

# Implementing Kinzi

*Martin Hosken,  
SIL Non-Roman Script Initiative (NRSI)*

## Executive Summary

This paper examines the storage and rendering of kinzi in more detail and suggests the following wording to appear in the Unicode Standard rubric discussion of Myanmar script on p250 directly following the example of storing kinzi.

In addition to rendering as a kinzi, a conjunct form of a U+1004 MYANMAR LETTER NGA, may appear as a full size U+1004 MYANMAR LETTER NGA with a visible killer above it. This would be stored as:

```
U+1004 MYANMAR LETTER NGA
U+1039 MYANMAR SIGN VIRAMA
U+200C ZERO WIDTH NON-JOINER
```

In addition, U+1004 MYANMAR LETTER NGA may appear as a base consonant with its own "medial" consonants. In such case, in order to distinguish from the kinzi case, the following encoding should be used:

```
U+1004 MYANMAR LETTER NGA
U+200C ZERO WIDTH NON-JOINER
U+1039 MYANMAR SIGN VIRAMA
...
```

## Introduction

Burmese has a letter called kinzi (ꯀ) which is used in some cases instead of င်. The encoding of this character is clearly stipulated in the Unicode Standard as being U+1004 U+1039. The problems arise when considering how to render various character sequences containing this character sequence.

## Kinzi or not?

Identifying when a kinzi should be present and when it should not can be somewhat controlled through the use of the ZWNJ character (U+200C). Thus the following character sequences would be rendered as shown:

```
... U+1004 U+1039 U+1000          ၎
```

... U+1004 U+1039 U+200C U+1000      عم

## First or Last?

When it is clear that the sequence U+1004 U+1039 represents a final consonant, then it is not difficult to decide whether to render a string using a kinzi or not. The problem is that it is not always clear whether the sequence U+1004 U+1039 represents a final consonant or whether it represents an initial consonant and the start of a following medial.

Consider the sequence U+1004 U+1039 U+101F ...<sup>1</sup> It is not clear whether this should be rendered as ع or as ق. In the former case, we consider the sequence as (U+1004 U+1039) U+101F and in the latter as U+1004 (U+1039 U+101F).

This is the real heart of the kinzi implementation problem since there is no way to disambiguate the two sequences. There are a number of approaches to addressing the issue.

## ZWNJ disambiguation

One approach is to use the ZWNJ to disambiguate. We can't use U+1004 U+1039 U+200C U+101F to identify the former case, since this would be rendered as عق which is neither. So, instead we must use the sequence U+1004 U+200C U+1039 U+101F to identify the second case. This is unfortunate, since in clear cases of ambiguity, as we shall see, the second rendering provides the better fallback.

## Language Rules

The first thing to consider is that kinzi (ع) is actually fairly rare. The second thing is that while both spellings are phonetically equivalent, only one is conventionally used. Therefore, on the assumption that there is a solution to the problem, we can find that solution<sup>2</sup>. I propose, therefore, that the solution is that kinzi is only used with characters that do not change shape (as opposed to size) when medialised. Thus we can state that:

In the sequence U+1004 U+1039 C, the sequence is interpreted as (U+1004 U+1039) C in all cases except when C is one of U+101A U+101B U+101D U+101F when it is interpreted as U+1004 (U+1039 C).

In addition,

The sequence U+1004 U+1039 C U+1039 U+200C shall be interpreted as U+1004 U+1039 U+200C C U+1039 U+200C.

The problem with this second rule is that you then end up with two different spellings being rendered identically (with and without the middle U+200C).

---

<sup>1</sup> This sequence is found as part of the word for 'extinguish' (U+1004 U+1039 U+101F U+1039 U+101B U+102D U+1019 U+1039 U+200C U+1038) عق:

<sup>2</sup> If we did not know that there was a solution, then the problem would be insoluble. But since we have the extra information that the solution is soluble, we can go about solving it. This is one of that strange class of problem that requires the one piece of information that a solution can be found to resolve an ambiguity in the solution.

## **Conclusion**

The question is whether the rules proffered here are appropriate and should be the basis for rendering kinzi or whether explicit marking is always required.