

# The Design and Research Potential of *CROW* for Language Research and Teaching



Jie Gao and Sherri Craig

Purdue Languages and Cultures Conference  
West Lafayette, IN

6 March 2016



# What is CROW?

- Started and imagined by Dr. Staples and Dr. Dilger to build a connection between the two programs which share similar goals and interests
- An alteration of C.R.O.W. (Corpus and Repository of Writing) chosen because of the connection to the Purdue OWL
- A team of undergraduate and graduate students and faculty from SLS and RC
- Goal: To create an online resource for students, teachers, and researchers with interests in and needs for corpus linguistics, pedagogical materials, and community.

# What Previous Projects Relate to *CROW*?

**Collaborative Online Instructor Network (COIN)**

**ICaP 2014-2015 Assessment and IMPACT Participation**

- PI: Jennifer Bay w/ Freddie deBoer and Sherri Craig
- Developed holistic rubrics to rate work (200+ samples)
- Determined more work was needed to understand ICaP
- Mostly 106 materials, 1 section of 106-I included

# Another Previous Project Related to *CROW*

## Purdue Second Language Writing Corpus (PSLW)

- PI: Shelley Staples w/ R. Scott Partridge and Heejung Kwon
- Collects student texts from 106-I courses
- Offers research and publication opportunities
- Offers professional and research development opportunities for graduate student

# How is Corpus Defined?

A body of naturally occurring language that is:

- authentic;
- machine-readable;
- sampled;
- representative (McEnery et al., 2006).

And, for the nature of our project, it should be assembled for a particular purpose.

# What is the Data Set?

Purdue Second Language Writing Corpus (PSLW)  
(Fall 2014 + Spring 2015)

Assignment	# of texts	# of words	Average word count
Narrative	888	710,356	799.95
Proposal	913	736,286	806.45
Interview Report	456	411,257	901.88
Synthesis Paper	881	813,123	922.95
Argumentative Paper	874	801,238	916.75
Total	4012	3,472,260	865.47

# Who is the Data Set?

Purdue Second Language Writing Corpus (PSLW)  
(Fall 2014 + Spring 2015)

Gender	%	Country	%	Class	%	Major	%
Male	57	China	76	Freshman	86	Other	22
Female	43	Other	9	Sophomore	7	Science	18
		Korea	5	Junior	2	Engineering	17
		India	6	Senior	5	Liberal Arts	17
		Malaysia	4			Explorers	14
						Management	12

# What are the TOEFL Scores in the Data Set?

Purdue Second Language Writing Corpus (PSLW)  
(Fall 2014 + Spring 2015)

Test Category	Mean Score
Listening	23.86
Reading	24.72
Speaking	21.14
Writing	23.12

TOEFL Score Range: 79-113

TOEFL Score Mean: 92.64



# PSLW Reporting Verb Project

- **Reporting verbs:** attribute information, feelings or actions, often to outside sources
  - Searches include terms such as, *argue, show, find, think*
- **Research Design:**
  - Data-driven learning: a list of reporting verbs generated from corpus based on frequency
  - Inductive and deductive: Students learn to identify reporting verbs, and then try to deduce the verb functions and options
  - Student-centered: After group discussions and self-revision of students' own writing, will the reporting verbs in students' last draft show difference in frequency?

# How does PSLW Compare to *CROW*?

## PSLW

**Source:**

ENGL 106-i: First-Year Composition - International Sections

**Content:**

Students' writing pieces

## COIN

**Source:**

ENGL 106: First-Year Composition

**Content:**

Pedagogical artifacts

## *CROW*

**Source:**

ENGL 106: First-Year Composition

ENGL 106-i: First-Year Composition - International Sections

ENGL 108: Accelerated First-Year Composition

**Content:**

Students' writing pieces

Pedagogical artifacts

# How does *CROW* Advance Corpus Work?

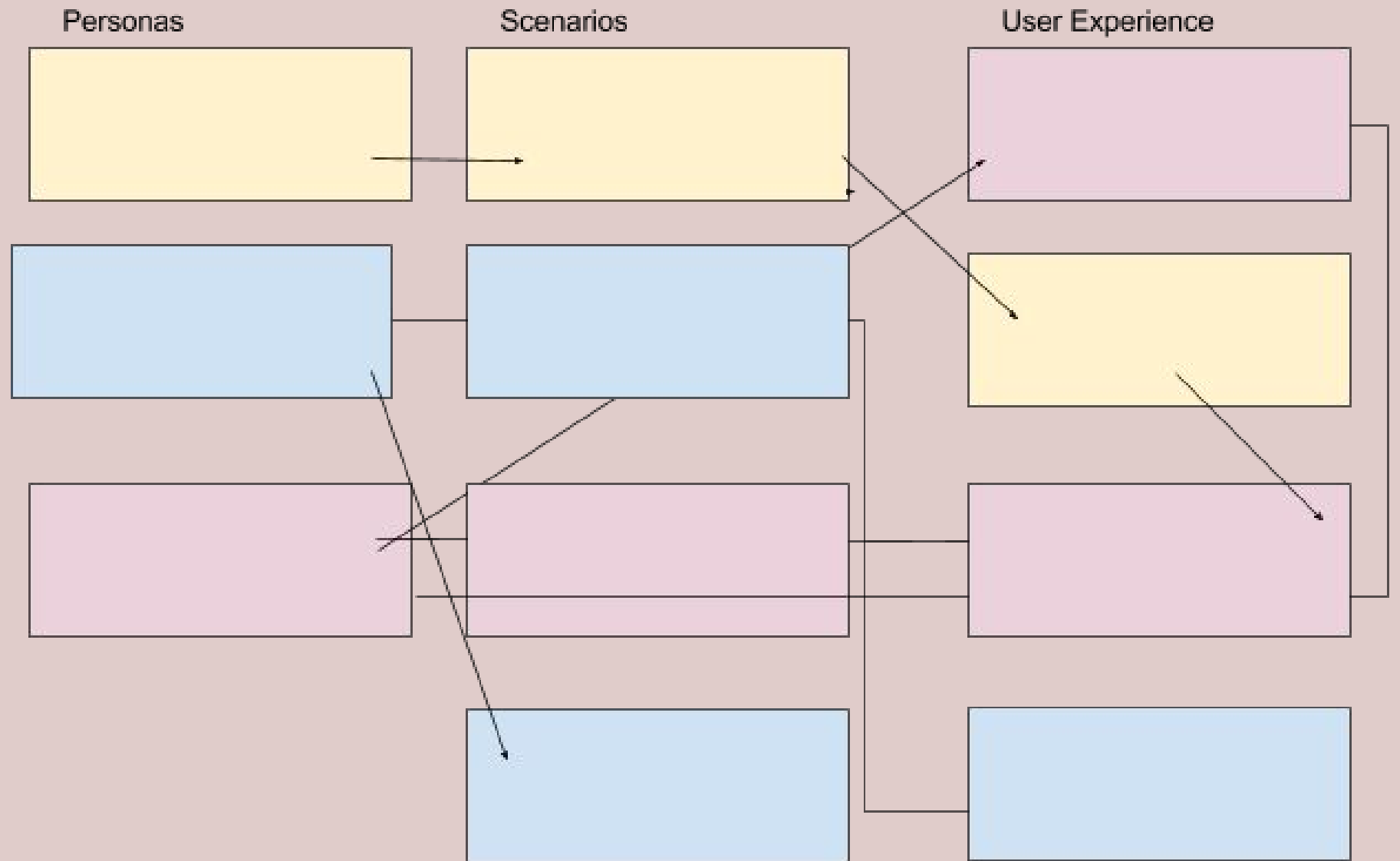
- **Development of an online interdisciplinary interinstitutional interface** where scholars can eventually submit their own texts and add to the repository and corpus
- **Inclusion of pedagogical artifacts** that accompany the production of the texts, including syllabi, assignment sheets, pre-writing readings, and schema building activities to contextualize corpus texts and improve learning
- **Advancement of corpus linguistics and language teaching research** through a unique combination of texts

# What Work Has Been Completed for *CROW*?

- There are many research questions based on the various *CROW* projects
- Purdue Department of English is the initial user
- Expanding to other universities in Fall 2016
- Prototyping website mock-ups
- Developing grants and conference presentation to expand ideas
- Creating user personas and scenarios
- Performed environmental scans



## Personas and Scenarios



# Personas + Scenarios Interaction

Mock up of how to potentially think about these elements at modular

**STUDENT LEVELS**

- No Restriction
- Senior Undergrad (S9)
- 1st Year Graduate (G1)
- 2nd Year Graduate (G2)
- 3rd Year Graduate (G3)

**NATIVENESS**

**TEXTUAL FEATURES**

**PAPER TYPES**

- No Restriction
- Argumentative Essay
- Creative Writing
- Critique/Evaluation
- Proposal
- Report
- Research Paper
- Response Paper

**DISCIPLINES**

- No Restriction
- Biology (BIO)
- Civil & Environmental Engineering (CLE)
- Economics (ECO)
- Education (LIX)
- English (ENG)
- History (HIS)
- Law (LAW)
- Mathematics (MTH)
- Medicine (MED)
- Music (MUS)
- Nursing (NUR)
- Philosophy (PHI)
- Political Science (POL)
- Psychology (PSY)
- Sociology (SOC)

**DISTRIBUTION ACROSS DISCIPLINES**

CLICK TO SELECT CLEAR SELECTION

**DISTRIBUTION ACROSS PAPER TYPES**

CLICK TO SELECT CLEAR SELECTION

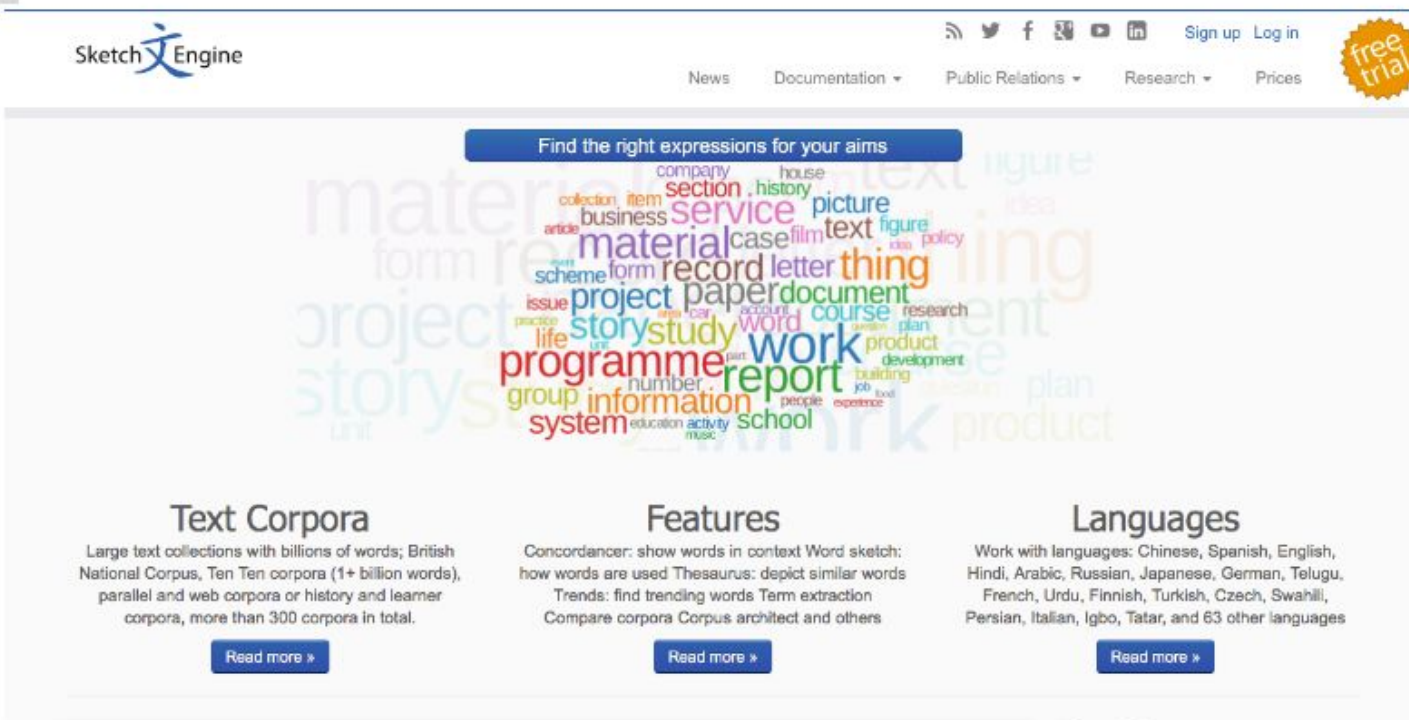
Report 44% Argumentative Essay 22% Creative Writing 1% Proposal 8% Research Paper 17% Critique 7% Response Paper 3%

Showing 1 to 20 of 829 papers

Paper ID	Title	Discipline	Paper Type
BIO_00_10.1	Inventing the Territory of Invasives: The Dangers of Biotic Disturbance	Biology	Argumentative Essay
BIO_01_04.1	The Evolution of Tenacity: A Look at the Factors that Drove Tetrapods to Move Out of the Water	Biology	Argumentative Essay
BIO_03_05.1	Intracellular Electric Field Sensing using Nano-sized Multimodes	Biology	Argumentative Essay
BIO_00_11.1	Exploring the Molecular Responses of Arabidopsis in Hypobiotic Environments: Identifying Possible Targets for Genetic Engineering	Biology	Proposal
BIO_01_01.1	V-Chowen: 1 and Steps towards a Specialty Tissue Model	Biology	Proposal
BIO_01_07.1	Zebrafish and PGC migration	Biology	Proposal
BIO_02_06.1	A Conserved Role of Cas-Spg System in Endoderm Specification during Early Vertebrate Development	Biology	Proposal
BIO_03_02.1	Linking scales to uniserial diversity	Biology	Proposal
BIO_00_01.1	The Ecology and Epidemiology of Phages	Biology	Report
BIO_00_02.1	Host-Parasite Interactions: On the Presumed Sympatric Speciation of <i>Vidua</i>	Biology	Report
BIO_00_02.2	Sensory Drive and Speciation	Biology	Report
BIO_01_02.3	Plant Pollination Systems: Evolutionary Trends in Generalization and Specialization	Biology	Report
BIO_00_02.4	Chromosomal Rearrangements, Recombination Suppression, and Speciation: A	Biology	Report

Category	Comments
Navigation	The categorized search menu system on the left, the search bar at the top and the hyperlinks near the essays create a simple navigation. Once users select an essay, the user demographic information and the essay text is revealed in a new tab, allowing the user to more easily move back and forth between the main page and the selected essay. But, in order to go back to the original home page, the user has to completely clear any selections they have made and then proceed from the beginning.
Intuitive Use	If all of the menus are collapsed, the site's intuitive use is reduced because of the overwhelming information provided on a single page. There is no page that provides information about the site, so MICUSP's site users must understand the purpose of it prior to interacting with the information. Without any kind of statement of purpose, anyone outside of this program, or anyone





Criteria	Evaluation
Users and core tasks	<p>Lexicographers, teachers, students, terminologist, translators, historians, and anyone who wants to research how words behave.</p> <p><b>More words for teachers:</b> Contemporary linguistics solves the difference between what we <i>can</i> say and what we <i>do</i> say. Thanks to corpora, you can find out how it deals with the problem to this</p>



# What are the Major *CROW* Projects?

- **Project 1:** Genre Analysis of First Year Writing
- **Project 2:** Writing Development in First Year Writing (L2 and potentially L1 as well)
- **Project 3:** Relationships between First Year Writing and Disciplinary Writing
- **Project 4:** Professional Development
- **Project 5:** Purdue-centric Assessment and Evaluation

# Thank you.



Jie Gao, Purdue University, [gao339@purdue.edu](mailto:gao339@purdue.edu)

Sherri Craig, Purdue University, [craig36@purdue.edu](mailto:craig36@purdue.edu)

