

Good morning everyone, my name is Yada Tsutomu; I work at Osaka University researching the history of the Japanese writing system. I would like to speak a little about the encoding of hentaigana as one of those involved in selecting the characters to be encoded, as well as one of those who eagerly anticipate the realization of this encoding.

Hiragana were created in the mid-9th century in Japan, and have continued to be one of the main components of the Japanese writing system for over one thousand one hundred years until this day. Hiragana are in principle syllabic characters. There are few examples of syllabic scripts used outside Japanese, apart from some script like Linear B and Nushu, though Japanese features two types of syllabic scripts—hiragana and katakana.

With syllabic letters, as a type of phonographic script, it would be logical for each syllable to be represented by one letter only. However in reality, this logical, rational relationship had been in effect for the syllabic scripts in the Japanese writing system for only one hundred years. For over a thousand years, Japanese people have been using several different letters to represent the same sound. (Slide 1) As you can see in this slide, the term hiragana in its original sense includes all variant letters, which are called hentaigana. (Slide 2) This slide shows an example of a Japanese text written in the Heian period. The letters that are marked in the slide all represent the sound “ka”; you can see that in this slide alone there are four different ways to represent the same syllable.

This can seem odd to the modern reader. I myself am also often asked why Japanese people continued such bothersome practices for so long. Strictly speaking, there is not enough evidence to answer this question without drawing from ones own imagination, though I would like to give a possible explanation about the mentality of Japanese people of the time and the root of such writing conventions.

Before the invention of hiragana and katakana, Japanese people used Hanzi

imported from China. As you all know, logographic Hanzi are vast in number, and you could say that there is no real end to the study of Hanzi. This can seem like a troubling idea, but if we look at things from a different perspective we might be able to find some sort of enjoyment in it all.

For example, Marty Friedman, a famous rock musician currently living in Japan, commented in a Japanese television program that the endless nature of the study of Japanese writing is fun. Here, Mr Friedman is pointing out the joy of having something that someone can continue to work at and improve their ability for the whole of their life.

When writing Japanese using Hanzi, several different characters can be used for the same word. Seeing that Japanese people accepted this diverse nature of Hanzi, in some ways it is natural to assume that they would accept a diverse style of writing in kana as well, or even that they would value the ability to write in a diverse manner.

In a time when there was no need to learn how to use a computer, or to learn languages like English, French or German, Japanese people looked at an individual's proficiency at writing in a diverse manner as a mark of culture.

However, on the other hand, humans are rational creatures. For example, in the case of hiragana, it appears that there was a desire at the time to utilize the diversity of hentaigana in order to convey information (Slide 3). As a result of this, there arose a trend where certain hentaigana were used for certain words, facilitating the recognition of those words in a string of letters, or using certain hentaigana for the start of words, the end of words or for particles, giving the usage of hentaigana a function similar to punctuation. Modern rules for Japanese punctuation were only established in the Meiji era, so this usage of hentaigana helped readers to correctly read texts.

The introduction of European printing press techniques into Japan brought with them a new way of seeing letters from a rational perspective—as a

result, the number of hentaigana used in general society sharply decreased after the latter half of the Meiji period. However, as I mentioned before, hentaigana have been used for over one thousand years—a significant span considering that one thousand one hundred years have passed since the birth of hiragana. However, today, when transcribing past texts into a digital format, we have no choice but to replace all hentaigana with the currently used hiragana letter of the same sound value. This replacement not only ignores the mindset of the original writer who would value a varied use of hiragana, but even leads to remove important linguistic information lying within the text.

Writing is fundamentally a tool used to pass on all human endeavours—knowledge, ideas, or feelings—from one generation to the next. It is vital to include hentaigana in encoding standards to prevent a part of Japanese culture from being lost in the digitization of Japanese writing.

I would like to offer my opinion as a researcher of linguistics. As was written in the proposal, one of the main goals for the inclusion of hentaigana in encoding standards was for them to be displayed correctly in electronic documents for governmental use, but the inclusion of hentaigana in encoding will also bring great benefit to linguistic research as well as any field of research that deals with past texts written in Japanese.

Researchers of the history of the Japanese writing system are required to display hentaigana when publishing their research results. However, without any encoding standards at present, researchers are forced to incorporate inline images in their work, (Slide 4) often resulting in an ugly appearance as you can see in this slide.

This slide shows an issue in the publishing of results, but there are also problems in methodology for researchers of the Japanese writing system.

Japanese linguistic research, along with research on other languages, is

moving from an age where individual researchers searched for linguistic evidence by hand to an age where digital texts and large-scale corpora are used to find linguistic evidence on a broader scale and with increased precision. In this transition to digital, it is the researchers of the history of the writing system that are currently excluded from the benefits of new research methods. This is because, as you understand, there is currently no way of digitizing a text whilst preserving its original usage of hentaigana. When researching the history of the Japanese writing system, there is no option other than to use the old method of checking manuscripts or photographs by hand for linguistic trends. If hentaigana were to be included into encoding standards, and digital texts indicating the original texts' use of hentaigana were to be shared among researchers, we could say that the foundations for this field of research will progress dramatically.

Also, seeing that hentaigana were used to some extent to convey linguistic information as I mentioned earlier, it is possible that new discoveries may be made in research on Japanese grammar which may have been overlooked previously. It goes without saying the impact this could have on study of literature, history, or any other field that investigates such texts.

For these reasons, I strongly desire the inclusion of standardized code points for hentaigana. I kindly request your consideration—thank you very much for listening.

大阪大学の矢田と申します。私は日本の表記体系の歴史を研究しております。今回の変体仮名の文字コード標準化に関して、その字種の選定作業に関わった者として、またその実現を切望する者の一人として、少しだけお話しさせていただきます。

平仮名は、9世紀中頃の日本で生み出され、現在まで1100年を超える期間にわたって、日本語表記体系の一つの柱を成しています。その基本的性質は「音節文字」です。日本語以外ではクレタ文字などごくわずかな例しか知られない音節文字ですが、日本語では、平仮名・片仮名という二種類の音節文字が使われ続けてきました。

音節文字も表音文字の一種ですから、記号としての働きという点からすれば、当然、一つの音節に対して一つの文字のみが対応するのが合理的なあり方です。しかし、現実には、日本語表記体系では、音節文字がそのようなあり方を実現したのは、わずか百年あまり前のことにすぎず、それ以前の千年以上の期間、日本人は一つの音に複数の表音文字を併用し続けていました。(スライド1) 平仮名とは、このように、本来、変体仮名を含めた全ての字体の集合として初めて完結するものでした。(スライド2) これは、平安時代に書かれた日本語テキストの実例の一つです。印をつけたのはいずれも同じ「カ」の音を表しますが、この写真の中だけで、4種類の字体が使用されています。

それは、現代人の目から見るといささか奇妙なことに映ります。私も、よく、「なぜ日本人はそんな面倒なことを長いあいだ続けたのか」という質問を受けます。そうした表記の習慣を続けた日本人の心性は、厳密に言えば想像するしかない部分もあるのですが、私はそれを次のように考えています。

日本人は、平仮名・片仮名の発明以前から、中国から移入された漢字を使っていました。ご存じの通り、表語文字である漢字は、膨大な字種を有しており、その学習は言ってみれば「きりがありません」。それは、大変やっかいなことであるようでいて、別の視点からは、そこにある種の楽しみのようなものを見いだすこともできます。

現在日本に在住しておられる著名なロック・ミュージシャンであるマーティ・フリードマンさんが、日本語の文字を学ぶことについて、その「endless」感が楽しいと、日本のテレビ番組で語ったことがあります。それは、上限無く人生の最後まで自分のリテラシーを高めていくことのできる楽しさを指摘したものです。

漢字で日本語を表記する場合、同じ語に数多くの漢字を当てることができます。そうした、多表記性という漢字表記のあり方を受け入れていた日本人にとって、仮名に対しても、一通りでない多様な書き方を認め、さらには、そうした多様な書き方をすることができる能力に価値を見いだすことは、ある意味で自然な成り行きであったと考えます。

コンピュータの使い方を覚える必要もない、英語もフランス語もドイツ語も覚える必要がない、そういった時代において、どれだけ多様な表記法を自在に操れるか、ということに教養の高さの一つの尺度を見いだしていたのが、当時の日本人でした。

しかし、一方でやはり人間は合理性を求める生き物でもあります。平仮名にしても、せっかく多様な変体仮名があるのだったら、それを何らかの形で情報伝達にも役立てたい、

という欲求もまた同時に生じてきました（スライド3）。その結果、ある特定の語彙はある特定の变体仮名で書いて、文字列の中からその語を特定しやすくするとか、ある变体仮名は語頭に特化して用い、一方である变体仮名は語尾とか助詞とかに特化して用いることで、变体仮名の選択に句読法の機能を持たせる、とかいった情報伝達上の規則が成立してきます。日本語に現代的な句読法が定着するのは明治時代以降のことで、それ以前には变体仮名を上手く使い分けることが、テキストの正しい読解のために大きく役立てられていたのです。

ヨーロッパの活版印刷術とともに、文字に関する合理的思想も日本に導入され、明治時代後期以降、变体仮名が使用される社会的な機会は急速に減少しました。ですが、先にも述べたように、平仮名が誕生して以降の1100年あまりの日本語の歴史の内、千年以上という圧倒的な長期間にわたり、变体仮名は使われてきたのです。しかし、現在、そうした過去のテキストを電子テキスト化しようとするれば、变体仮名はすべて、同じ音価を有する現行の平仮名に置き換えざるを得ません。それは、多様な变体仮名を使用することを良しとしていたその書き手の心性を無視することであるとともに、あまつさえ、テキストに潜在する重要な言語的情報を取りこぼすことにも他なりません。

文字は本来、過去の、知識や思想や感情やその他もろもろの人間の精神活動を後世に伝承するためのツールです。日本語文字の電子化が、日本の文化的継承のある部分の切り捨てにつながらないためにも、变体仮名の文字コード標準化はどうしても必要なことです。

言語研究者としての立場からも申し上げます。提案書にもありますように、变体仮名の文字コード標準化は、一つには現在もなお人名に使用されている变体仮名を行政上の電子的処理にきちんと載せられるようにする、というところに大きな目的の一つがありますが、言語研究を初めとして、日本語の過去の文献を研究資料とする諸研究分野にとっても、大きく裨益するところです。

日本語の文字・表記史研究者にとって、その研究成果を公表するに当たっては、当然变体仮名の提示が必要になります。しかしながら、その符号標準化が行われていない現状では、それを画像の混用によって行わざるを得ない場合も多く（スライド4）、ご覧のようなぶざまなありさまになることも珍しくありません。

これは、成果の公表に関わる場所ですが、研究方法にも、私たち文字・表記史研究者は大きな制約を抱えています。

日本の言語研究も、他の諸言語の研究と同様、個別研究者が手作業で用例を発掘していた時代から、電子化テキストや大規模コーパスを利用してより広汎に、かつ精緻に用例を探索する時代に移行してきています。そうした中で、そうした電子化の時代の恩恵にもっとも預かることが出来ていないのが、文字・表記史研究者です。それはすでにどなたにも明確であるように、变体仮名に関して、元のテキストの用字法を残した電子化が現時点では出来ないことに理由があります。日本語の文字・表記史研究では、今以て原本やその写真から一つ一つ例を探し出していくという、旧態依然とした研究手法以外のやりかたが導

入できておりません。変体仮名の文字コード標準化が行われ、原テキストの用字法を保存した電子化テキストが研究者に広く共有されれば、この分野の研究基盤は画期的に進歩するといえます。

加えて、先ほど述べたように、変体仮名が言語情報の伝達の一部も担っていたということからすれば、文法研究などでも、これまで見落とされてきたことの再発見を促す可能性があります。文学研究、歴史研究など、テキストの内容に関わる研究領域にも与える影響が大いにあるであろうことはいうまでもありません。

以上、変体仮名の文字コード標準化を強く願う立場からお話しさせていただきました。お聞き下さってありがとうございました。今日の私のお話が皆様のご理解を得られましたら、無上の幸いです。

About the inclusion of standardized codepoints for Hentaigana

YADA Tsutomu
(Graduate School of Letters,
Osaka University)

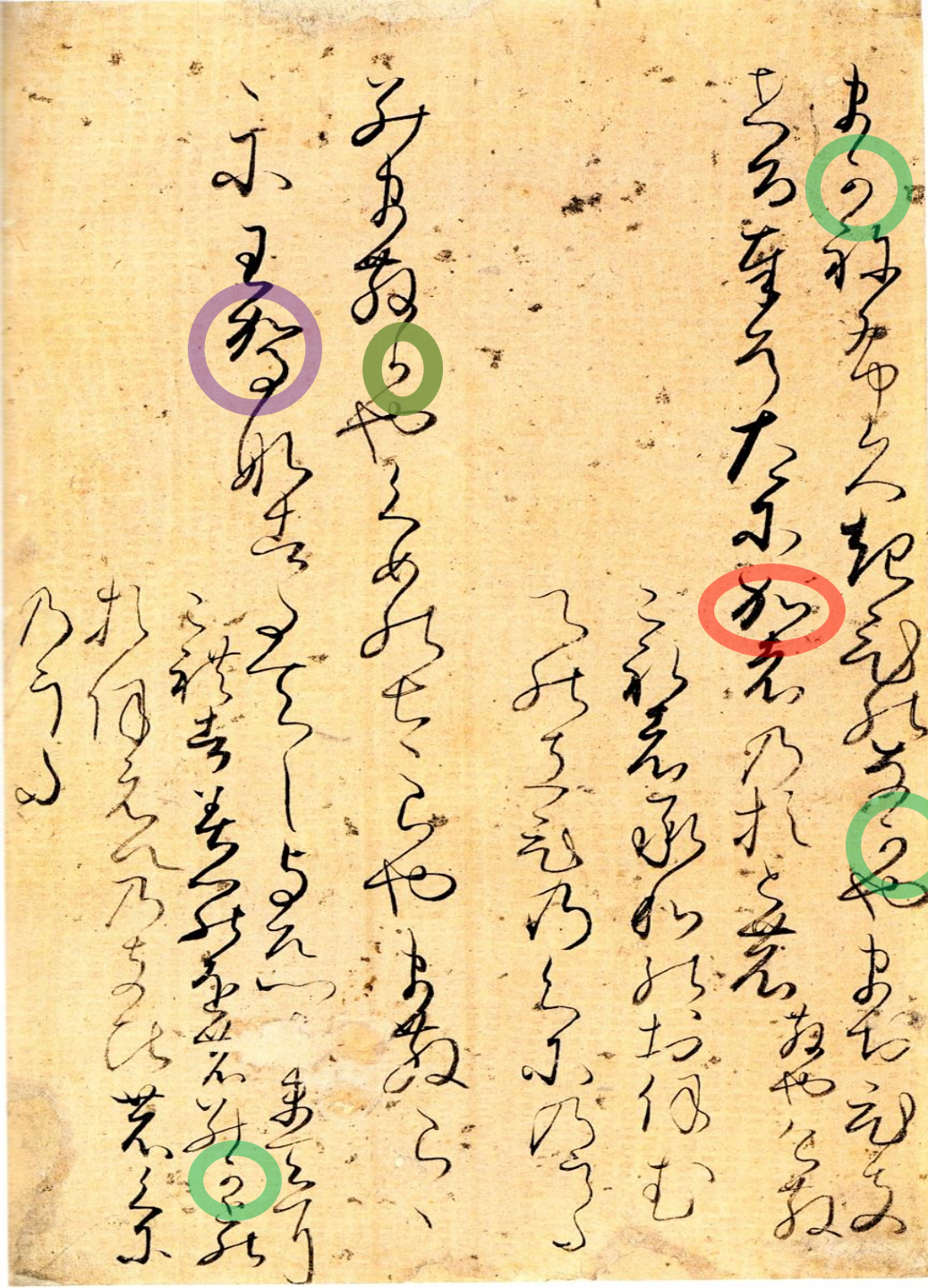
Syllables Katakana Hiragana

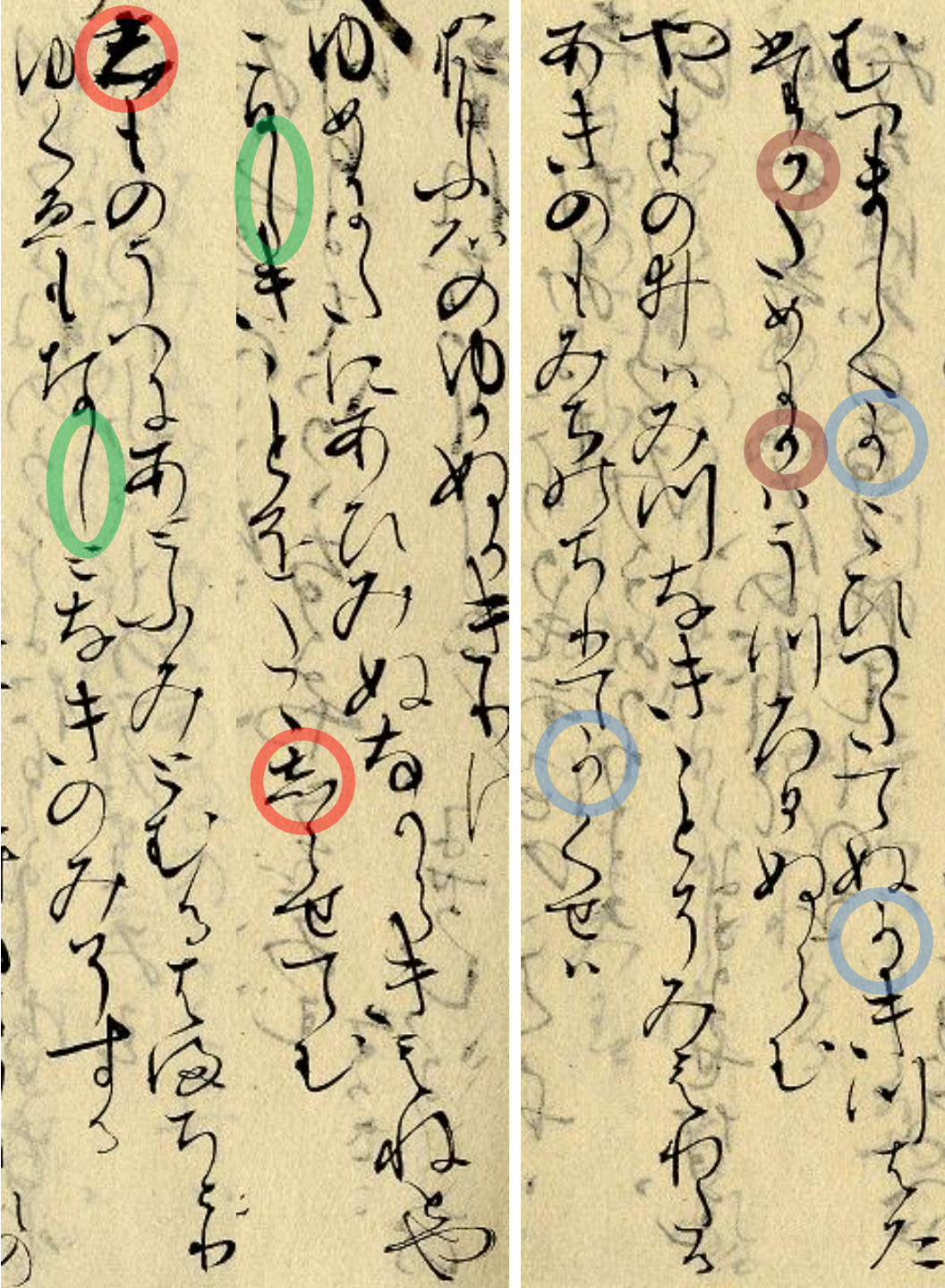
Current style Hentaigana

[a]	ア	あ	あ	阿	𐰇	𐰇	...
[ka]	カ	か	か	𐰇	𐰇	𐰇	...
[sa]	サ	さ	𐰇	𐰇	𐰇	𐰇	...
[ta]	タ	た	𐰇	𐰇	𐰇	𐰇	...
[na]	ナ	な	𐰇	𐰇	𐰇	𐰇	...
[ha]	ハ	は	𐰇	𐰇	𐰇	𐰇	...
[ma]	マ	ま	𐰇	𐰇	𐰇	𐰇	...

Variations of [ka]

か
か
かう
かう





[shi] ○ 志

mainly for the initial
syllable of a word

○ し

mainly for the non-
initial syllable of a word

[ka] ○ の

for the initial
syllable of a word

○ う

for the non-initial
syllable of a word

※ の and う are derived
from same Hanzi “可”

頃か)あたりから使用が見える後発の仮名字体であり、「お」に対して従の立場にあった。ところが、中世平仮名文書では、「れ」がオの主たる字体となっている。六四通の平仮名文書では、「れ」の用例のあるものが二五通、「お」の用例のあるものは二通であった。日蓮はオの音節を殆ど全て「を」で表すが、敢えてア行の仮名を用いる場合には「**れ**」である。⁽²⁵⁾また、親鸞はその消息に「**れ**」だけを用い、「お」は用いていない。⁽²⁷⁾

ソの仮名の「**ろ**」は、平仮名成立初期からの用例を持つが、「そ」に比べて補助的な字体であった。しかし、実用的平仮名文の世界では、「表制集紙背消息」「宣旨官牒紙背消息」の頃を境に、用例数で「そ」を圧倒するようになる。中世平仮名文書では、他に「**ろ**」という頻用仮名が現れるが、一通の中で「**ろ**」と「**ろ**」とが両方とも用いられた例の少ないことが特徴的である。因みに六四通での使用文書数は、「**ろ**」が二五、「**ろ**」が二三、「そ」が八である。

こうした一方で、中世平仮名文書では主たる字体でありながら、平仮名書いろは歌に採られていない字体があることにも触れておかねばならない。「**い**」がそれである。「**い**」は「表制集紙背消息」や「宣旨官牒紙背消息」に見えはじめる後発の字体で、中世に入って文書書写・文学書写の両方で多く用いられるようになった。六四通では、「**い**」が四七通、「つ」が三二通、「**は**」が七通に、

それぞれ用いられて「つ」も決して少「**い**」の方である。

必ずしもこれといわれるのは、『書史会要』。これは或いは、場合があったことを資料のいろは歌に見は、時代の差を反映できるのではないかと、そうした視点に文書所用の平仮名ときが浮かびあがって異なる字形がとり形の方が文書に実際

か。

正敷

四庫全書

嬌を申る色よりとまめいつくゝ恋の心の重のむく滑

大文

言方

橘花ありと風はまきつゝの神のおほや

風城

祁秀

うちへてまに
未^まはかよりまし
いせのめ
おの

東野

久城

五つらしはむかし
の時のちねはうき
なうきたね

松

守恭

[illegible]

立

長
弘

風をよきものとして立てゝおけの筆おのり

五

政惠

おはせし酒をたぐひある己祢のまゝ風

今

言文

来いといふて
 神を代の月をいふやの表

白山寺

周祐

ちうにちうふめつしとゆゑの寺にそひるにやう

補望

正基

人々を以て其の

張

光華

なるゆきもふまゝに
 戸匠に送るよとめ
 の實はゝや

...

信
楚

水木也。即乙午所書。

安政七年庚申春

申さず

*Thank you very much
for listening!*