

Identifying Patterns in Academic Publications on Frisian and Limburgish Through the Use of Digital Tools

Andreas Simons¹, Daan Hovens²

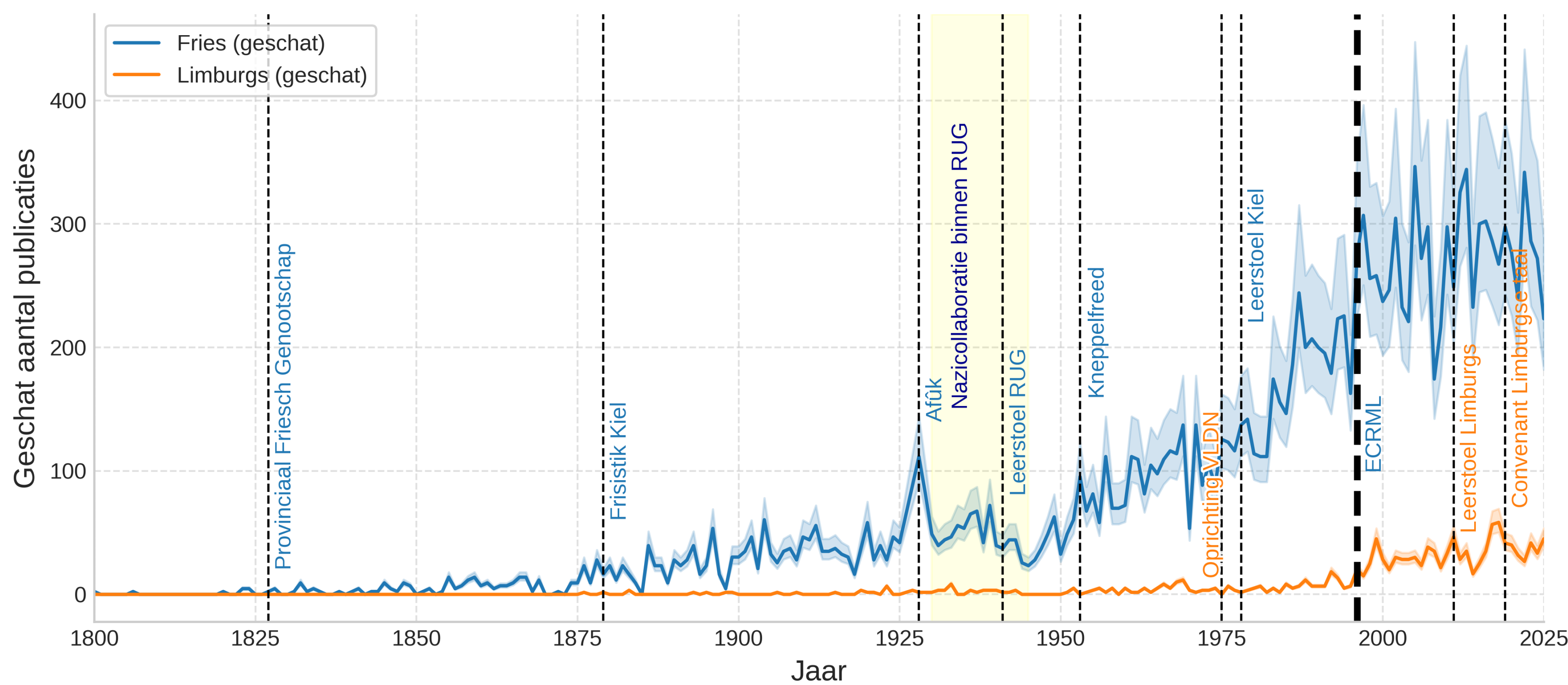
andreas.simons@uliege.be, d.hovens@maastrichtuniversity.nl

¹University of Liège, ²Maastricht University

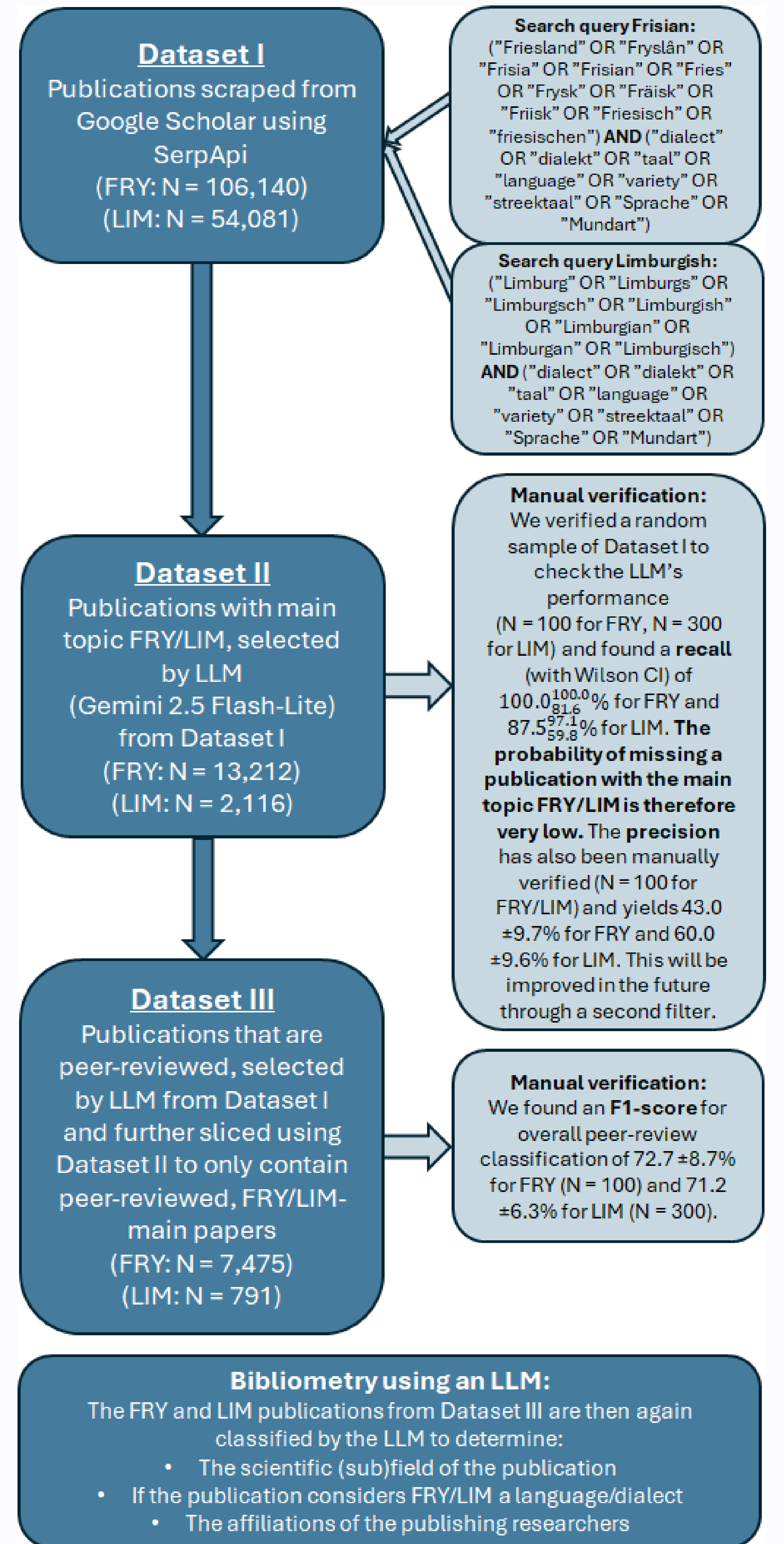
In a Nutshell

We conduct a bibliometric analysis of academic publications on Frisian (FRY) and Limburgish (LIM), two regional languages spoken in the Netherlands and Belgium/Germany. By scraping Google Scholar and using a Large Language Model (LLM), we attempt to obtain an exhaustive dataset of all (peer-reviewed) publications since 1800, from which we extract trends related to peer-review, quantity, authorship, scientific field, and language attitudes. We find that this methodology provides a cheap, fast, and reliable way of building comprehensive minority language literature reviews, as well as exposing long-term trends in research on these languages.

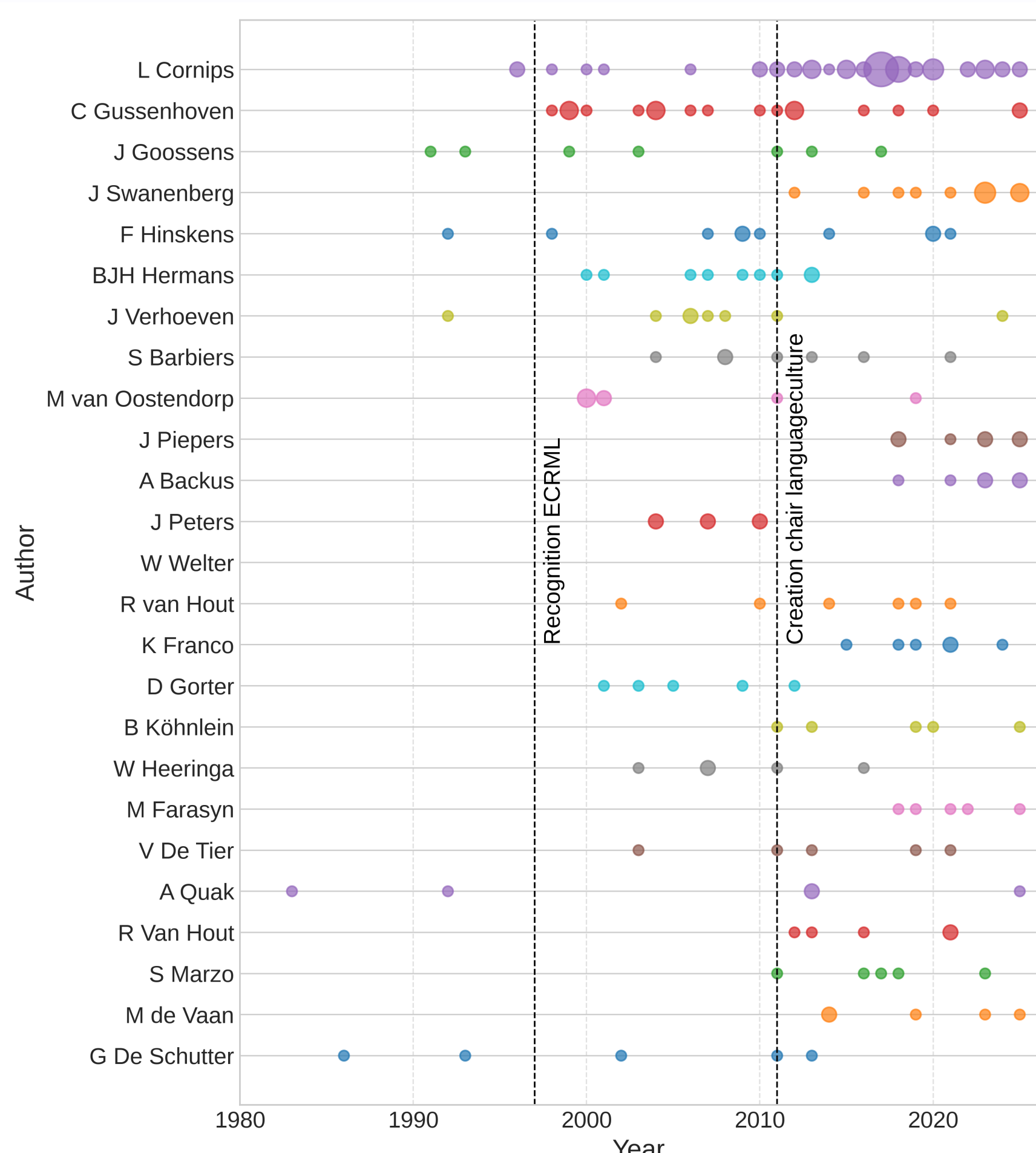
Estimated Peer-Reviewed Publications over Time (Dataset III)



Methodology



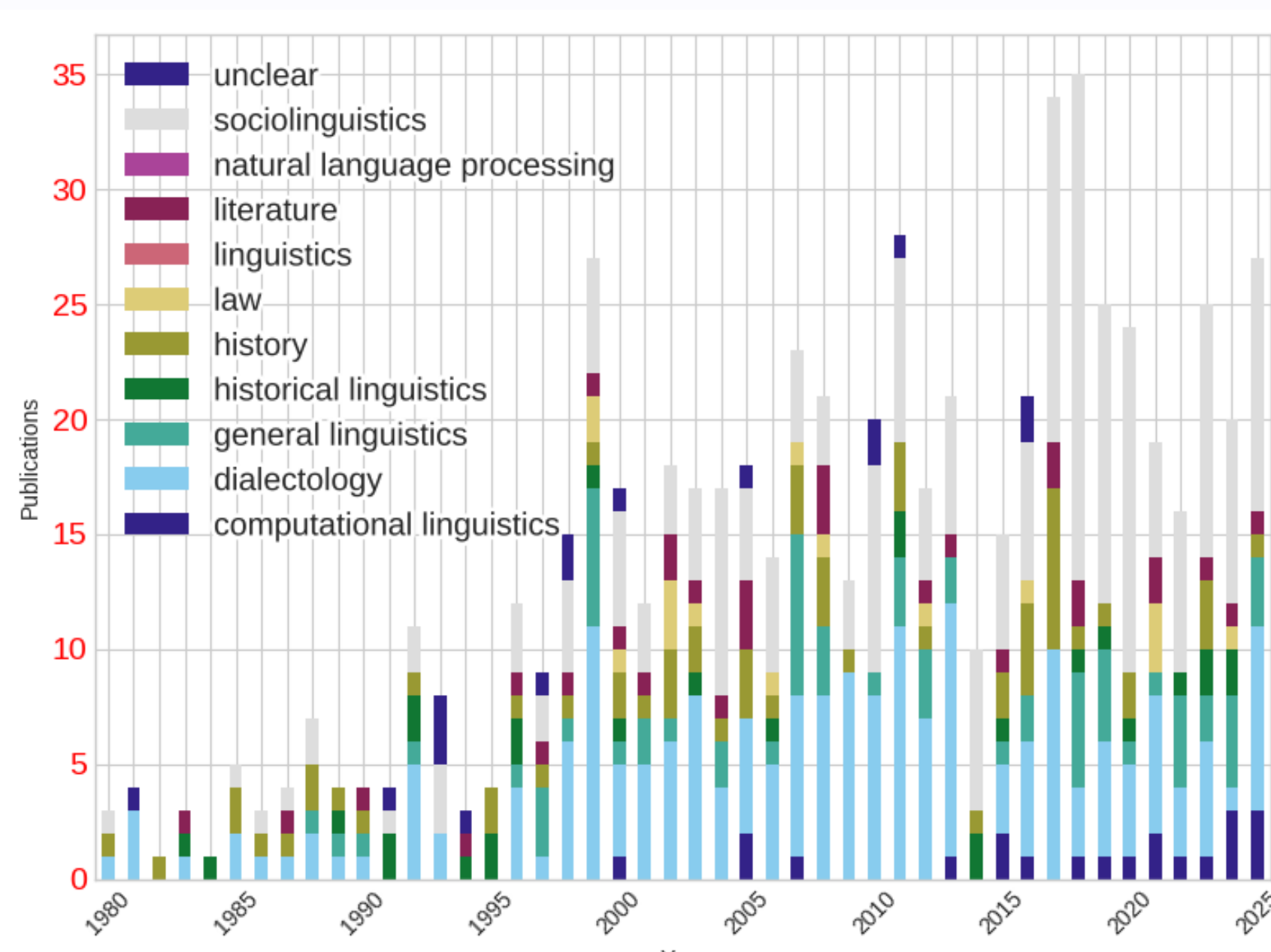
Most Prolific Authors (LIM)



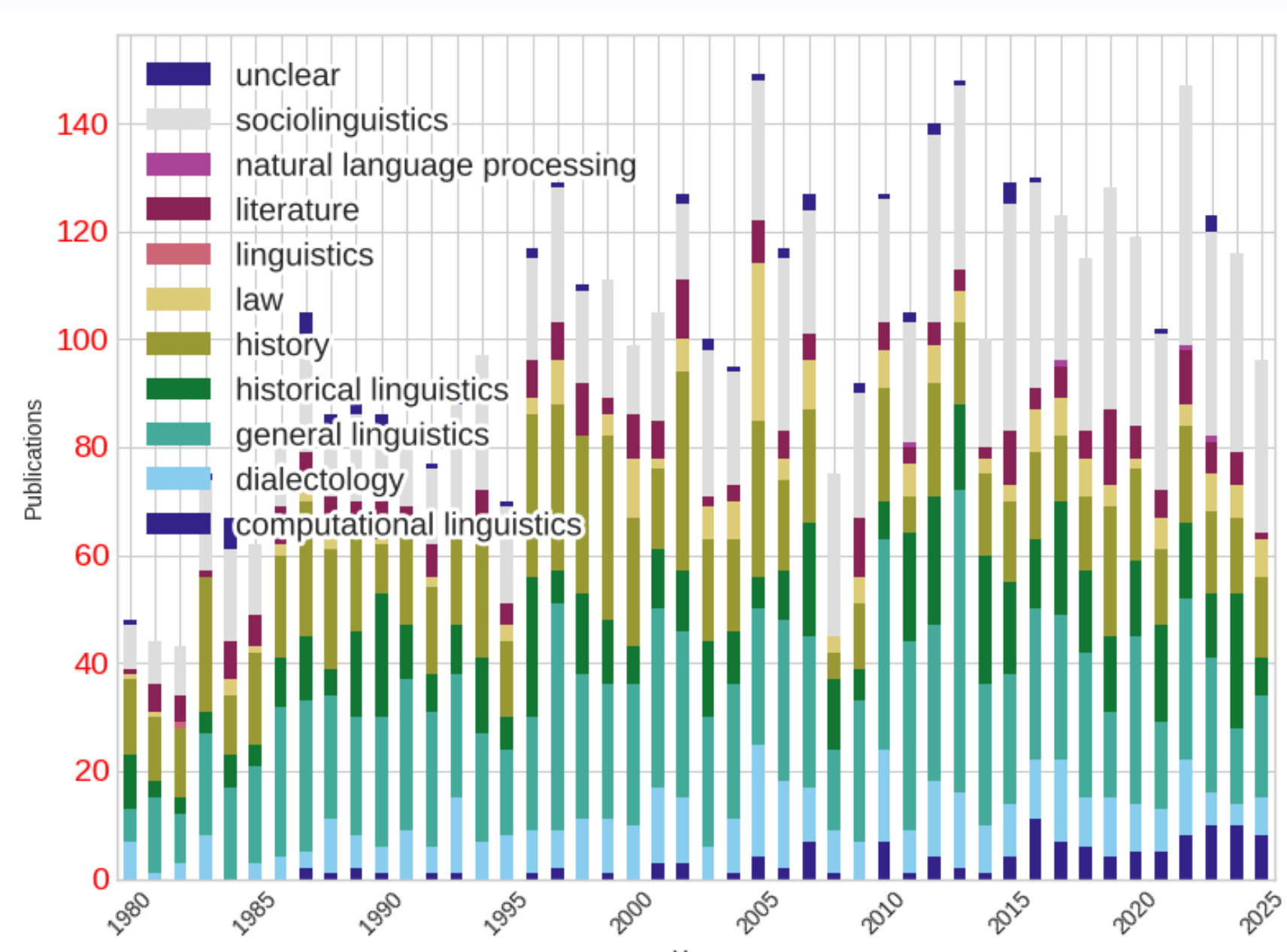
Most Prolific Authors (FRY)



Fields of Study (LIM)



Fields of Study (FRY)



Limitations & Ethics

- Google Scholar's exhaustiveness has been tested for other fields, but is unknown for linguistics.
- We did not distinguish between the Frisian dialects and e.g. East Frisian (Ostfriesisch); the term Limburg was not originally associated with Limburgish and still isn't in e.g. Germany.
- Not all publications are open access.
- Precision of main-subject classification remains low, but does not affect quantitative comparison.
- 1-5 kg CO₂ emitted as part of LLM inference for the entire project (equivalent to 10-50 km driving by car). Only publicly available metadata were provided as input to the LLM.