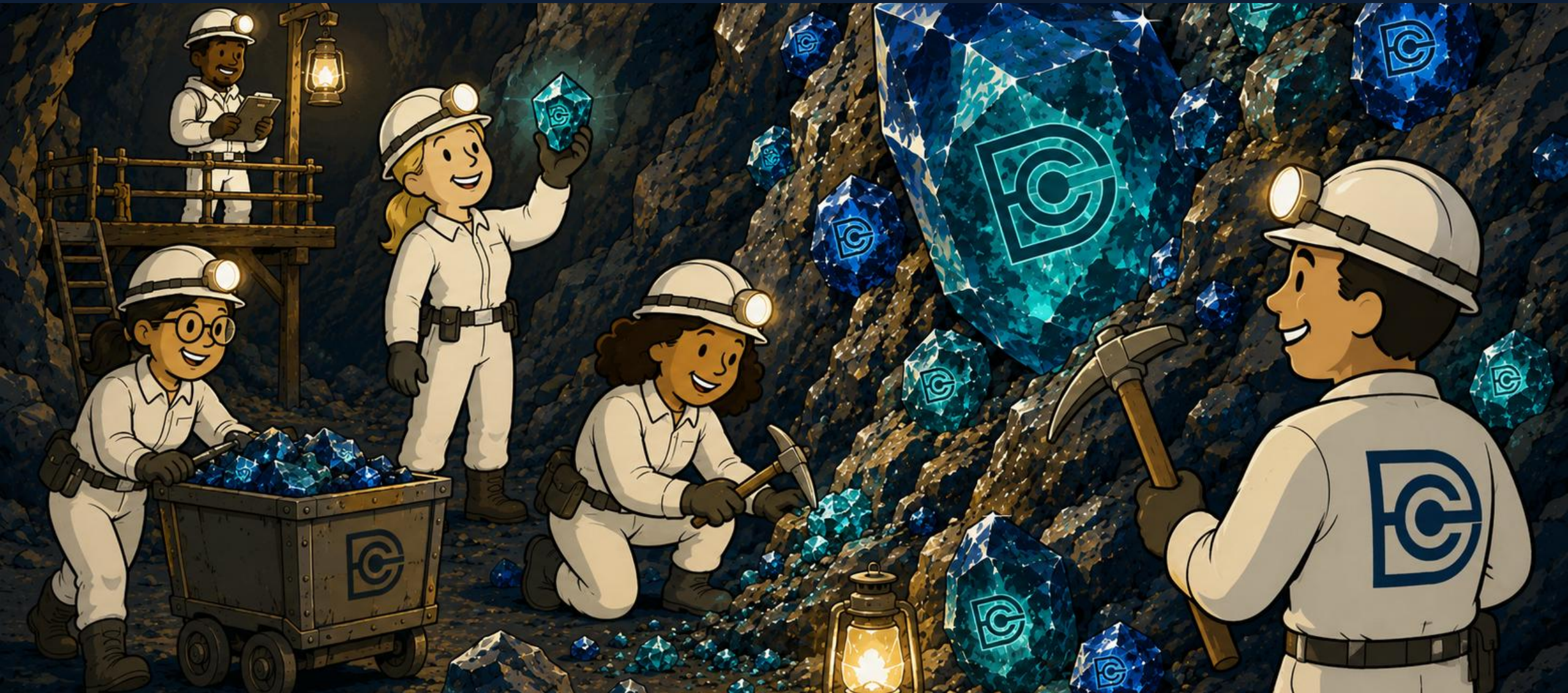


A Decade in Drama Research. Data Mining the DraCor Bibliography (2015–2025)



Luca Giovannini • Henny Sluyter-Gäthje • Daniil Skorinkin • Peer Trilcke • Ingo Börner • Frank Fischer • Tabeer Khilji
University of Potsdam and Free University of Berlin | DH2026, Daejeon, South Korea

This presentation:
plu.sh/decadrama



What is **DraCor**, and
why is it interesting to
mine its **bibliography**?

Introducing DraCor

- DraCor is a digital research infrastructure for European drama and beyond from the Antiquity to present
- Showcase for the “programmable corpora” concept (Fischer et al. 2019)
- Core idea: make dramatic texts machine-readable, comparable, and interoperable
- Currently hosting 23 corpora in 21 languages – more than 4500 plays encoded in TEI-XML
- Open data and community-driven development





DraCor – Open Infrastructure for Drama Analysis

Fre DraCor

French Drama Corpus

1,940

Number of plays

17,831

(M: 7958, F: 4247)

person + **personGrp**

Number of characters

text

17,701,440

Text tokens

509,958

(16,295,137)

sp

(Tokens)

48,755

(393,275)

stage

(Tokens)

dc9c208

Last update

14.7.2025, 06:35:29

Ger DraCor

German Drama Corpus

767

Number of plays

16,462

(M: 11632, F: 3430)

person + **personGrp**

Number of characters

text

11,969,043

Text tokens

472,952

(11,407,483)

sp

(Tokens)

225,126

(1,387,768)

stage

(Tokens)

e948b7c

Last update

19.2.2026, 08:35:23

Eng DraCor

English Drama Corpus

753

Number of plays

11,145

(M: 8564, F: 2080)

person + **personGrp**

Number of characters

text

16,448,904

Text tokens

550,693

(15,057,801)

sp

(Tokens)

97,646

(447,851)

stage

(Tokens)

dbc4fc6

Last update

1.2.2026, 22:14:19

Rus DraCor

Russian Drama Corpus

212

Number of plays

3,707

(M: 2608, F: 871)

person + **personGrp**

Number of characters

text

2,316,995

Text tokens

119,333

(2,191,657)

sp

(Tokens)

49,440

(215,112)

stage

(Tokens)

f602fad

Last update

1.2.2026, 16:39:25

Cal DraCor

Calderón Drama Corpus

205

Number of plays

3,406

(M: 1991, F: 1119)

person + **personGrp**

Number of characters

text

2,838,830

Text tokens

121,432

(2,736,034)

sp

(Tokens)

23,933

(115,495)

stage

(Tokens)

60316cb

Last update

1.2.2026, 02:54:39

ABOUT

[What Is DraCor?](#)
[Corpus Registry](#)
[Credits](#)
[Get in Touch](#)

DOCUMENTATION

[API](#)
[API v0 \(legacy\)](#)
[Encoding Guidelines \(ODD\)](#)
[FAQs](#)

TOOLS

[pydracor](#)
[rdracor](#)
[SPARQL](#)
[ezlinavis](#)

RESEARCH

[Research Bibliography](#)
[Posters](#)
[Card Games](#)
[DraCor in Science Communication](#)

The Fool Turned Critic

eng000485 1678

Eng DraCor

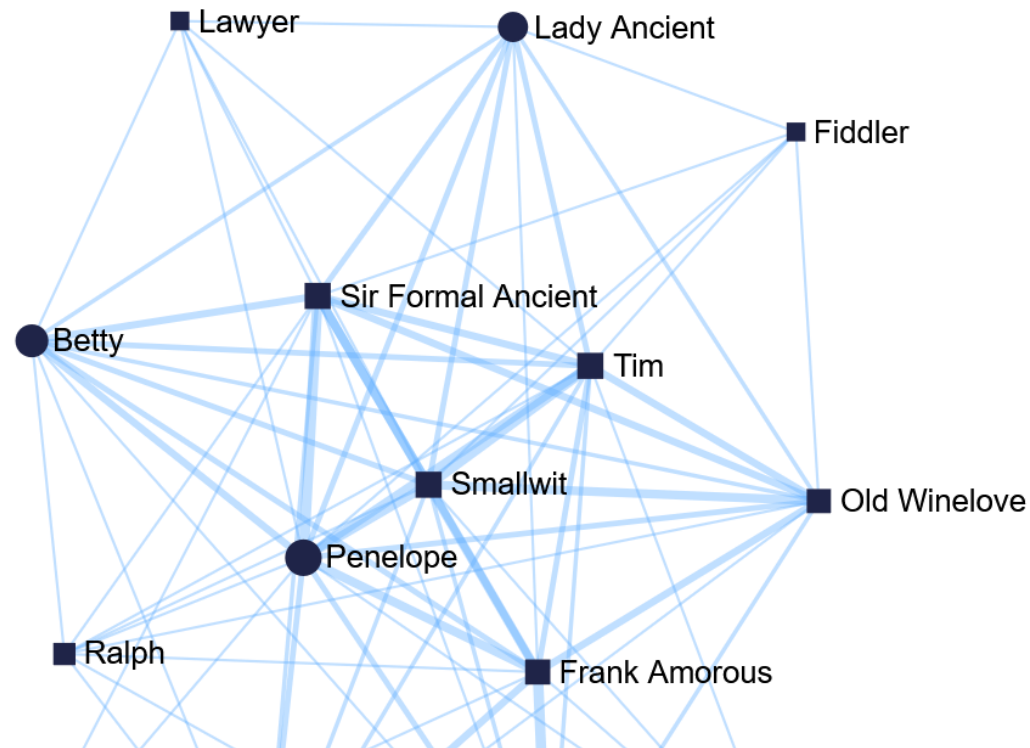


Thomas D'Urfey

Q2145838

b. 1653, Devon
d. 1723, London

Network Speech distribution Full text Downloads Tools



This tab shows a co-occurrence network. If characters appear in the same scene or act, they are linked.

Network properties

Segments: 17
All-in at segment 17 (at 100%)
Network size: 15
Density: 0.7
Diameter: 2
Average path length: 1.3
Average clustering coefficient: 0.86
Average degree: 9.87
Maximum degree: 14 (Penelope)

Characters (in order of appearan...

DraCor API (standardised, structured access to all corpus data via URL queries)

DraCor API

1.0.2 OAS 3.0

[/api/v1/openapi.yaml](#)

[Terms of service](#)
[Contact the developer](#)
MIT

Servers

https://dracor.org/api/v1

public

GET /info API info

GET /corpora List available corpora

GET /corpora/{corpusname} List corpus content

GET /corpora/{corpusname}/metadata List of metadata for all plays in a corpus

GET /corpora/{corpusname}/metadata/csv List of metadata for all plays in a corpus

GET /corpora/{corpusname}/plays/{playname} Get metadata and network metrics for a single play

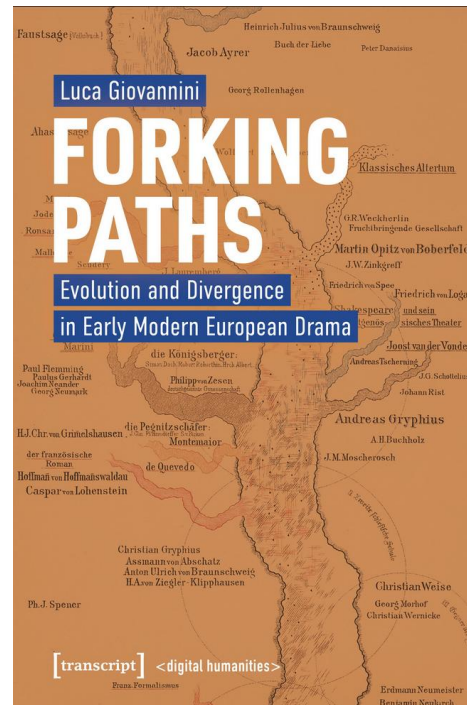
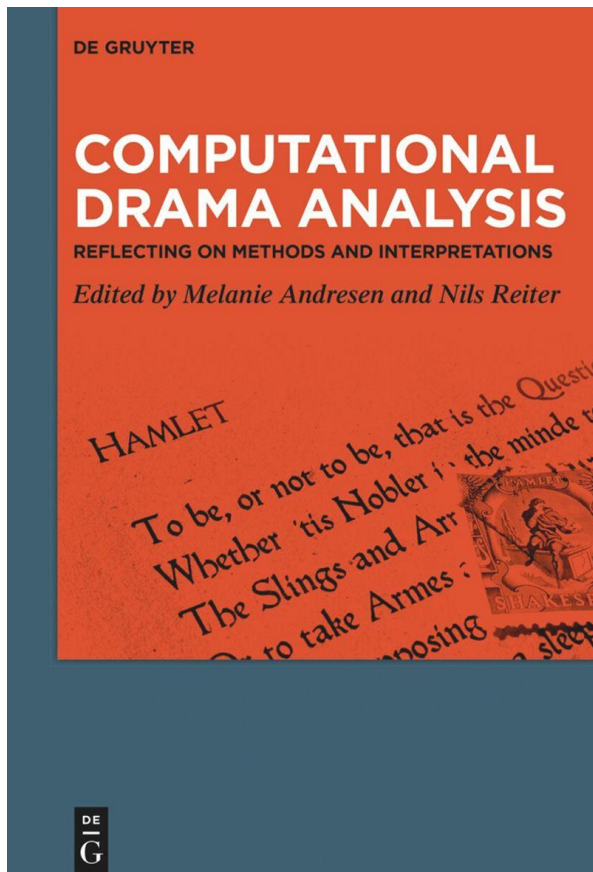
DraCor community



Left: core DraCor team (Vienna, 2024).

Above: participants to the first DraCor Summit (Berlin, 2025)

An empirical observation: in the last few years, a large portion of research in the field of computational drama analysis has been DraCor-powered...



We mine a DraCor-related bibliography as a proxy to understand which **topics**, **methods**, **tools**, and **scholars** have been shaping the field of computational drama analysis across the last decade.



Corpus

- Full texts of 114 papers from the DraCor bibliography (April 2025): journal articles, conference proceedings, book chapters, etc.
- Includes only works that actively use DraCor corpora or API (mere mentions are not enough)
- Some metadata:
 - Language: >50% in English, ~25% in German
 - Authorship: 35% written by teams of 3–5 authors
 - Most cited corpus: GerDraCor (35% of papers), followed by RusDraCor and FreDraCor



Case studies

- 1: Semantic Space
- 2: Methods, Tools & Formats
- 3: Citation Network

CS1: Semantic Space

Research question

- What is the semantic space of DraCor-related research? Do papers cluster by topic or method?

Method

- Full texts embedded with a multilingual transformer.
- Papers split into chunks, chunk embeddings mean-pooled into one vector per paper.
- PCA applied to reveal global structure in a language-agnostic way.

Infrastructure vs. research divide (PC1)

- ## Methodological gradient (PC2)

- Upper: ML-based papers (e.g. with classification tasks). Lower: ‘distant-reading’ approaches.



CS2: Methods, Tools & Formats

Research question

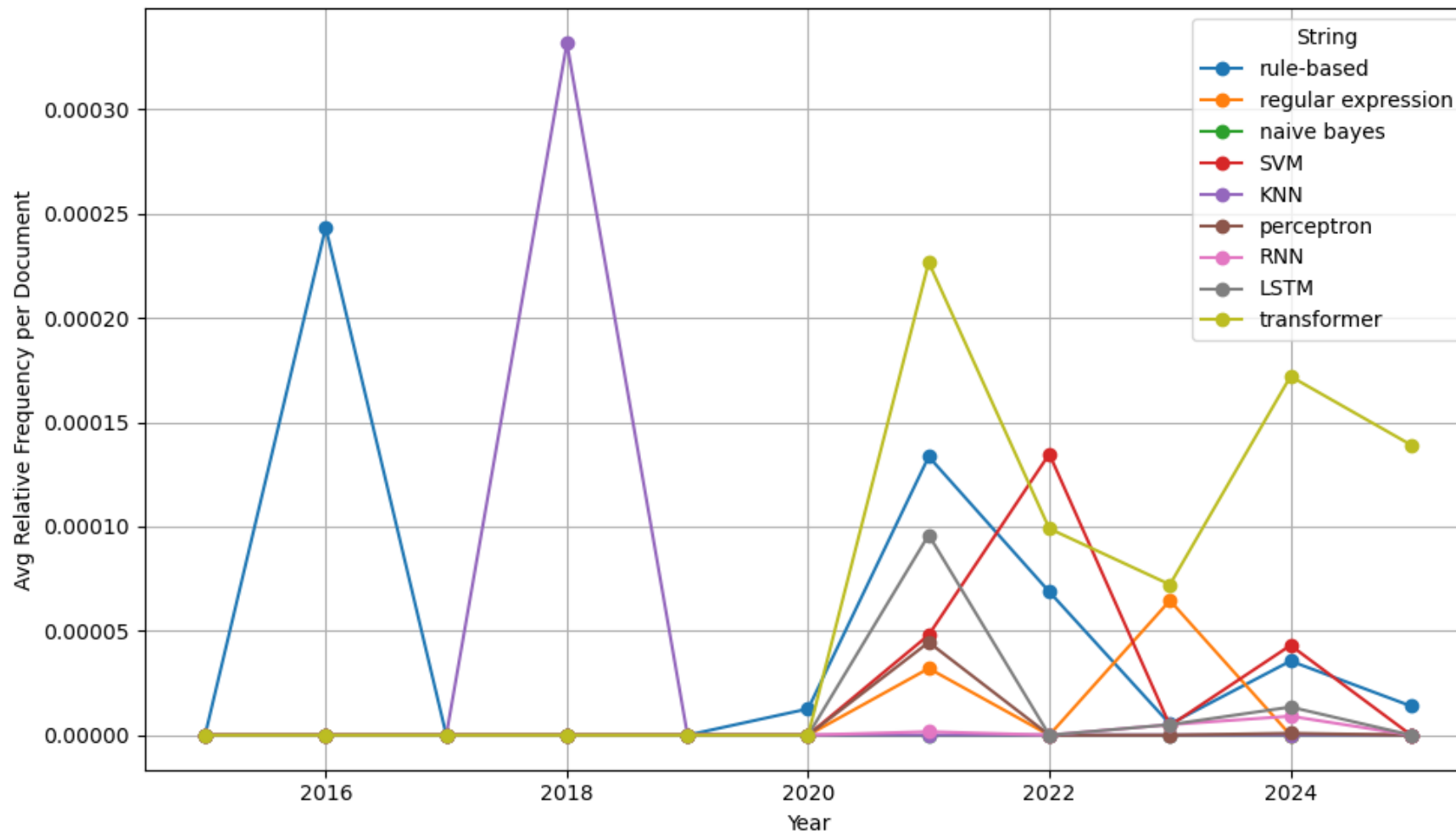
- What tools, formats, and methods recur across the DraCor bibliography? How have they evolved over time?

Method

- Multilingual word lists (Schöch, Fileva, Dudar 2022). String matching of n-grams (1–4 words), case-sensitive, accounting for spelling variants (e.g. lemmatisation/lemmatization) and translations.
- Counts aggregated per paper and diachronically to track trends 2015–2025.

CS2: Results

- **Top formats** (>70% of papers): TEI, XML
- **Top tools:** Python (40%) and Github (30%) signal engagement with DraCor's 'programmable corpora' concept.
 - API mentions peaked at dracor.org launch (2017).
- **Top methods:** annotation (45%); network analysis, classification, ML (30–40%)
 - Early papers used mainly network analysis; later contributions added classification and machine learning.



Algorithm shift
after 2021:
transformers
overtook
classical ML
and other
neural
architectures

CS3: Citation Network

Research question

- Which works are most influential within the DraCor bibliography itself? Can distinct research communities be identified?

Method

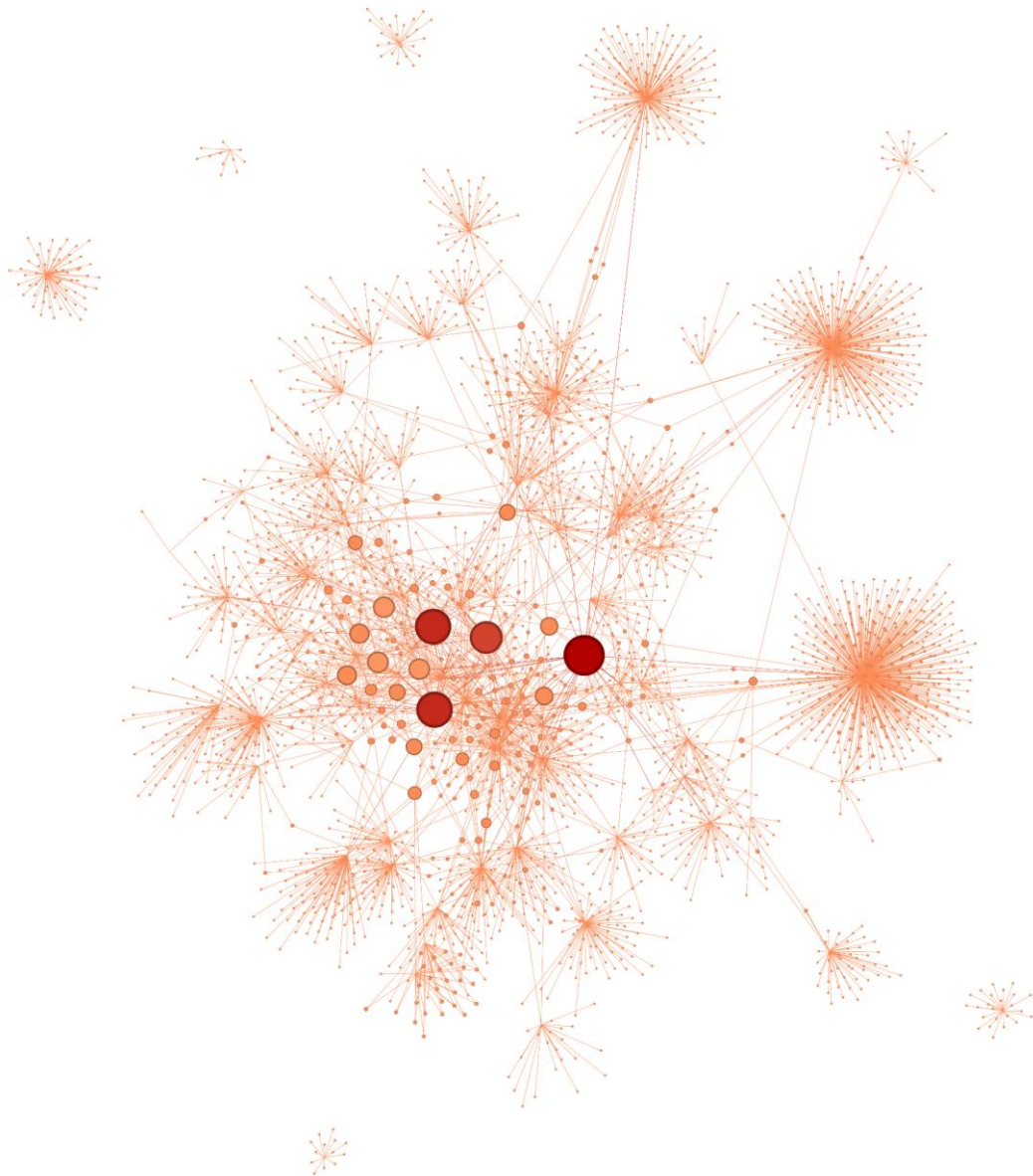
- Network of publications cited within individual items of the DraCor bibliography
 - More than 3,000 nodes (publications), edges = citation links
 - Node size scaled to in-degree (number of papers citing it).
 - Removed the standard DraCor citation paper (Fischer et al. 2019)



CS3: Citation Network — Results

Scale-free, hub-and-spoke topology ([live demo](#))

- Most papers have small citation clusters; a few highly-cited works become central hubs.

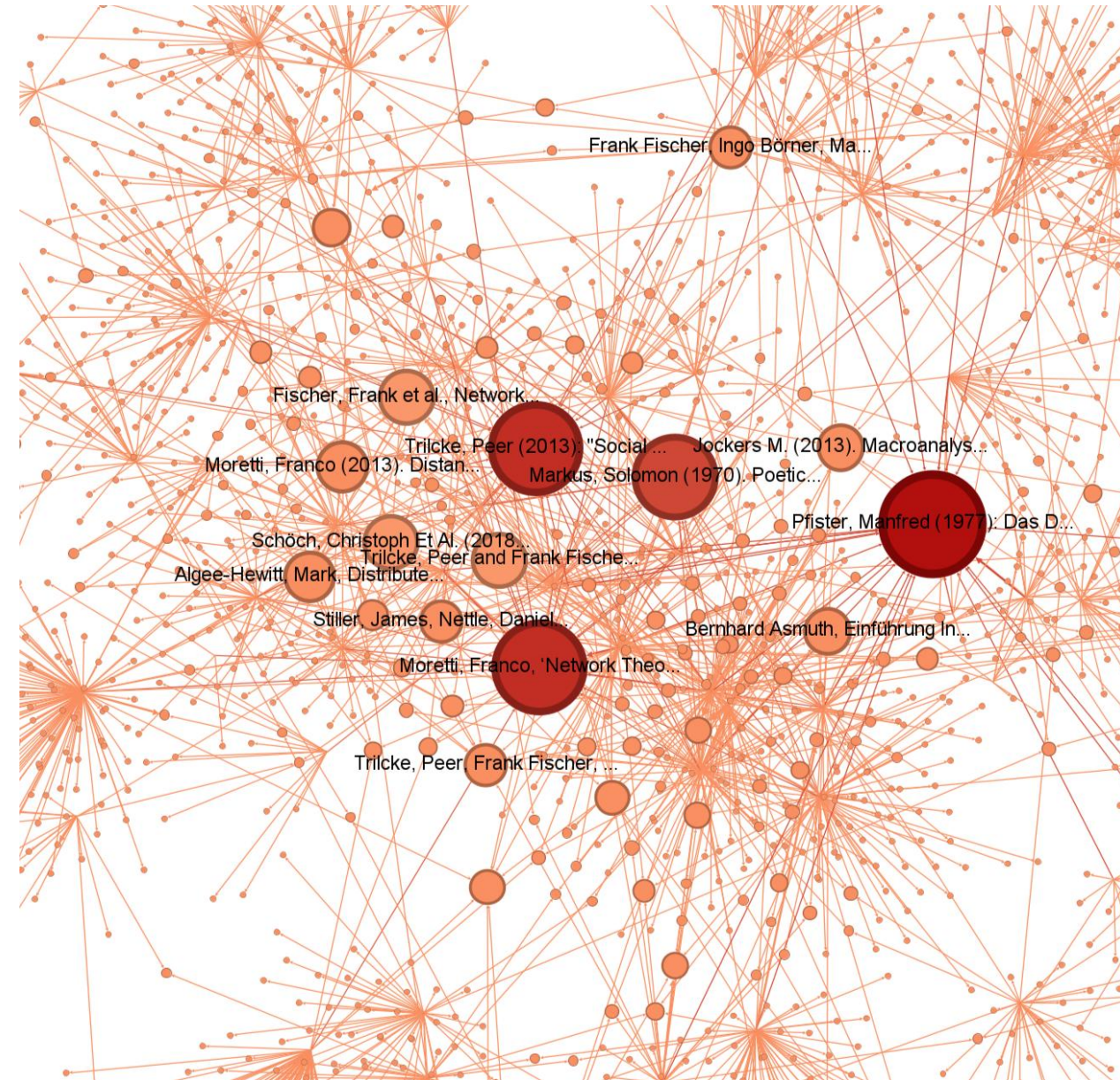


No major paradigm shifts: structuralist-formalist tradition dominates

- Most-cited: Pfister 1977, Marcus 1970 (early seminal works); Moretti 2011, Trilcke 2013 (literary network analysis).

Takeaway

- CDA has been methodologically responsive but theoretically stable, building on the same formalist foundations across a decade.



Summary & Outlook

The spectrum of DraCor-based scholarship has broadened

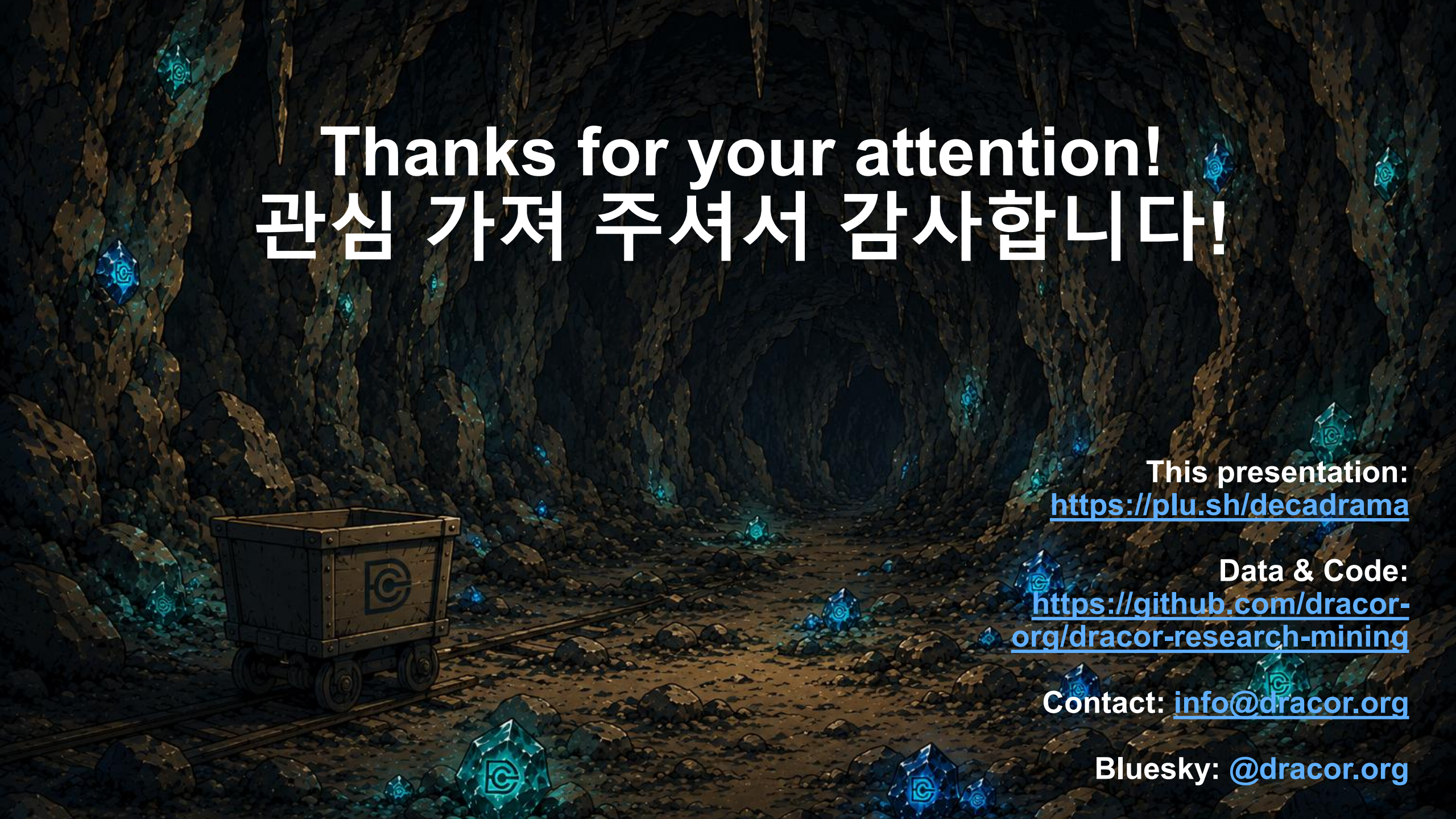
- E.g. network analysis (central in the early DraCor era) progressively flanked by other methods → DraCor is evolving into a more and more versatile resource for digital literary studies

Stable theoretical core

- Citation network shows consolidated clusters of key structuralist references; no major paradigm shifts detected over 10 years

Next steps: DraCorOS project

- Ontology-based modelling, Zenodo self-reporting routine for authors, Linked Open Data integration, knowledge graph for complex queries

The background is a detailed illustration of a dark, cavernous mine. The walls are made of jagged, dark brown and black rock. Numerous glowing blue crystals, each with a white 'C' logo inside, are embedded in the rock walls and scattered on the floor. In the lower-left foreground, a wooden mine cart with metal wheels and a 'C' logo on its side sits on a set of wooden tracks. The floor is uneven and covered with small rocks and debris. The lighting is dim, with the primary light source being the glowing crystals, creating a mysterious and industrial atmosphere.

Thanks for your attention!

관심 가져 주셔서 감사합니다!

This presentation:
<https://plu.sh/decadrama>

Data & Code:
<https://github.com/dracor-org/dracor-research-mining>

Contact: info@dracor.org

Bluesky: [@dracor.org](https://bsky.app/profile/@dracor.org)

Bonus content: the DraCor Agentic Research Lab

(a prototype by Ingo Börner 🤝 Claude)

Research Literature



152 publications in the DraCor research bibliography. **136** indexed in the RAG.

Stage Directions

All languages

All papers

Showing 4 of 152 entries

☐	Year ▼	Authors	Citation	Lang	Keywords ⓘ	Topics ⓘ	RAG
☐	2024	Katrin Dennerlein, Tho...	Katrin Dennerlein, Thomas Schmidt, Christian Wolff: Emotions in Stage Directions in German Drama of the Early Modern Period: Explorations via Computational Emotion Classification. In: Melanie Andresen, Nils Reiter (eds.): Computational Drama Analysis: Reflecting Methods and Interpretations. Berlin, Boston: De Gruyter 2024, pp. 189–215. (doi:10.1515/9783111071824-010) ↗	en	sd emotions emotion st period	Topic 19 Topic 1 Topic 4	✓
☐	2020	Peer Trilcke, Christoph...	Peer Trilcke, Christopher Kittel, Nils Reiter, Daria Maximova, Frank Fischer: Opening the Stage: A Quantitative Look at Stage Directions in German Drama. In: DH2020: »carrefours/intersections«. 22–24 July 2020. Book of Abstracts. University of Ottawa.	en	sd st epification topological pos	Topic 19 Topic 1	✓
☐	2019	Daria Maximova, Frank...	Daria Maximova, Frank Fischer: Using Machine Learning for the Automated Classification of Stage Directions in TEI-Encoded Drama Corpora. In: TEI2019: »What is text, really? TEI and beyond«. 16–20 September 2019. Book of Abstracts. University of Graz.	en	economics moscow school mid interference	Topic 4 Topic 9	✓

Backend: Chroma database, RAG implementation on indexed DraCor research literature
Frontend: bibliography browser importing methods from our paper

[← Bibliography](#)

The Schemer in German [🔗](#)

Krautter, Benjamin and Pagel, Janis (2024) en

Krautter_2024_Schemer

44 chunks indexed 5.147 tokens

Research this paper

Topics

Topic distribution based on chunk-level BERTopic clustering. Each topic shows the share of this paper's chunks assigned to it.

Method: BERTopic on chunk embeddings (nomic-embed-text-v2-moe)



- **Topic 7** feature features schemers character results 68%
- **Topic 1** proceedings url conference computational linguistics 30%
- **Topic 24** pass dramacoref mentions coreference passes 2%

Methods (6)

Computational and analytical methods detected in the paper text using curated multilingual word lists.

For each article, an
overview with
formats, tools,
methods and topics
(via BERTopic)

New Research Task

Research Question

e.g. What methods, tools, and software are used for network analysis of dramatic texts?

Which papers in the DraCor bibliography work on stage directions and how?

Papers (optional)

Search papers by identifier, title, or author...

Start Research

 Hide models

Research Agent

qwen2.5:14b (14.8B)



Tool calling required

Synthesis Agent

qwen2.5:14b (14.8B)



which-papers-in-the-dracor-research-literature-wor

2026-03-06T000949

Task

Which papers in the DraCor research literature work on stage directions? What approaches and methods do they use?

Research Report

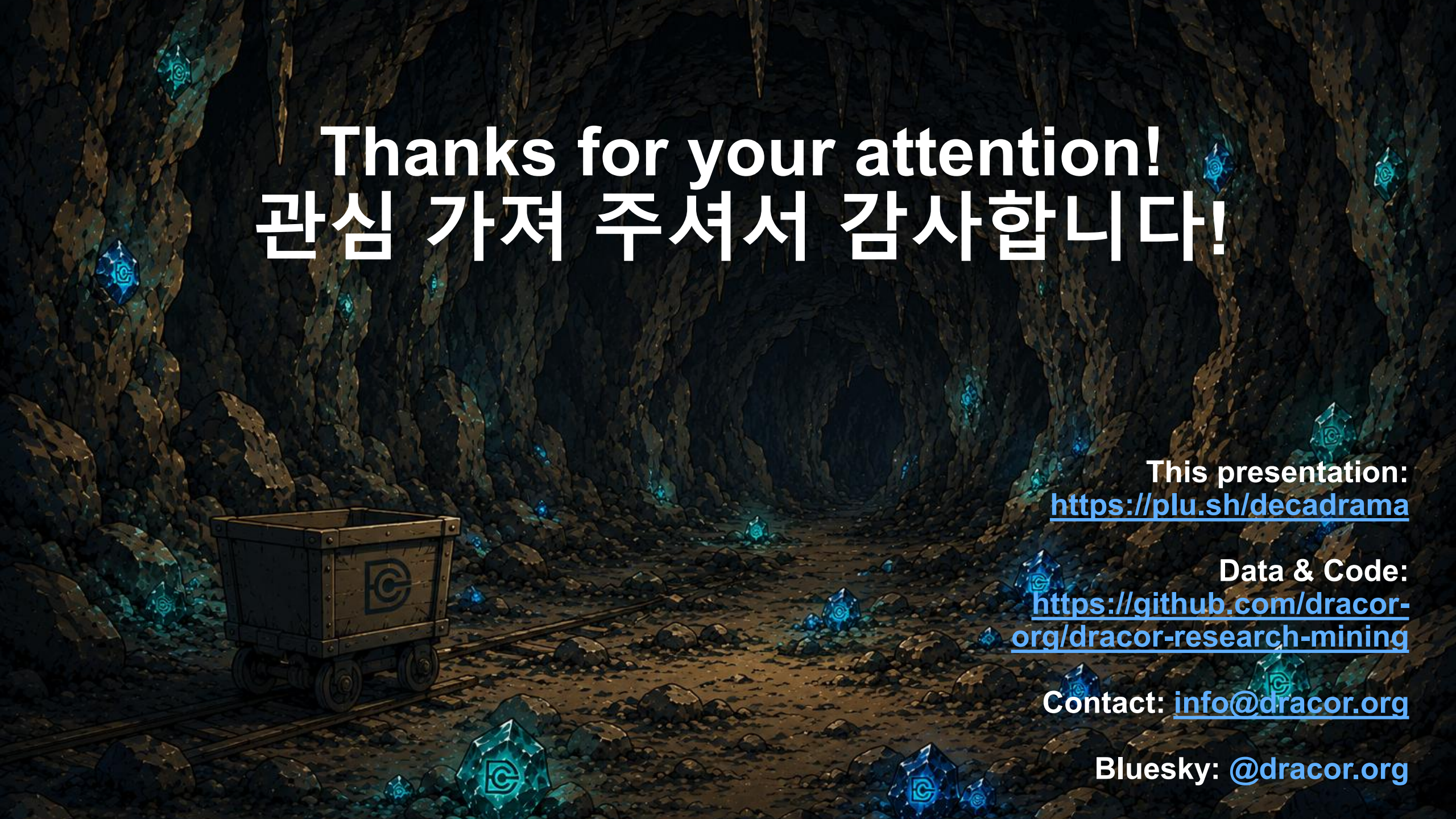
Introduction

The DraCor research literature contains several studies focusing on stage directions in drama texts, employing diverse methods to analyze and classify them. This report synthesizes key approaches and methodologies used by researchers in this domain.

Machine Learning for Stage Direction Classification

Maximova (2019)

Maximova's paper introduces a machine learning approach to automatically assign types to stage directions in TEI-P5-encoded Russian Drama Corpus (RusDraCor) (Maximova_2019_Machine [papers/Maximova_2019_Machine], p. 1). The study achieves an F1 score of approximately 0.75, indicating that about three out of four stage directions are classified correctly without human intervention. This method enables the analysis of larger corpora over extended periods and provides insights into the evolution of stage direction usage in drama literature.

The background is a detailed illustration of a dark, cavernous mine. The walls are made of jagged, dark brown and black rocks. Numerous glowing blue, crystalline structures are embedded in the walls and scattered on the floor. In the lower-left foreground, a wooden mine cart with metal wheels and a 'C' logo is on a set of tracks. The floor is uneven and covered with small rocks and debris. The lighting is dim, with the primary light source being the glowing crystals, creating a mysterious and industrial atmosphere.

Thanks for your attention!

관심 가져 주셔서 감사합니다!

This presentation:
<https://plu.sh/decadrama>

Data & Code:
<https://github.com/dracor-org/dracor-research-mining>

Contact: info@dracor.org

Bluesky: [@dracor.org](https://bsky.app/profile/@dracor.org)