The Matrix: Future Directions
Wrap up

Ling 567
March 9, 2020

Overview

- Wrap up/reflections
- Matrix: Future directions
- AGGREGATION

Goals: Of Grammar Engineering

- Build useful, usable resources
- Test linguistic hypotheses
- Represent grammaticality/minimize ambiguity
- Build modular systems: maintenance, reuse

Goals: Of this course

- Mastery of tfs formalism
- Hands-on experience with grammar engineering
- A different perspective on natural language syntax
- Practice building (and debugging!) extensible system
- Contribute to on-going research in multilingual grammar engineering

Big take-aways this term for me

- Automatically inferred choices files are still too noisy
- eng2yaq presents a case where the VPM framework possibly breaks down
- append-lists for long distance dependencies don’t get along (as implemented) with coordination
- … lots of specific bugs noted!

Reflections

- Where have the analyses provided by the Matrix (or suggested by the labs) seemed like a good fit?
- Where have they been awkward?
- What have you learned in this class about syntax?
- ... about knowledge engineering for NLP?
- ... about computational linguistics in general?
- ... about linguistics in general?
- What did working with a test corpus show you about the process of scaling to real-world text?

Feedback: Individual projects

- Was the workload reasonable?
- How would you have taken advantage of working with a partner?
- What was better about working alone?

More reflections

- Semantic representations are important
    - It’s easier to work on them if they serve as an interface to something
- Analyses of phenomena interact
    - The more streamlined/motivated the analysis of each phenomenon is, the smoother the interactions
    - What interactions did you encounter?

More reflections: model and modeling domain

- From 566: Distinction between the model (HPSG grammar fragment) and the modeling domain (there: English).

- How did this play out in 567?

Future directions overview

- More libraries (and semantic harmonization)
- How this class might evolve
- AGGREGATION

More libraries

- In progress: Wh questions
- Next up?
  - Pronouns, demonstratives, COG-ST
  - Other non-verbal predicates
  - Other intersective modifiers
  - Numeral classifiers
  - More verb subcategorization
  - Information structure in wh questions
  - Information structure in free word order languages

Creating a library for the customization system

- Choose phenomenon
- Review typological literature on phenomenon
- Refine definition of phenomenon
- Conceptualize range of variation within phenomenon
- Review HPSG (& broader syntactic) literature on phenomenon
- Pin down target MRSs
- Develop HPSG analyses for each variant
- Implement analyses in tdl
- Develop questionnaire
- Extend python backend
- Run regression tests
- Test with pseudo-languages
- Test with illustrative languages
- Test with held-out languages
- Add tests to regression tests
- Add to MatrixDoc pages

How to evaluate a library

- Pseudo-languages
- Illustrative languages
- Held-out languages
- Test suites
- Choices files
- Error analysis

More libraries/reflection from current class

- What do you most wish was available in the customization system, based on what came up in your test suite?
- In your test corpus?

Evolution of 567

- New phenomena: ~~Wh-questions~~, ~~possessives,~~ relative clauses, while-clauses ...?
- Ever bigger jump start --- reaching the limit on this one
- Time to “clean up” noisy inferred morphology early on
- How did these work out?:
  - Partnership with field linguists
  - Work with linguist-provided corpora
- Coverage-driven labs seem most satisfying (MT demo, corpus coverage). Is this true? Can the course be rebalanced to do more of this?

AGGREGATION Project: Motivation & overview

- Precision grammars are potentially useful for endangered language documentation (Bender et al 2012)

- Field linguists produce extremely rich annotations in the form of interlinear glossed text

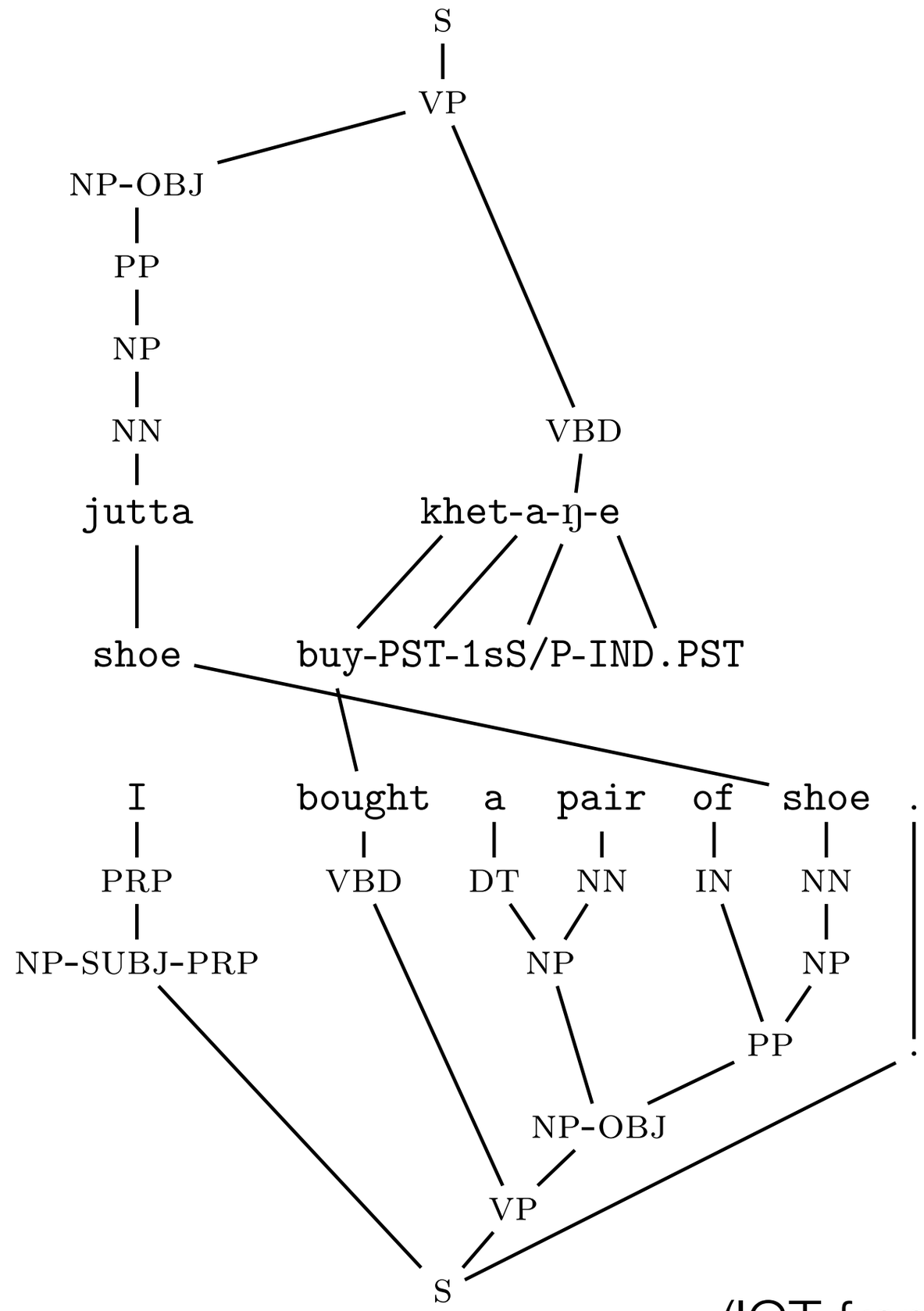
- The Grammar Matrix provides a mapping from grammar specifications to precision grammars

- Can we infer sufficiently accurate and complete grammar specifications from IGT?

RiPLes: Leveraging IGT (Xia & Lewis 2007, Lewis & Xia 2008, Xia & Lewis 2009, Georgi 2016)

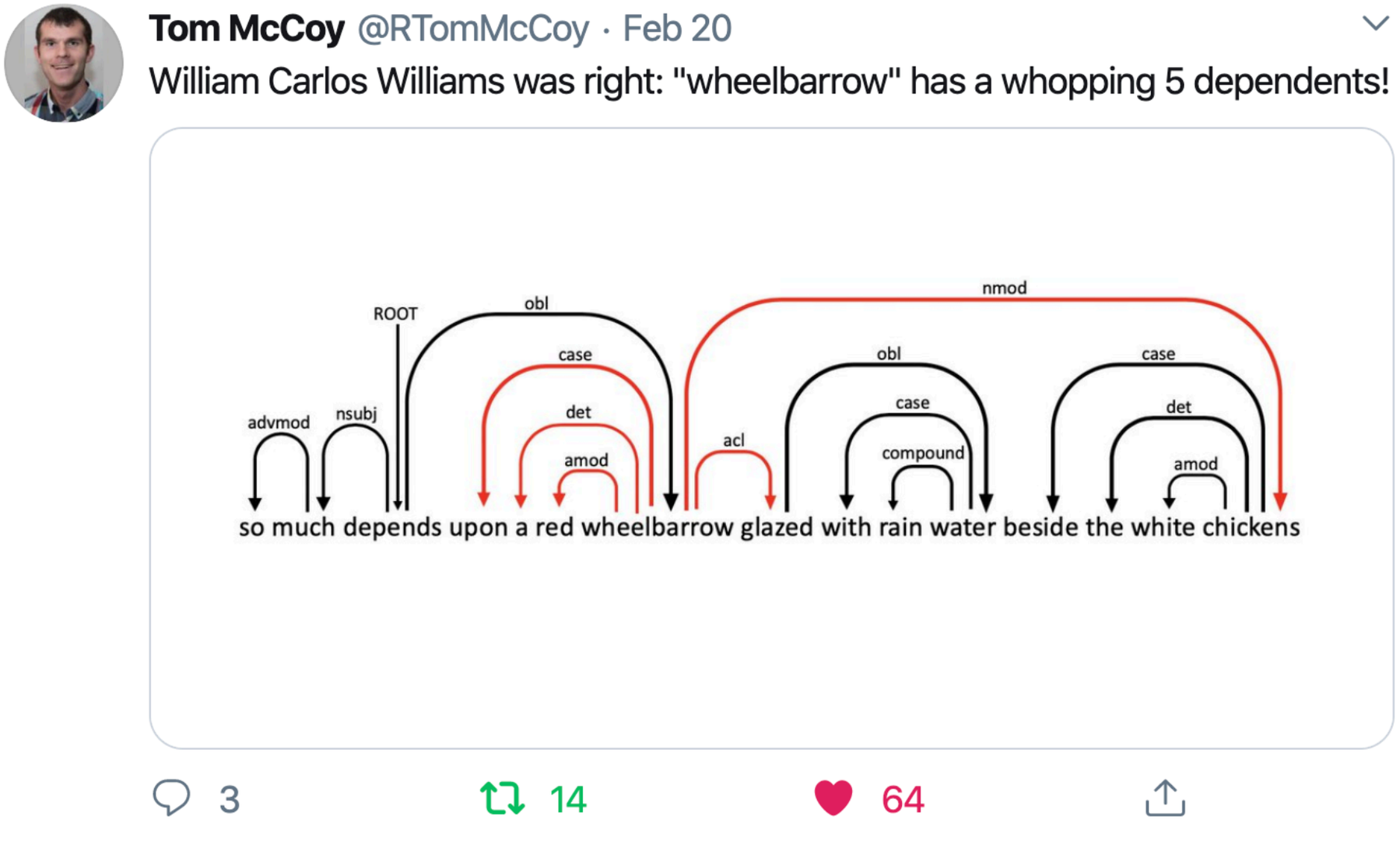
- Interlinear glossed text (IGT) is an extremely rich data type
- IGT exists in plentiful quantities on the web, even for low resource languages
- Example from Chintang [ctn]:

akka ita khurehẽ



(IGT from Bickel et al 2012)

Dependency relations



@emilymbender

Bender et al 2013: Inferring large-scale properties
Task 1: Major constituent word order

- Count word order patterns in projected trees
- Calculate ratios of OS:SO etc
- Plot points for each language in 3D space
- Compare to hypothesized canonical points for each word order
- V2 (and not free) if SVO,OVS >> SOV,OSV

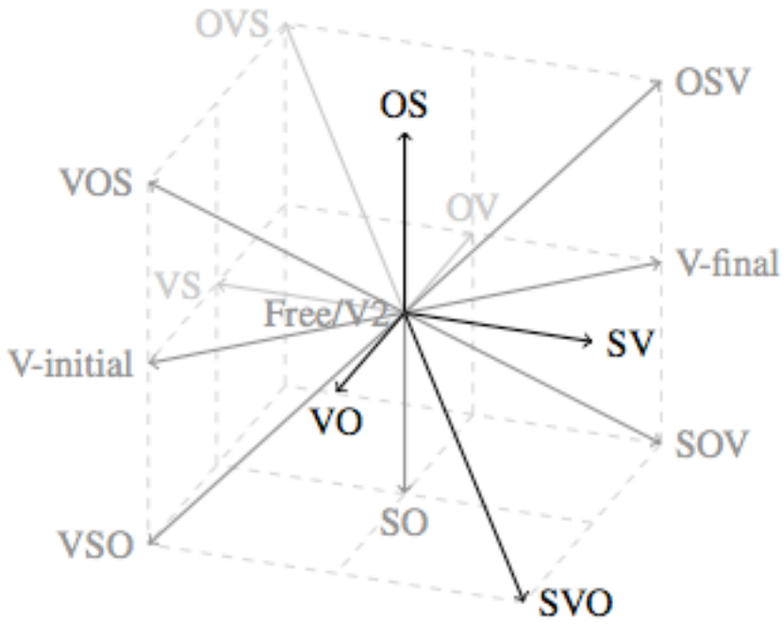


Figure 2: Three axes of basic word order and the positions of canonical word orders.

Dataset	Inferred WO	Baseline
DEV1	0.900	0.200
DEV2	0.500	0.100
TEST	0.727	0.091

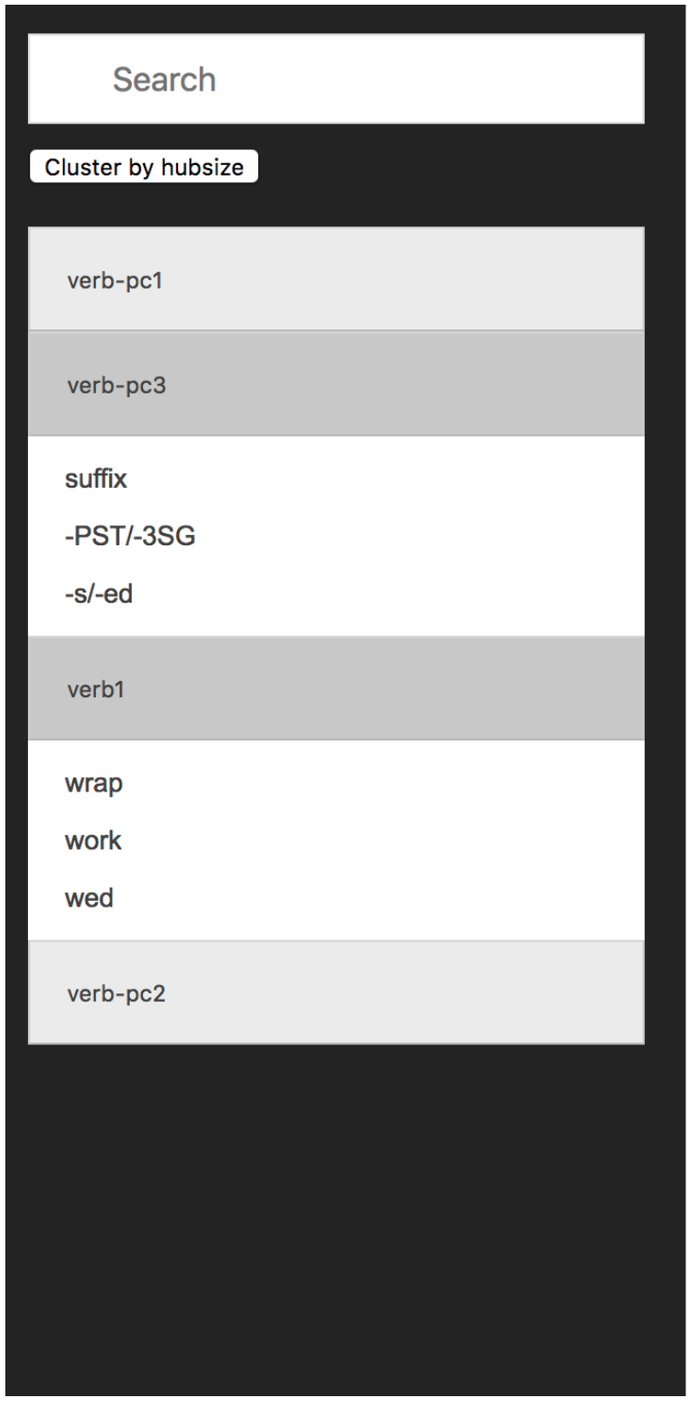
Table 2: Accuracy of word-order inference

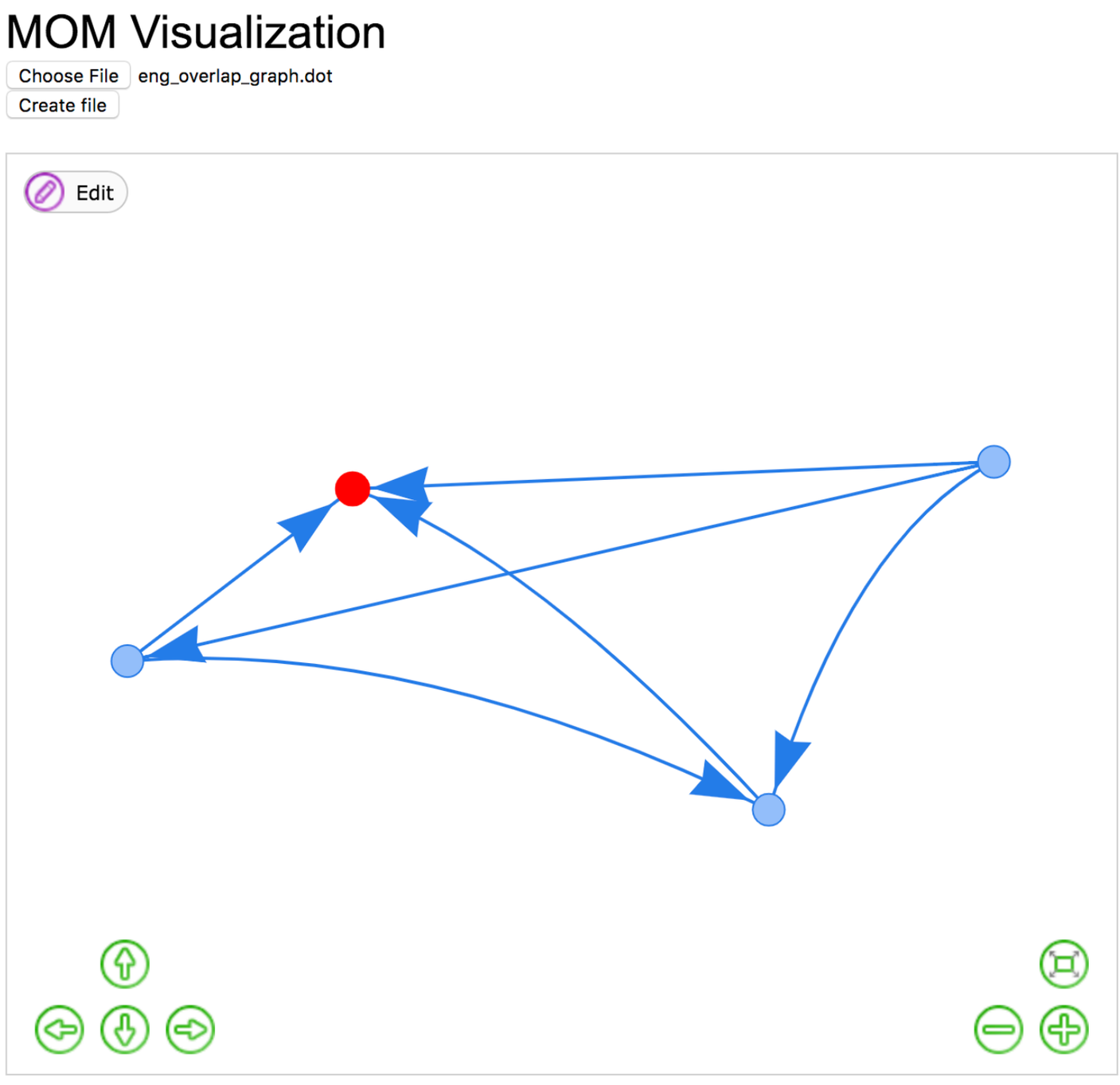
@emilymbender

Wax 2014, Zamaraeva 2016, Zamaraeva et al 2019: Learning lexicons & morphological systems

- General parameters like word order alone won't lead to a usable grammar
- Also required: lexicon and morphotactics (and morphophonology…)
    - Create lexical rules for each morpheme, with associated form and morphosyntactic and morphosemantic features
    - Group morphemes into position classes
    - Determine ordering relations
    - Lexicon: part of speech, case frame, argument optionality…

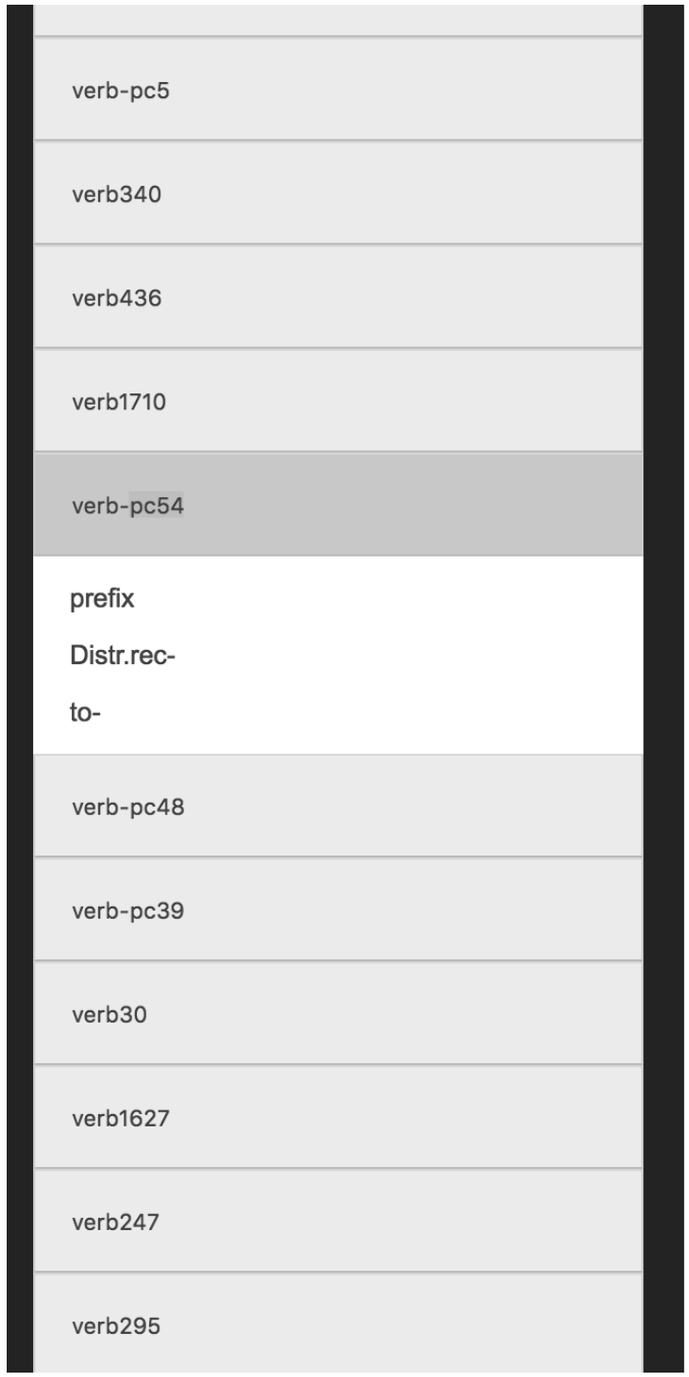
Lepp et al 2019: Visualizing inferred morphotactics

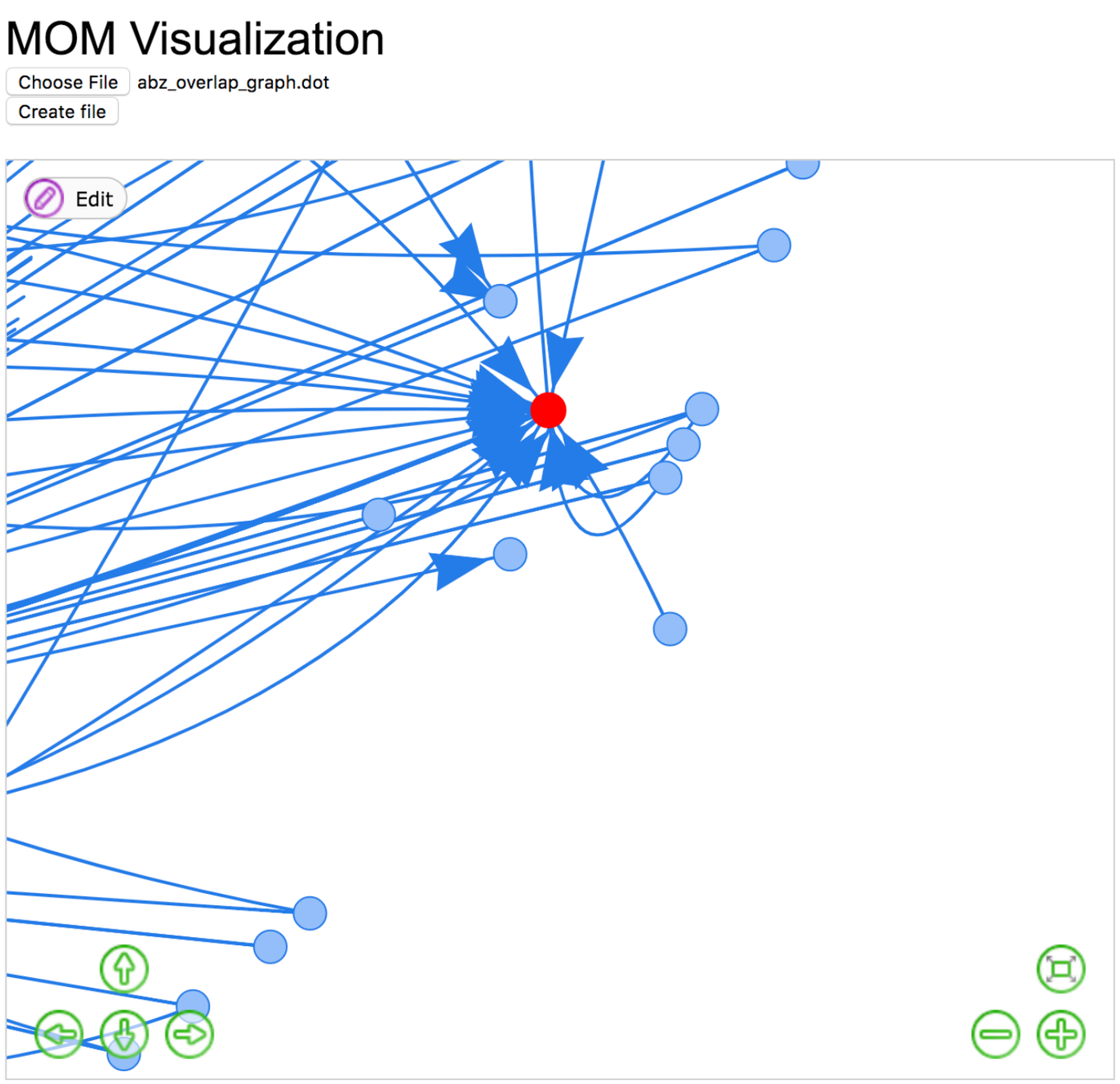




@emilymbender

Lepp et al 2019: Visualizing inferred morphotactics





@emilymbender

End-to-end evaluation with Chintang [ctn] (Zamareva et al 2019)

Choices file	# verb entries	# noun entries	# verb affixes	# noun affixes
ORACLE	899	4750	233	36
BASELINE	3005	1719	0	0
FF-AUTO-GRAM	739	1724	0	0
MOM-DEFAULT-NONE	1177	1719	262	0
INTEGRATED	911	1755	220	76

Table 3: Amount of lexical information in each choices file

choices file	lexical coverage (%)	parsed (%)	correct (%)	readings
ORACLE	116 (12.5)	20 (2.2)	10 (1.1)	1.35
BASELINE *	38 (0.4)	15 (1.6)	8 (0.9)	27.67
FF-AUTO-GRAM	18 (1.9)	4 (0.4)	2 (0.2)	5.00
MOM-DEFAULT-NONE	39 (4.2)	16 (1.7)	3 (0.3)	10.81
INTEGRATED	105 (11.3)	32 (3.4)	15 (1.6)	91.56

* We report slightly different results for lexical coverage and average readings for the baseline than Bender et al. (2014) because we removed determiners from the choices file.

Table 4: Results on 930 held-out sentences

Extending inference (Howell in progress)

- Previously available: major constituent word order, case systems, case frames for verbs, case values for nouns

- Adding: argument optionality, coordination, PNG on nouns and agreeing categories, tense/aspect/mood, sentential negation, auxiliaries, case-marking adpositions

- Initial system tested in Ling 567 as starting grammar specifications (noisy!)

- Testing on 15 languages: 10 dev, 5 held-out

    - Coverage, ambiguity, treebanked accuracy

Extending inference (Howell in progress)

Language	N	Lexical Cov.	Parse Cov.	Correct Pred-Arg Structure	Correct Features	Ambig.
Development1						
Abui [abz]	1569	54.75%	43.98%	8.28%*	7.01%*	2365
Chintang [ctn]	9785	22.24%	12.38%	2.15%*	1.02%*	5648
Nuuchahnulth [nuk]	641	6.08%	3.43%	1.09%	1.09%	18
Matsigenka [mcb]	349	20.59%	8.82%	0.57%	0.57%	6827
Wambaya [wmb]	818	13.81%	3.91%	2.32%	0.00%	6
Development2						
Haiki [yaq]	2235	21.39%	12.39%			93
Lezgi [lez]	1197	12.87%	10.03%			19121
Meithei [mni]	1717	8.56%	7.92%			7139
Tsova-Tush [bbl]	1611	30.11%	24.77%			2470

Table 1: Coverage and Ambiguity for Development Languages. Results are averages across 10 splits. * indicates results for only a single split.

External resources: WALS (Dryer & Haspelmath 2013)

- To what extent do the features in WALS map to Grammar Matrix grammar specifications? (Almeida et al 2019)
    - Some are mappable, but often without the degree of specificity required
- Where they do map, what is the best way to leverage them in inference of grammar specifications? (Zhang et al 2019)
    - Currently testing: use WALS info, when available, to constrain what inference looks for
- Could we do better with Autotyp? (Bickel & Nichols 2002)

Overview

- Wrap up/reflections
- Matrix: Future directions
- Next time: MMT extravaganza and course evals