

Critique textuelle et humanités digitales

Une introduction à CollateX

Frédérique Michèle Rey & Sophie Robert-Hayek

Université de Lorraine

Introduction

1 Introduction

2 What is a collation?

3 Automatic Collation - A Brief History of Research

4 Automatic Collation and Data Comparison

5 CollateX

6 bibliography

What is a collation?

What is a collation?

What is a collation?

collation

"In the Lachmannian method, *collatio* is part of *recensio*. It can be defined as the comparative examination of the witnesses in order to determine the variant readings in all witnesses."

M. Buzzoni, C. Macé, *Parvum lexicon stemmatologicum*

collation

"Comparaison des leçons de deux témoins ou plus."

Frédéric Duval, Les mots de l'édition de textes, Magister (Paris: École des Chartes, January 2015)

Aim of collation

- To track the history of text transmission
- To establish a critical edition
- To analyze variants
- To analyze language evolution
- To classify manuscripts
- etc.

The collation is the nodal point of every critical edition

Whether the method is (neo)Lachmanian or (neo)Bédieriste

- 1 List of witnesses (*recensio*)
- 2 Collation of witnesses (*collatio*) Analysis of variants (*examinatio*)
- 3 Text edition (*constitutio / emendatio*)

Paul Maas, Textkritik, 4. Auflage (Leipzig: Teubner Verlagsgesellschaft, 1960) or
D'Arco Silvio Avalle,

La doppia verità: fenomenologia ecdotica e lingua letteraria nel Medioevo romanzo, Archivio
romanzo 1 (Tavarnuzze (Firenze): SISMEL edizioni del Galluzzo, 2002)

List of witnesses and transcription

Sinaiticus

ΑΡΧΗΤΟΥ ΕΥΑΓΓΕΛΙΟΥ
ΟΥ ΙΥΧΥΚΛΘΩΣΤΕ

Alexandrinus

ΑΡΧΗΤΟΥ ΕΥΑΓΓΕΛΙΟΥ Η ΧΥΓΓΑΤΟΝ

Vaticanus

ΑΡΧΗΤΟΥ ΕΥΑΓΓΕΛΙΟΥ,
ΙΥΧΥΚΛΘΩΣΤΕ

Bèze

ΑΡΧΗΤΟΥ ΕΥΑΓΓΕΛΙΟΥ Η ΥΧΥΚΛΘΩΣΤΕ
Ο ΣΑΝΤΑΙ ΕΠΙΧΗΣΑΙ ΑΤΟΜΟΦΙΤΗ

Coridethianus

ΑΡΧΗΤΟΥ ΕΥΑΓΓΕΛΙΟΥ

Manual alignment of all the witnesses

| | | | | | | |
|---|---------------------|-----|-----|------|-----|----|
| ¶ | ἀρχὴ τοῦ εὐαγγελίου | ιω | χυ | υυ | του | θυ |
| A | ἀρχὴ τοῦ εὐαγγελίου | ιω | χυ | υιου | | θυ |
| B | ἀρχὴ τοῦ εὐαγγελίου | ιω | χυ | υιου | | θυ |
| D | ἀρχὴ τοῦ εὐαγγελίου | αηυ | χρυ | υιου | | θυ |
| Θ | ἀρχὴ τοῦ εὐαγγελίου | ιω | χυ | | | |

Table – Textual alignment of several witnesses of Mark 1:1

Edition and Apparatus

| | | | | | | | |
|---|---------------------|-----|-----|------|-----|----|--|
| ¶ | ἀρχὴ τοῦ εὐαγγελίου | ιω | χν | | | | |
| A | ἀρχὴ τοῦ εὐαγγελίου | ιω | χν | υν | τον | θυ | |
| B | ἀρχὴ τοῦ εὐαγγελίου | ιω | χν | υιου | | θυ | |
| D | ἀρχὴ τοῦ εὐαγγελίου | αην | χρυ | υιου | | θυ | |
| Θ | ἀρχὴ τοῦ εὐαγγελίου | ιω | χν | | | | |

Table – Alignement manuel des manuscrits grecs pour Marc 1:1

Αρχη του ευαγγελιου Ιησου Χριστου υιου Θεου

αιου Θεου B D | αιου Θεου *om.* ¶ θ | αιου του Θεου A

Critical Edition of the New Testament

Act 12,12! Kol 4,10!
1P,13

‘KATA MARKON’

Act 10,37s H 2,3-14s;
8,35; 10,29; 13,10; 14,9;
16,15-11; 3,11 5,7; 9,7;
14,6; 15,39
2-4 Mt 3,1-3 L 3,3-6
9,13; 14,21
2b: Mt 11,10 L 7,27 Ex
23,20 Mt 3,1 J 3,28 |
Is 40,3 B Jl,23

1 Ἀρχὴ τὸν εὐαγγελίου Ἰησοῦν Χριστοῦ [νυιοῦ θεοῦ]. ii
 2 Καθὼς γέρωπας ἐν τῷ Ὕστατῃ τῷ προφήτῃ.
 ἵδον τὸ ἀπόστελλα τὸν ἄγγελὸν μου πρὸ^τ
 προσώπου σου,
 δὲς κατασκευάσει τὴν ὁδὸν σου^τ. iii

St: Mt 3,4-6

3 φωνῇ βοῶντος ἐν τῇ ἑρήμῳ·
 ἔτομάστε τὴν ὁδὸν κυρίου,
 εὐθείας ποιεῖτε τὰς τρίβους γαύτον, τvi

Zch 13,4 L 7,25p JRg
1,8
Lv 11,21s
1Sm 14,25 Ps 81,17

4 ἐγένετο Ἰωάννης [ὁ] βαπτίζων ἐν τῇ ἑρήμῳ καὶ κηρύσσων βάπτισμα μετανοίας εἰς ὅφεσιν ἀμαρτιῶν. 5 καὶ ἔξεπορεύετο πρὸς αὐτὸν πάσα ή Ἰουδαία χώρα καὶ οἱ Ἱεροσολυμῖται πάντες, καὶ ἐβαπτίζοντο ὑπὸ αὐτοῦ ἐν τῷ Ἰορδάνῃ ποταμῷ ἔξοιλογόνεμοι τὰς ἀμαρτίας αὐτῶν.
 6 ‘καὶ ἦν ὁ Ἰωάννης’ ἐνδεδυμένος τρίχας καμῆλου ^θκαὶ ζώνην δερματίνην περὶ τὴν ὁσφὺν αὐτοῦ^γ καὶ ἐσθίων ἀκρίδας καὶ μέλι ἄγριον.

Inscriptio: ‘εναγγελιον κατα Μαρκον A D K L W Γ Δ Θ /¹³ l. 28. 33. 565. 700. 892.
1241. 1424. 2542 Θ lat | το κατα Μαρκον αγιον εναγγελιον 209. 579 (vglⁱⁱ) | – R* B* |
xxx R¹ B¹

¶ I,1 'υπο του θεου (κυριου 1241) A K P Δ f/^{1,13} 33. 565. 579. 700. 892. 1241. 1424.
2542. / 844 Θ | – R¹ Θ 28. / 2211 saⁱⁱⁱ; Or (er om) Ἰησον Χριστον Ιρ Επιρη | ttx R¹ B D
L W Γ latt sy co; Ir^{lat} • 2' 2-4 D Θ /¹ 700. / 844. / 2211; Ir Or¹ Επιρη | τοις προφηταις
A K P W Γ f/¹³ 28. 579. 1424. 2542 Θ vgm^{iv} sy^v (bom^{vi}); Ir^{lat} | ttx R B L Δ 33. 565. 892.
1241 sy^hmg co; Or¹ | τεγο Ρ A K L P W Γ Δ f/^{1,13} 28. 33. 579. 700. 892. 1241. 1424.
2542. / 844 Θ vg^h sy^h bom^{vi}; Or Eus | ttx B D Θ 28*. 565. / 2211 lat co; Ir^{lat} | τ⁷ (Mt
11,10) εμπροσθεν σου A Γ Δ f/^{1,13} 28. 33. 565. 579. 700*. 892. 1241. 1424. 2542. / 844 Θ f
ff⁷ | vg^h sy^h bom^{vi}; Eus | ttx R B D K L P W Θ 700*. / 2211 lat sy^h bom^{vi}; Ir^{lat}
• 3' του θεου υμον D (it) sy^h | τ⁷ (L 3,5s) add Is 40,4-8 W (c) • 4' 1-5 B 33 bom^{vi} | 2-5
892 | 2-6 A K P W Γ f/^{1,13} 565. 579. 1241. 1424. 2542. / 844 Θ sy^h sa^{vi} | 3-5 2 6
(D Θ 28). 700. / 2211 lat sy^h | ttx R L Δ bo | • 5' 3-6 1 2 A K P Γ Δ f/^{1,13} 579. 1424 Θ
(sy^h) | εν τω (-D*) λορδανην υτ αυτων D*! W Θ 700. 2542. / 2211 a | εις τον λορ-
δανην υτ αυτων 28. 565 | ttx R B L Δ 33. 892. 1241. / 844 f1 vgl co | • 6' τη δε ο (-A D W
Δ pm) Ιωαννης A K P W Γ Δ Θ f/^{1,13} (28). 565*. 579. 700. 1241. 1424. 2542. / 844 Θ it
sy^h sa bom^{vi} | ttx R B L 565. 892. (- o 33. / 2211) lat bom^{vi} | δερριν D a | ^oD it

In Bedierist Editions

- 1 Select your basis manuscript
- 2 Define the type of variants you want to collate (ex. ignore spelling variants)
- 3 Write on a table all the variants
- 4 Build your critical apparatus

Manual collation

Figure – Manual collation of Florilegium Coislinianum, cf. De Vos et al. 2010

Caroline Macé, "Chapter 3. Textual criticism and text editing," in *Comparative oriental manuscript studies: an introduction*, ed. Alessandro Bausi, Eugenia Sokolinski, and Comparative Oriental Manuscript Studies (Hamburg: COMSt, Comparative Oriental Manuscript Studies, 2015), 321–465

Automatic Collation - A Brief History of Research

VIII

VARIA

**LA COLLATION DES MANUSCRITS
A LA MACHINE ELECTRONIQUE**

INTRODUCTION

Pour étudier le mécanisme de la transmission des textes et en déduire les méthodes capables de les restituer à l'aide des variantes, j'avais pris comme terrain d'expérience une règle monastique anonyme composée probablement au VI^e siècle : la « Règle des quatre Pères », ainsi nommée parce qu'elle est le procès-verbal d'un synode où quatre abbés, Séraphon, Macaire, Paphnutius et un autre moine, tournent tour à tour la parole pour exprimer les divers points de la discipline monastique.

Cette règle réalisait toutes les conditions favorables : elle est courte (environ 250 lignes) ; elle est transmise par des manuscrits nombreux, qui s'étagent régulièrement depuis le VI^e-VI^e siècle jusqu'au XV^e et présentent une multitude de variantes de toutes sortes ; elle a été plusieurs fois publiée dans des éditions imprimées, notamment par Migne, *Patrologie latine*, tome CII, colonnes 435 et suivantes, et par M. François Massai, qui l'a placée en tête de son édition diplomaticale de la *Regula Magistri*.

Après avoir fait « à la main » le travail critique sur la Règle des quatre Pères, il m'a semblé intéressant de le faire exécuter par la machine électronique. C'était là une entreprise nouvelle. Jusqu'alors on avait employé la machine, dans le domaine littéraire et philologique, à des travaux de lexicologie, de stylistique et de traduction, mais personne ne l'avait encore utilisé pour les opérations de la critique textuelle en vue de collationner les manuscrits et de déterminer leurs relations. La Compagnie des Machines Bull (94, avenue Gambetta, Paris, 20^e) a compris l'intérêt que présentent de telles recherches et sur ma demande un groupe de techniciens s'est mis au travail, en 1960, pour établir des « programmes » et les expérimenter sur la machine.

Jacques Froger, “La collation des manuscrits à la machine électronique,” *Bulletin d'information de l'Institut de Recherche et d'Histoire des Textes* 13.1964 (1966): 135–71

Dom Jacques Froger (1964)

- Algorithm of text alignment (décalage, insertion, suppression, inversion)
- n-gram (considère des séquences de 2 ou 3 mots)
- identification des lieux variants (numéro de séquence du mot dans le texte)
- identification des variants (addition / omission / substitution)
- algorithm of reconstruction of the stemma

Automatic Collation - A Brief History of Research

Automatic Collation - A Brief History of Research

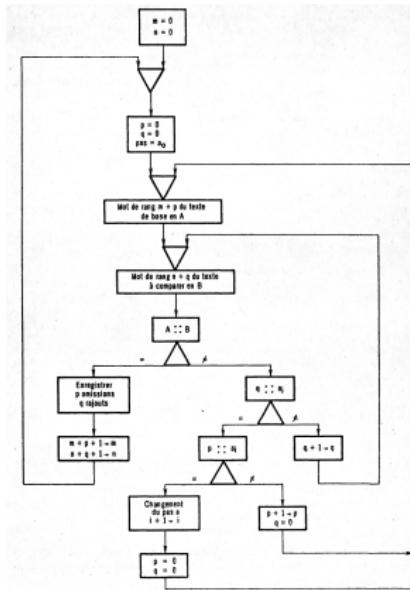


Figure – Algorithm of Mme Renaud used by Frogier

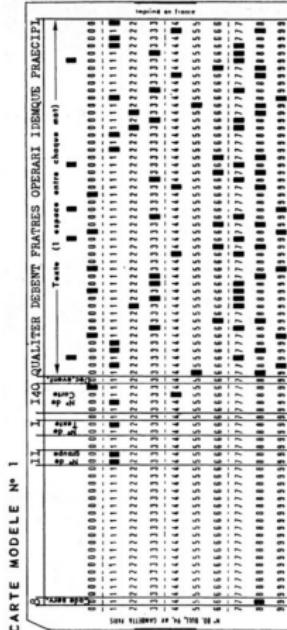
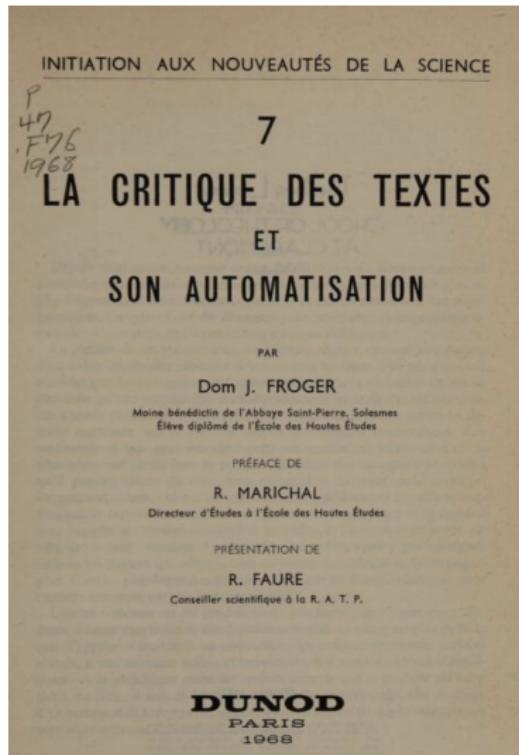


Figure – Carte perforée

Automatic Collation - A Brief History of Research



Automatic Collation - A Brief History of Research

Automatic Collation: A Technique for Medieval Texts

Penny Gilbert

A considerable amount has been published on the subject of computer-aided critical editions, i.e., the use of the computer to compare texts or manuscripts and to indicate variants. It is logical, therefore, before presenting my own work, to review this prior work in terms of its usefulness for medieval prose texts. Although another area of computerized activity, the determination of manuscript filiation has recently received some attention, that work depends on prior and independent determination of variants and does not appear here.¹ Because it is beyond the scope of the present subject, it will not be discussed here.

Jacques Froger's book *La Critique des textes et son automation*² touches briefly on the question of computer aids for locating variants. To check his theory, Froger made up test files of fourteen lines, hardly a sufficient amount to reproduce a complete manuscript in actual measure. The two texts are read word by word and cards are read in the computer which assigns a number to each word and to each space between words. In the output these numbers indicate the location of additions, deletions, and substitutions. The scholar specifies which text will serve as the base against which the other text will be compared. The computer, starting at the beginning of the text, compares the corresponding words from each, one at a time, until a mismatch occurs. Then a large block of words is ignored—Froger states usually about five, or twenty-five—until the texts match again. These intermediate discrepancies are saved and the texts are realigned at the place where the match was found. The printed results contain the variant, identification of the manuscript in which it was found, and a code indicating whether it is an addition, a deletion, or a substitution. The numbers already assigned to

“Computer-aided critical editions, i. e., the use of the computer to compare texts or manuscripts and to indicate variants.”

→ algorithme COLLATE
(base de CollateX)

Penny Gilbert, “Automatic collation: A technique for medieval texts,” *Computers and the Humanities* 7.3 (January 1973): 139–47, p. 139.

¹ W. Orr, “Computer Applications in Textual Criticism,” paper presented at the Symposium on the Use of Computers in Literary Research, Edinburgh, 27–30 March 1972; G. P. Zarr, “Algorithmi per critica codicum et les théories de Diane H. Quigley,” paper presented at the Symposium on the Use of Computers in Literary Research, Edinburgh, 27–30 March 1972; G. P. Zarr, “Automatic procedure di critica testuale: problemi e prospettive,” *“Lingu e Stile* 6 (Dec. 1971), pp. 397–415; G. P. Zarr, “Automatic procedure of textual analysis,” paper presented at the Second International Congress of Italian Linguistics, “Problemi della critica testuale e dei problemi degli intermezzi costitutivi,” paper presented at 13^{me} Congrès International de Linguistique et Philologie romanes, Québec, 1971. Jacques Froger, *La Critique des textes et son automatisation* (Paris: Dunod, 1970), pp. 12–20.

² Froger, pp. 230–3.

Penny Gilbert is a graduate student in the department of History at the University of Wisconsin conducting research at the University of Manitoba.

Automatic Collation - A Brief History of Research

DE GRUYTER OLDENBOURG

IT – Information Technology 2020, 44(2)

Elisa Nury* and Elena Spadini

From giant despair to a new heaven: the early years of automatic collation

<https://doi.org/10.1515/it-2019-0047>
Received November 17, 2019; revised January 30, 2020; accepted March 5, 2020

Abstract: This article presents a commented history of automatic collation, from the 1940s until the end of the twentieth century. We look at how the collation was progressively mechanized and automated with algorithms, and how the issues raised throughout this period carry on into today's scholarship. In particular, we examine the inner workings of early collation algorithms and their different steps in relation to the formalization of the Gothenburg Model. We also discuss what automatic collation allows after fascinating insights into the collaboration between Humanists and Computer Scientists, and the reception of computers by philologists.

Keywords: Collation, Textual Criticism, Philology, Digital Humanities, Algorithms

ACM CCS: Theory of computation → Design and analysis of algorithms; Social and professional topics → Professional topics → History of computing

1 Introduction

Whereas in the past, then, textual editors, finding the work rough, put over the side into By Path Meadow and, bengkuk and stone-boats, were captured by Glaz Despair, now the general process of association, like the sun in the Table, will cause us to have of what we want, and the work will bring POG, p. 1.

Today however, the reader finds a new type of editor, [...] friends with absolute boundary that have the meeting mass together and all the sons of God shoox for joy. – Düring 1962, p. 34.

In the early days of Humanities Computing, textual scholars saw the benefits of saving time to facilitate parts of their work when editing a text. Collation was a well-mixed

*Correspondence: Elisa Nury, Institut für Germanistik, Universität der Schweiz, CH-3012 Bern, Switzerland, e-mail: elisa.nury@unibe.ch, ORCID: <https://orcid.org/0000-0003-6488-6567>

Elena Spadini, Université de Lausanne, Centre des littératures en Suisse romande, 3015 Lausanne, Switzerland, e-mail: elena.spadini@unil.ch

candidate for automation, as a repetitive and often tedious task, during which the risk of making errors is high. Collation is the practice of comparing multiple versions of the same work, called witnesses, in search of the specific differences – the variants – which may shine a light on the history of the text, its genesis or its transmission.

Collation can be considered a form of intertextuality in the very large sense of the term, both involving comparison of different versions of a single work, and also allowing scholars on the margins of the text and influence its interpretation. In this article we argue that a history of automatic collation is relevant to understanding the numerous new developments in this area of philology.¹

Automatic collation makes use of computers or other devices to identify the variants in the texts, and display them on the screen. The *state of automatic collation*, described today, includes first a transcription of the witness' texts into machine-readable format, whether by hand or by OCR. Afterwards, these transcriptions are aligned where their texts match with the help of algorithms.

Since the 1940s, mechanical devices and later computer programs have been developed to assist scholars during collation; as a result, the name and definition of automatic collation has changed. The original definition is given by Gilbert [15, p. 139]: ‘computers aided critical editions, i.e., the use of the computer to compare texts or manuscripts and to indicate variants’. Interestingly, Gilbert refers to critical editions created with the help of a computer, but the actual definition describes very precisely the act of editing – compared texts and manuscripts – and its settings. In the beginning of the early tools were part of a larger infrastructure designed for the whole editing process, from gathering and collating various forms of a work, to printing and publishing the edited

¹ See for instance Spadini [25], chapter 5 and Mary [36] for a critical overview of automatic collation and lists of tools; at least 25 programs exist or have existed, eight of which were created in the past two years. Before going any further in this article, we would like to thank the anonymous reviewers and the colleagues that contributed their remarks.

To be fairly, we will use the term ‘automatic collation’, whether or not collation is fully automatic; we consider synonymous the expressions ‘semi-automatic collation’, ‘computer-supported collation’, ‘machine-assisted collation’.

Automatic Collation and Data Comparison

Automatic Collation and Data Comparison

Collation is, finally, a form of data/file comparison which is quite common in our digital common life through file synchronizations, backup or version control. It is generally performed using **diff utilities**.

Two examples

- Wikipedia “view history”
- Github commit diff

Diff, WDiff, Sequence alignments

- **Diff** compare only lines (alg. 1974)
- **GNU WDiff** compare lines and words (used by wikipedia)
- Sequence alignment in Bioinformatics for DNA and RNA (see for example this list of software and publication).

Jiannan Chao, Furong Tang, and Lei Xu, “Developments in Algorithms for Sequence Alignment: A Review,” Biomolecules 12.4 (April 2022): 546

Automatic Collation and Data Comparison

Numerous Algorithms

- COLLATE (Peter Robinson, 1989 and ff.)
- TUSTEP (Tuebingen System of Text Processing tools)
[https://www.tustep.uni-tuebingen.de/tustep_{ng}.html](https://www.tustep.uni-tuebingen.de/tustep_ng.html)
- MVD (Multi-Version Document) format ([D. Schmidt](#), “Graphical Editor for Manuscripts,” [Literary and Linguistic Computing 21.3 \(August 2005\): 341–51](#))
- TEIComparator (<https://tei-comparator.sourceforge.net/>)
- Juxta (<https://juxta.software.informer.com/>)
- CollateX
- Traviz

CollateX

- Website: <https://collatex.net/>
- Python Package: <https://github.com/interedition/collatex>
- Python Package Documentation:
<https://interedition.github.io/collatex/pythonport.html>

tutorials for CollateX

- <http://collatex.obdurodon.org/> -> The one-day
Computer-supported collation with CollateX
- Github repo of Leif-Jörn Olsson (University of Gothenburg):
<https://github.com/ljo/collatex-tutorial/>
- Github repo of Tara Andrews
- Github tutorial from DiXiT-eu:
<https://github.com/DiXiT-eu/collatex-tutorial>

CollateX

CollateX

Gothenburg model

- 1 Tokenization
- 2 Normalization/regularization
- 3 Alignment
- 4 Analysis
- 5 Visualization/output

1. Tokenization

Definition

“Tokenization is the division of a continuous text into units to be aligned (called tokens). Most commonly, tokens are whitespace-delimited words, but tokenization can be performed at any level of granularity, e.g., “letter, syllables, words, lines, phrases, verses, paragraphs, or text nodes”.”

(<https://wiki.tei-c.org/index.php/TextualVariance>)

1. Tokenization

Tokenization may be challenging

- for manuscript in scriptio continua (word separation ambiguity)
- for separating punctuation from the text (“cat” vs “cat,” or in hebrew בָּוֶל־אִישׁ)
- contractions, like English “doesn’t” or “can’t”
- in texts with special conventions (ex. reconstructed word Ev α[ρ]κη)
- for language like Hebrew that fusions different part of speech (ex. combines בְּ preposition “as” | שָׁמֵן name “human” | הַ his pronominal suffixe “his”)
- etc.

1. Tokenization

CollateX has a default tokenizer with the following properties

- It divides the text into tokens at white space.
- Punctuation is tokenized separately from alphanumeric characters.

It implies:

- “Peter’s cat.” → “Peter”, “”, “s”, “cat”, and “.”
- Ev α[ρ]κη → “Ev”, “α”, “[”, “ρ”, “]”, “κη”

2. Normalization

By default, the only normalization that CollateX performs is that it ignores trailing white space. I.e. “Peter’s cat” and “Peter’s cat” will be considered as similar.

Any more invasive normalization should be carried out beforehand.
For example (brackets deletion, diacritics characters, accentuation,
Hebrew vocalization, etc.)

3. Alignment

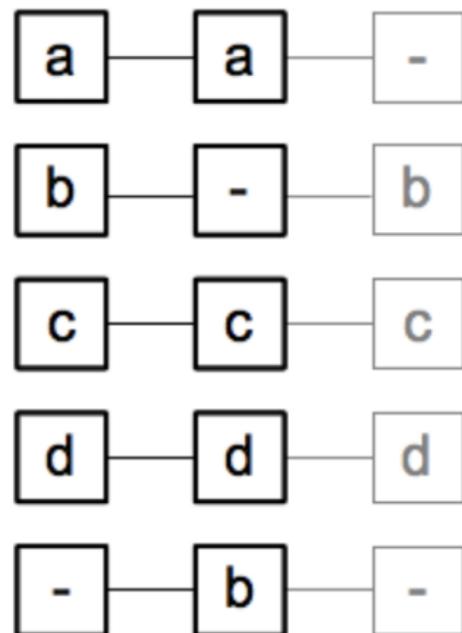
Consider three manuscripts

- MS A : “a b c d”
- MS B : “a c d b”
- MS C : “b c d”

3. Alignment

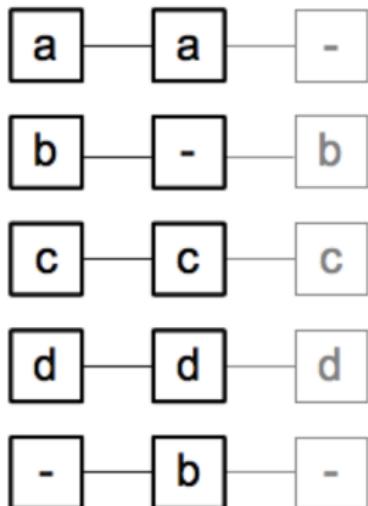
Consider three manuscripts

- MS A : “a b c d”
- MS B : “a c d b”
- MS C : “b c d”



An alignment of 3 versions

3. Alignment



```
<app>
  <rdg wit="#w1 #w2">a</rdg>
  <rdg wit="#w3" />
</app>
<app>
  <rdg wit="#w1 #w3">b</rdg>
  <rdg wit="w2" />
</app>
<app>
  <rdg wit="#w1 #w2 #w3">cd</rdg>
</app>
<app>
  <rdg wit="#w2">b</rdg>
  <rdg wit="#w1 #w3" />
</app>
```

Figure – XML TEI apparatus of the alignment sequence

3. Alignment

There are many different alignment algorithms (see, for example, Mme Renaud's algorithm in Foger's article from the 60s).

CollateX offers three alignment algorithms

- Dekker
- Needleman-Wunsch
- MEDITE

Dekker algorithm

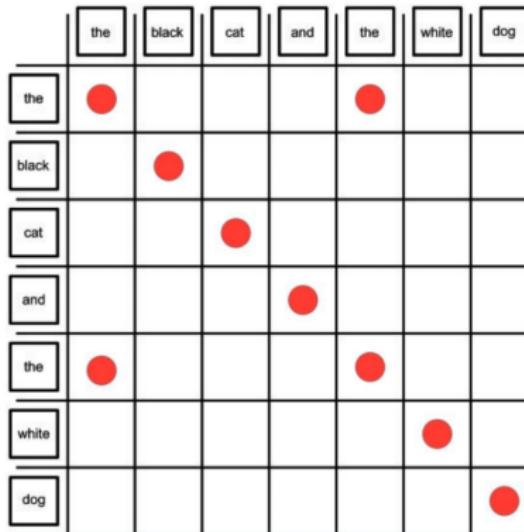


Figure – t_1, t_2 matrix

See Ronald Haentjens Dekker et al., “Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project,” *Digital Scholarship in the Humanities* 30.3 (September 2015):

Dekker algorithm

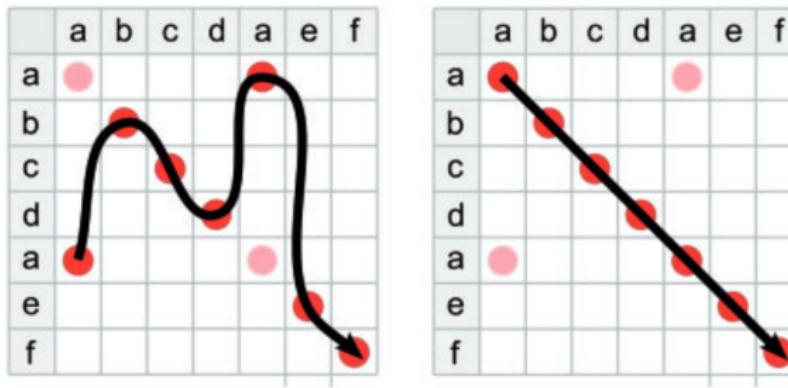


Fig. 5 a(left), and b (right): Unrealistic alignment conclusion (a) versus natural, elegant, or reader's common sense solution (b)

See [Haentjens Dekker et al., “Computer-supported collation of modern manuscripts,”](#)

Dekker algorithm

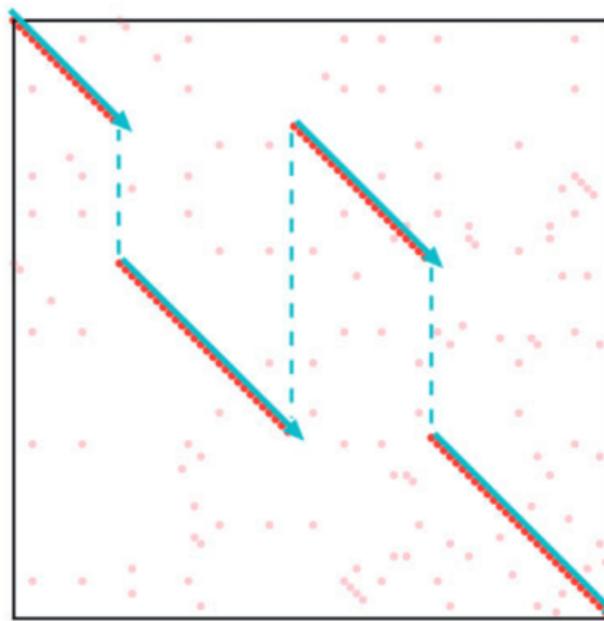


Fig. 7 Vectors describing alignment of two witnesses in a document-by-document matrix

Dekker algorithm

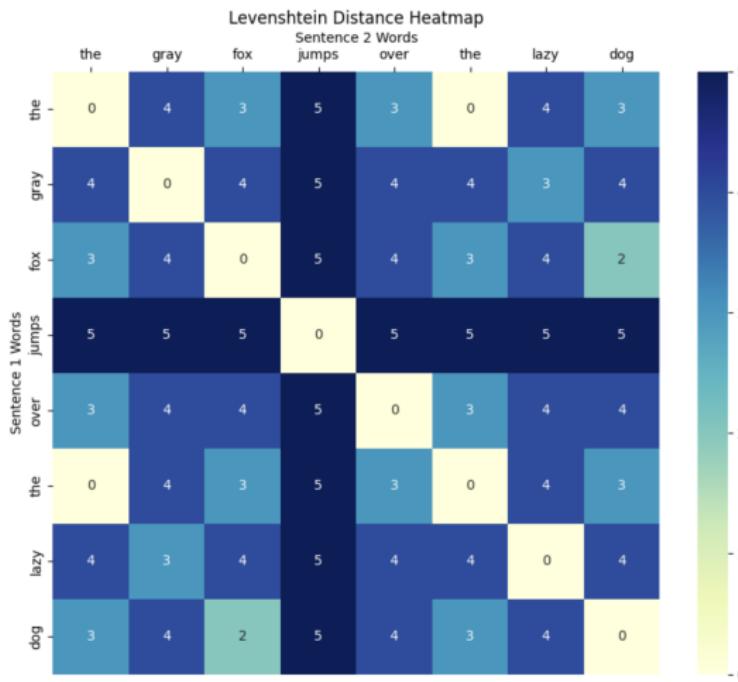
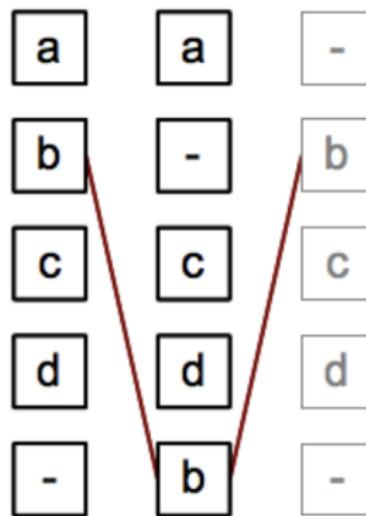


Figure – Levenshtein Distance Matrix

Analysis/Feedback



Analyzing an alignment

Figure – Addition, deletion or transposition

Visualization

Results

Variant Graph

Alignment Table

| | | | | |
|----|----|-------|-----------------|--------|
| W1 | Le | chat | mange la souris | grise |
| W2 | Le | chien | mange la souris | grasse |

GraphML

```
<?xml version="1.0"?><graphml graphdrawing.org/xmlns
  xmlns="http://graphml.graphdrawing.org/xmlns"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd"><key id="0" for="node"/>
  <attr name="number" attr.type="int"/></key>
```

GraphViz

```
graph G {
    v0 [label = ""];
    v1 [label = "Le"];
    v2 [label = "chien"];
    v3 [label = "chat"];
    v4 [label = "mange la souris"];
    v5 [label = "grise"];
    v6 [label = "grasse"];
    v7 [label = ""];
    v0 -> v1 [label = "W1, W2"];
    v1 -> v2 [label = "W2"];
    v1 -> v3 [label = "W1"];
    v2 -> v4 [label = "W2"];
    v3 -> v4 [label = "W1"];
    v4 -> v5 [label = "W1"];
    v4 -> v6 [label = "W2"];
    v5 -> v6 [label = "W1"];
    v6 -> v6 [label = "W2"];
}
```

TEI-P5

```
<?xml version="1.0"?><cx:apparatus
  xmlns:cx="http://interedition.eu/collatex/ns/1.0"
  xmlns="http://www.tei-c.org/ns/1.0"><Le><app><rdg
  wit="W1">chat</rdg><rdg
  wit="W2">chien</rdg></app><mange la
  souris><app><rdg wit="W1">grise</rdg></app>
  <rdg wit="W2">grasse</rdg></app>
</cx:apparatus>
```

Figure – Output on the CollateX website

Environnement de travail

Google Colab

<https://colab.research.google.com/>

bibliography



Avalle, D'Arco Silvio.

La doppia verità: fenomenologia ecdotica e lingua letteraria nel Medioevo romanesco
Archivio romanzo 1. Tavarnuzze (Firenze): SISMEL edizioni del Galluzzo, 2002.



Chao, Jiannan, Furong Tang, and Lei Xu. “Developments in Algorithms for Sequence Alignment: A Review.” Biomolecules 12.4 (April 2022): 546.



Duval, Frédéric. Les mots de l'édition de textes. Magister. Paris: École des Chartes, January 2015.



Froger, Jacques. “La collation des manuscrits à la machine électronique.” Bulletin d'information de l'Institut de Recherche et d'Histoire des Textes 13.1964 (1966): 135–71.



Gilbert, Penny. “Automatic collation: A technique for medieval texts.” Computers and the Humanities 7.3 (January 1973): 139–47.



Haentjens Dekker, Ronald, Dirk Van Hulle, Gregor Middell, Vincent Neyt, and Joris Van Zundert. “Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project.” Computers and the Humanities 53.2 (2019): 131–52.

