



**HAL**  
open science

## The benefits of increasing the digital availability of Alsatian theater

Pablo Ruiz Fabo, Carole Werner, Delphine Bernhard

► **To cite this version:**

Pablo Ruiz Fabo, Carole Werner, Delphine Bernhard. The benefits of increasing the digital availability of Alsatian theater. Digital Humanities 2022, Jul 2022, Tokyo (Online), Japan. pp.567-570, 10.5281/zenodo.7014965 . hal-03660481

**HAL Id: hal-03660481**

**<https://hal.science/hal-03660481>**

Submitted on 22 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

from the Archives of the National Center for Biological Sciences (NCBS).

There is sparse research that evaluates the socio-cultural undertones of an organization evidenced through its annual reports (Jagolinzer, 2021). Although the key purpose of annual reports is to provide public-domain disclosure of the operations and financial activities of the organization over the year, they are also opportune sites to understand institutional biases and silences. In focusing on the annual reports of NCBS over a 25-year period this project draws attention to this underappreciated genre that is often described as the most common medium for communicating strategies and plans, past performance, and future expectations of an organization (Srinivasan and Marques, 2017). In particular for research institutions, annual reports not only provide periodic research updates but also documents the subterranean and hegemonic policies, framing research frameworks and decisions. The analytical lens employed in this study employs the intersecting domains of Digital Humanities, Gender Studies, and STS (Science and Technology Studies). We employ CTDA (Critical Techno-cultural discourse analysis) as the guiding methodology to identify and recover the gendered institutional silences, biases, and underappreciated female subjectivities which often remain elided within normative research on and about scientific institutions. The CTDA triad of *technology*, *technology in practice*, and *socio-cultural beliefs* is operationalized with digital archives, the act of archiving and hegemonic masculinity representing the three facets respectively. With the use of this methodology attention is drawn towards the ways women scientists are perceived in institutional archival spaces.

By articulating significance and interpretation of archives in digital spaces, we examine the discourse of gender representation in digitally archived objects such as these annual reports. Further, to qualitatively assess the issue of women's representation in science, we operationalize techno-feminism (combining STS and Feminism) asserting that the technological advancements and social circumstances are not mutually exclusive. With an emphasis to understand gendered institutional silences, we also supplement our analysis with interviews of women scientists from NCBS. Our intervention combines a longitudinal computational analysis of digitized annual reports from the NCBS archive over a 25-year period with qualitative interviews to exemplify the potential of Critical Digital Humanities: to not only shed new light on the gendered and unexplored facets of scientific research in Global South spaces but also develop unique methodological approaches for insightful methods of knowledge creation in DH.

## Bibliography

- Brock, A. (2016). Critical technocultural discourse analysis: *Sage Publication*, 20(3), 1012–1030.
- Jagolinzer, A. (2021). Annual reports should inform society – not only those with a financial interest.
- Jasanoff, S. (2004). The Idiom of Co-Production. In *States of Knowledge: The Co-Production of Science and Social Order* (pp. 1–12). Taylor & Francis.
- Srinivasan, P., & Marques, A. (2017). *Narrative analysis of annual reports: A study of communication efficiency* (No. 486).
- Wajeman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics*, 34(1), 143–152.

## The benefits of increasing the digital availability of Alsatian theater

### Ruiz Fabo, Pablo

ruizfabo@unistra.fr  
Université de Strasbourg, France

### Werner, Carole

werner@unistra.fr  
Université de Strasbourg, France

### Bernhard, Delphine

dbernhard@unistra.fr  
Université de Strasbourg, France

Extensive efforts have been made to create digital corpora for major European dramatic traditions. A recent corpus hub is DraCor (Fischer & Börner, 2019), offering programmatic access to theater collections in 11 languages.

The creation of such corpora was accompanied by computational literary research on their traditions. Focusing on German and French drama (immediately relevant for our work), one may name projects like DLINA (Fischer et al., 2017),<sup>2</sup> QuaDramA and Q:TRACK (Reiter, 2020)<sup>3</sup> or *Emotions in Drama* (cf. Schmidt et al., 2021), for German; for French, studies like Schöch's (2017), or works in the *Revue Historiographique du Théâtre*'s special issue on Digital Humanities and theater (Galleron, 2017).<sup>4</sup>

*Alsatian* refers to Germanic varieties from Alsace (Eastern France), the main oral communication language

in the area for over 14 centuries, until their decline vs. French in the last third of the 20th century (Huck et al., 2007: 11–17). Despite Alsatian’s mainly oral status, it has a rich theatrical tradition, where comedic and popular genres predominate (Gall, 1974). Unlike major European traditions, digital resources for Alsatian theater are scarce, which precludes related computational literary research. To help make such research feasible, our project <sup>5</sup> is building a large TEI-encoded corpus of Alsatian theater, annotated with rich metadata, and making this corpus publicly available in formats appropriate for research and for the non-specialist community, covering mostly the 1870-1940 period. Our proposal discusses the relevance of developing such materials, in terms of opportunities for (computational) literary studies (CLS) and natural language processing (NLP), besides their potential to promote linguistic diversity through the digital medium.

Our corpus both complements and challenges existing knowledge of Alsatian theater. Earlier studies on the corpus period (Cerf, 1972, 1975; Huck, 1998, 2005; Hülsen, 2003) examined plots, plays’ social groups, and dialect representation in character speech, based on corpora with under 40 plays. Annotating the characters’ socio-professional groups in 109 plays (Ruiz Fabo and Werner, 2021) revealed the importance of artisans in plays set in small towns, contrary to claims in earlier studies, which presented agricultural professions as almost exclusive in such settings (Cerf, 1972: 340). The corpus also has potential for comparisons with the two hegemonic dramatic traditions that surround Alsatian theater: French and German. Hülsen (2003: 98 ff.) described the influence of German popular comedy in Alsatian theater, based on thirteen major Alsatian plays around 1900. Our corpus could be used for a larger-sample comparison, covering a longer period, with German and French plays from the same timeframe. A concrete first aspect to compare could be the *dramatis personae*, following our work on character annotation, cited above.

Digital literary collections for under-resourced languages also benefit NLP. Current NLP models like contextual embeddings (e.g. Devlin et al., 2018) require large amounts of text, yet unavailable for languages like Alsatian. However, the NLP community is interested in improving the viability of these technologies for low-resource languages, and in developing corpora in such languages (cf. Bernhard et al., 2019). Increasing the availability of Alsatian electronic text contributes to this effort and is a test-bed for new technology developments.

Towards our goals, we selected a varied corpus mixing major and lesser-known works. We performed optical character recognition (OCR) on National Library digitizations, <sup>6</sup> creating 25 TEI plays (300,000 tokens). OCR was also manually corrected for 28 further plays, awaiting

TEI conversion. Bibliographical metadata were transcribed for 359 plays, and the settings and *dramatis personae* for 257 plays (2,253 characters). We also annotated characters’ social variables when possible (gender, professional group, estimated social class, age). Synopses were written for 38 plays. TEI materials are publicly available <sup>7</sup> and accepted at the DraCor platform. Publication with a DOI on open repository Nakala was also carried out. <sup>8</sup>

A corpus exploration interface exposes all data. <sup>9</sup> Reiter et al. (2017) warn that a graphical user interface does not always help literary research. However, our project intends to promote public interest in Alsatian beyond research, and we believe that a web interface is a suitable means. Our interface gives access to all plays with corrected OCR output, TEI-encoded or not, searching by author, author gender, dates or publisher. It also allows filtering the corpus based on the presence of characters with different social attributes (professional group, social class, gender, see Figure 1). We believe that such navigation workflows are an attractive way for the public to engage with the corpus and with the social picture it conveys, and will assess this hypothesis with a user sample. Character co-occurrence is also relevant for research on plot and for sociolinguistic studies on character speech representation according to social variables.

Regarding limitations, male authors predominate, even if we sought variety. There are only five female authors, with 14 plays. Deeper research in non-digitized collections should counter this bias. Note also that sexism and racism are present in the corpus, and, even if the interface targets the general public, it does not at this point provide means that would help non-specialists be aware of these biases and critically question such social representations. This is important future work.

In sum, we present a digital corpus of plays in Alsatian varieties with rich metadata, including character annotations for social variables. TEI encoding is ongoing; where not available, corrected OCR output is provided. This material helps CLS research on Alsatian theater, impossible before given lack of a corpus, and has shown potential to challenge existing knowledge on the tradition. The texts also have relevance for NLP research on low-resource languages. To promote interest in the corpus’ language varieties, a corpus exploration interface opens up these resources to the community.



**Character cooccurrences in cast list per professional group. Top: Plays by female authors. Bottom: Plays by male authors. Proportion of female characters with profession in cast list is 45.71% with female authors and 11.30% with male authors.**

**Acknowledgements**

We thank further project members, Dominique Huck and Pascale Erhart, for advice on corpus selection and other project aspects. We also thank all interns who contributed to corpus or interface development: Nathanaël Beiner, Andrew Briand, Lena Camillone, Hoda Chouaib, Audrey Deck, Barbara Hoff, Valentine Jung, Salomé Klein, Audrey Li-Thiao-Te, Kévin Michoud and Vedisha Toory.

**Bibliography**

**Bernhard, D., Bras, M., Erhart, P., Ligozat, A.-L. and Vergez-Couret, M.** (2019). Language Technologies for Regional Languages of France: The RESTAURE Project. *International Conference Language Technologies for All (LT4All): Enabling Linguistic Diversity and Multilingualism Worldwide*. Paris, France <https://hal.archives-ouvertes.fr/hal-02418928> (accessed 14 January 2020).

**Cerf, E.** (1972). Le théâtre alsacien de Strasbourg, miroir d’une société (1898-1939). *Saisons d’Alsace*(43).

**Cerf, E.** (1975). Les contes merveilleux du théâtre alsacien de Strasbourg. *Revue des sciences sociales de la France de l’Est*, **4**: 3–30.

**Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.** (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

*ArXiv:1810.04805 [Cs]* <http://arxiv.org/abs/1810.04805> (accessed 6 December 2020).

**Fischer, F. and Börner, I.** (2019). Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama. *Digital Humanities 2019*. Utrecht, p. 5.

**Fischer, F., Göbel, M., Kampkaspar, D. and Kittel, C.** (2017). Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts. *Digital Humanities 2017*. Montreal, pp. 437–41.

**Gall, J.-M.** (1974). *Le théâtre populaire alsacien au XIXe siècle*. (Recherches et Documents XIX). Strasbourg: Istra.

**Galleron, I. (ed).** (2017). *Revue d’Historiographie Du Théâtre, Numéro 4 : Études Théâtrales et Humanités Numériques*.

**Huck, D.** (1998). D’r Herr Maire (1898) de Gustave Stoskopf. Entre ethnologie et littérature : les Alsaciens en auto-représentation. *Recherches Germaniques*, **28**: 163–90.

**Huck, D.** (2005). Le «Théâtre Alsacien de Strasbourg» et la production dramaturgique de ses fondateurs (1898-1914). *Culture et Histoire Des Spectacles En Alsace et En Lorraine : De l’annexion à La Décentralisation (1871-1946)*. Peter Lang, pp. 198–222.

**Huck, D., Bothorel-Witz, A. and Geiger-Jallet, A.** (2007). L’Alsace et ses langues. Eléments de description d’une situation sociolinguistique en zone frontalière. In Abel, A., Stuflessler, M. and Voltmer, L. (eds), *Aspects of Multilingualism in European Border Regions: Insights and Views from Alsace, Eastern Macedonia and Thrace, the Lublin Voivodeship and South Tyrol*. Bozen/Bolzano: EURAC Research (Europäische Akademie / Accademia Europea / European Academy), pp. 13–101 <http://ala.u-strasbg.fr/documents/Publication%20-%20L%27Alsace%20et%20ses%20langues.pdf>.

**Hülßen, B. von** (2003). *Szenenwechsel im Elsass: Theater und Gesellschaft in Strassburg zwischen Deutschland und Frankreich 1890-1944*. Leipziger Universitätsverlag.

**Reiter, N.** (2020). Möglichkeiten Quantitativer Dramenanalyse. *Comparatio. Zeitschrift für Vergleichende Literaturwissenschaft*, **12**(2): 39–52 doi: [10.1007/978-3-476-04516-4\\_16](https://doi.org/10.1007/978-3-476-04516-4_16). <https://comp.winter-verlag.de/article/COMP/2020/2/7> (accessed 7 December 2021).

**Reiter, N., Kuhn, J. and Willand, M.** (2017). To GUI or not to GUI?. *INFORMATIK 2017*. Gesellschaft für Informatik <https://dl.gi.de/bitstream/handle/20.500.12116/3880/B14-8.pdf?sequence=1&isAllowed=y> (accessed 13 November 2019).

**Ruiz Fabo, P. and Werner, C.** (2021). Exploration du théâtre alsacien à travers ses listes de personnages pendant la période 1870-1940. *Humanistica 2021*. Rennes

(online): Zenodo doi: [10.5281/zenodo.4762733](https://doi.org/10.5281/zenodo.4762733). <https://doi.org/10.5281/zenodo.4762732> (accessed 8 December 2021).

**Schmidt, T., Dennerlein, K. and Wolff, C.** (2021). Emotion Classification in German Plays with Transformer-based Language Models Pretrained on Historical and Contemporary Language. *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Punta Cana, Dominican Republic (online): Association for Computational Linguistics, pp. 67–79 <https://aclanthology.org/2021.latechclfl-1.8> (accessed 7 December 2021).

**Schöch, C.** (2017). Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. *Digital Humanities Quarterly*, **011**(2).

## Notes

1. <https://dracor.org/doc/credits> lists corpora and languages. Our first 25 TEI-encoded Alsatian plays were also accepted in DraCor.
2. <https://dlina.github.io/>
3. <https://quadrama.github.io/>
4. *Journal for Theater Historiography*. Special issue: <https://sht.asso.fr/revue/etudes-theatrales-et-humanites-numeriques/>
5. MeThAL, “Towards a macroanalysis of theater in Alsatian” (<https://methal.pages.unistra.fr/en.html>)
6. Numistral portal, BNU Strasbourg: <https://numistral.fr/fr/theatre-alsacien>
7. <https://git.unistra.fr/methal/methal-sources>
8. Nakala set <https://nakala.fr/collection/10.34847/nkl.feb4r8j9> contains our data
9. <https://methal.eu/ui/>

## Modeling Music for Musicologists: A Linked Open Data Approach

### Saccomano, Mark

mark.saccomano@uni-paderborn.de  
Paderborn University, Center for Music, Edition, Media (ZenMEM)

### Shibata, Elisabete

shibata@beethoven.de  
Beethoven-Haus Bonn, Research Centre “Beethoven-Archiv”

### Lewis, David

david.lewis@oerc.ox.ac.uk  
University of Oxford e-Research Centre

### Hankinson, Andrew

andrew.hankinson@rism.digital  
Répertoire International des Sources Musicales (RISM) Digital Center

### Page, Kevin

kevin.page@oerc.ox.ac.uk  
University of Oxford e-Research Centre

Modeling work in the digital humanities has traditionally focused on written texts; music, however, requires data models that can capture the varied, overlapping layers that are characteristic of its structure. Most encoding models for notated music—i.e., those presenting a precise representation of what can be performed by a musician—provide good coverage of the layers on the immediate music ‘surface,’ like measures, staves, and notes. Other, more analytical and less apparent structures are often not as well addressed. This includes formal objects such as musical themes and motifs, as well as properties that often lack explicit means for symbolic notation, such as texture and timbre. Our paper describes a model that includes a component to specifically address these types of musical structure, using different arrangements of the same musical work as examples.

Musical figures require descriptions that include both the beginning and end points that mark their extent, as well as a specification of the individual parameters lying at different structural layers that comprise that figure. For example, early arrangements of Beethoven’s Eighth Symphony contain a replacement of the novel triple forte dynamic marking with a simple fortissimo, or double forte, at the moment of the first movement’s recapitulation. We can point to the respective measures and state that one has a different dynamic marking than the other, but this does not include our identification of the measures as different expressions of the same music-theoretical structure: a recapitulation. Although digital annotations can reference passages in multiple works, these references do not make sense unless they are linked together as separate expressions of the same analytic object. In order to present such a comparison, we need to posit a distinct, abstract class to model it. Such parallelism is entailed in our common understanding of what an arrangement is, though as Flanders and Jannidis (2021) note, the modeling already inherent in such a term is ‘invisible through [its] very familiarity.’ Before beginning work on a data model,