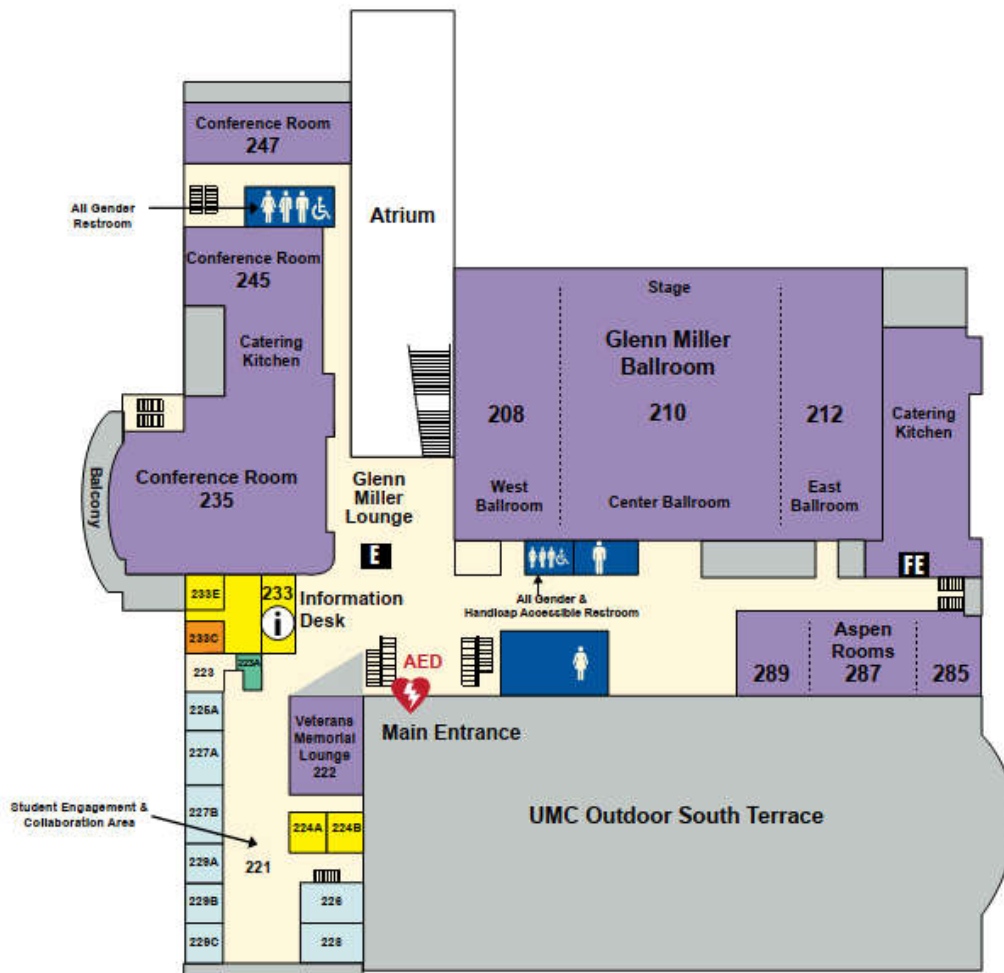




Second Floor



OFFICE.....	ROOM #
Aspen Rooms.....	285, 287, 289
Conference Rooms	235, 245, 247
CU NightRide	233C
Glenn Miller Ballroom.....	208, 210, 212
Quiet Space	223
Student Engagement & Collaboration Area	221
UMC Information Desk	233
UMC Security	233E
Veterans Memorial Lounge	222
Washroom (CU affiliates only)	223A

- Administrative Office
- Conference Room
- Restroom
- Washroom
- Student Organization Meeting Space
- Student Service
- ♥ Automated External Defibrillator (AED)
- E Elevator
- FE Freight Elevator
- i Information

Welcome to Boulder!

We are happy to see so many familiar and new faces at our first in-person meeting since 2019. This conference has been four years in the making, with a rather large hiatus in the middle when the pandemic forced us to cancel the original 2021 date and organize a virtual conference in its place.

The longer time frame has enabled us to gather the people and resources required to put together a conference unparalleled in our history with regard to its specific foci, both of which are very much in tune with the times.

Over the life of the DSNA (founded in 1975), lexicography has transitioned from an ink-and-paper-based endeavor to one that is digitally facilitated at every step of the way: research, compiling, editing, and publishing. Dictionaries and lexicons have benefited greatly from these advances in technology. They have also contributed significantly to the ways that language and technology interact, and that is the subject of Martha Palmer's workshop on Saturday. Additionally, we have several presentations that explore or highlight the use of digital resources and platforms for the making of dictionaries.

The University of Colorado Boulder, Colorado's flagship public university, honors and recognizes the many contributions of Indigenous peoples in our state. To quote from the university's Land Acknowledgement: "CU Boulder acknowledges that it is located on the traditional territories and ancestral homelands of the Cheyenne, Arapaho, Ute and many other Native American nations. Their forced removal from these territories has caused devastating and lasting impacts. While the University of Colorado Boulder can never undo or rectify the devastation wrought on Indigenous peoples, we commit to improving and enhancing engagement with Indigenous peoples and issues locally and globally."

To that end, our meeting this year, for the first time in its history, has a focus on the lexicography of Indigenous North American languages. We have two panels addressing challenges and opportunities in these languages: one from Canada, organized by Mark Turin and Christine Schreyer, and the plenary panel that has been organized by CU's own Andy Cowell, who is also the Faculty Director of Center for Native American and Indigenous Studies (CNAIS).

I hope you will enjoy this opportunity to meet like-minded colleagues from all over the continent and the world, while also taking time to explore some of the riches that Boulder has to offer.

Orin Hargraves
conference organizer

Wednesday 31 May

Registration and Accommodation Check-in 2 – 5 pm, University Memorial Center (UMC).

Opening reception: 5 pm – 6:30 pm, UMC South Terrace marquee

An opportunity meet and greet old friends and new, including:

Conference volunteers, all CU grad and undergrad students in linguistics

- Sarah Adams
- Sijia Ge
- Anushri Kartiknarayan
- Jaimie Jettmar

The conference committee

- Ed Finegan, DSNA president
- Kory Stamper, DSNA vice president and president-elect
- Lindsay Rose Russell, DSNA executive director
- Orin Hargraves, conference organizer

CU faculty participating in the conference

- Prof. Laura Michaelis-Cummings, Linguistics chair and conference faculty sponsor
- Prof. Martha Palmer, Arts and Sciences Professor of Distinction for Linguistics, and the former Helen & Hubert Croft Professor of Engineering in the Computer Science Department
- Prof. Andy Cowell, Professor of Linguistics and Faculty Director of the Center for Native American and Indigenous Studies (CNAIS)
- Prof. Alexis Palmer, Linguistics

Dinner on your own.

DSNA-24 Schedule

31 May – 3 June 2023

Thursday 1 June morning

Time	Conference Room (UMC 235)	Aspen Rooms (UMC 285-7)
8:30 – 8:50	Conference orientation	
9 – 10:30	<p><i>Session Chair: Rachel Stone</i></p> <p>Bill Poser, Special Features for Place Names in Electronic Dictionaries</p> <p>Robert Krovetz, Soup to Nuts: A Program to Automatically Induce a Lexicon of Multiword Expressions and Re-Tokenize a Corpus</p> <p>Dorian Cougias, A proposed architecture for a federated dictionary structure</p>	<p><i>Session Chair: Traci Nagle</i></p> <p>Anatoly Liberman, Idiom Dictionaries as a Genre</p> <p>Rod McConchie, Free dictionaries for a greater purpose: The use of promotional dictionaries</p> <p>Abolfazl Alamdar, The place of Persian light verbs in active bilingual dictionaries (video presentation)</p>
10:30 – 11	Break (refreshments at the back of the Aspen Rooms, 289)	
11 – 12:30	<p><i>Session Chair: Charles Carson</i></p> <p>Jesse Sheidlower, The Historical Dictionary of Science Fiction: An Introduction</p> <p>Grant Barrett, Defin-o-bots: Challenging AI to Create Usable Dictionary Content</p> <p>Ligeia Lugli, Democratizing digital lexicography: towards an accessible tool for building fully-customizable online dictionaries</p>	<p><i>Session Chair: Donna Farina</i></p> <p>Olga Menagarishvili, What Do Cultural Studies and Scientific and Technical Communication Have to Do with Dictionaries?</p> <p>Stefan Dollinger, DCHP-3 and the new “Consortium” Dictionary of Canadian English: projects, problems, prospects in the longue durée of language study in Canada, 1946–2023</p> <p>Lindsay Rose Russell, The Life and Lexicography of Frances H. Willard</p>
12:30 – 2	Lunch (the Hill, Alfred Packer Grill, or C4C (for on-campus residents))	

DSNA-24 Schedule

31 May – 3 June 2023

Thursday 1 June afternoon

Time	Conference Room (UMC 235)	Aspen Rooms (UMC 285-7)
2 – 3:30	Panel: Mark Turin, Christine Schreyer, Hlagñit Frances Brown, Ewa Czaykowska-Higgins, Mandy Jimmie, Megan Lukaniec, John Lyon, Kathleen Michel, Relational Lexicography: New Frameworks for Community-Informed Dictionary Work with Indigenous Languages	<i>Session Chair: Vicki Neufeldt</i> Nguyễn Văn Bền, English Version of a Media Dictionary Rachel Fletcher , EDIB terminology and meeting audience needs: (re)defining the terminology of diversity and inclusivity for language learners and publishing professionals Jason Siegel , Lexicography in education: Dictionary skills and knowledge for older students
3:30 – 3:50	Break (light refreshments at the back of the Aspen Rooms, 289)	
3:50 – 5:20	Relational Lexicography Panel continues.	<i>Session Chair: Jason Siegel</i> Ana Salgado et al , Domain labelling in the Morais dictionary: bringing structure to unstructured lexicographic data Daniele Trevelin Donato-Luz and Regiani Aparecida Santos Zacarias , Technical-Lexicographical Project For Selecting Examples For A Pedagogical Portuguese-English Dictionary Of Verbs

Dinner on your own.

DSNA-24 Schedule

31 May – 3 June 2023

Friday 2 June morning

Time	Conference Room (UMC 235)	Aspen Rooms (UMC 285-7)
9 – 10:30	Plenary Panel: Andy Cowell, Patricia Anderson, Nora Livesay, Sterling Martin, Lexical Resources for Indigenous Languages: Issues and Opportunities	
10:30 – 11	Break (refreshments at the back of the Aspen Rooms, 289)	
11 – 12:30	<p><i>Session Chair: Orión Montoya</i> Boban Dedović, Allo: A Modern Digital Dictionary Platform for Ancient Languages Félix Cortés and Iara Mantenuto, San Sebastián del Monte Mixtec Dictionary Project Anna Luisa Daigneault and Gregory D. S. Anderson, Living Dictionaries: A Platform for the Indigenous Languages of the Americas (video presentation)</p>	<p><i>Session Chair: Lisa Berglund</i> Linda Mitchell, Adapting Lexicons to Middle Class Eighteenth-Century England Kevin Rottet, Paronymic Attraction and Folk Etymology in Louisiana Daniele Franceshi, information session regarding the European Master in Lexicography course, Università degli Studi "Roma Tre"</p>
12:30 – 2	Lunch (the Hill, Alfred Packer Grill, or C4C (for on-campus residents))	

DSNA-24 Schedule

31 May – 3 June 2023

Friday 2 June afternoon

Time	Conference Room (UMC 235)	Aspen Rooms (UMC 285-7)
2 – 4	<p><i>Session Chair: Amber Frey</i> Ewa Czaykowska-Higgins, The digital and print dictionaries of nxaʔamxćin: challenges and responsibilities John Lyon and Kathleen Michel, The Upper Nicola Okanagan Talking Dictionary Project Helen Zhang, Collaborative dictionary for Hul'q'umi'num' using a low-code platform George Aaron Broadwell, Timucua, A language isolate of Florida: Constructing a dictionary from 17th century texts</p>	<p>Panel: Lynne Murphy, Marc Alexander, Fraser Dallachy, Daniele Franceschi, Rachel Stone, Perspectives on synonym relations</p>
4 – 4:15	<p>Break (light refreshments at the back of the Aspen Rooms, 289)</p>	
4:15 – 5	<p><i>Session Chair: Ed Finegan</i> David Skinner, The Most Controversial Dictionary (and the man who made it). Introduction to the exhibit.</p>	
5 – 5:30	<p>DSNA Business meeting in UMC 247. All are welcome.</p>	
6:30 – 8:30	<p>Banquet. Ed Finegan, A Life in Lexicography. UMC South Terrace</p>	

DSNA-24 Schedule

31 May – 3 June 2023

Saturday 3 June morning

Time	Conference Room (UMC 235)	Aspen Rooms (UMC 285-7)
9 – 10:30	<p>Workshop: Martha Palmer, Julia Bonn, Matt Buchholz, Andy Cowell, Jan Hajic, Pavlina Kalm, Skatje Myers, Alexis Palmer, James Pustejovsky, Nianwen Xue, The role of lexicons and lexicography in Natural Language Processing (see https://dsna-workshop.github.io/be for more detail)</p>	<p><i>Session Chair: Connie Eble</i> Beth Young, Using ElasticSearch in EEBO and ECCO to Identify Johnson’s Dictionary Quotation Sources Lisa Berglund, Biography and Character Assassination: Hester Lynch Piozzi in the OED Volker Harm, Between admiration and rejection: Grimm’s Deutsches Wörterbuch and its difficult relation with the Oxford English Dictionary</p>
10:30 – 11	<p>Break (refreshments at the back of the Aspen Rooms, 289)</p>	
11 – 12:30	<p>NLP Workshop continues</p>	<p><i>Session Chair: Steve Kleinedler</i> Brad Montgomery-Anderson, The Russian-English Dostoevsky Dictionary: The bilingual literary dictionary as corpus and learning tool Donna M.T.Cr. Farina et al, Investigating the Dictionaries and Reference Books for 20th-Century Slovenian Immigrants in the U.S. Joseph T. Farquharson, From possessive noun phrase to interjection: Mapping a lexicalization network in an English-related Creole</p>
12:30 – 2	<p>Lunch (the Hill, Alfred Packer Grill, or C4C (for on-campus residents))</p>	

Abstracts

Abstracts are alphabetized by the surname of the first (or only) author. Capitalization in titles follows the submitter's style in most cases.

Abolfazl Alamdar *The place of Persian light verbs in active bilingual dictionaries*

Dictionaries, especially in translation, can be divided into two types of passive (to decode or comprehend a text) and active (to encode or produce a text). It is clear that compiling an active dictionary is more difficult because in such dictionaries, it is necessary to provide extensive information about collocations and different linguistic structures in the target language.

Almost all Persian bilingual dictionaries are passive. In compiling an English-Persian active dictionary, the most challenging issue is the nature of Light Verbs (LVs) in Persian. In this language, a large part (approximately 90%) of the verbal meanings are made with noun + LV collocations. Even foreign nouns are used in combination with LVs and a compound verb is made, e.g. e-mail *kardan* 'to send an e-mail' (lit. e-mail doing).

The majority of the Persian verbs are used in complex forms which consist of an LV and a Non-verbal element (NV). NVs can be a noun, adjective, adverb or even a prepositional phrase. Persian LVs are generally known as semantically bleached verbs. The LVs *kardan* 'to do/make', *shodan* 'to become', *zadan* 'to hit', and *andāxtan* 'to drop/throw' are among the most frequent LVs.

With regard to the problems raised, about ten years ago, I started compiling a reference book entitled *A Comprehensive Corpus-based English-Persian Collocations Dictionary (CCEPCD)*. In this dictionary, LVs are the most important part. For example, in the entry **abeyance**, the expressions *ta'liq kardan* 'to put something in abeyance (lit. abeyance doing) and *ta'liq shodan* 'to fall/go into abeyance' (lit. abeyance becoming) have been of great importance, because Persian speakers use the word *abeyance* often in the above two expressions.

My main findings are as follows.

1. By ignoring the LVs in the target language, almost no dictionary can solve the text encoding problems.
2. Although the selection of entries is a function of valid English dictionaries, the expressions of that entry must be based on the foreign language.
3. Every active dictionary should be compiled according to the target language. Therefore, an English active dictionary will not be equally useful for Persian, Arabic, and Spanish speakers. In other words, an English passive dictionary can be useful for the speakers of all languages (for decoding an English text), but an active dictionary must be written for each language separately, which is rooted in the syntactic and semantic features of that language.

Patricia Anderson (Plenary) *Decentering the dictionary in practical lexicography with the case of the New Tunica Dictionary*

Today's Indigenous language communities, not content with making dictionaries that merely preserve a language for future generations, are actively creating dictionaries that facilitate ongoing language revitalization efforts. Whether they are designed for use in the classroom or at home, by native speakers or language learners, the dictionaries being made by and for Indigenous language communities are envisioned as multi-purpose tools with practical application to advance community revitalization efforts.

In the case of the New Tunica Dictionary, the Tunica community wanted a dictionary that bolstered several of their revitalization goals, such as codifying new language as legitimate, advancing Tunica grammar that encapsulates Tunica worldview over grammatical alternatives that embody a non-Indigenous worldview, and instilling a sense of pride among speakers and non-speakers alike.

In this session I argue that lexicographers who are invited into these contexts must decentralize the dictionary when engaged in the lexicographic process. Academic and non-community lexicographers focus far too much on the dictionary as the outcome, to the detriment of long-term revitalization efforts. Lexicographers and communities invest large amounts of time and effort into dictionary creation; and if such projects are undertaken with tunnel vision on the final outcome, you miss many opportunities to collaborate on lexicographic byproducts that can be extremely valuable in revitalization.

The easiest way to decenter the dictionary is to center something else. There are several candidates for a better focus during a dictionary project. In this talk, I will speak to the following three: (1) focusing on the goals of the larger revitalization project, (2) centering users both intended and unintended as well as users present and future, (3) spotlighting data structure, maintainability, and reusability. I will examine each of these through the lens of the New Tunica Dictionary project, highlighting areas we excelled in as well as areas we fell short of our intended outcome even as the project is ongoing.

Language revitalization is not a single-threaded enterprise. To be successful, language use needs to be fostered in many environments with many focuses. As lexicographers, we have an ethical obligation to widen our focus beyond dictionaries and dictionary making by centering a language community and its revitalization efforts, which the resulting dictionary can be designed to uplift and advance.

Grant Barrett *Defin-o-bots: Challenging AI to Create Usable Dictionary Content*

Although much of the current crop of artificial intelligence (AI) software is better termed "machine learning," its widespread adoption as a hoped-for precursor to true AI has in the last five years put sophisticated AI-like capabilities in the hands of the average software tinkerer,

and in some cases, in the hands of the average web user. The output of these software tools are very often indistinguishable from that of a writer or artist working alone.

In particular, as the text-based "AI" tools build on the natural-language processing that is at the core of machine learning, defining terms would seem to be a natural task for them.

Using these "AI" tools, such as ChatGPT (a language model developed by OpenAI) for text-generation and Stable Diffusion (a deep learning, text-to-image model developed by CompVis) for image-generation, I will demonstrate the level of sophistication of widely available off-the-shelf "AI" software and how it might be applied in lexicography.

One, I will show examples of its failures and success at generating definitions and example sentences for common, rare, and unusual lexical items.

Two, I will show examples of its failures and successes at generating illustrating images for common, rare, and unusual headwords.

Three, I will also quickly summarize basic intellectual property rights issues with the source material used in the machine learning that might stall uptake of these tools.

Lisa Berglund *Biography and Character Assassination: Hester Lynch Piozzi in the OED*

This paper was inspired by an article published in *Dictionaries* in 2011 by Roderick McConchie, entitled "Her Word had no Weight": Jane Austen as a Lexical Test Case for the OED." In that article, McConchie reviewed the presence (and absence) of quotations from Austen in OED illustrations of usage. He summarizes: "My own findings (still preliminary) have brought to light several specific problems with OED's representation of Austen: inaccurate or incomplete definitions of words that are found in her opus; missing senses, missing lexemes, and elusive meanings."

McConchie's article made me curious about how the writer Hester Lynch Piozzi, formerly Hester Lynch Thrale, is represented in OED. I initially expected that for this project I would review the words OED selects to be illustrated by Piozzi's authority, analyze the contexts of these quotations in her work, and explore how Piozzi's oeuvre and contributions to the English language are thus characterized in the dictionary. I wondered if she were credited as the original recorded source for any significations, and if so, which ones. Finally, I was interested in whether Piozzi's role as a lexicographer—she published the first original thesaurus in the English language, *British Synonymy*, in 1794—was in any way acknowledged in OED.

Today I will be sharing my discoveries and observations, which have diverged intriguingly from my expectations, as well as from the issues that McConchie regarded as central to the treatment of Austen. My principal conclusions can be summarized as follows: 1) yes, the representation of Piozzi has expanded since 1928; 2) yes, there is some thematic consistency among the words for which Piozzi is credited as the first source; 3) yes, OED does recognize,

briefly, her contributions to lexicography; and 4) Piozzi's (or Mrs. Thrale's) most remarkable appearances in OED have nothing to do with her use of words or her role as a lexicographer. In fact, as I will show, Mrs. Thrale is most conspicuous in OED as the subject of a fragmentary, sometimes tendentious, biography.

George Aaron Broadwell *Timucua, A language isolate of Florida: Constructing a dictionary from 17th century texts*

Timucua (ISO: tjm) is a language isolate formerly spoken in Northern Florida and adjacent parts of Georgia. Timucua speakers came into intensive contact with the Spanish colonial system in the 16th and 17th century, as Franciscans sought to evangelize them and convert them to Christianity. In the period from 1612-1635, seven religious books were published in Timucua, including bilingual Spanish-Timucua catechisms and confessionals. A brief *arte* and two secular letters also survive.

Because Timucua people were devastated by war and disease in the late 17th and early 18th century, the last known speakers of Timucua lived in the 18th century. However, despite the lack of speakers, the corpus of available material is extremely large.

In this paper, I describe the construction of a corpus of all available Timucua language material and the creation of a dictionary.

The corpus and dictionary development use Fieldworks Language Explorer (FLEX) as primary software. The corpus was transcribed by a team of graduate and undergraduate students over a period of four years, and currently contains ca. 148,000 words. The current Timucua dictionary has 4080 entries, and is based on careful interlinear analysis of the text corpus. An online version is available at webonary.org.

Because words vary widely in frequency, the dictionary adopts a 4 point confidence ranking of each entry, from Confident (for frequent and uncontroversial entries) through to Tentative (for infrequent words and hapax legomena). Each lexical entry provides reference to the portion of the corpus where examples of the entry can be found.

This methodology for studying an extinct isolate from a closed corpus of texts cannot answer all questions about diachronic connections, but helps to provide a firm basis for further work on the Timucua language.

Félix Cortés and Iara Mantenuto *San Sebastián del Monte Mixtec Dictionary Project*

Our talk offers a collaborative model and points of reflection derived from our experience in creating the San Sebastián del Monte Mixtec dictionary. We offer recommendations pertaining to the creation of a dictionary project that is community driven but with the collaboration of a university; we describe the roles of the different stakeholders, and we reflect on how we can make such projects sustainable.

San Sebastián del Monte Mixtec (ISO: mks), also known as Tò'on Nda'ví, is a language of the Mixtecan family, Otomanguan stock (Rensch 1976). San Sebastián del Monte is a town in the Santo Domingo Tonalá municipality of Oaxaca State, Mexico, in the district of Huajuapán de León, 45 km southwest of Huajuapán de León, with a population of approximately 2000 people (latitude: 17.677778, longitude: -98.021944). In 2019, we purposely decided to work on a Living Dictionary online, as the first step towards a printed dictionary, and as a way to have our work interactive and visible to everyone. We work in two teams that collaborate closely, one located in the U.S. and the other one in Mexico.

In this presentation, we explore how to make our collaborative writing process on the dictionary self-sustainable and long lasting. Thus, we want to contribute to the current conversations on online dictionaries for Indigenous languages (Garrett 2019, Harrison et al. 2019, Mosel 2011) by discussing our experience in collaborating on a community driven effort, which is part of a broader project and not the main project per se.

We offer examples on how to discuss and to think through the tasks of each collaborator. For example, in our experience the data entering portion is better led by students in the U.S. while brainstorming is better led by the community, not the outside linguists. We believe that the dictionary project better serves everyone if it does not stand alone, but is a project within a bigger revitalization plan that includes literacy workshops led by the community, pedagogical material creation, and cultural workshops. We propose that bigger projects that involve greater numbers of people interested in tasks related to linguistics revitalization have a greater probability of becoming self-sustainable, long lasting, and most importantly, community driven. A community dictionary project carried out in this manner also allows younger and older generations to work alongside each other, contributing to the revitalization project in a harmonious manner, through the transmission of information from older generations to younger ones. We are aware that this approach takes longer, but we are also aware that it will have far greater benefits in the long run. We will offer a detailed presentation of the model that we used, share some of the successes we achieved at different stages, comment upon how each project we have undertaken has fed into the dictionary creation, and reflect upon how we think the work will progress moving forward.

Dorian Cougias *A proposed architecture for a federated dictionary structure*

We propose an architecture for federating dictionary structures to provide for the exchange of information between heterogeneous dictionary systems, in different organizations, on different platforms. With the proliferation of dictionary APIs there is no reason users should be restricted to searching individual dictionaries when conducting term, definition, or thesaurus research. Likewise, when creating dictionary content, a methodology should exist to allow the dictionarist, lexicographer, or glossarist to research existing terminological entries or even corpora within any participating federated dictionary system. A proposal for federated

heterogeneous dictionary databases was put forward in 2000. It seems nothing came of it, probably because too much engineering and cooperation was needed to achieve the goal. The advances in APIs and corpora management allows us to more easily integrate multiple heterogeneous dictionary structures into a federated system without much need for each participant to change their native dictionary structure. Herein we not only propose the necessary data structure, but also the necessary licensing and attribution structure as well.

Andy Cowell (Plenary) *Lexical Resources for Indigenous Languages: Issues and Opportunities*

This panel could also have been titled “Beyond the Dictionary in Indigenous Language Lexicography.” With the shift to multimedia, on-line options for presenting lexical information, many new issues, challenges and opportunities have arisen for small indigenous languages. Challenges include things such as generating and maintaining large multimedia databases, in formats that will remain portable as technologies evolve. Opportunities include the rise of morphological parsers, which permits a user to enter a complex, agglutinated word into a lexical interface, and allows for the relevant core noun or verb stem to be automatically found and explained, when the user themselves might be unable to do this. Issues include data sovereignty in the on-line world. There are of course many other similar issues to discuss.

At the same time, many of the issues inherent in more traditional dictionaries remain crucial. Recent work in “revitalization lexicography” for example points to the weakness of traditional academic bilingual dictionaries of indigenous languages, which are heavily biased towards Indigenous > English directionality, with only schematic or index-like presentations of the English > Indigenous directionality. This is highly problematic in that most users in the future will likely be English speakers who need to get back to their heritage indigenous language. On-line databases face the same issues, and can even exacerbate them, as a simple query-based return of all forms in an indigenous language database whose definition contains reference to the English words ‘cut’ or ‘walk’ may return dozens or even hundreds of stems in some cases, completely overwhelming a beginning second-language learner.

Thus this panel will examine the emerging best practices in indigenous language lexicography, looking for ways in which past problems may be resolved by the new multi-media, on-line approaches, and ways in which we still face continuing challenges which should not be forgotten in the rush to often-attractive but sometimes over-sold new technologies.

A special focus of the panel will be how all of these issues further interact with the academic vs. community sides of the lexical resources issue. As the title of the panel attests, indigenous community needs are increasingly the main driver of Indigenous-language lexicography, yet the work continues to be done largely out of academia. Both sides face continuing challenges to adjust to the needs and expectations of each other (even as the dichotomy just evoked becomes increasingly blurred).

Andy Cowell and Julia Bonn (Workshop) *The Problem of Polysynthesis in UMR Annotations: Complexities in Handling Preverbal Modification and Noun Incorporation in Arapaho*

Arapaho, like many agglutinating polysynthetic languages, poses problems for lexicography, in that many "words" are partially made up of lexicalized elements, and are partially syntactic in nature. Arapaho in particular has many prefixes which cross-linguistically would be treated as auxiliary verbs, with meanings such as 'want to...' 'like to...' 'able to...' 'fail to...'. It also features verb-stem elements which can add vector elements to otherwise non-path verbs, such as nouut-ohwoo- 'dance to the outside of a location' and like classic polysynthetic languages, it includes extensive noun incorporation in verb stems. While these issues have been widely recognized by linguists and lexicographers, they pose a special new challenge in the context of attempts to design Uniform Meaning Representations, which seek to aid in cross-linguistic semantic and syntactic comparison and automated translation. In particular, the mismatch between verb stems and overall verb words can lead to loss of semantic content when UMR annotations are done, or conversely, attempts to maintain full content can lead to the creation of "lexicons" which contain thousands of additional words, beyond the stem level, whose features and meanings are highly predictable from the perspective of morpho-syntax, and which do not really need full semantic representations as single "words."

Anna Luisa Daigneault & Gregory D. S. Anderson *Living Dictionaries: A Platform for the Indigenous Languages of the Americas*

Our 20-minute tech demonstration is on how to create Living Dictionaries. They are interactive, mobile-friendly web tools that support endangered, under-represented and diasporic languages. Living Dictionaries are never out-of-print and infinitely expandable multimedia resources that combine language data (represented in diverse writing systems) with digital audio recordings of native speakers alongside photos and videos to expand cultural content well-beyond a static print dictionary.

Due to globalization, cultural assimilation, the long-term impacts of colonization, and official or de facto policies hostile to diversity, the vast majority of the indigenous languages of the Americas are threatened or endangered. There is an urgent need to address these issues with comprehensive, freely accessible tools that can assist in conservation efforts and revitalization programs while also creating safe spaces for these languages and their user communities. Living Dictionaries serve these needs by providing a free to use, cutting-edge solution to accessing language content with accurate pronunciations paired with engaging visual multimedia. Led by community activists and linguists around the globe, Living Dictionaries are collaborative projects that can help the languages of the Americas survive for generations to come. These digital dictionaries represent the future of these languages: vibrant and continually expanding.

Living Dictionaries promote connectivity over vast distances and support an online community of language learners who wish to hear (or see) and learn a language without close proximity to proficient speakers. They allow thousands of recorded words and phrases to be created easily and available at one's fingertips. High-quality audio and video recordings can be created through mobile phones and saved in the dictionary entries so that community members, new speakers and research scholars can listen to the pronunciation by a speaker. Living Dictionaries can serve multiple functions for different communities of stakeholders and provide meaningful opportunities for indigenous citizen science. User groups that could benefit from the development of such resources in the languages of the Americas include indigenous language activists, local and diasporic community stakeholders, cultural and linguistic advocates, educators and leaders, researchers, scholars and conservation scientists looking to understand ecological knowledge in local languages.

We will cover the following topics: how to register for an account on the platform, create a new Living Dictionary, add new entries, edit entries, add images, upload audio files, and how to record directly into the Living Dictionary using a smartphone or laptop. We have taught this software online to diverse global audiences during webinars many times. The web platform is accessible online, free of charge. The interface of the site is currently available for use in English, French, Portuguese and Spanish, among nine other dominant languages. We will show examples of multilingual Living Dictionaries for endangered languages spoken in North America, Mesoamerica, Central America and South America.

Boban Dedovic *Allo: A Modern Digital Dictionary Platform for Ancient Languages*
Ancient languages matter. They connect us to our ancestors and draw attention from many fields of study. Despite overwhelming interest, the practical study of ancient languages is usually limited to persons with access to training and resources. For introductory students, the dictionary is usually located in the appendix of the assigned grammar book. For advanced study, many ancient languages have one or more physically large, expensive, inaccessible, and difficult to use dictionaries. A systematic review of both print and digital dictionaries for eight ancient languages yielded three recurring challenges: accessibility, usability, and scope. To illustrate: the Oxford Latin Dictionary contains 2,400 pages with tiny font and weighs 9.1 pounds. The Brill Dictionary of Ancient Greek is similar; and, digital access to both requires institutional affiliation. The Chicago Assyrian Dictionary, although digitized, is comprised of 17 thick volumes. In languages like Ancient Egyptian, Akkadian, and Sumerian, yet another difficulty is added. Vital dictionaries for these languages require a separate publication of sign lists to understand the script. In many cases, these resources are hand-written in French or German. While the Internet has helped with respect to distribution of some of these materials, the challenges noted above—particularly usability—remain. This is not to say that these resources are inadequate or their creators inferior; rather, there exist hitherto unexplored

opportunities to leverage the Internet in order that trade-offs are mitigated. It is therefore necessary to reconsider the goals and priorities of digital dictionaries.

Enter Allo—a modern dictionary platform for ancient languages. Each dictionary contains 1,000 high-frequency words. It is designed to be accessible, user-friendly, and suitable for advanced users. To accessibility: we offer it online, charge nothing, and do not require registration. To usability: the interface is optimized for tablet and mobile devices via our design-oriented approach. Because we regarded speed and search as the most critical goals, features like autocomplete, caching, and a recommendation engine are fully integrated. These features may reduce the time it takes to find a given entry by orders of magnitude; based on a small user study, up to 50 times faster than a print book and 5 times faster than a PDF file. To content scope: all entries are aggregated such that they include both a beginner and advanced resource. The federated, or aggregated, model we chose allows us to make updates when new resources are released. Each entry contains between 30–50 individual data fields. Full page images of original source materials and cursor-zooming are built in. Advanced users can find grammatical details, tables of forms, usage frequency statistics, example sentences, and a comprehensive but compact grammar from authoritative sources. We demonstrate the benefits of this approach in our first complete language: Allo Latin, a digital Classical Latin–to–English dictionary kindly provided on <https://allo.conscious.ai/latin>.

Our experience suggests that digital dictionaries must be aggregations of information about a given entry and packaged in a clean user interface. Future innovation likely requires intense cooperation among three concerned parties: a technologist, lexicographer, and language expert.

Ewa Czaykowska-Higgins *The digital and print dictionaries of nxaʔamxčín: challenges and responsibilities*

For many Indigenous communities in North America and beyond, language maintenance, revitalization and reclamation play a central role in resisting and overcoming histories of oppression and colonization. Dictionaries can serve as one tool to support these goals and it is therefore a dictionary constructors' responsibility to take into account a language community's revitalization goals and needs. The purpose of this paper is 1) to present several challenges that have arisen in the construction of a digital and related print dictionary for nxaʔamxčín (Interior Salish; Washington State) based primarily on legacy recordings made by more than 20 fluent speakers working with a non-Indigenous linguist in the 1960s and 70s; and 2) to discuss how these challenges have been addressed in light of the responsibility to serve the language community.

The first challenge has involved deciding how to encode the language materials to meet the responsibility of ensuring the interoperability, portability, and long-term sustainability of a marked-up corpus from which dictionaries can be generated. As a Digital Humanities standard,

the Text-Encoding Initiative (<https://tei-c.org/>) is a tool that provides a large array of elements and attributes for language mark-up. The fact that TEI is a well-supported standard has been an important consideration for ensuring the digital longevity of the nxaʔamxčín corpus and digital dictionary. However, because it has been used in construction of very few dictionaries of Indigenous languages, using it has required customization challenges, some of which are discussed here.

A second set of primarily analytical challenges has involved decisions about 1) spelling; 2) lemmas; 3) encoding morphological relations; 4) appropriate glossing and definitions; and 5) creating an English-nxaʔamxčín index. For example, while the nxaʔamxčín Language Program has an established orthography, it has been essential to decide on how to spell words with the orthography, keeping in mind the responsibility of serving nxaʔamxčín learners: ultimately the spellings decided upon balance both pronunciation that reflects where schwa vowels appear in words and underlying morphological relationships.

Since language communities are increasingly using digital resources as tools for language revitalization, digital resources are valuable sources of information (Shepard 2016). However, as has been pointed out by many, digital resources are far more unstable than paper records. The third set of challenges facing the nxaʔamxčín dictionary project has thus been how to fulfill a responsibility to the language community to preserve the functionality of the digital application for decades of use. The nxaʔamxčín dictionary project follows the concept of Endings-compliance (<https://endings.uvic.ca/principles.html>): using tools and principles for digital longevity such as the creation of static sites, using open standards amenable to processing, validation, documentation, and no dependence on external services.

Constructing dictionaries for Indigenous languages that are being revitalized is a form of stewardship, requiring responsibility and accountability to the language community. As this paper illustrates, when responsibility to community guides the work, this affects all aspects of dictionary construction.

Stefan Dollinger *DCHP-3 and the new “Consortium” Dictionary of Canadian English: projects, problems, prospects in the longue durée of language study in Canada, 1946–2023*
“News from beyond the northern border.” This phrase mimics the default dominant perspective of our ALCS-chartered society, which is something I – usually – mildly commiserate. But not today. In the present paper, I’ll instead offer you a “once in a generation or two” update from Canada that is, hopefully, as important as Walter S. Avis’ one-page announcement in *American Speech*, “Important News from Canada”, that was the official note on the founding of the Canadian Linguistic Association for those working, “across the line”, in the States.

The goal of this talk is to link the events that led to the publishing of four fine dictionaries of Canadian English between 1962 and 1967 and the ensuing acceptance of “Standard Canadian English” as an entity. This Avis and Lovell period, begun in Charles Lovell’s private collection of Canadianisms as of 1946, will be linked with the failed attempts to update the DCHP-1 in the 1970s, the changes in linguistics in North America and English Studies, the developments around DCHP-2 (2006-2017). The real focus, however, will be the new work around DCHP-3, itself an academic “trick”, in the digital workspaces of today, and the long awaited, most welcome addition of a new project for a contemporary dictionary of Canadian English project, the “Consortium Dictionary” (precise name to be decided), as the present-day standard variety has been left undocumented since 2004.

The stars seem aligned, much like asked for in an obituary of Katherine Barber (Dollinger 2021), in which it is suggested that this time around stakeholders cannot and should not battle out lexicographical supremacy in a publishing framework that has led to the “Great Canadian Dictionary War” of the late 1990s, a war that knew – in the end – only losers, with all dictionaries indefinitely shut down. A consortium of Canadian stakeholders, the Editors Association of Canada, DCHP-3, Queen’s University Strathy Language Unit have gathered around chief editor designate John Chew III. to make a new dictionary a reality. On the problem side I will offer you the latest on what units do not seem to care much about our alignment of stars, for a common cause, not seen in Canadian lexicography since the mid-1960s. This talk is a compendium of sorts for Adams’ multi-part history of the DSNA (e.g. 2017) from “north of 49” and a critique of the neo-liberal publishing model that has driven our game since the 1920s; and to the ground, as I’d like to argue.

Daniele Trevelin Donato-Luz and Regiani Aparecida Santos Zacarias

Technical-Lexicographical Project for Selecting Examples for a Pedagogical Portuguese-English Dictionary of Verbs

Dictionaries are important for the foreign language learner, because they help, among other things, the acquisition of vocabulary and language use. This paper presents a technical-lexicographical project to select examples for a pedagogical Portuguese-English dictionary. It is part of a Bilingual Pedagogical Lexicography research, based on a theoretical and interdisciplinary study that comprises Bilingual Pedagogical Lexicography, Language Teaching, Contrastive Linguistics (Error Analysis) and Corpus Linguistics. This research is part of the project for the elaboration of the Portuguese-English Digital Pedagogical Dictionary of Verbs (PEDPDV) for Brazilian learners. The PEDPDV is supported by FAPESP (Foundation for Research Support of the State of São Paulo, Brazil) and by CNPq (National Council for Scientific and Technological Development, Brazil).

The PEDPDV is an innovative dictionary with emphasis on the needs of Brazilian Portuguese learners of English by taking Contrastive Linguistics and Corpus Linguistics as means of

investigation. The main dictionary project has been carefully designed and developed to meet learners' needs in language production. Accordingly, examples in the PEDPDV also deserved a well-designed project; therefore, examples have been carefully selected by taking into account the pedagogical nature of the dictionary. Current dictionaries are smart lexicography based, which is the cooperation between human lexicographers and machines. That considered, a technical project must be designed and developed so that all tasks that can be automated are delegated to the computer, while lexicographers focus on those points that need human expertise. In this context, the objective of this paper is to present the foundation research and part of the technical-lexicographical project that leads to the electronic process of selection and inclusion of examples in the PEDPDV entries, and some practical procedures that started and ended in the digital environment. First results have proven the project successful for pedagogical purposes.

Donna Farina, Marjeta Vrbinc, Alenka Vrbinc *Investigating the Dictionaries and Reference Books for 20th-Century Slovenian Immigrants in the U.S.*

As is well known, during the last quarter of the 19th century and the beginning of the 20th century an enormous immigration from Europe to the United States resulted in the establishment of multiple European ethnic groups in the (so-called) New World. The Slovenian immigration to the United States presents a compelling story. From 1905 until World War I, the greatest mass of Slovenians arrived, settling mostly in the midwestern and northeastern U.S., the largest concentration being in Cleveland, Ohio. In their new home, they worked in mining and smelting, in lumberyards and stockyards, and in manufacturing.

It is not possible to determine accurately how many Slovenians came to the U.S. during the high immigration periods, because they were often labeled as Austrian, Italian, Croatian, Slav, Slavic, Slavish, Slavonian, or Yugoslav—depending on when they arrived and the country they departed from. Due to anti-Slavic prejudice in the United States, Slovenians often self-labeled as Austrian; many of them were bilingual with German. They did this to avoid ethnic slurs such as "Grainers" or "Grenish" (a corruption of the term "Krainers," i.e., from the province of Krain, Kranjska, or Carniola).

This continuing project seeks to analyze the reference books and dictionaries of or for the Slovenian immigrant group. During Phase 1 of the project, the researchers identified, obtained, and began the analysis of numerous Slovenian–English, English–Slovenian dictionaries and reference works (grammar books, school textbooks, etc.) that were published mainly in the U.S. from the turn of the 19th century to the end of WWI. Phase 2 was a research trip to Cleveland; Joliet and Lemont, Illinois, during which the researchers examined newspaper archives, genealogical and other available materials that provide information about the dictionaries and reference books as well as about their authors, all prominent Slovenian–Americans.

These dictionaries and reference books were designed to help Slovenian immigrants learn English and acculturate to their new American environment; they also gave instructions on how to become a naturalized citizen of the U.S. Slovenian immigrants are said to have “acquired English with relative speed and facility”; this project seeks to understand how the dictionaries and reference books supported this effort.

This paper will provide an overview of initial work on this ongoing project. It presents a brief history of the Slovenian immigration and then focuses on a preliminary analysis of six of the dictionaries and reference books found. Emphasis will be on macrostructure and interesting features of these books. We will also briefly address microstructure which will be studied in greater detail during further research.

Joseph T. Farquharson *From possessive noun phrase to interjection: Mapping a lexicalization network in an English-related Creole*

The phenomenon addressed in this study is traditionally dealt with either in studies at the interface of syntax and semantics dealing with (in)alienable possession or investigations in lexicology dealing with terms of address. However, this paper employs an integrative approach to explore a construction- cum-word-formation-pattern that occurs in both the mainstream and slang registers of the Jamaican language (also known as Jamaican Creole/Patwa). The construction comprises either of the two available forms of the first person singular possessive determiner (*mi/mai*) followed by a noun, e.g. *mi brejrin* or *mai brejrin* 'my close (male) friend'.

As far as I am aware, the paper presents the first comprehensive study of this construction in any Creole language. It shows that it is fully productive in both slang and mainstream registers of Jamaican and argues that the *mi* variant is of older vintage than its *mai* counterpart. The paper provides a comparative description of the variant forms of the construction *mi N ~ mai N*, exploring the supra-segmental differences between the two and showing how the latter does not cover as much lexico-semantic and functional ground as the former. The focal points of the paper are the discussion of the range of functions that the construction performs (i.e., referring expression, term of address, pseudo-title, interjection, and swear word) and the argument that these functions represent different points in a single network of lexicalization. How far along this network a specific expression extends seems to depend on the semantic domain to which the N in the construction belongs (e.g. kinship, body part, affiliation). Therefore, an attempt is made to see to what extent the semantic domain of the N can be used to predict: (i) the likelihood of the expression advancing to the next node in the lexicalization network; and (ii) the differential productivity of the *mi* and *mai* variants.

Rachel Fletcher *EDIB terminology and meeting audience needs: (re)defining the terminology of diversity and inclusivity for language learners and publishing professionals*
The Cambridge Dictionary team has in recent years been implementing a series of targeted

editorial initiatives to review and improve the dictionary's representation of marginalized groups and its treatment of the language of race, gender, sexuality, disability, and similar topics that are the subject of rapidly developing public conversations about inclusivity and offensiveness. In this ongoing work, Cambridge lexicographers have faced many challenges familiar to those working on native-speaker dictionaries, as well as some challenges specific to its learner dictionary status, such as established ELT practices of avoiding potentially contentious content when writing for an international learner audience, and the necessity of doing justice to complex topics while keeping definitions accessible in vocabulary, style, and length to users who may not be fully confident in their English knowledge.

Many of these challenges were further highlighted when Cambridge Dictionary editors were asked to consult on wider diversity and inclusivity initiatives within Cambridge University Press & Assessment, the dictionary's publisher. These initiatives were aimed at improving awareness of best practice in EDIB (Equality, Diversity, Inclusion, and Belonging) for employees across the company, and involved the creation of documentation that in several cases included glossaries of relevant terminology. Although the dictionary editors were initially involved primarily to ensure that the groups working on these glossaries – who did not themselves have lexicographical experience – were using Cambridge Dictionary definitions correctly, the consultation process led to in-depth conversations and knowledge sharing that enriched both the EDIB documentation and the dictionary itself. However, it was not possible to reconcile all the needs, aims, and assumptions of these glossary projects with the established practices of the dictionary.

This paper presents some of the lessons learned by Cambridge Dictionary editors from this consultation process, focusing on how the differing perspectives of those involved, and the differing needs of their projects, led to both challenges and opportunities for improvement. It concludes by reflecting on what these challenges might mean not only for the lexicographical treatment of the terminology of diversity and inclusivity, but more generally for the future of learner dictionaries when their handling of certain topics is increasingly visible to – and scrutinized by – non-learner audiences.

Volker Harm *Between admiration and rejection: Grimm's Deutsches Wörterbuch and its difficult relation with the Oxford English Dictionary*

The great European dictionaries of the nineteenth and twentieth century such as Littré's *Dictionnaire de la langue française*, the *Deutsches Wörterbuch* of the brothers Grimm and the *Oxford English Dictionary*, however deeply embedded in their own national traditions, cannot be considered as isolated from each other; rather, they are part of a common European "knowledge space" in which they evolved in close connection to each other. As this network of mutual influences is still not fully understood, I will try to shed some light on the connections between two of the most prominent dictionary ventures of that time: the *Oxford English*

Dictionary (OED), founded in 1856/57, and the Deutsches Wörterbuch (DWB), which was inaugurated in 1838 by the brothers Grimm. It is well-known that the DWB provided inspiration for the OED in many respects. However, the other side of this relation has remained obscure so far, and the question as to which influence the OED had on the DWB throughout its history has not even been asked. In order to elucidate the OED's possible impact, I will try to reconstruct how the lexicographers working on and connected with DWB tried to take a position towards its younger and more successful English counterpart; thereby, I will rely on published and unpublished archival material as well as on the evidence of the dictionary itself. The focus of the investigation is on two crucial turning points in the history of DWB: First, the dictionary crisis around 1900, when the sharp contrast between the OED's rapid progress and the agony of DWB became more and more apparent, and, second, the period of the late 1950s, when, after the completion of the first edition, a revised version (DWB2) was prepared. In both periods, the question whether the OED could be a model was highly controversial: By some, it was considered as exemplary in terms of organization and effectiveness, but by other lexicographers, the OED was deemed superficial and diametrically opposed to the foundational principles laid by the Grimm brothers. When DWB2 was being designed in the 1950s, skepticism towards an OED-style dictionary still prevailed in the discussions. Against this background, it is surprising that, when the first installment was published in 1965, DWB2 looked similar to the OED in many respect, for example with regard to the entry structure, the dating of quotation evidence, and, most importantly, the reduction of the discursive style which has been one of the main hallmarks of DWB ever since its foundation. Thus, after more than a century of DWB history, the German Sonderweg in lexicography came to an end, and inspirations provided by the OED were eventually adopted without reservation.

Pavlina Kalm and William Croft (Workshop) *Enriching valency lexicons with constructional analyses: between verbal semantics and argument structure constructions*

One of the main objectives of valency lexicons is to identify argument structures evoked by predicates. The analysis used in these resources thus heavily rests on verbal semantics. However, it has been argued that lexically based representations of event structure lack an important layer of semantic information: a semantic analysis of argument structure constructions (or "constructions") in which verbs occur (Goldberg 1995, Kalm et al. 2019). Importantly, constructions carry meanings that are independent of particular verbs. In a standard dictionary, this type of semantic analysis is not immediately relevant to a description of lexical items as it is more pertinent to grammar; however, it is essential to the description of event structure, which is addressed in valency lexicons.

In many instances, the verbal semantics and the semantics of a construction in which it is used 'align', such as when a caused motion verb, e.g., *put*, occurs in a caused motion construction [SBJ V OBJ Loc-PP], as in *He put the book on the shelf*. The verb evokes an event in which an

agent causes a theme to change location. Similarly, the schematic description of the construction describes an externally initiated event of motion in which a theme moves to a different location. However, the caused motion construction is not restricted to motion verbs. Verbs from other verb classes that do not evoke motion events may occur in this construction. For example, a verb of contact by impact such as *hit* can describe a motion event if it occurs with a locative phrase: *He hit the baseball into the crowd*. The verb *hit* in this example evokes the same predicate argument structure as *hit* used in a forceful contact event, such as *He hit the dog*; however, the construals of the verb are different.

Existing valency lexicons infer the meaning of an utterance by pairing thematic roles evoked by a verb to syntactic arguments. Such analysis deals with the semantics of constructions only indirectly; constructions are not dealt with independently of verb meaning. As such, occurrences of verbs in syntactic contexts that describe different types of events than the verb itself pose a challenge. To deal with this issue, we propose that each verb entry in lexicons be supplemented by a semantic analysis of constructions in which the verb occurs. A representation in which verbal and constructional semantics are analyzed separately will enrich the description of event structure.

Our presentation touches on some of the practical issues having to do with generating mappings between constructional and verbal semantics, such as ways to identify which participant roles map to which grammatical roles in a construction and notation strategies for linking roles in these two types of representations. We also bring up challenges related to constructional analyses, such as defining which participants are essential to the description of constructional semantics.

Robert Krovetz *Soup to Nuts: A Program to Automatically Induce a Lexicon of Multiword Expressions and Re-Tokenize a Corpus*

Multiword expressions (MWEs) such as hot dog, Albert Einstein, make sure, and amino acid are an important topic in Computational Linguistics and Lexicography. They are often identified by statistical measures such as Pointwise Mutual Information (PMI) or Log-Likelihood (LL). This talk will discuss a program called soup-to-nuts that automatically induces a lexicon of MWEs and uses that lexicon to re-tokenize a corpus. The program is based on a lexical association metric called Mutual Rank Ratio (MRR) (Deane 2005). MRR differs from the other measures in that it is based not on deviation from chance occurrence, but deviation from a Zipfian distribution. Candidate bigrams and trigrams are ranked by MRR score, and then an algorithm is used to normalize candidates with regard to inflectional morphology and hyphenated form. The algorithm rejects ngrams that are chunks of longer expressions, such as osama bin and diego zoo, and it also rejects candidates based on linguistic patterns.

This talk will discuss how the program was developed, the reasons for different design decisions, and the impact of different parts of the algorithm. We found the best results by re-ranking the top candidates by dispersion (the number of different corpora that the candidate occurs in). This was used to create a common inventory of MWE candidates that is freely available to the community. The talk will conclude with initial results on creating MWE/MWE and MWE/unigram associations on the basis of re-tokenized corpora.

Anatoly Liberman *Idiom Dictionaries as a Genre*

In the post-Roman epoch, idioms have been collected in many countries, beginning at least with Erasmus. Modern phraseological dictionaries are numerous, and the literature on the structure of such dictionaries is rather rich. Now that my own explanatory and etymological dictionary of English idioms has been published by the University of Minnesota Press, I would like to throw a retrospective glance at this genre. In my talk, I intend to speak about the history of phraseological dictionaries, the selection of the material for them, the glosses accompanying individual idioms, and the numerous attempts to explain their origin. Such a state of the art report of the genre has, as far as I know, never been offered in the special literature, though the structure and the history of idioms have always been at the center of lexicologists' and lexicographers' attention.

Nora Livesay (Plenary) *The Ojibwe People's Dictionary*

The Ojibwe People's Dictionary (<https://ojibwe.lib.umn.edu>) was envisioned as a tool in Indigenous language reclamation, providing an major expansion of Nichols & Nyholm's 1995 Southwestern Ojibwe dictionary, a gateway into the Minnesota Historical Society's Ojibwe archives, and a hub for sharing pedagogical materials. The data structure of the Dictionary was modeled on the structure of Anishinaabemowin to assist in vocabulary building through the features of Word Parts, Word Families, Related Words, and thousands of native-speaker audio clips of inflectional forms of words and example sentences.

In the decade since the Ojibwe People's Dictionary launched in 2012 as an online talking dictionary, it's become one of the most-used online Anishinaabemowin dictionaries with more than 50,000 monthly visitors and widespread usage among language-learners and teachers throughout the US and Canada. Now, the Dictionary needs to improve website usability, and address underlying database and technology platform functionality to meet the needs of the language community. In particular, it needs to enable community-engaged editing of lexical entries. The community has also asked for a topical directory based on semantic domains, the ability to search for example sentence audio, better English and Ojibwe search with filtering capabilities, and a spell-checker app for word-processing, among other things.

The barriers encountered are a continuation of settler-colonial policies and priorities within a land-grant university (<https://landgrabu.org>) and a state with a history of erasure of

Indigenous cultures and languages. Despite recent high-profile gestures toward addressing tribal demands, the University of Minnesota continually fails to provide institutional support for the sustainability of Indigenous and American Indian Studies and its language programming initiatives.

Ligeia Lugli *Democratizing digital lexicography: towards an accessible tool for building fully-customizable online dictionaries*

This paper introduces a new project aimed at simplifying the creation and dissemination of digital dictionaries. With almost half of world languages at risk of extinction according to UNESCO (<https://www.iesalc.unesco.org/en/2022/02/21/a-decade-to-prevent-the-disappearance-of-3000-languages/>), and with an ever-expanding heritage of cultures and literatures waiting to be documented, facilitating electronic lexicography for low-resource languages is becoming increasingly important. Especially so, given that the lexicographic documentation of low-resource languages typically require more manual work than their well-resourced counterparts. This is because the language-models, lexical datasets and sheer amount of language data available for low-resource languages is insufficient for automating many of the tasks that can now be at least partly automated for majority languages. Moreover, the communities speaking those languages are often unfamiliar with the computational methods available for lexicography, and even when they have, through painstaking manual curation, prepared dictionaries, they are unable to package them in digital format and publish them online due to lack of coding skills and IT support.

This paper presents how our newly-started project intends to reduce the technical barrier that many lexicographers working on low-resource languages in 'low-tech' environments encounter. It opens with an overview of the needs of our target audience, as ascertained through preliminary focus groups with lexicographers engaging with a variety of languages worldwide (but mostly working in Latin America). The paper then explains how we intend to meet those needs by developing a Shiny application (Chang et al 2021) that allows users with no coding skills to build a digital dictionary-editing and dictionary-publishing pipeline. Special attention will be given to highlighting how our tool differs from the popular Lexonomy, produced within the eLexis project. The paper then moves on to a brief demo of a preliminary prototype of this application that we have prepared, and concludes by outlining the next steps that our project will take to better understand and meet the needs of our target audience.

John Lyon and Kathleen Michel *The Upper Nicola Okanagan Talking Dictionary Project*

This presentation describes the origins, history, and current state of the “Upper Nicola Okanagan Talking Dictionary Project”, a community-driven research project whose aim is to convert digitized archival recordings of word lists recited in the Upper Nicola dialect of Nsyilxcn (a.k.a. Okanagan, a Southern Interior Salish language spoken in British Columbia and

Washington State) into a digital format usable by the Upper Nicola indigenous community. The resulting tool will assist ongoing language revitalization efforts in the community.

The recorded content of the dictionary stems from recording sessions carried out by Dr. Yvonne Hebert between 1978 and 1980 with fluent elder Joe Albert Michel, now passed on, as part of her dissertation fieldwork (Hebert 1982). The recordings were archived at the Royal Museum of British Columbia in Victoria, and digitized by a team led by Sharon Lindley in 2010.

Michel recites lists of Okanagan words, phrases and sentences, typically pronouncing each word twice, followed by a colloquial English translation. There are currently 6821 sound files of individual word/phrase/sentence recitations and translations in the collection, with some duplicates. Many of the words and phrases in the collection are unique to the Upper Nicola dialect of Nsyilxcn, and are not found in existing dictionaries such as Mattina (1987). As of today, raw recordings have been divided into sound files, typically between 5-10 seconds in length each. Each sound file includes an individual word, phrase, or sentence recitation and Michel's accompanying translation. They have been saved as WAV files which are named according to English translation, or "keyword", resulting in a file structure of 26 folders corresponding to letters of the English alphabet.

Our team's current work involves listening to and transcribing the Nsyilxcn recordings into a spreadsheet. Our team crucially involves indigenous, Nsyilxcn language learners, who are currently enrolled in UBC Okanagan's Bachelor of Nsyilxcn Language Fluency (BNLF) degree. Our work proceeds with the permission of Oliver Lindley, Joe Albert's grand-nephew and Sharon Lindley's son, and the current owner of the recorded materials. We also gratefully acknowledge the support of grants from UBC and the Delmas Foundation.

In our presentation we share sample recordings and transcriptions, and highlight unique dialectal features of the dictionary. We also discuss formats that we envision the collection will eventually take, including possible root-based dictionaries with an eye to etymology (e.g. Mattina 1987), stem-based learner dictionaries with reversed concordance, and other more relational frameworks.

Through this work, Upper Nicola students gain experience with Salish language sounds, root identification, morphological structure, and pronominal paradigms. Students also contribute directly to their own language's revitalization through accessing audio recordings of their linguistic heritage, thus restoring a minimal semblance of the natural language transmission link that was broken by the Indian Residential School system.

Sterling Martin (Plenary) *Project ENABLE: an online resource for Enriching Navajo as a Biology Language for Education.*

To address the global decline of Indigenous languages, The United Nations has declared 2022-2032 "The International Decade of Indigenous Languages". One Indigenous language

undergoing rapid decline is *Diné bizaad* (Navajo language). Project ENABLE (Enriching Navajo as a Biology Language for Education) aims to tackle this issue head on. Over the last 40 years, the number of *Diné bizaad* speakers has dropped by 42 percent, and current estimates predict by 2030 only 10 percent of *Diné* (Navajo) will speak their language. The COVID-19 pandemic has made the situation even more dire. The virus has ravaged *Diné'tah* (Navajo Nation), causing the loss of many *Diné bizaad* speaking community members. One issue contributing to the difficulty controlling COVID-19 in *Diné'tah* is that there are almost no words for modern biology terms, despite *Diné bizaad* being a rich and expressive language. In this time of the COVID-19 global pandemic, accelerating climate change, and the prevalence of advancing technology, it is more urgent than ever to formally incorporate these biological terms into *Diné bizaad* so all speakers can understand the world around them.

For this session, I plan to discuss our experience creating modern *Diné* science neologisms and considerations for our website to disseminate these words to the *Diné* community. In partnership with high school teachers on *Diné'tah*, we identified 250 foundational biology terms that reflect foundational biology concepts to be translated into *Diné bizaad*. We then worked with a well-regarded *Diné* language expert to translate these English science words (including scientific definitions and example sentences) into *Diné bizaad* with a culturally respectful, *Diné* perspective. With the expert's help, we generated scientifically accurate words that would be understandable by community members that primarily speak *Diné bizaad*. These words and audio pronunciations (important so users can hear the pronunciations in both English and *Diné bizaad*) are curated on our website: <https://enablenavajo.org/dine/>. This website is designed to work on both cell phones and computers in areas with low internet connectivity, which is crucial as much of *Diné'tah* has scarce internet access, along with poor cell reception. Project ENABLE will allow teachers to supplement their biology lessons with terms in *Diné bizaad*, allowing the next generation of *Diné* students to discuss science with their families and communities in their native language.

Rod McConchie *Free dictionaries for a greater purpose: The use of promotional dictionaries*
The nineteenth century saw dictionaries become paragons of lexical rectitude and authority, an authority which was exploited by the US commercial world to sell a multitude of other products and services, often as giveaways and frequently with no relevance at all to the goods offered. These range in date from the eighteen-seventies to the mid-twentieth century and these promotional dictionaries can still be found and customized for a particular product.

The earliest so far identified is for the Burlington Railroad (1870). They began to increase in popularity, and in the 1890s were taken up by the appropriately named Webster Shoe Co., which gave them away with a purchase of school shoes. Children's shoes were the most popular, being given away by a large range of shoe-stores. By the nineteen-thirties, banks,

insurance companies, manufacturers of earth-moving equipment, carbon paper, waxes and solvents had all jumped on the lexicographical bandwagon, to mention only some.

This may seem like cynical exploitation of the dictionary simply for gain, but examples can be found thoroughly ink-stained and dog-eared, suggesting that children used and appreciated them. This paper will survey the dictionaries, their retain locations and their geographical distribution where possible.

Olga Menagarishvili *What Do Cultural Studies and Scientific and Technical Communication Have to Do with Dictionaries?*

As we all know, dictionaries are usually studied within an academic field called lexicography. In this presentation, I will propose an additional framework/lens that could be used to discuss dictionaries. More specifically, I will focus on why and how cultural studies (an interdisciplinary field that focuses on the questions of power and culture and, therefore, allows one to discuss “knowledge legitimation within cultural contexts”) and scientific and technical communication (an academic discipline that studies how people communicate about science and technology) could be used to explore dictionaries.

As cultural objects, dictionaries perform several types of cultural work. For example, they have the ability to influence knowledge legitimation: they exercise power as they assign value to knowledge by dictating what knowledge is valuable and trustworthy (the knowledge found in dictionaries) and what is not (the knowledge omitted by dictionaries). Dictionaries also create what Anderson calls “imagined communities” that can be defined as communities consisting of people who “will never know most of their fellow-members, meet them, or even hear of them, yet in the minds of each lives the image of their communion”. Such communities are present in our lives even though most of the time we are unaware of them, and this invisibility, perhaps, makes imagined communities even more powerful. Finally, each dictionary is created within a certain cultural context that affects that dictionary’s content, design, and usage.

As artifacts of scientific and technical communication, dictionaries are a genre with many subgenres that can be used to communicate about science and technology. Each dictionary has an audience, a certain purpose, and a structure and is used in a certain way, which are all aspects of documents discussed by scientific and technical communication. Moreover, definitions are a typical scientific and technical communication genre that is taught in every technical communication class. Further, many dictionaries combine text and visuals, which is another technical communication topic taught and studied by scientific and technical communication instructors/researchers.

In this presentation, I will use specific examples to discuss dictionaries from two angles: the cultural studies perspective and the scientific and technical communication perspective. This

presentation might be useful for lexicographers, instructors using dictionaries in their classes, and the general public interested in dictionaries.

Linda Mitchell *Adapting Lexicons to Middle Class Eighteenth-Century England*

Lexicons play an important role in eighteenth-century England as lexicographers shape them to meet the needs of the rising middle class. A key element is making these lexicons relevant and useful to a wider, working-class audience. In this paper, I look at how dictionaries become a means for the rising middle classes to move into a century of progress and literacy. Language was power, and people who were literate moved ahead in a changing world that brought new demands socially and economically. Dictionaries were key in preparing the rising middle classes to meet increased opportunities and aspirations. Literacy was necessary for shopkeepers and tradesmen to be successful in the British Empire. Dictionaries also functioned as a working-class person's finishing school, especially for topics that could make a person appear educated in social situations. For example, Nathan Bailey's drawings illustrated words in *Dictionarium Britannicum* (1730) that prompted interesting topics for conversation. On the title page he claims to include 500 drawings "for Giving a clearer Idea of those Figures, not so well apprehended by verbal Description."

Lexicons were particularly significant because they were a venue for codifying and standardizing the vernacular. They recorded inconsistencies in dialect, spelling, grammar, and pronunciation. As dictionaries transitioned from the late seventeenth century into the eighteenth century, definitions expanded with linguistic information that reflected the changes in the vernacular. Prefaces and introductions included lists of what a good dictionary should include, while ambitious dictionary authors attempted to outdo each other with expanding entries and new vocabulary words. Lexicons became useful tools for the middle classes as dictionary authors added parts of speech and pronunciation, grammar rules, letter-writing guides, terms of proper address, and content such as mythology or history. For example, in *A Guide to the English Tongue* (1709) and *A New English Dictionary* (1723) Thomas Dyche made several improvements in spelling and pronunciation guides. Words were often defined through etymologies or illustrated in sentences from literature or the Bible. Some lexicographers shamelessly borrow large amounts of text from other authors to expand their dictionaries.

The eighteenth century was an exciting time for lexicons; they assumed a dominant role over grammar texts because of the wide variety of information a dictionary included in entries. Some of the dictionaries were forerunners of encyclopedias. Dictionary authors who early on adapted their lexicons to the educational needs of the rising middle classes were more apt to find their dictionaries selling well at the bookshop. Households might have only two books: a bible and a dictionary. And today, many people have some kind of bible, and everybody has a dictionary, whether it is a physical book or an e-copy.

Brad Montgomery-Anderson *The Russian-English Dostoevsky Dictionary: The bilingual literary dictionary as corpus and learning tool*

This presentation describes the methodology and pedagogical motivation involved in the creation of a bilingual dictionary based on the work of a specific author. This in-progress work represents a new type of specialized dictionary: a bilingual Russian-English dictionary that is a comprehensive collection of all vocabulary in the fictional works of Russian author Fyodor Dostoevsky (1821-1881). I will first outline four special organizational and presentational features that distinguish such a work from other bilingual dictionaries. Second, I will discuss the place of this dictionary in the realm of specialized dictionaries and compare it with the dictionaries that most approach this type of project. Third, I will describe the pedagogical uses and motivations for the creation of such a dictionary by placing it in the larger context of a foreign-language teaching methodology in which the target language is taught through a literary work. Teaching the language through an important literary text immediately introduces students to authentic language and culture. Following Berardo (2006), authenticity promotes meaning-based language learning that engages students in a context-rich environment. This method is based on the belief that using culturally significant literary texts as the main vehicle for classroom language can ‘hook’ students and promote their own self-directed language learning beyond and outside the classroom. I base this belief in part on McRae’s (1994) observation that literary texts are representational and therefore appeal to the imagination and emotions. Following Kramsch (1998), this approach incorporates culture into language learning and allows students to learn Russian by interacting with an authentic and culturally significant literary work and grappling with its meaning.

Lynne Murphy, Daniele Franceschi, et al (Panel): *Perspectives on Synonym Relations*

The panel will consider a broad range of interactions between synonymy as a lexicological concept, lexicographical practice and dictionary/reference usability, with highly practical insights into the use and representation of synonym relations in lexicography.

- Introduction (MLM)
- Marc Alexander (University of Glasgow): Synonymy as an ordering principle: lessons from editing historical thesauruses
- Daniele Franceschi (Università Roma Tre): How do monolingual English learner’s dictionaries represent near-synonym relations? The case of Anglo-Saxon and Latinate ‘equivalents’
- Rachel Stone (Druide Informatique): Applications of a Synonyms Dictionary in a Bilingual Dictionary and Text Corrector
- Fraser Dallachy (University of Glasgow): Computational synonym disambiguation and semantic tagging using lexicographical resources

- M. Lynne Murphy (Sussex): Reaction paper: how synonymy shapes lexicography (and vice versa)
- General discussion (Q&A)

Skatje Myers, Martha Palmer and Ishan Jindal (Workshop) *Comparing English and Russian PropBank annotation*

Although semantic role labeling is useful in many downstream NLP tasks, these semantic representations are not available in many languages. We present our work on leveraging parallel sentences where the source language has automatic or manual SRL annotation to transfer annotations into the target language. The Universal PropBanks (UPB) 2.0 system uses unsupervised word alignments, filtering heuristics, and bootstrapping to automatically project English SRL annotations to 23 languages. Since this approach uses English-based semantic representations to create annotations in other languages, we provide a case study between English PropBank representations and the recently developed Russian PropBank. We evaluate the UPB 2.0 projecting English annotations onto sentences that have been annotated by Russian PropBank, developing language-agnostic and language-specific filters to improve the results.

Nguyễn Văn Bền *English version of a Media Dictionary*

Our Media Dictionary uses media materials instead of definitions to explain and to illustrate words. Media files such as .jpg, .png, .gif, .mp3, or .video such as .wav, mp4 are used. Common words like *tiger* and *duck* are illustrated by .jpg or .png visual files; *heartbeat* or *two-stroke* are illustrated by animated .gif files; whereas words with sounds such as *squeal*, *anthem* and *canticle* by .mp3 audio files. And other kinds of words by .mp4. Media is drawn from Wikimedia as well as copyright and author-permitted sources. Words are chosen from Webster's or Oxford's and the like.

Learners look up or listen to illustrations of English words in the online dictionary <http://Mediadic.com>. Words and vocabulary that appear in the media dictionary are those found in ordinary dictionaries. Now the dictionary makes use of 55000 media files and amounts to more than 200000 pairs of word and media.

The dictionary is useful. It is easy to learn even for children. They learn far more than words, they can learn lessons as well. It is said that a picture is worth a thousand words, that means, we can learn not only from main illustration but also from surrounding context. Learning time will be happier in class and out of class. It proves helpful to teachers in planning lessons. Media Dictionary. is also beneficial as a reference.

We are compiling the English version and find that we can make many more versions in other languages. It opens to new ways of learning and teaching, that means, learning and teaching with more media and less words, less sentences.

Media Dictionary has some advantages. It is a complement to common dictionaries. Sometimes, however rarely, it is confusing to guess the meaning from looking and hearing the relevant media. More work will be implemented to improve shortcomings. It offers also an opportunity to combine itself with a conventional dictionary. Besides, Media Dictionary includes a good database for other educational projects. Database involves many fields, and Media Dictionary makes abstruse ones more comprehensible.

Alexis Palmer and Matt Buchholz (Workshop) *Work with what we have: Bootstrapping from lexical resources for low-resource languages to AMR/UMR annotation*

Uniform Meaning Representation (UMR) is a framework for annotating the semantic content of a text. While UMR is designed to be cross-linguistically applicable, its annotation depends on valency lexicons and other materials which may not exist for low-resource languages. Using Arapaho as a case study, we demonstrate how support for UMR annotation can be developed automatically using existing data for low-resource languages, including lexicons and interlinear glosses.

Martha Palmer et al Workshop: *The role of lexicons and lexicography in Natural Language Processing*

Many Natural Language Processing rely on producing predicate argument structures as canonical representations of sentences that abstract away from syntactic variation.

Computational lexicons in the form of valency lexicons that can support the formation of such predicate argument structures therefore play an enduring role in Natural Language Processing (NLP) systems, as key enablers of such abstractions. The PropBank frame files (or their FrameNet and VerbNet cousins) provide the foundation for Semantic Role Labeling systems as well as English Abstract Meaning Representation (AMR) parsers. Pre-existing language-specific Frame Files, or other types of valency lexicons, simplify the task of spinning up AMR annotation projects for new languages such as Chinese and Arabic.

However, there are many unresolved fundamental questions having to do with the creation of such lexicons, and how to draw the boundaries between distinct senses and distinct word forms. This workshop will review existing computational valency lexicons that are widely used in NLP, the contributions they make and the lexicographic techniques used in their development. We want to address the complications that arise with polysynthetic languages and their incorporation of certain semantic concepts such as aspect or enablement as prefixes rather than as distinct lexical items such as verbs, as is preferred by analytic languages.

We are also interested in exploring the boundaries of compounds and multi-word expressions, and criteria for determining when they should have distinct entries. Issues that arise when mapping between valency lexicons in multiple languages, such as differences in argument structure, are also especially relevant to our investigation.

Bill Poser *Special Features for Place Names in Electronic Dictionaries*

For endangered indigenous languages, place names are of greater importance in dictionaries for most other languages due to their cultural importance and their role in demonstrating the occupation and use of the territory. We describe some special features for place names that have been implemented in several electronic bilingual dictionaries of severely endangered indigenous languages of Canada. One of these is the inclusion of maps in place name entries. For place names for which a precise location is known, not only are the latitude and longitude given but a small map icon is displayed. If an internet connection is available, clicking on this icon opens an interactive Google map centered on the location of the place name. This clarifies the location better than a purely verbal description.

The second is the provision of search by location. Place names, like other words, are of course listed in the alphabetical indices and findable via regular expression search for the name, but it is also possible to request a list of the place names located within a given distance of a point specified by latitude and longitude. This enables the user to discover the names of places unknown to him or her. This is helpful since it is not infrequently the case that indigenous people have names for places that have no English name since they are notable only to indigenous people.

Place name entries are also accessible via a separate interactive map. Places may be found by selecting the name from an alphabetical index in either language, which re-centers the map on that place and makes the marker bounce up and down. Different types of places (mountains, lakes, towns and villages, etc.) have markers of different colors, and the display of places of a particular type may be toggled on and off. It is, for example, to show only habitations or only lakes and rivers. Clicking on a marker pops up a balloon containing the name in both languages. Clicking on the name opens the entry for that place name in the appropriate dictionary (since the map may contain names from several languages or dialects) in another browser tab.

These features, though not particularly difficult to implement, appear to be novel and although of particular importance for indigenous languages, may be useful for other languages and in monolingual dictionaries as well.

Kevin Rottet *Paronymic Attraction and Folk Etymology in Louisiana*

One of the claims about folk etymology (FE) made—and contested—in the literature is that it is especially common among “the folk”, the masses with few or no literacy skills. Yet other questions remain in the study of FE, including what to call it, reflecting different perspectives on what exactly it is: popular etymology, paronymic attraction, remotivation, secondary motivation, analogical reformation? Whatever the label and emphases, FE is generally understood to be “motivated by a need to lend ‘semantic motivation’ to unfamiliar (and usually

long) words by replacing portions of them with elements that help ‘make sense’ of them”. Attempts to theorize FE have proposed typologies distinguishing cases of semantic motivation, phonological motivation, or both. A more elaborate model (Knappe 2004) proposes an expansion incorporating two other features: whether FE results in a change of meaning only or also a change of form, and whether it does or does not result in a change of denotational meaning; combined, these features give a total of six attested subtypes.

The historical lexicography of French and Creole in Louisiana reveals significant FE in many word-forms. Candidate lexemes include obscure learned words, as when *fluxion de poitrine*, ‘pneumonia’, becomes *frisson de poitrine* in Louisiana Creole, replacing *fluxion* with the more mundane *frisson* ‘shiver, chill’. Or they may be foreign loanwords, as when *Christkind*, a reference to the ‘Christchild’ brought to Louisiana by German immigrants, became *la Christine* among Francophones, with changes of interpretation of who (s)he was—thus a change of denotational meaning, in Knappe’s terms. FE also occurs frequently in the names of flora and fauna, a lexical field filled with transparently descriptive compounds, e.g. *goldenrod*, *cedar waxwing*, *yellow-bellied sapsucker*; unanalyzable polysyllabic names may thus invite (re)motivation, even if it is solely phonological, as when the Choctaw name for the mudfish (*shupik*) entered Louisiana French unanalyzed (*choupique*), but then made its way into regional English where FE turned it into the semantically nonsensical *shoe peg*.

Like the case of *shupik*, our particular focus will be on instances of FE resulting from loanwords. Indeed, Louisiana as a crossroads for contact between speakers of American Indian languages (such as Choctaw) and European languages (French, English, German, Spanish) gave rise to interesting instances of FE. In some cases, stories developed to “explain” a form by invoking the limited bilingualism of earlier times, as when Whatley (1983) cites anecdotes to semantically motivate *say-so*, a Louisiana French name for the ice cream cone which, in fact, started out as an American brand name from Toledo, Ohio.

Some data points to the usefulness of adding to Knappe’s phonological motivation and semantic motivation an additional category of morphological motivation that would apply to a case such as French *verrue* ‘wart’ which becomes *verrure* by analogy with several other names of maladies that end in *-ure* including *cassure* ‘break’ and *brûlure* ‘burn’.

The study shows that Louisiana French and Creole lexicography offers a fertile testing ground for Knappe’s model of FE, confirming the relevance of her subtypes.

Lindsay Rose Russell *The Life and Lexicography of Frances H. Willard*

This paper traces the life and lexicography of an indigenous Alaskan woman named Shik-Sha-Ni, later Frances H. Willard. Willard co-compiled some of the earliest dictionaries of Tlingit and English, publishing, for example, the 1904 *Grammar and Vocabulary of the Tlingit Language of Southeastern Alaska* with William A. Kelly. Shik-Sha-Ni was born in Alaska around 1870 and,

at a young age, taken from her home into a Christian missionary “ranch” at Fort Wrangel. There, she was educated and converted to Presbyterianism. Esteemed an extraordinarily “promising young scholar,” Shik-Sha-Ni was “rewarded” by receiving the first name of her headmistress, Fanny McFarland, and the surname of a mission sponsor from Auburn, New York, one “Miss Willard.” Hence, Shik-Sha-Ni was Frances H. Willard by the time she was taken to the mainland United States in 1885 where she was toured around the eastern seaboard as evidence that “barbarous heathens” could be saved. In spite of her tokenization, Willard took advantage of the opportunities available to her in the mainland: She finished her schooling in Baltimore, obtained a teaching post in New Jersey, picked up a nursing degree in California. Willard returned to Alaska in 1899 to become a teacher at the Sitka Training School. There, she embarked on a number of language revitalization projects, including the documentation of the Tlingit language. She worked with her cousin Tillie Paul on the first dictionary of Tlingit before collaborating with Kelly on the first Tlingit- English dictionary. The remainder of the talk will survey the value of this lexicographical work in what we might now term decolonial lexicography.

Ana Salgado, Rute Costa et al *Domain labelling in the Morais dictionary: bringing structure to unstructured lexicographic data*

This article provides a detailed analysis on the use of domain labels, i.e., special markers identifying a specialised field of knowledge, in successive editions of the Morais dictionary. Morais is a historical Portuguese language dictionary, commonly known by and disseminated under the name of António de Morais Silva. This monolingual dictionary has relevance for the Portuguese lexicographic tradition as it inaugurates modern Portuguese lexicography and serves as a model for all subsequent lexicographic production throughout the 19th and 20th centuries. The domain labels were retrieved from the abbreviation lists of its various editions. This work is part of an ongoing Portuguese national linguistic project. It has two goals: 1) to encode the first three editions of the Morais dictionary to make them available online (as well as publishing them as lexical resources using two different standards for structured lexicographic datasets) and 2) to provide a description of the lexicographic components of these editions following a rigorous linguistic treatment. This project is not merely of a lexicographic nature, but it also explores the convergence between lexicography and other research domains, such as terminology, ontologies, linked data, and digital humanities.

This article analyzes the domain labeling system in Morais from an evolutionary and diachronic perspective, in line with previous works that highlight the theoretical assumptions and methodological aspects of the lexicographical tradition around domain labeling. To organize lexicographic content, it is helpful to establish a hierarchical structure in general language dictionaries to systematize the included terminological information.

Each table of abbreviations has two distinct columns: one with the abbreviation and the other with the complete domain designations. Given the importance of domain labels, we conducted a survey of all domain labels found. We identify and demonstrate the previous and newly added domains. After reviewing the flat domain list, we evaluated whether there was a discernible knowledge organizational approach that identified possible generic domains and subdomains. In the organization of domains, we propose three possible levels: superdomain, domain, and subdomain. The superdomain corresponds to the broadest taxonomic grouping followed by a domain, whereas the subdomain is part of a broader domain. To facilitate the analysis and to focus on interoperability issues, we generated a metalabel, a tag that identifies the English equivalent of the corresponding domain.

The lists of domains included in general dictionaries' outside matter follow alphabetical ordering, without any concern for the relationships that can be established between those types of labels. This article describes both onomasiological and semasiological approaches to treating specialized lexicographic content. Following terminological principles and an onomasiological approach, we organize and conceptualize specialized knowledge using structured data formats, such as Text Encoding Initiative, also considering future alignments between different lexicographic resources.

The project will contribute towards a more significant presence of lexicographic digital content in Portuguese through open tools and standards.

Jesse Sheidlower *The Historical Dictionary of Science Fiction: An Introduction*

The Historical Dictionary of Science Fiction is a new, online-only dictionary on historical principles dedicated to the vocabulary of English-language science fiction, and related areas including science-fiction fandom and the critical study of the broader field of imaginative fiction (which includes fantasy, horror, and other subgenres). It is an outgrowth of an earlier Oxford English Dictionary project to crowdsource quotation research for specialty fields. As a genre, science fiction has very broad popular appeal and receives an increasing amount of scholarly attention. Its vocabulary is highly restricted, in that it is usually very clear whether a term is or is not a science-fictional one. A substantial proportion of the foundational texts have been digitized, and can thus be both searched (for research) and linked (for presentation). Science fiction is thus an ideal subject for historical lexicographic treatment.

This presentation will explore the HDSF in a variety of ways: its history, its editorial principles, its technical standards, its utility for users both academic and otherwise. Some issues are relevant only for this specific dictionary: defining the boundaries of science fiction; converting a research-gathering effort into a dictionary proper; transforming OED-specific markup language into something more appropriate for the HDSF. Some issues are similar to those of any dictionary: design and presentation of linguistic and bibliographic information; defining the

project's scope; establishing an inclusion policy for quotations; writing and presenting the front matter and other ways of explaining the project to its audience; handling ongoing updates. Other issues are more familiar to digital-humanities projects: deciding on various aspects of the technical infrastructure (programming languages, database and search backends, site architecture, etc.); linking to original, full-text sources; linking to other reference projects and to standard bibliographic resources; devising standards for sharing the data with other DH projects. Still other issues are even more broad: how to get publicity for a new scholarly endeavor; how to manage outside contributors; securing funding.

The decisions that had to be made are ones that have potential relevance to any similar project. Within the limited time available, the presentation, then, will explore not only the specifics of one particular dictionary, but also the factors relevant to developing a medium-size online historical dictionary for any specialty area of interest.

Jason Siegel *Lexicography in education: Dictionary skills and knowledge for older students*
Dictionary skills have long been a staple of primary school curricula in North America. Children learn skills such as using guide words, examining headwords to find correct spellings, and reading through senses to find the appropriate meaning for a sentence. These skills have been enshrined in US Common Core standards for elementary school students, and in other regional or national standards of the region as well, (e.g. Barbados, Ontario). However, when students reach secondary school, required dictionary skills instruction – if such a thing exists in a particular locale – does not tend to get more sophisticated. At this stage, students are asked to do little more than either create personal dictionaries for their own use or consult a dictionary for more diverse information; for example, instead of consulting a dictionary just for spelling or meaning, they may be asked to consult it for word origins and pronunciations. While these are laudable skills that should be preserved in the curriculum, Nesi (1999) shows that reference skills are highly complex. She proposes a taxonomy of 40 such skills needed for effective dictionary usage at the university level, including six skills regarding dictionary meta-knowledge, such as “Understanding principles and processes of dictionary compilation” and “Dictionary criticism and evaluation”. Although Nesi’s list is for university students, the skills that she identifies are appropriate for secondary school students as well, even for those not planning to continue to university. In this presentation, I propose a slightly revised version of the taxonomy to reflect developments of the last two decades. I further suggest ways to integrate these goals into the curriculum standards for secondary school students throughout North America. Specifically, I suggest age-appropriate scaffolding of dictionary skills and knowledge to fit in with and enhance existing language arts standards. Lastly, I provide examples of activities that are (or can be) used to develop these skills. These activities are adapted from those found in dictionary use literature or are my own work.

Mark Turin, Christine Schreyer, et al (Panel): *Relational Lexicography: New Frameworks for Community-Informed Dictionary Work with Indigenous Languages*

This panel brings together Indigenous language champions with anthropologists and linguists who are actively collaborating on dictionary projects for First Nations languages in British Columbia and Quebec, Canada.

For under-resourced languages spoken by Indigenous communities, dictionaries contain crucial historical, cultural and territorial information. When languages become endangered, dictionaries become primary tools for language learning. In language communities that have few written records, lexicography can be very time-consuming and labor intensive. In some cases, communities are building new dictionaries, the first of their kind for a language or a dialect, and in others, communities are using legacy language documentation to build new dictionaries to serve wider audiences.

In this double panel we highlight the specific needs and strategic goals that speakers of Indigenous languages have for their dictionary projects and reflect on the challenges that communities face in realizing these goals. All the projects and partnerships in the panel are examples of what we call ‘Relational Lexicography’, a shift towards dictionaries that are created by and with speakers and learners of under-resourced languages themselves, which recognize the relationships in community between speakers, between dialects, and also between community members and academics.

Confirmed participants:

- Ewa Czaykowska-Higgins (UVIC)
- ḥłagm̓il Frances Brown [Heiltsuk First Nation] (Hałzaqv̓la Revitalization Manager, Bella Bella)
- Megan Lukaniec [Huron-Wendat Nation] (University of Victoria)
- John Lyon (University of British Columbia - Okanagan)
- Kathleen Michel [Upper Nicola Indian Band] (En'owkin Centre, (University of British Columbia - Okanagan)
- Mark Turin (University of British Columbia - Vancouver)
- Christine Schreyer (University of British Columbia - Okanagan)

Nianwen Xue (Workshop) *UMR annotation of Chinese verb compounds and related constructions*

In this talk, we will examine the challenges of annotating the predicate-argument structure of Chinese verb compounds in Uniform Meaning Representation (UMR), a recent meaning representation framework that extends Abstract Meaning Representation (AMR) to cross-linguistic settings. The key issue is to decide whether to annotate the argument structure of a verb compound as a whole, or to annotate the argument structure of their component verbs as

well as the relations between them. We will examine different types of Chinese verb compounds, and propose how to annotate them based on the principle of compositionality, level of grammaticalization, and productivity of component verbs.

Beth Young *Using ElasticSearch in EEBO and ECCO to Identify Johnson's Dictionary Quotation Sources*

The quotations that Samuel Johnson includes in his Dictionary of the English Language have long been perceived as one of the more valuable features of the folio editions. In addition to their stated purpose of illustrating word meanings, these quotations have been used by researchers to explore topics including historical language change, Johnson's own reading habits and concerns, the development of a British literary canon, concepts of authorship and authority, and so on.

Researchers often enrich these studies by considering the original contexts from which Johnson drew these quotations. Indeed there is a long tradition of scholarship that focuses on identifying Johnson's sources; unfortunately, these sources have been remarkably challenging to identify. Johnson's citation information is often incomplete, or in some cases, wrong. And his quotations have often been significantly altered from their original form. Johnson provided so many quotations—more than 114,000 in the 1755 first folio edition—as to make even basic research a daunting task. Because of these challenges, the task is not complete.

Meanwhile, the need to verify Johnson's quotation sources grows more acute. The usability of digitized editions, with computerized search functions, is hampered by Johnson's inconsistent and easily confounded citations. Important resources are available online—from Wikipedia to the English Short Title Catalog—but for links to be enabled, the texts must be identified.

Today's readers become increasingly less familiar with Johnson's quoted authors and texts, making them less able to determine, e.g., whether Br. / Bro. / Brown / Browne likely refers to Thomas, Edward, or William. And when centuries-old passages are stripped of their context, readers grow increasingly less able to puzzle out their meanings.

For this reason, the Samuel Johnson's Dictionary Online project, which seeks to make Johnson's Dictionary easier to use and to study, has joined the effort to identify Johnson's sources. This project uses a custom-built ElasticSearch web interface to compare quotations to each other, and to two digital corpora (Early English Books Online (EEBO) and Eighteenth Century Collections Online (ECCO)), and to apply XML tags when we find a match. We plan to use the match data to link to other online databases, such as the English Short Title Catalog, providing contextual information and aiding further investigation.

This presentation will describe how this process works, outline some of its strengths and weaknesses, and share some of the discoveries we have made so far.

Helen Zhang *Collaborative dictionary for Hul'q'umi'num' using a low-code platform*

Hul'q'umi'num' is an indigenous language spoken on Vancouver Island in BC, Canada. The language is endangered, with approximately 40 L1 (mother-tongue) speakers remaining. In recent years there is growing momentum among the community to preserve and reclaim their

language. Currently there are 1,200 active L2 learners, including an estimated 90 fluent and 765 semi-fluent L2 speakers. The Hul'q'umi'num' Language and Culture Society (HLCS) supports the community in language revitalization efforts by offering various certificate and degree programs through a partnership with Simon Fraser University (SFU). Members of the community expressed the desire for an online "talking dictionary" for learners that would feature sound files, as well as a dictionary workflow that would facilitate collaboration and participation. The dictionary database is being maintained in Airtable, a low-code development platform (LCDP) which uses visual interfaces and menus to enable people with limited technical skills to develop applications.

The basis of the current database comes from previous lexicographic work, including several print dictionaries and a 1999 "talking dictionary" which was distributed with sound files on a disk. Spreadsheet software such as MS Excel and Google Sheets was useful in cleaning up and organizing this legacy data, but became difficult to manage when it came time to incorporate new words and sound files.

Airtable was chosen for several reasons. (1) It was crucial to have a relational database to properly handle one-to-many relationships in the data, such as one root being related to many words, or one word having multiple definitions. (2) Setting up views in Airtable is simple, allows multiple people to work on specific tasks and, most importantly, gives them the confidence to do so without being afraid to "mess up" the database. Specific views have been set up for tasks such as tagging lexical suffixes and linking sound files to words. (3) The Airtable interface designer and API makes it possible to make summary dashboards which allow the team to measure progress and see gaps that need to be filled in. (4) The database can be easily exported as a *.csv file, a portable format which is compatible with our website. (5) Finally, given that the HLCS has no permanent IT staff, it made sense to use software with existing documentation, so that the dictionary project is not dependent on a single person's expertise and coding style.

In addition to the above reasons for using Airtable, we discuss some potential downsides - in particular, the idea of "vendor lock-in". We attempt to mitigate this by maintaining an "everything" view in Airtable through linked table and lookup fields, which we can export at any time and have a complete CSV file of all the (text) data. Finally, we must remain wary about using Airtable-specific features such as the interface designer or certain extensions.

Presenter Biographies

Marc Alexander is Professor of English Linguistics at the University of Glasgow and is Director of the *Historical Thesaurus of English*. His research focuses on the study of words, meaning, and effect in English. He has published on historical lexicology, medical discourse, metaphor, astronomical names, color words in English, Parliamentary discourse from 1803 onwards, cognitive linguistics, and reader manipulation in detective fiction. He is a Fellow of the Royal Society of Arts and the Royal Historical Society, a trustee of the Marshall Trust, and a trustee and the Chair of the Board of Directors of Dictionaries of the Scots Language.

Abolfazl Alamdar is a PhD candidate in Linguistics at Allameh Tabataba'i University of Tehran, Iran. His interests include lexicography, corpus linguistics, phraseology, syntax, and exploring new ways of using English corpora to compile more efficient active bilingual dictionaries. He has been involved in compiling a dictionary entitled "An Active Corpus-based English-Persian Collocations Dictionary (ACEPCD)" since 2008, starting as a freelance researcher after finishing high school.

Gregory D. S. Anderson (PhD University of Chicago 2000), Founder and President, Living Tongues Institute for Endangered Languages has published numerous articles and books on endangered languages, language contact, linguistic typology and

historical linguistics. His areas of specialty include the Munda, Turkic, Native Siberian, Trans-Himalayan, Ogonoid, Central Sudanic, Chadic, Nilotic, Salish, Yeniseic and Oceanic languages and the isolate language Burushaski.

Patricia Anderson is a linguistic anthropologist and computer programmer who has worked with the Tunica Language Project since 2011. Working alongside Tunica-Biloxi language teachers, Anderson has contributed to projects such as the Beginning Tunica Textbook, immersion training workshops, translating holiday songs into Tunica, neologism creation extravaganzas, and the soon-to-be-unveiled Indigenous Language Manuscript Interface hosted by the American Philosophical Society. Anderson is the lead lexicographer for the New Tunica Dictionary and maintains the Tunica dictionary app. Anderson is passionate about the intersection of language revitalization and accessible technology.

Grant Barrett is a co-host and co-producer of *A Way with Words*, a national US radio show and podcast about language. He is also an officer of the American Dialect Society. <http://grantbarrett.com>

Lisa Berglund is Professor and Chair of English at Buffalo State University, where she teaches Shakespeare, Milton, eighteenth century studies, the history of the printed book, textual editing, and lexicography. Her

research focuses on Hester Lynch Piozzi, Samuel Johnson, and the history of English language lexicography. She has served as Executive Director of the American Society for Eighteenth Century Studies (2017-2021) and Executive Secretary of the Dictionary Society of North America (2007-2013). She is also President and principal stage director of Opera-Lytes, a community theater organization.

Julia Bonn is a PhD student in Linguistics and Cognitive Science at CU-Boulder following more than a decade as a Senior Research Assistant at CLEAR. She is a long term contributor to PropBank and the PropBank Lexicon, Verbnet, AMR, and UMR, as well as the creator of SpatialAMR, an extension to AMR annotation for fine-grained, multimodal annotation of spatially-rich corpora. Her research interests center on bringing multimodality and pragmatics into semantic annotation (gesture, embodiment, frame of reference, implicit argumentation), lexical resource development, and expanding semantic annotation to typologically diverse languages (Arapaho and Quechua UMRs).

George Aaron Broadwell is Elling Eide Professor of Anthropology at University of Florida. He is a linguist and linguistic anthropologist with research interests in language documentation and language history. He works with Native people in the US and Mexico to develop dictionaries and other language resources that are of value to both local communities and scholars. His

research has focused on the Choctaw, Timucua, Zapotec, and Copala Triqui languages and involves both fieldwork with contemporary speakers as well as study of historical documents in these languages.

Hłagñit (Frances D. Brown) is a descendent of the Heiltsuk and Kitsoo tribes. Frances's mother is Lágaxñáiyax known as Shirley Windsor (nee Mason) of the Qvúkva áitxv du Xaíxais tribes. Frances's late father is Qámas known as Charlie Windsor, of the Wúyálitxv du Yísdaítxv tribes. Frances has spent over two decades revitalizing the Heiltsuk language and culture. As Language Revitalization Manager for the Heiltsuk First Nation, Frances is unwavering in her commitment to improving Haítzaqvła programs and serves on the First Peoples Cultural Council Advisory Committee.

Matt Buchholz is a master's student in the Computational Linguistics, Analytics, Search and Informatics (CLASIC) program at the University of Colorado at Boulder. His research interests include information extraction, distributional lexical semantics, and the computational study of hip-hop lyrics.

Félix Cortés is a native speaker and a specialist of San Sebastián del Monte Mixtec and a self-taught linguist. He was the coordinator and director of the Indigenous interest groups focused on organizational fortification for the Mixteca/Indígena Community Organizing Project in Oxnard, California. Mr. Cortés has been the main

teacher and consultant on the language documentation and revitalization project since its inception. He is the co-author and the greatest contributor to the online dictionary, the co-author of linguistics articles, and the co-teacher of the language revitalization workshop. He is training interested community members in recording, transcription and translation.

Rute Costa is the Head of the Linguistics Centre of the Universidade NOVA de Lisboa and coordinator of the Lexicology, Lexicography and Terminology Group – LLT. She teaches Terminology, Terminology and Ontologies and Lexicography and Terminology in PhD Programs and Master's at NOVA FCSH. She is currently the PI of 2 projects: Mordigital – Digitisation of Dicionario da Lingua Portuguesa by António de Morais Silva [PTDC/LLT-LIN/6841/2020] and Linking Linguistics to Data Science – LL2DS [ERASMUS-EDU-2022-EMJM-DESIGN]. In 2011, she was honored with the title of “Chevalier de l'Ordre des Arts et des Lettres” by the Ministry of Culture and Communication of the French Republic.

Dorian J. Cougias is the Lead Analyst of the UCF and co-founder of Network Frontiers, a company focusing on the science of compliance, including harmonization methods, metrics, systems continuity, and governance. He has served as CIO of leading advertising agencies and has served as CEO of an international software company. He has written and spoken

extensively on all matters of information technology. He advises the Payment Card Security Council, Financial Technology Forum, and other industry organizations. He is also an Emeritus Professor of Technology at the University of Delaware, College of Human Services, Education, and Public Policy.

Andrew Cowell is Professor of Linguistics at the University of Colorado and Director of the Center for Native American and Indigenous Studies. His work focuses on linguistic anthropology, language documentation, and language maintenance and revitalization. He has created textual databases of several languages (Arapaho, Southern Sierra Miwok, Central Sierra Miwok) as well as an extensive lexical database of Arapaho, and another of Aaniiih (Gros Ventre). His recent work has focused on integrating computational and corpus methods with these databases of endangered indigenous languages to make them more accessible to a variety of users.

William Croft is Professor Emeritus at the University of New Mexico. Croft received his PhD at Stanford University under Joseph Greenberg. He has taught at the Universities of Michigan, Manchester and New Mexico, visited the Max Planck Institutes of Psycholinguistics (Nijmegen) and Evolutionary Anthropology (Leipzig), and the Center for Advanced Study in the Behavioral Sciences (Stanford), and given lectures throughout the world. He is the author of dozens of articles and nine books,

including Typology and Universals, Radical Construction Grammar, Explaining Language Change, Cognitive Linguistics (with Alan Cruse), Verbs, Ten Lectures on Construction Grammar and Typology and Morphosyntax.

Ewa Czaykowska-Higgins is a linguist in the Department of Indigenous Education at the University of Victoria whose lexicographic work has spanned 30 years. Working with the nxaʔamxčín Language Program (Colville Tribes), she has been involved in constructing nxaʔamxčín nwwawlxnt, a digital and print dictionary for nxaʔamxčín based on 1960s legacy documentation. She is also working with the WSÁNEĆ School Board on ÁŁŁEŁ SĆÁ: "Heading Out to Sea" a project to digitize, archive and mobilize, for SENĆOTEN language learning, the dictionary work of late PENAĆ Dave Elliott, creator of the SENĆOTEN orthography.

Anna Luisa Daigneault (MSc Université de Montréal 2009) is the Program Director of Living Tongues Institute for Endangered Languages. A linguistic anthropologist and TEDx speaker, her articles about protecting the world's linguistic diversity have been published by The Dominion, Global Voices, and SAPIENS. She specializes in documenting the indigenous languages of the Americas and creating technological tools such as Living Dictionaries for language activists.

Boban Dedović is the Executive Director and chief software architect of the OMNIKA

Foundation, a 501(c)(3) nonprofit dedicated to digitizing and making available all the world's mythological contents. As an interdisciplinary practitioner of psycholinguistics, he studied ancient languages and psychology at the University of Chicago and the University of Maryland en route to a BA, BS, and MA. Most recently, he presented his findings on mental language in the Homeric epics at the International Conference on the Mental Lexicon in Canada. His writings on and training in ancient languages include Sumerian, Homeric Greek, Classical Latin, Middle Egyptian, and Akkadian.

Stefan Dollinger is Professor of English Linguistics at UBC Vancouver. His monographs include *Creating Canadian English* (Cambridge UP, 2019), a biography of the – lexicographical – making of Canadian English. He has recently been probing cross-linguistic incompatibilities around standard notions with *The Pluricentricity Debate: On Austrian German and Other Germanic Standard Varieties* (Routledge, 2019) and *Österreichisches Deutsch oder Deutsch in Österreich? Identitäten im 21. Jahrhundert* (new academic press, 2021). He is chief editor of the Second edition of *A Dictionary of Canadianisms on Historical Principles* (UBC, 2017; www.dchp.ca/dchp2) and commenced, contrary to an announcement at DSNA Barbados, a third edition in 2022.

Daniele Trevelin Donato Luz is Ph.D. candidate at UNESP (São Paulo State

University - Brazil). Her current lexicography project is related to examples in the *Dicionário Digital de Verbos Português- Inglês* for Brazilian students. She has been involved in lexicography projects since 2013, starting as a researcher student as an undergraduate. She has a scholarship from the Brazilian Ministry of Education - CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*).

Donna M. Farina is a professor in the Department of Multicultural Education at New Jersey City University. She holds a Ph.D. in linguistics from the University of Illinois, Urbana. Her research focuses on dictionaries. She has examined topics as diverse as censorship in a 20th-century Russian dictionary, the philosophy of working lexicographers, dictionary front matter, and dictionary look-up among student English learners.

Joseph T. Farquharson is a Senior Lecturer in Linguistics at The University of the West Indies, Mona. Dr. Farquharson holds B.A. in Linguistics and Spanish and a Ph.D. in Linguistics from UWI, and an M.Phil. in European Literature (Spanish) from the University of Cambridge. He has been serving as the Coordinator of the Jamaican Language Unit (JLU) since August 2019 and is currently a member of the Communication & Information Advisory Committee of the Jamaica National Commission for UNESCO, and the Convenor of the subject panel for CAPE Communication Studies.

Rachel Fletcher is a Cambridge Dictionary editor at Cambridge University Press & Assessment. She holds a Ph.D. from the University of Glasgow, where her research focused on periodization in the lexicography of Old English from the seventeenth century to the present day. Rachel has published on topics including the Early Modern lexicographer William Somner and J.R.R. Tolkien's work on the Oxford English Dictionary. She also co-edits the DSNA Newsletter.

Daniele Franceschi is Associate Professor of English at Roma Tre University, Italy (Department of Foreign Languages, Literatures and Cultures), where he teaches at undergraduate, graduate and doctoral level. His publications are in the area of lexicography, lexicology, temporal semantics, ESP and cognitive studies. He is currently working on a taxonomy of semantic relations holding between near-synonyms of different origin in contemporary English as well as on the creation of a bilingual (English-Italian) dictionary of sports and games terminology.

María Teresa Fuentes Morán is Associate Professor in the Department of Translation and Interpreting (Faculty of Translation and Documentation), University of Salamanca (Spain). Her teaching and research focuses on applied aspects of Spanish Language and Linguistics, especially in the field of Lexicography, Lexical acquisition and accessibility. She has collaborated as co-author, advisor or reviewer of a large

number of dictionaries and her work is currently focused on the description of e-dictionary forms and their accessibility parameters.

Jan Hajič is professor of Computational Linguistics and deputy director of the Institute of Formal and Applied Linguistics at the School of Computer Science, Charles University, Prague. His field is NLP and building language resources with rich linguistic annotation. He is currently also the director of LINDAT/CLARIAH-CZ, a large research infrastructure on language resources. He has published more than 200 papers, a book, chapters, and encyclopedia and handbook entries. He is the chair of the Executive Board of META-NET, European research network in Language Technology, chair of the TACL Steering Committee, and a member of other international boards.

Volker Harm is a lexicographer at the Academy of Sciences in Göttingen and Associate Professor at Göttingen University, where he teaches general linguistics and linguistics of German. He was lexicographer at *Deutsches Wörterbuch von Jacob und Wilhelm Grimm* from 1998 to 2016, from 2011 as senior editor. Currently, he is senior editor of *Wortgeschichte digital* at the Göttingen Academy. He is author of several books about linguistics and lexicology (e.g. *Einführung in die Lexikologie*, 2015, and *Funktionsverbgefüge des Deutschen*, 2021). His most recent publication is *Historische Lexikographie des Deutschen* (2022), a co-edited volume

about current developments in historical lexicography of German.

Mandy Na'zinek Jimmie is Nl̓eʔkepmx from the Southern Interior of British Columbia and lives in her community (homeland). She is currently an Assistant Professor at the University of British Columbia-Okanagan campus in the Nl̓eʔkepmx Language Fluency Degree Program. In the early 1990's completed a Master of Arts Linguistic program at the University of British Columbia. Mandy was a post-secondary Nl̓eʔkepmx language instructor at the Nicola Valley Institute of Technology, an Indigenous Institute started by the five local Indigenous communities since 1994. She is actively involved in local, regional and provincial initiatives in Indigenous languages since then.

Pavlina Kalm completed her PhD in Linguistics at the University of New Mexico in August 2022. Her dissertation work titled *Social verbs: a force-dynamic analysis* proposes a semantic analysis of verbs that describes various types of social interactions between people. Kalm was the primary developer of the force-dynamic representations of event structure associated with VerbNet. Kalm has also published research that deals with event structure representation and the intersection of verbal semantics and the semantics of argument structure constructions. Her other research interests include typology and syntax. Kalm was a

recipient of the Bilinski Fellowship awarded to outstanding dissertation fellows in 2022.

Robert Krovetz is President of Lexical Research, a company doing research and development in natural language processing. His work focuses on word sense disambiguation, multiword expressions, and derivational morphology. He has published more than 30 papers, including one on morphology and information retrieval that has been cited by more than 1000 other papers and was ranked by Citeseer as one of the 100 most-cited papers of the year in computer science. Dr. Krovetz received his PhD from the University of Massachusetts at Amherst.

Anatoly Liberman emigrated to the US in 1975 with the highest academic degree available in Europe (Habilitation) from the Academy of Sciences of the USSR and has taught at the University of Minnesota ever since. His areas of research include but are not limited to historical Germanic linguistics. In lexicography, his focus has for several decades been on word origins and etymological dictionaries, mainly English.

Nora Livesay is Editor of the Ojibwe People's Dictionary, and an instructor for American Indian Studies at the University of Minnesota - Twin Cities. She has an M.A. in Linguistics and an M.Ed. in Indigenous Language Revitalization. Her current research project "Developing Indigenous scholars, curriculum and language documentation," (#2152631, NSF/NEH DLI-DEL) focuses on expanding

Anishinaabemowin pedagogical resources, and building capacity for community-directed dictionary editing.

Ligeia Lugli holds a PhD in Study of Religions from SOAS, where she worked on the Mahāyāna discourse on language. She is Head of Lexicography at the Mangalam Research Centre (Berkeley, California), where she directs the Buddhist Translators Workbench, a project aimed at creating digital resources for the study of the Buddhist Sanskrit vocabulary. She is currently leading two NEH-funded projects related to the development of computational tools for Buddhist Sanskrit Natural Language Processing and general lexicography.

Megan Lukaniec is a member of the Huron-Wendat Nation of Wendake, Québec and an Assistant Professor of Indigenous Language Revitalization in the Indigenous Studies Program at the University of Victoria. Since 2006, she has been working with and for her community to reawaken the Wendat language, which was dormant for over 150 years. Her most recent research, supported by a SSHRC grant and in collaboration with the University of Victoria's Humanities Computing and Media Centre, focuses on the development of online bilingual dictionaries of Wendat through building and analyzing a digital corpus of archival documentation encoded in XML.

John Lyon is an Assistant Professor in UBC Okanagan's Department of Community, Culture, & Global Studies, and teaches in

the Bachelor of Nsyilxcn Language Fluency Program, the first program of its kind in Canada. He is a trained linguist, whose research has long been with Interior Salish language communities, most recently with St'ait'imcets (Northern Interior Salish) and Nsyilxcn (Southern Interior Salish). He has completed dictionary projects for Coeur d'Alene Salish (Lyon & Greenwood 2008; Lyon 2010), and is currently working with the Upper Nicola Syilx community to produce a sound dictionary of heritage Nsyilxcn materials.

Iara Mantenuto is an Assistant Professor of Linguistics and the Program Coordinator for the Masters in Teaching of English as a Second Language at California State University Dominguez Hills in California. She is from Italy, and she earned a Ph.D. in Linguistics from UCLA. Dr. Mantenuto's research focuses on revitalization, reclamation and documentation, and morphosyntactic analysis of understudied languages – in particular Teramano (her heritage language) and San Sebastián del Monte Mixtec. Her research also explores the teaching of linguistics and the implementation of pedagogical techniques in language revitalization and reclamation contexts.

Sterling Martin (*Diné*), is a Postdoctoral Research Associate at Washington University in St. Louis studying regeneration in segmented marine worms. Sterling grew up in Shiprock, New Mexico where he graduated High School in 2008.

He continued his education and obtained his B.S. in Biochemistry from the University of Iowa in 2012 and then his Ph.D. in Biophysics from The University of Wisconsin - Madison in 2022. To bridge the gap between the languages of science and *Diné bizaad* (Navajo language), Sterling (amongst others) co-founded Project ENABLE (Enriching Navajo as a Biology Language for Education) to create modern Biology terms in *Diné bizaad*.

Kathy Michel is a member of the Upper Nicola Indian Band, and a 4th year student of the Bachelor Of Nsyilxcn Language Fluency Program at UBC Okanagan, and one of the inaugural graduating cohort. She is working with Dr. John Lyon and other students from the Upper Nicola to transcribe heritage sound recordings of her Uncle Joe Albert Michel made over 40 years ago, and to process these into a format which will culminate in a sound dictionary resource for the community.

Rod McConchie, PhD, Docent, is retired, having worked at Helsinki University, and is now an unaffiliated researcher. He is also a life member of Clare Hall, Cambridge. He is currently writing a book on Dr. Robert James (1703-1776), and is engaged more generally with the history of medical lexicography

Olga Menagarishvili (PhD in Rhetoric and Scientific and Technical Communication, University of Minnesota) is Assistant Professor in the Department of Technical Communication and Interaction Design at

Metropolitan State University in St. Paul, Minnesota. She has also taught scientific and technical communication at Georgia Institute of Technology in Atlanta, Georgia, the University of Minnesota in the Twin Cities, Minnesota, and the Royal Institute of Technology in Stockholm, Sweden. She specialized in lexicography as an undergraduate student and continued to explore dictionaries in graduate school and beyond. Her research interests include scientific and technical communication, lexicography, rhetoric, and multimodality.

Linda C. Mitchell is Professor of English and Comparative Literature at San José State University. She is also the author of *Grammar Wars: Language as Cultural Battleground in 17th- and 18th-Century England*; editor of *Letter-Writing Manuals and Instruction from Antiquity to the Present*; co-editor of "Study in the Cultural History of Letter Writing," *Huntington Library Quarterly*, Vol. 66 (2003); and editor of "Social History of Letter Writing," *Huntington Library Quarterly*, Vol. 78 (2016). Other works include John Milton's *Accedence Commenc't Grammar* with introduction and transcription in *Complete Works of John Milton* (Vol. IX),. Forthcoming, *Cultural History of English Lexicography, 1600-1800: The Authoritative Word*.

Brad Montgomery-Anderson specializes in nineteenth-century studies and Native American languages. He is the author of *Cherokee Reference Grammar*, which won

the Linguistic Society of America's 2017 Bloomfield Book Award. He spent four summers in Mexico with the Project for the Documentation of the Languages of Mesoamerica, during which he wrote a bilingual Chontal Mayan-Spanish dictionary. He has published articles in *Southwest Journal of Linguistics*, *International Journal of Applied Linguistics*, *Nineteenth-Century Contexts*, *Journal of American Indian Education*, and *Journal of Language Contact*. He teaches Latin, Greek, and linguistics at Colorado Mesa University in Grand Junction, Colorado. He grew up in Boulder.

M. Lynne Murphy is Professor of Linguistics at the University of Sussex, in Brighton, England. Her interests lie in lexicology, lexical semantics, pragmatics, and transatlantic Englishes, and her books include *Semantic Relations and the Lexicon* (2003), *Lexical Meaning* (2010) and *The Prodigal Tongue* (2018). Since 2020, Murphy is editor of *Dictionaries: The Journal of the Dictionary Society of North America*.

Skatje Myers has just earned her PhD in Computer Science and Cognitive Science at the University of Colorado at Boulder. Her interests include expanding semantic NLP to new domains and languages and exploring ways of using techniques such as active learning and annotation projection to make more efficient use of human annotation.

Nguyễn Văn Bền was born 1954, Vietnamese, retired urologist and freelance

linguist. He studies spoken English, English word stress, and phonological aspects of English-related dictionaries. His works including books, dictionaries are in papers and software. He also makes a Media Dictionary which works online. Its website is <http://mediadic.com/>. The dictionary uses pictures, sounds, animations to explain words, idioms. He would like to spread and share his works.

Alexis Palmer is an Assistant Professor of Linguistics at the University of Colorado, with a specialization in computational linguistics. Her research interests include computational semantics and discourse, as well as the development and application of computational methods for endangered and other low-resource languages. She is a co-organizer of the ComputEL workshop series on computational methods for the study of endangered languages and a co-founder of the new SIGEL special interest group. She has also served as a co-organizer for the SEMEVAL workshop and is currently program co-chair for the *SEM 2023 conference.

Martha Palmer is the past Helen & Hubert Croft Professor of Engineering in the Computer Science Department, and Arts & Sciences Professor of Distinction for Linguistics, at CU-Boulder, with over 300 peer-reviewed publications. She is a co-Director of CLEAR, an Association of Computational Linguistics Fellow, an Association for the Advancement of Artificial Intelligence Fellow, and a co-

editor of LiLT: Linguistic Issues in Language Technology. She has previously served as co-editor of the *Journal of Natural Language Engineering*, a member of the editorial board of Computational Linguistics and of TACL, President of ACL, Chair of SIGLEX, and Founding Chair of SIGHAN.

Bill Poser received his A.B. in Linguistics from Harvard College and his Ph.D. from MIT. He previously taught Linguistics at Stanford and First Nations Studies at the University of Northern British Columbia. His early work was on languages of Northeast Asia, especially Japanese. For thirty years he has worked on Carrier and Sekani, Athabaskan languages of British Columbia, for which he has created variety of bilingual dictionaries, paper and electronic. He has also written the software for dictionaries of Montana Salish, Kootenai, and Nahuatl.

James Pustejovsky is the TJX Feldberg Chair in Computer Science at Brandeis University. He works on computational and lexical semantics, is creator of Generative Lexicon Theory, and has worked to develop lexical semantic resources for the NLP community. He is chief architect for TimeML and ISO-TimeML, a recently adopted ISO standard for temporal information in language; the recently adopted standard ISO-Space, a specification for spatial information in language; and the co-creator (with N. Krishnaswamy) of the VoxML multimodal modeling framework, encoding

communication between humans and computers or robots utilizing speech, gesture, gaze, and action.

Kevin Rottet is professor of French Linguistics and adjunct professor in the Linguistics Department at Indiana University. His research focuses primarily on language contact, particularly in three settings: French and Creole in contact with English in Louisiana; English-Welsh contact in Wales; and French-Breton contact in Brittany. He is co-author of the *Dictionary of Louisiana Creole* (IUP, 1998), the *Dictionary of Louisiana French* (University Press of Mississippi, 2009), and the *Comparative Stylistics of Welsh and English* (University of Wales Press, 2018) and he is currently completing the collaborative *Dictionnaire étymologique, historique et comparatif du français de Louisiane* (DEHCFL).

Lindsay Rose Russell is the author of *Women and Dictionary Making: Gender, Genre, and English Language Lexicography* (Cambridge University Press, 2018). In addition to teaching in the Department of English and The Center for Writing Studies at the University of Illinois, Russell is Executive Director of the Dictionary Society of North America and Associate Editor of the *International Journal of Lexicography*.

Ana Salgado is a lexicographer and researcher at the Centro de Linguística da Universidade NOVA de Lisboa. She holds a PhD in Translation and Terminology (Universidade NOVA de Lisboa) and

graduated in Portuguese Studies (Faculdade de Letras da Universidade do Porto). She is the chair of Instituto de Lexicologia e Lexicografia da Língua of Academia das Ciências de Lisboa, where she is the chief editor of online dictionaries and one of the editors of *Thesaurus de Ciências da Terra*. She is also the project leader in revising the standard ISO 1951:2007. Her main research interests include lexicography, terminology and data modelling.

Christine Schreyer is an Associate Professor of anthropology at the University of British Columbia, Okanagan campus, where she teaches courses in linguistic anthropology. Her research focuses on language documentation and revitalization in Canada and in Papua New Guinea, with a specific focus on dictionary making practices and the use of technology for language documentation and revitalization. She is also interested in the relationship between endangered language communities and created language communities and she has conducted research with the Na'vi speech community (from the movie *Avatar*). Recently, she has been investigating babywearing terminology used in the online community of babywearers.

Jesse Sheidlower is the editor of the *Historical Dictionary of Science Fiction*. A former Editor at Large of the *Oxford English Dictionary*, he currently teaches a course in lexicography at Columbia University.

Jason Siegel is Research Fellow in Lexicography at The University of the West Indies-Cave Hill, in Barbados. He specializes in languages of the English and French Caribbean. He serves as the Barbados Focal Point for UNESCO's World Atlas of Languages, and is part of the founding committee for AmericaLex, DSNA's sister organization for Latin America and the Caribbean.

David Skinner is author of *The Story of Ain't: America, Its Language, and the Most Controversial Dictionary Ever Published* (2012), which is about *Webster's Third*. He was on the usage panel of the *American Heritage Dictionary* and wrote a widely noted essay about the panel after it was disbanded in 2018. He is editor of *Humanities* magazine, published by the National Endowment for the Humanities. (His views are his own and not those of NEH.) Before joining NEH, he was a staff editor at *The Weekly Standard*. He has written for many publications, including, recently, *American Purpose* and *Washington Free Beacon*.

Rachel Stone is a Montréal-based linguist and lexicographer for Antidote, a writing assistance application for French and English. She studied Linguistics and Classics at McGill University, where she conducted a study of residual Scottish Gaelic influence on the English of Maxville, Ontario. She previously interned at the Dictionary of Old English in Toronto. In her free time, she helms the SeaTalk.ca Nautical

Dictionary by the late sea captain Mike MacKenzie.

Mark Turin (PhD, Linguistics, Leiden University, 2006) is an anthropologist, linguist and occasional radio presenter, and an Associate Professor at the University of British Columbia. He is cross-appointed between the Institute for Critical Indigenous Studies and the Department of Anthropology. Mark is the principal investigator for the Relational Lexicography project through which a group of aligned researchers are developing a framework and toolkit for collaborative, community-informed dictionary work with marginalized languages. Mark writes and teaches on language reclamation, revitalization, documentation and conservation; language mapping, policies, politics and language rights; orality, archives, digital tools and technology. Twitter: @markturin

Alenka Vrbinč is professor of linguistics at the University of Ljubljana, Slovenia. She is primarily interested in lexicographic issues, the emphasis being on bi- and monolingual lexicography, learner's lexicography and e-lexicography and dictionary testing. She also does research on different phraseological topics such as culture-boundness of phraseology, contrastive idiom analysis and cross-linguistic research. She edited a German-Slovenian dictionary, coauthored an English-Slovenian dictionary and wrote a book entitled *A cross-linguistic and cross-cultural analysis of English and Slovene onomastic phraseological units*.

Nianwen Xue is a professor in the Computer Science Department and the Language & Linguistics Program at Brandeis University. His research interests include developing linguistic corpora annotated with syntactic, semantic, and discourse structures, as well as machine learning approaches to syntactic, semantic and discourse parsing. He is an action editor for *Computational Linguistics* and currently serves on the editorial boards of *Language Resources and Evaluation* (LRE). He also served as the editor-in-chief of the ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) from 2016 to 2019. He is currently the Chair-Elect of Sighan, an ACL special interest group in Chinese language processing.

Beth Rapp Young is Associate Professor of English at the University of Central Florida in Orlando, and the PI of the Johnson's Dictionary Online project (johnsonsdictionaryonline.com).

Regiani Zacarias is a Full Professor in English Language Teaching in the Department of Modern Languages at the São Paulo State University (UNESP), Brazil. She is also a professor at the Graduate Program in Linguistics and Portuguese and at the Professional Master's Program in Basic Education, in the same institution. She is a researcher with the São Paulo Research Foundation (FAPESP) and with the Brazilian National Council for Scientific and

Technological Development. Her primary research interests are pedagogical bilingual lexicography and metalexicography and the teaching of English as international language and Portuguese as foreign language.

Helen Zhang a master's student at Simon Fraser University on traditional Coast Salish territory in BC, Canada and is fortunate to work with members of the Hul'q'umi'num' Language & Culture Society on many things, including a dictionary project. She was raised in Rhode Island, USA and her parents are from Shanghai, China. While she could understand my parents regional dialect as a child, she always ended up responding in English. After a brief stint teaching English in Spain where I had a much easier time learning to speak Spanish compared to my heritage language, she decided to go back to school for linguistics.

Been there? Done that.

The conference schedule leaves few gaps for leisure but if you find an evening free or just feel like skiving, here are Boulder's best-known attractions, all easily reachable from campus.

Boulder Creek Path Boulder Creek flows out of the mountains through Boulder and has a paved walk and bike path all along the way.

Boulder Farmer's Market You don't dare miss any of Saturday morning's conference offerings, but if you do, there's the Boulder Farmer's market, on 13th between Arapahoe and Canyon. Local vendors and a small outdoor food court. Open till 1 pm.

The Colorado Chautauqua Celebrating its 125th anniversary this summer, the Chautauqua in Boulder is the only one still operating west of the Mississippi. There are numerous summer activities. Trailheads near the Flatirons start at the adjacent Chautauqua Park. The Dining Hall is also very good. It's at the top (west end) of Baseline Road.

Dushanbe Tea House Dushanbe (Душанбе), capital of Tajikistan, is a sister city to Boulder. In 1987 the people of Dushanbe presented Boulder with the gift of the Dushanbe Tea House, located on 13th between Canyon and Arapahoe. Amazing architecture for this part of the world, and good food to boot. Also delightful afternoon tea. Reservations suggested.

Museum of Boulder A history museum that documents the Boulder area from 300 million years ago to the present day. Broadway and Pine Street, just north of downtown.

Pearl Street Mall Pedestrian mall that runs from 11th to 15th Street on Pearl in downtown Boulder. Shopping and dining, good people-watching, often a few buskers, and the Boulder County Courthouse, a modest achievement in Art Deco.

Rent a Bike You can rent an e-Bike at multiple locations all over Boulder. You needn't return the bike to where you picked it up, just leave it at any Boulder BCycle location. There is one just outside the UMC (18th and Euclid) and one just outside our partner hotels (28th and Canyon). See <https://boulder.bcycle.com/>. Boulder is one of the country's most bike friendly cities. You can get nearly anywhere in Boulder on designated bikeways that largely avoid busy roads.

Take a Hike You can head West (towards the mountains) on Baseline Road, Arapahoe, Canyon, or Mapleton and you'll land, eventually, at trails and trailheads that take you up into the foothills and mountains. The Boulder Creek Path (see above) joins Canyon Street west of 6th Street and continues up Boulder Canyon on a paved path.

University of Colorado Museum of Natural History It's right next door to the UMC (west) and you can slip in easily for a short visit to bone up on dinosaurs and the like. There are also exhibits that highlight early Indigenous life in the Boulder Valley.