‘Tibetan in digital communication’

Nathan W. Hill and Edward Garrett

(SOAS, University of London)
Presentation overview

1. Current state of e-resources for Tibetan
2. Current state of e-resources for English
3. ‘Tibetan in digital communication’
Current Tibetan e-resources: OTDO

<table>
<thead>
<tr>
<th>Pt_0016 (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25v3</td>
</tr>
<tr>
<td>pas nI myI gso ba myed do / bod kyI lha btsan po myI rje lhäs mdzad pa khrI gtsug lde btsan gyI zha s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pt_0126 (87)</th>
</tr>
</thead>
<tbody>
<tr>
<td>002</td>
</tr>
<tr>
<td>u [---] [dga']? la / s[il]g pa ngan - dgu gtong myI phod / tshe cig btsan phyug mchog du 'dzin /</td>
</tr>
<tr>
<td>003</td>
</tr>
<tr>
<td>phyug mchog du 'dzin / - lha chos bzang po goig myI spyod / ma tshor bar du bdud - khyis khyer / g</td>
</tr>
<tr>
<td>004</td>
</tr>
<tr>
<td>cd / ma tshor bar du bdud khyis khyer / gne' myI 'dzangs shing che dgu yang / tshe ring da lta - gs</td>
</tr>
<tr>
<td>005</td>
</tr>
<tr>
<td>shing che dgu yang / tshe ring da lta - goig myI snang / shi zin phan cad sa phung yin / myI c</td>
</tr>
<tr>
<td>006</td>
</tr>
<tr>
<td>g myI snang / shi zin phan cad sa phung yin / myI cig - skyes pa'i tshe dus la ring por myi thogs d</td>
</tr>
<tr>
<td>007</td>
</tr>
<tr>
<td>yin / myI cig - skyes pa'i tshe dus la ring por myi thogs de'u re myed / khrel - dang gzungs myed m</td>
</tr>
<tr>
<td>008</td>
</tr>
<tr>
<td>gs myed mna' yang bsa' / mtha' ma yun du - legs myI spyod / tshe srog thung ba glog par myur / gs</td>
</tr>
<tr>
<td>010</td>
</tr>
<tr>
<td>par 'dre dang srin / gtan kyi grogs ngan skyid myI - myong / dge sdig myI shes nyon re mongs / n</td>
</tr>
<tr>
<td>011</td>
</tr>
<tr>
<td>tan kyi grogs ngan skyid myI - myong / dge sdig myI shes nyon re mongs / nang nub myI dal - phyug p</td>
</tr>
<tr>
<td>011</td>
</tr>
<tr>
<td>/ dge sdig myI shes nyon re mongs / nang nub myI dal - phyug por 'dod / lhag par yod kyang chog</td>
</tr>
</tbody>
</table>
'person'
<table>
<thead>
<tr>
<th>lha btsan po</th>
<th>myi rje lhas m</th>
</tr>
</thead>
<tbody>
<tr>
<td>la lta shung</td>
<td>myi snang</td>
</tr>
<tr>
<td>hung yin</td>
<td>myi cig - skye</td>
</tr>
<tr>
<td>la ring</td>
<td>myi thogs de'u</td>
</tr>
<tr>
<td>gun du</td>
<td>myi spyod</td>
</tr>
</tbody>
</table>
I want 'chair' as a verb.

<table>
<thead>
<tr>
<th>Corpus: British National Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits: 10103 (90.1 per million)</td>
</tr>
</tbody>
</table>

| Page 4 of 506 | Go | Next | Last |

| JA8 | you're sitting. If you can't bring your chair over. Kendal just bring your chair over.
| JA8 | your chair over. Kendal just bring your chair over. Come round and see the board. What
| JA8 | with you now okay? So can you move your chairs so that you can stand up quickly. I think
| JA8 | good. Now Jennifer if you'll just move your chair over to this side. Who's the best one at
| J42 | have to er withdraw and ask er to go to chair the meeting. If er if you don't mind. Do
| J42 | going into and you're going to take the chair for this. Well I hope it won't be a is
| J42 | I would agree. Chairman, may I'm sorry now chaired by Peter. Anyway erm I think we totally
| J9C | Wyle, I expect the courtesy towards the Chair, and I am now summing up, so please be
| J9C | back there, would you count, please. Eight, Chair And against. Six. Seven. Would would you
Part of speech tagging solves the problem

<table>
<thead>
<tr>
<th>Corpus: British National Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits: 1086 (9.7 per million)</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>HTK</td>
</tr>
<tr>
<td>HTK</td>
</tr>
<tr>
<td>HTK</td>
</tr>
<tr>
<td>HTK</td>
</tr>
<tr>
<td>HTK</td>
</tr>
<tr>
<td>HGN</td>
</tr>
<tr>
<td>HGS</td>
</tr>
<tr>
<td>HBN</td>
</tr>
</tbody>
</table>

They are all 'chair' as a verb!

- HTK: of the new panel, which he was asked to **chair**, and by virtue of his chairmanship he also
- HTK: presiding over, and Sir Denis Rooke was **chairing**, what was indeed by far the largest degree
- HTK: member of the visiting party, which was **chaired** by Edwin Kerr, reported back to the group
- HTK: resolved by the visiting party member who **chaired** the meeting allowing the head of department
- HTK: development was an agenda item for this meeting. **Chairing** the Working Party, Edwin Kerr called on
- HGN: if this is a meeting, she can bloody well **chair** it. I raise my glass and watch the bubbles
- HGS: `<p>` `That was the Lord Mayor's nephew **chairing** the panel, you know...' Molly announces, `</p>`
- HBC: the bottom. A powerful Steering Committee, **chaired** by Rod Turnbull, Managing Director, with
- HBN: were represented at the launch which was **chaired** by Wedgwood Group director of human resources
- HBN: Education, SCOTVEC and SED and was very ably **chaired** by Mr J Murray, Dean of Technology, Napier
But it gets even fancier...

- Sketch: typical subjects, objects, modifiers, etc.
- Thesaurus: words with similar meanings
**Sketch**

**chair (verb)**  
British National Corpus freq = 1086 (9.7 per million)

<table>
<thead>
<tr>
<th>object</th>
<th>freq</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub-committee</td>
<td>547</td>
<td>3.8</td>
</tr>
<tr>
<td>meeting</td>
<td>100</td>
<td>6.42</td>
</tr>
<tr>
<td>committee</td>
<td>106</td>
<td>6.21</td>
</tr>
<tr>
<td>seminar</td>
<td>6</td>
<td>5.7</td>
</tr>
<tr>
<td>panel</td>
<td>13</td>
<td>5.47</td>
</tr>
<tr>
<td>session</td>
<td>17</td>
<td>5.45</td>
</tr>
<tr>
<td>inquiry</td>
<td>8</td>
<td>4.96</td>
</tr>
<tr>
<td>commission</td>
<td>19</td>
<td>4.8</td>
</tr>
<tr>
<td>enquiry</td>
<td>5</td>
<td>4.73</td>
</tr>
<tr>
<td>forum</td>
<td>3</td>
<td>4.55</td>
</tr>
<tr>
<td>tribunal</td>
<td>3</td>
<td>4.45</td>
</tr>
<tr>
<td>conference</td>
<td>15</td>
<td>4.44</td>
</tr>
<tr>
<td>debate</td>
<td>8</td>
<td>4.18</td>
</tr>
<tr>
<td>board</td>
<td>12</td>
<td>3.57</td>
</tr>
<tr>
<td>talk</td>
<td>6</td>
<td>3.55</td>
</tr>
<tr>
<td>cabinet</td>
<td>3</td>
<td>3.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>subject</th>
<th>freq</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvey</td>
<td>3</td>
<td>7.3</td>
</tr>
<tr>
<td>Carrington</td>
<td>5</td>
<td>7.14</td>
</tr>
<tr>
<td>Sinclair</td>
<td>3</td>
<td>6.4</td>
</tr>
<tr>
<td>Fraser</td>
<td>4</td>
<td>5.8</td>
</tr>
<tr>
<td>Yeltsin</td>
<td>3</td>
<td>5.1</td>
</tr>
<tr>
<td>judge</td>
<td>11</td>
<td>4.96</td>
</tr>
<tr>
<td>Russia</td>
<td>3</td>
<td>4.07</td>
</tr>
<tr>
<td>president</td>
<td>7</td>
<td>3.54</td>
</tr>
<tr>
<td>minister</td>
<td>15</td>
<td>3.05</td>
</tr>
<tr>
<td>secretary</td>
<td>7</td>
<td>2.8</td>
</tr>
<tr>
<td>leader</td>
<td>4</td>
<td>1.93</td>
</tr>
<tr>
<td>person</td>
<td>5</td>
<td>1.73</td>
</tr>
<tr>
<td>director</td>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>manager</td>
<td>3</td>
<td>1.25</td>
</tr>
<tr>
<td>member</td>
<td>6</td>
<td>1.2</td>
</tr>
<tr>
<td>friend</td>
<td>4</td>
<td>1.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>and/or</th>
<th>freq</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>convene</td>
<td>3</td>
<td>6.59</td>
</tr>
<tr>
<td>arrange</td>
<td>3</td>
<td>3.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>part_intrans</th>
<th>freq</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>by</td>
<td>4</td>
<td>5.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>unary rels</th>
<th>freq</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>passive</td>
<td>450</td>
<td>10.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pp_by-p</th>
<th>freq</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvey</td>
<td>3</td>
<td>7.52</td>
</tr>
<tr>
<td>Carrington</td>
<td>5</td>
<td>7.25</td>
</tr>
<tr>
<td>Sinclair</td>
<td>3</td>
<td>6.52</td>
</tr>
<tr>
<td>Yeltsin</td>
<td>3</td>
<td>5.15</td>
</tr>
<tr>
<td>judge</td>
<td>8</td>
<td>4.51</td>
</tr>
<tr>
<td>Russia</td>
<td>3</td>
<td>4.09</td>
</tr>
<tr>
<td>president</td>
<td>8</td>
<td>3.74</td>
</tr>
</tbody>
</table>
# Thesaurus

**Chair**

British National Corpus freq = 1086

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Score</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>convene</td>
<td>0.108</td>
<td>528</td>
</tr>
<tr>
<td>attend</td>
<td>0.102</td>
<td>8993</td>
</tr>
<tr>
<td>host</td>
<td>0.088</td>
<td>1096</td>
</tr>
<tr>
<td>conclude</td>
<td>0.085</td>
<td>5552</td>
</tr>
<tr>
<td>adjourn</td>
<td>0.081</td>
<td>722</td>
</tr>
<tr>
<td>organize</td>
<td>0.078</td>
<td>3366</td>
</tr>
<tr>
<td>initiate</td>
<td>0.076</td>
<td>2064</td>
</tr>
<tr>
<td>conduct</td>
<td>0.074</td>
<td>5099</td>
</tr>
<tr>
<td>schedule</td>
<td>0.073</td>
<td>1602</td>
</tr>
<tr>
<td>approve</td>
<td>0.073</td>
<td>5241</td>
</tr>
<tr>
<td>address</td>
<td>0.071</td>
<td>6025</td>
</tr>
<tr>
<td>head</td>
<td>0.069</td>
<td>5885</td>
</tr>
<tr>
<td>propose</td>
<td>0.069</td>
<td>7341</td>
</tr>
</tbody>
</table>
and fancier still...

- Comparing the profile of different words
- Suggestions of example sentences
- Comparing usage across region, time period, genre
- Automatic translation
- Speech recognition
- … and on and on ...
Can we do anything like this for Tibetan?
Can we do anything like this for Tibetan?

Not yet, but we are working on it.
A second try with མི་ 'person'

74.170a

1. ... སེལ་|cv.imp བེས་|cl.quot བུས་|v.past.v.pres བོས་|cv.ela མི་|n.count དེ།|d.dem བོ།|case.agn ...

2. ... བི་|case.gen བེས་|n.count བུས་|v.past བོ།|cv.sem ||punc མི་|n.count དེ།|d.dem བོ།|case.agn བུ། བོ།|n.count ...

3. ... བཤུས་|v.fut.v.pres བོ།|cv.fin བེས་|cl.quot བུས་|n.v.invar བོ།|case.ass ||punc མི་|n.count བུ། བུ། བུ།|n.count ...

Analysis Pre-tagging

74.174a

1. ... བི།|case.term བོ།|case.loc མི་|n.count བེས་|d.indef ཁྱུད་|ཐོག་|n.count དེ།|case.term བུ། བུ། བུ།|n.v.past ...

2. ... བཞེས་|n.v.fut.v.pres བོ།|n.rel བོ།|case.term མི་|n.count དེ།|d.dem བོ།|case.gen བེས་|n.v.pres བི།|case.gen བུ། བུ།|n.count ...

3. ... བེས།|cl.quot བུ།།|v.past བོ།|cv.fin ||punc ||punc མི་|n.count དེ།|d.dem བོ།།|v.past བུ།།|n.v.invar ...

Analysis Pre-tagging
Tibetan in Digital Communication

Goals

1. 1,000,000 word part-of-speech tagged corpus of Tibetan texts
2. An automatic word breaker
3. An automatic part-of-speech tagger
Tibetan Word Breaking
and
Part of Speech Categories

Nathan W. Hill and Edward Garrett
(SOAS, University of London)
Part 1
Word Breaking
Section overview

1. The phonological word
2. The syntactic word
The Phonological Word

1. Phonotactics
   - Complex onset normally only in first syllable of two syllable words
1. Phonotactics
   - Complex onset normally only in first syllable of two syllable words
2. The domain of sandhi
   - Case markers, converses, clitics, are part of preceding word
     - \textit{rgyal-pos, rgyal-por}
     - \textit{khyod-kyis}
     - \textit{ri-rab-bas-kyan mtho} ‘higher even than Mount Meru’
The Phonological Word

1. Phonotactics
   - Complex onset normally only in first syllable of two syllable words

2. The domain of sandhi
   - Case markers, converbs, clitics, are part of preceding word
   - rgyal-pos, rgyal-por
   - khyod-kyis
   - ri-rab-bas-kyan mtho ‘higher even than Mount Meru’

3. The domain of tone (in modern languages)
   - One tone assigned to each lexical noun
   - One tone assigned to each verb phrase
The syntactic word

1. Cases and converses govern syntactic units larger than the phonetic word
The syntactic word

1. Cases and converbs govern syntactic units larger than the phonetic word

2. Useful to make verbs as small as possible
   - saṅs mi rgyas te, saṅs e rgyas
The syntactic word

1. Cases and converses govern syntactic units larger than the phonetic word

2. Useful to make verbs as small as possible
   - saṅs mi rgyas te, saṅs e rgyas

3. Useful to make nouns as large as possible
   - Our project is not interested in word internal derivation
   - Makes noun phrase syntax simper
The syntactic word

1. Cases and converses govern syntactic units larger than the phonetic word

2. Useful to make verbs as small as possible
   - saṅs mi rgyas te, saṅs e rgyas

3. Useful to make nouns as large as possible
   - Our project is not interested in word internal derivation
   - Makes noun phrase syntax simper
     - pha-ma ‘parents’
The syntactic word

1. Cases and converses govern syntactic units larger than the phonetic word

2. Useful to make verbs as small as possible
   - *sañs mi rgyas te, sañs e rgyas*

3. Useful to make nouns as large as possible
   - Our project is not interested in word internal derivation
   - Makes noun phrase syntax simper
     - *pha-ma* ‘parents’
     - *de-bzin-gsegs-pa-sini-po* ‘tathāgatagarbha’
The syntactic word

1. Cases and converses govern syntactic units larger than the phonetic word

2. Useful to make verbs as small as possible
   • saṅs mi rgyas te, saṅs e rgyas

3. Useful to make nouns as large as possible
   • Our project is not interested in word internal derivation
   • Makes noun phrase syntax simper
     • pha-ma ‘parents’
     • de-bzin-gsegs-pa-sni-po ‘tathāgatagarbha’
     • lus-ñag-sems gsum ‘the three, body speech, and mind’
The syntactic word

1. Cases and converbs govern syntactic units larger than the phonetic word

2. Useful to make verbs as small as possible
   - *saṅs mi rgyas te, saṅs e rgyas*

3. Useful to make nouns as large as possible
   - Our project is not interested in word internal derivation
   - Makes noun phrase syntax simper
     - *pha-ma* ‘parents’
     - *de-bzin-gség-pa-sñiṅ-po* ‘tathāgatagarbha’
     - *lus-ngag-sems gsum* ‘the three, body speech, and mind’
     - *khaṅ-žiṅ-gser-g.yu mdzo-rta gnag-lug* ‘house, fields, gold, turquoise, dzo, horses, yaks, sheep’
Compounds

• Two nouns in succession are understood as a compound;
  
  dandva like pha-ma 'parents' (pha 'father' and ma 'mother')
  
  tatpuruṣa like khyim-bdag 'householder' (khyim 'home' and bdag 'lord')
Compounds

- Two nouns in succession are understood as a compound;
  
  *dandva* like *pha-ma* 'parents' (*pha* 'father' and *ma* 'mother')
  
  *tatpuruṣa* like *khyim-bdag* 'householder' (*khyim* 'home' and *bdag* 'lord')

- When an adjective precedes its head this is also treated as a compound.
  
  *dug btsan-po* 'mighty poison'
  
  *btsan-dug* 'mighty poison'
Apposition

• Only case two nouns allowed to sit next to eachother.
  
  *bu khyeḥu ‘a child, a son’*

• Proper nouns require this treatment of apposition.

  *rgyal-po Gsal-rgyal gyi ‘King Prasenajit’*

  *btsun-mo chen-po Ḫbar-li ‘Ḥbar-li, the main queen’*
Part 2

Part of Speech Categories
POS Categories

1. Adjectives [adj]
2. Adverbs [adv.xxx]
3. Case markers [case.xxx]
4. Clitics [cl.xxx]
5. Converbs [cv.xxx]
6. Demonstratives, determiners, etc. [d.xxx]
7. Nouns [n.xxx]
8. Negation [neg]
9. Numbers [num.xxx]
10. Pronouns [p.xxx]
11. Verbs [v.xxx] and verbal nouns [n.v.xxx]
Section overview

1. The noun phrase
2. The verb phrase
3. Others
The Noun Phrase

1. Nouns [n.xxx]
2. Pronouns [p.xxx]
3. Numbers [num.xxx]
4. Adjectives [adj]
5. Determiners and demonstratives [d.xxx]
6. Case markers [case.xxx]
The Noun Phrase: Nouns [n.xxx]

1. Count nouns [n.count]
2. Mass nouns [n.mass]
3. Proper nouns [n.prop]
4. Relator nouns [n.rel]
## Count Nouns [n.count]

- Word can head a noun phrase.
- Dictionary meaning (i.e. gut feeling)
- (when possible) nominal suffixes -mo, -po, -bu etc.
- Only in special circumstances can a noun contain case and conversbial morphemes.

  - gaṅ-na-ba 'whereabouts'
  - bdag-gi-ba 'that which is mine'
  - kun-tu-rgyu 'parivrājaka'
  - rten-ciṅ-ḥbrel-bar-ḥbyurī-ba 'pratītyasamutpāda', etc.
1. When two nouns not in apposition would follow each other

   *nor-bu sbar gaṅ* 'a handful of jewels’
   *chu sñím-pa gaṅ* 'a handful of water’
   *nas khal ŋer gcig* ‘twenty one loads of barley

2. Knowing that there exists this syntactic difference, we tag all plausible mass nouns on the basis of their meaning (e.g. *zaṅs* 'copper').

3. A final list of mass nouns can only be securely put forward after the syntactic behavior of these words is better investigated.
Proper Nouns [n.prop]

1. Allowed to be grammatically complex
2. Otherwise not much to say about them
Relator Nouns [n.rel]

- one syllable, a genitive before it and a spatial case (allative, locative, terminative) after it
  
  *dehi nañ na* 'inside of that'
  
  *dehi druñ du* 'before him'
  
  *dehi ḥog tu* 'under that'
  
  *dehi tshe na* 'at that time'

- relator nouns not quantified, not suffixed with adjectives, determiners or demonstratives.
Relator Nouns [n.rel]

- tagged as relator nouns also when missing genitive to left or the spatial case to the right.

\[
de\text{hi } ts\text{he } \text{blon-po } \text{\daggerig phyi-rol nas na}\text{ñ } \text{du } \text{\daggerho\text{-pa las/ mi } btson } \text{du } \text{bzui\text{-ba } mtho\text{-ba } da}\text{ñ / 'Then the minister went inside from outside and saw the man who had been taken to prison'}
\]

\[
ts\text{he is not followed by a spatial case}
\]

\[
na\text{ñ is not preceded by a genitive}
\]
The Noun Phrase: Pronouns

1. Indefinite pronouns [p.indef]
2. Interrogative pronouns [p.interrog]
3. Personal pronouns [p.pers]
4. Reflexive pronouns [p.refl]
Indefinite Pronouns [p.indef]

- Expressions of indefinite quantity when used as pronouns.
  - *kha* as in *kha cig* ‘some’
Interrogative Pronouns [p.interrog]
Personal Pronouns [p.pers]

No distinction is drawn in person or number.
Reflexive Pronouns [p.refl]
Numbers [num.xxx]

- Cardinal numbers [num.card]
- Ordinal numbers [num.ord]
The Noun Phrase: Adjectives [adj]

- Type of nominal
  - *chen-po* not *che*
- Not be analyzable as a verbal noun
  - not *che-ba*
The Noun Phrase: Demonstratives, Determiners, etc.

1. Demonstratives [d.dem]
2. Determinates [d.det]
3. Emphatics [d.emph]
4. Indefinites [d.indef]
5. Plurals [d.plural]
Demonstratives [d.dem]
Determiners [d.det]
Indefinites [d.indef]

Used only for -cig and its allomorphs
<table>
<thead>
<tr>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ཕུགས་</td>
</tr>
</tbody>
</table>
The Noun Phrase: Case markers [case.xxx]

<table>
<thead>
<tr>
<th>Form</th>
<th>Name</th>
<th>POS tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>-kyi, -gyi, -gi, -yi, -ḥi</td>
<td>genitive</td>
<td>[case.gen]</td>
</tr>
<tr>
<td>-kyis, -gyis, -gis, -yis, -s</td>
<td>agentive</td>
<td>[case.agen]</td>
</tr>
<tr>
<td>-la</td>
<td>allative</td>
<td>[case.all]</td>
</tr>
<tr>
<td>-na</td>
<td>locative</td>
<td>[case.loc]</td>
</tr>
<tr>
<td>-du, -tu, -su, -ru, -r</td>
<td>terminative</td>
<td>[case.term]</td>
</tr>
<tr>
<td>-las</td>
<td>ablative</td>
<td>[case.abl]</td>
</tr>
<tr>
<td>-nas</td>
<td>elative</td>
<td>[case.ela]</td>
</tr>
<tr>
<td>-dañ</td>
<td>associative</td>
<td>[case.ass]</td>
</tr>
<tr>
<td>-bas</td>
<td>comparative</td>
<td>[case.comp]</td>
</tr>
<tr>
<td>-na-re</td>
<td>quotative</td>
<td>[case.nare]</td>
</tr>
</tbody>
</table>
The Verb Phrase

1. Finite verbs [v.xxx]
2. Verbal nouns [n.v.xxx]
3. Negation [neg]
4. Converbs [cv.xxx]
The Verb Phrase: Finite verbs

Apparent from the four stems we distinguish

1. Auxilliaries [v.aux]
2. Copulas [v.cop]
3. Negative verbs [v.neg]
4. Negattive copulas [v.cop.neg]
Auxiliary verbs

- A verb that occurs immediately after a verb stem
- Normally has a modal meaning
- Does not distinguish four stems
- We do not specifically tag the evidential auxiliaries of Middle Tibetan and Modern Tibetan.
Copulas and negative verbs

- Copulas are verbs that identify two absolutive noun phrases
- The only negative verb is *med*
- The only negative copula is *min*
The Verb Phrase: Verbal nouns

Each finite verb has a corresponding verbal noun. So far we recognize the following suffixes.

- -pa/-ba
- -rgyu
- -thabs
- -lugs
- -grags
- -tshul
- -tshad
- -mkhan
- -sa
The Verb Phrase: Negation

- With some reservation with put the preverbal morphemes *e* and *ye* here as well as the expected *ma* and *mi*.
- In particular this is a problem for *ye* because it often precedes negation.
  - *chos ye mi yon-ba* ‘no dharma at all has come’
# The Verb Phrase: Converbs [cv.xxx]

<table>
<thead>
<tr>
<th>Form</th>
<th>Name</th>
<th>POS tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>-kyi, -gyi, -gi, -yi, -ḥi</td>
<td>genitive</td>
<td>[cv.gen]</td>
</tr>
<tr>
<td>-kyis, -gyis, -gis, -yis, -s</td>
<td>agentive</td>
<td>[cv.agen]</td>
</tr>
<tr>
<td>-la</td>
<td>allative</td>
<td>[cv.all]</td>
</tr>
<tr>
<td>-na</td>
<td>locative</td>
<td>[cv.loc]</td>
</tr>
<tr>
<td>-du, -tu, -su, -ru, -r</td>
<td>terminative</td>
<td>[cv.term]</td>
</tr>
<tr>
<td>-las</td>
<td>ablative</td>
<td>[cv.abl]</td>
</tr>
<tr>
<td>-nas</td>
<td>elative</td>
<td>[cv.ela]</td>
</tr>
<tr>
<td>-dañ</td>
<td>associative</td>
<td>[cv.ass]</td>
</tr>
<tr>
<td>-te, -ste, -de</td>
<td>semi-final</td>
<td>[cv.sem]</td>
</tr>
<tr>
<td>-ḥo, -to, etc.</td>
<td>final</td>
<td>[cv.fin]</td>
</tr>
<tr>
<td>-ciñ, -śiñ, žiñ</td>
<td>imperfective</td>
<td>[cv.impf]</td>
</tr>
<tr>
<td>-cig, -śig, žig</td>
<td>imperative</td>
<td>[cv.imp]</td>
</tr>
<tr>
<td>-gin, -gyin, -kyin,</td>
<td>connecting</td>
<td>[cv.cont]</td>
</tr>
</tbody>
</table>
Other POS tags

1. Adverbs
2. Clitics
3. Dunno
4. Skt
Others: Adverbs

A sort of catch all for things we are not sure how to deal with.

1. Directional adverbs [adv.dir]
2. Intensive adverbs [adv.intense]
3. Mimetic adverbs [adv.mim]
4. Proclausal adverbs [adv.proclausal]
5. Temporal adverbs [adv.temp]
Directional adverbs [adv.dir]

• A category invented for adverbs that end in –cad, to which we have added a few similar words aong the way.
Intensive adverbs [adv.intense]

- Used for words that mean ‘very’
- We take śin and rab as probably etymologically verbs (cf. man du ‘muchly’), so understand the -tu as [cv.term]
- We put cuñ zad here also, when it functions adverbially
Mimetic adverbs [adv.mim]

- A category for ideophones, onomatopoeia, and the like.
Proclausal adverbs [adv.proclausal]

- Things that begin clauses and generally refer back adverbially to a previous clause.
- The second element is treated as it normally would be.

\begin{itemize}
  \item \textit{de} [adv.proclausal] \textit{nas} [case.ela] ‘then’
  \item \textit{yañ na} [case.loc] ‘alternatively’
  \item \textit{gal te} [cv.sem] ‘if’
  \item \textit{ḥo na} [case.loc] ‘in that case’
  \item \textit{de-ba s} [case.agn] ‘therefore’
  \item \textit{ḥon kyañ} [cl.focus] ‘nonetheless’
  \item \textit{khyad-pa} r [case.term] ‘in particular’
  \item \textit{la} r ‘moreover’
\end{itemize}
Temporal adverbs [adv.temp]

- Used for time expressions ‘now’, ‘today’, ‘previously’, etc.
- Often these words appear in nominal contexts, *saṅ gi gdugs la* ‘yesterday midday’
- But they can also occur as adverbial adjuncts without casemarking.
Others: Clitics

We use this tag for things that occur in places where nothing else appears to.

1. Focus clitics [cl.focus]
2. The clitic *lta* [cl.lta]
3. The quotatie clitic [cl.quot]
Focus clitics [cl.focus]

- A category for \textit{ni} and \textit{kyanî} \~ \textit{yanî}
- These words occur after anything
- They occur even after case marking

(i.e. those who see \textit{ni} as some kind of case marker are not paying enough attention)
Clitic *lta* [cl.lta]

- For some strange uses of *lta*

  ... *gyur na lta*

  ... *sdug-bsñal rnams myoñ-bar ḥoñ-ba lta /

  žiñ-khañ gi phyogs la ḥgro-ba dag e cani che ste

  (For *dag* in Mila other recensions have *lta*)
The quotatie clitic [cl.quot]

- For *ces* and *ces-pa* and its allomorphs.
- We now think it would be better to treat these as part of a new class of *verba dicendi* that take clauses as complements (cf. *zer*, *sñam*, etc.) but it is too late.

  (inclusion of *sñam* at left is an error)