MMT Extravaganza

Ling 567
March 10, 2022

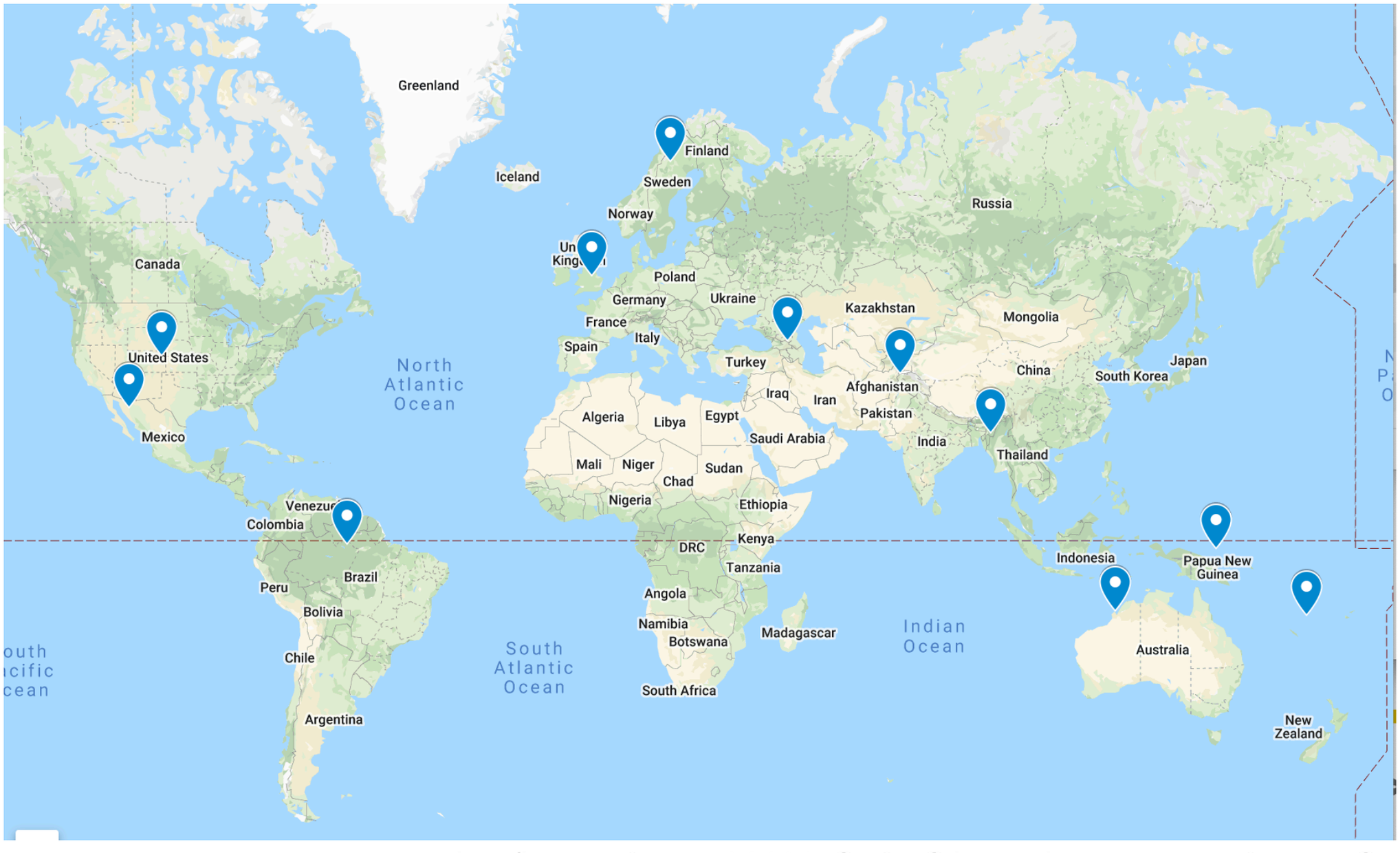
Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

arp	Arapaho	Algic	Kodama & Pendleton
bbl	Tsova-Tush	Nakh-Daghestanian	Briand & Brown
bcj	Bardi	Nyulnyulan	Nezakati & Wueger
erk	South Efate	Austronesian	Ban & Bansal
hix	Hixkaryana	Cariban	Lin & Soni
mni	Meitei	Sino-Tibetan	Liu & Zeng
ttv	Titan	Austronesian	Guo & Wang
wbl	Wakhi	Indo-European	Langford & Okada
yaq	Hiaki	Uto-Aztecan	Iyer & O'Connell
eng	Indo-European		
sje	Pite Saami		Nielsen & Spivey

Languages - mapped

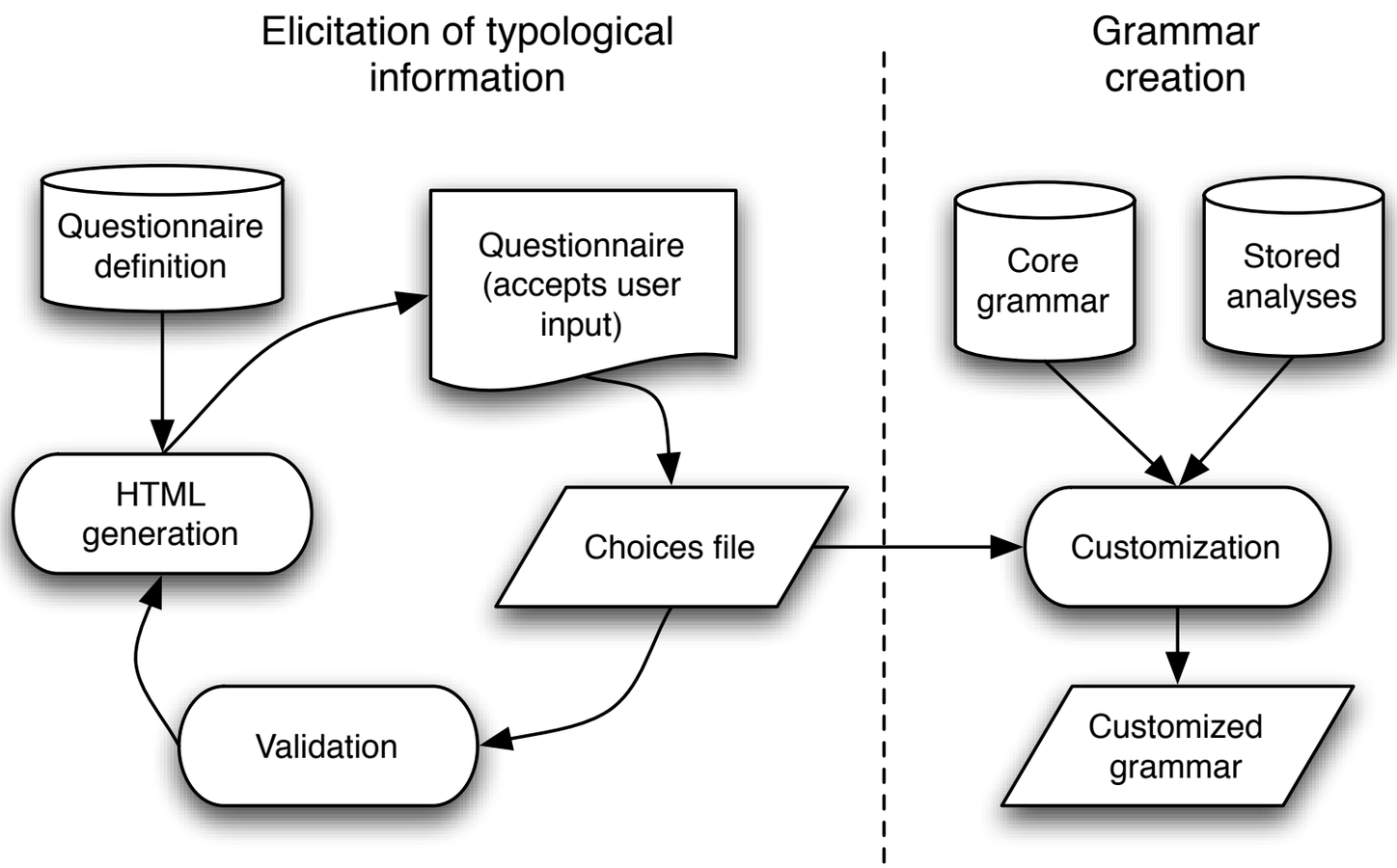


lat/long from WALS & Glottolog, map from Google

Grammar coverage (mostly shared)

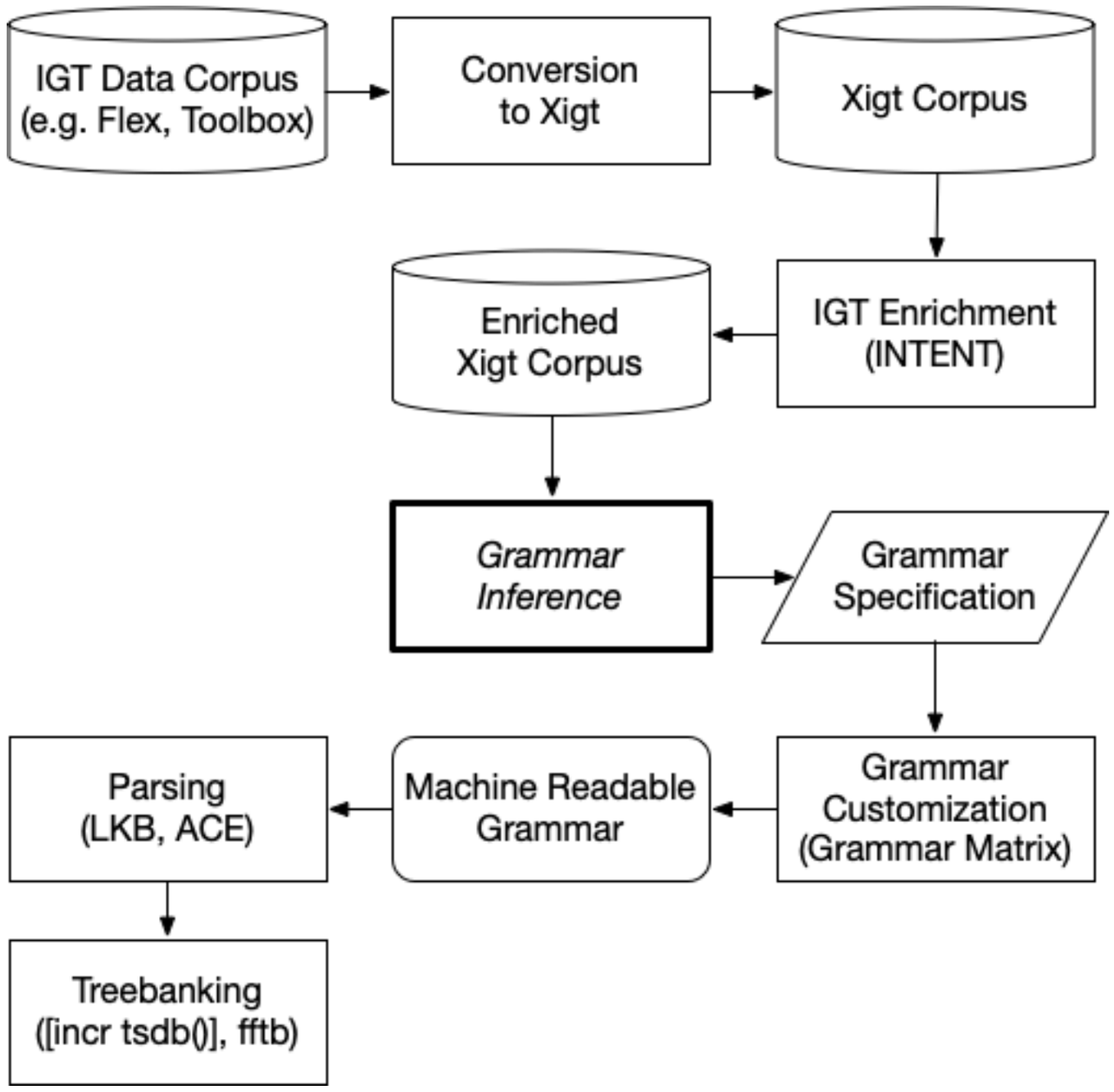
- Basic word order
- Case
- Agreement
- Personal pronouns
- Tense/aspect
- Sentential negation
- Argument optionality
- Matrix yes-no questions
- Coordination
- Modification (adjective, clausal mods)
- Clausal complements
- Wh questions
- Possessives

Grammar Development: Customization + Extension



(Bender et al 2010)

Starting point:
Automatically inferred grammar specifications



(Howell 2020, Conrad 2021)

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

arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
14	14	10	26	18	12	8	18	17	16	12

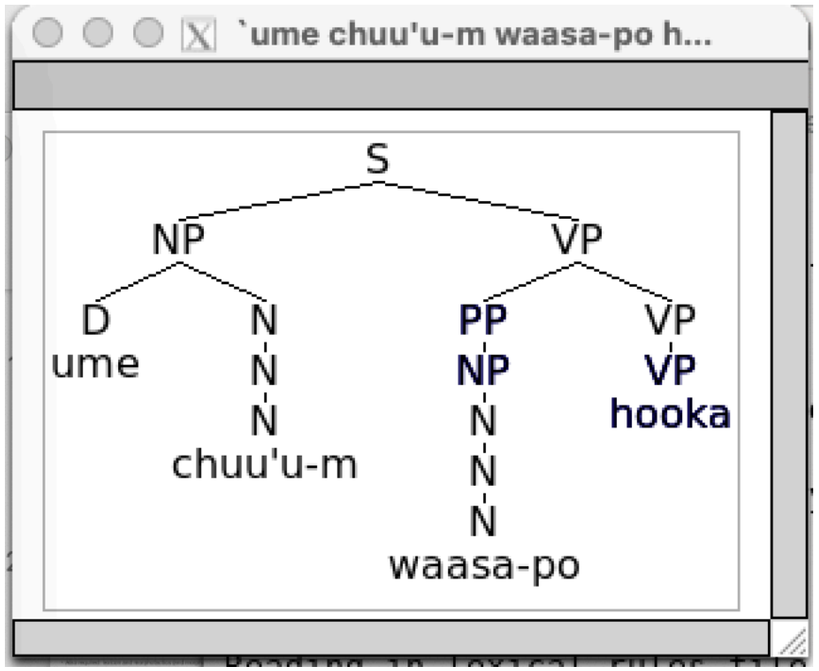
Items with end-to-end output (grammar & transfer rules as provided)

	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	14	11	8	8	10	8	9	10	12	9	7
bbl	9	14	8	10	14	9	8	11	9	8	6
bcj	6	9	9	5	5	6	6	6	7	6	6
eng	13	22	12	26	19	14	12	21	18	17	12
erk	10	12	8	13	17	9	10	15	13	12	8
hix	6	8	5	8	6	11	7	6	9	7	6
mni	5	5	7	4	3	4	8	5	5	4	3
sje	10	15	8	16	15	9	9	18	14	14	10
ttv	1	1	8	13	13	11	9	13	16	13	1
wbl	10	12	7	9	10	10	5	12	14	16	10
yaq	5	7	7	6	7	6	3	7	9	8	12

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

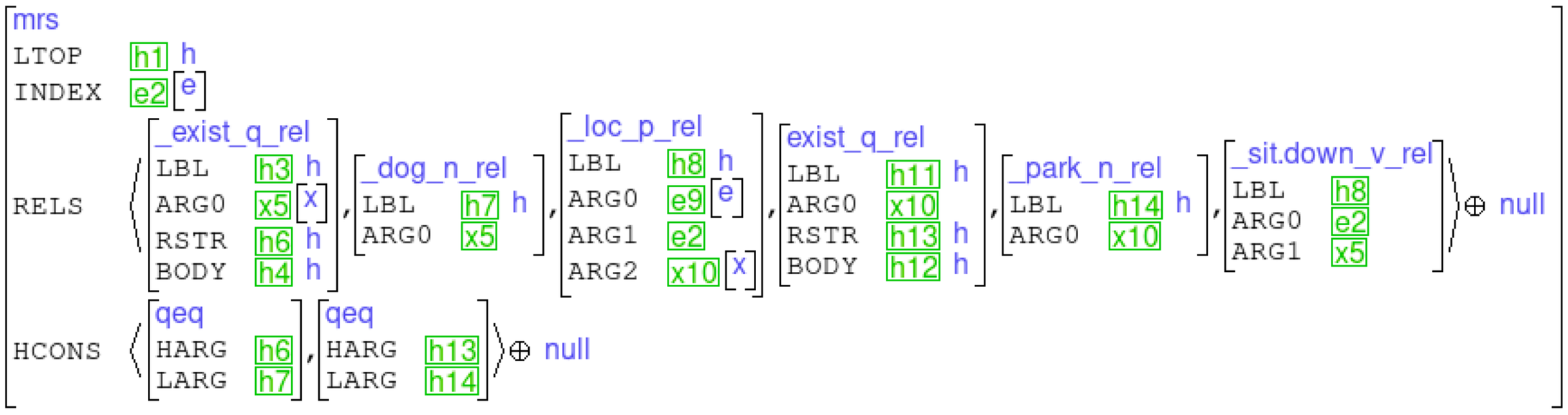
	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	14	12	8	11	11	9	9	12	12	12	7
bbl	9	14	8	13	14	11	8	11	9	9	6
bcj	6	7	9	7	5	7	6	7	7	8	6
eng	13	22	12	26	20	14	12	21	18	17	12
erk	10	12	8	14	17	10	10	15	13	13	8
hix	7	10	6	11	8	11	8	9	10	10	7
mni	5	5	7	5	3	5	6	5	4	5	3
sje	10	15	8	16	15	9	9	18	14	14	10
ttv	1	1	8	14	14	11	9	14	16	14	1
wbl	10	12	7	9	10	10	5	12	14	16	10
yaq	5	7	7	6	7	7	3	7	9	9	10

Need for transfer rules



chuu'u-m waasa-po hooka
dog-PL field-LOC be.sitting
dogs are (sitting) in the field (yaq)

NB: invented example!



Transfer rule example

```
in-to-loc-sit-mtr := monotonic_mtr &
[ INPUT [ RELS < [ PRED "_in_p_rel",
          LBL #lbl,
          ARG0 #event,
          ARG1 #sitter,
          ARG2 #location ] > ],
  OUTPUT [ RELS < [ PRED "_loc_p_rel",
                    LBL #lbl,
                    ARG1 #event,
                    ARG2 #location],
         [ PRED "_sit.down_v_rel",
           LBL #lbl,
           ARG0 #event,
           ARG1 #sitter] > ]].
```

Transfer rule example

```
pronoun-delete-mtr := monotonic_omtr &
  [ INPUT [ RELS < [ PRED "pron_rel",
                     ARG0 #x,
                     LBL #larg ],
                   [ PRED "exist_q_rel",
                     ARG0 #x,
                     RSTR #harg ] >,
           HCONS < qeq & [ LARG #larg,
                           HARG #harg ] > ],
  OUTPUT [ RELS < >,
           HCONS < > ]].
```

Total number of outputs
(grammars & transfer rules as provided)

	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	2830	802	212	78	10102	51	28	242	40	2880	19
bbl	1868	600024	12470	57	18442	17	20	48	18	104	25
bcj	4082	3546	1940	58	206	10	11	64	122	144	15
eng	912	20464	172	113	32653	21	3481	1019	24	2644	16
erk	44054	764	690	7369	31226	38	1174	2005	417	43509	310
hix	144	36	88	148	136	13	15	17	9	73	6
mni	1620	208	2636	88	232	28	21	68	148	241	52
sje	1260	644	84	111	14536	18	1171	577	21	2598	16
ttv	5400	6	398	95	363363	17	23	133	199	1034	1
wbl	4444	2782	422	1207	44244	23	4618	3578	177	18307	32
yaq	9998	1666	1924	30	500604	17	14	95	143	13789	1366

Total number of outputs (transfer rule propagation)

	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	2830	772	212	97	11682	55	28	278	36	2931	19
bbl	1868	1079881	12470	1917	49834	32	56	88	48	223	25
bcj	4082	2394	1940	69	240	12	11	70	122	174	15
eng	912	20463	172	113	33084	21	3481	1019	24	2644	16
erk	44054	770	690	7421	67130	42	1192	2039	431	43521	310
hix	180	68	90	173	762	13	52	51	40	208	7
mni	1620	1264	2636	276	684	36	85	196	256	414	52
sje	1260	644	84	111	14536	18	1171	577	21	2598	16
ttv	5400	6	398	133	368142	17	57	38595	117	1721	1
wbl	4444	2782	422	1207	44244	23	4618	3578	177	18307	32
yaq	9998	1990	1924	100	490985	18	46	95	117	16244	910

Maximum number of outputs (transfer rule propagation)

	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	1296	432	84	24	8192	36	4	144	8	2304	4
bbl	804	1073669	10598	944	41472	11	40	24	32	128	8
bcj	3888	2352	896	32	128	4	2	36	98	48	8
eng	324	12033	54	32	14116	6	3456	432	6	1536	4
erk	29049	432	168	4000	46080	12	1152	864	128	31246	128
hix	72	18	42	128	480	2	20	12	16	64	1
mni	1008	1188	1008	168	624	12	40	132	200	240	40
sje	792	204	24	32	10240	6	1152	432	5	1536	4
ttv	5400	6	126	48	247537	6	20	37884	34	1280	1
wbl	1944	1176	324	1152	40960	4	4608	1728	56	9216	8
yaq	7960	1176	1616	43	487205	6	42	64	49	13248	812

Items with exact match output (transfer rule propagation)

	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	7	6	3	10	9	7	5	9	1	1	2
bbl	6	10	2	9	7	1	4	6	3	3	5
bcj	2	4	6	6	4	4	3	5	0	3	4
eng	6	9	4	26	11	11	7	18	6	5	6
erk	5	8	3	13	13	9	5	10	5	3	5
hix	6	8	2	10	6	11	5	6	3	3	4
mni	4	3	1	4	2	5	4	2	0	0	1
sje	3	6	3	15	9	7	4	18	5	5	6
ttv	1	1	2	1	11	0	4	7	7	4	0
wbl	3	5	4	9	6	6	2	10	3	12	8
yaq	1	5	4	6	5	4	1	6	3	6	10

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

	arp	bbl	bcj	eng	erk	hix	mni	sje	ttv	wbl	yaq
arp	14	12	8	11	11	9	9	12	12	12	7
bbl	9	14	8	13	14	11	8	11	9	9	6
bcj	6	7	9	7	5	7	6	7	7	8	6
eng	13	22	12	26	20	14	12	21	18	17	12
erk	10	12	8	14	17	10	10	15	13	13	8
hix	7	10	6	11	8	11	8	9	10	10	7
mni	5	5	7	5	3	5	6	5	4	5	3
sje	10	15	8	16	15	9	9	18	14	14	10
ttv	1	1	8	14	14	11	9	14	16	14	1
wbl	10	12	7	9	10	10	5	12	14	16	10
yaq	5	7	7	6	7	7	3	7	9	9	10

Overview

- Background
- Overview results
- Interactive exploration
- Course evals