

MMT Extravaganza

Ling 567

March 12, 2026

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

hak	Hakka	Sino-Tibetan	Boonkongchuen, Gates
irk	Iraqw	Afro-Asiatic	Hollinger, Rosen
jje	Jejueo	Koreanic	Halatsis, Yeom
moh	Mohawk	Iroquoian	Afifi, Mamro
nep	Nepali	Indo-Aryan	Kurek, Song
tel	Telugu	Dravidian	Rao, Venugopal
sje	Pite Saami	Uralic	Nielsen, Spivey
eng	English	Indo-European	

Languages - mapped



lat/long from glottolog, map from Google

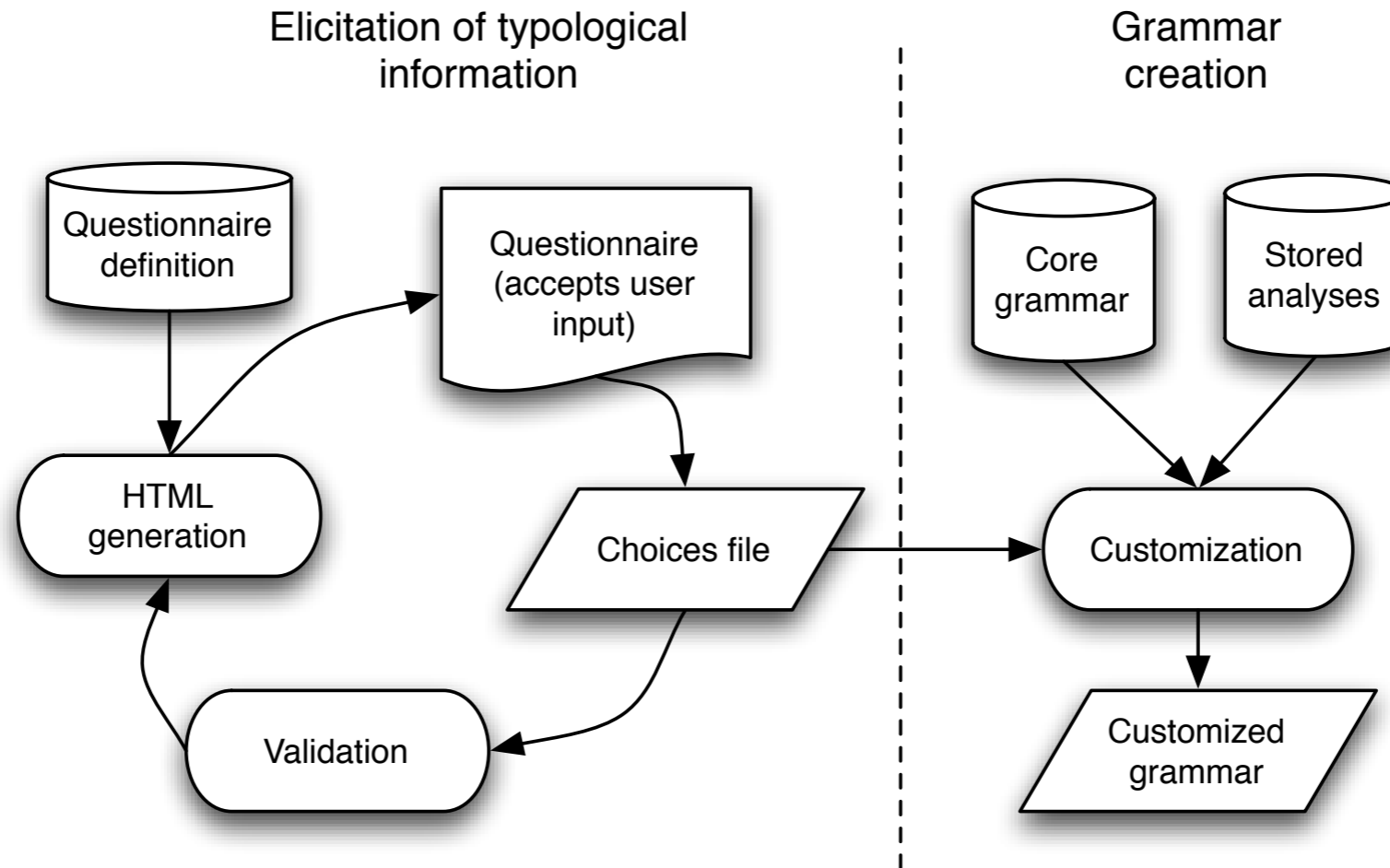
Languages - cupcaked & cookied



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
- 27.I ask whether dogs chase cars

Parse success

eng	hak	irk	jje	moh	nep	sje	tel
27	21	15	22	14	21	18	16

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

	eng	hak	irk	jje	moh	nep	sje	tel
eng	27	24	16	22	12	18	22	20
hak	18	21	12	16	9	11	17	14
irk	13	13	15	13	11	9	12	12
jje	18	18	10	22	8	11	15	0
moh	7	7	0	8	14	4	7	7
nep	13	11	10	11	7	18	13	10
sje	16	17	12	15	11	10	18	15
tel	2	2	0	1	1	2	2	11

Items with end-to-end output
(fixes to nonpast/non-past in semi.vpm)

	eng	hak	irk	jje	moh	nep	sje	tel
eng	27	24	16	22	12	18	22	20
hak	18	21	12	16	9	11	17	14
irk	13	13	15	13	11	9	12	12
jje	18	18	10	22	8	11	15	13
moh	7	7	7	8	14	4	7	7
nep	13	11	10	11	7	18	13	10
sje	16	17	12	15	11	10	18	15
tel	4	4	3	3	2	4	5	16

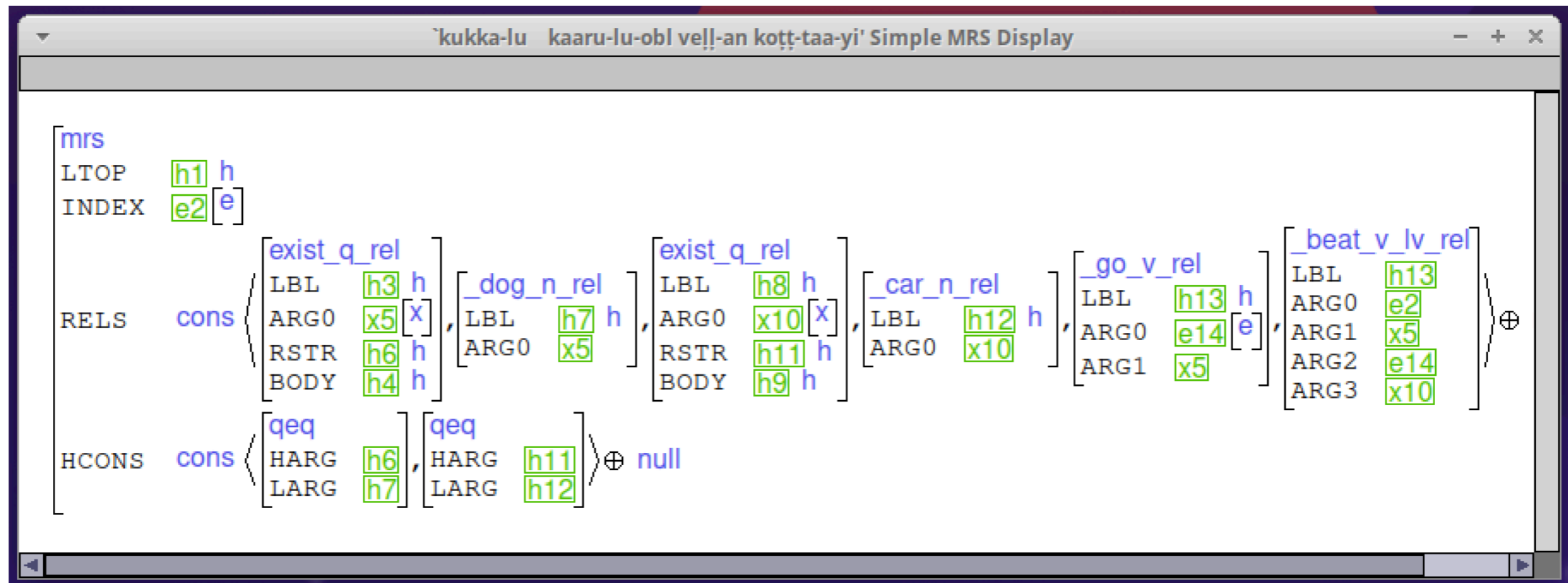
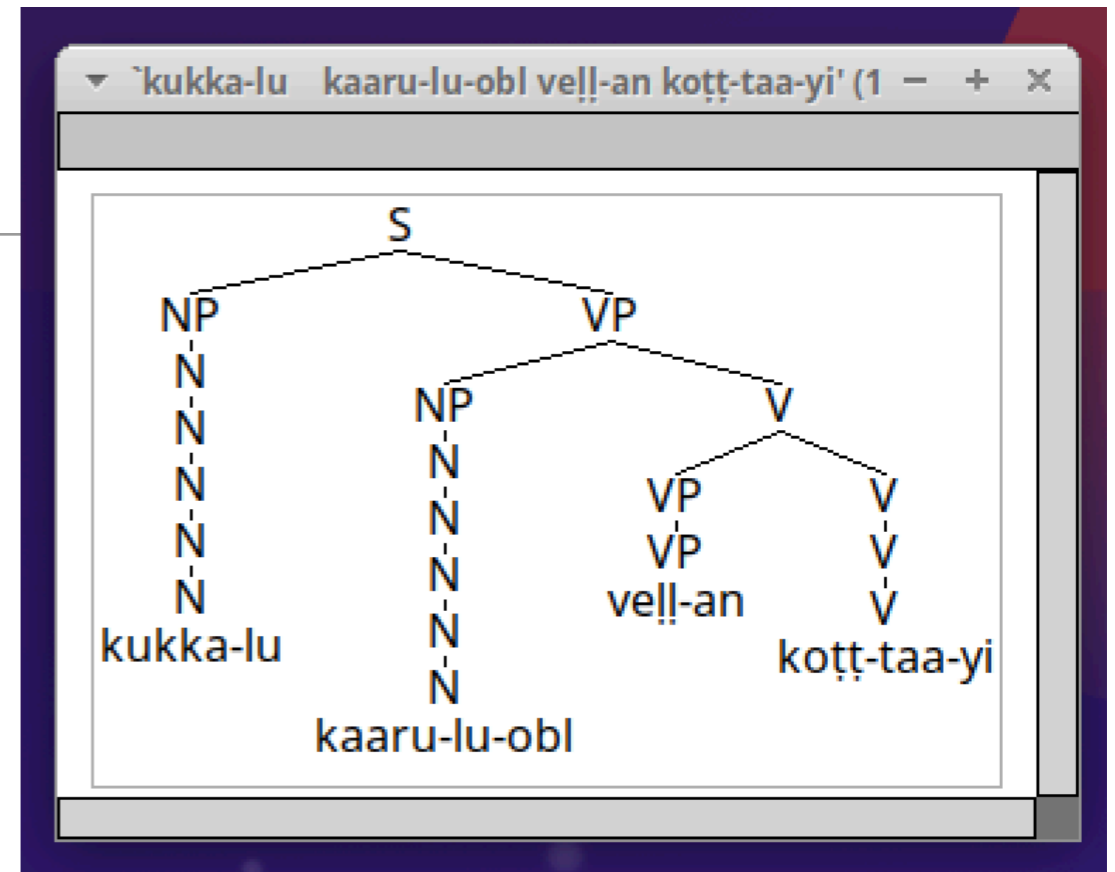
Items with end-to-end output (final run)

	eng	hak	irk	jje	moh	nep	sje	tel
eng	27	24	16	22	12	18	22	20
hak	21	21	15	19	9	13	17	14
irk	13	14	15	13	11	9	12	13
jje	19	18	11	22	9	11	15	14
moh	7	8	7	10	14	6	9	8
nep	13	11	10	11	7	18	13	10
sje	16	17	12	15	11	10	18	15
tel	14	15	11	14	9	11	16	16

Need for transfer rules

Telugu:

kukkalu kaarlu vellgotṭaayi
 kukka-lu kaaru-lu-obl vell-an koṭṭ-taa-yi
 dog-pl car-pl-obl go-INF beat-NPST-3.pl.nonhuman
 Dogs chase cars



Transfer rule example

```
go-beat-mtr := monotonic_mtr &
[ OUTPUT [ RELS < [ PRED "_beat_v_lv_rel",
    LBL #lbl,
    ARG0 #arg0,
    ARG1 #arg1,
    ARG2 #arg2 & e,
    ARG3 #arg3 ],
  [ PRED "_go_v_rel",
    ARG0 #arg2,
    ARG1 #arg1,
    LBL #lbl ] > ],
INPUT [ RELS < [ PRED "_chase_v_rel",
    LBL #lbl,
    ARG0 #arg0,
    ARG1 #arg1,
    ARG2 #arg3 ] >,
HCONS < > ]].
```

Total number of outputs (grammars as provided, w/semi.vpm fix)

	eng	hak	irk	jje	moh	nep	sje	tel
eng	121	85	269	3340	189	798	1055	47
hak	7586	402	4901	1980	1014	348	832	338
irk	303	132	2841	17828	1661	245	636	148
jje	837	71	262	3500	151	45	110	33
moh	23	10	171	4450	1349	5	98	25
nep	291	15	163	2260	137	1738	1382	13
sje	111	62	410	2980	271	462	577	35
tel	48	4	42	40	46	87	22	28

Total number of outputs (final run)

	eng	hak	irk	jje	moh	nep	sje	tel
eng	121	85	269	3340	189	798	1055	47
hak	8050	416	53513	2440	1014	444	832	338
irk	303	157	2841	17828	1661	245	884	173
jje	845	140	266	3500	153	71	120	42
moh	23	21	171	4534	1349	12	320	26
nep	291	25	163	2260	137	8218	1402	27
sje	111	62	410	2980	271	462	577	35
tel	153	41	361	1140	309	913	549	41

Maximum number of outputs (final run)

	eng	hak	irk	jje	moh	nep	sje	tel
eng	32	16	96	2048	72	480	432	9
hak	5184	138	31488	768	384	160	384	72
irk	112	36	1776	15360	648	160	432	48
jje	336	72	114	2048	48	36	36	9
moh	8	12	96	4096	600	6	216	16
nep	128	6	54	2048	72	5339	864	8
sje	32	16	224	2048	72	400	432	9
tel	32	8	216	512	128	800	432	8

Items with exact match output (final run)

	eng	hak	irk	jje	moh	nep	sje	tel
eng	27	21	14	19	8	14	18	6
hak	21	21	11	15	6	12	13	5
irk	12	11	15	9	6	8	10	5
jje	16	13	8	22	5	7	11	5
moh	7	6	6	5	12	5	8	2
nep	10	7	8	8	5	14	7	2
sje	15	13	10	11	9	9	18	5
tel	8	12	8	11	6	7	9	7

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

	eng	hak	irk	jje	moh	nep	sje	tel
eng	27	24	16	22	12	18	22	20
hak	21	21	15	19	9	13	17	14
irk	13	14	15	13	11	9	12	13
jje	19	18	11	22	9	11	15	14
moh	7	8	7	10	14	6	9	8
nep	13	11	10	11	7	18	13	10
sje	16	17	12	15	11	10	18	15
tel	14	15	11	14	9	11	16	16

Overview

- Background
- Overview results
- Interactive exploration
- Course evals