

This section shows all the available open-source datasets we found with both symbolic music and chord labels in the format of Roman numeral and chord letters.¹ Note that all datasets with Roman numerals can be translated into chord letters heuristically.

- Theme and Variation Encodings with Roman Numerals (TAVERN) (Devaney et al. 2015): 27 themes and variations for piano (10 by Mozart and 17 by Beethoven) in **kern format, divided into 1060 phrases, annotated with Roman numerals.²
- Annotated Beethoven Corpus (ABC) (Neuwirth et al. 2018): all Beethoven string quartets (16 string quartets, 70 movements) in MuseScore format, annotated with Roman numerals.³
- Roman Text (Tymoczko et al. 2019): contains 24 preludes from the first book of Bach's Well Tempered Clavier, and 48 romantic songs from France and Germany, annotated with Roman numerals.⁴
- Kostka-Payne Korpus: 46 excerpts in the workbook accompanying the theory textbook *Tonal Harmony* (Kostka and Payne 1995), annotated with Roman numerals.⁵
- “Sun” Quartets (López 2017): the Op. 20, six string quartets from Joseph Haydn in **kern format, annotated with Roman numerals.⁶
- Beethoven Piano Sonata with Function Harmony (BPS-FH) (Chen and Su 2018): consists of the symbolic musical data for the first movements from Beethoven's 32 piano sonatas, annotated with both Roman numerals and chord letters.⁷
- The Rameau dataset (Kröger et al. 2008): 371 Bach chorales in Lilypond format, and 156 of them are annotated with chord letters.⁸

¹ To the best of our knowledge, no available dataset with figured bass annotations is found.

² <https://github.com/jcdevaney/TAVERN>

³ <https://github.com/DCMLab/ABC>

⁴ <https://github.com/MarkGotham/When-in-Rome>

⁵ <http://davidtemperley.com/kp-stats/>

⁶ https://github.com/napulen/haydn_op20_harm

⁷ <https://github.com/Tsung-Ping/functional-harmony>

⁸ <https://github.com/kroger/rameau/tree/master/rameau-deps/genos-corpus>

- Band-in-a-Box Jazz standards (Choi, Fazekas, and Sandler 2016): 2486 Jazz songs with chord letters in Jazz standards, available in Band-in-a-Box file format.⁹
- Weimar Jazz Database: 456 songs with scores and chord letters in Jazz standards.¹⁰

Chen, Tsung-Ping, and Li Su. 2018. "Functional Harmony Recognition of Symbolic Music Data with Multi-Task Recurrent Neural Networks." In *Proceedings of 19th International Society for Music Information Retrieval Conference*, 90–97.

Choi, Keunwoo, George Fazekas, and Mark Sandler. 2016. "Text-Based LSTM Networks for Automatic Music Composition." *ArXiv:1604.05358 [Cs]*, April. <http://arxiv.org/abs/1604.05358>.

Devaney, Johanna, Claire Arthur, Nathaniel Condit-Schultz, and Kirsten Nisula. 2015. "Theme and Variation Encodings with Roman Numerals (TAVERN): A New Data Set for Symbolic Music Analysis." In *Proceedings of the 16th International Society for Music Information Retrieval Conference*. Vol. 728–734. <http://ai2-s2-pdfs.s3.amazonaws.com/5747/34ed2599a06e64150595b88cbf176a673945.pdf>.

Kostka, Stefan, and Dorothy Payne. 1995. "Workbook for Tonal Harmony." *New York*.

Kröger, Pedro, Alexandre Passos, Marcos Sampaio, and Givaldo De Cidra. 2008. "Rameau: A System for Automatic Harmonic Analysis." In *Proceedings of International Computer Music Conference*, 273–281. <http://www.academia.edu/download/3461326/cr1516.pdf>.

López, Néstor Nápoles. 2017. "Automatic Harmonic Analysis of Classical String Quartets from Symbolic Score." Master's Thesis, Universitat Pompeu Fabra.

Neuwirth, Markus, Daniel Harasim, Fabian Claude Moss, and Martin Rohrmeier. 2018. "The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets." *Frontiers in Digital Humanities* 5: 16.

Tymoczko, Dmitri, Mark Gotham, Michael Scott Cuthbert, and Christopher Ariza. 2019. "The Romantext Format: A Flexible and Standard Method for Representing Roman Numeral Analyses." In *Proceedings of the 20th International Society for Music Information Retrieval Conference*, 123–29.

⁹ <http://bhs.minor9.com/>

¹⁰ <https://jazzomat.hfm-weimar.de/dbformat/dbcontent.html>