

The Matrix: Future Directions

Wrap up

Ling 567

March 10, 2025

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

Reflections (break-out)

- 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 you learn about similarities and differences across languages?

Feedback: projects

- Was the workload reasonable?
- What helped in making it (more) manageable?
- How did you handle sharing the workload?
- What would have been 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

- Next up?
 - Serial verb constructions
 - Pronouns, demonstratives, COG-ST (definiteness)
 - Other non-verbal predicates
 - Other intersective modifiers
 - Numeral classifiers
 - More verb subcategorization
 - Information structure in wh questions & free word order languages
 - Other?
- Also, lots of bugs need fixing :) <https://github.com/delph-in/matrix/projects/1>

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, while-clauses,~~ relative clauses, ...?
- Ever bigger jump start --- ~~reaching~~ the limit on this one
- Use of inferred choices files from the AGGREGATION project (2019-2023)
- Coverage-driven labs seem most satisfying (MT demo). Is this true? Did you get enough of this this year?

AGGREGATION Project:

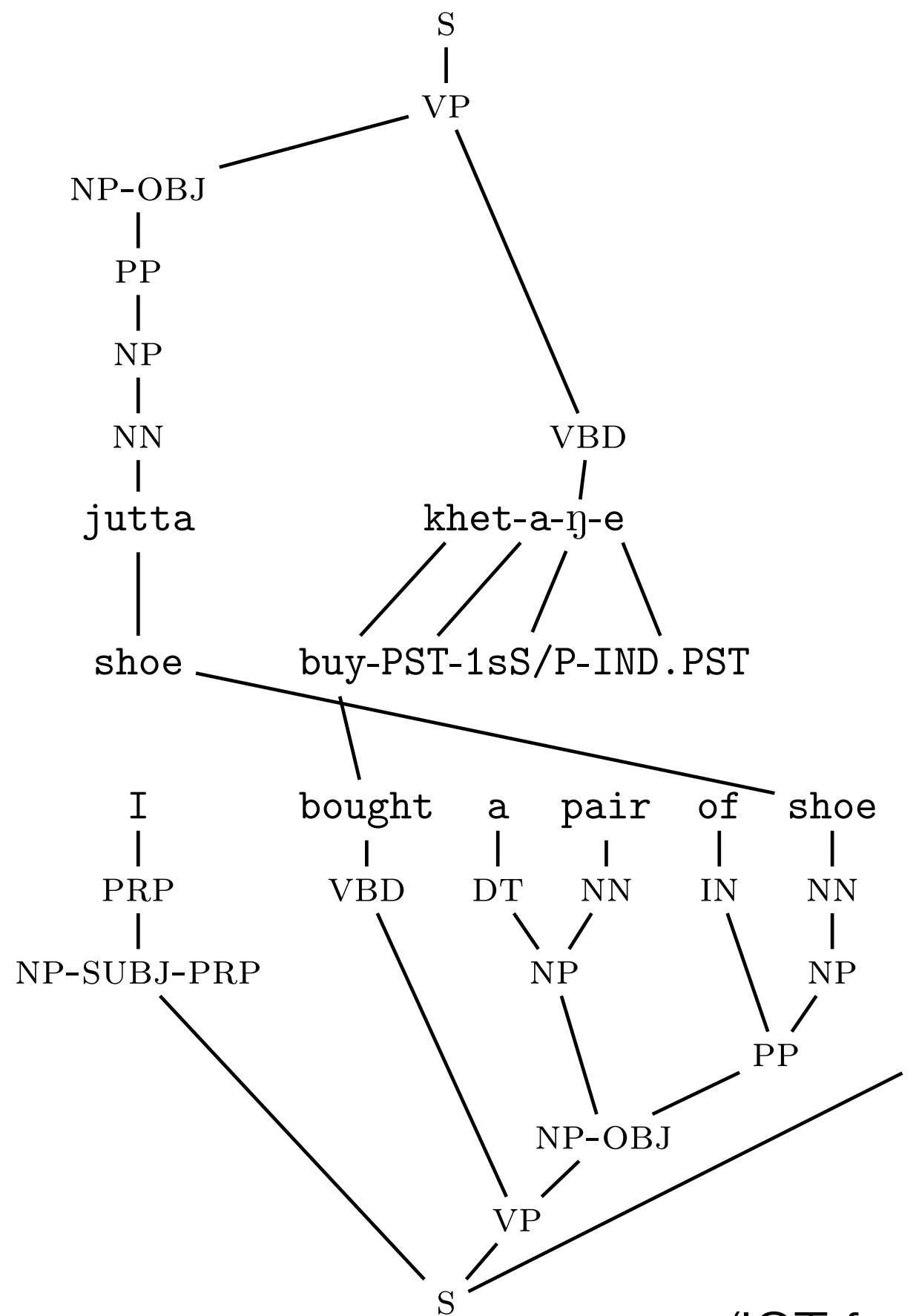
Motivation & overview

- 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)

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

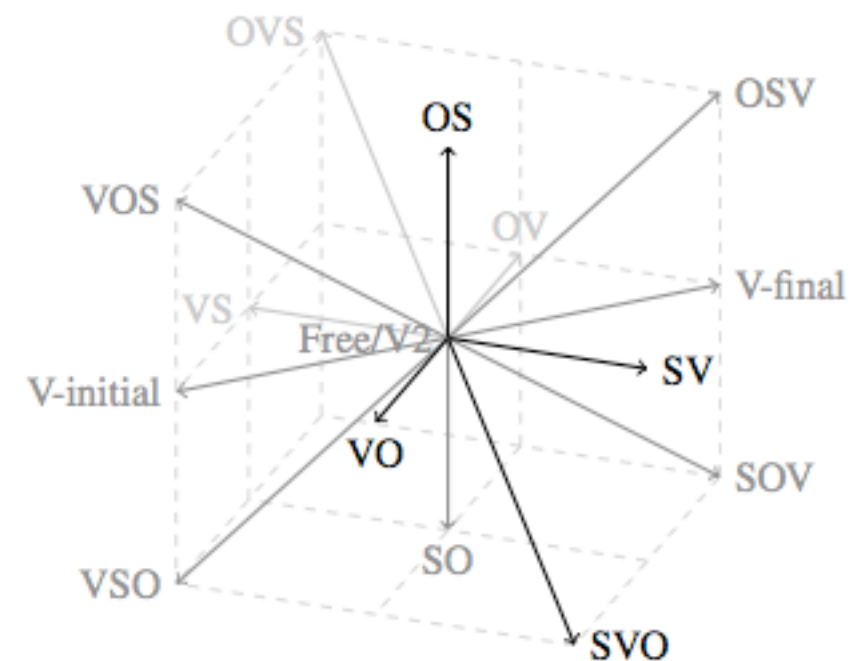


Figure 2: Three axes of basic word order and the positions of canonical word orders.

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

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

Cluster by hubsize

verb-pc1

verb-pc3

suffix
-PST/-3SG
-s/-ed

verb1

wrap
work
wed

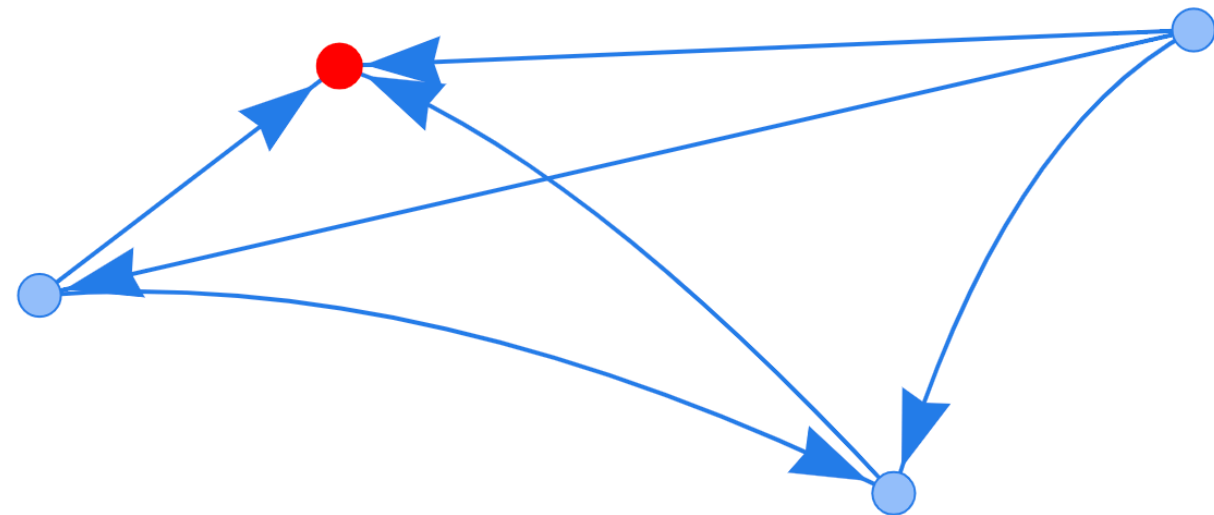
verb-pc2

MOM Visualization

Choose File eng_overlap_graph.dot

Create file

Edit



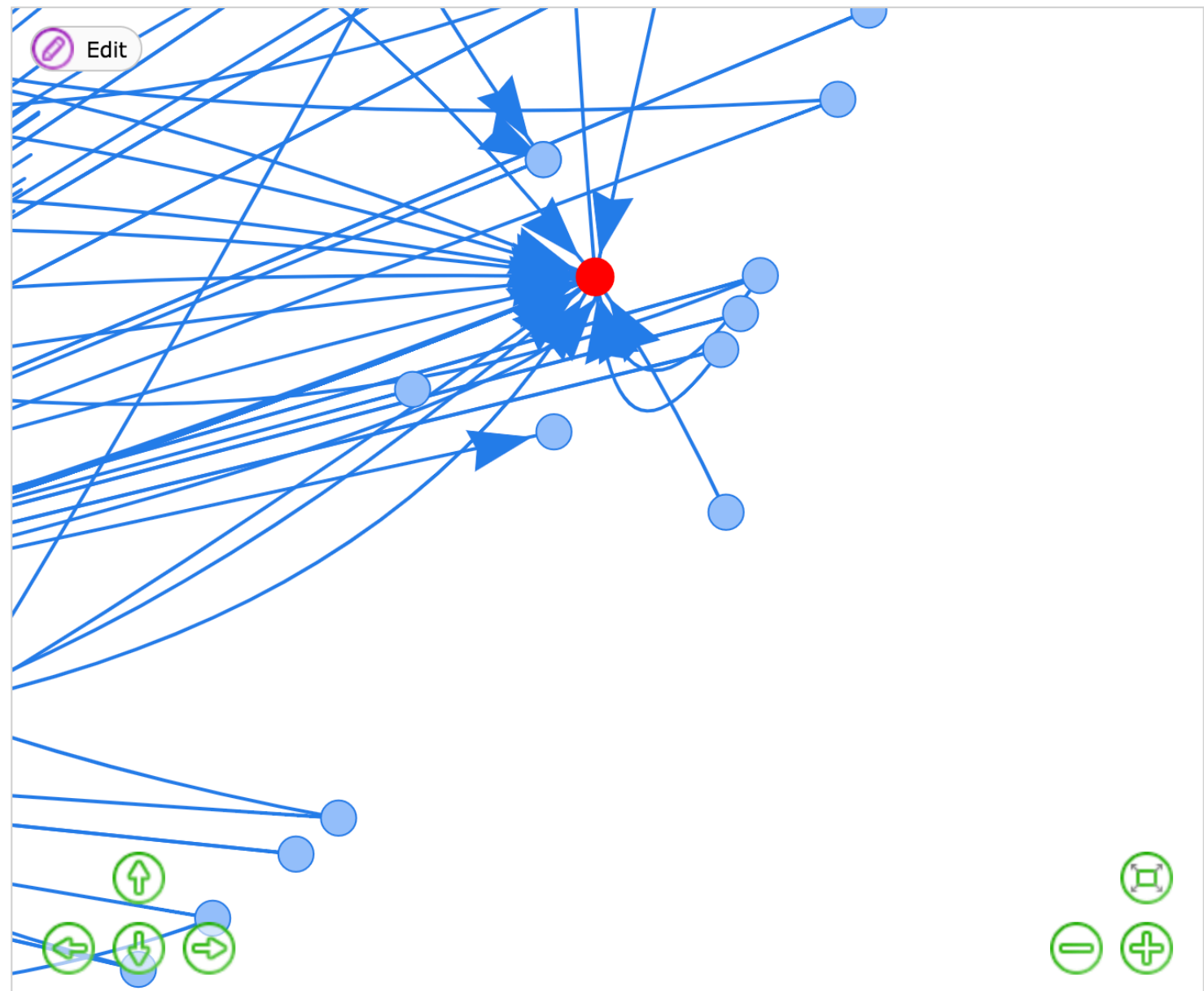
Lepp et al 2019: Visualizing inferred morphotactics

verb-pc5
verb340
verb436
verb1710
verb-pc54
prefix
Distr.rec-
to-
verb-pc48
verb-pc39
verb30
verb1627
verb247
verb295

MOM Visualization

Choose File abz_overlap_graph.dot

Create file



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 2020, Conrad 2021, Howell & Bender 2022, Dods 2022, Lin 2023

- 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, adverbial subordinate clauses, adnominal possession, valence changing morphology
- System tested in Ling 567 as starting grammar specifications (noisy!)
- Howell 2020, Howell & Bender 2022: Testing on 14 languages: 5 dev, 4 initial held-out, 5 more held-out
 - Coverage, ambiguity, treebanked accuracy

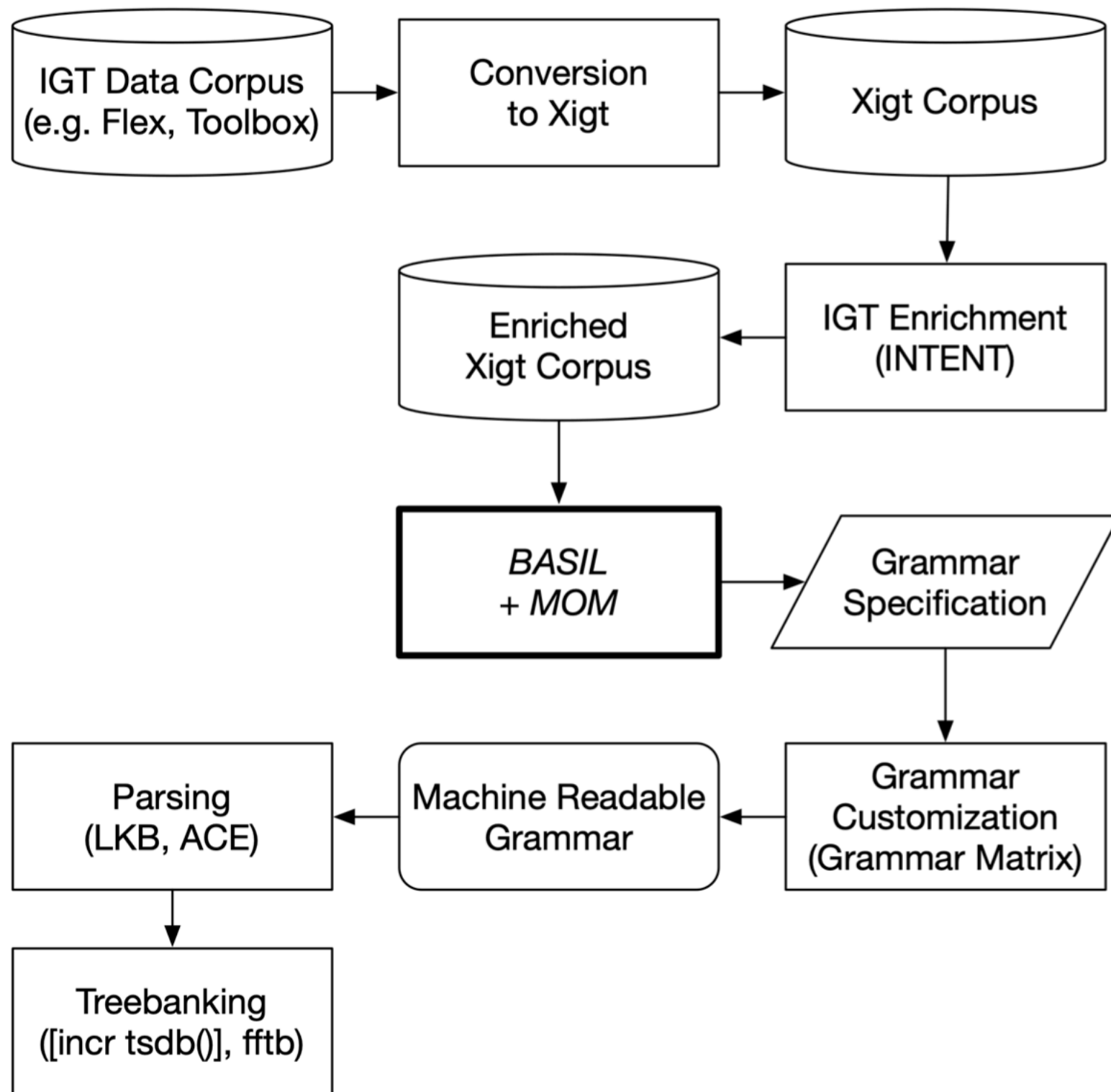
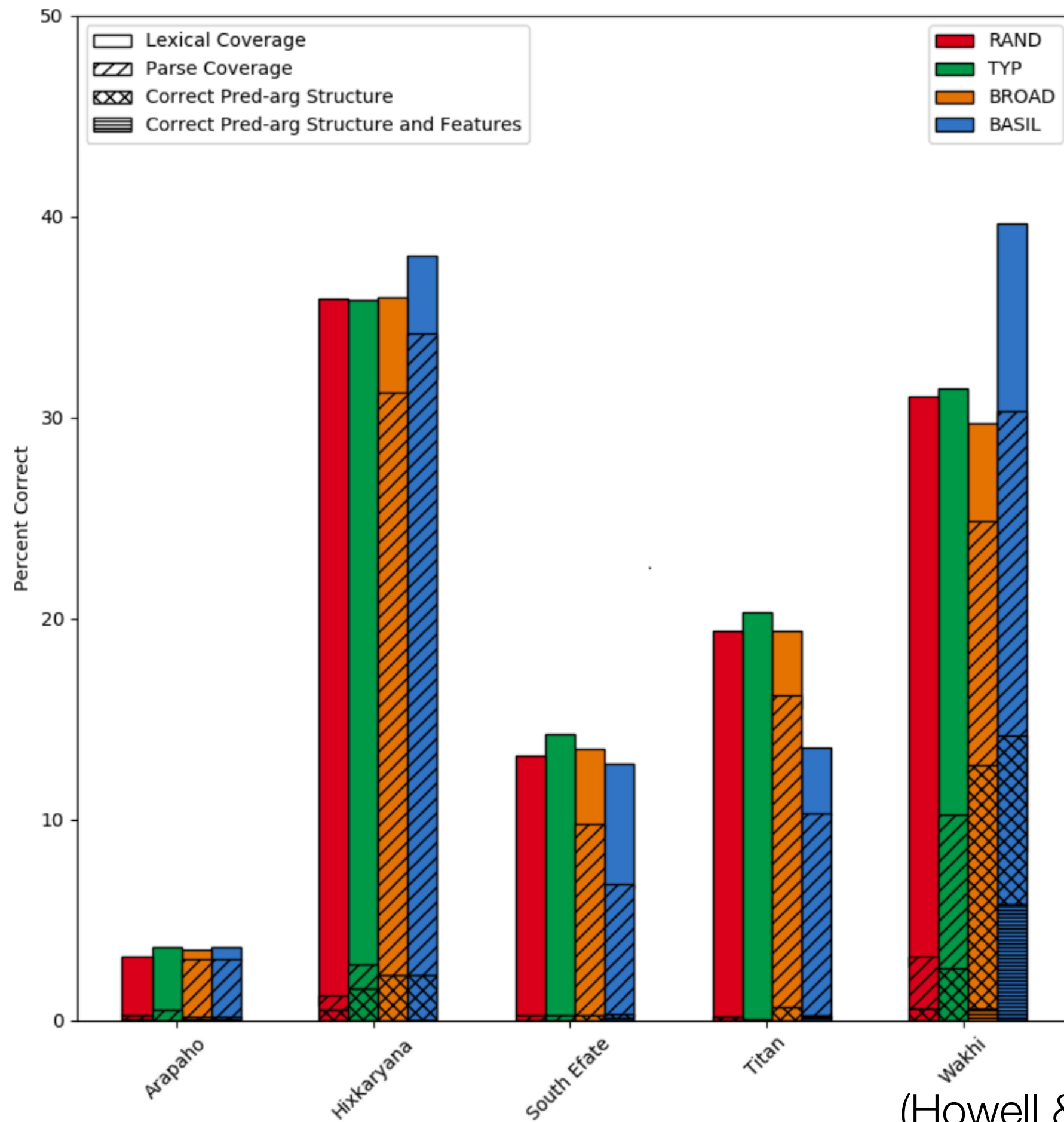


Figure 1: AGGREGATION Pipeline

(Howell & Bender 2022)



(Howell & Bender 2022)

Overview

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