

# MT Extravaganza

---

Ling 567

June 5, 2013

# Overview

---

- Background
- Overview results
- Interactive exploration
- Course evals

# Languages

---

iso	language name	language family	grammar authors
aaq(-pen)	Penobscot	Algic	Bolla & Trimble
eng	English	IE	
frr	Frisian	IE	Kilmer & Packard
hur	Halkomelem	Salish	Benak & Nichols
ita	Italian	IE	
lkt	Lakota	Siouan	Curtis & McHugh
lzh	Classical Chinese	Sino-Tibetan	Feng
mvi	Miyako	Japonic	Ledyard & Sato
ydd	Yiddish	IE	Chopra & Letcher

# 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

# New this year: MMT with ACE

---

- Faster system run times (though final run took 1h7m)
- More coverage (fewer system timeouts)
- Compatible with Condor (yay!)
  - => Possibility of parallelization (though not explored)
- Possibility of respecting ICONS representation of information structure

# New this year: MMT with ACE

---

- Faster system run times (though final run took 1h7m)
- More coverage (fewer system timeouts)
- Compatible with Condor (yay!)
  - => Possibility of parallelization (though not explored)
- 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

---

aaq	eng	frr	hur	ita	lkt	lzh	mvi	ydd
7	19	19	17	16	15	19	14	18

# Items with end-to-end output

---

# Items with end-to-end output

---

	aaq	eng	frr	hur	ita	lkt	lzh	mvi	ydd
aaq	6	6	6	6	6	5	7	6	5
eng	9	12	14	10	10	13	14	11	11
frr	8	11	16	8	10	15	12	10	12
hur	1	4	4	9	5	5	7	5	1
ita	8	10	11	7	11	10	9	8	9
lkt	7	9	12	9	9	12	11	8	9
lzh	8	10	11	12	9	11	14	10	10
mvi	9	11	10	9	9	9	11	12	8
ydd	8	11	14	8	9	13	12	10	13

# Items with exact match output

---

	aaq	eng	frr	hur	ita	lkt	lzh	mvi	ydd
aaq	0	0	0	11	10	0	13	0	0
eng	0	0	0	5	10	0	17	0	0
frr	0	0	0	6	10	0	12	0	0
hur	0	0	0	11	2	0	1	0	0
ita	0	0	0	4	14	0	6	0	0
lkt	0	0	0	3	8	0	8	0	0
lzh	0	0	0	6	9	0	19	0	0
mvi	0	0	0	6	11	0	15	0	0
ydd	0	0	0	4	8	0	12	0	0

# Total number of outputs

---

# Total number of outputs

---

	aaq	eng	frr	hur	ita	lkt	lzh	mvi	ydd
aaq	136466	68	224	408	240	65	76	66	4901
eng	92962	15	5305	3264	23	55	828	1238	522
frr	86943	82	231862	204	129	3210	148	1139	157
hur	4992	4	6	180	5	21	25	92	6
ita	83836	31	5255	3200	25	25	13	33	35190
lkt	61081	34	109	92	22	28	15	30	342
lzh	69013	334	579	588	360	427	494	1168	1180
mvi	71625	123	401	672	119	260	208	475	14476
ydd	98309	21	10527	78	15	65	138	1136	60

# Average number of outputs

---

	aaq	eng	frr	hur	ita	lkt	lzh	mvi	ydd
aaq	7182.42	3.58	11.79	21.47	12.63	3.42	4.0	3.47	257.95
eng	4892.74	0.79	279.21	171.79	1.21	2.89	43.58	65.16	27.47
frr	4575.95	4.32	12203.26	10.74	6.79	168.95	7.79	59.95	8.26
hur	262.74	0.21	0.32	9.47	0.26	1.11	1.32	4.84	0.32
ita	4412.42	1.63	276.58	168.42	1.32	1.32	0.68	1.74	1852.11
lkt	3214.79	1.79	5.74	4.84	1.16	1.47	0.79	1.58	18.0
lzh	3632.26	17.58	30.47	30.95	18.95	22.47	26.0	61.47	62.11
mvi	3769.74	6.47	21.11	35.37	6.26	13.68	10.95	25.0	761.89
ydd	5174.16	1.11	554.05	4.11	0.79	3.42	7.26	59.79	3.16

# Max number of outputs

---

	aaq	eng	frr	hur	ita	lkt	lzh	mvi	ydd
aaq	33820	21	90	234	180	35	42	54	3920
eng	18744	2	5184	1024	10	16	682	1120	216
frr	18744	48	184769	56	36	2304	110	1120	48
hur	4992	1	2	84	1	5	10	72	6
ita	26443	8	5184	1024	8	8	5	8	22932
lkt	18744	16	24	32	10	6	4	8	216
lzh	26443	98	150	240	240	144	330	1120	384
mvi	26443	48	240	240	63	175	110	385	9786
ydd	18744	4	10368	16	4	16	110	1120	12



# Overview

---

- Background
- Overview results
- Interactive exploration
- Course evals