



Natural language processing for literary studies: Graph Literary Exploration Machine (GoLEM)

Agnieszka Karlińska

Wiktor Walentynowicz

Jan Wieczorek

Maciej Maryl



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





1. Reference point: **Literary Exploration Machine (LEM)**
2. Design considerations
3. Architecture
4. Use cases
5. Resources: **Corpus of Literary Studies Discourse (KDL)**
6. GoLEM in action

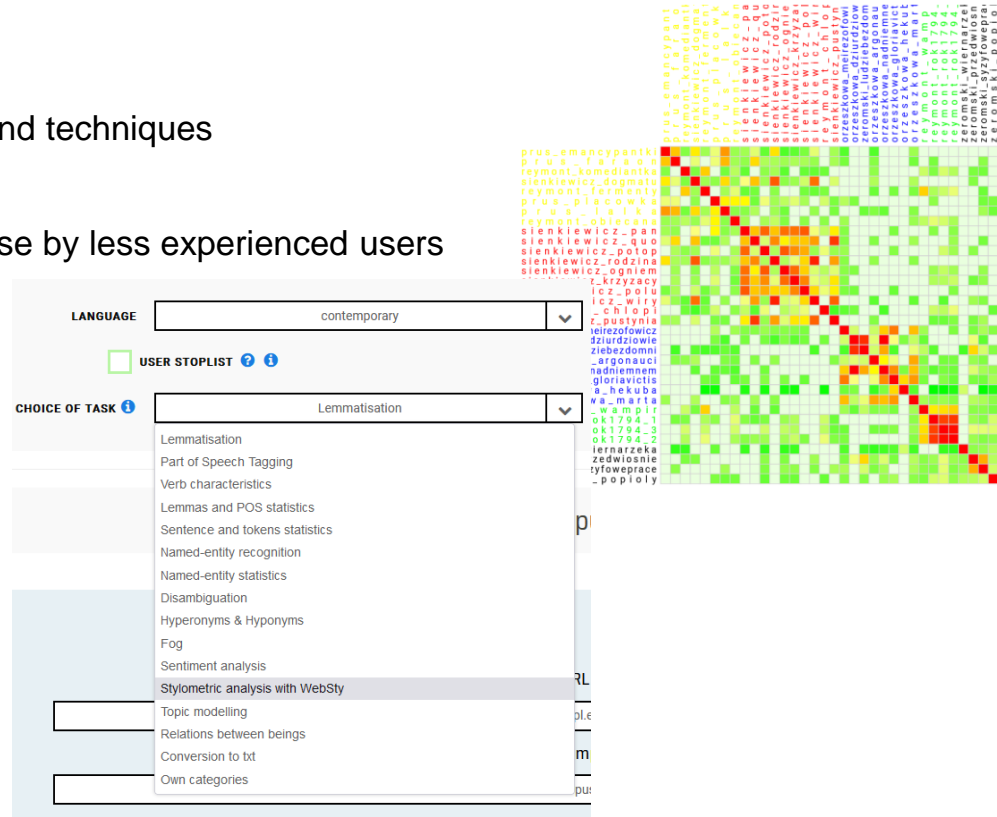
Source: Wikimedia Commons

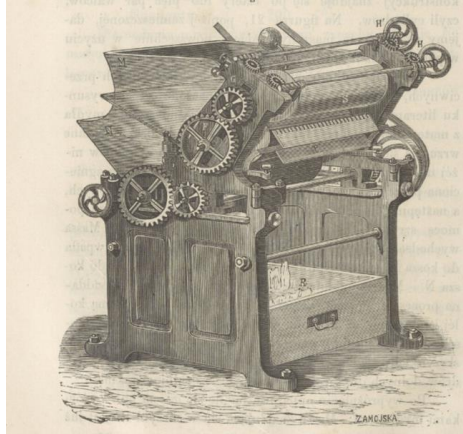
Pros

- wide, interdisciplinary range of research tools and techniques
- modular structure open to new applications
- CLARIN-PL tools in a simplified form, easy to use by less experienced users
- simple, three-step workflow
- possibility to apply default parameter settings or adjust them to the user's needs
- suitability for modern and 19th-century texts

Cons

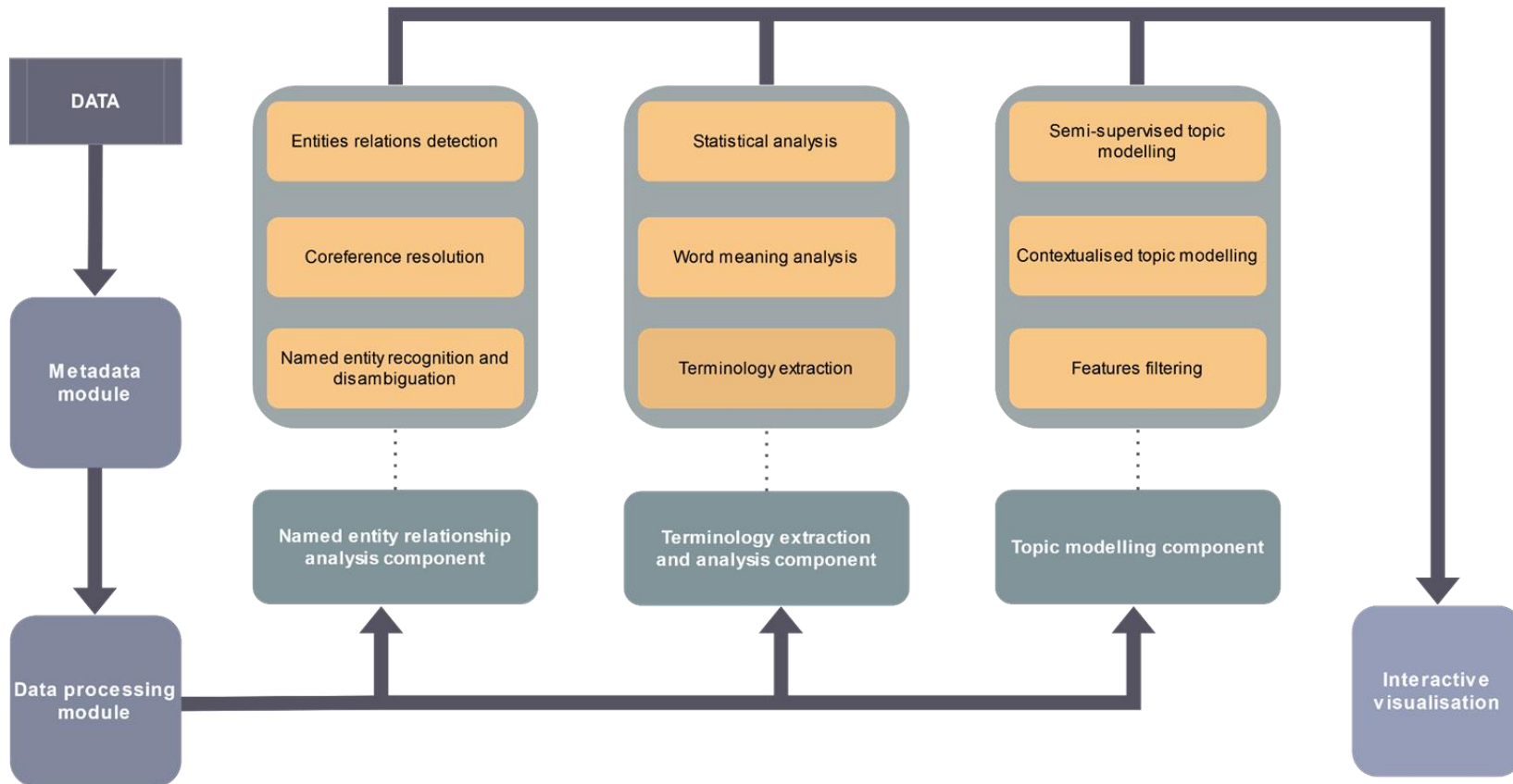
- black box
- no link between texts and metadata
- no integrated data processing
- limited size of files to process



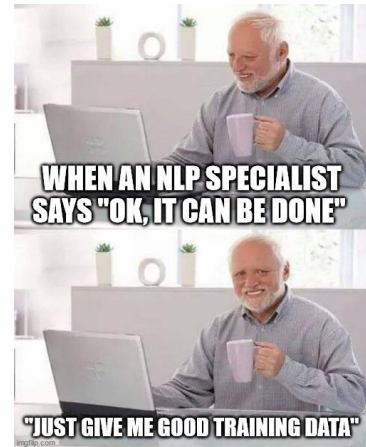


<https://polona.pl/item/cuda-przemyslu-opisanie-wazniejszych-wytworow-nowoczesnego-przemyslu-podlug,ODg4MzEwNTY/207/#item>

- bringing the state-of-the-art NLP solutions offered by **CLARIN-PL** with the resources developed at **Dariah.lab**
- creating a **comprehensive workflow**
- **close cooperation** between the research community and the IT professionals
- combining quantitative and qualitative **validation**
- tailoring to the specificity of **literary studies** and the current debates and trends in the **humanities research**
- adaptation to process **scientific papers**
- allowing for **user intervention** in sub-processes of the processing pipeline
- strong emphasis on **metadata** (filtering/grouping texts based on metadata)
- strong emphasis on **visualization** (e.g. graphs, maps, time series)



pisarze zacierają ślady przemian, dobrzy — właśnie je ujawniają. Po raz pierwszy zdarzyło mi się czytać **Brunona Schulza** w ostatniej klasie liceum. W tym samym czasie odkrywałem też muzykę **Gustawa Mahlera**. Odtąd dzieła tych twórców łączą się osobliwie w mojej świadomości, wzajemnie się przywołując lub odwołując do siebie. Na przykład, **Sklepy cynamonowe**, mają dla mnie za to i naturalne dopełnienie Pierwszą symfonię **Mahlera** z 1896 roku. Zwiastują jej trzecią część — pamiętną piosenką dziecięcia „**Ojciec Marcin**” — przemieniaczy się niepostrzeżenie w marsz pogrzebowy — zdaje się oddawać owo charakterystyczne dla **Schulza** rozpięcie między światem życia — dzieciństwem, a jego zmierzchem i śmiercią; ich ukryte podobieństwo, ich zagadkowy wspólny mianownik — punkt zero egzystencji. Chodzi mi jednak o dzieło pisarza z **Drohobycza**, a jednocześnie o to, że moje skojarzenia z młodością, choć niby zrodzone z przypadku, znajdują głębsze uzasadnienie w materii samych dzieł. Mam na myśli coś, co możnaby nazwać „pogańską aurą”, którą odnajduję w dziełach obu twórców, lecz którą u **Schulza** odkryłem poprzez **Mahlera**. Zanim zajmę się kwestią Metamorfozy, chciałbym się chwilę zatrzymać nad tym właśnie aspektem sztuki **Schulza** — aspektem, na który trudno zwraca się uwagę. Zacząć trzeba od tego, że o duchu pogańskim wspominał sam **Schulz** w napisanym po niemiecku **placuncie** „w 1937 roku”. **Expose** o książce **Brunona Schulza** **Sklepy Cynamonowe** — w szkicu mającym utworować drogę wydaniu książki we **Włoszech**. Pisał tam, między innymi: „Autor czuje się to bliższy odczutu Antyku sądzi, że swoją kreację, swoje fantazjowanie i czucie wywiłdł z pogańskich ogórnictwa życia”. Na warzach mieszkańców **Drohobycza** oślepiające słońce sierpnia maluje „barbarzyńską maskę kultu pogańskiego”. Tym znakiem roz- poczyna się wszystko. Przemiany, rozbudzone upałem, dają ujście „niechujnej, babkiej bujności sierpnia”. I kiedy „nikt nie wiedział, 149 że tam właśnie odprawiał sierpień tego lata swoją wielką, pogańską orgię”, skretylnia Ilują „podobna o bożka pogańskiego” — go ochrypla od krzyku, w konwulsji dzikiej uderza mięśnistym łonem z wściekłą zapalczywością w pierś bzu d z i k i e g o , zakłany całym tym nędzarskim chórem wynaturzonej, pogańskiej i plodności”. Rozbuchana przyroda antropomorfizuje się: rośliny, zwierzęta i dzwime istoty ludzkie, uchwycone w antycznym rytuale o silnych akcentach erotycznych, w nieustannej orgii życia i śmierci, migają jak w kalejdoskopie — jakby z wierszy **Bolesława Leśmiana**, z jego **Przemian**. Celebrantami tych rytuałów „zbieżnych z Naturą” są przede wszystkim jednostki ułonne fizycznie lub psychicznie. Podobnie jak u **Leśmiana**, nędzarc bez nóg z **Zalotów** — okaleczony żołnierz z wiersza pod tym tytułem, stary kulus z **Ballady dziadkowski** i miętny garbus, tak i u **Schulza** — jak zauważył **Czesław Karkowski** — takie postaci jak **Edzio**, **Dodo**, czy **Ilują** bliższe są Naturze niż inni, normalni mieszkańcy **Drohobycza**, a przez to bardziej podatne na jej przemiany. **Henry Levi** w swoim studium o **Schulzu** stawia pytanie, czy bożka (w pogańskim, naturalnie, rozumieniu boskości). I odpowiada: „Tutaj **Schulz** nie jest zdecydowany dostosowanie się do natury jest pogaństwem odnalezionym wraz z pre-Adamową niewinnością; czy przedchodnie podczas kanikuły nie mają mask bożków?” Mam tu do czynienia z wielkim rytualnym świętem, gdy **Chaos** odnajduje szalony Ład we wszystkich przejawach rzeczywistości, co symbolizuje postać bożka **Pana**. To temu właśnie poświęcony jest jeden z centralnych rozdziałów **Sklepów cynamonowych**. W opowiadaniu pod tytułem **Pan** towarzyszy przemianie ogrodu w „paroksyzm szaleństwa, wybuch wściekłości, cyniczny bezwstyd i rozpustę”. A spośród wysokich lipianów wyłania się przed bohaterem istota o „warzy włóczęgi lub pijaka (...) Pan bez flutu cofający się do swych ojczystych kniei”. Obraz barokowego bogactwa traw, roślin i kwiatałów zjawia i przekształca się w świat sam dla siebie — w wegetalną zmysłowość, gdzie element meksi miesza się z zeńskim, gdzie triumfuje dżeworództwo i zapytanie. To obóstwienie Natury, znane nam dzięki świadectwom antycznej kultury greckiej, przygotowuje pojawienie się centralnej postaci w twórczości **Schulza** — postaci ojca, **Janusa**, sprzedawcy fantasma- gonyrczych materiałów — maga, kabalisty, natchnionego kapłana pogańskiej religii, zanurzonej i spełnionej w Naturze. To postać dło- 150 niżjska, w sensie, jaki nadaje temu **Nietzsche** w **Narodzinach tragedii** — to bóg upojenia, który uruchamia spektakł ostatecznego upadku człowieka, by przemienić go w ekstatyczny akceptacją życia we wszystkich jego przejawach. Schulz Dzieciństwo i Mił utęgają ułożeniu: Dla starożytnych genealogia własnego



pisarze zacierają ślady przemian, do- brzy — właśnie je ujawniają. Po raz pierwszy zdarzyło mi się czytać **Brunona Schulza** w ostatniej klasie liceum. W tym samym czasie odkrywałem też muzykę **Gustawa Mahlera**. Odtąd dzieła tych twórców łączą się osobliwie w mojej świadomości, wzajemnie się przywołując lub odwołując do siebie. Na przykład, **Sklepy cynamonowe**, mają dla mnie za to i natural- ne dopełnienie Pierwszą symfonię **Mahlera** z 1896 roku. Zwiastują jej trzecią część — pamiętną piosenką dziecięcia „**Ojciec Marcin**” — prze- mieniaczy się niepostrzeżenie w marsz pogrzebowy — zdaje się odda- wać owo charakterystyczne dla **Schulza** rozpięcie między światem życia — dzieciństwem, a jego zmierzchem i śmiercią; ich ukryte podobień- stwo, ich zagadkowy wspólny mianownik — punkt zero egzystencji. Chodzi mi jednak ? a e dzieło pisarza z **Drohobycza**, a jednocześnie o to, że moje skojarzenia z młodością, choć niby zrodzone z przypadku, znajdują głębsze uzasadnienie w materii samych dzieł. Mam na myśli coś, co możnaby nazwać „pogańską aurą”, którą odnajduję w dziełach obu twórców, lecz którą u **Schulza** odkryłem poprzez **Mahlera**. Zanim zajmę się kwestią Metamorfozy, chciałbym się chwilę zatrzy- mać nad tym właśnie aspektem sztuki **Schulza** — aspektem, na któ- ry rzadko zwraca się uwagę. Zacząć trzeba od tego, że o duchu pogańskim wspominał sam **Schulz** w napisanym po niemiecku, w 1937 roku, **Expose** o książce **Brunona Schulza** **Sklepy Cynamonowe** — w szkicu mającym utworować drogę wydaniu książki we **Włoszech**. Pisał tam, między innymi: „Autor czu- je się to bliżsi odczutu Antyku, sądzi, że swoją kreację, swoje fanta- zjowanie i czucie wywiłdł z p o g a Ń s k i e g o o p o j m o w a n i a j a c i a”. Na warzach mieszkańców **Drohobycza** oślepiające słońce sierpnia maluje „barbarzyńską maskę kultu pogańskiego”. Tym znakiem roz- poczyna się wszystko. Przemiany, rozbudzone upałem, dają ujście „niechujnej, babkiej bujności sierpnia”. I kiedy „nikt nie wiedział, 149 że tam właśnie odprawiał sierpień tego lata swoją wielką, p o g a Ń s k i e g o , zakłany całym tym nędzarskim chórem wynaturzonej, p o g a Ń s k i e i plodności”. Rozbuchana przyroda antropomorfizuje się: rośliny, zwierzęta i dzwime istoty ludzkie, uchwycone w antycznym rytuale o silnych akcentach erotycznych, w nieustannej orgii życia i śmierci, migają jak w kalejdoskopie — jakby z wierszy **Bolesława Leśmiana**, z jego **Przemian**. Celebrantami tych rytuałów „zbieżnych z Naturą” są prze- de wszystkim jednostki ułonne fizycznie lub psychicznie. Podobnie jak u **Leśmiana**, nędzarc bez nóg z **Zalotów** — okaleczony żołnierz z wiersza pod tym tytułem, stary kulus z **Ballady dziadkowski** i pa- miętny garbus, tak i u **Schulza** — jak zauważył **Czesław Karkowski** — takie postaci jak **Edzio**, **Dodo**, czy **Ilują** bliższe są Naturze niż inni, normalni mieszkańcy **Drohobycza**, a przez to bardziej podatne na jej przemiany. **Henry Levi** w swoim studium o **Schulzu** stawia pytanie, czy **Tutaj** jest szalona, czy bożka (w pogańskim, naturalnie, rozumieniu boskości). I odpowiada: „Tutaj **Schulz** nie jest zdecydowany; dostosowanie się do natury jest pogaństwem odnalezionym wraz z pre-Adamową nie- winnością; czy przedchodnie podczas kanikuły nie mają mask boż- ków?” Mam tu do czynienia z wielkim rytualnym świętem, gdy **Chaos** od- najduje szalony Ład we wszystkich przejawach rzeczywistości, co symbolizuje postać bożka **Pana**. To temu właśnie poświęcony jest je- den z centralnych rozdziałów **Sklepów cynamonowych**. W opowia- daniu pod tytułem **Pan** towarzyszy przemianie ogrodu w „paroksyzm szaleństwa, wybuch wściekłości, cyniczny bezwstyd i rozpustę”. A spośród wysokich lipianów wyłania się przed bohaterem istota o „warzy włóczęgi lub pijaka (...) **Pan** bez flutu cofający się do swych ojczystych kniei”. Obraz barokowego bogactwa traw, roślin i kwiatów zjawia i przekształca się w świat sam dla siebie — w wegetalną zmysłowość, gdzie element meksi miesza się z zeńskim, gdzie trium- fuje dżeworództwo i zapytanie. To obóstwienie Natury, znane nam dzięki świadectwom antycznej kultury greckiej, przygotowuje pojawienie się centralnej postaci w twórczości **Schulza** — postaci ojca, **Janusa**, sprzedawcy fantasma- gonyrczych materiałów — maga, kabalisty, natchnionego kapłana pogańskiej religii, zanurzonej i spełnionej w Naturze. To postać dło- 150 niżjska, w sensie, jaki nadaje temu **Nietzsche** w **Narodzinach tragedii** — to bóg upojenia, który uruchamia spektakł ostatecznego upadku człowieka, by przemienić go w ekstatyczny akceptacją życia we wszy- stkich jego przejawach. Schulz Dzieciństwo i Mił utęgają ułożeniu: Dla starożytnych genealogia własnego

- author
- literary work
- fictional character
- place



- **Different levels of user experience**

- basic – users interested in processing a small number of texts and qualitative exploration of results
- advanced – users interested in the processing of larger corpora and further work in other tools (e.g. Gephi, neo4j)
- expert – users interested in using preprocessed data to test innovative methods and workflows

- **Theoretical framework**

- “travelling” concepts in the humanities
- historical semantics
- cultural analysis
- philosophy and history of science
- philosophy of science
- cultural transformation analysis



identification of constellations of terms and concepts and the comparison and visualisation of their flows over time, between literary and academic communities, within a single discipline, between different disciplines or from literary theory to artistic practice



<https://polona.pl/item/alegoria-matematyki,MTAzNDYwMTcw/0/#info:metadata>



<https://polona.pl/item/aeusserliche-figur-einer-grossen-kunst-uhr,MzI3ODIzMDg/>

GoLEM will allow for research aimed at addressing the following issues:

- showing how the **reception of a given author** has evolved
- mapping the discourse: reconstructing the **geographical horizon** of the collection of texts
- reconstructing and analysing the **terminology networks** of selected scholars
- grasping **differences in the understanding of key concepts** between scholars
- analysing the **semantic field** of a selected literary term
- detecting words with **usage change** across corpora
- tracing how the **topics** of selected researchers' papers have changed over time
- analysing the **links between authors** based on the topics discussed in their papers

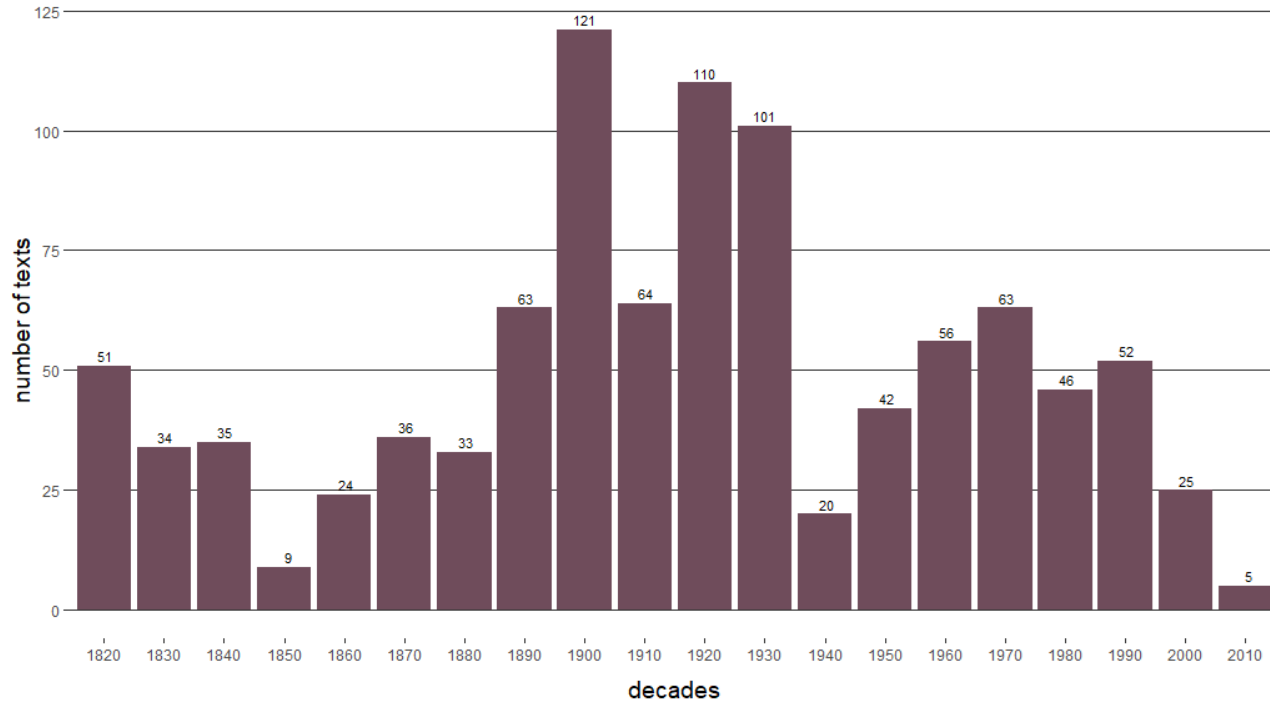


<https://polona.pl/item/aeusserliche-figur-einer-grossen-kunst-uhr,MzI3ODIzMDg/>

GoLEM will allow for research aimed at addressing the following issues:

- showing how the **reception of a given author** has evolved
- mapping the discourse: reconstructing the **geographical horizon** of the collection of texts
- reconstructing and analysing the **terminology networks** of selected scholars
- grasping **differences in the understanding of key concepts** between scholars
- analysing the **semantic field** of a selected literary term
- detecting words with **usage change** across corpora
- tracing how the **topics** of selected researchers' papers have changed over time
- analysing the **links between authors** based on the topics discussed in their papers

1. Scope: Polish literary studies discourse from 1822 to 2022
2. Patchwork corpus, created cumulatively/iteratively and consisting of three, related versions:
 - a. **KDL 1.0** (the initial, opportunistic corpus)
 - i. subcorpus of anthologies
 - ii. subcorpus of persons of literary studies
 - iii. subcorpus of journals
 - iv. subcorpus of monographs
 - b. **KDL 2.0** (the proper corpus): represents a selection from KDL 1.0 and is created by balancing and assessing the representativeness of the materials from the subcorpora
 - c. **KDL OA**: represents a selection from KDL 1.0 and contains texts that can be shared with a wider audience
3. In-depth and monitor corpus
4. Grounds for text selection: established expert knowledge (principle of authority), contained in handbooks, anthologies, selected writings of literary scholars and syllabuses of academic courses, citation analysis



GoLEM in action: mapping the literary studies discourse

Agnieszka Karlińska

ORCID: 0000-0002-4846-7086

✉ agnieszka.karlinska@ibl.waw.pl

🐦 @a_karlinska

Wiktor Walentynowicz

ORCID: 0000-0002-7672-8642

✉ jan.wieczorek@pwr.edu.pl

Jan Wieczorek

ORCID: 0000-0001-8709-7754

✉ wiktor.walentynowicz@pwr.edu.pl

Maciej Maryl

ORCID: 0000-0002-2639-041X

✉ maciej.maryl@ibl.waw.pl

🐦 @maciejmaryl

Thank you!

