

Les formats d’encodage de la notation musicale à l’épreuve d’un objectif de transcription manuscrite automatique

Hugo Scheithauer¹ , Gonzalo Romero-García² , Laurent Romary¹ , and Thibault Clérice¹ 

¹ Inria, ALMAAnaCH, Paris, France

² EPITA Research Laboratory, Image Processing and Pattern Recognition, Paris, France

Abstract

Le développement de la reconnaissance automatique de partitions (*Optical Music Recognition*, OMR) se heurte à une fragmentation des formats de représentation, répartie entre exigences éditoriales (MEI), impératifs de rendu graphique et de sémantique (MusicXML) et analyse computationnelle (Humdrum ***kern*). Cette tension est particulièrement saillante pour les sources manuscrites, où la sémantique musicale se double d’une matérialité documentaire complexe. S’appuyant sur les récents progrès de l’ATR (Automatic Text Recognition), cet article évalue la capacité des standards actuels à servir de vérités de terrain. En prenant pour cas d’étude les manuscrits de piano de Claude Debussy, nous défendons une approche graphématique de la transcription. Nous justifions le détournement de MusicXML vers un modèle « structurel » capable de préserver, dans une certaine mesure, l’ancrage physique et les phénomènes non-standards de la source. Cette proposition vise à participer à la stabilisation des pratiques d’annotation pour la création de jeux de données mutualisables, condition essentielle au développement d’un OMR historique robuste.

Mots-clés: reconnaissance optique de musique, transcription automatique de partitions manuscrites, transcription, jeu de données, standards, notation musicale occidentale

Keywords: optical music recognition, handwritten music recognition, transcription, dataset, standards, common western music notation

1 Introduction

Le développement des technologies de transcription automatique de texte (ou *Automatic Text Recognition*, ATR) a permis d’ouvrir de nouvelles voies d’accès à grande échelle pour les documents textuels, fruit d’un effort collectif de la communauté scientifique. Si les bibliothèques numériques offrent aujourd’hui un accès massif au patrimoine textuel, le patrimoine musical écrit n’est pas en reste. À titre d’exemple, la plateforme de la Bibliothèque nationale de France (BnF), Gallica¹, héberge un corpus de 76 105 documents rassemblés sous l’appellation « partitions », qu’il s’agisse de pièces uniques ou de recueils.

Le pendant de l’ATR pour la notation musicale est l’*Optical Music Recognition* (OMR). Cette tâche regroupe les méthodes visant à transformer automatiquement une image de partition imprimée ou manuscrite en une représentation exploitable par une machine. L’OMR se confronte à de nombreux défis inhérents à la nature de la notation musicale [9] : la non-linéarité de la notation, la superposition et les relations entre symboles, la richesse du vocabulaire et l’imbrication possible entre notation musicale et texte.

Hugo Scheithauer, Gonzalo Romero-García, Laurent Romary, and Thibault Clérice. “Les formats d’encodage de la notation musicale à l’épreuve d’un objectif de transcription manuscrite automatique.” *Actes de la Conférence Humanistica*, éd. par Serena Crespi, Simon Gabay, Martin Grandjean, Ariane Pinche, Marie Puren et Léa Saint-Raymond. Vol. 4. Anthology of Computers and the Humanities. 2026, 108–119. <https://doi.org/10.63744/K2fvu4Nzsv6t>.

© 2026 par les auteurs. Sous licence Creative Commons Attribution 4.0 International (CC BY 4.0).

1. Voir le site internet de Gallica. Consulté le 12/01/26.

Par conséquent, les formats de représentation divergent naturellement de ceux utilisés pour le texte. L'encodage de la notation musicale est marqué par une tension entre trois approches : la représentation des documents, portée par des systèmes favorisés par la musicologie et les humanités en général comme le format XML de la *Music Encoding Initiative* (MEI) [24]; des systèmes de représentation de la musique pour la partager et à destination des personnes qui la jouent avec MusicXML [16]; et des représentations fonctionnelles de la seule sémantique musicale avec le format Humdrum ***kern* [18], destiné à l'analyse musicologique computationnelle. Cette division entre formats et objectifs de représentation illustre l'opposition entre données d'usage et données d'entraînement, et les standards actuels, hérités de l'édition ou de la musicologie computationnelle, peinent à concilier précision documentaire, description graphique et validité musicale.

Après un état de l'art sur l'encodage de la musique et les corpus destinés à l'OMR, nous évaluerons d'abord la capacité des trois principaux formats cités précédemment à servir de vérité de terrain. Nous justifierons ensuite le choix de MusicXML comme compromis opérationnel, en précisant les limites inhérentes à ce format lorsqu'il est employé dans une transcription graphématique de source manuscrite.

Nos contributions sont les suivantes :

- **Évaluation critique des formats d'encodage** : une analyse comparative des formats MEI, MusicXML et Humdrum ***kern* sous l'angle spécifique de la production de données d'entraînement pour l'OMR de partitions manuscrites.
- **Justification méthodologique de MusicXML** : la démonstration du potentiel de MusicXML comme standard de compromis pour décrire une source manuscrite dans le contexte de l'OMR.
- **Ouverture vers l'OMR « de source »** : une proposition de changement de paradigme, passant de la seule reconnaissance du contenu sémantique musical à une transcription respectant la dimension documentaire et l'hétérogénéité des manuscrits.

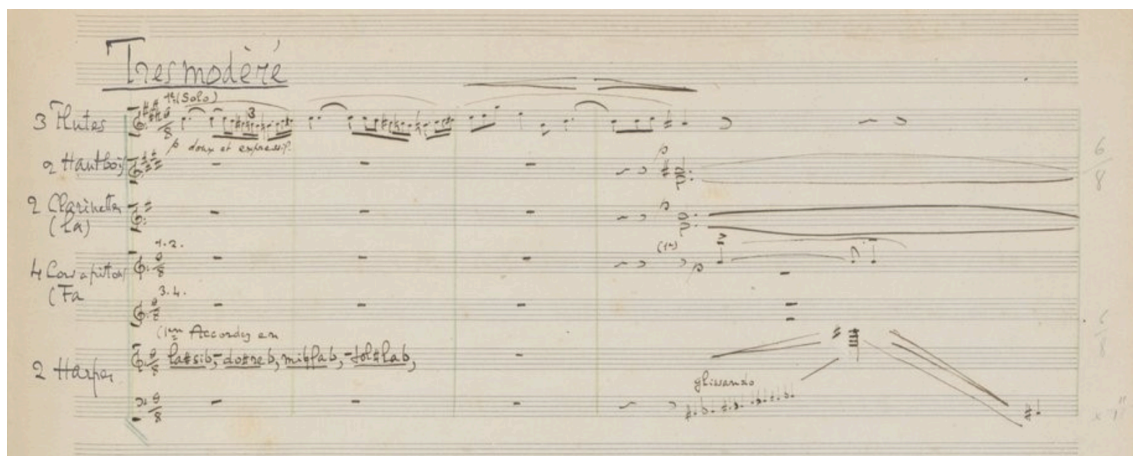


FIGURE 1 – Extrait du *Prélude « à l'après-midi d'un Faune »*, Claude Debussy, 1892-1894, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b55001035h/f9>.

Nous prendrons pour cas d'étude un extrait de la collection des manuscrits de partitions de piano attribués au compositeur Claude Debussy (1862 - 1918) [5], conservés à la BnF et numérisés sur Gallica (fig. 1). Contrairement aux partitions imprimées, ces sources présentent une hétérogénéité graphique s'écartant des standards de la *Common Western Music Notation*² (CWMN) au

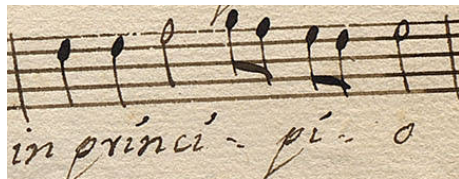
2. Système d'écriture occidentale en vigueur depuis le XVIIIe siècle, sur lequel nous nous concentrerons.

profit de pratiques idiosyncratiques. Le document n’est plus un simple support de contenu musical abstrait, mais le témoin matériel d’un processus créatif (ratures, hésitations, annotations), mettant à l’épreuve les formats actuels face aux variabilités d’une source.

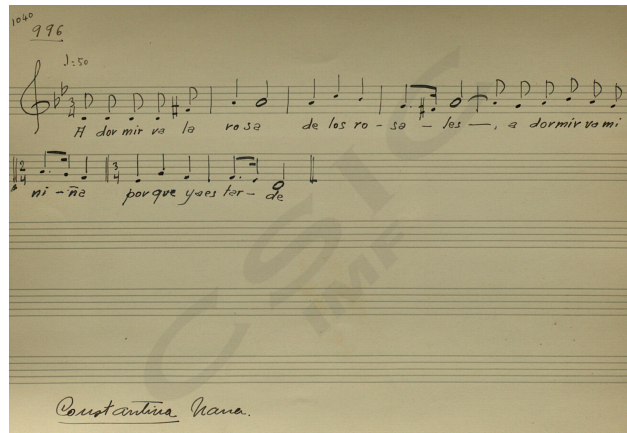
2 Travaux existants

L’adéquation des formats d’encodage de la musique à la transcription automatique reste peu explorée, alors qu’elle définit la nature des informations extraites par les systèmes d’OMR. García-Iasci, Rizo et Calvo-Zaragoza [15] et Castellanos, Gallego et Fujinaga [9] identifient la MEI, MusicXML et Humdrum ****kern** comme les standards prédominants, et soulignent l’efficacité des interfaces de transcription comme MuReT [33] et Sibelius [3] pour l’exportation de ces données.

Toutefois, cette multiplicité de formats témoigne d’une fragmentation du champ de recherche. TORRAS, BISWAS et FORNÉS [38] insiste sur la nécessité d’adopter un standard commun pour lever les freins technologiques de l’OMR et propose le *Music Tree Notation* (MTN). Ce format intermédiaire, basé sur une structure de graphe, vise à fédérer la communauté en définissant les objets qui doivent être ciblés par l’OMR, ouvrant ainsi la voie à la création raisonnée de nouveaux jeux de données permettant l’entraînement et une meilleure évaluation des modèles d’OMR.



(a) Extrait du Pau Llinás dataset [4].



(b) Exemple d’un document provenant du FMT dataset : Óscar David Castillo Pérez, *A dormir va la rosa*, Fondo de Música Tradicional IMF-CSIC, ed. E. Ros-Fábregas (consulté le 12/01/26), <https://musicatradicional.eu/piece/70690>.

FIGURE 2 – Exemples de jeux de données pour l’OMR manuscrit.

TORRAS, BISWAS et FORNÉS [38] souligne également que le manque de jeux de données larges et mutualisables demeure un obstacle majeur. Pour la CWMN manuscrite, et plus spécifiquement historique, les ressources exploitables sont rares, à l’exception du *Pau Llinás dataset*³ [4] (fig. 2). Cette pénurie s’explique par le coût de l’expertise humaine et le caractère chronophage de l’annotation manuelle, poussant notamment une partie de la recherche vers la création de données synthétiques [2 ; 21].

Enfin, les rares corpus structurés en MusicXML ou MEI (SEILS [26], DoReMi [35], PRIMuS [6]) s’avèrent inadaptés à notre étude car ils se concentrent soit sur la musique mensurale⁴

3. Le format d’annotation étant propriétaire, basé sur des chaînes de caractères, et dépourvu de coordonnées spatiales, cela limite fortement son interopérabilité.

4. La musique mensurale, ou notation mesurée, est un système de notation musicale utilisé en Europe entre le XIIIe et XVIIe siècles. Elle a fait l’objet de nombreuses études menées en OMR [8].

[7], soit sur des partitions numériques contemporaines. Si les jeux de données FMT-M et FMT-C [34] semblent plus proches de notre objet d'étude, leur inaccessibilité actuelle nous empêche d'évaluer la pertinence pour de l'OMR orienté vers la source.

3 Quels formats disponibles pour un objectif de transcription automatique de partitions manuscrites ?

3.1 Les trois formats principaux de l'OMR

La recherche actuelle repose sur trois formats dominants pour la CWMN : MusicXML, MEI et Humdrum ****kern**⁵. Si chacun permet de décrire le contenu musical sémantique, ils divergent par leurs priorités : l'échange et le rendu pour MusicXML, l'édition critique pour la MEI et l'analyse computationnelle pour Humdrum ****kern**.

Les formats MusicXML et MEI reposent sur le langage hiérarchique XML. MusicXML s'est imposé comme le standard d'échange entre logiciels d'édition (MuseScore [23], Sibelius), intégrant la représentation du contenu musical et des instructions de mise en page nécessaires au rendu visuel d'une partition. Il ne permet pas nativement de lier le contenu à une source externe, comme pourrait le faire le format ALTO XML pour le texte [1]. À l'inverse, la MEI, conçue sur le modèle de la TEI [37], est dédiée aux éditions⁶. Elle autorise une description fine des phénomènes musicaux et un lien granulaire avec la source originale, mais sa grande flexibilité et sa verbosité la rendent complexe à exploiter pour de l'OMR.

Humdrum ****kern** propose une structure textuelle tabulaire. Sa concision facilite sa tokenisation, ce qui explique sa prédominance dans les modèles d'apprentissage profond récents [11 ; 30 ; 31]⁷. Sa conception privilégie la sémantique musicale au détriment de la topographie documentaire, et, comme le souligne TORRAS, BISWAS et FORNÉS [38], il échoue à représenter la mise en page originale ou les éléments non structurels (groupements, nuances complexes), rendant impossible une restitution fidèle de la source manuscrite.

Nous considérons que deux conditions doivent être remplies pour qu'un format soit adapté à représenter une source dans le cadre d'une transcription basée sur une segmentation préalable : assurer l'ancrage spatial des objets à l'aide de coordonnées (x, y) et permettre leur encodage graphique sans les contraintes d'une grammaire musicale normalisatrice et d'une simplification excessive.

3.2 Fidélité de la représentation de la source et ancrage spatial de la transcription

Plutôt que l'édition scientifique, notre travail vise la création de vérités de terrain pour de l'apprentissage supervisé. En intégrant les phénomènes manuscrits et la matérialité du document, l'OMR pourrait ainsi devenir un levier pour l'histoire de la création musicale.

MusicXML, conçu pour garantir l'interopérabilité, l'échange de partitions structurées et leur reproduction (*engraving*), ne permet pas nativement d'aligner les objets musicaux aux pixels d'une source. Cette caractéristique est pourtant stabilisée en ATR grâce aux formats ALTO et PAGE XML [28], qui ancrent chaque caractère à une zone précise de l'image.

En ce qui concerne la représentation de la sémantique musicale, les schémas de MusicXML, Humdrum ****kern** et MEI sont suffisamment matures pour modéliser l'ensemble des objets séman-

5. Nous avons délibérément écarté certains formats de cette étude, comme Lilypond [25], PAEC [32] ou la ABC notation [39], dont l'usage reste marginal dans les travaux actuels sur l'OMR. D'autres structures de représentation, telles que les graphes (MuNG) [17] n'ont pas été explorées pour cet article. Sans nous avancer davantage, un format tel que MuNG a comme principal défaut de n'être adapté qu'à la représentation de la sémantique du contenu musical.

6. Voir, par exemple, la Digital-Interactive Mozart Edition, publiée par la International Mozarteum Foundation, Salzburg, ainsi que les autres projets listés sur le site officiel de la MEI, consultés le 12 janvier 2026.

7. Concernant la tokenisation des données d'entraînement pour l'OMR, MAYER et al. [22] proposent une linéarisation du format MusicXML pour l'optimiser aux tâches d'apprentissage machine. Cette approche a permis d'introduire le format *Linearized MusicXML* (LMX).

tiques de la CWMN contemporaine : par exemple les notes et leurs différentes propriétés (durée, rythme, hauteur, etc.); ainsi que les ornements, les dynamiques, les clés, les barres de mesure, etc. Comme expliqué par RETTINGHAUS et al. [29], le verrou n'est pas tant dans la richesse des schémas que dans l'incapacité des interfaces de transcription à exploiter pleinement ces structures. La question suivante est de savoir dans quelle mesure un format comme MusicXML peut représenter la matérialité de la source.

4 De la note au signe : enjeux paléographiques de la création de vérités de terrain pour de l'OMR de partitions manuscrites

4.1 Pour une approche graphématique de la transcription de la notation musicale

L'OMR souffre également d'une absence de consensus sur la modélisation des objets à extraire. Comme l'indique KIESSLING [20], la plupart des corpus de référence en ATR n'explicitent pas leur vocabulaire, rendant toute comparaison entre eux difficile. L'OMR ne fait pas exception à ce manque de transparence, comme le prouve l'hétérogénéité des classes entre les corpus de références [9]. À notre connaissance, seules deux tentatives de formalisation exhaustive existent : une première définissant 94 classes [35], et une seconde où 88 sont dénombrées [38].

La définition de ces objets soulève des problématiques musicologiques complexes qui dépassent le cadre de cet article et mériteraient une étude dédiée. Néanmoins, cette question est cruciale car elle conditionne notre capacité à représenter le contenu musical, et à s'accorder pour créer de nouvelles vérités de terrain interopérables. Comme l'explique PINCHE [27], la transcription est un acte de modélisation dont les choix techniques conditionnent les usages futurs de cet objet scientifique. Une telle approche suppose cependant l'existence de formats capables de porter cette modélisation. Il s'agit donc de déterminer la capacité de ces standards à représenter fidèlement la source originale et sa notation, au-delà de la seule abstraction du contenu musical.



FIGURE 3 – Extrait de *Children's Corner*, Claude Debussy, non daté, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b100762337/f13>. Cette mesure présente des omissions dans les motifs rythmiques : absence de doubles croches pointées dans le motif du deuxième temps, similaire à celui du premier temps, et omissions de la notation des triolets dans le dernier temps.

S'appuyant sur les travaux de PINCHE [27], nous adoptons une approche graphématique pour la transcription de la CWMN historique, et l'appliquons à MusicXML. Contrairement à une transcription normalisée ou graphétique (décrivant chaque allographe), ce niveau se focalise, dans notre cas, sur l'orthographe musicale et la représentation des phénomènes matériels. Ce choix constitue le meilleur équilibre entre « rigueur formelle et faisabilité pratique et technique » [20; 27] : il harmonise les pratiques collaboratives, limite le nombre de classes à apprendre lors de l'entraînement d'un modèle et préserve l'intégrité visuelle de la source [14; 27].

Suivant les conclusions de CLÉRICE et al. [14]; PINCHE [27], nous reconnaissons qu'une représentation parfaitement identique de la source n'est pas possible dans le cas d'une source musicale.

Notre modèle doit reposer sur une simplification raisonnée. Ce choix évite au modèle de devoir internaliser des règles complexes de syntaxe musicale (ou de normalisation pour l'ATR [20]) qui peuvent être source d'hallucinations [13]. L'approche graphématique garantit une performance computationnelle tout en respectant l'intégrité de la source (fig. 3, 4).

The figure consists of two side-by-side musical staves. Both are titled "Un peu retenu." and have the instruction "✖ (en conservant le rythme)" above them. The left staff shows a normalized transcription with corrected rhythmic markings, including a pink arrow pointing to a correction. The right staff shows a graphematic transcription with some elements hidden (indicated by grey boxes) and red circles highlighting specific rhythmic details. Both transcriptions include a piano (*pp*) dynamic marking and a treble clef.

FIGURE 4 – Illustration d'une transcription de la fig. 3 avec MuseScore. À gauche, une transcription normalisée, où les indications rythmiques ont été corrigées. À droite, une transcription graphématique avec MuseScore. Nous avons utilisé l'outil « invisible » de MuseScore qui permet de cacher certains éléments graphiques (et qui apparaissent ici grisés), et qui seront indiqués lors de l'export en MusicXML en tant qu'attribut XML. Ils peuvent ainsi être supprimés en post-traitement, cassant ainsi la rythmique de la mesure, mais préservant son aspect visuel.

4.2 Les phénomènes non-standards comme obstacles à la représentation de la source : MusicXML face à la transcription graphématique

Les manuscrits de partitions foisonnent de phénomènes « non-standards » : des éléments graphiques ou structurels s'écartant de la syntaxe normalisée pour s'inscrire dans une pratique manuscrite spécifique, à savoir des irrégularités de notation (rythmes implicites, omissions), d'artefacts physiques (ratures, annotations manuscrites) ou de contraintes de mise en page. Ces singularités révèlent une tension entre deux ontologies : l'approche sémantique centrée sur la logique musicale et ancrée dans la philosophie de MusicXML, et l'approche graphématique, centrée sur la morphologie des signes.

La gestion des indications textuelles illustre notamment cette limite. MusicXML traite le texte comme une métadonnée ancrée ponctuellement à l'échelle de la mesure, empêchant ainsi la représentation de son déploiement spatial, par exemple quand une indication s'étend sur plusieurs mesures ou systèmes. Faute de liens structurels entre le texte et la source, nous choisissons de limiter MusicXML aux symboles musicaux et aux symboles textuels de nature musicale. Les sources musicales présentent parfois des objets au statut hybride, principalement des indications de nuances et de dynamiques, où le signe textuel porte une instruction de jeu, et donc une information musicale. Si les indications de dynamique musicales conventionnelles (*p*, *f*) sont nativement gérées par MusicXML, les nuances idiosyncratiques du compositeur ne peuvent être représentées de façon standardisée (fig. 5). Afin de transcrire une source, cette approche impose donc une représentation binaire : une structuration MusicXML pour la donnée musicale et une modélisation tierce pour les éléments textuels. Cette fragmentation empêche une appréhension unitaire du document au sein d'un format unique.



FIGURE 5 – Extrait des *Reflète dans l'eau*, Claude Debussy, 1901-1905, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b55002573h/f9>. Au début de la première mesure de cet extrait, on observe un symbole de dynamique commun à l'ensemble de l'histoire de la CWMN : *ppp*, tandis qu'au-dessus des mesures, Debussy utilise une indication de tempo avec son propre vocabulaire : « Lent (dans une sonorité harmonieuse et lointaine) ». Le premier élément serait inclus dans notre proposition de transcription, tandis que le second relève des habitudes d'écriture de Debussy, et ne peut pas être représenté de façon normalisée avec MusicXML.

Nous pouvons utiliser MusicXML pour représenter un semblant de structure en faisant appel à ses éléments de mise en page. Par exemple, lorsque couplé à une étape préliminaire d'analyse et de segmentation de systèmes [10], nous pouvons sommairement représenter leur structure interne grâce aux sauts de ligne. Or, la fidélité graphique exige parfois de représenter des cas plus complexes, comme celui de mesures scindées entre deux systèmes (fig. 6). Dans ce cas, nous privilégions la représentation documentaire en fragmentant l'unité logique en deux entités distinctes, contrairement au schéma de MusicXML qui impose qu'une mesure soit une unité logique close. La cohérence musicale est ainsi sacrifiée au profit de la structure physique. Cela transforme l'usage du format où la hiérarchie sémantique s'efface devant la topographie du manuscrit.



FIGURE 6 – Extrait de *La Boîte à joujoux*, ballet en 3 tableaux, piano seul, Claude Debussy, non daté, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b108610504/f43>. Exemple d'une mesure scindée entre deux systèmes.

Enfin, MusicXML ne peut représenter nativement les ratures, les repentirs ou les symboles ambigus, comme par exemple ceux utilisés pour représenter des répétitions non-standardisées (fig.

7, 8, 9, 10). Leur intégration suppose un détournement du schéma original par l'usage de caractères de substitution.



FIGURE 7 – Extrait des *Images oubliées*, n° 1, pour piano, Claude Debussy, non daté, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b10073993b/f4.item>.

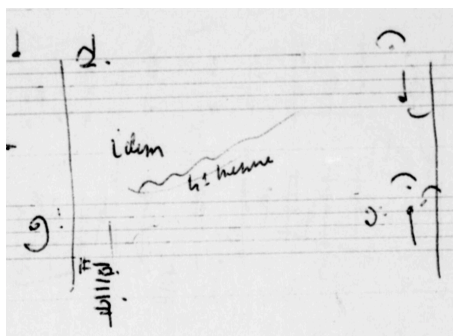


FIGURE 8 – Détail de la fig. 7 Marque de répétition inhabituelle, ici Debussy indique la répétition d'un motif isolé présent dans la quatrième mesure d'un système. Dans ladite mesure illustrée ici, le premier temps et le dernier temps sont cependant différents.



FIGURE 9 – Extrait de *Douze études pour le piano*, livres I et II, Claude Debussy, non daté, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b10860171f/f10.item>. On y observe de nombreuses ratures : mesures complètes, notes, dynamiques.

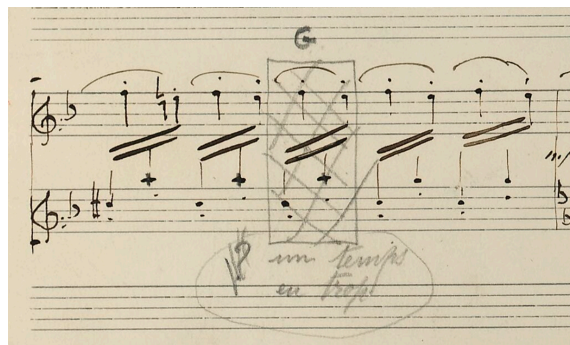


FIGURE 10 – Extrait de *Children's Corner*, Claude Debussy, 1908, manuscrit autographe, <https://gallica.bnf.fr/ark:/12148/btv1b52000863t/f17>. Lors de l'édition, un motif musical a été raturé car étant un temps en trop dans la mesure.

Ces détournements nous conduisent à définir un MusicXML « structurel ». Ce modèle de données n'est pas une partition fonctionnelle destinée à être affichée (*engraving*), mais une représentation où la structure matérielle du document prime sur sa validité théorique. En neutralisant les contraintes sémantiques les plus rigides au profit d'une vérité terrain fidèle à la source, ce MusicXML détourné devient un pivot pour l'entraînement de modèles capables de capturer son hétérogénéité.

5 Conclusion

L'état des lieux dressé révèle une forme de dette technique et ses conséquences scientifiques pour l'OMR. Entre le réductionnisme de Humdrum ***kern* et l'exhaustivité éditoriale de la MEI, MusicXML s'impose comme le pivot le plus pragmatique. Sa capacité à représenter la sémantique musicale et une mise en page sommaire, ainsi que son interopérabilité avec des éditeurs *open-source* tels que MuseScore en font un outil de production de données accessible, malgré son incapacité native à aligner le contenu musical sur la topographie de la source, qui échappe à la pure sémantique musicale.⁸

L'OMR orienté sources historiques semble souffrir d'un déficit d'infrastructure comparative-ment à l'ATR. Ce domaine a connu un saut qualitatif, porté non seulement par un progrès technologique, mais aussi par une profonde réflexion épistémologique sur ses standards. L'ATR offre à ce titre un miroir instructif : l'adoption généralisée du format ALTO XML pour l'échange de vérités de terrain, le développement de moteurs agnostiques comme Kraken intégrés dans des interfaces collaboratives telles qu'eScriptorium [19; 36], ont transformé la technologie en une véritable infrastructure de recherche. À l'image des initiatives pour la standardisation des *guidelines* et le partage des vérités de terrain [12; 14], l'OMR historique doit désormais stabiliser ses propres normes pour sortir de l'isolement des corpus spécifiques. En outre, la robustesse de MusicXML et celle de notre proposition graphématique demandent désormais à être confrontées à des structures de partition plus denses et plus complexes, comme des partitions d'orchestre.

En conclusion, l'avancement de l'OMR pour les sources historiques ne dépend pas seulement de la puissance des modèles, mais d'un accord collectif sur trois éléments :

1. **Une interface permettant le couplage partition-image** pour garantir un alignement strict entre le pixel et l'annotation ;

8. Notre choix a été motivé par l'absence d'accès public à l'interface MuRet, pourtant optimisée pour la recherche en OMR [15] et pour la transcription de sources musicales historiques [33]. Nous avons également écarté les solutions propriétaires (PhotoScore, Sibelius). MuseScore est un logiciel *open-source*, il offre notamment une compatibilité croissante avec MusicXML grâce à son développement continu [29].

2. **Un format d'encodage hybride** permettant une annotation granulaire, de la portée jusqu'à la page, inspiré d'ALTO XML ;
3. **La définition de *guidelines* communes**, capables de résoudre les ambiguïtés de la CWMN face aux idiosyncrasies des manuscrits.

Comme le démontre KIESSLING [20], si les modèles de transcription actuels affichent des performances satisfaisantes, leur fiabilité sur des données historiques reste limitée par la variabilité des sources et le manque de jeux de données de test représentatifs. La convergence vers des standards de représentation et une ontologie commune des signes musicaux, ainsi que vers des normes de transcription partagées, constitue ainsi les prérequis indispensables au développement d'un OMR robuste historique orienté sur la source.

Références

- [1] ALTO EDITORIAL BOARD. *ALTO (Analyzed Layout and Text Object) : Technical Metadata for Optical Character Recognition*. Library of Congress, 2024. URL : <https://www.loc.gov/standards/alto/>.
- [2] ASBERT, Gerard, TORRAS, Pau, KANG, Lei, FORNÉS, Alicia et LLADÓS, Josep. « GAN-based Content-Conditioned Generation of Handwritten Musical Symbols ». Oct. 2025. DOI : 10.48550/arXiv.2510.17869.
- [3] AVID TECHNOLOGY. « Sibelius ». 2024. URL : <https://www.avid.com/sibelius>.
- [4] BARÓ, Arnau, BADAL, Carles et FORNÉS, Alicia. « Handwritten Historical Music Recognition by Sequence-to-Sequence with Attention Mechanism ». In : *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. Dortmund, Allemagne, sept. 2020, p. 205-210. DOI : 10.1109/ICFHR2020.2020.00046.
- [5] BARRAQUÉ, Jean. *Debussy*, sous la dir. de François LESURE. Paris, France : Seuil, 1994.
- [6] CALVO-ZARAGOZA, Jorge et RIZO, David. « Camera-PrIMuS : Neural End-to-End Optical Music Recognition on Realistic Monophonic Scores ». In : *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018*. Paris, France, 2018. URL : http://ismir2018.ircam.fr/doc/pdfs/33_Paper.pdf.
- [7] CALVO-ZARAGOZA, Jorge et RIZO, David. « End-to-End Neural Optical Music Recognition of Monophonic Scores ». In : *Applied Sciences* 8, no. 4 (avr. 2018), p. 606. DOI : 10.3390/app8040606.
- [8] CASTELLANOS, FRANCISCO J., CALVO-ZARAGOZA, Jorge et IÑESTA, José M. « A Neural Approach for Full-Page Optical Music Recognition of Mensural Documents. ». In : *Proceedings of the 21th International Society for Music Information Retrieval Conference*. Montréal, Canada, 2020, p. 558-565. URL : https://program.ismir2020.net/static/final_papers/207.pdf.
- [9] CASTELLANOS, FRANCISCO J., GALLEGO, Antonio Javier et FUJINAGA, Ichiro. « Deep Learning for Optical Music Recognition : A Review ». In : *Authorea Preprints* (fév. 2025). DOI : 10.36227/techrxiv.174077177.78767136/v1.
- [10] CASTELLANOS, FRANCISCO J., GARRIDO-MUNOZ, Carlos, RÍOS-VILA, Antonio et CALVO-ZARAGOZA, Jorge. « Region-based Layout Analysis of Music Score Images ». Jan. 2022. DOI : 10.48550/arXiv.2201.04214.
- [11] CERVETO-SERRANO, Joan, RIZO, David, CALVO-ZARAGOZA, Jorge, LEWIS, David, PLAKSIN, Anna et STREMEL, Sophie. « kernpy : a Humdrum **Kern Oriented Python Package for Optical Music Recognition Tasks ». In : *Music Encoding Conference 2025*. Londres, Angleterre, 2025, p. 178-183. DOI : 10.17613/wg0x7-re949.

- [12] CHAGUÉ, Alix, CLÉRICE, Thibault et ROMARY, Laurent. « HTR-United : un écosystème pour une approche mutualisée de la transcription automatique des écritures manuscrites ». 2022. URL : <https://inria.hal.science/hal-04124743>.
- [13] CLÉRICE, Thibault, BAWDEN, Rachel, SMITH, David A. et PINCHE, Ariane. « PaRAMHTRS (Philology And Resolution of Abbreviations in Manuscripts obtained by HTR at Scale) ». In : *Journées du Datalab*. Paris, France : Datalab, jan. 2026. URL : <https://inria.hal.science/hal-05453316>.
- [14] CLÉRICE, Thibault et al. « CATMuS Medieval : A multilingual large-scale cross-century dataset in Latin script for handwritten text recognition and beyond ». In : *Proceedings of the 18th International Conference on Document Analysis and Recognition (ICDAR 2024)*. Athènes, Grèce, 2024, p. 174. DOI : 10.1007/978-3-031-70543-4_11.
- [15] GARCÍA-IASCI, Patricia, RIZO, David et CALVO-ZARAGOZA, Jorge. « Evaluating Music Encoding Approaches : An Accuracy Analysis of Tools and Standards. » In : *Music Encoding Conference 2025*. Londres, Angleterre, 2025, p. 189-194. DOI : 10.17613/ce18d-9kh25.
- [16] GOOD, Michael. « MusicXML for Notation and Analysis ». In : *The Virtual Score : Representation, Retrieval, Restoration*, sous la dir. de Walter B. HEWLETT et Eleanor SELFRIDGE-FIELD. T. 12. Computing in Musicology. Cambridge, MA : MIT Press, 2001, p. 113-124. DOI : 10.7551/mitpress/2058.003.0010.
- [17] HAJIČ JR., Jan et PECINA, Pavel. « In Search of a Dataset for Handwritten Optical Music Recognition : Introducing MUSCIMA++ ». Mars 2017. DOI : 10.48550/arXiv.1703.04824.
- [18] HURON, David. « Music Information Processing Using the Humdrum Toolkit : Concepts, Examples, and Lessons ». In : *Computer Music Journal* 26 (juin 2002), p. 11-26. DOI : 10.1162/014892602760137158.
- [19] KIESSLING, Benjamin. « The Kraken OCR system ». 2025. URL : <https://kraken.re>.
- [20] KIESSLING, Benjamin. « Transcription Guidelines for Generalized Automatic Text Recognition ». Déc. 2025. URL : <https://hal.science/hal-05429033>.
- [21] MAYER, Jiří et PECINA, Pavel. « Obstacles with Synthesizing Training Data for OMR ». In : *4th International Workshop on Reading Music Systems (WoRMS), Online*. 2022, p. 15-19. URL : <https://ufal.mff.cuni.cz/biblio/attachments/2022-mayer-p6357779747291544732.pdf>.
- [22] MAYER, Jiří, STRAKA, Milan, HAJIČ JR., Jan et PECINA, Pavel. « Practical End-to-End Optical Music Recognition for Pianoform Music ». In : *Proceedings of the 18th International Conference on Document Analysis and Recognition (ICDAR 2024)*. Athènes, Grèce, 2024, p. 55-73. DOI : 10.1007/978-3-031-70552-6_4.
- [23] MUSESCORE BVBA. « MuseScore ». 2024. URL : <https://musescore.org>.
- [24] MUSIC ENCODING INITIATIVE et al. « MEI 5.1 ». Jan. 2025. DOI : 10.5281/ZENODO.8319922.
- [25] NIENHUY, Han-Wen, NIEUWENHUIZEN, Jan et LILYPOND DEVELOPMENT TEAM. « LilyPond : The Music Notation Project ». 2024. URL : <https://lilypond.org>.
- [26] PARADA-CABALEIRO, Emilia, BATLINER, Anton, BAIRD, Alice et SCHULLER, Björn. « The SEILS dataset : Symbolically encoded scores in modern-early notation for computational musicology ». In : *18th International Society for Music Information Retrieval Conference*. Suzhou, Chine, 2017. URL : <https://archives.ismir.net/ismir2017/paper/000014.pdf>.

- [27] PINCHE, Ariane. « Editing in the age of AI : stratified establishment of medieval texts ». In : *International Journal for Digital Humanities* 7 (déc. 2025), p. 537-554. DOI : 10 . 1007 / s42803-025-00116-6.
- [28] PLETSCHACHER, Stefan et ANTONACOPOULOS, Apostolos. « The PAGE (Page Analysis and Ground-Truth Elements) Format Framework ». In : *2010 20th International Conference on Pattern Recognition*. Istanbul, Turquie, août 2010, p. 257-260. DOI : 10 . 1109 / ICPR . 2010 . 72.
- [29] RETTINGHAUS, Klaus, LEWIS, David, PLAKSIN, Anna et STREMEL, Sophie. « Pushing the standard to its limits : MuseScore as a feature-complete MusicXML editor ». In : *Music Encoding Conference 2025* (2025), p. 195-197. DOI : 10 . 17613/n8c9g-ywm52.
- [30] RÍOS-VILA, Antonio, CALVO-ZARAGOZA, Jorge et PAQUET, Thierry. « Sheet Music Transformer : End-To-End Optical Music Recognition Beyond Monophonic Transcription ». Avr. 2024. DOI : 10 . 48550/arXiv . 2402 . 07596.
- [31] RÍOS-VILA, Antonio, CALVO-ZARAGOZA, Jorge, RIZO, David et PAQUET, Thierry. « Sheet Music Transformer ++ : End-to-End Full-Page Optical Music Recognition for Pianoform Sheet Music ». Mai 2024. DOI : 10 . 48550/arXiv . 2405 . 12105.
- [32] RISM EDITORIAL CENTER. « PAEC : Pre-Authoritative Ensemble Cataloging Guidelines ». International Inventory of Musical Sources (RISM). 2021. URL : <https://rism.info/documentation/>.
- [33] RIZO, David, CALVO-ZARAGOZA, Jorge, MARTÍNEZ-SEVILLA, Juan C., ROSELLÓ, Adrián et FUENTES-MARTÍNEZ, Eliseo. « Design of a music recognition, encoding, and transcription online tool ». In : *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research*. Tokyo, Japon, nov. 2023, p. 18-29. DOI : 10 . 5281/zenodo . 10109915.
- [34] ROS-FÁBREGAS, Emilio. « Codified Spanish music heritage through Verovio : The online platforms Fondo de Música Tradicional IMF-CSIC and Books of Hispanic Polyphony IMF-CSIC. » In : *Music Encoding Conference 2021 - Proceedings*. Alicante, Espagne, mai 2021. URL : <http://hdl.handle.net/10261/285557>.
- [35] SHATRI, Elona et FAZEKAS, György. « DoReMi : First glance at a universal OMR dataset ». Juill. 2021. DOI : 10 . 48550/arXiv . 2107 . 07786.
- [36] STOKES, Peter A., KIESSLING, Benjamin, STÖKL BEN EZRA, Daniel, TISSOT, Robin et GARGEM, El Hassane. « The eScriptorium VRE for Manuscript Cultures, in Ancient Manuscripts and Virtual Research Environments ». In : *Classic @ Journal* , no. 18 (2021). URL : <https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/>.
- [37] TEI CONSORTIUM. *TEI P5 : Guidelines for Electronic Text Encoding and Interchange*, sous la dir. de Lou BURNARD et Syd BAUMAN. Oxford, Providence, Charlottesville, Nancy, 2024. URL : <https://tei-c.org/Guidelines/P5/>.
- [38] TORRAS, Pau, BISWAS, Sanket et FORNÉS, Alicia. « A Unified Representation Framework for the Evaluation of Optical Music Recognition Systems ». In : *International Journal on Document Analysis and Recognition (IJ DAR)* 27, no. 3 (sept. 2024), p. 379-393. DOI : 10 . 1007/s10032-024-00485-8.
- [39] WALSHAW, Chris, ALLENTUCK, Irwin et METHFESSEL, Guido. *The ABC Music Standard*. 2.1. 2011. URL : <http://abcnotation.com/wiki/abc:standard:v2.1>.