Research Report: Word Vectors for the Thoughtful Humanist

Sarah Connell and Julia Flanders
Northeastern University
December 2022

https://wwp.northeastern.edu/research/publications/reports/neh_2018/wvth_research_report.pdf

Background and Context

This white paper shares insights from the Women Writers Project's "Word Vectors for the Thoughtful Humanist" series, generously supported by the NEH's "Institutes for Advanced Topics in the Digital Humanities" program. In this project, the WWP team developed and taught a series of four workshops on teaching and research with word vector models at both introductory and intensive levels. The project team also published learning guides, code samples, case studies, and other materials that—along with a web-based interface for experimenting with trained models—provide a comprehensive suite of resources for both understanding and working with these methods.

Use of word embedding models in the humanities is still comparatively recent, and the mathematical concepts they rely on are complex and unfamiliar. Furthermore, the tools for working with these models are currently limited to the command line. To perform these analyses, a scholar currently must assemble a corpus of texts (of at least a few million words), use one of the existing suites of algorithms (such as Word2Vec or GloVe) to create a trained "model" (essentially, a very large block of data which represents the relationships between the individual words in the corpus as mathematical vectors), and then query that data concerning specific word relationships and clusters. While scholars such as Benjamin Schmidt, Ryan Heuser, and the researchers at the Stanford Literary Lab have posted useful software packages and detailed instructions to assist scholars in this work, nonetheless there are significant barriers to usage, particularly for scholars who are not already familiar with the command line. All of these challenges are magnified for teachers seeking to introduce these approaches in the (digital) humanities or social science classroom, where comprehensibility and ease of use are crucial to success.

When the original research paper on representing word relationships using vector space models was published by a team of researchers from Google in 2013,[1] it was hailed as a breakthrough in "deep-learning software designed to understand the relationships between words with no human guidance,"[2] with implications for natural language understanding, web searching, speech recognition, and other key areas of software development. In 2015, Benjamin Schmidt published a package of code that implemented Google's Word2Vec algorithms in R, a programming

[1] https://arxiv.org/pdf/1301.3781.pdf, with code published at https://code.google.com/archive/p/word2vec/
[2] https://gigaom.com/2013/08/16/were-on-the-cusp-of-deep-learning-for-the-masses-you-can-thank-google-later/

language increasingly familiar to digital humanists,[3] with the goal of making it easier for the digital humanities community to experiment with word-embedding models. He also published a series of reflections on the analytical value of these models,[4] and a set of compelling examples that illustrate their potential value for humanities research—a provocation taken up in turn by prominent scholars such as Ryan Heuser and others. This work showed that word vectors have value not only for helping machines infer semantics, but also for studying the conceptual and cultural frames of reference embodied in semantics: a key humanistic research domain. And as Schmidt notes in his original posting, word vectors are important to the humanities not only for their interpretive value, but also for the "methodological diversity" they offer to a digital humanities field that (he argues) is still impoverished in the fundamental algorithmic transformations it uses for the analysis of text.

Word embedding models represent connections between words as computable spatial relationships. Words which appear in similar semantic contexts are clustered in vector space so that if one could view the entire multidimensional cloud of words in a 20th-century document corpus one would expect to find "penny," "nickel," "dime," "quarter," "dollar," "coin," "buck," and so forth clustered together in one area, and "wonderful," "amazing," "awesome," "excellent," "fantastic," "superb," clustered together in another. (And, in a corpus of 18th-century texts, one would find somewhat different clusterings, reflecting the changing semantics of words like "awesome" and "fantastic.") These clusters do not reflect simple co-occurrence (although co-occurrence is part of the calculation): these are not words that appear *together,* but rather that occur *in the same kinds of contexts.* These models thus offer insight into how semantic concepts are verbalized and differentiated, and with appropriate corpus construction they can illuminate changes in those concepts across time, gender, genre, and other axes of comparison. As a diversification of the digital humanities tool set, they also provide a kind of methodological parallax: they add depth to our thinking about how the analysis of language works as a proxy for the analysis of culture.

Developing expertise with these tools helps humanities faculty and departments to build stronger teaching and research programs in the increasingly important cross-over areas between humanities and quantitative methods. For these methodological bridges to work in both directions, however, humanists need a strong understanding of tools like word embedding models and topic models, so that they can work as effective interdisciplinary partners. And to take these tools past a kind of mystified technological window-dressing that is only loosely connected to humanities questions, humanists need to learn more about how to build these methods persuasively into research agendas and publications, at the appropriate level of

[3] Especially following publication of Matthew Jockers' *Text Analysis with R for Students of Literature* (Springer, 2014).
[4] See http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html and https://github.com/bmschmidt/wordVectors/tree/master/vignettes.

abstraction. (Ben Schmidt's distinction between knowledge of "algorithms" and "transformations" is helpful here.[5])

While there exist good guides and tutorials on word vectors, at the start of this project we could not find any workshops that sought to take participants from complete novice-hood to self-sufficiency. It was clear that such workshops were much needed; for example, the Word2Vec algorithm is well documented but it relies on a complex set of code packages that may operate unpredictably. As noted above, the barriers for building a corpus, training models, and using those models in research or teaching are fairly high, with numerous technical dependencies and possibilities for confusion. Finally, we felt that these techniques would also gain from exploration embedded in discussion, to elicit reflection on what specific outcomes mean and how to articulate their significance. The lack of workshop opportunities was thus not simply an inconvenience but a real obstacle to humanists in gaining meaningful expertise.

These institutes sought to address an unmet need for well-scaffolded training opportunities in two ways. First, they offered participants a space to explore word embedding models intensively over a multi-day event through critical scrutiny and discussion of practical techniques, intellectual significance, and interpretive outcomes. After each event, participants received support and guidance in implementing these techniques in their home research and teaching environments. And second, the institutes provided two scaffolded environments for learning word embedding models. The first is a simple-to-use, open-access set of web tools hosted in the Women Writers Project Lab[6], to enable exploration and experimentation without the immediate barrier of learning command-line tools. The second is a thorough introduction to R and RStudio through commented code samples that could be adapted for use in participants' own research and teaching. The institutes also examined how to make arguments about and with text analysis data, and how to teach students to conduct such research. Similarly, these institutes explored ways of responding critically to published research that uses word vectors and similar techniques: how to assess the validity of methods, the preparation of the data, and the configuration of the tools in relation to the arguments being made.

Project Activities

This project sought to introduce humanities teachers, researchers, and students at varied levels of expertise to the set of text analysis methods and interpretive questions arising from word embedding models. More broadly, the project sought to develop an open-access curriculum and supporting materials for a series of institutes that could be offered initially as part of this award, and then repeated (by the Women Writers Project but also by others) in other contexts, broadening the dissemination and prompting increased critical consideration of these methods in a humanities and digital humanities context. Above all, we wanted to ensure that this

[5] See Ben Schmidt, "Do Digital Humanists Need to Understand Algorithms?" in *Debates in Digital Humanities,* ed. Matthew K. Gold and Lauren F. Klein (University of Minnesota Press, 2016).
[6] See http://www.wwp.northeastern.edu/wwo/lab/; a web-based Word2Vec interface is being developed with internal funding and will be launched in summer 2018.

curriculum would be accessible by learners (and teachable by teachers) with no prior experience of text analysis or programming.

To accomplish this goal, we offered four institutes in all: two focused on research and two focused on teaching applications, with two levels—introductory and intensive—for each audience. Introductory events were aimed at participants new to word vectors and interested in a conceptual orientation, and were taught using the Women Writers Vector Toolkit, a web-based interface for learning about and experimenting with word embedding models. Intensive events assumed no prior knowledge, but moved more quickly to cover both word vectors and the use of R/RStudio to train and analyze models, using a series of code "walkthroughs". The pedagogically-oriented events at both levels covered how to scope effective in-class activities, how to explain difficult concepts, how to design and assess meaningful assignments, and how to anticipate the necessary technical support. The research-oriented events focused on using word embedding models in textual interpretation and argumentation, including ways to connect the specific insights of word vectors with larger literary and cultural claims, constructing and testing a corpus in the context of an argument or avenue of textual research, and documenting and explaining results for a non-specialist audience. At all workshops, participants were encouraged to bring their own corpora, so that they could learn about word vectors using familiar materials, and so they could make genuine progress on their research and teaching goals during the events. They also had the option of working with corpora created by the WWP (including male-authored comparison corpora).

Following each institute, we offered followup check-in events for consultation with WWP staff and fellow participants as well as ongoing discussion via Slack, to help reinforce these challenging concepts and support participants in further experimentation. Participants were encouraged to share research and teaching outcomes (syllabi, assignments, blog posts, research papers) and were given the opportunity to post preliminary results and work in progress on the WWP blog.

To support these workshops, the WWP team developed a curriculum of guides, samples, glossaries, code tutorials, and other learning resources. These have now been adapted for standalone usage and are available at the Toolkit, in our GitHub repository, and on the WWP's site. The schedules for the four events contain links to all slides, lecture notes, handouts, and other learning resources:

- **July 2019:** Introductory institute focused on research uses of word vectors
- **May 2021:** Introductory institute focused on pedagogical uses of word vectors
- **July 2021:** Intensive institute focused on research uses of word vectors
- **May 2022:** Intensive institute focused on pedagogical uses of word vectors

The supporting materials we developed are aimed at a few different learning needs:

- Code walkthroughs include the code necessary to train, query, and visualize word embedding models, together with detailed comments that explain each command and how it can be modified. These were designed for both synchronous and standalone use.

- Glossaries, introductions, and explanations of key concepts provide clear and beginner-friendly reference points that can be returned to as needed while participants internalize new concepts.
- Case studies and sample teaching resources help to show applications for machine learning methods, offer examples for critique and analysis, and provide starting points for adaptation in participants' own teaching and research.
- Guides for corpus creation, evaluating research conducted with word vector models, and organizing research processes help to model best practices and support critical applications of these methods.

These materials are all released under a Creative Commons BY-NC license that encourages reuse and adaptation for non-commercial use. We actively encourage participants to use and adapt these materials in their own teaching, and we invite those who do so to share their results with us and with the WWP community via a guest blog post.

Pedagogical Design

The design for this series drew on the WWP's past experiences with teaching technical topics, primarily in the domains of text encoding and related skills. The WWP had an established model for introducing digital and technical skills, with a few key strategies, such as:

- Provide concrete examples and use analogies where appropriate to help make abstract or unfamiliar concepts more approachable. Wherever possible, connect technical explanations with things participants already know, and orient explanations around humanities research aims.
- Place a strong priority on pedagogical transparency: note where particular topics are difficult to teach and learn, acknowledge places where the instructors are still building their expertise, explain the rationale behind pedagogical design choices, and invite input from participants (who are themselves often experienced teachers) on pedagogical decisions.
- Structure workshops with ample hands-on time, beginning as early as possible. In the hands-on sections, encourage participants to work with their own materials, so that they can learn new skills with familiar materials, and draw on the motivation that comes from genuine research and teaching applications.
- Treat all concepts as worthy of explanation, even ones that might seem either very basic or very advanced; treat all questions as opportunities to connect what participants don't yet know to what they do already know, and to connect specific points of uncertainty to broader concepts and critical issues. Our strategy for answering all questions of any kind is to listen for what may be motivating the question, and try to address the broader uncertainty or gap it reveals.
- Provide extensive scaffolding in the form of lecture notes, slides, handouts/cribsheets, code samples, and templates; structure these materials for both in-class use and subsequent review. Comment all templates or code samples extensively with explanations and clear indications of where participants would modify the code.

- Make all course materials, including the workshop schedule, accessible indefinitely. We link from the workshop schedule to working versions of the slides, lecture notes, and other materials that are updated as we continue to teach, but we also keep each individual event schedule live. This ensures that participants can always access the most updated and improved versions of teaching materials, but can do so in the context in which they first learned them, as recorded in the schedules.

In teaching the institutes, we found that all of these strategies proved as helpful in teaching word vectors as they had for text encoding—in particular, the WWP's approach of building very substantial support infrastructures to scaffold learning was essential.

An example of how we implemented these teaching strategies for the word vectors institutes can be illustrated through the use of participant data during the workshops. We invited all participants to bring their own corpora, so that they could experiment with word embedding models using familiar texts, and so that they could use the institutes to explore genuine research and teaching questions. We provided guidance before the institutes on how to approach corpus design and data preparation, we worked directly with participants to help them build their corpora, and we trained models on each participant corpus and made them available with the web-based Women Writers Vector Toolkit sandbox. With this preparation in place, participants were able to begin hands-on exploration with models trained on their corpora on the first day of each institute—we used these corpora not just to demonstrate the kinds of insights that are possible with word vector models, but also to raise concrete and specific questions about choices in data preparation and model training. For participants in the "intensive" institutes, we loaded their data into RStudio Server, so that they could train a variety of models on their datasets, to gain a better understanding of how choices made in model training impacted outcomes. And, all participants were invited to attend optional sessions on downloading and installing R, RStudio, and the necessary code to train models on their own devices.

This approach accomplished several goals: during the institutes themselves, the participant models not only gave individuals the chance to learn using their own data, but also provided a collective body of knowledge, since participants were able to see each other's models and corpora. Several participants performed comparative research using models brought by others, and the collections of models and projects also gave the whole group a broad set of example applications to consider. Our discussions during the institutes raised many additional questions and opportunities for participants to continue improving their data preparation processes; we shared a "Data Preparation Guide and Checklist" to provide a post-institute roadmap for continued data work. Participant blog posts then offered more durable and public-facing descriptions of the work participants did with their corpora.

Despite the strong groundwork we were building on, there were still challenges in developing and teaching the curriculum for the institute series. In particular, we struggled with fine-tuning the level of detail in our explanations of more abstract or mathematical concepts. It was important for us that we work against the "black box" nature of machine learning methods like

word vectors, but there were some levels of detail that were genuinely out of scope for this particular audience and context. Over the course of our work on this series, we continued to improve our narratives and explanations for concepts like cosine similarities, vector space, vectors, embeddings, and so on, as we could better identify effective levels of detail and explanatory strategies. We also refined our approaches to the kinds of useful simplifications that are necessary at early stages in learning, collaborating with our colleagues in computer science to find ways for explaining very abstract concepts without leading to simplifications or distortions that would cause confusion down the line.

Another challenge was developing curricula that could reach learners with varying levels of experience in coding and machine learning. All of the sessions were taught with the understanding that no previous experience was required, but we did have many participants who had previously worked with R, Python, machine learning, or even word vectors. To address this challenge, we encouraged participants who had more experience to help support the newer learners, we used the chat in our remote sessions as a space for counter-discourses on more advanced topics, and we opened up our own pedagogical narratives for analysis in a meta-spirit, asking those who might be familiar with the material covered to think instead about *how* they were being taught, rather than on understanding any already-familiar concepts themselves.

Participant Research

One of the notable features of the institute series was the range and richness of participant research, which not only demonstrated the wide variety of investigative strategies and interpretive approaches that are possible with word embedding models, but also provided participants with a wealth of concrete examples to learn from, far more effectively than we could have done through invented scenarios. The corpora brought by participants spanned an extraordinary range of time periods, languages, disciplines, genres, and topics. A partial list can suggest the breadth of these collections:

- transcripts from policy debates in Congress from 1873 to 1896
- comments, transcripts, and interviews regarding Wikipedia deliberations
- documents from the Foreign Relations of the United States (FRUS) collection
- early Italian scientific, medical, and autobiographical essays and manuscripts
- journal articles and reviews from the International Journal for Philosophy of Religion
- transcriptions from upper-level undergraduate second language Spanish content-based courses
- novels related to school shootings published between 1977 and 2020
- British Gothic stories of the long 19th century
- 19th-century English colonial texts from the Persian Gulf

Participants also worked with corpora comprising the complete works of several authors, including William Shakespeare, Louise Erdrich, Ernest Hemingway, Charles Dickens, and Jane Austen.

While working on their own materials had clear pedagogical benefit, even for participants who were simply experimenting, several participants went on to develop more extensive versions of

their projects following the conclusion of the institute, and shared those outcomes via the WWP blog.

- Hayley Stefan created a corpus of popular fiction focused on school shootings in the USA, and her research revealed the significance of temporal structures and markers in organizing these narratives; as she notes in her blog post for the WWP, “time and dates are significant methods of structuring school shooting fiction, which…indicates that we have more to learn about how we understand temporality in regards to gun violence, childhood, distress, and collective trauma.” Her research also yielded a thoughtful analysis of the professional constraints that limit her ability to transfer her research skills with word embedding models into her classroom teaching, including departmental needs that affect course design and also the level of professional risk involved. You can find the full post at: “Struggling to Teach with Word Vectors.”
- Avraham Roos used word embedding models in his dissertation titled “Why is This Translation Different from All Other Translations? A Linguistic and Cultural-Historical Analysis of English Translations of the Passover Haggadah from 1770 to Now.” Working from a digitized collection of Passover Haggadot (in English translation), his research showed how cluster analysis of the trained model could bring to light significant patterns of language: for example, terms that describe the experiences of the Jews in Egypt, or the instructional language of the Haggadah itself. Although his corpus proved too small to support an analysis of how the translations change over time (which would have required splitting his corpus into chronological segments), he noted that this would be a fruitful direction for further work. You can find the full post at: “Exploring English Translations of the Passover Haggadah in Word2Vec.”
- Emily Miller developed a corpus of 19th-century novels for an ecocritical investigation focused on exploring the developments of language around, and attitudes towards, the natural world as industrialization and urban expansion progressed in the 19th century. Her blog post outlines a thoughtful process that serves as an example for researchers new to word embedding models, and discusses how she approached key activities like building a corpus, identifying terms to query, and analyzing results. Miller examines how discourse around terms like “gloom” as contrasted with “bright” can help expand our understandings of the ways that Victorian writers constructed their relationships with indoor and outdoor spaces, or with built and natural environments. You can find the full post at: “Are the Romantics to Blame? Exploring the WWP WordVectors Code as a Word Vectors Novice.”
- Becky Standard worked with a corpus of 19th- and 20th-century novels by women to explore descriptions of working roles and terms relating to employment and compare them based on gender. Using complex multi-word vectors to construct clusters of work-related terms, she developed a nuanced picture of the gendered language of work, including terms associated with the concept of work itself (for women: *drudgery*, *daily, industry*, *discipline*; for men, *profession*, *promotion*, and *advancement*). Examining the gendered valences of specific work terms yielded even more insight: for example, the word “factory” combined with the gender vector yielded *loom*, *spinning*, *cottage*, and *seamstress* for women, but *workmen*, *carpenter*, *city*, and *office* for men. Her study also used clustering algorithms to examine the relationships between terms for specific jobs

referenced in the corpus, revealing some provocative proximities (for instance, the fact that certain religious "jobs" (*chaplain*, *clergyman*, *priest*, *minister*) clustered together with terms for female caregiver roles (*servant*, *nurse*, *housekeeper*). You can find the full post at: "A Most Illustrious and Distinctive Career."

- James Clawson's work focused on Virginia Woolf's treatment of materialism and spiritualism and their manifestation in the descriptions of characters' inner and outer worlds. For an initial experiment, he developed two corpora comprising several hundred British novels written between 1836 and 1922: one "Victorian" corpus covering 1836–1901 and one "Georgian" corpus covering 1901–1922. Looking at descriptors that associate closely with references to servants, he compared these terms across the two corpora and discovered significant shifts in characterization that suggested an increased attention by novelists to depicting servants as human characters rather than stock figures. In a second experiment, he prepared a set of thirteen corpora representing overlapping chronological sections of the larger period, to enable him to study incremental changes over time rather than a simple before-and-after approach. He constructed complex vectors representing "materialism" and "spiritualism" (for each one, identifying a set of associated terms), and also vectors for different kinds of roles: spousal relationships, servants, parent-child relationships. He then plotted the terms that emerged at the intersection of the material/spiritual vectors and the vectors for each of these role groups, as a function of time. This complex analysis revealed some striking trends in which spiritualistic characterization (at least as represented by this analysis) seemed to reach a peak in novels around 1910. You can find the full post at: "A Word Embedding Model of One's Own: Modern Fiction From Materialism to Spiritualism."
- Caterina Agostini worked with a corpus of Italian scientific texts from the 1580s to the 1630s, including selections from Galileo's scientific works. Like Miller, Agostini shares insights into effective strategies for beginning a research project using these methods—she outlines the rationales for and impacts of the decisions as she explored word vector models as a method to analyze how scientific disciplines were standardized during her period. Agostini also discusses the particular challenges of working with a multilingual corpus (primarily in Italian, but with extensive Latin quotations), analyzing the impacts of textual heterogeneity on her results. You can find the full post at: "Explaining Words, in Nature and Science: Textual Analysis in Galileo's Works."

Lessons Learned

In the course of designing and teaching this series of institutes, the project team learned a tremendous amount about word embedding models, about effective approaches to teaching them, and also about larger design issues for institutes of this kind. As noted above, these institutes drew significantly on the WWP's well-tested approaches to digital humanities pedagogy, which proved their worth in this context as well. But there were several new insights we gained during this process that are worth describing here.

The first lesson concerns the kinds of explanatory strategies and metaphors used to support an early-stage understanding of word-embedding models: descriptions that structure the initial

accounts of how vectors are constructed, how the word embedding algorithm "reads" and processes the text, and what the model is actually modeling. In some senses, these explanations are necessarily provisional: they are not intended as rigorous or literal accounts but rather as a first, figurative approximation that helps orient the learner in the problem space and gives them some initial tools for thinking with. But their inexactness must still be deliberately managed: they can leave room for future refinement, but they should not create a misleading understanding that will later need to be corrected. Because the instructors' own understanding of these concepts was also developing during the course of the institute series, in some cases we started with an explanation that we later realized was imposing limitations on participants' understanding. For example, in the initial institutes we foregrounded the idea of a Cartesian coordinate space as a way of explaining dimensionality, based on the assumption that participants would be familiar with images of the x, y, and z axes from prior exposure to geometry or statistical diagrams. However, we came to realize that this explanatory method overliteralized the idea of vectors for individual words as being themselves the "dimensions" in question, which then made it difficult for participants to grasp what "embedding" and dimension reduction were and how they worked. For the last institute, we shifted to an explanation that introduced higher dimensionality earlier on, using a visual simplification of the observational process by which word vectors are actually constructed. This had the advantage of grounding the explanation of the *concept* of dimensionality in an understanding of the *process* of vector construction, rather than treating the concept as an abstraction, and it also gave us an initial visual metaphor that could later be refined without fundamentally altering how it worked, creating better explanatory continuity. In thinking about these explanations, we were careful to treat our own initial novice-hood as a useful index of where our participants would be coming from, and retained the metaphors (even when flawed) that helped us understand key concepts.

Another important lesson concerns the role of the "discussant" which we included in the design of all four institutes. In planning the events, we envisioned the discussants (two per event) as playing a dual role of providing extra assistance during the hands-on sessions, and of contributing experienced perspectives to the discussion sessions. Both of these roles were indeed valuable, but what we had not anticipated was how important the discussants would prove as both a "bridge" of mid-level expertise between the instructors and participants, and as practitioners with real-world projects to draw on in illustrating the consequences of specific design decisions or explanatory tactics. We also did not anticipate how effectively the discussants would function as additional voices during the more formal instruction portions of the institute; in particular, one discussant proved to have strong expertise in machine learning and became an informal additional instructor on specialized questions about the underlying mathematics. The value of the discussants was particularly striking in the virtual institutes (institutes 2–4), where the presence of a chat backchannel enabled them to address participant questions during the flow of the session, without interrupting the main presentation. Along similar lines, each institute included several participants who were already experienced with word embedding models or adjacent machine-learning and natural language processing techniques, but wanted to reinforce their skills or approach the topic from a humanities perspective. These participants proved enormously helpful in contributing explanations and answering more advanced questions which would otherwise have been out of scope for us to

address. They and the discussants served in effect as a parallel stream of instruction which reinforced rather than distracted from the planned curriculum.

This double stream of instruction was evident to some extent in the initial in-person institute, but it was prominent and transformative in the three virtual institutes, precisely because the videoconferencing technology supported the chat backchannel and participants embraced it as a way of enriching and expanding the learning process. The success of this approach also owed a great deal to the trust the instructors placed in the participants (and vice versa), that enabled such a backchannel to thrive as a space of learning without making the instructors' task more difficult or seeming to undercut their position. In their feedback on the events, participants in all three of the virtual institutes commented that they found the virtual design—surprisingly to them and to us—more effective than an in-person event, in large part because of the plurality of voices it supported and the way it enabled much fuller resolution of questions.

At the onset of the coronavirus pandemic, when it became clear that in-person events would be impossible, we initially assumed that we would need to postpone the institutes, because the idea of teaching such complex topics remotely went against all of our experiences with "hands-on" workshops on technical topics. However, we eventually decided to experiment with a virtual reimplementation and (as noted above) were very pleased with the outcomes. Several design points emerged as crucial to that success:

- we held the virtual institutes (which were originally designed as occupying three full days) as a series of five half-day events. This meant that participants had a sustained experience but without the exhaustion of spending several full days on Zoom. It also meant that participants could better accommodate real-life exigencies like meals and child care. Because the slides and lecture notes were available online, participants could also have some flexibility in skipping a session if necessary. We also planned generous breaks between sessions, and we included in our opening remarks a strong encouragement to participants to share responsibility for the success and vibrancy of the event rather than taking a more passive role. This exhortation seems to have struck a chord; participants were remarkably active in the discussion, responding both to the instructors and to other participants.
- we already had planned for multiple instructors as well as the discussants, but this choice was all the more crucial in a virtual event since it allowed the presenter to focus on the main channel of communication while the other instructor could manage discussion in the chat, make note of questions that should be addressed directly, and handle any troubleshooting of access and other technical details.
- we made all of the institute materials (including slide sets, lecture notes, code notebooks, and handouts) available online so that participants could load them locally rather than relying on our screen-sharing. This is our regular practice for all workshops, but as above, it proved especially valuable in the virtual environment where participants' modes of access might be limited by bandwidth or other constraints.

A final, minor lesson concerned our plans for supporting activities after the institute events. We had imagined (and participants seemed to agree) that it would be valuable to reconvene

informally during the course of the following months to share results and ask questions. However, participants on the whole found themselves too busy with their regular work to attend these followup events, and in many cases it also appeared that they didn't find they had the time to continue experimenting (and hence didn't feel that they needed continued support at that moment).

Future Directions

In the immediate term, the WWP team plans to focus on continuing to disseminate the outcomes from this institute series. We will add to and improve on our web resources, for example, publishing additional blogs posts from participants and sharing a set of tutorials that combine slides and lecture notes from key explanatory sessions such as: an introduction to core concepts of word embedding models, an end-to-end examination of the corpus building and model training process, and a session on evaluating research with word vector models.

To ensure that our work on this series can reach as broad an audience as possible, we are developing versions of our tutorials using Python and the Gensim implementation of word2vec. These tutorials are currently under development and will be published with our other code resources on GitHub when they are complete. The Python notebooks make use of our insights for teaching and learning word vectors from this project, but they also take advantage of some of the particular functionalities offered through the Gensim code, as well as the pedagogical capacity of Juypter Notebooks. We are adapting a version of these notebooks for a *Programming Historian* tutorial, in collaboration with Quinn Dombrowski (Stanford University), to help further broaden the reach of these learning resources. We also anticipate that we will continue sharing the outcomes of this series through future workshops, both offered at Northeastern and potentially in collaboration with other institutions, as well as through conference presentations and publishing opportunities.

In the longer term, the WWP team is eager to explore opportunities to broaden our engagements with machine learning, both because we have found that these methods are often most effective when used combinatorially, and also because we expect that the explanatory tactics developed for these institutes would be effectively transferred to other methods. In teaching the institutes, we found that participants often benefited from using word vectors in combination with other methods (for example keyword in context analysis); we also found that explaining word vectors by contextualizing them with other methods, especially topic modeling, provided a useful mechanism for helping many participants to understand the specific constraints and potentials of this method. With the framework for the "Word Vectors for the Thoughtful Humanist" series now established and tested, we hope to explore the possibilities of expanding the WWP's teaching and research using other machine-learning methods.

We encourage those with feedback, questions, or suggestions to contact us: wwp@northeastern.edu.