White Paper

Report ID: 2880668

Application Number: HD-228732-15

Project Director: Mark David LeBlanc

Institution: Wheaton College

Reporting Period: 5/1/2015-6/30/2017

Report Due: 9/30/2017

Date Submitted: 9/27/2017

1. Cover Page

Final Performance Report

HD-228732-15 – DH SUG Level II

"Easing Entry and Improving Access to Computer-Assisted Text Analysis for the Humanities"

Mark D. LeBlanc and Michael D.C. Drout, Wheaton College, Norton, MA
Scott Kleinman, California State University, Northridge

Submitted: September 30, 2017

2. Project Activities

We describe the successful completion of an NEH Digital Humanities Start-Up grant (Level II, 2015-2017) where we developed new functionalities for our *Lexos* text analysis software in order to position it to better serve a broader range of humanities users. ***Lexos* is a browser-based suite of tools that provides an easy entry to computational text analysis for humanities scholars and students**. But easing access does not by itself teach effective practices and workflows when learning to design and run computational experiments on texts. Sharing effective practices and workflows *while* using the tools is our attempt to "white box" the fundamental and many steps involved, for example, concerning document preparation and tokenization – two methods that must be transparent if we are to replicate experiments.[1] The *Lexos* workflow requires that the user be intentional while making the many choices that can affect the exploratory analyses to come.

We believe that future generations of scholars will turn, at least partially, to methods that allow a computational reading of their texts and corpora and will need more than just tools to turn their digitized texts into cool charts (although *Lexos* does produce some visualizations). Part of *Lexos*' contribution is **to teach effective practices, workflows, and limitations in performing computational text analysis**. Situated within a clean and simple interface, *Lexos* consolidates the common yet frustrating pre-processing operations that are needed for subsequent analysis, either in *Lexos*' internal workflow or for use with other external tools (e.g., scrub and cut in *Lexos* for later use in Stylo in R, Eder, *et al.* 2016). In our experience, scholars who might like to perform computational analysis in their areas of expertise and/or wish to teach their students how to do so become discouraged too early in the game. They quickly realize that they either lack the time for a manual preparation of texts, the skill sets needed to prepare their texts for a particular tool, or the intellectual contexts for situating computational methods within their work.

Prior to the grant period, we had identified three significant challenges for our intended users. The first challenge involves the **adoption** of computational text analysis methods. *Lexos* specifically targets "new-adopters" through a simple, browser-based interface that manages workflow through the three major steps of text analysis: pre-processing, generation of statistical data, and visualization.

The second challenge is the **opacity** of the procedures required to move between computational and traditional forms of text analysis. The well-known "black boxiness" of algorithmic methods represents part of the challenge for users in the humanities. In this work we have designed and built the system architecture for our "In the Margins" component: an aspiration for the text analysis community to share more of the "how to" and "why" stories of computational text analysis.[2] We now use the media-rich publishing platform Scalar (http://scalar.usc.edu/), including use of Scalar's built-in application programmer's interface (API), to embed content from the Scalar "book" within the *Lexos* interface itself, making it available as part of the workflow.

The third challenge is the **tension** between quantitative and computational approaches to text analysis and the traditions of theoretical and cultural criticism that dominate the humanities in the academy. A desired goal was to use *Lexos*' existing strengths to open a space for discussion of issues related to the opacity of algorithmic approaches and the limitations of visual representation of

---

[1] LeBlanc (2017) argued at DH2017 in Montréal, Canada in August 2017 that the stylometry community must do better moving forward if we are to replicate each other's experiments. We agree with David Hoover (NYU) that the time for a workshop on tokenization is ripe (see the section Continuation of the Project for a glimpse at our plans).

[2] See the Voyant Tools Help in Sinclair, Stéfan and Geoffrey Rockwell, 2016. *Voyant Tools*. Web: http://voyant-tools.org/.

humanities data. Our ITM prototype is one beginning toward reaching this goal and the Lexomics Research group's scholary publications form another contribution. Our dissemination efforts that have emerged from this project are contributing (by example) to this discussion.

Major Activities and Accomplishments:
Our Lexomics Research Group is comprised of two English professors, one computer science professor, and over the last three years during this funding cycle, dozens of undergraduate software developers and humanities scholars. The group's model of simultaneous scholarly activity and software development ensures that we continuously adapt *Lexos* to the needs of our core users. As with all software tools, *Lexos*' user base continues to request new functionality. The following list itemizes our Activities and Accomplishments, including those included in the original proposal that met the spirit of "high risk/high reward" within the timeframe of this proposal. We mention the tasks we completed and those items we did not, including some action items that we have moved to future funding plans. We share activities and accomplishments in four areas:

- *Lexos* Workflow and Tools
- "In the Margins" (ITM)
- Software Development
- Lexomics Research

Workflow and Tools
We met most of the original objectives in our grant proposal and completed others that emerged during our work which were not mentioned in the proposal. We offer an annotated list of completed action items:

a) A number of subtle but significant changes occurred in our code base, including moving our (User Interface) UI to the Bootstrap framework, moving many front-end functions to Ajax, updating our version of the DataTables Javascript library, improving error messages and warnings given to the user in the UI, and the addition of new Twitter and Feedback buttons.
b) For users who wish to run experiments locally, for example to avoid transferring large collections of texts over the web to our public server, we implemented MacOS and Windows installers.
c) Our old Select tool was replaced with a new Manage tool. Users now have a more natural way to select from those uploaded files that are to be active in a current analysis.
d) The new Manage tool allows users to tag collections of text with a class label, e.g., all of Austen's novels marked with an "Austen" class label. Class labels are used in our TopWord tool, as well as to position *Lexos* for future tools that perform classification analyses.
e) Improvements to *Lexos'* pre-processing ("scrubbing") and tokenization features. Our attention here supports our call for better reproducibility in computational methods, that is, we submit that the stylometry community must do a better job of sharing just how texts are pre-processed in a published work (*cf.* LeBlanc, 2017):

    (i) Enable user to remove whitespace (helpful when working in non-Western languages).
    (ii) Wiser handling of ampersands (&) which may or may not be HTML, SGML, or XML entities.

(iii) Added a "Keep Words" option; this is the opposite of "Stop Words" where users upload the words they want removed; Keep Words allows them to filter out text not present in their pre-determined list of terms.

(iv) Added full support for scrubbing XML/TEI texts. Users can selectively handle specific tags, allowing them to take advantage of embedded markup in preparing their texts (e.g. by choosing between diplomatic and critical readings in a TEI text). The function also handles well-formed XHTML.

(v) Added new character entity conversion rule sets based on the Medieval Unicode Font Initiative (MUFI) to allow ingest of texts encoded according to this standard. These rule sets provide proof of concept for other rule sets to be developed for languages containing many encoded special characters.

(vi) Added a new server-side tokenizer (e.g., counting instances of each word). This better handles large Document Term Matrices (DTMs).

(vii) Added a show/hide columns feature in the Tokenize tool.

(viii) *Lexos* users can now enter "milestones" in their uploaded document to signify a "cut here" marker. This adds a user-centric choice to *Lexos*' multiple options for segmenting a text into smaller segments. Cutting by milestones helps users to cut a text into arbitrary segment sizes. Detecting milestones also features in subsequent tools like Rolling Windows to display segment boundaries that correspond to chapters or other user-defined structural divisions. This addresses the opacity of the relationships between source texts and data analysis.

f) Added new features to the workflow, including "culling" (e.g., uses only the top-N words and/or words that appear in at least in L of M documents).

g) Scott Kleinman completed work on visualizing topic modeling data produced by MALLET (McCallum, 2002). Topic modeling is another method of exploring the semantic properties that can supplement the algorithmic methods available within the *Lexos* tool set. However, its adoption has been limited by the difficulty of interpreting MALLET's output data. We added to the *Lexos* Multicloud tool the ability to generate "topic clouds" using a "topic-word-counts" file created by MALLET. Our future work will continue to seek to determine how *Lexos* can visualize data from other topic modeling programs, particularly other forms of MALLET data output.

h) Our four visualization tools were all improved. WordCloud, Multicloud, and Bubbleviz were given improved tooltips.

i) The WordCloud visualization tool was upgraded so that the term counts table was switched to DataTables with search and sort functions.

j) The Multicloud visualization has a new toggle switch for topic clouds.

k) Our Rolling Windows tool and associated visualizations have benefited from a reorganized UI display.

l) Our Statistics tool was switched to display data using the DataTables Javascript library with client-side download buttons for multiple formats.

m) The Hierarchical Clustering tool now by default creates a dendrogram as a PNG file; a new Newick format export button is included. This allows users to download dendrogram architectures to use with other dendrogram visualization tools.

n) K-Means Clustering has new interactive tooltips for Voronoi cells for better readability. Its Principal Component Analysis view is now generated with the Plotly graphing library, which provides pan and zoom, as well as export features.
o) The Similarity Query tool was rewritten with more efficient code. Output now uses the DataTables Javascript library.
p) TopWords tool has more robust comparison options (e.g., compare a document to other documents, compare a document to other classes of documents, and compare classes of documents to other classes). TopWords also has a clearer interface.

We did not add our goal of one supervised classification method to our existing unsupervised options (hierarchical and K-means clustering). A decision was made in the second year that we must limit new functionality if we were to complete our other priorities. We have moved this work to future funding plans. However, our new feature that allows the user to set class labels is part of our continued work to add classification techniques and moves *Lexos* closer to this goal.

"In the Margins" (ITM)

We have designed and prototyped the system architecture for our "In the Margins" (ITM) component using the media-rich publishing platform Scalar (http://scalar.usc.edu/works/lexos), including use of Scalar's built-in application programmer's interface (API) to embed content from the Scalar "book" within the *Lexos* interface itself, making it available as part of the workflow. "In the Margins" is an innovative plan to express the "how to" and "why" documentation for *Lexos* in a form that explores the tensions between quantitative methods often created in the disciplines and theoretical traditions prevalent in the humanities. The ITM Scalar book also contains the fully-documented user manual for *Lexos*.

Work completed during the time of this grant includes the inclusion of an ITM-tab (on the left-hand side of the screen) on each page, thereby providing commentary and suggestions in context, for example, while inside the Scrubber tool the ITM pages illuminate a discussion of how *Lexos* handles each of the user's preprocessing directives.

Although we were able to produce a number of short 3-10 minute "talking head" videos by disciplinary experts (we dedicated one student videographer in Year 2), we were unable to integrate the "talking head expert" clips within the ITM as originally proposed. In short, we underestimated the videography expertise needed to produce quality materials. We address our continued plans to weave video content into ITM in the section Continuation of the Project below.

ITM was intended to be a space for discussion of best practices in the field of computational text analysis, a need that has been recognized elsewhere in the Digital Humanities community, as indicated by the creation of a Special Interest Group (SIG) for Digital Literary Stylistics by the Alliance of Digital Humanities Organizations. We have incorporated efforts from elsewhere in the community by embedding significant Zotero bibliographies in ITM. Our own contributions have been largely devoted to exploring the use of hierarchical clustering, with some materials based on experiments undertaken in the summer of 2017 yet to be incorporated in the online text. We expect to continue adding materials to ITM on an ongoing basis.

Software
This grant funded significant improvements to our code base and the adoption of new software practices. Software development of Lexos now follows a new suite of effective software engineering practices, including code linting, unit testing, type hinting, PEP-8 documentation standards, and continuous integration.

Our open source software is publically available at GitHub:

https://github.com/WheatonCS/Lexos

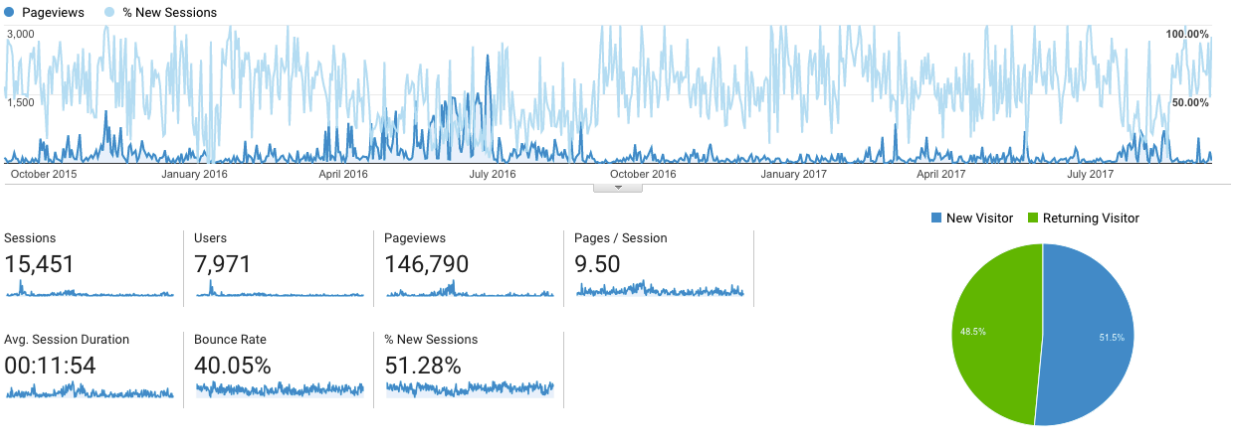
Since 2015, we have released **Lexos v2.0, v3.0, and v3.1.1** [using Python v2.7]**.** *Lexos* v3.1.1 runs on the live public server (http://lexos.wheatoncollege.edu), a new 2017 Ubuntu Linux server funded by and hosted at Wheaton College [as of 09/30/2017].

With an eye towards sustainability, we have continued development past the time of this funding to produce *Lexos* v3.2, which includes a refactoring effort in Python v3.6.x. We address our additional next steps in the section Continuation of the Project below.

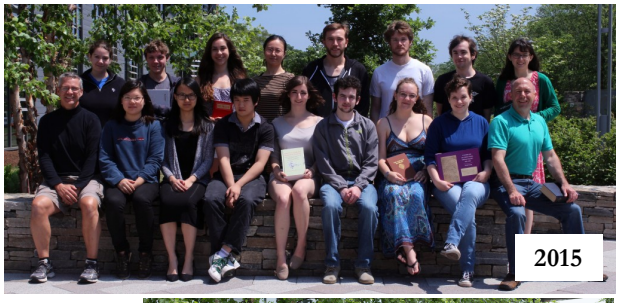
Our GitHub repository tells a good tale of software development as well as a measure of the range and number of our community of developers. An undergraduate senior and lead developer Cheng Zhang '18 responded to a request for stats on our **GitHub** (software) repository with the following post on Slack (our communication channel as we work):

> chantisnake [3:08 PM]
> in the last 4 years we have worked on this for 1586 days, there are 322 active days; we have a total of 608 files, with 1,219,095 lines of code; We have (presently) 59 Python files and 18,348 lines of Python source code; we have 53 contributors (mostly undergraduates) and we have had a total of 5110 commits.
>
> **2017**: 1518 commits, 1,323,466 lines added and 1,490,002 lines removed
> **2016**: 1615 commits, 2,772,011 lines added and 1,673,556 lines removed

Server Analytics
Server analytics show over 15,000 sessions by nearly 8,000 users over the course of the grant period. Nearly 50% of new visitors return to the site. Visitor flow was active throughout the entire grant period. (Note: Some users have also presumably installed the local version of *Lexos* for use on their own machines.)



We have tried to analyze use via online surveys, but these did not return enough responses to be useful. By word of mouth and communications with hands-on workshop attendees, we know that *Lexos* is used for both teaching and research at a variety of institutions of higher education. We have had users approach us from different regions of the United States, as well as Canada and Mexico. We are aware of sustained usage by participants in the NEH-funded Workshop on Building and Strengthening Digital Humanities through a Regional Network (2015), and *Lexos* is now being used by participants in the DH Nepal summer institute where Scott Kleinman taught in Kathmandu in 2017. Advisory Board member Alan Liu also reports regular use of *Lexos* in his classes, and it contributed to the early development of the 4Humanities WhatEvery1Says project (recently funded by the Mellon Foundation), of which Liu and Kleinman are co-PIs. We expect further collaboration between the two projects in the future.


2015

2016

2017

The Lexomics Research Group

Our Lexomics Research Group continues to serve as a motivating example for interdisciplinary work in the Digital Humanities at Wheaton College, in the very DH-active New England region, and beyond. For the last eight years (six of those with NEH funding), Wheaton College has funded 60^{+} undergraduate researchers. In addition to work during the academic year, each summer at Wheaton is a merge of "grad school" and "start-up company" as 8-12 computer science, mathematics, and majors in the Humanities work in one lab together with project faculty, including visiting guest faculty and their students. Our dissemination successes during this funding period speaks to our commitment to active scholarship in the Digital Humanities: 1 book, 10 published articles, 13 talks, 9 hands-on workshops, 2 articles forthcoming, 3 articles in review, and 4 articles and books in progress. These are listed on the Lexomics

website: https://wheatoncollege.edu/academics/special-projects-initiatives/lexomics/publications-grants/, and in the Dissemination Totals section below. In addition, LeBlanc continues to teach and disseminate materials for the "Computing for Poets" course, an active part of the Wheaton connected curriculum, featuring English and Computer Science course connections (Boese, *et al.*, 2017; LeBlanc, 2016, 2017).

3. Continuation of the Project

The development of *Lexos* is an integral portion of our larger Lexomics research program, which has been very productive, especially in the fields of medieval and early modern literature. New features are continuously added to *Lexos* in response to the needs of this scholarly program, and we expect development to continue as part of the ongoing Lexomics program. We have had some interest from users developing their own tools to work together to embed components of *Lexos* (particularly the scrubbing functions), and we are exploring possibilities for developing in this direction.
As part of the Lexomics program, Wheaton College will continue to provide hardware and access to a hosted installation of *Lexos* available to the public for the near future.

Strong potential for collaboration has come from our "client" project, the 4Humanities WhatEvery1Says project, which is developing a schema to document the components and relations between different parts of a Digital Humanities research workflow and support emerging publication standards for transparent data provenance and reproducible research. WhatEvery1Says recently received a Mellon Foundation Scholarly Communications grant, and we are exploring both integrating its schema into *Lexos*' workspace component and making *Lexos* a tool available within the WhatEvery1Says suite.

We have continued software development past the time of this funding. Whereas we are anticipating a move away from Python v2 to Python v3, our current build (*Lexos* v3.2) includes a refactoring effort in Python v3.6.x. In addition, our group has adopted a number of effective practices to help ensure the long-term stability and usability of our software, including unit testing, type hinting, PEP-8 documentation standards, continuous integration, and object-based design that facilitates APIs with various parts of the *Lexos* code base.

We are encouraged by our small success with generating "topic clouds" using a "topic-word-counts" file created by MALLET. Our future work will continue to seek to determine how *Lexos* can visualize data from other topic modeling programs, for example, other forms of MALLET data output.

We underestimated the videography expertise that is needed to produce quality materials. Therefore, we continue to seek best avenues for providing tools and materials that address the opacity of the procedures required to move between computational and traditional forms of text analysis, as well as foreshadow the tension between quantitative and computational approaches to text analysis. We very much seek a collection of materials that share effective practices, especially valuable for those new to the game of computational text analysis.

We agree with David Hoover (NYU) that the time for a workshop on tokenization (pre-processing) is ripe. Our present work on our pre-processing software modules (and subsequent APIs) is

positioned to play a key role if we are to teach (a) effective practices and (b) experimental methods and workflows that are reproducible.

Audiences

Lexos has two primary audiences: entry-level users and users of small to medium data sets. Our experience continues to show that entry-level users, especially those "new to the game", often become discouraged too early. Whereas undergraduate and graduate students represent the next generation of scholars, our strategy during this funding cycle was to reach as many professors as possible, as we did in the nine (9) hands-on workshops and from our own scholarly output. Users of small to medium data sets comprise the second segment of our audience, especially those using early and non-Western languages (who are by definition underserved when so much text analysis deals with modern European languages). We attempt to support historical and non-Western languages by integrating functions to support the use of those languages, such as the MUFI rule sets or tokenization features that allow the study of Chinese. The *Lexos* GitHub repository contains a test suite which contains substantial numbers of test cases using different languages.

Long-Term Impact

Lexos is now used at multiple institutions in the classroom, though we have been unable to measure the extent of its use. The development of the *Lexos* workflow was fundamental to the development of the intellectual basis for the workflow system proposed to the Mellon Foundation which led to the WhatEvery1Says grant. We expect that will have an impact beyond *Lexos* current user base.

This NEH funding has contributed to our growing reputation of an active DH@Wheaton College. Our summer research camps represent an innovative model of interdisciplinary research, akin to NSF's Research Experience for Undergraduates (REU) program. While "labs" have not been traditionally associated with the Humanities, the rate of return when practicing active collaboration in a common setting (lab or not) is even more obvious and needed today than when we started eight years ago.

While DH work is still in its infancy at state comprehensive universities, the grant helped establish a strong reputation for DH innovation at CSUN, which recently has been asked to sponsor the Digital Humanities Summer Institute and will be sending a large cohort there.

4. Grant Products

Lexomics website: **http://lexomics.wheatoncollege.edu**

Publically available Lexos website: **http://lexos.wheatoncollege.edu**

GitHub Repository (open source): **https://github.com/WheatonCS/Lexos**

1M lines of code, 53 contributors (mostly undergraduates), and a total of 5110 commits.

Dissemination Totals:

A. Summary:

Books: 1 published, 1 in progress

Articles: 10 published, 2 forthcoming, 3, in review, 4 in progress

Presentations Given: 13 talks, 9 hands-on workshops

B. Books

Drout, Michael D.C., Yvette Kisor, Leah Smith, Allison Dennett and Natasha Piirainen. *Beowulf Unlocked: New Evidence from Lexomic Methods* (New York: Palgrave Pivot, 2016).

C. Articles

[2015]

Berger, R. and Michael D.C. Drout (2015). "A Reconsideration of the Relationship Between *Víga-Glúms Saga* and *Reykdœla Saga*: New Evidence from Lexomic Analysis." *Viking and Medieval Scandinavia* 11 (2015): 1-32.

Drout, M. and Elie Chauvet (2015). "Tracking the Moving Ratio of *þ* to *ð* in Anglo-Saxon Texts: A New Method, and Evidence for a lost Old English version of the 'Song of the Three Youths.'" *Anglia* 133.2 (2015): 278-319.

LeBlanc, M.D. and Drout, M.D.C. (2015). "DNA and 普通話 (Mandarin): Bringing introductory programming to the Life Sciences and Digital Humanities. *Procedia Computer Science: International Conference on Computational Science*, 51, 1937-1946. http://www.sciencedirect.com/science/article/pii/S187705091501266

[**2016**]

LeBlanc, M.D. (2016). Computing and the Digital Humanities. An NCWIT Teaching Paper: National Center for Women & Information Technology. http://www.engage-csedu.org/sites/default/files/LeBlanc_EngageCSEdu-TeachingPaper.pdf

[**2017**]

Boese, E.S., LeBlanc, M.D., and Quinn, B.A. (2017). "EngageCSEdu: Making interdisciplinary connections to engage students."*ACM Inroads*, v8(2), 33-36. https://www.ncwit.org/sites/default/files/file_type/making_interdisciplinary_connections_to_engage_students.pdf

LeBlanc, M.D. (2017). "Toward Reproducibility in DH Experiments: A Case Study in Search of Edgar Allan Poe's First Published Work." Presented at and short paper published in Digital Humanities 2017, Montreal, Canada, August 8-10, 2017. https://dh2017.adho.org/abstracts/027/027.pdf

[**forthcoming**]

Scott Kleinman, R. Nichols, K. Nielbo, E. Slingerland, U. Bergeton, and C. Logan. "Topic Modeling Ancient Chinese Texts: Knowledge Discovery in Databases for Humanists." *Journal of Asian Studies*.

Michael D. C. Drout. "Adapting Lexomic Analyses to Old English Gloss Corpora: A general strategy and some preliminary results," in *Studies on Late Antique and Medieval Germanic Glossography and Lexicography in Honour of Patrizia Lendinara.* ed. Loredana Teresi et al. (Pisa: ETS, 2017).

[**Articles under review**]

Mary Dockray-Miller and Michael D.C. Drout with Sarah Kinkade and Jillian Valerio, "The Author and the Authors of the *Vita Ædwardi Regis*." *Speculum*.

Michael D.C. Drout and Leah Smith, "A Pebble Smoothed by Tradition: Lines 607-661 of *Beowulf* as a formulaic set-piece," *Oral Tradition*.

Michael D.C. Drout and Wenzhuo Shi. "Rolling Window Analysis of *The Dream of the Red Chamber* (紅樓夢): Some promising early results from new "Lexomic" methods." *Chinese Literature: Essays, Articles, Reviews*.

[**Articles and Books in progress**]

Michael D.C. Drout *A (Rolling) Window on the Past: Lexomic Evidence for the Compilation and Transmission of Beowulf*. (New York: Palgrave Pivot, 2017).

Michael D.C. Drout, Elizabeth Peterson, Ann Marie Brasacchio, Yun Meng, and Elizabeth Oliveira, "Lexomic Analysis of *The Two Noble Kinsmen*, *Henry VIII*, *Pericles* and Shakespeare's Other Collaborations."

Michael D.C. Drout and Chloë Urbanczyk, "The Transmission History of the Vercelli Book."

Michael D.C. Drout, Jonathan Gerkin and Scott Kleinman. "Relationships Between and Among *St. Erkenwald* and the Cotton Nero A.x. Poems: New Evidence from Lexomic Analysis."

D. Talks and Workshops

[2015]

Michael D.C. Drout. "*Beowulf*," Intensive 5-Day Seminar, Schooling for Life, Los Angeles, CA. December 28, 2014-January 2, 2015.

LeBlanc, M.D. "Exploring Digitized Texts: the Digital Humanities as Makers." Presented at Denison University, Granville, OH. September 2015.

S. Kleinman: "Text Analysis with Lexos." Workshop on Building and Strengthening Digital Humanities through a Regional Network, San Diego State University, San Diego, October 2015.

[2016]

S. Kleinman: "Digital Humanities Projects with Small and Unusual Data: Some Experiences from the Trenches," Invited lecture at the UC Irvine Data Sciences and Digital Humanities Symposium, February 2016. The transcript (http://scottkleinman.net/blog/2016/03/15/digital-humanities-projects-with-small-and-unusual-data/) was selected as an Editor's Choice by Digital Humanities Now (15 March 2016).

M. Drout, M. LeBlanc, and J. Lund (Wheaton reference librarian): "Using Research Projects to Promote Information ." Presented at The Council of Independent Colleges *Information Fluency in the Disciplines: Workshop on English and American Language and Literature*, March 10-12, 2016, Louisville, KY.

Michael D.C. Drout. "Lexomic Analysis of *Beowulf* and J.R.R. Tolkien's Scholarship on the Poem: A Confluence." Signum University, March 22, 2016.

Digital Britain: New Approaches to the Early Middle Ages, March 25, 2016, Harvard University, Cambridge, MA.

LeBlanc: "Using *Lexos* to Mine Medieval Texts: The Lexomic Workflow."

S. Kleinman: "Lexomics Evidence and the Composition of the Early Middle English *South English Legendary*."

Michael D.C. Drout: "Two Kinds of Lexomic Evidence and the Composition and Transmission of *Beowulf."*

Michael D.C. Drout. "...and then the Liberal Arts take over everything: Convergence and the growing power of our intellectual tradition," Benedictine College, Atchison, KS, March 30, 2016.

S. Kleinman: "Sex, Space, and Sanctuary: Translating Freedom in Laȝamon's *Brut* and the *South English Legendary*'s Life of St Mary of Egypt," Presented at the 51st International Congress on Medieval Studies, Kalamazoo, May 2016.

Michael D.C. Drout. "Linguistics and Lexomics," Weekend Seminar, Schooling for Life, Los Angeles, CA. May 13-15, 2016.

Michael D.C. Drout. "Lexomic Approaches: An Introduction," Johnson C. Smith University (JCSU)'s Faculty Summer Institute for Technology & New Media. Charlotte, NC. June 6-10, 2016.

[2017]

Kleinman, S. "Reducing Barriers to Participation in Automated Text Analysis in the Humanities MLA convention, Philadelphia, PA (January 2017).

Kleinman led a hands-on workshop "Text Analysis with *Lexos*" at UCLA Digital Humanities Infrastructure Symposium, 23 February 2017.

LeBlanc led a hands-on workshop using *Lexos* at **NeMLA 2017**: "Using *Lexos* to Explore your Digital Texts" (Baltimore, MD, March 23, 2017).

LeBlanc led a hands-on workshop using *Lexos* at **CCSCNE 2017**: "Bringing Computational Thinking to the Digital Humanities: Introducing Students to Explorations of Digitized Texts" at the College of Saint Rose, Albany, NY, April 7, 2017.

Mike Drout, Mark LeBlanc, and Kate Boylan presented "*Lexos*: Easing Entry to Computational Studies with Digitized Texts" [Boston College Libraries Coffee & Code series, March 20, 11-12:30 pm., O'Neill Library, Digital Studio, Room 205].

Scott Kleinman led a hands-on workshop "Text Analysis with *Lexos*" at the Institute of Advanced Communication, Education and Research, Kathmandu, Nepal, June 2017.

[**upcoming sessions**]

Mark LeBlanc and Kate Boylan will present "From coding to curating: a decade of building tools for close reading of digitized texts" at Bucknell University Digital Scholarship Conference, Oct. 6-8, 2017.

Mark LeBlanc will lead a hands-on workshop "Introducing Students to Explorations of Digitized Texts" at Bucknell University Digital Scholarship Conference, Oct. 6-8, 2017.

Michael D. C. Drout. "A (Rolling) Window on the Verse: Lexomic Analysis of Word-Foot Patterns in Anglo-Saxon Poetry," 53rd International Congress on Medieval Studies, May 2018, Western Michigan University, Kalamazoo, MI.

Websites

Lexomics website:	**http://lexomics.wheatoncollege.edu**
Publically available Lexos website:	**http://lexos.wheatoncollege.edu**
GitHub Repository (open source):	**https://github.com/WheatonCS/Lexos**