MT Extravaganza

Ling 567
June 6, 2019

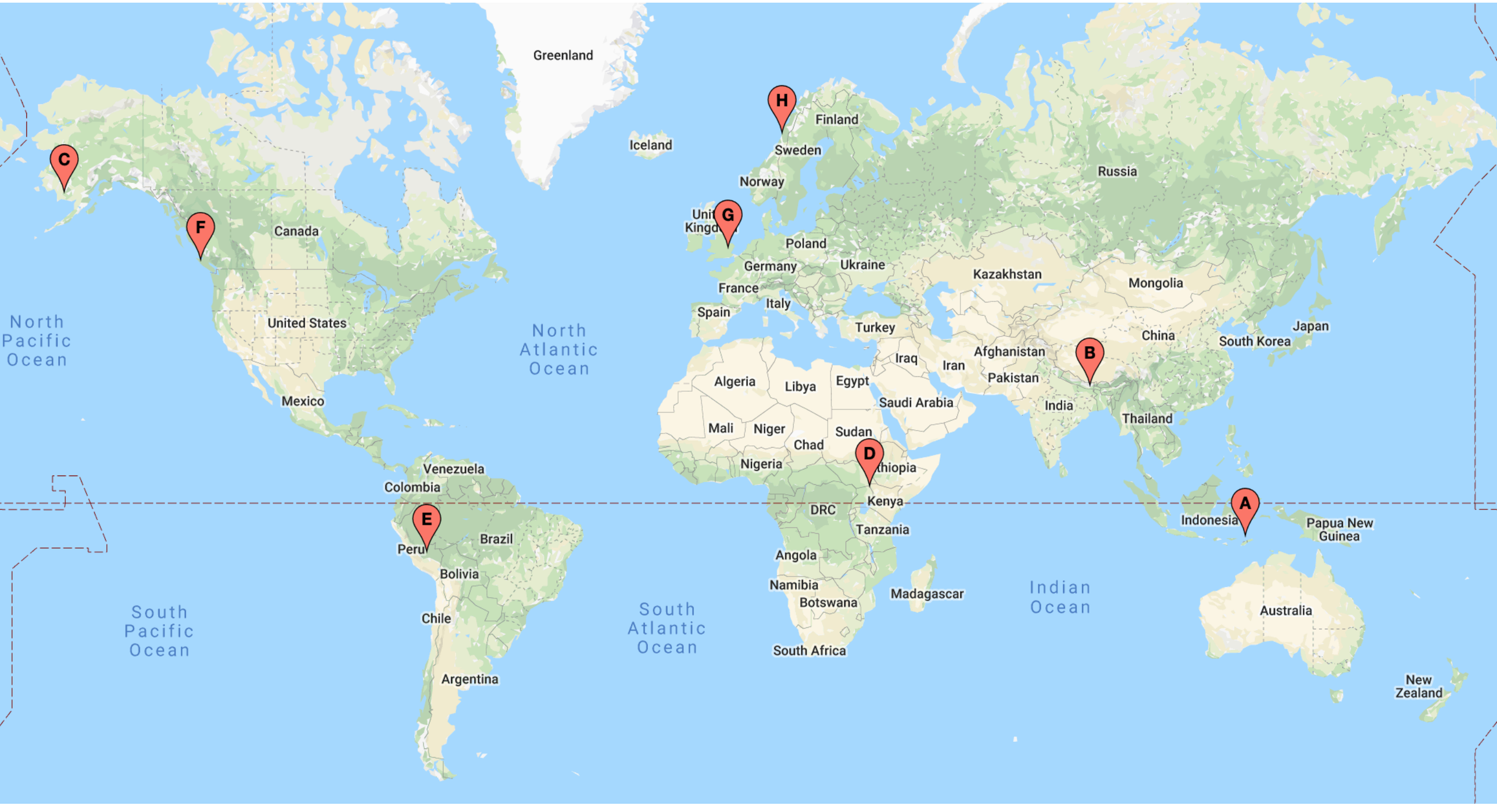
Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

abz	Abui	Alor-Pantar	Dods, Pompeo
ctn	Chintang	Kiranti	Vogel, Zhang
esu	Central Alaskan Yup'ik	Eskimo-Aleut	Martinez, Strunk
eng	English	Indo-European	
ikx	Ik	Kuliak	Sampath, Tien
mcb	Matsigenka	Arawakan	Gioannini, Mena
nuk	Nuuchahnulth	Wakashan	Downey, Nielsen
sje	Pite Saami	Uralic	Nielsen, Spivey

Languages - mapped



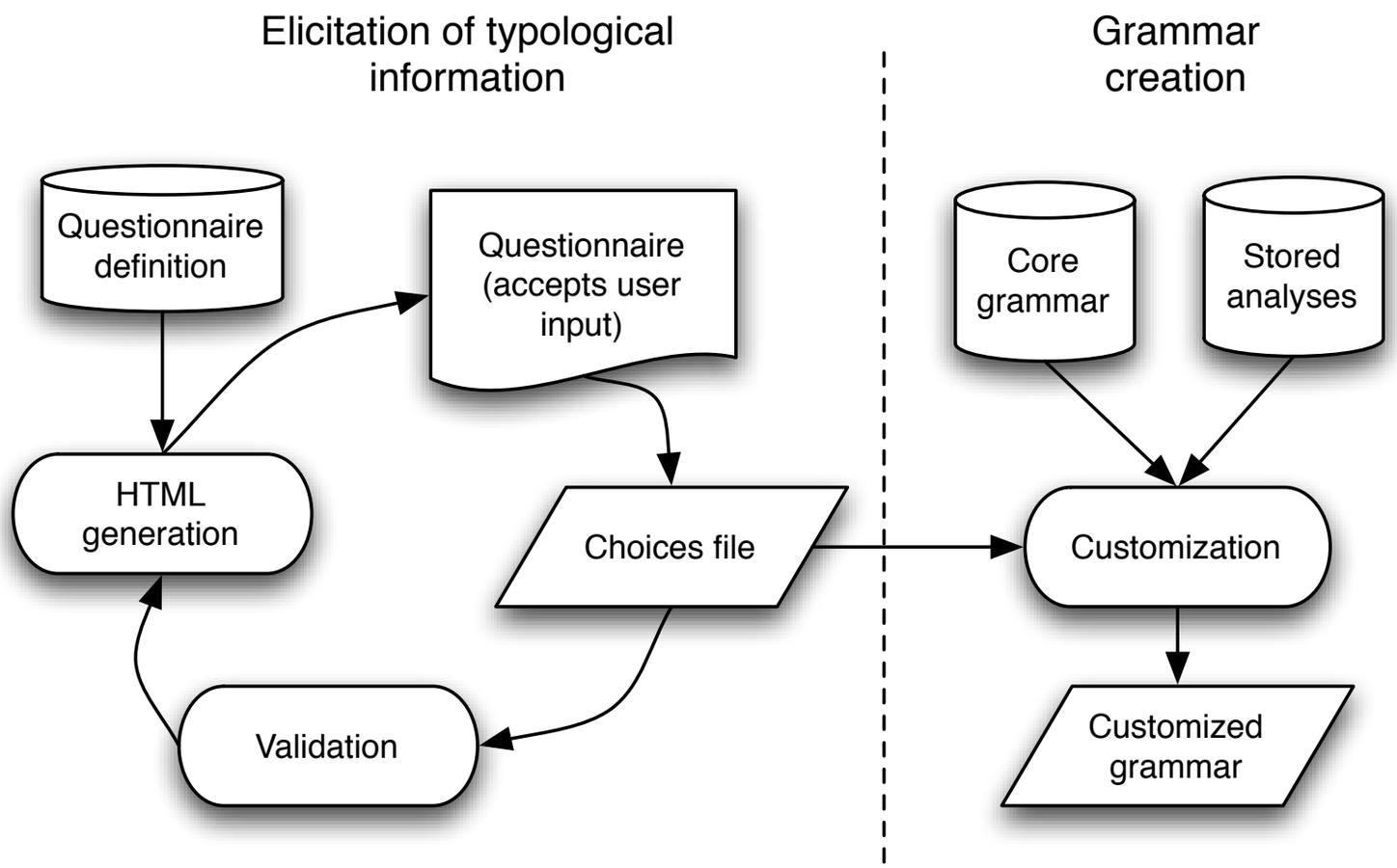
lat/long data mostly from wals.info; map by batchgeo.com

Languages - cupcaked

Grammar coverage (mostly shared)

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

Grammar Development: Customization + Extension



Added chaos: Started with automatically inferred grammar specifications

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

MMT with ACE

- Faster system run times
- More coverage (fewer system timeouts)
- Compatible with Condor (yay!)
- Possibility of respecting ICONS representation of information structure

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.I asked what the dogs chased
19.The dog s car sleeps
20.My dogs sleep
21.The dog sleeps because the cat sleeps
22.The dog sleeps after the cat sleeps
23.Who sleeps
24.What do the dogs chase
25.What do you think the dogs chase
26.Who asked what the dogs chase
27.I can eat glass
28.It doesnt hurt me

Parse success

abz	ctn	eng	esu	ikx	mcb	nuk	sje
20	12	28	16	13	8	12	17

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

	abz	ctn	eng	esu	ikx	mcb	nuk	sje
abz	17	9	16	10	13	12	9	17
ctn	7	12	10	6	7	6	6	9
eng	22	12	28	14	17	14	15	21
esu	10	10	11	16	12	9	11	10
ikx	12	10	13	11	13	8	9	11
mcb	6	5	6	7	7	8	6	6
nuk	10	7	7	7	9	8	12	8
sje	16	7	15	9	13	10	9	17

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

	abz	ctn	eng	esu	ikx	mcb	nuk	sje
abz	17	9	16	10	13	12	10	17
ctn	9	12	12	6	9	6	8	9
eng	22	12	28	14	17	14	15	21
esu	14	10	16	17	13	9	12	13
ikx	12	10	13	11	13	8	9	11
mcb	8	5	7	7	7	8	6	7
nuk	12	9	10	8	11	8	12	11
sje	16	7	15	9	13	10	9	17

Transfer rule example

```
eat-ini-mtr := monotonic_mtr &
[ CONTEXT [ RELS < [ PRED "_eat_v_rel",
                ARG2 #arg2 ] > ],
  FILTER [ RELS < [ ARG0 #arg2 ] > ],
  INPUT [ RELS < >,
          HCONS < > ],
OUTPUT [ RELS < [ PRED "_food_n_rel",
                LBL #larg,
                ARG0 #arg2 & [ COG-ST type-id,
                               NUM sg] ],
              [ PRED "exist_q_rel",
                ARG0 #arg2,
                RSTR #harg ] >,
         HCONS < [ HARG #harg,
                   LARG #larg ] > ],
FLAGS [ EQUAL < #arg2 > ]].
```

Total number of outputs (transfer rules as provided)

	abz	ctn	eng	esu	ikx	mcb	nuk	sje
abz	90	1794	7463	324	25699	24432	20	4644
ctn	14	1959	140	150	152	3530	26	50
eng	115	989	117	500	51161	20120	50	1019
esu	115	3910	5510	7905	39916	10296	56	239
ikx	65	1045	1021	414	4313	4520	41	469
mcb	24	2430	919	1064	10690	7848	27	414
nuk	76	2027	312	671	5210	6982	61	2624
sje	66	517	110	139	4453	19876	24	571

Total number of outputs (transfer rule propagation)

	abz	ctn	eng	esu	ikx	mcb	nuk	sje
abz	90	1610	7495	352	25777	25392	23	4652
ctn	21	1959	366	150	620	3530	33	50
eng	115	989	117	500	51161	20120	50	1019
esu	243	3910	23941	7909	81923	10296	96	545
ikx	65	1045	1345	414	4729	4520	41	469
mcb	151	2430	1432	1064	12145	7848	27	738
nuk	112	3557	393	993	8716	6982	61	2942
sje	66	517	110	139	4453	19876	24	571

Maximum number of outputs (Final)

	abz	ctn	eng	esu	ikx	mcb	nuk	sje
abz	36	296	6912	210	18432	18432	4	3456
ctn	6	532	168	112	288	1248	15	12
eng	38	198	32	336	49152	9216	15	432
esu	124	1536	22048	6632	66096	3328	40	252
ikx	41	262	864	350	4096	2880	17	432
mcb	126	888	576	630	10530	2496	6	324
nuk	29	1040	128	414	3564	1920	16	2304
sje	16	130	32	96	4096	9216	4	432

Items with exact match output (Final)

	abz	ctn	eng	esu	ikx	mcb	nuk	sje
abz	17	2	15	7	0	7	7	14
ctn	6	5	9	6	1	4	6	6
eng	16	4	28	10	1	8	9	17
esu	11	2	12	14	1	6	8	9
ikx	9	3	11	8	1	6	6	8
mcb	7	0	5	5	1	8	4	6
nuk	6	2	8	5	1	6	11	6
sje	12	1	14	6	0	6	6	17

Overview

- Background
- Overview results
- Interactive exploration
- Course evals