

MMT Extravaganza

Ling 567

March 7, 2024

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

| | | | |
|-----|--------------|---------------|--------------------|
| amh | Amharic | Afro-Asiatic | Xu, Zhou |
| kal | Kalaallisut | Eskimo-Aleut | Hsu, Ngae |
| nxq | Naxi | Sino-Tibetan | Chen, Deng |
| ono | Onondaga | Iroquoian | Brooks, Schifferer |
| pnk | Paunaka | Arawakan | Silva, Slusarczyk |
| yua | Yucatec Maya | Mayan | Morrison, Stevens |
| sje | Pite Saami | Uralic | Nielsen, Spivey |
| eng | English | Indo-European | |

Languages - mapped



lat/long from glottolog, map from Google

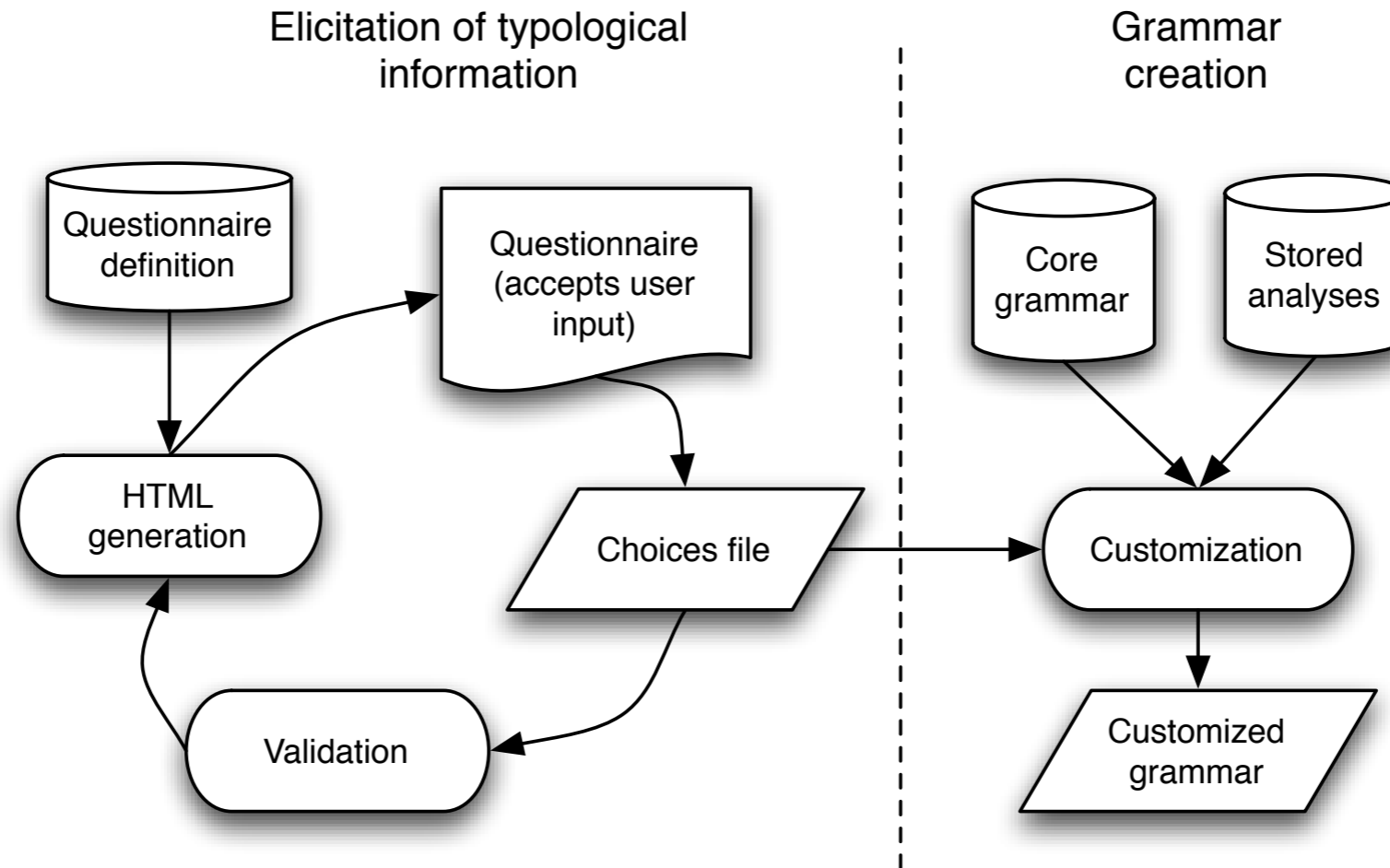
Languages - cookied (Thank you Liz!)



Grammar coverage (mostly shared)

- Basic word order
- Case
- Agreement
- Personal pronouns
- Tense/aspect
- Sentential negation
- (Argument optionality)
- Matrix yes-no questions
- Coordination
- Modification (adjective, clausal mods)
- Clausal complements
- Wh questions
- Possessives

Grammar Development: Customization + Extension



(Bender et al 2010)

Set up

- Transfer-based MT: Grammars parse and generate, mapping surface strings to semantic representations in MRS
- Grammars developed on the basis of the Grammar Matrix, facilitating harmonized semantic representations
- Quasi-lexical interlingua (English lemmatas as PRED values)
- ‘semi’ (Semantic Interface) maps variable properties (PNG, TAM, COG-ST) from grammar internal space to interlingual space. Lossy mapping, provides defaults
- One ‘accommodation’ transfer grammar per language, instantiating shared transfer rules

Input sentences

- 1.Dogs sleep
- 2.Dogs chase cars
- 3.I chase you
- 4.Dogs eat
- 5.The dogs dont chase cars
- 6.I think that you know that dogs chase cars
- 7.I ask whether you know that dogs chase cars
- 8.Cats and dogs chase cars
- 9.Dogs chase cars and cats chase dogs
- 10.Cats chase dogs and sleep
- 11.Do cats chase dogs
- 12.Hungry dogs eat
- 13.Dogs in the park eat
- 14.Dogs eat in the park
- 15.The dogs are hungry
- 16.The dogs are in the park
- 17.The dogs are the cats
- 18.The dog s car sleeps
- 19.My dogs sleep
- 20.Who sleeps
- 21.What do the dogs chase
- 22.What do you think the dogs chase
- 23.Who asked what the dogs chase
- 24.I asked what the dogs chased
- 25.The dog sleeps because the cat sleeps
- 26.The dog sleeps after the cat sleeps

Parse success

| | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|
| amh | eng | kal | nxq | ono | pnk | sje | yua |
| 14 | 26 | 21 | 15 | 11 | 26 | 18 | 12 |

Items with end-to-end output
(grammar, semi.vpm & transfer rules as provided)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| amh | 14 | 12 | 10 | 9 | 6 | 11 | 10 | 12 |
| eng | 16 | 26 | 19 | 12 | 12 | 18 | 21 | 15 |
| kal | 2 | 12 | 19 | 13 | 5 | 2 | 12 | 12 |
| nxq | 10 | 9 | 9 | 12 | 6 | 11 | 8 | 10 |
| ono | 6 | 6 | 7 | 7 | 9 | 9 | 5 | 0 |
| pnk | 13 | 12 | 15 | 13 | 11 | 26 | 9 | 16 |
| sje | 12 | 18 | 12 | 10 | 10 | 12 | 18 | 11 |
| yua | 7 | 7 | 10 | 8 | 5 | 10 | 7 | 12 |

Items with end-to-end output
(addressed aspect in yua)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| amh | 14 | 12 | 10 | 9 | 6 | 11 | 10 | 12 |
| eng | 16 | 26 | 19 | 12 | 12 | 18 | 21 | 15 |
| kal | 2 | 12 | 19 | 13 | 5 | 2 | 12 | 12 |
| nxq | 10 | 9 | 9 | 12 | 6 | 11 | 8 | 10 |
| ono | 6 | 6 | 7 | 7 | 9 | 9 | 5 | 7 |
| pnk | 13 | 12 | 15 | 13 | 11 | 26 | 9 | 16 |
| sje | 12 | 18 | 12 | 10 | 10 | 12 | 18 | 11 |
| yua | 7 | 7 | 10 | 8 | 5 | 10 | 7 | 12 |

Items with end-to-end output (final run)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| amh | 14 | 14 | 11 | 9 | 6 | 11 | 12 | 13 |
| eng | 16 | 26 | 19 | 12 | 12 | 17 | 21 | 15 |
| kal | 2 | 17 | 19 | 14 | 5 | 2 | 15 | 14 |
| nxq | 10 | 10 | 10 | 12 | 6 | 12 | 9 | 11 |
| ono | 6 | 8 | 8 | 8 | 9 | 10 | 7 | 8 |
| pnk | 15 | 14 | 15 | 14 | 11 | 23 | 12 | 17 |
| sje | 12 | 18 | 12 | 10 | 10 | 12 | 18 | 11 |
| yua | 7 | 10 | 11 | 9 | 5 | 10 | 9 | 12 |

Need for transfer rules

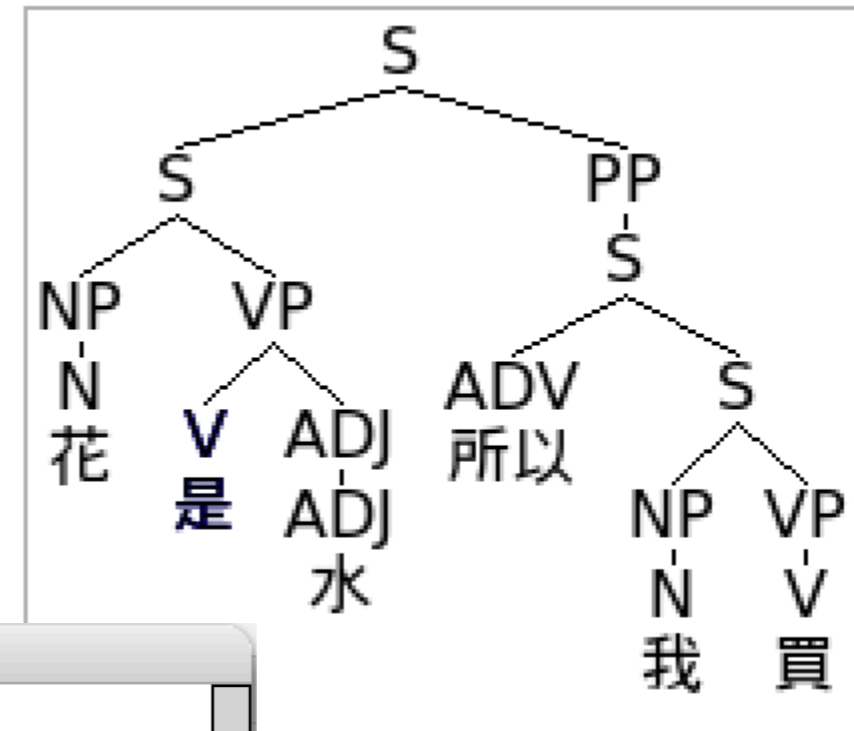
花 是 水 所以 我 買

hue1 si4 sui3 soe3i3 gua3 bue3

flower be pretty so 1SG buy

The flower is pretty, so I bought it

nan / invented example



```
Indexed MRS Display
h1 e2{ prop-or-ques tense no-aspect mood }
{ h3:exist_q_rel(x5{ bool cog-st }, h6, h4)
  h7:_flower_n_rel(x5)
  h8:_pretty_a_rel(e2, x5)
  h9:_so_subord_rel(e10{ iforce tense no-aspect mood }, h11, h12)
  h13:exist_q_rel(x15{ bool cog-st first sg }, h16, h14)
  h17:pron_rel(x15)
  h18:_buy_v_rel(e20{ prop-or-ques tense no-aspect mood }, x15, x19{ bool cog-st }) }
{ h6 =q h7 h11 =q h8 h12 =q h18 h16 =q h17 }
{ e2 !f x21 }
```

Transfer rule example

```
so-because-mtr := monotonic_mtr &
[ OUTPUT [ RELS < [ PRED "_because_subord_rel",
                    LBL #lbl,
                    ARG0 #arg0,
                    ARG1 #arg1,
                    ARG2 #arg2 ] > ],
  INPUT [ RELS < [ PRED "_so_subord_rel",
                  LBL #lbl,
                  ARG0 #arg0,
                  ARG1 #arg2,
                  ARG2 #arg1 ] > ] ].
```

Total number of outputs (starter grammar + yua aspect fix)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|--------|-----|-----|-----|-------|----------|-----|-------|
| amh | 7976 | 114 | 43 | 11 | 260 | 7800 | 62 | 6218 |
| eng | 2274 | 84 | 67 | 18 | 1004 | 39227 | 587 | 1620 |
| kal | 28 | 74 | 44 | 20 | 123 | 120 | 106 | 963 |
| nxq | 9147 | 267 | 118 | 18 | 268 | 11436 | 112 | 431 |
| ono | 550458 | 596 | 216 | 43 | 62023 | 455472 | 140 | 3361 |
| pnk | 783194 | 180 | 180 | 297 | 15038 | 10674196 | 130 | 37315 |
| sje | 12570 | 127 | 45 | 16 | 3332 | 42307 | 361 | 4697 |
| yua | 482 | 43 | 148 | 32 | 299 | 19332 | 72 | 4465 |

Total number of outputs (final run)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|---------|------|-----|--------|-------|----------|------|-------|
| amh | 7976 | 816 | 44 | 11 | 260 | 7632 | 806 | 6222 |
| eng | 1970 | 79 | 63 | 18 | 1004 | 39121 | 587 | 1620 |
| kal | 28 | 375 | 44 | 26 | 123 | 120 | 388 | 1539 |
| nxq | 9188 | 345 | 121 | 20 | 268 | 11579 | 134 | 447 |
| ono | 550454 | 1063 | 219 | 366 | 62599 | 496851 | 352 | 5161 |
| pnk | 3014704 | 986 | 180 | 107699 | 15614 | 18405818 | 6070 | 40875 |
| sje | 12570 | 127 | 45 | 16 | 3332 | 42307 | 361 | 4697 |
| yua | 898 | 1786 | 188 | 37 | 299 | 27541 | 3558 | 5009 |

Maximum number of outputs (final run)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|---------|------|-----|--------|-------|---------|------|-------|
| amh | 5840 | 648 | 16 | 2 | 96 | 3069 | 648 | 3256 |
| eng | 608 | 32 | 14 | 3 | 480 | 20885 | 216 | 1024 |
| kal | 20 | 128 | 8 | 4 | 96 | 72 | 216 | 512 |
| nxq | 8276 | 140 | 44 | 4 | 144 | 6830 | 24 | 128 |
| ono | 296841 | 576 | 72 | 320 | 61440 | 304037 | 246 | 4096 |
| pnk | 2908108 | 464 | 64 | 103917 | 7296 | 7467569 | 2808 | 21504 |
| sje | 9216 | 24 | 8 | 4 | 2688 | 19055 | 216 | 2304 |
| yua | 768 | 1728 | 64 | 12 | 192 | 13074 | 3456 | 4608 |

Items with exact match output (final run)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| amh | 14 | 14 | 1 | 7 | 4 | 9 | 10 | 9 |
| eng | 13 | 26 | 1 | 11 | 7 | 11 | 18 | 10 |
| kal | 1 | 2 | 1 | 11 | 3 | 2 | 9 | 9 |
| nxq | 7 | 9 | 0 | 11 | 3 | 9 | 7 | 7 |
| ono | 4 | 7 | 1 | 5 | 9 | 8 | 5 | 4 |
| pnk | 8 | 9 | 1 | 9 | 5 | 20 | 7 | 9 |
| sje | 9 | 17 | 1 | 8 | 5 | 8 | 18 | 7 |
| yua | 6 | 10 | 1 | 8 | 3 | 7 | 7 | 12 |

Items with end-to-end output (final run)

| | amh | eng | kal | nxq | ono | pnk | sje | yua |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| amh | 14 | 14 | 11 | 9 | 6 | 11 | 12 | 13 |
| eng | 16 | 26 | 19 | 12 | 12 | 17 | 21 | 15 |
| kal | 2 | 17 | 19 | 14 | 5 | 2 | 15 | 14 |
| nxq | 10 | 10 | 10 | 12 | 6 | 12 | 9 | 11 |
| ono | 6 | 8 | 8 | 8 | 9 | 10 | 7 | 8 |
| pnk | 15 | 14 | 15 | 14 | 11 | 23 | 12 | 17 |
| sje | 12 | 18 | 12 | 10 | 10 | 12 | 18 | 11 |
| yua | 7 | 10 | 11 | 9 | 5 | 10 | 9 | 12 |

Overview

- Background
- Overview results
- Interactive exploration
- Course evals