Converting Phake Dictionary to Unicode

# List of fields in use – Tai Phake Dictionary

|  |  |  |
| --- | --- | --- |
| field code | field name | script |
| \lx | Lexeme | Tai Phake |
| \le | Lexeme Alternative Spelling | Tai Phake |
| \ph | Phonetic form | IPA (Banchob font) |
| \so | Source (listing of sound file link) |  |
| \hm | Homonym number |  |
| \ps | Part of speech |  |
| \sn | Sense number |  |
| \de | Definition (English) |  |
| \ge | Gloss (English) |  |
| \pc1 | (Picture 1) |  |
| \pc2 | Picture 2 |  |
| \pl | Couplet form [We use this to help distinguish words in this language, which has otherwise a lot of homography] | Tai Phake |
| \pd | Couplet form phonetic | IPA (Banchob font) |
| \pde | Couplet form English |  |
| \pdn | Couple form Assamese | Assamese |
| \dn | Definition Assamese | Assamese |
| \rf | Reference [For example sentences taken from texts] |  |
| \xv | Example Phake | Tai Phake |
| \xr | Example Phonetic | IPA (Banchob font) |
| \xe | Example free translation English |  |
| \xn | Example free translation Assamese | Assamese |
| \notes | Notes |  |
|  |  |  |
| then |  |  |
| \se | Subentry |  |

# Conversion table for the Banchob font

|  |  |
| --- | --- |
| Banchob font | Unicode |
| N | ŋ |
| M | ñ |
| j | ɛ |
| v | ü |
| z | ə |
| q | ɔ |
| I | ī |
| E | ē |
| J | ɛ̄ |
| V | ǖ |
| Z | ə̄ |
| A | ā |
| U | ū |
| O | ō |
| Q | ɔ̄ |
| 1 | ¹ |
| 2 | ² |
| 3 | ³ |
| 4 | ⁴ |
| 5 | ⁵ |
| 6 | ⁶ |
| 7 | ⁷ |
| 8 | ⁸ |
| 9 | ⁹ |

# Changes of font within fields

Sometimes there are changes of fonts to the IPA within fields marked as <banchob kAN4 nz6>

Unfortunately this has not been done consistently, as we see from these examples

\de an intensifying word for <yqm6> emacited, thin

\de deer, (a corrupted sound of < kwaaN2 > )

\de a hall for worshiping Lord Buddha, <used with kyauN2>

\pde an administrative unit in a geographically lower area, such as for example Lower Assam,

whereas Upper Assam would be termed as <banchob kAN4 nz6>

\de even, equal in number, amount, or value, size, used in negative form

with <vN5 > as < uN5 kA2>

\de to give light with the torch, or lantern as<kai3phai2>

However all of these have < and >

I think the change could be implemented by saying change everying within the < > brackets according to the the conversion table above and removing < and > (and any consequent double spaces.