

Innovators of the Past: Modelling Novelty and Resonance in Dutch Historical Language Records

Lassche, Alie

a.w.lassche@hum.leidenuniv.nl
Leiden University, Netherlands, The

Ros, Ruben

r.s.ros@hum.leidenuniv.nl
Leiden University, Netherlands, The

Introduction

Studies in the humanities are increasingly incorporating methods from information theory, which are based on dynamics in word usage and measures of relative entropy. We know from previous work that word usage in newspapers and chronicles is sensitive to the dynamics of socio-cultural events. (Wevers, Kostkan, and Nielbo 2021; Gao et al. 2012; Daems et al. 2019) Entropy measures have for example been used to detect differences between opposite political movements in French politics, and between distinct historical periods. (Barron et al. 2018; Galdi 2019) More specifically, several studies have applied windowed relative entropy to thematic text representations to generate signals that capture information novelty, which is the content difference from the past, and information resonance, which is the degree to which future information conforms to novelty. The methods have been successfully applied to parliamentary debates from the French Revolution, to Dutch newspapers from the second half of the 20th century, to Danish newspapers from the COVID-19 pandemic, and to early modern Dutch chronicles. (Barron et al. 2018; Wevers, Kostkan, and Nielbo 2021; Nielbo et al. 2021; Lassche, Kostkan, and Nielbo 2022)

This study builds further on these existing studies, by using methods for novelty and resonance detection to measure innovation of individual historical actors. To explore to what extent this method can be applied to corpora with different characteristics, we use two historical Dutch corpora from different periods and with different features. We investigate the novelty and resonance scores on the aggregated actor-level in our corpora to identify historical figures responsible for innovative and resonant texts. We ask who is responsible for innovation, how the actor-level distribution of novelty behaves (over time) and to what extent the ‘novelty bias’ observed in earlier studies holds on the level of actors. Comparing different corpora helps us to critically interrogate the information theoretical measures. Building on existing applications of the method and our comparison, we hope to work towards more general arguments about novelty and resonance in historical language data.

Corpora

In our study, we distinguish the *Chronicles Corpus* and the *Parliament Corpus*, of which the characteristics are included in Table 1. The *Chronicles Corpus* consists of early modern Dutch chronicles, written between 1500 and 1850. Chronicles are manuscripts written by middle-class people, in which they report on current events in their communities, and on what they consider interesting or relevant. These texts are ordered chronologically, providing both the date and a report of a certain event. They were rarely printed in the lifetime of their authors, but, despite their scribal form, they could still circulate in the localities, be read and continued by other authors, influencing future generations. (Pollmann 2016, 2011) The *Chronicles Corpus* was scanned and transcribed with the HTR tool Transkribus. (Kahle et al. 9AD) The corpus is enriched with a date label that is essential in this study in order to create time series of chronicle fragments.

Previous research in which novelty and resonance signals were computed over the *Chronicles Corpus* already showed how changes in the novelty signal coincide with times of political uncertainty. The resonance signal, however, remained flat over time, which suggested that future information did not conform to the introduced novelty. (Lassche, Kostkan, and Nielbo 2022) An investigation of actor-level novelty and resonance scores, as will be done in this paper, will provide more insight in the level of innovation of individual chroniclers.

Table 1:

Corpus	# Texts	# Authors	# Tokens
Chronicles	144	109	6.103.765
Parliament	1493	410	18.050.303

The Parliament Corpus consists of proceedings of the Dutch Lower House from the period between 1945 and 1963. The proceedings contain speeches delivered in parliament by members of parliament, the chair, and government ministers. They record the day-to-day practices of legislation, the filing of motions, and question answering. Proceedings are not verbatim reports, but edited transcriptions of the speeches. The digitized proceedings are tokenized and lemmatized and contain metadata on speaker party affiliations and debate topics. The corpus consists of all sessions held between 1945 and 1963. In this period, parliament would meet several days a week, with the exception of the summer months and campaign periods. From the perspective of parliamentary language, actor-level novelty and resonance appear as promising metrics to evaluate the standard depiction of the period as one during which parliament was ‘docile’ and largely overshadowed by the expanding administration. (Hemerijck 1994; Daalder and Berg 1990) Especially the differences in novelty and resonance between government members and representatives of coalition and opposition parties would be an innovative way of looking at postwar parliamentary history.

Methods

We use Top2Vec as a means to generate document (chronicle or speech) representations. (Angelov 2020) Top2Vec learns joint word and document representation and uses document cluster centroids and word embeddings to generate topics. The cosine similarity between a document and the identified cluster centroids can

be used to generate document-topic distributions that serve as the basis for novelty detection.

With novelty, we refer to an event's $S(j)$ reliable difference from past events $S(j-1)$, $S(j-2)$, $S(j-n)$ in window w and resonance as the degree to which future events $S(j+1)$, $S(j+2)$, $S(j+n)$ conform to an event's $S(j)$ novelty. Where T is the transience of $S(j)$. This model for novelty and resonance was originally proposed by Barron et al., but here we use the symmetrized and smooth version with the Jensen-Shannon divergence.

(Unfortunately, the Digital Humanities 2023 software is not able to properly display mathematical formulae)

Preliminary Results

As an example of our approach, Figures 1 and 2 show the aggregated novelty scores on the actor level plotted against the resonance scores, both for the Chronicles Corpus and the Parliament Corpus. Figure 1 shows a decrease in novelty over time, suggesting that earlier chronicles consist of more novel content than later chronicles. Figure 2 shows the variation in novelty and resonance across parliamentary speakers, with additional information on parliamentary roles indicated by colours and information on the number of speeches represented by scatter size. The figure points to the relatively low resonance of government members compared to members of parliament and the relatively high novelty of chairs (grey). The latter observation can be explained by the institutional function of the parliament chair, who is responsible for opening and introducing debates and their topics. However, the low resonances of government members as well as high-scoring members are not so easily explained. Our paper will discuss these and other findings in more detail.

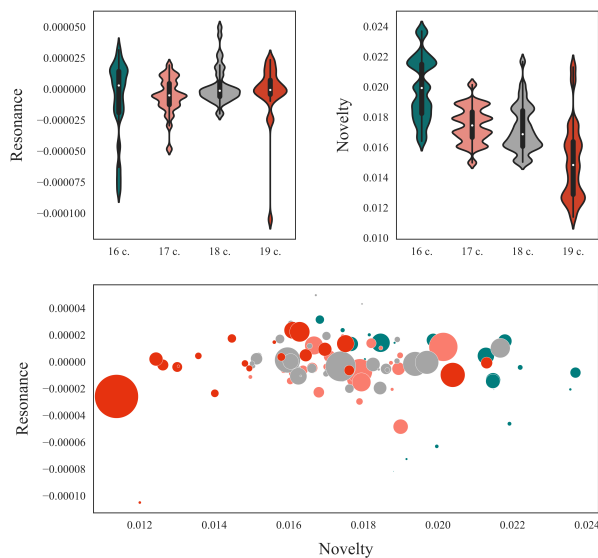


Figure 1. Novelty and Resonance in the Chronicles Corpus. Top Left: Violin plots of the resonance actor resonance distribution grouped by century. Top Right: Violin plots of the resonance actor novelty distribution grouped by century. Bottom: Scatter plot of novelty vs. resonance on the aggregate level of chronicles at scale =30. The colours correspond to the century in which the chronicle was finished, the size of the scatters refers to the size of the chronicle.

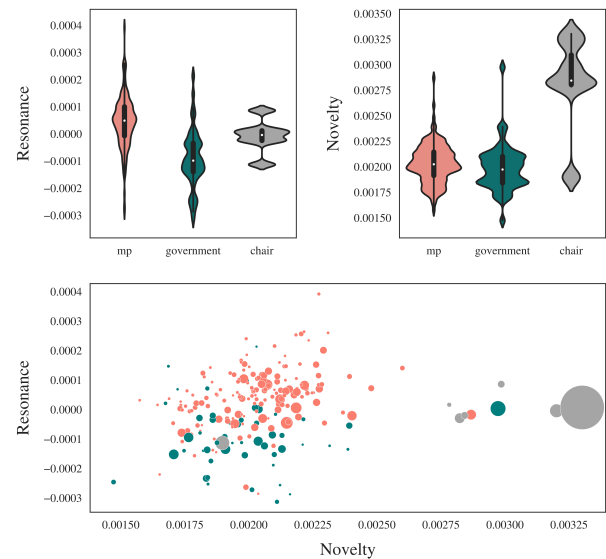


Figure 2. Novelty and Resonance in the Parliament Corpus. Top Left: Violin plots of the resonance actor resonance distribution grouped by role. Top Right: Violin plots of the resonance actor novelty distribution grouped by role. Bottom: A scatter plot of novelty vs. resonance on the aggregate level of parliamentary speakers at scale =30. The colours correspond to the role metadata tag in the parliamentary proceedings. Roles can be "government" (blue), "mp" (red), and "chair" (gray). The plot shows the relatively low resonance of government speakers compared to members of parliament (mp's) as well as the high novelty of chairs. The size of the scatters refers to the number of unique speeches by the speakers

Bibliography

- Angelov, Dimo.** 2020. "Top2Vec: Distributed Representations of Topics," nos. arXiv:2008.09470 [cs, stat] (August). <http://arxiv.org/abs/2008.09470>.
- Barron, Alexander T. J., Jenny Huang, Rebecca L. Spang, and Simon DeDeo.** 2018. "Individuals, Institutions, and Innovation in the Debates of the French Revolution." *Proceedings of the National Academy of Sciences* 115 (18): 4607–12. <https://doi.org/10.1073/pnas.1717729115>.
- Daalder, Hans, and Johannes T. van den Berg.** 1990. *Politiek en historie: opstellen over Nederlandse politiek en vergelijkende politieke wetenschap*. Amsterdam: Bakker.
- Daems, Joke, Thomas D'haeninck, Simon Hengchen, Teclé Zere, and Christophe Verbruggen.** 2019. "'Workers of the World'? A Digital Approach to Classify the International Scope of Belgian Socialist Newspapers, 1885." *Journal of European Periodical Studies* 4 (1): 99–114. <https://doi.org/10.21825/jep-s.v4i1.10187>.
- Gao, Jianbo, Jing Hu, Xiang Mao, and Matjaž Perc.** 2012. "Culturomics Meets Random Fractal Theory: Insights into Long-Range Correlations of Social and Natural Phenomena over the Past Two Centuries." *Journal of the Royal Society Interface* 9 (73): 1956–64. <https://doi.org/10.1098/rsif.2011.0846>.
- Guldi, Jo.** 2019. "The Measures of Modernity: The New Quantitative Metrics of Historical Change over Time and Their Critical Interpretation." *International Journal for History, Culture and Modernity* 7 (1): 899–939. <https://doi.org/10.18352/hcm.589>.
- Hemerijck, A. C.** 1994. "Hardnekkigheid van Corporatistisch Beleid in Nederland." *Beleid En Maatschappij* 21 (1-2): 23–47.
- Kahle, P., S. Colutto, H. Hackl, and H. Mühlberger.** 9AD. "Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents." In 2017 14th IAPR In-

ternational Conference on Document Analysis and Recognition (ICDAR), 04:19–24. <https://doi.org/10.1109/ICDAR.2017.307>.

Lassche, Alie, Jan Kostkan, and Kristoffer Nielbo. 2022. “Chronicling Crises: Event Detection in Early Modern Chronicles from the Low Countries.” In *Computational Humanities Research*. Antwerp.

Nielbo, Kristoffer L., Frida Haestrup, Kenneth C. Enevoldsen, Peter B. Vahlstrup, Rebekah B. Baglini, and Andreas Røpstorff. 2021. “When No News Is Bad News – Detection of Negative Events from News Media Content.” arXiv. <http://arxiv.org/abs/2102.06505>.

Pollmann, Judith. 2011. *Catholic Identity and the Revolt of the Netherlands, 1520-1635. The Past & Present Book Series*. Oxford; New York: Oxford University Press.

Pollmann, Judith. 2016. “Archiving the Present and Chronicling for the Future in Early Modern Europe.” *Past & Present* 230 (suppl 11): 231–52. <https://doi.org/10.1093/pastj/gtw029>.

Wevers, Melvin, Jan Kostkan, and Kristoffer Nielbo. 2021. “Event Flow - How Events Shaped the Flow of the News, 1950-1995.” In *Computational Humanities Research Conference*, 62–76. Amsterdam.