Encoding Matters

Néstor Nápoles, Gabriel Vigliensoni, and Ichiro Fujinaga

McGill University, CIRMMT

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Introduction

In previous studies of computational musicology, we have observed that modifying the workflow for obtaining digital scores may also affect the result of the studies

- Using a different version of the same music notation software
- Importing/Exporting the same file in a different music notation software
- Translating the digital score from one symbolic format to another

The best practices learned from this experience contributed to the development of a methodology for creating symbolic music corpora (Cumming et al., 2018)

Encoding Matters

Encoding
discrepancies
in existing data

Encoding Matters

Investigate the **discrepancies** between symbolic music files that (intend to) represent the same music score document, which are studied in two different contexts

- A person encodes a music score document using a music notation software (Transcription)
- A symbolic music format is translated into a different symbolic music format (Translation)

Encodings

Selected work

Beethoven, Op.18 No.1 - I. Allegro con brio

Encodings

- *Finale encoding*
  - Source: the Gutenberg project
  - Encoder: Geof Pawlicki
- *MuseScore encoding*
  - Source: musescore.org
  - Encoder: Gavin Ailes
- *Sibelius encoding*
  - Source: tes.com
  - Encoder: submitted by user *dunhallin*

Translations

Each of the previous encodings has been exported as a **MusicXML** file using the latest stable version of the music notation software, afterwards, these files are translated into **two** other symbolic music formats

Translations

- MusicXML ----> Music Encoding Initiative (MEI)
  - Parser: Verovio v2.0.0 (https://www.verovio.org/)

- MusicXML ----> Humdrum(**kern)
  - Parser: humlib (https://github.com/craigsapp/humlib - commit *b71f716*)

Translations

Each of the previous encodings has been exported as a **MusicXML** file using the latest stable version of the music notation software, afterwards, these files are translated into **two** other symbolic music formats

Translations

- MusicXML ----> **Music Encoding Initiative (MEI)**
  - Parser: Verovio v2.0.0 (https://www.verovio.org/)

- MusicXML ----> Humdrum(**kern)
  - Parser: humlib (https://github.com/craigsapp/humlib - commit *b71f716*)

Translations

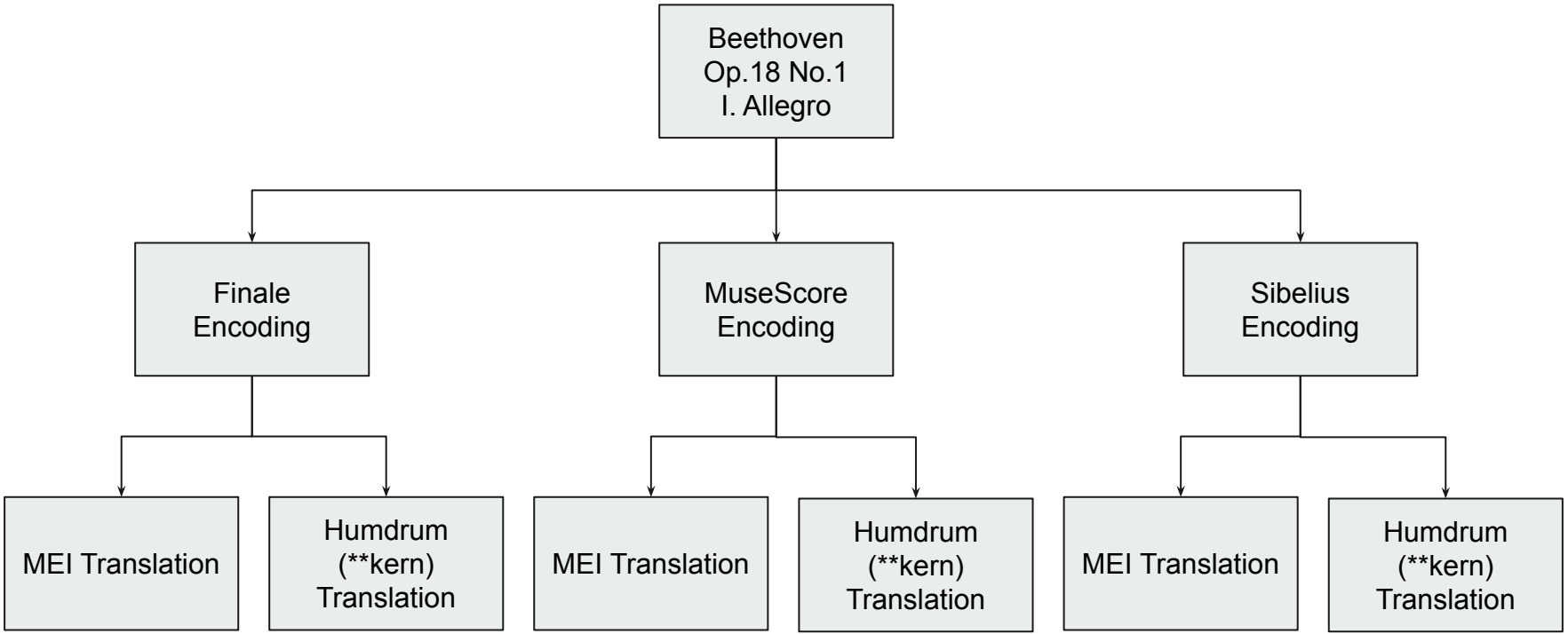
Each of the previous encodings has been exported as a **MusicXML** file using the latest stable version of the music notation software, afterwards, these files are translated into **two** other symbolic music formats

Translations

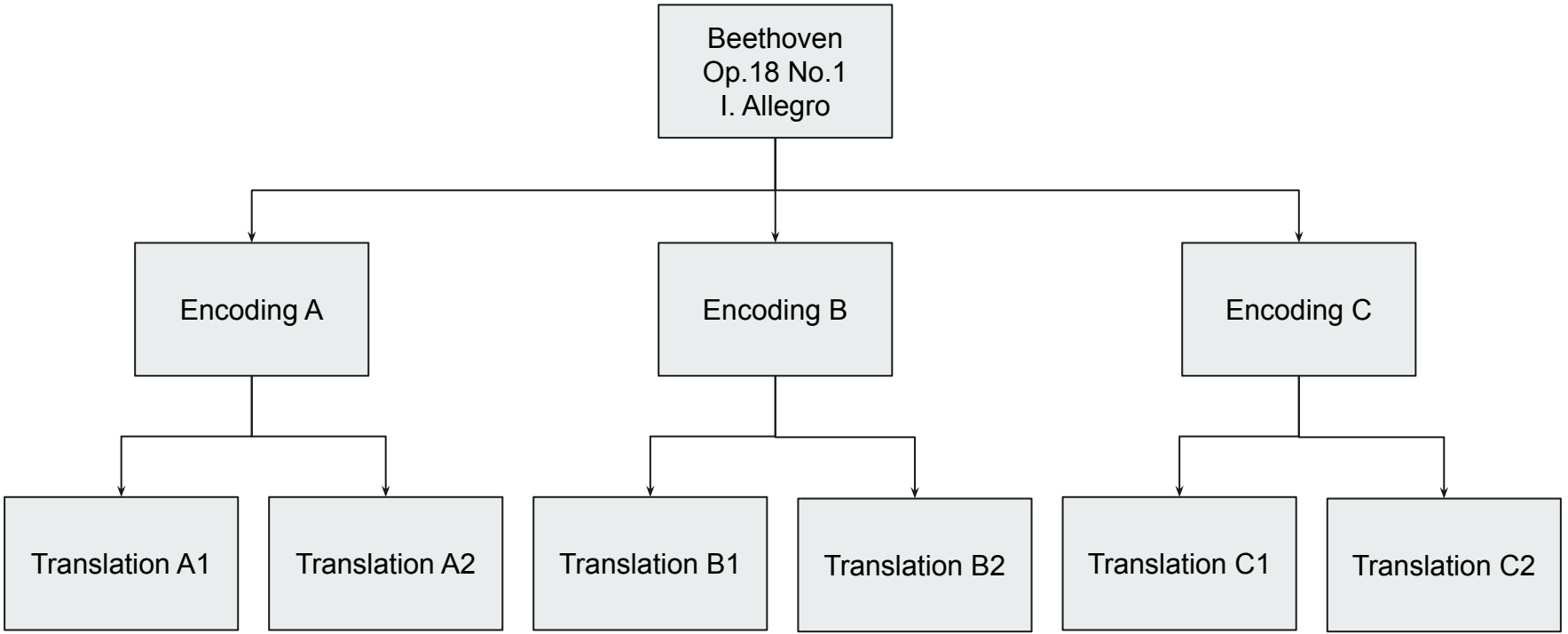
- MusicXML ----> Music Encoding Initiative (MEI)
  - Parser: Verovio v2.0.0 (https://www.verovio.org/)

- MusicXML ----> **Humdrum(**kern)**
  - Parser: humlib (https://github.com/craigsapp/humlib - commit *b71f716*)

Encodings and Translations



Encodings and Translations (anonymized)



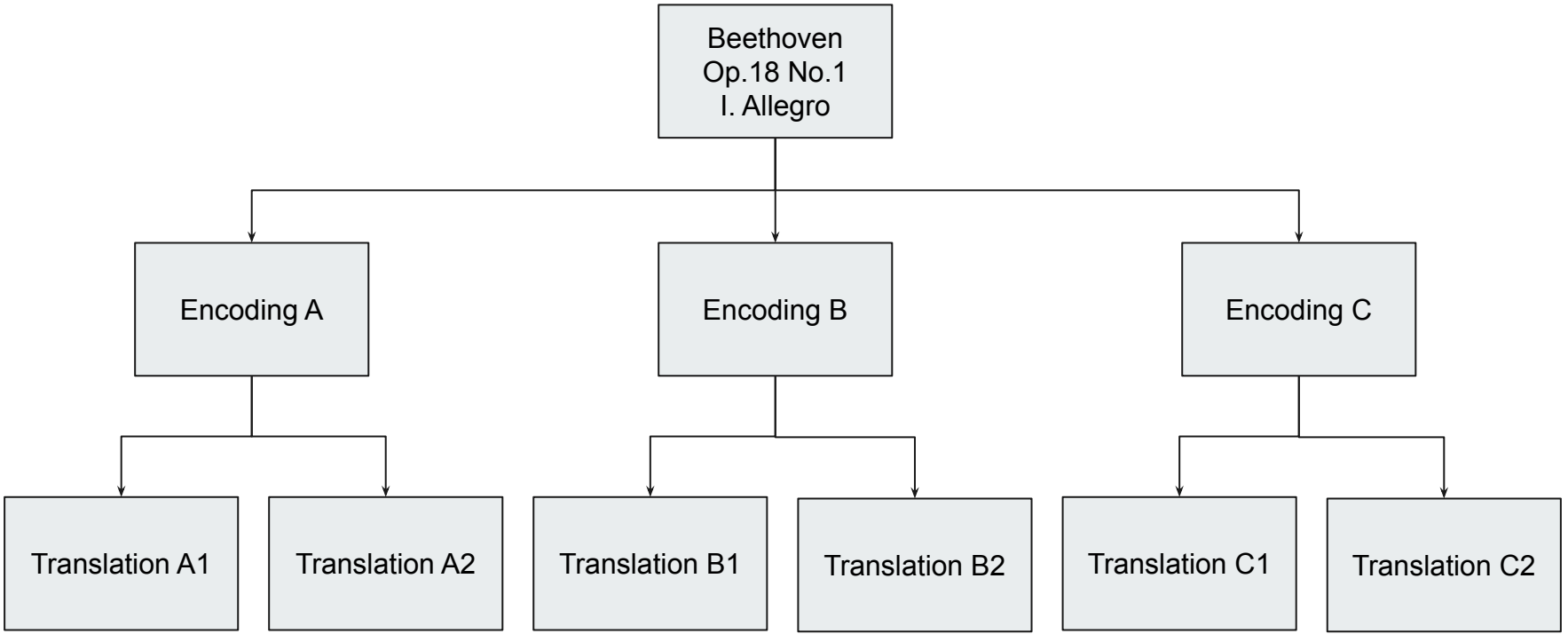
Method

Studying discrepancies in two different contexts

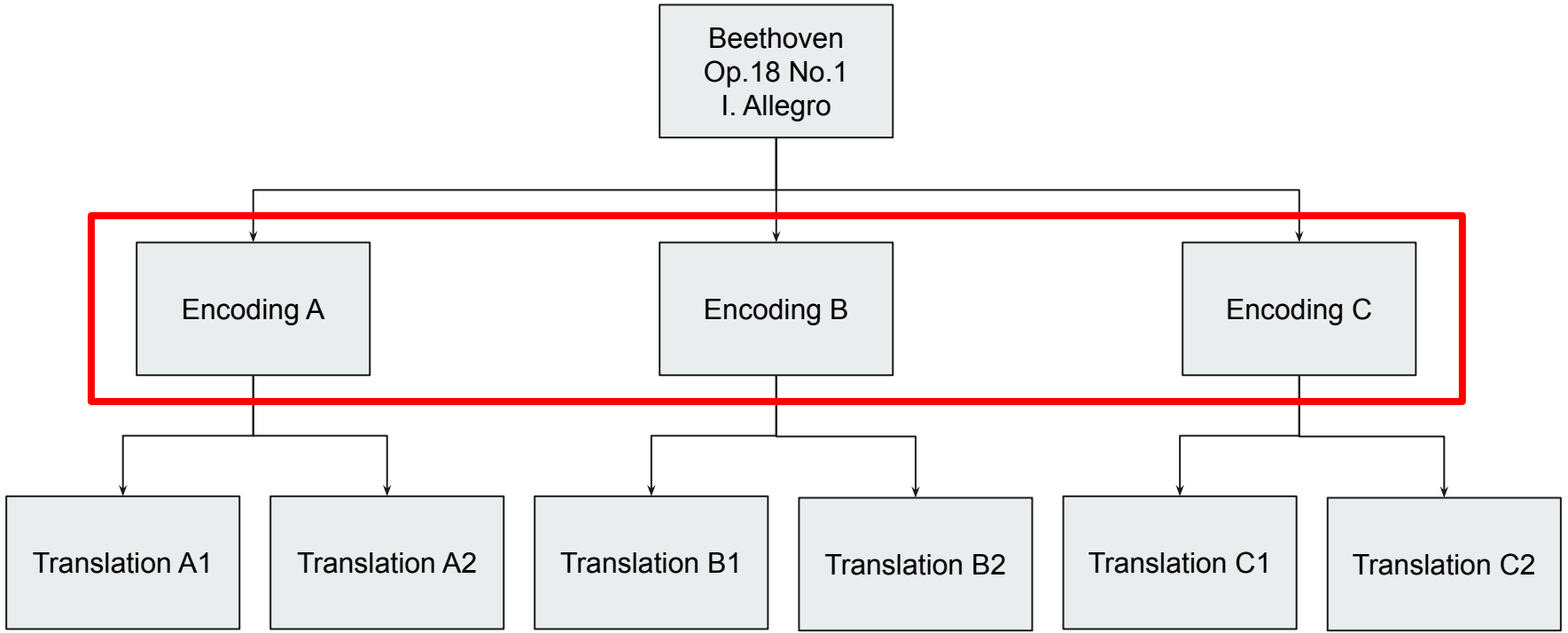
- **A person encodes a music score document using a music notation software (Transcription)**

- A symbolic music format is translated into a different symbolic music format (Translation)

Encodings and Translations (anonymized)



Encodings and Translations

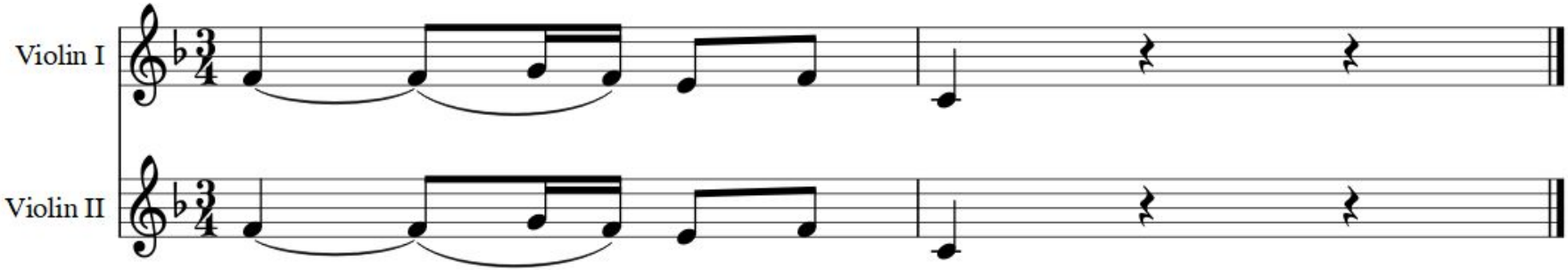


What is *equivalent*?

Having the same **note/rest attacks**, starting at the same **offset**

What is *equivalent*?

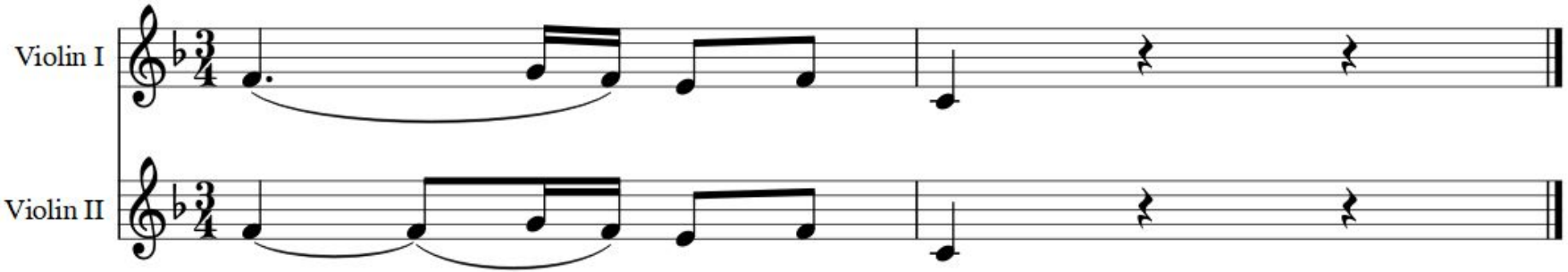
Having the same **note/rest attacks**, starting at the same **offset**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

What is *equivalent*?

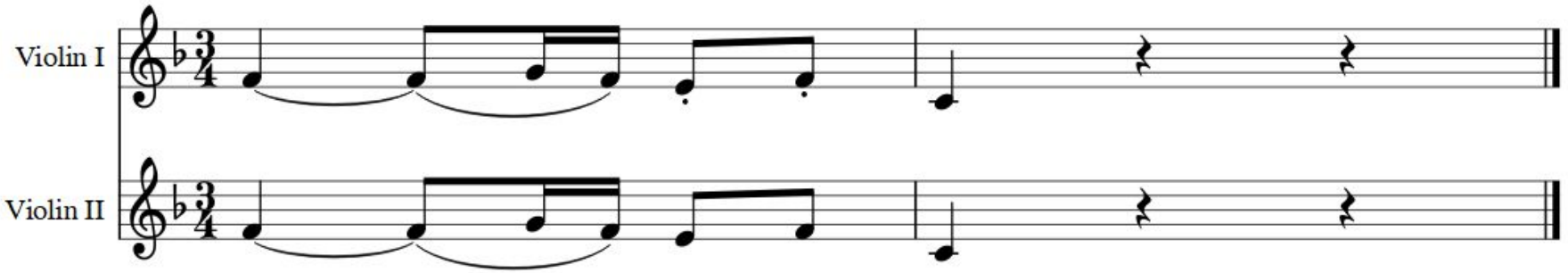
Having the same **note/rest attacks**, starting at the same **offset**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

What is *equivalent*?

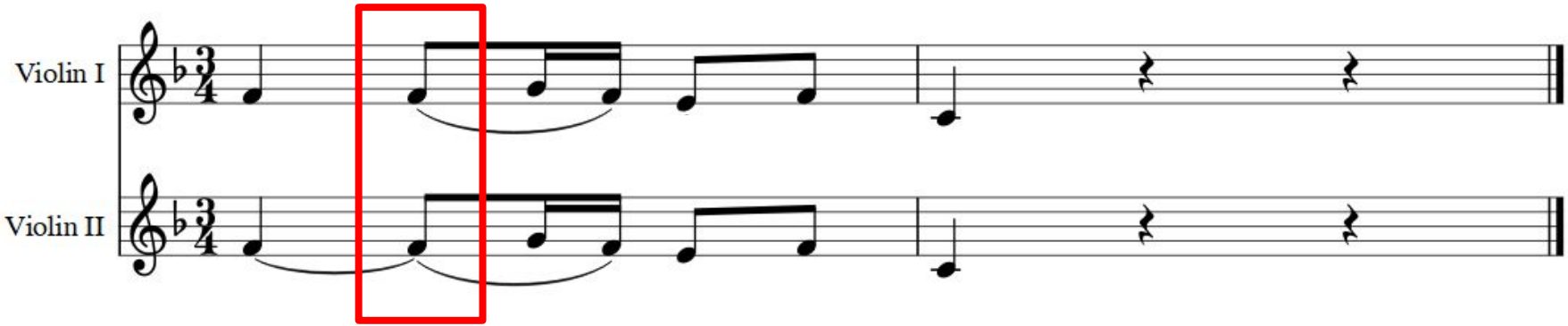
Having the same **note/rest attacks**, starting at the same **offset**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

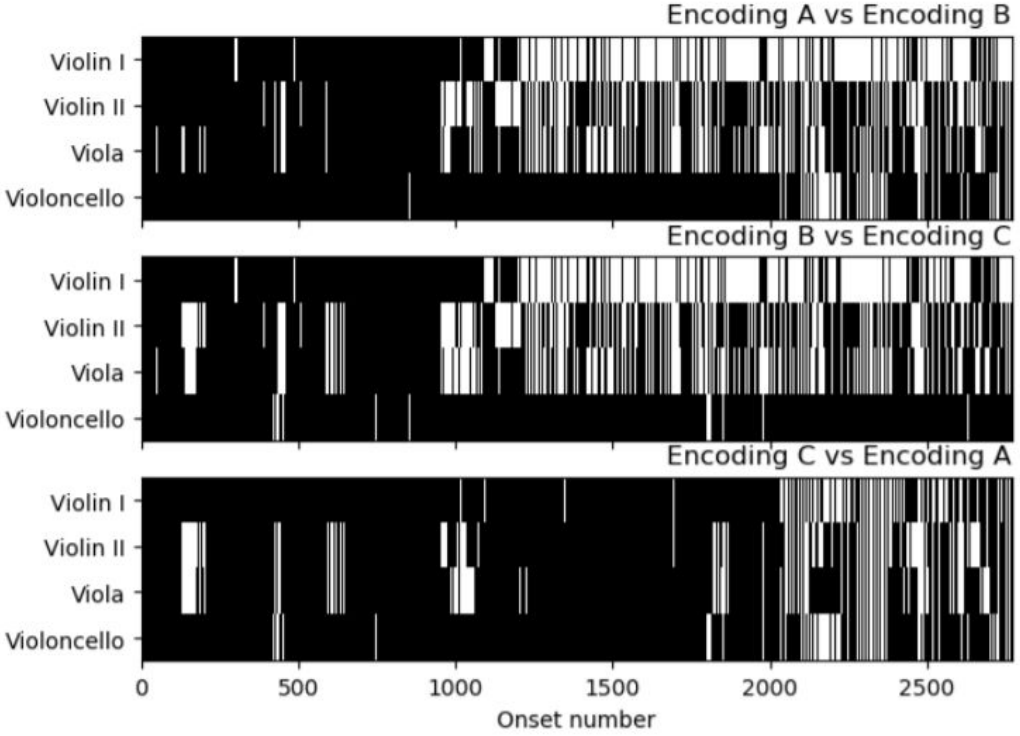
What is *equivalent*?

Having the same **note/rest attacks**, starting at the same **offset**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

Comparison of Encodings



Discrepancies

Introduced by *Music Notation Software*

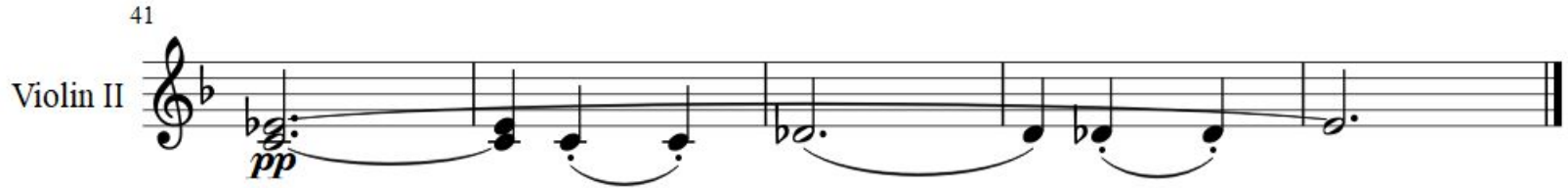
Introduced by *Human Encoders*

Discrepancies (Music Notation Software)

Encoding B, measures 41-45

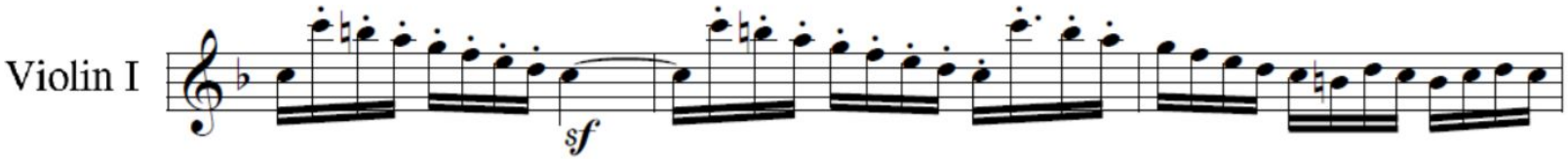
Discrepancies (Music Notation Software)

Unclosed tie (imported in a different music notation software)



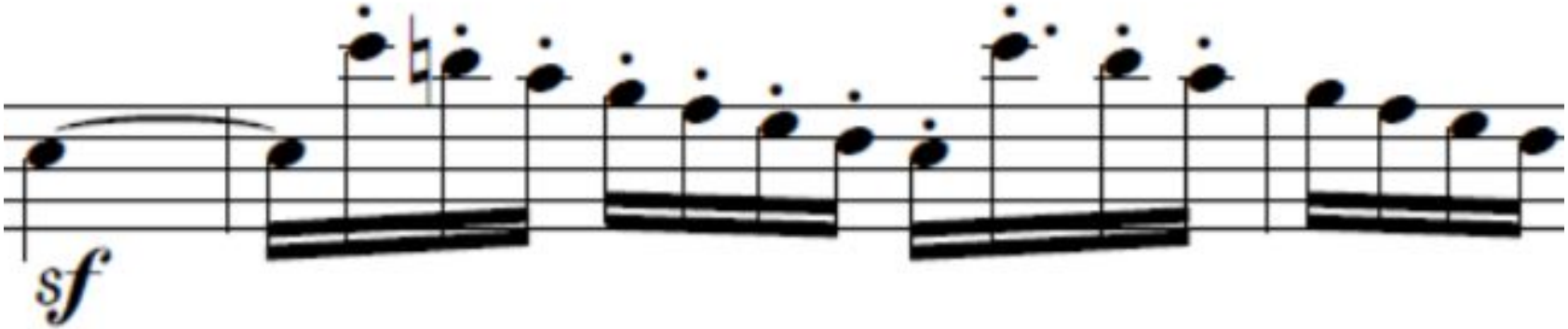
Encoding B, measures 41-45

Discrepancies (Music Notation Software)



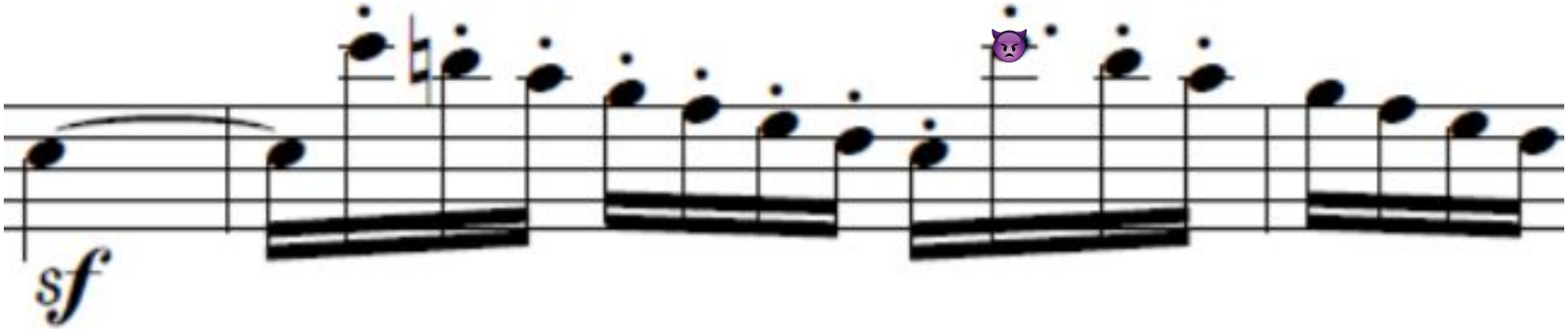
Encoding B, measures 168-170

Discrepancies (Music Notation Software)



Encoding B, measures 168-170

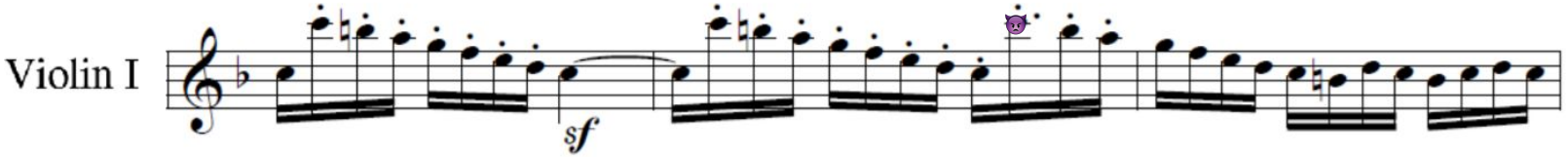
Discrepancies (Music Notation Software)



Encoding B, measures 168-170

Discrepancies (Music Notation Software)

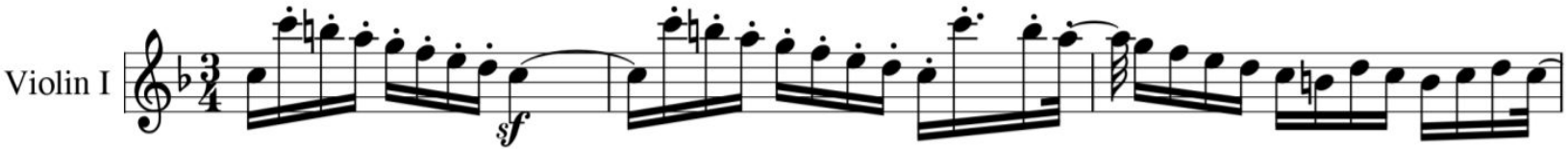
Overrun measure



Encoding B, measures 168-170

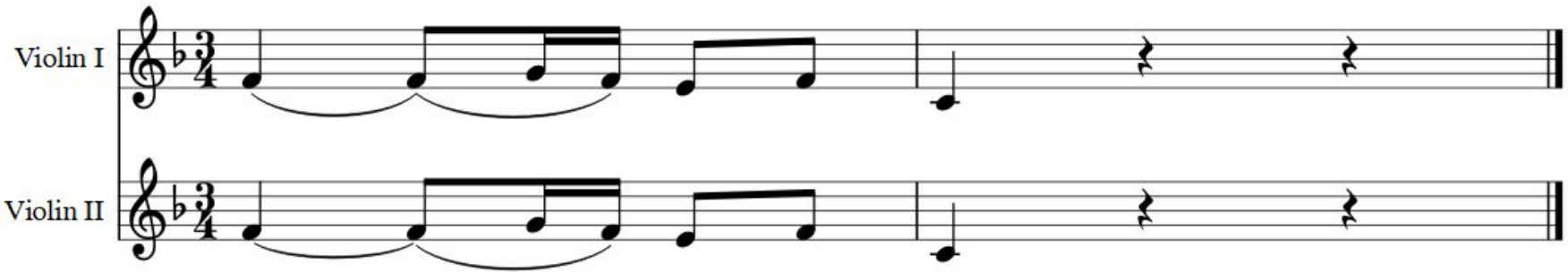
Discrepancies (Music Notation Software)

Overrun measure (imported in a different music notation software)



Encoding B, measures 168-170

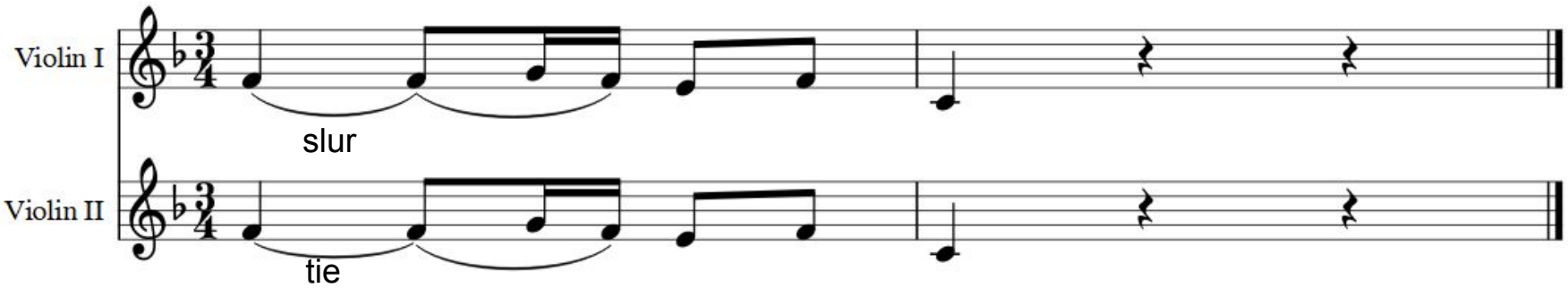
Discrepancies



Encoding C, measures 1-2

Discrepancies (Human Encoders)

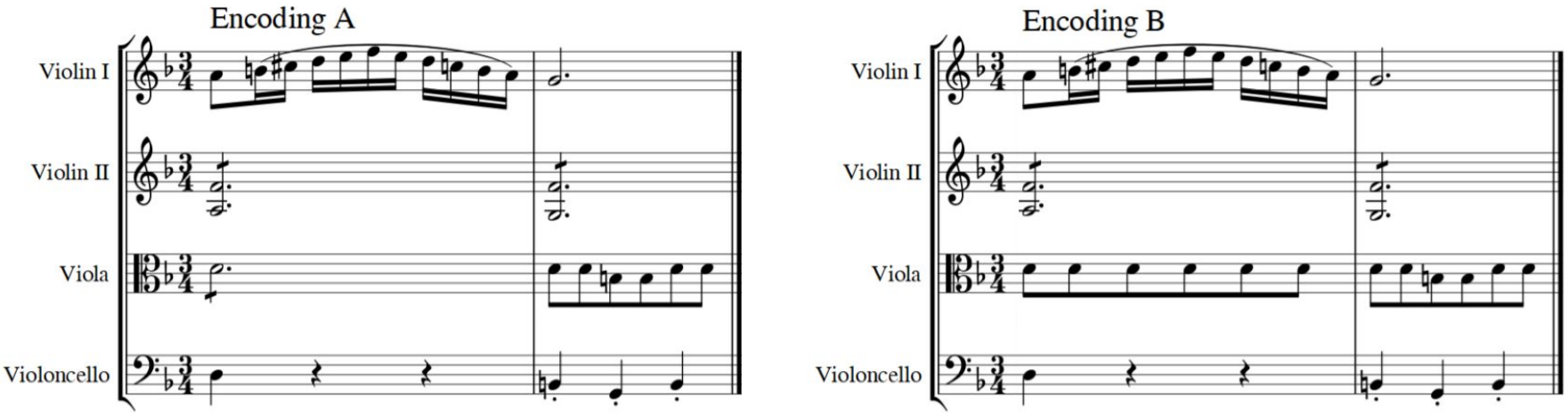
Slurs instead of ties



Encoding C, measures 1-2

Discrepancies (Human Encoders)

Repeated notes, trills, and grace notes



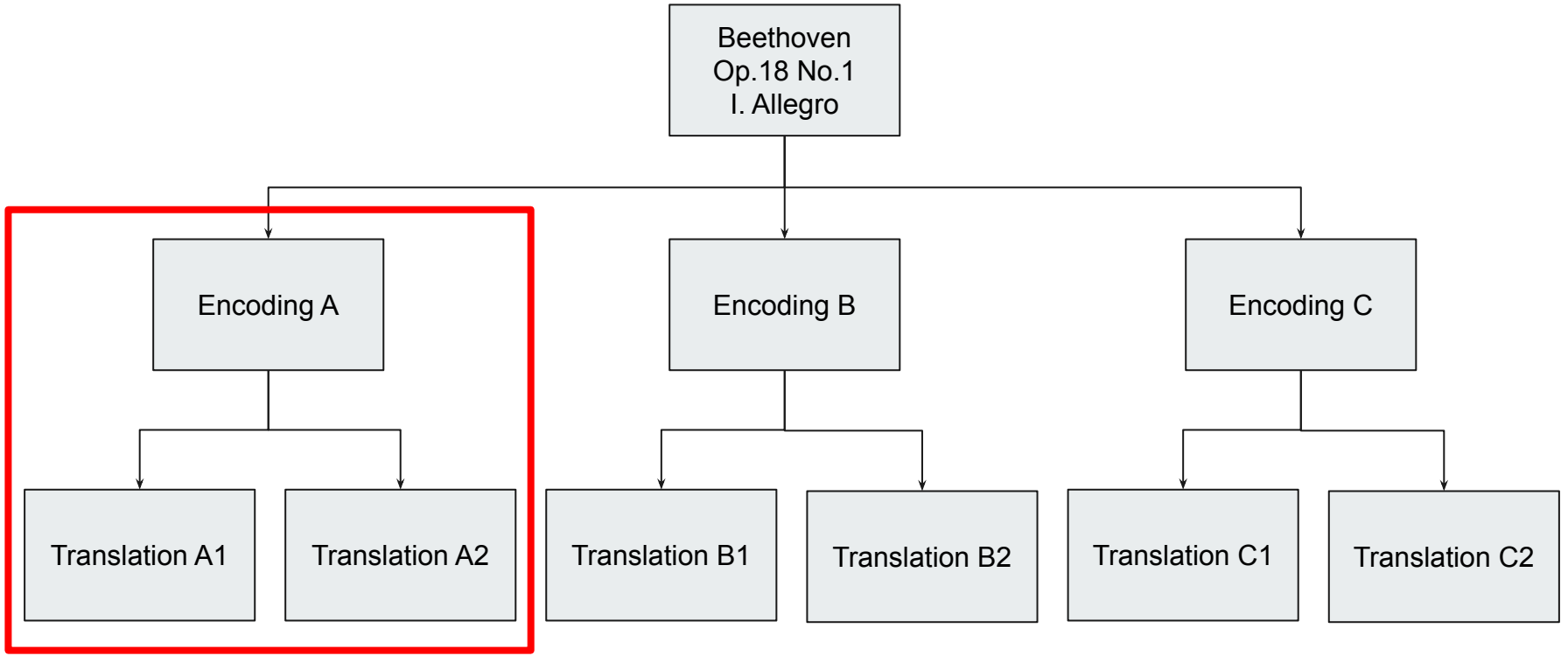
Encodings A and B, measures 30-31

Encoding Matters

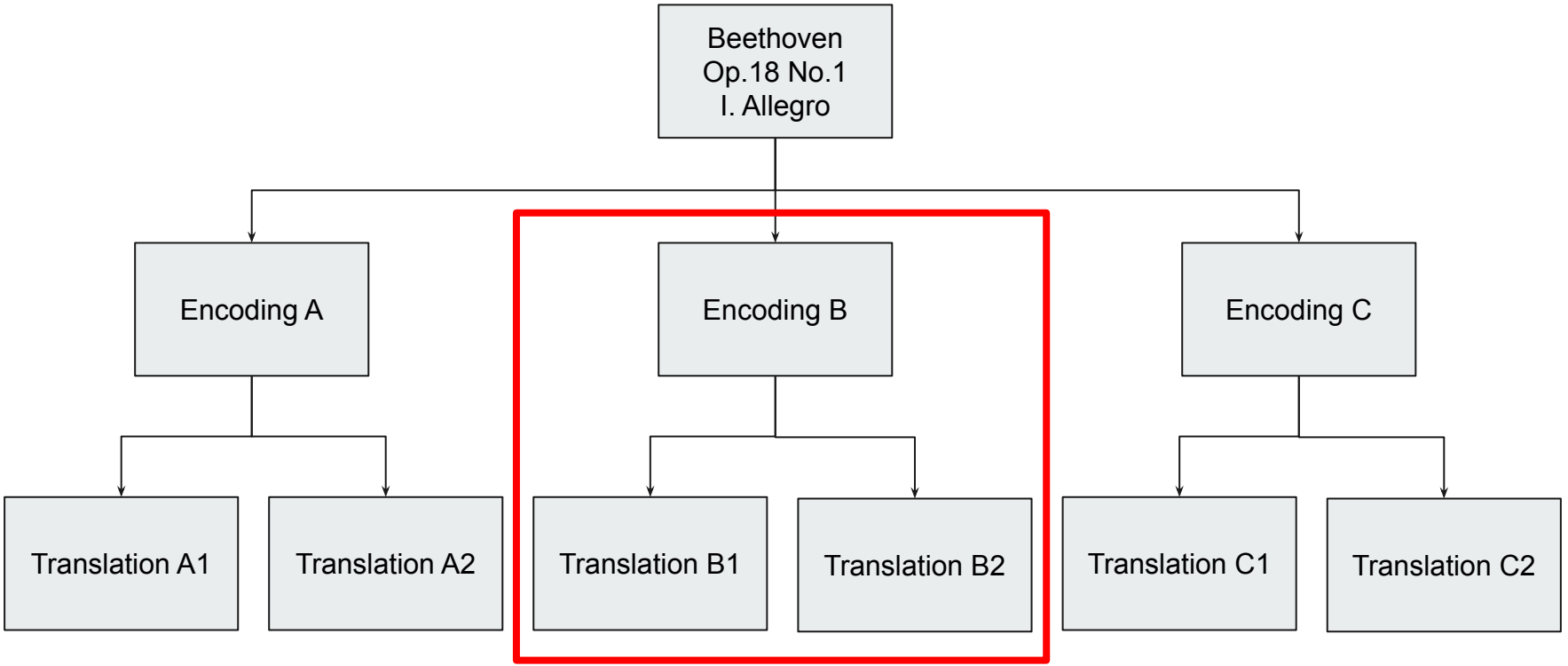
Studying discrepancies in two different contexts

- A person encodes a music score document using a music notation software (Transcription)
- **One symbolic music format is translated into a different symbolic music format (Translation)**

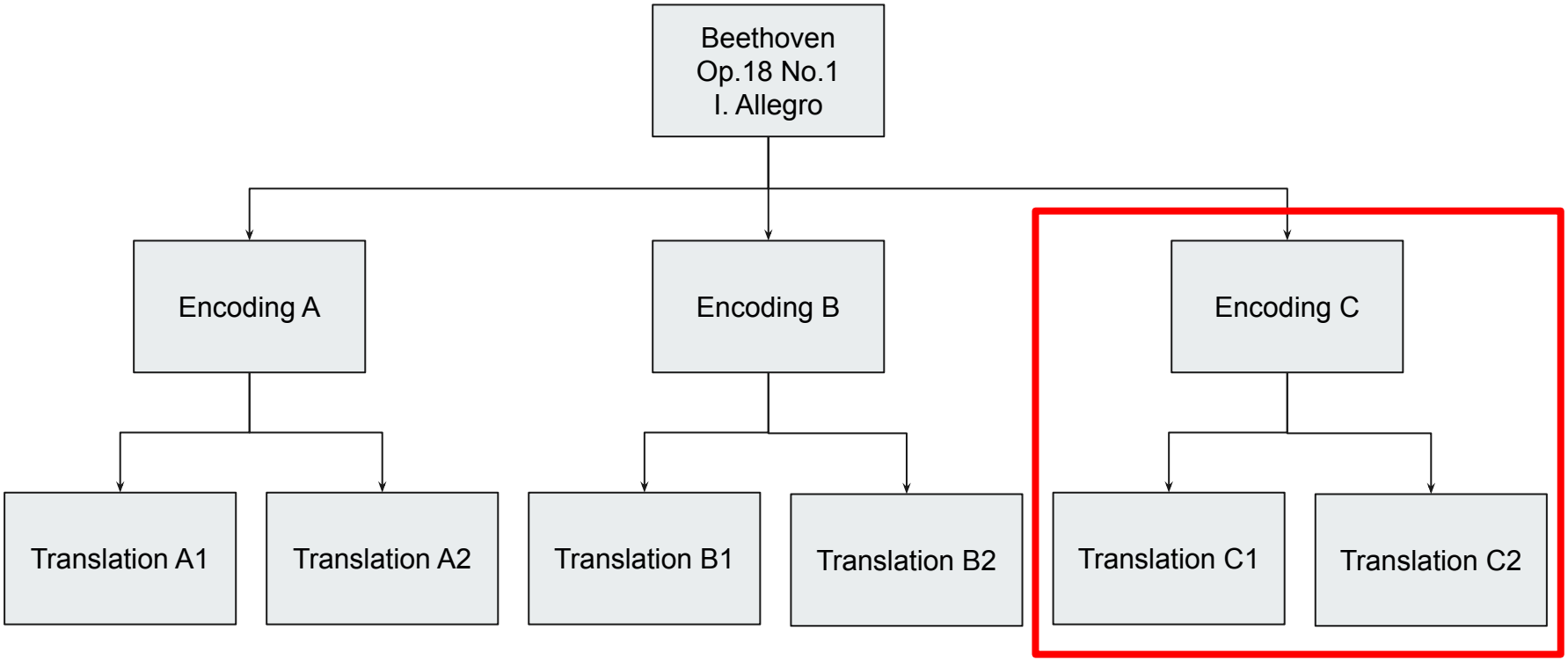
Encodings and Translations



Encodings and Translations



Encodings and Translations

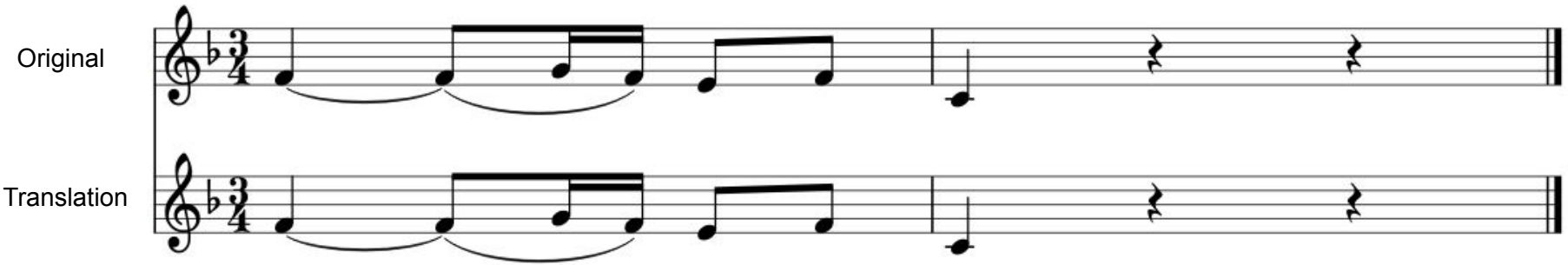


What is *equivalent (in a translation)*?

Having the same note/rest **events**, starting at the same offset **and preserving the same attributes (duration, articulation, and ornaments)**

What is *equivalent (in a translation)*?

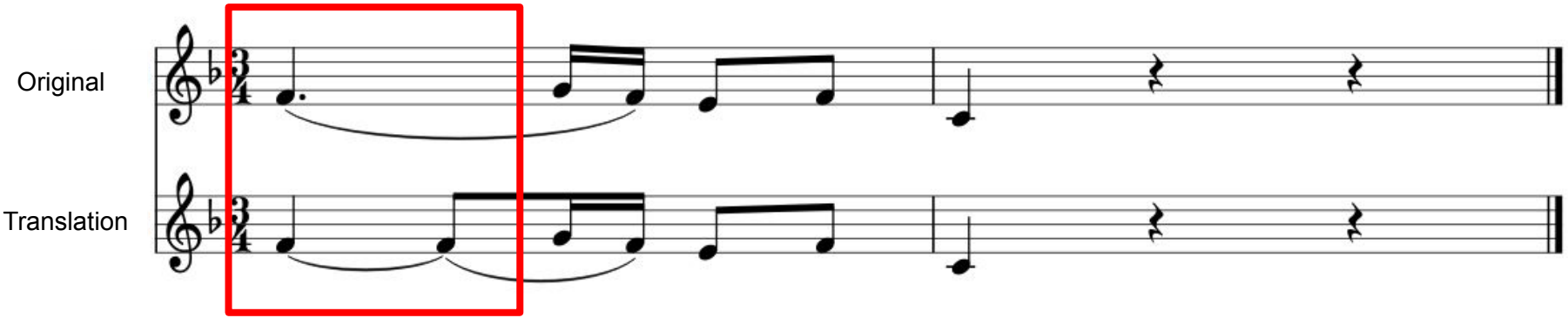
Having the same note/rest events, starting at the same offset **and preserving the same attributes (duration, articulation, and ornaments)**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

What is *equivalent (in a translation)*?

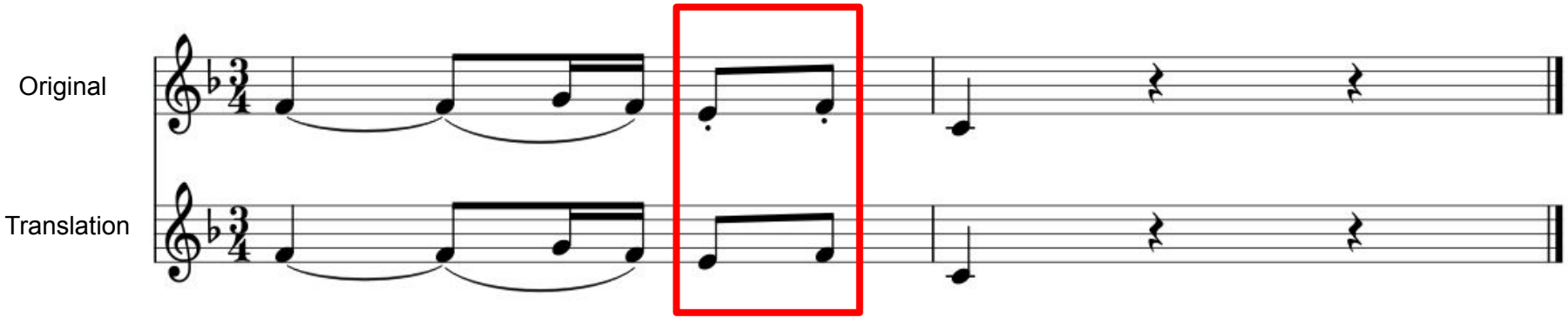
Having the same note/rest events, starting at the same offset **and preserving the same attributes (duration, articulation, and ornaments)**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

What is *equivalent (in a translation)*?

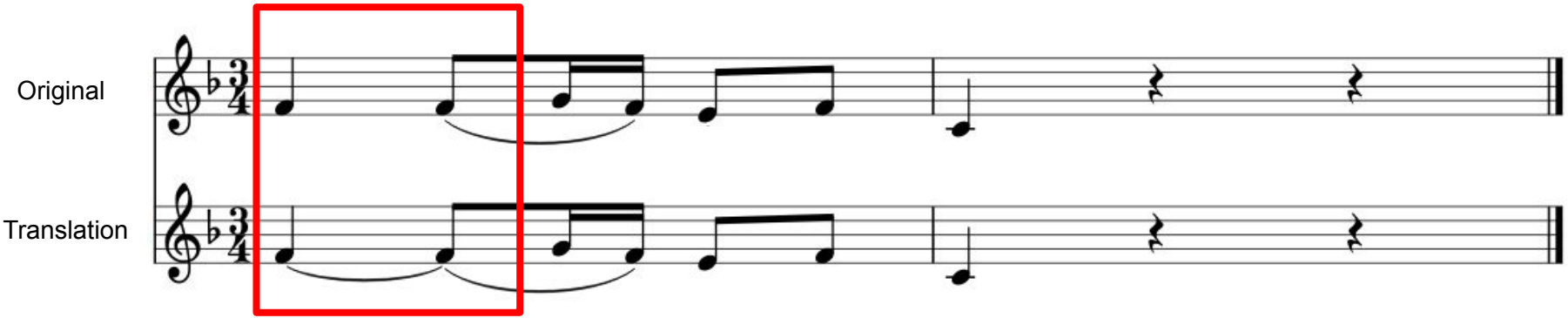
Having the same note/rest events, starting at the same offset **and preserving the same attributes (duration, articulation, and ornaments)**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

What is *equivalent (in a translation)*?

Having the same note/rest events, starting at the same offset **and preserving the same attributes (duration, articulation, and ornaments)**



Beethoven Op.18 No.1 - I. Allegro, measures 1-2

Discrepancies (Translation)

Synchronized: note/rest events in the original score that **were found** in the translation

- Identical: note/rest events that have preserved their properties
- Different: note/rest events that have **not** preserved their properties

Non-sync: note/rest events in the original score that **were not found** in the translation, at least not in the offset of the original

Encoding vs. Translation	Synchronized Identical / Different	Non-sync
A vs. A1	95.2% / 4.8%	0.0%
A vs. A2	6.1% / 48.5%	45.4%
B vs. B1	95.1% / 4.9%	0.0%
B vs. B2	18.6% / 5.3%	76.1%
C vs. C1	94.4% / 5.6%	0.0%
C vs. C2	26.1% / 2.3%	71.6%

Discrepancies (Translation)

Synchronized: note/rest events in the original score that **were found** in the translation

- Identical: note/rest events that have preserved their properties
- Different: note/rest events that have **not** preserved their properties

Non-sync: note/rest events in the original score that **were not found** in the translation

Found in the translation

Encoding vs. Translation	Synchronized Identical / Different	Non-sync
A vs. A1	95.2% / 4.8%	0.0%
A vs. A2	6.1% / 48.5%	45.4%
B vs. B1	95.1% / 4.9%	0.0%
B vs. B2	18.6% / 5.3%	76.1%
C vs. C1	94.4% / 5.6%	0.0%
C vs. C2	26.1% / 2.3%	71.6%

Discrepancies (Translation)

Synchronized: note/rest events in the original score that **were found** in the translation

- Identical: note/rest events that have preserved their properties
- Different: note/rest events that have **not** preserved their properties

Non-sync: note/rest events in the original score that **were not found** in the translation

Not found in the translation

Encoding vs. Translation	Synchronized Identical / Different	Non-sync
A vs. A1	95.2% / 4.8%	0.0%
A vs. A2	6.1% / 48.5%	45.4%
B vs. B1	95.1% / 4.9%	0.0%
B vs. B2	18.6% / 5.3%	76.1%
C vs. C1	94.4% / 5.6%	0.0%
C vs. C2	26.1% / 2.3%	71.6%

Conclusions

Investigate the **discrepancies** between symbolic music files that (intend to) represent the same music score document

We hope that the better understanding of these discrepancies may help to

- Compare symbolic music files to find discrepancies
- Evaluate the quality of symbolic music corpora out there
- Mitigate *known* patterns of discrepancies automatically

Thank you

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

McGill

Schulich School of Music
École de musique Schulich

DDMAL DISTRIBUTED DIGITAL MUSIC ARCHIVES & LIBRARIES LAB

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

Fonds de recherche
Société et culture
Québec

Future work

Search for some of these patterns in large corpora to find their relevance in the existing corpora

Write routines to attempt to auto-correct these discrepancies when they are unambiguous (e.g., if a tie starts in XML-based formats, it must conclude)