

# Modeling clausal complementation in the LinGO Grammar Matrix

## Presentation for the DELPH-IN summit

Olga Zamaraeva, Kristen Howell and Emily M. Bender

Department of Linguistics, University of Washington

June 18 2018

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

Error Analysis

Conclusion

References

Extra slides

- ▶ LinGO Grammar Matrix (Bender et al. 2002, 2010)
  - ▶ Grammar Engineering toolkit
  - ▶ Typological breadth + syntactic theory (HPSG) depth and precision
  - ▶ Libraries: Word order, Case, Morphotactics, and other
  - ▶ Previously no subordinate clauses
- ▶ **My contribution:** Clausal Complements library implementation, with underlying cross-linguistic account in HPSG
  - ▶ *I know [that Kim left]*

## Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

Error Analysis

Conclusion

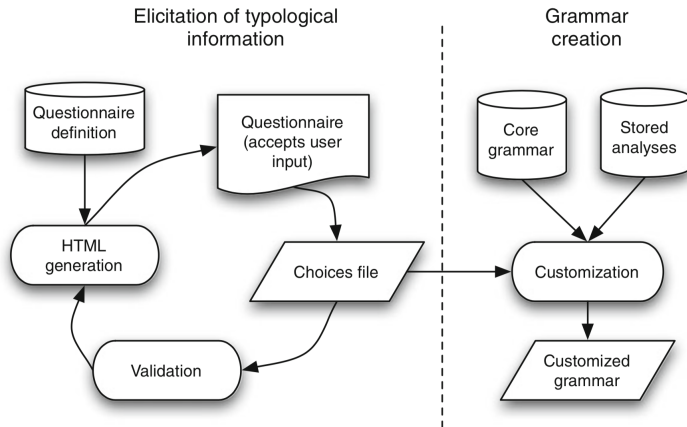
References

Extra slides

# Grammar Matrix: A customization system

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit



from Bender et al. (2010)

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages  
Development languages

Evaluation

Held-out languages  
Error Analysis

Conclusion

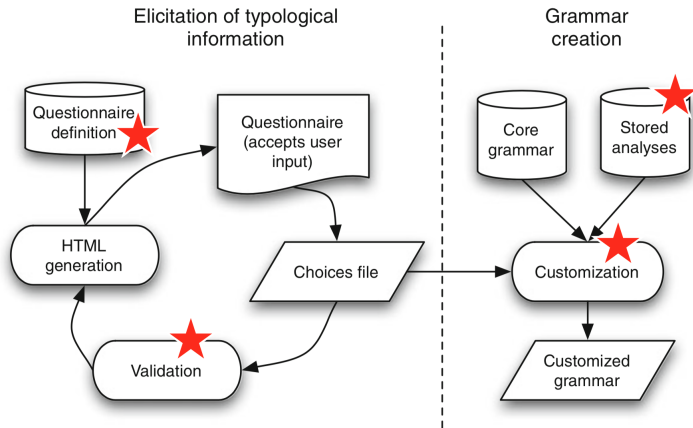
References

Extra slides

# Grammar Matrix: A customization system

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit



Bender et al. (2010)

from

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages  
Development languages

Evaluation

Held-out languages  
Error Analysis

Conclusion

References

Extra slides

# Library: typological scope

- ▶ The library covers:
  - ▶ Clauses appearing as objects of verbs
    - ▶ finite, full propositions
  - ▶ Complement clauses marked by complementizers
    - ▶ appearing before or after the clause
    - ▶ regardless of the basic word order
  - ▶ Extraposition
    - ▶ SOV, VOS, V-initial, OVS, OSV
  - ▶ Complement clauses marked morphologically
    - ▶ Morphotactics library machinery
  - ▶ Joint work with Howell:
    - ▶ German-like word order in sentences with subordinate clauses (incorporate analysis by Fokkens (2014))
    - ▶ Nominalized clausal complements

# Typology: Objectival clausal complements

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

Error Analysis

Conclusion

References

Extra slides

Finite, marked by a complementizer:

(1) Kim thinks [that Sandy left] [eng]

# Typology: Extraposition

Examples from Malagasy (Dryer 1980):

(2) na-mono an-dRabe Rakoto  
PAST-hit ACC-Rabe Rakoto  
'Rakoto hit Rabe.' [mlg]

(3) Mihevitra Rabe [fa mitady ny zaza Raso]a  
thinks Rabe [COMP look.for the child Raso]a  
'Rabe thinks that Raso is looking for the child.' [mlg]

# Order of subordinator and subordinate clause

Examples from Uzbek (Noonan 2007):

(4) Men bilamen [ki bu odam joʻja-ni oʻgʻirladi]  
I know-1sg [COMP this man chicken-OBJ stole-3sg]  
'I know that the man stole the chicken.' [uzb]

(5) Xotin [bu odam joʻja-ni oʻgʻirladi deb] dedi  
woman [this man chicken-OBJ stole COMP] said.3sg  
'The woman said that the man stole a chicken.' [uzb]



# Lexical types: Clausal complement-taking verb

(6) 
$$\left[ \begin{array}{l} \textit{clausal-verb-lex} \\ \text{MOD} \langle \rangle \\ \text{SUBJ} \langle \boxed{1} \rangle \\ \text{COMPS} \langle \boxed{2} \left[ \begin{array}{l} \text{SPR} \langle \rangle \\ \text{SUBJ} \langle \rangle \\ \text{COMPS} \langle \rangle \end{array} \right] \rangle \\ \text{ARG-ST} \langle \boxed{1} \left[ \text{HEAD } \textit{noun} \right], \boxed{2} \rangle \end{array} \right]$$

- ▶ The complement of this verb can be further specified:
  - ▶ entity (nominalized clausal complement)
  - ▶ event (non-nominalized clausal complement)

# Phrase structure rules

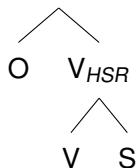
- ▶ Each Matrix-generated grammar will have:
  - ▶ at least one Head-Subject Rule (HSR)
  - ▶ at least one Head-Complement Rule (HCR)
  - ▶ ...to account for the basic word order
- ▶ Need **additional** rules to accommodate for variation associated with clausal complements
  - ▶ use **features** to not overgenerate:
  - ▶ **INIT +/-** to account for order of complementizer and clause and extraposition in V-final languages
  - ▶ **EXTRA +/-** for extraposition in V-initial languages

# Sample analysis: Extraposition in OVS

(7)

$HCR1$			
COMPS	$\boxed{1}$		
H-DTR	INIT	-	
	SUBJ	$\langle \rangle$	
	COMPS	$\langle \boxed{3} \oplus \boxed{1} \rangle$	
NH-DTR	$\boxed{3}$		
ARGS	$\langle \boxed{3}, \boxed{2} \rangle$		

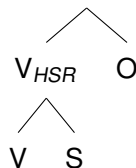
$Top_{HCR1}$



(8)

$HCR2$			
COMPS	$\boxed{1}$		
H-DTR	INIT	+	
	SUBJ	$\langle \rangle$	
	COMPS	$\langle \boxed{3} \oplus \boxed{1} \rangle$	
NH-DTR	$\boxed{3}$		
ARGS	$\langle \boxed{2}, \boxed{3} \rangle$		

$Top_{HCR2}$



Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

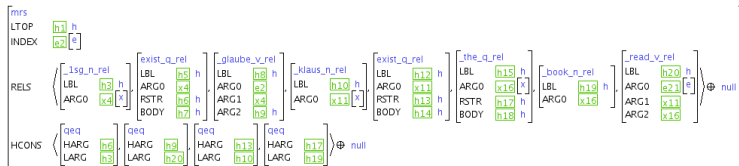
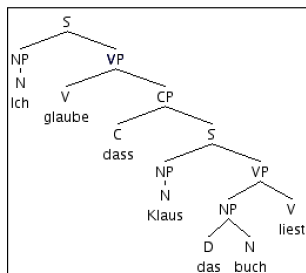
Error Analysis

Conclusion

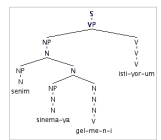
References

Extra slides

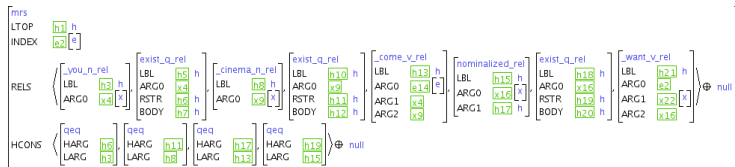
# Semantics of clausal complements



# Semantics of nominalized clausal complements



- ▶ Nominalized clauses in LinGO Grammar Matrix (Howell et al. to appear)
- ▶ Main clause verb is looking for a complement of type *ref* (x, individual)
- ▶ Embedded verb still needs to have a subject
- ▶ A new HSR required: *non-event-subject-head-phrase*



# Combinatorics of user choices

dimension	choices
word order (extraposition)	6
complementizer position and form	10
extraposition	3
nominalization	5
verb feature	2
subtotal	+1800
word order (no extraposition)	3
complementizer position and form	10
nominalization	5
verb feature	2
subtotal	+300
word order=V2	1
subordinate word order	2
complementizer position and form	10
nominalization	5
verb feature	2
subtotal	+200
<b>total</b>	<b>2300</b>

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

Error Analysis

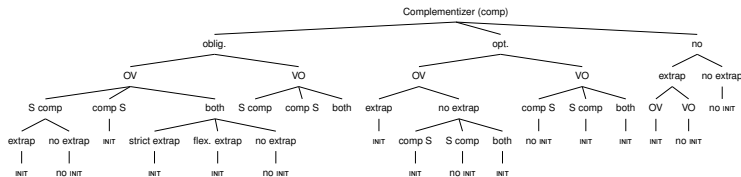
Conclusion

References

Extra slides

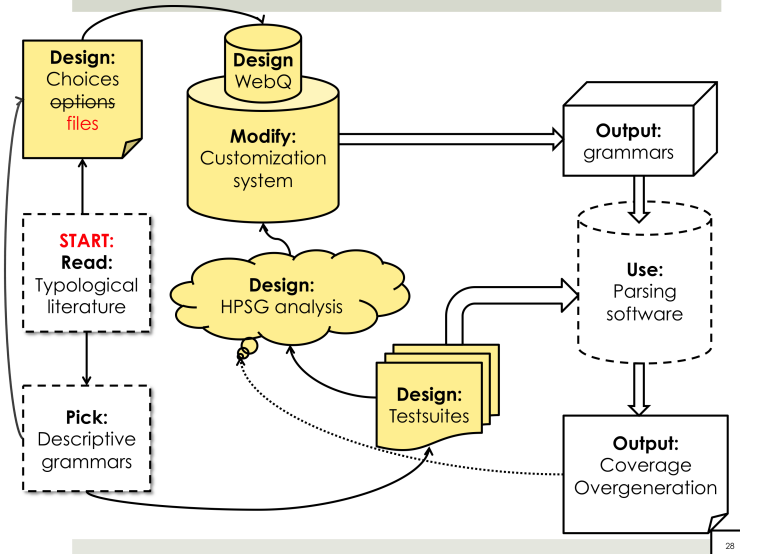
# Sample customization logic

- ▶ FORM, ASPECT, MOOD, VC (verb cluster), NMZ (nominalized) - infer directly from user choices, per strategy
- ▶ INIT, EXTRA: infer from combination of choices



Decision tree illustrating the logic of using the INIT feature based on user choices

# Test-driven development



Modeling clausal  
 complementation  
 in the LinGO  
 Grammar Matrix

Presentation for the  
 DELPH-IN summit

Introduction

Typological scope

Syntactic analysis

Semantic  
 representations

Test-Driven  
 development

Pseudolanguages  
 Development languages

Evaluation

Held-out languages  
 Error Analysis

Conclusion

References

Extra slides



# Coverage and overgeneration: Pseudolanguages

- ▶ 50 languages sample, each choice value used at least once
  - ▶ E.g.: *OVS/OVS, obligatory complementizer before or after clause, no nominalization, no extraposition*
  - ▶ E.g.: *V-final/V-final, obligatory complementizer before or after clause, no nominalization, obligatory extraposition*
- ▶ 100% coverage and 0% overgeneration
  - ▶ ...but this is not evaluation yet; this is test-driven development

# Development languages

Coverage=100%, overgeneration=0% across the board (still not yet proper evaluation)

Language	iso639	fam	WO	comp	order	morph	extrap	# strat	pos	neg
Russian	rus	IE	free	opt	comp S	nmz,form	-	3	6	11
German	deu	IE	V2/V-fin	oblig <sup>1</sup>	comp S.	-	-	1	6	4
Tagalog	tgl	Astrn.	V-in	oblig	comp S	-	flexible	1	3	4
Lango	laj	NS	SVO	oblig	comp S	sbjunct.	-	3	4	4
Turkish	tur	Tur	SOV	opt	both	nmz, sbjunct.	strict	4	7	9

# Held-out languages

This is true evaluation:

Language	iso639	fam	WO	comp	order	morph	extrap	# strat.	Cov.	Overgen.
Jalkunan	bxl	NC	SOV	opt	comp S	-	strict	1	4/8	0/12
Paresi-Haliti	pab	Awk	SOV	-	-	nmz	strict	1	4/4	0/6
Yakima Sahaptin	yak	PP	free	-	-	nmz	-	1	10/10	0/6
Modern Hebrew	heb	AA	SVO	oblig	comp S	-	-	1	2/2	0/9
Wangkangurru	wgg	PN	free	-	-	aspect	-	1	10/10	0/3

- ▶ Jalkunan
  - ▶ a strategy not seen in typological survey
  - ▶ a “dummy” in-situ pronoun **along with** extraposition
- ▶ otherwise perfect scores
  - ▶ but testsuites were compiled with severe limitations:
    - ▶ not enough examples with simple clausal complements
    - ▶ sources can be incomplete or vague

(9) ma n so [[ma je] see]  
1Sg **3SgNonhObj** know.Pfv [[1Sg father] come.Pgv]  
'I know that my father has come.' [bxl] (Heath 2017)

# Insufficient integration: Information Structure Library

- ▶ Information Structure library (Song 2014) constrains head-subj rule's NHD to be MC +
- ▶ Clausal complements and Clausal Mods libraries want the subordinate clause to be MC -
- ▶ Clash!

# Insufficient integration: Auxiliaries and Word Order

- ▶ Word order bugs: INIT constraints not assigned correctly in some languages with auxiliaries
- ▶ Need to fully(?) integrate Clausal Complements library with the Word Order library

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

Error Analysis

Conclusion

References

Extra slides

# Conclusion

- ▶ New Grammar Matrix library for clausal complements
- ▶ Main challenge: output streamlined grammars operating in large space of typological possibilities
- ▶ Future work:
  - ▶ Better interaction with other libraries
  - ▶ Subject sharing
  - ▶ Wh-complements
  - ▶ ...leading into my next project\*: Wh-questions in the Grammar Matrix

# References

- Emily M. Bender, 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 John Carroll, Nelleke Oostdijk, and Richard Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan, 2002.
- Emily M Bender, Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation*, 8(1):23–72. ISSN 1570-7075. URL <http://dx.doi.org/10.1007/s11168-010-9070-1>. 10.1007/s11168-010-9070-1.
- Antske Sibelle Fokkens. 2014. *Enhancing Empirical Research for Linguistically Motivated Precision Grammars*. PhD thesis, Department of Computational Linguistics, Universität des Saarlandes.
- Jeffrey Heath. 2017. *A grammar of Jalkunan (Mande, Burkina Faso)*. Language Description Heritage Library.
- Kristen Howell, Olga Zamaraeva, and Emily M. Bender. to appear. Nominalized clauses in the grammar matrix. In *The 25th International Conference on Head-Driven Phrase Structure Grammar*.
- Michael Noonan. 2007. In Timothy Shopen, editor, *Language Typology and Syntactic Description*, volume 2, chapter Complementation. Cambridge University Press, Cambridge, UK.
- Sanghoun Song. 2014. *A grammar library for information structure*. PhD thesis, University of Washington.

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages  
Development languages

Evaluation

Held-out languages  
Error Analysis

Conclusion

References

Extra slides



Verb-final in V2 German:

- (10) Ich glaube [dass Klaus das Buch liest]  
I know [COMP Klaus the book reads]  
'I think that Klaus is reading the book.' [deu] (Fokkens,  
2014)

# GM Library development steps

(established previously)

- ▶ Typological survey
  - ▶ Defining the scope
- ▶ Designing the questionnaire subpage
  - ▶ reducing typological descriptions to *choices*
- ▶ HPSG analysis
  - ▶ lexical types, phrase structure, rules, features...
- ▶ Mapping the analysis to TDL
  - ▶ TDL is HPSG-based machine-readable formalism
- ▶ Python implementation
  - ▶ Produce correct TDL based on a combination of choices
- ▶ Testing
  - ▶ pseudo, illustrative, and **held-out** languages (testsuites)
- ▶ Documenting

Modeling clausal  
complementation  
in the LinGO  
Grammar Matrix

Presentation for the  
DELPH-IN summit

Introduction

Typological scope

Syntactic analysis

Semantic  
representations

Test-Driven  
development

Pseudolanguages

Development languages

Evaluation

Held-out languages

Error Analysis

Conclusion

References

Extra slides

# Adding types

- ▶ ...From the customization system's point of view!
  - ▶ I develop the algorithms which make decisions which types/constraints to emit based on user choices
- ▶ Sometimes the algorithm is simple:
  - ▶ always add a clausal complement-taking verb
  - ▶ e.g. user said there is a complementizer:
    - ▶ check that the Polar Questions library hasn't already added a complementizer supertype and add it
  - ▶ add subtypes, one per strategy
- ▶ With additional HSR/HCR, it is more complex:
  - ▶ need to check **combinations** of choices
  - ▶ general idea: Add additional rule when either the complementizer or the clausal comp.-taking verb cannot use the basic rule
  - ▶ e.g. an additional HSR for VOS orders with extraposition