

# Summary of Assumptions for Determining the Root Letter when Machine Sorting Dzongkha

Compiled by the Dzongkha Development Authority  
Ministry of Education  
Royal Government of Bhutan

April 2004

## **Overview**

Sorting Dzongkha poses a number of challenges. Proper sorting of Dzongkha requires that the root letter be identified for each syllable. This is relatively straightforward for (i) syllables having one or more consonants stacked in vertical combination with the root letter and (ii) syllables whose root letter is marked by an explicit vowel sign. However, it can be very difficult to ascertain the root letter in Dzongkha syllables which are comprised of a sequence of simple consonants.

To complicate matters further a tailored collation that conforms to the Unicode Collation Algorithm and ISO/IEC 14651 will not necessarily be aware of when a particular character sequence occurs at the beginning of a syllable versus when that same sequence occurs in the middle of a syllable or at the end of a syllable. As a result certain Dzongkha syllables / character sequences that are not ambiguous to a human reader will be potentially ambiguous from the perspective of a computerized sorting algorithm.

On the basis of exhaustive lexical research the Dzongkha Development Authority has drafted the following roster of assumptions which enable Dzongkha to be reliably machine sorted using a general purpose (multi-weight) collation algorithm.

Also note that this document specifies conventions for data entry (in subsections C.1, C.4 and C.5) that must be followed in order to ensure proper collation of Dzongkha materials.

## Roster of Assumptions

### A. Potentially ambiguous items ending with མིང་མཐའ་ ལ་

A.1. Assume that the following 13 items do not exist:

མག་ལ་	བད་ལ་	འབ་ལ་
འག་ལ་	མད་ལ་	གས་ལ་
དང་ལ་	མན་ལ་	བས་ལ་
མང་ལ་	དཔ་ལ་	
གད་ལ་	དབ་ལ་	

A.2. Assume that the following 5 items have the second letter as the root letter:

དག་ལ་	འད་ལ་	དམ་ལ་
བག་ལ་	གན་ལ་	

**B. Potentially ambiguous items ending with མིང་མཐའ་ པ་ རྟེན་ རྩོམ་**

B.1. Assume that the following 29 items **do not exist**:

རྟེན་ལོ་	བཟུ་	རྟེན་ལོ་
བཟུ་	བཟུ་ལོ་	རྟེན་ལོ་
བཟུ་ལོ་	མཟུ་	རྟེན་ལོ་
མཟུ་	མཟུ་ལོ་	འབུ་
མཟུ་ལོ་	འབུ་	འབུ་ལོ་
འབུ་	འབུ་ལོ་	རྟེན་ལོ་
འབུ་ལོ་	གཟུ་	གཟུ་ལོ་
རྟེན་ལོ་	གཟུ་ལོ་	བཟུ་
མཟུ་	མཟུ་ལོ་	བཟུ་ལོ་
གཟུ་ལོ་	རྟེན་ལོ་	

B.2. Assume that the following 7 items have the **first** letter as the root letter:

རྟེན་ལོ་	གཟུ་	གཟུ་
རྟེན་ལོ་	མཟུ་	
མཟུ་	རྟེན་ལོ་	

**C. Potentially ambiguous items ending with either suffix ན་ or**

མེང་མཐའ་ ན་ རྟ་ རྟི

C.1. The following item is **ambiguous** as to which letter is the **root letter** -- since the second letter is the root letter in Choekay but the first letter can be the root letter in Dzongkha:

མངན་

The **second** letter will be treated as the root letter since this accords with Choekay.

Therefore, when the first letter needs to be treated as the root letter (in Dzongkha), this item must be written as: མངན་

C.2. Assume that the following 19 items **do not exist**:

བགན་	བདན་	དཔན་
བགནི	བདནི	དཔནི
མགན་	མདན་	འབན་
མགནི	འདནི	བམན་
འགན་	གནན་	བམནི
དངན་	མནན་	
གདན་	དཔན་	

C.3. Assume that the following 2 items have the **first** letter as the root letter:

དགན་

དགནི

C.4. Assume that the following 11 items have the **second** letter as the root letter:

འགན'	འདན'	འཇན'
དངན'	གནན'	དམན'
གདན'	མནན'	གསན' **
མདན'	དཇན'	

\*\*Dzongkha syllables **ending** in གསན should be written as ending with གྲ (i.e. གྲན ) in order to assist in proper sorting. Thus:

མགྲོན' འདོན' རློན' ལྷོན' རྫོན' རྩོན' བཟུན' གཟིན' ལེན'

C.5. The following 3 items, with the **first** letter as the root letter, will not always sort properly -- due to the existence of a syllable ending in ན' (shown in brackets) in which the **second** letter is the root letter:

མངན' [མངན']	དམན' [དམན']	གསན' [གསན'] **
-------------	-------------	----------------

\*\*Dzongkha syllables **ending** in གསན should be written as ending with གྲི (i.e. གྲིན ) in order to assist in proper sorting. Thus:

མགྲོན' འདོན' རློན' ལྷོན' རྫོན' རྩོན' བཟུན' གཟིན' ལེན'

Therefore, when the first letter needs to be treated as the root letter (in Dzongkha), these above items must be written as:

མངྲི	དམྲི	གསྲི
------	------	------

**D. Potentially ambiguous items ending with either *suffix* མ' or**

མིང་མཐའ་ མ' རྣམ་ མོ'

D.1. Assume that the following 25 items **do not exist**:

དགམ'	བདམོ'	དབམོ'
དགམོ'	མདམ'	འབམོ'
བགམོ'	མདམོ'	དམམ'
མགམ'	འདམོ'	དམམོ'
མགམོ'	གཞམོ'	གསམ'
འགམོ'	མཞམོ'	གསམོ'
དངམོ'	དཔམ'	བསམོ'
མངམོ'	དཔམོ'	
གདམོ'	དབམ'	

D.2. Assume that the following 2 items have the **first** letter as the root letter:

དངམ'                      མངམ'

D.3. Assume that the following 9 items have the **second** letter as the root letter:

བགམ'                      འགམ'                      གདམ'

འདམ'

གཞམ'

འབམ'

འདམ'

མཞམ'

བམམ' \*\*

\*\*An ambiguity would result for any Dzongkha syllables ending in བམམ . Assume that no Dzongkha syllables end with བམམ .

**E. Potentially ambiguous items ending with *suffix* ག' or མ'**

followed by མིང་མཐའ་ རྟི

E.1. To resolve this potential ambiguity, assume that there does not exist any Choekay or Dzongkha syllable with root letter རྟི and having prefix letter ག' or མ' (with or without any suffix letters).

**F. Potentially ambiguous items ending with *suffix* ག' or གས', ང' or ངས', བ' or བས', མ' or མས'.**

F.1. Assume that the following 2 items do not exist:

མགས'

དྲངས'

F.2. Assume that the following 2 items have the first letter as the root letter:

བགས'

མངས'

F.3. Assume that the following 5 items have the second letter as the root letter:

དགས'	དབས'	དམས'
འགས'	འབས' **	

\*\* འབས' occurs only rarely as contraction of syllable འབའ' with grammar particle ས .

G. Assume that ཐུ', ཐོ', ཐོཾ' do not exist as possible མིང་མཐུན་ for any syllables that begin with the following two-letter sequences (and lacking an explicit vowel):

དག_	གད_	དམ_
བག_	བད_	དབ_
མག_	མད_	འབ_
འག_	འད_	དམ_
དང_	གད_	གས_
མང_	མད_	བས_

That is to say, assume that དགཐུ', དགཐོ' and དགཐོཾ' do not exist; and so on for each of the remaining 17 items in the list -- with each combination likewise assumed to not exist.