Close and distant reading of music theory

Frans Wiering, f.wiering@uu.nl
Paris, 16 November 2019
Abstract

In the late 1990s I embarked on a project to digitise Gioseffo Zarlino’s music-theoretical writings. They formed one of the cores of the Thesaurus musicarum italicarum (http://tmiweb.science.uu.nl/), which now contains over 30 digitised music treatises from the 16th-18th centuries. The general idea of the project was to make the treatises better accessible by providing corrected, lightly edited texts shown in various ‘views’, illustrations, linking, annotations, and search functionality. In effect, we created scholarly hypertexts. Even though the hope was that treatises would be less selectively studied, it is easy to observe with hindsight that we were enhancing close reading. With the advent of large text corpora, however, a new approach to source study has emerged as well, which Franco Moretti dubbed ‘distant reading’. Distant reading takes as its starting point that it is not humanly possible to read all relevant materials in their entirety. Nor is this necessary, because by using techniques from computational linguistics and big data analysis we may find meaningful patterns that would otherwise go unnoticed. While I haven’t systematically done any such analysis on the (not so big) corpora of digitised music treatises, I will report a few initial experiments and speculate about their potential for the study of the history of music-theoretical concepts.
Between modes and big(gish) data

- my talk is mainly about technology for studying theoretical sources
- initial motivation came from modes of polyphony
- fascinated by the writings of Bernhard Meier (1923-1983)
  - hermeneutic approach, close-reading of music and theory
  - aimed towards a reappraisal of ‘the modal system’ as it functioned at the time
Between modes and big(gish) data

- I had some issues with Meier
  - ignores often considerable differences
  - static view, no historical change
  - selective evidence
- more of an empirical approach
  - larger timespan (mid-13th to mid-18th century)
  - collect and analyze as much evidence as possible: music, theory, literature, diaries, art...
- outcomes
  - then as well as now, modes were often a contested topic
  - viewpoints expressed in text and in music
  - substantial amount of new evidence

finished 1995
Between modes and big(gish) data

polyphonic modal cycles

---

Illud prius dicemus / Tonus... esse quandam canendi normam / quae in fine per ascensum atque descensum cognoscitur (Aaron 1516, chap. i:26).


Tonus regula dicitur, quae per ascensum et descensum omnes descriptas ac etiam pernotabiles modulationes in fine diiudicat (Vanneo 1533, fol. 29°).

Tonus est cognitio principij, medij ac finis cuiuslibet cantus, ascensus & descensus iudicans (Cerone 1613, p. 350).

Tonus est regula per ascensum & descensum, omnes descriptas ac etiam pernotabiles modulationes in fine diiudicans (Cerone 1613, p. 350).
Between modes and big(gish) data

• ‘big data’ research enabled by latin treatises digitised in *Thesaurus Musicarum Latinarum*

• but at the time no Italian treatises were available...

• so why not digitise these as well?
Thesaurus musicarum italicarum

- founded in 1996
- ‘multimedial’ editions of Italian-language music treatises
  - searchable hypertext
  - illustrations, music, facsimile
  - textual criticism, linking, enrichment, annotation, collaboration
- use state-of-the-art technology of 1990s
- Zarlino’s works as guinea pigs
Pietro Aaron  
5 treatises

Agostino Agazzari  
2 treatises each

Valerio Bona  
Giovanni Del Lago  
Pietro Pontio  
Rocco Rodio  
Orazio Vecchi  
1 treatise each

Gioseffo Zarlino  
3 treatises

Bernardino Baldi  
2 treatises

Girolamo Mei  
1 treatise

Fedele Fenaroli  
1 treatise

Orazio Tigrini  
1 treatise

TmiWeb (2000)

Vincenzo Galilei  
2 treatises

Hercole Bottrigari  
1 treatise

Giovanni Maria Artusi  
6 treatises

Nicola Vicentino  
1 treatise
What is TmiWeb?

TmiWeb is the online version of Thesaurus musicarum italicarum, an encyclopaedia of Italian music treatises from the Renaissance and Baroque periods. It contains the works of two of Italy’s most important and influential composers: Aaron (c. 1480-c.1545) and Gioseffo Zarlino (c. 1517-1590) and their writings by their contemporaries.

Nearly 30 works are online on TmiWeb in a fully searchable format. A list of texts can be found [here](#).

All texts are encoded using a slightly-extended variant of the TEI P5 standard. TmiWeb employs the Kiln publication environment developed at King’s College, London.
Scholars reading sources

- common behaviours
  - close reading
  - complex reading
  - browsing, exploratory search
- obviously TMI enables these

- distant reading (Morretti 2013)
  - there is just too much to read
  - discover patterns computationally
- how could TMI enable distant reading?
Exploring technologies

1. of-the-shelf text mining product
   - Voyant, https://voyant-tools.org/
   - TEI markup a bit of a nuisance

2. home-baked Python programs
   - programming a bit of a nuisance
   - can make good use of TEI markup

3. if time, dubious bonus with TML

NB these are *not* attempts at doing musicology
Attempt 1

TMI: Aaron, Pietro, Compendiolo

TMI Menu Texts Search Admin

Aaron, Pietro, Compendiolo (after 1545)

Navigation

DEL CANTO FERMO.


Dello letterato ordinate nella mano. Cap. 2.

Dichiarazione delle note. Cap. 3.

This corpus has 34 documents with 1,923,141 total words and 63,996 unique word forms. Created about a day ago.

Document Length:
- Longest: TMI: Zacconi, Lodovico, Pr... (259039); TMI: Zacchino, Gioseffo, Is... (241088); TMI: Zerino, Gioseffo... (21268); TMI: Zacchino, Giosseffo, So... (192965); TMI: Zacchino, Gioseffo,D... (158379)
- Shortest: TMI: Agazzari, Agostino, D... (2628); TMI: Vecchi, Orazio, Most... (3068); TMI: Agazzari, Agostino,D... (3095); TMI: Artusi, Giovanni... (3841); TMI: Aaron, Pietro, Tratt... (5269)

Vocabulary Density: 1

items: 12,610 context 0 expand ? Scale ?
aartos

toscanello in musica
di messer piero aron fiorentino del ordine hierosolimitano et
canonico in rmini, nuovamente stampato con l'aggiunta da lui fatta
et con diligentia corretto.
ad totus italae juventae de petri aron egregii musici laudibus io.
gazoldi epiq.
si vis scire modum generousa juventae cantando
petrus aron clarus musicius arte docet.
edocet et coeli non nula minuita tanta
organum pulsino.

Summary

This corpus has 10 documents with 1,011,411 total words and 33,613 unique
words, about 21 hours ago.

Document Length:
- Longest: zarains58 (394389); zarins58 (263954); zarsof (181818); zarim (153519); vicant (100197)
- Shortest: aartm (4456); aart (13936); aarcom (14503); aartos (44779); aaruic (44806)

Vocabulary Density:
- Highest: aartos (0.140); aarcom (0.138); aarim (0.134); aartos (0.113); aaruic (0.085)
- Lowest: zardim (0.034); zarins58 (0.060); vicant (0.061); zarins58 (0.065); zarsof (0.077)
Demo information

- A video of my demo can be found at https://bit.ly/2MpQk13

- You can experiment yourself with Voyant to analyse this corpus at https://bit.ly/2Mp1Tpg
Exploiting the markup with Python
Zarlino reading the classics

<table>
<thead>
<tr>
<th>Title</th>
<th>Zarins58</th>
<th>Zarins89</th>
<th>Zardim89</th>
<th>Zarsop88</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boet.anic-ins-mus</td>
<td></td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>Ptol.CLAuer</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eucid-ede</td>
<td>70</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bible</td>
<td>40</td>
<td>30</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Vergil-aen</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pline-giuri-nat</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Suda</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Horace-arco-poe</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Aristot-met</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Piet-aban-pro</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Aristot-philo</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Aristot-ani</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Cleori-mus</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Boet.anic-ins-ari</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Aristoi-mus</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Aristoi-ana-pos</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Vitruv-arc</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Plato-rep</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Aristoi-pol</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Plutar-mus</td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Zarlino reading Boethius' *Musica*
Chapters Zarlino was interested in

1.1 Introduction: Music forms a part of us through nature

1.20 Concerning the additions of strings and their names

1.23 Which ratios are between pitches in each genus
Chapters Zarlino was interested in

4.3-4: note names and diatonic monochord partition

4.13-17 species of consonances and modes

5.11 The manner in which Ptolemy set out the consonances
What about Vicentino?
What about Vicentino?

1.21 Concerning the genera of song

1.23 Which ratios are between pitches in each genus

4.6 Partition of the monochord of the netai hyperboleon through the three genera

4.12 [Rationale of the monochord partition diagram]
Different reading practices

NB percentages per author, not absolute numbers
Bonus: TopicModellingTool on TML

1. ut numerus proportio lib siglum numeri maior gf vi pars
2. modal fa css search copyright concordances submenu aria hidden about
3. antiphona gf versus domine dominus respensorium amen deus delta gloria
4. diapason diapente ut diatessaron proportio semitonium tonus consonantia toni proportione
5. gf fa tenor breuis tertia discantus quinta semibreuis sexta semibreues
6. gf ut cantus exempla quatuor cantum unum tenor csm tres
7. ut fa gf sol sqb la cantus Toni tonus voces
8. vt sv mv svcd staff vna on pe duo quatuor
9. diatessaron diapason diapente meson species hypaton diezeugmenon prima tono hyperboleon
10. ut punctus natura anima corpus corporis mundi motus naturae animae
11. lig ut longa staff on brevis pt tres duas prima
12. ut diapente diapason voces lib siglum diatessaron vocum quatuor cantus
13. hec musice musica dyapason semitonium tercia eciam dyapente due figura
14. ut gf caput fa gladod sol la exemplum finpra uocum
15. musica sonus sonum libro musicae ratione genus musicam omega sonorum
What have we learned?

**Text**
- markup may be an obstacle
- what elements (not) to select from docs
- importance of stopword removal
- deal with language variation (spelling, dialects, change over time)
- multiple meanings
- multilinguality

**Software**
- importance of language support
  - stemming
  - POS tagging
  - named entity recognition
- parameter tuning
- document or corpus scope
- tool transparency
- complex questions tend to require programming
Interpreting the outcomes

- not an easy thing...
- it takes a while to understand the potential of most tools: practice tool criticism
- usually large amounts of ‘noise’ in results
- patterns are everywhere: they don’t mean that much by themselves
- minimally, one should start with an intuition or exploratory question
- do the outcomes make sense as arguments in musicological reasoning?
Final questions

• what research problems are suited to distant reading approaches?

• what tools are needed and what properties should these have?

• what data are needed and in what shape?

• who is going to produce these?