Interdisciplinary Research (IDR) Origination Awards

Cover Page

Project Title

Increasing Access to Marginalized African Languages through Historical Dictionaries

Principal Investigator(s) (full-time faculty)

Name (PI listed first)	Department	College
Daren Ray	History	FHSS
Jim Law	French and Italian	Humanities
Earl Brown	Linguistics	Humanities

Track

Track two

Abstract: More than 2,000 languages are spoken in Africa, but Africans rely on languages of wider communication from Europe to communicate among their diverse communities. The dominance of European languages has at least three negative consequences. First, "elite closure" ensures that the 85-90% of Africans who do not attain fluency in a European language have less access to knowledge, which decreases social and economic mobility. Second, continuous elaboration of language is necessary to preserve the cultures they encode. Third, the production of knowledge in all disciplines is impeded when the concepts, values, and perspectives that are unique to African languages are marginalized. We aim to increase access to marginalized African languages by compiling historical African language dictionaries into a translingual database. Gaining access to concepts in the "semantic archive" of African languages will multiply opportunities for Africans to coin new words from their linguistic heritages. Although the linguistic data in historical dictionaries can be problematic, they are the only comprehensive data on African languages before the elevation of European languages in African societies. Compiling this database will require the team to resolve difficulties in using Optical Character Recognition (OCR) for non-standardized fonts printed in the 19th century, develop a schema for structuring disparate dictionary data, and design a relational database to analyze the data. Increasing access to a wider set of African language data will enable businesses, NGOs, and governments in developing the language resources they need to reach the vast number of Africans who are not fluent in European languages.

Summary of Plans for External Funding

Following completion of the database creation process, we will apply for an NEH Digital Humanities Advancement Grant (which focuses on the development of scalable work that enhances scholarly research, teaching, and public programming) or an NEH Collaborative Research Grant (which has previously funded a more limited linguistic database project on African languages based on contemporary data.) In addition to greatly expanding the number of dictionaries included in the database, the external grants would allow us to add dozens of additional dictionaries, enable registered scholars to compare dictionaries and other linguistic data, and create a public-facing website to access the database. If we are able to identify dictionaries of endangered African languages, we will also apply for Endangered Languages Documentation Program and NSF grants for research that preserves endangered languages. Once complete the database will also support external grants that members of the team can apply for so they can pursue research in the fields of history, historical linguistics, and corpus-based translation studies.

PROJECT NARRATIVE

1. Introduction

The Problem: Over 2,000 languages are spoken in Africa (Dalby 2006), but both colonial and post-colonial governments in Africa have supported only a few of them (e.g. Swahili, Hausa, Twi, Zulu) as national or majority languages (Ansre 1974, Ufomata 1999). Only 10% of African languages have been adopted as the medium of instruction in schools, where they are limited to primary education levels (Ouane and Glanz 2010). The formal and informal marginalization of most African languages impedes their intellectualization, the process by which languages are developed as a valid medium of communication in academic, social, technical, and public domains (Kaschula and Nkomo 2019).

Despite dissatisfaction with relying on European languages, African states have retained European languages as the medium of official communication and instruction. In addition, schooling and social pressures condition Africans to practice diglossia: they reserve conversations in their mother tongue for private situations but use European languages in the public sphere. Speakers of most African languages often use words from European languages to refer to new technical and cultural terms rather than using terminology derived from their own linguistic heritage. Scholars have identified multiple reasons for the failure of independent African governments to promote the widespread use of African languages following decolonization, including: 1) avoidance of ethnic tensions; 2) perceived expenses of developing curricula in African languages; 3) retention of ties to former imperial states; and 4) monolingual and bilingual learning paradigms that are inappropriate for the multi- and translingual contexts of African societies, among others (Clayton 1998; van Pinxteren 2021).

The dominance of European languages in African societies has at least three negative consequences. First, "elite closure" ensures that the 85-90% of Africans who do not attain fluency in a European language do not have equal access to knowledge, which decreases social and economic mobility (Scotton 1993; Ouane and Glanz 2010). Second, continuous elaboration of language is necessary to preserve the cultures they encode (Bunyi 1999, Kresse 2007). Third, the production of knowledge in all disciplines is impeded when the concepts, values, and perspectives that are unique to African languages are marginalized (Ngugi 1986; Wiredu 1995).

Some African institutions have sought to remedy the marginalization of African languages in higher education by developing specialized handbooks, curricula, and discipline-specific dictionaries. Since these resources focus narrowly on "mere translation" of terms developed in other languages, they are rarely adopted in daily conversation (Nkomo 2019). Even specialists with access to these resources continue to use terms derived from European languages. This problem extends also to commonly used vocabulary elaborated from the "semantic archives" of African languages. For example, Kiswahili speakers routinely use the word *televisheni* (television), even though it has a Kiswahili equivalent (*runinga*) coined by a Swahili poet and promoted by institutions such as the Kenya Broadcasting Corporation (Kresse 2007). Improving access to Africa's extensive linguistic heritage will help speakers of marginalized languages without access to higher education to use their African linguistic heritages to elaborate their vocabularies.

Our Solution: We will contribute to the goal of making marginalized African languages more accessible by compiling linguistic data from historical African language dictionaries created from 1840 to 1940 into a translingual database that scholars, translators, and speakers can use to elaborate African languages. It may seem counterintuitive to rely on century-old dictionaries to develop African languages since their deficiencies are well known. The missionaries who compiled the first African language dictionaries (with the indispensable assistance of Africans) were mostly amateur linguists. They maligned African religious, social, and political concepts by translating them into words with negative connotations in European

languages (Peterson, 1997). They often limited their translation efforts to the narrow goal of facilitating proselytization, biblical translation, and biblical literacy (Ansre 1974). They also disagreed about how to write even similar languages (Chimhundu 1992). In addition, scholars have demonstrated that the process of compiling dictionaries and restricting education to certain dialects diminished the diversity of African languages. These efforts promoted monolingualism and bilingualism at the expense of multilingualism and translingualism (the simultaneous use of multiple languages) (Harries, 1988; van Pinxteren 2021).

Nevertheless, this century of haphazard dictionary production created a relatively untapped source of data on historical African languages. This data is unique because it was collected before the elevation of European languages in African societies. Competition among missionaries ensured that they recorded data from multiple dialects. Their disagreements about which African languages they should promote led them to record data about many more African languages than those which governments invest in today. Since missionaries compiled these dictionaries before most of their language consultants would have been fully conversant in European languages, there is a distinct likelihood that they captured words for concepts that have since been displaced by European loanwords. Demonstrating this hypothesis is one of the many applications of the proposed database that will strengthen our application for external funding. Since missionaries compiled these dictionaries prior to standardization, they recorded linguistic data that is relatively free of influence from European languages.

For Africans to elaborate new words from their linguistic heritage, they need access to words and concepts that have gone into disuse. Compiling multiple dictionaries into a single database will enable researchers, writers, poets, and other wordsmiths to plumb historical African vocabularies, multiplying opportunities for them to craft new terminologies derived from a translingual "semantic archive" of African languages. Beyond the possibility for linguistic elaboration and intellectualization, making a wider set of African language data available will enable businesses, NGOs, and governments in the localization efforts they need to reach the vast number of Africans (85-90%) who are not fluent in European languages (e.g. Dizolele et al. 2022). Historical language data will also support the development of Human Language Technologies for those many African languages with less accessible corpora (Roux and Ndinga-Koumba-Binza, 2019).

Collaboration: We have assembled an interdisciplinary team particularly suited to overcome the challenges of compiling a database of historical African languages from non-standardized, century-old dictionaries, including Dr. Daren Ray (a historian), Dr. Jim Law, and Dr. Earl Brown (both of whom are linguists). Dr. Ray, has conducted interdisciplinary research in historical linguistics, oral traditions, and ethnography in Africa. He worked with undergraduate research assistants over the past year (2022-23) to identify dictionaries for a pilot study and will present research about a 19th century network of dictionary makers at the Rocky Mountain Workshop in African History with one of his students in April 2024.

The proposed study will focus initially on Bantu languages in eastern and southern Africa compiled by speakers of different European languages (English, French, German, Italian, and Portuguese). Dr. Ray will supervise undergraduate research assistants as they compile information on the historical and cultural contexts for the collection, publication, and dissemination of each dictionary included in the database. Dr. Jim Law, a historical linguist with research interests in semantic change will apply his experience designing linguistic databases and applying Optical Character Recognition (OCR) to non-standardized texts to process scanned images of the pilot dictionaries into a text file. He and Dr. Earl Brown, a corpus linguist, will develop computer code to parse the data in each dictionary entry for inclusion in the database. Dr. Earl Brown will also design a relational database that will enable researchers to analyze the historical linguistic data. Combining the expertise of linguists and a historian ensures that technical linguistic data conforms with disciplinary norms for language description and that the contexts in which these data were initially collected as dictionaries is well understood.

2. Data description

Dr. Ray has identified eighteen historical dictionaries, published between 1857 and 1924 that will form the foundation of the database. The dictionaries vary in the information they contain, but most include the following linguistic data: lemmas (African word entries), translations into a European language, definitions in a European language, and sample phrases in the dictionary's African language. Some dictionaries also include notes on grammar, parts of speech, word variants, and related words.

We have divided them into three sets. Set 1 includes seven bilingual dictionaries (with short definitions in English) of Bantu languages spoken in eastern and southern Africa. Set 2 includes six Kiswahili dictionaries with definitions in different European languages (English, French, German, Italian). Set 3 includes a single bilingual dictionary of a non-Bantu language (Maa) as well as five multilingual dictionaries that present data from multiple African Bantu and non-Bantu languages in the same volume. Bantu languages form the largest single group of African languages (450+) and Dr. Ray's familiarity with Kiswahili and related Bantu languages will aid in the design of the project database. Designing a database that is flexible enough to accommodate Bantu and non-Bantu African languages, as well as multiple European languages, will make it possible to scale the dictionary to accommodate many more languages, which is a primary consideration for the NEH grants we will apply for.

All of the pilot dictionaries are available in pdf format, but they were published using fonts which are no longer standard. Although some of the dictionaries have been scanned using generic, automated OCR models, Dr. Ray has observed that the accuracy is often unsuitable for computational analysis. We will verify the accuracy of the machine-readable text for dictionaries that have been scanned, and train custom OCR models to make the remaining dictionaries machine readable before creating the relational database.

3. Methodology

The methodology can be organized into four sequential steps: (1) Produce an OCR model capable of converting the page images into machine readable text, (2) using this model, produce accurate digital transcriptions of each dictionary, (3) convert the raw text into structured data within a relational database, and (4) create a front-end system to access the data. We will prioritize completing these steps with the Set 1 dictionaries. As we complete each step of the methodology with Set 1 (the most typical dictionaries of the collection), we will then make adjustments to account for the complexities introduced by the other dictionary sets (e.g., additional languages or atypical data structures).

To ensure the highest possible accuracy, we will test multiple OCR engines (e.g., Tesseract, Kraken, ABBYY), which the researchers have used previously with other data sets. Existing OCR models can be fine-tuned through a training process to increase accuracy on specific document types. Dr. Ray, with the help of research assistants, will create transcriptions of portions of the documents for this training. Dr. Law and Dr. Brown will then use these corrected transcriptions to fine-tune different OCR models in different engines until an appropriate model is achieved (at least 95% accuracy). It is anticipated that different models may need to be used for different dictionaries.

After producing digital transcriptions of the dictionaries with our fine-tuned OCR model(s), our transcriptions will need to be corrected. This will involve a combination of by-hand correction (performed by research assistants overseen by Dr. Ray) and model retraining. We may also use successful machine-transcribed pages as further input to fine-tune the model.

As we correct the raw text to achieve maximum accuracy, we will compare the entry structure of each dictionary and identify a set of standard data keys that can be used across the entire dataset. As mentioned earlier, there are commonalities and differences in the data provided within each dictionary's entries, so care will need to be taken to organize the data in such a way that it can be combined in a single dataset. Research assistants will be helpful in verifying that all data contained within each dictionary is accounted for within this structure. We will ensure that these keys organize the data so that it is appropriate for researchers in our respective fields. For example, the dataset should allow for queries based on historical

information such as the background of each dictionary's creators, linguistic information such as the morphological and phonological forms of each word, and lexicographic information such as listed synonyms. We will convert the raw text into a structured data format (JSON) that applies these standardized keys. The data will then be imported into a relational database.

Finally, we will develop a front-end system to access the data. In its early form, this will be an internal system to be used by the researchers. It will form the basis for an eventual public-facing website (funded with a subsequent, external grant) that will allow external researchers to query the data. Each of the three researchers will contribute to the design of this tool so that it is suited to historical and linguistic research.

4. Schedule

Project Timeline

	Year 1 2024-2025												Year 2 2025-2026											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Historical Research																								
ASA Conference																								
OCR Transcriptions																								
Set 1																								
Set 2																								
Set 3																								
TXT > JSON Conversions																								
Set 1																								
Set 2																								
Set 3																								
Database (Re)design and Data entry																								
Set 1																								
Set 2																								
Set 3																								
Digital humanities conference																								
Applications for External Grants																								

- **Historical research:** Dr. Ray and the research assistants will research the historical contexts for the collection and publication of one dictionary each month during each academic school year of the grant.
- **ASA Conference:** Dr. Ray will present initial research and a summary of the project at the African Studies Association Annual Meeting in November 2024 (Chicago, IL). The conference will provide opportunities to discuss the project with scholars who have successfully received NEH grants for similar linguistic database projects.

- **OCR Transcriptions:** We will perform optical character recognition (OCR) of the PDF files of the dictionaries and create text files (TXT), one text file per dictionary.
- **TXT > JSON:** Dr. Law and Dr. Brown will use the text files created though the OCR process to create structured data files in JSON format. The creation of the JSON files for Sets 1, 2 and 3 will be completed in Jun 2024, December 2025, and April 2026, respectively.
- **Database (Re)design and Data entry:** Dr. Brown will complete a working prototype of a relational database for the Set 1 dictionaries by December 2025. We plan to complete entry of Set 2 by April 2025 and Set 3 by August 2025.
- **Digital humanities conference:** We will present an academic paper at the African Digital Humanities Conference held in Ghana in Feb. 2026. The precise location in 2026 for this annual conference hasn't yet been determined, but if for some reason this conference isn't feasible, we will present our paper at another academic conference held in Africa that year.
- **Applications for External Grants:** We will apply for a National Endowment for the Humanities Digital Humanities Advancement Grant to create a public facing, scalable database in June 2026. If unsuccessful we will apply again in June 2027 or shift focus to a NEH Collaborative research grant, whose deadline is in November 2027. NEH rules prevent simultaneous application for these grants. See the list of potential external grants for a full list and timeline of our intended applications.

5. Expected Outcomes

By the end of the IDR funding period, we aim to achieve the following outcomes. First, we will complete a prototype database containing all eighteen dictionaries. Second, we will present the database to stakeholders who are actively supporting marginalized African languages. For example, we will present the database at an African digital humanities conference held annually in Ghana. Stakeholders at this conference include representatives from African universities and scholars working in the field of digital humanities who specialize in designing database that account for the distinctive features of African languages. The conference will allow us to receive feedback on the database to maximize its usefulness for increasing access to marginalized African languages for scholars and the general public. Third, we will develop a streamlined workflow to process additional dictionaries for incorporation into the database. This will allow us to demonstrate to external funders that we can quickly scale the database to include many more dictionaries. Fourth, we will submit grant proposals for external funding to expand our prototype database to a publicly accessible database. (See below for descriptions of these external grant opportunities).

These outcomes will support our long-term goals for later stages of the project funded by external grants. First, we will expand the prototype database of eighteen dictionaries to a large database that includes historical dictionaries and related resources for an estimated 200 African languages. This expansion will increase the usefulness of the database for researchers in history and linguistics. In addition, the external funding will support the creation of a public-facing website that allows user-friendly access to the database, including communities with minimal access to computing resources. Providing a public portal for this larger database will aid a wider variety of communities outside of academic circles to investigate concepts from multiple African languages that would otherwise be difficult to access.

Once the database is functional, each member of the team will also be positioned to apply for external grant applications that will support multiple publications in their respective fields. Dr. Ray will pursue historical research on how colonization impacted Africans' religious and social concepts, categories, and terms. Dr. Law will use the database to investigate trends in lexical semantic change. Dr. Brown will pursue research on corpus-based translation.

BUDGET NARRATIVE

Item	Year 1	Year 2	Total	Notes
Two Undergraduate Research Assistants	$14,400	$14,400	$28,800	10 hrs/wk, 48 weeks $15/hr, 2 assistants, annually
Computer equipment	$1,700		$1,700	Dell, Core I-7, 16 GB Ram, 512 SSD hard drive, with dedicated GPU, monitor, and external hard drive for backup
ABBYY Fine Reader Subscription	500	500	1000	$100/ annual subscription, 5 researchers, including assistants
Consulting Services	250	250	500	5/hrs a year, 50$/hr
Faculty Travel	$2,000	$9000	$11,000	
Total	$18850	$24,150	$43,000	

We request **$40,000** over the two-year period for this project. **$28,800** will be distributed as wages to the undergraduate research assistants supervised by Dr. Ray. They will assist in two primary activities 1) historical research into the compilation of each dictionary included in the database; 2) hand checking OCR transcripts to refine and train the OCR model(s). Undergraduate students will complete their work in the History Department's student research laboratory. Two assistants will be necessary to optimize data validation. **$1700** will purchase a dedicated desktop computer for use by the research assistants.

An additional **$1,000** will be spent on subscriptions for software and cloud services to support Optical Character Recognition: $1000 for two years of ABBYY subscriptions for researchers and assistants. Kraken and Tesseract are open-source software, and the Office of Digital Humanities will provide server spaces for hosting those services at no cost to the project through its own application process.

$500 ($250 annually) are reserved to consult with Dr. Troy Spier, Assistant Professor of English and Linguistics, who previously designed a database for Bantu African languages with funding from the National Endowment for the Humanities.

$2000 will cover Dr. Ray's travel to one domestic conference (the African Studies Association Annual Meeting in November 2024).

$9000 will enable three team members to attend an international conference in Africa (ideally, the Symposium on African Digital Humanities in Ghana) in February 2026.

We will seek international travel grants from BYU's David M. Kennedy Center for International Studies, to cover the shortfall of $3000 ($1000 per person).

References

Ansre, Gilbert. 1974. “Language Standardisation in Sub-Saharan Africa.” In *Advances in Language Planning*, edited by Joshua A. Fishman, 368–89. Contributions to the Sociology of Language 5. Boston: De Gruyter Mouton. https://doi.org/10.1515/9783111583600.

Bunyi, Grace. 1999. “Rethinking the Place of African Indigenous Languages in African Education.” *International Journal of Educational Development* 19 (4): 337–50. https://doi.org/10.1016/S0738-0593(99)00034-6.

Chimhundu, Herbert. 1992. “Early Missionaries and the Ethnolinguistic Factor During the ‘Invention of Tribalism’ in Zimbabwe.” *Journal of African History* 33 (1): 87–109.

Clayton, Thomas. 1998. “Explanations for the Use of Languages of Wider Communication in Education in Developing Countries.” *International Journal of Educational Development* 18 (2): 145–57. https://doi.org/10.1016/S0738-0593(98)00002-9.

Dalby, Andrew. 2006. *Dictionary of Languages: The Definitive Reference to More Than 400 Languages*. London: Bloomsbury Publishing Plc.

Dizolele, Mvemba Phezo, Jacob Kurtzer, and Hareem Fatima Abdullah. 2022. “Localizing Humanitarian Action in Africa » Philanthropy Circuit.” Philanthropy Circuit. August 16, 2022. https://philanthropycircuit.org/blog/localizing-humanitarian-action-in-africa/.

Harries, Patrick. 1988. “The Roots of Ethnicity: Discourse and the Politics of Language Construction in South-East Africa.” *African Affairs* 87 (346): 25–52.

Kaschula, Russell H., and Dion Nkomo. 2019. “Intellectualization of African Languages: Past, Present, and Future.” In *The Cambridge Handbook of African Linguistics*, edited by H. Ekkehard Wolff, 601–22. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108283991.021.

Kresse, Kai. 2007. *Philosophizing in Mombasa: Knowledge, Islam, and Intellectual Practice on the Swahili Coast*. International African Library. Edinburgh: Edinburgh University Press.

Nkomo, Dion. 2019. “Theoretical and Practical Reflections on Specialized Lexicography in African Languages.” *Lexikos* 29 (1): 96–124.

Ouane, Adama, and Christine Glanz. 2010. “Why and How Africa Should Invest in African Languages and Multilingual Education: An Evidence- and Practice-Based Policy Advocacy Brief.” Hamburg: UNESCO Institute for Lifelong Learning. https://unesdoc.unesco.org/ark:/48223/pf0000188642.

Peterson, Derek. 1997. “Colonizing Language? Missionaries and Gikuyu Dictionaries, 1904 and 1914.” *History in Africa* 24 (January): 257–72. https://doi.org/10.2307/3172029.

Pinxteren, Bert van. 2022. “Language of Instruction in Education in Africa: How New Questions Help Generate New Answers.” *International Journal of Educational Development* 88 (January): 102524. https://doi.org/10.1016/j.ijedudev.2021.102524.

Roux, Justus C., and H. Steve Ndinga-Koumba-Binza. 2019. “African Languages and Human Language Technologies.” In *The Cambridge Handbook of African Linguistics*, edited by H. Ekkehard Wolff, 623–44. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108283991.022.

Scotton, Carol Myers. 1993. “Elite Closure as a Powerful Language Strategy: The African Case.” *International Journal of the Sociology of Language* 103 (1). https://doi.org/10.1515/ijsl.1993.103.149.

Thiong’o, Ngugi wa. 1986. *Decolonising the Mind: The Politics of Language in African Literature*. Oxford: Heinemann.

Ufomata, Titi. 1999. “Major and Minor Languages in Complex Linguistic Ecologies: The Nigerian Experience.” *International Journal of Educational Development* 19 (4): 315–22. https://doi.org/10.1016/S0738-0593(99)00031-0.

Wiredu, Kwasi. 1995. *Conceptual Decolonization in African Philosophy*. Edited by O. Oladipo. Lagos: Hope Publishers.

Plans for External Grant Applications

	Request Amount	Planned Submission	Lead PI
Database development grants			
NEH Digital Humanities Advancement Grant	$75,000	June 2026	Daren Ray
Legacy Material Grant by the Endangered Languages Documentation Program	$10,880 (10,000 Euros)	June 2026	Jim Law
NSF Dynamic Language Infrastructure – NEH Documenting Endangered Languages (DLI-DEL)	$450,000	February 15, 2027	Jim Law
NEH Collaborative Research Grant	$250,000	November 2027	Daren Ray
Grants using the database			
Bernadotte E. Schmitt Grant for Research in European, African, or Asian History (American Historical Association)	$1500	February 15, 2027	Daren Ray
NSF Linguistics Program	$300,000	July 15, 2027	Jim Law or Earl Brown

The current award will support development of a database consisting of 18 African language dictionaries. Completion of the database will make our team competitive to apply for two categories of funding: 1) grants to increase the scale and access of the database, 2) grants for research using the database.

We will apply for a NEH Digital Humanities Advancement Grant in June 2026, once we can demonstrate the scalability of our database. NEH policy prohibits us from seeking funds for a single project simultaneously. So we will apply for an NEH Collaborative Research Grant in November 2027 to focus on the larger aim of improving access to marginalized languages through our translingual database. This NEH grant has funded a more limited linguistics database project based on contemporary Bantu African language data. It is specifically designed to support research by interdisciplinary teams from multiple institutions. Seeking input from African stakeholders early in our design process will help us identify potential partners for an expanded team, including language specialists in Africa and experts in "minimal computing" who will help make the database accessible to communities in Africa that have minimal information technology infrastructure.

If we are able to identify dictionaries of African languages that are currently endangered, we will also apply for a Legacy Material Grant by the Endangered Languages Documentation Program in June 2026 and a NSF Dynamic Language Infrastructure – NEH Documenting Endangered Languages (DLI-DEL) in February 2027. Both grants support the digitization of materials that describe endangered languages, as well as linguistic research in those languages.

We will also pursue individual grants to support more specialized research in our respective disciplines from the American Historical Association (Dr. Ray) and the NSF Linguistics Program (Dr. Law and Dr. Brown).

Biographical Sketch

Daren Ray
Department of History
Brigham Young University, Provo, UT 84602
daren.ray@byu.edu, 801-422-0510

1) Professional Preparation

PhD, African History, University of Virginia, May 2014
MA, African History, University of Virginia, December 2008
BA, History, Brigham Young University, April 2006

2) Professional Appointments

Aug. 2020 –	Assistant Professor, History Dept., Brigham Young University
Aug. 2015 – Aug. 2020	Assistant Professor, History Dept., Auburn University
July 2014 – June 2015	Visiting Assistant Professor, History Dept., Connecticut College
July 2013 – June 2014	Visiting Assistant Professor, History Dept., The American University in Cairo
February – May 2011	Instructor, History Dept. Roanoke College

3) Publications

1. *Ethnicity, Identity, and Conceptualizing Community in Indian Ocean East Africa* (Ohio University Press, 2023)

2. "Recycling Interdisciplinary Evidence: Abandoned Hypotheses and African Historiologies in the Settlement History of Littoral East Africa" *History in Africa* 49, June 2022, https://doi.10.1017/hia.2022.7

3. "Celebrating Swahili New Year: A Performative Critique of Textual Islam in Coastal Kenya," *Muslim World Journal* 105:4, October 2015, https://doi.10.1111/muwo.12112

4. Defining the Swahili" in *The Swahili World*, ed. Stephanie Wynne-Jones and Adria LaViolette (Routledge) 2018, https://doi.10.4324/9781315691459-5

5. "From Constituting Communities to Dividing Districts: The Formalization of a Cultural Border between Mombasa and its Hinterland," in *Borderlands in World History, 1700-1914*, ed. Cynthia Radding, Paul Readman, Chad Bryant (Palgrave MacMillan) 2014 https://doi.org/10.1057/9781137320582_6

4) Relevant Peer-reviewed presentations

1. "Muyaka's Lament: Poetic Memory and the Surrender of Mombasa, 1815-1840," Archive and Public Culture Research Lab, University of Cape Town, South Africa, 12 November

2. "Swahili Swords and Sabaki Words: Expressions of Warfare in Nineteenth-Century Kiswahili Literature," African Studies Association Annual Meeting, 1 December

5) Professional Memberships

American Historical Association
African Studies Association

Biographical Sketch

James Law
Department of French and Italian
Brigham Young University, Provo, UT 84602
jimlaw@byu.edu, 801-422-2782

1) Professional Preparation

PhD, French Linguistics, University of Texas at Austin, 2020
MA, French Linguistics, University of Texas at Austin, 2017
BA, French Studies, Brigham Young University, 2015

2) Professional Appointments

Assistant Professor, Department of French and Italian, Brigham Young University, 2020- present

3) Publications

1. Law, J. (2023). Constructional change and frame element selection: Insights from the French Spending frame. *Constructions and Frames* 15(1), 120-145

2. Law, J. (2023). Regularity of semantic change in Romance anatomical terms. *Journal of Historical Linguistics* 13(2), 295-325

3. Law, J. (2022). Reflections of the French nasal vowel shift in orthography on Twitter. *Journal of French Language Studies* 32(2), 197-215

4. Law, J. (2022). Metonymy and argument alternations in French communication frames. *Cognitive Linguistics* 33(2), 387-413

5. Law, J. (2022). Frame-based metonymy in teaching L2 vocabulary. In H. C. Boas (ed.) Directions for Pedagogical Construction Grammar: Learning and Teaching (with) Constructions. Berlin: De Gruyter, 305-332

6. Law, J., D. Barny & R. Poulin (2020). Patterns of peer interaction in multimodal L2 digital social reading. *Language Learning and Technology* 24(2), 70-85

7. Law, J. (2019). Diachronic frame analysis: The Purpose frame in French. *Constructions and Frames* 11(1), 42-77

8. Law, J. (2018). Conceptualizations of time in French depuis 'since, for' constructions. *Cognitive Linguistics* 29(2), 163-195

4) Peer-reviewed presentations

1. Law, J. Metaphor and pragmaticalization of Romance motion verbs. 16th International Cognitive Linguistics Conference, Heinrich Heine University, Düsseldorf, Germany. August 7-11, 2023.

2. Law, J. & A. F. McBride. Variable articulations of the fricative /ʒ/ in the Poitevin-Saintongeais language of France. 53rd Linguistic Symposium on Romance Languages, Paris, France. June 26-30, 2023.

3. Law, J. Text genre and participant role alternations in the French Spending frame. 7th International Conference on Grammar & Text, Universidade NOVA, Lisbon, Portugal. November 18-20, 2021.

4. Law, J. A lexicogrammatical approach to participant role alternations in the French Spending frame. Corpus Approaches to Lexicogrammar, Edge Hill University, Omskirk, UK. July 3, 2021.

5. Law, J. Revealing the Secret of a French valency pattern alternation. 94th Annual Meeting of the Linguistic Society of America, New Orleans, LA. January 2-5, 2020.

6. Law, J. Iconic and indexical semantic change in Romance body-part terms. 24th International Conference on Historical Linguistics, Australian National University, Canberra, Australia. July 1-5, 2019.

7. Brozovsky, E., L. Hinrichs, B. Kaufman, J. Law, L. Orjuela, M. Wahl, and J. Wolfgang. How whom retreated against the advice of prescriptive grammarians: A multivariate analysis of written English corpus data since 1800. 93rd Annual Meeting of the Linguistic Society of America, New York, NY, January 3-6, 2019.

8. Law, J. Diachronic frame analysis: The Purpose frame in French. 10th International Conference on Construction Grammar, Université Paris 3: Sorbonne Nouvelle, Paris, France, July 16-20, 2018.

9. Law, J. The frames approach to metonymic change: The Purpose frame in French. 48th Linguistic Symposium on Romance Languages, York University, Toronto, ON, April 25-28, 2018.

10. Law, J. Frame-based metonymy in teaching L2 vocabulary. Constructionist Approaches to Language Pedagogy 3, The University of Texas at Austin, Austin, TX, February 15-17, 2018.

11. Law, J. Les jons fon skil veulent: Reflections of the French nasal vowel shift in variant orthography. New Ways of Analyzing Variation 46, University of Wisconsin-Madison, Madison, WI, November 2-5, 2017.

12. Law, J. A corpus-based analysis of the semantic change chercher ‘look for’ > ‘provoke’ in French. Pacific Ancient and Modern Language Association, Pasadena, CA, November 11-13, 2016.

5) Professional Memberships

International Cognitive Linguistics Association (ICLA)

Biographical Sketch

Earl Kjar Brown
Department of Linguistics
Brigham Young University, Provo, UT 84602
ekbrown@byu.edu, 801-422-3970

1) Professional Preparation

PhD, Hispanic Linguistics, University of New Mexico, 2008
MA, Spanish Linguistics, Brigham Young University, 2003
BA, Spanish, Brigham Young University, 2001

2) Professional Appointments

Professor of Linguistics, Brigham Young University, 2022- present
Associate Professor of Linguistics, Brigham Young University, 2017-2022
Associate Professor of Spanish, Kansas State University, 2015-2017
Assistant Professor of Spanish, Kansas State University, 2012-2015
Assistant Professor of Spanish, California State University, Monterey Bay 2008-2012
Visiting Instructor of Spanish, Auburn University, 2007-2008

3) Publications

1. Woods, Kelly, Brett Hashimoto & Earl K. Brown. (2023). "A multi-measure approach for lexical diversity in writing assessments: Considerations in measurement and timing." Assessing Writing, 55: 100688. doi:10.1016/j.asw.2022.100688
2. Brown, Earl Kjar. (2023). "Cumulative exposure to fast speech conditions duration of content words in English." Language Variation and Change, 35(2): 153-173. doi:10.1017/S0954394523000157
3. Brown, Alan V., Stayc DuBravac, Michael Toland, Earl Kjar Brown. (2023). "Metalinguistic awareness in Spanish partial immersion and English-only elementary students." Journal of Spanish Language Teaching. doi:10.1080/23247797.2023.2202003.
4. Brown, Earl Kjar. (2023). "Repetition and procedural knowledge of sound patterns." *The Handbook of Usage-Based Linguistics*, ed. by Manuel Díaz-Campos y Sonia Balasch. 127-144. Wiley Blackwell
5. Gradoville, Michael S., Earl Kjar Brown, Richard J. File-Muriel. (2022). "The phonetics of sociophonetics: Validating acoustic approaches to Spanish /s/." *Journal of Phonetics*, 91: 101- 125.
6. Heilpern, James A., Earl Kjar Brown, Zachary D. Smith & William G. Eggington. (2022). "Generic Ab Initio." *Buffalo Law Review*, 70: 613-694.
7. Brown, Esther L., William D. Raymond, Earl Kjar Brown, Richard J. File-Muriel. (2021) "Lexically specific accumulation in memory of word and segment speech rates." *Corpus Linguistics and Linguistic Theory*, 17(3): 625-651.
8. Linford, Bret, Alicia Harley, & Earl K. Brown. (2021). "Second Language Development of Phonetic Reduction in Spanish During Study Abroad: The Case of /s/-Weakening." *Studies in Second Language Acquisition*, 43(2): 403-427.
9. Eddington, David Ellingson & Earl Kjar Brown. (2021). "A production and perception study of /t/ glottalization and oral releases following glottals in the US." *American Speech,* 96(1): 78- 104.
10. Brown, Earl K., Richard J. File-Muriel & Michael S. Gradoville. (2021). "The last stronghold of word-final /s/ in Barranquillero Spanish: Prevocalic word-final /s/ in cohesive bigrams." *The Routledge Handbook of Variationist Approaches to Spanish*, ed. by Manuel Díaz-Campos. 113- 124. Routledge.

11. Richard J. File-Muriel, Earl K. Brown & Michael S. Gradoville. (2021). "A sociophonetic approach to /s/-realization in the Colombian Spanish of Barranquilla." *Sociolinguistic Approaches to Sibilant Variation in Spanish*, ed. by Eva Núñez-Méndez, 246-261. Routledge.
12. Brown, Alan V., Yanira Paz & Earl Kjar Brown. (2021). *El léxico-gramática del español: Una aproximación mediante la lingüística de corpus* 'The Lexical Grammar of Spanish: A Corpus-based Approach'. Routledge.

4) Peer-reviewed presentations

1. Brown, Earl Kjar, Brett Hashimoto, Scott Jarvis. "Operationalizing sub- and multi-word units in measures of lexical diversity." To be presented at the annual conference of the American Association for Applied Linguistics in Houston, Texas in March 2024.
2. Brown, Earl Kjar, Brett Hashimoto, Alan V. Brown. "Exploring the predictive power of multiple lexical diversity measures for L2 Spanish writing proficiency." *Hispanic Linguistics Symposium.* Brigham Young University, Provo, Utah, October 2023.
3. Heilpern, James, Earl Kjar Brown, Zachary Smith, William Eggington. "Going Generic: A Corpus Linguistics Approach to Genericity Determinations in Trademark Law." *8th Annual Law & Corpus Linguistics Conference.* Brigham Young University, Provo, Utah, October 2023.
4. Brown, Earl Kjar, Brett Hashimoto, Alan V. Brown. "Exploring the predictive power of multiple lexical diversity measures for L2 Spanish writing proficiency." *Corpus Linguistics 2023.* Lancaster University, England, July 2023.
5. Hashimoto, Brett, Catherine Marshall, Earl Kjar Brown. "Lexical Complexity in the United States Code." *American Association for Applied Linguistics.* Portland, Oregon, March 2023.
6. Brown, Alan V., Troy Cox, Greg Thompson, & Earl K. Brown. "Fluency and the linguistic complexity of Spanish OPIcs by level." *Hispanic Linguistics Symposium.* Online, hosted by Arizona State University, November 2022.
7. Hashimoto, Brett and Earl Kjar Brown. "Lexical complexity in the US Code." *American Association for Corpus Linguistics.* Northern Arizona University, September 2022.
8. Woods, Kelly, Brett Hashimoto, Earl Kjar Brown. "Predicting writing proficiency using various lexical diversity measures." *American Association for Applied Linguistics.* Pittsburgh, Pennsylvania, March 2022.
9. Brown, Alan, Yanira Paz, Earl Kjar Brown. "Una aproximación a la gramática del español a través de la lingüística de corpus" 'A corpus-linguistics approach to Spanish grammar.' *XVIII Encuentro Internacional de GERES.* Online, hosted by Pontificia Universidad Católica Argentina (Buenos Aires). June 2021.
10. Brown, Alan, Yanira Paz, Earl Kjar Brown. "Understanding Spanish L2 Acquisition through L2 Corpora: Initial Results of Automated Analysis." *American Association for Applied Linguistics.* Online, March 2021.
11. Brown, Alan, Yanira Paz, Earl Kjar Brown. "Automated Analyses of Spanish L2 Corpora: Initial Results." *Second Language Research Forum.* Online, hosted by Vanderbilt University in Nashville, Tennessee, October 2020.
12. Smith, Zachary, James Heilpern, Bill Eggington, Earl K. Brown. "Trademark Genericness and Corpus Linguistics." *5th Annual Law & Corpus Linguistics Conference.* Provo, Utah, February 2020.

5) Professional Memberships

American Association for Applied Linguistics (AAAL)

Current and Pending Support

None.