

MT Extravaganza

Ling 567

March 12, 2015

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

iso	language name	language family	grammar authors
bcj	Bardi	Australian	Van Anh & Xue
ckt	Chukchi	Chkukotko-Kamchatkan	Dermer & Peterson
col	Columbian-Wenatchi	Salish	Gahler & Mills
eng	English	IE	
frr	Frisian	IE	Kilmer & Packard
khm	Khmer	Austro-Asiatic	Topping & Whelan
lit	Lithuanian	IE	Chan & LaBuda
quc	K'iche'	Mayan	Calderwood & Monaco
shu	Chadian Arabic	Afro-Asiatic	Heath & Marsh
tku	Upper Necaxa Totonac	Totonacan	Myhre & Vedder
yak (yak)	Yakima Sahaptin	Plateau Penutian	Lockwood
yak (zak)	Yakima Sahaptin	Plateau Penutian	Chewning

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

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 3rd year)

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

MMT with ACE (for the 3rd 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.These dogs sleep
- 5.Dogs eat
- 6.The dogs chase cars
- 7.The dogs dont chase cars
- 8.I think that you know that dogs chase cars
- 9.I ask whether you know that dogs chase cars
- 10.Cats and dogs chase cars
- 11.Dogs chase cars and cats chase dogs
- 12.Cats chase dogs and sleep
- 13.Do cats chase dogs
- 14.Hungry dogs eat
- 15.Dogs eat quickly
- 16.The dogs are hungry
- 17.The dogs are in the park
- 18.The dogs are the cats
- 19.Who sleeps
- 20.What do the dogs chase
- 21.What do you think the dogs chase
- 22.Who asked what the dogs chase
- 23.I asked what the dogs chased

Timing

- Submit jobs to translate 23 sentences from src2tgt to Condor
- Most run within a minute: e.g: shu2eng, eng2shu finish within 60 seconds of being submitted.
- Longest:
 - khm2ckt: 13 minutes
 - quc2tku: 15 minutes
 - khm2zak: 23 minutes

Parse success

bcj	ckt	col	eng	frr	khm	lit	quc	shu	tku	yak	zak
18	15	21	23	18	23	21	17	21	18	20	18

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

	bcj	ckt	col	eng	frr	khm	lit	quc	shu	tku	yak	zak
bcj	14	11	10	12	2	11	11	11	12	8	13	11
ckt	0	12	9	10	4	11	1	10	1	1	10	9
col	8	18	21	19	7	17	0	15	19	0	19	11
eng	10	20	20	23	8	19	17	11	22	10	21	13
frr	5	15	14	18	10	14	15	10	15	9	14	11
khm	5	20	21	19	4	19	17	16	21	7	20	12
lit	9	17	16	15	7	15	20	15	15	8	16	13
quc	4	11	15	12	7	14	2	15	15	7	14	10
shu	5	17	17	20	6	17	0	10	21	12	20	13
tku	4	8	8	7	1	8	8	9	9	9	9	7
yak	13	16	17	17	7	17	0	15	16	1	19	14
zak	8	13	14	12	5	14	0	13	13	3	15	16

(‘run4’)

Items with end-to-end output
(more semi.vpm fixes; added in transfer rules)

	bcj	ckt	col	eng	frr	khm	lit	quc	shu	tku	yak	zak
bcj	14	11	10	12	2	11	13	9	12	8	13	11
ckt	0	12	9	10	4	11	1	7	1	1	10	10
col	10	18	21	19	7	17	0	14	19	0	19	12
eng	12	21	21	23	8	20	22	16	22	10	21	15
frr	7	16	15	18	10	15	15	10	16	9	14	13
khm	6	20	21	19	4	19	22	13	21	7	20	13
lit	10	17	16	15	7	15	20	12	15	8	16	14
quc	5	11	15	12	7	14	2	15	15	7	14	11
shu	6	18	18	20	6	18	0	14	21	12	20	14
tku	4	8	8	7	1	8	8	7	9	9	9	7
yak	13	17	18	17	7	18	0	12	16	1	19	15
zak	10	13	14	12	5	14	0	12	13	3	15	16

(‘run7’)

Transfer rule propagation

[illegible]

Transfer rule example

```
pro-insert-arg1-mtr := monotonic_mtr &
[ INPUT.RELS < >,
  CONTEXT.RELS < [ ARG0.SF prop-or-ques,
    ARG1 #x & x ] >,
  FILTER.RELS < [ ARG0 #x ] >,
  OUTPUT [ RELS < [ PRED "pron_rel",
    ARG0 #x,
    LBL #larg ],
    [ PRED "exist_q_rel",
    ARG0 #x,
    RSTR #harg ] >,
  HCONS < qeq &
    [ HARG #harg,
    LARG #larg ] > ],
  FLAGS.EQUAL < #x > ].
```

Items with end-to-end output: Final
(more semi.vpm fixes, transfer rule propagation)

	bcj	ckt	col	eng	frr	khm	lit	quc	shu	tku	yak	zak
bcj	17	17	17	17	14	15	16	14	17	17	16	14
ckt	12	13	12	13	12	12	12	9	13	11	11	11
col	16	19	21	21	14	17	20	14	20	19	19	13
eng	17	21	22	23	16	20	22	18	23	21	21	15
frr	16	17	18	18	17	16	17	12	18	16	16	14
khm	16	20	22	22	16	19	22	14	22	20	21	13
lit	18	19	20	20	16	17	19	15	19	18	18	16
quc	14	11	15	15	13	14	15	15	15	14	14	11
shu	15	18	19	21	14	18	20	15	21	20	20	14
tku	15	14	14	15	14	14	15	12	15	18	15	13
yak	16	18	19	20	14	19	19	13	16	17	18	15
zak	14	14	16	16	12	15	15	13	14	15	14	16

(‘run16’)

Items with exact match output (run16)

	bcj	ckt	col	eng	frr	khm	lit	quc	shu	tku	yak	zak
bcj	13	11	13	16	10	15	1	9	11	15	14	10
ckt	7	12	9	1	7	10	0	8	6	8	8	7
col	8	11	21	19	1	15	1	11	19	13	13	11
eng	11	12	19	23	15	20	1	12	22	14	16	11
frr	9	10	15	17	16	15	0	10	16	14	12	10
khm	11	11	19	19	1	19	1	11	14	13	14	12
lit	13	11	15	18	12	17	1	10	11	16	14	12
quc	7	7	8	9	8	9	0	14	9	9	9	4
shu	10	10	18	21	13	17	1	10	21	13	0	0
tku	10	9	11	15	13	13	1	8	9	17	13	9
yak	11	13	15	19	6	18	1	10	0	0	16	14
zak	7	8	12	13	7	11	1	9	0	0	10	12

Average number of outputs (truncated) (run16)

	bcj	ckt	col	eng	frr	khm	lit	quc	shu	tku	yak	zak
bcj	4180	2451	114	25	2861	62	1588	7241	74	2385	6928	478
ckt	906	395	310	5	47	42	373	8	2	2541	72	97
col	459	9015	103	8390	761	68	4728	56	555	10060	18490	7009
eng	934	1752	46	5	235	10	604	39	19	11876	10023	25
frr	4126	5087	1126	174	6567	204	2819	31	180	24278	50155	2710
khm	5272	9640	7623	1490	754	135	19302	184	4761	8337	18386	6192
lit	13838	1881	221	184	10877	121	3125	15100	27	6093	21649	2668
quc	62486	16101	773	23	53	467	5500	104	13	28477	7965	2793
shu	449	1033	528	43	233	88	376	8	12	8785	10026	10
tku	11945	4525	309	29	2000	102	15755	318	6	3686	835	352
yak	1505	2287	107	157	71	65	4595	29	8	466	2451	194
zak	5310	378	52	52	87	9	498	67	6	229	104	713

Max number of outputs (run16)

	bcj	ckt	col	eng	frf	khm	lit	quc	shu	tku	yak	zak
bcj	53403	43200	2048	256	62208	588	15552	.17m	1536	49152	.14m	8208
ckt	18432	6912	6144	50	768	441	3528	36	18	54016	1152	1044
col	7968	.15m	864	.19m	15552	882	94338	288	6912	.11m	.21m	.15m
eng	18432	12288	360	32	5184	54	3528	640	192	73481	.13m	324
frf	73728	43200	12288	2580	.10m	2583	27168	144	2304	.31m	.57m	56957
khm	.10m	64794	.14m	12500	15552	1260	.10m	1584	.10m	58191	.17m	.11m
lit	.14m	19440	3072	3360	.25m	882	41472	.35m	384	78336	.44m	39516
quc	118m	.34m	13824	256	1024	7056	93312	2048	144	.55m	.17m	46656
shu	9216	6912	5832	500	5184	756	2592	128	128	77009	.13m	42
tku	.24m	98304	6144	480	31104	2160	.35m	6912	64	58368	12352	7648
yak	28064	27648	1536	2240	768	690	49968	162	55	5184	35328	3132
zak	.12m	2960	224	328	648	52	7056	732	48	1696	640	10920

Overview

- Background
- Overview results
- Interactive exploration
- Course evals