Partially Automated Grammar Engineering and Chintang

Emily M. Bender
University of Washington/Aarhus Universitet

Seminar für Allgemaine Sprachwissenschaft
Universität Zürich

10 July 2014

Acknowledgments

- This material is based upon work supported by the National Science Foundation under Grant No. 1160274. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

- Collaborators: Balthasar Bickel, Joshua Crowgey, Michael Goodman, Robert Schikowski, Fei Xia, Olga Zamaraeva

- The Chintang and Puma Documentation Project!

High-level Goals

- Create precision, implemented, morphosyntactic grammars in the context of language description
  - Both grammars and eventually treebanks (Bender et al 2012)
- Speed up the development of precision grammars, though:
  - Grammar code re-use (Grammar Matrix; Bender et al 2002, 2010)
  - Leveraging analyses encoded by field linguists, in Toolbox lexica and in IGT

Precision Grammars

- Map (near-)surface strings to semantic representations, and vice versa
  - Syntactic structures mostly a means to this end
- Encode a sharp notion of grammaticality
- Require models *où tout se tient*
- In our case, use HPSG (Pollard & Sag 1994) and MRS (Copestake et al 2005)

Precision Grammars: Example

din khiptukum
din khipt-u-kV-m
day count-3P-IND.NPST-1/2nsA
'(We) count days.' [ctn] (Bickel et al., 2013b)

$$\left\langle h_1, \left| \begin{array}{l} h_3\text{:\_din\_n\_day}(x_4), \\ h_5\text{:\_exist\_q\_rel}(x_4, h_6, h_7), \\ h_6\text{:\_khipt-u-kv-m\_v\_count}(e_2, x_9, x_4) \end{array} \right| \{ h_6 =_q h_3 \} \right\rangle$$

=> demo

Precision Grammars in Language Documentation: Wambaya [wmb]

Precision Grammars in Language Documentation: Wambaya [wmb]

- Wambaya [wmb] is a Mirndi language spoken in the West Barkly Tablelands region (Northern Territory) of Australia (Nordlinger 1998; Green & Nordlinger 2004)

Precision Grammars in Language Documentation: Wambaya [wmb]

- Wambaya [wmb] is a Mirndi language spoken in the West Barkly Tablelands region (Northern en & Nordlinger 2004)

Precision Grammars in Language Documentation: Wambaya [wmb]

- Wambaya [wmb] is a Mirndi language spoken in the West Barkly Tablelands region (Northern Territory) of Australia (Nordlinger 1998; Green & Nordlinger 2004)

Precision Grammars in Language Documentation: Wambaya [wmb]

- Wambaya [wmb] is a Mirndi language spoken in the West Barkly Tablelands region (Northern Territory) of Australia (Nordlinger 1998; Green & Nordlinger 2004)

- Developed by hand (with jump start from the Grammar Matrix) on the basis of the analyses in Nordlinger 1998 (incl. IGT)

Precision Grammars in Language Documentation: Wambaya [wmb]

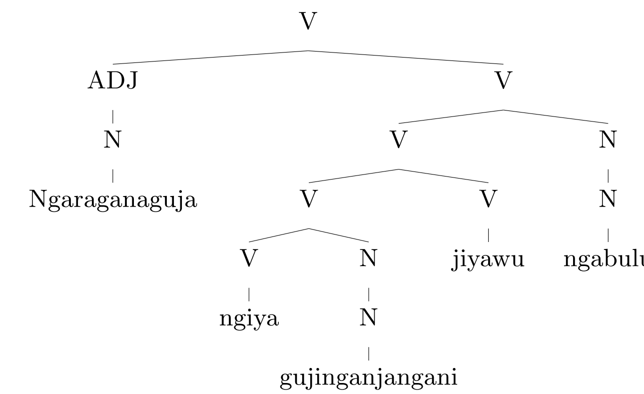
- Wambaya [wmb] is a Mirndi language spoken in the West Barkly Tablelands region (Northern Territory) of Australia (Nordlinger 1998; Green & Nordlinger 2004)

- Developed by hand (with jump start from the Grammar Matrix) on the basis of the analyses in Nordlinger 1998 (incl. IGT)

- Development set: 794 examples from Nordlinger 1998

Precision Grammars in Language Documentation: Wambaya [wmb]

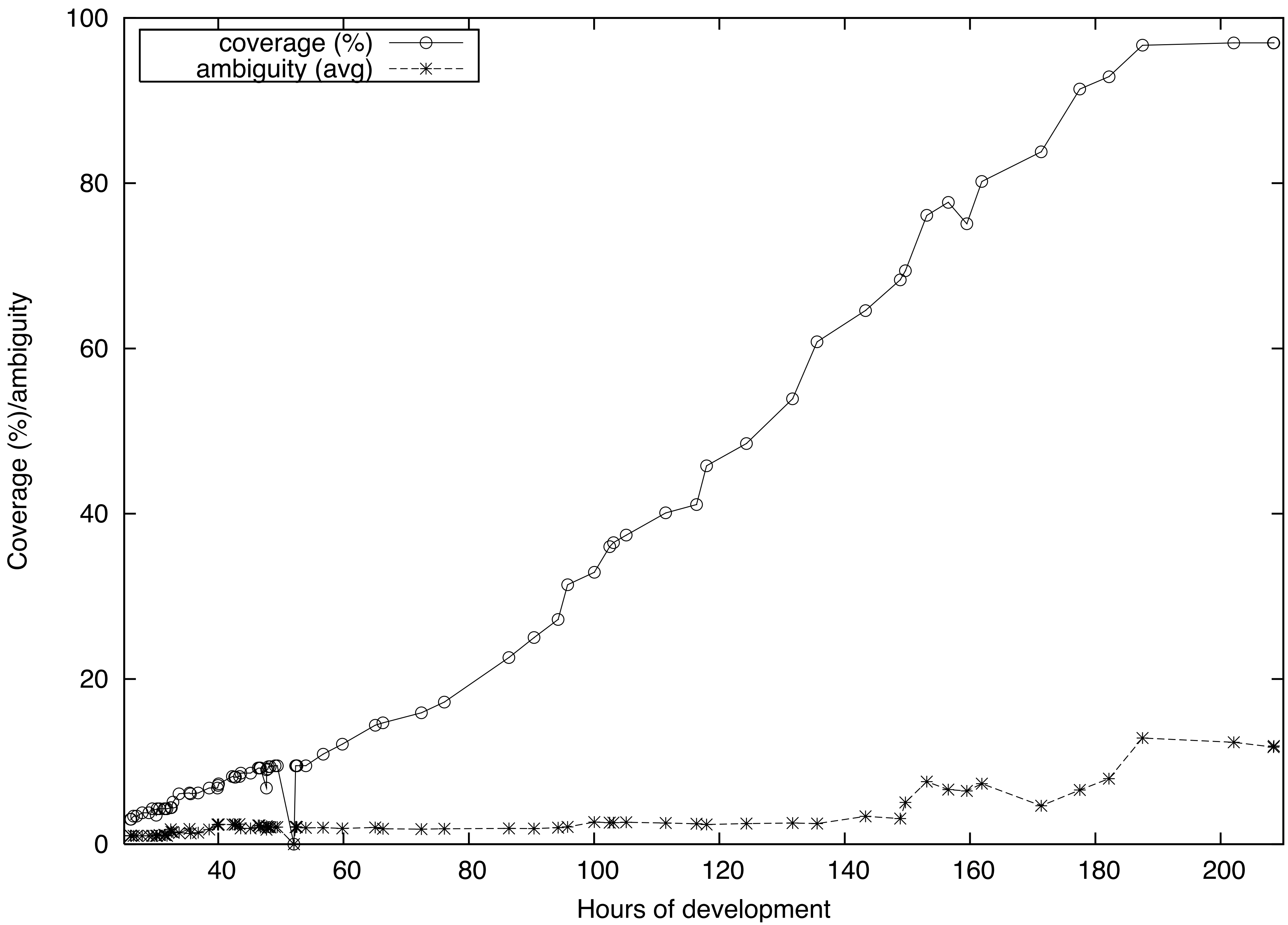
- Wambaya [wmb] is a Mirndi language spoken in the West Barkly Tablelands region (Northern Territory) of Australia (Nordlinger 1998; Green & Nordlinger 2004)
- Developed by hand (with jump start from the Grammar Matrix) on the basis of the analyses in Nordlinger 1998 (incl. IGT)
- Development set: 794 examples from Nordlinger 1998
- Held-out test data: Narrative from Nordlinger 1998



Wambaya grammar scope

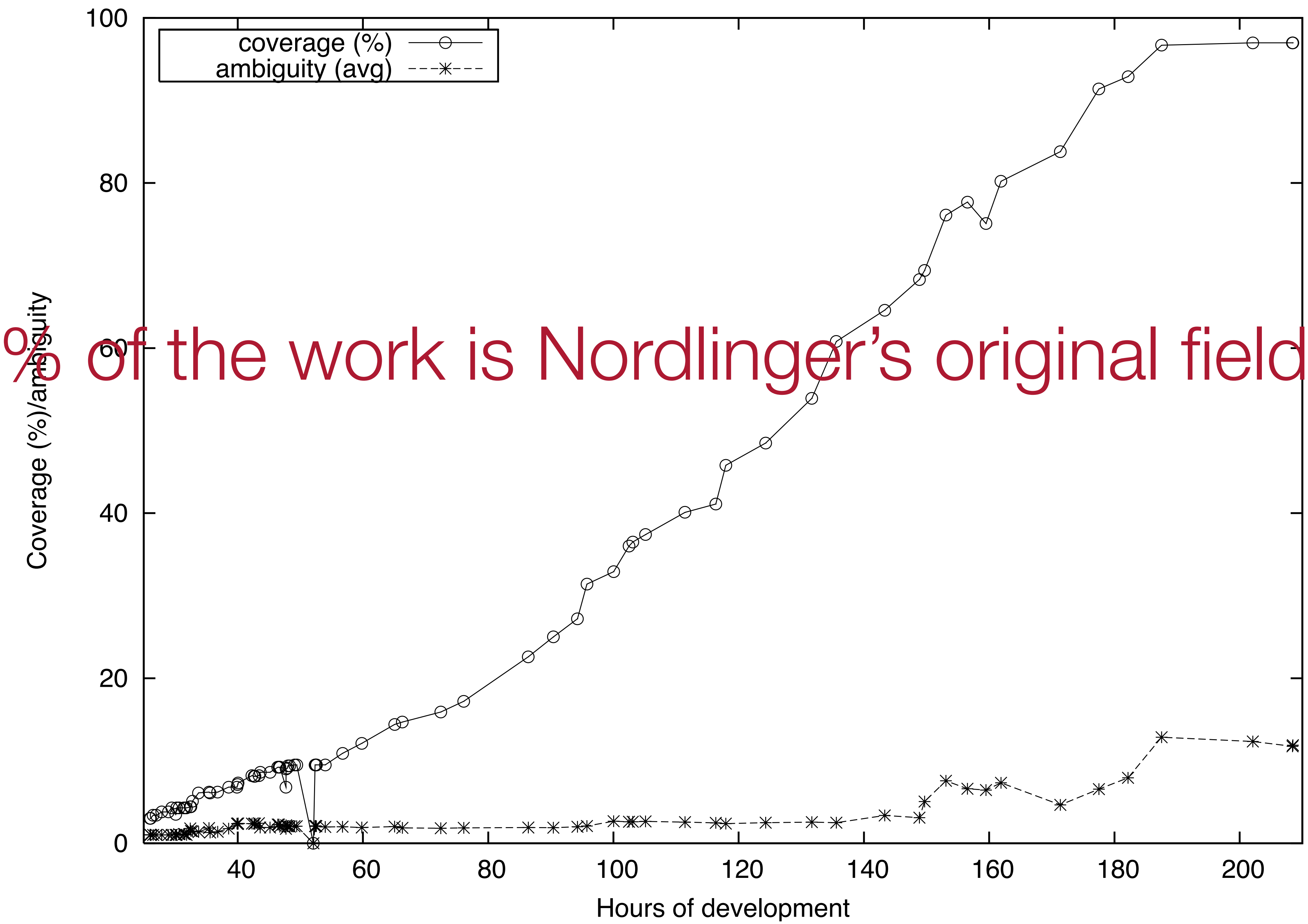
- Word order (2nd position auxiliary, discontinuous noun phrases)
- Argument optionality
- Linking of syntactic to semantic arguments
- Case (split ergativity)
- Agreement (verb-subject, verb-object, adj-noun)
- Lexical adverbs, including manner, time, location, and negation
- Derived event modifiers
- Derived nominal modifiers
- Lexical adjectives (demonstratives, possessives, numerals, others)
- Subordinate clauses (clausal complements, purposives, simultaneous and prior events)
- Verbless clauses: adjective, nouns, and adverbs functioning as predicates
- Illocutionary force: imperatives, declaratives and interrogatives, including *wh*-questions
- Coordination: of clauses and noun phrases
- Inalienable possession construction
- Secondary predicates
- Causatives of verbs and adjectives

Wambaya grammar development



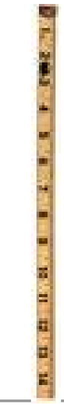
(Bender 2008)

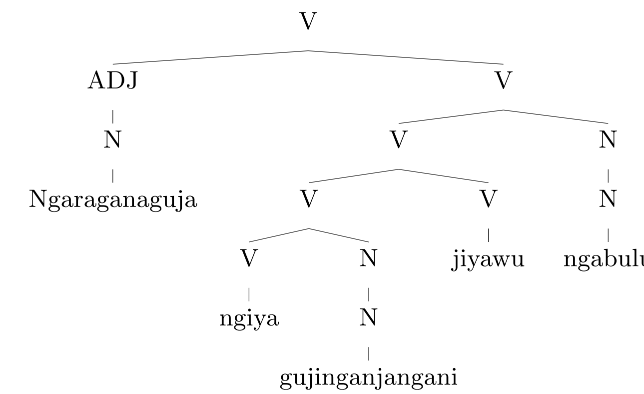
Wambaya grammar development



>95% of the work is Nordlinger's original field work

(Bender 2008)





Wambaya grammar evaluation

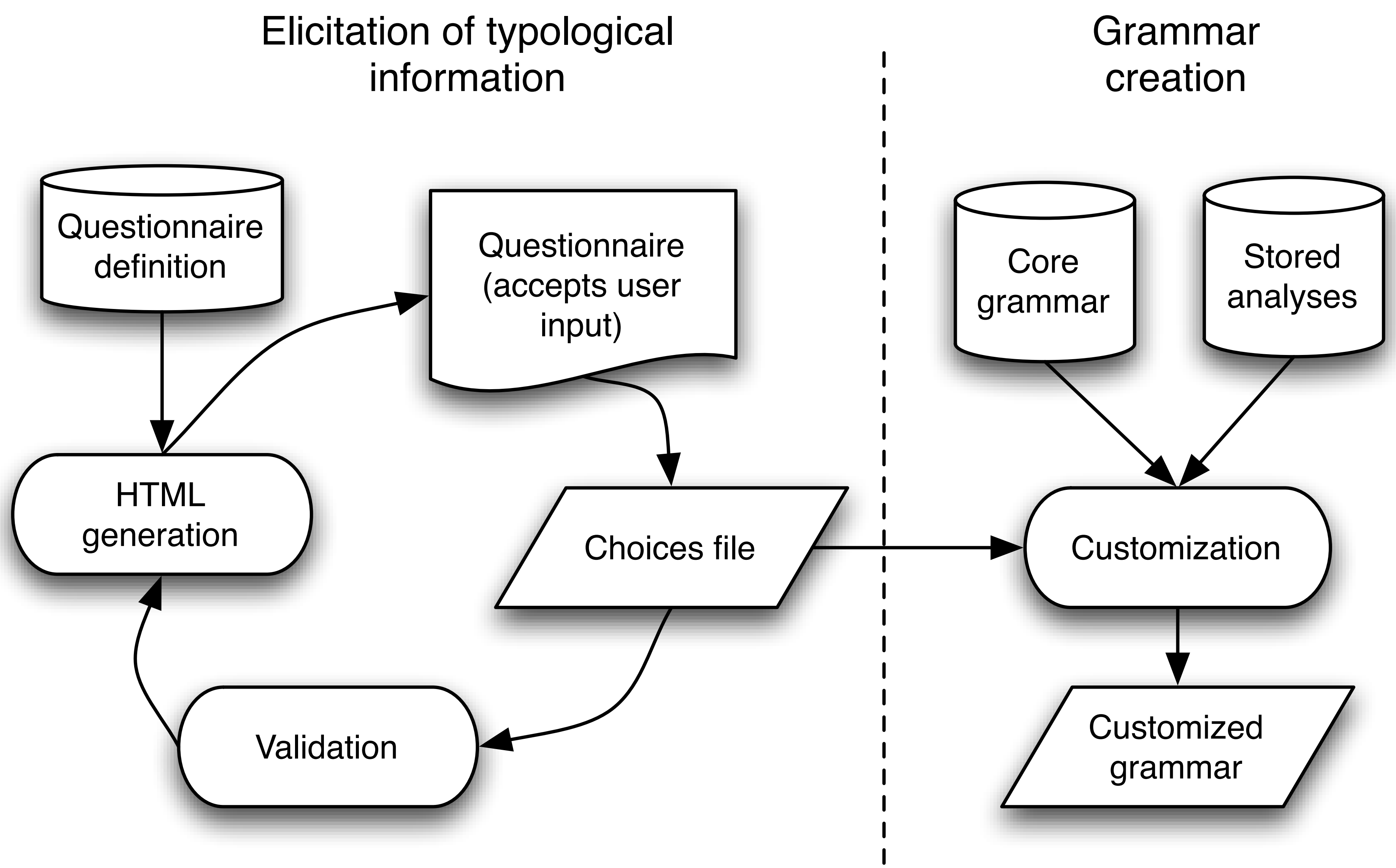
- Held out test data “The two Eaglehawks”
- 72 sentences (orig text: 92, removed 20 seen sentences)
- Run twice: before and after adding lexical entries and adjusting morphophonology only

	correct	parsed incorrect	unparsed	average ambiguity
Existing vocab	50%	8%	42%	10.62
w/added vocab	**76%**	8%	14%	12.56

Grammar Matrix

- Facilitate grammar development & grammar interoperability through code-reuse
- Combine breadth of typological analysis with depth of grammar implementation ("computational linguistic typology")
- Core grammar: Information hypothesized to be cross-linguistically useful
- Libraries: Sets of analyses of cross-linguistically variable phenomena
- Accessed through a web-based questionnaire
- => Demo Grammar Matrix questionnaire

Grammar Matrix



Morphophonology and Morphosyntax

- Amenable to different kinds of formal analysis
- Should be kept separate (but linked) (Bender & Good, 2005)
- Grammar Matrix derived grammars assume the existence of morphophonological analyzers, and target the morpheme-segmented line of IGT:

Morphophonology and Morphosyntax

- Amenable to different kinds of formal analysis
- Should be kept separate (but linked) (Bender & Good, 2005)
- Grammar Matrix derived grammars assume the existence of morphophonological analyzers, and target the morpheme-segmented line of IGT:

unisaa	*khatte*	*mo*	*kosi*	*moba*
u-nisa-a	khatt-e	mo	kosi-i	mo-pe
3sPOSS-yngr.bro-ERG.A	take-IND.PST	DEM.DOWN	river-LOC	DEM.DOWN-LOC

The younger brother took it to the river. [ctn] (Bickel et al., 2013c)

Morphophonology and Morphosyntax

- Amenable to different kinds of formal analysis
- Should be kept separate (but linked) (Bender & Good, 2005)
- Grammar Matrix derived grammars assume the existence of morphophonological analyzers, and target the morpheme-segmented line of IGT:

unisaa	*khatte*	*mo*	*kosi*	*moba*
u-nisa-a	khatt-e	mo	kosi-i	mo-pe
3sPOSS-yngr.bro-ERG.A	take-IND.PST	DEM.DOWN	river-LOC	DEM.DOWN-LOC

The younger brother took it to the river. [ctn] (Bickel et al., 2013c)

Chintang as a case study language: Why?

- Morphologically rich
- Active descripti
- Relatively matu
- Treasure trove c

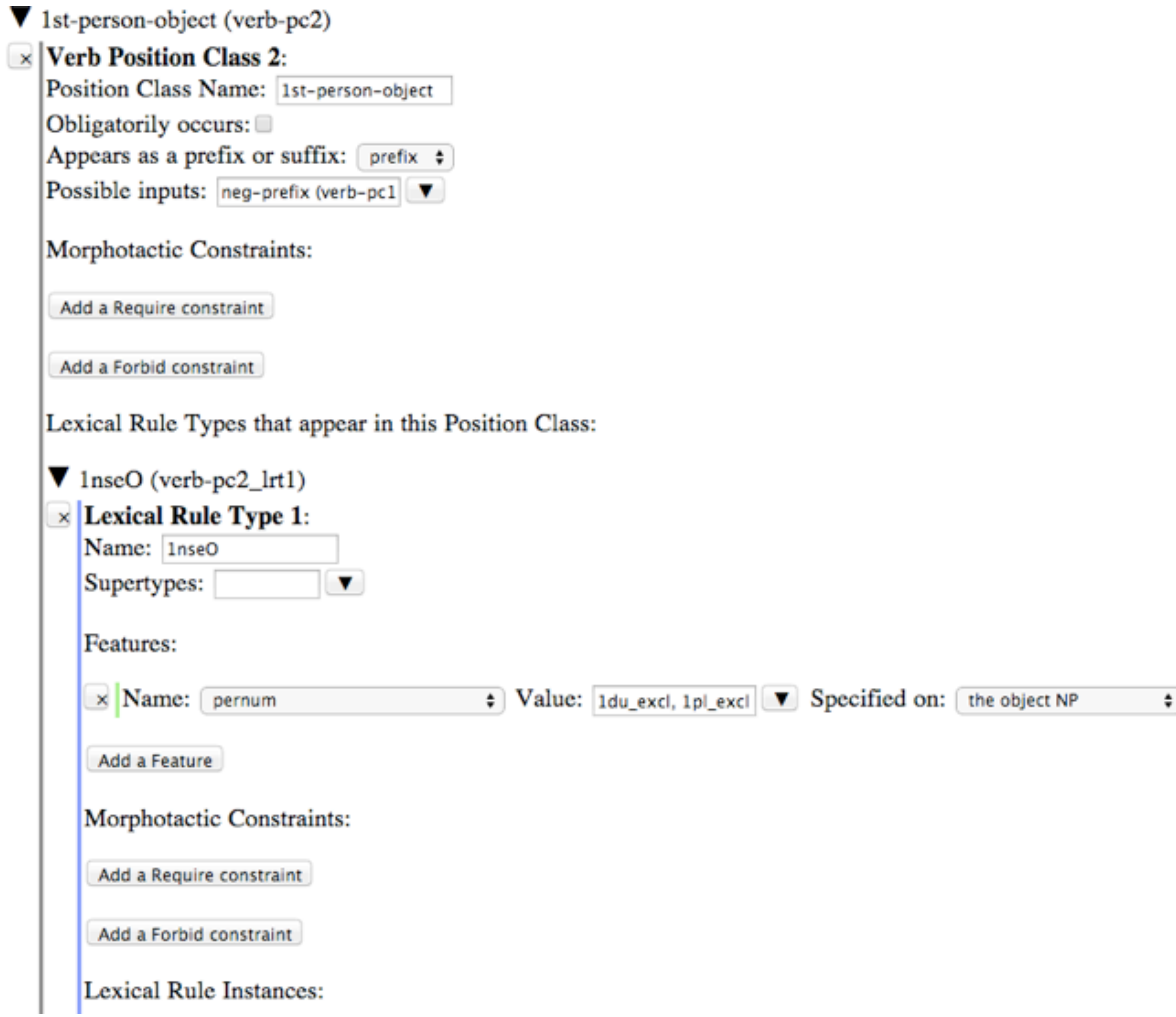
Chintang as a case study language: Why?

- Morphologically rich
- Active descriptive/documentation work
- Relatively mature descriptive materials
- Treasure trove of IGT!

Defining Morphosyntactic rules for Chintang
(Bender, Schikowski & Bickel 2012)

- Done collaboratively with Robert Schikowski, on the basis of Schikowski 2012
- 160 verbal lexical rules grouped into 54 position classes
  - Including duplicated prefix/suffix rules to handle the multi-root verbs
- 24 nominal lexical rules grouped into 6 position classes
- Elsewhere in the choices file, define values for case, tense/aspect/mood, person/number/gender
  - => Partial modeling of the syntactico-semantic effects of 131/160 rules

Questionnaire Snippet



=> demo?

Implementation of Morphology

Feature	# Rules	Notes
PERNUM	103	Person and/or number of verb's dependent or of noun
CASE	17	
FORM	35	Form (finite/non-finite) of verb
TENSE	27	
ASPECT	5	
MOOD	17	
NEGATION	8	

Reuse of Resources I: Import from Toolbox Lexicon
(Bender, Schikowski & Bickel 2012)

- Define lexical types in Grammar Matrix customization system
- Use new ‘Import Toolbox lexicon’ subpage to map Toolbox lexical entries to types
    - Toolbox tags are user-defined; Import page allows user-specified tag-value pairs
    - Lexical types, import types, import tag-value constraints are all unbounded

Chintang Toolbox lexicon

```
\lex kond
\id 179
\psrev v
\val A-ERG P-NOM V-a(A).o(P)
\ge search; look.for
\dt 22/Feb/2011
```

Chintang Toolbox lexicon

```
\lex kond
\id 179
\psrev v
\val A-ERG P-NOM V-a(A).o(P)
\ge search; look.for
\dt 22/Feb/2011
```

Grammar Matrix lexical entry: source code

```
khad := trans-verb-lex &
  [ STEM < "kond" >,
    SYNSEM.LKEYS.KEYREL.PRED "_search;look.for_v_rel" ].

trans-verb-lex := erg-abs-transitive-verb-lex.

erg-abs-transitive-verb-lex := transitive-verb-lex &
  [ ARG-ST < [ LOCAL.CAT.HEAD noun &
                              [ CASE erg ] ],
             [ LOCAL.CAT.HEAD noun &
                              [ CASE abs ] ] > ].
```

```
[ trans-verb-lex
  STEM        ⟨ kond ⟩
  SYNSEM      [ lex-synsem
                CAT   [ cat
                        HEAD  [ verb
                                AUX   –
                                FORM  form ]
                        VAL   [ SUBJ   ⟨ [0] [ CAT   [ HEAD [ noun
                                                              CASE  erg ] ]
                                               CONT  [ INDEX [1] ] ] ⟩
                                COMPS  ⟨ [8] [ CAT   [ HEAD [ noun
                                                              CASE  nom ] ]
                                               CONT  [ INDEX [9] ] ] ⟩ ] ]
                CONT  [ INDEX [14] [ event
                                     TENSE   tense
                                     ASPECT  aspect
                                     MOOD    mood ]
                        RELS  ⟨ [ arg12-ev-relation
                                  PRED  _search;look.for_v_rel
                                  ARG0  [14]
                                  ARG1  [1]
                                  ARG2  [9] ] ⟩ ] ]
  INFLECTED   [ TRANSITIVE-VERB-FLAG  + ] ]
```

Final import types: entries imported

- Common nouns: 4,741
- Native Chintang intransitive verbs: 282
- Borrowed Nepali intransitive verbs: 142
- Native Chintang transitive verbs: 285
- Borrowed Nepali transitive verbs: 190
- Total: 5,640/9,034 (62%) of all entries; 899/1,440 (62%) of the verbs
- Most frequent unimported POS tages: adverbs (866), adjectives (515), interjections (377), affixes (286)

Evaluation procedure

- Define choices file, including: lexical types, lexical rules (as above), import classes, supporting grammatical information, 10 lexical entries for pronouns, 1 lexical entry for an auxiliary
- Create grammar using Grammar Matrix customization system
- Load grammar into LKB grammar development environment (Copestake 2002) with [incr tsdb()] (Oepen 2001) for test suite management
- Refine choices file based on parsing results over three sample narratives (2,906 word tokens)
- Test lexical coverage of resulting grammar on four held-out narratives
- Evaluate precision/recall of morphosyntactic and morphosemantic features

Test data narratives

- 1,453 total word tokens
- Varied genres: biographical monologue, Pear Story (Chafe 1980), traditional story, recipe
- Example test item (target is second line):

thupro	wassace	uyuwakte	pho
thupro	wassak-ce	u-yuŋ-a-yakt-e	pho
many	bird-NS	3NSS/A-live-PST-IPFV-IND.PST	REP

‘There lived many birds.’ [ctn] story_rabbit.005

Lexical coverage results

Narrative	total		# analyzed		% analyzed		avg ambiguity	
	type	token	type	token	type	token	type	token
Durga_Exp	206	489	120	265	58	54	1.24	1.14
choku_yakkheng	152	331	89	184	59	56	1.26	1.20
pear_6-1	206	433	105	203	56	51	1.20	1.62
story_rabbit	85	200	43	69	51	35	1.37	1.23
All	568	1453	324	721	57	50	1.40	1.27

Morphosyntactic/morphosemantic features: precision/recall

- 10 randomly selected word types from each of four test narratives
- Compared morphosyntactic and morphosemantic features in grammar's analysis to gloss provided in Chintang corpus
- In case of ambiguity, select the best match from the grammar

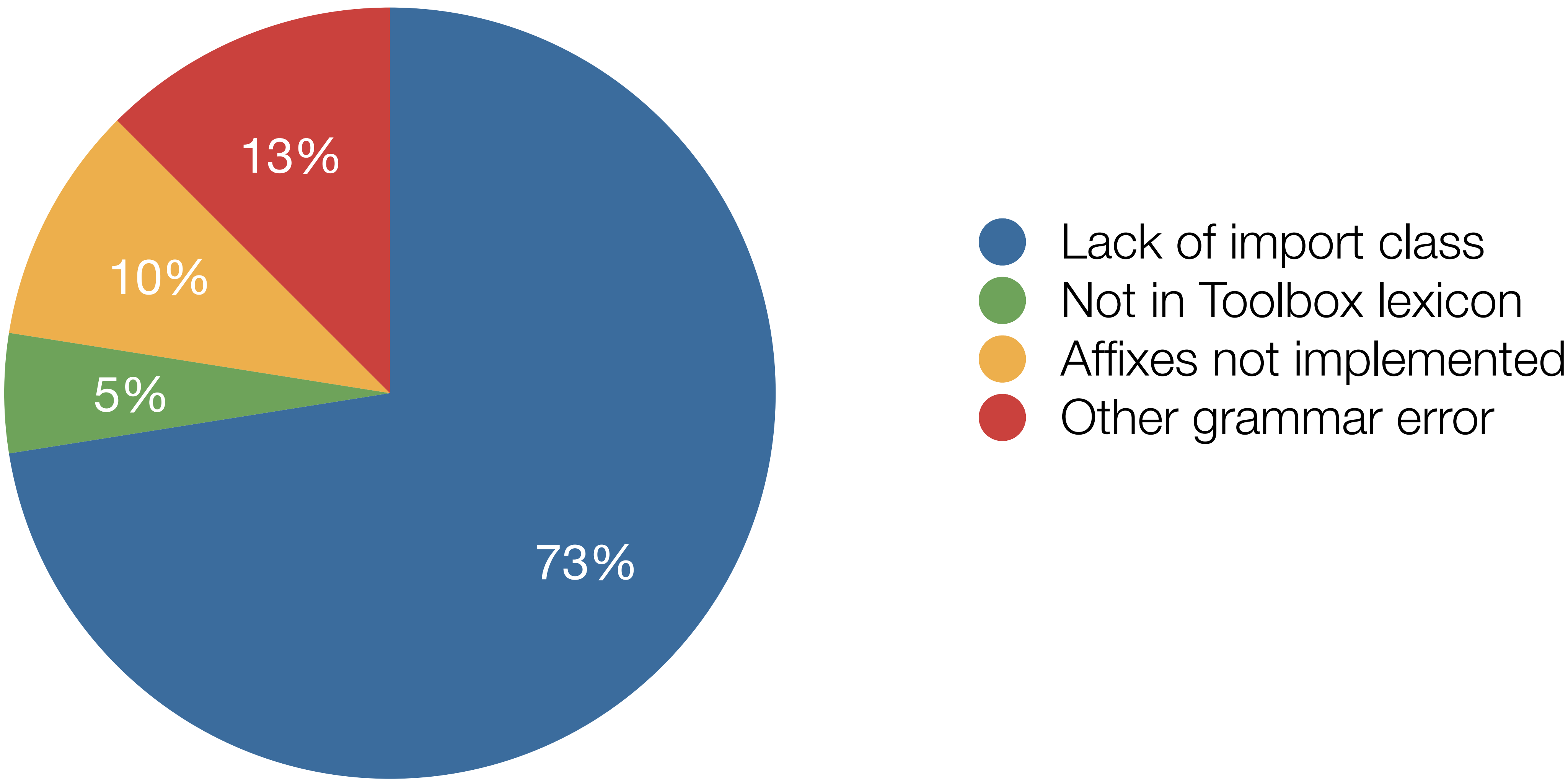
Narrative	Total gold attributes	Precision	Recall	F-measure
Durga_Exp	16	.48	.88	.62
choku_yakkheng	21	.57	.62	.59
pear_6-1	31	.71	.92	.80
story_rabbit	29	.62	.83	.71
Total	97	.61	.82	.70

Error analysis: features

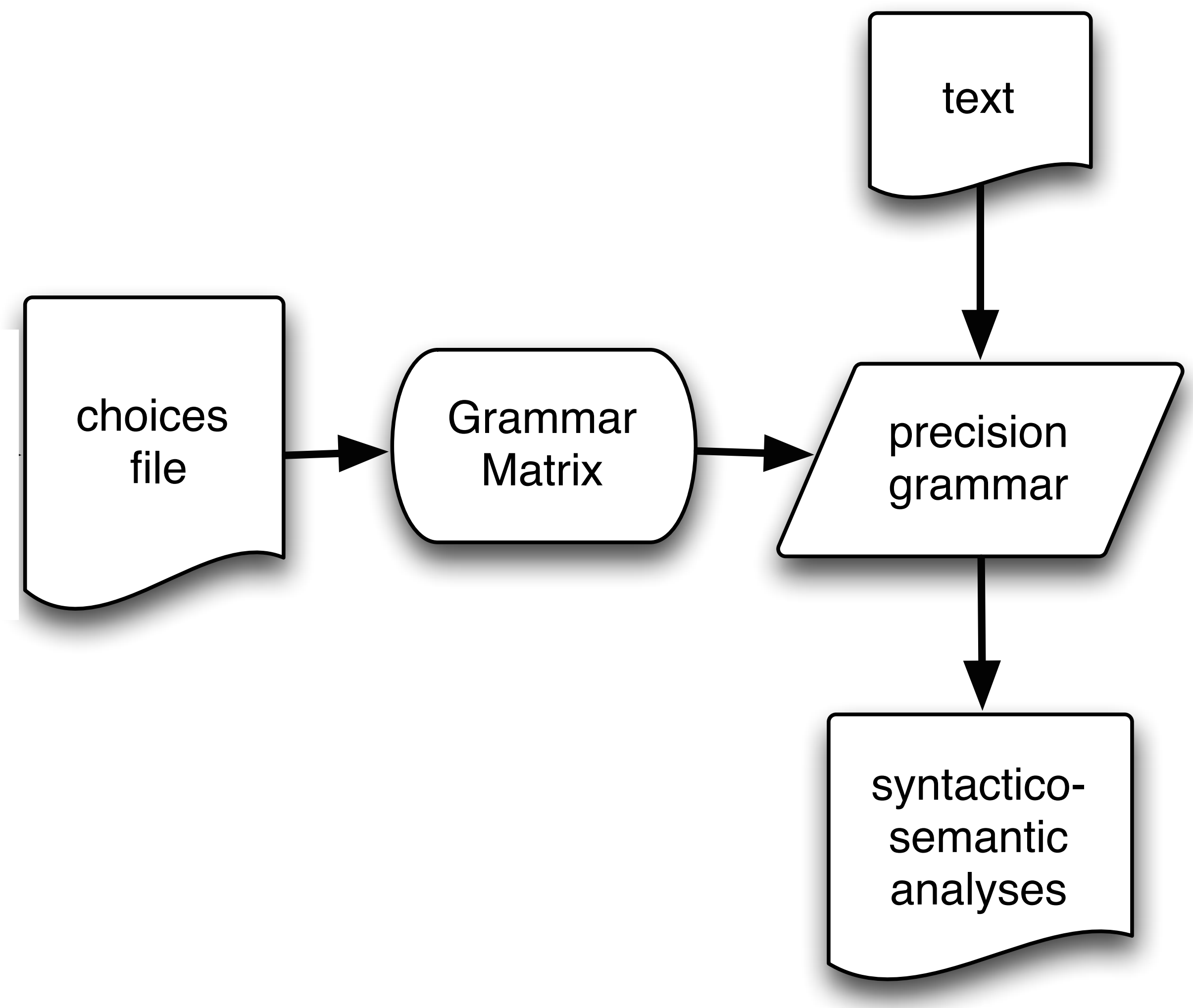
- Glosses in Chintang corpus represent a systematic policy to not gloss zero morphs
- Grammar Matrix-derived grammar includes rules for zeroes
    - Default singular number: 19 errors
    - Default nominative case: 15 errors
    - Default [FORM fin]: 9 errors
    - Total errors in precision attributable to default values: 52 errors
- 18 recall errors are primarily due to cases where the intended lexical root is not available

Error analysis: Lack of coverage

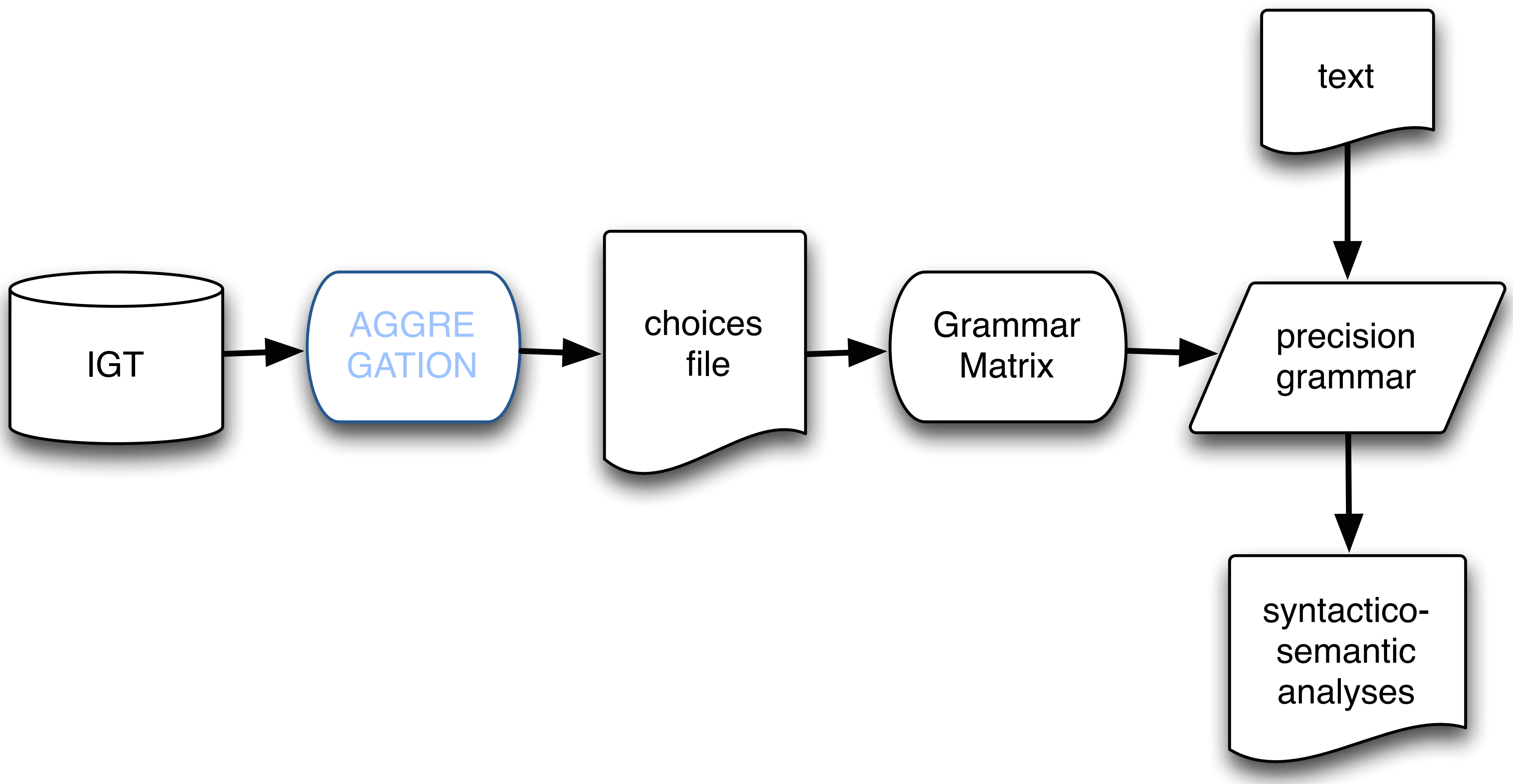
- Randomly selected 10 unanalyzed word forms from each of the four narratives
-



AGGREGATION Project: Overview



AGGREGATION Project: Overview



Choices files

```
word-order=v-final
has-dets=yes
noun-det-order=det-noun
...
case-marking=erg-abs
erg-abs-erg-case-name=erg
erg-abs-abs-case-name=abs
...
verb4_valence=erg-abs
  verb4_stem1_orth=sams-i-ne
  verb4_stem1_pred=_sams-i-ne_v_re
...
verb-pc3_inputs=verb-pc2
  verb-pc3_lrt1_name=2nd-person-subj
    verb-pc3_lrt1_feat1_name=pernum
    verb-pc3_lrt1_feat1_value=2nd
    verb-pc3_lrt1_feat1_head=subj
    verb-pc3_lrt1_lri1_inflecting=yes
    verb-pc3_lrt1_lri1_orth=a-
```

Choices files

```
word-order=v-final
has-dets=yes
noun-det-order=det-noun
...
case-marking=erg-abs
erg-abs-erg-case-name=erg
erg-abs-abs-case-name=abs
...
verb4_valence=erg-abs
  verb4_stem1_orth=sams-i-ne
  verb4_stem1_pred=_sams-i-ne_v_re
...
verb-pc3_inputs=verb-pc2
  verb-pc3_lrt1_name=2nd-person-subj
    verb-pc3_lrt1_feat1_name=pernum
    verb-pc3_lrt1_feat1_value=2nd
    verb-pc3_lrt1_feat1_head=subj
    verb-pc3_lrt1_lri1_inflecting=yes
    verb-pc3_lrt1_lri1_orth=a-
```

} Word order

Choices files

```
word-order=v-final
has-dets=yes
noun-det-order=det-noun
...
case-marking=erg-abs
erg-abs-erg-case-name=erg
erg-abs-abs-case-name=abs
...
verb4_valence=erg-abs
  verb4_stem1_orth=sams-i-ne
  verb4_stem1_pred=_sams-i-ne_v_re
...
verb-pc3_inputs=verb-pc2
  verb-pc3_lrt1_name=2nd-person-subj
    verb-pc3_lrt1_feat1_name=pernum
    verb-pc3_lrt1_feat1_value=2nd
    verb-pc3_lrt1_feat1_head=subj
    verb-pc3_lrt1_lri1_inflecting=yes
    verb-pc3_lrt1_lri1_orth=a-
```

} Word order

} Case system

Choices files

```
word-order=v-final
has-dets=yes
noun-det-order=det-noun
...
case-marking=erg-abs
erg-abs-erg-case-name=erg
erg-abs-abs-case-name=abs
...
verb4_valence=erg-abs
  verb4_stem1_orth=sams-i-ne
  verb4_stem1_pred=_sams-i-ne_v_re
...
verb-pc3_inputs=verb-pc2
  verb-pc3_lrt1_name=2nd-person-subj
    verb-pc3_lrt1_feat1_name=pernum
    verb-pc3_lrt1_feat1_value=2nd
    verb-pc3_lrt1_feat1_head=subj
    verb-pc3_lrt1_lri1_inflecting=yes
    verb-pc3_lrt1_lri1_orth=a-
```

} Word order

} Case system

} Lex types & entries

Choices files

```
word-order=v-final
has-dets=yes
noun-det-order=det-noun
...
case-marking=erg-abs
erg-abs-erg-case-name=erg
erg-abs-abs-case-name=abs
...
verb4_valence=erg-abs
  verb4_stem1_orth=sams-i-ne
  verb4_stem1_pred=_sams-i-ne_v_re
...
verb-pc3_inputs=verb-pc2
  verb-pc3_lrt1_name=2nd-person-subj
    verb-pc3_lrt1_feat1_name=pernum
    verb-pc3_lrt1_feat1_value=2nd
    verb-pc3_lrt1_feat1_head=subj
    verb-pc3_lrt1_lri1_inflecting=yes
    verb-pc3_lrt1_lri1_orth=a-
```

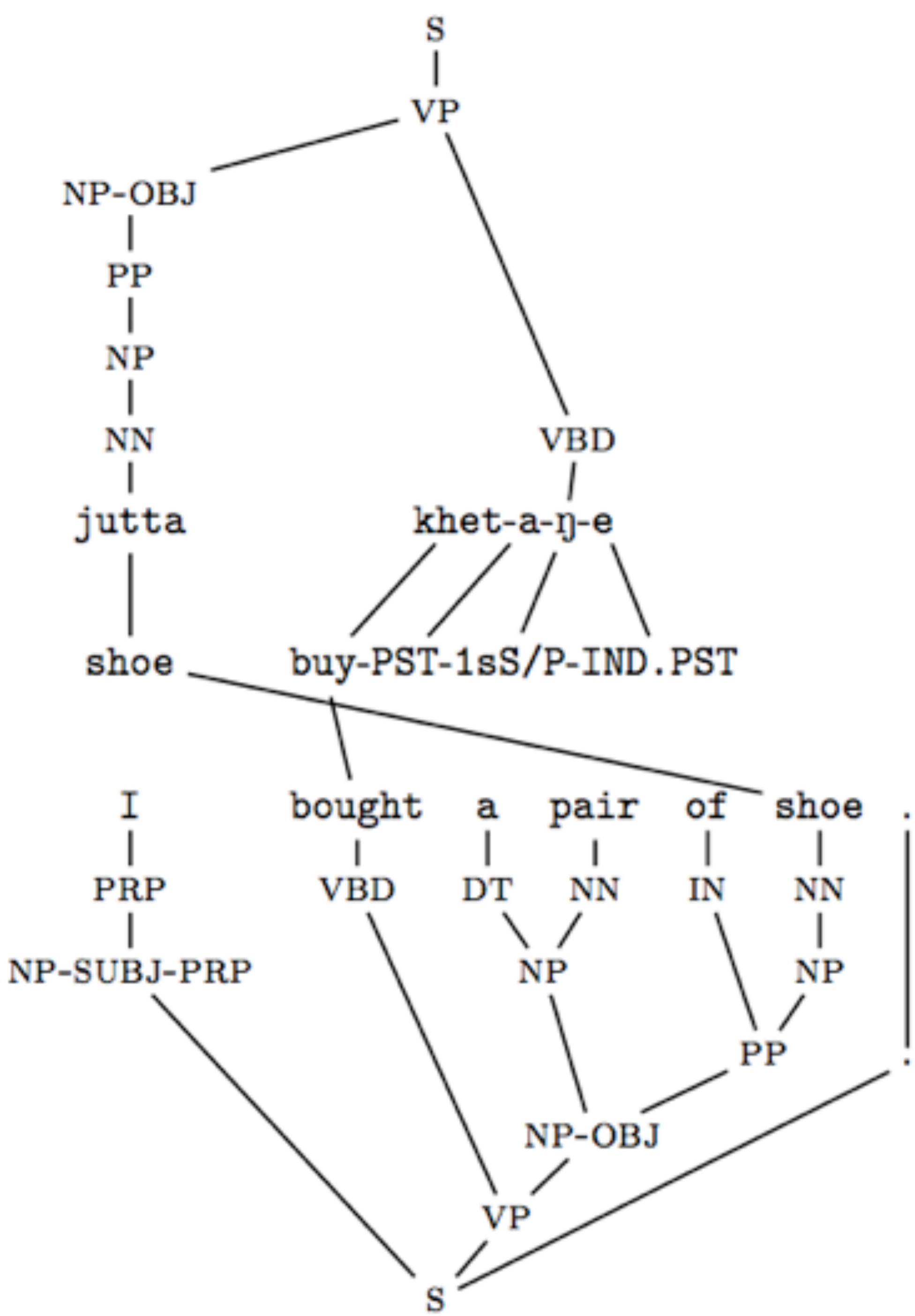
} Word order

} Case system

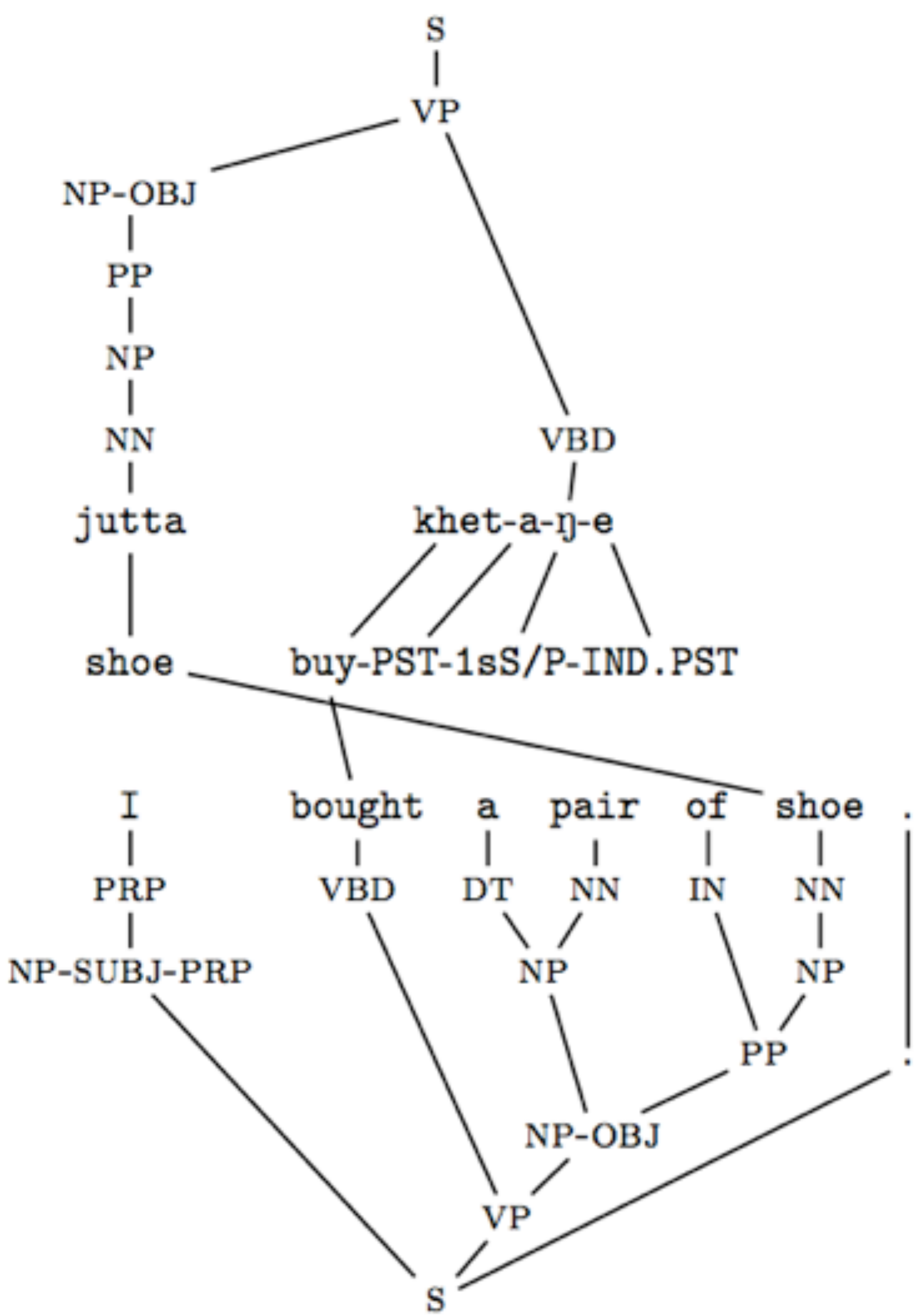
} Lex types & entries

} Lex rules

Projected structure

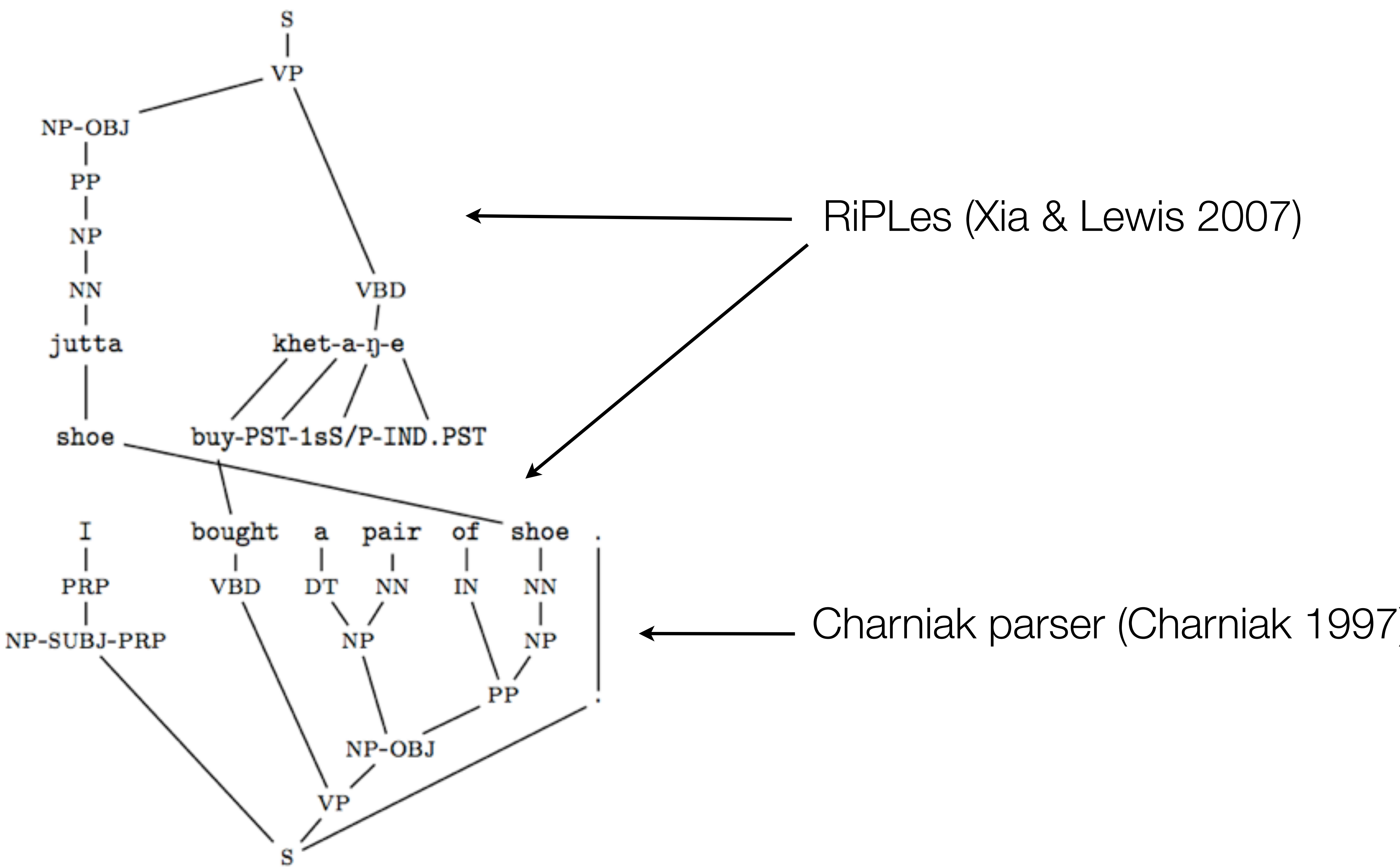


Projected structure



← Charniak parser (Charniak 1997)

Projected structure



Learning General Choices from IGT: Word order
(Bender et al 2013)

Learning General Choices from IGT: Word order
(Bender et al 2013)

- Count S-V and O-V orders in projected trees

Learning General Choices from IGT: Word order
(Bender et al 2013)

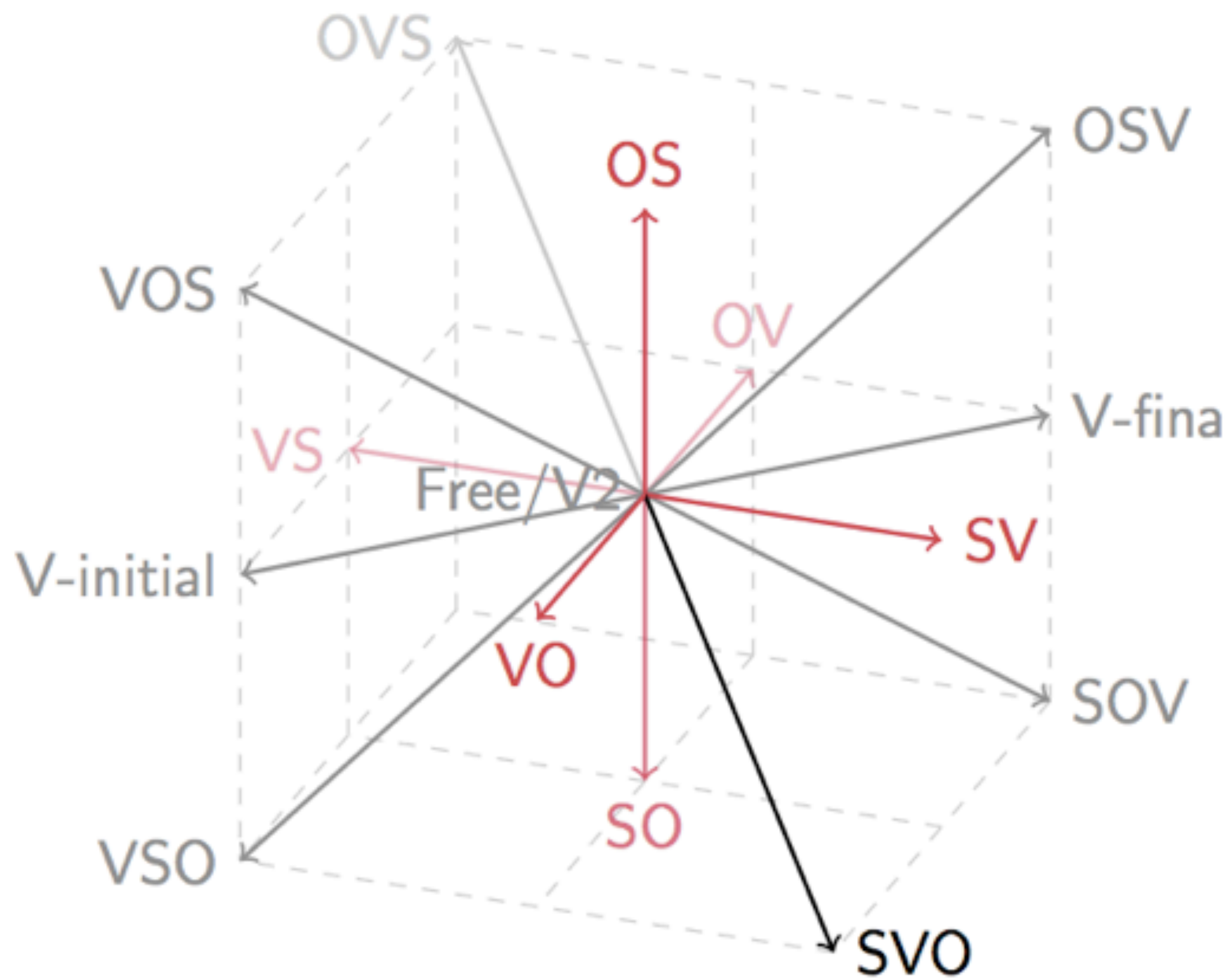
- Count S-V and O-V orders in projected trees
- Calculate observed word order on three axes (SV-VS, OV-VO, SO-OS)

Learning General Choices from IGT: Word order
(Bender et al 2013)

- Count S-V and O-V orders in projected trees
- Calculate observed word order on three axes (SV-VS, OV-VO, SO-OS)
- Compare to canonical points

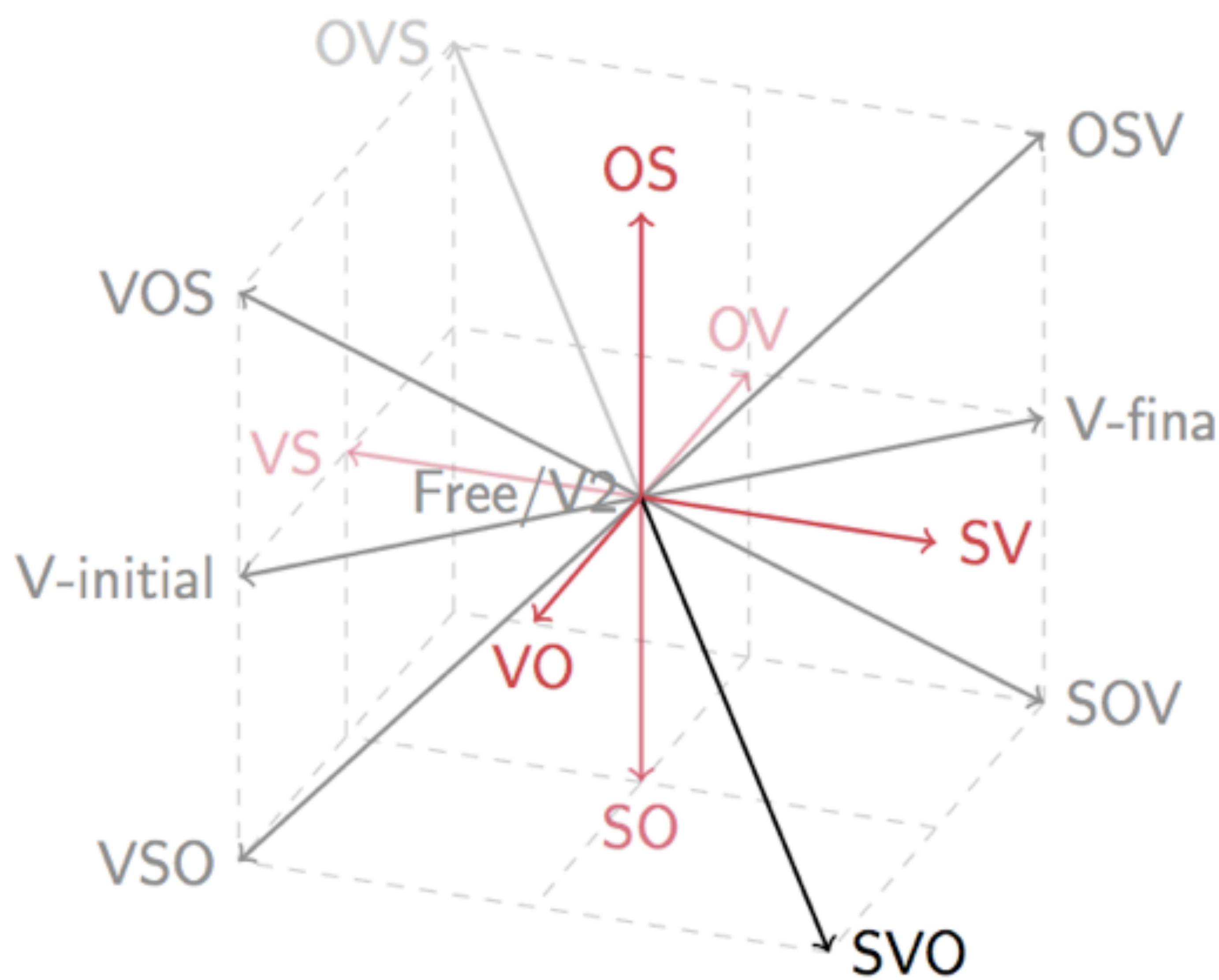
Learning General Choices from IGT: Word order
(Bender et al 2013)

- Count S-V and O-V orders in projected trees
- Calculate observed word order on three axes (SV-VS, OV-VO, SO-OS)
- Compare to canonical points



Learning General Choices from IGT: Word order
(Bender et al 2013)

- Count S-V and O-V orders in projected trees
- Calculate observed word order on three axes (SV-VS, OV-VO, SO-OS)
- Compare to canonical points
- Answer for Chintang: v-final



Learning General Choices from IGT: Case System
(Bender et al 2013)

Case system	Case grams present	
	NOM ∨ ACC	ERG ∨ ABS
none		
nom-acc	✓	
erg-abs		✓
split-erg (conditioned on V)	✓	✓

Learning General Choices from IGT: Case System
(Bender et al 2013)

Case system	Case grams present: NOM ∨ ACC	Case grams present: ERG ∨ ABS
none		
nom-acc	✓	
erg-abs		✓
split-erg (conditioned on V)	✓	✓

- “Gram” method

Learning General Choices from IGT: Case System
(Bender et al 2013)

Case system	Case grams present	
	NOM ∨ ACC	ERG ∨ ABS
none		
nom-acc	✓	
erg-abs		✓
split-erg (conditioned on V)	✓	✓

- “Gram” method
- Results for Chintang: erg-abs

Learning General Choices from IGT: Case System
(Bender et al 2013)

Case system	Top grams
none	$S_g{=}A_g{=}O_g$, or $S_g{\neq}A_g{\neq}O_g$ and S_g, A_g, O_g also present on the other argument types
nom-acc	$S_g{=}A_g$, $S_g{\neq}O_g$
erg-abs	$S_g{=}O_g$, $S_g{\neq}A_g$
tripartite	$S_g{\neq}A_g{\neq}O_g$, and S_g, A_g, O_g absent from others
split-s	$S_g{\neq}A_g{\neq}O_g$, and A_g and O_g both present on S list

Learning General Choices from IGT: Case System
(Bender et al 2013)

Case system	Top grams
none	$S_g{=}A_g{=}O_g$, or $S_g{\neq}A_g{\neq}O_g$ and S_g, A_g, O_g also present on the other argument types
nom-acc	$S_g{=}A_g$, $S_g{\neq}O_g$
erg-abs	$S_g{=}O_g$, $S_g{\neq}A_g$
tripartite	$S_g{\neq}A_g{\neq}O_g$, and S_g, A_g, O_g absent from others
split-s	$S_g{\neq}A_g{\neq}O_g$, and A_g and O_g both present on S list

- “SAO” method

Learning General Choices from IGT: Case System
(Bender et al 2013)

Case system	Top grams
none	S_g=A_g=O_g, or S_g≠A_g≠O_g and S_g, A_g, O_g also present on the other argument types
nom-acc	S_g=A_g, S_g≠O_g
erg-abs	S_g=O_g, S_g≠A_g
tripartite	S_g≠A_g≠O_g, and S_g, A_g, O_g absent from others
split-s	S_g≠A_g≠O_g, and A_g and O_g both present on S list

- “SAO” method
- Results for Chintang: None :(

Learning Case Frames from IGT
(Bender et al 2014)

- For each IGT with overt subject and/or object,
- use grams to find case assigned to subject and object NPs
- no case gram => use default case value for that position

```
# find 1-arg verbs (not dropped subj)
@S < (@VP < @VB=verb !< @OBJ) < @SBJ=sbj !< @OBJ;
# find 1-arg verbs for OSV or VSO langs
@S < @VB=verb < @SBJ=sbj !< @OBJ;

# find 2-arg verbs, both args overt
@S < (@VP < @VB=verb < @OBJ=obj) < @SBJ=sbj;
# 2-arg verbs for OSV or VSO
@S < @VB=verb < @SBJ=sbj < @OBJ=obj;

# find 2-arg verbs w/dropped sbj
@S < (@VP < @VB=verb < @OBJ=obj) !< @SBJ=sbj;
# 2-arg verbs for OSV or VSO
@S < @VB=verb !< @SBJ=sbj < @OBJ=obj;
```

Putting It All Together: Interim Results
(Bender et al 2014)

Choices file	# verb entries	# noun entries	# det entries	# verb affixes	# noun affixes
ORACLE	900	4751	0	160	24
BASELINE	3005	1719	240	0	0
FF-AUTO-NONE	3005	1719	240	0	0
FF-DEFAULT-GRAM	739	1724	240	0	0
FF-AUTO-GRAM	739	1724	240	0	0
MOM-DEFAULT-NONE	1177	1719	240	262	0
MOM-AUTO-NONE	1177	1719	240	262	0

Putting It All Together: Interim Results
(Bender et al 2014)

	Training Data (N = 8863)				Test Data (N = 930)			
choices file	lexical coverage (%)	items parsed (%)	items correct (%)	average readings	lexical coverage (%)	items parsed (%)	items correct (%)	average readings
ORACLE	1165 (13)	174 (3.5)	132 (1.5)	2.17	116 (12.5)	20 (2.2)	10 (1.1)	1.35
BASELINE	1276 (14)	398 (7.9)	216 (2.4)	8.30	41 (4.4)	15 (1.6)	8 (0.9)	28.87
FF-AUTO-NONE	1276 (14)	354 (4.0)	196 (2.2)	7.12	41 (4.4)	13 (1.4)	7 (0.8)	13.92
FF-DEFAULT-GRAM	911 (10)	126 (1.4)	84 (0.9)	4.08	18 (1.9)	4 (0.4)	2 (0.2)	5.00
FF-AUTO-GRAM	911 (10)	120 (1.4)	82 (0.9)	3.84	18 (1.9)	4 (0.4)	2 (0.2)	5.00
MOM-DEFAULT-NONE	1102 (12)	814 (9.2)	52 (0.6)	6.04	39 (4.2)	16 (1.7)	3 (0.3)	10.81
MOM-AUTO-NONE	1102 (12)	753 (8.5)	49 (0.6)	4.20	39 (4.2)	10 (1.1)	3 (0.3)	9.20

Next Steps

- Combine MOM and gram approaches
- Refine search patterns to capture more lexical entries in gram approach
- Extend MOM to noun morphology
- Extend approach to learning other morphosyntactic properties (both lexically specific, and language-level)
- Extend Grammar Matrix customization system to handle more phenomena
- Test on more languages

References

Bender, Emily M. 2008. Evaluating a crosslinguistic grammar resource: A case study of Wambaya. In *Proceedings of ACL-08: HLT*, 977–985, Columbus, Ohio, June. Association for Computational Linguistics.

Bender, Emily M., Joshua Crowgey, Michael Wayne Goodman, and Fei Xia. 2014. Learning grammar specifications from igt: A case study of chintang. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, 43–53, Baltimore, Maryland, USA, June. Association for Computational Linguistics.

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation* 1–50. 10.1007/s11168-010-9070-1.

Bender, Emily M., Dan Flickinger, and Stephan Oepen. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In J. Carroll, N. Oostdijk, and R. Sutcliffe (Eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, 8–14, Taipei, Taiwan.

Bender, Emily M., Sumukh Ghodke, Timothy Baldwin, and Rebecca Dridan. 2012. From database to treebank: Enhancing hypertext grammars with grammar engineering and treebank search. In S. Nordhoff and K.-L. G. Poggeman (Eds.), *Electronic Grammaticography*, 179–206. Honolulu: University of Hawaii Press.

Bender, Emily M., and Jeff Good. 2005. Implementation for discovery: A bipartite lexicon to support morphological and syntactic analysis. In *Proceedings from the Panels of the Forty-First Meeting of the Chicago Linguistic Society: Volume 41-2*.

Bender, Emily M., Michael Wayne Goodman, Joshua Crowgey, and Fei Xia. 2013. Towards creating precision grammars from interlinear glossed text: Inferring large-scale typological properties. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 74–83, Sofia, Bulgaria, August. Association for Computational Linguistics.

Bickel, Balthasar, Martin Gaenszle, Novel Kishore Rai, Vishnu Singh Rai, Elena Lieven, Sabine Stoll, G. Banjade, T. N. Bhatta, N Paudyal, J Pettigrew, and M Rai, I. P.and Rai. 2013. Tale of a poor guy. Accessed: 15 January 2013.

Chafe, Wallace. 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood NJ: Ablex Pub. Corp.

Charniak, Eugene. 1997. Statistical parsing with a context-free grammar and word statistics. In *Proceedings of AAAI-1997*.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford, CA: CSLI Publications.

Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language & Computation* 3(4):281–332.

Goodman, Michael Wayne. 2013. Generation of machine-readable morphological rules with human readable input. *UW Working Papers in Linguistics* 30.

Green, Ian, and Rachel Nordlinger. 2004. Revisiting Proto-Mirndi. In C. Bowern and H. Koch (Eds.), *Australian Languages: Classification and the Comparative Method*, 291–311. Amsterdam: John Benjamins.

Nordlinger, Rachel. 1998. *A Grammar of Wambaya, Northern Australia*. Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Oepen, Stephan. 2001. [incr tsdb()] — Competence and performance laboratory. User manual. Technical report, Computational Linguistics, Saarland University, Saarbrücken, Germany.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL and Stanford, CA: The University of Chicago Press and CSLI Publications.

Schikowski, Robert. 2012. Chintang morphology. Unpublished ms., Unpublished ms, University of Zürich.

Wax, David. 2014. Automated grammar engineering for verbal morphology. Master's thesis, University of Washington.

Xia, Fei, and William Lewis. 2007. Multilingual structural projection across interlinearized text. In *NAACL-HLT 2007*, Rochester, NY.