

Taking Flight with MACAWS: Learner Corpus Data from and into the Classroom



May 18, 2020

Mariana Bertho
Aleksey Novikov
Adriana Picoral
Bruna Sommer-Farias
Shelley Staples

marianabertho@email.arizona.edu
alekseynovikov@email.arizona.edu
adrianaps@email.arizona.edu
brunasommer@email.arizona.edu
slstaples@email.arizona.edu



THE UNIVERSITY OF ARIZONA
CERCLL
Center for Educational Resources in
Culture, Language and Literacy



Corpora & Corpus-Based Pedagogy

- Corpus:
 - a collection of texts, spoken or written (plural = corpora)
 - representative of a particular domain of language use
 - balanced across genres or registers
- Learner corpus:
 - corpus of learner production, defined broadly (students, L2 writers, etc.)



Corpora & Corpus-Based Pedagogy

- Corpus-based pedagogy--Use of corpus or corpus findings to:
 - Inform curriculum
 - Create materials (e.g., textbook or by teachers)
 - Engage students directly in examining corpus data



Corpora & Corpus-Based Pedagogy

- Finding corpora:
 - Appropriate to your learner's level
 - Appropriate to your learner's needs/goals
 - An advantage of learner corpora
 - List of corpora (bit.ly/corpora_list)



Corpora & Corpus-Based Pedagogy

Corpus

- 9,759 samples of college writing, containing 8,347,760 words
- Texts from Fall 2014 to Spring 2018
- Filter by country, gender, assignment, date, TOEFL scores
- Search by word, word family, or phrase

[Search corpus](#)

Repository

Downloadable assignment prompts, syllabi, rubrics, lesson plans, and other course material.

- Filter by material, document format, course, and date
- Search the full text of resources; export to markdown

[Browse repository](#)

What's new?

- November 2: Arizona Teachers of English to Speakers of Other Languages, "Bridging L2 writing and academic vocabulary through corpus-based activities"
- November 15: Symposium on Second Language Writing, "Using a learner corpus and a repository of pedagogical materials for L2 writing research and teaching"
- March 31, 2020: TESOL 2020

Resources & Tutorials

Writing Models

- Compare how other writers use an English idiom

Lesson Planning

- Find sample sentences for a lesson on signal phrases

Research & Analysis

- Export Crow data for statistical analysis in R
- Perform machine-driven queries with the API

Corpus Building

- Develop a similar corpus using the source code

What is MACAWS?

**Multilingual Academic Corpus of
Assignments: Writing & Speech**

Second Language (L2)	Number of files	Number of tokens	Number of students
Russian	812	111,705	100
Portuguese	3,000	536,168	255



Accessing, Searching, and Filtering MACAWS



Accessing MACAWS

Please contact collaborate@writecrow.org for access to the MACAWS corpus.

Click on “Search corpus”

Corpus

- 1,037 samples of college writing, containing 614,781 words
- Texts in Russian & Portuguese, from 2017-2018
- Filter by L1, course, topic, draft, assignment mode
- Search by word or phrase

Search corpus

MACAWS demonstration

Data-Driven Learning

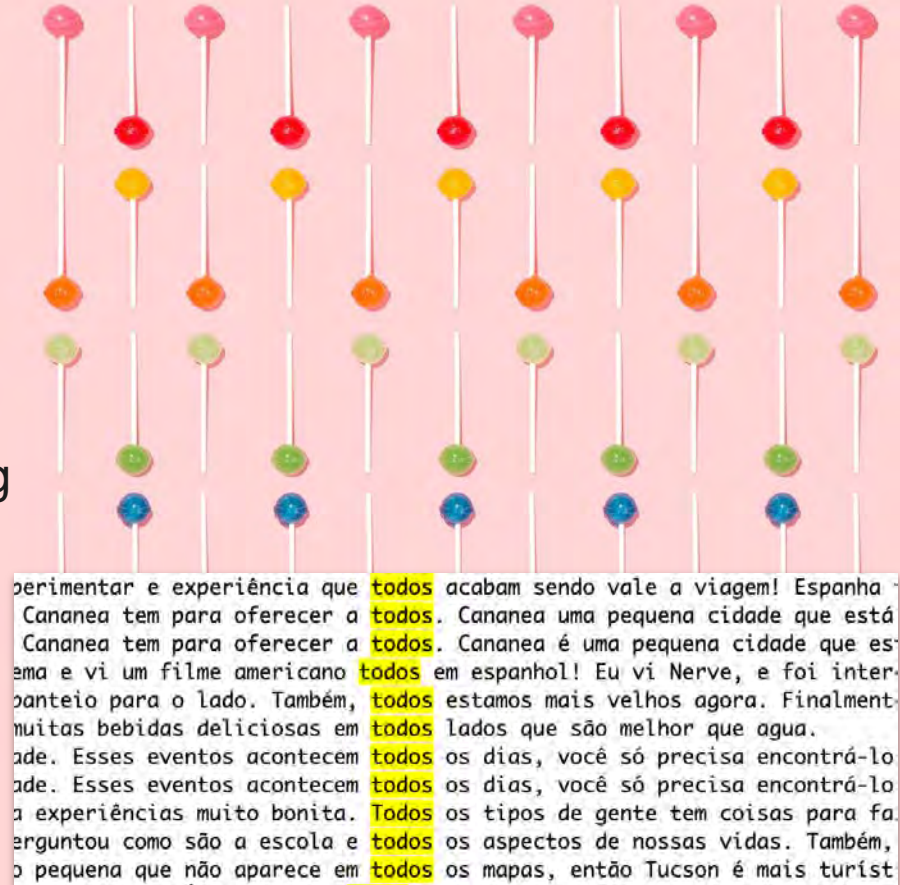
Incorporation of **authentic examples** of language use into classroom activities

Main goal: Pattern noticing

Pedagogical approach: Inductive learning

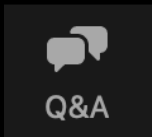
Concordance lines: a line of text that shows a search term highlighted with its surrounding context

What is **iDDL**?



iDDL demonstration

Questions?
Write your question
in the “Q&A”



Your turn!

What other searches would you implement for your classroom?

Write your answers in the chat to share your ideas.

Examples:

- “todos os” for intermediate Portuguese
- “учиться” vs “изучать” for beginner’s Russian
- “important” vs. “significant” for academic English writing



List of Resources

How to embed iDDL into LMS, sample materials, webinars, tutorials:

<http://macaws.flock.group>

How to embed iDDL to Google Sites: <https://youtu.be/Lo1Mz5LiEQ>

Corpus and Repository of First Year Writing in English: <http://crow.corporaproject.org>

List of Corpora: bit.ly/corpora_list

iframe generator to embed Google Docs & others: <https://www.iframe-generator.com/>

Fill in the blanks activity generator: https://bit.ly/fill_in_the_blank_generator

Articles

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 359–370). London: Routledge.

Vyatkina, N. (2020). Corpus-Informed Pedagogy in a Language Course: Design, Implementation, and Evaluation. In *New Technological Applications for Foreign and Second Language Learning and Teaching* (pp. 306-335). IGI Global.

Thank you! Questions?



Q&A

Mariana Bertho
Aleksey Novikov
Adriana Picoral
Bruna Sommer-Farias
Shelley Staples

marianabertho@email.arizona.edu
alekseynovikov@email.arizona.edu
adrianaps@email.arizona.edu
brunasommer@email.arizona.edu
slstaples@email.arizona.edu



<https://macaws.corporaproject.org/>



THE UNIVERSITY OF ARIZONA
CERCLL
Center for Educational Resources in
Culture, Language and Literacy