



HAL
open science

A mathematical formulation of textual criticism

Norman Simon Rodriguez

► To cite this version:

| Norman Simon Rodriguez. A mathematical formulation of textual criticism. 2024. hal-04747165

HAL Id: hal-04747165

<https://hal.science/hal-04747165v1>

Preprint submitted on 21 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A mathematical formulation of textual criticism

Norman Simon Rodriguez

Textual criticism is the discipline that aims to reconstruct the texts of ancient works from their surviving copies. Even though modern textual criticism has been practiced for nearly two centuries, no comprehensive mathematical formalization of its goals and logic has been developed to date. This paper provides a systematization of the main intuitions that form the basis of the discipline. Basic concepts like “manuscript”, “witness”, “collation”, “textual unit” and “reading”, among others, are first defined and then used to create more complex textual-transmissional constructs, like the notions of local and non-local genealogies, archetypes, contamination, confluents, evolutions, and retroversions. We also formalize the task of a textual critic in terms of probability spaces. In order to showcase the expressive potential of the model, we use it to provide definitions of three common Lachmannian heuristics (*stemma codicum*, local *stemma*, and variant *stemma*), as well as a framework for understanding majority rules. Moreover, we describe the model’s potential for allowing researchers to fully specify and develop computer simulations in an internally consistent way. The present mathematical formulation offers a way to mathematically characterize the textual transmission of works and the various strategies that could be leveraged to recuperate their original wordings. It can benefit textual critics by providing increased rigor in discussions about variants and also by facilitating the creation of new computational methods and software packages.

Most literary and scientific works from Antiquity did not survive in their original form, but rather as copies of copies of the originals or of early translations. Those copies were created by scribes who, over the centuries, introduced variation each time they transcribed a manuscript, with the result that any given work is not extant in a single wording, but rather in dozens or potentially even hundreds of different wordings, all of them differing from one another to varying degrees. Despite this, the variation is usually not so extreme so as to disallow the recuperation of an accurate

form of the original wording. It also tends to follow patterns derived from the way that human manual copying works. The discipline that studies how to combine the surviving copies of a given work in order to produce urtexts that approximate the original (or “autograph”) as devised by the author(s) is called textual criticism.

Textual critics often start by transcribing the texts (or “witnesses”) of the surviving manuscripts of a given work (which are collectively referred to as “the extant tradition”) and then aligning them in tables called “collations”, which allow them to easily see the differences between the texts. The differences are shown at discrete segments of text (called “textual units”, “variant units”, “textual locations”, or “places of variation”), at each one of which the textual critic chooses the wording (or “reading”) that she considers is most likely original, and relegates the remaining ones to a footnote. The final, reconstructed text, is called the “critical text”, and the footnotes are called the “apparatus”. A critical text with apparatus is known as a “critical edition”. In a sense, the job of textual critics consists of creating critical editions.

Early textual criticism was carried out by directly comparing manuscripts and using subjective judgment to select the reading that seemed to be the original at every textual unit (Haverling 2020). Starting from the nineteenth century, with the foundational work of Karl Lachmann and others, the genealogical metaphor gained popularity as a means to create heuristics for the design of approximate genealogical trees (known as “*stemmata*”) that could be used to recuperate the original text of a work, or at least the earliest recoverable form of the text underlying the extant manuscripts (the “archetype”) (see Timpanaro 2005; Bordalejo 2015; Maas 1958; Trovato 2014). According to this metaphor, a manuscript serves as a “parent” to another manuscript if a scribe used the former as a source to create the latter. In modern times, the genealogical metaphor has found an important expression in the borrowing of phylogenetic methods from biology into textual criticism (Howe et al. 2004).

Textual transmission can indeed be understood as a genealogical process, with witnesses engendering other witnesses, in what could be modeled as a dynamic, time-varying complex network whose evolution is driven by its previous states and exogenous historical factors¹. The literature on textual transmission, however, tends to presume an intuitive understanding of this complexity on the part of the reader, and emphasizes not the description of how textual transmission works, but rather the development of both qualitative and computational methods for *stemma* creation (see Hoenen (2020) for a panoramic view) or the creation of models for the estimation of macroscopic manuscript population parameters using birth-and-death processes (Weitzman 1974, 1987; Camps and Randon-Furling 2022). Compounding the difficulty is the fact that textual criticism as it is currently practised suffers from some vagueness in its core concepts, similarly to Economics and Linguistics before they

¹The textual transmission and evolution of a work can also be compared to other non-genealogical metaphors like the spreading of rumors in complex social networks or the spread of evolving virus strains in human populations, metaphors that have been extensively studied in complexity science, epidemiology and sociology (see Arruda, Rodrigues, and Moreno 2018; Moreno, Nekovee, and Pacheco 2004; Pastor-Satorras et al. 2015).

were mathematized in the 19th and 20th centuries. A notable early attempt at providing rigorous definitions of aspects of textual transmission, Greg’s *Calculus of Variants* (Greg 1927), did not gain traction and was unfortunately limited in scope, focusing mainly on small single-autograph genealogies that do not display contamination. Greg’s approach aimed to represent genealogical relationships by means of an algebra-like notation. His work appeared in the early twentieth century, shortly before the advances of graph theory that make complex network analysis possible started to become widely known.

To our knowledge, the only model to date that engages mathematically with the copying of documents and allows for an understanding of genealogies as complex networks is the one developed by del Gratta et al. (2021), who define “evolutions” of “documents” (roughly equivalent to genealogies of witnesses) as time-dependent genealogical graphs that are mapped to documents by means of a topological bundle and fiber construct². Their model is an important contribution to the understanding of textual transmission in computational terms, but it only formalizes a relatively small subset of the notions commonly used by textual critics in practice. The present article provides many definitions not included in their model. Additionally, our way of understanding genealogical relationships between witnesses is different in that we make use of the language of graph theory, which is already part of the vocabulary found in philological literature (Manafzadeh and Staedle 2020; Hoenen 2019), instead of the language of bundles. This may facilitate adoption on the part of textual critics who, given their professional training in stemmatology, are more naturally familiar with thinking in terms of graphs and networks.

In this vein, the purpose of the present model is not to formulate new techniques for *stemmata* creation nor to propose ways of estimating the probability of a given reading of being original, but rather to propose a “meta-language” that can be used to formulate any textual criticism problem, model or heuristic using a shared set of primitive, mathematical concepts. We aim to provide a general vocabulary in which all models of textual criticism can be expressed. Much like algebra gave Economics a shared language to formulate economic theories (eg, Neoclassical, Marxian, Keynesian) and allowed for their rigorous, systematic comparison, the present system (or others similar to it) could be adopted in the future in order to facilitate textual criticism discussions.

The article starts by presenting the basic definitions of the model (Section 1), including the concepts of traditions, genealogies, autographs, archetypes, and scribal error, and many others. Taking these basic notions as a foundation, we provide a formal definition of the task of the textual critic, framing it in terms of a probability space (Section 2). We will demonstrate (Section 3.1) how three common Lachmannian heuristics can be expressed in the mathematical language introduced by our model and reflect on the need to keep them both internally and externally consistent. We will also explain majority rules (Section 3.2) as alternative heuristics and how their

²See also Del Gratta et al. (2022) and Del Gratta et al. (2023).

effectiveness depends on the topologies of the latent genealogies, using the Greek New Testament as an example. A simulation approach (Section 3.3) is finally proposed as a technique with significant potential for the choice of heuristics for reconstructive purposes. It is important to notice that although the examples we offer throughout the article are taken from the textual criticism of the Hebrew Bible and the Greek New Testament, the model’s logic, being general, is not restricted to biblical textual criticism and is applicable to any textual tradition.

1 Basic concepts

Definition 1.1 (Alphabet, string). Let there be a set \mathcal{A} , called an ALPHABET, whose elements are symbols. A STRING \vec{s} is a finite sequence of elements from \mathcal{A} . The number of elements in \vec{s} is called its length and is denoted by $|\vec{s}|$, and $\vec{s}(i)$ is the i -th symbol of \vec{s} . The string when $|\vec{s}| = 0$ is called the EMPTY STRING and is denoted by ϵ . Two strings are considered the same string if they have exactly the same symbols in the same order.

Informally, a string is a sequence of characters. Suppose that the alphabet set is the English alphabet. Then, a string of length 6 could be, for instance, the word *scroll* (it has 6 letters). Its 4th element is the letter *o*. The alphabet can also include the blank space, and thus *the scroll* considered as a whole can also be a string. The symbols of the alphabet are not restricted to the characters normally used to write. Obeli, for instance, can be included in an alphabet.

Definition 1.2 (String concatenation). Let \vec{s}_1 and \vec{s}_2 be strings. Their CONCATENATION is denoted $\vec{s}_1\vec{s}_2$ and is the string that is equal to the sequence of symbols in \vec{s}_1 followed by the sequence of symbols in \vec{s}_2 . Note that in general $\vec{s}_1\vec{s}_2 \neq \vec{s}_2\vec{s}_1$ (ie, the concatenation operation is not commutative).

For example, let $\vec{s}_1 = \text{“the”}$ and $\vec{s}_2 = \text{“scroll”}$. The concatenation $\vec{s}_1\vec{s}_2$ is *“thescroll”*.

Definition 1.3 (Substring, substring coordinate). Let \vec{s} be a string. The string \vec{v} is a SUBSTRING of \vec{s} if there exist strings \vec{u} and \vec{w} such that $\vec{s} = \vec{u}\vec{v}\vec{w}$. Note that the strings \vec{u} and \vec{w} could be empty. The substring \vec{v} can be represented as $\vec{s}(x : x + h)$ where $x = |\vec{u}|$ and $h = |\vec{v}|$. The number x marks the position of the substring in the string \vec{s} and h is its length. x and h are the substring’s COORDINATES.

Definition 1.4 (Manuscript, witness, tradition). Let M and Λ be sets such that each $m \in M$ and each $\lambda \in \Lambda$ is a set of attribute-value pairs. Each element $m \in M$ is called a MANUSCRIPT. A tuple (m, \vec{s}, λ) where \vec{s} is a string is called a WITNESS. A set of witnesses is called a TRADITION.

Informally, witnesses are strings that belong to manuscripts. There may be multiple witnesses associated to the same manuscript, like what happens

with palimpsests or when a correction is added to an existing witness. Manuscripts are sets of attribute-value pairs. A manuscript, for instance, could be $\{\text{name: 'Leningrad Codex', location: 'Saint Petersburg', material: 'parchment'}\}$, indicating that its name is “Leningrad Codex”, that it is currently located in Saint Petersburg and that it is made of parchment. Thus, we differentiate between a manuscript and its texts, which we call witnesses.

Attributes of a witness (the set of attribute-value pairs λ) can include, for instance, the type of hand it was written in (formal, informal, etc.), the script (uncial, minuscule, etc.), the language (e.g. Hebrew, Greek, Latin), etc.

In colloquial usage, a tradition is sometimes understood as a collection of related manuscripts. However, since a manuscript can have more than one witness, a tradition is better understood as a set of witnesses (eg, a “tradition of the book of Ruth” is a collection of witnesses of the book of Ruth).

Definition 1.5 (Gap, lacuna). A GAP is a special symbol in an alphabet that represents a gap of length 1 in the transcription of the string of a witness. It is denoted by \mathfrak{g} . A LACUNA is a string made up entirely of gaps. For instance, a lacuna $\vec{\mathfrak{g}}$ of size 1 is $\vec{\mathfrak{g}} = \mathfrak{g}$, a lacuna of size 4 is $\vec{\mathfrak{g}} = \mathfrak{g}\mathfrak{g}\mathfrak{g}\mathfrak{g}$, etc. The size of a lacuna may be unknown. In that case, it is notated by $\vec{\mathfrak{g}}^*$ (using the Kleene star operator).

Lacunae are different from the empty string because they signal a place where there was a string that later disappeared from the witness, ie, they do not represent the absence of characters but are rather a placeholder for a lost string.

Definition 1.6 (Fragmentary witness, complete witness). Let $w = (m, \vec{s}, \lambda)$ be a witness. w is a FRAGMENTARY WITNESS if \vec{s} contains at least one lacuna. It is COMPLETE otherwise.

Definition 1.7 (Segment). Let $w = (m, \vec{s}, \lambda)$ be a witness. A SEGMENT of witness w is a tuple $(w, (x, h))$ where $x, h \in \mathbb{N}_0$ and $x + h \leq |\vec{s}|$. The position of the segment is the number x and h is the segment’s length³.

Definition 1.8 (Overlapping segments). Let $(w, (x_1, h_1))$ and $(w, (x_2, h_2))$ be two segments of the same witness w . The segments are overlapping if $(x_1 \leq x_2 < x_1 + h_1)$ or $(x_2 \leq x_1 < x_2 + h_2)$.

Definition 1.9 (Configuration). Let $w = (m, \vec{s}, \lambda)$ be a witness and let $C_w = ((w, (x_1, h_1)), (w, (x_2, h_2)), \dots, (w, (x_n, h_n)))$ be a tuple of segments of it. C_w is a CONFIGURATION of w if $\vec{s} = \vec{s}(x_1 : x_1 + h_1)\vec{s}(x_2 : x_2 + h_2) \dots \vec{s}(x_n : x_n + h_n)$. The size of a configuration C is denoted by $|C|$ and the configuration’s i -th segment is denoted by $C(i)$. The set of all the elements of a configuration C is denoted by $\cup C$.

³A set of segment attributes (attribute-value pairs) could optionally be attached to the definition of segment, depending on the modeling needs.

Intuitively, configurations are ways of splitting the string of a witness in such a way that if the substrings are concatenated together, the original string is obtained. As an example, take a witness w whose string is “the scroll”. One possible configuration of this string is the tuple $((w, (0, 4)), (w, (4, 6)))$, because the substrings implied by it, namely “the” and “scroll”, recreate the original string if concatenated. Another possible configuration of the same witnesses could be the tuple $((w, (0, 1)), (w, (1, 4)), (w, (4, 6)))$, implying “t”, “he ” and “scroll”. An example of a tuple of segments that is not a configuration is $((w, (4, 6)), (w, (0, 3)))$, which implies “scroll” and “the”, because it has the segments in the wrong order and, additionally, because it left out one of the characters (the blank space).

Definition 1.10 (Collation, textual unit). A COLLATION \mathcal{C} is a set of configurations such that $|C| = n, n > 0$ (n finite) for all $C \in \mathcal{C}$. That is, a collation is a set of configurations of the same (non-zero) size n . The set $\mathcal{C}(i) = \{C(i) : C \in \mathcal{C}\}$ is the i -th TEXTUAL UNIT (or text location) of the collation \mathcal{C} . Since all the configurations have the same number of elements n , such collation can also be called an n -collation.

Collations could be represented as matrices. For instance, consider two witnesses $w_1 = (m_1, \text{Aaron and Moses}, \lambda_1)$ and $w_2 = (m_2, \text{Aaron and Israel}, \lambda_2)$. We can create segments that imply the sequences (Aaron and , Moses) and (Aaron and , Israel) and arrange the sequences in a matrix like 1 (we show the implied substrings instead of the coordinates for ease of visualization; the order of the witnesses in the matrix is not important). The segments corresponding to the first textual unit of this collation are represented as the vector $[(w_1, \text{Aaron and }), (w_2, \text{Aaron and })]^T$ and those of the second textual unit as the vector $[(w_1, \text{Moses}), (w_2, \text{Israel})]^T$.

$$\begin{bmatrix} (w_1, \text{Aaron and }) & (w_1, \text{Moses}) \\ (w_2, \text{Aaron and }) & (w_2, \text{Israel}) \end{bmatrix} \quad (1)$$

Definition 1.11 (Reading, reading set). Let $\mathcal{C}(i)$ be a textual unit. Let $R_{\mathcal{C},i} = \{\vec{s}(x : x + h) : ((m, \vec{s}, \lambda), (x, h)) \in \mathcal{C}(i) \text{ for all } m, \vec{s}, \lambda, x, h\}$. The elements of the set $R_{\mathcal{C},i}$ are called the READINGS of the textual unit and the set itself is called the textual unit’s READING SET.

In other words, readings are the unique strings that appear in a given textual unit. In our previous example, the unique substrings of the first and second textual units are, respectively, {Aaron and } and {Moses, Israel}. This means that the first textual unit has only one reading, while the second textual unit has two.

Definition 1.12 (Base text, apparatus, critical edition). Given an n -collation \mathcal{C} , let $R_{\mathcal{C},i}$ be the reading set of its i -th textual unit. Choose exactly one arbitrary configuration $C_w \in \mathcal{C}$, which will be called the collation’s BASE TEXT. Let \vec{s} be the string of the witness w . Define $\mathfrak{A}_{\mathcal{C},i} = \{\vec{\rho} \in R_{\mathcal{C},i} : \vec{\rho} \neq \vec{s}(x : x + h)\}$, where x, h are the coordinates of the segment $C_w(i)$. The tuple $\mathfrak{A}_w = (\mathfrak{A}_{\mathcal{C},1}, \mathfrak{A}_{\mathcal{C},2}, \dots, \mathfrak{A}_{\mathcal{C},n})$ is an

APPARATUS of witness w over the collation \mathcal{C} . A CRITICAL EDITION of w is a tuple (C_w, \mathfrak{A}_w) ⁴.

In words, a base text is a configuration of a witness that was singled out to serve as the collation basis and an apparatus is a tuple that, for each textual unit, contains the readings that are present in the collation and are different from the reading of the base text. A critical edition consists of a base text and an apparatus.

1.1 Scribes, exemplars and copies

Definition 1.13 (Kleene closure). Let \mathcal{A} be an alphabet. Its KLEENE CLOSURE is the set $\mathcal{A}^* = \{\epsilon\} \cup \mathcal{A} \cup (\mathcal{A}\mathcal{A}) \cup (\mathcal{A}\mathcal{A}\mathcal{A}) \cup \dots$, where $(\mathcal{A}\mathcal{A})$ represents the set of strings formed by concatenating two symbols from \mathcal{A} , $(\mathcal{A}\mathcal{A}\mathcal{A})$ represents the set of strings formed by concatenating three symbols from \mathcal{A} , and so on.

Definition 1.14 (Null witness, null segment). The null witness is the tuple $\epsilon = (\emptyset, \epsilon, \emptyset)$. The null segment is the tuple $\mathfrak{e} = (\epsilon, (0, 0))$.

Definition 1.15 (Scribe, copying event, complete copying event). Let $L \subset \mathcal{A}^*$ be a finite language where \mathcal{A}^* is an alphabet's Kleene closure. Let M be a set of manuscripts and Λ a set of attribute-value pairs. The set $M \times L \times \Lambda$ contains all the possible witnesses that can be created from M , L and Λ . Let Σ be a proper (finite) subset of $(M \times L \times \Lambda) \times (\mathbb{N}_0 \times \mathbb{N}_0)$. The set Σ is a set of segments. Let C_w be a configuration of a witness w such that $\cup C_w \subset \Sigma$ (all the segments of the configuration are contained in the set of segments Σ ; there is no restriction on segment length, and it is not required that all the segments have the same length). Let Z be a set of attribute-value pairs, where each $\zeta \in Z$ is called a SCRIBE. The attributes of scribes can be names, dates and places of birth, genders, etc. A COPYING EVENT is a relation $K \subset (\Sigma \cup \{\mathfrak{e}\}) \times \cup C_w \times Z$ such that i) no $(\sigma, c, \zeta) \in K$ exists where σ and c are overlapping segments (this is to avoid the situation where a segment is a copy of itself) and ii) for all $(\sigma, c, \zeta), (\eta, c, \mu) \in K$ it holds that $\zeta = \mu$ (ie, each segment c can be matched to one and only one scribe in the copying event, even when the segment c was matched to different segments σ and η). A COMPLETE COPYING EVENT is a copying event where for all $c \in \cup C_w$ there exists at least one $(\sigma, c, \zeta) \in K$.

Copying events match arbitrary segments to the segments of a given witness (the copy) and then to scribes. They model the idea of human scribes reading segments from their exemplars and copying them into their copies.

Definition 1.16 (Copy, exemplar). Let $K \subset (\Sigma \cup \{\mathfrak{e}\}) \times \cup C_w \times Z$ be a copying event. The witness w is called the COPY. Let $E = \{e : ((e, (x, h)), c, \zeta) \in K\}$. The witnesses contained in the set E are called the EXEMPLARS.

⁴Some philologists consider that a critical edition is such only if it contains, apart from the base text and apparatus, additional materials like explanations, manuscript analyses and *stemmata*. Depending on the use case, these elements can be added to the definition of a critical edition as, possibly, a set of attribute-value pairs.

Definition 1.17 (Composed copy, properly composed copy). Let w be a copy. w is a COMPOSED COPY if at least one of its exemplars is the null witness. A composed copy is a PROPERLY COMPOSED COPY if the null witness is its only exemplar.

Definition 1.18 (Perfect copy, imperfect copy). Let $w = (m_1, \vec{s}, \lambda_1)$ be a copy from a copying event such that $\vec{s} \neq \epsilon$. w is a PERFECT COPY if it only has one exemplar and the string of the exemplar is identical to \vec{s} . A copy that is not perfect is IMPERFECT.

As we said above, a copying event can be considered as a matching between segments and between them and scribes. Let's say that we want to represent the situation of a human scribe who was copying from two witnesses e_1 and e_2 , one saying "Adam ate from the tree" and the other one "Eve ate from the fruit", respectively. The witnesses e_1 and e_2 are the scribe's exemplars. The scribe decides to write down "Adam and Eve ate from the fruit" in a copy w .

One of the possible copying events can be visualized as in Figure 1, where the elements of $(\Sigma \cup \{\epsilon\})$ are to the left, the elements of $\cup C_w$ are in the middle and the scribes are to the right (we show the implied substrings of the segments instead of their coordinates). Here all the segments of the copy were copied by the same scribe ζ_1 , but the definition of a copying event allows a witness to be copied by more than one scribe (as long as each individual segment is assigned to only one scribe). As we can see, the first substring of the copy was produced by combining two segments from different exemplars.

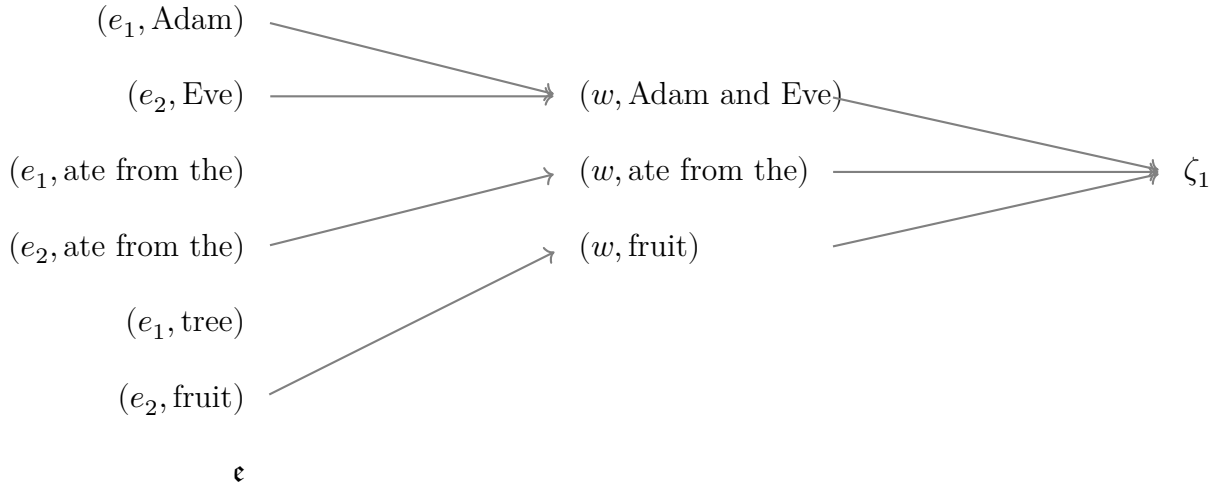


Figure 1: Representing a copying event.

Note that a witness can be both exemplar and copy in the same copying event, provided that no segment was copied from an overlapping segment (this constraint follows from the definition of a copying event). We can imagine, for example, a scribe copying the repetitive records of Numbers 7, where the offerings of the leaders of Israel are listed in detail for each one of them even though all the offerings included the

same elements. The scribe, instead of copying from another manuscript, can look at his previously copied text and copy from there. That means that the latter segments of the witness were copied from the earlier segments of the same witness. No segment, however, can be copied from itself, nor can a segment be copied from an overlapping segment.

Copying events can be given probability distributions in order to simulate scribal propensities and other copying patterns. For instance, let K^* be the set of all possible copying events for given $(\Sigma \cup \{\epsilon\})$ and Z . A probability distribution $\Pi(K^*)$ could be defined that assigns a probability to the elements of each potential copying event. Say that we want to characterize the propensity of a scribe to commit omission errors (like homoteleuton). We can define a subset of K^* that includes only the tuples that correspond to that particular scribe and further group the tuples into those that contain omission errors and those that do not. Given the probabilities assigned by $\Pi(K^*)$, we can calculate the conditional probabilities for that particular scribe and that particular type of scribal error. This feature of the model can be useful when creating computer simulations of real-world textual traditions, with the distribution Π providing a way to incorporate randomness in a controlled way. For example, a greater copying precision could be assigned to the scribes of a certain locale relative to scribes in other geographical areas in such a way that, on average, those scribes make fewer mistakes over many runs of a Monte Carlo simulation.

Definition 1.19 (Author, autograph, urstelle). A scribe is an *AUTHOR* if it contains an attribute that identifies it as such⁵. Let w be a copy from a copying event K . The copy w is an *AUTOGRAPH* if there exists at least one $(\sigma, c, \zeta) \in K$ such that ζ is an author. The segments of an autograph are called *URSTELLEN* (singular *urstelle*).

As we know from Definition 1.15, a scribe ζ is a set of attribute-value pairs. One of those attributes is whether a scribe is an author or not. To a large extent, the task of a textual critic involves recovering the author(s)'s original words, so it is crucial that authors are clearly identified.

Definition 1.20 (First autograph, autographic copy). Let w be an autograph. w is a *FIRST AUTOGRAPH* if it is also a properly composed copy. It is an *AUTOGRAPHIC COPY* otherwise.

Definition 1.21 (Autographic edition). Let $w = (m, \vec{s}, \lambda)$ be an autographic copy. w is an *AUTOGRAPHIC EDITION* if the set of attribute-value pairs λ contains an attribute that identifies w as such.

⁵One could use the author construct to model both real-world authors who penned their works themselves and amanuenses acting under their direct supervision and approval. Holmes (2013, 657–58) has argued for a definition of autograph that we could summarize as a witness such that i) it was written by at least one author (or by an amanuensis under the direct supervision, revision and approval of at least one author) and ii) copies of it were produced by someone else than the author(s) without the direct supervision, revision and final approval of an author. In terms of the present model, the definition of an autograph is more inclusive as it only requires that at least one of the scribes associated to it be an author.

Informally, a witness is called an autograph if at least one author participated in its creation. If the author(s) created the autograph using only one exemplar and the exemplar was the null exemplar, the autograph is called a first autograph. A first autograph in real life is the creation of author(s), and it is in this sense that the word ‘autograph’ is almost always understood in professional discussions about textual criticism. Strictly speaking, however, not all autographs are first autographs. If the autograph was copied from an ordinary exemplar(s), it is called an autographic copy. Additionally, although one could define an autographic edition as an autographic copy that contains significant deviations from all of its exemplars, the decision on whether an autographic copy is a new edition of a given work or not is subjective, and therefore we prefer to define an autographic edition as one that has been marked as such.

Definition 1.22 (Latent tradition). Let T be a tradition. If T contains all the witnesses that ever existed of a given work or pre-specified set of works, including at least one first autograph, and all the witnesses are complete, then T is called a LATENT TRADITION.

Definition 1.23 (Extant tradition). Let T be a latent tradition. An EXTANT TRADITION $T^\times \subseteq T$ is a subset of T that contains only the witnesses available to the researcher.

Normally, the latent tradition is unknown to the textual critic and only T^\times is known. Many manuscripts have disappeared from history due to climatic conditions, wars and normal wear and tear, taking their witnesses with them.

Definition 1.24 (Lost witness, extant witness). Let T be a latent tradition and $T^\times \subseteq T$ an extant tradition. A witness $w \in T$ is called LOST if $w \notin T^\times$. It is EXTANT otherwise.

1.2 Local genealogies

Definition 1.25 (Textual history). Let T be a tradition and let \mathcal{K} be a set of complete copying events such that there exists exactly one copying event in \mathcal{K} for each copy $w \in T$ and all the exemplars of all the copying events are elements of the set $T \cup \{\varepsilon\}$. Let $\cup\mathcal{K} = \{(\sigma, c, \zeta) \in K : K \in \mathcal{K}\}$. Define $\mathcal{S} = \{\sigma, c : (\sigma, c, \zeta) \in \cup\mathcal{K}\}$ and $\mathfrak{K} = \{(\sigma, c) : (\sigma, c, \zeta) \in \cup\mathcal{K}\}$ (notice the parenthesis in the definition of \mathfrak{K} , which makes \mathfrak{K} a set of tuples). The directed graph $H = (\mathcal{S}, \mathfrak{K})$ is called the TEXTUAL HISTORY (or history for short) of the tradition T if it is acyclic and there are no directed paths connecting any $\sigma_i, \sigma_j \in \mathcal{S}$ where σ_i and σ_j are overlapping segments.

Textual histories can accommodate relationships between segments at any granularity level, as illustrated by Figure 2. Here, witness w_1 produced two copies, w_2 and w_3 , and w_3 was itself copied into w_4 . Notice that while a witness can appear several times in a history *as an exemplar* (like w_1), it cannot appear *as a copy* more than once. This is because while witnesses are copied into only once, they can be

copied from many times. Thus, we see for example that w_3 appears once as a copy and once (but potentially more times) as an exemplar. Importantly, w_3 appears in two configurations, (“A hedgehog”, “and a house”) and (“A hed”, “gehog”, “and a house”), but the last segment of the two is the same, so it is only included once. The possibility of including the same witness several times as an exemplar (in different configurations) allows us to split its string in many different ways, thus allowing for different granularities. For instance, in the copying event that created w_2 from w_1 , “doghouse” gave rise to “A dog” and “and a house”, while in the event that created w_3 , the scribe split the string of w_1 in a different way and created “A hedgehog” from “A hedge and a dog” and “and a house” from “house”.

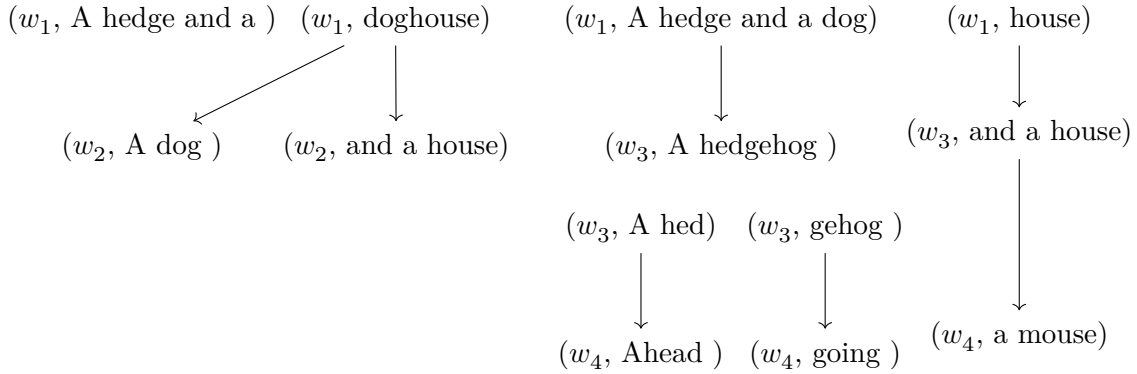


Figure 2: An example of a simple history.

Definition 1.26 (Latent history). Let $H = (\mathcal{S}, \mathfrak{R})$ be a history of tradition T . H is called the LATENT TEXTUAL HISTORY of T if T is a latent tradition and \mathfrak{R} contains every and all the correct real-world parent-child relationships between any two elements of \mathcal{S} .

Definition 1.27 (Ancestor segment, descendant segment). Let $H = (\mathcal{S}, \mathfrak{R})$ be a history, and let $\sigma_i, \sigma_j \in \mathcal{S}$. If there is a directed path from σ_i to σ_j in H , then σ_i is an ANCESTOR SEGMENT of σ_j (and σ_j is a DESCENDANT of σ_i). We denote “ σ_i is an ancestor of σ_j ” as $\sigma_i \prec \sigma_j$ and the equivalent relation “ σ_j is a descendant of σ_i ” as $\sigma_j \succ \sigma_i$. The symbols \preceq and \succeq represent the relations “is the same as or is an ancestor of” and “is the same as or is a descendant of”, respectively.

We can think of latent textual histories as the all-encompassing genealogies that tell us exactly which segments gave rise to exactly which other segments, for all the segments that were copied in a tradition. If we could take an autograph and split it into textual units, we could exactly find out which segments of which witnesses were descended from each segment in the autograph by just consulting the latent textual history. Needless to say, real-world traditions do not come with their latent textual histories attached, and thus this ideal scenario is not feasible (if it were, there would be no need for textual criticism).

We require that histories be acyclic in order to prevent a segment from being an ancestor of itself, a situation that is disallowed by the nature of the copying process in the real world. For the same reason, we also forbid that a segment be an ancestor of an overlapping segment. For example, let “Israel and Egypt” and “Israel” be the substrings of two overlapping segments (which, by definition, come from the same witness). The textual history cannot include any path that allows “Israel” to be descended from “Israel and Egypt”, nor vice versa, because “Israel” cannot be an ancestor of itself.

Definition 1.28 (Local genealogy, latent local genealogy). Let $H = (\mathcal{S}, \mathfrak{K})$ be a history. Let $\sigma \in \mathcal{S}$ be a segment and let \mathcal{S}_σ be a subset of \mathcal{S} such that $\sigma \in \mathcal{S}_\sigma$. Let \mathfrak{K}_σ be a set of ordered pairs such that for any $\pi, \nu \in \mathcal{S}_\sigma$, $(\pi, \nu) \in \mathfrak{K}_\sigma$ implies $(\pi, \nu) \in \mathfrak{K}$. The directed graph $G_\sigma = (\mathcal{S}_\sigma, \mathfrak{K}_\sigma)$ is a LOCAL GENEALOGY of segment σ . It is the subgraph of H that is induced by \mathcal{S}_σ . If H is a latent history and \mathcal{S}_σ is the union of $\{\sigma\}$ and the set of the vertices of H that are reachable from σ regardless of edge direction, then G_σ is the LATENT LOCAL GENEALOGY of the segment σ .

Local genealogies are the genealogies of specific segments. If one joins together all the local genealogies of a tradition, the result is the textual history. In this sense, textual histories are the collection of all the local genealogies of a tradition.

Note that if two segments are genealogically related, then they share the same latent local genealogy. That is because if a given segment π is reachable from another segment ν over a latent history, then π belongs in the latent local genealogy of ν but, following the same reasoning, that also means that ν is reachable from π and that ν belongs in the latent genealogy of π . They share the same latent genealogy. That latent genealogy includes them and all their ancestors, all their descendants and all the segments that are genealogically related to them to any degree.

Regarding local genealogies that are not latent, our definition allows the local genealogy to include segments that are unrelated to the segment to which the genealogy belongs. Given that there is no genealogical relationship between those nodes and the segment, those segments are not reachable from it.

Definition 1.29 (Conflation). Let $G_\sigma = (\mathcal{S}_\sigma, \mathfrak{K}_\sigma)$ be a local genealogy. A segment $\iota \in \mathcal{S}_\sigma$ is a CONFLATION if its in-degree is greater than 1.

Consider the three local genealogies of Figure 3 (as usual, the substrings are shown instead of the coordinates). The local genealogy in the middle depicts a conflation.

Remark 1.1 (Conflations do not necessarily imply contamination, nor vice-versa). When a conflation was created from segments that came from different witnesses, this alone suffices to make the copy a contaminated one, but this is not the case if the conflation was created from segments that came from the same witness. On the other hand, the presence of contamination does not directly imply that the copy contains conflations, because in a copy there can be segments that come from different witnesses all having in-degree equal to 1.

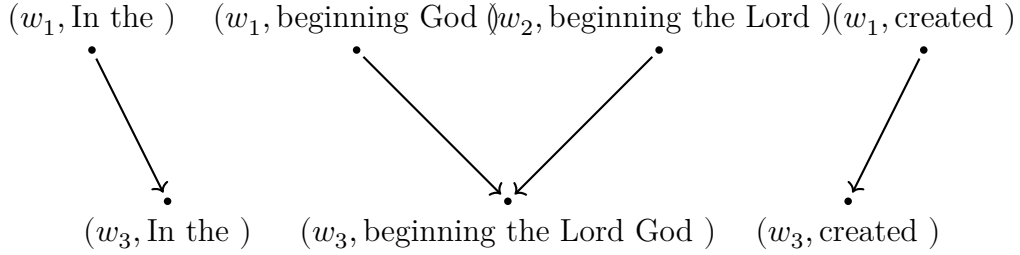


Figure 3: Example of conflation.

1.2.1 Evolutions

Local genealogies can be compressed so that they do not show relationships between segments but rather between *readings*. This simplifies the graphical representation of the history of the development of a reading, that is, of its “evolution”.

Definition 1.30 (Evolution). Let $X = (S, E)$ be a directed acyclic graph where S is a set of segments. Create a partition \mathfrak{P} of its vertices in such a way that if an arbitrary segment is reachable (ignoring edge direction) from any other arbitrary segment that has the same reading, without passing through a segment with a different reading, then both segments belong in the same cluster. The segments in each cluster all have the same reading and form a weakly connected component. Define a labeled directed graph $\mathcal{X} = (\mathfrak{P}, \mathfrak{H}, \mathfrak{R})$ where the vertices (\mathfrak{P}) are the clusters of the partition, the edges (\mathfrak{H}) are ordered pairs of clusters and \mathfrak{R} is a labeling function that assigns labels to the vertices of the graph in such a way that each cluster is labeled with the reading it attests to. Let \mathfrak{a} and \mathfrak{b} be clusters in the partition \mathfrak{P} . The edge $(\mathfrak{a}, \mathfrak{b})$ exists in \mathfrak{H} iff there is at least one segment $\iota \in \mathfrak{a}$ and at least one segment $\xi \in \mathfrak{b}$ such that $(\iota, \xi) \in E$. The labeled directed graph \mathcal{X} is called the **EVOLUTION** of the readings of the graph X .

We can create evolutions of any directed acyclic graph whose vertices are segments. We will provide two examples of evolutions of local genealogies for now, but we will later on mention evolutions of other types of graphs (see Definition 3.5, Definition 3.6).

Consider the local genealogy of Figure 4. The letters are readings and the subscripts are the witnesses they belong to. We can group all the segments that have the same reading such that there is an undirected path between all the segments in each group (that is, such that they form a weakly connected component), like in Figure 5.

Note in Figure 5 that segments B_8 and B_{10} are in different sets even though they share the same reading. This is because there is no undirected path between them that passes exclusively through segments that also have the reading B .

We can create a directed graph in which the vertices are the segment sets. Two sets are connected by an edge iff there is at least one edge in the local genealogy joining one element of one set with one element of the other. This procedure takes us to Figure 6.

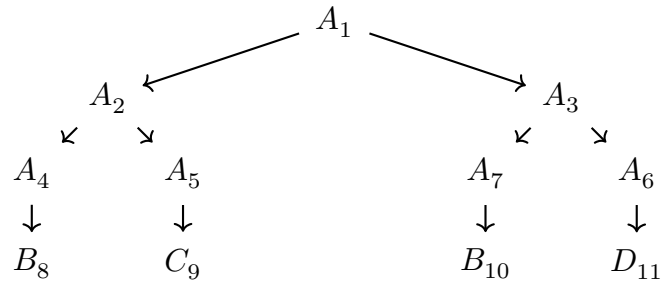


Figure 4: A local genealogy.

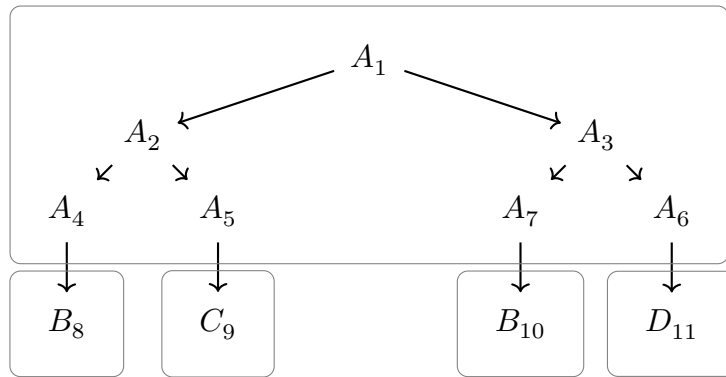


Figure 5: Grouping the readings in the local genealogy.

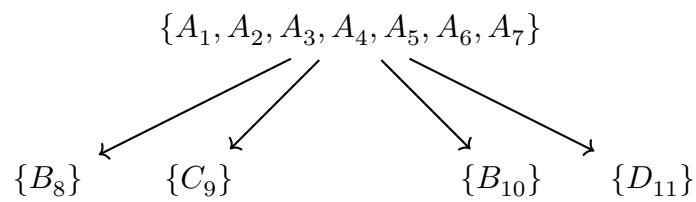


Figure 6: A directed graph whose vertices are sets of segments.

After this, we can create the evolution of the readings by labeling the vertices with the readings common to all the segments in each node. The resulting graph is in Figure 7. This is the evolution of the readings. As we can see, the text directly evolved from reading A into three different readings, B , C , and D . The fact that B appears twice means that the local genealogy arrived at that reading in two occasions via different paths.

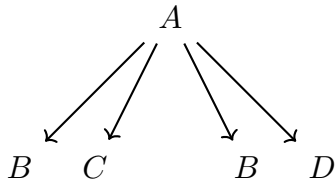


Figure 7: An evolution of readings.

Figure 8 shows another example. In it, reading A evolved into readings C , D , B , and H directly, and those readings gave rise to new readings (E and F) and to readings C and D via different paths.

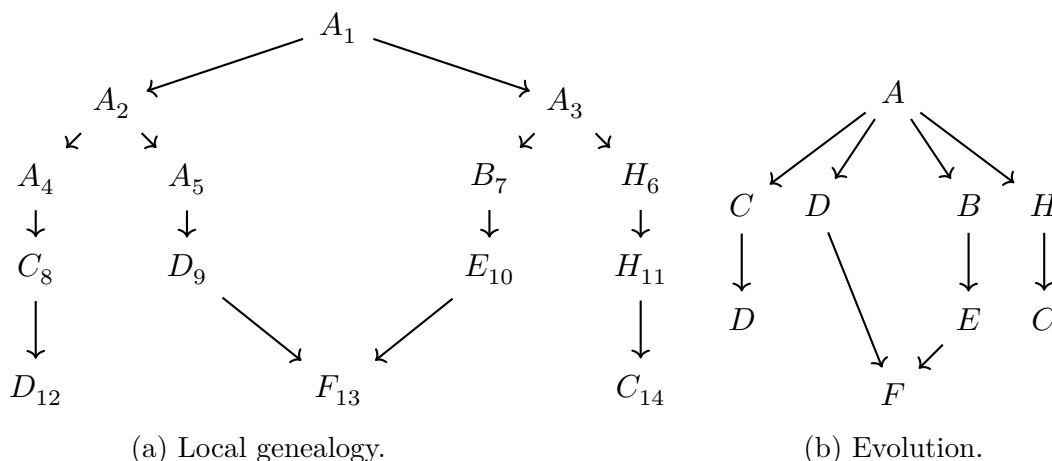


Figure 8: A more complex local genealogy and the evolution of its readings.

Despite their usefulness to summarize local genealogies, evolutions can be misleading if not interpreted correctly as one reading with a rich and complex tradition of hundreds of witnesses can be compressed to only one vertex in the evolution, making it visually similar to another reading that only appeared once in the genealogy.

1.3 Non-local genealogies

Definition 1.31 (Non-local genealogy). Let $H = (\mathcal{S}, \mathfrak{R})$ be a history over a tradition T . Let $D = \{(e, w) : ((e, (z, g)), (w, (x, h))) \in \mathfrak{R}\}$ (ie, the edges in D represent

exemplar-copy relationships). The directed graph $\mathcal{G} = (T, D)$ is a NON-LOCAL GENEALOGY (or genealogy for short) of the tradition T .

Non-local genealogies are representations of how witnesses descend from other witnesses. A graphical illustration of a computer-simulated genealogy is Figure 9 (by convention, first autographs are considered to have in-degree 0 and thus their null exemplars are not shown). It must be pointed out that non-local genealogies are not the same thing as *stemmata codicum*. *Stemmata* purport to reconstruct non-local genealogies (see Section 3.1.3).

Definition 1.32 (Contamination). Let $\mathcal{G} = (T, D)$ be a non-local genealogy. A witness $w \in T$ is CONTAMINATED if its in-degree is greater than 1.

For example, $(e_1, w_1), (e_2, w_1) \in D$ indicates that w_1 is a contaminated copy, as it was created from at least two exemplars, namely e_1 and e_2 ⁶.

When a scribe corrects a previous witness, the new witness (the correction) is considered to be a descendant of the previous witness (see Andrews and Macé 2013, 510). If the scribe corrected the witness using an additional auxiliary exemplar, then the new witness is said to be contaminated because it has two exemplars: the original witness and the auxiliary witness. For instance, let's consider the example in figure 10. Here, a “junior” scribe copied a witness w_0 from a certain manuscript and produced a copy w_1 on a manuscript m_1 (the manuscript m_1 is represented as a square that encloses the witnesses it contains). Then, a “senior” scribe corrects the copied witness w_1 to produce a new witness w_2 on m_1 that includes all the segments of w_1 that were correct as well as new segments that replace the ones that were incorrect. If the senior scribe took the corrected segments from another exemplar, say a witness w_3 from another manuscript, then w_2 is said to be a contaminated descendant of both w_1 and w_3 .

Definition 1.33 (Ancestor, descendant). Let $\mathcal{G} = (T, D)$ be a non-local genealogy, and let $w_i, w_k \in T$. If there is a directed path from w_i to w_k in \mathcal{G} , then w_i is an ANCESTOR of w_k (and w_k is a DESCENDANT of w_i). We denote “ w_i is an ancestor of w_k ” as $w_i \prec w_k$ and the equivalent relation “ w_k is a descendant of w_i ” as $w_k \succ w_i$. The symbols \preceq and \succeq represent the relations “is the same as or is an ancestor of” and “is the same as or is a descendant of”, respectively.

More informally, if a witness is a parent, a grandparent, or a great-grandparent (to any degree) of another, it is its ancestor.

Remark 1.2 (The null witness is the ultimate ancestor of all the witnesses in a non-local genealogy). A non-null witness cannot exist without an exemplar and thus it is either a properly composed copy or a descendant of a properly composed copy. Properly

⁶See Hoenen (2019) for a treatment of contamination from a Lachmannian perspective. Hoenen defines contamination not on non-local genealogies directly as we do, but rather on *stemmata codicum* (see Section 3.1.3 for the definition of *stemma codicum*).

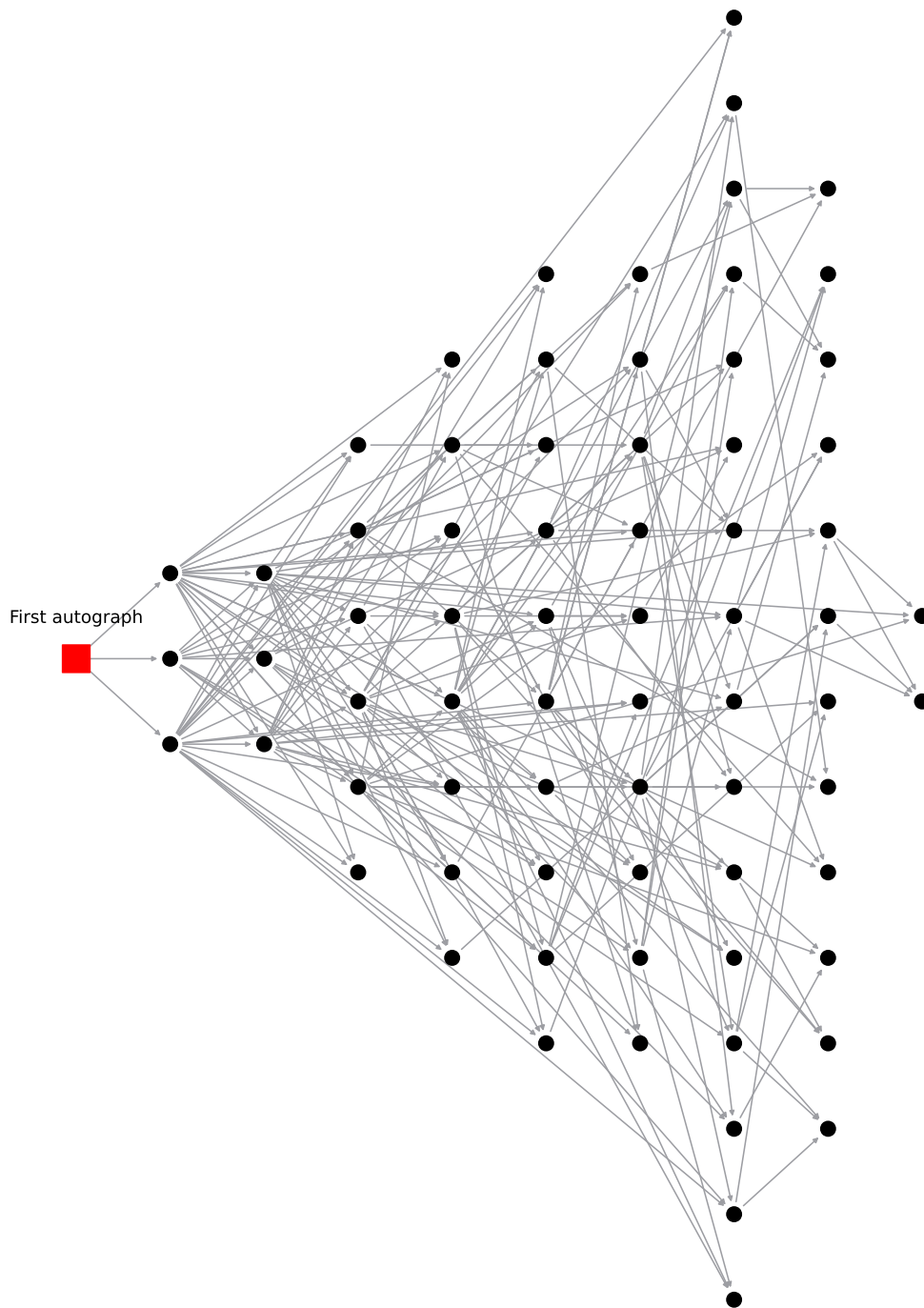


Figure 9: An example of a genealogy. The graph is topologically sorted and the red square is a first autograph.

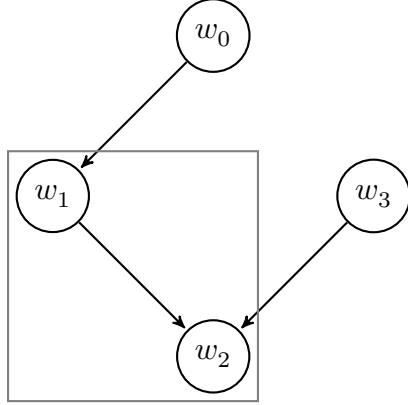


Figure 10: An example of a correction. Witnesses w_1 and w_2 are enclosed by a square to indicate that they belong to the same manuscript (m_1).

composed copies are children of the null witness, and therefore all non-null witnesses in a genealogy descend from the null witness.

Definition 1.34 (Latent non-local genealogy). Let T be a tradition. $\mathcal{G} = (T, D)$ is called the LATENT NON-LOCAL GENEALOGY of T if T is a latent tradition, D contains every and all the correct real-world exemplar-copy relationships between any two elements of T and \mathcal{G} forms exactly one weakly connected component⁷.

Intuitively, a latent genealogy is the “perfect” genealogical tree that is unknown to the researcher and contains all the witnesses that ever existed of a given work or set of works, as well as perfect information about their parent-child relationships. In an ideal world, we would like to know the latent non-local genealogy.

Latent genealogies are similar to what is sometimes referred to as *arbres réels* or “real trees” in stemmatology (see Hoenen 2020, 232). *Arbres réels*, however, have manuscripts, not witnesses, as their vertices. An implication of this is that if there is so-called “circular contamination”, a situation in which a manuscript contains a witness that is the indirect ancestor of another witness in the same manuscript (see Mink 2004, 50), the *arbre réel* is forced to include a cycle because the manuscript is an ancestor of itself (Roelli 2020, 213–14). Such a situation can be represented in a non-local genealogy without using cycles because the vertices in a genealogy are witnesses, not manuscripts. The example of figure 11 illustrates this. If we were to create an *arbre réel*, we would have that the manuscript at the top is an ancestor of the manuscript at the bottom because w_1 was copied from w_0 , but the manuscript at the bottom is an ancestor of the first manuscript, because w_2 was copied from w_1 .

⁷This constraint is added in order to guarantee that all the witnesses in the genealogy are genealogically related with one another in some way, either by being ancestors or descendants. Dropping the requirement that the graph form a weakly connected component would allow for the existence of clusters of witnesses that are not genealogically related to other witnesses.

The non-local genealogy, whose vertices are witnesses rather than manuscripts, does not contain a cycle.

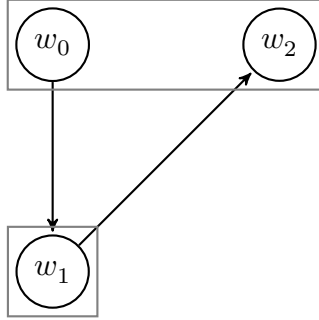


Figure 11: An example of circular contamination (the rectangles represent manuscripts).

It must be pointed out that even though some scholars have called this phenomenon “circular contamination”, it is not true contamination as we define the term in our model, given that it does not entail the creation of witnesses whose in-degree is greater than 1. We do not provide a formal definition of this kind of pattern, but one could be created from the existing definitions of the model if needed.

Non-local genealogies may contain cycles, however. This is because a scribe can use a segment from an already copied section of a witness as input to create another segment in the same witness. This effectively means that the witness is an ancestor of itself. The cycles can be mediated by other witnesses as well.

Definition 1.35 (Generation number). Let $\mathcal{G} = (T, D)$ be a non-local genealogy. Let $w, v \in T$ be two arbitrary witnesses such that $w \prec v$. A GENERATION NUMBER for w and v (or the number of generations between w and v) is the length of a directed path from w to v in \mathcal{G} (the length of a path is the number of its edges).

Generation numbers are often invoked when counting the generations between a first autograph and an extant witness. Our definition allows us to count the number of generations between any two witnesses that share an ancestor-descendant relationship.

Remark 1.3 (There may be multiple generation numbers between two witnesses). Let $\mathcal{G} = (T, D)$ be a non-local genealogy. There can exist $w, v \in T$ such that $w \prec v$ and the number of directed paths from w to v is greater than one. Consider for example the edges $\{(w, e), (e, f), (f, v), (w, g), (g, v)\} \subset D$. The directed paths implied by them are $w \rightarrow e \rightarrow f \rightarrow v$ and $w \rightarrow g \rightarrow v$. The generation number for the first path is 3, while the generation number for the second is 2. Therefore, there are multiple generation numbers between w and v . Following a similar logic, we could also show that there may be multiple generation numbers between any two *segments*.

Definition 1.36 (Most recent common ancestor). Let $A = (B, C)$ be a directed graph and let $B' \subset B$ be an arbitrary set of vertices. Define $B^{\text{comm}} = \{b \in B : b \prec$

b' for all $b' \in B'$ where $b \prec b'$ means that there is a directed path from b to b' in A . That is, B^{comm} is the set of the common ancestors of the vertices of the set B' over the graph A . Let $A^{\text{comm}} = (B^{\text{comm}}, C^{\text{comm}})$ be a subgraph of A such that for any $x, y \in B^{\text{comm}}$ it holds that $(x, y) \in C^{\text{comm}}$ iff $(x, y) \in C$ (i.e., A^{comm} is the subgraph induced in A by B^{comm}). The vertices of A^{comm} that have out-degree equal to zero are the MOST RECENT COMMON ANCESTORS⁸ of the set B' .

Definition 1.37 (Archetype, latent archetype). Let $\mathcal{G} = (T, D)$ be a non-local genealogy and let $T' \subset T$ be an arbitrary set of witnesses. The most recent common ancestors of T' over \mathcal{G} are called the ARCHETYPES of T' . If \mathcal{G} is a latent non-local genealogy, the archetypes are called LATENT ARCHETYPES.

The idea behind our definition of archetype is straightforward. An archetype is a specific witness that can be traced as a most recent ancestor of all the witnesses in a given set of witnesses. Note that these are different (but similar) to Lachmannian archetypes (Definition 3.10). Depending on the specific shape of the non-local genealogy and on which witnesses are included in the set of descendants, there may be more than one archetype, or even none. Figure 13 shows an example of a set of witnesses with one archetype and one example of a set of witnesses with two archetypes. In each pane, all the magenta vertices are descended from the blue vertex (or vertices) and the highlighted edges show all the directed paths between the archetypes and the magenta vertices. A genealogy where the subset of witnesses (the magenta vertices) has no archetype is shown in Figure 12 (there are two first autographs and the magenta vertices do not have a common ancestor).

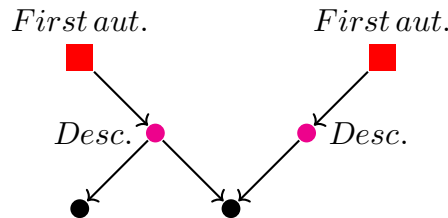
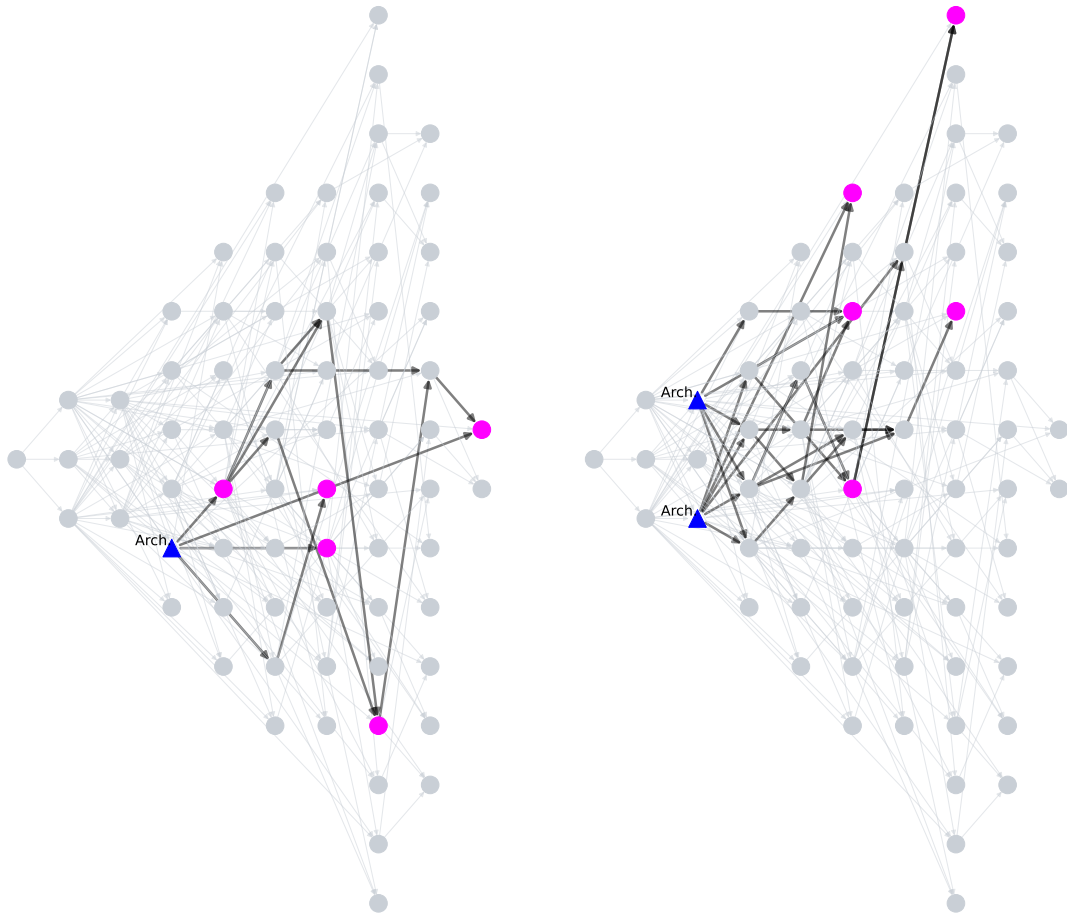


Figure 12: An example of a witness set with no archetypes. The first autographs are the red squares and the descendants are the magenta vertices.

If $T' = T^\times$ (ie, if it is the extant tradition), the archetypes are the most recent sources from which all the surviving copies are directly or indirectly derived. They are not necessarily the author's autographs, but rather the latest witnesses in the genealogy that gave rise to all the extant copies. The archetypes may already contain errors or alterations introduced during the previous copying process.

⁸This is an adaptation of the traditional concept of lowest common ancestor, which is defined for pairs of descendants, to apply to larger sets of descendants (see Bender et al. 2005).



(a) Single archetype

(b) Multiple archetypes

Figure 13: Examples of archetypes. The magenta vertices are arbitrary witnesses and the blue triangles are their archetypes. All the directed paths between the archetypes and the magenta vertices have been highlighted.

Definition 1.38 (Hyparchetype, latent hyparchetype). Let $\mathcal{G} = (T, D)$ be a non-local genealogy and let $T' \subset T$ be an arbitrary set of witnesses. Let $T^{\text{arch},'} \subset T$ be the archetypes of the set T' over \mathcal{G} . Define $\mathcal{W}_{T'} = \{W \in \wp(T') : |W| > 1\}$, where $\wp(T')$ is the power set of T' . Define $T^{\text{hyparch},'} = \{h \in T : h \text{ is an archetype of } W \text{ over } \mathcal{G} \text{ for } W \in \mathcal{W}_{T'} \text{ and there exists at least one } a \in T^{\text{arch},'} \text{ such that } a \prec h\}$. The elements of the set $T^{\text{hyparch},'}$ are the HYPARCHETYPES of T' over \mathcal{G} . If \mathcal{G} is a latent non-local genealogy, the hyparchetypes are called LATENT HYPARCHETYPES.

The most common case is when $T' = T^\times$, ie, when it is the extant tradition. Textual critics are often interested not only in the archetype(s) of the extant tradition considered as a whole, but also in the intermediate archetypes (or “hyparchetypes”) of smaller witness sets. The hyparchetypes are descendants of the archetypes of T' but are, at the same time, archetypes themselves of any of all the possible combinations of the witnesses of T' . Note that these are different (but similar) to Lachmannian hyparchetypes (Definition 3.11).

Definition 1.39 (Text type). Let $\mathcal{G} = (T, D)$ be a latent non-local genealogy. Let $a \in T$ be an arbitrary witness and let $T_{w \succ a} = \{w \in T : w \succ a\}$, that is, it is the set that contains all of the descendants of a over \mathcal{G} . The set that contains all the witnesses in the tradition except for a and its descendants is $T - (T_{w \succ a} \cup \{a\})$. The set $T_{w \succ a} \cup \{a\}$ is the TEXT TYPE of the archetype a if for each $w \in T_{w \succ a} - \{a\}$ it does not hold that $w \succ q$ for all $q \in T - (T_{w \succ a} \cup \{a\})$.

In words, a text type is a set of witnesses that is “self-contained”, in the sense that the ancestors of the elements of the set are also in the group (ignoring the ancestors of the witness a itself which, by definition, are also ancestors of its descendants). The idea of text types has been around for a long time, especially in the textual criticism of the New Testament. Its origins can be traced to the early eighteenth century, when scholars started to group manuscripts along geographical lines, qualitatively clustering them according to perceived similarity in order to facilitate the processing of the Greek New Testament’s proverbially abundant extant tradition (Epp 2013, 523–53). The term caught on with time, with most critics using it freely despite there being no agreed upon definition. Unsurprisingly, scholars in the twenty-first century have been proposing a move away from fixed types and towards understanding witness relationships in terms of clusters (Epp 2013, 553–71), in order to avoid the frequent misunderstandings that surround the term. This is also true of textual critics working on the non-Greek traditions, like the Latin New Testament (Burton 2013, 177–86), where text types have also been used in the past. Scholars specializing in the textual criticism of the Hebrew Bible have too brought to attention the vagueness with which the concept has been used after it was imported from the field of Greek New Testament studies (Martín-Contreras and Miralles-Maciá 2014, 18–20).

Generally speaking, the practical implication of using text types is that once the textual critic has “detected” one, the (reconstructed) ancestor a can be used as a representative of its descendants in such a way that the number of witnesses to collate

becomes lower. Entire sets of witnesses are collapsed this way into single, hypothetical witnesses. It is the author’s impression that, in practical terms, textual critics have traditionally equated the term text type with a very vague notion of a set of witnesses that are descended from a single archetype, with no presence of readings from the archetypes of the other text types. That is, the key assumption appears to be that text types descend from distinct, individual archetypes. Whenever a witness is thought to have descended from more than one such archetype, the witness is said to be “mixed” and thus cannot be sharply assigned to one text type or the other but rather its “distinctive readings” are counted in order to calculate rough proportions, so that the witness can be classified as, say, “mostly Alexandrian, with some Byzantine readings”, or “mostly Masoretic with some Qumran readings”, etc. Figure 14 illustrates. In the figure, we see two direct descendants of the autograph, which are the ancestors of their respective text types (they are shown as triangles with thicker perimeters). The witnesses that descend from these two ancestors (plus the ancestors themselves) are the two text types, namely the yellow (left) and the blue (right) text types, shown as coloured shades. The green witness to the top descends from both the yellow and blue ancestors and is therefore not given a blanket assignment but is rather considered “mixed”.

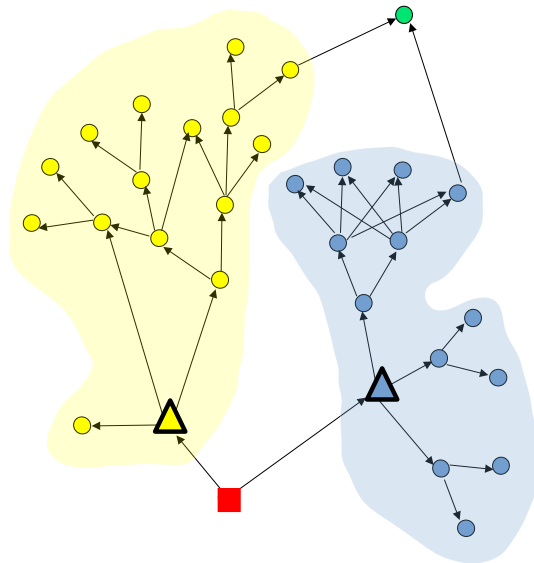


Figure 14: An illustration of text types. The triangles are the ancestors of the other witnesses in their respective text types.

Two related notions are textual families and manuscript groups. There is significant confusion in the way these terms are understood and used, not only by themselves but also in relation to text types. Generally speaking, text types are considered to include many witnesses and manuscript groups (really, “witness groups” if one wanted to be

more precise) to include much fewer. Families would lie somewhere in between. A way to solve the nomenclature puzzle could be to define them as “nested text types”, that is, a witness group would be a text type that is a subset of a family, and a family would be a text type that is a subset of a “real” text type.

The following remarks apply to latent non-local genealogies with only one first autograph and no other composed copies. This topology is important because it is the most often presumed topology for latent non-local genealogies in textual criticism discussions.

Remark 1.4 (Single-first-autograph latent traditions that descend entirely from the first autograph). In a latent non-local genealogy with only one first autograph and no other composed copies, the entire tradition (with the exception of the first autograph) is descended from the first autograph. Let $\mathcal{G} = (T, D)$ be a latent non-local genealogy. If \mathcal{G} has no composed copies apart from the first autograph a , then all witnesses except for a must be copied from a non-null witness, ie, there does not exist $(\varepsilon, w) \in D$ such that $w \neq a$. Thus, for a non-null copy $u \in T$ either $u = a$ or $u \leftarrow v$, where v is a non-null witness (the arrow represents an edge in the genealogy). Assume that $v \neq a$. We know that all non-null copies descend from the null witness (Remark 1.2), so we would have that there exists at least one $w \preceq v$ such that $(\varepsilon, w) \in D$ and $w \neq a$, but this contradicts our statement that the only composed copy in \mathcal{G} is a . Thus, $v \succeq a$. In consequence, for any non-null witness $u \in T$ it holds that either $u = a$ or $u \succ a$.

Remark 1.5 (Single-first-autograph text-types). In a latent tradition with only one first autograph and no other composed copies, the entire tradition is at the same time the first autograph’s text type. Let $\mathcal{G} = (T, D)$ be a latent non-local genealogy and a the only first autograph in it. The definition of a text type requires that for each descendant w of a the proposition $w \succ q$ does not hold, for all $q \in T - (T_{w \succ a} \cup \{a\})$. We know from Remark 1.4 that all the members of T except for a are descendants of a , that is, that $(T_{w \succ a} \cup \{a\}) = T$, and therefore $T - (T_{w \succ a} \cup \{a\}) = \emptyset$. Thus, the proposition $w \succ q$ is trivially false for all w and the latent tradition $T = (T_{w \succ a} \cup \{a\})$ is the text type of the first autograph a .

1.4 Scribal error

In real-world textual transmissions, scribal error happened when a scribe read a segment from his or her exemplar and produced a segment that was not identical to it. For example, the scribe could have read “Delilah said to Samson” in the exemplar and written down “She said to Samson” in the copy (substitution of “she” for “Delilah”).

Definition 1.40 (Perfectly copied segment, scribal error). Let $G_\sigma = (\mathcal{S}_\sigma, \mathfrak{K}_\sigma)$ be a local genealogy and let $((m, \vec{s}, \lambda_i), (x, h)) \in \mathcal{S}_\sigma$ be a segment such that its in-degree is equal to one and its parent segment is not the null segment. Let $((l, \vec{q}, \delta), (z, g)) \in \mathcal{S}_\sigma$ be the parent segment. $((m, \vec{s}, \lambda), (x, h))$ is a PERFECTLY COPIED SEGMENT if $\vec{s} = \vec{q}$. Any non-null segment in \mathcal{S}_σ is a SCRIBAL ERROR if it is not a perfectly copied segment.

Intuitively, we can understand a scribal error as a segment where the scribe failed to copy the text correctly, thus producing a different string, or when the scribe “invented” a new reading (this happens when the parent segment is the null segment). Our definition also implies that all conflations are scribal error, because conflations cannot be perfectly copied segments (a perfectly copied segment must have in-degree equal to 1). To illustrate why this is so, consider a scribe that takes one segment from one witness and another segment from another witness. Let the first segment be “beginning God” and let the second be “beginning the Lord”. The scribe could produce, for example, “beginning the Lord God”, but also “beginning God the Lord”. Given that neither of the two outputs appears verbatim in either of the parent segments (and in fact the scribe has to decide the order of the segments, thus adding information not present in the parent segments), none of the two outputs can be said to be identical to the inputs and, therefore, both outputs are scribal errors.

Remark 1.6 (An imperfect copy does not necessarily contain scribal error). Let $w = (m, \vec{s}, \lambda)$ be a contaminated copy and let e and d be its exemplars such that the strings of e and d are not equal. w is imperfect because it is contaminated. There can exist an $\vec{s} = \vec{f}_1 \vec{g}_2 \vec{f}_3 \vec{g}_4 \dots \vec{f}_{n-1} \vec{g}_n$ where \vec{f}_i is the string of a segment of e that was perfectly copied for the i -th segment of \vec{s} , and, similarly, \vec{g}_i is the string of a segment of d that was perfectly copied for the i -th segment of \vec{s} . The string \vec{s} is made up of strings from perfectly copied segments, which means that w does not contain scribal errors, even though it is an imperfect copy.

The above observation simply highlights the fact that even when scribes are “perfect” and do not commit any transcription mistakes, the resulting copies may still be imperfect, due to the effect of contamination. Scribal error happens at the segment level, while imperfect copying is defined for the witness’ string as a whole.

Scribal errors are sometimes classified according to their perceived causes. For instance, “omissions” happen when a scribe is thought to have dropped a segment from the exemplar, as in “and David came” for “and David and Joab came”. Other common error types are additions, transpositions, substitutions, fissions and fusions.

1.5 Translations

Ancient works often survive not only in their original languages but also in ancient translations. These translations provide valuable information that the textual critic can use in order to reconstruct the earliest stages of the work.

Before a textual critic can start considering translational evidence, she needs to be aware that ancient translations have traditions and genealogies of their own, and that their archetypes are usually no longer extant. The Septuagint and the Vulgate, for example, have extraordinarily complex textual histories that span across geographies and eras (Kreuzer 2015; Houghton 2016; Linde 2015). Thus, a large part of the work associated to using translational evidence in textual criticism has to do with reconstructing the archetype(s) of the translation from its extant tradition.

Definition 1.41 (Translation, translated segment). Let $L \subset \mathcal{A}^*$ be a finite language where \mathcal{A}^* is an alphabet’s Kleene closure. Let $O \subseteq L$ and $Q \subseteq L$ be languages. Given a non-local genealogy $\mathcal{G} = (T, D)$ such that the strings of its witnesses are all elements of L , and given an edge $((m, \vec{s}, \lambda), (l, \vec{q}, \delta)) \in D$, the copy (l, \vec{q}, δ) is a TRANSLATION of the exemplar (m, \vec{s}, λ) if $\vec{s} \in O$, $\vec{q} \in Q$ and $\vec{q} \notin O$. A segment from a translation is called a TRANSLATED SEGMENT.

Otherwise said, a translation is a copy that is in a language different from the language of at least one of its exemplars. Translations, like any copy, are associated to scribes⁹, and the same scribe can be a normal scribe or a translator depending on the copying event. Of course, nothing forbids a scribe from being associated to a translation made from an exemplar that is itself a descendant of a translation. For example, the Latin Vulgate’s Gallican Psalter was translated from a Greek Psalter which, in turn, was translated from a Hebrew Psalter.

As translations are copies, they can also be contaminated if they have more than one exemplar.

1.6 Topological complications

Our model allows for many topological complications. For example, one can define latent genealogies with more than one author and more than one first autograph, autographic copy or autographic edition. It can also help represent genealogies with more than one work, which could be particularly useful for florilegia, chronicles, lectionaries, psalters and other similar situations. Some genealogies may contain witnesses in which a segment is imported from a parallel passage from the same work or from another work and then added to the copy either as a conflation or a standalone segment, creating harmonizations. One example is scroll 4QDeutⁿ, which contains passages of Deuteronomy that have been conflated with parallel passages in Exodus (Eshel 1991).

These complications apply to traditions of witnesses copied in the original language as well as to translational traditions. A complication specific to the latter is when new translations are incorporated into an existing translational tradition. Taking the Septuagint as an example, one could imagine a scribe making a new translation of a Hebrew Psalm into Greek and then another scribe creating a contaminated copy by taking half of the Psalm from a Septuagint exemplar and the other half from the new translation. Figure 15 illustrates, showing a synthetic complex topology in which a genealogy of Hebrew witnesses (gray vertices and edges, with first autographs in red) and a translational genealogy of Greek witnesses (in purple) have been joined. Notice that this genealogy has three first autographs and that the translational genealogy was created from two Hebrew exemplars (shown in orange). One exemplar (the one to the right) gave rise to the translational tradition and the other one was incorporated later. Real-world genealogies are of course much more complex than this toy example.

⁹See Screnock (2017) for a recent exploration of the idea of translators as scribes from a textual criticism perspective.

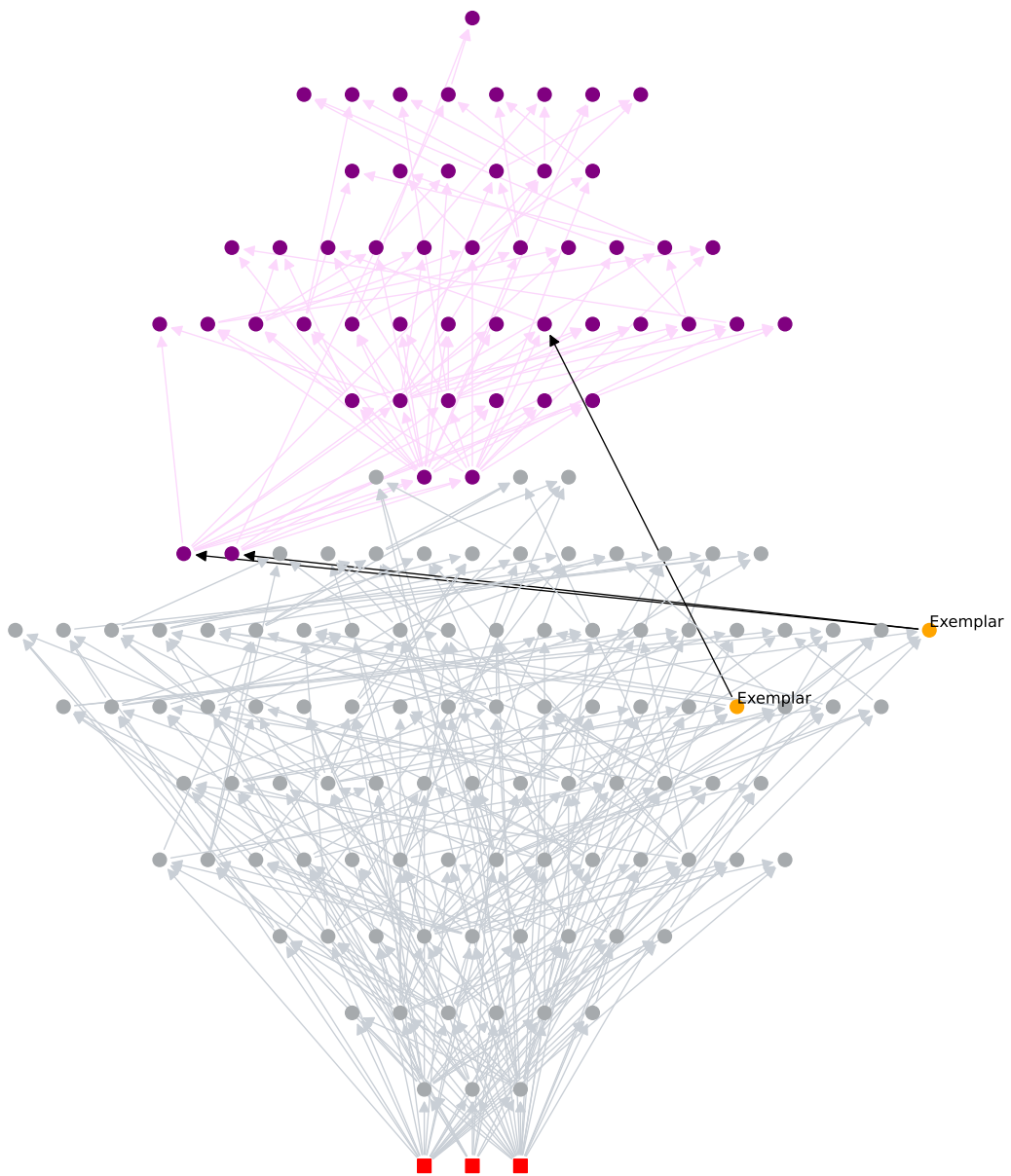


Figure 15: A simulated latent genealogy in the original language (gray, bottom) with a versional genealogy attached to it (purple, top). The red squares are first autographs and the orange vertices are the exemplars used by the translators.

2 Reconstruction of the original text

Armed with the previous conceptual apparatus, we can examine the traditional task of the textual critic, which is the reconstruction of the original text. This is best done with an example. Psalm 145 verses 13-14 presents an interesting problem. The psalm is an acrostic that is missing the line corresponding to the Hebrew letter *nun* in the majority (Masoretic) text. The Masoretic text is reflected in the Jewish Publication Society translation of 1917, which reads:

13. Thy kingdom is a kingdom for all ages, and Thy dominion endureth throughout all generations.
14. The LORD upholdeth all that fall, and raiseth up all those that are bowed down.

Here verse 13 starts with the Hebrew letter *mem* and verse 14 with the letter *samekh*. In the Hebrew alphabet, the letter *nun* is in between *mem* and *samekh*.

Ancient Qumran scroll 11QPS^a has a line between these two that says (Abegg, Flint, and Ulrich 1999, 571):

13. [...] God is faithful in his words, and gracious in all his deeds.

This line starts with the letter *nun*, thus completing the *mem-nun-samech* sequence. The Greek Septuagint and the Syriac Peshitta also have this line, which suggests that their translators used exemplars that had it. For the Latin Vulgate, the Psalter that was translated from the Hebrew agrees with the Masoretic text, while the Gallican Psalter, which was translated from the Greek, agrees with the Septuagint.

Major modern translations normally include this line¹⁰, although some scholars have expressed doubts about its originality on contextual grounds (Kimelman 1994). For example, the *nun* verse is an almost verbatim repetition of verse 17, which may be seen as unusual, and is present in a Qumran scroll that had signs of having been copied rather freely, given that it adds at various places a constant refrain¹¹ which is not present in any of the other extant witnesses. Additionally, the phenomenon of incomplete acrostics is not unusual in the Psalms, as all the other davidic Psalms that are acrostics (Ps. 25, 34, and 37) are missing a letter or two (Kimelman 1994, 50).

Although we are not going to solve the issue of the originality of the *nun* line of Psalm 145 nor discuss it in depth, we will use it to illustrate the problem that the textual critic has to face. In the exposition that follows, we will uphold simplifying assumptions for the sake of clarity.

¹⁰So the New International Version, the Christian Standard Bible, and the English Standard Bible (in brackets), among others. The New King James Bible does not include it but mentions it in a footnote. The New Jewish Publication Society Tanakh omits it.

¹¹“Blessed be the Lord and blessed be his name forever and ever” (Abegg, Flint, and Ulrich 1999, 570).

Definition 2.1 (Hypothetical witness, hypothetical segment). Given a latent tradition $T \subset \mathcal{T}$ and an extant tradition $T^\times \subset T$, where \mathcal{T} is the universal witness set, a HYPOTHETICAL WITNESS is any non-null witness $u \in (\mathcal{T} - T)$. The segments of witness u are called HYPOTHETICAL SEGMENTS.

Note that hypothetical witnesses not only are non-extant, but they also cannot be part of the latent tradition. Hypothetical witnesses are “imaginary” witnesses that a textual critic constructs, possibly in order to hypothesize what an actual non-extant witness looked like. For instance, a witness that actually existed but is no longer extant (so it is in $T - T^\times$) cannot be as such recovered, but a new witness (a hypothetical one) can be designed to approximate it. That new witness is not part of the latent tradition because it is an invention created by the textual critic, it never existed within the scribal transmission process.

Definition 2.2 (Reconstruction, reconstructed segment). Let $T \in \mathcal{T}$ be a latent tradition where \mathcal{T} is the universal witness set. Let $w \in \mathcal{T}$ be a fragmentary witness. Define a new witness $\hat{w} \in (\mathcal{T} - T)$ such that it is a copy of w according to a copying event K . Notice that \hat{w} is a hypothetical witness by Definition 2.1. Let \vec{s} be the string of w and let $\vec{\hat{s}}$ be the string of \hat{w} . \hat{w} is called a RECONSTRUCTION of witness w if there exists at least one $((w, (x, h)), (\hat{w}, (z, g)), \zeta) \in K$ such that i) $\vec{s}(x : x + h)$ is a lacuna and $\vec{\hat{s}}(z : z + g)$ is not and ii) there does not exist a $((w, (a, b)), (\hat{w}, (c, d)), \zeta) \in K$ such that $\vec{s}(a : a + b) \neq \vec{\hat{s}}(c : c + d)$ for all $\vec{s}(a : a + b)$ that are not a lacuna. A segment $(\hat{w}, (z, g))$ is called a RECONSTRUCTED SEGMENT if \hat{w} is a reconstruction and it is the case that $\vec{s}(x : x + h)$ is a lacuna and $\vec{\hat{s}}(z : z + g)$ is not for all $((w, (x, h)), (\hat{w}, (z, g)), \zeta) \in K$.

For instance, let’s say that the string of a witness w is “And David brought the **g** to **ggggggggg**”. A textual critic can create a reconstruction \hat{w} that says “And David brought the ark to Jerusalem” by replacing the lacunae with the strings “ark” and “Jerusalem”. Another option could have been “And David brought the army to **ggggggggg**”, in which the first lacuna was replaced but the second one was not. From our definition above, a reconstruction can only alter lacunose substrings. Any non-lacunose substrings must be left unchanged (for instance, “And Solomon brought the ark to Jerusalem” would not qualify as a reconstruction).

Definition 2.3 (Available reading, extant reading, conjecture, lost reading). Let $T^\times \cup \hat{T}$ be a tradition where T^\times is an extant tradition and \hat{T} is a set of hypothetical witnesses. Let $\mathcal{C}(i)$ be a textual unit from a collation made from $T^\times \cup \hat{T}$ and let $R_{\mathcal{C},i}$ be its reading set. A reading in $R_{\mathcal{C},i}$ is AVAILABLE if it does not contain a lacuna. An available reading is called EXTANT if it is attested by at least one extant witness. It is called a CONJECTURE if it is not extant and is therefore only attested to by hypothetical segments. All the readings in $R_{\mathcal{C},i}$ that are not available and those readings that are attested exclusively by lost witnesses (ie, witnesses in the latent tradition but not present in $T^\times \cup \hat{T}$ and therefore not present in $R_{\mathcal{C},i}$ either) are called LOST.

Our classification, thus, splits all possible readings into two classes, available and lost, and further splits available readings into extant and conjectural:

1. *Available*: Readings that can be used by the textual critic. They are in the reading set $R_{\mathcal{C},i}$ and are not lacunose.
 1. *Extant*: At least one extant witness attests to the reading.
 2. *Conjectural*: The reading does not appear in any extant witness; it was created by textual critics.
2. *Lost*: The reading existed at some point in time in at least one witness but is currently not attested.
 1. Readings that are in the reading set $R_{\mathcal{C},i}$ but are lacunose.
 2. Readings attested only in lost witnesses (they are not in the reading set $R_{\mathcal{C},i}$).

Definition 2.4 (Target witness, originality status, sample space, hypothesis). Let T be a latent tradition, $T^\times \subset T$ an extant tradition and \bar{T} a set of reconstructions. Let $q \in \mathcal{T}$ be an arbitrary witness, where \mathcal{T} is the universal set of witnesses. We will call q the TARGET WITNESS. Create an n -collation \mathcal{C} with configurations from the witnesses in $T^\times \cup \bar{T}$. Let \mathcal{O} be a set of mutually exclusive and collectively exhaustive states called ORIGINALITY STATUSES such that \mathcal{O} contains exactly one element that represents the idea of a reading being attested by the target witness at a given segment. Let $R_{\mathcal{C},i}$ be a reading set and $\mathbf{o} \in \mathcal{O}$ the element that represents attestation in q . Define $\Omega_{\mathcal{C},i} \subset (\{\vec{\rho}_1\} \times \mathcal{O}) \times (\{\vec{\rho}_2\} \times \mathcal{O}) \times \cdots \times (\{\vec{\rho}_l\} \times \mathcal{O})$ for $\vec{\rho}_1, \vec{\rho}_2, \dots, \vec{\rho}_l \in R_{\mathcal{C},i}$ such that each element of $\Omega_{\mathcal{C},i}$ contains at most one $(\vec{\rho}, \mathbf{o})$, that is, no element of $\Omega_{\mathcal{C},i}$ can contain more than one reading marked as original. The set $\Omega_{\mathcal{C},i}$ is called the SAMPLE SPACE of textual unit $\mathcal{C}(i)$ and each of the elements $\omega \in \Omega_{\mathcal{C},i}$ is called a HYPOTHESIS for the textual unit.

We can understand an original reading as the reading that is attested by the witness that the textual critic is trying to reconstruct. Normally, it is first autographs which are singled out as target witnesses, but this need not always be the case.

Coming back to our example, the first step in a textual criticism analysis is normally the gathering of witnesses and their collation. Our extant tradition will be the Masoretic text \mathfrak{M} (as represented by the Leningrad Codex) and the 11QPS^a scroll (Q). After collating the witnesses, we extract the textual unit that corresponds to verse 13 and find the readings¹², namely the empty string, which we will denote by $\vec{\rho}_{\mathfrak{M}}$, and the string “God is faithful in his words and gracious in all his deeds” ($\vec{\rho}_Q$). Both readings are available and are extant, which means that our textual unit does not contain conjectures. Let $R_{\mathcal{C},i} = \{\vec{\rho}_{\mathfrak{M}}, \vec{\rho}_Q\}$ be the reading set of our textual unit.

Let’s define a set of originality statuses in order to classify the readings:

¹²The Leningrad Codex and 11QPS^a are in Hebrew, but we show the readings in English translation for simplicity.

1. *Original* (notated as **o**).
2. *Intentional variant* (notated as **i**). Examples are pious expansions and euphemisms.
3. *Unintentional variant* (notated as **u**). One example is *homoteleuton*.
4. *Lost* (notated as **l**). Only applicable to readings that contain lacunae.

The set $\mathcal{O} = \{\mathbf{o}, \mathbf{i}, \mathbf{u}\}$ will therefore be our originality status set (we do not include the originality status **l** in \mathcal{O} as none of the competing readings are lacunose).

Using this notation, we can say that for Psalm 145:13, Kimelman (1994) considers that the originality status of $\vec{\rho}_{\mathfrak{M}}$ is **o** and that that of $\vec{\rho}_Q$ is **i**, while the translators of most modern versions consider that the originality status of $\vec{\rho}_Q$ is actually **o**, and that the status of $\vec{\rho}_{\mathfrak{M}}$ is **u** as the *nun* line could have been accidentally dropped from a (proto-)Masoretic archetype.

Let's now combine the reading set and the originality status set to determine all the potential hypotheses about the original text of the textual unit. Let's call the textual unit's sample set $\Omega_{e,i}$. We can create a sample set like Equation 2.

$$\Omega_{e,i} = \left\{ \begin{array}{l} ((\vec{\rho}_{\mathfrak{M}}, \mathbf{o}), (\vec{\rho}_Q, \mathbf{i})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{o}), (\vec{\rho}_Q, \mathbf{u})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{i}), (\vec{\rho}_Q, \mathbf{o})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{i}), (\vec{\rho}_Q, \mathbf{i})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{i}), (\vec{\rho}_Q, \mathbf{u})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{u}), (\vec{\rho}_Q, \mathbf{o})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{u}), (\vec{\rho}_Q, \mathbf{i})), \\ ((\vec{\rho}_{\mathfrak{M}}, \mathbf{u}), (\vec{\rho}_Q, \mathbf{u})) \end{array} \right. \quad (2)$$

We can also represent $\Omega_{e,i}$ in tabular form, as in Table 1.

Table 1: A tabular representation of $\Omega_{e,i}$

Original	Intentional variant	Unintentional variant	Comments
$\vec{\rho}_{\mathfrak{M}}$	$\vec{\rho}_Q$		Kimelman (1994)
$\vec{\rho}_{\mathfrak{M}}$		$\vec{\rho}_Q$	
$\vec{\rho}_Q$	$\vec{\rho}_{\mathfrak{M}}$		
	$\vec{\rho}_{\mathfrak{M}}, \vec{\rho}_Q$		
	$\vec{\rho}_{\mathfrak{M}}$	$\vec{\rho}_Q$	

Original	Intentional variant	Unintentional variant	Comments
$\vec{\rho}_Q$		$\vec{\rho}_M$	Most modern translators
	$\vec{\rho}_Q$	$\vec{\rho}_M$	
		$\vec{\rho}_M, \vec{\rho}_Q$	

Definition 2.5 (Textual theory). Given a sample space $\Omega_{\mathcal{C},i}$, let $\mathcal{F}_{\mathcal{C},i}$ be a partition of it. Let $\cup f = \{(\vec{\rho}, y) : (\vec{\rho}, y) \in \omega, \omega \in f\}$ for any $f \in \mathcal{F}_{\mathcal{C},i}$. Constrain $\mathcal{F}_{\mathcal{C},i}$ in such a way that for all $f \in \mathcal{F}_{\mathcal{C},i}$ it holds that $\vec{\rho}_i = \vec{\rho}_j$ for all $(\vec{\rho}_i, \mathbf{o}), (\vec{\rho}_j, \mathbf{o}) \in \cup f$. This restriction guarantees that in each cluster of the partition at most one reading can be original. Let $(\Omega_{\mathcal{C},i}, \mathcal{F}_{\mathcal{C},i}, \mathbb{P}_{\mathcal{C},i})$ be a probability space where $\mathbb{P}_{\mathcal{C},i}$ is the probability function. $(\Omega_{\mathcal{C},i}, \mathcal{F}_{\mathcal{C},i}, \mathbb{P}_{\mathcal{C},i})$ is a TEXTUAL THEORY of the textual unit $\mathcal{C}(i)$.

We will provide an example of a textual theory momentarily.

Textual theories are the way we model the beliefs held by textual critics about the originality statuses of the competing readings of a given textual unit. Textual critics have different ways of creating textual theories, including a whole array of heuristics that take into account many types of evidence, ranging from the stylistic features of the scribe's handwriting, the spelling and grammatical profile of the text, the carefulness or lack thereof of the scribe, the ideological affiliation of the scriptorium where the manuscript was produced, etc.

Remark 2.1 (The probability function in a textual theory can be made to depend on the probabilities assigned to the readings of other units). A reading can be considered more likely to be original depending on if a certain set of readings is also considered likely original in other textual units of the collation. This is often seen with harmonizations, where the presence of harmonizations earlier in a witness can make an unusual segment later on be considered likely to be a harmonization as well. Our example from the Psalms provides another illustration when Kimelman argues that the *nun* line is not original because 11QPS^a has a spurious refrain present at other textual units, which could indicate that the witness or one of its ancestors were created by a scribe who took too many creative liberties. Thus, even though we are modeling textual theories as happening for each textual unit separately, this does not mean that the probabilities of the readings of different textual units cannot affect one another.

We have 8 outcomes in our example, and we can group them into 3 events: either $\vec{\rho}_M$ is original, or $\vec{\rho}_Q$ is original, or none are original (see Table 2). Each event has a probability associated to it and the sum of the probabilities is equal to 1.

Table 2: Events from the sample space.

Event	Outcomes
$\vec{\rho}_{\mathfrak{M}}$ alone is original	$((\vec{\rho}_{\mathfrak{M}}, \mathbf{o}), (\vec{\rho}_Q, \mathbf{i}))$ and $((\vec{\rho}_{\mathfrak{M}}, \mathbf{o}), (\vec{\rho}_Q, \mathbf{u}))$
$\vec{\rho}_Q$ alone is original	$((\vec{\rho}_{\mathfrak{M}}, \mathbf{i}), (\vec{\rho}_Q, \mathbf{o}))$ and $((\vec{\rho}_{\mathfrak{M}}, \mathbf{u}), (\vec{\rho}_Q, \mathbf{o}))$
None are original	$((\vec{\rho}_{\mathfrak{M}}, \mathbf{i}), (\vec{\rho}_Q, \mathbf{i})), ((\vec{\rho}_{\mathfrak{M}}, \mathbf{i}), (\vec{\rho}_Q, \mathbf{u})),$ $((\vec{\rho}_{\mathfrak{M}}, \mathbf{u}), (\vec{\rho}_Q, \mathbf{i}))$ and $((\vec{\rho}_{\mathfrak{M}}, \mathbf{u}), (\vec{\rho}_Q, \mathbf{u}))$

Definition 2.6 (Textual thesis). Let $(\Omega_{\mathcal{C},i}, \mathcal{F}_{\mathcal{C},i}, \mathbb{P}_{\mathcal{C},i})$ be a textual theory. Let $\Theta_{\mathcal{C},i} = \{f \in \mathcal{F}_{\mathcal{C},i} : \mathbb{P}_{\mathcal{C},i}[f] = \max(\mathbb{P}_{\mathcal{C},i}[f'], f' \in \mathcal{F}_{\mathcal{C},i})\}$. The set $\Theta_{\mathcal{C},i}$ is a TEXTUAL THESIS for textual unit $\mathcal{C}(i)$.

A textual thesis is thus the set of events that are the most probable according to a given textual theory. Ideally, the textual thesis should include only one event, but sometimes, like when there is significant uncertainty about the statuses of the readings, a tie can appear between two or more events. To illustrate this, let's suppose that our analysis was inconclusive and therefore the probability function assigned the probabilities as in Table 3. The thesis contains two equiprobable events, namely “ $\vec{\rho}_{\mathfrak{M}}$ alone is original” and “ $\vec{\rho}_Q$ alone is original”.

Table 3: Assigning probabilities to the events.

Event	Probability
$\vec{\rho}_{\mathfrak{M}}$ alone is original	0.4
$\vec{\rho}_Q$ alone is original	0.4
None are original	0.2

Definition 2.7 (Critical text). Let $\Theta_{\mathcal{C},i}$ be a textual thesis that has q as its target witness. Let $\mathfrak{C}_{\mathcal{C},i} = \{\vec{\rho} : (\vec{\rho}, \mathbf{o}) \in \cup \Theta_{\mathcal{C},i}\}$ if $\{\vec{\rho} : (\vec{\rho}, \mathbf{o}) \in \cup \Theta_{\mathcal{C},i}\} \neq \emptyset$ and $\mathfrak{C}_{\mathcal{C},i} = \{\vec{\mathfrak{g}}\}$ otherwise, where $\mathbf{o} \in \mathcal{O}$ is the originality status that represents originality, $\vec{\mathfrak{g}}$ is a lacuna and $\cup \Theta_{\mathcal{C},i} = \{(\vec{\rho}, y) : (\vec{\rho}, y) \in \omega, \text{ for all } \omega \in f \text{ and all } f \in \Theta_{\mathcal{C},i}\}$. Let $\mathfrak{C}_{\mathcal{C}} = \mathfrak{C}_{\mathcal{C},1} \times \mathfrak{C}_{\mathcal{C},2} \times \dots \times \mathfrak{C}_{\mathcal{C},n}$ (there are n textual units). Define a set $\aleph_q = \{\vec{s} : \vec{s} = \vec{c}_1 \vec{c}_2 \dots \vec{c}_n, \text{ for all } \vec{c}_j \in \mathfrak{c} \text{ and all } \mathfrak{c} \in \mathfrak{C}_{\mathcal{C}}\}$. A string $\vec{s} \in \aleph_q$ is called a CRITICAL TEXT of q if there exists a witness $\hat{q} = (m, \vec{s}, \lambda)$ such that \hat{q} is a reconstruction of q .

A critical text is the reconstructed string of the target witness. There may be more than one possible critical texts, depending on how many potentially original readings there are in each textual unit. In our example, given that we gave both competing

readings the same probability of being original, we have two critical texts of Psalm 145, one including the *nun* line and the other omitting it.

2.1 Retroversions

One of the most important stages of a textual criticism project is the retroversion, or back-translation, of ancient translations into the original language. This allows the textual critic to treat the retroverted witness as a witness in the original language and include its readings in textual theories. It is important to keep in mind that the retroverted witnesses of an ancient translation are a reconstruction of *its exemplar(s)* exclusively. In consequence, instead of using general statements such as that a certain translation was translated from “the Masoretic text”, it should instead be said that it was translated from *a specific witness* (or set of witnesses) that was probably a descendant of the Masoretic archetype or archetypes (and not from a Masoretic archetype itself, unless there are historical reasons to believe that).

Definition 2.8 (Retroverted witness, retroversion). Let w and q be witnesses such that w is hypothesised to be a translation of q in a given latent non-local genealogy. Let \hat{q} be a translation of w such that the language of the string of \hat{q} is the same as the language of the string of q . \hat{q} is a RETROVERTED WITNESS of w if it is a reconstruction of q . Note that w and q can be reconstructions themselves. If w is thought to have been translated from more than one exemplar, then more than one retroversion can be created from it. If \hat{q} is a retroverted witness, its reconstructed segments are called RETROVERSIONS.

3 Heuristics

In the practice of textual criticism there is no standard way of creating textual theories, but there are heuristics that are intuitively thought to be reasonably effective at creating hypothetical reconstructions. Famous sets of heuristics were created in the 18th and 19th centuries, some of which are still in use today. Two notable ones are *lectio brevior potior* (“the shorter reading is the stronger one”) and *lectio difficilior potior* (“the more difficult reading is the stronger one”), which we have illustrated diagrammatically in Figure 16. The graphs are evolutions of latent local genealogies. *Lectio brevior potior* rests upon the general assumption that scribes were more prone to add text than to remove it. *Lectio difficilior potior* assumes that autographic readings were rougher than later ones, ie, that scribes tended to smooth out the text to make it more sensible, palatable or understandable over time.

Both *lectio brevior potior* and *lectio difficilior potior* are simply loosely binding guidelines, and when applied they are qualified by many considerations. Even in their original formulations, they were not held to be absolute but rather to be used in conjunction with other maxims of common sense (Epp 2005, 127). The reconstructive power of these heuristics and others that are similar is therefore hard to surmise and,

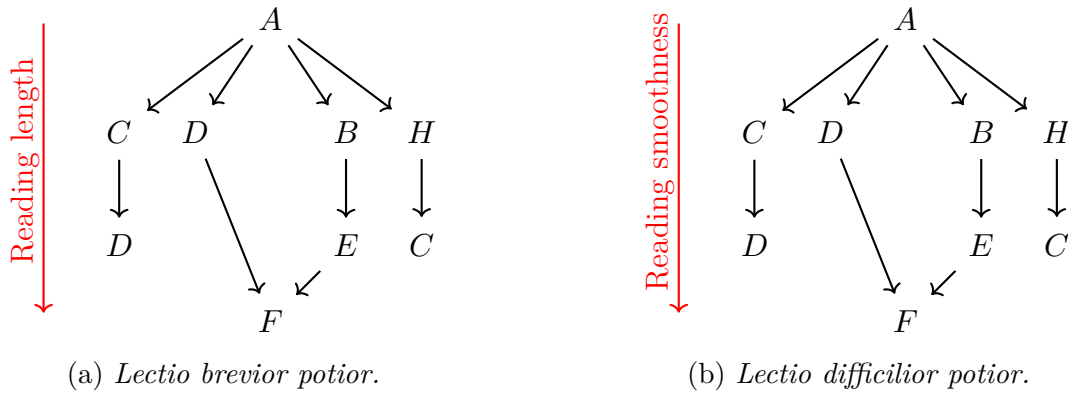


Figure 16: Illustration of *lectio brevior potior* and *lectio difficilior potior*.

given the many exceptions that are found when trying to apply them consistently, one could wonder if they are heuristics at all (see Housman 1972).

Another rule of thumb often seen especially in the textual criticism of the Hebrew Bible is the detection of harmonizations by an assessment of the possibilities that they arose by random chance. We see, for instance, that many stereotypical readings present in the Samaritan Pentateuch, as well as in the Septuagint, happen at very predictable places, like in passages that recount a previously narrated event, where there is an alleged tendency to harmonise the first and second narratives so that both have the same details as opposed to the shorter Masoretic text (see vgr Gen 30:36 in the Samaritan Pentateuch, which inserts text from Gen 31:10-13). The Samaritan Pentateuch also tends to “complete” some formulaic lists of nations, like at Gen 15:21, where it adds “and the hivites” (it does a similar thing at Exo 13:5, 23:23, 23:28). This is often taken as indication that these differences likely arose from intentional alteration rather than from accidental scribal error. In the presence of potential harmonizations like these, scholars can presuppose that the incomplete text is more original than the one that is more fulsome. This is because they consider that it was more likely that a scribe systematically updated an exemplar that he or she perceived as incomplete than the opposite scenario, when the scribe debased an exemplar perceived as too complete, especially if the alleged debasements do not show a clear pattern. For instance, consider the following sequence of letters, which we will call outline 1: *Aa Bb C D E F f Gg*. Each uppercase letter represents a semantic unit (a narrative, for instance) and each lowercase letter represents a repetition or restatement of that unit. Now consider outline 2: *Aa Bb Cc Dd Ee Ff Gg*. It can be held as more likely that a scribe took outline 1 and “completed” it to create outline 2 than the opposite. Otherwise, it would be hard to explain why the scribe debased only the repetitions *c*, *d* and *e* and not the repetitions *a*, *b*, *f* and *g*. Thus, unless there is a clear pattern in the hypothetical debasement, the opposite hypothesis that outline 2 was harmonized is preferred. The weakness of this heuristic is that the opposite hypothesis can be argued for just as easily. It could be the case that a scribe

was copying from an exemplar that contained outline 2 but, due to either accidental or intentional omissions, did not incorporate repetitions c , d and e in the copy.

These considerations aside, there are at least two heuristic approaches that have indeed been applied with some consistency over the years, and we shall occupy ourselves with them. These are Lachmannian *stemmata* and the majority rule¹³. A meta-heuristic approach, that of complex network simulations, will be proposed and discussed thereafter. The purpose of this section is not to provide an exhaustive catalog of textual criticism methods, but rather to illustrate by way of example how different theories and techniques can be formulated in a clear and flexible way by leveraging the present mathematical model’s expressiveness.

3.1 Lachmannian heuristics

The most important notion used by Lachmannian textual critics is the idea of *stemma*.

Definition 3.1 (Target tradition). Given a latent tradition T , let $T^{\text{comp}} \subset T$ be the set that contains all the composed copies in it. Define an extant tradition $T^\times \subseteq T$, its archetype set $T^{\text{arch},\times}$ and its hyparchetype set $T^{\text{hyparch},\times}$. The TARGET TRADITION of the tradition T^\times is the set $T^\times \cup T^{\text{arch},\times} \cup T^{\text{hyparch},\times} \cup T^{\text{comp}}$.

3.1.1 Local *stemmata*

Latent non-local genealogies are not accessible to textual critics, and thus a natural impulse is to attempt their reconstruction using the evidence provided by their extant witnesses. One of the ways of approaching this problem has been since at least the nineteenth century the creation of hypothetical genealogical trees of witnesses (Chiesa 2020; see also Trovato 2020). These genealogical trees are called *stemmata codicum* (Maas 1958, 20). Before we define them, we will define a similar concept, that of local *stemmata*, which provides genealogical trees between *segments*.

Definition 3.2 (Latent local ancestries of a history). Let $H = (\mathcal{S}, \mathfrak{K})$ be a latent history. Let $(\mathcal{S}, \mathcal{L})$ be its transitive closure. The set \mathcal{L} contains all the ancestry relationships between the segments. The graph $(\mathcal{S}, \mathcal{L})$ is called the LATENT LOCAL ANCESTRIES OF THE HISTORY.

Definition 3.3 (Target local *stemma*). Let $H = (\mathcal{S}, \mathfrak{K})$ be a latent textual history of a tradition T . Let $G_\sigma = (\mathcal{S}_\sigma, \mathfrak{K}_\sigma)$ be the latent local genealogy of a segment $\sigma \in \mathcal{S}$. Given an extant tradition $T^\times \subset T$ and its target tradition $T^{\text{target},\times}$, define $\mathcal{S}'_\sigma = \{(w, (x, h)) \in \mathcal{S}_\sigma : w \in T^{\text{target},\times}\}$. Let $(\mathcal{S}, \mathcal{L})$ be the latent local ancestries of the history H . Define a directed graph $(\mathcal{S}'_\sigma, \mathcal{L}'_\sigma)$ where $(\iota, \pi) \in \mathcal{L}'_\sigma$ iff $(\iota, \pi) \in \mathcal{L}$ for $\iota, \pi \in \mathcal{S}'_\sigma$. The graph $\text{tranr}((\mathcal{S}'_\sigma, \mathcal{L}'_\sigma))$ is called a TARGET LOCAL STEMMA of segment σ over the extant tradition T^\times , where tranr represents the transitive

¹³Other heuristics, not covered in the present article, have been proposed by scholars. An example is offered by Weitzman (1978) (see also Weitzman (1985) for a critique of Lachmannian heuristics).

reduction operation. Optionally, the definition of the *stemma* can be modified with an additional arbitrary set of edges that can be joined to the edges of $\text{trnr}((\mathcal{S}'_\sigma, \mathcal{L}'_\sigma))$ in order to represent relationships of interest to the textual critic that would otherwise be removed by the transitive reduction operation (for instance, to represent the genealogy of a particularly complex conflation).

Intuitively speaking, target local *stemmata* are genealogical trees that connect segments where an edge from one segment to another means that the first is an ancestor of the second. They are transitive reductions to allow them to represent the genealogical relationships in a compressed manner. They contain the extant segments as well as the relevant segments from the non-extant composed copies (like those of the first autographs) and the non-extant archetypes and hyparchetypes.

As the target local *stemma* is never known, Lachmannian textual critics create approximations, hypothetical local *stemmata*.

Definition 3.4 (Hypothetical local *stemma*). Let $H = (\mathcal{S}, \mathfrak{K})$ be a latent history for a tradition T and let $G_\sigma = (\mathcal{S}_\sigma, \mathfrak{K}_\sigma)$ be the latent local genealogy of a segment $\sigma \in \mathcal{S}$. Let $S_\sigma^\times = \{(w, (x, h)) \in \mathcal{S}_\sigma : w \in T^\times\}$ for an extant tradition $T^\times \subset T$ and let $\hat{S} = \{(w, (x, h)) : w \in \hat{T}, (w, (x, h)) \text{ is a reconstructed segment}\}$ for a set of reconstructions \hat{T} . Define a directed acyclic graph $(\mathcal{S}_\sigma \cup \hat{S}, \hat{B}_\sigma^\times)$ where \hat{B}_σ^\times is a set of ordered pairs where each pair represents an ancestry relationship over an arbitrary latent history other than H ¹⁴. The graph $\text{trnr}((S_\sigma^\times \cup \hat{S}, \hat{B}_\sigma^\times))$ is called the HYPOTHETICAL LOCAL STEMMA of the segment σ . Optionally, the definition of the *stemma* can be modified with an additional arbitrary set of edges that can be joined to the edges of $\text{trnr}((S_\sigma^\times \cup \hat{S}, \hat{B}_\sigma^\times))$ in order to represent relationships of interest to the textual critic that would otherwise be removed by the transitive reduction operation.

Some latent local genealogies may display a situation in which the segments become “atomized”, thus detaching the descendants of a segment from that segment’s ancestors, as in Figure 17. In Figure 17a, “doghouse” gave rise to “house”, and then a substring of “house”, the string “se”, gave rise to another string “so”. In the latent history, “se” is connected to “so”, but “so” cannot be traced back to “doghouse”, because “doghouse” is not an ancestor of “se” strictly speaking, but of “house”. Similarly, in Figure 17b, “housemate” descends from “house and a bat”, not from “doghouse and a bat”, which means that, in the genealogy, it cannot be traced back to “doghouse” and “bat”. Any Lachmannian technique aiming to be comprehensive should be able to handle edge cases like these. One option could be to redefine ancestry so that it includes the ancestors of segments that contain another segment as a substring (e.g. merging the genealogies of “se” and “house” in Figure 17a) or adjusting segment length on a case-by-case basis (e.g. making “housemate” derive not from “house and a bat” but from “doghouse and a bat” in Figure 17b).

¹⁴We require that the ancestries be defined over a history other than H because H is not defined for $\mathcal{S}_\sigma \cup \hat{S}$.

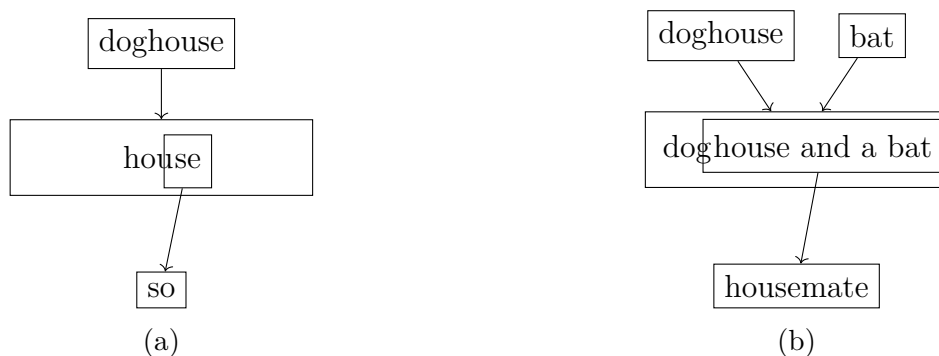


Figure 17: Two edge cases.

3.1.2 Variant *stemmata*

Definition 3.5 (Target variant *stemma*). A TARGET VARIANT *STEMMA* is an evolution of a target local *stemma*.

Definition 3.6 (Hypothetical variant *stemma*). A HYPOTHETICAL VARIANT *STEMMA* is an evolution of a hypothetical local *stemma*.

Variant *stemmata* are meant to be approximations of evolutions of latent local genealogies. Hypothetical variant *stemmata* show the readings found in an extant tradition (potentially with some additional hypothetical readings) and the hypothesized evolutive relationships between them, where each edge represents a “shift event”, i.e., a change from a reading to a different one (Hoenen 2020, 233–34).

Variant *stemmata* are often implicit in discussions where textual critics compare competing readings. When a reading is sensible and another one is nonsensical, the nonsensical reading is frequently hypothesized to be secondary. An example is Habakkuk 2:5, where the Masoretic Text has “and yet, because *the wine* is treacherous, the strong man is arrogant and does not last” (likely nonsensical) and the Qumran scroll 1QpHab has “and surely *wealth* has betrayed the arrogant man, who will not last”. The key difference between both readings is the word *hyyn* (“the wine” in the Masoretic text) or *hwn* (“wealth” in 1QpHab). The variant *stemma* of this particular textual unit could be “wealth” → “wine”, given that “wealth” has much better semantic fit than “wine”. This variant *stemma*, however, is not inevitable, and other *stemmata* could be proposed. The variant *stemma* $?^{orig} \rightarrow ?^{early} \rightarrow$ “wealth” → “wine” in particular could represent the hypothesis of a sensible autographic reading that evolved into an early lost nonsensical reading. That nonsensical reading could have been replaced with the sensible (but non-original) reading “wealth” by a well-meaning scribe, and that reading could have been in turn replaced with the nonsensical reading “wine” by a later scribe by accident. One could go further and come up with another variant *stemma* that makes “wealth” later than “wine”, like $?^{orig} \rightarrow$ “wine” → “wealth”. According to this hypothesis, “wine” is the early nonsensical reading and “wealth” is a pious correction. In absence of meaningful evidence

about the topology of the *latent local genealogy*, all hypothetical variant *stemmata* are speculative, and all that can be said is that the nonsensical reading is probably not original (if one assumes that authors normally tried to make sense). The alternative sensible reading may have to be made part of the critical text *faute de mieux*.

The choice between two readings, one nonsensical and one sensible, may not seem too difficult, however. When more readings, including multiple sensible ones, are part of a local genealogy, the number of possible hypothetical variant *stemmata* starts to grow, along with the difficulty of making a choice. An example is Habakkuk 2:18, where there are three competing sensible readings in the extant tradition¹⁵. The Masoretic text has “a teacher of lies” (in reference to an idol). The Septuagint and the early Greek scroll 8 $\dot{\text{H}}\text{evXII}$ gr say “deceitful appearance”. Finally, 1QpHab has “rebellion of falsehood”¹⁶. Three of the many possible hypothetical variant *stemmata* that a textual critic could formulate are shown in Figure 18.

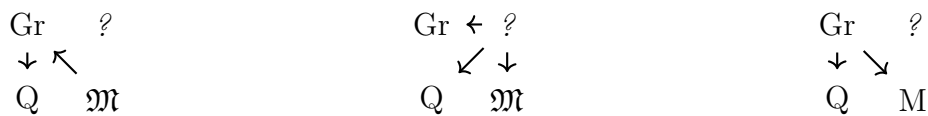


Figure 18: Three possible hypothetical variant *stemmata* for Habakkuk 2:18. \mathfrak{M} : Masoretic reading; Gr: Septuagint and 8 $\dot{\text{H}}\text{evXII}$ reading; Q: 1QpHab reading; ?: Lost reading.

Remark 3.1 (Ideal characteristics of hypothetical local *stemmata*). The purpose behind creating hypothetical local *stemmata* is recuperating the target local *stemmata*. Therefore, the ideal hypothetical local *stemma* must fulfill at least these two characteristics: i) it is isomorphic to the target local *stemma* that it aims to recover, and ii) its hypothetical variant *stemma* mirrors the target variant *stemma*, meaning that it is isomorphic to it and preserves the positions of the labels. If the hypothetical local *stemma* is not isomorphic to its target counterpart and is therefore not ideal, then the textual critic should strive to at least fulfill the second condition. Of course the target local *stemma* is never known, so the two conditions are typically asserted only by assumption. It is important to highlight the presumptive nature of hypothet-

¹⁵Another example in Habakkuk is verse 1:17, where the two competing readings make sense. The reading favored by the Masoretic text and the Septuagint is “shall (the king of Babylon) therefore empty *his net* (...)?”, while the reading favored by 1QpHab and the early Greek scroll 8 $\dot{\text{H}}\text{evXII}$ gr is “unsheathe *his sword*”. The difference is in the word *chrmw* (“his net”) or *chrww* (“his sword”). The context is a prayer in which the prophet decries the cruelty and pride of the Babylonians, who “fish people with their nets” (1:15) and then “offer sacrifice to their net” (1:16), in reference to idolatry. The text of verse 1:17 says “shall he therefore empty his net / unsheathe his sword, and not spare to slay the nations continually?”, with the verb “to slay” being semantically closer to “sword” than to “net” elsewhere in the Hebrew Bible. There are many textual units in the Hebrew Bible where several of the competing readings make sense.

¹⁶Martínez (1994, 202) translates it as “sham oracle”. The Hebrew text says *wmry shqr*, that can be translated as “and rebellion of falsehood”.

ical *stemmata*, especially because the presumption of isomorphy is quite strong and potentially difficult to justify in some cases.

3.1.3 The *stemma codicum*

Definition 3.7 (Latent non-local ancestries of the history). Let $\mathcal{G} = (T, D)$ be a latent non-local genealogy created from a history H . Let (T, \mathfrak{L}) be the transitive closure of \mathcal{G} . The set \mathfrak{L} contains all the ancestry relationships between the witnesses. The graph (T, \mathfrak{L}) is called the LATENT NON-LOCAL ANCESTRIES of history H .

Definition 3.8 (Target *stemma codicum*). Let $\mathcal{G} = (T, D)$ be a latent non-local genealogy and let (T, \mathfrak{L}) be its associated latent non-local ancestries. Let $T^\times \subseteq T$ be an extant tradition and $T^{\text{target}, \times}$ its target tradition. Define a set $\mathfrak{L}' = \{(w_i, w_j) : w_i \prec w_j; w_i, w_j \in T^{\text{target}, \times}\}$, where \prec is defined over \mathcal{G} . In other words, \mathfrak{L}' contains all the correct ancestry relationships between the extant witnesses, their archetypes, their hyparchetypes and the composed copies. A graph $\text{tranr}((T^{\text{target}, \times}, \mathfrak{L}'))$ is called a TARGET *STEMMA CODICUM* of the tradition T^\times . Optionally, the definition of the *stemma* can be modified with an additional arbitrary set of edges that can be joined to the edges of $\text{tranr}((T^{\text{target}, \times}, \mathfrak{L}'))$ in order to represent relationships of interest to the textual critic that would otherwise be removed by the transitive reduction operation.

A target *stemma codicum* is sometimes referred to in the literature as a “true *stemma*” (see Hoenen 2020, 232). An example target *stemma codicum* for a simulated single-first-autograph tradition is in Figure 19. In the figure, the *stemma* comprises the magenta, blue, yellow and red vertices, and the green arrows. The magenta vertices are the extant tradition, the red vertex is the (non-extant) first autograph, the blue vertices are the (non-extant) archetypes of the extant tradition, the yellow vertices are the hyparchetypes, and the green arrows are the edges of the *stemma codicum* (the vertices of the *stemma* are shown in a larger size than the other vertices of the latent genealogy for contrast). Note that the green edges do not represent exemplar-copy relationships (these are the gray edges), but rather ancestor-descendant relationships (i.e., any two vertices connected by a green edge could be mediated by many intermediate copies).

Textual critics often want to reconstruct the ancestry relationships between the extant witnesses, that is, they want to reconstruct a target *stemma codicum*. Typically, however, target *stemmata codicum* are impossible to recover with certainty, and thus textual critics create approximations. It is in this hypothetical sense that the term *stemma codicum* is most often used. A textbook example of a *stemma codicum* is shown in Figure 20. It places a hypothetical witness x at the top and adds a set of hypothetical witnesses denoted by Greek letters, which are presumed ancestors of the extant witnesses (shown in capital Latin letters). The actual ancestor-descendant relationships between the extant witnesses are unknown and therefore the image depicts only hypothetical ones. The hypothetical witness that is imagined to be the

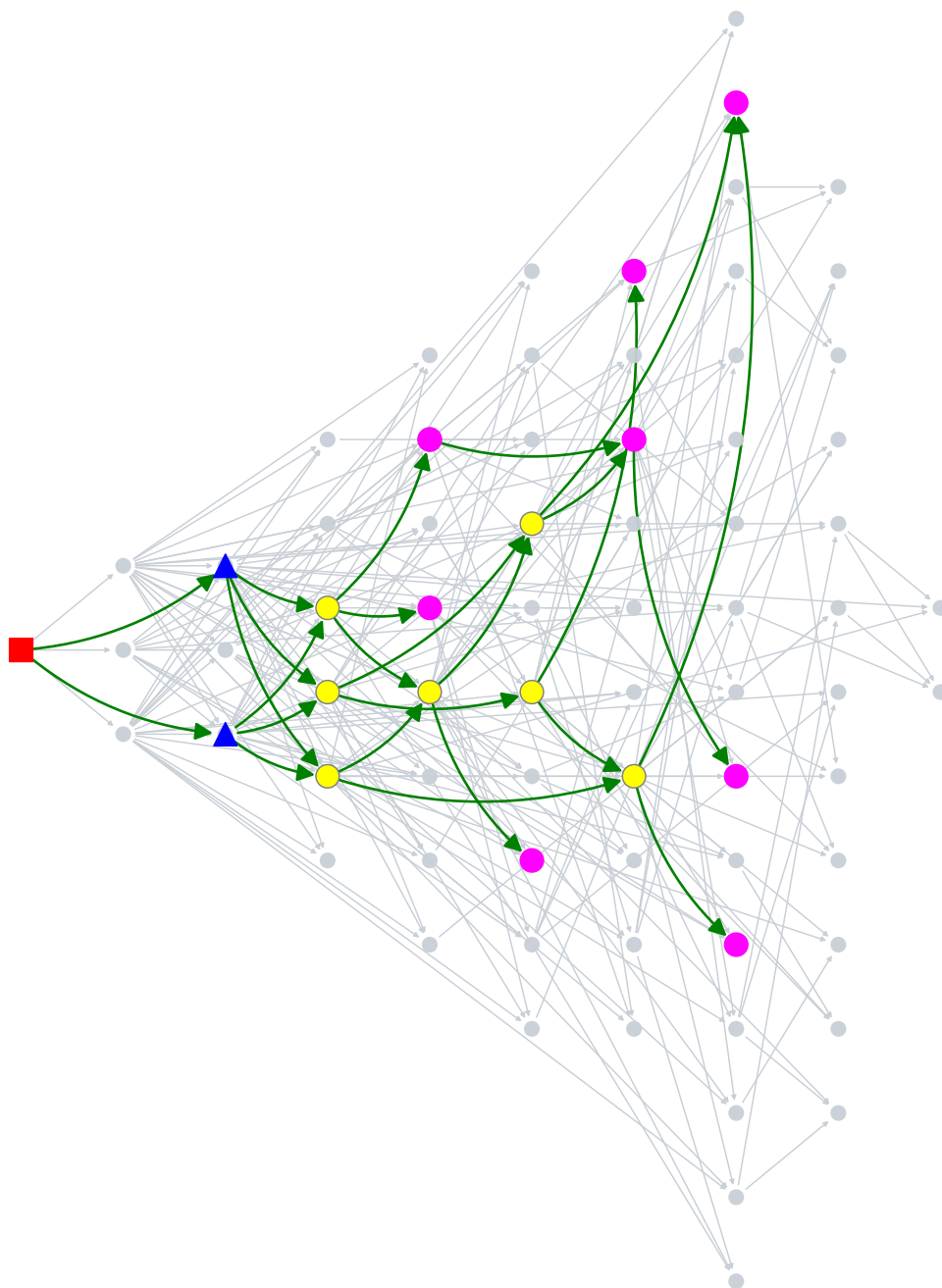


Figure 19: A latent genealogy with a superimposed target *stemma codicum*. The red square is the first autograph, the blue triangles are the archetypes, the yellow vertices are the hyparchetypes and the magenta vertices are the extant tradition.

archetype of the extant tradition is placed below x (but separated from it by a ticked line in order to show that there are possibly many intermediate ancestors).

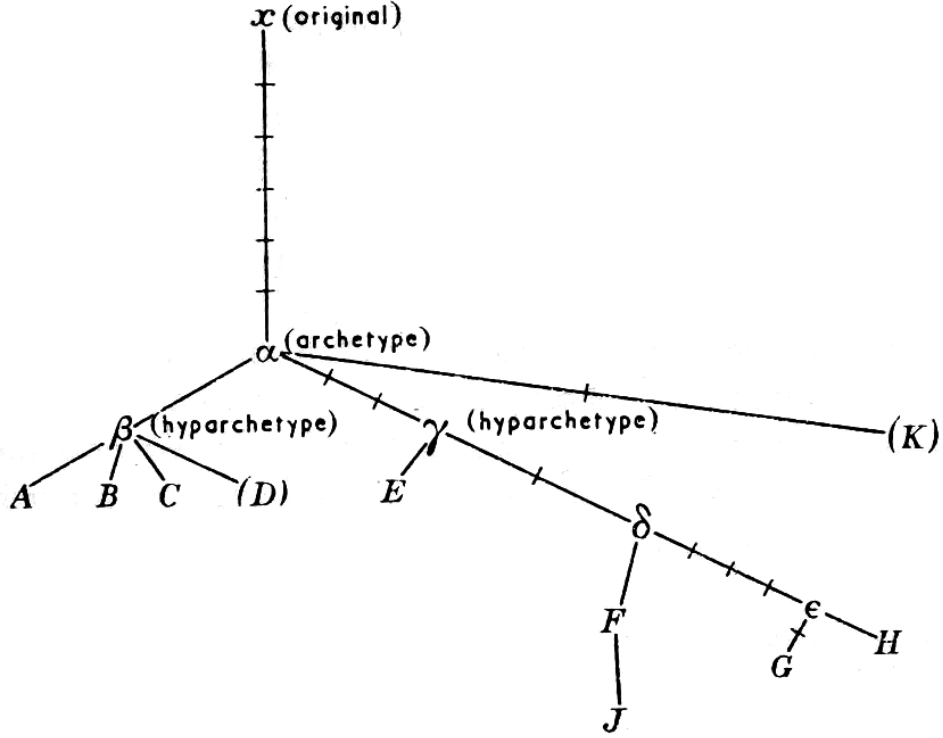


Figure 20: A hypothetical *stemma codicum*. Taken from Maas (1958, 5).

Definition 3.9 (Hypothetical *stemma codicum*). Let $T^\times \subseteq T$ be an extant tradition and let \hat{T} be a set of reconstructions. Let $\mathcal{G} = (T, D)$ be the latent non-local genealogy of T . Define a directed graph $(T \cup \hat{T}, \hat{A}^\times)$ where \hat{A}^\times is a set of ordered pairs where each pair represents an ancestry relationship over an arbitrary latent non-local genealogy other than \mathcal{G} ¹⁷. A graph $\text{transr}((T^\times \cup \hat{T}, \hat{A}^\times))$ is called a **HYPOTHETICAL STEMMA CODICUM** of the extant tradition T^\times . Optionally, the definition of the *stemma* can be modified with an additional arbitrary set of edges that can be joined to the edges of $\text{transr}((T^\times \cup \hat{T}, \hat{A}^\times))$ in order to represent relationships of interest to the textual critic that would otherwise be removed by the transitive reduction operation.

Remark 3.2 (Ideal characteristics of hypothetical *stemmata codicum*). Similar to the case of hypothetical local *stemmata* (Remark 3.1), the ideal hypothetical *stemma codicum* must be isomorphic to its corresponding target *stemma codicum*. On top of this requirement, for any two witnesses $(m_i, \vec{s}, \lambda_i), (m_j, \vec{u}, \lambda_j)$ that are matched in the isomorphism, it must hold that $\vec{s} = \vec{u}$. In other words, the hypothetical *stemma* must have both the same structure and the same textual content as the target *stemma*.

¹⁷We require that the ancestries be defined over a genealogy other than \mathcal{G} because \mathcal{G} is not defined for $T \cup \hat{T}$.

In a hypothetical *stemmata codicum*, many assumptions need to be made on the part of the textual critic. In Figure 20, for instance, there are six hypothetical reconstructions and all the edges are also hypothetical. Furthermore, the textual critic assumed that the latent genealogy had only one first autograph (x) and that the extant tradition derives from a single archetype (α), which does not have to obligatorily be the case, as an extant tradition may also have zero or more than one archetypes, as we saw earlier.

Definition 3.10 (Lachmannian archetype). A Lachmannian archetype is a most recent common ancestor of an extant tradition over a hypothetical *stemma codicum*.

Definition 3.11 (Lachmannian hyparchetype). Let $LACH = \text{tranr}((T^\times \cup \hat{T}, \hat{A}^\times))$ be a hypothetical *stemma codicum* and let $T^{\text{Larch}, \times}$ be the set of the Lachmannian archetypes of the extant tradition T^\times over the *stemma*. Define $\mathcal{V}_{T^\times} = \{V \in \wp(T^\times) : |V| > 1\}$, where $\wp(T^\times)$ is the power set of T^\times . Define $T^{\text{Lhyparch}, \times} = \{h \in T^\times \cup \hat{T} : h \text{ is a Lachmannian archetype of } V \text{ over } LACH \text{ for } V \in \mathcal{V}_{T^\times} \text{ and there exists at least one } a \in T^{\text{Larch}, \times} \text{ such that } a \hat{\succ} h\}$, where $a \hat{\succ} h$ means that there is a directed path from a to h in $LACH$. The elements of the set $T^{\text{Lhyparch}, \times}$ are called the LACHMANNIAN HYPARCHETYPES of the extant tradition T^\times over the hypothetical *stemma codicum* $LACH$.

In other words, while the “true” archetypes and hyparchetypes are defined on non-local genealogies, their Lachmannian counterparts, which are hypothetical, are defined over hypothetical *stemmata codicum*. True archetypes were illustrated in Figure 13 in non-local genealogies, while an example Lachmannian archetype is the hypothetical witness α in Figure 20. Ideally, the Lachmannian archetypes and hyparchetypes should correspond to their real-world counterparts (ie, their strings should match the strings of the actual ones, and their ancestry relationships should mirror those present in the latent non-local genealogy \mathcal{G} , by Remark 3.2).

Remark 3.3 (Hypothetical local *stemmata* and hypothetical *stemmata codicum* need to be consistent with one another). We know from Definition 1.31 that a witness w is a copy of e if there exists at least one segment from e that is a parent of a segment of w in the latent history. In consequence, and from Definition 3.3 and Definition 3.8, we have that if $(w, (x, h)) \prec (v, (g, z))$ (in the target local *stemma*), then $w \prec v$ (in the target *stemma codicum*)¹⁸. Given that hypothetical *stemmata* purport to reconstruct the target *stemmata*, the existence of at least one $(w, (x, h)) \hat{\succ} (v, (g, z))$ should imply $w \hat{\succ} v$, where $\hat{\succ}$ represents the “is an ancestor of” relationship over a hypothetical *stemma*. Therefore, a hypothetical *stemma codicum* and a hypothetical local *stemma* that violate this condition cannot be consistent with each other.

¹⁸Notice that $(w, (x, h)) \prec (v, (g, z))$ implies $w \prec v$ but not vice versa. Consider the paths $(w, (x, h)) \rightarrow (b, (f, j))$ and $(b, (k, l)) \rightarrow (v, (g, z))$ (the segments are not overlapping). These two paths imply that $w \prec b$ and $b \prec v$, which means that $w \prec v$. However, there is no path such that $(w, (x, h)) \prec (v, (g, z))$.

3.1.4 Internal consistency of Lachmannian models

We have proposed a way of understanding *stemmata* that sees them as approximations of latent genealogies. That, is, hypothetical *stemmata codicum* are approximations of target *stemmata codicum*, which are approximations of latent non-local genealogies. Similarly, hypothetical local *stemmata* aim to reconstruct the target local *stemmata* that, in turn, aim to reconstruct the latent local genealogies. The same logic applies to hypothetical variant *stemmata*, target variant *stemmata* and evolutions of latent local genealogies.

This implies that the freedom a textual critic has when imagining hypothetical variant *stemmata* is limited by her other decisions regarding the hypothetical *stemma codicum* and the hypothetical local *stemmata*. That is because all the local *stemmata* need to be perfectly consistent with the *stemma codicum* (Remark 3.3) and any variant *stemmata* must of course be consistent with the local *stemmata* from which they were derived. For example, if a local *stemma* contains a conjecture, then the *stemma codicum* must contain a witness that contains that conjecture. This is because no segment can exist in a history if it is not part of a witness.

Local *stemmata* can affect other local *stemmata* too in an indirect fashion. If, for instance, due to having too many conjectural segments in too many local *stemmata*, the *stemma codicum* is becoming bloated with implausible hypothetical witnesses, then in order to include an additional conjectural reading in a given local *stemma*, the textual critic may decide to give up another conjectural reading in another local *stemma*.

Since variant *stemmata* are just reflections of local *stemmata*, any changes to them have to be supported by changes in the topology of the associated local *stemmata*, with the potential upstream impact on the *stemma codicum* and on the other local *stemmata*. The three kinds of *stemmata* (*codicum*, local, and variant, both when they are target *stemmata* or hypothetical *stemmata*) are aspects of the same conceptual mechanism and cannot be incoherent with one another. They are inherently interdependent and changes in one affect the others. In consequence, a textual critic creating a Lachmannian model of a real-world tradition needs to be mindful of preserving the internal consistency of her *stemmata*.

In a similar vein, any *stemma*-creating techniques derived from genetics research should be expressed in a mathematical framework specific to textual criticism, after which they should be tested both algebraically and by means of numerical simulations in order to determine the extent to which they can work well on manuscript datasets¹⁹. Phylogenetic methods have been embraced by many in the field thanks to the perceived qualitative similarities between scribal error and gene inheritance and mutation (Manafzadeh and Staedle 2020; Macé and Baret 2006; Howe et al. 2001; Howe et al. 2004). The direct application of phylogenetic methods to witness tradi-

¹⁹Spencer et al. (2003) provides an example of a statistical assessment of the advantages and shortcomings of phylogenetic techniques for the tradition of the Miller's Tale in Chaucer's *Canterbury Tales*.

tions has not been unambiguously established from a theoretical standpoint, however, and part of the reason is that a comprehensive mathematical comparison between genetics and textual transmission, to the best of our knowledge, has not been carried out to date.

3.2 Majority rules

Majority rules are another important heuristic used in textual criticism, especially in the field of Greek New Testament studies, but with potential applications in many other traditions. Majority rules are posited as effective for recovering original readings in genealogies that show a certain “normality” of transmission, ie, traditions in which a reasonable independence between the witnesses can be assumed or, in other words, traditions in which the extant witnesses can be treated as descendants of distinct independent branches all originating in the autograph that is being reconstructed (Robinson 2005). Rather than trying to reconstruct variant *stemmata*, majority rule practitioners forego most genealogical preoccupations (for considering them unnecessary or unsolvable) and make use of frequency counts to directly arrive at textual theories instead.

Consider the latent local genealogy of Figure 21, where an autographic segment σ_0 is the direct ancestor of three different segments σ_1 , σ_2 and σ_3 . The probabilities p_1 , p_2 and p_3 represent how likely it is that the reading of the child segment is different from the reading of the parent segment. Thus, for instance, $p_1 = 0.05$ means that the probability of σ_1 having a reading different from that of σ_0 is 5%. We can alternatively call these probabilities “propensities to err”, as they would represent the propensity of a scribe to commit scribal error when copying the segment.

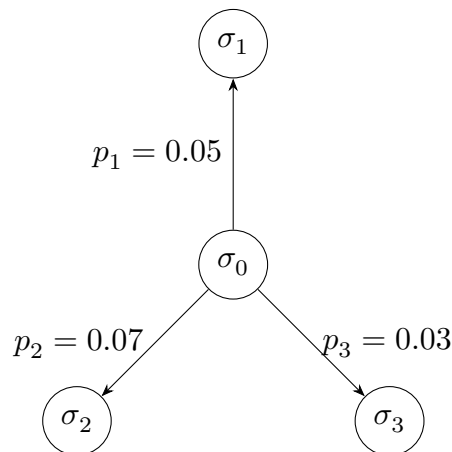


Figure 21: A latent local genealogy with radiating segments.

Assuming that the autographic segment is not extant, let’s define a majority rule that selects a reading if it is present in at least 2 of the 3 descendant segments. We can conceptualize the genealogy as a set of three experiments, each with two outcomes

(either preserves the original reading or not), which means that there are $2^3 = 8$ outcomes, as shown in Table 4 (the probabilities of the fourth column are calculated assuming independence). The event in which at least two outcomes are original has a probability of 99.3%. This means that for this particular local genealogy, a 2:3-majority rule would have almost perfect effectiveness at recovering the original reading.

Table 4: The outcomes of the local genealogy

σ_1	σ_2	σ_3	Probability of outcome
Original	Original	Original	$(1 - p_1) \times (1 - p_2) \times (1 - p_3) = 85.67\%$
Original	Original	Error	$(1 - p_1) \times (1 - p_2) \times p_3 = 2.65\%$
Original	Error	Original	$(1 - p_1) \times p_2 \times (1 - p_3) = 6.45\%$
Original	Error	Error	$(1 - p_1) \times p_2 \times p_3 = 0.19\%$
Error	Original	Original	$p_1 \times (1 - p_2) \times (1 - p_3) = 4.51\%$
Error	Original	Error	$p_1 \times (1 - p_2) \times p_3 = 0.14\%$
Error	Error	Original	$p_1 \times p_2 \times (1 - p_3) = 0.34\%$
Error	Error	Error	$p_1 \times p_2 \times p_3 = 0.01\%$

Adding new direct copies of σ_0 can only improve the effectiveness of the majority rule. If we assume for simplicity that all the propensities are equal to a constant β , we can derive a formula to calculate the probability of obtaining exactly k original readings in n copies (Equation 3) and then a formula to calculate the probability of half plus one copies having the original reading (the minimum number of required copies would in reality be $b = \frac{n}{2} + 1$ for even n and $b = \frac{n+1}{2}$ for odd n), as in Equation 4. If we plug in $n = 10$ and $\beta = 0.05$, for example, we get an effectiveness of 99.99% already.

$$\mathbb{P}[k] = \binom{n}{k} (1 - \beta)^k \beta^{n-k} \quad (3)$$

$$\mathbb{P}[k \geq b] = \sum_{k=b}^n \binom{n}{k} (1-\beta)^k \beta^{n-k} \quad (4)$$

Of course, real-world local genealogies do not look like Figure 21, as extant copies are typically several generations removed from the autograph (but remember that generation numbers may not be unique, see Remark 1.3). One step above in complexity could separate the extant segments from the autographic segments by a constant number of generations, like in Figure 22, where the extant segments are σ_4 , σ_5 and σ_6 .

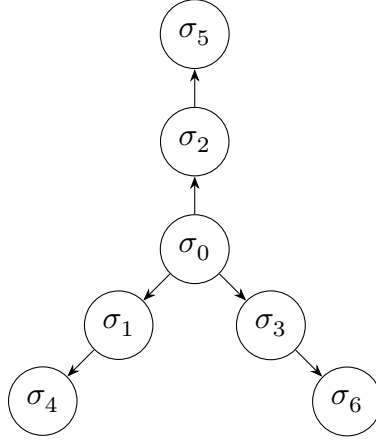


Figure 22: A local genealogy in which the extant segments are not direct descendants of the autographic segment.

A formula can be created for a situation like this, assuming again a constant propensity β . The probability that the outermost segment of each branch agrees with the autographic segment is equal to $(1-\beta)^j$ where j is the number of generations between the segments. This is because while the propensities to err (β) are all independent, the probability of agreement with the autograph is not. It depends on the probabilities of agreement of the ancestors. Thus, while the probability of agreement of a second generation is equal to $(1-\beta)$, the probability that the third generation segment agrees with the autographic segment depends on the probability of agreement of its parent, i.e., it is equal to $(1-\beta)(1-\beta)$. Given that all the branches of the local genealogy have the same number of generations, all the outer copies share the same probability of agreement with the autographic segment, namely $(1-\beta)^j$. Let's label the probability of disagreement as $q = 1 - (1-\beta)^j$. We can then adapt Equation 4 to incorporate this information by replacing β with q and obtain Equation 5. In this new equation, h means the number of outer copies. Here, for $\beta = 0.05$ as before (which implies $q = 0.0975$ for $j = 2$), and for $h = 3$, the majority rule has an effectiveness of 97.33%. If we set $h = 10$, the effectiveness is 99.85%. Compared to the simpler topology, this new topology reduces the effectiveness of the majority rule, but not so much so as to

render it invalid, especially if h is large. This is particularly relevant for the Gospels, where there are often hundreds of witnesses supporting the majority reading against one or two that attest to minority readings.

$$\mathbb{P}[k \geq b] = \sum_{k=b}^h \binom{h}{k} (1-q)^k q^{h-k} \quad (5)$$

Proponents of the majority rule rarely assume that latent local genealogies look like the previous examples. In fact, the most commonly held stereotypical latent genealogy is slightly more complex, potentially displaying fractal properties. It was proposed by Zane Hodges (1968) and is sometimes referred to as “Hodges’ truism”:

The manuscript tradition of an ancient book will, under any but the most exceptional conditions, multiply in a reasonably regular fashion with the result that the copies nearest the autograph will normally have the largest number of descendants.

We can create a graphical version of the truism on a latent local genealogy, as in Figure 23. In the illustration, the autographic segment (the red vertex) was copied four times. The genealogy has five generations, with the first generation (the autographic segment) having 60 descendants, as shown in the figure. Each subsequent generation is connected by a green “orbit” and the numbers on the orbits indicate how many descendants each of the segments in the orbit has. So, for example, each one of the second-generation segments has itself 14 descendants. As can be seen, the further a segment is from the autographic segment, the fewer descendants it has. Following Hodges’ reasoning, an erroneous reading first introduced, say, in the second generation, would be copied 14 times, thus appearing in only $\frac{14}{60} = 23.3\%$ of the copies. That is to say that the majority reading (with 76.7% of the evidence in its favor) would still reflect the original. If the error appeared in the *third* generation instead, the number of erroneous descendants would get down to 6, that is, to 10% of all the descendants, making the majority rule more efficacious. The further down the line a reading appears in the genealogy, the weaker the numerical support in its favour. The argument then goes that the odds are against erroneous readings becoming the overwhelming majority. Thus Hodges’ truism implies that the number of segments that support a given reading is, *ceteris paribus*, correlated with its genealogical antiquity and that majority readings should be given default originality status until there is forceful evidence to the contrary.

3.3 Monte Carlo simulations

Both Lachmannian *stemmata* and majority rules have strengths and weaknesses and must therefore be used judiciously. As for the first, beyond being internally consistent (as we discussed in Section 3.1.4), any Lachmannian model needs to also be *externally* consistent with the data available about the textual tradition. This implies

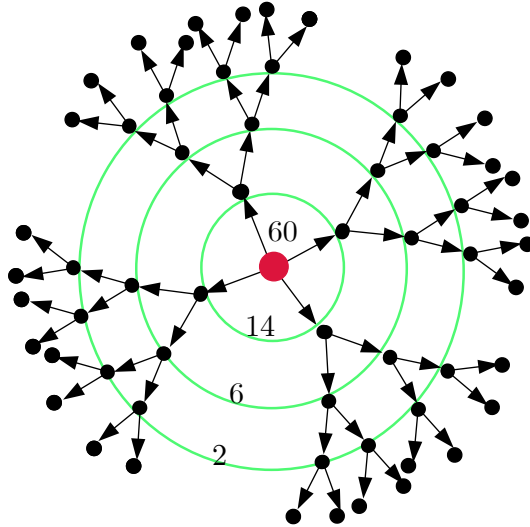


Figure 23: An instance of Hodges' truism.

that, when applying Lachmannian methods to textual traditions, the appropriateness of the methods must be *justified* rather than assumed. It has been admitted that *stemmata* created from purely textual data cannot be reasonably expected to “give exact representations of a known or suspected historical reality” (Bordalejo 2015, 5), a caveat that calls for the incorporation of extra-textual information (like historically-informed assumptions about the topologies of the latent genealogies) and general caution when interpreting the outputs produced by phylogenetic software (Spencer et al. 2003; Guillaumin 2020; see also Macé 2015, 356). Even in evidence-rich traditions like the Greek New Testament, the idea of relying on Lachmannian heuristics for reconstructive purposes has been called into question due to the extant evidence being potentially insufficient (Robinson 2005 esp. p. 556-557). Whenever possible, therefore, justifications should be proffered to demonstrate that the Lachmannian heuristics can reasonably be expected to reconstruct the autograph(s) of a given tradition to a reasonable degree, especially in comparison to naïve baseline models like majority rules or random selection of readings, in a philological parallel to the benchmark analyses that are common in computer science for the comparison of machine learning algorithms.

On the other hand, real-world local and non-local genealogies do not resemble self-similar structures perfectly and there are many ways in which, due to the vagaries of history, genealogies could deviate from the normality assumption. While small deviations would not invalidate majority rules immediately, large disturbances or an accrual of smaller ones could indeed²⁰. Figure 24 shows an example local genealogy

²⁰See the seminal 1978 debate between Zane Hodges and Gordon Fee: Fee (1978b); Hodges (1978a);

where the majority rule fails, with the red vertex being the autographic segment, the black vertices representing segments attesting to the original reading and white vertices being the segments that contain erroneous readings. If we assume that all segments but the autographic one are extant, a majority rule would declare the erroneous reading as the original one. Thus, when it comes to using majority rules for text critical purposes, having plausible assumptions about the topology of the latent local and non-local genealogies is paramount.

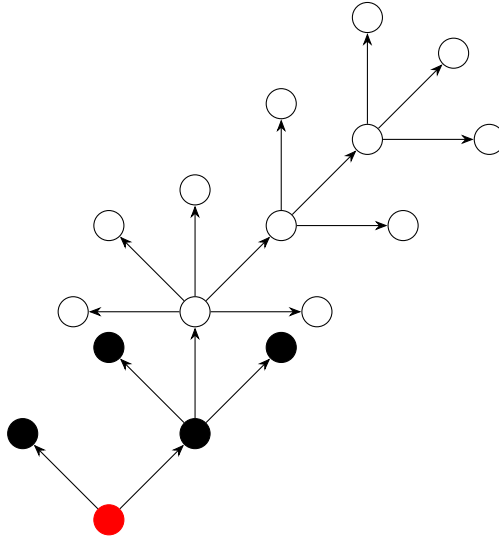


Figure 24: An example of a local genealogy where the majority rule fails.

One possible way to help researchers choose if a certain heuristic is appropriate for a given tradition or not is by leveraging computer simulations²¹, especially in traditions known to have had many witnesses. A researcher could create a simulated textual history that tries to mimic its real-world counterpart in its general features (as far as the available data about the latter permit it) and use it to compare an array of heuristics (Lachmannian, subjective mixture of early witnesses, diplomatic editions, etc) against several majority rules. After one method has been shown to be the most promising for a particular tradition, a textual critic could rely on it to analyze the data and propose critical texts. While artificial traditions have a history of being used in textual criticism (see Andrews and Macé 2013, 511–16 for an example), we propose that complex artificial *genealogies* can also be modeled computationally according to theoretical presuppositions in order to serve as digital twins of real-life ones.

One advantage of using computer simulations is their ability to easily accommodate genealogies with arbitrary topologies. Non-trivial genealogies can become alge-

Fee (1978a); Hodges (1978b).

²¹These simulations can be framed within the conceptual apparatus of complex network theory. Since complex networks are based on graph theory (Latora, Nicosia, and Russo 2017; Estrada 2011) and genealogies and histories as we have defined them are directed graphs, the model we have presented is already fully compatible with a complex networks approach.

braically intractable very rapidly, not only because of the greater number of segments they must contain, but also because of the presence of contamination and conflation, which may greatly complicate the calculations. If idiosyncratic propensities to err are included (that is, if we do not assume a constant β but rather give each scribe or copying event its own implied propensity to err), the complexity soars beyond what is computationally feasible. And this is without taking into account the need to model geographical features like the distances between scribal centers or disruptive events like wars or the destruction of libraries (like the fire of the Library of Alexandria). For models that cannot be solved using algebraic methods, Monte Carlo simulations can provide a convenient way of estimating of the probabilities of agreement of the competing simulated critical texts with the simulated autographs.

A second advantage of this approach is its reliance on first principles, which forces the researcher to lay out his or her assumptions explicitly, therefore reducing the tendency to create theories with the purpose to defend a favorite desired result. At a general level, the simulation can be implemented as a discrete-time model in which a set of suitable copying events K^* is sampled at each tick given a probability distribution $\Pi(K^*)$. Many real-life features can be incorporated in the distribution $\Pi(K^*)$. It can be designed in such a way that, for instance, particular scribes in particular geographical locations are more or less accurate than their peers, or it can be made to incorporate preferential attachment by making certain witnesses produce more direct offspring if they contain fewer scribal errors (this would model the real-life phenomenon of scribes preferring early exemplars created by reputed scribes over exemplars made by apprentices). The probabilities can be also adjusted in order to reflect the presuppositions of different reconstructive schools.

Simulations can be set up so that each run creates an artificial latent history. Reconstructive methods can then be automatically tested on that history, and the resulting critical texts can be compared to the (known) autograph in order to determine the success of each method. This can be done over many thousands or millions of Monte Carlo runs in order to reduce noise in the results and produce narrow confidence intervals. The final verdict would indicate which reconstructive method is better suited for the latent history. The artificial history need not perfectly match the real one in all the details or in its way of dealing with specific readings, because what is more important is its topology rather than the actual textual content of the segments. What is crucial is that the artificial latent histories provide a good enough mirror image of the topology of the unknown real latent history.

As was mentioned above, simulations could work better for traditions with many witnesses and highly complex topologies, where there is enough evidence to be fed into the simulation about the textual transmission of a work. Genealogies with narrower evidence bases could still benefit from being modeled this way, but the insights provided by the simulations would inevitably be more limited. The Greek New Testament, famed for the large size of its extant tradition, is a good example of a tradition where simulations have great potential. Most scholars today adhere to the view that majority rules are not appropriate for recovering the texts of the books of the New

Testament and, in consequence, prefer the readings that have the earliest attestation, typically those present in Codex Sinaiticus and Codex Vaticanus (both from the fourth century) and in a small group of witnesses taken from ancient Egyptian papyri. The opposite position, that majority rules can be used for textual theories, is held by a minority of scholars²². Despite the apparent scholarly consensus favoring the first view, the reality is that the controversy has not been satisfactorily settled to date. The issue has been called a “deadlocked debate” (Boogert 2015) and is arguably one of the greatest methodological conundrums of the discipline (Holmes 1983, 14).

The method currently in use for the production of the *Editio Critica Maior* (ECM), the standard academic edition of the Greek New Testament, follows a neo-Lachmannian approach called the Coherence-Based Genealogical Method, or CBGM (Pessoa 2022). Instead of aiming to create hypothetical archetypes, the CBGM’s goal is to reconstruct a hypothetical witness, the “initial text” (*Ausgangstext*)²³, which is imagined as an ancestor of an archetype. In order to do so, it makes use of qualitative analyses, supported by descriptive statistics of full-text collations of early witnesses, in order to construct hypothetical variant *stemma* (see Wasserman and Gurry 2017 for a detailed explanation of the method). Most of the extant witnesses are not included in the collations, as only those that are not “Byzantine” (roughly those that are present in manuscripts earlier than the sixth century) are considered valuable. Since the bulk of the tradition is made up of, precisely, later “Byzantine” witnesses²⁴, the *Editio Critica Maior* is effectively a minority text. Although some “Byzantine” readings have occasionally been adopted, the *Editio*’s critical text overwhelmingly favors Sinaiticus and Vaticanus. On the other hand, the majority-text position has also produced critical texts, with the most salient being the Antoniades edition (Kleovulos of Sardis, Christodoulou of Stavroupoli, and Antoniades 1904; see also Karavidopoulos 2012), the Hodges-Farstad Majority text (Farstad and Hodges 1985) and the more recent Robinson-Pierpont Byzantine textform (Robinson and Pierpont 2018). All these propose a combination of majority rules with subjective judgement (subjective decisions being normally confined to breaking ties) and, in the case of the Hodges-Farstad text, also with Lachmannian techniques.

Apart from the obvious difference in reconstructive approaches, both schools differ in a way that, despite being crucial, is not often acknowledged, namely the presumed topologies of the underlying latent genealogies. The ECM presupposes a latent tradition that started from the autographs in a highly uncontrolled way, giving birth to a

²²Some excellent descriptions of the *status questionis* are Govier (1996); Wisselink (1989), ch. 1; Boogert (2015); Lanier (2018); and Andersen (2014). Other informative works are Robinson (2005); Pickering (2014), esp. chapters 1-4; Zeolla (2012) chapters 6-11; and the online resource by Snapp (2010).

²³There is unfortunately significant ambiguity around what the concept of *Ausgangstext* actually means. This prevents us from offering a mathematical definition of it in the present work. As a matter of fact, the CBGM and its techniques, and the notion of the *Ausgangstext* itself, have yet to be fully understood and embraced by scholars (see Pessoa 2022, 19–20; Epp 2020).

²⁴We use quotation marks to enclose the word “Byzantine” given that there is no agreed upon set of criteria to distinguish a “Byzantine” witness from one that is not.

great number of erroneous readings early on in the first and second centuries. These traditions would be represented relatively well in the Egyptian uncials of the fourth century and in the third-century Egyptian papyri. At some point after the fourth century, the pre-existing “Byzantine” readings (presumed to be erroneous and originating in what is today Turkey and the Balkans around the second century) would have started to become more and more popular, displacing the previous uncontrolled families. This takeover would have occurred as the result of a long decentralized process driven by scribes who doubled as textual critics over the centuries, as German scholar Klaus Wachtel has proposed (1995, 2009). As a consequence, the safest way to reconstruct the *Ausgangstext* would be to look to the earliest manuscripts from Egypt. These manuscripts, however, are but a handful, and there are textual units where the adopted reading is supported by only one or two manuscripts against many dozens or hundreds that attest to a majority reading (Robinson 2009).

The topology presumed by majority-text advocates on the other hand is based on the idea of normality, ie, that scribes worked largely independently from one another and that the genealogies branched in an approximately self-similar manner as in Hodges’ truism. According to this view, the genealogies started to branch early, but the general care of scribes still kept the witnesses fairly close to the autographs. Any divergence caused by a single scribe would have been cancelled out by the sheer inertia of the majority of scribes not making the same mistake at the same textual unit. At some point in the second century, a number of manuscripts made it to the Christian communities of Alexandria, where they were probably altered by scholars trained in the Alexandrian techniques of textual criticism (Robinson 1993, 2003). These witnesses were the ancestors of the third and fourth century papyri and uncials from Egypt. The overwhelming majority of witnesses circulating in the world in the fourth century, however, would have been free from those editions and, therefore, a majority-rule critical text created at that time would have produced a text very similar to the autographs. The hypothesis also states that the earliest manuscripts from Asia Minor and the Balkans (including the autographs, which were kept by the churches in those regions) were lost due to climatic conditions, which left the modern textual critic with only the few Egyptian manuscripts that survived thanks to the favorable desert weather. It is only after the move from papyri to the more durable parchment took place that the descendants of the earlier Asiatic witnesses started to show up in the extant tradition. The fact that these descendants, the “Byzantine” witnesses, appear in large numbers in different locations and do not show signs of descending from a single late archetype would indicate that they harken back to the autographs following relatively independent streams of transmission (Robinson 2005). In all, the theory aims to provide a justification for using late manuscripts for reconstructing the autographs (or at least for reconstructing very early archetypes) by means of majority rules.

The two theories can be described using the conceptual apparatus proposed in the present article. In fact, we can see that the typical latent local topology imagined

by the editors of the *Editio Critica Maior* resembles the one of Figure 24²⁵, while the one imagined by the majority text advocates looks more like Figure 23. Figure 24 and Figure 23 are of course extremely simplified diagrams. The actual latent history of any given New Testament book would have included millions of segments and edges²⁶, connected in non-trivial ways. It is impossible to fully reconstruct the latent histories, and the closest approximation, the hypothetical local *stemmata*, could be too simple to represent them, due to the fact that the extant tradition is but a fragment of the actual latent tradition. One could make up for this limitation by including many hypothetical witnesses in the *stemmata*, but each additional witness would add assumptions to the theory that could be too specific and, therefore, highly falsable. The more hypothetical witnesses are added, the greater the falsability of the *stemmata* as a whole.

In a simulation approach, the problem can be divided into two phases. The first would try to imagine the general topology of the latent history of a given book. The second phase would test a number of reconstructive heuristics against several majority rules on the simulated history. A history resembling Hodges' truism would favor the majority rule, while an imagined history resembling Figure 24 would favor an antiquity-based strategy. In order to create the simulated histories, the researcher would have to utilize insights from historical research and adopt reasonable postulates about the transmission process over the centuries. The kind of evidence needed for this would include not only the extant manuscripts, but also knowledge about supply and demand for manuscripts in ancient times, the politics and geography of religious persecution, the literacy rates of Christian populations in the Mediterranean basin, the evolution of the demographics of Christianity, the price of parchment and papyri before and after the fall of the Roman Empire, the impact of wars on rates of manuscript destruction, etc. Although the creation of the hypothetical latent histories necessitates the adoption of assumptions to compensate for significant gaps in the available evidence, these assumptions could in certain cases be comparatively less risky than those of someone who seeks to create hypothetical local *stemmata*. This is because while *stemmata* aim to trace back the segments associated with a specific textual unit and recover the evolution of its latent local genealogy, the simulation

²⁵Incidentally, a topology similar to this one is normally presupposed for the Masoretic tradition, with all of the Medieval witnesses, which make up the majority of the extant tradition, deriving from late Masoretic or proto-Masoretic archetypes. Applying majority rules to those witnesses could at most recover the texts of their archetypes, but not necessarily the texts of the ancestors of the archetypes in the earliest centuries of transmission. This is why it is often acceptable to treat the Masoretic corpus as a single witness in apparatuses. This is likely not the case for the books of the Greek New Testament, because no "Byzantine" corpus for any book has been proven to derive from a late "Byzantine" archetype forming a text type (Wachtel 1995).

²⁶There are about 5,800 *extant* manuscripts across all the books of the New Testament (Leggett and Paulson 2023), which means that there were many more in the latent non-local genealogy, with numbers possibly in the dozens of thousands. A New Testament book can have several thousand textual units and therefore, it is not at all unlikely that latent histories had millions of segments and edges.

approach does not have to make any claims about the readings of a textual unit in particular, *but rather about a reconstructive method*. Thus, instead of claiming “this is how the reading evolved over time” as *stemmata* do, the simulation would say “for this particular book or textual unit, a majority approach is recommended over an antiquity-prioritizing one, or vice versa, or no preference can be justified”, regardless of the actual readings. The simulation’s recommendation, in conjunction with internal evidence and other subjective elements, would then be taken into consideration by the textual critic to produce the critical text. This method could offer fruitful insights into extremely complex latent local genealogies like those of the *Pericope Adulterae* and the Longer Ending of Mark²⁷. In the end, the focus becomes the choice of the most plausible transmissional theory (ie the most plausible hypothetical topology) given the known features of the transmissional process.

The Greek New Testament is of course only one example tradition where simulations could be applied, but it is by no means the only tradition that can benefit from it. Within Biblical textual criticism, the traditions of the Vulgate/Vetus Latina and the Septuagint are obvious candidates, but other traditions like those of the Homeric epics and some Roman poets could also be modeled this way.

4 Conclusion

This article provided a formulation of the main notions used in modern textual criticism using the language of graph theory and probability theory, with the purpose of providing a common mathematical framework for the description and evaluation of textual criticism theories, models and methods. In this framework, the textual critic is tasked with assigning probabilities to outcomes within sample spaces composed of the competing readings of textual units. Additionally, by embracing the concept of genealogies, represented as directed graphs where witnesses and segments serve as vertices and parent-child relationships define the edges, we propose an approach that can be used to study the intricate dynamics of textual transmission in a rigorous and at the same time flexible way. Further building upon that foundation, we have introduced additional constructs like evolutions, scribal error, translations and retroversions, among many others. A benefit of this mathematical approach is its potential for reducing semantic ambiguity in academic discussions about readings, traditions and genealogies.

We also demonstrated the capabilities of our formulation by representing two important heuristics families, namely Lachmannian *stemmata* and majority rules, using our system of definitions. A simulation-based meta-heuristic approach useful for choosing which heuristics to adopt for given traditions was outlined. The required virtual scenarios can be created by directly encoding the present system in any programming language suitable for running Monte Carlo simulations.

²⁷For an overview of the discussion about these two passages, see Black and Cerone (2016); Black (2008).

The ideas presented in this paper can be used to further the field of computer-assisted textual criticism. The probabilistic foundation offers a systematic approach to analyzing and interpreting textual variants, enabling computational methods to assign probabilities to different readings and originality statuses, possibly even within a Bayesian framework in which an automated program creates prior distributions and is updated as the human researcher incorporates new evidence. This can pave the way for advanced techniques that could aid textual critics in their decision-making process by providing quantitative assessments and guiding their evaluations of variant readings. For scholars working with Lachmannian heuristics in particular, the present system can also help them find ways to ensure the internal and external consistency of their *stemmata* using automated inconsistency detection software.

References

- Abegg, Martin, Peter Flint, and Eugene Ulrich. 1999. *The Dead Sea Scrolls Bible: The Oldest Known Bible Translated for the First Time into English*. HarperOne.
- Andersen, David. 2014. "Arguments for and Against the Byzantine and Alexandrian Text Types." In *Digging for the Truth: Collected Essays Regarding the Byzantine Text of the Greek New Testament. A Festschrift in Honor of Maurice a. Robinson*, edited by Mark Billington and Peter Streitenberger. FYM.
- Andrews, Tara, and Caroline Macé. 2013. "Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation Through Graph Analysis of Texts and Stemmata." *Literary and Linguistic Computing* 28 (4): 504–21. <https://doi.org/10.1093/lc/fqt032>.
- Arruda, Guilherme Ferraz de, Francisco A. Rodrigues, and Yamir Moreno. 2018. "Fundamentals of Spreading Processes in Single and Multilayer Complex Networks." *Physics Reports* 756 (October): 1–59. <https://doi.org/10.1016/j.physrep.2018.06.007>.
- Bender, Michael A., Martín Farach-Colton, Giridhar Pemmasani, Steven Skiena, and Pavel Sumazin. 2005. "Lowest Common Ancestors in Trees and Directed Acyclic Graphs." *Journal of Algorithms* 57 (2): 75–94. <https://doi.org/10.1016/j.jalgor.2005.08.001>.
- Black, David Alan, ed. 2008. *Perspectives on the Ending of Mark. 4 Views*. Nashville, Tennessee: B&H Academic.
- Black, David Alan, and Jacob N. Cerone, eds. 2016. *The Pericope of the Adulteress in Contemporary Research*. Library of New Testament Studies 551. London: Bloomsbury T&T Clark.
- Boogert, Ernst. 2015. "The Origin of the Byzantine Text: New Perspectives in a Deadlocked Debate." PhD thesis.
- Bordalejo, Barbara. 2015. "The Genealogy of Texts: Manuscript Traditions and Textual Traditions." *Digital Scholarship in the Humanities* 31 (3): 563–77. <https://doi.org/10.1093/lc/fqv038>.
- Burton, Phillip. 2013. "The Latin Version of the New Testament." In *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis*, edited by Bart D. Ehrman and Michael Holmes. Brill.
- Camps, Jean-Baptiste, and Julien Randon-Furling. 2022. "Lost Manuscripts and Extinct Texts : A Dynamic Model of Cultural Transmission." *CHR 2022: Computational Humanities Research Conference, Computational Humanities Research, Dec 2022, Antwerp*,

- Belgium, December.
- Chiesa, Paolo. 2020. "Principles and Practice." In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- del Gratta, Riccardo, Federico Boschetti, Luigi Bambaci, and Francesco Sarnari. 2021. "Document Analysis and Textual Philology: A Formal Perspective." *International Journal of Information Science and Technology*, September, Vol 5 No 1 (2021). <https://doi.org/10.57675/IMIST.PRSM/IJIST-V5I1.192>.
- Del Gratta, Riccardo, Federico Boschetti, Angelo Mario Del Grosso, Luigi Bambaci, and Simone Zenzaro. 2022. "La Filologia come sistema dinamico." *Umanistica Digitale*, October, 1–20 Pages. <https://doi.org/10.6092/ISSN.2532-8816/13684>.
- Del Gratta, Riccardo, Simone Zenzaro, Angelo Del Grosso, and Federico Boschetti. 2023. "2023 7th IEEE Congress on Information Science and Technology (CiSt)." In, 605–10. Agadir - Essaouira, Morocco: IEEE. <https://doi.org/10.1109/CiSt56084.2023.10409916>.
- Epp, Eldon Jay. 2005. "The Eclectic Method in New Testament Textual Criticism: Solution or Symptom?" In, edited by Eldon Jay Epp. Leiden, Boston: Brill.
- . 2013. "Textual Clusters: Their Past and Future in New Testament Textual Criticism." In *The Text of the New Testament in Contemporary Research. Essays on the Status Quaestionis*, edited by Bart D. Ehrman and Michael Holmes, 2nd ed. Leiden ; Boston: Brill.
- . 2020. "In the Beginning Was the New Testament Text, but Which Text? A Consideration of 'Ausgangstext' and "Initial Text"." In, 369405. Brill. https://doi.org/10.1163/9789004442337_012.
- Eshel, Esther. 1991. "4QDeutⁿ — a Text That Has Undergone Harmonistic Editing." *Hebrew Union College Annual* 62: 117–54.
- Estrada, Ernesto. 2011. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press.
- Farstad, Arthur, and Zane Hodges, eds. 1985. *The Greek New Testament According to the Majority Text*. 2nd ed. Thomas Nelson.
- Fee, Gordon D. 1978a. "Modern Textual Criticism and the Majority Text: A Rejoinder." *Journal of the Evangelical Theological Society* 21 (2).
- . 1978b. "Modern Textual Criticism and the Revival of the Textus Receptus." *Journal of the Evangelical Theological Society* 21 (1).
- Govier, Clive E. 1996. "The Majority Text Debate: A Study in New Testament Text-Critical Method." PhD thesis.
- Greg, W. W. 1927. *The Calculus of Variants: An Essay on Textual Criticism*. Oxford.
- Guillaumin, Jean-Baptiste. 2020. "Criticisms of Digital Methods." In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- Haverling, Gerd V. M. 2020. "Background and Early Developments." In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- Hodges, Zane C. 1968. "The Greek Text of the King James Version." *Bibliotheca Sacra* 125 (500).
- . 1978a. "Modern Textual Criticism and the Majority Text: A Response." *Journal of the Evangelical Theological Society* 21 (2).
- . 1978b. "Modern Textual Criticism and the Majority Text: A Surrejoinder." *Journal*

- of the *Evangelical Theological Society* 21 (2).
- Hoenen, Armin. 2019. "An Open Problem in Computational Stemmatology - a Model for Contamination." *Umanistica Digitale*, May, No 5 (2019). <https://doi.org/10.6092/IS.SN.2532-8816/8555>.
- . 2020. "The Stemma as a Computational Model." In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- Holmes, Michael W. 1983. "The 'Majority Text Debate': New Form of an Old Issue." *Themelios* 8 (2): 13–19.
- . 2013. "From 'Original Text' to 'Initial Text': The Traditional Goal of New Testament Textual Criticism in Contemporary Discussion." In *The Text of the New Testament in Contemporary Research: Essays on the Status Quaestionis*, edited by Bart D. Ehrman and Michael W. Holmes. Brill.
- Houghton, H. A. G. 2016. *The Latin New Testament*. Oxford University Press Oxford. <https://doi.org/10.1093/acprof:oso/9780198744733.001.0001>.
- Housman, A. E. 1972. "The Application of Thought to Textual Criticism." In, edited by J Diggle and F. R. D. Goodyear. Vol. 3. Cambridge University Press.
- Howe, Christopher, Adrian C. Barbrook, Matthew Spencer, Peter Robinson, Barbara Bordalejo, and Linne R. Mooney. 2001. "Manuscript Evolution." *Endeavour* 25 (3): 121–26. [https://doi.org/10.1016/s0160-9327\(00\)01367-3](https://doi.org/10.1016/s0160-9327(00)01367-3).
- Howe, Christopher, Adrian Barbrook, Linne Mooney, and Peter Robinson. 2004. "Parallels Between Stemmatology and Phylogenetics." In *Studies in Stemmatology II*, edited by Pieter van Reenen, August den Hollander, and Margot van Mulken. John Benjamins Publishing Company.
- Karavidopoulos, Ioannis D. 2012. "The Ecumenical Patriarchate's 1904 New Testament Edition and Future Perspectives." *Sacra Scripta* x (1).
- Kimelman, Reuven. 1994. "Psalm 145: Theme, Structure, and Impact." *Journal of Biblical Literature* 113 (1): 37. <https://doi.org/10.2307/3266309>.
- Kleovulos of Sardis, Christodoulou of Stavroupoli, and Vasileios Antoniadis, eds. 1904. *The New Testament, Approved by the Great Church of Christ*. Constantinople: The Patriarchate's Typographer.
- Kreuzer, Siegfried. 2015. *The Bible in Greek: Translation, Transmission, and Theology of the Septuagint*. SBL Press.
- Lanier, Gregory R. 2018. "Taking Inventory on the 'Age of the Minuscules': Later Manuscripts and the Byzantine Tradition Within the Field of Textual Criticism." *Currents in Biblical Research* 16 (3): 263–308. <https://doi.org/10.1177/1476993X18760519>.
- Latora, Vito, Vincenzo Nicosia, and Giovanni Russo. 2017. *Complex Networks. Principles, Methods and Applications*. Cambridge University Press.
- Leggett, Katie, and Greg Paulson. 2023. "How Many Greek New Testament Manuscripts Are There Really? The Latest Numbers," September. <https://web.archive.org/web/20231007205220/https://ntvmr.uni-muenster.de/intfblog/-/blogs/how-many-greek-new-testament-manuscripts-are-there-really-the-latest-numbers>.
- Linde, Cornelia. 2015. *How to Correct the Sacra Scriptura? Textual Criticism of the Latin Bible Between the Twelfth and Fifteenth Century*. SSMLL.
- Maas, Paul. 1958. *Textual Criticism*. Oxford.
- Macé, Caroline. 2015. "Textual Criticism Is Not Only an Art. A Response to B. Alexan-

- derson.” *Revue D’histoire Des Textes* xiv (October): 353–58.
- Macé, Caroline, and Phillippe Baret. 2006. “Why Phylogenetic Methods Work: The Theory of Evolution and Textual Criticism.” In *The Evolution of Texts: Confronting Stemmatological and Genetical Methods: Proceedings of the International Workshop Held in Louvain-La-Neuve on September 1-2, 2004*, edited by Caroline Macé, Phillippe Baret, Andrea Bozzi, and Laura Cignoni. *Linguistica Computazionale*, Vol XXIV-XXV. Istituti editoriali e poligrafici internazionali.
- Manafzadeh, Sara, and Yannick M. Staedle. 2020. “Terminology and Methods.” In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- Martín-Contreras, Elvira, and Lorena Miralles-Maciá. 2014. “Interdisciplinary Perspectives for the Study of the Text of the Hebrew Bible: Open Questions.” In *The Text of the Hebrew Bible. From the Rabbis to the Masoretes*, edited by Elvira Martín-Contreras and Lorena Miralles-Maciá. Göttingen: Vandenhoeck & Ruprecht.
- Martínez, Florentino García. 1994. *The Dead Sea Scrolls Translated: The Qumran Texts in English*. Brill.
- Mink, Gerd. 2004. “Problems of a Highly Contaminated Tradition: The New Testament: Stemmata of Variants as a Source of a Genealogy for Witnesses.” In *Studies in Stemmatology II*, edited by Pieter van Reenen, August den Hollander, and Margot van Mulken. John Benjamins Publishing Company.
- Moreno, Yamir, Maziar Nekovee, and Amalio F. Pacheco. 2004. “Dynamics of Rumor Spreading in Complex Networks.” *Physical Review E* 69 (6). <https://doi.org/10.1103/physreve.69.066130>.
- Pastor-Satorras, Romualdo, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. “Epidemic Processes in Complex Networks.” *Reviews of Modern Physics* 87 (3): 925–79. <https://doi.org/10.1103/revmodphys.87.925>.
- Pessoa, Leonardo. 2022. “The CBGM and Lachmannian Textual Criticism.” *TC: A Journal of Biblical Textual Criticism* 27: 17–31.
- Pickering, Wilbur N. 2014. *The Identity of the New Testament Text IV*. Wilbur N. Pickering.
- Robinson, Maurice A. 1993. “The Recensional Nature of the Alexandrian Text-Type: A Response to Selected Criticisms of the Byzantine-Priority Theory.” *Faith and Mission* Fall (1.0): 46–74. <https://www.galaxie.com/article/fm11-1-07>.
- . 2003. “In Search of the Alexandrian Archetype: Observations from a Byzantine-Priority Perspective.” In, edited by C.-B. Amphoux and J. K. Elliot, 45–67. Lausanne, Switzerland: Éditions du Zèbre.
- . 2005. “The Case for Byzantine Priority.” In *The New Testament in the Original Greek. Byzantine Textform*, edited by Maurice A. Robinson and William G. Pierpont, 533–87. Chilton Book Publishing.
- . 2009. “Rule 9, Isolated Variants, and the ”Test-Tube” Nature of the NA27/UBS4 Text: A Byzantine-Priority Perspective.” In, edited by Mark J. Boda and Stanley E. Porter. Grand Rapids, Michigan / Cambridge, UK: William B. Eerdmans Publishing Company.
- Robinson, Maurice A., and William Pierpont, eds. 2018. *The New Testament in the Original Greek. Byzantine Text Form*. 2nd ed. <https://byzantinertext.com/study/editions/robinson-pierpont/> <https://archive.org/details/robinson-pierpont-2018-gnt-edition>.

- Roelli, Philipp. 2020. "Definition of Stemma and Archetype." In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- Screnock, John. 2017. *Traductor Scriptor: The Old Greek Translation of Exodus 1-14 as Scribal Activity*. Leiden, The Netherlands: Brill. <https://doi.org/10.1163/9789004336568>.
- Snapp, James. 2010. "Equitable Eclecticism: The Future of New Testament Textual Criticism." https://www.academia.edu/12549707/Equitable_Eclecticism_The_Future_of_New_Testament_Textual_Criticism.
- Spencer, Matthew, Barbara Bordalejo, Peter Robinson, and Christopher Howe. 2003. "How Reliable Is a Stemma? An Analysis of Chaucer's Miller's Tale." *Literary and Linguistic Computing* 18 (4): 407–22. <https://doi.org/10.1093/lc/18.4.407>.
- Timpanaro, Sebastiano. 2005. *The Genesis of Lachmann's Method*. Translated by Glenn W. Most. Chicago: University of Chicago Press.
- Trovato, Paolo. 2014. *Everything You Wanted to Know about Lachmann's Method: A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text*. *libreriauniversitaria.it edizioni*.
- . 2020. "Neo-Lachmannism: A New Synthesis?" In *Handbook of Stemmatology: History, Methodology, Digital Approaches*, edited by Philipp Roelli. DeGruyter.
- Wachtel, Klaus. 1995. *Der Byzantinische Text Der Katholischen Briefe : Eine Untersuchung Zur Entstehung Der Koine Des Neuen Testaments*. Berlin: Walter de Gruyter.
- . 2009. "The Byzantine Text of the Gospels: Recension or Process?" In. Conference at New Testament Textual Criticism Session 23-327. SBL. http://www.uni-muenster.de/NTTextforschung/cbqm_presentation/ByzEvvPDF.zip.
- Wasserman, Tommy, and Peter J. Gurry. 2017. *A New Approach to Textual Criticism: An Introduction to the Coherence-Based Genealogical Method*. SBL Press.
- Weitzman, Michael. 1974. "A Statistical Approach to Textual Criticism, with Special Reference to the Peshitta of the Old Testament." PhD thesis, University of London.
- . 1978. "The Tradition of Manuscripts: A New Approach." *The Heythrop Journal*, 28–45.
- . 1985. "The Analysis of Open Traditions." *Studies in Bibliography* 38: 82–120.
- . 1987. "The Evolution of Manuscript Traditions." *Journal of the Royal Statistical Society. Series A (General)* 150 (4): 287. <https://doi.org/10.2307/2982040>.
- Wisselink, W. F. 1989. *Assimilation as a Criterion for the Establishment of the Text*. Kampen: Uitgeversmaatschappij J.H. Kok - Kampen.
- Zeolla, Gary F. 2012. *Differences Between Bible Versions Translation Principles, Greek Text-Types, and KJV-Onlyism*. 3rd ed.