Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

Gaetan Rappo*

Abstract

This article uses character n-grams methods to assess the authorship of three religious texts written in medieval Japan, comparing them to the works of Monkan (1278-1357), a Shingon monk active during the first half of the 14th century. Such texts belong to a literary genre called shōgyō that presents many challenges that render the use of traditional authorship attribution methods inappropriate. They were composed in Japanese kanbun (classical Chinese read in Japanese word order), and their contents are closer to cumulative work than to original creations by a single author. This article thus draws on previous research on the translators of the Chinese Buddhist canon, which have proven to be far more effective than traditional methods developed for modern languages for the analysis of this type of literature. Concretely, it proposes a workflow proceeding from the preparation of the corpus (with manuscript edition, encoding,...) to a concrete data analysis using the variable length n-gram method. The last part of the article deals with future perspectives that would further refine the results, such as taking into account the various speaking voices found inside the shōgyō and their relationship to the author, as well as stylistic analysis based on grammatical patterns. As a whole, the experiment succeeded in showing global trends in the texts of the Japanese Shingon schools, finding stylistic differences between the works of Kūkai, of monks from the 12th century, and of Monkan. Combined with a rigorous historical enquiry into the redaction context of the texts and their manuscripts, the data analysis also demonstrates that one of them was almost certainly written by Monkan, and another either by him or by one of his close disciples.
During the late Heian and medieval periods in Japan, the esoteric schools of Buddhism produced an extremely large corpus of religious texts, which specialists call “sacred teachings” (shōgyō 聖教).¹ This category includes an extremely diverse array of texts, including doctrinal commentaries on Buddhist texts, ritual manuals, records of oral transmission, and initiation certificates.

This little-studied literature presents several challenges when attempting to identify the author of any given text. The first of these challenges is the language. While these texts are based on the Buddhist canon—which is itself written in classical Chinese—they were actually composed in ancient Japanese, using the Chinese way of writing (this form of composition is called kanbun 漢文). This unfortunately makes it nearly impossible to reliably utilize Western word-based methods of authorship attribution. Even character-based methods (see Peng et al. 2003) were designed largely for modern Chinese, and do not take into account the linguistic characteristics of this genre. The second major challenge is stylistic, as these texts often consist of a succession of quotes from previous works, and use a well defined manner of expressing themselves that leaves little room for originality.²

However, within the last few years, new methods for identifying the translators—rather than the authors—of Chinese Buddhist texts have achieved great success through a careful analysis of recurrent sequences of characters (n-grams) and of their relative frequencies (see Hung, Bingenheimer, and Wiles 2010). Such sophisticated methods have not yet been applied to shōgyō, however. This article will apply these methods in analyzing the work of one monk in particular—who happened to have been largely misunderstood, in a manner that mirrors authorship attributions in an interesting way.

¹Graduate school of Letters, Nagoya University
²This is partly why a previous experiment presented at the JADH conference in 2016 using the character n-gram methods proposed by Stamatatos (2009) for plagiarism detection was inconclusive. Optimized versions of this method, as in Kuta and Kitowski 2014, could lead to better results, but as this paper demonstrates, the variable length n-gram method appears more suited to the type of texts used here. Moreover, Stamatatos uses a supervised algorithm, which implies training datasets with preclassified samples. This approach can be seen as orienting the data in a certain way. On this issue, see Bingenheimer, Hung, and Hsieh 2017, 4.
Monkan (1278–1357), a monk of the Shingon school of esoteric Buddhism, was a major figure active during the first half of the fourteenth century. Originally from Saidaiji 西大寺 in Nara, Japan, he moved to Daigoji 醍醐寺 during the 1310s, and became an influential figure at the imperial court. A close advisor to Emperor Go Daigo 後醍醐 (1288–1339), he retained an influential role in the regime of the southern court, as both a ritual master and a confidant of its sovereigns. However, after his death, Monkan was widely criticized by members of his own school. Sources depict him as a horrible heretic, a member of the sinister Tachikawa lineage (Tachikawa-ryū 立川流), who had seduced the emperor with his mastery of dark, sexual rites. The rediscovery of his work over the last decade has allowed researchers to reevaluate this depiction. These efforts showed that—far from being a heretic—he was actually a follower of the broader doctrinal and ritual trends of the medieval Shingon school, and especially of Daigoji Temple.3

Monkan’s reputation demonstrates that traditional perceptions of certain historical events often differ from the facts. Similar discrepancies are found in the intellectual history of the esoteric schools in medieval Japan. In fact, aside from a lack of interest until recently, several factors contributed to blurring the image of this particular period of Japanese religion. These include the limited availability of sources—most of which are manuscripts within the custody of difficult-to-access monastic libraries; the later emergence of a sectarian view of Japanese Buddhism that would largely determine how each current viewed its own fabricated past; and last, but not least, the tendency of medieval esoteric monks to attribute their own work to other, more prestigious masters.

The main objective of this article is to explore the potential of using digital methods as an auxiliary tool for researchers attempting to recover such history. Concretely, it will use n-gram methods to assess the authorship of three religious texts that are traditionally believed to have been composed by different monks between the twelfth and fourteenth centuries. However, at least two of these texts were likely the work of Monkan. After presenting the results of this analysis, I will also suggest possible ways to further refine these results.

1. Previous Research on the Authorship of the Target Texts

3. See Abe 2013; Quinter 2015, 181–85; and Rappo 2017b.
The three target texts of this study were not selected randomly. Previous scholarship has either identified Monkan as their author, or demonstrated that they were produced by people fairly close to him.

This attribution is indeed the case with the first text, the Goyûigō hiketsu 御遺告秘決. It is included in the famous collection of edited Shingon texts called the Zoku Shingonshû zensho (vol. 28) and attributed to the monk Jichiun 実運 (1105–1160). However, the same colophon used as a basis for this claim dates the text to 1162, two years after Jichiun’s death. This has led several researchers to doubt the attribution. Satoshi Itô, in his research on medieval Shintô and its links to Shingon Buddhism, has suggested that the text could have been authored by the little-known Dōjun 道順, Monkan’s master at Daigoji, or by someone close to him (Itô 2011, 275). More recently, Yasurô Abe has—through his familiarity with Monkan’s texts—strongly argued that this attribution was erroneous and that Monkan should be seen as the text’s true author (first in Abe 2006b, but also Abe 2009 and Abe 2013, 244).

The second text is the Daijingu honji 大神宮本地, also known as the Daishi gongyûjôki 大師御入定記. This text contains many similarities to Monkan’s work, especially on the topic of the Joint Ritual of the Three Worthies (Sanzon gôgyôhô 三尊合行法). Three known manuscripts of this text are extant. One conserved in the Hôbodai’in of the Tôji temple includes a colophon by a certain Kôken 光賢, who claims that he copied the text from an earlier source. His identity has been debated, however. Kamakura period texts mention one Kôken, a disciple of the Daigoji monk Kenjin 憲深 (1193–1262) who was active during the mid-thirteenth century. However, Satoshi Itô (2011, 260) has demonstrated that the author was most likely another Kôken, born in 1344. Linked to the esoteric lineage of Mount Miwa, he was a monk of Uchiyama Eikyûji 内山永久寺, a temple known for its links to the southern court and to Monkan. Moreover, Kôken also copied a text called Tenku-ji-sho 天口事書 in 1372, at Amanosan Kongôji temple 天野山金剛寺, which was for a brief time the capital of the southern court (Nagashima 1942, 253). Kôken enjoyed a strong association with this regime, as he remained close to emperors Chôkei 長慶 and Go-Kameyama 後亀山.

All of this points to a powerful connection with Monkan, who himself died at Kongôji in 1358, after having transmitted his knowledge to the monk Zen’e 禪惠. The link

4. On this text, see the edition and commentary in Rappo 2019. See also Itô 2011.
5. On this lineage, see Andreeva 2017.
between Kōken and Monkan is also confirmed through another document, since Kōken copied one of Monkan’s texts—the Saigoku himitsušō 最極秘密抄—in 1392. Kōken based this copy on a Daigoji manuscript he probably gained access to with the help of Kyōken 敦賢, his master at the Eikyūji (Abe 2009, 123). This Kyōken was in fact directly linked to Monkan’s disciple Hōren 宝蓮 (1302–after 1365), as he is mentioned in the colophon of the Yuga dentō-shō 瑜伽伝灯鈔, a text compiled in 1365 to showcase the teachings of Monkan’s lineage (Rappo 2017b, 121–22). All of this suggests a close proximity between the Daijingū honji and figures closely related to Monkan. The idea of Monkan as the original author of this text is thus historically possible.

The last text is called the Ben’ichisan himitsuki. It describes a ritual called the Ichibutsu nimyōō 一仏二明王 (One Buddha and Two Wisdom Kings), which may well have been a prototype of Monkan’s Three Worthies ritual (see Abe 2006b). It is believed to have been written by the eleventh-century monk Ningai 仁海 (951–1046), but research by Kazuhiro Fujimaki and Keiichi Uchida has demonstrated that it was authored by monks from the same temple as Monkan and was completed during the thirteenth century. This can be seen in the colophons of the manuscript held at Hikone-ji. The first claims that it was copied from Kenjin’s personal copy in 1250. This might indicate that the text was initially composed by Kenjin.

However, it is also possible that Monkan was linked to the text’s conception. Another colophon mentions a copy made by a certain “Ono sōjō” 小野僧正, a title usually attributed to Ningai. However, as Keiichi Uchida has shown, it can just as easily refer to Monkan. In this case, the date given by the colophon (1290) would be far too early, but Uchida mentions the possibility of a copyist error, which would make 1301 the more likely date. This would still be a little early for Monkan to be the author of this text, but it demonstrates that he most certainly used it.

Taken together, historical evidence thus relates that the Ben’ichisan himitsuki is

---

6. On this text, see Fujimaki 2002a and Fujimaki 2002b. I used my own edition of the Hikone-ji manuscript.

7. A copyist may have misread year 3 of Shōan 正安三年, 1301, and written year 3 of Shōō 正応三年, 1290. See Uchida 2014, 44.

8. There are other texts with colophon bearing the name Ono sōjō that could be attributed to Monkan. On this subject see Uchida 2015.
probably further from Monkan’s own work and might have been one of his sources. Our analysis confirms this, and demonstrates that it was probably not one of his own works. However, in doing so, our analysis also sheds new light on its true place in the history of medieval Shingon Buddhism and Shintō.

2. Analysis Method and Previous Research
The method used in this article is directly inspired by the NGSM (N-gram-based System for Multiple Documents), using character n-grams. This method has been used in Buddhist studies in Japan since the 1990s, and variations of this method have been utilized by scholars following the pioneering work of Kōsei Ishii and Shigeki Moro.9 While word-based n-grams are common in the analysis of Western languages, such studies focus on character-based n-grams because the original texts lack spaces, and are thus very difficult to parse automatically.

As stated in the introduction, this article relies especially on the research of Hung, Bingenheimer, and Wiles on the identity of the translators of several canonical texts. Their 2010 article analyzed n-grams within the Madhyamaka Āgama collection, from the Buddhist canon. They were able to give additional weight to a hypothesis first proposed by Kōgen Mizuno on the actual translators of a group of 24 out of 72 texts related to this canonical source. In a 1969 publication, Mizuno judged through his extensive reading of the texts that the actual translators of the 24 works were not those attested by the traditional catalogs, but rather the famous Nie Daozhen, with the help of Dharmanandin (Mizuno 1969; see Hung, Bingenheimer, and Wiles 2010).

Their method—to which I will return in section 5—proved to be extremely useful in assessing recurring vocabulary or phraseology in Buddhist texts, but it has not yet been applied to the analysis of non-translated texts, and especially to a genre as composite and complex as the shōgyō.

3. Corpus
The research described in this article applies the aforementioned method to a corpus of texts that remain untranslated. To properly assess the authorship of these texts, I selected an initial

group consisting of texts known to have been written by Monkan.

**Table 1: Corpus**

<table>
<thead>
<tr>
<th>Contents</th>
<th>Date and manuscripts</th>
<th>Length (in characters)</th>
<th>(in) Edition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Saigyokushō</strong> 西玉鈔</td>
<td>Record of Monkan’s esoteric initiation at Saidaiji</td>
<td>1314</td>
<td>7655</td>
</tr>
<tr>
<td><strong>Goyuigō shichika daiji</strong> 御遺告七箇大事</td>
<td>Seven secret rituals described as taken from the posthumous Testament of Kūkai as rites of the Jewel</td>
<td>Written in 1321; held in the Chizan bunko library of Chishaku’in Temple.</td>
<td>3678</td>
</tr>
<tr>
<td><strong>Goyuigō hiketsu (long version)</strong> 御遺告秘決</td>
<td>Secret teachings on the Joint Ritual of the Three Worthies, based on the Testament</td>
<td>1327</td>
<td>7217</td>
</tr>
<tr>
<td><strong>Goyuigō daiji</strong> 御遺告大事</td>
<td>Oral teaching on the two versions of the Joint Ritual, and images of the deities</td>
<td>1327, on the same day as the previous document</td>
<td>8913</td>
</tr>
<tr>
<td><strong>Himitsu gentei kuketsu</strong> 秘密源底口決</td>
<td>Oral teachings on the Joint Ritual of the Three Worthies, especially the separated form</td>
<td>Recopied in 1338, probably written in the late 1320s</td>
<td>3930</td>
</tr>
<tr>
<td><strong>Ichinisun gōgyō hishidai</strong> 一二寸行秘次第</td>
<td>Ritual procedure for the separate version of the Joint Ritual of the Three Worthies</td>
<td>Recopied in 1338, probably written in the late 1320s</td>
<td>3385</td>
</tr>
<tr>
<td><strong>Saigoku himitsu shō</strong></td>
<td>Large doctrinal</td>
<td>Around 1337</td>
<td>8422</td>
</tr>
</tbody>
</table>
Of all Monkan’s work, two known and available texts were omitted, the first being the *Chū rishukyō* 注理趣経. This text was included in the *Nihon Daizōkyō*, but its edition is too unreliable for authorship attribution testing. The same can be said of the *Kinpusen himitsuden* 金峯山秘密伝, which may not have been conserved in its original version (see Satō 1966). Other texts by Monkan exist, but they are either unedited or unavailable.
The second group of texts will be used as a basis for comparison. This consists of the work of Kūkai 空海 (774–835) contained in the Taishō canon. The clearly apocryphal Goyuigō 御遺告 was also included, as it served as one of the most important sources of Monkan’s thought. Furthermore, it also demonstrates if the model is able to identify texts not written by Kūkai.

The third group consists of texts written by monks of the Daigoji lineage of the Shingon school during the late Heian and early Kamakura periods. This includes three particular texts by Jichiun, who was—as said above—said to have written the Goyuigō hiketsu (short version).

The last group is designed to better assess the place of the Daijingū honji and Ben’ichisan himitsuiki within the realm of medieval Shintō. It contains major Shintō texts, such as the Tenshō daijin kuketsu 天照大神口決, a work by Kakujō 觉乗 (ca. 1273–1368), the third abbot of the Saidaiji temple, who was also very close to Monkan.

4. Data Preparation
Before explaining the methods used and issues encountered, it is necessary to describe how the data was gathered. Until recently, no edition of Monkan’s works existed, including for two of the three target texts. The first step was to read the manuscripts and manually copy their contents.

Japanese kanbun manuscripts contain much more information than the base text. This includes indications for the Japanese reading order (yomikudashi), as well writing in the margins and on the reverse. The Japanese reading order can be quite helpful as it restores the potential original reading of the text in its native language. However, indications tend to differ in each manuscript, and unless access to the original versions is available, the manuscripts remain extremely subjective. Writings on the reverse (uragaki 裏書), however, were not necessarily written by the same author, and are not generally useful in authorship attribution. I eventually intend to include all of this in future studies, as it would open many possibilities. However, for now, following previous scholarship on the author attribution of Sino-Japanese Buddhist texts, I will use only the main text in its Chinese reading order.

Each text was turned into a plain-text file containing only the Chinese characters, without punctuation or spaces. The following workflow was used:

12. On this, see Rappo 2018h.
A. Choosing the most reliable manuscript edition of each text, including the original colophons (*hon’okugaki 本奥書*).
B. Normalizing variant characters into their standard versions.\textsuperscript{13}
C. Using a simple Perl script to convert Japanese simplifications, when used in the original text, into traditional Chinese characters (*kyūkanji 旧漢字*).
D. Standardizing the character encoding to Unicode (UTF-8).
E. Replacing Siddham script (*bonji 梵字*) with corresponding Chinese characters. While Siddham has been recently added to Unicode, a comprehensive Unicode Siddham font has yet to be created.

\textbf{5. N-gram Extraction Method: The Variable-Length N-grams}

Generally, this study follows the workflow proposed by Hung, Bingenheimer, and Wiles (2010). Using a Python script,\textsuperscript{14} the n-grams of all the texts were extracted first. Then, a threshold of occurrences was determined in order to construct a feature set of n-grams. This means that n-grams not appearing within a sufficient number of texts have not been selected. Since the texts being analyzed contain great similarities in vocabulary, the threshold was set to a higher level than by Hung, Bingenheimer, and Wiles (2010), meaning that the selected n-grams appeared in at least 20\% of the texts. Although the texts share several common characteristics, this still gave feature sets of a size ranging from around 2,000 to 9,993 grams.

The second step was to apply another script in order to utilize the variable length n-gram method mentioned above. Variable length n-grams are advantageous in that the same expression is not counted twice when it appears in another, longer n-gram. So, for example, if a given text contains seven occurrences of the term “sanzon” 三尊 and two of “Sanzon gōgyōhō” 三尊合行法, this means that the seven occurrences of “sanzon” already contain the two “sanzon” appearing in the term “Sanzon gōgyōhō.” In order to avoid exaggerating the importance of the term “sanzon” alone, its frequency has to be corrected—in this example, by subtracting two (see Table 2). This applies to all other n-grams included within any larger

\textsuperscript{13} For example, “収”—a variant of the character “sai” 最 frequently used in medieval Japanese manuscripts—was replaced with its more common counterpart.

\textsuperscript{14} The complete script was designed with the help of IT engineer Stéphane Lapie. It is based on the work of Joey Hung and Markus Bingenheimer for their 2010 article. I thank them for providing the scripts and data from their experiments.
n-grams. The principle can be summarized in the following table, adapted from Hung, Bingenheimer, and Wiles (2010):

**Table 2: Example of variable length n-gram method**

<table>
<thead>
<tr>
<th>Gram</th>
<th>Text1</th>
<th>Text2</th>
<th>After adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>三尊合行法</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>三尊合行</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>三尊合</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>尊行法</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>三尊</td>
<td>7</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>尊合</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>合行</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

After having extracted all the n-grams of the feature set, their relative frequency inside a given text was calculated using the above formula, which divides the number of occurrences by the length of the text. This can be seen in the following formula (see Hung, Bingenheimer, and Wiles 2010):

\[
F(d, f_{S_x}) = \frac{C(d, f_{S_x})}{L_d}
\]

Here \(f_{S_x}\) describes an n-gram of the feature set and \(F(d, f_{S_x})\) is the frequency of the n-gram \(f_{S_x}\) in text \(d\). \(L_d\) corresponds to the length of document \(d\), and \(C(d, f_{S_x})\) is the frequency of the n-gram \(f_{S_x}\) in \(d\). In this experiment, the final number was further multiplied by 100 in order to avoid working on too small a scale.

Three analyses were performed with different corpuses. The first (A) included just Monkan’s work and the three target texts. The second (B) added Kūkai’s works. The third (C) compared Monkan’s and the target texts to works of Shingon monks from the medieval period. The last (D) included Shintō works.

A sample of the most common n-grams in Monkan’s work and their compared frequencies in Kūkai and Daigoji texts can be found below. The table shows that some of
Monkan’s most frequently used expressions (such as hōshu 寶珠) are completely absent from Kūkai, and he shares a larger part of his vocabulary with the Daigoji texts. As the Daigoji texts are closer to him both temporally and doctrinally, this is not surprising.

Table 3. Most common n-grams in Monkan’s texts

<table>
<thead>
<tr>
<th>N-gram</th>
<th>Average Monkan *100</th>
<th>Average Kūkai *100</th>
<th>Average Daigoji masters *100</th>
</tr>
</thead>
<tbody>
<tr>
<td>寶珠</td>
<td>0.190393877</td>
<td>0.000247561</td>
<td>0.014943543</td>
</tr>
<tr>
<td>大師</td>
<td>0.131475375</td>
<td>0.005479055</td>
<td>0.014818168</td>
</tr>
<tr>
<td>本尊</td>
<td>0.110283609</td>
<td>0.004632978</td>
<td>0.068879229</td>
</tr>
<tr>
<td>不二</td>
<td>0.097648525</td>
<td>0.005708021</td>
<td>0.000477561</td>
</tr>
<tr>
<td>愛染</td>
<td>0.088629724</td>
<td>0.000247562</td>
<td>0.008694612</td>
</tr>
<tr>
<td>六大</td>
<td>0.078585432</td>
<td>0.044482079</td>
<td>0.036832980</td>
</tr>
<tr>
<td>々々</td>
<td>0.07521041</td>
<td>0</td>
<td>0.019586369</td>
</tr>
<tr>
<td>不動</td>
<td>0.074470096</td>
<td>0.009612421</td>
<td>0.033855667</td>
</tr>
<tr>
<td>法界</td>
<td>0.070224635</td>
<td>0.058678932</td>
<td>0.01685722</td>
</tr>
<tr>
<td>大事</td>
<td>0.068710886</td>
<td>0.000495123</td>
<td>0.010444794</td>
</tr>
<tr>
<td>也即</td>
<td>0.064564609</td>
<td>0.001739984</td>
<td>0.002425489</td>
</tr>
<tr>
<td>兩部</td>
<td>0.064171434</td>
<td>0.009490394</td>
<td>0.013159857</td>
</tr>
<tr>
<td>故云</td>
<td>0.06379245</td>
<td>0.040224162</td>
<td>0.001380701</td>
</tr>
<tr>
<td>三尊</td>
<td>0.06309266</td>
<td>0.002249213</td>
<td>0.005873264</td>
</tr>
<tr>
<td>也此</td>
<td>0.060212947</td>
<td>0.019166717</td>
<td>0.032013403</td>
</tr>
<tr>
<td>無量</td>
<td>0.060176532</td>
<td>0.050585765</td>
<td>0.012697104</td>
</tr>
<tr>
<td>一心</td>
<td>0.059312078</td>
<td>0.024107653</td>
<td>0.000604654</td>
</tr>
<tr>
<td>灌頂</td>
<td>0.058989385</td>
<td>0</td>
<td>0.037581233</td>
</tr>
<tr>
<td>塔印</td>
<td>0.057515801</td>
<td>0</td>
<td>0.002472249</td>
</tr>
<tr>
<td>基秘</td>
<td>0.05637847</td>
<td>0.01485802</td>
<td>0.020616306</td>
</tr>
</tbody>
</table>

Taken together, these data suggest that some differences exist between the various groups of texts analyzed. However, data alone can be difficult to interpret. In order to evaluate the meaning of these results, graphics were created using the PCA method through a factor reduction. Concretely, this means that the millions of possible relations between the frequencies of each n-gram in each of the target texts were reduced to two-dimensional...
graphs, providing a better look at the data from a more meaningful perspective. This produces two components—which are simply abstractions of the differential between the frequencies established by the PCA calculation—that form the two graphical axes.\footnote{A factor dimension reduction with a varimax rotation was used. On the use of PCA in this context, see Hung, Bingenheimer, and Wiles 2010.}

### A. Monkan’s Works and the Target Texts

The results are already quite interesting when looking only at Monkan’s texts. They show that the model is able to distinguish between the various types of works that Monkan authored. For example, the four dots in the upper side on Figure 1 are all ritual procedures, *shidai* 次第. The works belonging to this category describe the precise steps of a ritual ceremony, from the construction of the altar to the gestures and words spoken to summon, greet, and honor the deities. They thus share several common characteristics. This commonality appears in every stage of this experiment, and such ritual works can thus be considered to form a group of their own.

**Component 1**

Figure 1. Monkan and the target texts (T4)

One of these ritual texts, the *Ichinisun gōgyō hishidai* 一二寸合行秘次第, describes a ritual called the Joint Ritual of the Three Worthies (*Sanzon gōgyō-hō*).\footnote{An edition can be found in Abe 2006a. On the Three Worthies, see Abe 2006b, 2009, and 2013, 242–50; Dolce 2010, 176–89; Trenson 2016, 406–14; and Rappo 2018a. For the iconography, see Abe 2011a, Rappo 2017a, and Dolce 2008.} The ritual consists of three steps, dedicated to two Wisdom Kings—Aizen and Fudō—flanking a central figure: either Kannon or the Wish-fulfilling Jewel (*nyoi hōshu* 如意宝珠). Another variation honors a combined form of the three, incarnated by Kūkai, the founder of the Shingon school. It can be found in the *Jigyō shidai* (which is, however, a very brief text, describing only the ritual itself).

The Joint Ritual is the central theme of a large part of Monkan’s work. More importantly for us, it is the main subject of a large proportion of the works appearing in the right part of our graph. It is thus significant that they can be automatically distinguished from
other texts by Monkan that appear in the center. This is another—and more subtle—differentiation inside Monkan’s work that appears in every stage of this analysis.

The position of the Daijingū honji, on the left of this group, shows that it is at least closely related. In fact, it seems even closer to the fundamental texts of the Three Worthies\textsuperscript{17} than the Sanzon gōgyō hiketsu, a text composed by Monkan, but probably later than the others.\textsuperscript{18} It also comes as no surprise that the Goyuigō hiketsu (short version) is firmly placed within the same group.

This is not the case for the Ben’ichisan himitsuki, which appears on the left next to an early work by Monkan, the Goyuigō shichikadaiji. Both texts largely share the same subjects: the rituals and deities of Murōji 寺室生 Temple (called Benichi’san in esoteric texts). However, this seems to confirm the idea that the Ben’ichisan himitsuki is more distinct from Monkan’s texts than the other two target works used in this article.

\textbf{B. Monkan and Kūkai}

The data presented in Figure 1 would be meaningless if the model were not able to identify 17. For a partial list, see Abe 2013, 246. The Daigoji version of the Goyuigō hiketsu probably also belongs to that group.

18. On this text, see Rappo 2010b and 2018b.
authorship attributions. Adding an author with a clear writing style as a basis of comparison brings several interesting issues to light. In this experiment, I used the Taishō editions of the *Jūjūshinron* 十住心論 (T. 2426), the *Benkenmitsu nikyōron* 辨顯密二教論 (T. 2427), the *Sokushin jōbutsugi* 即身成仏義 (T. 2428), the *Shōjijissōgi* 聲字實相義 (T. 2429), the *Unjigi* 吞字義 (T. 2430), and the *Goyuigō* 御遺告 (T. 2431). As can be seen in Figure 2, works attributed to Kūkai—with the exception of the *Goyuigō*—display a certain homogeneity.

Kūkai’s texts each appear after around 0.25 on the second component line, and they form a group clearly distinct from Monkan’s works. This time, Monkan’s texts are divided in two groups, with the Joint Ritual of the Three Worthies on the right and the others near the center.

The only anomaly is the *Goyuigō*, which appears in the middle of Monkan’s doctrinal texts and ritual procedures. This is not a misidentification, but rather, a very logical and precise placement. While attributed to Kūkai, the *Goyuigō* is actually a famous apocryphal work, probably written during the tenth century. It also happens to serve as one of the major foundations of Monkan’s own doctrine and rituals.

Figure 2 also demonstrates that there is a difference between the way Kūkai wrote during the eighth century, and the way Monkan wrote during the fourteenth. While the *Goyuigō* was written much closer to Kūkai from a strict chronological point of view, it still shares many common characteristics with medieval or pre-medieval texts.\\(^{19}\)

---

\(^{19}\). On the influence of the *Goyuigō*, see Tomabechi 2010.
C. Monkan and Works from the Early Medieval Shingon School

The third stage of our analysis involved a larger corpus. It included not only Monkan and Kūkai, but also works by famous monks of the early medieval Shingon school. Among them we find the prince-monk Shūkaku 守覚法親王 (1150–1202), one of the most influential figures of the Shingon school during the twelfth century. Since Shūkaku was not directly a member of the Daigoji lineages, texts by masters of this branch—and especially of its Sanbōin sub-branch—were added. One person of particular interest in this subset is Jichiun, the alleged author of the Goyuigō hiketsu (short version). Three of his works—the Shoson yōshō 諸尊要抄 (T. 2484), the Hizō konpōshō 祕蔵金寶抄 (T. 2485) and the Genpishō 玄祕抄 (T. 2486)—were added, as well a text by Shōken 勝賢 (1138–1196), the Jishōki 治
The results are harder to interpret on Figure 3. The model does succeed in distinguishing clearly defined groups of texts, such as Monkan’s Three Worthies texts on the upper part of component 2, and the works of the Daigoji masters on the right of component 1. However, Monkan’s ritual procedures and doctrinal treatises are mixed among Kūkai’s works, which the model fails to identify here as an independent ensemble.

This problem can be explained by the fact that most of the analyzed texts are longer than Monkan’s and thus represent a larger set of data. Another reason is that the method designed by Hung, Bingenheimer, and Wiles (2010) was intended to compare stylistic differences between two groups of texts.

However, these difficulties can be solved with a careful reduction of the number of texts used in the comparison. In another variable n-gram and PCA analysis (Figure 4), the inquiry was narrowed to Monkan’s texts and the works by the masters of the Daigoji temple mentioned above. Works by Kūkai and Shūkaku were removed from the dataset, but the three texts by Jichiun, as well as two documents by Shōken and Seigen, were retained. This particular dataset was selected for two reasons: first, to verify if the model can find differences between Monkan’s work and similar, but earlier, texts; and second, to see where the Goyuigō hiketsu (short version) appears compared to Jichiun’s work.
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

With this better-calibrated data, the picture becomes clearer. Monkan’s work remains grouped with two large ensembles at the center: the top one being the texts related to the Joint Ritual of the Three Worthies, and the other, in the center, composed of earlier texts, doctrinal treatises, and ritual procedures outside the Three Worthies. Texts by Jichiun appear to the far right, while works by Shōken and Seigen are located closer to the center. This means that they all have a higher first component rating than Monkan’s work. In fact, a clear line dividing Monkan’s work and texts by other masters of Daigoji can be drawn at around component 1, 0.25.

The three target texts are positioned relatively similarly to where they fall in Figure 2, with Kūkai. This confirms that the Goyuigō hiketsu (short version) stands firmly in the middle of the Three Worthies group. It thus appears to be very distant from Jichun’s other known works. This also demonstrates that the attribution to Jichun is most certainly a fabrication, probably by Monkan himself.

The Daijingū honji stands between the two groups of Monkan’s works on the line of the second component. While this does not prove that Monkan was its author, it certainly
demonstrates that Daijingū honji is closer to him than to earlier Shingon texts. This gives further weight to Itō Satoshi’s identification of the first copyist of the text, Köken, as a fourteenth-century monk.

The position of the Benichi’san himitsuki is still similar. This stage, like the previous one, does not provide any further information on this text. More sources on the topic of Murōji and its rituals are probably needed to further assess its position relative to Monkan’s work.

D. Monkan and Medieval “Shintō” Texts

To better ascertain its position, an analysis was performed with two Shintō-related texts that happen to be quite similar to those of Monkan and also to the two target texts. One is the Benichi’san hiki, a text about the legends of the Murōji temple that derives from the Zuishin’in, a different Shingon lineage than Monkan’s.20 The second is the Tenshō daijin kuketsu, a text written in 1327 by the Saidaiji monk Kakujō, one of Monkan’s close disciples.21

In combining the previous three stages with a new n-gram extraction including both Monkan’s work and important medieval Shintō texts, the relative position of our target texts does not change much. Figure 5 demonstrates that adding those Shintō texts to our dataset of Monkan’s work creates a crowded center. The explanation is simple. Each of these texts—including the Goyuigo shichika daiji by Monkan—mainly describes the deities and rituals of the Murōji temple, or closely related subjects (Tenshō daijin kuketsu). This shows that the model is not very apt at analyzing more than two groups of texts from a single dataset. In a way, one can say that the pure lexical differences inherent in the extremely predetermined genre of ritual procedures have a large enough impact on the calculations that they disguise eventual stylistic discrepancies between Monkan’s more conventional work and our target texts.

In order to resolve this problem, another analysis was performed without the ritual procedures (see Figure 6). Only the Ichinisun gōgyō hishidai—which contains a lot of oral transmission and no ritual techniques—was unremoved. These results become more

20. On this text, see Fujimaki 2002a. His edition was used in this study.
21. Yasurō Abe’s edition (Abe 1997) was used in this study.
enlightening. This model shows that the *Ben’ichisan himitsuki* differs substantially from Monkan’s texts, and that it is far closer to the *Ben’ichisan hiki*, a work produced by a rival lineage of the Shingon school.

One could argue that this is just a matter of a similar vocabulary found in both texts, but our previous experiment has shown that the similitudes in n-gram frequencies between those texts and Monkan’s doctrinal or narrative work (especially texts mentioning the Murōji rituals, gods, and legends) were actually higher than among his rituals procedures and the rest of his texts. This suggests that the model actually shows a deeper difference, which is in line with the traditional dating of the *Ben’ichisan himitsuki* by specialists who see it as having been written in the mid-thirteenth century.

The model also suggests that the *Tenshō daijin kuketsu* shares many similarities with Monkan’s work, especially on the thematic level, while not being completely identical on the stylistic side. This should not be a surprise, as it was written by someone close to Monkan, but not by Monkan himself.
This characterization of the *Tenshō daijin kuketsu* helps us better understand the place of the *Daijingu honji* in his model, which is in fact consistent with each of the analyses. This work stands directly between two groups of texts: those describing the Joint Ritual of the Three Worthies along with Monkan’s earlier work, and Monkan’s broader doctrinal treatises (with the exception of the *Ichinisun gōgyō hishidai*, in which the writing style is very different from the other Three Worthies texts).

While this does not prove conclusively that Monkan was the author of the *Daijingu honji*, it adds further likelihood to this very probable hypothesis. In his analysis of the text, Satoshi Itō had already suggested that it was quite close to the *Goyuigō hiketsu* (short version), and that he sees it as having been transmitted within the same circles (Itō 2011, 260). His research has also shown that the *Daijingu honji* is in fact a composite text, constructed of nine different sections. Many of them relate to Monkan’s ideology. One of them is almost

---

22. For a summary of the first four parts of this text and their links to the Three Worthies, see Itō 2011, 260–63. For the text itself, see Rappo 2019.
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

identical to his version of the Joint Ritual of the Three Worthies. Another uses a metaphor of Venus entering Kūkai’s mouth found in the *Goyuigō* to assert the master’s identity as a fundamental incarnation of nondualism (*funi 不二*). This is the main idea behind the version of Monkan’s ritual that uses Kūkai as a main object of worship.23 Yet another section discusses a very interesting image of a triad formed by Kūkai flanked by the deities Niu and Kōya myōjin. This pattern is absent from other known documents written by Monkan, but it remains quite close to the trifold structure of his ritual. An even more striking section provides a long discussion of the idea of the three minds of awakening (*sanbodaishin 三菩提心*) in association with the sun, the moon, and the Wish-fulfilling Jewel, a concept used often by Monkan.24 Itō describes it, quite aptly, as the summary of a very similar chapter found in the *Goyuigō daiji*, one of Monkan’s most important texts.25

Aside from the textual and data-based evidence, historical reasons further support the idea that Monkan was the author of the *Daijingū honji*. As stated in the introduction, this text was copied by Kōken, a monk who is known to have copied several other texts authored by Monkan. Kōken’s master was also the monk known as Kyōken 敎賢 of Sairinji 西輪寺—a monk who was very close to Monkan’s disciple Hōren, and who also copied several of Monkan’s texts.26 This shared transmission route between Monkan’s work and a text that Kōken claims to have copied—but not written—is another hint to its real author.

All of this firmly suggests either that Monkan was the real author of the *Daijingū honji*, or that it was composed by someone very close to him, as well as being based on his teachings. While this data-based analysis cannot establish complete proof of this attribution, it provides a strong argument that—combined with the aforementioned historical reasons—makes a compelling case for Monkan as its author.

24. On Monkan and the three minds of awakening, see Dolce 2010, 182.
25. Itō 2011, 262. In this book, Itō again sees Dōjun as the probable author of this text. Subsequent studies have shown that Monkan was most certainly the real author, on the basis of his master’s teachings. For an edition of and commentary on this text, see Rappo 2018b.
26. On Kyōken and Kōken, see Rappo 2017b, 121–23. Kyōken was probably initiated into Monkan’s lineage, but he is not included in Hōren’s list of Monkan’s direct disciples (compiled in 1365). For the complete list, see Uchida 2003, 6–7. See also Uchida 2004.
5. Limitations of the Results and Future Research Possibilities

While this analysis provides compelling evidence to assist in identifying the real author of at least one, if not two, of our target texts, it remains difficult to prove beyond doubt the authorship of a given text within the context of medieval Japanese monastic literature.

An additional important element in Monkan’s authorship is related to the very nature of this literature, and to its peculiar writing style. In fact, an analysis of common characters in Monkan’s Three Worthies-related texts shows that nearly 75% of their contents are identical (see Figure 7, below). The Goyuigō hiketsu (short) is 85% identical. Such results are quite impressive, given the variety of Chinese characters. Adding the Daijingu honji and the Ben’ichisan himitsuki does lower these percentages to around 68% for them and 70% for Monkan’s texts.

When compared with random Shingon texts (Taishō 2411, 2413, 2414, 2415, 2416), the lowest number is 58%. This demonstrates the homogeneity of Monkan’s Three Worthies-related texts, but also suggests that the genre of shōgyō—and especially the writings of the medieval Shingon school—maintained a limited variety of style and vocabulary. The shōgyō are cumulative works that rely heavily on citations and do not possess real individuality.

In fact, Monkan used language that was fairly restricted to a certain group of people. Unfortunately, this precise language—used by the monks of Daigoji from the thirteenth through the fourteenth centuries—is not well represented in the Taishō canon. Most of their work also takes the form of ritual manuals. Such texts are full of quotations and predetermined formulas, and they leave even less space for individual inventiveness. However, while the very notion of authorship is debatable in this context, this experiment has demonstrated that Monkan’s texts indeed contained some characteristics that distinguished them from other Shingon texts from a similar period, even when they shared a common subject.

Figure 7 (the columns are displayed in the same order as the text names’ below, from left to right)

6. Levels of Discourse and Speaking Voices in Shōgyō Literature

To achieve better results in future work, it will be necessary to obtain more data—for
example, a text version of the *Shingon shū zensho* and other similar anthologies of Shingon texts. Better editorial practices for manuscripts—which are still conceived with paper editions in mind—are also crucial.

Another potentially fruitful way to refine these results would be to implement a method that takes into account a very important feature in the *shōgyō* literature: the level of discourse. Monkan rarely presents his texts as his own voice, but relies heavily—like most of his contemporaries—on the authority of previous masters or canonical sources. Therefore, sentences in his texts are frequently introduced by regular figures of speech.\(^{27}\) Table 4 demonstrates an example based on Monkan’s work.

**Table 4. Speaking voices in *shōgyō***

<table>
<thead>
<tr>
<th>Type of discourse</th>
<th>Introduced by</th>
<th>Contents</th>
<th>Ends with</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical source</td>
<td>経（文）云（generally provides the title of the sūtra or commentary）</td>
<td>Usually verifiable quote from canonical texts.</td>
<td>文</td>
</tr>
<tr>
<td>Quotes from Kūkai</td>
<td>大師云（something gives the title of the work）</td>
<td>Quotes from Kūkai. Alleged quotes are possible.</td>
<td>文</td>
</tr>
<tr>
<td>Quotes from previous masters</td>
<td>先師（or the monk’s name）云（something gives the title of the work）</td>
<td>Words of the previous masters of Daigoji. Mostly verifiable, but an author may attribute his own ideas to a previous master.</td>
<td>文（omitted more frequently）</td>
</tr>
<tr>
<td>Oral traditions</td>
<td>口伝云（may mention the name of Oral traditions of the Shingon school, mainly from Daigoji. Same as above.</td>
<td>Usually does not conclude the citation with 文</td>
<td></td>
</tr>
</tbody>
</table>

---

\(^{27}\) On the structure of this type of text, see, for example, Hashimoto 2010.
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

<table>
<thead>
<tr>
<th>Authorship Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secret precepts</td>
<td>Citation of secret oral teachings of the person at the origin of the tradition</td>
<td>秘決云</td>
</tr>
<tr>
<td>Secret transmissions</td>
<td>Secret transmissions, quite hard to verify (but sometimes possible).</td>
<td>秘伝云</td>
</tr>
<tr>
<td>The author’s own opinions</td>
<td>Ideas the author presents as his own. Surprisingly rare.</td>
<td>私云</td>
</tr>
</tbody>
</table>

To give a concrete example, in the case of the *Sanzon gōgyō hiketsu* by Monkan, such citations represent 62% of the text. A large part of the rest consists of commentaries and paraphrases of the citations, using identical vocabulary. However, the parts presented as Monkan’s direct voice are extremely rare, representing less than 1% of the text.

Another good way to refine these results would be to implement tags directly into the editions of the texts. XML, as well as TEI, would be quite useful, but would also present difficulties with the rendering of both *kanbun* and the Siddham script. Implementation of such tags would be extremely useful, as it would permit the isolation of different speech levels and present a better picture of the relative importance of the n-grams or patterns found in texts. However, using such tags to eliminate all the reported quotations from the text would be disastrous. In fact, a writer’s style can still mark the way that quotations are introduced. Their presence also provides indications about the sources and the doctrinal traditions to which the text belongs. Furthermore, some might not be authentic citations. While keeping such issues in mind, one of the future stages of this project will be the implementation of such tags into a larger statistical analysis and the assessment of their usefulness in the study of other authors.
7. Recurring Grammatical Patterns
To provide a better analysis of writing styles, tags would be useful to mark grammatical elements and assess word functions. Using word n-grams would also open up further possibilities. In this case, the main problem is the segmentation of words and phrases. This has been attempted by Bingenheimer, Hung, and Hsieh (2017). For this project, an R script was also used to extract grammatical patterns from Monkan’s texts.

More specifically, the script assessed the relative frequency of the very common sequence of “motte … nasu” 以〜為 in his work (which means that something becomes or is used as something else), after which it was compared to a dataset similar to those used in the n-gram experiments.

The results can be seen on the next table. Following the method given by Bingenheimer, Hung, and Hsieh 2017, the average frequency of 以〜為 was also calculated. The data was also grouped, when possible, by author. Individual texts, such as the Ben’ichisan himitsuki and Daijingū honji, were left alone.

When left untouched, the impact of average frequencies remains difficult to evaluate. However, when the data is grouped, following the ensembles defined by the previous n-gram experiments, it displays some interesting characteristics. It seems that the sequence 以〜為 is especially frequent in Monkan’s discursive or doctrinal texts. For example, the Himitsu gentei kuketsu uses it quite frequently, and the average number of uses is higher in this particular part of Monkan’s work than in any other text or group of texts.

These results also seem to confirm that the ritual procedures have their own grammatical characteristics, which probably have a larger impact (as seen in most of the previous PCA calculations) than individual style. They employ very discursive sentences, and are often limited to lists of items, mantras, or gestures used in the ritual.

Interestingly, the Goyuigō hiketsu (short version) again is situated closer to Monkan’s work than to Jichiuin’s, as it displays a higher frequency of this pattern. However, while the Daijingū honji uses 以〜為 more frequently than the Ben’ichisan himitsuki, it is difficult to draw any conclusion about these two texts based on this analysis alone.

The number for Kūkai’s texts is lower than the average of Monkan’s, but it is still comparable. This demonstrates the role of the founder as a sort of stylistic model for the

---

28. The script was developed with the help of Mark Ravina of Emory University.
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

subsequent members of the Shingon school. In this context, grammatical patterns were clearly taken from previous works that the monks read. While further investigations using a larger corpus remain necessary, the relatively high frequency of the pattern “motte… nasu” in discursive or doctrinal texts seems to suggest that it is a characteristic of this genre.

This can be confirmed through a similar experiment with a text by Eison (叡尊) (1202–1290), a major figure of the Saidaiji temple, who authored texts that were largely independent from the esoteric tradition. The target text is one of them, and it displays a very low frequency of this pattern.29

However, a more thorough analysis—including a larger dataset of medieval Shingon sources and other Buddhist texts—is necessary to obtain more refined results. Including other grammatical patterns, as in Bingenheimer, Hung, and Hsieh 2017, will also be necessary. This was not possible in this article, owing mainly to a lack of usable sources from a similar period as points of comparison to Monkan’s work, but also to a lack of groundwork on the frequency of use of Chinese grammatical particles in medieval Japanese Buddhist *kanbun*. This area will be pursued in future studies.

<table>
<thead>
<tr>
<th>Text</th>
<th>Occurrences of &quot;Motte...nasu&quot;</th>
<th>Length of the texts (in characters)</th>
<th>Frequency (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goyuigō hiketsu (Daigoji)</td>
<td>13</td>
<td>7087</td>
<td>0.183434457</td>
</tr>
<tr>
<td>Himitsu gentei kuketsu</td>
<td>26</td>
<td>3930</td>
<td>0.661577608</td>
</tr>
<tr>
<td>Goyuigō daiji</td>
<td>21</td>
<td>8913</td>
<td>0.235610905</td>
</tr>
<tr>
<td>Goyuigō hiketsu</td>
<td>4</td>
<td>2024</td>
<td>0.197628458</td>
</tr>
</tbody>
</table>

29. T 2357. On this work, see Ōno busho kaisetsu jiten, Tōkyō: Daitō shuppansha, 1933, vol. 1, 392.
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

<table>
<thead>
<tr>
<th>Title</th>
<th>Count</th>
<th>Frequency</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goyuigō shichikadaiji</td>
<td>5</td>
<td>3678</td>
<td>0.135943448</td>
</tr>
<tr>
<td>Sanzon gōgyō hiketsu</td>
<td>5</td>
<td>3925</td>
<td>0.127388535</td>
</tr>
<tr>
<td>Saigoku himitsushō</td>
<td>25</td>
<td>8422</td>
<td>0.296841605</td>
</tr>
<tr>
<td>Tōryū saigoku hiketsu</td>
<td>9</td>
<td>4810</td>
<td>0.187110187</td>
</tr>
<tr>
<td>Saigokushō</td>
<td>26</td>
<td>7655</td>
<td>0.339647289</td>
</tr>
<tr>
<td>Gomashidai</td>
<td>3</td>
<td>1946</td>
<td>0.154162384</td>
</tr>
<tr>
<td>Ichinisun gōgyō hishidai</td>
<td>2</td>
<td>3385</td>
<td>0.059084195</td>
</tr>
<tr>
<td>Jigyōshidai</td>
<td>0</td>
<td>1155</td>
<td>0</td>
</tr>
<tr>
<td>Jūichimen Kannon hihō</td>
<td>0</td>
<td>518</td>
<td>0</td>
</tr>
<tr>
<td>Daijingū honji</td>
<td>2</td>
<td>1229</td>
<td>0.16273393</td>
</tr>
<tr>
<td>Benichisan himitsuki</td>
<td>3</td>
<td>2239</td>
<td>0.133988388</td>
</tr>
<tr>
<td>Jūjūshinron (Kūkai)</td>
<td>118</td>
<td>81222</td>
<td>0.145280835</td>
</tr>
<tr>
<td>Benkenmitsu nikiyōron (Kūkai)</td>
<td>10</td>
<td>9236</td>
<td>0.108236822</td>
</tr>
<tr>
<td>Sōkushin jōbutsugi (Kūkai)</td>
<td>43</td>
<td>26676</td>
<td>0.161193582</td>
</tr>
<tr>
<td>Shoshon yōshō (Jichiun)</td>
<td>26</td>
<td>46829</td>
<td>0.055521151</td>
</tr>
<tr>
<td>Hizō konpōshō (Jichiun)</td>
<td>43</td>
<td>36908</td>
<td>0.116505906</td>
</tr>
<tr>
<td>Genpishō (Jichiun)</td>
<td>49</td>
<td>34900</td>
<td>0.140401146</td>
</tr>
<tr>
<td>Ōrishū kaizu shakumonshō 應理宗戒圖釋文鈔 (Eison)</td>
<td>6</td>
<td>8065</td>
<td>0.074395536</td>
</tr>
</tbody>
</table>

| Frequency average Monkan (doctrinal, discursive texts) | 0.262798055 |
| Frequency average Monkan (procedures)                  | 0.053311645 |
| Frequency average Kūkai                                | 0.13823708  |
| Frequency average Jichiun                              | 0.104142734 |
Conclusion

The variable length n-gram method was originally designed to identify translators, rather than authors, of Buddhist texts. However, our experiment has proven that it can be quite useful for shōgyō texts produced by the medieval Japanese esoteric schools. The cumulative texts produced by monks of the Shingon school, full of quotations from various sources, are in fact closer to compiled translations of canonical sources than to modern literary texts.

This is why the method designed by Hung, Bingenheimer, and Wiles (2010) proved to be quite effective. While it garnered limited results using all the texts at the same time, successive analysis of carefully selected groups of documents provided gradual clarification of the relative position of the three target texts with Monkan’s work, Shingon texts, and medieval Shintō texts.

This provided very interesting results, and allowed probable identification of the author of two of the texts. Regarding the Goyuigō hiketsu, the experiment allowed a further confirmation of Monkan as its true author through raw data. Regarding the Daijingū honji, it provided concrete evidence of its proximity—in subjects, vocabulary, and even grammar—to Monkan. While a lack of historical documentation prevents any certainty regarding the text’s authorship, the differences in style that it displays in comparison with a text by Kakujō seem to suggest that it was likely composed by Monkan.

The case of the Ben ‘ichisan himituski was less conclusive, but it provided reasoning against Monkan’s authorship of this text. However, he used ideas similar to what can be found in it, and the affiliation is clear. Previous attributions were probably correct in identifying its author with Kenjin or another thirteenth-century monk.

However, as we have seen with Monkan’s ritual procedures, results can vary dramatically within the same author’s work, depending on the genre or nature of the text. This, along with the very nature of such texts, could also indicate that the idea of an original author might be misleading.

In fact, this modern concept of authorship is certainly not the most adequate tool for understanding this type of literature.³⁰ Most of the sacred teachings are ritual manuals or

³⁰ This is, of course, inspired by Michel Foucault’s criticism of the notion of an original author,
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

records of oral traditions, full of quotes from various sources. The writers themselves do not try to emphasize their status as original creators. They position themselves as the members of a long tradition, and even attribute their works—just as Monkan probably did—to previous monks of greater prestige.

For the historian, both of these attributions can be important when one tries to study the image that a given religious group had of itself. However, when one tries to surpass traditional presentations and to understand how knowledge spread in medieval Japanese temples, digital tools, and especially data mining, can be extremely helpful. Combined with a solid historical analysis, these tools provide raw data that allow scholars to restore their understanding of this still little-studied period of Japanese religious history, providing newer and more historically accurate models. This may also free Buddhist studies scholars from the widespread traditional sectarian views that have dominated academia for so many years.

especially in “L’ordre du discours” (Gallimard: Paris; 1971).
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

Bibliography


Ishii Kōsei 石井公成. 2001. “N-Gram riō no kanōsei—bukkyō bunken ni okeru ihon hikaku to yakusha-sakusha hantei N-Gram 利用の可能性——仏教文献における異
Data Mining in the Works of the Shingon Monk Monkan (1278–1357): Using Digital Methods to Assess the Contested Authorship of Three Religious Texts from Medieval Japan

本比較と訳者・作者判定。” Kanji bunken jōhō shori kenkyū 漢字文献情報処理研究 2:59–61.


