MT Extravaganza

Ling 567
March 13, 2014

Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

iso	language name	language family	grammar authors
bcj	Bardi	Australian	Ho
cox	Nanti	Arawakan	Inman & Morrison
eng	English	IE	
frr	Frisian	IE	Kilmer & Packard
glk	Gilaki	IE	Antoniak & Lockwood
mar	Marathi	IE	Baer & Bothale
mni	Meithei	Sino-Tibetan	Peters (& Wong)
mya	Burmese	Sino-Tibetan	Woo (& Wibel)
nya (nya)	Chichewa	Niger-Congo	Bhalla
nya (nyb)	Chichewa	Niger-Congo	Hrynkevich
uig	Uyghur	Altaic	Sugar
yux	Yukaghir	Uralic-Yukaghir(?)	Troelsen & Zamaraeva

Languages

iso	language name	language family	grammar authors
bcj	Bardi	Australian	Ho
cox	Nanti	Arawakan	Inman & Morrison
eng	English	IE	
frr	Frisian	IE	Kilmer & Packard
glk	Gilaki	IE	Antoniak & Lockwood
mar	Marathi	IE	Baer & Bothale
mni	Meithei	Sino-Tibetan	Peters (& Wong)
mya	Burmese	Sino-Tibetan	Woo (& Wibel)
nya (nya)	Chichewa	Niger-Congo	Bhalla
nya (nyb)	Chichewa	Niger-Congo	Hrynkevich
uig	Uyghur	Altaic	Sugar
yux	Yukaghir	Uralic-Yukaghir(?)	Troelsen & Zamaraeva

Languages - mapped

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

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
- Information structure

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 2nd 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 2nd 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. I can eat glass
7. It doesn't hurt me
8. The dogs chase cars
9. I think that you know that dogs chase cars
10. I ask whether you know that dogs chase cars
11. Cats and dogs chase cars
12. Dogs chase cars and cats chase dogs
13. Cats chase dogs and sleep
14. Do cats chase dogs?
15. Hungry dogs eat
16. Dogs eat quickly
17. The dogs are hungry
18. The dogs are in the park
19. The dogs are the cats
20. ~~The dogs-FP chase the cats~~
21. ~~The dogs chase-FP the cats~~
22. ~~The dogs chase the cats-FP~~

Parse success

bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux
18	16	19	19	19	17	12	17	17	16	15	15

Items with end-to-end output
(post vpm fixes, pre transfer rule propagation)

src \ tgt	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux	avg
bcj	18	12	7	7	13	12	12	11	12	12	11	10	11.42
cox	10	16	10	8	13	11	10	10	2	10	12	10	10.17
eng	10	16	19	18	18	16	13	18	12	12	16	17	15.42
frr	11	15	18	19	17	16	14	16	12	12	16	16	15.17
glk	11	15	18	16	19	16	14	17	13	12	15	16	15.17
mar	4	8	10	10	10	17	7	10	6	6	9	10	8.92
mni	8	11	12	12	12	12	12	11	10	10	11	11	11.00
mya	8	13	13	11	15	13	11	17	9	9	14	16	12.42
nya	10	14	12	12	14	13	13	14	13	14	13	14	13.00
nyb	11	16	12	12	15	15	13	15	13	15	15	14	13.83
uig	10	14	15	14	15	15	13	15	11	11	15	14	13.50
yux	8	11	12	12	11	11	11	11	8	8	12	15	10.83
avg	9.92	13.42	13.17	12.58	14.33	13.92	11.92	13.75	10.08	10.92	13.25	13.58	

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

src \ tgt	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux	avg
bcj	18	12	7	7	7	6	7	7	12	12	11	6	9.33
cox	10	16	12	9	12	9	8	8	2	10	14	8	9.83
eng	10	16	19	18	18	16	13	18	12	12	16	17	15.42
frr	11	15	18	19	17	16	14	16	12	12	16	16	15.17
glk	11	15	18	16	19	16	14	17	13	12	15	16	15.17
mar	11	14	17	17	15	17	12	15	12	12	15	15	14.33
mni	8	11	12	12	12	12	12	11	10	10	11	11	11.00
mya	8	13	13	11	12	10	8	13	9	9	14	13	11.08
nya	10	14	12	12	12	11	11	11	13	14	13	10	11.92
nyb	11	16	12	12	12	12	11	11	13	15	15	10	12.50
uig	10	14	15	14	15	15	13	15	11	11	15	14	13.50
yux	8	13	15	15	14	13	12	14	9	9	14	15	12.58
avg	10.50	14.08	14.17	13.50	13.75	12.75	11.25	13.00	10.67	11.50	14.08	12.58	

Transfer rule propagation

```
chase := chase-mtr. ; except mar
feel-think := feel-think-mtr. ; except mar
hungry-v-to-a := hungry-v-to-a-mtr. ; excpet cox, yux
hungry-sv := hungry-sv-to-a-mtr. ; except cox, yux
hungry-sv-v := hungry-sv-to-v-mtr. ; yux only
hungry-v-sv := hungry-v-to-sv-mtr. ; cox only
be-located := be-located-mtr. ; except yux
catch-pain-inv := catch-pain-inv-mtr. ; except bcj, frr
make-harm-catch-pain := make-harm-catch-pain-mtr. ; bcj only
catch-pain-make-harm := catch-pain-make-harm-mtr. ; frr only
```

Items with exact match output
(post vpm/iso.txt fixes, pre transfer rule propagation)

	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux
bcj	12	5	5	4	7	7	6	6	8	9	7	6
cox	5	8	9	0	8	6	7	5	2	8	5	7
eng	5	8	19	17	9	15	12	17	10	11	14	15
frr	5	8	17	18	14	14	11	15	10	10	14	13
glk	4	8	10	14	16	14	12	15	9	10	8	10
mar	3	5	10	10	8	16	7	10	5	5	8	9
mni	6	6	12	11	11	11	12	10	8	9	10	10
mya	4	7	5	3	12	12	10	17	8	8	12	13
nya	6	6	12	11	12	10	9	10	12	13	9	9
nyb	6	7	12	11	12	11	9	10	11	15	10	9
uig	7	8	15	14	8	14	11	14	10	10	14	13
yux	4	6	2	10	5	9	9	9	7	6	9	15

Total number of outputs (post transfer rule propagation)

	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux
bcj	12	5	5	4	5	5	4	4	8	9	7	5
cox	5	8	11	0	9	7	7	6	2	8	6	8
eng	5	8	19	17	9	15	12	17	10	11	14	15
frr	5	8	17	18	14	14	11	15	10	10	14	13
glk	4	8	10	14	16	14	12	15	9	10	8	10
mar	7	8	17	16	13	16	11	14	10	11	14	14
mni	6	6	12	11	11	11	12	10	8	9	10	10
mya	4	7	5	3	9	9	7	13	8	8	12	11
nya	6	6	12	11	12	10	9	10	12	13	9	9
nyb	6	7	12	11	12	11	9	10	11	15	10	9
uig	7	8	15	14	8	14	11	14	10	10	14	13
yux	4	8	2	13	8	11	10	12	7	7	11	15

Average number of outputs (truncated) (post transfer rule propagation)

	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux
bcj	589946	145770	1612	3760	45712	0	22	837	14381	15070	415	0
cox	5567	2436	1577	52	2284	2	18	1	14	47	7	4
eng	6227	3109	26	286	388	1	3	31	296	23	93	229
frr	12516	8331	27	7877	5738	9	8	49	207	136	369	118
glk	2956	6963	784	39	506	2	3	15	67	34	2	33
mar	3088	2273	3170	67	82	3	2	13	75	36	4	1
mni	18024	18915	1238	195	2170	4	6	9	163	68	4	2
mya	1838	1824	2760	842	1088	2	4	35	410	62	494	305
nya	65807	49698	2537	736	2384	12	21	10	940	641	90	3
nyb	17423	4010	698	87	1337	1	3	3	96	76	8	1
uig	6049	1718	705	21	97	7	3	12	187	141	4	2
yux	606	9046	704	559	36	1	2	10	12	4	209	39

Max number of outputs
(post transfer rule propagation)

	bcj	cox	eng	frr	glk	mar	mni	mya	nya	nyb	uig	yux
bcj	7.3m	2.4m	30360	70720	.86m	4	392	15616	.26m	.26m	7680	4
cox	65013	30720	29084	768	34074	32	256	14	182	768	32	16
eng	60476	30720	448	5184	4096	6	16	256	4356	243	1728	2880
frr	.12m	82944	312	99574	98062	100	64	768	1936	1052	6912	2160
glk	31104	99919	13165	384	6400	8	16	96	484	264	8	600
mar	31104	30720	60136	768	798	16	8	104	484	262	12	8
mni	.24m	.34m	22345	3072	38944	50	64	96	1936	972	16	8
mya	31752	30626	43724	15552	14000	12	32	256	6732	1024	9216	2880
nya	.93m	.47m	47699	12288	39200	150	256	56	7744	8736	1536	24
nyb	.26m	34560	13165	768	19477	6	20	20	704	546	96	6
uig	.11m	19456	13165	138	800	100	12	96	1944	2402	32	4
yux	10584	.17m	13165	10368	200	14	4	96	132	48	3936	720

Overview

- Background
- Overview results
- Interactive exploration
- Course evals