

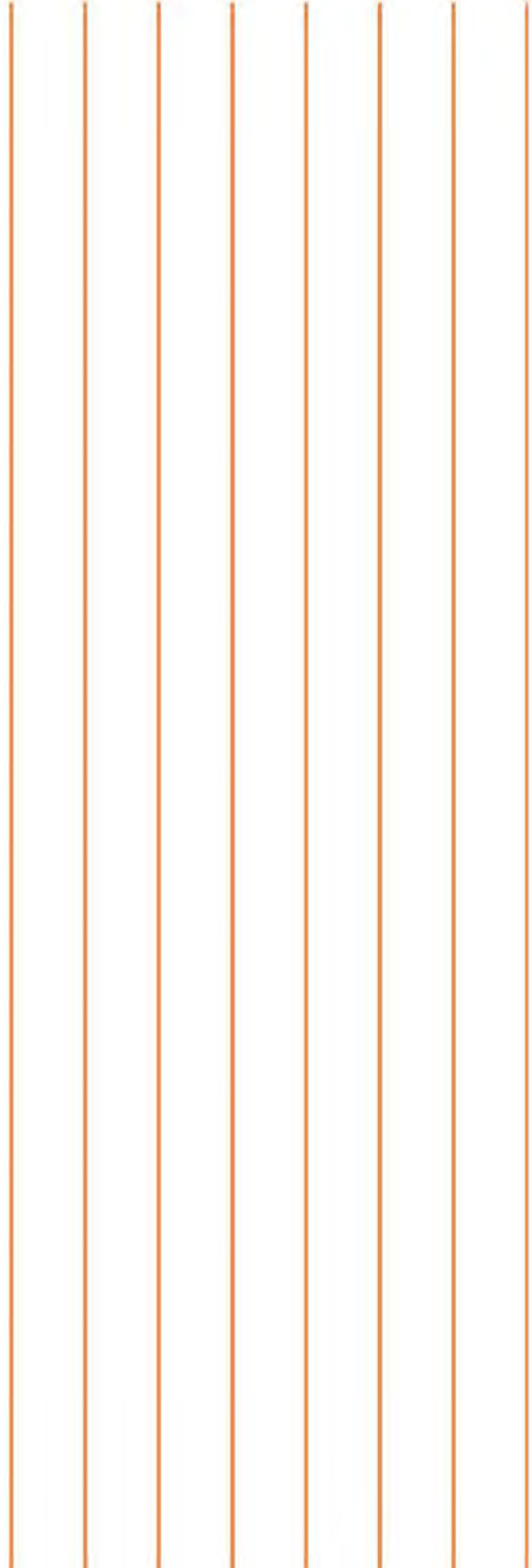


2019

Preprint N°499

How to Generate a Fingerprint

Victoria Beyer



How to Generate a Fingerprint

Victoria Beyer



CONTENTS

Table of Contents

<i>1</i>	<i>Introduction</i>	<i>1</i>
<i>2</i>	<i>Flow Chart</i>	<i>2</i>
<i>3</i>	<i>Counting Pages</i>	<i>3</i>
<i>4</i>	<i>Choosing Lines</i>	<i>4</i>
<i>5</i>	<i>Choosing Symbols</i>	<i>6</i>
<i>6</i>	<i>Sets 1 to 4</i>	<i>8</i>
<i>7</i>	<i>Page Suffix</i>	<i>12</i>
<i>8</i>	<i>Date and Date Suffix</i>	<i>13</i>

INTRODUCTION

Introduction

The purpose of this manual is to provide an overview of one of the methods we used to classify the books included in the *Sphaera* database.¹ To do this, we have slightly adapted a method used in libraries to catalogue an inventory of early modern printed texts. Our approach is based specifically on the process outlined in the EDIT16 database.²

Each of our fingerprints consists of sixteen symbols, separated into four sets of four. A suffix indicates the way in which the sets were generated. This is followed by the year of publication and a further suffix which indicates the format in which the date appears in the publication. The choice of symbols is guided by a very specific set of rules—presented in detail on pages three to fourteen, and in an overview on page two of this manual. The strictness of the process allows us to detect similarities between the completed fingerprints, which in turn makes it possible to compare books from our corpus without reference to metadata. This means that we can quickly flag publications which are likely to be of similar or identical content and physical composition. The major advantage of this method over others is the ease with which books can be classified, as well as its relatively good adaptability to a fully digital workflow.

However, some issues have emerged in our application of the process, particularly due to this digital environment as well as the composition of our body of publications.

As we have generated the fingerprints using only PDF scans of the actual books, flaws in the sometimes already compromised prints were exacerbated.

We have not introduced a specific symbol to distinguish these cases from cases where the print in the original book is illegible, as it is not always clear whether the issue stems from the original or the scan. Such particularities are therefore not reflected in the fingerprints. The same goes for the extreme variety of symbols present in the publications, which are not transcribed in detail. It is therefore possible, in theory, for two fingerprints to be identical even though the text from which they were generated is different.

Another significant difficulty arises from the fact that we analyse only a single copy from an entire print run. As damage has occurred with many copies, and their individual history is not always traceable, this means that we are not sure whether all exemplars from the same run are identical. Digitization is a further potential source of inaccuracy as pages may be missed, misordered, rendered illegible, or the integrity of the original book otherwise compromised in the process.

Nevertheless, the fingerprints have allowed us to classify the large number of publications in our database and will be an essential tool in detecting and investigating relations among the editions which constitute the corpus.

¹ db.sphaera.mpiwg-berlin.mpg.de

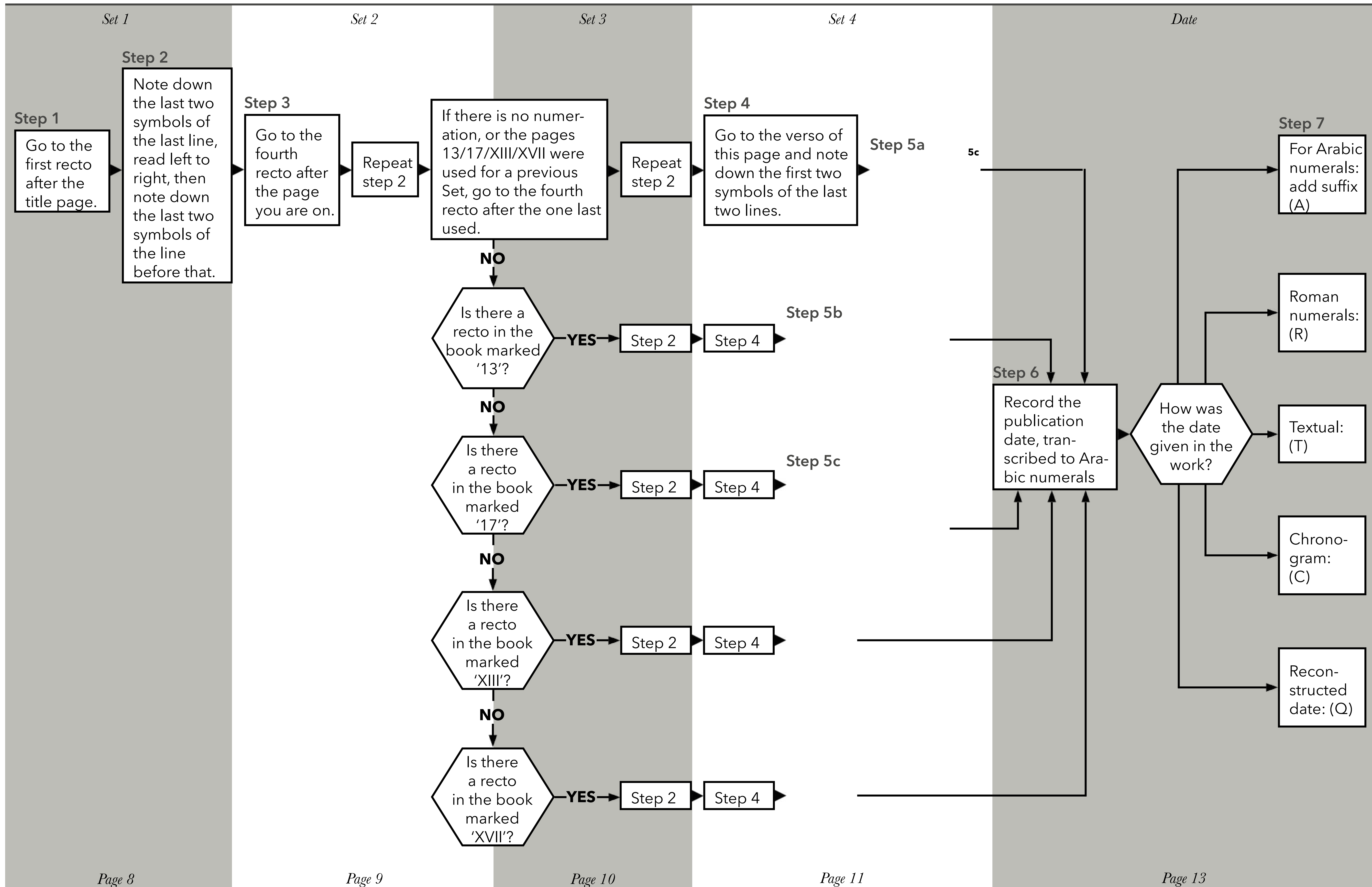
² edit16.iccu.sbn.it

How to generate the Fingerprint

01 Note down the last two symbols from the bottom two lines of the first recto [Set 1] then do the same for the fourth recto following this [Set 2]

02 Find a page marked 13, 17, XIII, XVII, or count to the fourth recto and record the last two symbols of the bottom two lines [Set 3] then do the same for the verso [Set 4]

03 Transcribe the date of publication to Arabic numerals and note it down, then specify the way in which it was given with the appropriate suffix [Date]



*Specific Issues:
Counting Pages*

A Pages are counted from the first recto after the title page

B Leaves which are entirely blank are not counted

C For books that are too short, the final sets are generated from the two lines directly above those used for the preceding Set

XXXX XXXX XXXX XXXX (C) 1500 (Q)

If the book does not have the number of pages required to generate all four sets in the usual manner, the remaining symbols are taken from the two lines above those used for the previous Set.

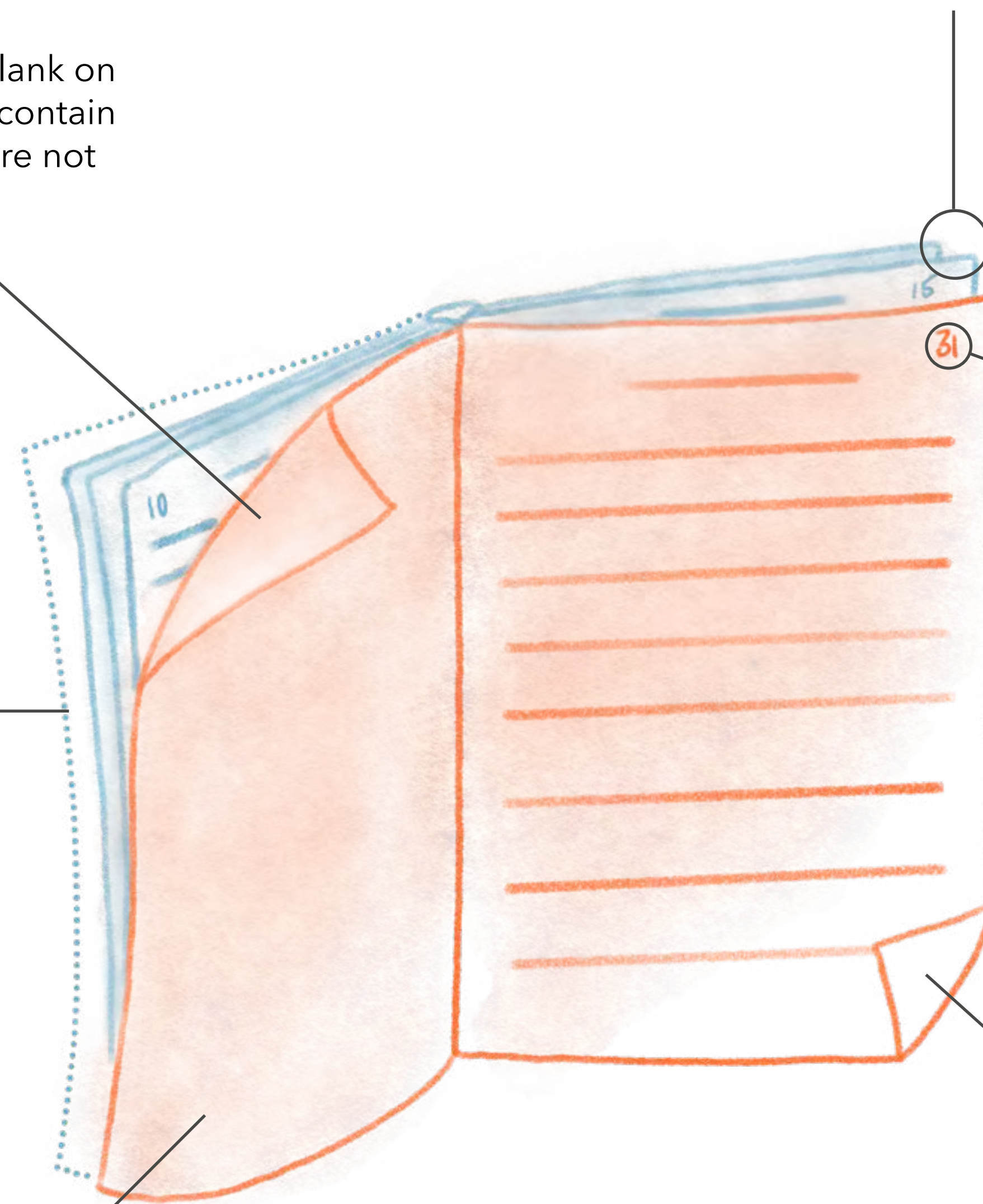
Leaves which are blank on recto and verso or contain only handwriting, are not counted.

If multiple pages appear to be missing from the beginning of the book, Sets 1 and 2 are given as +++++ +++++.
If it is certain that only the title page is missing, the Sets are generated as usual.

If the numeration on the actual pages 13, 17, XIII, or XVII is evidently wrong or illegible, do not use them. If 13 is numbered incorrectly, use 17. If the same is the case for 17, count to the fourth recto after the recto used for Set 2.

If the verso of the page used for Set 3 contains no letterpress, the symbols for Set 4 are taken from the two lines above those used for Set 3.

If the pages usually used for Sets 1 to 3 do not contain letterpress—because they are blank, contain only images, engraved text, or handwriting—the Set in question is generated from the next recto.



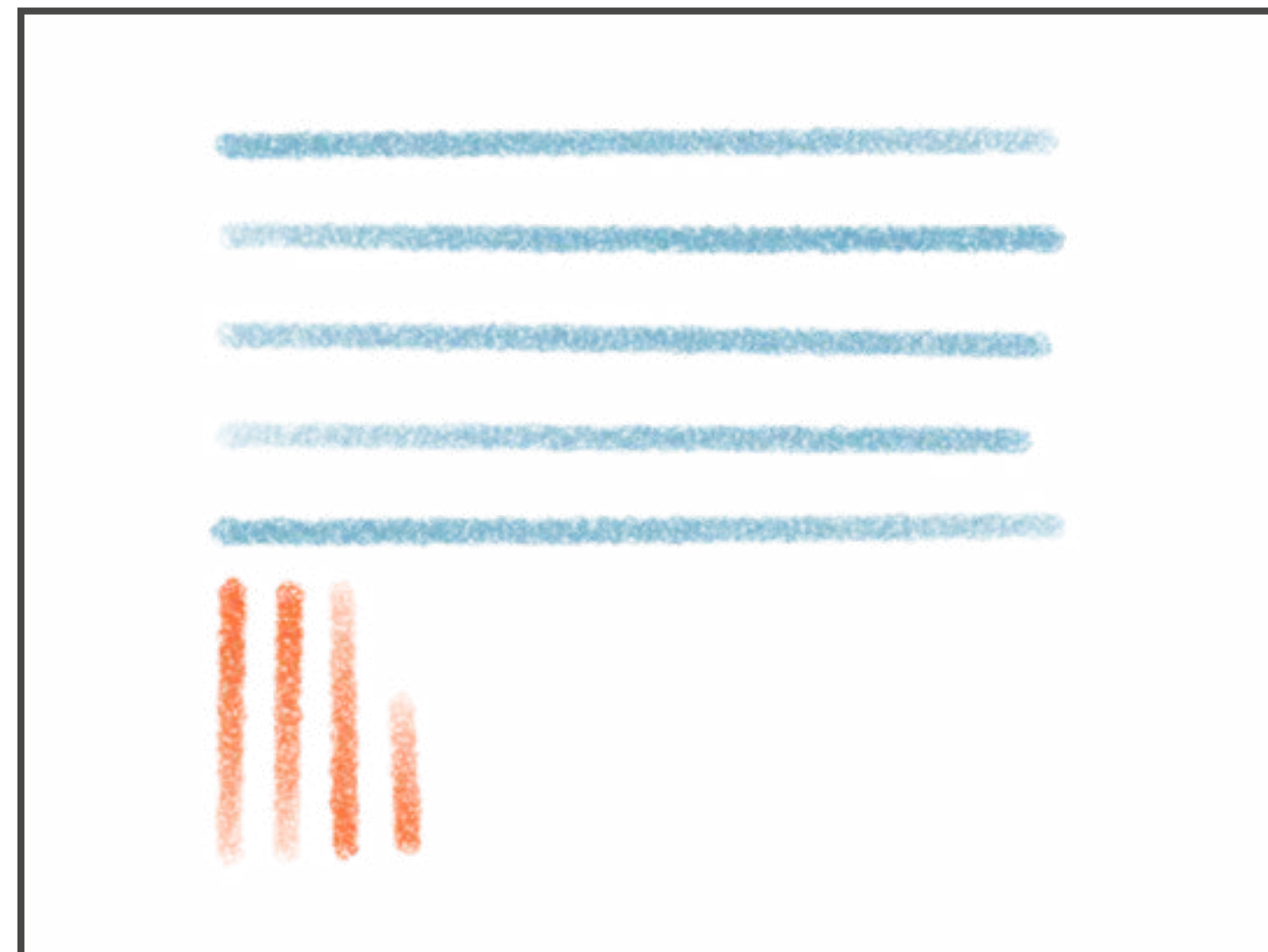
*Specific Issues:
Choosing Lines*

A Only lines containing letterpress are used to generate the Sets

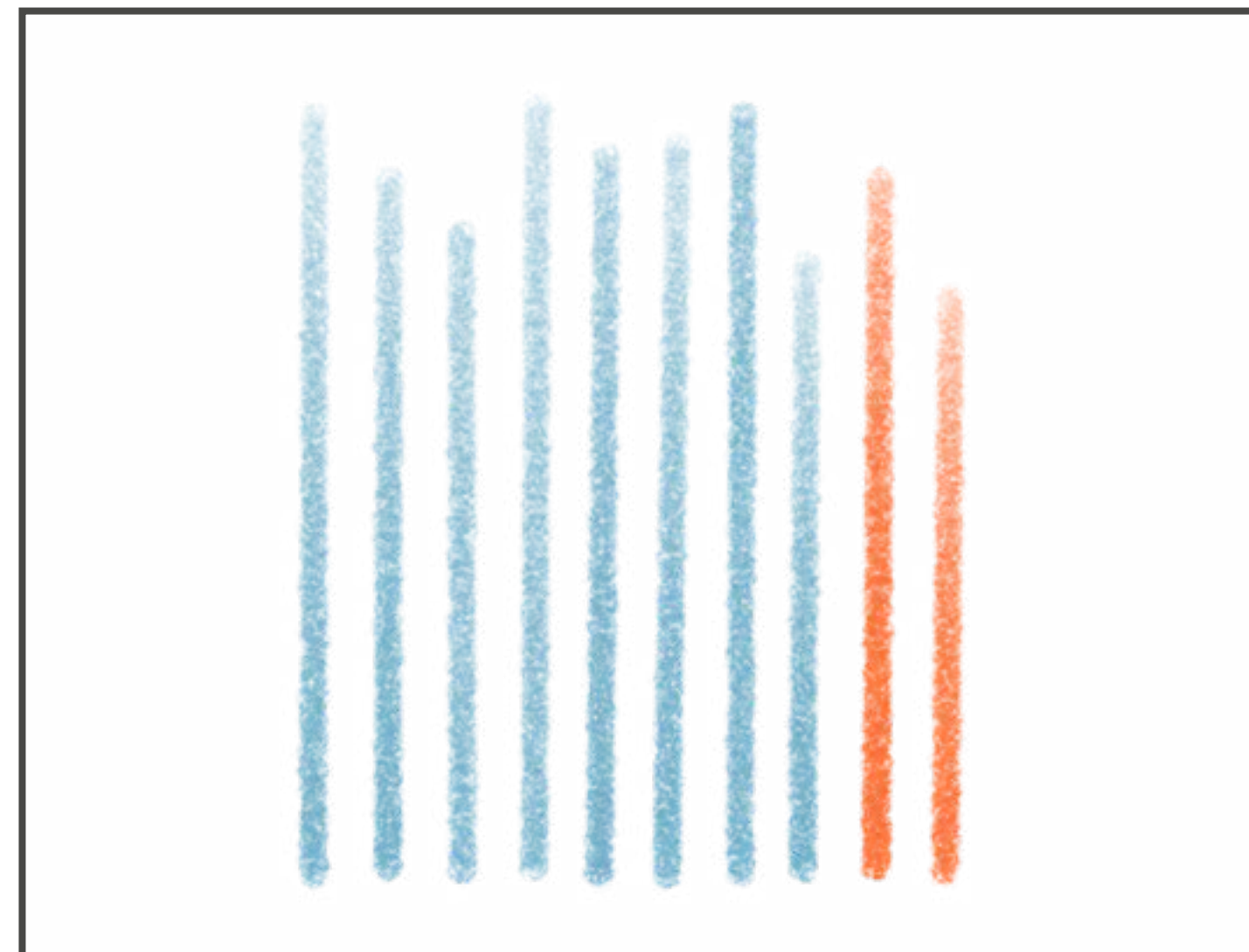
B Marginalia, line numeration, printer's guides, etc. are disregarded

C For text laid out in columns, Sets are generated from the column furthest to the left

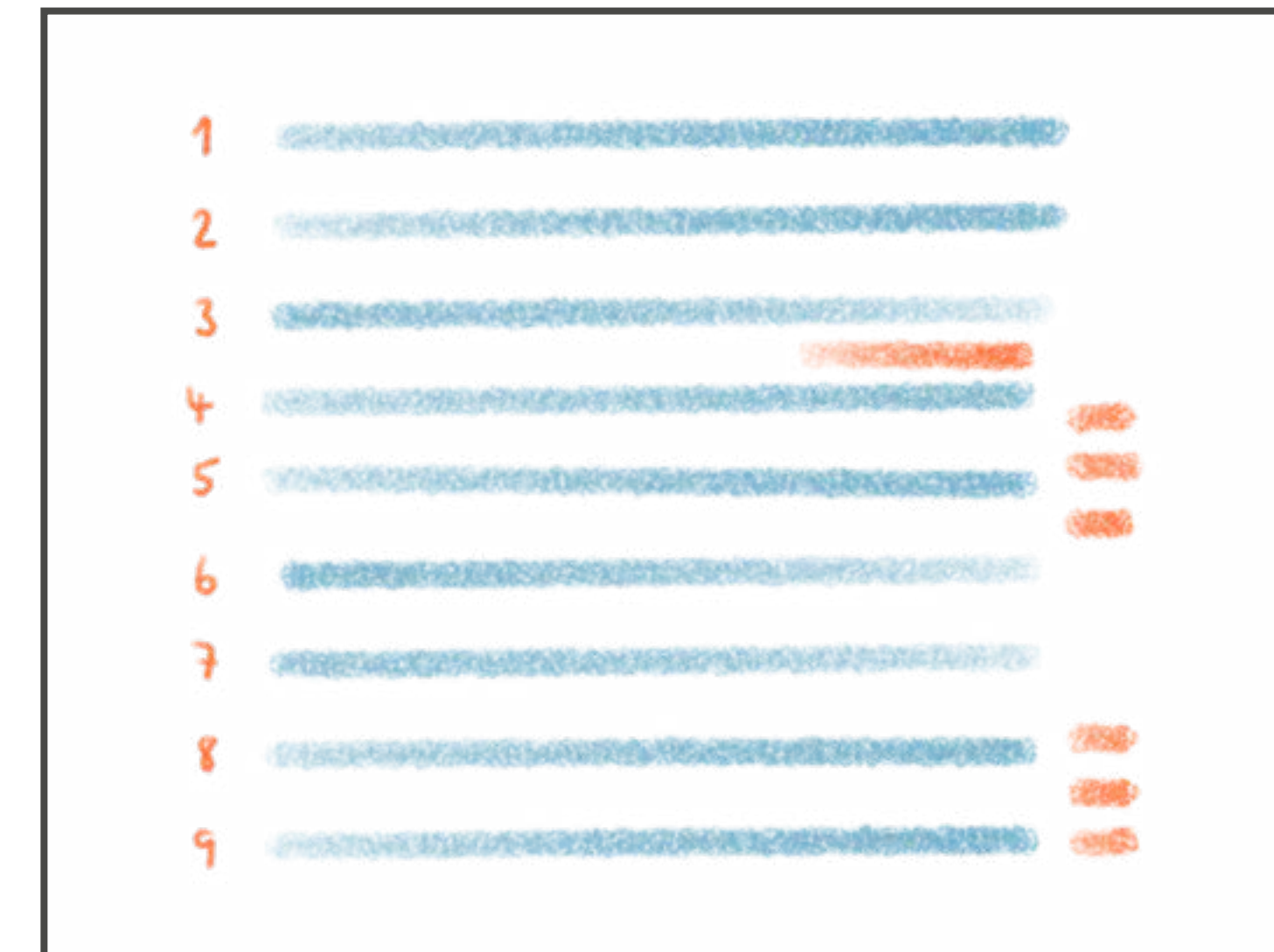
XXXX XXXX XXXX XXXX (C) 1500 (Q)



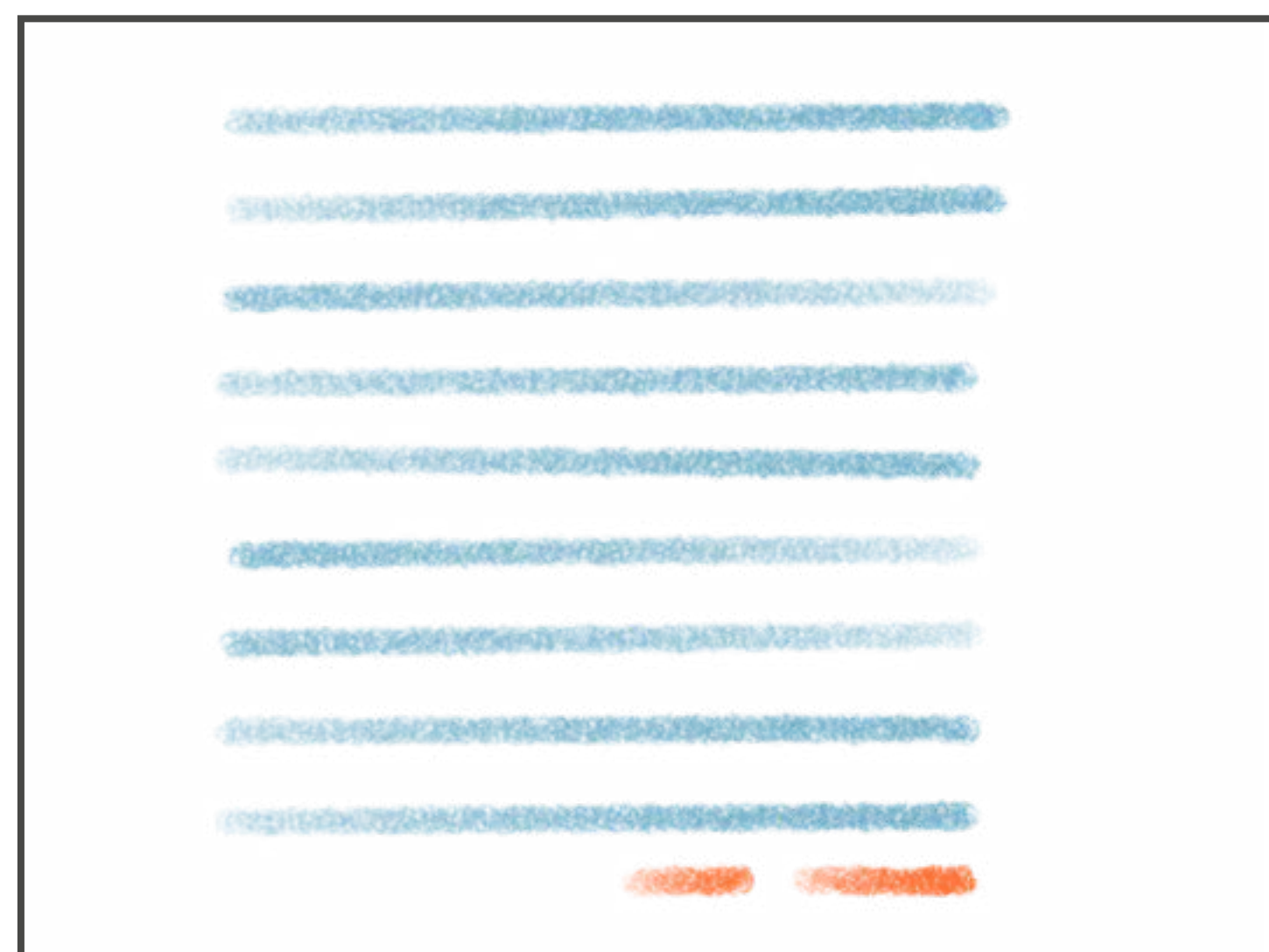
Text laid out at 90° to the main text body on the page is disregarded.



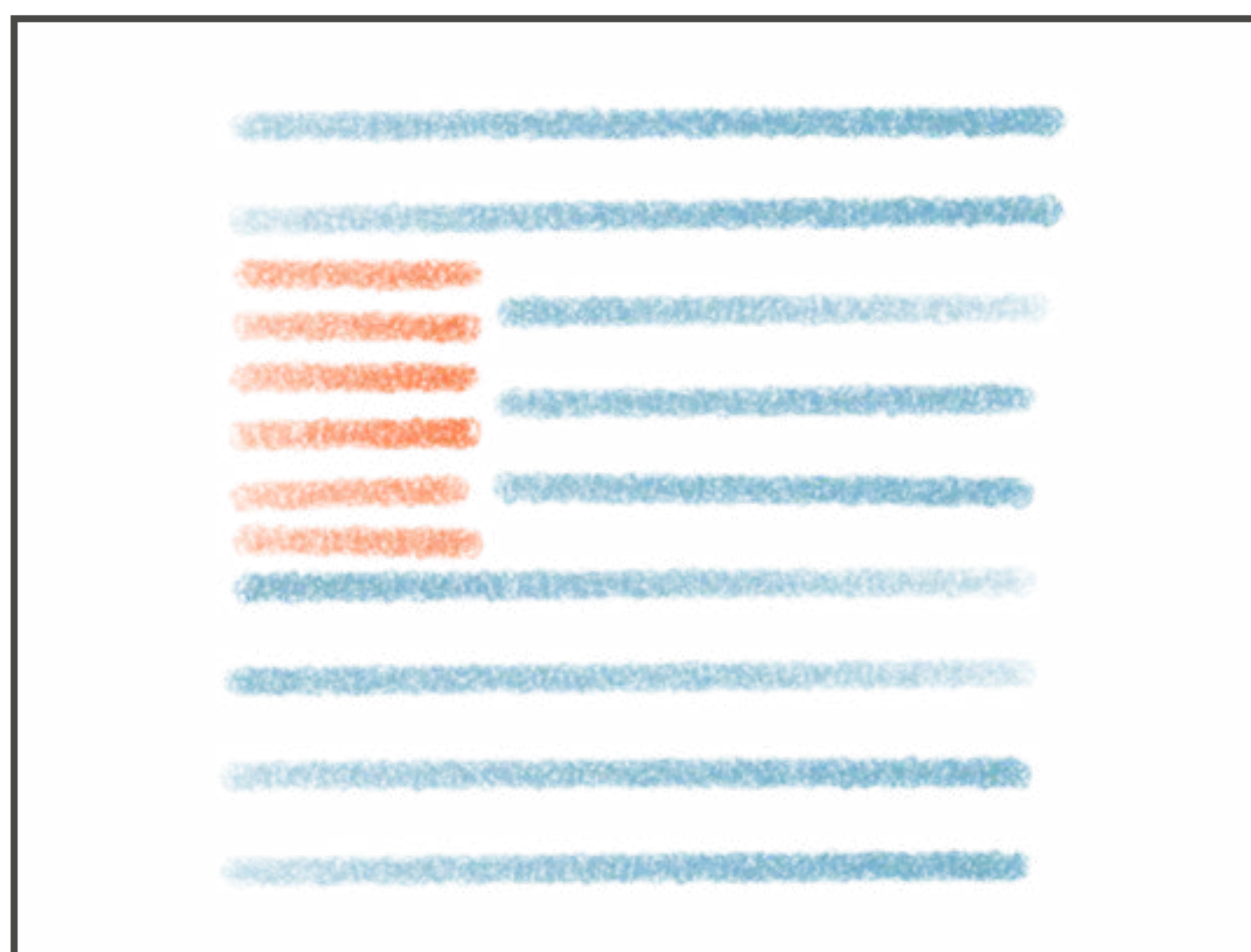
If the entire text on a page is laid out 90° to the dominant orientation in the book, the page is turned according to the standard orientation and the Set generated as usual.



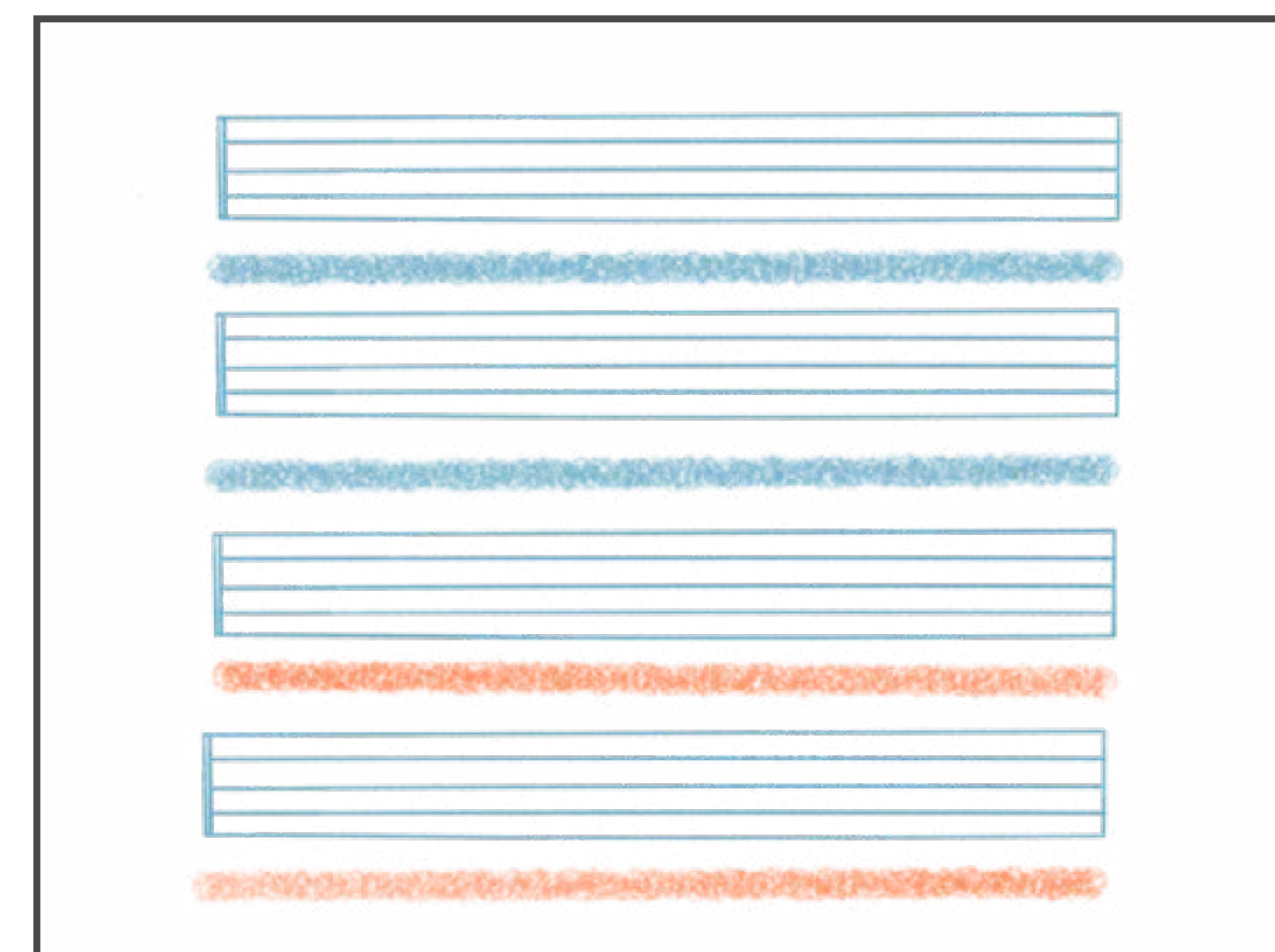
Text in margins, line numeration, and text between lines are disregarded.



Text occurring below the main body of text—denoting the page's location on the folio, the first word on the next page, or similarly meant for the assembly of the book rather than the reader—are disregarded.



Any boxes of text embedded in a larger flow of text are disregarded. The Set is generated as usual, from the main flow of text.



Letterpress text set below or in between lines of musical notation is treated as standard text.

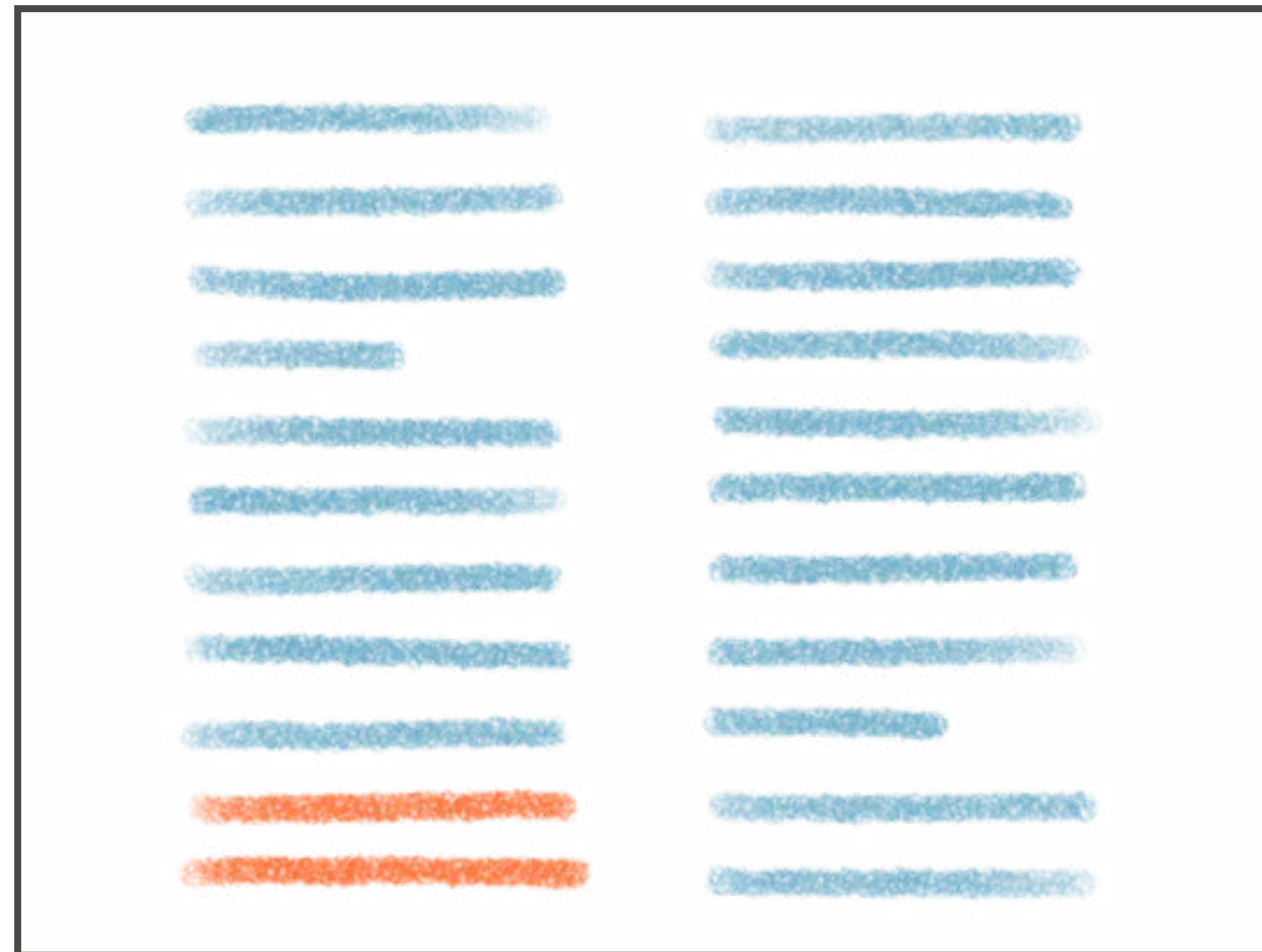
*Specific Issues:
Choosing Lines*

A Only lines containing letterpress are used to generate the Sets

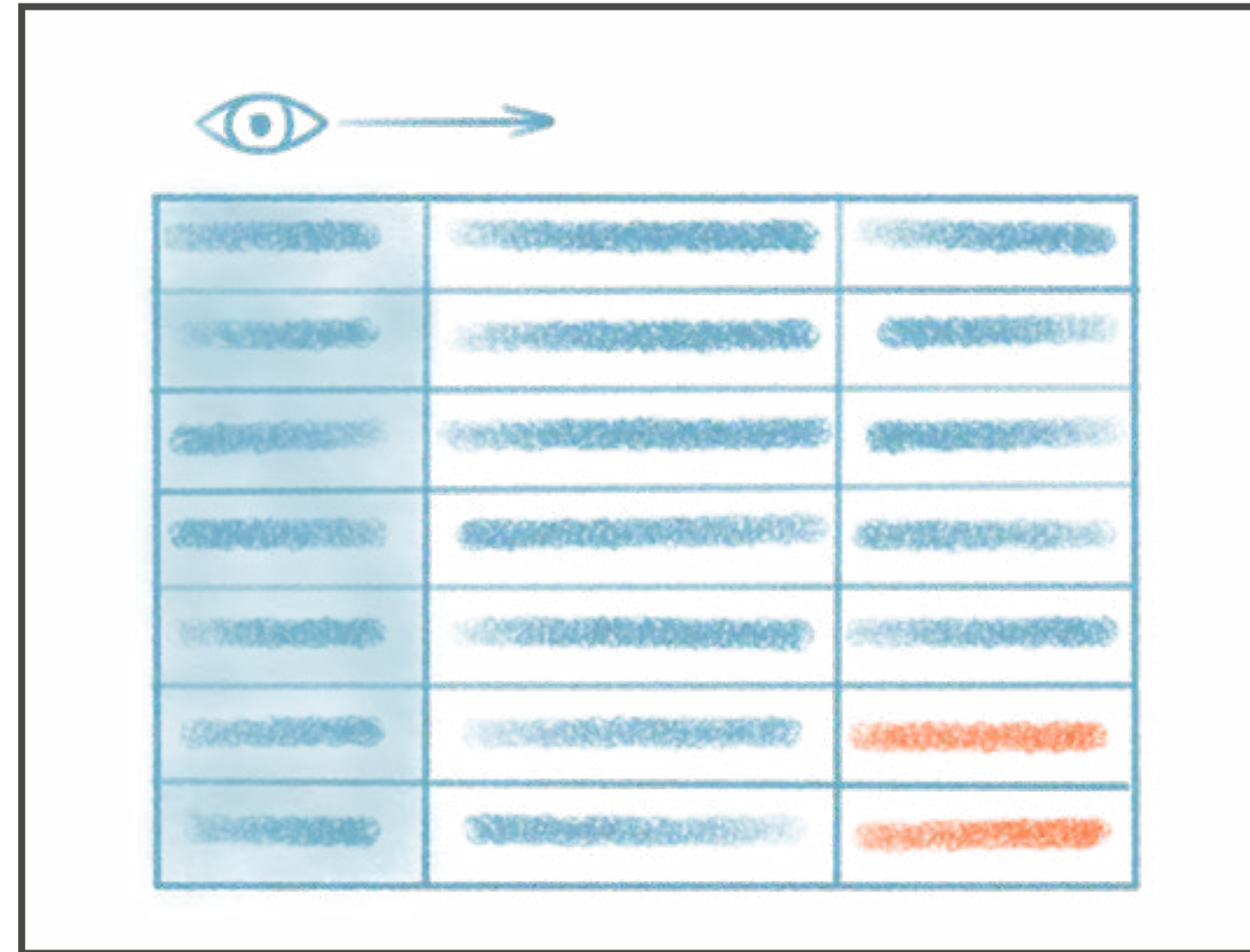
B Marginalia, line numeration, printer's guides, etc. are disregarded

C For text laid out in columns, Sets are generated from the column furthest to the left

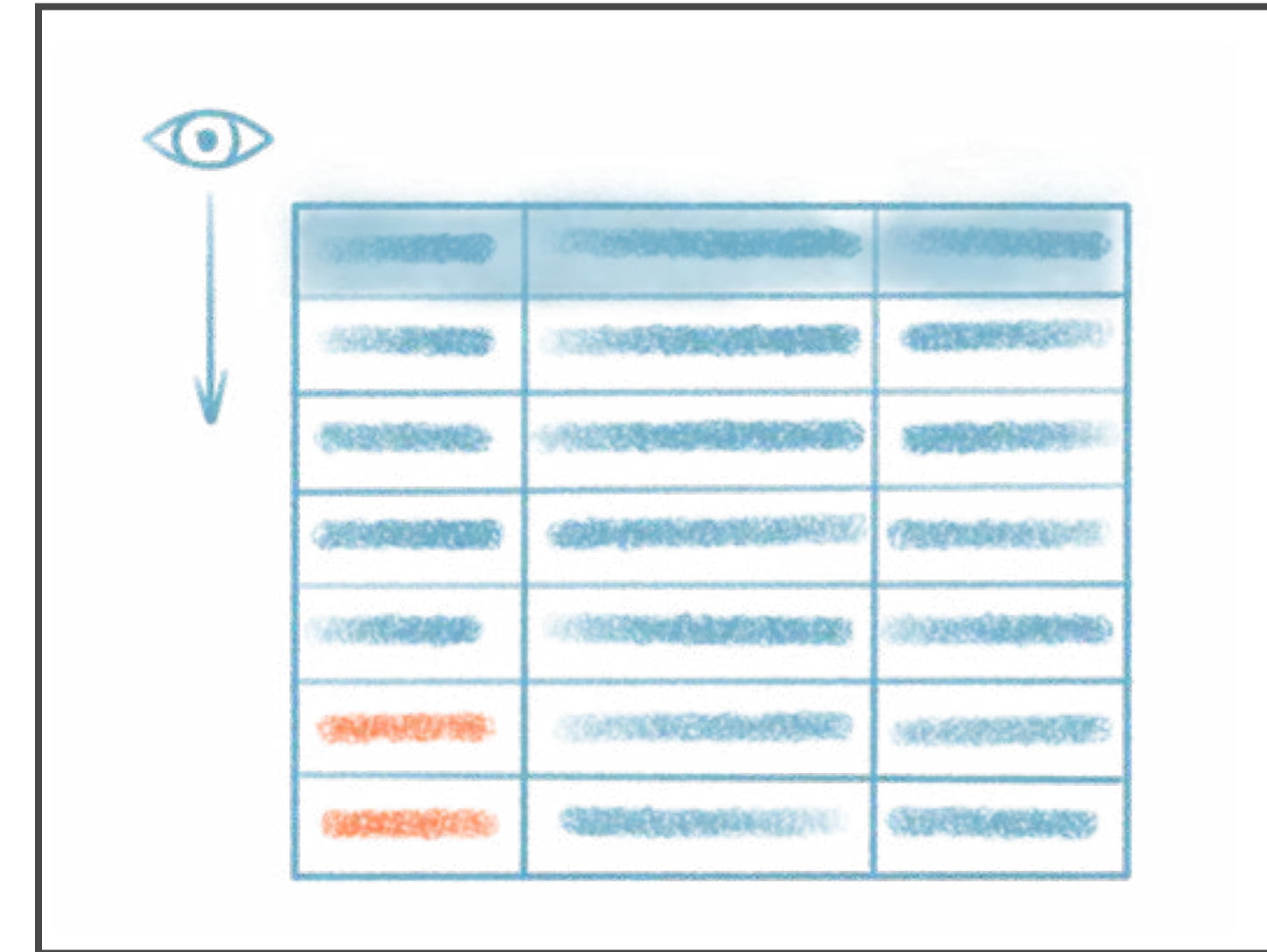
XXXX XXXX XXXX XXXX (C) 1500 (Q)



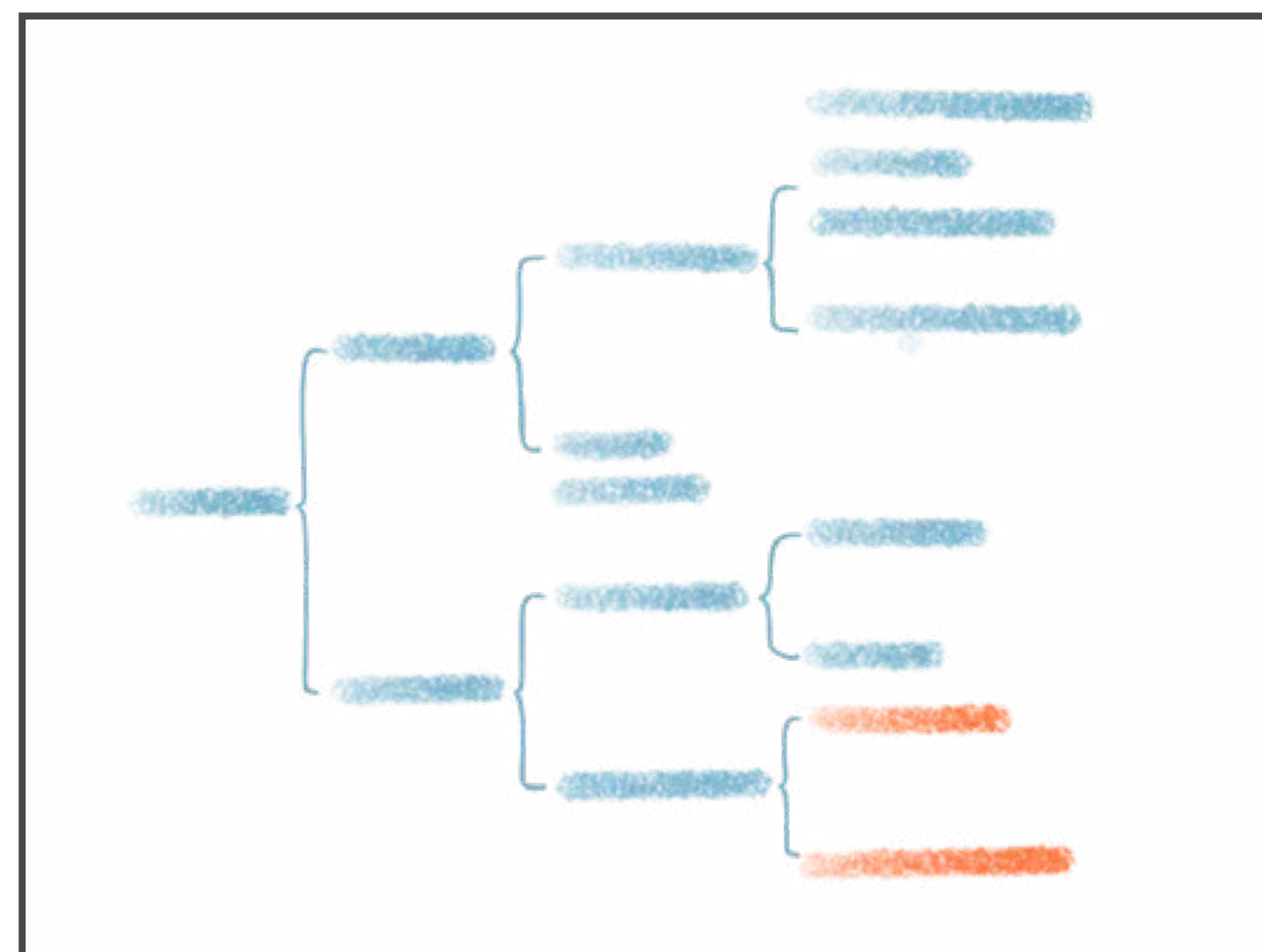
If all text on a page or that part of text relevant for the Set is laid out in columns, the Set is generated from the column furthest to the left.



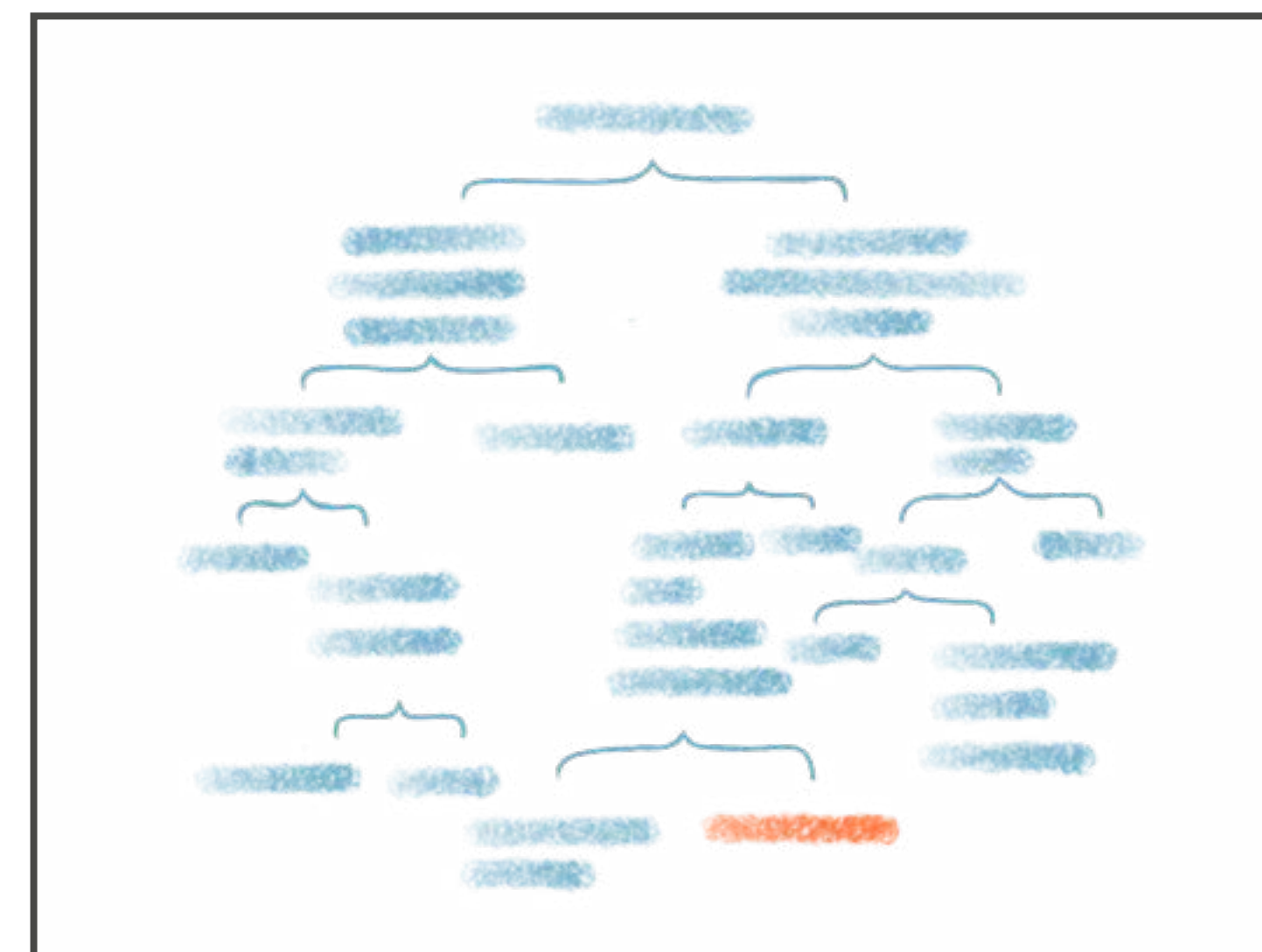
If text is laid out in a table designed to be read from left to right, symbols are taken from the cell or cells at the bottom right, equivalent to normal text. Set 4 is generated from the first symbols in those cells.



If text is laid out in a table designed to be read from top to bottom, the symbols are taken from the cell or cells at the bottom left, equivalent to text in columns.



If letterpress text forms part of a tree graph designed to be read from left to right, symbols are taken from the node at the bottom right.
If the node consists of a single line only, the second half of the Set is generated from the node exactly above.



If letterpress text forms part of a tree graph designed to be read from top to bottom, symbols are taken from the node at the bottom right.
If the node consists of a single line only, the second half of the Set is given as ++.

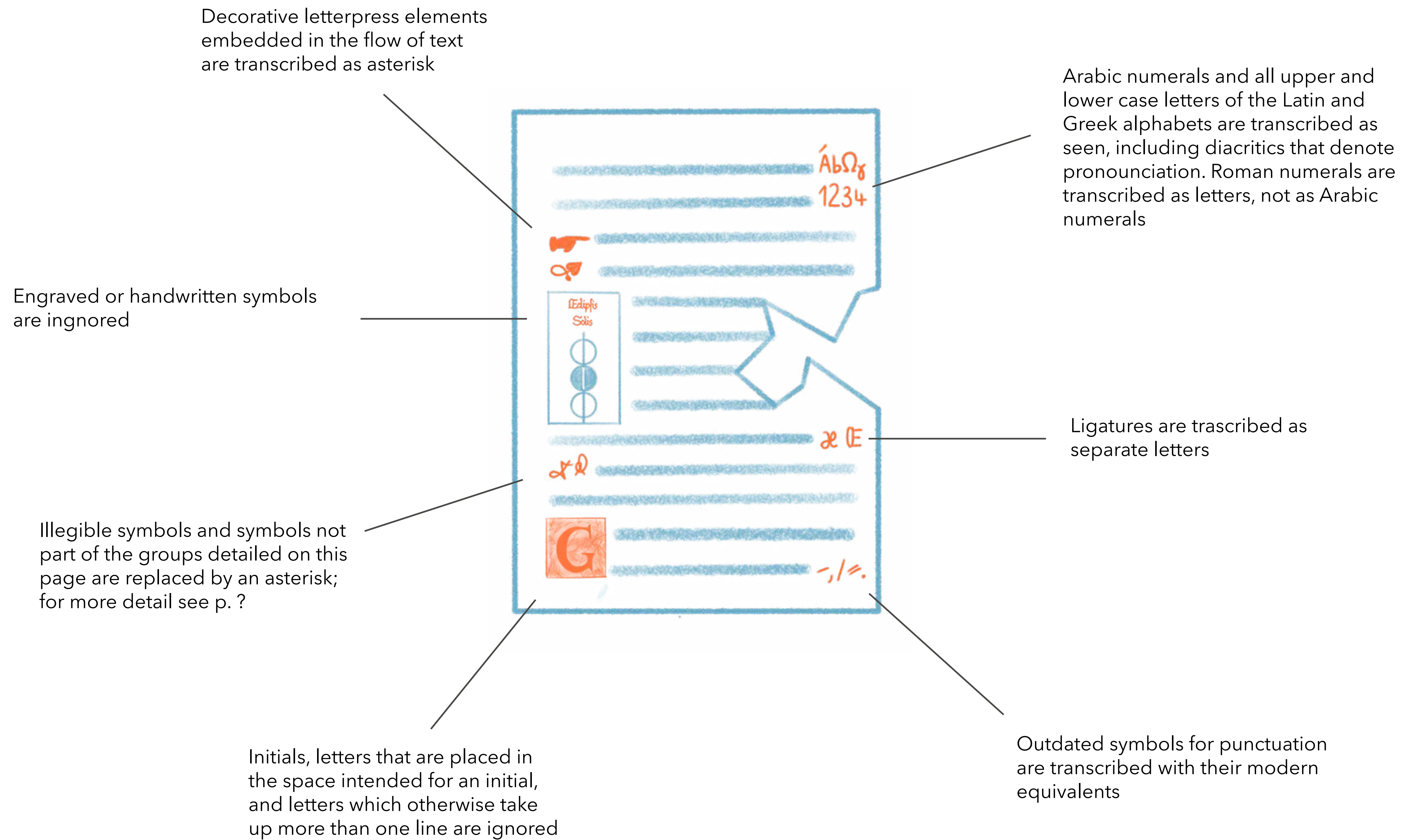
*Specific Issues:
Choosing Symbols*

A The Latin and Greek alphabets as well as Arabic numerals are transcribed as seen, ligatures are separated

B Initials and letters acting as their placeholders are ignored

C Illegible symbols are transcribed with an asterisk

XXXX XXXX XXXX XXXX (C) 1500 (Q)



*Specific Issues:
Symbols*

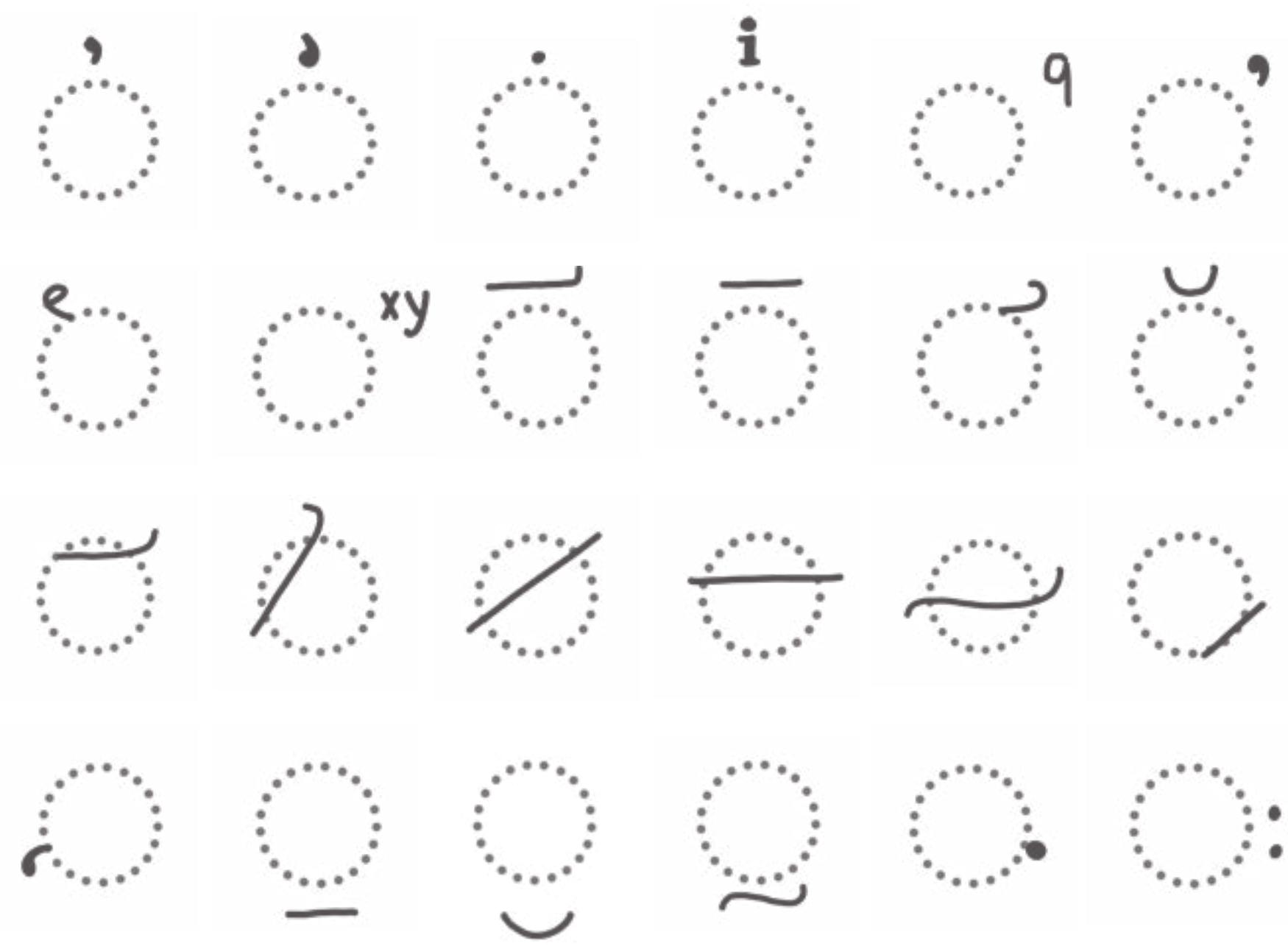
A The Latin and Greek alphabets as well as Arabic numerals are transcribed as seen, ligatures are separated

B Initials and letters acting as their placeholders are ignored

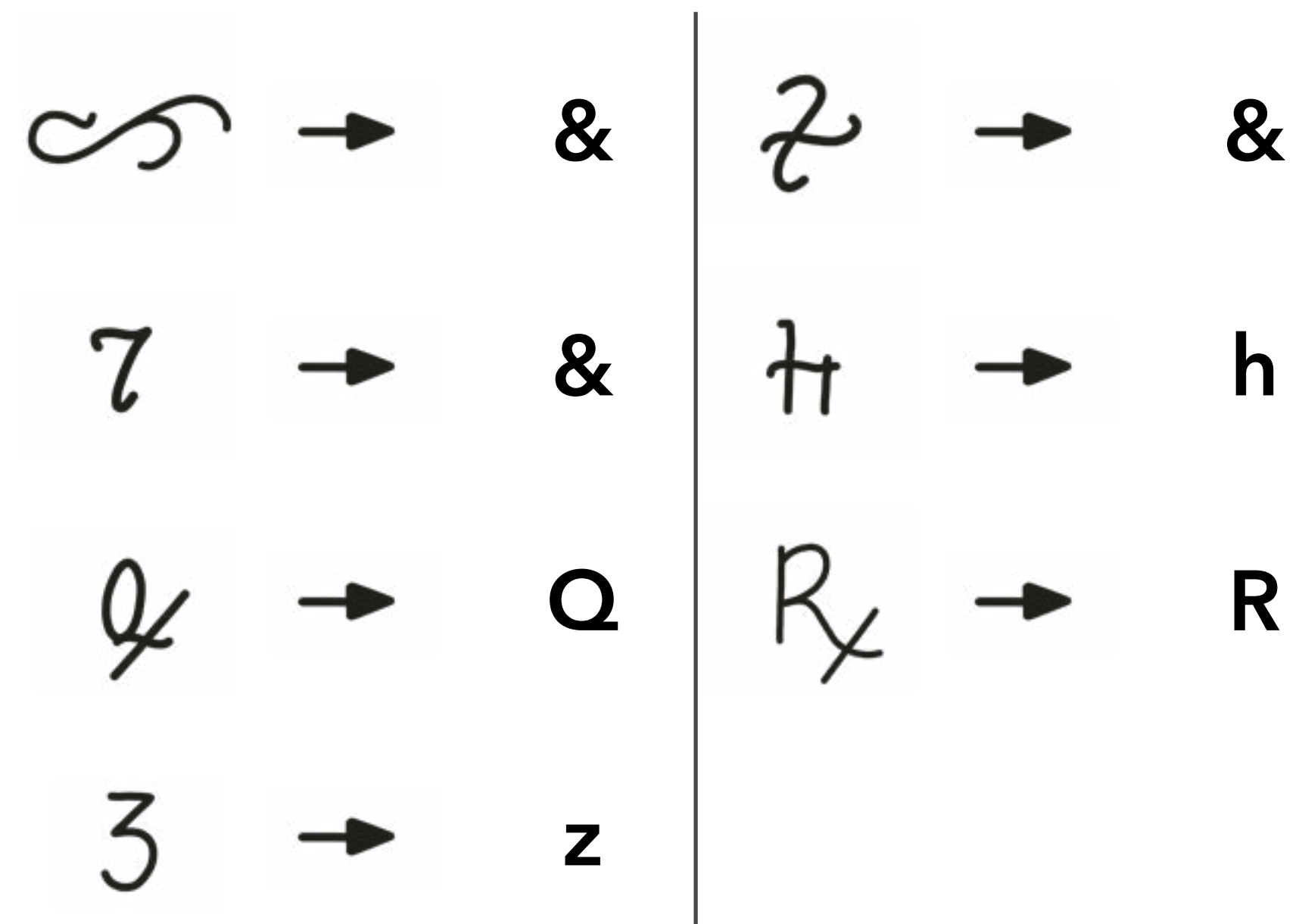
C Illegible symbols are transcribed with an asterisk

XXXX XXXX XXXX XXXX (C) 1500 (Q)

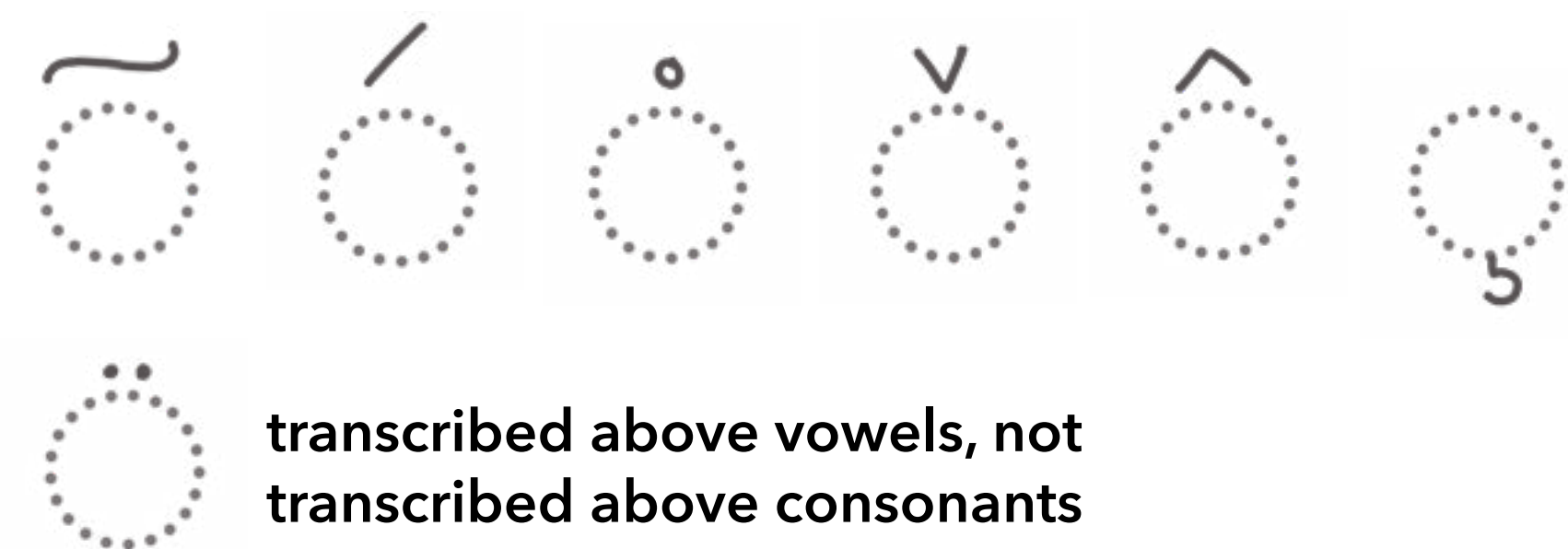
Accents and additions that are not transcribed:



Special characters and transcriptions:



Accents that are transcribed if they serve to denote pronunciation in the source:



Set 1

Title Page

If the first recto after the title page is itself a title page, the Set is generated from the recto following that page.

If the document never had a title Page, Set 1 is generated from the first recto in the book containing letterpress.

Missing Pages

If it appears that the title page is missing but the following pages are obviously complete, the fingerprint is generated as above.

If multiple pages are, or may be, missing, Sets 1 and 2 are recorded as +++++ +++++. If the book contains no recto marked '13', '17', 'XIII', or 'XVII', Sets 3 and 4 are also recorded as +++++ +++++. In this case, the appropriate suffix is (C).

No Letterpress

If the first recto after the title page contains no letterpress, either because it is blank or because it contains only images or engraved text, it is not counted. The set is instead taken from the next recto, or the first recto to contain letterpress.

Index

The Index is treated as a table, with the symbols taken from the bottom right cells.

Damage

Is there damage to the page at the lines needed to generate the Set, the entire affected pair is transcribed as ++, even if it seems that only one symbol is missing. Reconstruction of the symbols should not be attempted.

Not enough Symbols

If the line or page does not contain enough symbols to complete the Set, the missing symbols of each pair are transcribed as +.

Can't read/ weird Symbols

Illegible symbols and symbols that do not belong to those characters detailed on page 7 are transcribed as *.

*How to generate
Set 2*

01 Go to the fourth
recto after the page
used for Set 1

02 Note down the last two
symbols of the last line,
read from left to right

03 Note down the last two
symbols of the penultimate
line, read from left to right

XXXX XXXX XXXX XXXX (C) 1500 (Q)



Missing Pages

If it appears that the title page is missing but the following pages are obviously complete, the fingerprint is generated as above.

If multiple pages are, or may be, missing, Sets 1 and 2 are recorded as +++++ +++++. If there the book contains no recto marked '13', '17', 'XVII', or 'XVII', Sets 3 and 4 are also recorded as +++++ +++++. In this case, the appropriate suffix is (C).

No Letterpress

If the fourth recto after the one used for set one contains no letterpress, either because it is blank or because it contains only images or engraved text, it is not counted. The set is instead taken from the next recto, or the first recto to contain letterpress.

Index

The Index is treated as a table, with the symbols taken from the bottom right cells.

Damage

Is there damage to the page at the lines needed to generate the Set, the entire affected pair is transcribed as ++, even if it seems that only one symbol is missing. Reconstruction of the symbols should not be attempted.

*Not enough
Symbols*

If the line or page does not contain enough symbols to complete the Set, the missing symbols of each pair are transcribed as +.

*Can't read/
weird Symbols*

Illegible symbols and symbols that do not belong to those characters detailed on page 7 are transcribed as *.

*How to generate
Set 3*

01 Go to the recto marked 13, 17, XIII, XVII, or count to the fourth recto after the one used for Set 2

02 Note down the last two symbols of the last line, read from left to right

03 Note down the last two symbols of the penultimate line, read from left to right

XXXX XXXX XXXX XXXX (C) 1500 (Q)



No Letterpress

If the recto marked '13' contains no letterpress, either because it is blank or because it contains only images or engraved text, the recto marked '17' is used. If the same is the case here, the Set is generated from recto 'XIII', 'XVII', or the fourth recto after the one used for Set two. If the page still does not contain letterpress, the Set is generated from the next recto to contain letterpress.

Damage

Is there damage to the page at the lines needed to generate the Set, the entire affected pair is transcribed as ++, even if it seems that only one symbol is missing. Reconstruction of the symbols should not be attempted.

Missing Pages

If pages are missing from the middle or end of the book, each remaining Set is recorded as ++++. This is the case for any and all Sets that would usually be generated from pages following the last correctly present page.

*Not enough
Symbols*

If the line or page does not contain enough symbols to complete the Set, the missing symbols of each pair are transcribed as +.

Faulty Pagination

If the pagination is faulty—if, for instance, recto '14' follows recto '12', recto '13' occurs twice, or similar—the Set is generated from the recto marked '17'. If the same is the case here, the Set is generated from recto 'XIII', 'XVII' or the fourth recto following the one used for Set two.

*Can't read/
weird Symbols*

Illegible symbols and symbols that do not belong to those characters detailed on page 7 are transcribed as *.

*Not enough
Pages*

If there are not enough pages to complete Sets 3 and/or 4 in the usual manner, each remaining Set is generated from the last symbols of the two lines directly above the ones used for the preceding Set.

*How to generate
Set 4*

01 Go to the verso of
the page used for
Set 3

02 Note down the first two
symbols of the last line,
read from left to right

03 Note down the first two
symbols of the penultimate
line, read from left to right

XXXX XXXX XXXX XXXX (C) 1500 (Q)



Missing Pages

If pages are missing from the middle or end of the book, all Sets that remain are recorded as +++. This is the case for any and all Sets that would usually be generated from pages following the last correctly present page.

No Letterpress

If the verso of the page used for Set 3 contains no letterpress, either because it is blank, or because it contains only images or engraved text, the symbols are taken from the two lines immediately above those used to generate Set 3.

Damage

Is there damage to the page at the lines needed to generate the Set, the entire affected pair is transcribed as ++, even if it seems that only one symbol is missing. Reconstruction of the symbols should not be attempted.

*Tables and
Graphs*

The choice of lines is the same as for any other Set. If Set 4 is generated from a horizontal table or a graph, the symbols used are the leftmost symbols of the bottom right cells or nodes.

*Not enough
symbols*

If the line or page does not contain enough symbols to complete the Set, the missing symbols of each pair are transcribed as +.

*Can't read/
weird symbols*

Illegible symbols and symbols that do not belong to those characters detailed on page 7 are transcribed as *.

*Not enough
pages*

If there are not enough pages to complete Sets 3 and/or 4 in the usual manner, each remaining Set is generated from the last symbols of the two lines directly above the ones used for the preceding Set.

*How to generate the
Page Suffix*

01 If Set 3 was generated from a page numbered 13 or XIII, the suffix is (3).

02 If Set 3 was generated from a page numbered 17 or XVII, the suffix is (7).

03 If Set 3 was generated from a page counted from the one used for Set 2, the suffix is (C).

XXXX XXXX XXXX XXXX (C) 1500 (Q)

Suffix

Missing Pages

If pages are missing from the middle or end of the book, and the last two sets have been recorded as +++++ +++++, the correct suffix is (C).

If Sets 1 and 2 were recorded as +++++ +++++ because pages were missing from the beginning of the book, and Sets 3 and 4 were recorded as +++++ +++++ because the book contains no recto marked '13', '17', 'XIII', or 'XVII', the appropriate suffix is also (C).

Not enough Pages

If Sets 3 and/or 4 were generated from the last symbols of the lines directly above those used for the preceding Set, the correct suffix is (C).

*Pages used/
No Letterpress*

If neither page 13, 17, XII, nor XVII could be used for Set 3—because they contained no letterpress, or because they were already used for previous Sets—the pages are counted. The correct suffix in this case is (C).

Damaged pages

If page 13/17/XIII/XVII is so badly damaged that Set 3 is recorded as +++++, the suffix is nevertheless given as (3) for page 13/XIII and (7) for page 17/XVII.

This remains the case even if Set 4 is equally damaged and given as +++++.

Date

No Date given

If no date is given in the work, the date is reconstructed and the case marked with the suffix (Q). If reconstruction only yields a date range, the latest date of the range is used.

Partial Date

If the date is only partially given, for instance because a symbol was omitted in print, the date is reconstructed as above and marked with the suffix (Q). The same applies to dates which are only partially visible due to damage.

Century Ranges

If information is so sparse that only a century range can be reconstructed for the publication date, the date transcribed is the last year of that century, followed by the suffix (Q). For the sixteenth century this would mean *1599 (Q)*.

Composite Formats

If a date is given in a composite format, suffixes are combined to reflect this. A date given as *MD45* would be transcribed as *1545 (RA)*, *15XLV* as *1545 (AR)*, etc.

Multiple Formats

If the date is given more than once in the book and appears in multiple formats, the suffix is chosen according to the hierarchy presented on page 2. If the date is given in Arabic numerals anywhere in the book, it is transcribed with the suffix (A). If there is no arabic numeral for the date but it appears in Roman numerals somewhere in the book, the suffix is (R). If there is neither, the date is taken from the textual spelling, suffix (T). A Chronogram (C) is only used if the date is given in no other way.

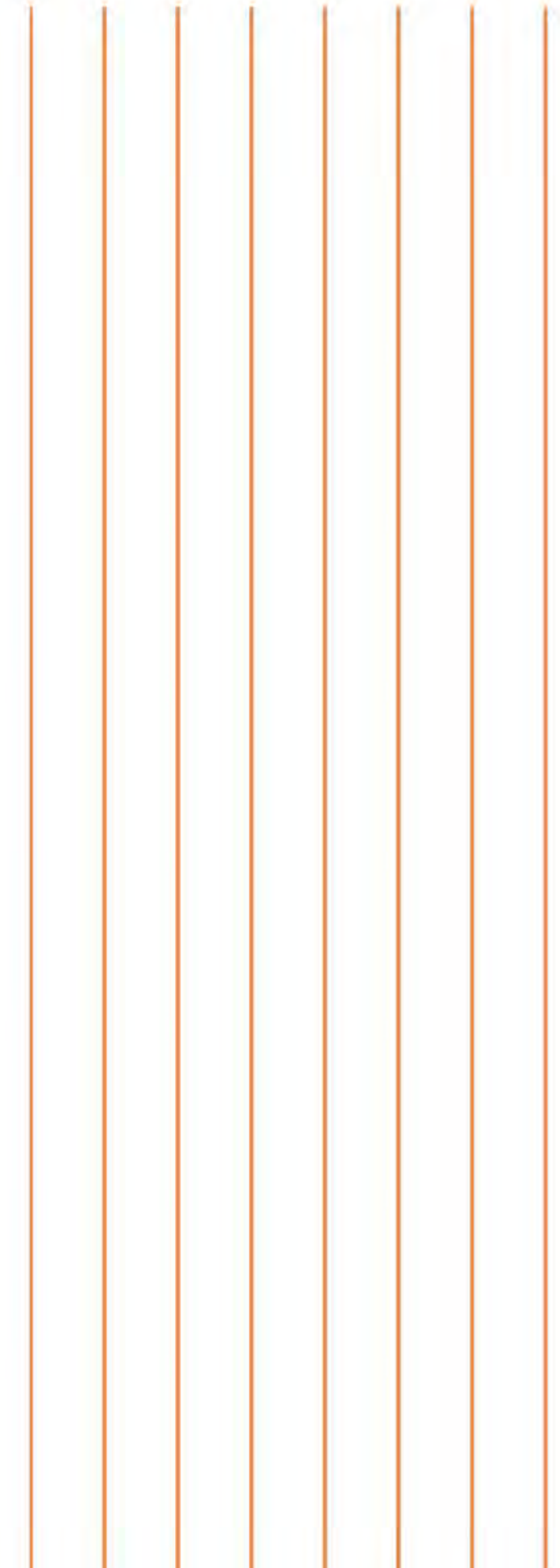
Conflicting Dates

If the work gives two conflicting dates for publication, the later of the two is used for the fingerprint. Only if there is absolute certainty – through context or further sources – that the later of the dates is an error should the earlier date be used.

*The fingerprint of this
booklet* is:*

1312 d.he e.s- toth (3) 2019 (A)

Victoria Beyer
2019



Max Planck Institute for the History of Science

Preprints since 2014 (a full list can be found at our website)

- 454 Klaus Geus and Mark Geller (eds.) **Esoteric Knowledge in Antiquity** (TOPOI - Dahlem Seminar for the History of Ancient Sciences Vol. II)
- 455 Carola Sachse **Grundlagenforschung. Zur Historisierung eines wissenschaftspolitischen Ordnungsprinzips am Beispiel der Max-Planck-Gesellschaft (1945–1970)**
- 456 David E. Rowe and Robert Schulmann **General Relativity in the Context of Weimar Culture**
- 457 F. Jamil Ragep **From Tūn to Turun: The Twists and Turns of the Tūsī-Couple**
- 458 Pietro Daniel Omodeo **Efemeridi e critica all'astrologia tra filosofia naturale ed etica: La contesa tra Benedetti e Altavilla nel tardo Rinascimento torinese**
- 459 Simone Mammola **Il problema della grandezza della terra e dell'acqua negli scritti di Alessandro Piccolomini, Antonio Berga e G. B. Benedetti e la progressiva dissoluzione della cosmologia delle sfere elementari nel secondo '500**
- 460 Stefano Bordoni **Unexpected Convergence between Science and Philosophy: A debate on determinism in France around 1880**
- 461 Angelo Baracca **Subalternity vs. Hegemony – Cuba's Unique Way of Overcoming Subalternity through the Development of Science**
- 462 Eric Hounshell & Daniel Midena **"Historicizing Big Data" Conference, MPIWG, October 31 – November 2, 2013 (Report)**
- 463 Dieter Suisky **Emilie Du Châtelet und Leonhard Euler über die Rolle von Hypothesen. Zur nach-Newtonschen Entwicklung der Methodologie**
- 464 Irina Tupikova **Ptolemy's Circumference of the Earth** (TOPOI - Towards a Historical Epistemology of Space)
- 465 Irina Tupikova, Matthias Schemmel, Klaus Geus **Travelling along the Silk Road: A new interpretation of Ptolemy's coordinates**
- 466 Fernando Vidal and Nélia Dias **The Endangerment Sensibility**
- 467 Carl H. Meyer & Günter Schwarz **The Theory of Nuclear Explosives That Heisenberg Did not Present to the German Military**
- 468 William G. Boltz and Matthias Schemmel **Theoretical Reflections on Elementary Actions and Instrumental Practices: The Example of the Mohist Canon** (TOPOI - Towards a Historical Epistemology of Space)
- 469 Dominic Olariu **The Misfortune of Philippus de Lignamine's Herbal or New Research Perspectives in Herbal Illustrations From an Iconological Point of View**
- 470 Fidel Castro Díaz-Balart **On the Development of Nuclear Physics in Cuba**
- 471 Manfred D. Laubichler and Jürgen Renn **Extended Evolution**
- 472 John R. R. Christie **Chemistry through the 'Two Revolutions': Chemical Glasgow and its Chemical Entrepreneurs, 1760-1860**
- 473 Christoph Lehner, Helge Wendt **Mechanik in der Querelle des Anciens et des Modernes**
- 474 N. Bulatovic, B. Saquet, M. Schlender, D. Wintergrün, F. Sander **Digital Scrapbook – can we enable interlinked and recursive knowledge equilibrium?**
- 475 Dirk Wintergrün, Jürgen Renn, Roberto Lalli, Manfred Laubichler, Matteo Valleriani **Netzwerke als Wissensspeicher**
- 476 Wolfgang Lefèvre **„Das Ende der Naturgeschichte“ neu verhandelt**
- 477 Martin Fechner **Kommunikation von Wissenschaft in der Neuzeit: Vom Labor in die Öffentlichkeit**
- 478 Alexander Blum, Jürgen Renn, Matthias Schemmel **Experience and Representation in Modern Physics: The Reshaping of Space** (TOPOI - Towards a Historical Epistemology of Space)
- 479 Carola Sachse **Die Max-Planck-Gesellschaft und die Pugwash Conferences on Science and World Affairs (1955–1984)**
- 480 Yvonne Fourès-Bruhat **Existence theorem for certain systems of nonlinear partial differential equations**
- 481 Thomas Morel, Giuditta Parolini, Cesare Pastorino (eds.) **The Making of Useful Knowledge**

- 482 Wolfgang Gebhardt Erich Kretschmann. **The Life of a Theoretical Physicist in Difficult Times**
- 483 Elena Serrano **Spreading the Revolution: Guyton's Fumigating Machine in Spain. Politics, Technology, and Material Culture (1796–1808)**
- 484 Jenny Bangham, Judith Kaplan (eds.) **Invisibility and Labour in the Human Sciences**
- 485 Dieter Hoffman, Ingo Peschel (eds.) **Man möchte ja zu seinem Fach etwas beitragen**
- 486 Elisabeth Hsu, Chee Han Lim **Enskilment into the Environment: the *Yijin jing* Worlds of *Jin* and *Qi***
- 487 Jens Høyrup **Archimedes: Knowledge and Lore from Latin Antiquity to the Outgoing European Renaissance**
- 488 Jens Høyrup **Otto Neugebauer and the Exploration of Ancient Near Eastern Mathematics**
- 489 Matteo Valleriani, Yifat-Sara Pearl, Liron Ben Arzi (eds.) **Images Don't Lie(?)**
- 490 Frank W. Stahnisch (ed.) **Émigré Psychiatrists, Psychologists, and Cognitive Scientists in North America since the Second World War**
- 491 María Sánchez Colina, Angelo Baracca, Carlos Cabal Mirabal, Arbelio Pentón Madrigal, Jürgen Renn, Helge Wendt (eds.) **Historia de la física en Cuba (siglo XX)**
- 492 Matthias Schemmel **Everyday Language and Technical Terminology: Reflective Abstractions in the Long-term History of Spatial Terms**
- 493 Barbara Wolff „Derartige kolossale Opfer ...“ **Der Nobelpreis für Physik für das Jahr 1921 – was geschah mit dem Preisgeld?**
- 494 Thomas Horst **The Reception of Cosmography in Vienna: Georg von Peuerbach, Johannes Regiomontanus, and Sebastian Binderlius**
- 495 Markus Asper **Science Writing and Its Settings: Some Ancient Greek Modes**
- 496 Dagmar Schäfer, Zhao Lu, and Michael Lackner (eds.) **Accounting for Uncertainty: Prediction and Planning in Asian History**
- 497 Joachim Nettelbeck **Verwalten von Wissenschaft, eine Kunst**
- 498 Carla Rodrigues Almeida **Stellar equilibrium vs. gravitational collapse**
- 499 Victoria Beyer **How to Generate a Fingerprint**