

MT Extravaganza

Ling 567

March 10, 2016

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

ISO 639-3	Lg Name	Lg Family	Grammar Authors
dtg	Toro Tegu	Niger-Congo	Garnick & Mansfield
eng	English	Indo-European	
gug	Paraguayan Guaraní	Tupí-Gurananí	Tittle & Wicke
hdn	Haida	Na-Dene	Pearah & Piepgrass
max	Ternate Malay	Austronesian	Preddy & Sen
pbv	Pnar	Austro-Asiatic	Jarvis & McCranie
shu	Chadian Arabic	Afro-Asiatic	Heath & Marsh
ybh	Yakkha	Sino-Tibetan	Arshad & Dai

Languages - mapped



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

Languages - cupcaked



Grammar coverage (shared)

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

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, INFO-STR) from grammar internal space to interlingual space. Lossy mapping, provides defaults
- One ‘accommodation’ transfer grammar per language, instantiating shared transfer rules

MMT with ACE (for the 4th year)

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

A big thank you to
Woodley & Sanghoun!

Input sentences

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

Timing

- Submit jobs to translate 24 sentences from src2tgt to Condor: 64 lg pairs
- 32 finish within 5 seconds
- 5-9s : dtt2dtt, hdn2max, max2eng, pbv2pbv, eng2dtt, eng2hdn, dtt2gug, shu2dtt
- 10-19s: pbv2ybh, gug2ybh, shu2hdn, max2shu, max2hdn, gug2max, max2dtt
- 20-29s: hdn2hdn, shu2gug, hdn2dtt, gug2pbv, eng2gug, gug2eng
- 30-59s: pbv2shu, max2gug, gug2gug

Timing: Winners!

- pbv2max 1:09
- hdn2gug 1:15
- gug2dt 2:17
- pbv2gug 2:55
- pbv2dt 4:40
- gug2hdn 5:40
- pbv2hdn 6:47
- pbv2eng 7:41
- hdn2gug (no pro-drop hack):
3:42:58

Parse success

dtb	eng	gug	hdn	max	pbv	shu	ybh
18	24	24	23	24	22	22	16

	dtb	eng	gug	hdn	max	pbv	shu	ybh
Total parses	19	25	85	94	29	146	24	34
Avg parses	1.06	1.04	3.54	4.09	1.21	6.35	1.09	2.13

Items with end-to-end output
(initial vpm fixes, no transfer rules)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	18	17	15	16	17	16	14	13
eng	20	24	22	23	24	22	21	15
gug	13	16	24	16	16	21	20	11
hdn	18	23	22	23	23	23	15	14
max	20	24	22	23	24	23	21	15
pbv	17	21	20	21	21	22	20	12
shu	17	22	20	20	21	20	22	11
ybh	13	14	12	13	14	13	10	16

(‘run13’)

Items with end-to-end output
(added in transfer rules, take1)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	18	17	17	16	17	16	14	14
eng	20	24	24	23	24	23	21	16
gug	13	16	24	21	16	21	20	11
hdn	17	23	23	23	23	23	15	14
max	20	24	24	23	24	23	21	16
pbv	16	21	22	21	21	22	20	13
shu	18	22	22	21	22	21	22	12
ybh	13	14	13	13	14	13	10	16

(‘run14’)

Total number of outputs (added in transfer rules, take1)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	5731	208	76853	7256	2217	389	517	896
eng	18280	80	125490	10056	4857	117	115	4472
gug	22964	1864	125076	3237511	13663	9409	4009	3018
hdn	20660	585	611965	26004	12028	1690	1023	5790
max	58840	13830	160162	42608	19263	1513	3467	5104
pbv	94121	5513	650714	254470	32704	1099	930	13122
shu	18282	835	118057	32392	2764	93	93	9603
ybh	2460	86	1915	780	253	128	24	4690

(‘run14’)

Transfer rule example

```
hungry-have-hunger-mtr := monotonic_mtr &
[ INPUT [ RELS < [ PRED "_hungry_a_rel",
    LBL #lbl,
    ARG0 #arg0,
    ARG1 #arg1 ] >,
  HCONS < > ],
OUTPUT [ RELS < [ PRED "_have_v_rel",
    LBL #lbl,
    ARG0 #arg0,
    ARG1 #arg1,
    ARG2 #hunger],
  [ PRED "_hunger_n_rel",
    LBL #larg,
    ARG0 #hunger ],
  [ PRED "exist_q_rel",
    ARG0 #hunger,
    RSTR #harg ] >,
  HCONS < [ HARG #harg,
    LARG #larg ] > ]].
```

Transfer rule example: hack for gug2hdn

```
be_v_id-pronoun-delete-mtr := monotonic_mtr &
  [ CONTEXT [ RELS < [ PRED "_be_v_id_rel",
                      ARG1 #x ] > ],
    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 < > ]].
```

Items with end-to-end output (added in transfer rules, take1)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	18	17	17	16	17	16	14	14
eng	20	24	24	23	24	23	21	16
gug	13	16	24	21	16	21	20	11
hdn	17	23	23	23	23	23	15	14
max	20	24	24	23	24	23	21	16
pbv	16	21	22	21	21	22	20	13
shu	18	22	22	21	22	21	22	12
ybh	13	14	13	13	14	13	10	16

(‘run14’)

Items with end-to-end output: Final
(further misc fixes, transfer rule propagation)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	18	18	18	17	18	17	16	14
eng	21	24	24	23	24	23	22	17
gug	20	23	24	22	23	23	22	14
hdn	18	23	23	23	23	23	21	15
max	21	24	24	23	24	23	22	17
pbv	18	23	23	22	23	23	22	15
shu	19	22	22	21	22	21	22	14
ybh	14	16	14	14	16	14	13	16

(‘run19’)

Total number of outputs (added in transfer rules, take1)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	5731	208	76853	7256	2217	389	517	896
eng	18280	80	125490	10056	4857	117	115	4472
gug	22964	1864	125076	3237511	13663	9409	4009	3018
hdn	20660	585	611965	26004	12028	1690	1023	5790
max	58840	13830	160162	42608	19263	1513	3467	5104
pbv	94121	5513	650714	254470	32704	1099	930	13122
shu	18282	835	118057	32392	2764	93	93	9603
ybh	2460	86	1915	780	253	128	24	4690

(‘run14’)

Total number of outputs: Final (further misc fixes, transfer rule propagation)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	5731	209	76879	7260	2225	390	1602	896
eng	18296	80	125491	10056	4857	125	314	4484
gug	102816	31503	125076	348439	71673	92275	4093	7484
hdn	20724	684	611966	25396	12388	1717	3095	5298
max	58876	13830	160162	42608	19319	1713	50332	5122
pbv	101986	37225	650783	214092	95981	16425	4304	15378
shu	18314	835	118059	32392	2764	102	228	9645
ybh	2640	96	1943	788	277	129	85	4690

(‘run19’)

Items with exact match output (run19)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	18	17	18	17	18	13	15	13
eng	17	24	24	21	24	21	21	14
gug	15	20	24	16	23	17	15	12
hdn	15	19	23	18	19	14	20	11
max	17	23	22	18	24	17	21	13
pbv	12	19	19	11	15	19	18	11
shu	15	22	22	19	22	16	22	13
ybh	13	14	13	13	14	12	13	16

Maximum number of outputs (run19)

	dtc	eng	gug	hdn	max	pbv	shu	ybh
dtc	2336	128	42680	5376	1372	125	864	216
eng	9600	32	51830	7168	2744	32	144	3744
gug	50717	13671	80704	164289	43233	65625	1296	6552
hdn	9280	256	203454	14336	5488	480	1104	2856
max	34629	8000	42770	21504	12348	625	20736	2448
pbv	22848	13519	300118	113339	36298	12517	1152	5712
shu	10080	500	42682	24576	1372	24	96	8736
ybh	1140	20	392	224	80	96	16	2448

Overview

- Background
- Overview results
- Interactive exploration
- Course evals