Classification of New Testament Manuscripts: potential computational insights

Sophie Robert-Hayek

University of Lorraine
Écritures - *Maison des Sciences de l'Homme*

12/09/2024 - Aarhus

Motivation

Outline

Classification of New Testament manuscripts

Classification of manuscripts allow:

- Reduction of the mass of considered manuscripts for the edition of the text;
- Better understanding of the transmission of the New Testament text.

But New Testament present many challenges:

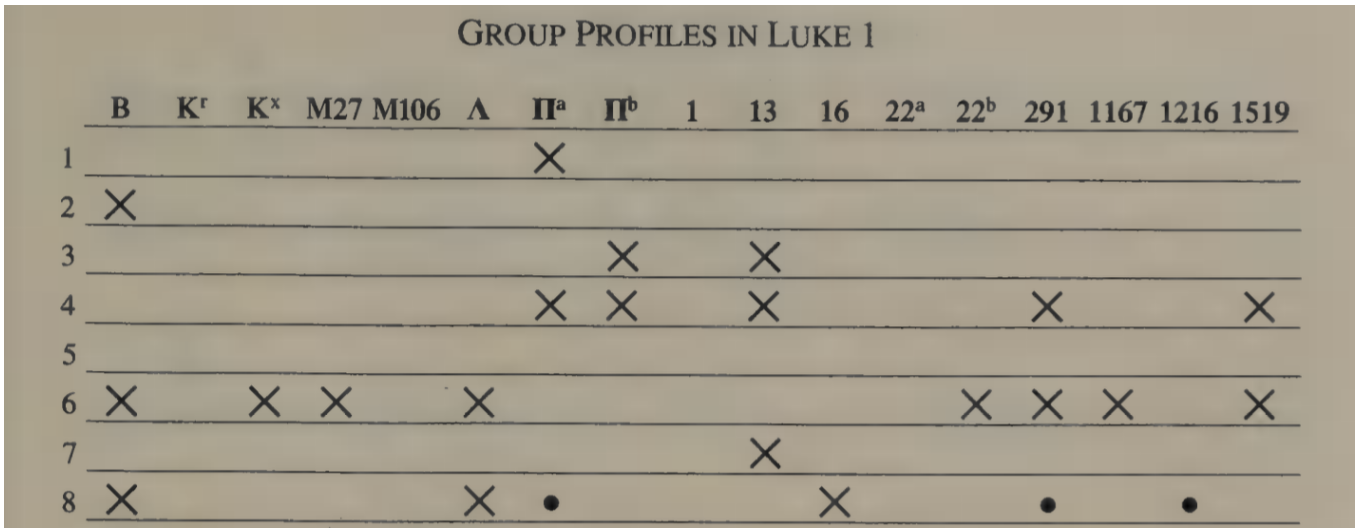
1. Severe contamination of traditions;
2. Very large volume of manuscripts, rendering systematic collation impossible.

Current classification approaches

- *Text-type*
- *Categories* (Alands): check the presence/absence of readings to **measure proximity to Byzantine manuscripts** (*Test Tellen*);
- *Quantitative Analysis* (Colwell and Tune): compute agreement rates **between manuscripts and major class representatives**;
- *Claremont Profiling Method*: compute absence/presence of **readings compared to the *Textus Receptus***.

The Claremont profiling method

Example Claremont profile for Luke 10:



Each manuscript is represented as a binary vector, **very practical for computational approaches!**

Potential computational insights

Outline

Computational insights

Computational approaches applied to humanities problem can help us:

- **Visualize data interactively** for better understanding;

Computational insights

Computational approaches applied to humanities problem can help us:

- **Visualize data interactively** for better understanding;
- **Systematize** existing manually found established findings;

Computational insights

Computational approaches applied to humanities problem can help us:

- **Visualize data interactively** for better understanding;
- **Systematize** existing manually found established findings;
- Suggest **new methods relying on heavier computation** than what can be done manually (with different parameters for example).

Computational insights

Computational approaches applied to humanities problem can help us:

- **Visualize data interactively** for better understanding;
- **Systematize** existing manually found established findings;
- Suggest **new methods relying on heavier computation** than what can be done manually (with different parameters for example).

Computational approaches to **manuscript studies** have the potential to:

1. Analyze the relevance of existing methods (Wisse, Aland categories, Quantitative Analysis...);
2. Suggest new approaches;
3. Better understand manuscript relationships and transmission of the New Testament manuscripts.

Transforming the text into numbers

Numerical data is required to rely on **computational approaches** we need to **compute the distance between texts**.

Transforming the text into numbers

Numerical data is required to rely on **computational approaches** we need to **compute the distance between texts**.

Wisse approach:
Reading per reading comparison:

$$\begin{array}{lcccccc} MS1: & 1 & 0 & 1 & 1 & 1 & 0 \\ MS2: & 1 & 0 & 0 & 1 & 0 & 0 \end{array}$$

Calculation: Number of differences: 3

$$d_H(A, B) = 3$$

Transforming the text into numbers

New possible approach:

***Distances between full text*:**

- MS 1: Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ Χριστοῦ
- MS 2: Ἀρχὴ τοῦ εὐαγγελίου Ἰησοῦ

Jaccard Distance Formula:

$$d_J(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

Calculation:

$$|A \cap B| = 4 \quad |A \cup B| = 5$$

$$d_J(A,B) = 1 - \frac{4}{5} = 0.2$$

Dimensionality reduction

Dimensionality reduction is the process of mapping high-dimensional data into three dimensions while **preserving pairwise distances** as much as possible.

For manuscript studies:

- Manuscripts that are **close together** in the projection are **close** to each other in text;
- Manuscripts that are **far together** in the projection are **far** to each other in text.

Computational insights: visualization

Visualization:

- Interactive visualization of the Wisse profiles;
- 3D projections of the manuscripts according to their similarity, **to visualize the distances between the manuscripts**;
- Display of the similarity between manuscripts, also **as the verse level**.

Textual clustering

Clustering is the process of grouping a set of objects into clusters, where objects within the same cluster **are more similar to each other** than to those in different clusters.

The process consists in:

1. Transforming the texts into numbers;
2. Measuring the distance between the texts;
3. Applying an algorithm that group together texts that are similar to each other.

Computational insights: new clustering

Computational insights:

- Perform clustering using Wisse profiles to **check its systematicity**;
- Perform clustering using **the whole content of the text**;
- Check the **homogeneity** between automatic clustering and existing knowledge on manuscripts.

Next steps:

- Derive new knowledge on textual relationships;
- Taking into account a wide array of Gospel chapters.

Live demonstration

Outline

Demonstration :-)

Prototype Web application to apply the different methods on:

- 18 manuscripts;
- Transcriptions taken from the Münster NTVMR;
- Chapter 10 of Luke.