

Knowledge Representation and Phonological Rules for the Automatic Transliteration of Balinese Script on Palm Leaf Manuscript

Made Windu Antara Kesiman^{1,3}, Jean-Christophe Burie¹, Jean-Marc Ogier¹, Philippe Grangé²

¹ University of La Rochelle,
Laboratoire Informatique Image Interaction (L3i), La Rochelle,
France

² University of La Rochelle,
Departement of Indonesian Language, La Rochelle,
France

³ Universitas Pendidikan Ganesha,
Laboratory of Cultural Informatics (LCI), Singaraja, Bali,
Indonesia

{made_windu_antara.kesiman, jean-christophe.burie, jean-marc.ogier, pgrange}@univ-lr.fr

Abstract. Balinese ancient palm leaf manuscripts record many important knowledges about world civilization histories. They vary from ordinary texts to Bali's most sacred writings. In reality, the majority of Balinese can not read it because of language obstacles as well as tradition which perceived them as a sacrilege. Palm leaf manuscripts attract the historians, philologists, and archaeologists to discover more about the ancient ways of life. But unfortunately, there is only a limited access to the content of these manuscripts, because of the linguistic difficulties. The Balinese palm leaf manuscripts were written in Balinese script in Balinese language, in the ancient literary texts composed in the old Javanese language of Kawi and Sanskrit. Balinese script is considered to be one of the most complex scripts from Southeast Asia. A transliteration engine for transliterating the Balinese script of palm leaf manuscript to the Latin-based script is one of the most demanding systems which has to be developed for the collection of palm leaf manuscript images. In this paper, we present an implementation of knowledge representation and phonological rules for the automatic transliteration of Balinese script on palm leaf manuscript. In this system, a rule-based engine for performing transliterations is proposed. Our model is based on phonetics which is based on traditional linguistic study of Balinese transliteration. This automatic transliteration system is needed to complete the optical character recognition (OCR) process on the palm leaf manuscript images, to make the manuscripts more accessible and readable to a wider audience.

Keywords. Knowledge representation, phonological rules, automatic transliteration, Balinese script, palm leaf manuscript.

1 Introduction

In many Southeast Asian countries, the literary works were mostly recorded on dried and treated palm leaves (Figure 1). The palm leaf manuscripts record many important knowledges about world civilization histories. It attracts the historians, philologists, and archaeologists to discover more about the ancient ways of life. Although the existence of ancient palm leaf manuscripts in Southeast Asia is very important both in term of quantity and variety of historical contents, unfortunately the access to their content is limited, because of the linguistic difficulties and the fragility of the documents.

In order to make the manuscripts more accessible, readable and understandable to a wider audience, an optical character recognition (OCR) system and also an automatic transliteration system are urgently needed for this kind of ancient manuscript collection. In the domain of document image analysis, the handwritten character recognition has been the subject of

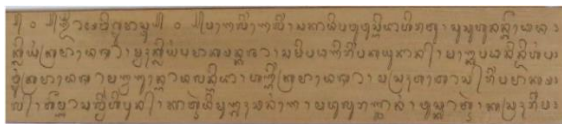


Fig. 1. Sample of palm leaf manuscript from Bali, Indonesia

intensive research during the last three decades. Some methods have already reached a satisfactory performance especially for Latin, Chinese and Japanese script. However, the development of handwritten character recognition methods for other various Asian script, such as Balinese script [1], Devanagari script [2], Gurmukhi script [3], Bangla script [4], and Malayalam script [5], always presents many issues.

Moreover, to give more access to the important information and knowledge in palm leaf manuscript only with an OCR system is not enough, because normally in Southeast Asian script the speech sound of the syllable change related to some certain phonological rules. Therefore, using an OCR system which is completed by an automatic transliteration system will help to transcribe these ancient documents, and to develop the automatic analysis and indexing system of the manuscripts. By definition, transliteration is defined as the process of obtaining the phonetic translation of names across languages [6].

Transliteration involves rendering a language from one writing system to another¹. In [6], the problem is stated formally as a sequence labeling problem from one language alphabet to other. A transliteration engine for transliterating the Balinese script of palm leaf manuscript to the Latin-based script is one of the most demanding systems which has to be developed for the collection of palm leaf manuscript images. It will help to index and to access quickly and efficiently to the content of the manuscripts.

For this purpose, we present in this paper an implementation of knowledge representation and phonological rules for the automatic transliteration of Balinese script on palm leaf manuscript. Many transliteration model have been proposed [6]–[9].

¹ <https://www.accreditedlanguage.com/2016/09/09/what-is-transliteration/>

In this system, a rule-based engine [9] for performing transliterations is proposed. Our model is based on phonetics which is based on traditional linguistic study of Balinese transliteration.

This paper is organized as follow: section 2 gives a brief description about Balinese palm leaf manuscripts in Bali and the Balinese script. Section 3 presents the detail description of knowledge representation for the glyph segment image collection in the experimental study. The phonological rules for Balinese script transliteration is formally described in section 4. The result and evaluation of the experimental results are presented in Section 5. Finally, conclusions with some prospects for the future works are given in Section 6.

2 Balinese Palm Leaf Manuscript and Balinese Script

2.1 Balinese Palm Leaf Manuscript

In Bali, Indonesia, the palm leaf manuscripts were called *Lontar*. The epic of *lontar* varies from ordinary texts to Bali's most sacred writings. Apart from the collection at the museum (Museum Gedong Kertya Singaraja and Museum Bali Denpasar), it was estimated that there are more than 50,000 *lontar* collections which are owned by the private families.

Unfortunately, many discovered *lontars*, part of the collections of some museums or belonging to private families, are in a state of disrepair due to age and due to inadequate storage conditions. *Lontars* were written and inscribed with a small knife-like pen (a special tool called *Pengerupak*). It is made of iron, with its tip sharpened in a triangular shape so it can make both thick and thin inscriptions. The writings were incised in one (and/or both) sides of the leaf. The manuscripts were then scrubbed by the natural dyes to leave a black color on the scratched part as text.

2.2 Balinese Script

The Balinese palm leaf manuscripts were written in Balinese script and in Balinese language,

composed in the old Javanese language of *Kawi* and *Sanskrit*. Balinese language is a Malayo-Polynesian language spoken by about more than 3 million people mainly in Bali, Indonesia². Balinese script (*Aksara Bali*) is a descendant of *Brahmi* script, and is considered to be one of the more complex scripts from Southeast Asia. The type of writing system is syllabic alphabet. The alphabet and numeral of Balinese script is composed of ± 100 character classes including consonants, vowels, and some other special compound characters. Some characters are written on upper baseline (Ascender) or under the baseline of text line (Descender). Writing in Balinese script, there is no spaces between words from left to right in a horizontal text line. According to the Unicode Standard 9.0³, Balinese script has actually the Unicode table from 1B00 to 1B7F. In general, in Balinese script⁴:

- The vowel "A" is implicit after all consonants and consonant clusters and should be supplied in transliteration, unless: (a) another vowel is indicated by the appropriate sign; or (b) the absence of any vowel is indicated by the use of an "ADEG-ADEG" sign.
- Vowels are almost always indicated by one of a class of agglutinating signs (*panganggesuara*) added above, below, before, or after the consonant or consonant cluster which they affect.

To initialize the work on document image analysis of palm leaf manuscripts, Kesiman et al. [10] have already built the *AMADI_LontarSet*⁵, the first handwritten Balinese palm leaf manuscript dataset. It includes three components of dataset as follows: binarized images ground truth dataset, word annotated images dataset, and isolated character annotated images dataset. In the isolated character annotated images dataset, there are 133 glyph classes which were collected as the glyph segment image collection. The study of isolated Balinese character recognition on palm leaf manuscript was also already done by Kesiman et al. [1].

² www.omniglot.com/writing/balinese.htm

³ <http://www.unicode.org/versions/Unicode9.0.0/>

⁴ <https://www.loc.gov/catdir/cpso/romanization/balinese.pdf>



Fig. 2. Consonant basic glyphs (from left to right: glyph "Na", "NA TEDONG", "GANTUNGAN NA", "TA", "TA TEDONG", and "GANTUNGAN TA")

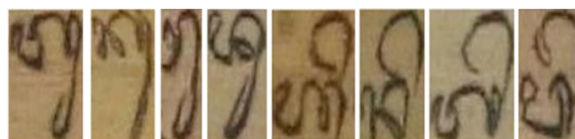


Fig. 3. Compound glyphs (from left to right: glyph "TU", "KU", "RU", "DU", "I", "NI", "TI", "WI")



Fig. 4. Vowel glyphs (from left to right: glyph "TALENG", "TEDONG", "ULU", "SUKU", "CECEK", "PEPET", "SURANG")

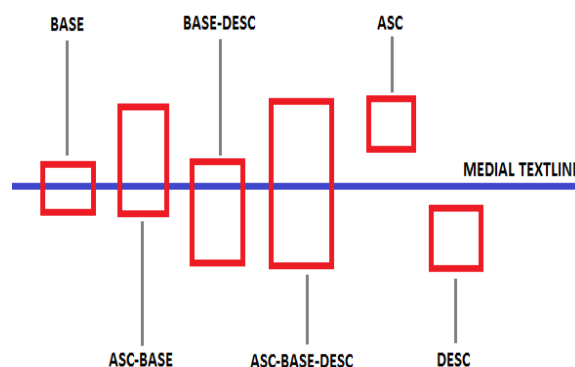


Fig. 5. Spatial position of the glyphs related to the medial text line

The transliteration engine which is developed in this work is based on the need of a transliteration system to complete the OCR process of the palm leaf manuscript.

⁵ http://amadi.univ-lr.fr/ICFHR2016_Contest/index.php/download-123

Table 1. Consonant basic glyphs and their second glyph form (conjunct form)

| Consonant basic glyph name | Second glyph form name | Speech sound | Onset | Nucleus |
|----------------------------|------------------------|--------------|-------|---------|
| NA and NA TEDONG | GANTUNGAN NA | NA | N | A |
| TA and TA TEDONG | GANTUNGAN TA | TA | T | A |
| KA and KA TEDONG | GANTUNGAN KA | KA | K | A |
| WA and WA TEDONG | SUKU KEMBUNG | WA | W | A |
| YA and YA TEDONG | NANIA | YA | Y | A |

Table 2. Consonant compound glyphs

| Consonant compound glyph name | Glyph component | Speech sound | Onset | Nucleus | Coda |
|-------------------------------|------------------|--------------|-------|---------|------|
| TU | TA + SUKU | TU | T | U | - |
| KU | KA + SUKU | KU | K | U | - |
| NI | NA + ULU | NI | N | I | - |
| DI | DA + ULU | DI | D | I | - |
| NING | NA + ULU + CECEK | NING | N | I | NG |

Table 3. Numeral, Punctuation, Vowel, and Special Consonant Glyphs

| Glyph name | Type | Speech sound |
|------------|-----------------------------------|--|
| 0 - 9 | Numeral | 0-9 |
| PAMADA | Punctuation | . (point) |
| TALENG | Vowel | E |
| TEDONG | Vowel | Combined with TALENG, it gives speech sound of "O" |
| ULU | Vowel | I |
| SUKU | Vowel | U |
| CECEK | Special Consonant and Punctuation | NG and , (comma) |
| PEPET | Vowel | E |
| SURANG | Special Consonant | R |
| ADEG-ADEG | Special Consonant | - |

3 Knowledge Representation

3.1 Glyph Segment Image Collection

As already explained in the previous section, the transliteration engine which is proposed for this work is based on and is developed for the glyph segment image collection of the *AMADI_LontarSet* [10]. From the 133 classes of glyph segment image collection which were used in the isolated Balinese

character recognition task [1], [10], some of them represent the consonant of the basic character of Balinese alphabet. In their base form, they will produce a speech sound of a syllable which will be ended with a vowel "A".

For example, glyph "NA", glyph "CA", glyph "RA". etc. Each of those consonant basic character glyph has their associate second glyph form (conjunct form⁶) which is normally called *Gantungan* or *Gempelan*, or with another specific

⁶ In other words, each consonant has two forms, the regular and the appended form
(<https://www.loc.gov/catdir/cpsd/romanization/balinese.pdf>)

glyph name. For example, glyph “*GANTUNGAN NA*” for glyph “*NA*”, glyph “*GUWUNG*” for glyph “*RA*”, glyph “*GEMPELAN SA*” for glyph “*SA*”. The *Gantungan* will be written under other consonant glyph, and the *Gempelan* will be written aside other consonant glyph. The conjunct form will be used when one consonant follows another without a vowel in between⁷.

These *Gantungan* and *Gempelan* will be used to annihilate the vowel “A” (the nucleus of the syllable) of their previous (upside or left side) written consonant glyph (Figure 2). For the reason of limited number of pages, Table 1 listed only 5 examples of the consonant of the basic character glyphs, their associate second glyph forms (conjunct form), speech sound, and their component of syllable (onset and nucleus) which were found in the glyph segment image collection. Interested reader can contact the authors of this paper to get the complete table. In some cases, it was found in the collection, the consonant of the basic character glyphs which were written in their cursive form (optional ligatures) with a glyph called “*TEDONG*”. For example, glyph “*KA TEDONG*” for glyph “*KA*”, glyph “*MA TEDONG*” for glyph “*MA*”, glyph “*NA TEDONG*” for glyph “*NA*”.etc.

But the speech sound of the consonant will not be changed. In addition to the basic glyphs, in the glyph segment image collection, there are also the compound glyphs. A compound glyph is actually a glyph segment which is composed by more than one basic glyph, but the glyph segmentation process cannot separate them. The collection of compound glyph can facilitate the OCR process to deal with the improper segmentation task. The compound glyphs produce a speech sound of a syllable which will be ended with a different vowel than vowel “A” or will be combined with other consonant (Figure 3). For the reason of limited number of pages, Table 2 listed only 5 examples of the consonant of the compound glyphs, their glyph components, and their component of syllable (onset, nucleus and coda) which were found in the glyph segment image collection. Interested reader can contact the authors of this paper to get the complete table.

⁷ www.omniglot.com/writing/balinese.htm

The rest of the glyphs consist of the numeral glyphs, punctuation glyphs, the vowel glyphs and special consonant glyph which can be used to change the speech of sound of other basic or compound glyphs (see the examples in Figure 4 and Table 3).

3.2 Glyph Properties and Categorizations

Seven properties to categorize the glyphs were specifically defined based on the existing glyph segment image collection. Property “**Id**” defined the identity number of the glyphs. The value of “**Id**” ranges from 1 to 133. Property “**Level1**” defined the name of the glyphs. The name of the glyphs normally represents their speech sound. But for some glyphs, the name is totally different. Property “**Level2**” is categorized in six groups: **CON** for consonant, **VOC** for vocal, **GAN** for gantungan, **GEM** for gempelan, **NUM** for numeral, and **PUN** for punctuation. **GAN** groups all second glyph form (conjunct form) of consonant basic glyphs (see Table 1) which should be written as a descender of other glyph (below other glyph).

GEM groups all second glyph form (conjunct form) of consonant basic glyphs (see Table 1) which should be written aside (right side) of other glyph. A special consonant called “*ADEG-ADEG*” is also categorized as **GEM** because, if it is needed, it can also annihilate the vowel “A” (the nucleus of the syllable) of their previous glyph. **VOC** groups not only the vowel glyph on the Table 3. The special consonant glyphs were also categorized as **VOC** because they can change the speech sound of other consonant glyph. One special case is glyph “*CECEK*”. It has two functions, as a special consonant and as a punctuation. In this system, glyph “*CECEK*” is considered as special consonant so it has the property of **VOC**.

All other basic and compound consonant glyphs are considered as **CON**. Property “**Level3**” defined the spatial position of the glyphs related to the medial text line on the manuscript. It will be used to construct the phonological rules for all glyphs from OCR results. Six categories of spatial position for the glyph were defined (Figure 5). **ASC**

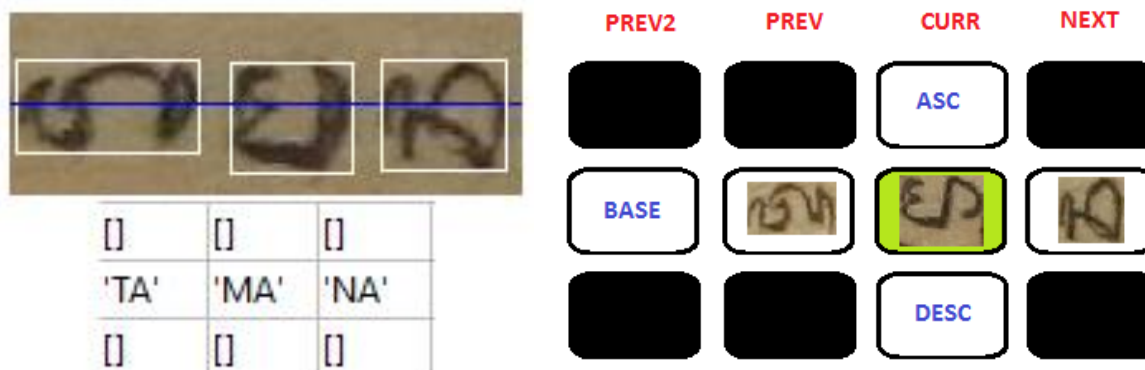


Fig. 6. Example of RULE1 and RULE8 which are applied to an OCR result

(ascender) glyphs and **DESC** (descender) glyphs do not intersect the medial text line, **ASC** glyphs are written above the medial text line, and **DESC** glyphs are written below the medial text line. **BASE** glyphs are normally the basic glyph segments written exactly on the medial text line, while **ASC-BASE** glyphs and **BASE-DESC** glyphs are normally the compound glyph segments between a **BASE** glyph with an **ASC** glyph or with a **DESC** glyph. In the glyph collection, there is only one **ASC-BASE-DESC** glyph class which is called glyph "ADEG-ADEG".

Property "**StartSyllable**" keeps the information of the onset of syllable for the consonant basic glyphs or the speech sound for the consonant compound glyphs, numeral, punctuation, and special consonant glyphs. Property "**EndSyllable**" keeps the information of the nucleus of syllable for the consonant basic glyphs (see Table 1).

Property "**SplitSyllable**" keeps the information of the onset, nucleus, and coda of syllable for the consonant compound glyphs (see Table 2). All information about glyph properties and categorizations are saved in a XML file and will be loaded on a list of glyph with a linked list pointer data structure.

4 Phonological Rules

The method of transliteration depends on the characteristics of the source and target languages [7]. Our model is based on phonetics which is based on traditional linguistic study of Balinese

transliteration. Based on the glyph segment image collection, the glyph properties and categorizations, the conditional scheme of phonological rules for the transliteration of Balinese script were finally identified and formally defined.

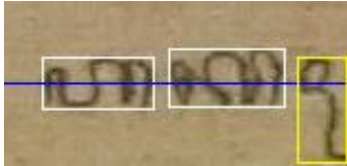
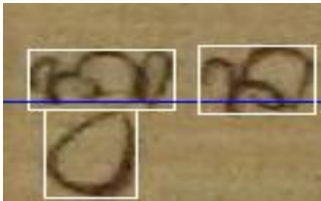
The phonological rules defining the transliteration are provided and are applied in sequential conditional checking order. The engine check all conditional phonological rules that might be applied to the given sequential glyph position and apply them in sequential order one by one. The OCR module for Balinese glyph recognition feeds the transliteration module with a sequential glyph data structure which is presented in Figure 6.

In Balinese script, the final speech sound for a syllable of a current (**CURR**) base (**BASE**) glyph will be determined by the ascender (**ASC**) of current glyph, the descender (**DESC**) of current glyph, the **BASE** of the **NEXT** glyph, the **BASE** of the previous (**PREV**) glyph, or even in some certain phonological rules, it can also be influenced by the **BASE** of the two previous (**PREV2**) glyphs.

The phonological rules for Balinese script transliteration should be finally built and be formally defined based on that OCR output data structure. Only a few examples of the phonological rules are described in this paper. Interested reader can contact the authors of this paper to get the complete rules.

```
RULE1: IF CURR.BASE.LEVEL1~==EMPTY AND
CURR.BASE.LEVEL2=CON/GEM AND
CURR.BASE.LEVEL3=BASE =>
SPEECH_SOUND=CURR.BASE.STARTSYLLABLE
```

Table 4. OCR output and phonological rules for transliteration of some sample word segments from Balinese palm leaf manuscript

| OCR Output | Phonological Rules Output | | | | | | | | | | | | |
|---|---------------------------|------------|------------|------|----------------|------------|---|------------|--|---|--|--|---|
|  <table border="1" style="margin-top: 10px;"> <tr><td> </td><td> </td><td> </td></tr> <tr><td>'A'</td><td>'KA'</td><td>'BISAH_...</td></tr> <tr><td> </td><td> </td><td> </td></tr> </table> | | | | 'A' | 'KA' | 'BISAH_... | | | | <pre> RULE1:..... RULE8:.....a a RULE1:.....K RULE8:.....Ka RULE15:.....KaH aKaH RULE32:.....* aKaH* Final Output: AKAH </pre> | | | |
| | | | | | | | | | | | | | |
| 'A' | 'KA' | 'BISAH_... | | | | | | | | | | | |
| | | | | | | | | | | | | | |
|  <table border="1" style="margin-top: 10px;"> <tr><td> </td><td> </td></tr> <tr><td>'KA'</td><td>'NA'</td></tr> <tr><td>'SUKU KEMBUNG'</td><td> </td></tr> </table> | | | 'KA' | 'NA' | 'SUKU KEMBUNG' | | <pre> RULE1:.....K RULE2:.....KW RULE12:.....Kwa Kwa RULE1:.....N RULE8:.....Na KwaNa Final Output: KWANA </pre> | | | | | | |
| | | | | | | | | | | | | | |
| 'KA' | 'NA' | | | | | | | | | | | | |
| 'SUKU KEMBUNG' | | | | | | | | | | | | | |
|  <table border="1" style="margin-top: 10px;"> <tr><td> </td><td> </td><td> </td><td> </td></tr> <tr><td>'NI'</td><td>'TALENG'</td><td>'WA'</td><td>'BISAH_...</td></tr> <tr><td> </td><td> </td><td> </td><td> </td></tr> </table> | | | | | 'NI' | 'TALENG' | 'WA' | 'BISAH_... | | | | | <pre> RULE17:.....NI NI RULE32:.....* NI* RULE1:.....W RULE3:.....WE RULE15:.....WEH NI*WEH RULE32:.....* NI*WEH* Final Output: NIWEH </pre> |
| | | | | | | | | | | | | | |
| 'NI' | 'TALENG' | 'WA' | 'BISAH_... | | | | | | | | | | |
| | | | | | | | | | | | | | |
|  <table border="1" style="margin-top: 10px;"> <tr><td> </td><td> </td><td> </td><td> </td></tr> <tr><td>'TALENG'</td><td>'NA'</td><td>'TEDONG'</td><td>'RA'</td></tr> <tr><td> </td><td> </td><td> </td><td> </td></tr> </table> | | | | | 'TALENG' | 'NA' | 'TEDONG' | 'RA' | | | | | <pre> RULE32:.....* * RULE1:.....N RULE4:.....No *No RULE32:.....* *No* RULE1:.....R RULE8:.....Ra *No*Ra Final Output: NORA </pre> |
| | | | | | | | | | | | | | |
| 'TALENG' | 'NA' | 'TEDONG' | 'RA' | | | | | | | | | | |
| | | | | | | | | | | | | | |

```

RULE8: IF PREV.BASE.LEVEL1~="TALENG" AND
CURR.ASC.LEVEL1=EMPTY AND
CURR.BASE.LEVEL1~=EMPTY AND
CURR.BASE.LEVEL2=CON AND
CURR.BASE.LEVEL3=BASE AND
CURR.BASE.ENDSYLLABLE="A" AND
CURR.DESC.LEVEL1=EMPTY AND
NEXT.BASE.LEVEL1~="ADEG-ADEG" AND
NEXT.BASE.LEVEL2~="GEM =>
SPEECH_SOUND=SPEECH_SOUND+"A"

```

For the example in Figure 6, when the transliteration engine proceeds the glyph “MA” as the **CURR BASE** glyph, *RULE1* will be applied first, because glyph “MA” is a **CON BASE** type glyph. In this case, the transliteration engine will take first the **STARTSYLLABLE** of glyph “MA” as the speech sound output, which is “M”. The second rule which can be applied of this sequential glyph position for glyph “MA” is the *RULE8*.

RULE8 describes the concatenation of speech sound “A” with the last speech sound output, because the **CURR BASE** glyph does not have an **ASC** glyph and/or a **DESC** glyph. The **PREV BASE** glyph is not a glyph “TALENG” and the **NEXT BASE** glyph is not a glyph “ADEG-ADEG” or **GEM** type glyph. The final speech sound output will be “MA”. If all the same sequential rules can be applied to glyph “TA” on the left and glyph “NA” on the right of glyph “MA”, the final transliteration output will be “TAMANA”.

5 Experimental Results

To evaluate the correctness of phonological rules, the experiments of transliteration process for sample word segments of Balinese palm leaf manuscript were performed for all rules (see some examples in Table 4). Before the transliteration process, the OCR for Balinese script was performed.

6 Conclusions and Future Works

The knowledge representation and phonological rules for Balinese script which were presented in this paper already perform an effective and correct transliteration process. To improve the accuracy of OCR for Balinese script on palm leaf manuscripts, a lexicon-based statistical approach is needed. It should be useful to integrate the phonological rules

in the OCR post correction step when the physical feature description is failed to do the recognition process. Our future interests are to build an optimal lexicon dataset for our system, in term of quantity and completeness of the dataset and to define an appropriate lexicon information (characters, syllables and words level) for Balinese script.

Acknowledgements

The authors would like to thank Museum Gedong Kertya, Museum Bali, and all families in Bali, Indonesia, for providing us the samples of palm leaf manuscripts, and the students from Department of Informatics Education and Department of Balinese Literature, Ganesha University of Education for helping us in ground truthing process for this experiment. This work is also supported by the DIKTI BPPLN Indonesian Scholarship Program and the STIC Asia Program implemented by the French Ministry of Foreign Affairs and International Development (MAEDI).

References

1. Kesiman, M.W.A., Prum, S. Burie, J.C., & Ogier, J.M. (2016). Study on Feature Extraction Methods for Character Recognition of Balinese Script on Palm Leaf Manuscript Images. *23rd International Conference on Pattern Recognition*, Cancun.
2. Ramteke, R.J. (2010). Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition. *International Journal Comput. Appl.*, Vol. 1, No. 18, pp. 1–5.
3. Aggarwal, A., & Singh, K. (2015). Use of Gradient Technique for Extracting Features from Handwritten Gurmukhi Characters and Numerals. *Proceeding Comput. Science*, Vol. 46, pp. 1716–1723.
4. Zahid-Hossain, M., Ashraful-Amin, M., & Hong Y. (2012). *Rapid Feature Extraction for Optical Character Recognition*. Vol. abs/1206.0238.
5. Ashlin-Deepa, R.N., & Rajeswara-Rao, R. (2014). Feature Extraction Techniques for Recognition of Malayalam Handwritten Characters: Review. *Journal Adv. Trends Comput. Sci. Eng. IJATCSE*, Vol. 3, No. 1, pp. 481–485.
6. Shishtla, P., Ganesh, V. S., Subramaniam, S., & Varma, V. (2009). *A language-independent transliteration schema using character aligned models at NEWS 2009*. pp. 40.

7. **AbdulJaleel, N. & Larkey, L. S. (2004).** *English to Arabic Transliteration for Information Retrieval: A Statistical Approach.*
8. **Finch, A. & Sumita, E. (2010).** Transliteration using a Phrase-based Statistical Machine Translation System to Re-score the Output of a Joint Multigram Model. *Proceedings of the 2010 Named Entities Workshop, Sweden*, pp. 48–52.
9. **Pretkalinina, L., Paikens, P., Gruzitis, N., Rituma, L. & Spektors, A. (2012).** Making Historical Latvian Texts More Intelligible to Contemporary Readers. *Proceedings of the LREC Workshop on Adaptation of Language Resources and Tools for Processing Cultural Heritage Objects.*
10. **Kesiman, M. W. A., Burie, J.C., Ogier, J.-M., Wibawantara, G. N. M. A., & Sunarya, I. M. G. (2016).** AMADI_LontarSet: The First Handwritten Balinese Palm Leaf Manuscripts Dataset. *15th International Conference on Frontiers in Handwriting Recognition, Shenzhen, China*, pp. 168–172.

*Article received on 23/12/2016; accepted on 20/02/2017.
Corresponding author is Made Windu Antara Kesiman.*