A Modest Proposal for Operationalising Dramatic Texts

Luca Giovannini

(University of Potsdam, Germany)

This presentation: shorturl.at/duttD

It all starts with operationalisation

"Operationalizing means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts" (Moretti 2013: 13, cf. Pichler and Reiter 2022, Jacke 2025).

With **much** approximation, we can operationalise drama by

- identifying and quantifying **dramatic features**
- modelling their **relations**

What are the ʹkey componentsʹ of a play?

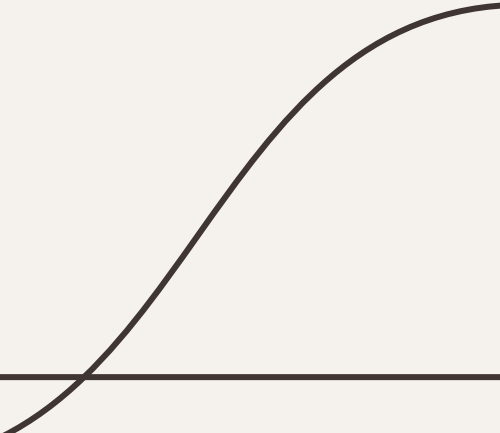| Aristotle (*Poetics* 1450a, 5-10)	Russian Formalists (e.g. Boris Yarkho)	Modern literary theory (e.g. drama handbooks)
six "constituent parts" (μέρη) of tragedy: plot, character, diction, thought, spectacle, song	dialogue (диалог) characters (персонажи) action (действие)	characters dialogue plot

From 'components' to actual metrics

Two common approaches:

- Using basic metrics readily extractable from the dramatic text
- Developing new measures tailored to specific research questions

⚠️ "Measures can be taken on any quantifiable aspect of a text, but finding out the significance of that metric for understanding the text, or better, mapping that metric onto a pre-existing critical concept (such as style, plot, or theme), is crucial to making sense of what is being measured". (Algee-Hewitt, 2017, 759)

Notable attempts at finding features for drama: a quick tour

Boris Yarkho

(1889–1942)

- Member of the Moscow formalist circle
- Believer in the "possibility and necessity of a »complete quantification« of philological knowledge" (Fischer et al. 2019a)
- *Methodology for the Exact Study of Literature* (2006)
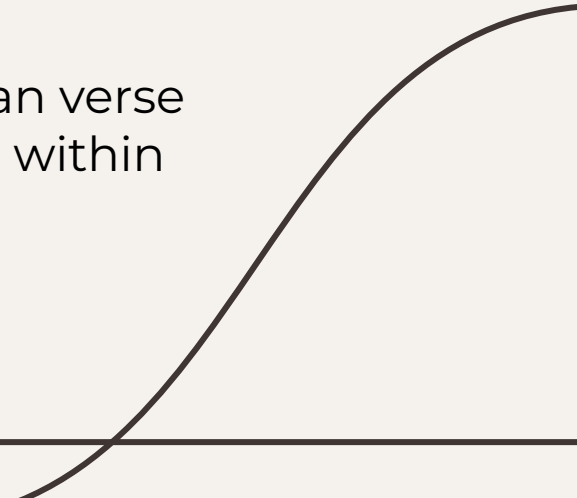
Yarkho's two studies on drama

- "Speech Distribution in Five-Act Tragedies. A question of Neoclassicism and Romanticism"

    1. the percentage of scenes with 1, 2, 3… speakers (monologues, dialogues, trialogues…);
    2. the uniformity in the distribution of this percentage in the plays of each period (sigma);
    3. the average number of scenes in a play (the coefficient of mobility of action);
    4. the number of characters in a play.

- "The Comedies and Tragedies of Corneille. A Study in the Theory of Genre"

the liveliness of dialogue (the average length of lines); the types of stylistic connection between speeches and the dialogue's connectivity coefficient; the sensory tone of visual images (sensory, i. e. colour, light, sound, etc. epithets); the types of verbally named emotions (anger, horror …); the accumulation of characters per act; the amount of staged (prescribed by the plot) gestures (murder, transfer of objects, reading letters, etc.): (a) on stage, (b) off stage; the number of collective characters and their contribution to dialogue;
the participation of the servants; (a) in dialogue, (b) in action; **justified and unjustified stage entrances and exits**; the types of endings….

More metrics from Russian Formalists

- Revzina and Revzin (1971): idiosyncratic measures such as 'character connectedness', 'dramatic engagement rank', and 'probabilistic link'
- Sapogov (1974): measures for computing speech distribution patterns
- Sperantov (1998) on stage directions in Russian verse tragedy: frequency, average length, presence within verses, lexical diversity, etc.

Modern examples

Netzwerkdynamik, Plotanalyse – Zur Visualisierung und Berechnung der ›progressiven Strukturierung‹ literarischer Texte

Trilcke, Peer
trilcke@uni-potsdam.de
Universität Potsdam, Deutschland

Fischer, Frank
ffischer@hse.ru
Higher School of Economics, Moskau, Russland

Göbel, Mathias
goebel@sub.uni-goettingen.de
Staats- und Universitätsbibliothek Göttingen, Deutschland

Kampkaspar, Dario
kampkaspar@hab.de
Herzog-August-Bibliothek Wolfenbüttel, Deutschland

Kittel, Christopher
contact@christopherkittel.eu
Universität Graz, Österreich

measures for investigating plot development:

"event-based"
(*all-in index*, *final scene size*)

"progression-based"
(*drama change rate*, *beat chart*)

Distributed Character: Quantitative Models of the English Stage, 1550–1900

Mark Algee-Hewitt

New Literary History, Volume 48, Number 4, Autumn 2017, pp. 751-782 (Article)

advanced network centrality metrics:
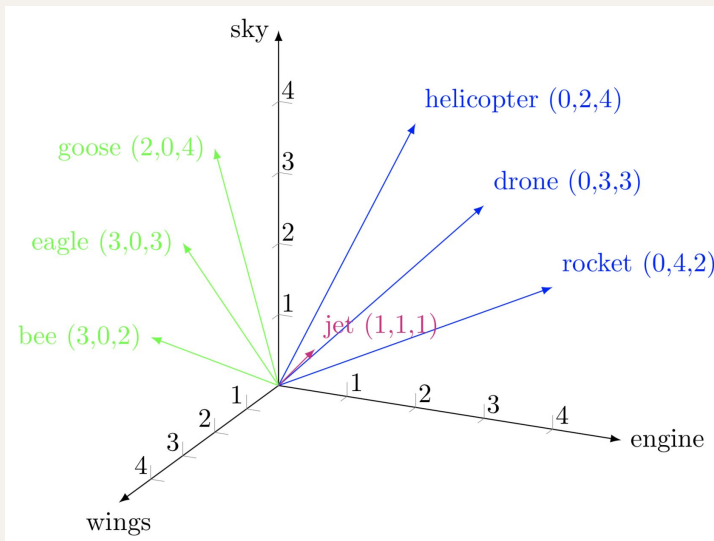*protagonism* and *mediatedness*

A possible paradigm shift?

Examining and comparing individual dramatic features between plays

→

Constructing and using an holistic representation of a play via its structural features

The (obvious?) answer: vectors

- In mathematics and physics: a quantity that has both magnitude (size) and direction
- In computer science and DH: an **ordered** sequence (tuple/list/array) of numbers (scalars)

Vectors/embeddings in DH research

Different textual units have been vectorised so far:

Word // Sentence // Document

… within different methods and research paradigms (stylometry, topic modelling, cultural analytics, etc.)

… but what about vectors based on non-linguistic textual properties, i.e., **vectors of latent features**?

A modest proposal:

For the purposes of computational drama analysis, let's build »**play embeddings**« (i.e. vectors composed of features that embody different <u>formal</u> aspects of the text) and use them as proxies for the plays themselves.
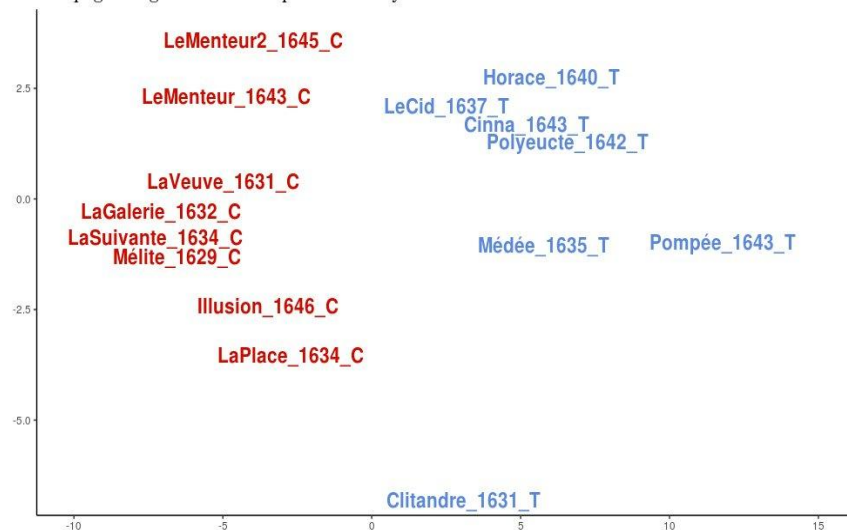
A brand new idea? Not really
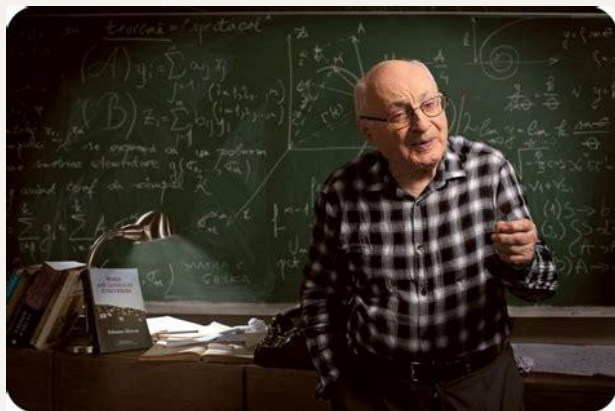
Formal embedding *of characters* has already been experimented with by S. Marcus in 1970 (one-hot encoding of character appearances)

Character = {0,1,1,1,0,1,0,0,1}



Comedies and **Tragedies** of Pierre Corneille

Data come from large-scale quantiative study on distinistive features of classic dramatic genres in Corneille **done by Boris I. Yarkho in 1920s**. Each text was represented across 15 features that Yarkho tried to synthesise into clear 'comedy' vs. 'tragedy' cut. This study served as a general demonstration of Yarkho's grand project of quantiative methodology for literary studies. 120 pages long work was first published only in 2006.

LeMenteur2_1645_C

LeMenteur_1643_C

Horace_1640_T

LeCid_1637_T

Cinna_1643_T

Polyeucte_1642_T

LaVeuve_1631_C

LaGalerie_1632_C

LaSuivante_1634_C

Mélite_1629_C

Médée_1635_T

Pompée_1643_T

Illusion_1646_C

LaPlace_1634_C

Clitandre_1631_T

by @artjomshl 2020-07-01

Digital reproductions of Yarkho's study on Corneille (Shelya 2020)

A easy computational implementation of vectorisation
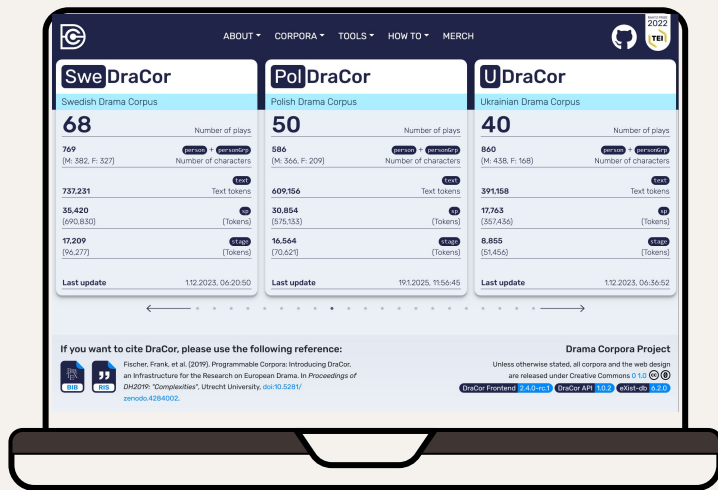
- Via the **DraCor** Project (Fischer et al. 2019)
- An open digital infrastructure for the computational study of drama
- Hosting +4000 plays in 17 languages
- Made of **programmable corpora**, i.e. corpora that expose an open, transparently documented and research-driven API
- For each play, the DraCor **metrics service** computes a range of textual features

New API release (1.1.0) out now!

(Numeric) features counted or computed for each play by the DraCor API

Character network	"averageClustering", "density", "averagePathLength", "maxDegreeIds", "averageDegree", "diameter", "numConnectedComponents", "numEdges"
Cast statistics, speech distribution	"numOfCoAuthors", "numOfSpeakers", "numOfSpeakersFemale", "numOfSpeakersMale", "numOfSpeakersUnknown", "numOfPersonGroups", "wordCountText", "wordCountSp", "wordCountStage"
Play size	"numOfActs", "numOfScenes", "numOfSegments", "numOfP", "numOfL"

A play

© Biblioteca Nacional de España

Its vector representation

=

V = {41, 3, 14, 11, 3, 0, 0, 1, 3306, 16919, 16461, 473, 6.571428571428571, 0.505494505494505, 0.7699649556792411, 14, 11, 1, 46, 0.0494505494505494, 0.6608466660481829, 0.2431203018946239, 0.926829268292683, 0.5714285714285714, 0.6391666666666669, 3.0, 0.5974025974025974, 0.8461538461538461, 0.3571428571428571, 0.2857142857142857, 0.3571428571428571, 4.99, 0.3571428571428571, 0.2857142857142857, 0.3571428571428571, 0.0175417232479634, 0.5493558858726955}

Using the DraCor API [0.2.0 OAS3] /api.yaml to get vector data

- ❏ 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 one play
- ❏ GET /corpora/{corpusname}/plays/{playname}/characters → Get metadata for characters

	name	id	size	averageClustering	density	averagePathLength	averageDegree	diameter	datePremiered	yearPremiered	yearPrinted	maxDegree
2	clarin-teresa	span000018	9,00	0.9523809523809526	0.944444444	10555555555555500,00	7555555555555550,00	2,00			1895,00	1895,00	8,
3	dicenta-juan-jose	span000001	16,00	0.8348439754689755	0.483333333	16083333333333300,00	45839,00	3,00			1895,00	1895,00	12,
4	echegaray-arrastrar	span000003	15,00	0.8443007733007734	0.609523809	14095238095238000,00	8533333333333330,00	3,00			1905,00	1901,00	13,
5	echegaray-mancha	span000002	9,00	0.9153439153439152	0.888888888	11111111111111100,00	7111111111111110,00	2,00			1895,00	1895,00	8,
6	galdos-casandra	span000005	16,00	0.741504329004329	0.516666666	15416666666666600,00	27576,00	3,00				1910,00	12,
7	galdos-electra	span000006	15,00	0.7921386021386021	0.533333333	14761904761904700,00	7466666666666660,00	3,00				1901,00	13,
8	galdos-perfecta	span000004	16,00	0.8261984543234543	0.575	14333333333333300,00	8625,00	3,00			1896,00	1896,00	14,
9	lorca-bernarda	span000042	18,00	0.8625000000000002	0.718954248	12941176470588200,00	12222222222222200,00	3,00				1945,00	16,
10	lorca-bodas	span000041	25,00	0.8671424124849889	0.596666666	14033333333333300,00	14.32	2,00				1938,00	24,
11	lorca-rosita	span000007	20,00	0.8651508295625943	0.342105263	17421052631578900,00	45783,00	3,00			1935,00	1938,00	17,
12	lorca-yerma	span000008	16,00	0.832042957042957	0.616666666	12952380952380900,00	45901,00	2,00			1934,00	1938,00	14,
13	lorca-zapatera	span000040	15,00	0.8977441077441078	0.561904761	14761904761904700,00	7866666666666660,00	3,00				1938,00	12,
14	munoz-conferencia	span000032	10,00	0.8126984126984127	0.688888888	13111111111111100,00	45694,00	2,00			1916,00	1916,00	9,
15	munoz-ortiz	span000053	16,00	0.8887529137529139	0.825	10571428571428500,00	12375,00	2,00			1927,00	1927,00	14,
16	munoz-pergaminos	span000030	26,00	0.8304698930044186	0.513846153	14984615384615300,00	12846153846153800,00	3,00			1918,00	1918,00	23,
17	munoz-refugio	span000031	17,00	0.8471417471417473	0.647058823	13529411764705800,00	10352941176470500,00	2,00			1933,00	1933,00	16,
18	unamuno-esfinge	span000056	12,00	0.8119107744107744	0.651515151	13484848484848400,00	7166666666666660,00	2,00				1898,00	11,
19	unamuno-fedra	span000012	6,00	0.8388888888888889	0.733333333	12666666666666600,00	36666666666666600,00	2,00				1924,00	5,
20	valera-asclepigenia	span000027	8,00	0.7833333333333333	0.535714285	45778,00	27454,00	3,00			1928,00	1878,00	6,
21	valera-atahualpa	span000028	12,00	0.6527777777777778	0.409090909	17272727272727200,00	45781,00	3,00				1878,00	8,
22	valle-aguila	span000043	71,00	0.8413618618829599	0.125955734	2503420523138830,00	8816901408450700,00	5,00			1907,00	1907,00	30,

Metadata CSV for the Spanish DraCor (snippet)

All 4.318 DraCor plays, vectorised
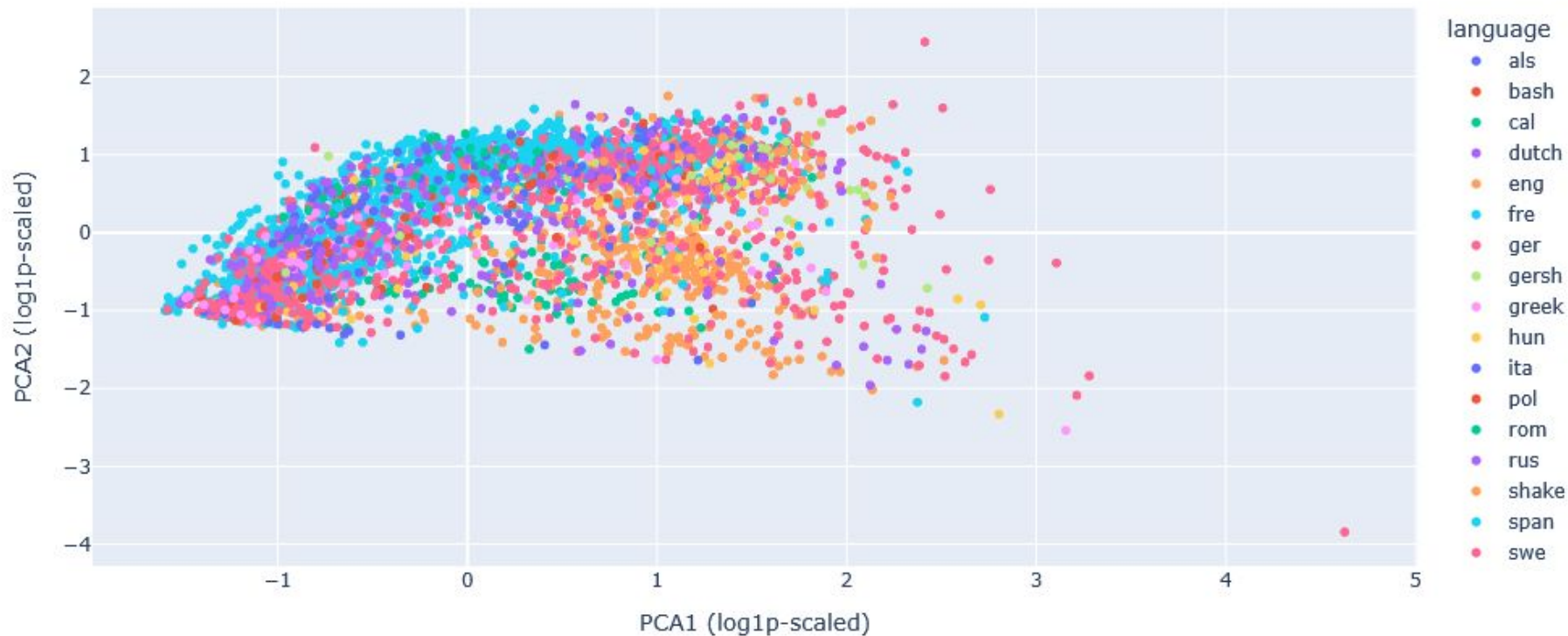
All 63.431 DraCor characters, vectorised
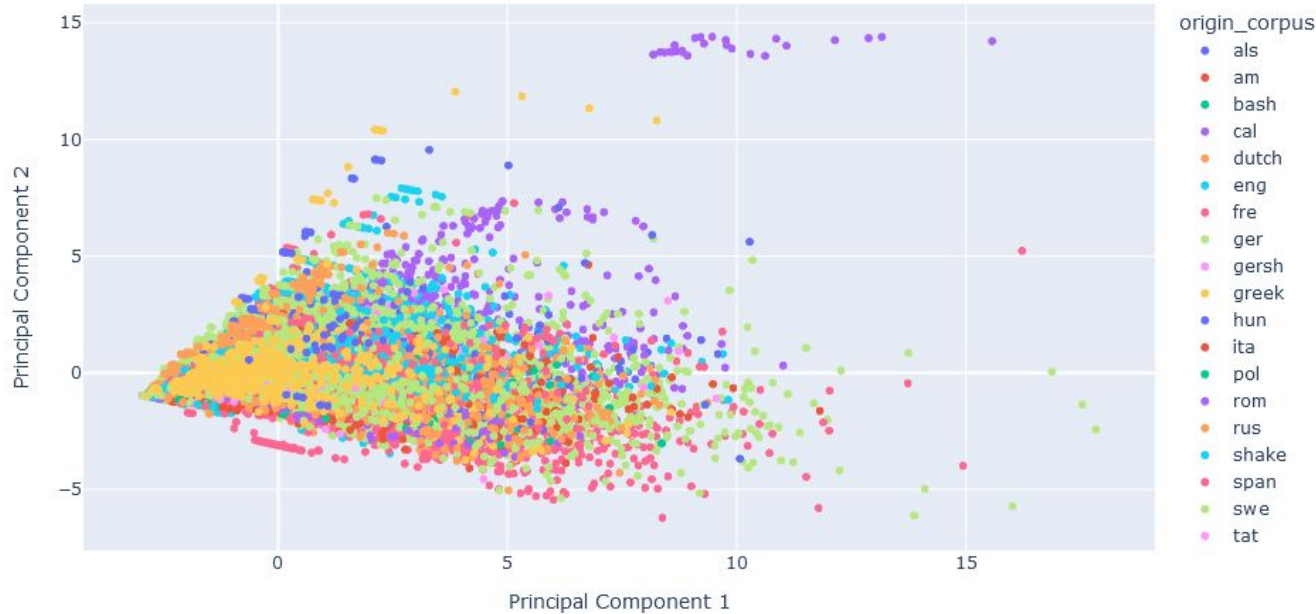
From: Luca Giovannini and Daniil Skorinkin, "Computational Modelling of Literary Characters", talk at DH Kolloquium Erlangen-Nuremberg, 27.01.2025

Link: **plu.sh/cmcerlangen**

We can do distant reading via structural vectors

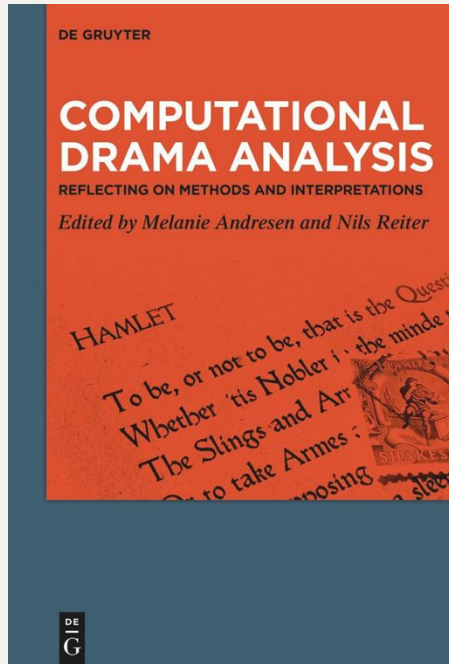Early modern French and English plays are structurally quite different

↓

Distinction between 'regular' vs. 'irregular' theatre

To conclude,
let's take a look at
some examples of
drama vectorisation

Exhibit 1: Tragic and comical networks
(Szemes & Vida 2024)

DE GRUYTER

COMPUTATIONAL DRAMA ANALYSIS

REFLECTING ON METHODS AND INTERPRETATIONS

Edited by Melanie Andresen and Nils Reiter

HAMLET

To be, or not to be, that is the Question:
Whether 'tis Nobler i: the minde to
The Slings and Arr
Or to take Armes :

"Our aim is to create a method that is able to cluster texts with similar structures on the basis of the play's well-interpretable and simple properties" (165)

Research goal: distinguishing comedies and tragedies in GerDraCor and ShakeDraCor on the basis of metrics related to network theory and speech/stage time distribution

Result: "measures of network theory and the distribution of speech and stage time between characters can be helpful in identifying a genre fingerprint that can be used to distinguish between comedies and tragedies"

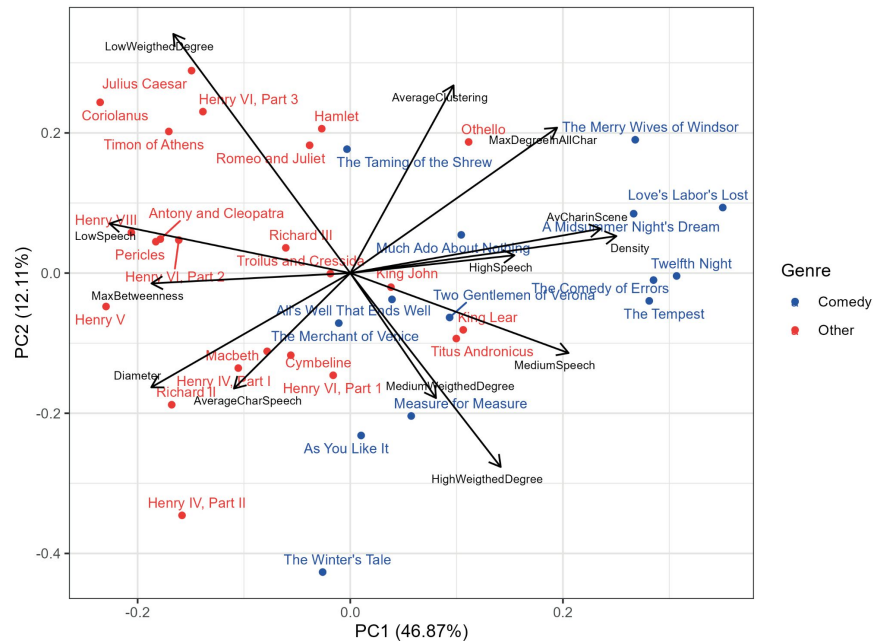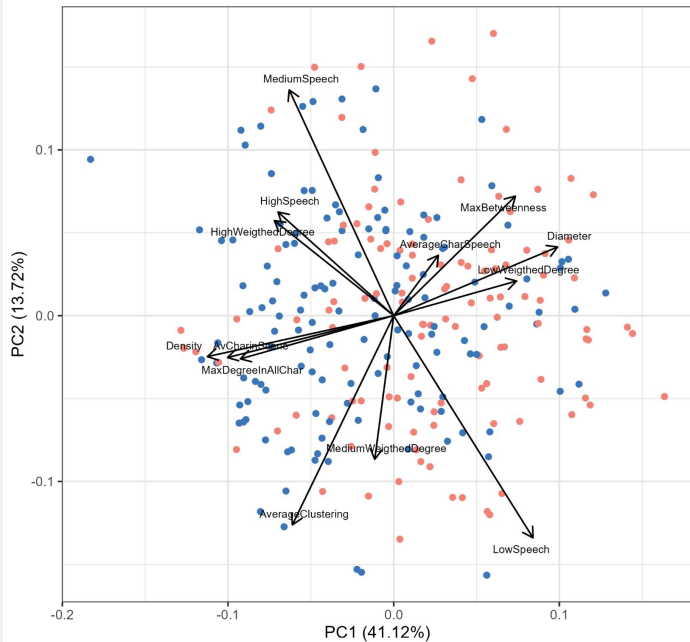# GerDraCor vs ShakeDraCor, vectorised

Exhibit 2: Opera libretti (Giovannini & Skorinkin 2024)

jcls. Journal of Computational Literary Studies

Article

Computational Approaches to Opera Libretti
An Experiment on DraCor Corpora

Luca Giovannini [1,3]
Daniil Skorinkin [2]

OPEN ACCESS

1. Institute for German Studies, University of Potsdam, Potsdam, Germany.
2. Digital Humanities Network, University of Potsdam, Potsdam, Germany.
3. Department of Linguistic and Literary Studies, University of Padova, Padua, Italy.

Research question:
Are there any peculiar formal features which set libretti apart from comedies and tragedies from the same period?

Outcome: it is possible, to some extent, to identify traits clearly distinctive of libretti (comic and non-comic), and such traits do not always align with the ones characterising comedies and tragedies.
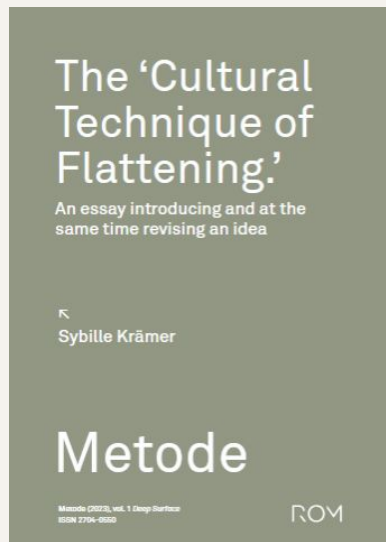
Exhibit 3: Formal development of early modern European drama (Giovannini 2024)

- **Idea**: Using vectorisation to follow large-scale cultural evolution
- **RQ**: Is it possible to trace the development of autonomous national dramatic traditions across Europe using formal features?
  - Well, kind of, maybe ask me later at the coffee break
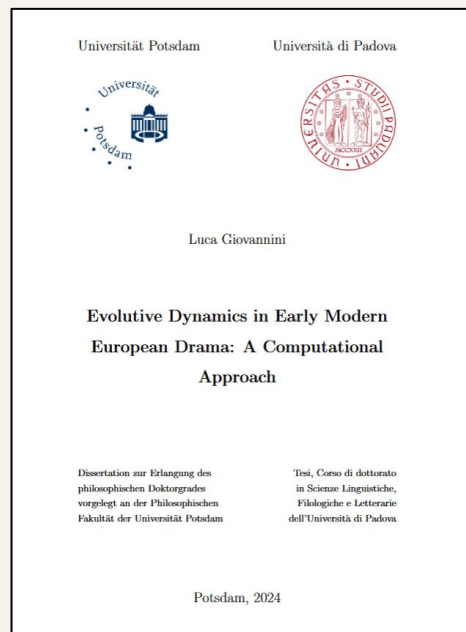
Is this method sound?

- Just another example of the simultaneously '**hubristic**' and '**reductionist**' mindset often ascribed to DH?

- Drama vectorisation is part of the broader "**cultural technique of flattening**" (Krämer 2023a, 2023b), fundamental not only to DH projects, but also to humanistic inquiry in general.
  - The common opposition between traditional "deep" hermeneutics and data-driven, "shallow" digital approaches is **misleading**, since humanities have **always** relied on surface-level representations of cultural objects (think of concordances, library signatures, catalogs of works, historical timelines, dating tables, and diagrammatic inscriptions)

The 'Cultural Technique of Flattening.'

An essay introducing and at the same time revising an idea

↖
Sybille Krämer

Metode

Metode (2023), vol. 1 Deep Surface
ISSN 2704-0550

ROM

Epistemic practices like **drama vectorisation** should be seen not as an attempt to overtly simplify the complexity of theatrical texts, but rather as the **continuation of formalist morphological thinking** with **computational tools**.

Obrigado!
Alguma pergunta?

giovannini@uni-potsdam.de
@lucagiovannini.bsky.social

This presentation:
https://shorturl.at/duttD

References

❑ Algee-Hewitt, Mark (2017). "Distributed Character: Quantitative Models of the English Stage, 1550–1900". New Literary History, 48(4):751–782. doi: https://doi.org/10.1353/nlh.2017.0038.

❑ Fischer, Frank, Akimova, Marina. and Orekhov, Boris. (2019a) "Preface: Data-Driven Formalism". Journal of Literary Theory, Vol. 13 (Issue 1), pp. 1-12. https://doi.org/10.1515/jlt-2019-0001.

❑ Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019b). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama". In DH 2019 Book of Abstracts, Utrecht. University of Utrecht. doi: https://doi.org/10.5281/zenodo.4284002.

❑ Giovannini, Luca (2024). "Evolutive Dynamics in Early Modern European Drama: A Computational Approach". Unpublished PhD thesis, University of Potsdam and University of Padova.

References

❏ Giovannini, Luca, and Daniil Skorinkin (2024). "Computational Approaches to Opera Libretti". Journal of Computational Literary Studies, 3. doi: https://doi.org/10.48694/jcls.3595.

❏ Jacke, Janina. (2025) 'Operationalization and Interpretation Dependence in Computational Literary Studies', Journal of Computational Literary Studies. 4(1). DOI: https://doi.org/10.48694/jcls.3959.

❏ Krämer, Sybille (2023a). "Should We Really 'Hermeneutise' the Digital Humanities? A Plea for the Epistemic Productivity of a 'Cultural Technique of Flattening' in the Humanities". Journal of Cultural Analytics, 7(4). doi: https://doi.org/10.22148/001c.55592.

❏ Krämer, Sybille (2023b). "The Cultural Technique of Flattening". Metode, 1. URL: https://metode.r-o-m.no/en/articles/essay/the-cultural-technique-of-flattening.

❏ Marcus, Solomon (1970). Poetica matematica. Academiei, Bucuresti.

❏ Moretti, Franco (2013). "'Operationalizing': or, the Function of Measurement in Modern Literary Theory", December. URL: https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf.

References

❏ Revzina, Olga G., and Isaak I. Revzin (1971). "Nekotorye Matematicheskie Metody Analiza Dramaturgicheskogo Postroeniia" ["Some Mathematical Methods for Analyzing Dramatic Constructions"]. Tochnye Metody v Issledovaniiakh Kul'tury i Iskusstva (Materialy k Simpoziumy), 291–300.

❏ Pichler, Axel, and Nils Reiter (2022). "From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities". Journal of Cultural Analytics, 7(4). doi: https://doi.org/10.22148/001c.57195.

❏ Revzina, Olga G., and Isaak I. Revzin (1971). "Nekotorye Matematicheskie Metody Analiza Dramaturgicheskogo Postroeniia" ["Some Mathematical Methods for Analyzing Dramatic Constructions"]. Tochnye Metody v Issledovaniiakh Kul'tury i Iskusstva (Materialy k Simpoziumy), 291–300.

❏ Sapogov, Vyacheslav A. (1974). "Nekotorye Harakteristiki Dramaturgicheskogo Postroeniia Komedii A. N. Ostrovskogo »Les«" ["Some Characteristics of the Dramatic Construction of A. N. Ostrovsky's Comedy »The Forest«"]. In Vyacheslav Alexandrovich Sapogov, editor, A. N. Ostrovskij i Russkaia Literatura [A. N. Ostrovsky and Russian Literature], 60–69. Yaroslavl State Pedagogical Institute, Kostroma.

References

❏ Sperantov, Vladimir V. (1998). "Poetika Remarki v Russkoy Tragedii XVIII - Nachala XIX vv. (K Tipologii Literaturnykh Napravleniy)" [The Poetics of Stage Directions in Russian Tragedy of the Eighteenth and Early Nineteenth Centuries (On the Typology of Literary Movements)]. Philologica, 5(11/13): 9–48. URL: https://rvb.ru/philologica/05/05sperantov.htm.

❏ Szemes, Botond, and Bence Vida (2024). "Tragic and Comical Networks: Clustering Dramatic Genres According to Structural Properties". In Melanie Andresen and Nils Reiter, editors, Computational Drama Analysis: Reflecting on Methods and Interpretations. De Gruyter, Berlin; Boston. URL: https://doi.org/10.1515/9783111071824-009.

❏ Trilcke, Peer, Frank Fischer, Mathias Göbel, Dario Kampkaspar, and Christopher Kittel (2017). "Netzwerkdynamik, Plotanalyse – Zur Visualisierung und Berechnung der Progressiven Strukturierung Literarischer Texte". In DHd2017 Book of Abstracts., Bern. doi: https://doi.org/10.5281/ZENODO.4622799.

❏ Yarkho, Boris (2006). Metodologiia Tochnogo Literaturovedeniia: Izbrannye Trudy po Teorii Literatury [Methodology for the Exact Study of Literature: Selected Works in Literary Theory]. Jazyki Slavianskikh Kul'tur, Moskva.