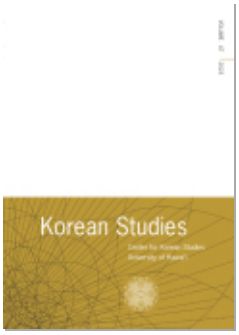
PROJECT MUSE

Epilogue: What Counts as Deep Learning in Korean Studies?

Wayne de Fremery



Korean Studies, Volume 47, 2023, pp. 300-311 (Article)

Published by University of Hawai'i Press
DOI: *https://doi.org/10.1353/ks.2023.a908626*

For additional information about this article
https://muse.jhu.edu/article/908626

[172.71.254.32] Project MUSE (2025-04-05 00:05 GMT)

Epilogue
What Counts as Deep Learning in Korean Studies?

Wayne de Fremery

What counts as deep learning in Korean studies? Certainly, what appears in this special section. How then might these articles help us to think about Korean studies and deep learning? This is a usefully tricky question. The phrase *deep learning* has become an important double entendre in our time, suggesting both artificial forms of "intelligence" and deeply engaged forms of human knowing. *What counts* is similarly plural, entailing processes associated with counting (who or what does it) and its consequences, especially who and what are made to count (i.e. matter). The meaning of *Korean studies* is as usefully amorphous as ever.

What follows is meditative rather than expository. A central hypothesis will hold my attention. It is uncomfortably simple: copies and practices related to copying are foundational infrastructure in the humanities, digital or otherwise, as practiced in Korean studies (and elsewhere). That is, as the articles in this special section demonstrate, a great deal of what we do as Koreanists and humanists concerns copying. Learning, especially the kind we call deep, is formulated through interactions with and as a function of producing copies.

A corollary to this hypothesis, one that I will take up briefly in my conclusion, is that bibliography, that old discipline which can never quite

Wayne de Fremery is a bibliographer, student of Korean poetry, and Professor of Information Science and Entrepreneurship at Dominican University of California where he directs the Françoise O. Lepage Center for Global Innovation (wayne.defremery@dominican.edu).

Korean Studies © 2023 by University of Hawai'i Press. All rights reserved.

[172.71.254.32] Project MUSE (2025-04-05 00:05 GMT)

decide if it is an art or a science, provides tools for counting and considering copies, as well as doing the generative work of copying and making people, places, and things count. Bibliography can help us to think about copies, how we count them and make them count, as well as how we use them to learn. If anything, my meditation suggests an attention to the material objects and processes that formulate some of the infrastructures that support our work as Koreanists and as humanists helps situate us in our community and among others. My hope is that this situational awareness will be useful as we collectively consider the tremendous contributions made by the authors presented in this volume, as well as the ways that we might support and extend their work.

Korean Studies

Benedict Anderson has made the case that nations can, at least in part, be understood as opportunities for individuals to imagine themselves as part of a community.[1] He identifies a material mechanism that facilitates this kind of imaginative process: print capitalism, especially the production of newspapers. Implicit in Anderson's analysis is the idea that engagements with copies created with fidelity at regular intervals and at industrial scale can enable individuals to collectively imagine national communities. Korean studies, I've come to think, can be understood in a similar way, as an imagined community. Rather than daily newspapers, copies of journals like this one allow us to image a community of people who share an interest in the contested ideas and geographies that formulate and are formulated by *Korea*.

Similarly, and perhaps more pertinent to this special section of *Korean studies*, Korean studies is supported and shaped, I would suggest, by shared practices of copying and considering copies. The deep learning displayed by the essays here is a prime example. Despite their disciplinary diversity—a diversity not dissimilar to the eclectic news items of Anderson's community building mechanism—the research in each is premised on and supported by the collection, creation, and consideration of digital representations of historical phenomena: i.e. digital copies. The digital materiality of these copies and their similarity to the phenomena they copy facilitate the arguments about Korea. This obvious fact helps to make plain how copies serve as infrastructure for the kind of learning displayed by these articles. The digital copies, the specifics of their materiality, together with the creativity and insightfulness of the authors, help to formulate what

could be asserted about Korea and what we, as readers, can learn. And indeed we learn so much!

Just as copies produced with clay, bamboo, stone, or paper have powerfully shaped (and continue to shape) what can be formulated as knowledge, digital copies now powerfully contribute to formulating what can be known and learned. This special section is thus a marker in evolving knowledge practices as they pertain to the study of Korea, one that relies predominately (but not entirely!) on digital inscription, transcription, and transformation. It is a marker of how Korean studies as a community is constituted and is changing in our historical moment, as the kinds of copies with which we work change.

"The" Humanities

Considering copies can also help us to consider the plurality of our community and the ways it intersects with others, including humanist communities, however we might conceive of them. Copies and copying have long played a central role in formulating what and how we as humanists learn. Whether we are considering the formulation of the Confucian classics in Han China, the revival of Classical learning in Renaissance Europe, or the revival and reformulation of Confucianism in Song China and Chosŏn Korea—to say nothing of how these revivals and reformations have informed the various literary, historical, and socio-political revolutions we are currently living through—the assessment and production of reproductions have been central. We know Confucius and Aristotle as we do because what they said or wrote, or what we imagine they said or wrote, has been copied and recopied. The depth of our knowledge of them is a function of how many different copies of each we have encountered and how we have become intimate with each.

Since graduate school, a great deal of my research has concerned early twentieth-century Korean poetry. If I can claim any expertise, any "deep learning" in the subject, it will be derived largely from the hours and years I've spent exploring, considering, and comparing a diverse variety of copies of Korean poems. My interest in poets and poiesis shapes the ways that I have become close with the many copies I have studied. The historians, literary specialist, linguists, and curators that have contributed to this volume are diversely intimate with what represents the objects of their interest—Barbara Wall with varied digital copies of *The Journey to the West*,

[172.71.254.32] Project MUSE (2025-04-05 00:05 GMT)

Hyeok Hweon Kang and Michelle Suh with digital representations of the *Chosŏn wangjo sillok*, Sol Jung with digital copies of four sixteenth-century Japanese diaries, Jamie Jungmin Yoo, Kiho Sung, and Changhee Lee with digital simulacra of Du Fu's poetry, Jing Hu with digital data associated with the *Collected Works of Kang Wi*, Shoufu Yin with digital representations of varied anthologies from fourteenth-to-seventeenth century China and Chosŏn Korea, Benoit Berthelier with digital copies of diverse publications organized into large North and South Korean corpora, Jacob Reidhead with digitized versions of South Korean newspapers, Liora Sarfati and Guy Shababo with digital media concerning the Sewŏl disaster, and Javier Cha with the enormity of what is being produced as a result of our newfound ability to copy using digital technologies. This intimacy with digital copies and the production of copies as unimaginably big data, large-scale digital corpora, or individually produced digital transcriptions and translations, grounds the learning on display and articulates its intersections with diverse disciplinary practice.

Information Science

Considering copies similarly helps us to examine how Korean studies as a diverse disciplinary field intersects with other, similarly diverse fields. Copies are, of course, central to other imagined communities of study, ones that, in the context of digital humanities, we encounter and inhabit if only marginally and intermittently. As I have argued recently,[2] copies and copying are central to the sprawling community associated with information science. Claude Shannon's seminal 1948 paper, "A Mathematical Theory of Communication," from which we get his definition of information as entropy, is, for example, essentially a theory of how to copy messages. Shannon's theory of communication concerns the copying of information at a source so that it is available at a destination (see Fig. 1). Importantly, Shannon acknowledges that what is being copied will not be produced with anything approaching exact fidelity. "Since, ordinarily, channels have a certain amount of noise, and therefore a finite capacity, exact transmission is impossible. This, however, evades the real issue. Practically, we are not interested in exact transmission . . . , but only in transmission to within a certain tolerance."[3] As copies, no two messages can be exactly the same when transposed in space or time. The real issue, Shannon asserts, is to create a reproduction of a message that is sufficiently similar to its source for a given purpose.

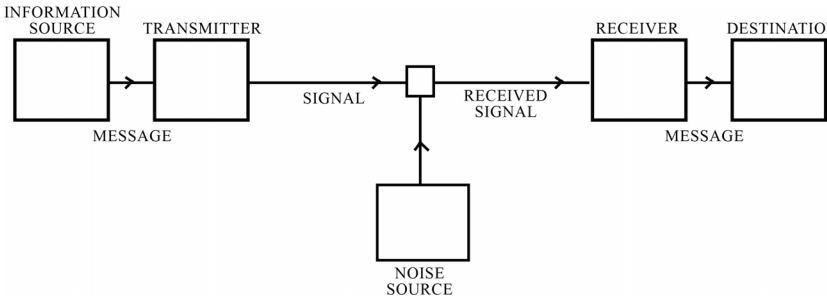


Fig. 1. "Schematic diagram of a general communication system," redrawn by author based on Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* (July, 1948): 381, accessed 3 January 2020, https://archive.org/details/bellsystemtechni27amerrich/page/n9

For Shannon and for many who identify as members of information science as an intellectual community, meaning and purpose are situated in the means of communication, in the engineering problems associated with copying and preserving copies that enable and formulate human knowledge. This instead of attempting to establish the significance of particular copies. Tefko Saracevic, for example, has written, "The domain of information science is the transmission of the universe of human knowledge in recorded form, centering on manipulation (representation, organization, and retrieval) of information, rather than knowing information."[4]

To acknowledge Shannon, Saracovic, and information science more broadly is, of course, to acknowledge the powerful role played by information scientists (loosely defined) in formulating the kinds of copies that serve as the basis for research presented in this special section. It was Shannon, for example, with his binary definition of information, that helped to enable the creation of digital documents. Others working in the academy, government, and industry created the document formats of the digital reproductions leverage by the authors in this volume. Yet others created the analytical algorithms used to recopy, manipulate, and compare the digital copies with which the authors in this volume worked as part of their efforts to hypothesize and formulate the learning with which we are presented.

Infrastructures

To recognize the ways that information theory and the practices of information scientists have shaped the research in this special section is not

[172.71.254.32] Project MUSE (2025-04-05 00:05 GMT)

to lend the knowledge produced by those in information science pride of place. For what good is a record of human knowledge if there are no humans capable of interpreting it? Rather, acknowledging the ways that information scientists have considered copies helps to reveal copies and copying as infrastructure that shapes (but does not predetermine) what has come to count as learning and knowledge about Korea.

Acknowledging the work of information scientists helps us to see how copies and copying, and increasingly digital copies and copying, are embedded in our learning practices. Embeddedness is a key element in Leigh Star's well-known framework for thinking about infrastructure.[5] Indeed, copies and processes associated with copying fit neatly into Star's framework. Copies and copying become transparent, which is to say they are difficult to see because they are so obviously important to practice in Korean studies and the humanities. Copies have reach and scope. They are formulated to be elsewhere and to suggest what they have been made to represent, i.e. what they copy. The ways that we as humanists and Koreanists attend to copies and copying are learned as part of membership in our community and have links with our evolving conventions of practice. Methods for considering and producing copies become standards of practice, standards that are fashioned incrementally using established bases of knowledge. Like other infrastructures, and as this special section demonstrates, previous forms of copying "break down" and new ones are built on top of them. Print and manuscript copies, while they function as an established base for humanities research and will always support rich and diverse modes of scholarly investigation, as infrastructure they alone cannot facilitate the kind of inquiry pursued by the authors in this volume.

Deep Learning and Artificial Intelligence

Acknowledging the infrastructural role played by copies in human learning makes it somewhat easier to see and conceptualize the infrastructural role that copies and copying play in artificial forms of "learning" and "intelligence."[6] To suggest that digital copies are integral to evolving forms of artificial intelligence and deep learning is to state the obvious. It is also to acknowledge the difficulty of seeing obvious infrastructures that support diverse work in diverse communities of practice.

The terms artificial intelligence (AI), machine learning (ML), and deep learning (DL) are often used interchangeably. In the popular press, AI can mean "almost any kind of computerized analysis or automation."[7]

But AI experts will make a distinction between "general artificial intelligence" or "strong AI," and "narrow" or "weak AI." "General AI is the Hollywood kind of AI,"[8] writes Meredith Broussard. It does not exist yet. It is "anything to do with sentient robots (who may or may not want to take over the world), consciousness inside computers, eternal life, or machines that 'think' like humans."[9] Narrow AI does exist. But it is not nearly so "human-like" or "intelligent." It is simply a "mathematical method for prediction"[10] that is enabled, yes, by some fancy copying. In the brief discussion that follows, *artificial intelligence* will refer to "narrow" or "weak" AI.

The relationship between AI and ML has been an intimate one since AI was formulated as its own field. As data scientists John Kelleher and Brendan Tierney suggest, the term "machine learning" was being used to "describe programs that gave a computer the ability to learn from data"[11] in the early stages of artificial intelligence's development. Machines "learn" by comparing data and keeping a record of their comparisons. A machine has "learned" something when it has created a record of comparisons that usefully describes the relationship between categories of data. These descriptions are often called "models," which is what I mean by the term when I use it. In the more formal language of Kelleher and his colleagues, "Machine learning algorithms automate the process of learning a model that captures the relationship between the descriptive features and the target feature in a dataset."[12]

These "descriptive" and "target" features can be anything. For some recent work for the National Library of Korea I did with my colleague, Kim Sanghun, that concerned the development of deep learning models that could be used to help automate the transcription of images of the library's rare periodical holdings,[13] we created three sets of descriptive and target features using forms of deep learning associated with convolutional neural networks. The first set was associated with identifying specific regions in images of periodicals held by the library, colophons on colophon pages, in our case, because we were tasked with developing more robust descriptive metadata for the periodicals in addition to simply transcribing the periodicals. The second set was associated with identifying meaningful elements in the colophons, *han'gŭl* glyphs or *hancha* as opposed to smudges, for example. We also created descriptive and target features for categorizing the meaningful elements of the broader bibliographical systems embodied by the libraries periodicals. That is, we created descriptive and target features for individual *han'gŭl* syllables and Sino-Korean glyphs, as well as punctuation marks, so that an image such as the following could be associated with an appropriate target: 印. The

[172.71.254.32] Project MUSE (2025-04-05 00:05 GMT)

obvious but none-trivial point to be made is that the deep learning models we built were built using digital copies (digital images) with the aim of producing new copies (encoded text).

Where machine learning enables computers to automatically identify (i.e. learn) patterns that map descriptive features to targets, deep learning is a specific kind of machine learning that enables machines to identify which patterns in features discern the features that have been identified. The process is called "deep learning" because each representation of a feature is recursively represented by simpler representations to identify which part of the descriptive feature is best associated with a "target." Deep learning automates the process of nesting and networking increasingly simpler representations (copies of less fidelity) inside of more complex representations in order to build complex descriptions that can be associated with particular targets. The process creates copies (rather than turtles) all the way down, copies that facilitate the predictive powers of deep learning.

Philology in a New Key with a Bibliographical Bent

Deep learning as a form of artificial intelligence is not discussed at length by the deeply learned scholars who provided articles for this special section. But I hope my brief description of some of the algorithmic processes associated with deep learning suggests how copies and copying are essential to forms of artificial intelligence called deep learning and the human modes of deep learning on display here. When copies and the process of copying are acknowledged as essential infrastructure supporting both human and artificial modes of learning, we are positioned better to consider how deeply entangled both modes of learning have been and are becoming. We are presented an opportunity to debate the mechanisms that formulate Korean studies as a community, the community's evolving relationships with others, and, of course, what counts as deep learning in Korean studies. We are situated to consider *how* the infrastructures that support deep learning in both its human and artificial forms are entangled with—and contribute to—formulating what can be known about *Korea* by counting and accounting for diverse representations of *Korea*. We can investigate how both forms of learning are likely to unequally affect the lives of those who think of themselves as Korean or live in places we associate with Korea. We are positioned to see that, just as tools and methodologies from information and computer science have facilitated the humanistic

explorations found in this special section, deep learning can also facilitate our humanistic explorations, as well as benefit from our humanistic modes of inquiry and critique. We are situated to undertake these investigations with the knowledge that our old philological tools will serve us well since they were built to help us investigate copies and the socio-mechanical process that produce them.

Indeed, philology in a "new key" with a bibliographical bent may provide a useful framework for exploring the new scholarly horizons brought into view by our evolving relationship with what represents *Korea*, i.e. what presents *Korea* again. Where philology can be thought of as the "multifaceted study of texts, languages and the phenomenon of language itself,"[14] philology in a new key suggests "procedures for investigating the 'implicate order' of human memory and its material representations."[15] The bibliographical bent entails globally diverse and historically informed recursive practices of enumeration, description, analysis, and critique[16] brought to bear on what has been cared enough about to be made available elsewhere through copies. The recursive nature of bibliography, of counting and accounting for what count as meaningful representations in the contexts provided by *Korea*, along with an engagement with how different communities have attended to received representations associated with *Korea* and the ways that they described and recopied these representations, can help us to be better informed about what and how we are learning. Analytical bibliography's attention to mechanical processes of reproduction can be leveraged to enable even deeper engagements with the mechanical processes of digital copying and the ways that they shape learning. Critical bibliography provides a rich and contentious discourse about what should be copied and how. It is a discourse that makes plain just how fraught and consequential decisions always are when they concern who or what will be made available elsewhere and to the future through reproductions. Philology in a new key with a bibliographical bent presents Koreanists, no matter the media through which they investigate *Korea*, an opportunity to learn how and more deeply.

Notes

1. See Benedict Anderson, *Imagined Communities: Reflections on the Origins and Spread of Nationalism*, 2nd ed. (New York and London: Verso, 2006).

2. See Wayne de Fremery and Michael Buckland, "Copy Theory," *Journal for the Association of Information Science and Technology* 73, no. 3 (2022): 407–418. I have also discussed some of these ideas in Wayne de Fremery, "Twenty-First-Century Pleasures: Some Notes on

[172.71.254.32] Project MUSE (2025-04-05 00:05 GMT)

Form, Media Transformations, and Korean Literary Translation," *Translation Review* 108 (2021): 78–103.

3. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal* (October, 1948): 646.

4. Tefko Saracevic, "Information Science," in *Encyclopedia of Library and Information Science*, 4th ed., ed. John McDonald and Michael Levine-Clark (Boca Raton: CRC Press, 2018), 2216.

5. Infrastructure, according to Star, is embedded and transparent. It has "reach or scope" and is "learned as part of membership." It has "links with conventions of practice." It facilitates standards but is also an "embodiment of standards." It is built on what Star calls "an installed base" and becomes "visible upon breakdown." It can be fixed in "modular increments," but "not all at once or globally." Geoffrey Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA and London, England: MIT Press, 1999), loc. 572 of 4690, Kindle, citing Susan Leigh Star and Karen Ruhleder, "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces," in *Information Systems Research* 7 (1996): 111–134.

6. Portions of this section have appeared elsewhere, such as in Wayne de Fremery, "Teaching Computers to Read Korean: Big Data and Artificial Intelligence at Adan Mun'go," *Muncha wa Sasang* 3 (2018): 107–115. Portions of this passage were also presented as part of a paper presented at the "New Perspectives on the History of Books and Reading in Korea" conference, Harvard University, December 8, 2022.

7. Mariya Yao, Adelyn Zhou, and Marlene Jia, *Applied Artificial Intelligence: A Handbook for Business Leaders* (NP, Topbots, 2018), 8.

8. Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (Cambridge, Massachusetts and London, England: The MIT Press, 2018), loc. 604 of 4633. Kindle.

9. Meredith Broussard, *Artificial Unintelligence*, locs. 606–607 of 4633. Kindle.

10. Ibid.

11. John D. Kelleher and Brendan Tierney, *Data Science* (Cambridge, MA and London, England: MIT Press, 2018), 14. Kindle.

12. John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies* (Cambridge, Massachusetts and London, England: MIT Press, 2015), locs. 475–476 of 13053. Kindle.

13. Wayne de Fremery et al., Han'gukhyŏng ingong chinŭng kwanghak muncha insik (AI OCR) palchŏn ŭihan yŏn'gu 한국형 인공지능 광학적문자인식 (AI OCR) 발전을 위한 연구 (Toward the development of a Korean AI optical character recognition system) (Seoul: National Library of Korea, 2021).

14. James Turner, *Philology: The Forgotten Origins of the Modern Humanities* (Princeton: Princeton University Press, 2014), loc. 115 of 19247, Kindle.

15. Jerome McGann, *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction* (Cambridge, MA and London, England: Harvard University Press, 2014), 3. Kindle

16. For a description of these key elements of bibliography see Wayne de Fremery, *Cats, Carpenters, and Accountants: Bibliographical Foundations of Information Science*, forthcoming in 2024 from MIT Press.

References Cited

Anderson, Benedict. *Imagined Communities: Reflections on the Origins and Spread of Nationalism*, 2nd ed. New York and London: Verso, 2006.

Bowker, Geoffrey, and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA and London, England: MIT Press, 1999. Kindle.

Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge, Massachusetts and London, England: MIT Press, 2018. Kindle.

de Fremery, Wayne. "Teaching Computers to Read Korean: Big Data and Artificial Intelligence at Adan Mun'go." *Muncha wa Sasang* 3 (2018): 107–15.

de Fremery, Wayne. "Twenty-First-Century Pleasures: Some Notes on Form, Media Transformations, and Korean Literary Translation." *Translation Review* 108 (2021): 78–103.

de Fremery, Wayne. *Cats*, *Carpenters, and Accountants: Bibliographical Foundations of Information Science*, forthcoming in 2024 from MIT Press.

de Fremery, Wayne, and Michael Buckland. "Copy Theory." *Journal for the Association of Information Science and Technology* 73, no. 3 (2022): 407–418.

de Fremery, Wayne, et al., *Han'gukhyŏng ingong chinŭng kwanghak muncha insik (AI OCR) palchŏn ŭihan yŏn'gu* 한국형 인공지능 광학적문자인식 (AI OCR) 발전을 위한 연구 (Toward the development of a Korean AI optical character recognition system). Seoul: National Library of Korea, 2021.

Kelleher, John D., and Brendan Tierney. *Data Science.* Cambridge, MA and London, England: MIT Press, 2018. Kindle.

Kelleher, John D., Brian Mac Namee, and Aoife D'Arcy. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies.* Cambridge, Massachusetts and London, England: MIT Press, 2015. Kindle.

McGann, Jerome. *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, MA and London, England: Harvard University Press, 2014. Kindle.

Saracevic, Tefko. "Information Science." In *Encyclopedia of Library and Information Science*, 4th ed., edited by John McDonald and Michael Levine-Clark, 3105–3116. Boca Raton: CRC Press, 2018.

Shannon, Claude E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* (July and October, 1948): 379–423, 623–656, https://archive.org/details/bellsystemtechni27amerrich/page/n9

Star, Susan Leigh, and Karen Ruhleder. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7 (1996): 111–34.

Turner, James. Philology: *The Forgotten Origins of the Modern Humanities*. Princeton: Princeton University Press, 2014. Kindle.

Yao, Mariya, Adelyn Zhou, and Marlene Jia. *Applied Artificial Intelligence: A Handbook for Business Leaders*. NP, Topbots, 2018.