

MMT Extravaganza

Ling 567

March 7, 2024

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

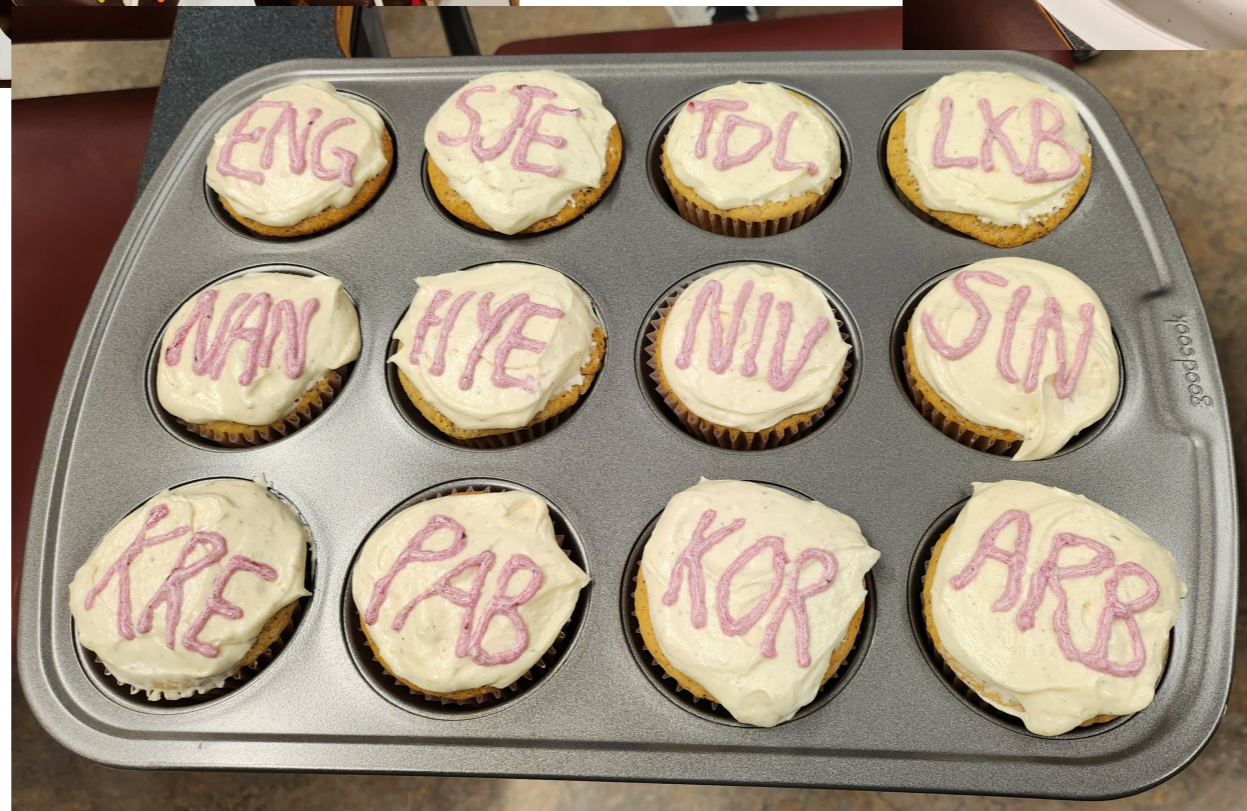
arb	Modern Standard Arabic	Afro-Asiatic	Pollack & Pong
hye	Eastern Armenian	Indo-European	Abagyan & McKibbin
kor	Korean	Koreanic	Nash & Wáng
kre	Panará	Nuclear-Macro-Je	Crowther & Luedke
nan	Southern Min	Sino-Tibetan	Kim & Wang
pab	Paresi	Arawakan	Bellamy & Tosolini
sin	Sinhala	Indo-European	Bechler & Cassell
eng	English	Indo-European	
sje	Pite Saami	Uralic	Nielsen & Spivey

Languages - mapped



lat/long from WALS, map from Google

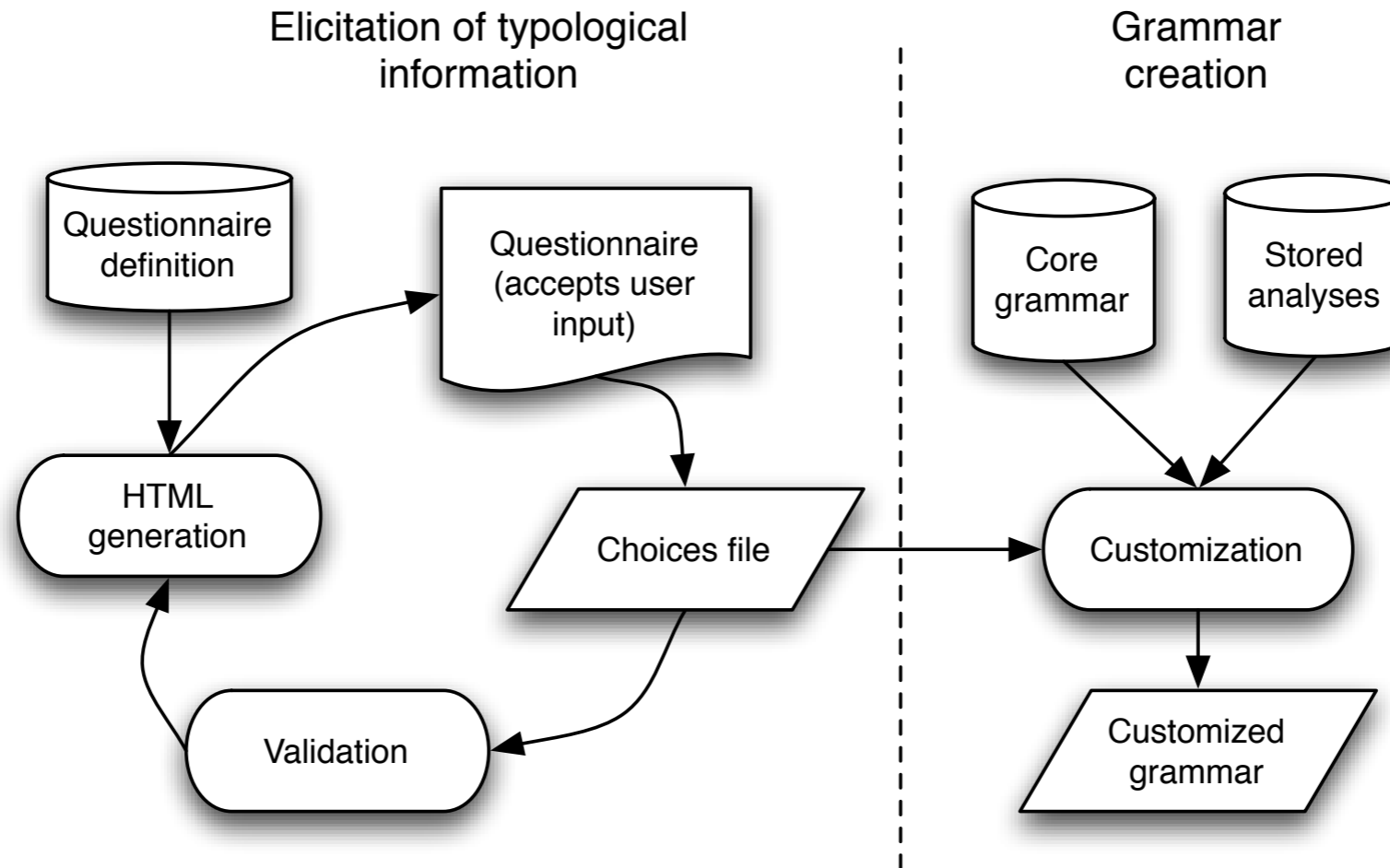
Languages - cupcaked (Thank you!)



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

arb	eng	hye	kor	kre	nan	pab	sin	sje
11	26	22	18	16	25	24	16	18

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

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	7	6	8	8	8	9	8	6
eng	14	26	13	18	19	23	19	17	21
hye	10	16	12	0	17	18	16	16	18
kor	10	13	7	18	12	13	13	11	13
kre	9	12	7	12	14	11	11	13	11
nan	11	18	8	0	14	22	13	12	19
pab	13	13	10	0	12	12	21	11	10
sin	8	14	10	0	15	14	12	15	15
sje	8	16	9	14	14	17	13	11	18

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

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	7	6	8	8	8	9	8	6
eng	14	26	13	18	19	23	19	17	21

arb	eng	hye	kor	kre	nan	pab	sin	sje
11	26	22	18	16	25	24	16	18

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

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	7	6	8	8	8	9	8	6
eng	14	26	13	18	19	23	19	17	21
hye	10	16	12	0	17	18	16	16	18
kor	10	13	7	18	12	13	13	11	13
kre	9	12	7	12	14	11	11	13	11
nan	11	18	8	0	14	22	13	12	19
pab	13	13	10	0	12	12	21	11	10
sin	8	14	10	0	15	14	12	15	15
sje	8	16	9	14	14	17	13	11	18

Items with end-to-end output: intermediate
(transfer rule propagation)

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	9	7	8	8	8	9	8	6
eng	14	26	13	18	19	23	19	17	21
hye	12	20	12	0	17	20	18	16	17
kor	11	14	8	18	13	14	14	12	13
kre	12	16	7	12	14	14	14	13	10
nan	13	21	8	0	14	20	13	12	19
pab	14	14	10	0	13	13	21	12	11
sin	10	16	10	0	15	16	14	15	14
sje	8	16	9	14	14	17	13	11	18

Some fixes & remaining issues

- Fixed:

- kor: remove ASPECT from semi.vpm
- L/R-HNDL in VP and S coordination across languages

- Open:

- Gen search space in hye
- Generation with adverb-marked clausal mods
- Some transfer rules (nominalization in pab; locatives in nan; other?)

Items with end-to-end output

Final

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	10	7	8	8	8	9	8	6
eng	14	26	13	18	19	23	19	17	21
hye	12	22	12	16	17	20	18	16	17
kor	11	15	8	18	13	14	14	12	13
kre	12	16	7	12	14	14	14	13	10
nan	13	22	8	16	14	20	13	12	19
pab	14	15	10	12	13	13	21	12	11
sin	10	16	10	13	15	16	14	15	14
sje	8	18	9	14	14	17	13	11	18

Need for transfer rules

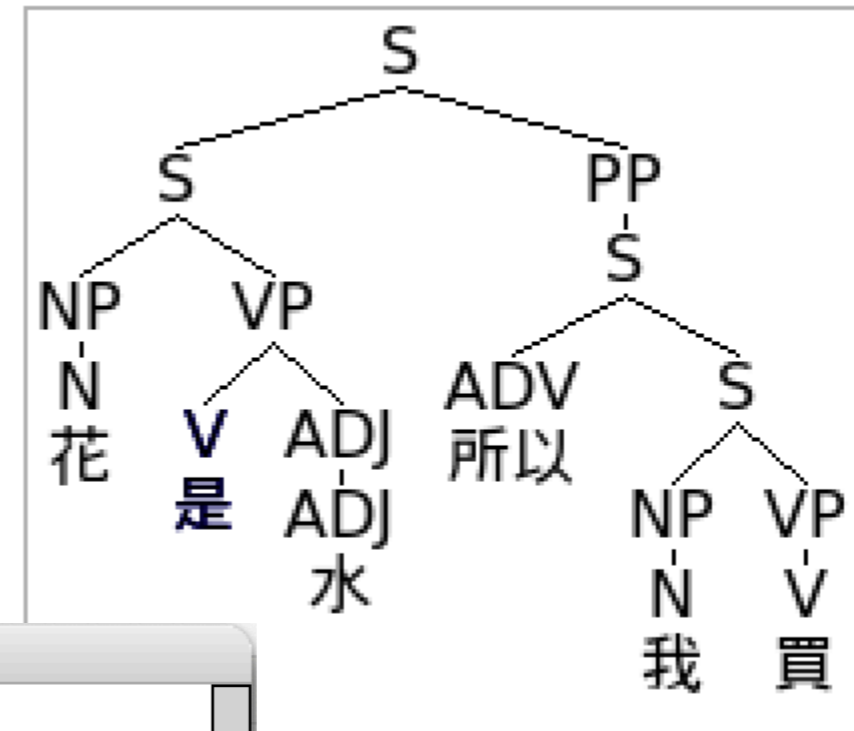
花 是 水 所以 我 買

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 (grammars & transfer rules as provided)

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	465	37	56	2824	60	28	185	3872	34
eng	181	84	152	874	1553	173	650	1661	587
hye	325	71	346	0	291	193	1071	2445	277
kor	9342	736	5280	10380	718	50	1116	739	202
kre	181	475	708	32115	516	33	204	1934	39
nan	11533	11610	7584	0	2056	230	1201	1856	2474
pab	85397	240	1118	0	596	151	11412	2170	404
sin	125	100	244	0	490	50	182	633	112
sje	781	99	496	1746	398	120	318	302	361

Total number of outputs (final run)

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	465	157	68	2824	60	29	185	3872	34
eng	181	84	152	874	1553	173	650	1661	587
hye	425	294	1426	710	291	409	1249	2445	271
kor	9346	1468	5812	10380	721	115	1120	931	202
kre	429	631	708	32120	516	112	422	1934	33
nan	11821	22114	7975	69146	2056	1050	1201	1856	2474
pab	85385	370	1118	1416	597	154	11412	2172	410
sin	187	298	1232	1456	490	494	448	633	106
sje	781	127	496	1746	398	120	318	302	361

Maximum number of outputs (final run)

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	192	108	32	1536	24	18	72	3557	12
eng	72	32	72	384	1080	36	216	1296	216
hye	192	96	1232	144	80	176	512	1728	144
kor	8640	288	4000	7680	324	42	720	288	84
kre	144	200	368	25272	160	45	128	592	6
nan	8640	8232	3680	24576	1296	798	720	576	864
pab	82756	128	464	768	204	60	9632	680	288
sin	48	84	1056	936	264	237	144	144	24
sje	640	24	256	768	216	60	128	72	216

Items with exact match output (final run)

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	4	4	2	2	3	3	3	3
eng	5	26	11	15	13	20	16	15	18
hye	4	17	10	14	12	16	14	12	15
kor	3	13	7	18	10	12	11	12	11
kre	4	16	6	10	14	12	11	12	8
nan	4	19	7	14	10	17	11	12	14
pab	4	15	8	10	8	9	21	10	10
sin	5	13	7	11	12	12	11	15	12
sje	4	17	9	11	8	16	10	10	18

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

	arb	eng	hye	kor	kre	nan	pab	sin	sje
arb	11	10	7	8	8	8	9	8	6
eng	14	26	13	18	19	23	19	17	21
hye	12	22	12	16	17	20	18	16	17
kor	11	15	8	18	13	14	14	12	13
kre	12	16	7	12	14	14	14	13	10
nan	13	22	8	16	14	20	13	12	19
pab	14	15	10	12	13	13	21	12	11
sin	10	16	10	13	15	16	14	15	14
sje	8	18	9	14	14	17	13	11	18

Overview

- Background
- Overview results
- Interactive exploration
- Course evals