже. Я он очень хорошо понимаю. Я хорошо пишу по-русски, но я он плохо говорю. Я понимаю немного по-русски, когда
те тоже. Я тожо понимаю по-испански нимного. Я читаю и пишу по-русски плохо, а по-испански хорошо. Я говорю и понимаю по-русски
по- английски и хорошо говорю по-испански. Я немного пишу по-русски. Я говорю по-английски дома. Моя мама и мой папа свободно
т, что это очень красивый язык. пишет и читает хорошо по-русски.
у что она хочет учиться в России. Она говорит и пишет хорошо по-русски, но понимает и читает не очень хорошо.
читать оригинальные документы, если она не знает как читать по-русски. Она думает, что она плохо говорит, читает, пишет, и поним
аю Русский язык потому что я хочу знать как писать и читать по-русски. Я люблю Русский язык и физику. Я не люблю математику. Я
ский язык. Я тоже неплохо знаю русский язык. Я бистро читаю по-русски, но я медленно говорю. Моя мама не понимает русский язык.

Other Resources

- CIABATTA
  - Corpus in a Box: Automated Tools, Training, and Advising
  - https://github.com/writecrow/ciabatta/wiki
- YouTube Channel
  - https://writecrow.org/youtube
- Crow for Teachers
  - https://writecrow.org/crow-for-teachers/
- Crow Workshop Series
  - https://writecrow.org/youtube (playlists: Workshops)

Thank you!

slstaples@arizona.edu

http://writecrow.org

https://sites.google.com/email.arizona.edu/macawswebinar

CROW
Corpus & Repository of Writing

MACAWS
Multilingual Academic Corpus of Assignments - Writing and Speech