

MMT Extravaganza

Ling 567

March 9, 2023

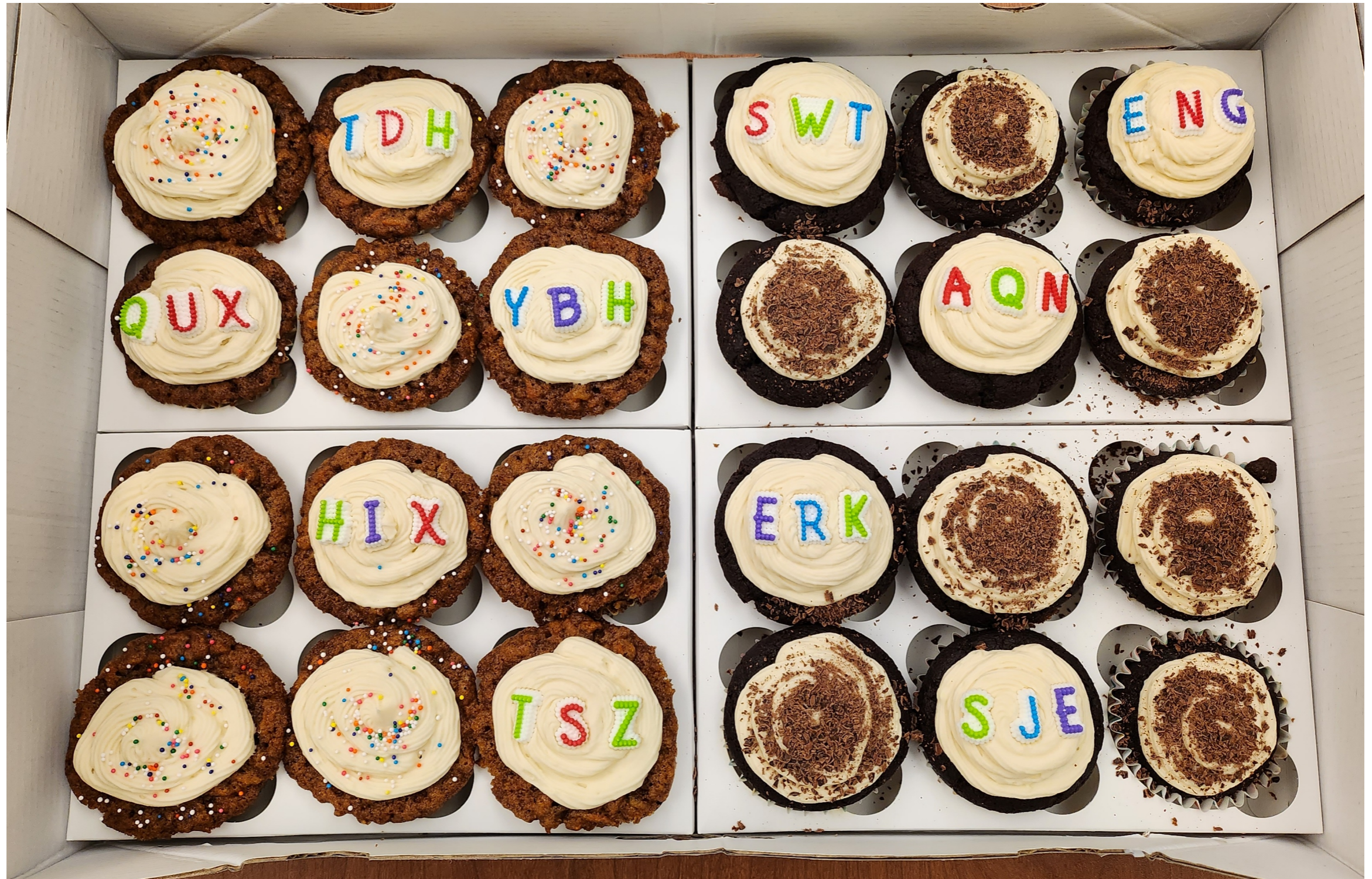
Overview

- Background
- Overview results
- Interactive exploration
- Course evals

Languages

aqn	Northern Alta	Austronesian	Anaya & Davenport
erk	South Efate	Austronesian	Koh & Saechao
hix	Hixkaryana	Cariban	Ruditsky & Wang
qux	Yauyos Quecha	Quechuan	Skinner
swt	Sawila	Timor-Alor-Pantar	Baumgartner-Zhang & Bi
tdh	Thulung	Sino-Tibetan	Fang & Martins
tsz	Purepecha	Tarascan	Haberland & Patil
ybh	Yakkha	Sino-Tibetan	Dhungana
eng	English	Indo-European	
sje	Pite Saami	Uralic	Nielsen & Spivey

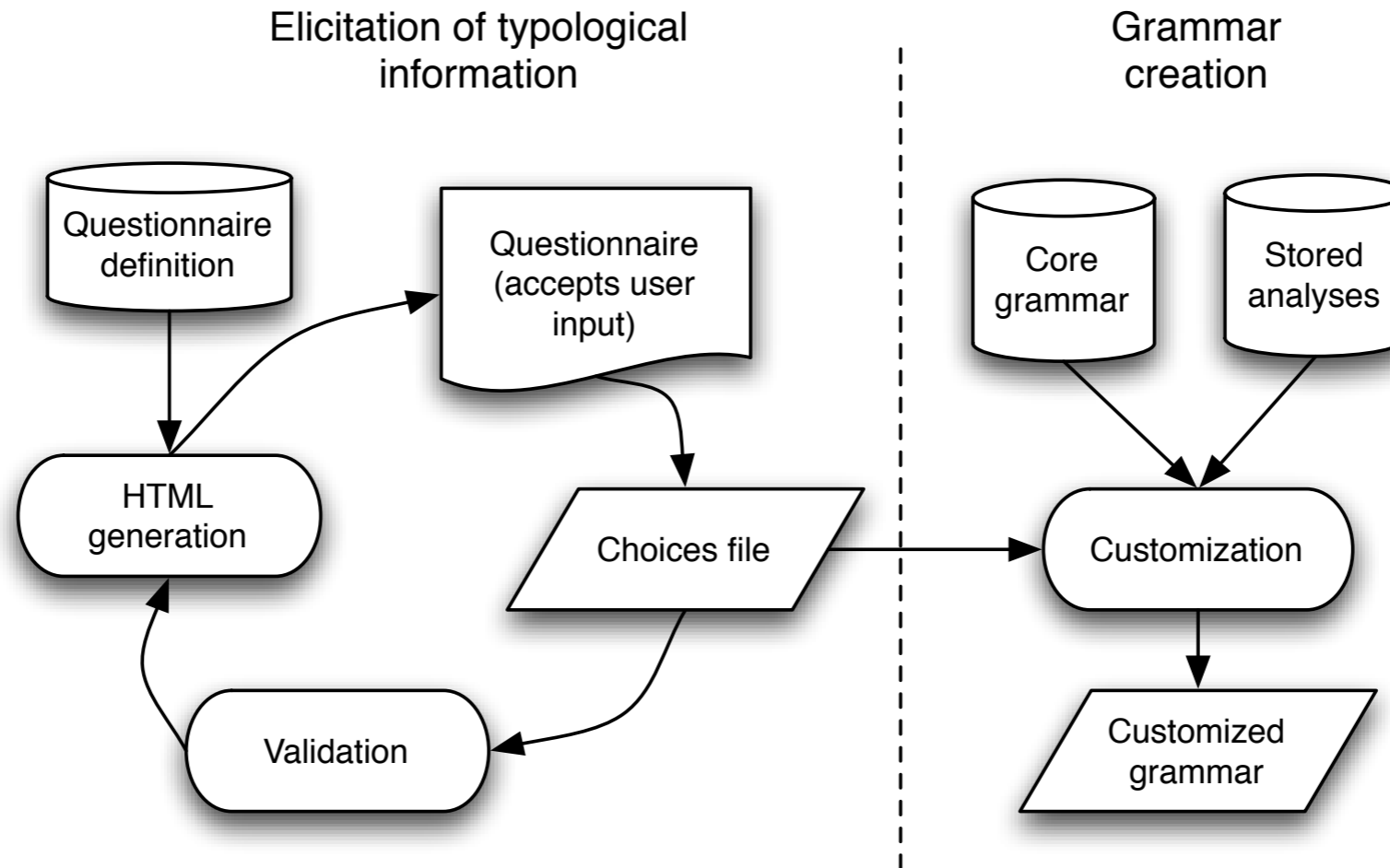
Languages - cupcaked (Thank you Lindsay!)



Grammar coverage (mostly shared)

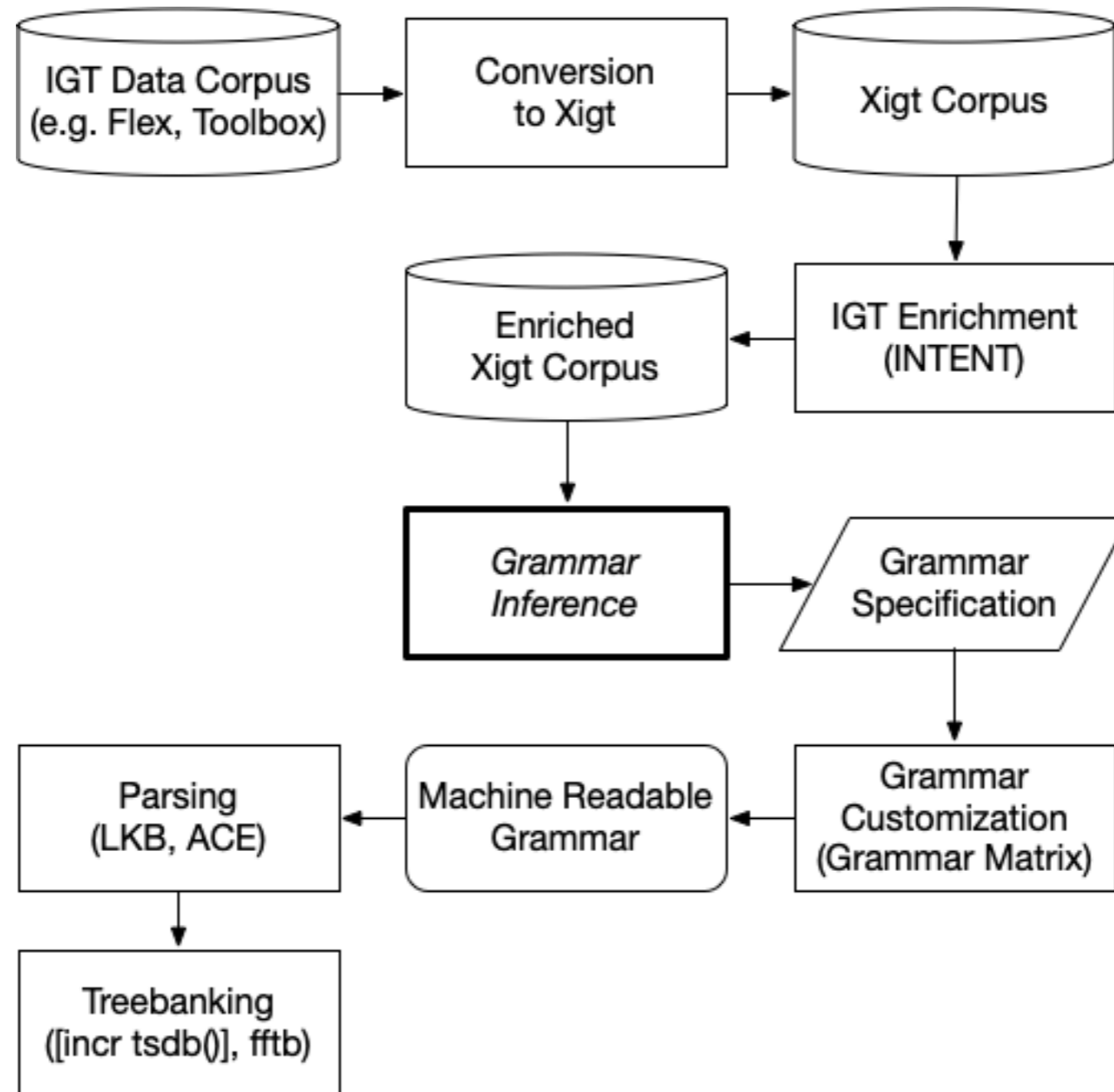
- Basic word order
- Case
- Agreement
- Personal pronouns
- Tense/aspect
- Sentential negation
- (Argument optionality)
- Matrix yes-no questions
- Coordination
- Modification (adjective, clausal mods)
- Clausal complements
- Wh questions
- Possessives

Grammar Development: Customization + Extension



(Bender et al 2010)

Starting point: Automatically inferred grammar specifications



(Howell 2020, Conrad 2021, Dods 2022)

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

Input sentences

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

Parse success

aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
15	26	12	17	7	18	10	12	17	6

Items with end-to-end output
(grammar, semi.vpm & transfer rules as provided)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	15	9	9	9	9	9	8	7	6	3
eng	15	26	16	19	13	21	13	15	13	9
erk	9	9	12	7	8	10	9	7	0	2
hix	10	12	1	17	10	11	11	9	0	1
qux	4	4	4	4	7	4	3	4	0	3
sje	12	16	13	14	10	18	10	11	9	8
swt	7	8	0	0	7	9	10	0	0	0
tdh	11	9	0	8	10	11	8	12	0	0
tsz	10	9	0	7	9	12	8	12	15	0
ybh	5	6	6	6	5	6	6	5	0	6

Items with end-to-end output
(grammar, semi.vpm & transfer rules as provided)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	15	9	9	9	9	9	8	7	6	3
eng	15	26	16	19	13	21	13	15	13	9

aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
15	26	12	17	7	18	10	12	17	6

Items with end-to-end output
(grammar, semi.vpm & transfer rules as provided)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	15	9	9	9	9	9	8	7	6	3
eng	15	26	16	19	13	21	13	15	13	9
erk	9	9	12	7	8	10	9	7	0	2
hix	10	12	1	17	10	11	11	9	0	1
qux	4	4	4	4	7	4	3	4	0	3
sje	12	16	13	14	10	18	10	11	9	8
swt	7	8	0	0	7	9	10	0	0	0
tdh	11	9	0	8	10	11	8	12	0	0
tsz	10	9	0	7	9	12	8	12	15	0
ybh	5	6	6	6	5	6	6	5	0	6

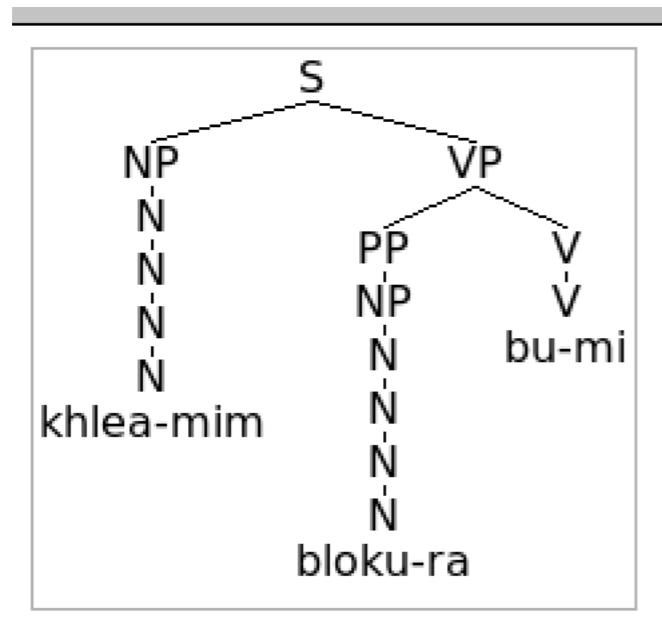
Items with end-to-end output
(semi.vpm updates to tsz, erk, ybh, hix, tdh)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	15	9	10	9	9	9	8	7	6	3
eng	15	26	19	19	13	21	13	15	13	9
erk	10	9	12	9	8	11	9	7	6	2
hix	10	12	11	17	10	11	11	9	6	4
qux	4	4	4	4	7	4	3	4	3	3
sje	12	16	15	14	10	18	10	11	9	8
swt	7	8	9	9	7	9	10	7	5	3
tdh	11	9	10	8	10	11	8	12	6	3
tsz	10	9	10	7	9	12	8	12	15	3
ybh	5	6	6	6	5	6	6	5	5	6

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

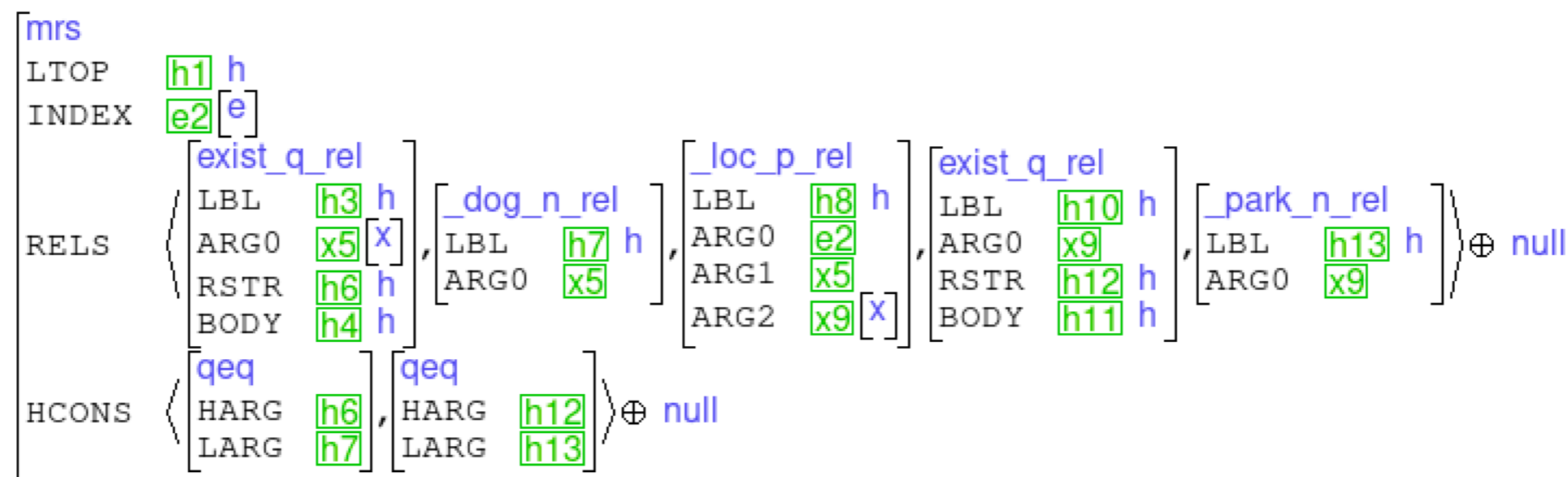
	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	15	10	10	9	10	9	8	8	7	3
eng	15	26	19	19	13	21	13	15	13	9
erk	11	11	12	10	10	11	10	9	7	2
hix	14	16	13	17	11	15	12	10	7	4
qux	6	6	6	6	7	5	5	5	4	3
sje	12	16	15	14	10	18	10	11	9	8
swt	8	10	9	10	8	9	10	8	6	3
tdh	12	10	11	9	10	11	9	12	6	3
tsz	11	12	11	8	9	12	9	12	15	3
ybh	5	6	6	6	5	6	6	5	5	6

Need for transfer rules



khlea-mim bloku-ra bu-mi
 dog-PL.NOM river-SG.LOC be-3PL
 dogs are in the river (tdh)

NB: invented example!



Transfer rule example

```
in-to-loc-sit-mtr := monotonic_mtr &
[ INPUT [ RELS < [ PRED "_in_p_rel",
    LBL #lbl,
    ARG0 #event,
    ARG1 #sitter,
    ARG2 #location ] > ],
  OUTPUT [ RELS < [ PRED "_loc_p_rel",
    LBL #lbl,
    ARG1 #event,
    ARG2 #location],
    [ PRED "_sit.down_v_rel",
    LBL #lbl,
    ARG0 #event,
    ARG1 #sitter] > ]].
```

Transfer rule example

```
pronoun-delete-mtr := monotonic_omtr &
  [ 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 (grammars & transfer rules as provided)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	54472	2639	454	75	2068	1782	67	3312	36237	54
eng	16200	101	616	87	43	587	75	54	4278	352
erk	3216	1263	254	44	36	580	100	20	654	28
hix	384	218	86	71	34	82	61	37	460	138
qux	168	46	31	20	40	34	24	30	60	36
sje	16176	101	342	52	45	361	74	39	8231	346
swt	1664	252	855	53	316	104	227	372	660	26
tdh	13504	87	193	31	54	100	53	47	3750	58
tsz	14152	110	180	23	39	108	50	82	16228	84
ybh	232	32	38	26	22	32	28	9	270	480

Total number of outputs (transfer rule propagation)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	54471	2259	454	59	2072	1786	67	3313	36528	54
eng	16200	93	616	87	43	587	75	54	4278	352
erk	3248	1283	254	38	44	584	102	25	664	28
hix	17544	4158	280	71	38	2728	63	38	470	138
qux	208	56	37	24	40	40	28	31	70	36
sje	16176	101	342	52	45	361	74	39	8231	346
swt	1696	292	859	47	324	118	227	392	680	26
tdh	13512	95	197	33	54	100	55	47	3750	58
tsz	14168	218	188	27	39	108	54	82	16228	84
ybh	232	32	38	26	22	32	28	9	270	480

Maximum number of outputs (transfer rule propagation)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	49982	1728	352	20	1840	864	32	3040	35254	24
eng	12288	32	144	16	8	216	16	18	3600	144
erk	1536	864	64	10	8	432	32	4	300	16
hix	11264	3456	192	16	4	1728	16	16	180	120
qux	64	30	15	8	18	24	16	24	40	24
sje	12288	24	96	10	16	216	16	14	4487	144
swt	988	96	486	28	228	24	48	170	180	24
tdh	12288	24	64	10	12	36	12	15	3600	24
tsz	12288	48	64	10	8	36	16	31	14400	36
ybh	64	16	8	16	8	12	12	2	120	216

Items with exact match output (transfer rule propagation)

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	10	10	1	9	5	6	7	6	3	1
eng	7	26	2	16	6	18	10	8	4	1
erk	6	10	2	9	6	9	7	7	2	0
hix	8	15	1	16	6	12	10	6	3	0
qux	3	5	0	6	7	3	4	3	1	0
sje	6	15	2	11	4	18	7	7	8	0
swt	4	10	1	10	4	7	10	6	2	0
tdh	4	10	2	9	5	8	8	9	2	1
tsz	4	11	0	7	5	9	6	7	15	1
ybh	3	5	0	5	2	4	5	4	2	1

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

	aqn	eng	erk	hix	qux	sje	swt	tdh	tsz	ybh
aqn	15	10	10	9	10	9	8	8	7	3
eng	15	26	19	19	13	21	13	15	13	9
erk	11	11	12	10	10	11	10	9	7	2
hix	14	16	13	17	11	15	12	10	7	4
qux	6	6	6	6	7	5	5	5	4	3
sje	12	16	15	14	10	18	10	11	9	8
swt	8	10	9	10	8	9	10	8	6	3
tdh	12	10	11	9	10	11	9	12	6	3
tsz	11	12	11	8	9	12	9	12	15	3
ybh	5	6	6	6	5	6	6	5	5	6

Overview

- Background
- Overview results
- Interactive exploration
- Course evals