MT Extravaganza

Ling 567
June 1, 2017

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

abz	Abui	Trans-New Guinea	Emerson, Nguyen
eng	English	Indo-European	
ilo	Ilocano	Austronesian	Blackburn, Shintani
kaz	Kazakh	Turkic	Haeger, McMillan-Major
khr	Kharia	Austro-Asiatic	Sim, Woldenga-Racine
kkk	Kokota	Austronesian	Rebollo
flr	Kifuliiru	Niger-Congo	Clark, Watkins
shu	Chadian Arabic	Afro-Asiatic	Heath & Marsh
sje	Pite Saami	Uralic	Nielsen, Spivey
tur	Turkish	Altaic	Gokcen, Riggen

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

- 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 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 (‘run 18’)

- Submit jobs to translate 24 sentences from src2tgt to Condor: 100 lg pairs
- 29 finish within 5 seconds
- 72 finish within 10 seconds
- 10-14s: tur2flr, eng2kaz, ilo2shu, khr2flr, eng2khr, eng2kkk, flr2abz, kaz2abz, kkk2khr, eng2shu, flr2shu, flr2tur, kaz2tur, shu2flr, eng2flr, khr2khr, kkk2shu
- 15-19s: abz2shu, abz2eng, abz2khr, khr2tur
- 20-29s: sje2tur, abz2kaz, shu2khr
- 30-59s: abz2sje, abz2tur, sje2khr
- Winner! ilo2khr 2m23s

Timing: Winners! (2016)

- pbv2max 1:09
- hdn2gug 1:15
- gug2dtt 2:17
- pbv2gug 2:55
- pbv2dtt 4:40
- gug2hdn 5:40
- pbv2hdn 6:47
- pbv2eng 7:41
- hdn2gug (no pro-drop hack): 3:42:58

Parse success

abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
18	24	19	22	19	17	15	22	22	24

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
Total parses	49	25	24	35	22	95	18	24	26	29
Avg parses	2.72	1.04	1.26	1.59	1.16	5.59	1.20	1.09	1.18	1.21

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

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	16	14	14	14	15	12	13	14	13	12
eng	15	24	15	20	19	16	14	21	23	23
flr	10	10	19	9	11	8	8	12	11	12
ilo	11	18	12	22	14	11	13	17	18	18
kaz	11	14	10	13	19	10	12	12	13	13
khr	9	13	10	13	12	17	10	13	14	14
kkk	12	13	11	14	13	9	14	13	13	13
shu	14	22	13	17	17	14	14	22	20	20
sje	13	19	14	17	15	16	12	17	22	20
tur	14	23	15	17	18	20	14	20	24	24

(‘run11’ [19])

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

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	16	16	14	15	14	14	13	14	15	15
eng	17	24	18	22	19	20	14	22	24	24
flr	10	13	19	10	11	11	8	12	11	13
ilo	12	18	15	22	14	13	13	18	18	18
kaz	13	14	11	14	19	11	12	13	14	14
khr	11	13	12	13	12	17	10	14	14	14
kkk	13	14	12	14	13	12	14	14	13	13
shu	15	22	16	19	17	18	14	22	22	22
sje	14	19	16	17	15	16	12	17	22	20
tur	15	23	18	18	18	21	14	21	24	24

(‘run14’ [20])

Total number of outputs
(added in transfer rules, as provided)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	203	3964	61	43	195562	490	706	3642	667	69826
eng	74	78	43	27	92	2998	236	410	601	188
flr	60	137	51	45	141	4567	174	625	55	60
ilo	186	713	34	41	533	90239	687	1461	2058	14264
kaz	39	143	32	37	9940	429	234	121	357	111
khr	67	139	56	61	295	5910	297	299	295	666
kkk	122	1553	28	62	6274	84	11730	37791	1334	40
shu	72	83	32	24	114	10618	561	292	621	167
sje	70	134	49	292	1597	9590	320	3715	383	23242
tur	73	295	53	82	383	5815	861	1215	642	252

(‘run14’ [20])

Transfer rule example

```
hungry-add-emote-mtr := monotonic_mtr &
[ INPUT [ RELS < [ PRED "_hungry_a_rel",
           LBL #lbl,
           ARG0 #event,
           ARG1 #arg1 ] >,
      HCONS < > ],
  FILTER [ RELS < [ PRED "_emote_v_rel",
            ARG2 #harg ] >,
      HCONS < [ HARG #harg,
            LARG #lbl ] > ],
OUTPUT [ RELS < [PRED "_hungry_a_rel",
          LBL #lbl,
          ARG1 #arg1 ],
         [PRED "_emote_v_rel",
          ARG0 #event,
          ARG1 #arg1,
          ARG2 #harg ] >,
      HCONS < [HARG #harg,
           LARG #lbl ] > ],
FLAGS.EQUAL < #harg, #lbl > ].
```

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

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	16	16	14	15	14	14	13	14	15	15
eng	17	24	18	22	19	20	14	22	24	24
flr	10	13	19	10	11	11	8	12	11	13
ilo	12	18	15	22	14	13	13	18	18	18
kaz	13	14	11	14	19	11	12	13	14	14
khr	11	13	12	13	12	17	10	14	14	14
kkk	13	14	12	14	13	12	14	14	13	13
shu	15	22	16	19	17	18	14	22	22	22
sje	14	19	16	17	15	16	12	17	22	20
tur	15	23	18	18	18	21	14	21	24	24

(‘run14’ [20])

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

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	16	16	14	17	14	14	13	14	16	15
eng	17	24	18	22	19	20	14	22	24	24
flr	10	13	19	13	11	11	8	12	14	13
ilo	12	19	16	22	15	14	13	19	19	18
kaz	14	15	12	14	19	12	13	14	16	15
khr	14	17	15	16	15	17	10	15	17	17
kkk	13	14	12	14	13	12	14	14	14	13
shu	15	22	16	20	17	18	14	22	22	22
sje	15	20	16	17	16	16	12	18	22	20
tur	16	24	18	19	19	21	14	22	24	24

(‘run18’ [18])

Total number of outputs
(added in transfer rules, as provided)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	203	3964	61	43	195562	490	706	3642	667	69826
eng	74	78	43	27	92	2998	236	410	601	188
flr	60	137	51	45	141	4567	174	625	55	60
ilo	186	713	34	41	533	90239	687	1461	2058	14264
kaz	39	143	32	37	9940	429	234	121	357	111
khr	67	139	56	61	295	5910	297	299	295	666
kkk	122	1553	28	62	6274	84	11730	37791	1334	40
shu	72	83	32	24	114	10618	561	292	621	167
sje	70	134	49	292	1597	9590	320	3715	383	23242
tur	73	295	53	82	383	5815	861	1215	642	252

(‘run14’ [20])

Total number of outputs (transfer rule propagation)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	203	3964	61	116	195451	490	786	3642	20795	69826
eng	74	78	43	27	92	2998	236	410	601	188
flr	60	137	51	66	141	4567	174	625	973	60
ilo	186	721	38	41	535	90192	687	1464	2060	14264
kaz	56	145	39	37	9940	439	240	161	397	117
khr	85	155	64	73	305	5910	213	281	307	682
kkk	122	1553	28	66	6274	84	11730	37791	1766	40
shu	72	83	32	25	114	10618	561	292	621	167
sje	73	138	49	292	1598	9590	320	3719	383	23242
tur	76	299	53	92	384	5815	861	1219	666	252

(‘run18’ [18])

Items with exact match output (run18)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	6	13	6	1	9	2	9	12	14	0
eng	5	24	8	14	11	15	12	21	21	0
flr	4	13	9	8	3	8	6	9	13	0
ilo	4	18	8	16	8	10	12	18	14	0
kaz	4	15	5	11	15	10	12	13	14	0
khr	5	16	8	9	8	16	9	14	16	0
kkk	4	12	6	1	9	2	14	12	10	0
shu	4	20	8	14	10	12	12	22	17	0
sje	5	19	7	11	9	12	11	17	22	0
tur	4	22	8	13	11	15	13	19	19	0

Maximum number of outputs (run18)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	52	1288	16	24	183397	144	384	2052	18830	60123
eng	17	32	10	3	64	1200	81	192	216	32
flr	17	64	16	16	81	4320	108	512	864	16
ilo	40	500	8	9	296	45622	324	768	1296	7200
kaz	17	102	12	8	6048	192	81	44	134	28
khr	12	56	16	12	210	3072	36	64	96	512
kkk	60	1264	8	16	5760	46	11008	36288	1024	12
shu	17	32	4	3	64	4832	324	128	216	32
sje	17	24	10	72	486	3456	126	3072	216	18240
tur	20	128	10	16	108	2400	324	512	216	64

Total number of outputs (transfer rule propagation)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	203	3964	61	116	195451	490	786	3642	20795	69826
eng	74	78	43	27	92	2998	236	410	601	188
flr	60	137	51	66	141	4567	174	625	973	60
ilo	186	721	38	41	535	90192	687	1464	2060	14264
kaz	56	145	39	37	9940	439	240	161	397	117
khr	85	155	64	73	305	5910	213	281	307	682
kkk	122	1553	28	66	6274	84	11730	37791	1766	40
shu	72	83	32	25	114	10618	561	292	621	167
sje	73	138	49	292	1598	9590	320	3719	383	23242
tur	76	299	53	92	384	5815	861	1219	666	252

(‘run18’ [18])

Item 16: ambiguity in abz

It is the dogs and is in parks
She is the dogs and is in parks
We are the dogs and are in parks
They were the dogs and were in parks
I was the dogs and was in parks
They are the dogs and were in parks
She was the dogs and was in parks
It was the dogs and was in parks
They are the dog and were in parks
It was the dogs and is in parks
We were the dogs and are in parks
It is the dogs and was in parks
It was the dog and was in parks
We are the dogs and were in parks
We were the dog and are in parks
He was the dogs and was in parks
We were the dogs and are in a park
They were the dogs and are in parks
We are the dog and are in parks
We were the dogs and were in parks
We are the dogs and are in a park
We were the dog and were in parks
We are the dog and are in a park
We were the dogs and were in a park
We are the dogs and are in the parks
We were the dog and were in a park
We are the dog and are in the parks
We were the dogs and were in the park
We are the dogs and are in the park
We were the dog and were in the park

I am the dog and am in a park
You are the dog and are in the parks
You are the dog and were in a park
It was the dog and is in a park
They were the dog and were in the park
He was the dogs and was in a park
He is the dogs and is in a park
They were the dogs and were in the parks
He was the dogs and is in the park
It was the dogs and is in the parks
We are the dog and were in a park
He was the dog and is in the park
We are the dogs and were in the park
You were the dogs and are in a park
I was the dog and am in a park
You are the dogs and are in the park
You are the dogs and were in the park
They were the dog and were in the parks
You are the dog and are in the park
He was the dog and was in a park
She is the dogs and was in the park
I was the dogs and was in the park
She is the dog and was in the park
I was the dog and was in the park
She was the dogs and is in a park
You are the dog and were in the park
He was the dogs and was in the park
She was the dog and is in a park
You are the dogs and were in the parks
They were the dog and are in the park

We are in the dogs and the park
You are in the dog and parks
We are in the dog and parks
You are in the dogs and a park
We are in the dogs and a park
You are in the dog and a park
We are in the dog and a park
You are in the dog and the parks
We are in the dog and the parks
He is in the dogs and parks
She is in the dogs and the parks
He is in the dogs and the parks
She is in the dogs and a park
He is in the dogs and a park
She is in the dog and the parks
He is in the dog and the parks
She is in the dog and the park
He is in the dog and the park
She is in the dogs and the park
He is in the dogs and the park
She is in the dog and parks
He is in the dog and parks
She is in the dog and a park
He is in the dog and a park
They are in the dogs and parks
It is in the dogs and a park
They are in the dogs and the parks
It is in the dog and the parks
They are in the dog and the park
It is in the dog and the park
They are in the dogs and the park

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 < > ]].
```

Total number of outputs (transfer rule propagation)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	203	3964	61	116	195451	490	786	3642	20795	69826
eng	74	78	43	27	92	2998	236	410	601	188
flr	60	137	51	66	141	4567	174	625	973	60
ilo	186	721	38	41	535	90192	687	1464	2060	14264
kaz	56	145	39	37	9940	439	240	161	397	117
khr	85	155	64	73	305	5910	213	281	307	682
kkk	122	1553	28	66	6274	84	11730	37791	1766	40
shu	72	83	32	25	114	10618	561	292	621	167
sje	73	138	49	292	1598	9590	320	3719	383	23242
tur	76	299	53	92	384	5815	861	1219	666	252

(‘run18’ [18])

Total number of outputs (be_v_id hack)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	203	884	61	89	195579	490	786	3642	949	69826
eng	74	78	43	27	92	2998	236	410	601	188
flr	60	137	51	66	141	4567	174	625	973	60
ilo	186	721	38	41	535	90201	687	1464	2060	14264
kaz	56	145	39	37	9940	439	240	161	397	117
khr	85	155	64	73	305	5910	213	281	307	682
kkk	122	1553	28	66	6274	84	11730	37791	1766	40
shu	72	83	32	25	114	10618	561	292	621	167
sje	73	138	49	292	1598	9590	320	3719	383	23242
tur	76	299	53	92	384	5815	861	1219	666	252

(‘run21’ [21])

Maximum number of outputs (run18)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	52	1288	16	24	183397	144	384	2052	18830	60123
eng	17	32	10	3	64	1200	81	192	216	32
flr	17	64	16	16	81	4320	108	512	864	16
ilo	40	500	8	9	296	45622	324	768	1296	7200
kaz	17	102	12	8	6048	192	81	44	134	28
khr	12	56	16	12	210	3072	36	64	96	512
kkk	60	1264	8	16	5760	46	11008	36288	1024	12
shu	17	32	4	3	64	4832	324	128	216	32
sje	17	24	10	72	486	3456	126	3072	216	18240
tur	20	128	10	16	108	2400	324	512	216	64

Maximum number of outputs (run21)

	abz	eng	flr	ilo	kaz	khr	kkk	shu	sje	tur
abz	52	256	16	20	183525	144	384	2052	192	60123
eng	17	32	10	3	64	1200	81	192	216	32
flr	17	64	16	16	81	4320	108	512	864	16
ilo	40	500	8	9	296	45631	324	768	1296	7200
kaz	17	102	12	8	6048	192	81	44	134	28
khr	12	56	16	12	210	3072	36	64	96	512
kkk	60	1264	8	16	5760	46	11008	36288	1024	12
shu	17	32	4	3	64	4832	324	128	216	32
sje	17	24	10	72	486	3456	126	3072	216	18240
tur	20	128	10	16	108	2400	324	512	216	64

Overview

- Background
- Overview results
- Interactive exploration
- Course evals