Manuscript genetics and Perl

or, How I made sense of 101011 manuscripts

Tara L Andrews
Departmental Lecturer
Faculty of History, University of Oxford
tla@mit.edu
• How do we reconstruct the past?
• How do we reconstruct the past?

• Reading what they wrote is a good start.
• How do we reconstruct the past?
• Reading what they wrote is a good start.
• But their originals don’t tend to survive.
• How do we reconstruct the past?

• Reading what they wrote is a good start.

• But their originals don’t tend to survive.

• So we have to read what someone else read of what someone else read of what someone else read of what they wrote.
• How do we reconstruct the past?
• Reading what they wrote is a good start.
• But their originals don’t tend to survive.
• So we have to read what someone else read of what someone else read of what someone else read of what they wrote.
• So we have to compare the versions.
This is not a talk about Perl.
This is about a cool use of Perl.
"The collation of manuscripts requires the infuriating accuracy of a pedant and the obsessive stamina of an idiot. It is therefore an ideal task for a computer."

—Peter Robinson, “Collation and Textual Criticism”, LLC vol. 4 no. 2, 1989
Looking at app App452

Context:  երբեմն երեխի համար եռանդի տեսանելու կարգի հոդված, ի հայտ է տալիս իր գրիմնագր 2. Following context:

<table>
<thead>
<tr>
<th>HD</th>
<th>HW</th>
<th>HX</th>
<th>HY</th>
<th>HF</th>
<th>HI</th>
<th>HJ</th>
<th>HK</th>
<th>HZ</th>
<th>NV</th>
<th>HA</th>
<th>HB</th>
<th>H0</th>
</tr>
</thead>
</table>

Available readings:
1: Երեխի համար եռանդի տեսանելու կարգի հոդված (HJ #K #B #I #A #D #J #O #V #Y #Z )
2: Երեխի համար եռանդի տեսանելու կարգի հոդված (HX )
3: Երեխի համար եռանդի տեսանելու կարգի հոդված (HF )
6: Երեխի համար եռանդի տեսանելու կարգի հոդված (HW )

Manuscript details:
3: Witness(es) #X contain a, after Երեխի համար եռանդի տեսա
5: Witness(es) #F contain a, after Երեխի համար եռա
--- h

Available commands:
accept (#)
accept detail (#)
subst detail (#) (value)
spelling (# alternate) (# canonical)
orth (# alternate) (# canonical)
emend (comment text)
note (note text)
next
help
save
quit
---
Manuscript genetics

- Not invented here
- But pretty cool anyway
- A way to reconstruct the past, a guide to good editing decisions
A helpful stemma
A helpful stemma

Thursday 20 March 14
How to make a stemma
How to make a stemma

• Know the answer in advance
How to make a stemma

• Know the answer in advance

• Assume a ‘right’ text that you can compare the existing ‘wrong’ texts to
How to make a stemma

• Know the answer in advance

• Assume a ‘right’ text that you can compare the existing ‘wrong’ texts to

• Assume that the wrong is the child of the right
How to make a stemma

• Know the answer in advance
• Assume a ‘right’ text that you can compare the existing ‘wrong’ texts to
• Assume that the wrong is the child of the right
• Doesn’t really work for texts like mine.
To get the answer, I have to have the answer.
Wasn’t this supposed to be labor-saving?
How to make a stemma?

• I have way too many manuscripts to go through them by hand

• I don’t know in advance what the ‘right’ text is

• So I can’t say what is ‘wrong’

• And I don’t believe that all change is error
Pretend the manuscripts are organisms.
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thursday 20 March 14
Choosing your data
Choosing your data

• Not all variations are equal
  • Some variations don’t come from copying
  • Some variations will get ‘corrected by a later scribe
  • Some variations are singularities, so not very useful for judging relationships
Choosing your data

- Not all variations are equal
  - Some variations don’t come from copying
  - Some variations will get ‘corrected by a later scribe
  - Some variations are singularities, so not very useful for judging relationships
- But how choosy do you need to be?
A little experiment
A little experiment

• Use all the data, corrected for spelling variation.
A little experiment

- Use all the data, corrected for spelling variation.
- Get rid of a common scribal variation (-🏞 at the end of a word.)
A little experiment

• Use all the data, corrected for spelling variation.

• Get rid of a common scribal variation (-unakan at the end of a word.

• Only use data where there are at least two witnesses for each variant.
A little experiment

• Use all the data, corrected for spelling variation.

• Get rid of a common scribal variation (-û at the end of a word.

• Only use data where there are at least two witnesses for each variant.

• See if there is any difference.
1. All readings

2. All readings, strict pairs only

3. Selected readings

4. Selected readings, strict pairs only
Sampling techniques
Sampling techniques

- Three samples, one from each book of the Chronicle
Sampling techniques

• Three samples, one from each book of the Chronicle
• 5,000 words in total for an 80,000-word text
Sampling techniques

• Three samples, one from each book of the Chronicle
• 5,000 words in total for an 80,000-word text
• Needed to give a single answer
Sampling techniques

• Three samples, one from each book of the Chronicle
• 5,000 words in total for an 80,000-word text
• Needed to give a single answer
• Needed to give consistent answers across samples
Possible trees from the text of Book 2

Possible trees from the text of the prophecy
Possible trees from the text of Book 2

Possible trees from the text of the prophecy
Possible trees from the text of Book 2

Possible trees from the text of the prophecy
“In this year there appeared a hairy star [i.e. comet] moving toward the west in the month of Areg and its tail gave off a little of its light; and it was there for 15 days and then became invisible to viewers. O kind and friendly brothers, pardon the lack of words, because the exemplar ended here because the text was not finished. Remember to Jesus Christ the sinful and foolish scribe Khachik, and my parents, and you shall also be remembered by the same Lord.
Non-fragmentary manuscripts omitted:
- Paris 191, 200
- Jerusalem 3651
- Matenadaran 2855, 2899, 3380, 6605, 8159, 8232, 8894
- Rome 25
- Vienna 243, 246
How academics see their jobs

- Data, facts, figures, observations, etc.
- Mighty thoughts of the academic
- Scholarship, success, riches
How academics see their jobs

Data, facts, figures, observations, etc.

Mighty thoughts of the academic

Scholarship, success, riches

*warning: gross oversimplification and/or wishful thinking*
How academics see computers in their jobs

- Data, facts, figures, observations, etc.
- Computer program
- Scholarship, success, riches
- Irrelevance and penury
How academics should see computers

Data, facts, figures, observations, etc.

Computer program

Mighty thoughts of the academic

Scholarship, success, riches
Non-fragmentary manuscripts omitted:
Paris 191, 200
Jerusalem 3651
Matenadaran 2855, 2899, 3380,
6605, 8159, 8232, 8894
Rome 25
Vienna 243, 246

*Based on Jerusalem mss. 1051, 1107
Non-fragmentary manuscripts omitted:

Paris 191, 200
Jerusalem 3651
Matenadaran 2855, 2899, 3380, 6805, 8159, 8232, 8894
Rome 25
Vienna 243, 246

*Based on Jerusalem mss. 1051, 1107
Non-fragmentary manuscripts omitted:
Paris 191, 200
Jerusalem 3651
Matenadaran 2855, 2899, 3380,
6605, 8159, 8232, 8894
Rome 25
Vienna 243, 246

*Based on Jerusalem mss. 1051, 1107
Non-fragmentary manuscripts omitted:
Paris 191, 200
Jerusalem 3651
Matenadaran 2855, 2899, 3380, 6605, 8159, 8232, 8894
Rome 25
Vienna 243, 246

*Based on Jerusalem mss. 1051, 1107
Thank you
Obrigada
[shnorhakalutiu]