Engaging Language in Foundations Writing: Harnessing Computational Models to Advance Excellence in the Teaching, Learning, and Assessment of Writing

Introduction

Foundations writing (FW) is an integral part of the first-year experience for college students across the university, with the Writing Program as a whole serving approximately 12,000 students per year. Multilingual writers are a growing population at the university, particularly with the University's Strategic Plan Pillar 4 (UA Global), which aims to "Increase our international student population to move into the top 20 of national research universities for percent of international students". In addition, with the growth in microcampuses, such as Ocean University, there is an increasing demand for FW courses that can meet the needs of these new students, most of whom are multilingual.

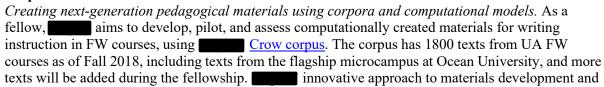
The curriculum in the current sequence designed for international students at UA (ENGL 106, 107, and 108) has created new opportunities for multilingual writers to participate in a genre-based approach to writing instruction, in which writers learn to understand writing across the curriculum and to engage in both public facing and academic writing styles to prepare them for further study as well as the workforce (Tardy, 2009). However, current textbooks that provide language support (e.g., Language Power, Ferris, 2014) are decontextualized from the course content. They also don't leverage newer research on the computational power of corpora (singular, corpus), which are large bodies of texts (in this case student writing). Corpus linguistics, a method to computationally examine language use in context, has been widely recognized as a way to identify patterns of language use that are associated with particular writing contexts and as a way to inform writing curricula in ways that are integrated with genre conventions and contexts (Aull, 2015; Cotos, 2014; Swales & Feak, 2013; Staples & Reppen, 2016). The incorporation of corpus linguistics into curricula has been shown to be effective in language classrooms (Bolton & Cobb, 2017). However, there are few existing materials and empirical studies that engage the FW context, and there is not a clear model for how improvement in FW can be measured using computational models.

The proposed project incorporates corpus linguistic approaches into FW courses, dovetailing with the Grand Challenge of leveraging data and computing for the 4th Industrial Revolution (IR). It aims to:

- 1) advance excellence in teaching by involving instructors in ethical, evidence-based and computationally driven approaches to improve writing instruction;
- 2) advance excellence in learning, by identifying needs of learners and develop materials to improve learning using computational methods;
- 3) *advance excellence in assessment*, by using the same computational methods to evaluate effectiveness of teaching materials.

The CUES project provides a significant impact in the area of teaching, learning, and assessing writing by introducing an innovative method (corpus linguistics) in tandem with existing approaches from educational research (quasi-experimental designs to test writing improvement, student and teacher evaluations, focus groups, and classroom observation) to evaluate the computational models. Fellowship support is requested for 3-years, to allow for materials development and assessment in each of the 3 FW courses that focus on preparing international students for writing at the university.

Proposed Activities and Methods



evaluation harnesses corpus linguistics, her specialization that she has published widely on, in major journals in applied linguistics and writing studies, as well as 2 books. Over the 3-year period, a cycle of materials development and evaluation will be repeated for the 3 courses. It is anticipated that 7 sections in each of the 3 courses will be reached during the CUES fellowship (400 students and approximately 18-21 instructors). In the 3 years that has been collecting data for the Crow project, 21 teachers have participated, with no extra incentives, suggesting that participation will be high for this project. The integration of materials will start with ENGL 108 since developed pilot materials for this course in Summer 2018. As the most advanced class, it is also easiest to incorporate innovations into this course. The following topics, drawn from previous research and experience as a writing administrator/instructor, are likely to be included, but these will be further developed through focus-group consultation with practicing teachers (GTAs and lecturers): 1) Argument (e.g., expressing writers' stance explicitly and implicitly); 2) Informational discourse (e.g., information packaging); 3) Source use (e.g., citation patterns, including forms and functions).

Involvement of instructors in ethical, evidence-based and computationally grounded research and practice. will train practicing teachers (including micro-campus faculty) in an evidence-based practice model that allows teacher-scholars to 1) identify hypothesized areas of need based on reflective practice, a widely used approach to professional development that encourages teachers to use their own experiences as a catalyst for instructional change (Mann & Walsh, 2017); 2) engage in computational, corpus-based analysis to research that area of need; 3) provide feedback on the materials created in collaboration with 4) use those materials in their classrooms; and 5) provide evaluation of the materials to supplement the assessment of students' writing and students' evaluation. Throughout this process, discussion of ethical decisions in the use of student data, such as the informed consent process, and critical engagement with computational tools will be emphasized to help instructors and graduate student researchers to address the challenges of the 4th IR, including the use of big data to promote excellence in teaching and learning.

Assessing student learning using computational models. Using a pre/post-test quasi-experimental design, language features will be assessed in multiple drafts of a writing assignment, to see whether there is improvement in the language students are using in the specific genres they are writing after they engage with the computationally created pedagogical materials. Since texts will be available from the Crow corpus for the same semesters, same courses, and same assignments, students who did not use the computationally created materials will serve as a control group. Students' initial drafts will serve as a pre-test. Instructors who participate in the experimental group will incorporate the computationally created materials into their course sequence directly after students submit their first draft. The (revised) final draft will serve as the post-test. Computational analysis will form the basis of the assessment, targeting the language taught through the computationally created materials, but also drawing on previous research on the language used in academic and professional writing to determine whether students in the experimental group use more of the language associated with academic and professional genres after instruction (Biber et al., 1999; Staples & Reppen, 2016; Staples et al., 2016). It is expected that outcomes will show improvement in the experimental group's use of language in instructional materials as well as language that they will need for future writing in their majors and in their professions.

Triangulating computational data with human-centric methods of evaluation. While experimental designs are considered the gold standard in educational research, many educational researchers have argued for modifications to address ethics (e.g., quasi-experimental) and paring experimental designs with more human-centric approaches, including classroom observations, teacher and student evaluations of the materials, and focus groups (Hanley et al., 2016). Following this approach, participating classes will be observed (2 times/semester starting in Spring 2020) just prior to the interventions to establish a baseline understanding of classroom interaction and then during the

interventions to see how students and teachers are engaging with materials and questions that arise. Teachers and students will fill out evaluations directly after using the materials in their classes. Teacher focus groups (at least 1/semester) will be used during the initial stages of materials development and after the first round of intervention with the materials. Data from all of these sources will be used to evaluate the effectiveness of the materials and to fine-tune the intervention.

Outcomes. Outcomes will include: 1) the computationally created materials; 2) evaluation of and (anticipated) improvement of instruction on language choices appropriate for the FW context as well as models of computationally driven assessment; 3) training of instructors and graduate student teacher-researchers in a model of computationally driven ethical practice; 4) publications. The materials will initially be available to FW instructors at both the UA campus, UA's microcampuses, other units on campus that focus on the teaching of writing, and teacher-researchers at other institutions who have access to Crow. The existing Crow platform, which houses the corpus that will be leveraged to create the pedagogical materials, will provide initial dissemination of the materials (through the repository). Eventual publications will include research articles, a solo-authored monograph, and a textbook (co-authored with colleague Christine Tardy).

Timeline

Period	Activities/Outcome	Budget items

References

- Aull, L. (2015). First-year university writing: A corpus-based study with implications for pedagogy. New York: Palgrave.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.
- Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. ReCALL, 26(2), 202-224.
- Ferris, D. (2014). Language power: Tutorials for writers. New York: Bedford/St. Martin's Press.
- Hanley, Chambers, & Haslam (2016). Reassessing RCTs as the 'gold standard': Synergy not separatism in evaluation designs. *International Journal of Research & Method in Education*, 39(3), 287-298.
- Mann, S. & Walsh, S. (2017). *Reflective practice in English language teaching: Research-based principles and practices*. New York: Routledge.
- Shin, J., Velazquez, A., Swatek, A., Staples, S., & Partridge, R. S. (2018). Examining the effectiveness of corpus-informed instruction of reporting verbs in L2 first-year college writing. *L2 Journal*, 10(3), 31-46.
- Staples, S., & Reppen, R. (2016). Understanding L2 writing in first-year composition: A lexicogrammatical analysis across L1s, assignments, and writing quality. *Journal of Second Language Writing*, 32, 17-35.
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university: Grammatical complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149-183.
- Swales, J. & Feak, C. (2013). *Academic writing for graduate students: Essential tasks and skills* (3rd Ed.). Ann Arbor: University of Michigan Press.
- Tardy, C. (2009). Building genre knowledge. West Lafayette: Parlor Press.