



Introducing the Corpus and Repository of Writing: Ethically Compiling and Analyzing Student Data

writecrow.org

Shelley Staples¹, Aleksandra Swatek², Bradley Dilger³

1. University of Arizona 2. Adam Mickiewicz University in Poznań, Poland 3. Purdue University

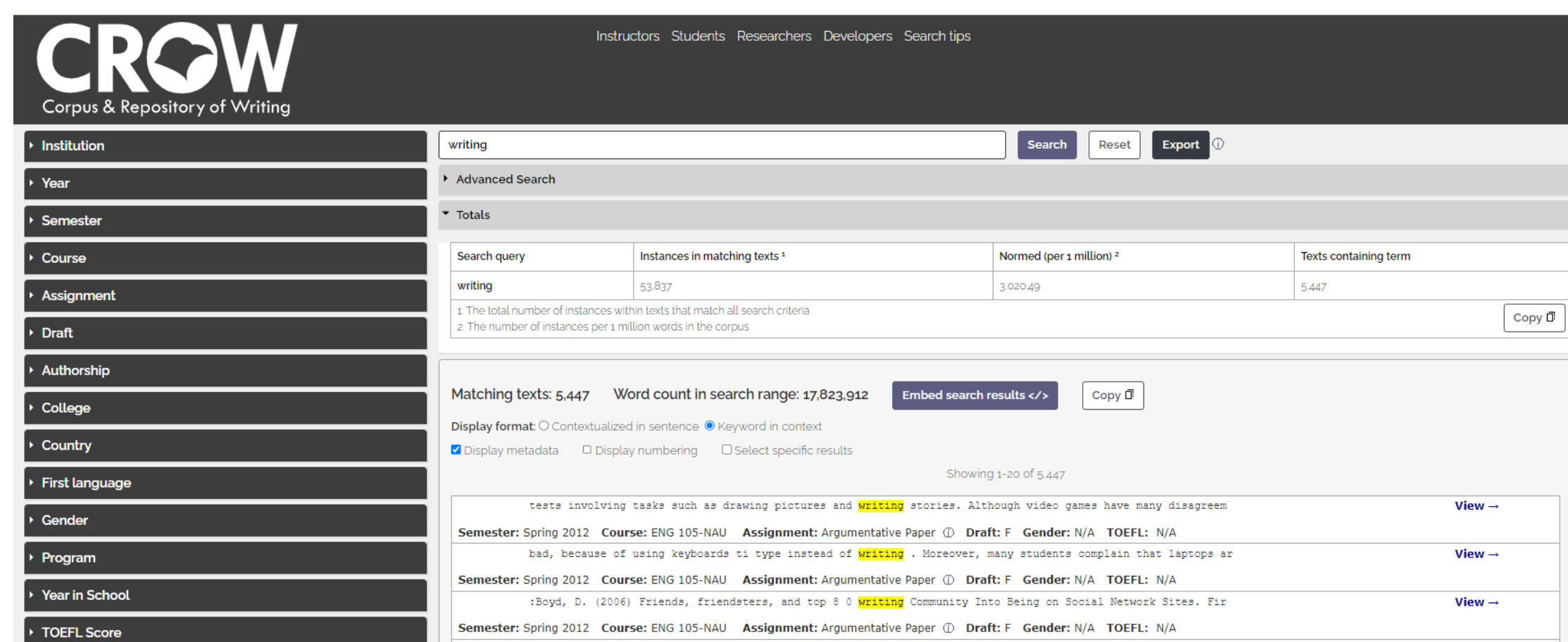
OVERVIEW OF THE CORPUS AND REPOSITORY OF WRITING (CROW)

- **Corpus and repository:** a web-based platform developed using common web frameworks and providing access upon request.
- **Research:** provides diverse data sets to examine student writing and classroom-based instruction using corpus tools.
- **Professional development:** a sustainable platform for writing instructors to share work with others.
- **Mentoring:** provides mentors an archive of sample learner writing and assignments to share with new instructors.
- **Assessment:** supports implementation of common assignment outcomes across 100+ sections of First Year Writing (L1 + L2 Writing).
- **Sustainability:** from day one, embed sustainable practices in project decision-making, development, research design and team-building.

DATASET

- **Learner corpus** (a collection of texts by students or other learners) representing 38 writing genres, totaling 17,832 texts and 17,823,912 words.
- **Corpus metadata:** student first and second language, majors, year in school.
- **Repository of diverse pedagogical artifacts** includes such items as syllabi, assignment sheets, assessment rubrics, activity sheets, and group work activity sheets. Currently over 500 items.

CORPUS AND REPOSITORY INTERFACE



Keyword in Context (KWIC) Search Results View
<https://crow.corporaproject.org>

Registered users can access Crow data based on their needs:

- **Standard access:** allows to see excerpts with searched keywords.
- **Full access:** designed for teachers who want to use students' texts in their classes.
- **Excerpt export:** users can export search results for import into spreadsheets.
- **Offline curated corpus access:** training required for this access to a subcorpus.

No commercial use is allowed.

ETHICS OF CARE AND STUDENT WORK

Data collection process includes varying levels of opt in/opt out, depending on institutional ethics board recommendations. For example:

- **Purdue University and Northern Arizona University: Opt Out** – students receive emails at the end of the semester offering the opportunity to exclude their course materials from the corpus. Repository materials are requested from the instructors who were interested in sharing their materials.
- **University of Arizona: Opt In** – students are asked at the end of semester to opt-in to sharing their data.
- **PI Opt Out:** PIs in leadership positions are removed from the process if they are in a position of power with regards to the courses in which data is collected.

Incentives

Gift cards for students and instructors participating at the University of Arizona.

METHODOLOGICAL CHOICES AND SAMPLE STUDY

Motivation

First year writing courses at some universities in the USA are designed to respond to needs of various student populations. The question remains if the design of the course based on those needs impacts the outcome of the course.

Curated Data for Preliminary Study

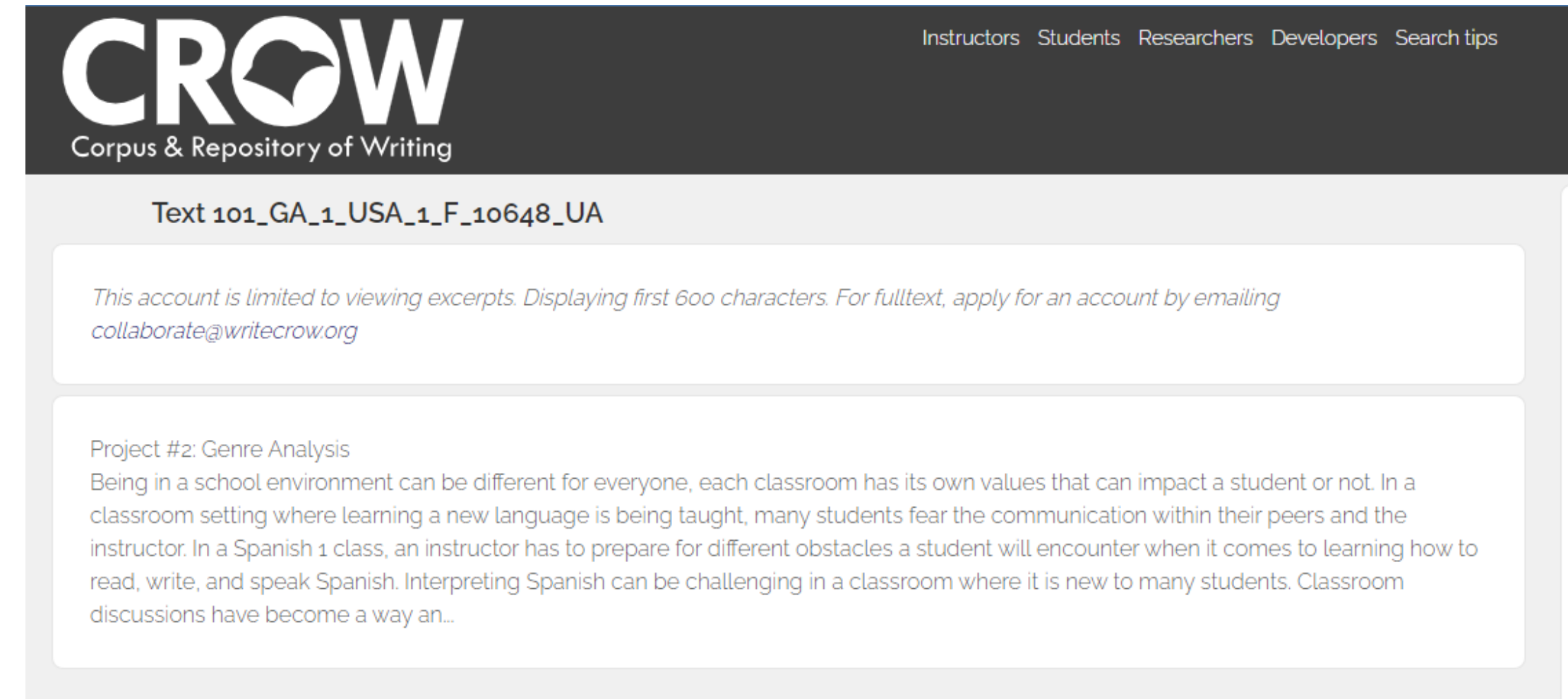
Data source (course #)	ENGL 106 (Course 1)	ENGL 107/108 (Course 2/3)	ENGL 101/102 (Course 2/3)
Learning outcomes and assignments	Academic literacy	Similar	Similar
Community of practice	Multilingual writers from outside USA	Multilingual writers from outside USA	Both monolingual (L1) & multilingual (L2) writers, most from USA

SAMPLE EXCERPTS FROM STUDENTS' TEXTS

Student 1: "However, this source's content is extremely useful, as it provides background information, provides an **example** of a direct derivative style, and can support the supplementary sustainability argument. Information related to **the design elements** of Native American Architecture serves as background information, and the proposal itself is a direct derivative of Cheyenne and Arapaho architecture. Additionally, **the author's assessment** of Native American Architecture serves as an **analysis** of indigenous vernacular architecture through a lens of sustainability."

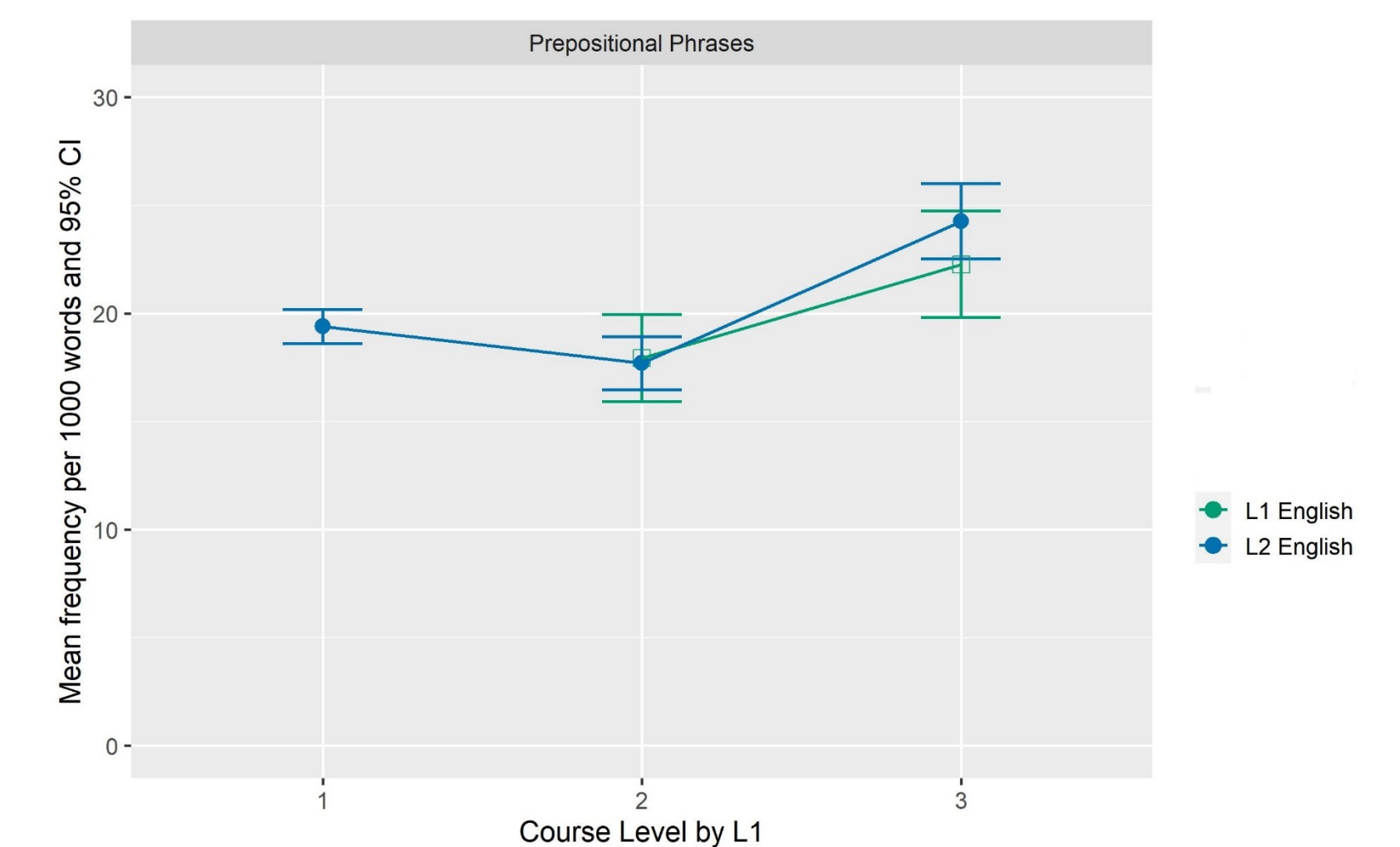
Student 2: "**The primer** of the book provides information regarding the obstacles that women face from a **young** age to going to college and continuing in the faculty and workplace. **Negative experiences** of women who are in pursuit in a STEM major for **a long period** of time are depicted, providing a strong foundation to support **the argument** of women being disadvantaged in the STEM field."

EXAMPLE CORPUS TEXT (Standard excerpt view)



Analysis

Staples analyzed features of linguistic complexity (prepositional phrases) in both samples. Means and confidence intervals across the three courses are plotted below.



RESULTS

- L1 and L2 students in both courses **develop along the same trajectory** with the same assignments/tasks.
- The courses can be considered **equivalent** in terms of outcomes related to language development.
- Students can **more confidently choose a course of interest**, without fear of a course being considered less challenging.