

Comparative metadata for popular theater in Alsatian, German and French

Pablo Ruiz Fabo, Fanny Boisnard, Alexia Schneider - Université de Strasbourg, France

Project goals and methods

We describe ongoing work to create metadata for popular theater subgenres in Alsatian, German and French.¹ Alsatian theater (written in Germanic varieties spoken in Alsace, Eastern France) has been influenced by both German and French theater, and large-scale metadata will help compare the traditions. As we work with lesser-studied subgenres, the data adds variety to Computational Literary Studies (CLS).²

We transcribe and annotate the following:

- Bibliographic metadata based on the frontmatter (title, author, publisher and its location, publication year for our source), also searching for the first edition and premiere year
- Plays' genre based on their subtitle; if absent, secondary literature is consulted
- The *dramatis personæ*, annotating variables to describe characters (gender, social class, socioprofessional group, etc.) besides character relations
- The plays' setting, including time and place information, retrieving coordinates from OpenStreetMap when possible
- Number of acts and scenes
- We retrieve Wikidata IDs for plays, authors and publishers; adding missing ones to Wikidata is a project goal

For French, we work on vaudeville, a comedic subgenre; corpus selection is based on Gidel (1986) and Matthes (1983). For German, we work on two popular comedy subgenres called *Posse* and *Schwank* (Wilms, 1969; Klotz, 2007). These three subgenres have been identified as the main influences of Alsatian theater (Hülßen, 2003). For Alsatian, our corpus includes *Schwank* and more refined comedy types, besides about 25 popular dramas (*Volksstück*), ca. 25 Christmas tales (*Weihnachtsstück*), and a handful of serious dramas. Data includes 6,834 characters from 589 plays by 200 authors, between 1800 and 1929.

Rationale and related work

Alsatian theater is at the crossroads of French and German drama (Hülßen, 2003), but no systematic comparison has been performed between Alsatian and these traditions. We recently created a prosopography for Alsatian characters (Ruiz & Bermúdez, 2022), obtaining findings (Ruiz et al., 2024) that complement earlier knowledge of Alsatian theater, based on smaller samples (Cerf, 1972; Gall, 1974; Huck, 2005; Hülßen, 2003). Since the corpus overview afforded by such annotations proved useful, here we extend the methodology to German and French popular subgenres, to explore subgenre evolution across traditions.

One of our inspirations is Wiedmer et al. (2020), who insist on the value of the *dramatis personæ* as an indicator of possible plots and conflicts. DraCor Einakter (Çakir & Fischer, 2020) also develops similar annotations to ours, covering one-act plays in German (1740-1850). More generally, the literary polysystems approach (Even Zohar, 1990), which

¹ The work is carried out within the Thealtres project: <https://thealtres.pages.unistra.fr/>

² Data for this submission is at <https://git.unistra.fr/thealtres/dh2024-data>

pays attention to interactions between peripheral traditions (Alsatian in our case) and core ones (German and French) is also an inspiration (see López-Iglésias, 2010; Rybicki & Woźniak, 2018 for other polysystems applications). We approach popular, lesser-studied historical subgenres, including non-canonical authors. Such work is valuable as it adds variety to CLS. As studies by Jannidis et al. (2019) suggest, the richness of popular literature can be underestimated by critical literature based on small samples.

Corpus trends

Table 1 shows current annotation volume (at ca. 60% of the targeted corpus, i.e. ca. 300 plays per tradition), and figure 1 shows its temporal distribution; this is still uneven across traditions and will be made more homogeneous.

Table 2 shows the socioprofessional groups we defined, drawing from professional taxonomies like HISCO (van Leeuwen et al., 2004), but modifying them to reflect groups of potential dramatic relevance.³ The colors reflect a more abstract grouping, and figure 2 displays the distribution of these more abstract groups. The largest groups are elementary workers (often domestic workers) and the bourgeoisie (often the former’s employers), besides authority figures; disputes between these groups can be a source of (comedic) conflict in the plays. The aristocracy is more present in the other traditions than in Alsatian (trend already identified by Hülsen, 2003), as are artists. In Alsatian, agriculture professions are more present than in the other traditions, likely due to the presence of popular dramas (not just comedies) in the corpus. This is also a likely reason for a skew towards lower classes in Alsatian (figure 3).

Limitations and outlook

The study has several limitations. First, genre is mainly determined by “self-identification” (subtitles); the subgenre-label coherence could be verified. Second, there is a large gender imbalance in authors. Countering this would require archive searches to find women writers absent from printed and digitized sources. Third, the study is limited to characters described with a socioprofessional group or relation in the *dramatis personæ* (42% to 76% depending on the tradition).

We will complete transcription and annotation in 2024. Thanks to representing plays as feature vectors, we will attempt statistical modeling to identify the most defining characteristics of subgenres diachronically among the features under study. By approaching lesser-known subgenres and lesser-studied language varieties, the dataset adds empirical variety to CLS.

Table 1: Annotation volume

	plays	authors	author gender (f, m, u) [*]	characters	character gender (f, m, u) [*] (%)	characters per play
als	302	78	4, 69, 5	2,959	26, 68, 6	9,8
ger	143	50	1, 48, 1	2,329	34, 59, 7	16,27
fre	144	72	9, 62, 1	1,546	37, 59, 4	10,74

^{*} Female, male, unknown

³ As for annotation reliability, two of the authors doubly annotated the French subcorpus characters, with an overlap of 93% for fine-grained socioprofessional groups and 84.5% for social classes.

Annotated plays per decade and tradition

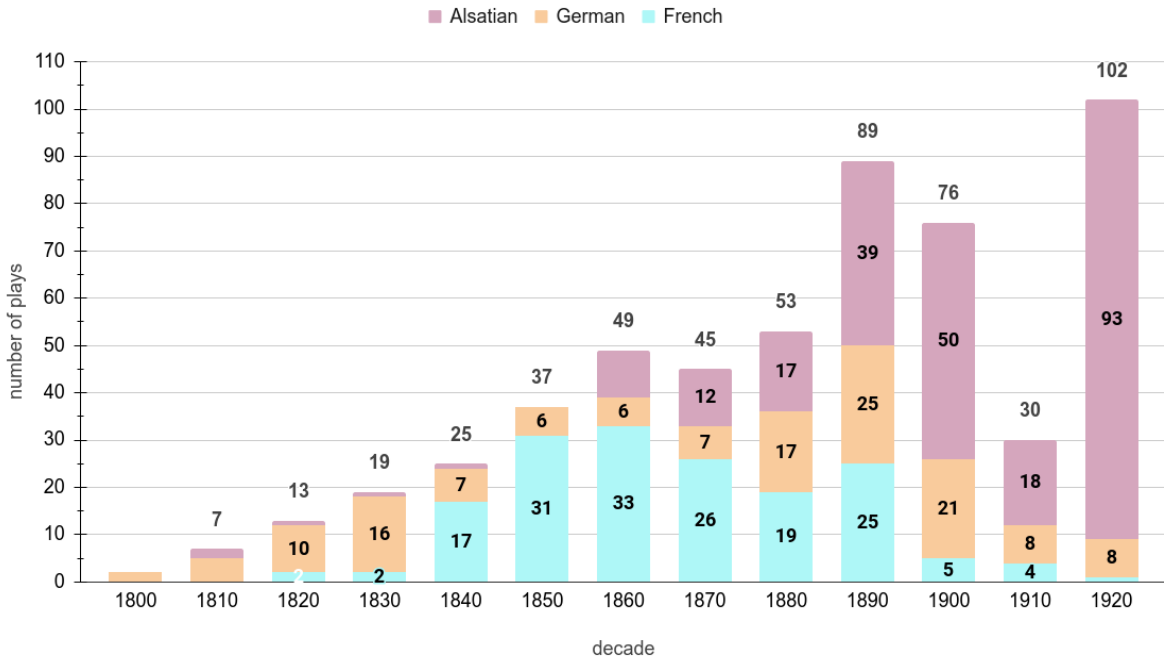
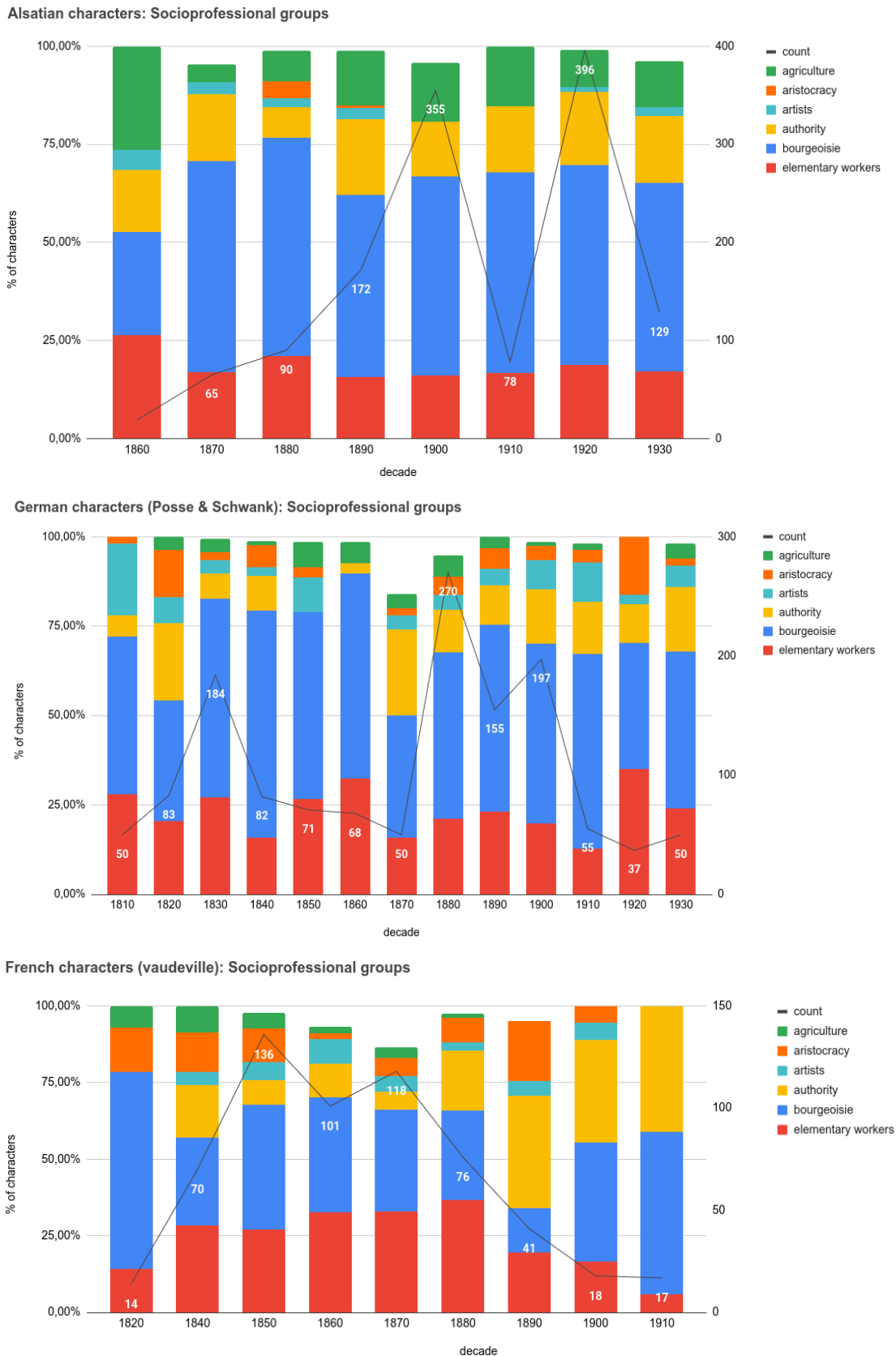


Figure 1: Plays temporal distribution

Table 2: Socioprofessional groups

fine-grained classification			coarse classification
1. agriculture	7. government officials	13. police & security	1. agriculture
2. aristocrats	8. hobbies	14. professionals, scientific, technical	2. aristocracy
3. artists	9. industry & transportation	15. rentiers	3. artists
4. clergy	10. intermediate professions	16. service & sales	4. authority figures
5. crafts	11. military	17. unknown	5. bourgeoisie
6. elementary professions	12. outlawed	18. volunteer positions	6. elementary professions
			7. other

Figure 2: Coarse-grained socioprofessional group distribution of characters per tradition, diachronically; Alsatian (top), German (mid), French (bottom)



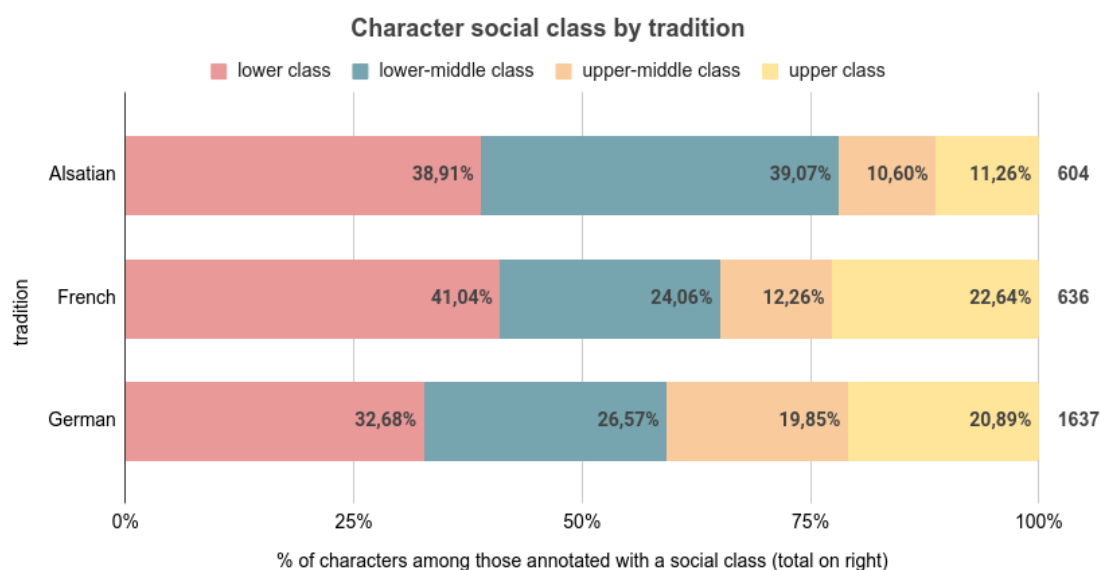


Figure 3: Character social class distribution in the three traditions

Acknowledgements

We thank other project members (Benjamin Krautter and Lara Nugues) for early discussions on corpus selection and annotation criteria.

This work was supported by MISHA (Maison des Sciences de l'Homme - Alsace)

References

- Çakir, D. C., & Fischer, F.** (2022). Dramatische Metadaten: Die Datenbank deutschsprachiger Einakter 1740-1850. *Digital Humanities Im Deutschsprachigen Raum*. DHd 2022.
- Cerf, E.** (1972). Le théâtre alsacien de Strasbourg, miroir d'une société (1898-1939). *Saisons d'Alsace*, 43.
- Even-Zohar, I.** (1990). Polysystem theory. *Polysystem Studies [=Poetics Today]*, 11(1), 9–26.
https://www.tau.ac.il/~itamarez/works/books/Even-Zohar_1990--Polysystem%20studies.pdf
- Gall, J.-M.** (1974). *Le théâtre populaire alsacien au XIXe siècle*. Istra.
- Gidel, H.** (1986). *Le vaudeville*. Presses universitaires de France.
- Huck, D.** (2005). Le «Théâtre Alsacien de Strasbourg» et la production dramaturgique de ses fondateurs (1898-1914). In J.-M. Leveratto, J. Benay, O. Thomas, & S. Wuttke, *Culture et histoire des spectacles en Alsace et en Lorraine: De l'annexion à la décentralisation (1871-1946)* (pp. 198–222). Peter Lang.
- Jannidis, F., Konle, L., & Leinen, P.** (2019). Makroanalytische Untersuchung von Heftromanen. *Digital Humanities Im Deutschsprachigen Raum*, 167–173.
- López-Iglésias Samartim, R.** (2010). *O processo de construção do sistema literário galego entre o franquismo ea transição (1974-1978)*. *Margens, relações, estrutura e*

estratégias de planificação cultural. PhD Thesis, Universidade de Santiago de Compostela.
https://minerva.usc.es/xmlui/bitstream/handle/10347/2858/1/9788498874549_content.pdf

- Ruiz, P., & Bermúdez Sabel, H.** (2022). Feature structures for character social variable annotation and an application to Alsatian theater. *Text Encoding Initiative 2022 Conference (TEI 2022)*. [Virtual poster session].
<https://doi.org/10.5281/zenodo.7110070>
- Ruiz, P., Bernhard, D., Briand, A., & Werner, C.** (2024). Computational drama analysis from almost zero electronic text. In M. Andresen & N. Reiter (Eds.), *Reflecting Computational Drama Analysis: Reflecting Methods and Implementations*. De Gruyter. <https://univoak.eu/islandora/object/islandora%3A157880>
- Rybicki, J., & Woźniak, M.** (2018). Polysystem theory and macroanalysis. A case study of Sienkiewicz in Italian. *Digital Humanities 2018*. Digital Humanities 2018, Mexico City. <https://dh2018.adho.org/en/polysystem-theory-and-macroanalysis-a-case-study-of-sienkiewicz-in-italian/>
- Van Leeuwen, M. H. D., Maas, I., & Miles, A.** (2004). Creating a Historical International Standard Classification of Occupations An Exercise in Multinational Interdisciplinary Cooperation. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 37(4), 186–197. <https://doi.org/10.3200/HMTS.37.4.186-197>
- von Hülsen, B.** (2003). *Szenenwechsel im Elsass: Theater und Gesellschaft in Strassburg zwischen Deutschland und Frankreich 1890-1944*. Leipziger Universitätsverlag.
- Wiedmer, N., Pagel, J., & Reiter, N.** (2020). Romeo, Freund des Merkurio: Semi-Automatische Extraktion von Beziehungen zwischen dramatischen Figuren. *DHd2020 Spielräume. Digital Humanities Zwischen Modellierung Und Interpretation. Konferenzabstracts.*, 194–200. <https://doi.org/10.5281/zenodo.3666689>
- Wilms, B.** (1969). *Der Schwank Dramaturgie und Theatereffekt: Deutsches Trivialtheater 1880-1930*. Freie Universität Berlin.