

Analysis of Language Resources

Eighth Lecture
Hiroyuki Akama

From *Variables* to *Variants*

- *We propose here...*
- A quantitative model, which uses lexical occurrence and co-occurrence data to analyze the similarity and the dissimilarity among a set of parallel text collections and to presume lateral or mutual dependence in their relationship.
- A set of web-based software tools that will facilitate latent context detection and the precise matching of variant texts. (Miyake et al., 2002~2004)



Ex. Research on the Synoptic Gospels

4-VAR System

t_i : text

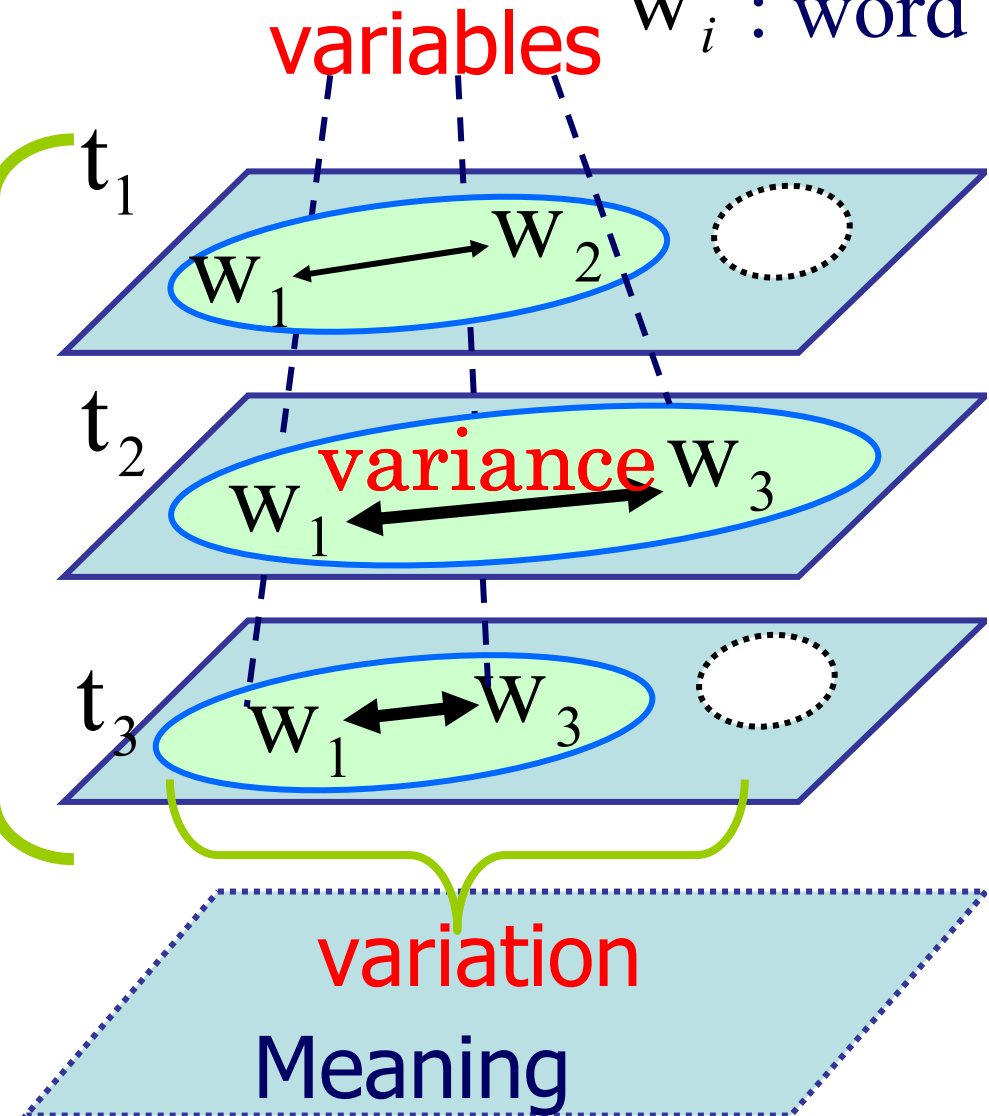
w_i : word

① **variable**
for words instances

② **variance**
for lexical data

③ **variant**
for text creation

④ **variation**
for latent applicability



From the researches made in Akama Laboratory

- Biblical Research:
- *The Frame Problem in Text Analysis*
- *Development of NLP based Synoptic Software for Text Analysis*
 - (With Miyake, Nakagawa and Makoshi)
- French Grammar:
- *Probabilistic Language Processing in the Form of a Decision Tree* → Next time
 - (With, Shimizu and Shimizu)

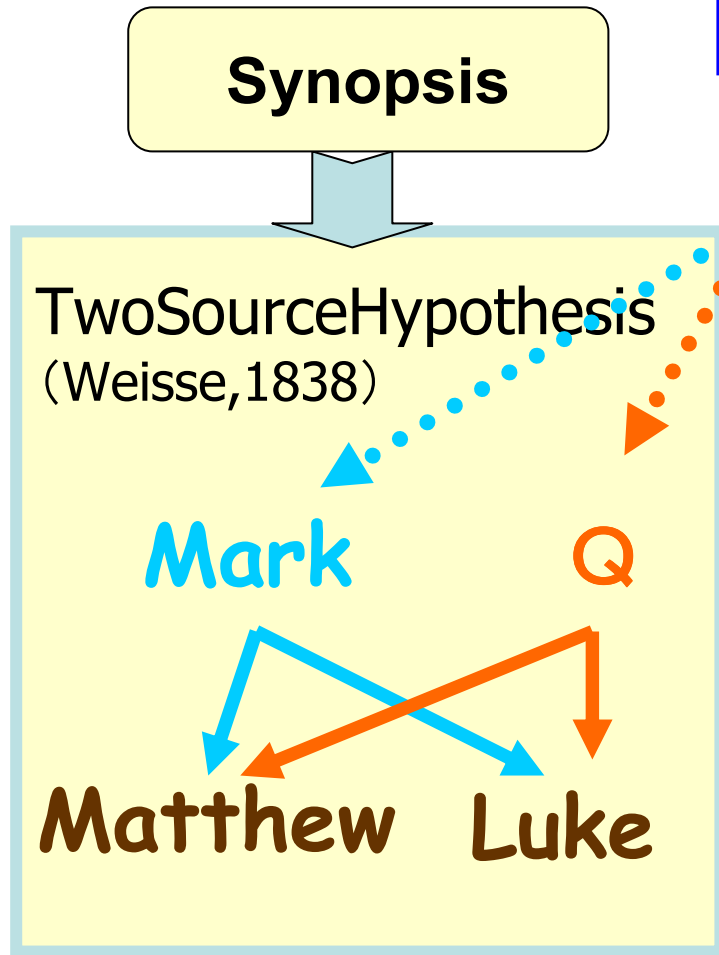
Examples of the computational humanities technology --Taken from the **Biblical research**

Tele-Synopsis

Tele-COEX

[illegible]

Synoptic Problem Solution with frames

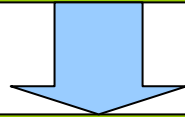


	Matthew	Mark	Luke
1	Matthew 1:1-11	Mark 1:1-11	Luke 1:1-11
2	Matthew 1:12-23	Mark 1:12-15	Luke 1:12-17
3	Matthew 2:1-12	Mark 1:16-20	Luke 1:18-25
4	Matthew 2:13-23	Mark 1:21-28	Luke 1:26-38
5	Matthew 3:1-12	Mark 1:29-34	Luke 1:39-45
6	Matthew 3:13-17	Mark 1:35-45	Luke 1:46-55
7	Matthew 4:1-11	Mark 1:46-54	Luke 1:56-66
8	Matthew 4:12-23	Mark 1:55-64	Luke 1:67-79
9	Matthew 5:1-48	Mark 2:1-12	Luke 2:1-14
10	Matthew 6:1-16	Mark 2:13-17	Luke 2:15-20
11	Matthew 6:17-34	Mark 2:18-22	Luke 2:21-28
12	Matthew 7:1-12	Mark 2:23-28	Luke 2:29-34
13	Matthew 7:13-27	Mark 3:1-6	Luke 2:35-40
14	Matthew 8:1-13	Mark 3:7-12	Luke 2:41-51
15	Matthew 8:14-17	Mark 3:13-19	Luke 2:52-58
16	Matthew 8:18-26	Mark 3:20-27	Luke 2:59-64
17	Matthew 9:1-13	Mark 3:28-35	Luke 2:65-70
18	Matthew 9:14-17	Mark 3:36-39	Luke 2:71-76
19	Matthew 9:18-26	Mark 3:40-45	Luke 2:77-82
20	Matthew 9:27-34	Mark 3:46-51	Luke 2:83-88
21	Matthew 9:35-38	Mark 3:52-57	Luke 2:89-94
22	Matthew 10:1-42	Mark 3:58-62	Luke 2:95-100
23	Matthew 11:1-19	Mark 4:1-11	Luke 2:101-106
24	Matthew 11:20-24	Mark 4:12-20	Luke 2:107-112
25	Matthew 11:25-30	Mark 4:21-25	Luke 2:113-118
26	Matthew 12:1-14	Mark 4:26-34	Luke 2:119-124
27	Matthew 12:15-21	Mark 4:35-41	Luke 2:125-130
28	Matthew 12:22-31	Mark 4:42-51	Luke 2:131-136
29	Matthew 12:32-40	Mark 4:52-54	Luke 2:137-142
30	Matthew 12:41-50	Mark 4:55-64	Luke 2:143-148
31	Matthew 13:1-17	Mark 4:65-71	Luke 2:149-154
32	Matthew 13:18-23	Mark 4:72-81	Luke 2:155-160
33	Matthew 13:24-30	Mark 4:82-91	Luke 2:161-166
34	Matthew 13:31-35	Mark 4:92-101	Luke 2:167-172
35	Matthew 13:36-43	Mark 4:102-111	Luke 2:173-178
36	Matthew 13:44-52	Mark 4:112-121	Luke 2:179-184
37	Matthew 13:53-58	Mark 4:122-131	Luke 2:185-190
38	Matthew 14:1-12	Mark 4:132-141	Luke 2:191-196
39	Matthew 14:13-21	Mark 4:142-151	Luke 2:197-202
40	Matthew 14:22-36	Mark 4:152-161	Luke 2:203-208
41	Matthew 15:1-20	Mark 4:162-171	Luke 2:209-214
42	Matthew 15:21-28	Mark 4:172-181	Luke 2:215-220
43	Matthew 15:29-34	Mark 4:182-191	Luke 2:221-226
44	Matthew 15:35-38	Mark 4:192-201	Luke 2:227-232
45	Matthew 16:1-12	Mark 4:202-211	Luke 2:233-238
46	Matthew 16:13-28	Mark 4:212-221	Luke 2:239-244
47	Matthew 17:1-13	Mark 4:222-231	Luke 2:245-250
48	Matthew 17:14-21	Mark 4:232-241	Luke 2:251-256
49	Matthew 17:22-27	Mark 4:242-251	Luke 2:257-262
50	Matthew 18:1-11	Mark 4:252-261	Luke 2:263-268
51	Matthew 18:12-14	Mark 4:262-271	Luke 2:269-274
52	Matthew 18:15-20	Mark 4:272-281	Luke 2:275-280
53	Matthew 18:21-35	Mark 4:282-291	Luke 2:281-286
54	Matthew 19:1-12	Mark 4:292-301	Luke 2:287-292
55	Matthew 19:13-15	Mark 4:302-311	Luke 2:293-298
56	Matthew 19:16-22	Mark 4:312-321	Luke 2:299-304
57	Matthew 19:23-30	Mark 4:322-331	Luke 2:305-310
58	Matthew 19:31-39	Mark 4:332-341	Luke 2:311-316
59	Matthew 19:40-42	Mark 4:342-351	Luke 2:317-322
60	Matthew 19:43-45	Mark 4:352-361	Luke 2:323-328
61	Matthew 19:46-48	Mark 4:362-371	Luke 2:329-334
62	Matthew 19:49-51	Mark 4:372-381	Luke 2:335-340
63	Matthew 19:52-54	Mark 4:382-391	Luke 2:341-346
64	Matthew 19:55-57	Mark 4:392-401	Luke 2:347-352
65	Matthew 19:58-60	Mark 4:402-411	Luke 2:353-

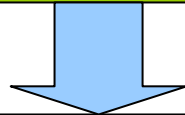
Feature Example of the *Parallel Synoptic Table*

Quantitative Analysis

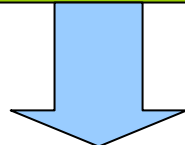
**Count word frequencies
by a Parallel table**



**Distribute to 7 categories
by using Tele-Synopsis**

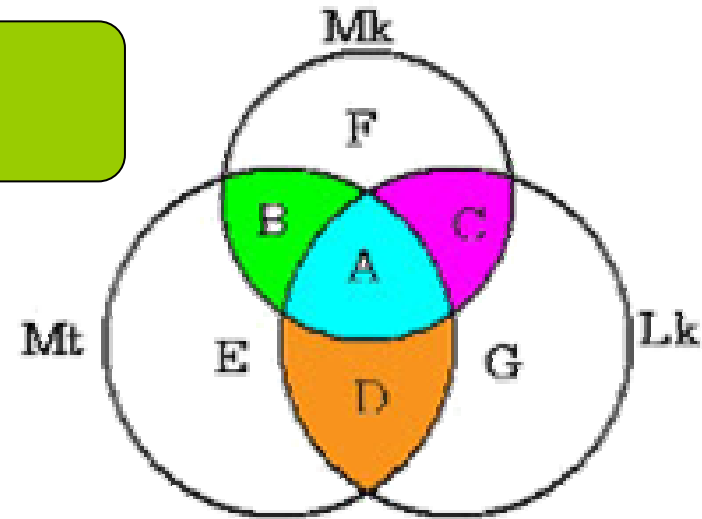


**Principal Component Analysis
(Varimax rotation)**



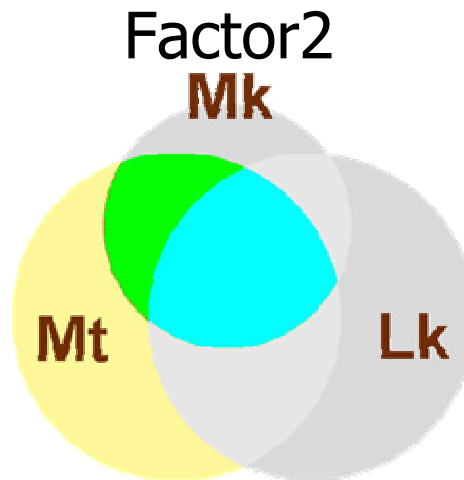
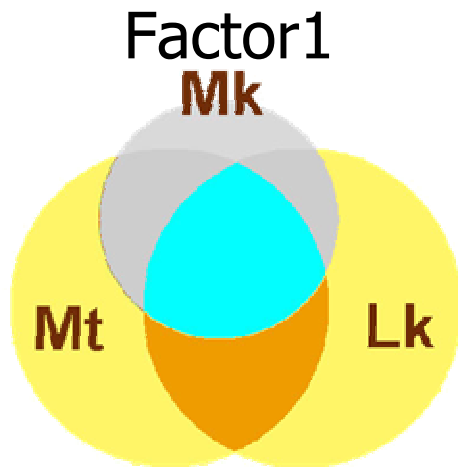
Correlation among 7 categories

Examine the hypothesis

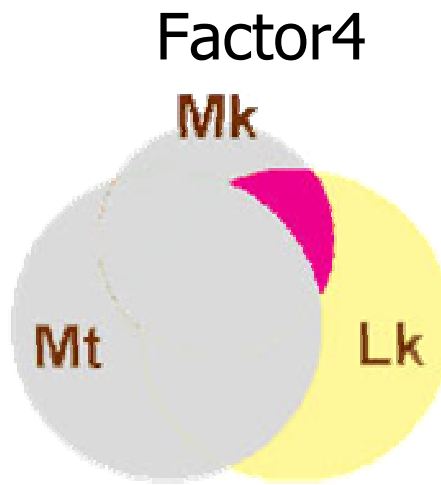
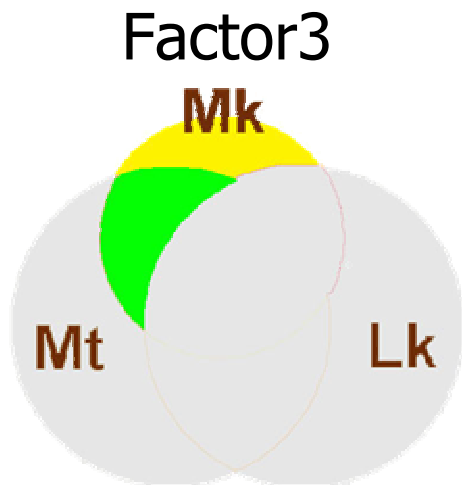


Mk: Mark
Mt: Matthew
Lk: Luke

Factor Pattern



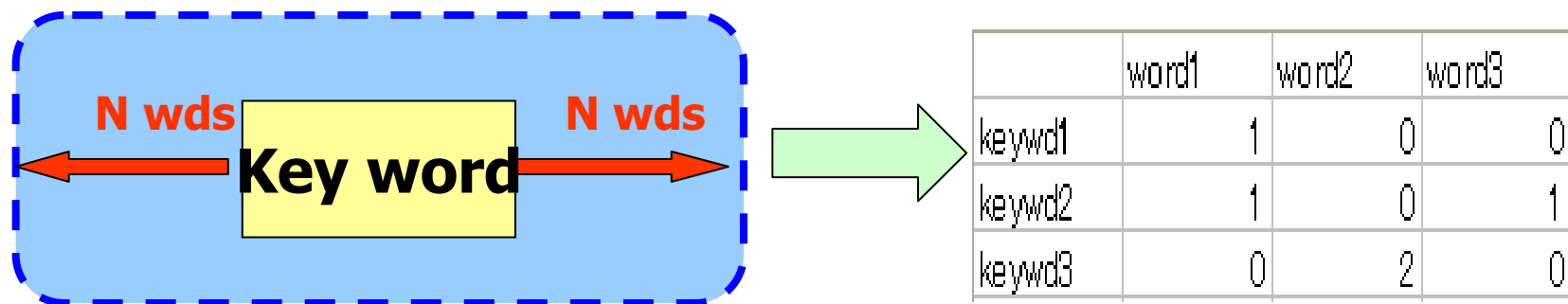
Rejection of
the Hypothesis



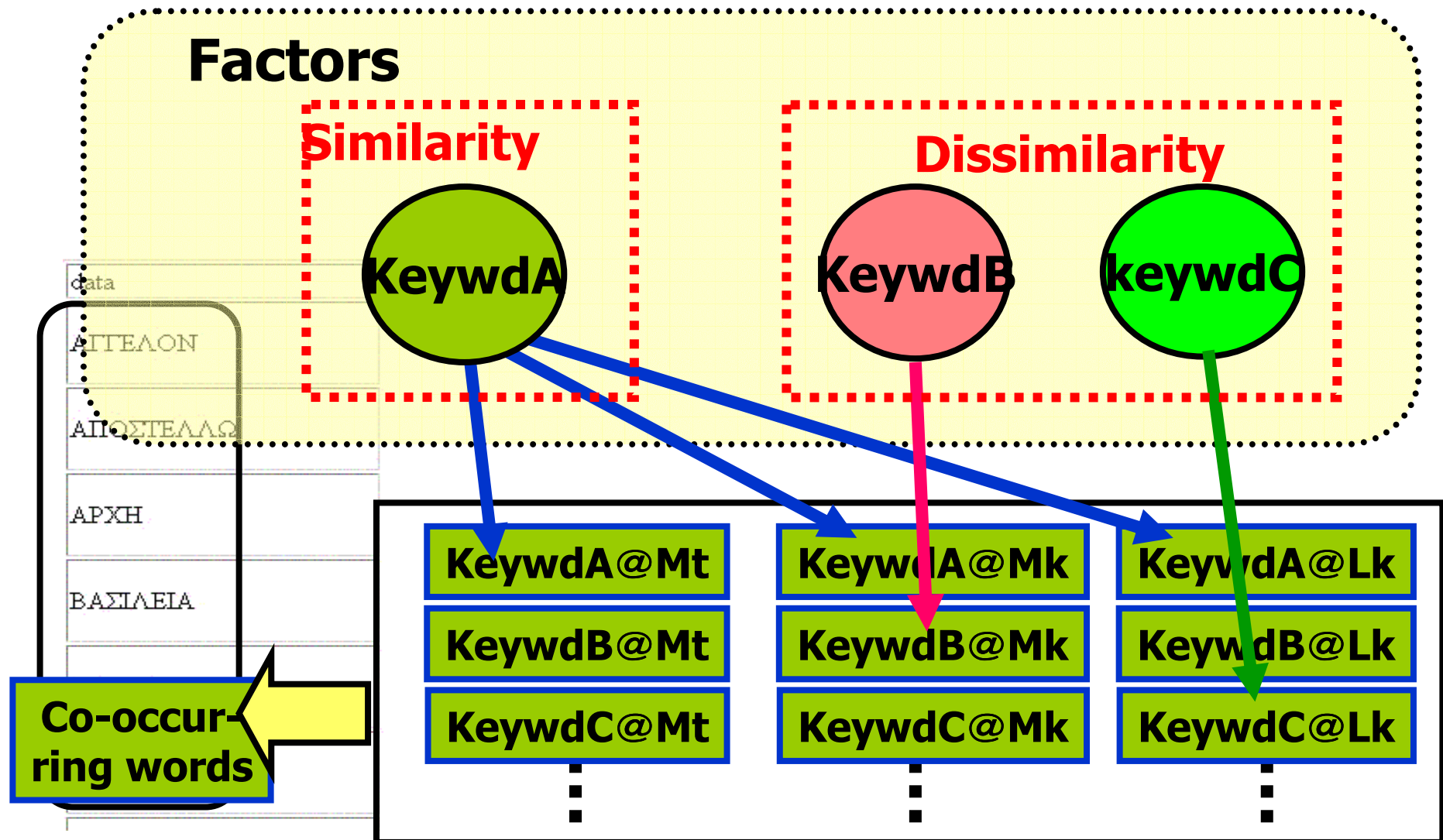
Mk: Mark
Mt: Matthew
Lk: Luke

Windowing Method

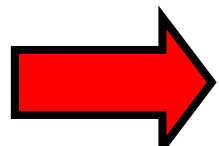
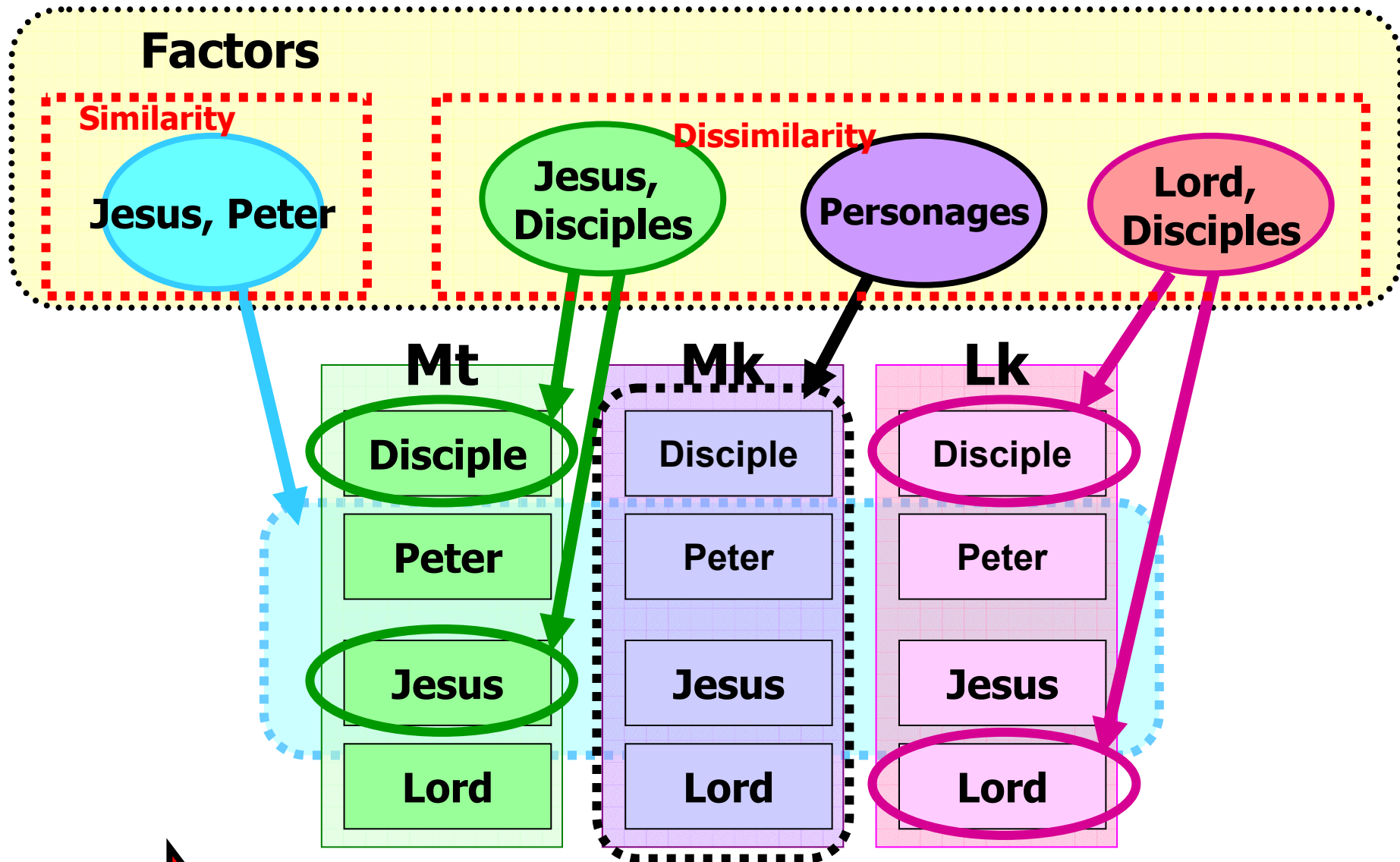
- While sliding through every variant text, the window in size n of Tele-COEX stops to get the information about the neighbor words inside the window when it encounters with the instance of the keyword.



Factor Analysis of Windowing Data



Result of Analysis



The features of the documents are extracted

Summary

- Quantitative analysis with computational Synopsis
 - Comparing the results with Synopsis
- Genealogical analysis
 - Text variants ; Notes of Saussure's Course, for example...
- NLP based text framing software
 - Powerful tool to break parallel text for genealogical analysis
- Knowledge frames
 - the synoptic problem with the frames