

PACIFIC WORLD

Journal of the Institute of Buddhist Studies

Third Series Number 11
Fall 2009

Special Issue Celebrating the
Sixtieth Anniversary of the
Institute of Buddhist Studies
1949–2009



Pacific World is an annual journal in English devoted to the dissemination of historical, textual, critical, and interpretive articles on Buddhism generally and Shinshu Buddhism particularly to both academic and lay readerships. The journal is distributed free of charge. Articles for consideration by the *Pacific World* are welcomed and are to be submitted in English and addressed to the Editor, *Pacific World*, 2140 Durant Ave., Berkeley, CA 94704-1589, USA.

Acknowledgment: This annual publication is made possible by the donation of BDK America of Berkeley, California.

Guidelines for Authors: Manuscripts (approximately twenty standard pages) should be typed double-spaced with 1-inch margins. Notes are to be endnotes with full bibliographic information in the note first mentioning a work, i.e., no separate bibliography. See *The Chicago Manual of Style* (15th edition), University of Chicago Press, \$16.3 ff. Authors are responsible for the accuracy of all quotations and for supplying complete references. Please e-mail electronic version in both formatted and plain text, if possible. Manuscripts should be submitted by February 1st.

Foreign words should be underlined and marked with proper diacriticals, except for the following: bodhisattva, buddha/Buddha, karma, nirvana, samsara, sangha, yoga. Romanized Chinese follows Pinyin system (except in special cases); romanized Japanese, the modified Hepburn system. Japanese/Chinese names are given surname first, omitting honorifics. Ideographs preferably should be restricted to notes.

Editorial Committee reserves the right to standardize use of or omit diacriticals. Conventionalized English form of sutra title may be used if initially identified in original or full form in text or note. Editorial Committee reserves the right to edit all submissions. Upon request, page proofs may be reviewed by the author.

Include institutional affiliation and position, or present status/occupation and place. All manuscripts submitted for publication become the property of *Pacific World*. By agreeing to publication in the *Pacific World*, authors give the journal and the Institute of Buddhist Studies an unlimited license to publish and reprint the essay. This license includes, but is not limited to, any media format (hard copy, electronic, etc.), and international rights. Copyright remains the author's.

EDITORIAL COMMITTEE

Richard K. Payne, Chair

David Matsumoto

Eisho Nasu

Natalie Quli

Scott Mitchell

ASSISTANT EDITOR

Natalie Quli

BOOK REVIEW EDITOR

Scott Mitchell

PACIFIC WORLD

Journal of the Institute of Buddhist Studies

Third Series, Number 11
Fall 2009

CONTENTS

The Lord of All Virtues HUDAYA KANDAHJAYA	1
Bianhong, Mastermind of Borobudur? HIRAM WOODWARD	25
Theravāda in History PETER SKILLING	61
Tsongkhapa on Tantric Exegetical Authority and Methodology DAVID B. GRAY	95
Nāgārjuna's Worldview: Relevance for Today KRISTIN LARGEN	119
Pattern Recognition and Analysis in the Chinese Buddhist Canon: A Study of "Original Enlightenment" LEWIS LANCASTER	141
Basing Our Personhood on the Primal Vow JUNDO GREGORY GIBBS	183
Shinjin and Social Praxis in Shinran's Thought TAKAMARO SHIGARAKI, TRANS. DAVID MATSUMOTO	193
The Metaphor of "Ocean" in Shinran TAKANORI SUGIOKA, TRANS. MARK UNNO	219
World Macrohistory and Shinran's Literacy GALEN AMSTUTZ	229

The Daoist Facet of Kinpusen and Sugawara no Michizane Worship in the <i>Dōken Shōnin Meidoki</i> : A Translation of the <i>Dōken Shōnin Meidoki</i> TAKUYA HINO	273
The Taoist Priest (Daoshi) in Comparative Historical Perspective: A Critical Analysis RUSSELL KIRKLAND	307
<i>Pì xiè jí 闢邪集: Collected Refutations of Heterodoxy</i> by Ouyi Zhixu (藕益智旭, 1599–1655) CHARLES B. JONES	351
Initiation and the Chinese <i>Hevajra-tantra</i> (T. 18, 892) CHARLES WILLEMEN	409
A Comparison of the Tibetan and Shingon <i>Homas</i> RICHARD K. PAYNE	417
BOOK REVIEW	
<i>Readings of the Lotus Sūtra</i> , eds. Stephen Teiser and Jacqueline Stone TAIGEN DAN LEIGHTON	451
BDK ENGLISH TRIPITAKA SERIES: A PROGRESS REPORT	459

Pattern Recognition and Analysis in the Chinese Buddhist Canon: A Study of “Original Enlightenment”

Lewis Lancaster

University of California, Berkeley, Emeritus

With the assistance of Howie Lan and Ping Auyeung

University of California, Berkeley

THE FOLLOWING DISCOURSE is based on a somewhat different approach to the study of a doctrinal term in Buddhist texts. The information presented here has been gathered through a software interface developed at the University of California, Berkeley by team members Lewis Lancaster, Howie Lan, and Ping Auyeung. A two-year grant of support (2007–2009)¹ was given by the National Science Foundation for the development of this tool. We have collaborated with the Institute of Tripitaka Koreana² in Seoul, and they have generously shared scanned images of rubbings taken from the printing blocks at Hae-in Monastery. The “software” makes use of a digital version of the previous publication by Lancaster and Sungbae Park, *The Korean Buddhist Canon: A Descriptive Catalogue*.³ The digital version of the catalogue is the work of Charles Muller of Tokyo University, who has made it freely available on the Internet.⁴ The interface project has also been a part of the Electronic Cultural Atlas Initiative (ECAI)⁵ and received support from that group’s *Atlas of Chinese Religions* research, which was funded by the Luce Foundation and is in collaboration with the GIS Center at Academia Sinica in Taiwan.⁶ Continued research on developing the interface used for this article is being pursued in cooperation with The School for Creative Media,⁷ ALIVE,⁸ and the Halliday Centre at the Department of Chinese Translation and Linguistics at City University of Hong Kong.⁹ Additional support comes from collaboration with research in the University of California at Berkeley College of Engineering.¹⁰ Future expansion of the analytics is being

done through cooperation with staff and faculty members at Carnegie-Mellon University,¹¹ Rutgers University,¹² and UCLA¹³ through a separate grant from the National Science Foundation.¹⁴ This is an indication of the need for teamwork and collaboration as we find ways to make use of technology and computation in the humanities. The approach followed below could not have been possible without the technical help of Howie Lan and additional support from Ping Auyeung, and therefore they are rightly listed as co-authors.

The intention of the search and retrieval strategy described below is focused on the history of the appearance of the term/compound 本覺, often translated as “original enlightenment.” It is an expression that has been studied in great detail and was chosen for this reason. The works of Jacqueline Stone¹⁵ and Robert Buswell¹⁶ contain valuable accounts of the ways in which the term was used and interpreted from the seventh century onward. There is no attempt to reconstruct the careful discourse laid out in those volumes. Rather, in this article, I have chosen to focus on the earliest appearances of the term as an attempt to trace the history of a word in Buddhist texts. This follows up on a lecture that I gave more than two decades ago under the title of “The Question of Apocryphal Words in the Chinese Buddhist Texts.”¹⁷ At that time, I suggested that the term 本覺 should be considered “apocryphal” since no Sanskrit equivalent could be determined. In this paper, I attempt to address the issue once again in greater detail, using some of the tools that are under construction for tracing the patterns of occurrences of vocabulary in the canonic texts.

As we test the effectiveness of a new approach, it is important to match the computer and computational results with previous knowledge. Because of the need to compare different strategies of research, we are focusing on this particular compound that has a long history within the scholarly literature of the Buddhist tradition and studies.

The initial and crucial question that arises from the method is whether quantification of data has a role to play in the study of doctrinal matters. What can be accomplished by computation of occurrences of a term and displaying a report in visual form? In part, the motivation behind the development of this interface software has been the awareness that the deluge of data created by digital technology requires new ways of retrieval and analysis. Scholars need tools that can help them quickly and efficiently use thousands of items identified by digital search.

The “work flow,” that is, the procedure by which a scholar approaches a digital search and the results of that search, forms the structure of this article. The steps that have been used are one method of handling the technology and the data. The “work flow” procedure used for this software is somewhat different from the ordinary way of research, and thus it is necessary to outline the particular strategies in some detail.

When approaching a research problem, we have to make immediate decisions about how to proceed. Academic training is structured along the lines of procedural steps. A major component of our past training and practice has been directed toward use of library reference assistance, based on codex collections of the data. However, in the digital age, where thousands or even millions of data are available, our former methods have begun to falter. Today, a pressing question is how a scholar should arrange, classify, and analyze search results from large data sets and/or the Internet. The older library reference system of using aids for research that point to sources of information is not so helpful in the digital arena, where we can go directly to the data without the intermediate step of consulting collateral documents. It is often difficult, without the considered judgment of the compilers of reference works, to determine the nature of the data that we access through the Internet and digital sets of information. Can software and new methods of approaching data through computation help us deal with these issues of verifying the accuracy of retrieved information? In other words, can we use computation of the data itself to solve problems such as determining accuracy of the data?

In the example being described in this paper, we turn attention to the digital version of the thirteenth-century Korean printing block edition of the Buddhist texts. It contains more than 52 million characters carved onto nearly 166,000 wooden surfaces, each producing a page of text when transferred to paper. There are other digital data sets that incorporate the readings of this edition, such as SAT and CBETA (see below). While it is recognized that volumes 1–55 of the *Taishō shinshū daizōkyō*¹⁸ print edition of the twentieth century are primarily based on the Koryō woodblock version, the editors of the Tokyo edition added dozens of texts known in Japan but not found in the Northern Sung corpus as recorded in the Koryō. For this reason, the study of terminology between SAT and CBETA, each based on the *Taishō*, will not be identical to the patterns found in the Koryō version. The conclusions

reached in this article are limited to the results obtained from the Korean block prints, which constitute the oldest complete set of original printing blocks for a version of the Chinese canon.

As mentioned above, in the past, scholars have approached a corpus such as the Korean canon through references in the form of catalogues, dictionaries, glossaries, concordances, and bibliographies. This type of research has been little changed in the field of Buddhist studies since the nineteenth century. A change is occurring in the contemporary world because a revolution in technology allows us to search and retrieve from the whole of material in the digital format. As a result, a complete inventory of every word or phrase is available; sometimes the examples number in the thousands. The older references based on codex publications are ill-suited to deal with this superfluity of data. In the comments below, the computational approach combined with visual analytics is explored as one way of handling reference questions in the digital age.

For the chosen example of how the new approach might be used, we start with the term 本覺. Our interest is in all occurrences of this combination of glyphs adjacent to one another. The goal of the research is to see the term in its total context within the Hae-in Monastery version of the canon and to determine the origin and uses of the glyphs in 1,514 texts. The previous approach of having scholars do a manual page by page reading and collecting has serious limitations when applied to more than 160,000 pages. In the last decade of the twentieth century, search and retrieval of target words and phrases was transformed for Buddhist scholars using Chinese language texts. Digital versions for the canonic material make it possible to find all references and display the results in a menu listing each line in which a target word occurs. In many cases, such a menu contains thousands of line references. While the current search for a term identifies all of these lines, the numbers can be large enough to require days or weeks of study by a scholar to understand the pattern of the occurrences. We now need tools that take us beyond this current state of the art. The functions of one such tool are described in the search and computation discussed in this article. We have not given a name to this tool since it is still in development. It will be referred to as the “software” (in quotation marks) for the present.

When we use the “software” to make a search in the Korean edition of the Buddhist canon for our target glyphs 本 and 覺 we find that they

occur adjacent to one another 763 times. Even though the number is large, it is a great advance over having to deal with the total number of glyphs in the whole of the corpus. Identifying 763 specific sites within 52,000,000 glyphs is a major accomplishment. Nonetheless, 763 occurrences is still a significant amount of data to handle, and the effort required to go through those hundreds of lines and analyze them is time consuming. The effort being made through the new “software” is directed toward taking these 763 examples and helping scholars analyze and classify so that significant occurrences and patterns will be identified in the shortest time possible.

As a first step with the “software” interface, we look for the number of times that each of the two glyph/characters appears in the corpus (see fig. 1). As the search is made, a report appears in visual form on a “ribbon” of blue dots, where each of the blue dots represents one of the 52 million glyphs. The dots are arranged by “panes” that correspond to the more than 160,000 pages of the version preserved in Korea. The dot is an abstract image that permits the user to see patterns of occurrence without the barrier of complex display of natural language glyph constructions such as in the Google report.

It is at this first step that we note the distinct shift in methodology. The initial move on the part of the scholar is to turn directly to the data itself rather than to reference works. As mentioned above, this is accomplished because the “software” provides a process of searching through the entire corpus at once. We have not gone through a reference work that points to data residing in another volume located at a separate site.

In order to proceed with the “work flow,” the user is shown the visual pattern of occurrence of 本覺 on the “ribbon of blue dots” (see fig. 2). This pattern is made into a visual one by changing the color of the dots that represent the target word from blue to red. Across the blue background, a pattern of red dots alerts us to the occurrences of the target word. This visual becomes the first factor in the scholar’s “work flow” planning. It shows that the glyphs are adjacent to one another in a scattering, marked by heavy concentration, in a few places and single isolated ones throughout the canon. Securing this much information within a few seconds can be compared to the hours of effort it would take to construct such a pattern, even with an Internet search that returns all examples of the term. In other words, an enormous amount of data is being displayed quickly and visually. We can “see”

the occurrences of our target search within the 52 million glyphs and immediately understand the nature of the pattern. This is very different from the current Internet search based on Google algorithms where we have hundreds of individual items listed in a long series of “pages” (see fig. 3). This is not a criticism of the present technology. It has been a great boon to Buddhist scholars that the digital versions of the Chinese are freely supplied by the CBETA¹⁹ and SAT²⁰ sites. These efforts have advanced our research many fold. We are deeply indebted to Fagu Buddhist College in Taiwan and Tokyo University for providing this service.

As with all digital technology, there is no point at which it can be said to be complete or finished. Data in the computer is always dependent on our continued efforts to preserve, disseminate, and access it. The new “software” interface being described here is an attempt to take the search and retrieval function to another level of speed and analysis. Visualizations of data can take many forms. In the window below (fig. 3), the glyphs have been shown as individual “blue dots” and the search, retrieval, and display was constructed based on “place of appearance” for each of the 52 million glyphs.

In the next visual, we explore the pattern of occurrence from the perspective of each text rather than each word and page. In this window (fig. 4) the 1,514 texts that make up the corpus of the Hae-in Monastery printing blocks are represented by a grid of squares. If the term in our search is present in any one of the texts, the square which represents it changes color to show the presence of the term (once or multiple times) in that text. Such a view of the 1,514 divisions is in contrast to the “ribbon of blue dots” that displays 52 million characters and 166,000 pages (fig. 2). Rather than looking at an image based on characters or pages, we have the possibility of a relatively smaller image exhibiting the search results showing only a report from each of the texts that make up the corpus. These visuals, whether based on each glyph or each text, are intended to provide different lens for viewing the pattern of occurrence for our two glyphs.

Attention turns to the “work flow,” which is determined based on the illustrations shown above. In order to understand the factual basis for the visual patterns, the “software” can provide the user with the following computations. A visual can be presented with all words counted and displayed by number of occurrences in a bar graph (see fig. 5).

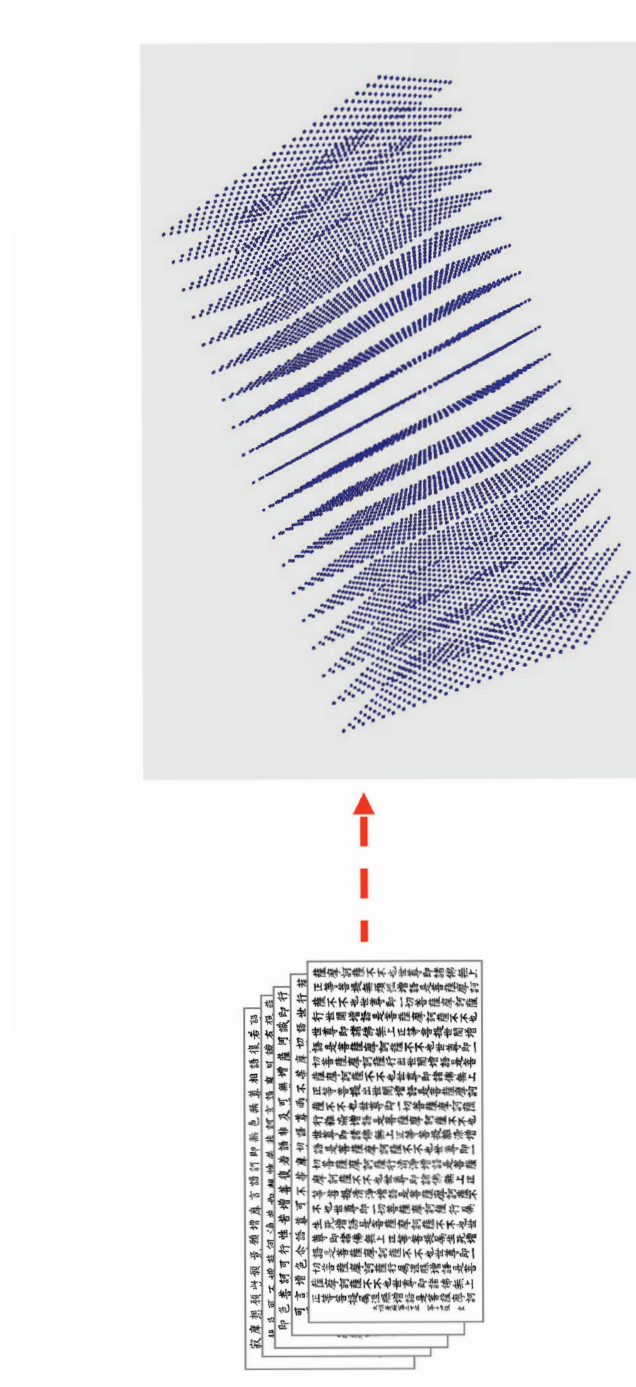


Figure 1. Rubbings of printing blocks shown abstracted into pane of “blue dots”



Figure 2. New interface showing all pages of the canon abstracted into a ribbon of “blue dots”

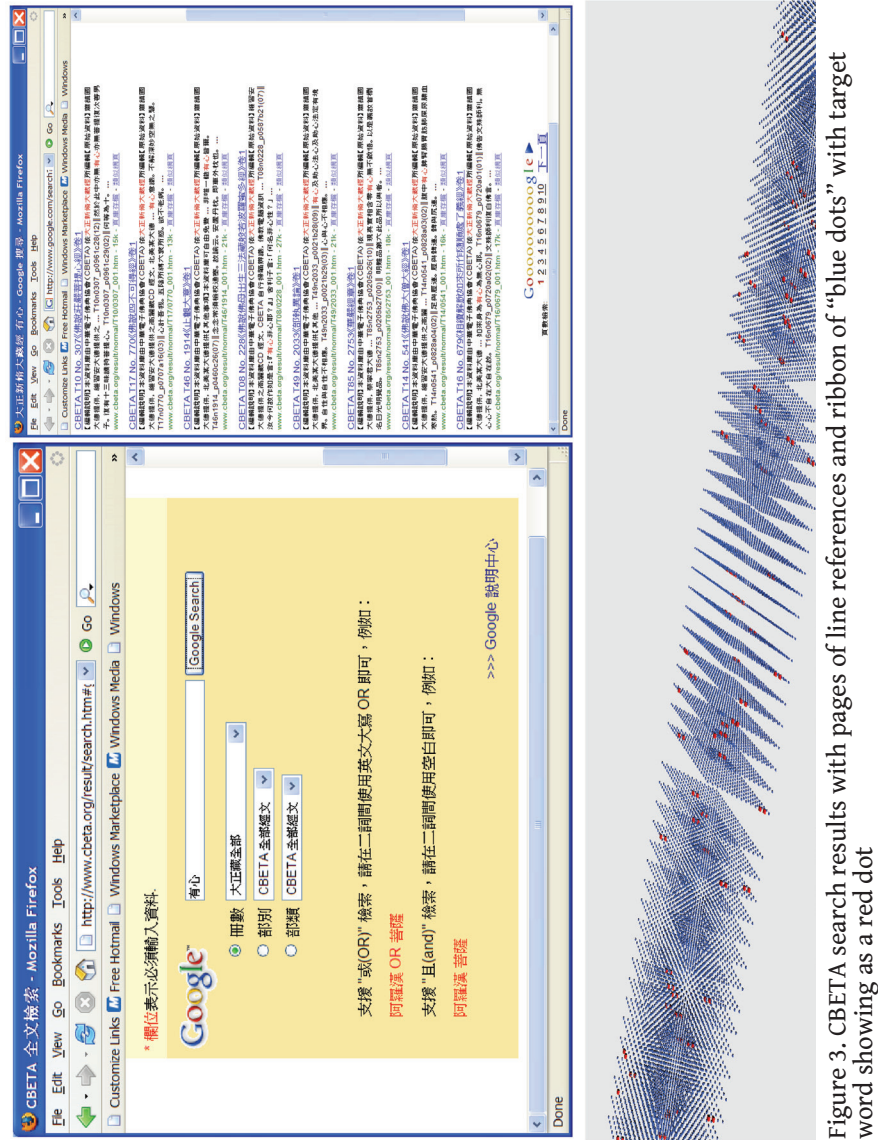


Figure 3. CBETA search results with pages of line references and ribbon of “blue dots” with target word showing as a red dot

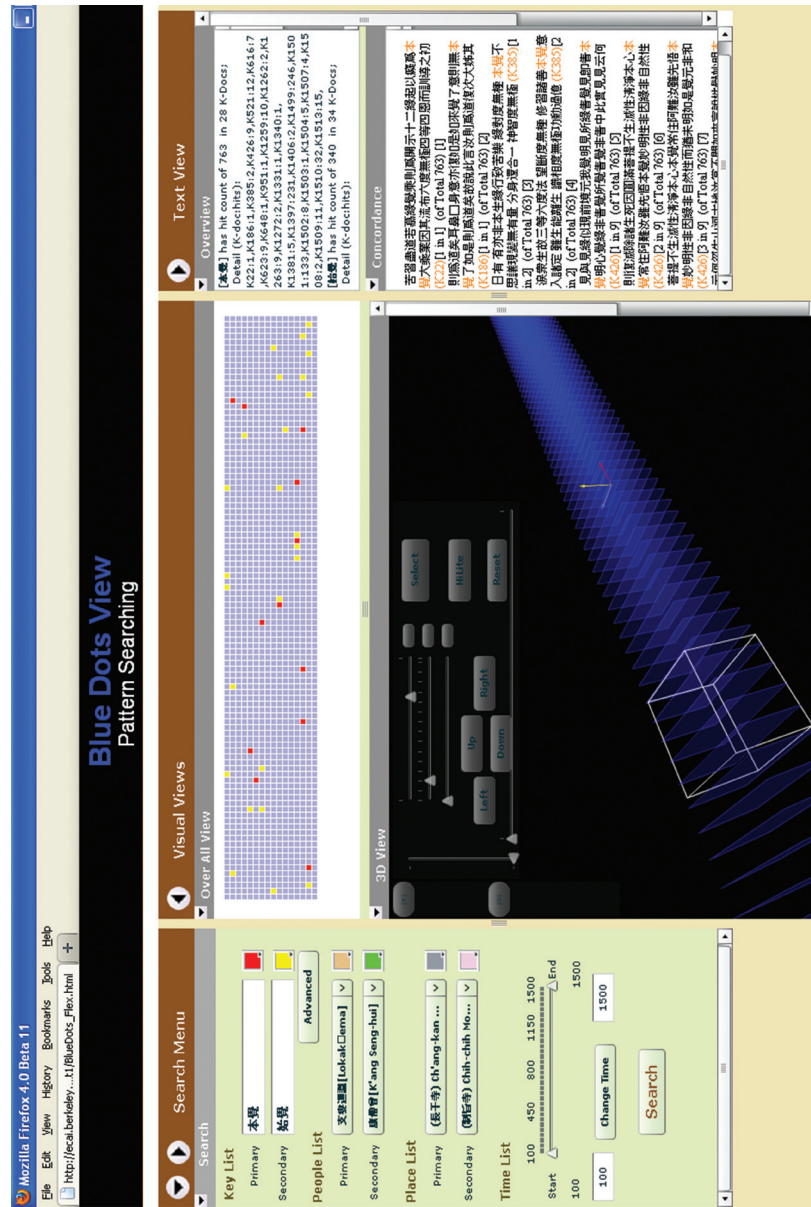


Figure 4. Full view of current interface of “software” showing multitude of search results

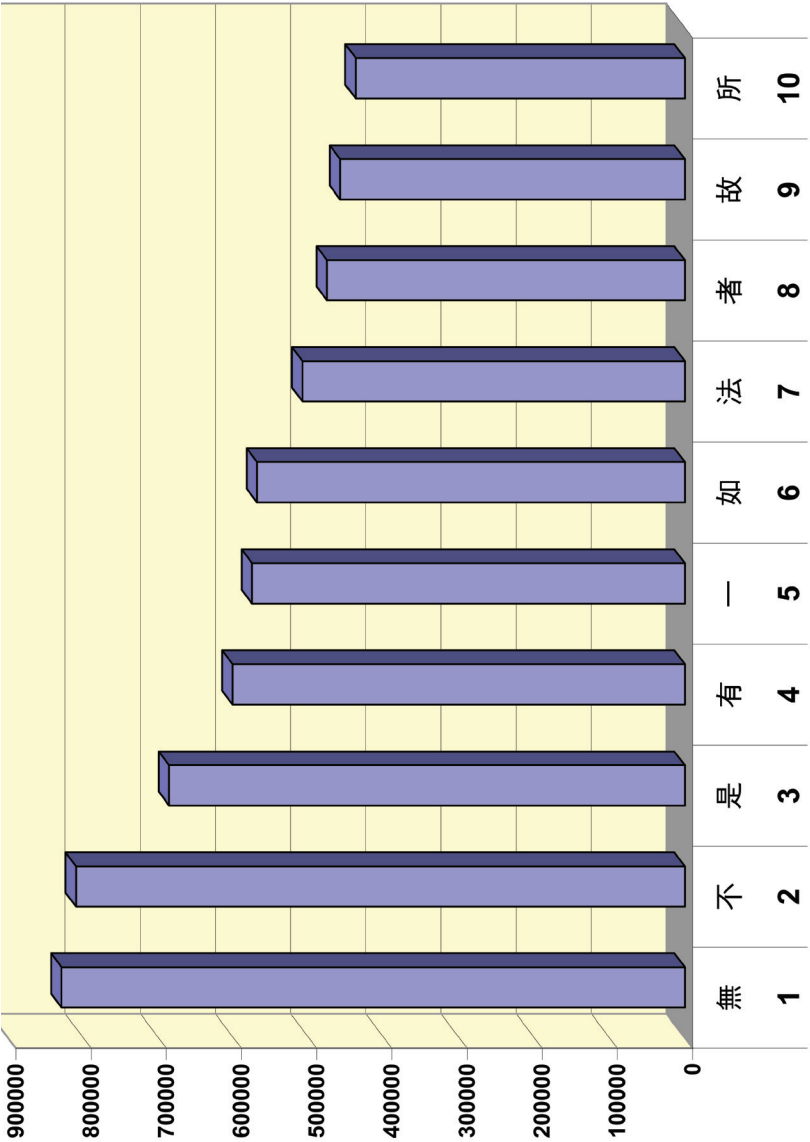


Figure 5. Computation of graph of total count of each character/ glyph in the canon

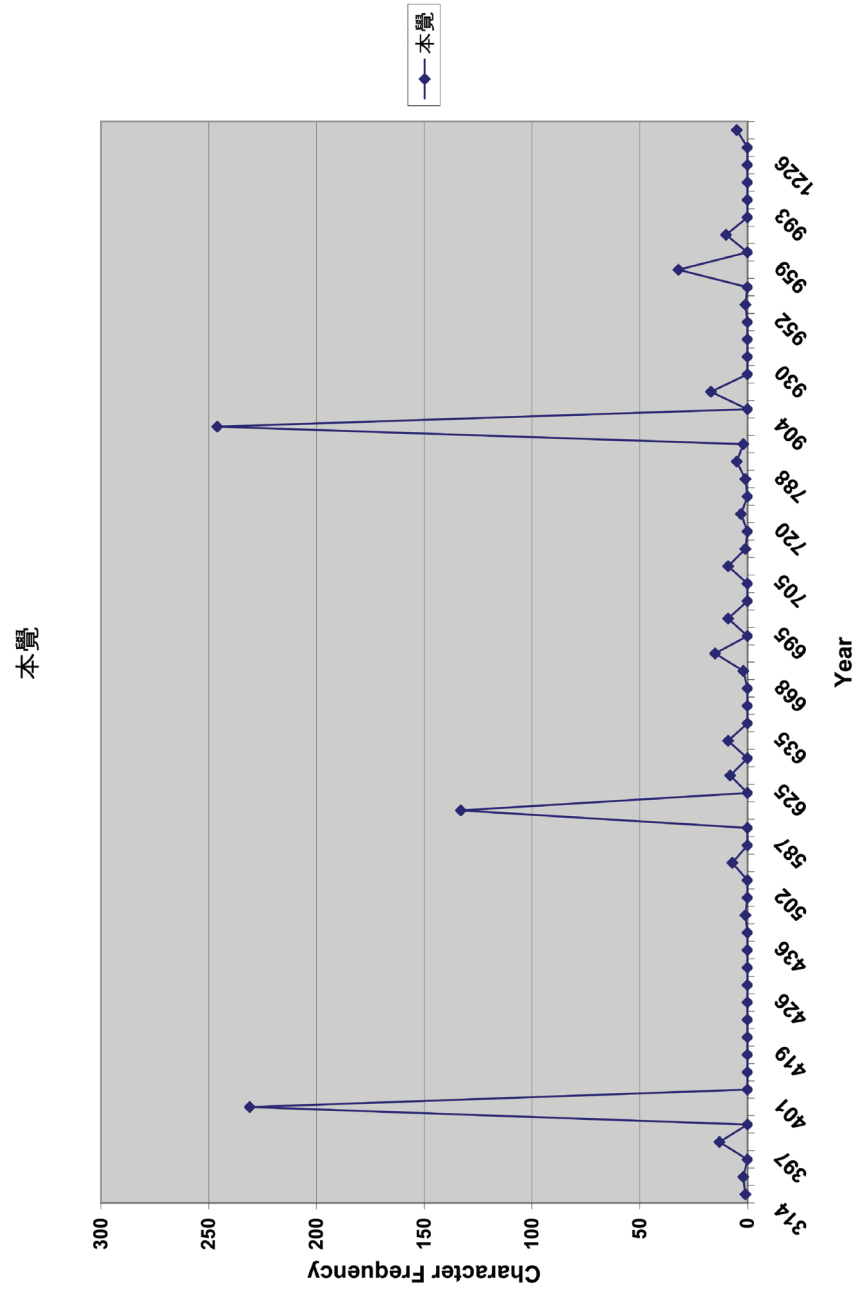
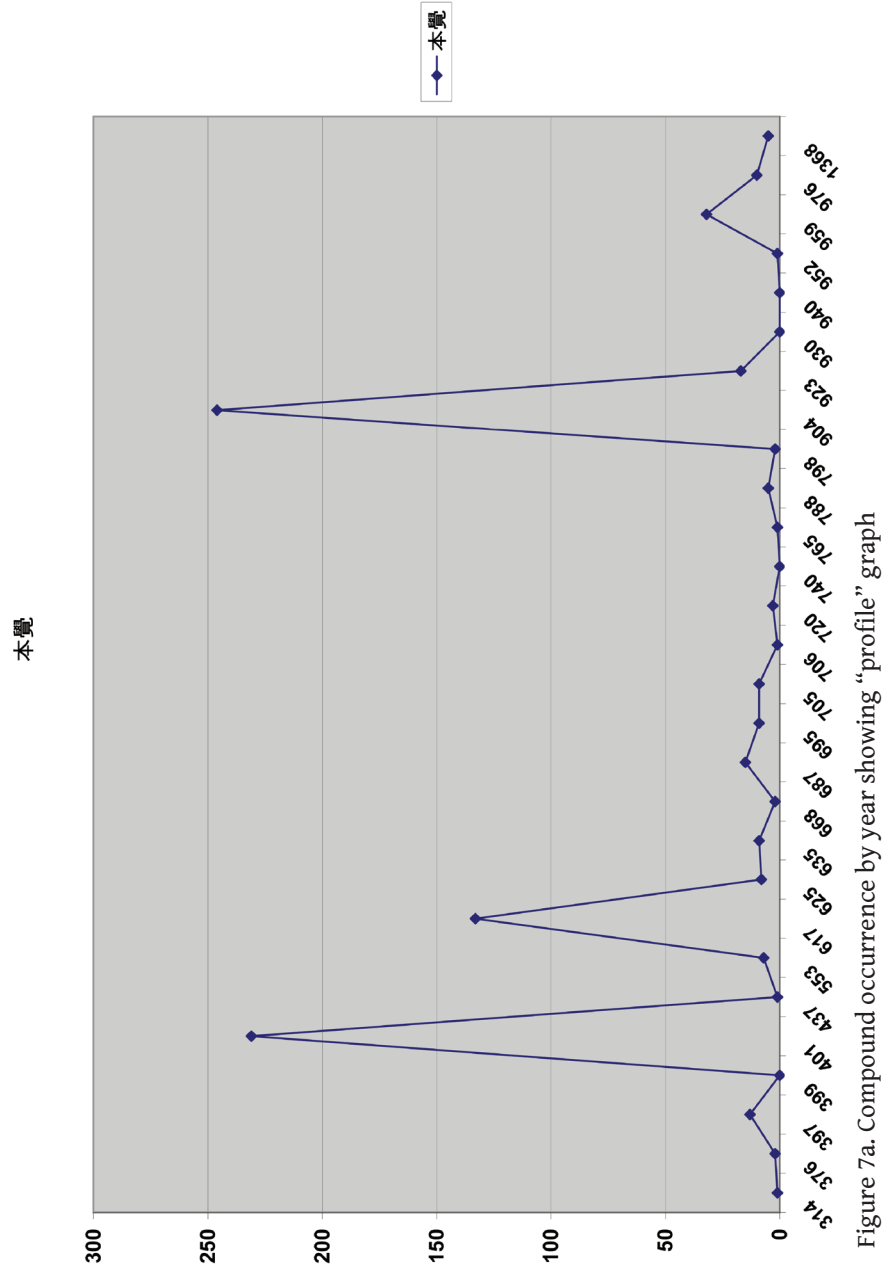


Figure 6. Computation of occurrence by date of translation



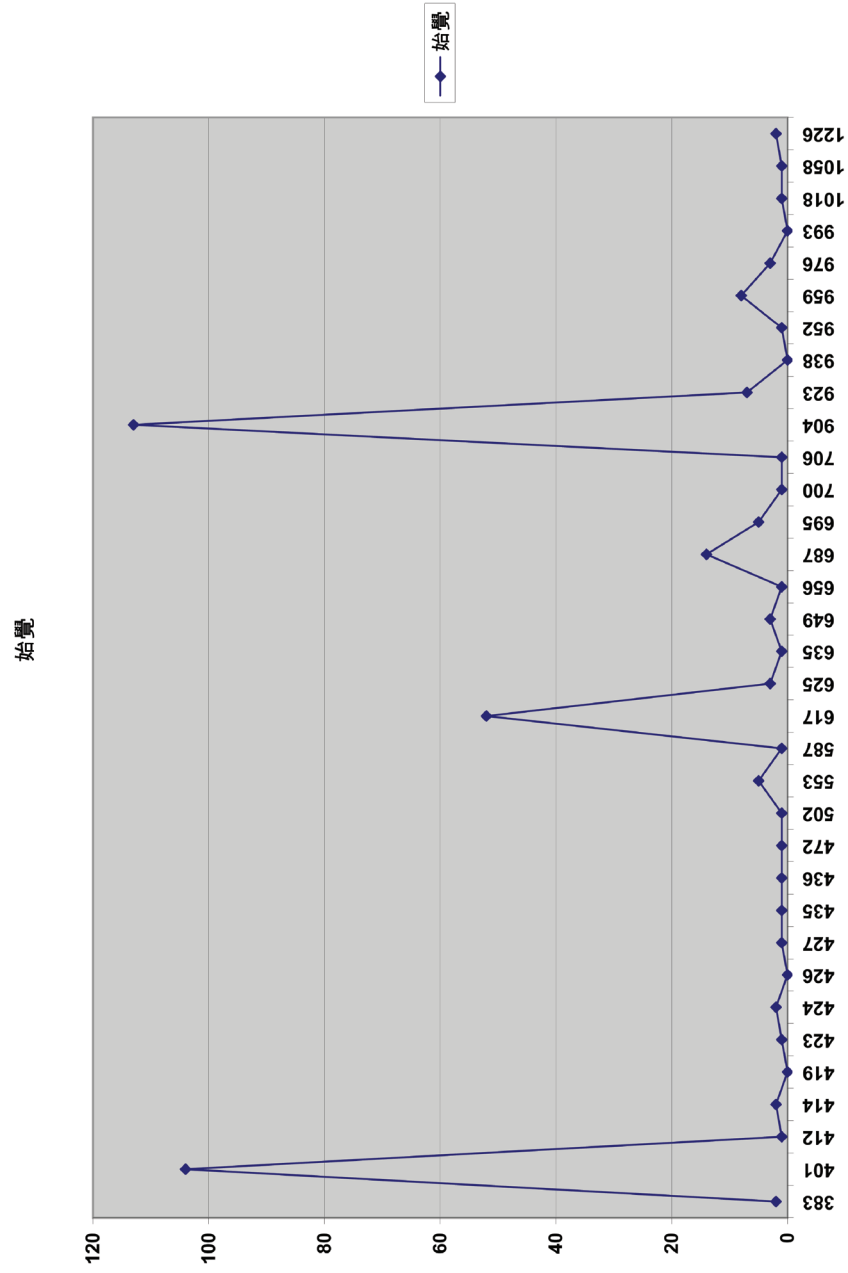


Figure 7b. "Profile" graph of companion compound

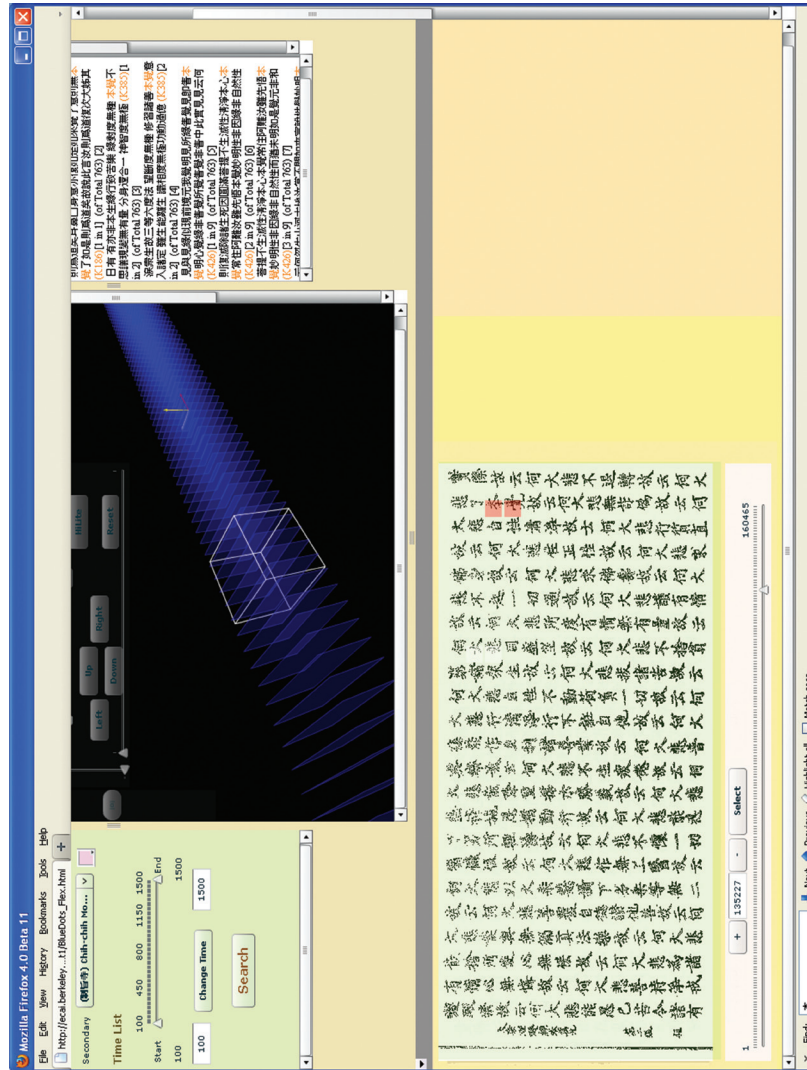


Figure 8. Scanned image of rubbing block appearing with target word highlighted

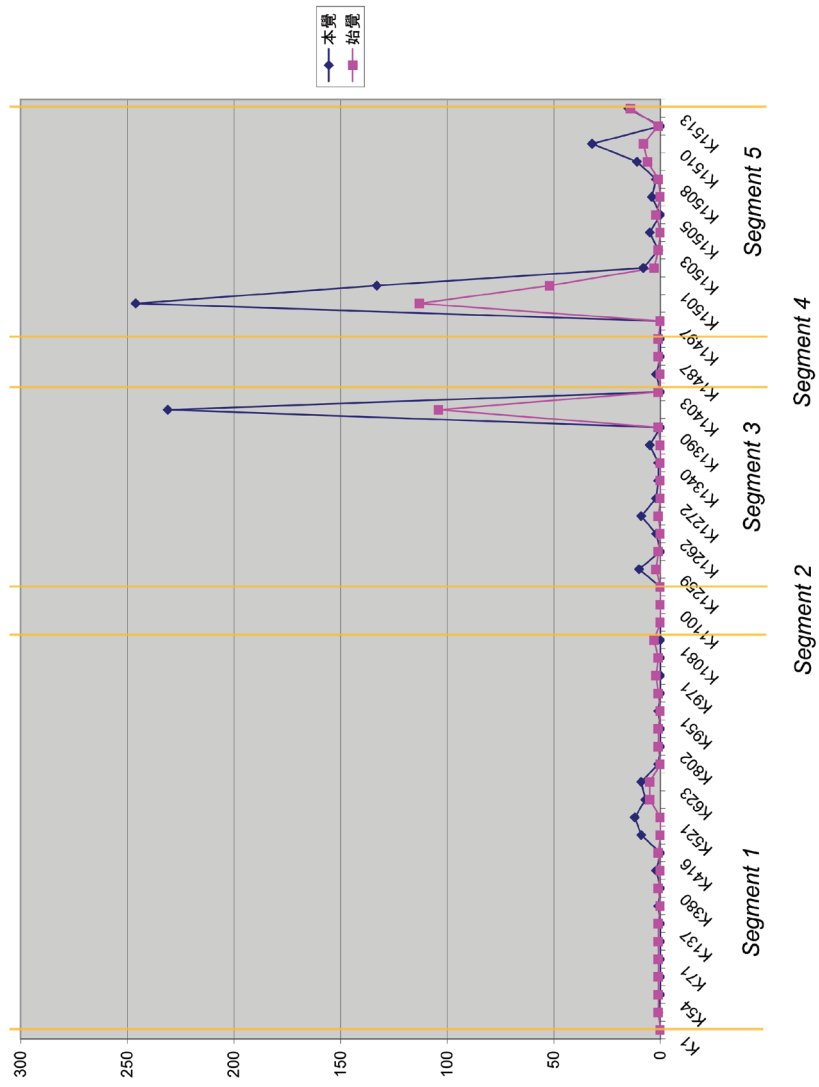


Figure 9. Computation and analysis graph showing five distinct segments of canon

COMPUTATION STEP ONE:

COUNTING ALL OCCURRENCES OF EACH GLYPH 本 AND 覺

Now that we have the overall pattern of occurrence (763 places scattered throughout the whole of the set with sizable clustering at a few points), the next step is to discover the significance of that visual pattern. The “software” provides assistance in the following fashion: the visual pattern based on computation can be used to give us a determination of the inner relationships of the glyphs that are the constituent elements of the data. Our inquiry for both glyphs gives the information:

本 is found in 1,180 texts with 71,833 hits

覺 is found in 1,182 texts with 69,527 hits.

In this count, we have determined that the two glyphs appear individually in large numbers throughout the 1,514 texts. Another computation reports that these individual glyphs are contained in 78% of the texts. Thus, the visual view of the text squares (fig. 4), with a large number of these squares colored to report occurrence, is based on this numerical computation.

Work Flow Analysis: The fact that these two glyphs are so widely used alerts us to the possibility that there will be a number of variables in the function and meaning of any adjacent position of the glyphs.

COMPUTATION STEP TWO:

COUNTING ALL OCCURRENCES OF THE
TWO GLYPHS 本覺 STANDING ADJACENT TO ONE ANOTHER

The next search is to combine the glyphs and search for every occurrence of the two in adjacent positions. This is at the heart of the research. We need to know when these two glyphs form the compound that means “original enlightenment.” The report comes back with the statistic that the adjacent pair can be found in 763 places in the 166,000 pages. We also receive the information that the 763 occurrences of the adjacent pair of glyphs appear in 28 of the 1,514 texts. This computation can be further refined to show that the 28 occurrences represent about 2% of the 1,514 texts of the canon. While the number of hits for the adjacent glyphs number in the hundreds, this is far smaller than the numbers for the occurrence of each of the glyphs alone:

本覺 763 hits compared to
 本 by itself 71,833
 覺 by itself 69,527

The adjacent occurrence is less than half of 1% compared to the separate individual examples of the glyphs that form the compound.

Work Flow Analysis: Since the compound appears in only 2% of the texts and the combination of two glyphs is less than half of 1% of the times when the single glyphs occur, it seems that the adjacent glyphs form a specialized term that has limited range in the text corpus.

COMPUTATION STEP THREE:

COUNTING THE NUMBER OF OCCURRENCES IN EACH OF THE 28 TEXTS

The “software” displays a new feature: a read-out of the catalogue K. (Korea) number of the texts where the words occur followed by the number of occurrences within the particular document. As we will see later, this count for each text is a crucial element in understanding the patterning of the 763 examples of 本覺. The report of the texts gives them in the sequential order of their appearances in the printing blocks at Hae-in Sa, i.e., K. 22–K. 1513. Thus we find that in K. 22 there is one hit (K. 22:1) for 本覺, and in K. 1397 there are 231 (K. 1397:231), etc.

Table 1. Occurrence count of target word listed by each text

K. 22:1, K. 186:1, K. 385:2, K. 426:9, K. 521:11, K. 616:7, K. 623:9, K. 648:1, K. 951:1, K. 1258:10, K. 1262:2, K. 1263:9, K. 1272:2, K. 1331:1, K. 1340:1, K. 1381:5, K. 1397:231, K. 1406:2, K. 1499:246, K. 1501:133, K. 1502:8, K. 1503:1, K. 1504:5, K. 1507:4, K. 1508:2, K. 1509:11, K. 1510:32, K. 1513:15.

Work Flow Analysis: The distribution of the adjacent glyphs involves a relatively small number of texts with a wide range of difference in enumeration. In order to judge the occurrences of the adjacent glyphs, we must search for characteristics of the 28 texts and determine if there are patterns that help explain the history of the adjacent glyphs 本覺.

COMPUTATION STEP FOUR:

COUNTING THE NUMBER OF OCCURRENCES OF 本覺
BASED ON TIME OF TRANSLATION OR COMPILATION OF EACH TEXT

Because the “blue dots” are not just pictures but each contains many fields of metadata behind the image, it is possible, for example, to compute the occurrences of the 763 adjacent glyphs based on the time of translation. The ancient catalogues of China, as well as the colophons attached to texts, give us temporal information about the translation or compilation/authorship of each text. This time-stamped data can also be used to look for patterns of occurrence. The profile of the image, representing “time,” indicates that there are specific “peaks” of activity. When we look at the image by the arrangement of the canonic texts compared to the image that shows the arrangement adjusted to time of translation, there are questions about the resulting patterns (see fig. 6).

In order to make this more meaningful, the computation that formed the basis for the imagery can be expressed in tabulation:

- (1) The text *K.* number in which the adjacent glyphs occur.
- (2) The number of examples found in each text.
- (3) The percentage of the occurrences in the text compared to the total number of 763.
- (4) The time of translation or compilation.

In table 2, we have a numerical report reflecting the information that underlies the visual pattern of occurrence as seen in the “blue ribbon” (fig. 2). Similar to the visual patterning, the tabulation indicates that the term is widely used throughout the corpus. In the tabulation, 本覺 is shown as appearing in sutras said to have been translated from the fourth century CE up to the Northern Sung dynasty and Koryŏ works of the tenth and fourteenth centuries CE (and one additional text dated to the Ming dynasty, seventeenth century CE).

The “software” now provides the next step, which displays the texts according the number of occurrences within them. The information shows that the range of hits is from a single one in a text (e.g., *K.* 1503) up to an impressive 246 (e.g., *K.* 1499). By this method, the texts can be clustered into units based on the number of times the adjacent glyphs occur (see table 3).

Table 2. Computation of occurrences arranged by order to text with date and percentages

Index	K-doc	HitsInDoc	Total Hits	%	Date range, CE
1	K. 22	1	763	0.13%	706
2	K. 186	1	763	0.13%	314
3	K. 385	2	763	0.26%	376
4	K. 426	9	763	1.18%	705
5	K. 521	12	763	1.57%	397–439
6	K. 616	7	763	0.92%	553
7	K. 623	9	763	1.18%	695–700
8	K. 648	1	763	0.13%	397–398
9	K. 951	1	763	0.13%	437–439
10	K. 1258	10	763	1.31%	976–997
11	K. 1262	2	763	0.26%	798–798
12	K. 1263	9	763	1.18%	635–730
13	K. 1272	2	763	0.26%	720–720
14	K. 1331	1	763	0.13%	720–774
15	K. 1340	1	763	0.13%	765–765
16	K. 1381	5	763	0.66%	788–788
17	K. 1397	231	763	30.28%	401–401
18	K. 1406	2	763	0.26%	668–668
19	K. 1499	246	763	32.24%	904–975
20	K. 1501	133	763	17.43%	617–686
21	K. 1502	8	763	1.05%	625–702
22	K. 1503	1	763	0.13%	952–952
23	K. 1504	5	763	0.66%	1368–1644
24	K. 1507	4	763	0.52%	923–973
25	K. 1508	2	763	0.26%	923–973
26	K. 1509	11	763	1.44%	923–973
27	K. 1510	32	763	4.19%	959–960
28	K. 1513	15	763	1.97%	687–695

Table 3. [本覺] Occurrences ranked numerically

Number of occurrences	K-doc
1	K. 22, K. 186, K. 648, K. 951, K. 1330, K. 1340, K. 1503
2	K. 385, K. 1262, K. 1272, K. 1406, K. 1508
4	K. 1507
5	K. 1381, K. 1504,
7	K. 616
8	K. 1502
9	K. 426, K. 623, K. 1263
10	K. 1258
11	K. 1509
12	K. 521
15	K. 1513
32	K. 1510
133	K. 1501
231	K. 1397
246	K. 1499

The “software” reports the data in percentages as well as exact count, i.e., of the 763 hits, 610 or 80% are from three of the 28 texts: K. 1397, K. 1499, and K. 1501.

Work Flow Analysis: The term is distributed unevenly. The pattern based on counts can indicate where attention should be directed. This alerts us to fact that any study of the term must take special note of three texts (K. 1397, K. 1499, and K. 1501, where 80% of the hits occur).

If we turn to the three texts that contain 80% of the examples (i.e., K. 1397, K. 1501, and K. 1499) and ask for a display of the time of translation/compilation for the three, we can see the temporal pattern placed alongside the numerical occurrence pattern. When the “software” returns the report on the dating of the three major texts, it shows us:

fourth/fifth century, K. 1397 (231 occurrences) (Yao-Chin “translation”)
 seventh century, K. 1501 (133 occurrences) (Korean compilation)
 tenth century, K. 1499 (246 occurrences) (Sung compilation)

Work Flow Analysis: The dating of K. 1397 with 231 examples of the glyph in the fourth century is an anomaly since the other texts with more than 100 entries are later compilations. Dating of the text K. 1397 is in question.

The anomaly of K. 1397 can be further explored by looking at the computation by date for all of the texts that contain the target word. We find that the attributions show examples of the adjacent glyphs (本覺) from the fourth to the eleventh centuries. The earliest and subsequent dates of translation as indicated in the catalogues and colophons are shown in table 4.

Table 4. Cluster of occurrences by century

Date of translation	Text(s)
4th/5th century CE (5 texts)	K. 186, K. 385, K. 521, K. 951 and K. 1397
6th century CE (1 text)	K. 616
7th century CE (3 texts)	K. 1406, K. 1501, K. 1502
8th century CE (11 texts)	K. 22, K. 1338, K. 1340, K. 1262, K. 1381, K. 426, K. 623, K. 1263, K. 1272, K. 1330, K. 1340
10th century CE (4 texts)	K. 1503, K. 1507, K. 1508, K. 1509
11th century CE (2 texts)	K. 1510, K. 1499
13th century CE (1 text)	K. 1500

Work Flow Analysis: The texts that display of the adjacent glyphs can be seen to cluster:

314–553 CE (7 texts) (314–439 CE and 553 CE)
 635–798 CE (13 texts)
 904–997 CE (6 texts)

We can note in the clustering of text by date of translation given in the catalogues that there are two periods of no recorded use of the adjacent glyphs. That is 553–635 CE and 798–904 CE. It is just as important to note the lack of occurrences as to record examples of them. The first temporal hiatus is made more impressive when we note that except

for one text in 553 CE, there was no example found after 439 CE until regular appearance seems to “resume” in 635 CE. This means that with one exception, our target word or adjacent position of the two glyphs disappeared from the corpus for nearly two centuries from 439 CE to 635 CE. It is hard to explain why a word would have been so popular in the fourth century and then fade from the texts only to reoccur in large numbers centuries later.

Work Flow Analysis: The composition of the temporal clustering needs to be further refined.

314–439 CE (6 texts)

Except for 553 CE, there was no example of the term over a period of more than two centuries (439–635 CE).

635–798 CE (13 texts)

Two of these are Korean compilations and the remaining attributed to Tang dynasty translations.

We have to note that translations of Buddhist texts in China ceased in 794 CE and did not resume until the Northern Sung once again started a bureau in 982 CE. Therefore, the hiatus of any pattern between 794 and 982 has to take this into account. The fact that there is a substantial break in the use of the term from 439 to 635 CE (apart from one example in 553 CE) must be considered when noting that the period of 314–439 CE stands out as an “island” of occurrences.

904–997 CE (6 texts)

All of these are compilations written in Korea, except for K. 1499, which was done in 961 CE in the Sung.

Work Flow Analysis: With the larger groupings in later periods, the fourth-century cluster that is so distant in time from the other examples must be classified as an anomaly since there are few examples until the seventh century. The number of occurrences within each text may provide a method to determine the nature of the pattern.

As a result of repeated pattern searching, the “software” makes a discovery that changes the workflow. Having created a number of graphical displays of the occurrence pattern for 本覺, a secondary search is made for any other nearby glyphs that have a similar pattern. In the case of 本覺 the “software” reports that there is such a “companion word,” 始覺. The two terms march side by side in a surprising profile match as shown in figs. 7a and 7b.

Work Flow Analysis: It is important to note those texts where both terms appear. Does the appearance of the “companion word” (始覺) assure us that the adjacent glyphs 本 and 覺 form the compound that we are researching?

The search for the origin of the compound 本覺 and its companion word 始覺 has the “work flow” now focused on six texts that have a dating as early as the fourth/fifth century CE. Our computation gives us the number of occurrences in each of them. The result is something of a surprise.

Texts that lack the “companion word”:

- K. 186 (1 example)
- K. 648 (1 example)
- K. 385 (2 examples)
- K. 521 (11 examples)
- K. 951 (1 example)

Text that shows a similar occurrence pattern of “companion word”:

- K. 1397 (231 examples) (shares profile with 始覺)

Work Flow Analysis: The list of fourth-/fifth-century texts can be split into two divisions by virtue of the count of occurrences of the adjacent glyphs.

- K. 186, K. 385, K. 648, and K. 951 have one or two examples.
- K. 521 and K. 1397 have multiple examples (11 and 231).

K. 1397 shows co-occurrence with the “companion word.” K. 521 does not have the “companion word,” which implies it may belong to a different tradition than K. 1397. Examination of the early texts will give “ground truth” of the “software” suggestion that K. 186, K. 385, K. 648, and K. 951 may not be early examples of 本覺.

The most noticeable difference is seen in K. 1397, which has a large number of occurrences unmatched by any of the other texts that date to the fourth/fifth century. In the computation, we attempt to understand this anomaly.

K. 1397: IS THE DATING CORRECT?

The “software” now runs some very specific searches automatically. One of these is to find all the places where the title of a text is mentioned by another text. We have already noted that the large number of occurrences of the glyphs in K. 1397 was a distinct variation of the pattern found in fourth-century texts.

The title search detects another anomaly. *K.* 1397 has three places where it mentions the title for *K.* 616. This is a problem because *K.* 1397 is dated as early as 384 CE, but *K.* 616 is dated as being completed in 553 CE. We assume from this information that the date of 384 CE for *K.* 1397 must be seriously questioned since it lists the title of a text that is dated in the sixth century. In the section below dealing with the dating of segments of the Hae-in corpus, we note yet another problem with the dating of this text.

This computation on the information found in the canonic data is an example of how the user can check on accuracy without relying on a separate reference work.

K. 521: IS THE DATING CORRECT?
“COMPANION WORD” NOT PRESENT

This text is problematic from several points of view. First, the number of occurrences of the target word (11) places it apart from the other texts dated to the fourth/fifth century, except for *K.* 1397. The tie between the two texts (*K.* 521 and *K.* 1397) is made more evident when the “software” reports on examples of titles appearing in other texts. The title of *K.* 521 appears in *K.* 1397 three times. This implies that *K.* 521 was composed prior to *K.* 1397. The logic then follows that if *K.* 1397 must be dated later than the fourth/fifth century, the appearance of the title for *K.* 521 in *K.* 1397 cannot be used to set the date of *K.* 521 in the fourth/fifth century. The title search shows that we do not find an example of the title of *K.* 521 until the catalogue entry in *K.* 1053 (compiled 502–557 CE). Thereafter, the title is found in a number of sources:

- K.* 1053 (502–557 CE)
- K.* 1054 (594 CE)
- K.* 1059 (664–665 CE)
- K.* 1062 (730 CE)
- K.* 1397 (388–417 CE)
- K.* 1499 (961 CE)

Such catalogue searching alerts us to the fact that the *K.* 521 title is only found in the sixth- and seventh-century catalogues apart from the *K.* 1397 examples.

This leads to another internal search strategy that can help to establish the nature of the data without reference beyond it. While titles can be used as examples of the existence of a text, the simple

occurrence of a title does not guarantee that a similarity of title words is proof of identity. The same title may have been used for more than one document. A stronger argument can be mounted when we have not only a title, but also a quote from the text. If the quotations found in one text can be linked to the words found in another, then we have assurance that the title does, in fact, refer to a specific document.

We do have quotations of K. 521 in other texts. For example, in 禪源諸詮集都序 (a text identified as belonging to the Tang dynasty with no specific year designated) there is the following quotation said to be from K. 521:²¹

金剛三昧經云。

即是動不動不禪是無生禪

In K. 521 we find the passage:²²

即是動不動不禪是無生禪

This establishes that K. 521 is the text being referenced, and, therefore, we know that 金剛三昧經 was known in the literature possibly as early as 618 CE, the first year of the Tang dynasty.

Another quote is found in K. 1499 done in 961 CE, indicating K. 521 as the source for the expression:²³

知諸名色唯是癡心分別癡心分別

This quotation does indeed appear in the 金剛三昧經, K. 521:²⁴

知諸名色唯是癡心癡心分別分別

This look at the quotations gives us proof that the 金剛三昧經 was known in the Tang dynasty and that knowledge of it also appeared in the Sung dynasty of the tenth century.

At this point, the quantitative approach of counting the number of occurrences of a glyph and looking at dating has indicated that there are problems with K. 1397 (*Shi mo ho yen lun*) and K. 521 (*Vajrasamādhi-sūtra*). These texts have been previously questioned, and much serious study has been given to the nature of their origins. The software has quickly alerted us to the fact that these two texts are problematic in terms of the date as well as the sequence with other texts. This identification of anomalies was constructed from only computation and pattern searching within the corpus of the Korean Buddhist texts. The data has yielded evidence of these problems without reference to any collateral documentation.

Our search for the origin of 本覺 continues. We turn again to the fourth-/fifth-century texts as the earliest examples and ask the interface to bring up the text for each occurrence of the adjacent glyphs. The scanned images of the printing blocks from Hae-in Monastery appear in the window with the target glyphs highlighted in color (fig. 8).

Workflow Analysis: Removing K. 1397 and K. 521 from the list of early texts, we are left with the adjacent glyphs occurring in K. 186, K. 648, and K. 656 in the fourth/fifth centuries and K. 951 in the sixth century. If the compound is to be found in the first centuries of Buddhist translations in China, it would have to be in one of the three fourth-/fifth-century texts. We look for the presence of the “companion word.”

K. 186: “COMPANION WORD” NOT PRESENT

K. 186, a text attributed to the well-known translator Dharmarakṣa, is said to have been translated January 28, 314 CE. If the compound 本覺 is found this early, it would give substance to the assertion that it is an ancient expression in Buddhism and belongs to an Indic origin. The text on the Korean printing blocks for the earliest dated example of adjacent occurrence of the glyphs in K. 186 reads:²⁵

如來覺了意則無本覺了

The question is how to determine if this passage contains the adjacent glyphs as a compound. One approach is to run parsing software to see if there are linguistic reasons for making the decision. While this is certainly a good approach, there are difficulties for scholars to make use of such sophisticated programs. Therefore, we have included in the “blue dots” “software” another approach that is based on computation only. The “software” searches the sentence and computes whether or not there is evidence that the characters preceding and following the glyphs are compounded with the glyphs. For example, 無本 is entered for the search since 無 is the glyph that precedes the first glyph. Another search is made for 覺了, based on the fact that 了 is the glyph that follows the second member of the compound 本覺. The report notes that 無本 and 覺了 are both compounds that are found hundreds of times in the corpus. Therefore, from counting instances of occurrence, the software suggests that 本 and 覺 should be seen as members of different compounds, not constituting a compound by their adjacent status. Having been alerted to this

possibility, we can directly consult the text to see if we agree with the software prediction that 本覺 in this phrase is not a compound.

如來覺了意則無本覺了

Here the first use of 覺了 represents a verb with the attendant particle to indicate that the action is completed or in the past tense. A similar example occurs four characters later. It would appear to be a denial of the teaching about 本覺 to take it as a compound since the negative 無 preceding it would read:

The Tathāgata has realized that the mind consequently is lacking original enlightenment. . . .

If this were the case, there would be no need to discuss 本覺 further. The sentence can be read differently:

The Tathāgata has realized that the mind consequently lacks any basis and [he also] realized. . . .

如來覺了意則無本.....覺了

If this is a correct interpretation of the use of the adjacent glyphs, then 本覺 is not a compound but the placement of two characters next to one another, with each having a separate grammatical function. The “software” has correctly predicted that in this phrase 如來覺了意則無本覺了, the adjacent glyphs 本覺 are not a compound.

K. 648: “COMPANION WORD” NOT PRESENT

We see a similar use of the adjacent glyphs in K. 648:²⁶

當知是覺因覺習覺本覺緣者

In this case, the software searches for 覺本 and for 覺緣 and finds them both to have many examples of appearing as compounds. Therefore, we receive the prediction that it is also a passage where the adjacent glyphs 本覺 do not comprise a compound but are last and first members of other compounds.

Here the structure of the sentence is obvious with verb-object phrases being repeated.

當知是.....覺因.....覺習.....覺本.....覺緣.....者

This shows us that K. 648, said to be translated by Gautama Sanghadeva in 397–398 CE, does not use the adjacent glyphs 本覺 as a compound.

K. 951: “COMPANION WORD” NOT PRESENT

Another of the early dated adjacent glyphs is found in K. 951, a translation attributed to Buddhavarman and dated to 437–439 CE.²⁷

以彼地有言語根本覺觀法故

Here once again, the glyphs have different functions. For example, 根本 is a much-used compound, as is 覺觀. Thus the phrase should be parsed as follows:

以彼地有言語.....根本.....覺觀,,,,,法故

The adjacent characters do not form a compound.

K. 385

The same holds true for the second adjacent example in K. 385:²⁸

修習諸善本 覺意入諸定

This is a line of verse, and it splits in the same fashion as the one listed above.

修習諸善本.....覺意入諸定

Work Flow Analysis: The two characters 本 and 覺 can occur next to one another without being a compound. Therefore, adjacent pairs (in K. 186, K. 385, K. 648) are not examples of an early occurrence of the compound. The fact that the “companion word” is not present in any of these texts indicates that it is an important indicator of the function and use of the two characters.

From this computation, we can reach the following conclusions: (1) The texts dated from 384–471 CE containing the adjacent glyphs (i.e., K., etc.) can all be rejected as containing early examples of the use of 本覺 as a compound. (2) We can next note that the earliest references that can be supported as true use of the compound 本覺 are in K. 426 (9 occurrences), K. 521 (11 occurrences), K. 616 (7 occurrences), and K. 623 (9 occurrences). These documents carry the following dates:

K. 426, listed as 700 CE (lacks the “companion word”)

K. 521, listed as 397–439 CE (lacks the “companion word”)

K. 616, listed as 553 CE (“companion word” occurs)

K. 623, listed as 695–700 CE (“companion word” occurs)

The pattern has now suggested that K. 521 and K. 616 are the early attested sites for the use of 本覺 as a compound.

Work Flow Analysis: The relationship of one text to another is crucial in the dating. We need to run the search for examples of the titles of each of the texts to see if this can help with the arrangement of the order.

This “software” search for titles quoted in other texts reports that K. 521 mentions the title of K. 616 three times. This alerts us to fact that K. 521 cannot be prior to K. 616. Therefore, the software computation and search leaves us with K. 616 as the earliest example of the compound.

K. 616: “COMPANION WORD” OCCURS

When the text is put into the title “software” search, we find that the title is quoted a number of times

K. 1404 (502–557 CE)

K. 1055 (597 CE)

K. 1501 (617–686 CE)

K. 1075 (649 CE)

K. 1058 (695 CE)

K. 1498 (788–810 CE)

This indicates that the date of 553 CE is possible and that it is well-known title that was often used over the next century.

Work Flow Analysis: Since the computation is able to find profiles of words, can it also provide a method of seeing structures in the organization of the corpus as a whole? We must ask for the “software” to give us a visual that detects if there are discernable segments in the canonic arrangement based on our computational searches.

The “software” alerts us to possibility of five segments of the entire corpus (see fig. 9).

Work Flow Analysis: Can we determine if the five segments in the visual can be documented?

The discovery of the five segments accords with the history and structure of the corpus (see table 5).

Table 5. Descriptive analysis of segments of the canon

- I. First set of block prints sent to Koryŏ from Northern Sung
K. 1–1087
 - Translations (ca. 148–711 CE)
 - Compilations (562–730 CE)
- II. Supplemental set of prints from the Northern Sung representing new translations being made after the carving of the original printing blocks
K. 1088–1256
 - Translations (982–999 CE)
- III. A supplement of blockprints that represent older translations of compilations that had not been included in the original set of printing blocks of the Northern Sung
K. 1257–1406
 - Translations (730–798 CE)
 - Compilations (730–997 CE)
- IV. A supplement of blockprints that represented the continued translation project of the Northern Sung
K. 1407–1497
 - Translations (1000–1090 CE)
- V. A supplement of compilations made in East Asia with authorship given in the colophons
K. 1498–1513
 - Compilations (China) (668–1360/1644 CE)
 - Compilations (Korea) (617–1251 CE)

Work Flow Analysis: “Software” should now report on which texts are found in each of the five identified segments.

The “software” returns the following report of how the texts with 本覺 are found in the segments listed above for the way in which the printing blocks have been assembled (see table 6). In these texts, we find the earliest examples of the glyphs as a compound (see table 7).

Table 6. First segment of blockprints (excluding K. 22, K. 186, K.385, K. 426, K. 648, and K. 951 as not containing the compound)

Text	Occurrences	Percentage	Date range (CE)
K. 521	12	1.57%	397–439
K. 616	7	0.92%	553
K. 623	9	1.18%	695–700

Table 7. Third segment of supplemental translations and compilations in the Tang dynasty

Text	Occurrences	Percentage	Date range (CE)
K. 1258	10	1.31%	976–997
K. 1262	2	0.26%	798–798
K. 1263	9	1.18%	635–730
K. 1272	2	0.26%	720–720
K. 1331	1	0.13%	720–774
K. 1340	1	0.13%	765–765
K. 1381	4	0.66%	788–788
K. 1397	231	30.28%	401–401
K. 1406	2	0.26%	668–668

Work Flow Analysis: 本覺 does not appear in the second or fourth segments of the canon, that is, in any of the translations made during the Northern Sung period. In other words, the compound is not found in the late Sanskrit texts that were available in Kaifeng representing translations made from 982–1090 CE.

This gives further evidence that the presence of the term 本覺 belongs to texts of East Asian origin rather than those that are translations from known Sanskrit documents.

Work Flow Analysis: While we have rejected the earliest examples and noted that the compound is not found in the second or fourth segments, the issue regarding the Sanskrit equivalent is not fully resolved since we have a number of occurrences in Tang dynasty translations.

This third supplement is of great interest because it brought to Korea a set of compilations and translations that were not included in the first set of rubbings from the Northern Sung. It is understandable that they were not part of the initial group because that segment only contained materials produced prior to 730 CE when the catalogue of the *Kaiyuanlu* was used as a guide for canonic listings. The texts found in the third segment (translations: 730–798 CE; compilations: 730–997 CE) that have examples of the compound include three compilations from the Tang: K. 1258, K. 1263, and K. 1406. Of the remaining texts, three are attributed to Amoghavajra, and it is in these that we can seek examples that might point to a Sanskrit equivalent for the term. We note the anomaly of K. 1397 once again. It is included in the segment of the Hae-in set that is a collection of texts mainly attributed to the Tang dynasty. In that segment it is the only text that carries a fourth-/fifth-century date. Even in the Tang list it stands out as an anomaly, with far more examples than any other text of the period.

If we explore the Tang translations, we look at the works attributed to Amoghavajra and Prajña. The texts are not listed in the *Kaiyuanlu* since they were translated after 730 CE when additional materials were rendered in Chinese. The “companion word” is not found in either text, i.e., K. 1381 and K. 1262.

Work Flow Analysis: Does the “companion word” occur in any of these Tang examples? If not, it suggests that the glyphs do not represent the compound.

K. 1340: “COMPANION WORD” DOES NOT OCCUR

This is the well-attested 仁王護國般若波羅蜜多經 that was translated by Kumārajīva, and this gives the possibility of a Sanskrit equivalent if we find the compound in both versions. The passage in this work of Amoghavajra reads:²⁹

名爲佛自性清淨名本覺性

We find the equivalent passage in Kumārajīva’s translation of the text:³⁰

名爲佛自性清淨名覺薩婆若性衆生本業

名爲佛自性清淨名本覺性 = K. 1340

The hope for a Chinese equivalent is dashed when we note that 覺薩婆若性, “comprehends the nature of all knowledge,” in Kumārajīva has been replaced with 本覺性, “the nature of original enlightenment,” in Amoghavajra. This suggests that the original form of the Indic text did

not have the compound 本覺, but it was added to the eighth-century translation, most probably as a Chinese element.

K. 1331: “COMPANION WORD” DOES NOT OCCUR

In this text we find the expression:³¹

顯明本覺真

The “software” reports that 覺真 is a well-attested compound (Skt. *prabuddha-tattva*). At the other side of the first glyph, 顯明 is also widely used as a verb “to make known.” It is reasonable to assume that 本 is a modifier for the compound 覺真 and not just for 覺 alone. This would give us the reading:

made known the basis for the reality of enlightenment.

It does not appear that we can list this statement as proof of our compound in a Sanskrit work of Amoghavajra.

K. 1272: “COMPANION WORD” DOES NOT OCCUR

K. 1272: First Occurrence

This work by Amoghavajra has the glyphs in adjacent position. As part of the search algorithm, the “software” indicates when the target glyphs are also used adjacent to one another in reverse order. This can be seen in this text of Amoghavajra where we find 是名本覺覺本 and 覺觀覺本. It is important to take note of such a reversal of the glyphs, since it tells us much about the changing functions of the glyphs in the compound 本覺.

: 佛言本來清淨故是名本覺。覺本淨 徹無處。是故名爲法身智身滿足故。佛言一切衆生當用覺觀。覺本心性體靜無生。離衆生垢故。覺本無 寂離涅槃性故³²

In the first phrase we find:

佛言本來清淨故是名本覺

The Buddha said: “Because of the original purity this is, namely, the basis for enlightenment.”

The next phrase 覺本淨 徹無處是故名爲法身 reverses the order of the glyphs and can be rendered:

Because of the full comprehension that the basis is pure without any foundation it is taken to be the *dharmakāya*...

K. 1272: Second Occurrence

There is another example in the text, where we find the expression³³

觀本覺體

Here is another example of the “software” predicting that the glyphs are standing in relationship to those that follow and precede them. Both 觀本 and 覺體 are compounds that occur hundreds of times in the corpus. Thus, we would translate this as

comprehends the basis for the reality of enlightenment.

Work Flow Analysis: We can eliminate the works of Amoghavajra as having examples of the compound. The last remaining possibility will be the works of Prajña. First, we will look for the “companion word.”

These two texts of Prajña were translated in 788 and 793 CE, respectively.

K. 1381: “COMPANION WORD” DOES NOT OCCUR

大乘理趣六波羅蜜多經 is the lone translation of this text in the canon. In it, we find four examples of our glyphs standing adjacent to one another.³⁴

K. 1381: Occurrence 1

云何大悲了本覺故

In this phrase, we find the pattern of a question marked by the function character 云何 with the answer appearing before the function character 故. This would result in the translation:

What is the completion of Great Wisdom? . . . It is the basis for enlightenment.

There is no reason to discount this occurrence as the compound. However, a major problem with using this example for establishing a Sanskrit equivalent is the fact that no extant Sanskrit passage or any of the Chinese translations match the reading we find here.

I have translated the compound as “basis for enlightenment” rather than “original enlightenment.” There is a major difference in these two renderings. It would appear that 本覺 should be taken as an answer to the process of completing the Great Wisdom. The phrase 大悲了 occurs many times in the canon and always implies a completion of a process. The answer should be a phrase related to a process question. Therefore, the compound 本覺 is translated as “basis

for enlightenment,” i.e., the process of bringing to fruition the Great Wisdom as the “basis.” If 本覺 is an already-existing entity in the mind, it would not be described as occurring in reference to a specific contemporary action. The Great Wisdom is described in many places as an ongoing process, and the meaning in this text seems to accord with that idea. In other words, it is not some primordial task that resulted in 本覺, which lies buried in the mind, but it is an ongoing activity that forms the basis for enlightenment. In this way the adjacent glyphs constitute a phrase more than a nominal compound.

A similar example from the same translation by Prajña helps with this translation:³⁵

於大悲了達諸法

In the completion of the Great Wisdom one encompasses all the mental states (*dharma*).

K. 1381: Occurrence 2

Another example for this text of Prajña is:

四念處以爲守護本覺心王 第一義禪定

The four stations of mindfulness, considered as protectors of the basis of the preeminent thought of enlightenment, are the highest meaning of meditation.

We find the expression 應堅守護本尊三昧耶 in: 頂一切如來真實攝大乘現證大教王經³⁶

... 應堅守護本尊三昧耶

... should maintain the basis for cherishing *samādhi*.

In keeping with the discussion of the process required for the creation of the “basis of enlightenment,” we note the focus on the importance of meditation in these examples.

K. 1381: Occurrence 3

了真本覺是名智慧菩薩摩訶薩³⁷

... comprehends completely that the true basis of enlightenment is namely the wisdom of the Bodhisattva Mahāsattva. ...

The “software” alerts us to the fact that the term 真本 is found in hundreds of examples. I translate it as meaning the “true basis.”

K. 1381: Occurrence 4

不見本覺心自覺智現前真性常不動³⁸

... do not see that present true nature, (whether it be of) the basis of enlightened thoughts or the wisdom of one's own enlightenment, is always unmoving. ...

It would seem that 真性 is the object of the verb form 不見 and the intervening words are modifications of the object. The section does not exclude the possibility of the glyphs forming a compound.

Two occurrences of the adjacent glyphs continue in Prajña's version of the 大方廣佛華嚴經. This is a promising source for a potential Sanskrit equivalent. The *Gaṇḍavyūha* has six other translations in the main body of Koryō main blocks and another three in the supplement. This well-known text, which also forms one part of the *Avatamsaka-sūtra*, is known in extant Sanskrit manuscripts. It is important to note that of all the translations of the *Gaṇḍavyūha* in the Koryō blocks, i.e., K. 102, K. 104, K. 1029, K. 1282, and K. 1262, it is only in K. 1262 that we find the compound.

K. 1262

K. 1262: Occurrence 1

諸佛菩薩自證悟時轉阿賴耶得本覺智³⁹

All the buddhas and bodhisattvas at the time of their own realization transform the *ālaya* [-vijñāna] and attain the wisdom that is the basis of enlightenment.

K. 1262: Occurrence 2

善知識者心如明燈順本覺性而覺了故善知識者如⁴⁰

... because the Good Teacher's thought is just like a bright lamp, it is in accord with the nature of the basis of enlightenment and is completely enlightened.

Table 8. Fifth segment compilations made in East Asia

Text	Occurrences	Percentage	Date range (CE)
K. 1499	246	32.24%	904–975
K. 1501	133	17.43%	617–686
K. 1502	8	1.05%	625–702
K. 1504	5	0.66%	1368–1644
K. 1509	11	1.44%	923–973
K. 1510	32	4.19%	959–960
K. 1513	15	1.97%	687–695

Work Flow Analysis: The largest numbers of occurrences of the adjacent glyphs are found in the texts identified as compilations made in Korea or China. They are not sources for proof of a Sanskrit equivalent. We need to see where the “companion word” occurs in these compilations.

The computation suggests that the popularity of the compound is primarily found in the compilations, not the texts that are said to be translations. The “companion word” appears in all of the listed compilations except for K. 1504, a Ming document that has been added to the Hae-in Sa blocks at a much later date and hardly belongs in the explorations of the Koryŏ block prints. The important point is to note that most of the occurrences of the compound with its “companion word” are found in East Asian compilations. It would be difficult to establish the Sanskrit equivalent for the compound based on documents of this sort.

CONCLUSIONS

From this extensive computational study, the ideas that I first put forward in the AAR paper⁴¹ have now been established with much more precise data. The evidence as found in the Koryŏ edition provides no definitive proof that a Sanskrit original can be established for the term. The oldest use of the term seems to be in the sixth-century K. 616. At that time, it was combined with a “companion word,” and thereafter the two compounds appear with a similar profile. K. 616 is probably an East Asian compilation, and we have no extant Sanskrit equivalent for it. It was a surprise to me that just by using rather simple computation I could identify anomalies and spot problems in dating and attribution.

My conclusion is that the “companion word” 始覺 may be the crucial and important link to the origin of the compound and its meaning. The compound 本覺 seems to have been used initially in the sense “basis of enlightenment” as a process rather than a name for an artifact eternally existing in the mind.

Final Work Flow Analysis: There are many questions still awaiting research. What of the occurrence of the compound and “companion word” in the texts that were added to Taishō and the supplements that appear in CBETA? Do other doctrinal terms have “companion words,” and can the “software” find these in other examples? The “companion word” needs to be explored in the same way that I have done here. The interface must be developed into a robust and open source tool for free distribution to others. Another term, “buddha-nature,” also should be similarly researched.

The future development of the “software” and the ways in which computation will be used and accepted by scholars in the field opens up new vistas for understanding Buddhist words, their history of use, and the network analysis for how words interact. In this regard, Buddhist studies has the potential of taking the lead in developing new methods of research for the humanities.

NOTES

1. National Science Foundation grants 0929851 and 0840061, “SGER: Text Analysis and Pattern Detection.”
2. See http://kb.sutra.re.kr/ritk_eng/index.do for a description of the work of this institute.
3. Lewis Lancaster and Sungbae Park, comp., *The Korean Buddhist Canon: A Descriptive Catalogue* (Berkeley: University of California Press, 1979).
4. http://www.acmuller.net/descriptive_catalogue/index.html.
5. <http://www.ecai.org>.
6. <http://www.ascc.sinica.edu.tw/en/service/gis.html>, under the guidance of Professor Fan I-chun.
7. <http://www.cityu.edu.hk/scm>, led by the Dean Jeffrey Shaw.
8. <http://www.cityu.edu.hk/scm/alive>, Applied Laboratory for Interactive Visualization and Embodiment under the direction of Sarah Kenderdine.
9. <http://ctl.cityu.edu.hk>, Halliday Centre for Intelligent Application of Language Studies directed by Chair Jonathan Webster and his colleague John Lee.

10. <http://www.me.berkeley.edu/faculty/frenklach>, Professor Michael Frenklach.
11. <http://www.cs.cmu.edu/~christos>, Professor Christos Faloutsos.
12. <http://eliassi.org>, Dr. Tina Eliassi-Rad.
13. <http://www.humnet.ucla.edu/humnet/alc/faculty/tangher.html>, Professor Timothy Tangherlini.
14. National Science Foundation grant 0970179, "EAGER: Network Pattern Recognition Project."
15. Note that the publication of Jacqueline Stone, *Original Enlightenment and the Transformation of Medieval Japanese Buddhism* (Honolulu: Kuroda Institute/University of Hawaii Press, 1999), contains a bibliography and description of the studies of this topic in Japan and elsewhere.
16. See the work of Robert Buswell, *Cultivating Original Enlightenment: Wōnhyō's Exposition of the Vajrasamādhi-Sūtra* (University of Hawaii Press, 2008). A review of the book by Charles Muller is found at http://www.acmuller.net/reviews/cultivating_original_enlightenment.html, originally published in *International Journal of Buddhist Thought and Culture* 10 (February 2008).
17. Lewis Lancaster, "The Question of Aprocryphal Words in the Chinese Buddhist Texts," paper presented at the Annual Meeting of the American Academy of Religion, Atlanta, GA, November 8, 1986.
18. Known as simply *Taishō*, this work was edited by Junjiro Takakusu and Kaikyoku Watanabe and published in a set of 100 volumes from 1924–1934 with the publishing house Daizo shuppan kabushiki kaisha in Tokyo.
19. Chinese Buddhist Electronic Text Association (CBETA), <http://www.cbeta.org>. See a description of this data set at <http://international.uiowa.edu/centers/caps/documents/HuiminBhiksuFromCBETAtBIP-IBA.pdf>.
20. The Saṃgaṇikīkṛtaṃ Taiśotripiṭakaṃ: The SAT Daizōkyō Text Database, <http://21dzk.l.u-tokyo.ac.jp/SAT/index.html>.
21. Even though the "software" presents these reports in terms of the page and columns of the Koryō blocks, most users do not yet have access to that data. Therefore, for this article, I list the *Taishō* references for convenience. 禪源諸詮集都序, T. 2015, 405b:28.
22. 金剛三昧經, T. 273, 368a:12. Also quoted in Wōnhyō's commentary, 金剛三昧經論, T. 1730, 976b:27.
23. 宗鏡錄, T. 2016, 443a:24.
24. 金剛三昧經, T. 273, 372b:23.
25. 佛說大淨法門經, T. 817, 818c:17.
26. 中阿含經, T. 26, 579c:2.

27. 阿毘曇毘婆沙論, *T.* 1546, 161a:1.
28. 菩薩瓔珞經, *T.* 656, 69b:1.
29. 仁王護國般若波羅蜜多經, *T.* 246, 836c:20 = *K.* 1340.
30. 佛說仁王般若波羅蜜經, *T.* 245, 826b:20 = *K.* 19.
31. 受菩提心戒儀, *T.* 915, 941b:19.
32. 大乘瑜伽金剛性海曼殊室利千臂千鉢大教王經, *T.* 1177A, 757c:7.
33. 大乘瑜伽金剛性海曼殊室利千臂千鉢大教王經, *T.* 1177A, 742b:24 = *K.* 1272.
34. 大乘理趣六波羅蜜多經, *T.* 261, 904c:5.
35. 大乘理趣六波羅蜜多經, *T.* 261, 902a:4.
36. *T.* 874, 310c:2–3.
37. *T.* 261, 906b:18.
38. *T.* 261, 912a:1.
39. 大方廣佛華嚴經, *T.* 293, 688a:8.
40. *T.* 293, 812a:18.
41. See note 17.

