



HAL
open science

Editing and Attributing Musical Texts: the Chansonnier du Roi and the MARITEM Project

Jean-Baptiste Camps, Christelle Chaillou, Viola Mariotti, Federico Saviotti

► To cite this version:

Jean-Baptiste Camps, Christelle Chaillou, Viola Mariotti, Federico Saviotti. Editing and Attributing Musical Texts: the Chansonnier du Roi and the MARITEM Project. EADH2021: Interdisciplinary Perspectives on Data, 2nd International Conference of the European Association for Digital Humanities, Krasnoyarsk, 2021, 2021, Krasnoyarsk, Russia. halshs-03260116

HAL Id: halshs-03260116

<https://shs.hal.science/halshs-03260116>

Submitted on 14 Jun 2021

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Editing and Attributing Musical Texts: the *Chansonnier du Roi* and the MARITEM Project

Jean-Baptiste Camps¹

Christelle Chaillou²

Viola Mariotti²

Federico Saviotti³

1. École nationale des chartes, Université Paris Sciences & Lettres
2. CESCO / CNRS, Poitiers
3. Università degli Studi di Pavia

This contribution examines the case of the *Chansonnier du Roi*, a very important 13th century lyrical manuscript, in the context of the ongoing MARITEM project. It presents a workflow for the edition and stylometry of musical texts, from text acquisition and encoding to the analysis of text and musical notations.

1 The *Chansonnier du Roi* and the MARITEM Project

Compiled during the second half of the 13th century, the *Chansonnier du Roi* (MS Paris, BnF, fr. 844) contains 602 lyrical compositions from different musical and literary traditions: profane songs of French *trouvères* and Occitan *troubadours*, French motets, instrumental works and Latin sacred compositions. Moreover, shortly after the original compilation, some additional pieces also from multiple origins (French *rondeaux* and *motets entés*, Occitan *dansas* and *descortz*) were transcribed in many of the blank pages and columns. Assembling several different repertoires in a uniform plan (further developed through additions), the *Chansonnier du Roi* is an ideal resource for the study not only of different and multilingual lyric traditions *per se*, but also of their unitary reception in the late 13th century Gallo-Romance area. This, together with a small but significant presence of lyrics from the 14th century, explains why we chose this manuscript as the object of our research.

The first purpose of the MARITEM project (ANR) is to produce a dataset and digital edition of text and music, encoded in XML/TEI and XML/MEI. If the digital scholarly edition of texts is a well established practice, the field of edition of medieval music is on the other hand relatively new. The *Corpus Monodicum* project (Haug and Puppe, 2020) developed a software, Monodi+, for the musical transcription of medieval latin songs in MEI. We plan to work with the software with a few adaptations and optimisations with the cooperation of the *Corpus Monodicum* team. The first challenge will be to link the textual edition (TEI) and the musical edition (MEI). In this aspect, the project is pioneer. The choice to work with only one manuscript was made with the perspective to create a prototype and to develop a methodology for the edition and the indexation of all musical *chansonniers* in different medieval languages (mainly Old French,

Old Occitan and Old German). The edition of music and text will prove very useful to both musicologists and philologists, as both aspects will be addressed with equal accuracy.

It is our belief that involving both complementary constituents of medieval lyric songs will bring new results about the history of the codex, the languages, the link between musical composition and language, the history and link between the different traditions, and notably the attribution of the songs.

2 Data pipeline

Acquisition of the text is done with a pipeline that aims to fully integrate the contributions of human and artificial intelligence, in the spirit of digital philology (Andrews, 2012). It builds on the workflow developed for another 13th century French manuscript, the *Légendier* BnF, fr. 412 (Camps, Clérice and Pinche, 2020).

Layout analysis was performed with Transkribus (Kahle, 2017) default model, with some success for text regions and baselines (estim. F measure for baselines according to Transkribus: 0.71). Zones for music and illumination were added and typed manually. We plan to train a more specific model in the future to better detect musical notations and illumination.

The **handwritten text recognition** was performed using a model trained on data from MS fr. 412, with good results concerning the main hand (CER around 8%, WER 25%). The prediction was then fully corrected by a human expert, inside Transkribus. In the next steps, we aim to reuse and adapt the pipeline for automatic text segmentation, normalisation and lemmatisation (cf. Camps, Clérice and Pinche, 2020).

For now, the **melodies** are human-transcribed, but we hope to be able to train a model for music recognition.

3 Towards a scholarly digital edition

The future edition of the *Chansonnier du Roi* will be an electronic and interactive edition where it will be possible to consult and to query two different textual levels at the same time: the first one will be the allographic and graphematic transcription of the manuscript (Robinson et Solopova 1993; Stutzmann, 2011; Camps, 2016), while the second one will be the normalized edition. Both levels will be accompanied by high definition images of all the pages of the manuscript.

Both transcriptions, allographic and graphematic, are conceived to reflect different aspects of the text as it was written by medieval copyists. An allographic transcription is a transcription whose goal is to “give access to every form of every letter or sign” (Stutzmann, 2011). We followed the recommendations of the Medieval Unicode Font Initiative (Haugen, 2015) in order to reproduce the formal Medieval letter variants and abbreviations marks. Allographic transcriptions respect as well the Medieval word segmentation, and punctuation. We decided to transcribe allographically about one page and a half per copyist (about 20 copyists for the entire manuscript) in order to study the very special *usus copianti* of each scribe of the *Manuscrit du Roi*, and appreciate the very different hands, both French and Italian, which have composed this

precious manuscript. Compared to the former, graphematic transcriptions simply normalise variant letter forms.

The second level of the electronic edition of the *Chansonnier du Roi* is the editorial one, focused on the normalised edition of the lyrical texts, in order to propose a text approachable even for non specialised readers. It modernises the modern word segmentation, use of punctuation, accents and capital letters, and of course the resolution of all the abbreviation marks.

The edition follows the guidelines of the TEI (TEI Consortium, 2020) for the text and MEI for musical notations (Music Encoding Initiative, 2020). The melodies are transcribed with the software Monodi+ (Eipert and al., 2019). The musical transcription is easier than the text because the notation is clear and simple. The music edition has two levels: a diplomatic transcription with the keys, the signs and the presentation of the manuscript; and a modernised transcription with a G key and a verse alignment.

4 Stylometric analysis of text and music

The availability of a complete transcription of this manuscript makes possible the stylometric analysis of the songs of the *trouvères* and *trouveresses*, at a level impossible until now. This is a critical issue, because disputed attributions are very numerous inside the Old French Lyrical tradition (Gatti, 2019), yet it poses specific challenges to both traditional and stylometric approaches:

1. individual components are very short (fig. 1);
2. lyrical idiolect, based on a shared and elitist cultural tradition, is rather homogeneous, apparently leaving little space for personal features, though such a situation is not unheard of in the stylometry of Medieval and Early Modern texts (e.g. Camps and Cafiero, 2013; 2019), but still poses significant challenges.
3. attribution of the text and of the melody have rarely been addressed as a whole. Indeed, the musical elements (for example: musical curve, modality, intervals), in connection with the texts could be an important contribution for the knowledge of songs attributions.
4. the manuscript tradition creates noise, with both linguistic and substantial variants due to subsequent copy steps upstream of the manuscript which are quite difficult to retrace.

An automatic quantitative analysis of the scriptological data (Goebel, 1975; Dees, 1987), such as graphic and phonetic allotropes in different texts attributed to the same author or in different authorial corpora, will provide useful hints in the linguistically stratified French *scriptae* of the two main copyists. Thus, not only will it be possible to try to better determine their origins (it's already acknowledged that one of them is generally Italian, but no previous study could determine with any plausibility the origin of the other, responsible for the main part of the *chansonnier*), but also to recover valuable insights about the linguistic habits of the authors themselves.

Specifically, the *Chansonnier du Roi* offers important documentation on the works of Thibaut de Champagne, perhaps the most prominent of all *trouvères*

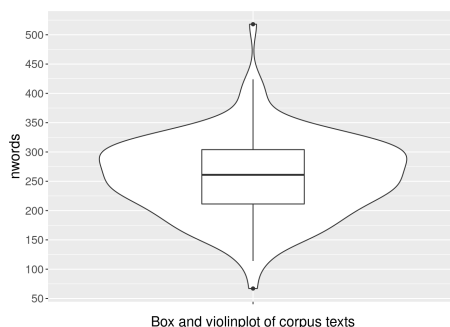


Figure 1: Distribution of the length in words of the texts

(Barbieri, 1999). The attribution of several songs to Thibaut is still disputed (Wallensköld, 1925; Callahan, 2010).

In order to give new insights into these disputed attributions, we performed several stylometric analyses, using features robust to noise and short text length, in particular character 3-grams (Camps, Clérice and Pinche, 2020). Both exploratory and supervised analyses, the latter using SVM, were performed to shed more light on the attribution of these components.

For instance, we trained models on a corpus of 140 songs to distinguish Thibaut’s hand from a group of contemporary *trouveres* (Gace Brulé, Gautier de Dargies and Blondel de Nesle), with a leave-one-out approach. While the models do not attain a perfect accuracy (global 80.6%), precision reaches 100% for attribution to Thibaut (0.46% recall; table 1). These preliminary results (that we plan to substantially extend for the conference) seem to confirm the attribution to Thibaut of two very famous songs, *Ausi com l’unicorne sui* (Linker 240,3; RS 2075) and *Li dous penser et li dous souvenir* (Linker 240,35; RS 1469) (table 2).

	precis	recall	F_1	GT \ Pred	Not Thib.	Thib
Not Thib.	76.79	1.00	0.87	Not Thib.	86	0
Thibaut	1.00	0.46	0.63	Thibaut	26	22

Table 1: Metrics and confusion matrix for the leave-one-out training

title	RS	Thibaut
<i>Quant fine Amours me proie que je chant</i>	306	
<i>Sans atente de gueredon</i>	1867	
<i>Dame, li vostres fins amis</i>	1516	
<i>Ausi com l’unicorne sui</i>	2075	X
<i>Tres haute amours, qui tant s’est abessie</i>	1098	
<i>Li dous penser et li dous souvenir</i>	1469	X

Table 2: Model results for a sample of disputed pieces

Concerning musical stylometry, all the composers seem to integrate some little elements consciously or unconsciously in their compositions. A recent study observed that the troubadour Bernard de Ventadorn used in most of this songs

three ascending notes at the beginning of verses (Chaillou-Amadiou 2016). The percentage calculation and the comparison with a witness corpus highlighted that it was really a stylistic trait by Bernard. Conversely the absence of this element suggests that a few melodies could have been created by others. Such methodology can be automated and show promise for the joint stylometry of text and music.

References

- Andrews, T.** (2012). The third way: philology and critical edition for a digital age. *Variants: The Journal of the European Society for Textual Scholarship*, **10** <http://boris.unibe.ch/43071/>.
- Barbieri L.** (1999). Note sul 'Liederbuch' di Thibaut de Champagne, *Medioevo romanzo*, **23**, pp. 388-416.
- Barton L. W. G.** (2002). The neumes project: digital transcription of medieval chant manuscripts. In *Second International Conference on Web Delivering of Music, WEDELMUSIC 2002*. Darmstadt, p. 211–218, doi: 10.1109/WDM.2002.1176213.
- Callahan, C.** (2010). Thibaut de Champagne and Disputed Attributions: The Case of MSS Bern, Burgerbibliothek 389 (C) and Paris, BnF fr. 1591(R). *Textual Cultures*, **5**(1). Indiana University Press: 111–32 doi: 10.2979/tex.2010.5.1.111.
- Camps, J.-B.** (2016). La 'Chanson d'Otinél': édition complète du corpus manuscrit et prolégomènes à l'édition critique. Paris: Paris-Sorbonne PhD thesis, dir. Dominique Boutet, doi: 10.5281/zenodo.1116735. <https://halshs.archives-ouvertes.fr/tel-01664932>.
- Cafiero, F. and Camps, J.-B.** (2019). Why Molière most likely did write his plays. *Science Advances*, **5**(11). American Association for the Advancement of Science: eaax5489 doi: 10.1126/sciadv.aax5489.
- Camps, J.-B. and Cafiero, F.** (2013). Setting bounds in a homogeneous corpus: a methodological study applied to medieval literature. *Revue Des Nouvelles Technologies de l'information (RNTI)*, **SHS-1**: 55–84, <https://halshs.archives-ouvertes.fr/halshs-00765651/>.
- Camps, J.-B., Clérice, T. and Pinche, A.** (2020). Stylometry for Noisy Medieval Data: Evaluating Paul Meyer's Hagiographic Hypothesis. *ArXiv:2012.03845 [Cs]* <http://arxiv.org/abs/2012.03845> (accessed 8 January 2021).
- Chaillou-Amadiou, C.** (2016). Le chant du poète, entre innovation et tradition. In *L'expression de l'identité dans la poésie lyrique romane, entre texte et musique*, ed. F. Saviotti. Milan: Presses universitaires de Milan, p. 101-114.
- Chaillou-Amadiou, C.** (2017). Philologie et musicologie. Les variantes musicales dans les chansons de troubadours. In *Les noces de Philologie et de musicologie. Textes et musiques du Moyen Âge*, ed. C. Cazaux-Kowalski and al. Paris: Classiques Garnier, 2017, p. 69-95.
- Haug, A. and Puppe, F.** (2020). *Corpus monodicum: Die einstimmige Musik des lateinischen Mittelalters, Digitale Ausgabe, Alpha-Version*, Mainz and Würzburg, <https://corpus-monodicum.de>.
- Dees, A.** (1987). *Atlas des formes linguistiques des textes littéraires de l'ancien français*, Tübingen, Niemeyer.
- Eipert T. and al.** (2019). Editor Support for Digital Editions of Medieval Monophonic Music. In *Proceedings of the 2nd International Workshop on Reading*

Music Systems. Delft, <https://sites.google.com/view/worms2019/proceedings>.

Gatti, L. (2019). Repertorio delle attribuzioni discordanti nella lirica trovierica, Roma, Sapienza Università Editrice, http://www.editricesapienza.it/sites/default/files/5850_Gatti_Lirica_trovierica_interior_0A.pdf).

Goebel, H. (1975). Qu'est-ce que la scriptologie? *Medioevo romanzo*, 2, pp. 3-43

Haugen, O. E. (ed). (2015). *MUFI Character Recommendation v. 4.0*. Medieval Unicode Font Initiative <https://bora.uib.no/bora-xmlui/handle/1956/10699> (accessed 8 January 2021).

Kahle, P., Colutto, S., Hackl, G. and Mühlberger, G. (2017). Transkribus-A service platform for transcription, recognition and retrieval of historical documents. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 4. IEEE, pp. 19–24.

Lavrentiev, A. (2007). Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies : actes de la journée d'étude de l'ENS LSH, 6 juin 2005. Chambéry: Université de Savoie, 2007.

Lyrik des deutschen Mittelalters, database, <http://www.ldm-digital.de>

Linker. Robert White Linker, *A Bibliography of Old French Lyrics*, University, MS: Romance Monographs, 1979.

Music Encoding Initiative (2020). *Guidelines*. Mainz <https://music-encoding.org/> (accessed 8 January 2021).

Refrain, database, <http://refrain.ac.uk>

Robinson, P. and Solopova, E. (1993). Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue. *The Canterbury Tales Project Occasional Papers*, 1: 19–52 <<http://www.canterburytalesproject.org/pubs/transguide-MI.pdf>>.

RS. *G. Raynauds Bibliographie des altfranzösischen Liedes* neu bearbeitet und ergänzt von Hans Spanke, Leiden: Brill, 1955.

Stutzmann, D. (2011). Paléographie statistique pour décrire, identifier, dater... Normaliser pour coopérer et aller plus loin?. In Franz, F., Fritz, C. and Vogeler, G. (eds), *Kodikologie Und Paläographie Im Digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2*. (Schriften Des Instituts Für Dokumentologie Und Editorik, 3). Norderstedt, pp. 247–77 <https://halshs.archives-ouvertes.fr/halshs-00596970/>.

TEI Consortium (2020). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. <https://tei-c.org/Guidelines/> (accessed 9 May 2015).

Troubadour Melodies Database, <http://www.troubadourmelodies.org/melodies>

Wallensköld, A. (1925), ed. Thibaud IV (1201-1253 ; comte de Champagne): *Les chansons de Thibaut de Champagne, roi de Navarre*. (Ed.) , A. <https://gallica.bnf.fr/ark:/12148/bpt6k53236> (accessed 8 January 2021).

Wick C., Puppe, F. (2019). OMMR4all - a Semiautomatic Online Editor for Medieval Music Notations. In *2nd International Workshop on Reading Music Systems WoRMS*.