Conceptual History Using Digital Methods

Harald Kümmerle, History of Science / Digital Humanities

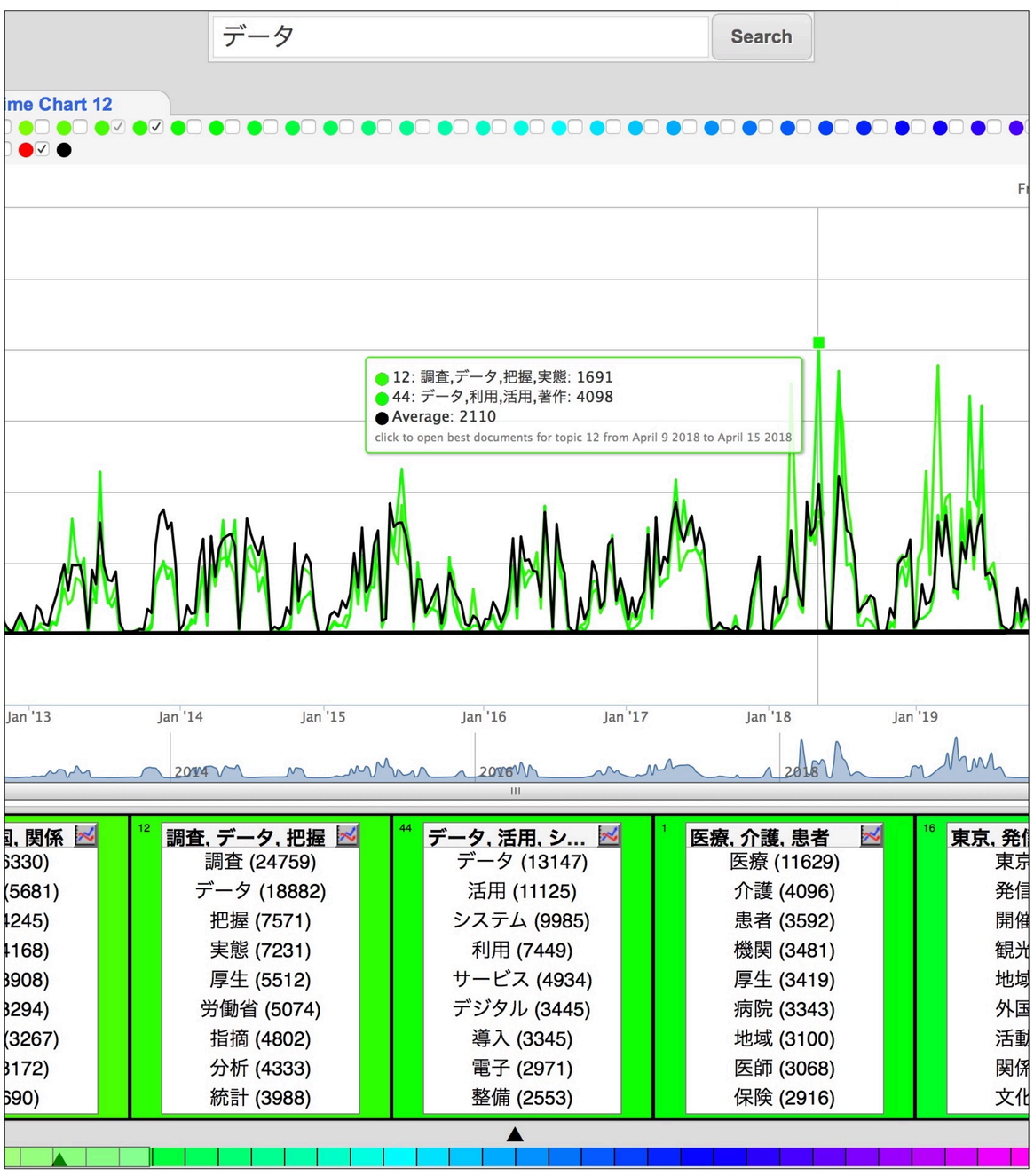
Computer-aided text mining facilitates the systematic examination of large amounts of texts. This is a very promising method for the history of concepts, since it can eliminate the need to limit oneself to a small number of key texts. If the functionality of the algorithms and the corresponding limitations are disclosed, a connection to more classical methods of textual analysis enables a combination of the advantages of quantitative and qualitative research.

LDA topic models

Since the 2000s, the method of modeling topics in a text corpus using so-called Latent Dirichlet Allocation (LDA) has emerged as an alternative to older algorithms such as dictionary-based classification and cluster recognition. The topic models that emerge are characterized by the property that they inherently take into account the ambiguity of words (Blei et al., 2003).

Case study: the concept of data

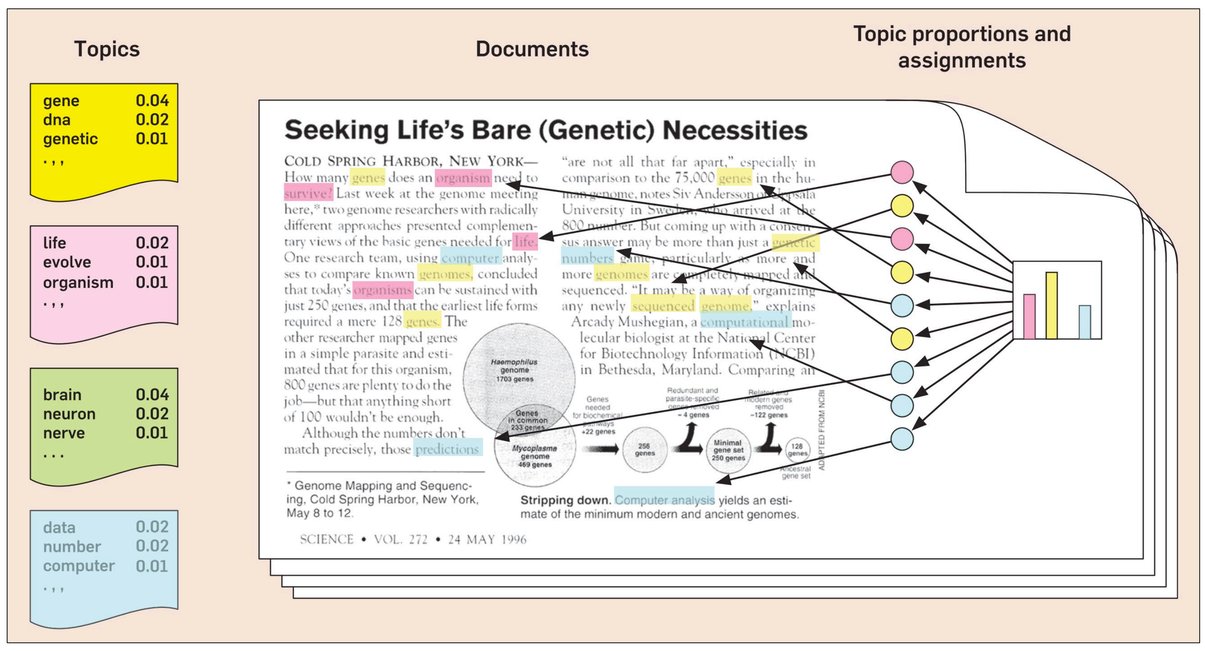
This project also provides the methodology for a study of the digital transformation in Japan (e.g. Kümmerle, 2022) based on the concept of data. The main corpus consists of over 160.000 speech contributions in the Japanese National Diet from 2011 to 2021 that contain the word *dēta* (the common translation of "data" into Japanese) or other words closely connected to its reception history, e.g. *shiryō* or *yoken*. With the right parameters, the topic modeling algorithm succeeds in identifying the two distinct meanings of *dēta* that correspond with the two different meanings in dictionaries: (1) facts for devising an argument and (2) encoded material for computation. This provides the basis for a systematic analysis.



Two meanings of *dēta* are identified and tracked over time

Operationalization

A direct evaluation of the topic models can be understood as *distant reading* of the analyzed text corpus in the sense of Franco Moretti (2013). Usually, this is supplemented with a *close reading* of individual texts from the corpus. By systematically linking these steps with the Grounded Theory Method (Baumer et al., 2017; Nelson, 2020), the research project connects the analysis using LDA topic models to the history of concepts. While Reinhart Koselleck (1972) has developed conceptual history for the study of social history, Hans-Jörg Rheinberger (2008) contributed to its adoption in the history of science placing an emphasis on hybridity. The situation for "data" is similar.



Topic models can handle hybrid concepts (Blei, 2012)

Layered traditions

Here, both the specifics of text genres and the diversity of the metadata which can guide the creation of categories must be considered. Working with Japanese-language text corpora is particularly informative not only because of the logogram character of the Chinese characters. It is also significant that many concepts were newly created during the reception of knowledge from China and Europe or were influenced hereby, but then developed a substantial momentum of their own, often with connotations that are in part independent from their origins. The concept of information is instructive: In Japan, the word 情報 (*jōhō*) was coined in the 19th century in context of intelligence-gathering by the military and the state. Taken over into Chinese and read as *qíngbào*, this connotation is still dominant in China. In Japan, however, the word became neutral during the postwar democracy (Nakamoto, 2002). Japan's attempts to establish a system of "information banks" (*jōhō ginkō*) for sharing private data (Kümmerle, 2023) has to be understood against this background and creates fruitful tensions with the concept of data.

Literature

Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, *68*(6), 1397–1410.

Blei, D. M. (2012). Probabilistic Topic Models. *Commun. ACM*, *55*(4), 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Koselleck, R. (1972). Begriffsgeschichte und Sozialgeschichte. In P. C. Ludz (Ed.), *Soziologie und Sozialgeschichte: Aspekte und Probleme* (pp. 116–131). Opladen: Westdt. Verl.

Kümmerle, H. (2022). Japanese data strategies, global surveillance capitalism, and the "LINE problem." *Matter: Journal of New Materialist Research*, *3*(1), 134–159.

Kümmerle, H. (2023). More Than a Certification Scheme: Information Banks in Japan Under Changing Norms of Data Usage. In A. Khare & W. W. Baber (Eds.), *Adopting and Adapting Innovation in Japan's Digital Transformation*. Singapore: Springer Nature.

Moretti, F. (2013). *Distant Reading*. London: Verso.

Nakamoto H. (2002). Yōgo "jōhō"—Tāminojorīteki kōsatsu -. *Jōhō no kagaku to gijutsu*, *52*(6), 339–342.

Nelson, L. K. (2020). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*, *49*(1), 3–42.

Rheinberger, H.-J. (2008). Begriffsgeschichte epistemischer Objekte. In E. Müller & F. Schmieder (Eds.), *Begriffsgeschichte der Naturwissenschaften: Zur historischen und kulturellen Dimension naturwissenschaftlicher Konzepte* (pp. 1–9). Berlin: de Gruyter.

An Institute of the
Max Weber Foundation

Deutsches Institut für Japanstudien
German Institute for Japanese Studies
ドイツ日本研究所

More about this project