This talk aims to:
- Provide some historical perspective on documentation activities and outputs, with a focus on Nepal
- Matisoff (1991: 498): “It is high time to “mainstream” S(ino-)T(ibetan) linguistics”
- It’s also time to mainstream documentation methods & outputs within the context of 21st century digital & informatics scholarship
- A case study of how this is being done with four language communities of Manang District, Nepal

Language Documentation in Nepal in a traditional perspective
- B.H. Hodgson and G. Grierson in late 19th/early 20th centuries
- Linguistic surveys of Nepal intensified ca. 1980’s: Werner Winter, now LinSuN at Tribhuvan University (Regmi 2010)
- Ongoing documentation initiatives by Summer Institute of Linguistics
- Energies skewed particularly to eastern Nepal, but this is beginning to change
- In Manang and surrounds (‘Tamangic’): Georg, Glover, Hildebrandt, Honda, Mazaudon, Noonan

Documentation outputs in Nepal:
- Growing number of grammars published in mainstream venues: Brill, Mouton, sketches through Routledge edited volumes, Lincom Europa
- Even greater amount of information as unpublished mimeos, handouts or else as limited-release publications
- Many outputs were concerned largely with issues of genealogical affiliation and shared lexico-grammatical correspondences
- So, content more focused on paradigmatic patterns, comparative glossary-building and contrastive (-emic) analysis

Newer initiatives have brought methods and outputs within this particular field into the 21st century:
- Archives: Digital Himalaya (University of Virginia, University of Cambridge), Tibetan Himalayan Library (U of Virginia), LACITO
- Documentation blogs and web pages: CPDP, Nar-Phu, etc.
- A-V companions to grammars: van Driem and Tshering’s 1998 Dzhongkha practical grammar
- The online journal Himalayan Linguistics now has a “field reports” component
- But there is still room for more work and development
Field reports can legitimize certain types of (descriptive) studies or specific phenomena, allowing for replication or incorporation into theoretical platforms as necessary.

- The linguistic diversity across the geographically changeable and compact/bounded inhabitable regions of Nepal, combined with within-family and across-family contact, along with varying degrees of threat/maintenance to these languages should all shape the methods of documentation too.
- Historically, this would be a tall order for any purely paper-bound output.
- But existing grammars already hint at the possibilities of what a multi-variable approach to documentation on any given language/in any area might reveal (handout, appendix).
**Ongoing Need**

- This kind of information is essential; not only does it contribute to/challenge theories of natural human language...
- E.g. phonemic vs. sub-phonemic, conditioned vs. free variation, structure-preserving vs. structure altering; lexical vs. post-lexical; lexically general vs. specified (Kiparsky 1982; Mohanan 1986; Blevins 2004; Nespor and Vogel 2007)
- An added bonus is that the variation frequently appears to have spatial and sociolinguistic motivations
- These observations open up possibilities for revisiting and expanding methods and outputs of language documentation & description, enriching analysis by factoring in other variables

---

**Manang Languages**

**Case 1: Lexical Comparison**

- My first field trip to Nepal in 1998 involved gathering lexeme information following the model of STEDT

---

**STEDT**

- Another good example of digital innovation in comparative method outputs

---

**Manang Languages**

**Case 1: Lexical Comparison**

- Facilitates—problematizes etyma reconstruction; reveals conservative/innovative variants; provides leads to contact effects & dialect variation

---

**Manang Languages**

**Case 1: Lexical Comparison**

- Nâr & Phu spoken to northeastern
- Nêshangte-Manang spoken in northern & central VDCs
- Manang-Gurung in southern & central VDCs
- Gyalumdo around Tal & Chame

---

**Manang Languages**

- Another good example of digital innovation in comparative method outputs
• Case 1: Lexical Comparison

Which then feeds into other comparative initiatives on a larger typological scale...

Currently the only systematic comparative analysis I have is based on (Indic) loanwords in the context of relative endangerment (Hildebrandt 2004, 2009; Wilbur 2005)

<table>
<thead>
<tr>
<th>Lg.</th>
<th>Loans or Loan Free Variation w/Indic</th>
<th>PTB or PST etyma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ghashok Gurung</td>
<td>149 (29%)</td>
<td>360 (71%)</td>
</tr>
<tr>
<td>(n = 509)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manang Gurung</td>
<td>30 (12%)</td>
<td>220 (88%)</td>
</tr>
<tr>
<td>(n = 250)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manange</td>
<td>35 (9%)</td>
<td>377 (91%)</td>
</tr>
<tr>
<td>(n = 412)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nar-Phu</td>
<td>15 (4%)</td>
<td>345 (96%)</td>
</tr>
<tr>
<td>(n = 360)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Tone Systems in 19 S-T languages](image)

Manang Languages

- Case 2: Sino-Tibetan tonogenesis:
  - Relatively “recent” diachronic phenomenon, and many S-T languages still incipiently tonal in terms of phonetic correlates, domains of contrast and perceptual functional load:

<table>
<thead>
<tr>
<th>No tone</th>
<th>The languages are a-tonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amalgam</td>
<td>For at least half of the tones, the phonetic correlates include F0 and phonation differences. The number of contrastive tones is often four or fewer</td>
</tr>
<tr>
<td>Pure</td>
<td>Tone is (almost) entirely a function of F0. It applies to only one tone. The number of tones is often greater than four</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dolakha Newar</th>
<th>Gare/Nepal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klimauri, Limbu, Qiang</td>
<td></td>
</tr>
<tr>
<td>Burmese, Dege Tibetan, Kham, Kyirong, Tibetan, Manange</td>
<td></td>
</tr>
</tbody>
</table>

![Tamangic Tonogenesis](image)

Tamangic Tonogenesis

PROTO

<table>
<thead>
<tr>
<th>*HI</th>
<th>*p</th>
<th>*p'</th>
<th>*b</th>
<th>*m</th>
<th>MODERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>*p</td>
<td>p</td>
<td>p'</td>
<td>m</td>
<td>m</td>
<td>/1/ modal</td>
</tr>
<tr>
<td>*p'</td>
<td>p</td>
<td>p'</td>
<td>m</td>
<td>m</td>
<td>/2/ modal</td>
</tr>
<tr>
<td>*LOW</td>
<td>--</td>
<td>--</td>
<td>p</td>
<td>m</td>
<td>/3/ non-modal</td>
</tr>
<tr>
<td>*p'</td>
<td>p</td>
<td>p'</td>
<td>m</td>
<td>m</td>
<td>/4/ non-modal</td>
</tr>
</tbody>
</table>

![Map of Tibet/China](image)

Issues of Representation

- The different diachronic paths and currently varied systems of these languages have some significant consequences for representation of tone
  - Featural approach: 2 tones + initial C [VOICE] feature (Kjellin 1975); but: in some lgs., voicing differences part of the consonant, part of the vowel, or else both (in particular, Tamang)
  - Separate tone & phonation: /1, 2/ tone, /3, 4/ phonation/register (Maddieson 1984); but: across lgs., /3, 4/ different trajectories

![Map of Tibet/China](image)

Resulting Systems

<table>
<thead>
<tr>
<th>Tamang</th>
<th>Gurung</th>
<th>Thakali</th>
<th>Manange</th>
</tr>
</thead>
<tbody>
<tr>
<td>*HI</td>
<td>/1/</td>
<td>54 ± asp</td>
<td>33 ± asp</td>
</tr>
<tr>
<td></td>
<td>/2/</td>
<td>55 ± asp</td>
<td>54 ± asp</td>
</tr>
<tr>
<td>*LOW</td>
<td>/3/</td>
<td>33/22 ft. ± asp</td>
<td>11 fl. ± asp</td>
</tr>
<tr>
<td></td>
<td>/4/</td>
<td>211 fl. ± asp, [b]?</td>
<td>12 fl. ± asp, [b]?</td>
</tr>
</tbody>
</table>

(β = ‘breathy/murmur phonation; [b] = possible phonetic voicing effect of onset; Chao numbering system where 5 = high, 1 = low)

- However, Mazaudon & Michaud (2006, 2008), Hildebrandt (2007), Mazaudon (2005): high degrees of idiolectal & dialectal variation, phonetic correlates differently weighted across languages, varied role of F0 (pitch) in defining the systems
Mazaudon & Michaud (2008) suggest a ‘panchronic’ approach:

- It’s possible that Gurung, Tamang, Thakali → Manange-type system
- If so, we are observing tonogenesis still in-action, with gradual de-linking (& possibly re-linking) of non-F0 correlates
- In particular, re-linking may occur via contact with Indic languages (non-tonal, true register-based systems, dominant, lingua-franca presence in Nepal)
- And we are likely to witness a great deal of inter-speaker & regional variation

Issues in Morpho-Syntax

- Case 3: “Optional” Case-Marking
- “Case” is traditionally thought of as a structural/grammatical correlate to verb-argument relations, or else lexically specified by a case governor, or else semantically constrained (e.g. with spatial cases)

<table>
<thead>
<tr>
<th>Issues of Representation</th>
<th>The Problem of Acoustic Correlates</th>
</tr>
</thead>
<tbody>
<tr>
<td>• My past research on phonetic correlates has revealed more questions than answers</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F0</th>
<th>Initial C VOT</th>
<th>Medial C VOT</th>
<th>Spectral Tilt (modal v. non-modal)</th>
<th>V Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 tones (rural)</td>
<td>4/3 (urban)</td>
<td>Male: /3/ vs. others</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Nar(-Phu) (9 spkrs, 4 communities)</td>
<td>/4/ +asp allowable for all tones</td>
<td>/3/ v. /4/</td>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td>Gyalsamdo (2 speakers): 3-way (2 high, 1 low)</td>
<td></td>
<td>Insignificant data</td>
<td>n.s.</td>
<td></td>
</tr>
</tbody>
</table>

Issues in Morpho-Syntax

- Optional Ergative Case Marking (OEM)
- Well documented within T-B (Chelliah & Hyslop eds. 2011)
- Examples from Manang-Gurung

<table>
<thead>
<tr>
<th>Ergative case marking in Nepali</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ma*-ma pa-aɗari ke tā</td>
</tr>
<tr>
<td>2. ma*-ma pa-aɗari ke-dādi-yu</td>
</tr>
<tr>
<td>3. s'unga-horo ke s'unga-horo dyj</td>
</tr>
<tr>
<td>4. s'unga-horo ke s'unga-horo dyj</td>
</tr>
</tbody>
</table>

Issues in Morpho-Syntax

- A kind of differential subject marking
- The behavior is linked to semantic & information-structural factors
- The task is to catalogue the relevant conditions that correlate with OEM (& other optional case expressions) in these languages
- Sociolinguistic factors, including the growing role and influence of Nepali, are probably at play in the observed variation across the different linguistic levels
- Add to this the recent, significant development & settlement-impacting activities going on in Manang (and Lamjung, where other Gurung varieties are spoken)
Road-Building:
Blasting a road out of cliff walls, near Tal village, lower Manang (summer 2010)

What results from this does not always resemble an actual road! (Lamjung, 2009)

Manang Languages

- Am I seeking to do the impossible?
  - I want to obtain a comprehensive, representative data-set of lexico-grammatical, phono-semantic & discourse strategies found in these four (+) languages (along with data on dialect variation)
  - I want to better understand how development initiatives interrupt or preserve lg. attitudes, usage & transmission
  - Actually, this type of study is attractive to a cross-linguistic, geo-spatial perspective
  - GIS (Geographic Information Systems) representation of Manang

GIS

- GIS is a system for storing and displaying geo-spatial information on the web or in other digital formats
- It integrates software, hardware & programming to answer questions involving geographically referenced data

GIS and Documentation

- GIS-documentation link-ups are increasingly employed, with some interesting and compelling exemplars
  - DELAMAN network (spatial representation of metadata from endangered language archives)
- Berkeley Linguistics Mapping Project (BeLMaP): Studies the role of space in the spread of linguistic features via diffusion/borrowing in areas of intense contact (Michael 2010)

GIS & Documentation

- Anju Saxena and collaborators: “Digital Areal Linguistics: A Lexical View of the Himalayan Micro-Area” (Swedish Research Council)
**A GIS Perspective of Manang Languages**

- NSF CAREER: “Documenting the Languages of Manang, Nepal for Local and International Impact”
- Visualize patterns and usage scenarios in micro-level spatial perspectives

**Mapping Languages of Manang**

- Map designed by Shunfu Hu (SIUE), with G.A. assistance from Prita Malla & Kanchan Karki
- http://www.siue.edu/~shu/nepal7.html

**Mapping Languages of Manang**

- The geo-points may themselves become “hotlinks” to downloadable data, or transcribed A-V files in formats amenable to acoustic research

**Acoustic Findings**

- 80-90 words elicited from speakers in each village throughout Manang
- Organized by a range of phonetic, phonological & lexical factors

- If village is represented by more than one language, we attempt representation from each lg.
- So far: data from 10 M-G; 6 Gyalsumdo; 1 Manange; 2 Nar (more Manange & Nar data to be gathered in 2014)

- Words were recorded in isolation (three repetitions) & frame-medial or final context (three repetitions)
- Gurung kwe ‘bee’ & la-pa ‘drive.away-NOM’
  - For nouns: toso ɲə-e kwe mро-e-po [now 1SG-ERG bee see-ASP-NOM] ‘Now I see a bee.’
- Gyalsumdo ʈo ‘stone’ & ʈo ‘walk/go’
  - For nouns: ʈa to tʰoŋ-so [1SG stone see-TAMEVID] ‘I saw the stone.’
  - For verbs: ʈa tənta to-ke (re) [1SG now walk/go-TAMEVID (EVID)] ‘I am walking now/I walk now.’

**Acoustic Findings**

- What can we look to as modern reflexes, or as features to the tonogenetic developments in these languages?
- Just what kind of variation is possible amongst any generalizations?
- Pitch-melody (within/across the registers)
- Behavior of initial obstruents (VOT)
- Voicing of vowels (and consonants) with respect to Electroglottographic measurements
Acoustic Findings

• Pitch-melody: Four Manang-Gurung speakers (5 measurement points)

Observations & Analysis

• Pitch-melody: Two Nar-Phu speakers (3 measurement points)

• Pitch-melody: Four Gyalsumdo speakers (5 measurement points)

Observations & Analysis

• Another possible cue: Voice Onset Time (VOT) differences on initials in different registers (or tone groupings within the H-L division) may emerge as a reflex of older (obstruent) voicing contrasts (Manange, 2 speakers)

Observations & Analysis

• VOT differences on initials, Nar (2 speakers)
Observations & Analysis

- In word-medial position, passive plosive voicing assimilation (V_V) is common, but not regular, and does not align with tone.

- VOT differences on initials, Gyalsumdo (4 speakers)

- The male speakers show a strong tendency towards pre-voicing or else breathy onsets with words in WT 3 & WT 4 (those cases are not reflected in these bar-graphs)

- EGG CQ for Manang-Gurung:

- When either the Kaski or the Manange tone models are considered, a weak correlation between /4/ vowels & lower CQ, but it is not consistent across speakers

- Observations & Analysis

- Electroglottographic cues: Mazaudon and Michaud (2008) observed for Tamang that the open-quotient (Oq) values were significantly higher, with a dipping then rising pattern through time for the LOW tones vs. HI.

- I.e. for the LOW tones, they observed an overall rise in airflow rate in the nucleus

- My study: all words recorded through a EG2-PCX2 two-channel EGG assembly (see slide #1)

- I measured Closed Quotient (CQ): difference between time of v.f. closure in relation to total time of voicing cycle; Non-modal predicted to carry lower CQ values than modal, as v.f. closure is shorter in time & opening portion lags for longer

- I took only one measurement point

- EGG CQ for Gyalsumdo:

- Three of four speakers: initial vowels of WT 3 and/or 4 show lowered CQ values, but this is not always significant

- Summary

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Manange</th>
<th>Manang Gurung</th>
<th>Nar(-Phu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch-melody</td>
<td>High-low &amp; level-falling/contour</td>
<td>High-Low emerges only when compared to Manange model</td>
<td>High-Low</td>
</tr>
<tr>
<td>Onset voicing</td>
<td>No voicing; phonetic aspiration rare</td>
<td>No voicing; aspiration split in tones /3/ &amp; /4/</td>
<td>Lower VOT in low register</td>
</tr>
<tr>
<td>Other cues</td>
<td>Possibly jitter</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tamangic Languages of Manang
### Summary

#### Two Registers

<table>
<thead>
<tr>
<th>Characterized by</th>
<th>“High” (WT 1/2)</th>
<th>“Low” (WT 3/4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher F0 (no evidence for contour diffs. yet), ±asp obs.</td>
<td>Lower F0, tendency towards obstructive voicing, particularly by males, weak evidence for shorter vocal fold closure for vowels, obstructive aspiration rare</td>
<td></td>
</tr>
</tbody>
</table>

### OEM Findings

<table>
<thead>
<tr>
<th>Variables associated with OEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal/lexical properties</td>
</tr>
<tr>
<td>predicate valence</td>
</tr>
<tr>
<td>clause polarity</td>
</tr>
<tr>
<td>aspect</td>
</tr>
<tr>
<td>objectivity</td>
</tr>
<tr>
<td>focus</td>
</tr>
<tr>
<td>definiteness</td>
</tr>
<tr>
<td>specificity</td>
</tr>
<tr>
<td>referentiality</td>
</tr>
<tr>
<td>agent control</td>
</tr>
</tbody>
</table>

### OEM Findings

- Note that non-marking of overt A arguments is much more common than marking, and that arguments themselves are also frequently optional

<table>
<thead>
<tr>
<th>Verbs with overt A/ε NPs</th>
<th>4/5 NPs</th>
<th></th>
<th>5/6</th>
<th>37</th>
<th>13</th>
<th>100%</th>
<th>100%</th>
<th>100%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>With ERG S</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.33%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>With ERG S</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8.33%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Transitive</td>
<td>13</td>
<td>2.77%</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>18.57%</td>
<td>2</td>
<td>7.41%</td>
<td>0%</td>
</tr>
<tr>
<td>With overt A</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8.33%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>With overt A</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.33%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>WHIRG A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Unclear</td>
<td>3</td>
<td>2.12%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.33%</td>
<td>3</td>
<td>2.12%</td>
<td>0%</td>
</tr>
</tbody>
</table>

### Future Plans

- These are languages with low referential density (cf. Bickel 2003)
- And ergative-marking appears to be as much information-structural related as strictly determined by grammatical function of the argument
- These are still early days, and we plan to align our tonal & morphosyntactic observations with sociolinguistic usage & attitude surveys gathered alongside the phonetic & discourse data

### Final Considerations

- A spatial perspective is not a substitute for intensive, comprehensive documentation of systems as they are used in everyday settings, across genres; the methods of investigation must remain rigorous
- There is also the non-trivial matter of community permission, input and collaboration in an endeavor resulting in linguistic mapping at a micro-level (cf. Penfield et al 2008, Rice 2011)
- Such initiatives also rely on intense cross-disciplinary (and cross-institutional) collaboration with experts on hardware, software and programming, on larger budgets, and on longer timelines (e.g. NSF CAREER, ELD Large Grants, U.S./U.K./EU cross-council collaborations, etc.)
Final Considerations

• Following guidelines advocated by ELAR, by DoBeS and by Bird and Simons (2003), all of this collaboration and expertise must all ultimately be open-source (to the extent possible), transportable, cross-platform (non-proprietary), available to/learnable by a wide range of users, must find a long-term home for storage/access/archive, and must use mark-up languages available for long-term access.

• However, spatial representations of structure and usage in such multi-lingual, heavy-contact, endangerment-prone areas provide an additional, more intuitive visual perspective of ‘what’s going on’.

• Such representations are particularly illuminating in areas where multiple features are considered simultaneously, or where structural variables are paired with socio-cultural/attitude/usage-scenario ones.

• They also open up linguistic documentation and analysis to wider audience numbers and types (van Uytvanck et al.’s ‘curiosity factor’).