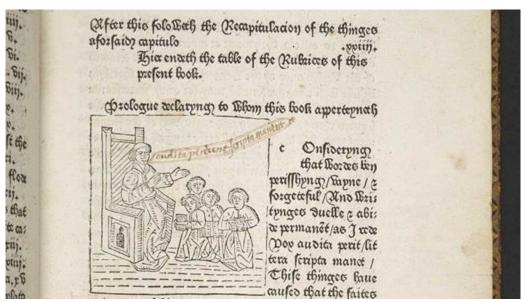
Incunabula catalogue data workshop

DHNB, Reykjavik 29 May 2024



William Caxton's 1481 edition of the 'Mirror of the World'. BL shelfmark IB.55040.

Agenda

Welcome

08:30-09:15 Intro to Catalogues as Data

Transforming Printed Catalogues into Data: Use case and data preparation (training data, transcription with <u>Transkribus</u>) Code development and data outputs

09:15-10:00

Hands-on Exercise:

Catalogue entry extraction from transcribed text using a <u>Jupyter notebook</u>

10:00-10:30 Comfort Break

10:30-10:50

Introduction to the corpus analysis tool AntConc: Set up, Data import, Tools and Setting

10:50-11:45

Hands-on Exercise: Analysis with AntConc

11:45-12:00

Closing discussion

LIBRARY

British Library Digital Research Team



Neil Fitzgerald Head of Digital Research



Stella Wisdom Contemporary British



Nora McGregor Europe & Americas



Dr Mia Ridge Western Heritage



Dr Adi Keinan-Schoonbaert Asia & Africa



Dr Rossitza Atanassova Digitisation



Deirdre Sullivan Business Support Officer



Harry Lloyd Research Software Engineer



Erin Burnand Universal Viewer Product Owner



Saira Akhter Research Software Engineer



James Misson Research Software Engineer



Lanie Okorodudu Senior Testing Engineer

bl uk

Catalogues as Data: The GLAM Context

- Understand better library and archival collections (publishing trends)
- Digitisation practice and resources (<u>LWM map/newspapers tools</u>)
- Metadata analysis and enhancement at scale (language, subjects, genres, named entities)
- Linked data
 - Some great examples:
- English Short Title Catalogue (Tolonen, Mikko et al.)
- British and Irish Newspapers Title List and Handbook (Ryan, Yann)
- <u>Automated Language Identification</u> (Morris, Victoria)

Harnessing Machine Learning and Al

- Use with new methods and tools
- Train language models
- Extract information and use crowdsourced information
- Build resources for GLAM professionals
- Facilitates computational analysis of collections

Some great examples:

<u>Introduction — Classifying 19th Century British Library books using</u>
<u>Crowdsourcing and Machine Learning (living-with-machines.github.io)</u>

Convert a Card project and web app

Corpus linguistic analysis with catalogue data

- Examine the outputs from automated text recognition (errors, inconsistencies, variants, punctuation, special characters/scripts)
- Understand the language used by cataloguers and curators over time
- Address any problematic or offensive terminology
- Identify patterns in the data and research questions
- Identify curatorial voices and any biases
- Explore the relationships between the descriptions (items, named entities)
- Explore the transmission of descriptions between catalogues and the relation to other resources
- Inform the development of inclusive descriptive curatorial and cataloguing practice

Corpus linguistic method tested by https://cataloguelegacies.github.io/

"Given a set of guidelines for producing curatorial descriptions, corpus techniques could be used to check the extent to which guidelines are being followed at a macro-level, e.g. by identifying what aspects of objects tend to be referred to or not, and by gauging the overall extent of description versus interpretation and evaluation.

Further, such analysis could form a basis for plans to edit and enhance a catalogue by providing areas to focus on and estimates of the person time required. It could also be that a corpus-based characterization of the language used in an exemplary catalogue could be used to develop or refine guidelines by identifying that catalogue's distinctive linguistic features."

Salway, Andrew; Baker, James (2020). *Investigating curatorial voice with corpus linguistic techniques: the case of Dorothy George and applications in museological practice.*University of Sussex. Journal contribution. https://hdl.handle.net/10779/uos.23475305.v1

Incunabula catalogue: Project background

AHRC-RLUK funded Professional Practice Fellowship Project (2022-23): <u>Legacies of Curatorial Voice in the Descriptions of Incunabula Collections at the British Library</u>

Project aims to demonstrate new ways of working with digitised catalogues that would also improve the discoverability and usability of the collections they describe.

The focus of the research is the *Catalogue of books printed in the 15th century now at the British Museum* (or *BMC*) published between 1908 and 2007 which describes over 12,700 volumes from the British Library incunabula collection. By using computational approaches and a corpus analysis tool with the catalogue data we gain new insights into this valuable resource and enable its future reuse in contemporary online resources.

Project summary is available at https://app.transkribus.org/sites/BL-Incunabula

Incunabula catalogue: Data preparation steps

Transkribus:

- Training the layout recognition models
- Optical Character Recognition
- Transkribus Outputs

Our text data is noisy. This is due to some errors in the layout recognition and special characters/abbreviations in the transcriptions of the titles.

Early experiments

LEIPZIG

prunknofleri; inside the last cover; Sum Dinci Adamir Persii Anno 1563 a Dño Valentino Waldenburg Parrocho Phichoulensi tc; and on 1^a: Collegii Soc. Jesu Vien . . . Anno 1634. Old boards, backed with stamped pigskin. Bought in January, 1878. On the core: IA. 12224 Gragorius Parsins 1563.

ARISTOTELES. Paruulus philosophiae moralis in Ethicam Aristotelis introductorius.

18. TITLE: Paruulus philo zophie Moralis. 28. Par-uulus philozophie Mo ralis in Ethicam Arestotilis Introductorius. 203. END: ... qui sine fine viuit et regnat Amē. || Deo gratias.

Quarto. aaa bbb*ccc*ddd*. 22 leaves, the last two blank. 3*: 20 lines, 148×88 mm. Types: 160, title, heading on 2*, and *Deo gratias' at end; 8t, text (leaded). Capital spaces. 212×158 mm. Bound with IA. 22560, IA. 6700, and other tracts. Old boards, backed with stamped pigskin,

with remains of clasps. Bought in May, 1864. IA. 12229.

ARISTOTELES. Paruulus philosophiae naturalis in Physicam Aristotelis introdu-

1*. TITLE: Paruulus philozofiphie naturalis. 2*. Paruulus philozophie nalturalis in phisicam Arestotilis Introductorius. 15th. COLOPHON: Finit Paruulus | philozophie naturalie.

Quarto. Aa Bb*Cc*. 16 leaves, the last blank. 3*: 19 lines, 40 x 87 mm. Types: 160, title, heading on 2*, and first line of colophon; 81, text (leaded). Capital spaces. Hein 25337 202 × 142 mm. Bound with IA, 4646 in old boards backed with stamped leather.

Bought in November, 1888.

Inside the first cover is a note: Ex ibris Cassiani 3 column HANUS DE MONTE. Campus sophi-*14 August, 1496. starum. 18. TITLE: Campus sophistarum. 18. Laurentiu

Bernsprunck | Zwicaulensis; artium magister lectori Salutem dicit. 28. Campus sophistarum Stephani de monte: viri ingenio || acutissimi / ordinis beate Marie de monte Carmeli : quodlis bet scibile sophistice insolu-biliterg interimere docens p Lau rentiu Bernschprunk Zwicauiensem in florentissimo Lips czensi studio repetitus / ad vnguemca castigatus incipit. 5b. END: . . positionis respondentis. 68. [Diagram.]

Quarto. at. 6 leaves. 28: 35 lines and a space, 150 × 95 mm. Types: 160, title and headings: 81, text. Woodcut diagrams, each with a woodcut letter. Capital spaces, with guide-letters.

210×153 mm. Capitals and initial-strokes supplied in red. Industry games Good c. RF 3851 Bought in March, 1874. IA. 12237.

MONTE RUTILO, SAMUEL DE. Synonyma aduerbiorum.

1º. TITLE: Sinonoma partium indecli nabilium mastri Samuelis de monte rutilo epistolari nors||ma ontexta. loquendica ornată ac maioreimodă efficaciam | eloquentie coducentia. 22ª. COLOPHON: Partium indeclinabilium Sinonima per magistrum Sallmuelem de nonte Outilo congesta feliciter finiunt | Laus deo.

Quarto. A⁴ B⁴C D⁶. 22 leaves. 3⁸: 37 lines, 150×92 mm. Types: 160, first line of title and headings; 81, body of title and text. Capital spaces. Hain *11500. The paragraph-mark does not appear in this book, but the upright

205 x 140 mm. Capitals, paragraph-marks, colons,

nitial-strokes, and underlines supplied in red. Bought in March, 1906.

IACOBUS THANNER

DATES: according to Maittaire's Supplementum, p. 60, an edition of Aquinas De proprio motu cordium, &c., was signed by Thanner and dated 1495. After this no book of his is known till 1498, in which year he printed sixteen books, the first with a precise date being Horace, Carminum liber quartus, completed 7 July. Thanner continued printing till 1529; a change of paragraph-mark made between 1501 and 1503 helps to distinguish between his undated incunabula and his books of the 16th century. (513 (2f Bückerfe, Murch, 1920, 4-278)

TYPES (see Plate LXII): 88 [P. 1], German text type, apparently the same as Hochfeder, Nuremberg, 88 [P. 9] and Schenck, Erfurt, 86° [P. 5], distinguished from Landsberg 88 [P. 1] by the hyphens (both single and short double, whereas Landsberg is is short and double and is reversed), smaller C, larger E with shorter middle stroke to the right, O and Q with single loop, where Landsberg has the Latin diamonded forms, round T instead of square, also by the fantastic paragraph-mark, which was superseded between 1501 and 1503.

175 [P. 2], title type, of pointed church style, very like Landsberg 172 [P. 2], but B has a longer stroke at the

top and is not 'thorned'; the hyphen also is not so steeply sloping.

DEVICE: in white, on black ground, an incomplete circle surmounted by a cross flanked by the letters i t, with two transverse bars, the upper one bent into a semicircle in the middle, in frame, 3,5 × 3,5 mm.

column3ARIS. Epistolae. *31 May, 1498. 6 1* TITLE: Epistole Phalaridis per || Franciscum | CImpressum Lipczk per Iacobum Thanner Her||bipolenser, Anno salutis christiane. M. cccc lxxxxviji || die aretinum tra ducte. | [Woodcut.] 1b. Iohannes Honorius | Ienses, Anno salutis christian | Crisous Cubitensis | Jectori felicitatem. 40b E.ND: bat | Nune Phalaris doctum protulit ecce caput. 41* | Counto. A-F'-C'. 44 leaves, the last blank. 2*: 33 lines, 145 | xds mm. Types: 755, tide; 65, seat. Capital spaces, with guide-levers. Ham 1750-C. 656

us deo. 43°. C Ad lectorem ... COLOPHON:

COLOGNE

156 [P. 3], not reproduced. Small church type on large body used for signatures of IB. 4477. Dr. Haebler's

measurement.

33° [P. 4] semi-cursive German text type. In use in 1484 and 1485. Two forms of D. Differs from 93° in D. L. Q. S. V. and d. Indistinguishable from Gritinger 93 [P. 5] except in having two forms of D and in florid waved H, which is that of Grinninger 93. T from a different fount occasionally found. The control of the property of the p

IA. 4489, which is probably after 1500.

81 [P. 6], Latin text type. In use in 1492. Indistinguishable from Grüninger 80 [P. 13], except L, P, Q, T, V, and perhaps K, which are indistinguishable from Grüninger 86 [P. 7].

40? [P. 7], not reproduced. Minute type, used for guide-letters in IA. 4469-70, dated 1492.

40? [P. 7], not reproduced. Minute type, used for guide-letters in IA. 4469-79, dated 1494.
85° [P. Div. B, x²], not reproduced. Large text type. In use in 1455. Indistinguishable from Grüninger 90 [P. 2], except possibly by slight difference in serif of V.
85° [P. Div. B, x²], not reproduced. In use in 1457. The same type as the preceding, with the addition of the N of 94°. Majuscules from a different fount mixed in IA. 4514.
92 [P. 8], Dutch type. Dr. Hachber's measurement: 97, Very like Bumgart [P. 6], but A and T do not seem to belong to the type. Found only on title-page of IA. 4489, which is probably later than 1500.
180 [P. Div. B, 3], large type for headings, õc. In use in 1485] and 1495°. Similar to the earlier form of Grüninger 160 [P. 1], but straight is common, and lower-case (e.g. e and o) seems to have more forms. Mr. Proctok's upper 1-5° IZ of 1493 Vit. Pata: apparat to be this type, created or cut down to 160. See IR. 4357.

ant. record style 1 of r. o. 14 of 3 vit. rat. appears to be two type; read of cut down to 10. See 11. 43 pt.;

10. P.F. Div. C. 1 is large text type. Is found of A with Species 10. Co. 13 and P are all found in Pr. of

10. In the Co. 1 is large text type. The found of the control of the course, is therefore uncertain (see

10. P. Pr. of 148 Vit. Patr. 4], text type, found of by in 13. 45 pt. Tall of a level with line. Apparently

a pure form of 81 [P. 6], without the admixture as from Grininger 85. Differs from Grininger 86 [P. 13] in

Seelentrost. [In Low German.] *1484.

1b. [Woodcut.] 2h. [L]Iber iste col||lectus est exdiuersis | libris De biblia. De | passionali . . . 1108. Hyr na volget der || clevn selen troist vii || leret vns van den seuen sacrame ten der hillygen kirchen. [1518. Hyr begynnet de | tafel ouer den selen troyst . . . 154%. COLOPHON: Dyt buoch ys ges druckt to Collen durch ludewich | van renchen | ym yair vnses herê | M. cocc.

Folio. $[a-o^{0.6}p q^6 r^6 st^6 v^6 x^6; y^6]$ 154 leaves, 3-:50 so numbered, with some errors. 2 columns. 4*:43 lines, 200 \times 125 mm. Types: 158, headings; 93°, text. Spaces left for capitals. The woodcut on 1b and the first page of text have woodcut borders first used by Knoblochtzer and afterwards passing to Koehoff (see IB. 3605, p. 228). With woodcuts. Voullième 1065.

285 x 207 mm. Wanting the four last leaves. Capitals and initial-strokes supplied in red. Traces of early manuscript signatures. The border and woodcuts are

Bought in July, 1859.

IACOBUS DE VORAGINE. Dat duvtsche

des duyd||schen passionails mit synë tzogesetze||ten ||impressa ad laudem sum||me monadis. xvij. kaleno. historien mit groissem vlijss gear beit. In de iar vns here. M. cccc. lxxxy | vp sent Marien magdalenen auent. || gedruckt durch mich Lodovuich van || renchen, burger tzo coellen - 4 2518. Dat register des eijrsten deils des duydschen | passionails . . . 251b. END: Van der verderfnisse ind verstoe | ring der stat jerusalem. xeix.

PART II. 18. TITLE: Dat ander deil des | duythschen passionails : 250*, COLOPHON : Hier eyndet dat ander deyl | des duydschen passionails van der hylli|ghen leuc durch dat gantze jair mit etzillighe tzogesetzden historien in dem iair | vns herê. M. cccc. lxxxv. vp alre hillighë | auent. Ind hait gedruckt tzo der eeren | gods Ludovuich

END: . . . Xj. dusent jonffrouwë, c.li.

Folio. Each part: [n-2 A-I'A K-N'A]. 252 leaves, the las From Each part: [a=2 A-1 to K-Not]. 22 leaves, the last blank, 3-25 numbered ii]-CCli in pt. 1 and ii]-CCli [omitting xcvii) in pt. 11. a columns. Pt. 1, 3*; 43 lines, 158 x 13 mm. Types: 325, title; 158, headings; 93, text. Spaces, many with guide-letters, left for capitals. The first page of text in each part has Knoblochter's woodcub broders, those on a*being the same as on 25 of IB. 4485, but with the lower piece reversed. With four woodcut capitals and many woodcuts. Voollième 626.

283 x 197 mm. Capitals supplied in alternate red and blue; paragraph-marks, initial-strokes, and underlines in red. Dr. Kloss's copy. Old stamped leather (Manstricht? Bought in 1835. IB. 4455-4456.

DESTRUCTORIUM VITIORUM.

16 August, 1485.

[18, TITLE: Summa que Destructori||um Viciorum appellatur.] 2^a, Tabula ppendiosa pm alphabeti ordinem... 21^a. Explicit tabula. || Sequitur capitulatio huius libri. 276. [O]Mne peccatū vt dicit bea||tus augustinus cotra fau| .. 203*. COLOPHON: Insignis notabilisos | compilatio haud modicum cuica statui conferens || omne genus Passionail. [Legenda Aurea in Low Ger: vitiorum suis cum speciebus clarissis me euidenteres *21 July-31 October, 1485. eradicans, ob id non immerito des structorium vitiorum nuncupata. a cuiusdam fabri | lignarij filio, maximam ad 1*, TITLE: Dat duytsche passionail. || → Dat eyrste | ecclesie vtilitatem Anno || M. ccc. xxix. collecta. de dell → 250°. COLOPHON: Hyer eyndet dat eijrste deyl | nouo Colonie exactissime || correcta. ac summo studio septembris. Anno domilni Millesimo quadringentesimo octuagesimoquin|to finita.

Folio. **Pit***c!*, d=*p***e**p**e**a-O**. 294 leaves, the last blank. 2 columns. 22**: 59 lines and head-line, 265 (281)×182 mm. Types: 126, head-lines and headings; 80*, text. Spaces left for capitals, that on 21* measuring 90 mm. The head-lines give the parts and chapters. Volliliem; 95.8. Hain *550.

357 × 271 mm. Imperfect, wanting the title and blank leaf. Ornamental capital in blue, red, and yellow on 278; other capitals supplied in alternate red and blue, paragraph-marks and initial-strokes in red. On 2* is written Simon Gunton'; on 293*, 'Commitati see trinitatis ex

MARTIN LANDSBERG

space on 28. Hain *11456. 205 × 140 mm.

Bought in February, 1876,

PROCLUS DIADOCHUS. Sphaera.

18. TITLE: Procli Diadochi Sphera | Astronomiam discere inci||pientibus vtilissima Thoma Linacro Bri-tanno || Interprete: ad Arcturum: Cornubie Valliegi ils lustrissimum Principem. 15th. . . . supra horizon-tem extante. 15th Cintroductoria additio ad precedentia. 16º. (to right) END: € Terti9 numer9 subter. ex | duobus superioribus distras [hendo relinquitur.

Quarto. A⁶ B⁶ C⁸. 16 leaves. 3^b: 17 lines, 144×95 mm. Types: 172, lines 1, 2 of title, and heading on a^b; 88, test (leaded), body of title, preface on a^b, and addition on 14^b and 15^b. Woodcut diagrams. Capital spaces, with guide-letters. Hain 13387. First published at Venice by Aldus in the Astronomici Veteres,

A woodcut on 1b is apparently that cut for the Sphaera mundi of

200 × 147 mm. Bought in October, 1871.

WIMPINA, CONRADUS. Palillogia de theologico fastigio.

1*. TITLE: Palillogia de Theologico fasti gio Ex nobilitate obiecti eius || Christi reparatoris et glorificatoris nostri sumpta. In || qua tribus codicillis nobilitas Christi ostenditur Quou Primus de Christi parentela : natura et gratia | C Secundus de Christi virtute et sapientia. €Tercius de cius priuilegij singularitate. 2º. €Vniuersis bonarum artium / s sacre theologie studio||sis Conradus ex fagis cognomento Wimpina / S . . . 34°. END: . . . can do lectori exhibendum curaui.

Quarto. A-C⁴ D⁴ E F⁶. 34 leaves. 5⁸: 34 lines and marginal headings, 145 x 93 (120) mm. Types: 172, lines 1, 2 of title, and heading on 3⁸; 88, body of title and text. Capital spaces. Hain

214 × 149 mm. Capitals, paragraph-marks, initial-strokes, and underlines supplied in red. Bought in April, 1871.

WIMPINA, Conradus, Panegyricus de

18. FIRST TITLE: Panegyricus de Christi dollmini nostri ecclesie immacu||late Sponse zelosissimi.|| ¶ Primus de christi Sublimitate || ¶ Secundus de christi admirabilitate || € Tertius de cius Bonitate || € Quartus de cius clementia ∥ ¶ Quintus de eius amabilitate. 1^b. ¶ Vni-uersis sacretheologie studiosis Conradus Wim∥pine ex Fagis Salutem. 6^b. Finis. [7^b. SECOND TITLE: Panegyricus de Christi do|mini reparatoris et glori||ficatoris nostri pientissimi: et ecclesie immaculate | sponsi dnirabilitate, 484 END: Finis, | DEVICE.]

A⁶[B-E⁶F⁸]. 38 leaves. 2⁶: 34 lines, 147×91 mm.
2, lines 1, 2 of title; 88, body of title and text. Capital
spaces. Hain *16213. (cond. 12.306*). 193 x 139 mm. A fragment, consisting of the first

IA. 12006. quire only. Bought in January, 1868,

SIBUTUS, Georgius, Siluula in Albiorim

illustratam. : *. TITLE : Georgij Sibuti Daripini || Poete et oratoris ureas ti : Siluula in Albiorim il llustratam. [[Woodcut.]]

28°. COLOPHON: Impressum Lipcz per Baccalaureum | Martinü lantzperg Herbipolitanü: | DEVICE. This, though recognized as an incunable both by Hain (14717) and Proctor (2083), cannot have been printed in the 15th century, as the University of Wittenberg, to which it refers, was not founded until

ARNOLDUS DE COLONIA

DATE: in 1492 Arnold signed an edition of Alexander Gallus, Tractatulus de accentibus, but without naming the place of printing; the Epistolae of Seneca, finished oil 26 june of the next year, are 'impresse Liptak p Arnoldi de Colonia'. The last books signed by him are another edition of Senecas' epistles, Guilelmus de Saliceto, De salute corporis, and Jacobus de Clusa, De arte bene moriendi, all dated 1495; but an edition of Johannes de Lapide, sante corpora, and piscolos yet canal, the titre user including a minimum, and a minimum to minimum the foot of the last page, wherever the end of the text might fall; he also printed the first line (and that only) of title and colophon in his larger type.

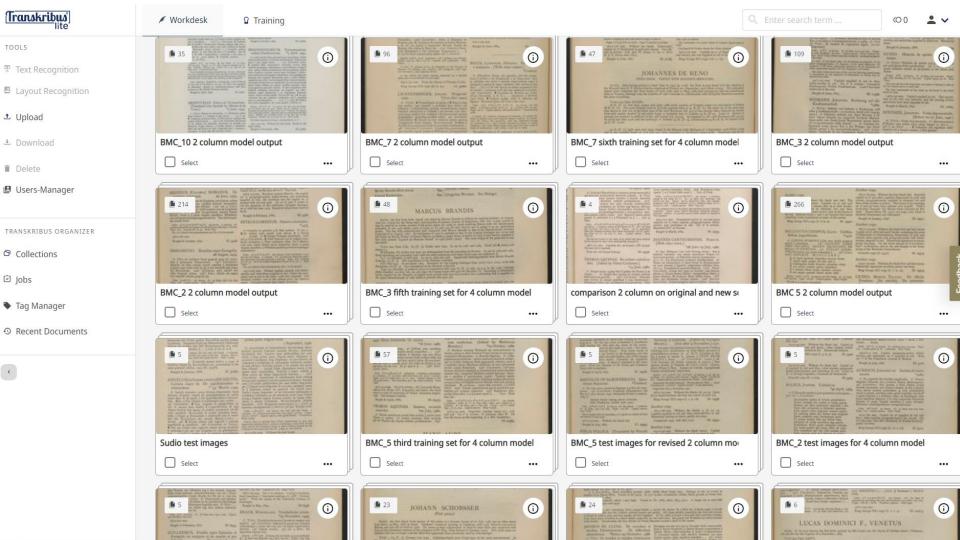
Tyrus: 81 [F. 1], text type of the common Strandburg style; square C (multilated E), L-without thoms, narrow diamonded N, plain V surplish choldle hyphen for two lines) and Ø. Inditinguishable from Cologne, Quentell, 86 [F. 2], by hyphen and L, N,V, and by absence of the alternative M (Isabele M⁽⁰⁾); from Strandburg, Printer of Causa, 80 [F. 1], by hyphen and L, N,V, and by absence of the alternative M (Isabele M⁽⁰⁾); from Strandburg, Printer of Causa, 80 [F. 1], by hyphen, N, and V; from Strandburg, Husser, 90 [F. 3], by hyphen and A, which is short. Not reproduced.

Strassurg, rates, so tr. o per pripages and st, which is not.

C. Bier T, is round, also E, which has a broken downstroke productionally supplied by Londardic forents.

C. Bier T, is round, also E, which has a broken downstroke, create D, I, V, N with a diamond, crossed P. According to Dr. Haebler R is supplied by the right-hand half of M, but this does not occur in the books described below; large double ungiful typhen. Not reproduced.

T is recorded as an ina-keeper in



BMC_8 4 column model output - Copy - Page 3

< 3 182

FLORIUS, FRANCISCUS. De amore Camilli iliae.—Leonardi Aretini libellus de samantibus Guiscardo et Sigisa ex Boccaccio transfiguratus.

Undated

1º. Francisci Florii Florētini.- de amore Camil·lli et Emilie aretnorum / ad Gullermum tar-lldiuum prologus feliciter incipit; || в la dhuc diucius qtardum ulixem ca-|sta expectauerti penelopez 4. / Francisci Florii Florētini de amore Camilli | et Emilie aretinorū liber feliciter incipit; 42°, СОСОРІЮХ: Francisci Florii Florentini / de || duobus mantibus liber felici-|ter expletus est turonis. editus || in domo.- domini Gullermi || archiepiscopi turonensis/pri-|die kalendas ianuarii. Anno do||mini.- millesimo-quadringente ||simosexagesimoseptimo; 42°, Incipit alius libellus de duobus amatibus || per Leonardia aretini in latinū ex boccacio || transfiguratus; || eVm sepius mecum egisses / . . . 43°. Sequit transfigurato; /| quā Leonardus de || Guiscardo et Sigrismunda Tancredi filia.- || in latinū ex Boccacio conuertiti: coº, Finis:

Quarto. [a-e¹⁶.] 50 leaves. 5^a: 24 lines, 133×80-2 mm. Type: 110 GR. Capital spaces, mostly with guide-letters. The verso of leaf 3 is blank.

The type of this book is very fresh and in what is here taken to be its first state.

178×125 mm. Rubricated. On the last page are written two Latin epitaphs, one by Pontanus on Petrus his 'compater' at Naples, the other from Brescia. Eighteenth-century blue French morocco, with the label of the Musseum Pauli Girardot de Prefond, and label of the Bibliotheca Heberiana, with a note in Heber's hand '100 Hibberts Sale 1829, 4 10, 0'.

Grenville copy (G. 10468). IA. 39:

RODERICUS [SANCIUS] ZAMORENSIS. Speculum vitae humanae. *Undated.

1°. Ad sanctissimū et beatissimum || dominū / dominū Paulum secum|dum pontuhcē maximū. bler inici|pit dictus Speculū humane uite || . . . || (I. 14) [S]Anctissimo ac || clementissimo || in christo patri || . . . 116°, COLOPHON:

Edidit hoc lingue clarissima norma latine.

Excelsi ingenii uir rodoricus opus.

Qui norma angelica è custos bene fidus in arce; Sub pauli ueneti nomine pontificis.

Claret in italici zamorensis episcopi ausis Eloquii.- it superos gloria parta uiri;

119*. De materiis pertractandis in || primo libro. et de tabula capitulo||rum eius. 122b. Incipit Reptoriū siue Tabu-||la per alphabetū... 124b, PRINTERS' COLOPHON:

Hos lege diuinos lector studiose libellos.

Vnde trahes uite commoda multa tue.

Hoc speculū clarū manibus gestare memēto.

Que tua sit uita noscere quisquis amas,
Nam tibi distinctum punctis / uirisqa relectū

Viris / perfinxit regia parisius.

Presserunt petrus cesaris / simul atq iohānes

Stol / qbus ars quod habet omne retulit eis;

Folio. [a-1¹⁶ m³; aa⁸] 126 leaves, 117, 118, 125, and 126 blank. 2 columns. 6⁵:33 lines, 183 × 125-7 mm. Type: 110 GR. Capital spaces, nearly all with guide-letters, except the one-line spaces in the alphabetical table. Hain 12018.

Reprinted from the edition of Friburger, Gering, and Crantz, 1472 (IB. 39018), with the addition of a table of contents (119*– 22*) and a second metrical colophon.

In this and the two following books the type is in its second

 277×202 mm. Without the last blank. At the foot of $124^{\rm b}$ is the formal signature of I, Defontemarie (?). Eighteenth-century French red morocco.

Bought in March, 1847. IB. 39202.

PIUS II. De duobus amantibus. Undated

[1*, Ence Siluii Senensis poete laureati uiri cla-|riasimile de duobus amantibus historia ad Ga-[sparum Silch milië ut cā legat feliciter incipit.] 2^b, I. 2: Prefatio. || [E]Neas siluius poeta / imperialises secretari-|lus Salutern plurimam dicit.* mariano ossi|no. ... 3^b, I. 20: Incipit opusculīt de duobus audītibus. 37^b, Colophon: Explicit opusculum Ence Siluii || de duobus amantibus;

Quarto. [ab10 c* d10,] 38 leaves, the last blank. 40: 24 lines, $132\times81-3$ mm. Type: 110 GR. Capital spaces.

This edition does not contain the Epistola iuueni non esse negandum amorem and the Epistola amatoria composed in the name of Duke Sigismund of Austria which are sometimes inserted after the Praefatio to Sozinus.

195 × 145 mm. Imperfect, wanting the first leaf and the blank. Leaf 3 is bound between leaves 7 and 8. Formerly part of a tract-volume.

Bought in July, 1888.

IA. 39205.

PIUS II. De miseria curialium. Undated.

1*. eNeas siluius poeta. S.P. dicit dño io||hāni de aich. pspicaci / et claro iurū cō-||sulto. Stultos esse / q regibus seruiūt. / || 30*, Солоном: Liber de miseriis curia-|| lium feliciter finit;

Quarto. [a-c¹⁰.] 30 leaves. 3° : 24 lines, 132×80^{-1} mm. Type: 110 GR. Capital spaces, with guide-letters.

The type in this book is very fresh. 196 × 144 mm. Formerly part of the same tract-volume

196 × 144 mm. Formerly part of the same tract-volum as the preceding (IA. 39205).

Bought in July, 1888.

IA. 39207.

DUUYIIL III WIAILII, 104/.

IB. 39202.

PIUS II. De duobus amantibus. Undated.

1a. Enee Siluii Senensis poete laureati uiri cla rissimi. de duobus amantibus historia ad Ga-sparum Slich militem ut ea legat feliciter incipit.) 2b, l. 2: Prefatio. ENeas siluius poeta/ imperialisque secretari us Salutem plurimam dicit. mariano sosi no ... 3b, l. 20: Incipit opusculu de duobus auiatibus. 376, COLOPHON: Explicit opusculum Enee Siluii de duobus amantibus;

Quarto. Jab'° ce d'°. 38 leaves, the last blank. 42: 24 lines, 132X 81-3 mm. Type: 110 GR. Capital spaces.

This edition does not contain the Epistola iuueni non esse negandum amorem and the Epistola amatoria composed in the name of Duke Sigismund of Austria which are sometimes inserted after the Praefatio to Sozinus.

195 X 145 mm. Imperfect, wanting the first leaf and the blank. Leaf 3 is bound between leaves 7 and 8.

ps://app.transkribus.org/sites/BL-Incunabula/browse

Incunabula catalogue data: Code development

Entry detection algorithm

Language detection

GitHub Repository

Exercise:

Follow the link to the <u>Binder</u> notebook to understand the catalogue entry extraction process.

Reusable datasets

<u>Incunabula catalogue dataset</u> published on the British Library Shared Research Repository include:

- 3500 Images (JPG) and corresponding XMLs (ALTO and PAGE)
- Catalogue entries in full and English-only descriptions (TXT)
- Metadata (CSV)
- Code

DH2023 Poster

COFFEE



Computational Analysis with AntConc

Intro to AntConc

<u>Set up</u> v.4.2.4

Our files are in this folder Download files for vols.1-10.

<u>Data import</u> - File-Open Corpus Manager. Raw Files. Add files. Create.

Tools and Settings (also described in the AntConc Help pages)

Global Settings - Restore Settings - Yes - saves your Settings for the next session

Word lists

Word lists are a useful starting point for getting an overview of the linguistic features of a corpus.

- Select the Word Tool and press Start to generate a word list, using the default settings.
- Note the top 30 word types (over 30% of the words in the corpus)? Do all the words/strings make sense?
- Select the case setting and run the query again. Note the difference.
- What named entities appear in the top 200 word types?
- Click on King to explore how the word is used in the corpus.
- Select the <u>Invert Order</u> setting to browse the long tail of infrequently used word types. Eg. mostly Latin words some names, strings of letters. Note some of the transcription errors!
- Sort by Type to browse word types alphabetically (capitalised hits at top). Eg. exchange (25) vs. exchanged (1)

For the domain-specific terms you may refer to this Glossary of Incunabula

If saving the outputs of your searches, ensure the filename includes information about your corpus, query, settings, or AntConc version. You can copy and paste the results or save results from the File menu (Ctrl+S)

KWIC (Key-Word-in-Context)

Search for a word or a 'string' to narrow your enquiry to a subset of the corpus. KWIC displays the results in a concordance.

Search for *appear* (case selected) as a word and then as a string (deselect word)

Sorted by pattern frequency in which the word appears (by default Sort to Right)

Use the Sort to left option to examine how 'appear' has been used by cataloguers, eg. with modal verbs, with negative words, the consistency of use.

Adjust the <u>Sort Options</u> according to which word relative to the search term you want to inspect.

Order by value to browse the word alphabetically (to the left or right).

This tool can help with comparative analysis of catalogue data, eg. language patterns, curatorial voices, across collections and over time.

For example the cataloguer of vol.7 and 8 adopted the phrase 'there appear to be no (sure) means of determining/deciding ...'

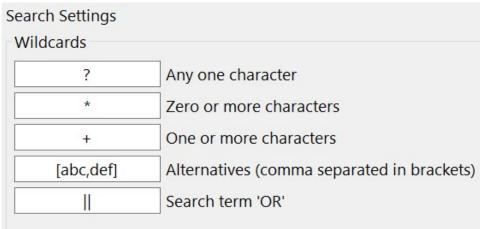
Expand the Context Size to see more of the lines and click on the word to go to the relevant File for close reading of the text.

Use the <u>Case Setting</u> to search with/without capitalisation - eg. 'Wanting the blank leaf' vs 'Imperfect, wanting...' These patterns reflect the rules of cataloguing, others - the cataloguers' linguistic choices - eg. 'was presumably printed **not later than/long after/very long after**....

Wildcards

- determin+ results show it is only used as a verb/verb form and not as a noun
- wear* (query as a word or a string)
- *Iy brings up mostly adverbs but also names and nouns, July, Lilly, fly. Note some transcription errors, hyphenated words
- Be||being||was||were||am||been
- In the <u>Word Tool</u> you can refine your search by defining the min <u>Frequency</u> or <u>Range</u> (size of the corpus)
- You can also use regex commands
- Choose a word type to explore in context
- Discuss with your neighbour

Wildcards help to explore known variants

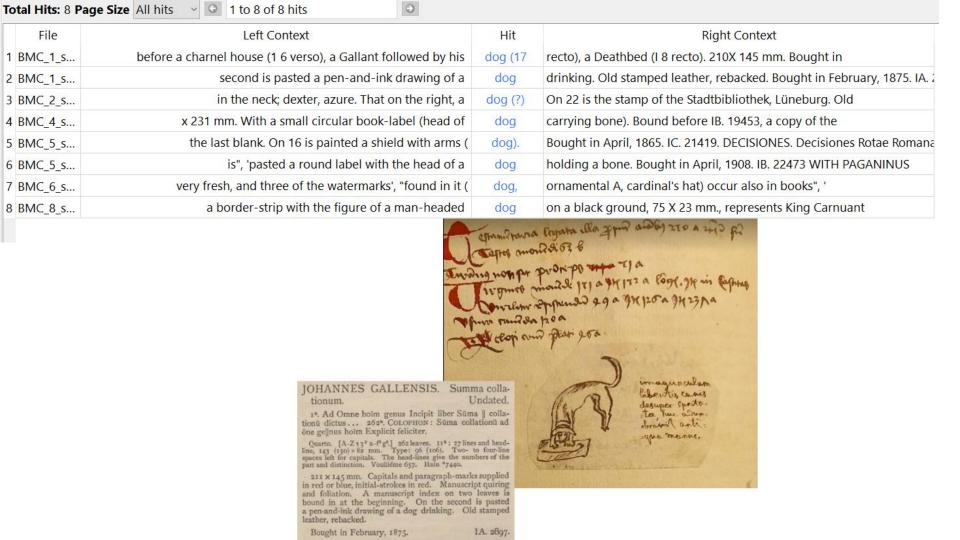


Unusual or historical language

	File	Left Context	Hit	Right Context
1	BMC_1	green cap, wearing a black sword, and one in a blue	jerkin	and dull crimson hose and cap. Title. De uidicta libertate
2	BMC_1	and cap and green hose, wearing sword; one in dark blue	jerkin,	hose, and cap; one in grey jerkin, green hose white
3	BMC_1	cap. Before them, with back to spectator, a man in blue	jerkin	and hose and scarlet cap. On the left a man
ı	BMC_1	trimmed with green. Before him four men: one in dull crimson	jerkin	and cap and green hose, wearing sword; one in dark
	BMC_1	hose white apron, and black hat, and one in dull crimson	jerkin,	blue hose, and green cap. Title. De seruis fugitiuis et
5	BMC_1	green-clad figure of the Prologue miniature. A man in grey	jerkin	and scarlet hose is leaving the room. Title. De edendo.
,	BMC_1	one in dark blue jerkin, hose, and cap; one in grey	jerkin,	green hose white apron, and black hat, and one in
3	BMC_1	in blue robe trimmed with grey fur, the other in green	jerkin	and hose, and holding a golden rod. Both wear dull
)	BMC_1	in his hand a docu ment with seals; one in red	jerkin,	grey hose, and green cap, wearing a black sword, and
10	BMC_1	robe and dull crimson cap; behind him a man in yellow	jerkin	and hose and dull crimson cap shows a book to

Appropriate use

File	Left Context	Hit	Right Context
1 BMC_1_s	capital on 22 supplied in red, green, and yellow, with a	grotesque	head, other capitals (with some omissions), paragraph-marks, initial-
2 BMC_1_s	page cuts. A five-line woodcut D on 22 and 4°; 15-line	grotesque	A on 2b ; spaces, with guide-letters, left for
3 BMC_1_s	phia, and the Muse Meretricule' on 12, 4°, and", "1092. Thirteen-line	grotesque	figure capitals. Hain '3359.", '183X 130 mm. On the title-page
4 BMC_8_s	evet, 1493). From the Old Library. Undated. Type: 114 G. Calligraphic	grotesque	L beginning title, Lom¬ bard Q on 22.', "The grotesque
5 BMC_8_s	Calligraphic grotesque L beginning title, Lom¬ bard Q on 22.', "The	grotesque	L'on 1e was also used in an unsigned
6 BMC_8_s	be earlier than this. The calli- graphic L with three	grotesque	faces which starts the title apparently occurs again, with
7 BMC_8_s	Passion, etc. are those used by Jehannot. The borders, representing	grotesque	beasts and faces and in one case fleur-de-
8 BMC 8 s	in this style used at Lyons. Calligraphic L, with two	grotesque	faces to the right, the neck of the upper



Collocates

Discover the words (collocates) that appear frequently near the search term.

Our corpus includes words for animals as part of descriptions of illustrations/woodcuts/watermarks, but mainly used in bookbindings descriptions.

Search for *sheep* (as a string, no case) and explore its collocate types. Results rankings show frequency to the left or right of search term.

Use the <u>Span</u> (proximity to search term) and <u>Min. Freq</u>. to refine the results. Adjust the <u>Range</u> (across how many volumes)

Note *sheep* and *sheepskin* are used interchangeably! This can inform revision of cataloguing rules.

Find the collocates of *before* and explore the <u>Sort</u> by options.

<u>Likelihood</u> and <u>Effect</u> are different ways of showing statistically significant relationships.

You can adjust the measures and thresholds in the <u>Tool Settings</u> for Collocates

Collocates

Shows words that appear frequently within a certain distance of the search term.

Find the collocates of *before* and explore the <u>Sort</u> by options.

<u>Likelihood</u> and <u>Effect</u> are different ways of showing the strength of the word relationship.

You can adjust the measures and thresholds in the <u>Tool Settings</u> for Collocates Refine the search by using the <u>Min. Frequency</u> and <u>Window Span</u> options.

Cluster and N-Gram Tools

Find word patterns based on the search term.

What 2-word clusters (sequences) are there for blind in the corpus?

Change the Min. Frequency to display the top 5 clusters. Click on the results.

Change the <u>Search Term Position</u> to <u>Right</u> and increase the <u>Cluster Size</u> to 3.

What observations can you make about the cluster 'gold and blind'?

The cataloguers of vols. 9 and 10 are most consistent in their phraseology to describe bookbindings. The detail seemed important to them and they had the expertise (and the time!)

You get similar results with the N-Gram Tool.

Plot Tool

Visualise where the search term appears in the corpus.

Useful for comparison between different search terms and examining large corpora.

Find the mentions of *morocco* (goatskin used in bindings) in the corpus

Sort by <u>Dispersion</u> to see how widely spread it is across the file

You can change the Dispersion statistics measures in the <u>Tool Settings</u>

Normalised Frequency takes into account the total number of tokens in the file.

Plot the use of colour patterns in the corpus. Use the <u>Overlay</u> option to compare the use of different patterns.

Search for *blue and red*, then select the Overlay option, choose another colour and search for *red and gold* Note that the majority of mention of *red and gold* is in volume 8. Click on the outliers to see the result in the file.

To save the graph as png file, select the <u>Graphic View</u> option in the Tool Settings

Keyword Tool

Compare words that appear unusually frequently in a target corpus in comparison with a reference corpus (eg. a specialised vs a generalised corpus) and can help characterise or understand the features of the target corpus.

Uses a number of statistical measures and thresholds that can be customised in the Tool Settings.

Likelihood - measures what is 'statistically significantly unusual' in the target corpus compared to the reference corpus **Effect** - measures 'the degree of unusualness', and the higher the effect, the higher the unusualness

You can import a new corpus into the Corpus Manager and open it using the Reference Tab

You can switch easily between the Target and Reference Tabs in the Manager or from the File menu.

Selecting the negative keywords setting in Tool Settings would list the most unusually infrequent word in the target corpus.

<u>This sheet</u> lists the positive and negative <u>Keyword Types</u> characteristic of our target corpus with reference to a corpus describing historical photographs, and vice versa. Note the similarities in terms of function words, verbs, object information. The photographic descriptions refer to people, places, events, but also provide measurements and archival information.

See the Keyness lesson which compares the historical photographs corpus with a wordlist generated from the British National Corpus

Discussion

Feedback on the approach and tool. Will you apply it to your data?

What are your research questions (printing culture, ownership and provenance, named entities, outdated or problematic language)

Explore further:

NER, punctuation in the descriptions - eg. Original?, quotation marks, referenced sources, annotations for digitised corpora

Ground truth for incunabula titles transcriptions

Combine data with MSS annotations in the curators' copy of the catalogue and Contextual information from the digitised catalogue

Recommendations for inclusive descriptive policy and practice

Thank you!

digitalresearch@bl.uk

https://blogs.bl.uk/digital-scholarship/

