The Computed Synoptic Table —Tele-Synopsis for Biblical Research

Maki Miyake *(mmiyake@dp.hum.titech.ac.jp)*

Department of Human System Science, Tokyo Institute of Technology

Hiroyuki Akama *(akama@dp.hum.titech.ac.jp)*

Department of Human System Science, Tokyo Institute of Technology

Masanori Nakagawa
(nakagawa@nm.hum.titech.ac.jp)

Department of Human System Science, Tokyo Institute of Technology

Nobuyasu Makoshi *(makoshi@gsic.titech.ac.jp)*

Global Scientific Information Center, Tokyo Institute of Technology

I. Introduction

While over the last two centuries, 'the synoptic problem' has been one of the controversial subjects in the studies of the *New Testament*, only a few studies so far have attempted to give an objective, statistical explanation of the mutual relationships between the synoptic Gospels, Matthew, Mark and Luke (in abbreviation, Mt, Mk and Lk, respectively) (Conzelmann and Lindemann 45-53). Furthermore, even though a large number of studies have made various assumptions of their genealogical interdependence, there still seems to remain a lack of the computational humanities technology enabling the Gospel researchers to present valid arguments based on a huge amount of biblical text data. As the first step of our study, there is a need to develop some specific applications to automatically collect the thorough data of the lexical usage patterns from the electronic bible(Miyake, Akama, Sato and Nakagawa 2002), thus the web-based biblical software, named *Tele-Synopsis* (<http://nerva.dp.hum.titech.ac.jp/tele-synopsis/parallel>), is designed to gather information of the word usage under various conditions and to help further statistical approach to the origin of the variant texts.

II. Tele-Synopsis — Web-based biblical software

The basic concept design of *Tele-Synopsis* is founded upon the possibilities of natural language processing (NLP) for mediating Thesaurus creation and Conceptual mapping, dual problematic fields whose key concept is always cognition of 'frame' (Minsky; Winston 211-277). *Tele-Synopsis*, which allows us to manipulate lexical data of parallel and variant texts (Miyake, Akama, Sato, Nakagawa and Makoshi 2004), uses the NA27th version of the texts (Nestle-Aland) and for the parallels, the *Synopsis Quattuor Evangeliorum* by Kurt Aland, recognized as the most reliable parallel synoptic table (PST) to date. This system has a merit to make it possible for users to independently add and remove each sentence so as to customize their own synoptic table by changing the temporary segmentation of pericope, yet the challenges are still left on the optimum solutions available to the users, and so we need a sort of 'TextTiling' algorithm that allows us to break parallel texts into units the most suitable for biblical research.

III. Segmentation Problem

Although there are traditionally two types of synoptic tables covering a lost source called the 'Q' (Mt and Lk) and Mark (Mt, Lk and Mk) respectively, few trials have been done to produce synoptic tables treating other combinations of two Gospels, such as Mt and Mk, Lk and Mk. This kind of inexhaustiveness is due to the raison d'etre of the synoptic tables that is to consolidate *Two-Source Hypothesis*, according to which Mk and the 'Q' are the origins of quotations (Kloppenborg et al. and Reader). In addition, we have to note that the two traditional synoptic tables were solely made by using a *Form Criticism* which divided the texts into parts by the arbitrary unities coming from tradition or reduction. It can be recognized that the two traditional synoptic tables 'mesh' the world of the Gospels too roughly (as is the case for the Markan triptych table) or too finely (for the bilateral table of the 'Q'). As long as the problem of text segmentation remains unresolved, any experiment in quantitative text analysis will be still a long way from being realized. For our goal of the scientific examination of the Two-Source Hypothesis, we propose a new statistical method of generating the segmentation criteria of the synoptic Gospels, a sort of 'TextTiling' methodology enabling a computed synoptic table (CST) with an objective segmentation based on objective criteria.

IV. The Computed Synoptic Table(CST)

The computed synoptic tables (CST) are produced by using the algorithm called Synoptic Patch (Figure 1) that consists of the combination of 1) N-gram calculation, 2) Windowing data gathering and 3) TextTiling method.

1) Data from the n-gram model

We calculated for the 3 parallel texts (Mk,Mt,Lk) all the cases of n-gram models, thus made an exhaustive list of the instances where words co-occurred across texts. These overlaps were classified by the four combination patterns (D:Mt-Lk, C:Mk-Lk, B:Mk-Mt, A:Mk-Mt-Lk) (Figure 2), and the longest matched strings of words can be thought of as proofs of cross-citation. Having in view the occurrence probability of N-gram instances, we extracted the overall data under the condition of ($N>3$) because the significance of the bi-gram data is relatively low. This process will allow us to build a more objective synoptic table to replace the traditional one.

2) Data obtained by a windowing method

It is well-known that there has been in the realm of Information Retrieval (IR) remarkable progress owing to the elaboration of what we call vector space model or concept-based IR. This method, that consists of collecting the information about term i occurring n times in document j, allows us to identify a word (or a document) using a k-dimensional vector representation. Each entry of the vector corresponds to the frequency of each of k co-occurring words. Then the similarity between documents will be computed by the cosine of the angle between these vectors in a k-dimensional Euclidian space. Taking into consideration the principle that a context-sensitive word (or string of words) is categorized by the neighbor words appearing within a certain distance from it, we implemented some functions to set up a set of synchronized windows changing in size for each parallel n-gram instance (longest matched strings of words) to be centered in. The rule of the window operation for recording one by one and simultaneously in the parallel texts the frequency data of the co-occurring words is that each window must stop the extension if the border meets that of the previous (when moving leftward) or the next (when moving rightward) pericope.

3) Application of TextTiling

Synoptic Patch as a method of partitioning off the texts allows us to calculate at every step of the window extension the correlation coefficient between the word frequency vectors generated from each corresponding window instance. Before the extending operation, the cosine similarity value remains 1, but as different words are being distributed in the parallel setting, this value begins to decline and continues to fall down until another parallel N-gram instance is met in the window extension (cohesion score graph used in 'TextTiling' (Hearst 33-64)). However, in each pericope, there may be several instances of centered key strings (a series of the longest matching words) that are supposed to produce an overlap of windows and descending similarity curves, so that we computed at each word position the mean of the correlation coefficients obtained from all the pairs of parallel word vectors inside a pericope. The threshold is determined by us at 0.5 to properly resegment the periscope because the traditional synoptic tables with the three Gospels tends to include in each frame many divergent passages making the parallel word vectors nearly non-correlated or sometimes too highly correlated. That is why we fixed the segmentation point by using the threshold for the cohesion score graph instead of selecting, just as Hearst recommends it, the steepest part of the descending curve.

V. Result and Conclusion

The Synoptic Patch allows us to produce by fulfilling the identical criteria two remaining bilateral synoptic tables allocating Mk and Mt for one and Mk and Lt for the other. The index of difference between the traditional Synoptic Tables (ST) and the Computed Synoptic Table (CST) can be defined by the distribution of the words into the 7 categories as shown in Figure 2. The effects of the new combinations are clearly revealed by the diminution in quantity of some textual overlaps. The ratio of the common parts (A+B+C+D) is 60% in the PST and 42% in the CST (Figure 3). Figure 4 shows the drop in number of the words belonging to the categories A and D whose considerable weights would support the two source hypothesis. It cannot be denied that the new balance between the original parts E, F and G (increasing) and the common parts A+B+C+D (decreasing) will influence the verification regarding the historical formation of the synoptic Gospels. We can instinctively grasp the changing features of the parallels attachment by horizontally comparing the two tables in Figure 5. It will be left for the future investigations to completely evaluate the effi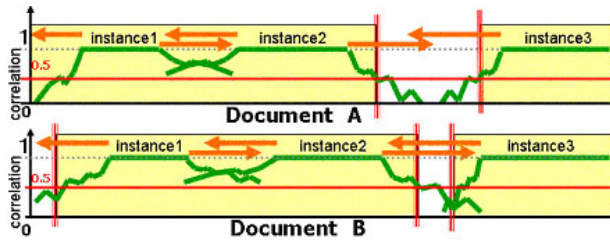cacy of the CST. Further information will be obtained at : <http://nerva.dp.hum.titech.ac.jp/tele-synopsis/synopsis.html> .
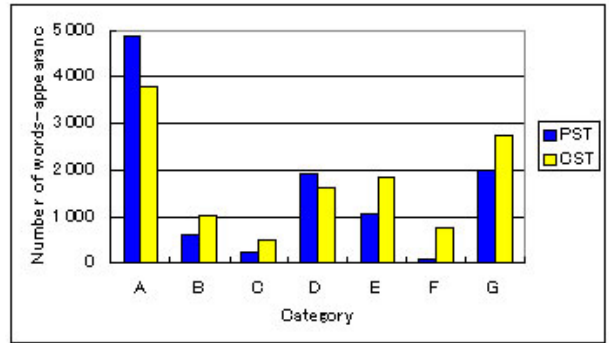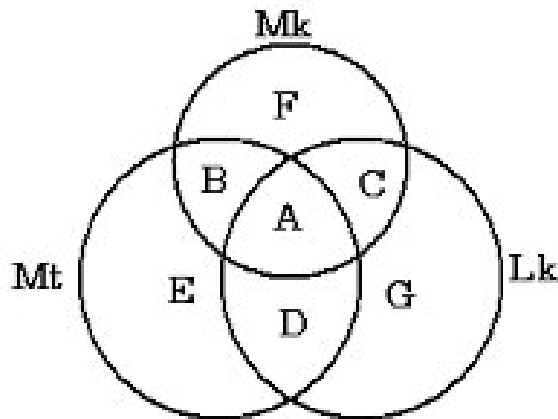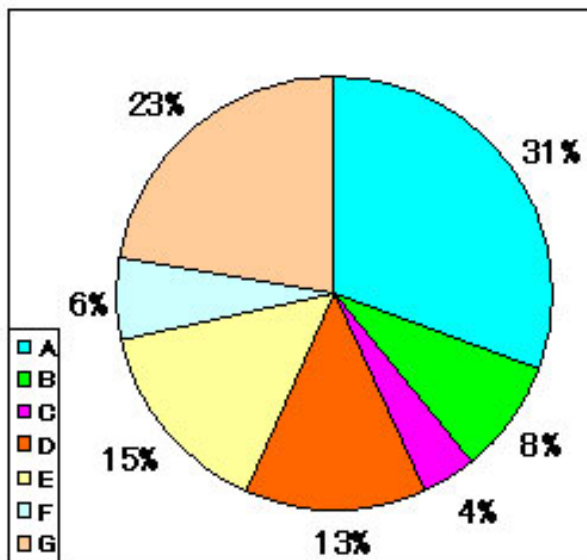
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Bibliography

Winston, Patrick Henry, and Berthold Horn. *The Psychology of Computer Vision.* New York: McGraw-Hill, 1975.

Aland, Kurt. *Synopsis of the Four Gospels.* 9th ed. Stuttgart: German Bible Society, 1989.

Conzelmann, H., and A. Lindemann. *Interpreting The New Testament.* Trans. Siegfried S. Schatzmann. Peabody, Mass.: Hendrickson Publishers, 1988.

Hearst, Marti A. "Segmenting text into multi-paragraph subtopic passages." *Computational Linguistics* 23 (1997): 15-36.

Kloppenborg, John S. *Q Thomas Reader.* Sonoma, Calif: Polebridge Press, 1990.

Minsky, M.L. *A Framework for representing knowledge.* Cambridge: Massachusetts Institute of Technology A.I. Laboratory, 1974.

Miyake, M., H. Akama, M. Sato, and M. Nakagawa. "Approaching to the Synoptic Problem by Factor Analysis." *Proceedings of the Institute of Statistical Mathematics* 48.2 (2002): 327-337.

Miyake, M., H. Akama, M. Sato, and M. Nakagawa. "Tele-Synopsis for Biblical Research." *Proceedings of the IEEE ICALT.* , 2004. 931-935.

Nestle, Erwin, and Kurt Aland, et al., eds. *Nestle-Aland Novum Testamentum Graece.* 26th ed. Stuttgart: Deutsche Bibelstiftung, 1979.