Modeling causative complexity across languages with the Interclausal Relations Hierarchy

RRG 2019

UB, August 19-21, 2019

Erika Bellingham[1], Pia Järnefelt[2], Kazuhiro Kawachi[3], Yu Li[1], Alice Mitchell[4], Guillermo Montero-Melis[5], Sang-Hee Park[1], Anastasia Stepanova[1], Manne Bylund[2], and Jürgen Bohnemeyer[1]

University at Buffalo The State University of New York 1

Stockholm University 2

National Defense Academy 1952 3

University of Bristol 4

Max Planck Institute for Psycholinguistics 5

NSF

SYNOPSIS

- Introducing CAL
- A new study design for semantic typology
- Variables and stimuli: the CAL Clips
- Preliminary findings
- Summary

INTRODUCING CAL

- Causality Across Languages
  - NSF Award #BCS-1535846; PI J. Bohnemeyer
- a new horizon in semantic typology: causality
  - first ever large-scale *meaning-based* crosslinguistic study of the representation of causality

- subprojects

  - *The semantic typology of causality*
    - how are causal chains semantically categorized across languages for the purposes of linguistic encoding?
  - *The representation of causality in discourse*
    - how are causal chains represented in narratives across languages?
  - *Causality at the syntax-semantics interface*
    - how much variation is there across languages in form-to-meaning mapping in the representation of causal chains?
  - *Causality in language and cognition*
    - how are causal chains cognitively categorized across cultures and what role does language play in this variation?

- the sample

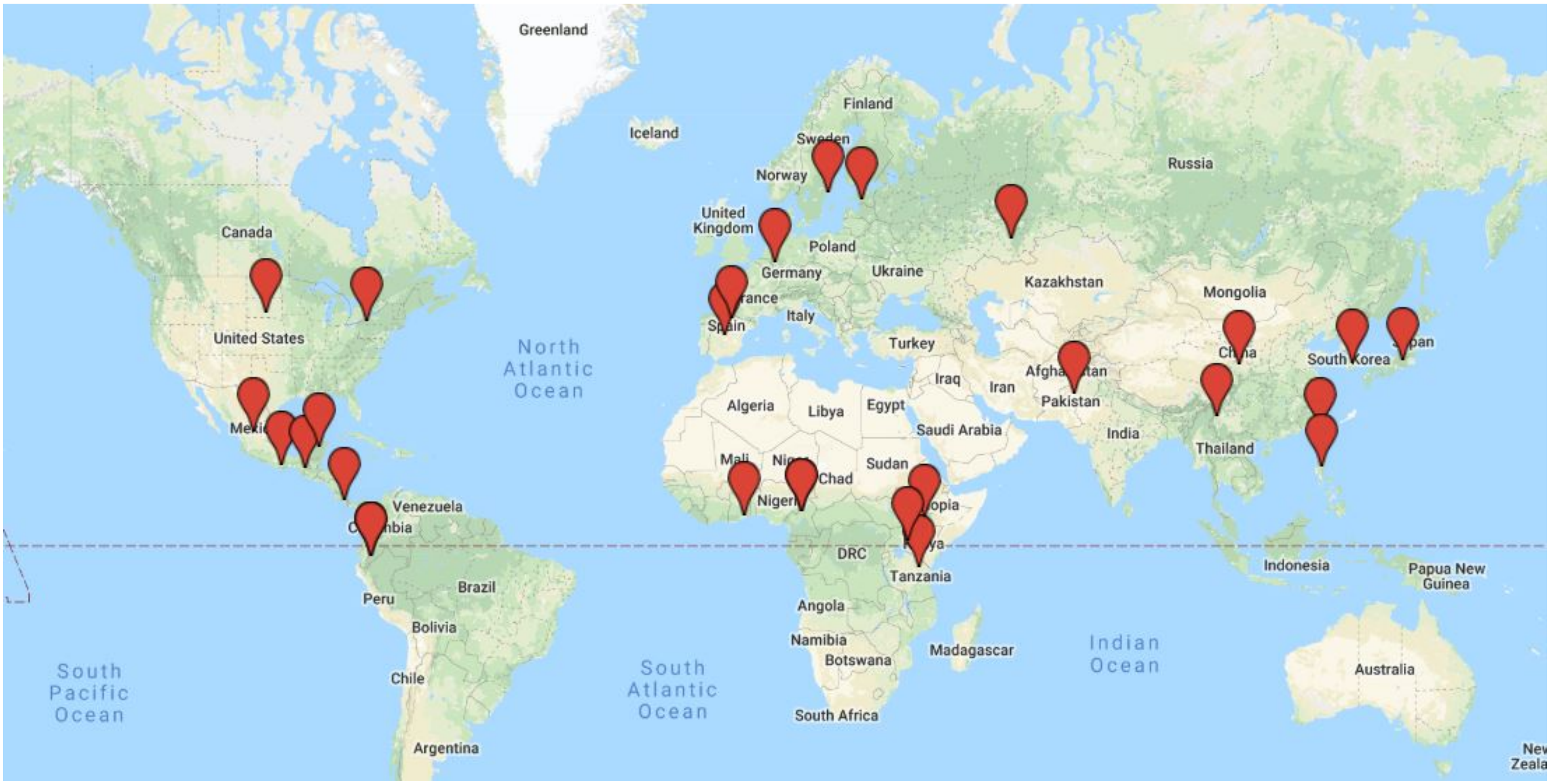


Figure 1.1. Big map, lotsa languages, southern void
Figure 1.1. *Big map, lotsa languages, southern void*

SYNOPSIS

- Introducing CAL
- A new study design for semantic typology
- Variables and stimuli: the CAL Clips
- Preliminary findings
- Summary

A NEW STUDY DESIGN FOR SEMANTIC TYPOLOGY

- domain: form-meaning mapping in causatives
  - the 'Iconicity Principal' (Haiman 1983): simple 'direct' causal chains favor simple causative constructions

(2.1) Le=máak=o' t-u=nik-ah le=bàaso-s-o'b=o'
YUC DEF=person=D2 PRV-A3=scatter-CMP(B3SG) DEF=cup-PL-PL=D2
'The man, he scattered the cups'

Figure 2.1. *HO5_cuptower*

- the Iconicity Principle (cont.)
  - while more complex constructions/descriptions are preferred for more complex, 'indirect' chains
    - e.g. Bohnemeyer et al (2010); Comrie (1981); Dixon (2000); Haiman (1983); Haspelmath (2008); Kemmer & Verhagen (1994); Levin & Rappaport-Hovav (1995); Levshina 2015, 2016, 2017; McCawley (1976, 1978); Shibatani ed. (1976); Shibatani & Pardeshi (2002); Talmy (1976); Verhagen & Kemmer (1997); *inter alia*

(2.2) YUC

a. #Le=x-ch'úupal=o' t-u=nik-ah le=bàaso-s-o'b=o'
DEF=female:child=D2 PRV-A3=shatter+slap-APP-CMP(B3SG) DEF=cup-PL-PL=D2
'The girl, she scattered the cups'

b. Le=x-ch'úupal=o' t-u=mèet-ah
DEF=F-female:child=D2 PRV-A3=make-CMP(B3SG)
u=nik-ik le=bàaso-o'b le=máak=o'
A3=scatter-INC(B3SG) DEF=cup-PL DEF=person=D2
'The girl, she made the man scatter the cup'

Figure 2.2. *HUO2_cups*

- our research question: what exactly does ‘simple’ or ‘direct’ mean - and does it mean the same thing across languages?
  - some candidate variables (cf. Bohnemeyer et al 2010; Dixon 2000)
    - **mediation** - the presence/absence of an intermediate subevent b/w cause and effect
      - ≈ an intermediate participant (CE) b/w CR and AF
    - **prototypicality** - the extent to which the causal chain conforms to the prototypical agent-patient schema
      - hypothesized to be associated with simple transitive causative clauses (Hopper & Thompson 1980)
      - in particular, **agentivity:** the extent to which the causer is a prototypical intentional human agent

- some candidate variables (cont.)
  - **domain** - physical/biological vs. psychological vs. social causation
  - **force dynamics** - causation vs. letting/enabling (Talmy 1988)
  - **contiguity** of subevents - absence/presence of temporal/spatial gaps b/w subevents

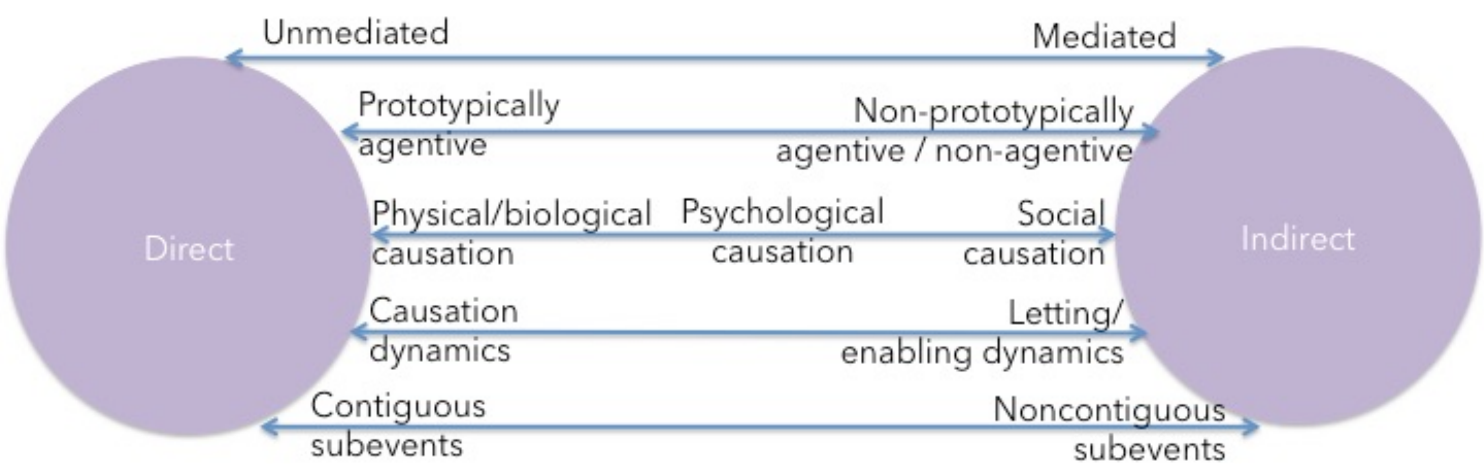


Figure 2.3. *A multidimensional continuum model of causation directness*

- previous quantitative studies
  into the form-meaning mapping in causatives
  - typological “library” studies: Escamilla 2012
  - elicited production studies: Bohnemeyer et al 2010
  - corpus-based studies:
    Haspelmath 2008: 22-23; Levshina 2015, 2016, 2017

- a new approach

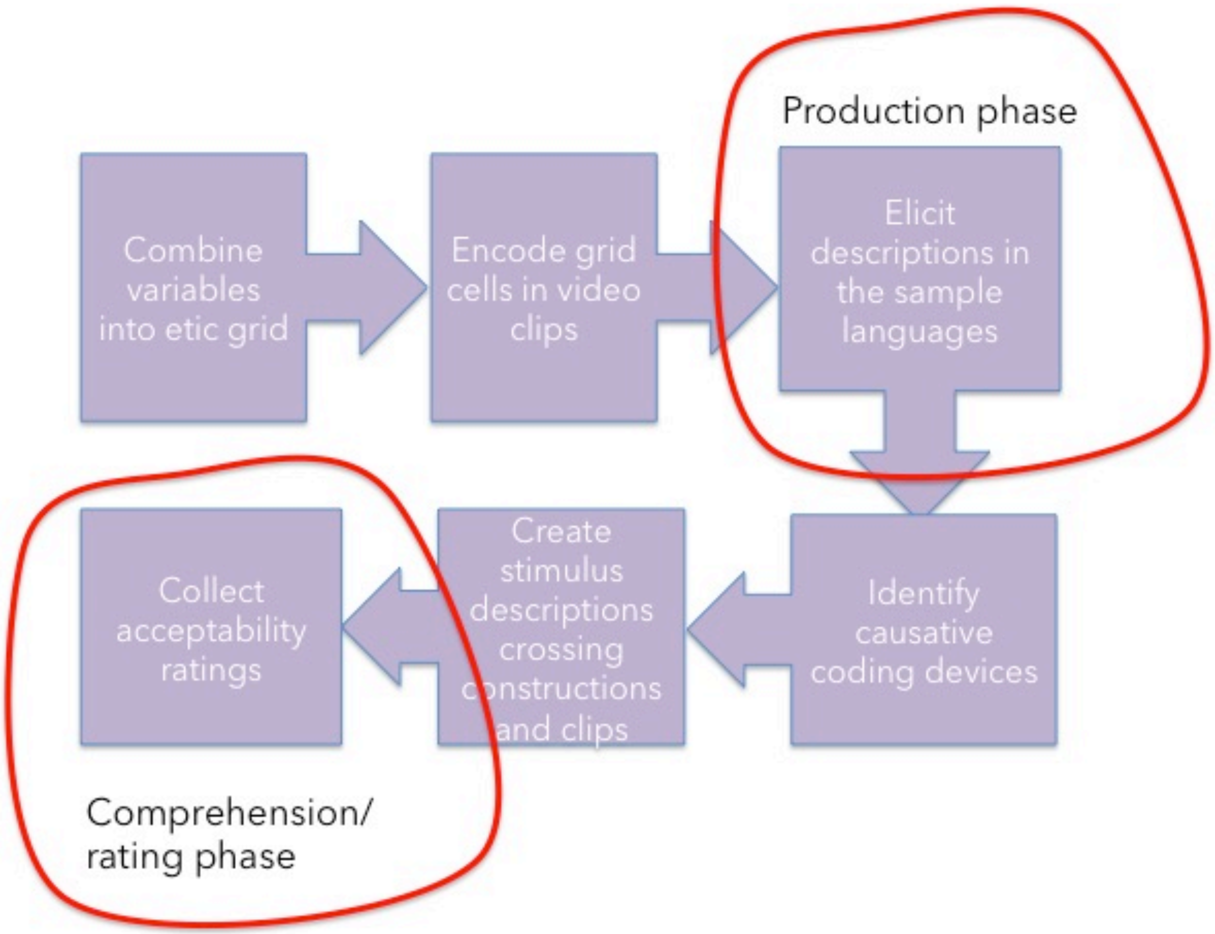


Figure 2.4. *A hybrid study design for semantic typology*

- advantages of this hybrid design type
  - vis-à-vis corpus studies
    - applicable to languages
      for which (large) corpora are unavailable
    - provides both positive and negative evidence
    - gives direct access to the scene being described
  - vis-à-vis traditional elicited production studies
    (the staple in contemporary semantic typology)
    - allows rapid data collection and analysis
      from a larger number of speakers
    - provides both positive and negative evidence

- we used the **Layered Structure of the Clause** (LSC) model of Role and Reference Grammar (Van Valin 2005)
  - to assign a complexity level to each construction type

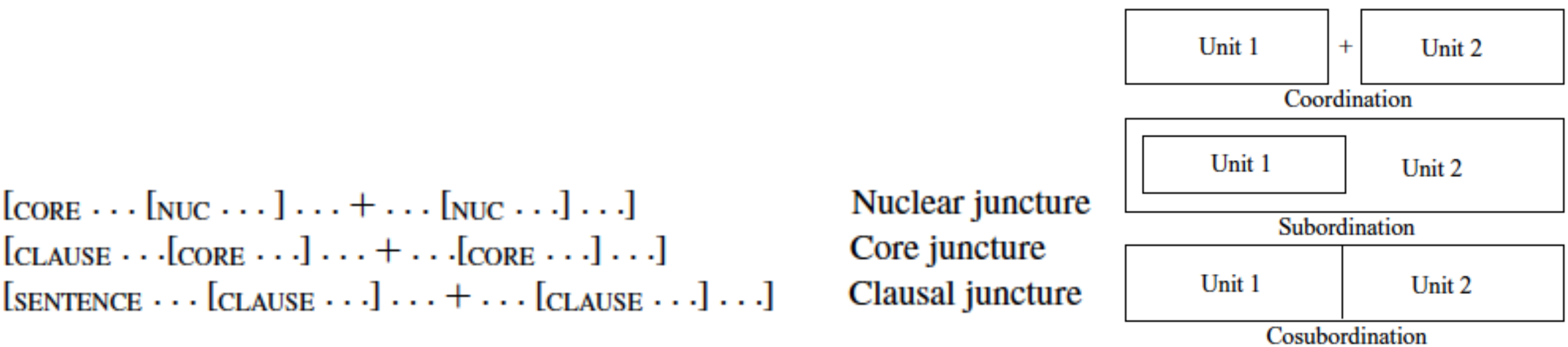


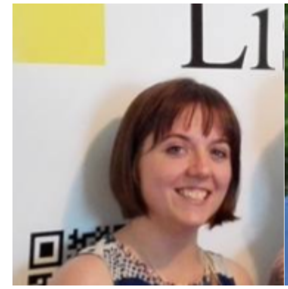
Figure 2.5. *Juncture (left) and nexus types in the Layered Structure of the Clause model (Van Valin 2005: 188)*

- why the LSC model?
  - because it gives us a single scale
    - on which to rank the relative complexity level of any causative coding device
      - namely, the morphosyntactic side of the Interclausal Relations Hierarchy
  - in contrast, in phrase structure grammars, one would have to assess separately
    - the complexity of the causing event representation
    - the complexity of the resulting event representation
    - the complexity of the construction that relates the two

SYNOPSIS

- Introducing CAL
- A new study design for semantic typology
- Variables and stimuli: the CAL Clips
- Preliminary findings
- Summary

VARIABLES AND STIMULI: THE CAL CLIPS



- design: E. Bellingham; J. Bohnemeyer
- 58 short video clips featuring everyday causal chains
    - most staged/enacted, a few found on the internet
- variables manipulated
    - **causer** (**CR**) type: volitional vs. accidental vs. force
    - **causee** (**CE**; = intermediate participant in the chain) type
        - volitional/controlled
        - vs. involuntary response to psychological impact
        - vs. involuntary response to mechanical impact
        - vs. no CE

- **affectee** (**AF**) type
  - volitional/controlled
  - vs. involuntary response to psychological impact
  - vs. involuntary response to mechanical impact
  - vs. physical object
- **resulting event type**
  physical state change vs. location change vs. process
- **force dynamics**
  - causation (43 core + 10 sup.) vs. letting (5 sup. scenes)

- stimuli: the CAL Clips (cont.)
  - examples
    - CR = force; CE = none; AF = mechanically impacted; resultant event = location change; FD = causation

Figure 3.1. *NM2_reporter*

- stimuli: the CAL Clips (cont.)
  - examples (cont.)
    - CR = accidental; CE = volitional/controlled; AF = object; resultant event = location change; FD = letting

Figure 3.2. *UCO1_ball*

- stimuli: the CAL Clips (cont.)
    - examples (cont.)
        - CR = volitional; CE = psychologically impacted; AF = object; resultant event = physical change; FD = letting

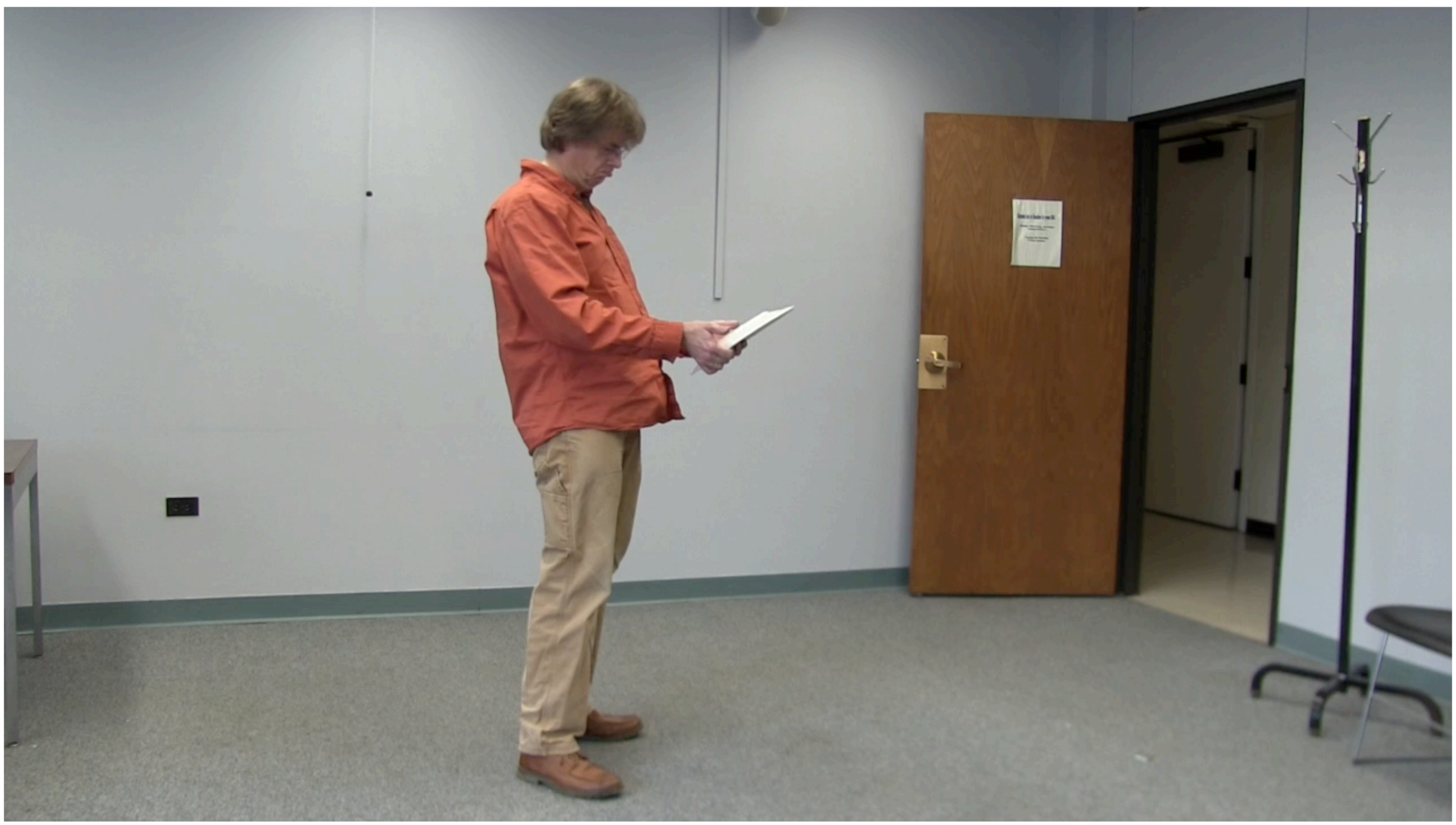


Figure 3.3. *HUO1_plate*

- stimuli: the CAL Clips (cont.)
  - examples (cont.)
    - CR = volitional; CE = volitional/controlled; AF = object; resultant event = process; FD = causation

Figure 3.4. *HCOproc1_swing*

SYNOPSIS

- Introducing CAL
- A new study design for semantic typology
- Variables and stimuli: the CAL Clips
- Preliminary findings
- Summary

PRELIMINARY FINDINGS

- the languages from which data has been collected for the Semantic Typology subproject so far

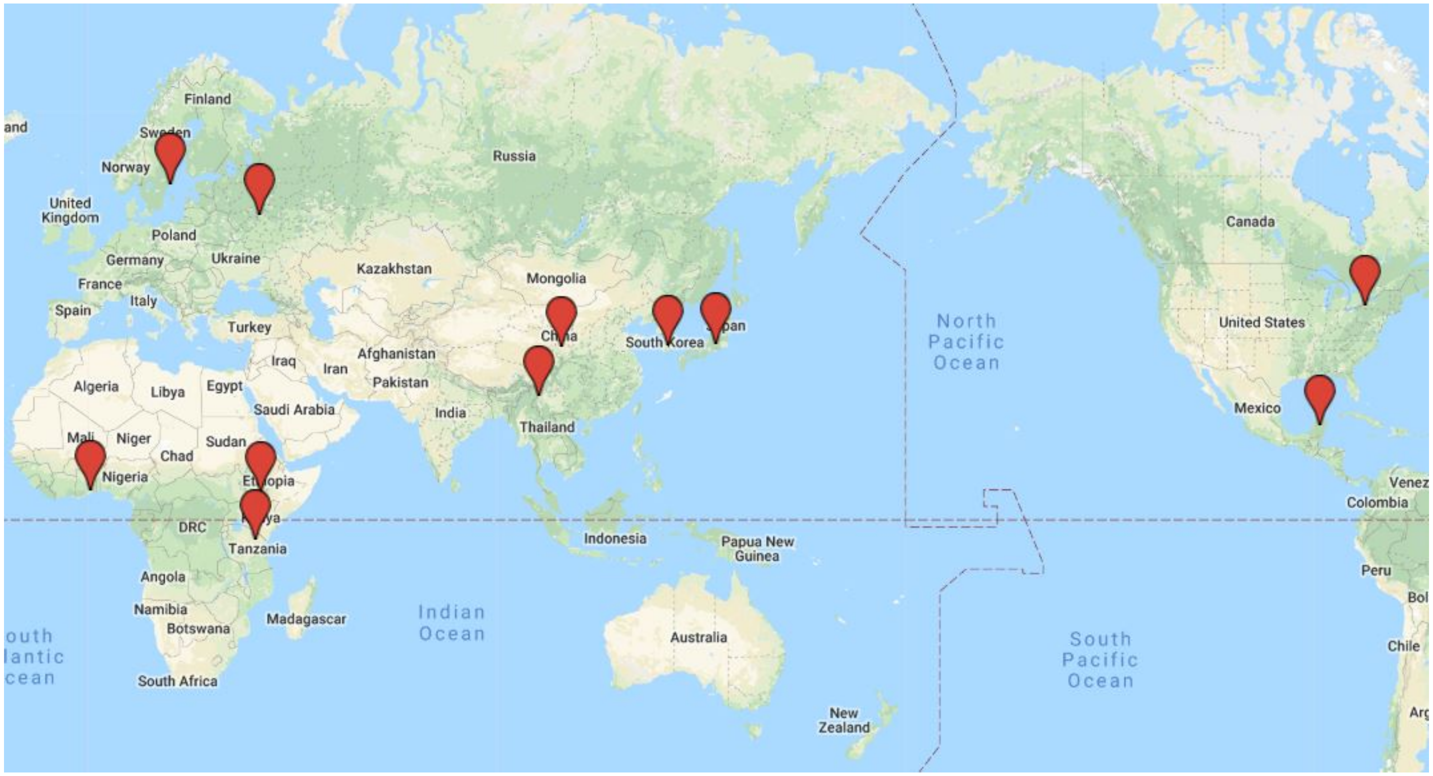


Figure 4.1. *The current sample of the CAL Semantic Typology subproject*

- populations included in the analysis so far and researchers

Language	Genus	Field site	Participants	Researcher	Affiliation
Datooga	Nilotic	Tanzania	12	A. Mitchell	U of Bristol
English	Germanic	U.S.A.	13	E. Bellingham, S. Evers	UB
Japanese	Japonic	Japan	14	K. Kawachi	National Defense Academy of Japan
Korean	Isolate	R.O.K.	12	S. Park	UB
Russian	Slavic	Russia	12	A. Stepanova	UB
Sidaama	Cushitic	Ethiopia	12	K. Kawachi	National Defense Academy of Japan
Swedish	Germanic	Sweden	12	P. Järnefelt, G. Montero-Melis, E. Bylund	Stockholm U
Yucatec	Mayan	Mexico	12	J. Bohnemeyer	UB
Zauzou	Lolo-Burmese	P.R.C.	12	Y. Li	UB

Table 4.1. *The current sample of the CAL Semantic Typology subproject*

- waiting in the wings:
  Ewe (J. Essegbey, UFL); Mandarin (J. Du, F. Li, Beihang U)

- causative coding devices included in the analysis

Table 4.2. *Causative coding devices in the sample languages that were included in the analysis*

Construction	Datooga	English	Swedish	Japanese	Korean	Russian	Sidaama	Yucatec	Zauzou
Transitive causative verbs	✓	✓	✓	✓	✓	✓	✓	✓	No
Morphological causatives	✓	No	No	✓	✓	No	✓	✓	No
Resultative constructions	No	✓	✓	No	✓	No	No	No	✓
Periphrastic causatives	✓	✓	✓	No	✓	✓	No	✓	✓
Single-core constructions augmented by an oblique causer PP/NP	✓	✓	No	✓	✓	No	✓	No	No
Event nominalizations used as causer arguments	No	No	No	No	✓	✓	✓	No	No
Causal converb constructions	No	No	No	✓	✓	No	✓	No	No
Causal connective constructions	✓	✓	✓	✓	No	✓	✓	✓	✓
'So X that Y'-type constructions	No	✓	✓	No	No	✓	No	No	No

- distribution of construction types over LSC juncture levels

Table 4.3. *Construction types by language and juncture (AC – Adjunct causer/reason ('because of x'), CC – Causal connective, CV – Converb, MC – Morphological causative, PC – Periphrastic causative, RV – Resultative construction (incl. resultative-type serial verb construction), SC - Scalar Connective construction ('So x that y'), TC – Transitive causative verb)*

Juncture level Language	Field site	Simplex or nuclear-layer	Core-layer	Clause-layer
Datooga (Nilotic)	Tanzania	MC, TC	AC, PC, SC	CC
English (Germanic)	United States	RV, TC	PC	AC, CC, SC
Japanese (Japonic)	Japan	MC, TC	AC	CC
Korean (isolate)	South Korea	MC, RV, TC	PC	CC, CV
Russian (Slavic)	Russia	TC	PC	AC, CC, SC
Sidaama (Cushitic)	Ethiopia	MC, TC	AC, PC	CC
Swedish (Germanic)	Sweden	RV, TC	PC	CC, SC
Yucatec (Mayan)	Mexico	MC, TC	PC	CC
Zauzou (Loloish)	China	RV	CC, CV, PC	CC

- analysis I: a descriptive look at the data
  - Figure 4.2 breaks down the data by clip, population, and number of participants who rated a given juncture
    - as the most compact acceptable for the particular clip

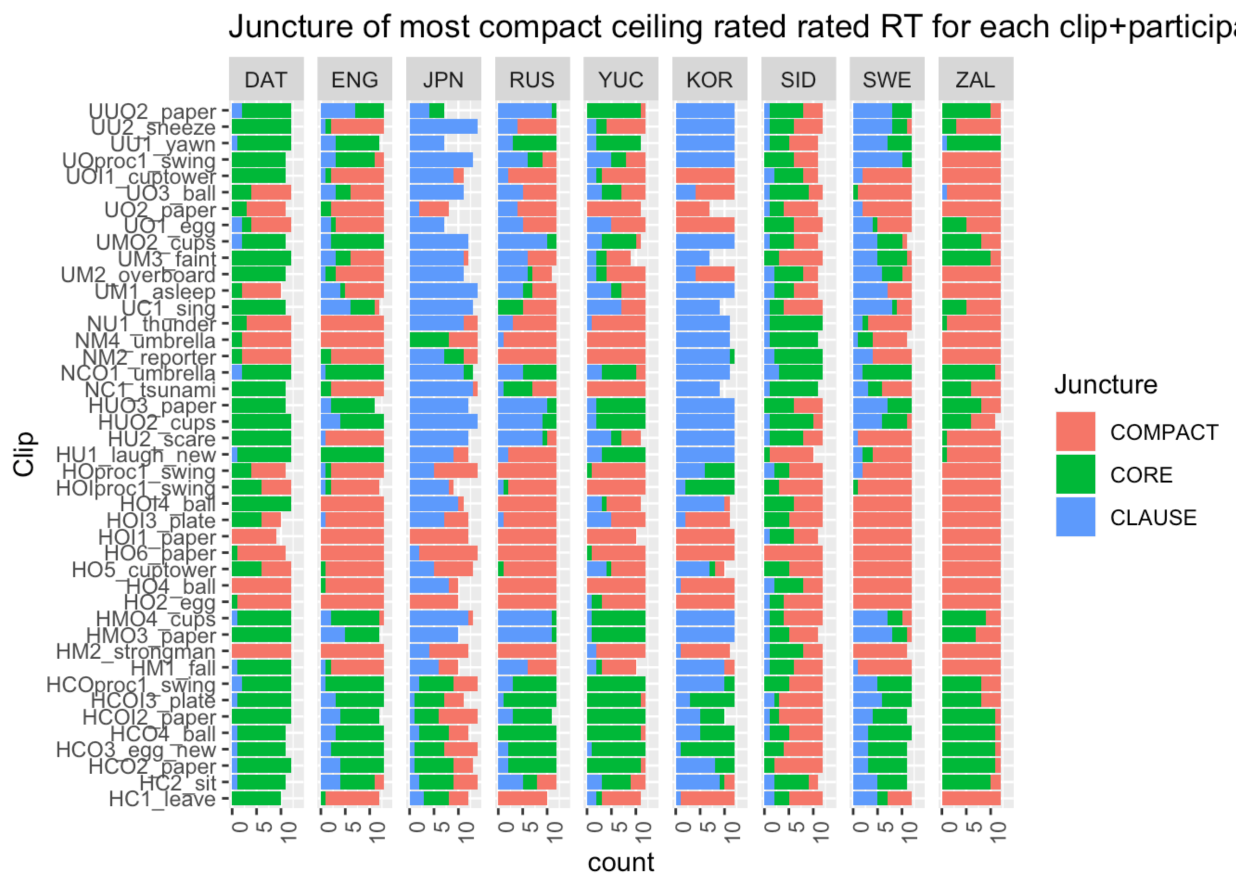


Figure 4.2. *Most compact ceiling-rated juncture level by clip, population, and number of participants*

- analysis I: a descriptive look at the data (cont.)
  - most Japanese and Korean speakers accepted only clausal junctures for more than half of the clips
  - in contrast, very few speakers of Datooga, Sidaama, Yucatec, and Zauzou *required* clausal junctures for *any* clip
  - the speakers of European languages fell in between these extremes

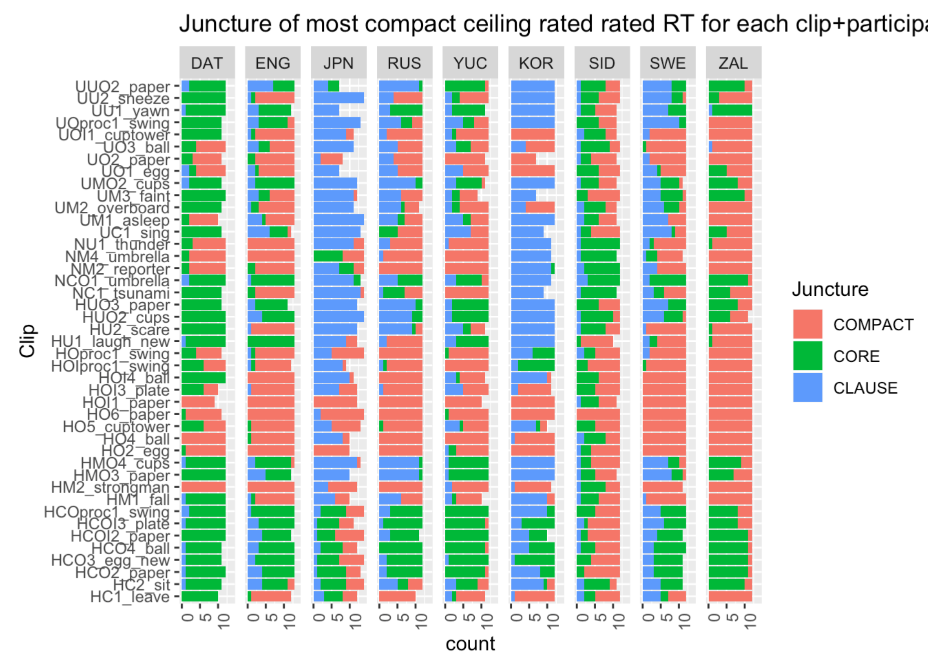
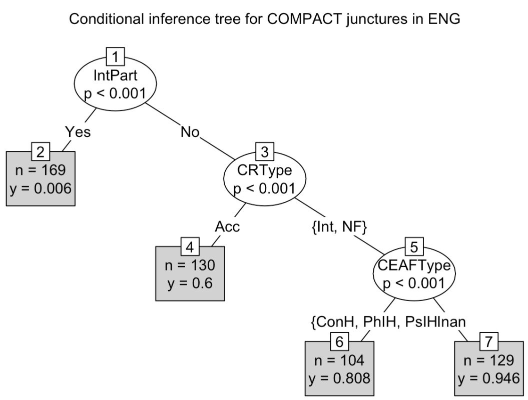
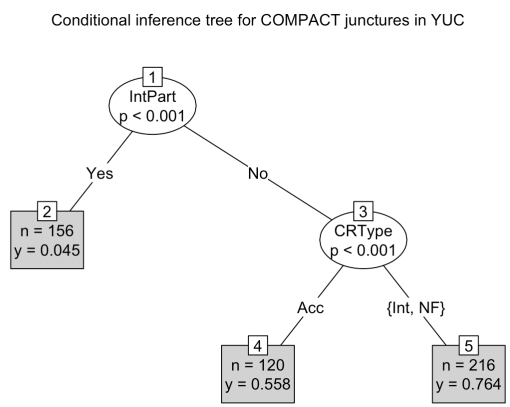
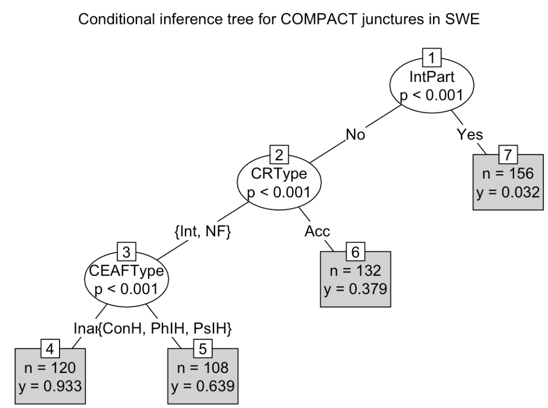


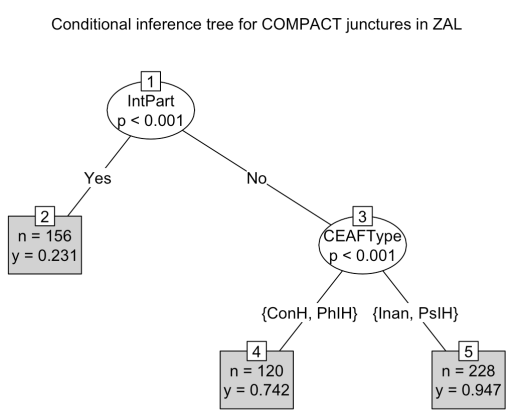
Figure 4.2. *Most compact ceiling-rated juncture level by clip, population, and number of participants*

- analysis II: predictive models - conditional inference trees (Hothorn, Hornik, & Zeileis 2006; Tagliamonte & Baayen 2012)
    - compact response types only: mediation is the most powerful predictor in most languages









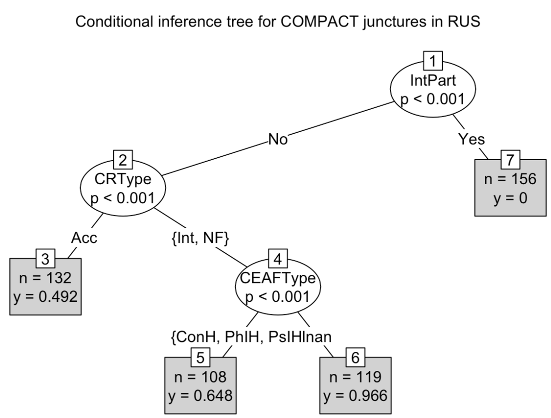


Figure 4.3. *Conditional inference trees predicting ceiling rating for compact responses in English, Yucatec, Swedish, Zauzou, and Russian (left to right and top to bottom). IntPart - Mediation; CRType - Causer Type; CEAFType - Causee/Affectee Type)*

- analysis II: predictive models - conditional inference trees (cont.)
  - exceptions occur in Japanese and Korean due to specific properties of morphological (Japanese) and syntactic (Korean) causatives in these languages
  - the Datooga and Sidaama data could not be modeled due to paucity of observations (Datooga) and rampant inter-speaker variation (Sidaama)

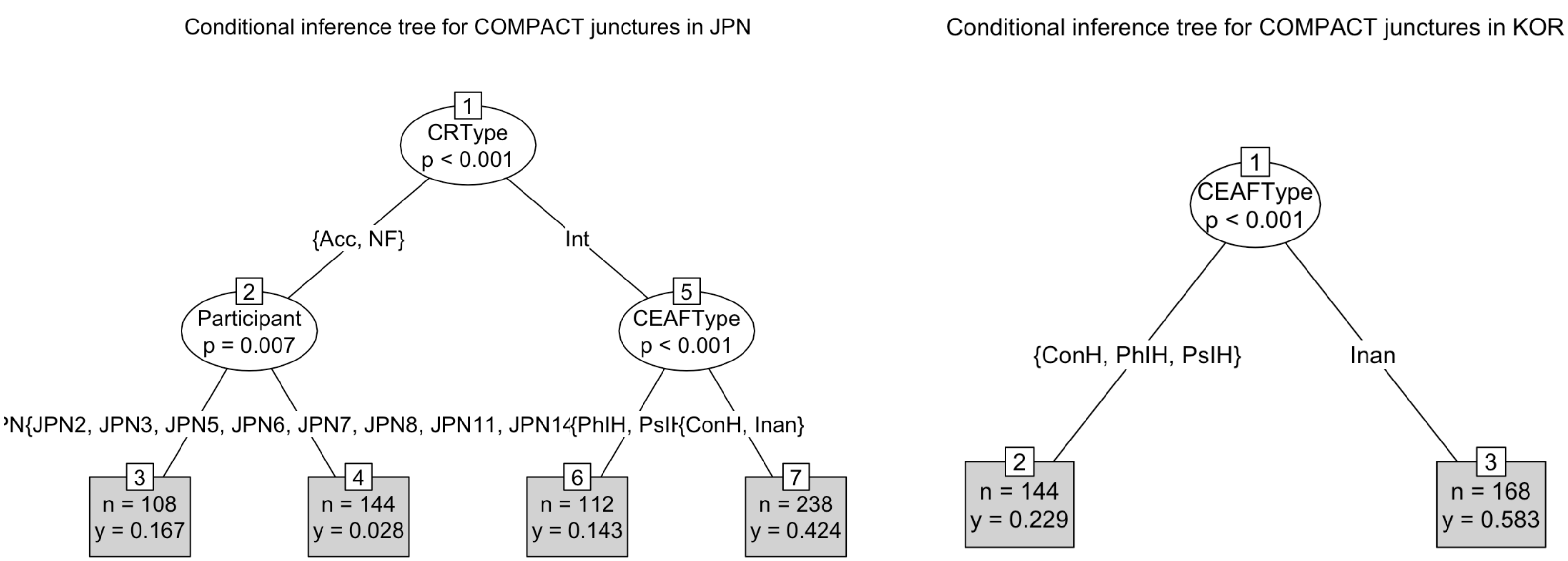


Figure 4.4. *Conditional inference trees predicting ceiling rating for compact responses in Japanese (left) and Korean (IntPart - Mediation; CRType - Causer Type; CEAFType - Causee/Affectee Type)*

- analysis III: predictive models - random forests (Breiman 2001; Tagliamonte & Baayen 2012)
  - rank order scores of variable importance for predicting the most compact ceiling-rated juncture for each clip

Table 4.4. *Variable importance scores from random forest models predicting the most compact ceiling-rated junctures for each clip and population. Each model based on 500 conditional inference trees.*

Predictor / Population	Causer type	Causee/affectee type	Mediation	Participant
Datooga	0.04841989	0.14564641	0.02896133	0.10604420
English	0.018603960	0.034118812	0.228742574	0.003722772
Japanese	0.14597872	0.16267021	0.02238298	0.06090426
Korean	0.03346667	0.14488889	0.05260000	-0.03765556
Russian	0.04628723	0.10455319	0.17526596	0.04520213
Sidaama	0.047586957	0.009000000	0.005663043	0.118826087
Swedish	0.07028877	0.15290909	0.10286631	0.09253476
Yucatec	0.04135135	0.03478919	0.25851892	0.06817297
Zauzou	0.01001058	0.04094180	0.16625397	0.03135450

- preliminary conclusions
  - the Iconicity Principle is borne out quantitatively across languages
  - however, the preferred structural complexity level of causatives is driven not only by Mediation
    - but also by Causer Type and Causee/Affectee Type
    - and in some languages, those competing variables dominate over Mediation

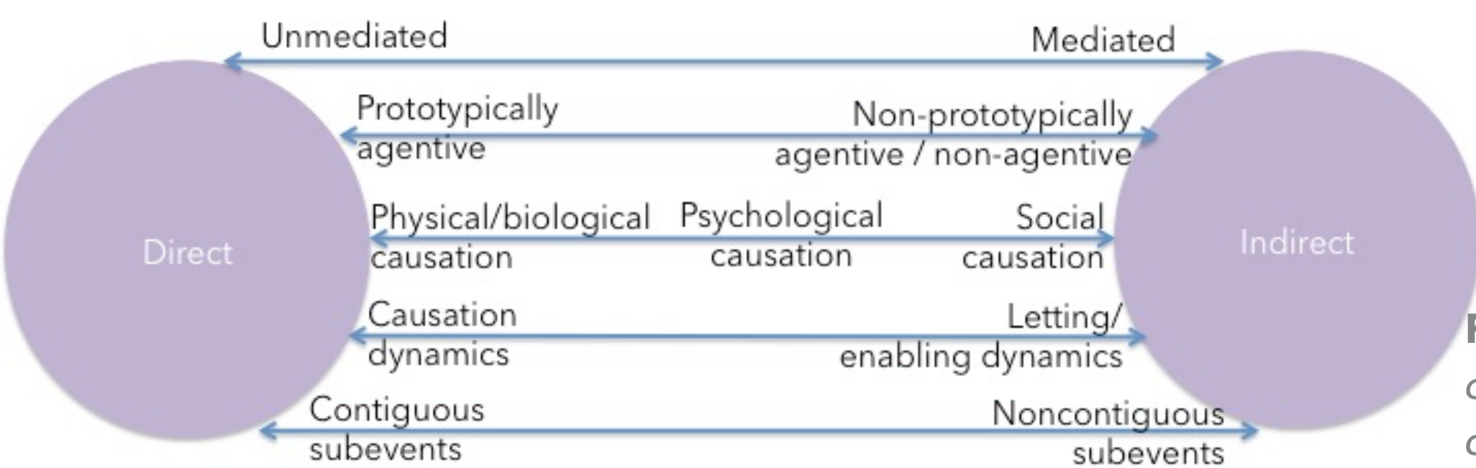


Figure 4.5. *A multidimensional continuum model of causation directness*

- preliminary conclusions (cont.)
  - in Japanese and Korean, agentivity and patientivity are stronger predictors than mediation
  - in these languages, clause-layer junctures are preferred for low-agentivity/low-patientivity scenes
      - i.e., scenes that do not conform to the Transitivity Hypothesis (Hopper & Thompson 1980)
    - core junctures - periphrastic causatives - are either not available (Japanese)
      - or are dispreferred for low-agentivity/low-patientivity scenes (Korean)

SYNOPSIS

- Introducing CAL
- A new study design for semantic typology
- Variables and stimuli: the CAL Clips
- Preliminary findings
- Summary

SUMMARY

- the Iconicity Principle is empirically confirmed
    - contrary to Escamilla (2012)
  - across languages, speakers prefer
    - morphosyntactically simpler representations for semantically simpler (more direct) causal chains
    - morphosyntactically more complex representations for semantically more complex (less direct) causal chains
- however, directness of causation is sensitive not only to mediation, but also to a host of other factors
  - including agentivity, patientivity, and force dynamics

- languages differ in the primary semantic variable that governs complexity of causatives
  - in most languages in our sample, this is mediation
    - i.e., the presence/absence of an intermediate participant in the causal chain
  - however, in Japanese, the dominant variable is agentivity
    - compact descriptions (incl. morphological causatives) are acceptable with mediated chains,
      - but not with accidental human causers or natural force causers
  - in Korean, patientivity is the dominant factor

- our study also showcases the usefulness of the LSC model
  - as a tool for measuring morphosyntactic complexity
    - including in, but not restricted to, typological research

ACKNOWLEDGMENTS

- epic thanks to the CAL researchers
  who contributed to the studies presented here

Clockwise from top left: Erika Bellingham, Pia Järnefelt, Yu Li, Guillermo Montero-Melis, Anastasia Stepanova, Sang-Hee Park, Alice Mitchell, Kazuhiro Kawachi

- massive thanks also to
  - colleagues who have provided advice:
    Dare Baldwin; Dedre Gentner; Beth Levin; Gail Mauner; Eric Pederson; Robert D. Van Valin, Jr., Phillip Wolff
    - all of whom shall be held blameless for any foolish and harebrained claims in this presentation
  - our sponsor
    - the material presented here is based upon work supported by the National Science Foundation under Grant No. BCS153846 and BCS-1644657, 'Causality Across Languages'; PI J. Bohnemeyer.

REFERENCES

Atlas, J. D., & S. C. Levinson (1981). It-clefts, informativeness, and logical form: Radical pragmatics (revised standard version). In P. Cole (ed.), *Radical pragmatics*. New York, NY: Academic Press. 1-62.

Bellingham, E., S. Evers, K. Kawachi, A. Mitchell, & J. Bohnemeyer (2017). An experimental approach to the semantic typology of causative constructions. Poster, *12th Association for Linguistics Typology Conference (ALT 2017)*.

Bohnemeyer, J., N.J. Enfield, J. Essegbey, & S. Kita. (2010). The Macro-Event Property: The segmentation of causal chains. In *Event representation in language: Encoding events at the language-cognition interface*, eds. Jürgen Bohnemeyer and Eric Pederson, 43–67. Cambridge: Cambridge University Press.

Bohnemeyer, J., Benedicto E., A. Capistrán Garza, K. T. Donelson, A. Eggleston, N. Hernández Green, M. Hernández Gómez, J. S. Lovegren, C. K. O'Meara, E. Palancar, G. Pérez Báez, G. Polian, R. Romero Méndez, & R. E. Tucker. (2012). Marcos de referencia en lenguas mesoamericanas: un análisis multivariante tipológico [Frames of reference in Mesoamerican languages: a typological multivariate analysis]. In N. England (Ed.), *Proceedings of the Conference on Indigenous Languages of Latin America-V*. Austin, TX: The Archive of the Indigenous Languages of Latin America.

Bohnemeyer, J., K. T. Donelson, R. E. Tucker, E. Benedicto, A. Eggleston, A. Capistrán Garza, N. Hernández Green, M. S. Hernández Gómez, S. Herrera Castro, C. K. O'Meara, E. Palancar, G. Pérez Báez, G. Polian, & R. Romero Méndez. (2014). The cultural transmission of spatial cognition: Evidence from a large-scale study. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Bohnemeyer, J., K. T. Donelson, R. E. Moore, E. Benedicto, A. Capistrán Garza, A. Eggleston, N. Hernández Green, M. S. Hernández Gómez, S. Herrera Castro, C. K. O'Meara, G. Pérez Báez, E. Palancar, G. Polian, & R. Romero Méndez. (2015).The contact diffusion of linguistic practices: Reference frames in Mesoamerica. *Language Dynamics and Change* 5(2):169-201.

Bohnemeyer, J., K. T. Donelson, Y.-T. Lin, R. Moore, H.-S. Hsiao, J. A. Jódar Sánchez, J. Lovegren, J. Olstad, G. Pérez Báez, & J. Seong. (In prep a). Language, culture, and the environment shape spatial cognition. Manuscript, University at Buffalo.

Bohnemeyer, J., E. Benedicto, K. T. Donelson, A. Eggleston, C. K. O'Meara, G. Pérez Báez, R. E. Moore, A. Capistrán Garza, N. Hernández Green, M. S. Hernández Gómez, S. Herrera Castro, E. Palancar, G. Polian, & R. Romero Méndez. (In prep b). The linguistic transmission of cognitive practices: Reference frames in and around Mesoamerica. Manuscript, University at Buffalo.

Breiman, L. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth International Group.

Breiman, L. (2001). Random forests. *Machine Learning* 45(1): 5-32.

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. Journal of Phonetics, 37(1):79-96.

Comrie, B. (1981). *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.

Dixon, R.M. (2000). A typology of causatives: form, syntax and meaning. In *Changing valency: Case studies in transitivity*, eds. Robert M. W. Dixon and Alexandra Y. Aikhenvald, 30--83. Cambridge: Cambridge University Press.

Escamilla Jr, R.M. (2012). *An updated typology of causative constructions: Form-function mappings in Hupa (Californian Athabaskan), Chungli Ao (Tibeto-Burman) and Beyond*. PhD Dissertation, University of California, Berkeley.

Fausey, C. M., B. L. Long, A. Inamori, & L. Boroditsky. (2010). Constructing agency: the role of language. *Frontiers in Psychology* 1: 162. doi:10.3389/fpsyg.2010.00162.

Haiman, J. (1983). Iconic and economic motivation. *Language* 59(4):781–819.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1): 1-33.

Hopper, P. J., & S. A. Thompson. (1980). Transitivity in grammar and discourse. *Language*, 251-299.

Ikegami, Y. (1991). 'DO-language' and 'BECOME-language': two contrasting types of linguistic representation. In Y. Ikegami (ed.), *The empire of signs: Semiotic essays on Japanese culture*. Amsterdam: Benjamins. 285-326.

Kemmer, S. & A. Verhagen. (1994). The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics* 5(2):115–156.

Levin, B. & M. Rappaport-Hovav. (1995). *Unaccusativity: At the syntax-semantics interface*. Cambridge, MA: MIT press.

Levinson, S. C.; S. Meira; & the Language and Cognition Group. 2003. 'Natural concepts' in the spatial topological domain—adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79.485–516.

Levshina, N. (2015). European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2): 487–520.

Levshina, N. (2016). Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages. *Folia Linguistica* 50(2): 507–542.

Levshina, N. (2017). Measuring iconicity: A quantitative study of lexical and analytic causatives in British English. *Functions of Language* 24(3): 319–347.

Majid, A., J.S. Boster, & M. Bowerman. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*(2), 235-250.

McCawley, J. (1976). Remarks on what can cause what. In *Syntax and Semantics VI: The grammar of causative constructions*, ed. Masayoshi Shibatani, 117–129. New York, NY: Academic Press.

McCawley, J. (1978). Conversational implicature and the lexicon. In *Syntax and semantics IX: Pragmatics*, ed. Peter Cole, 245-258. New York, NY: Academic Press.

Rappaport-Hovav, M. (2014). Lexical content and context: The causative alternation in English revisited. *Lingua*, *141*, 8-29.

REFERENCES (CONT.)

Shibatani, M. (ed.) (1976). *The grammar of causative constructions*. New York: Academic Press (Syntax and Semantics; 6).

Shibatani, M. & P. Pardeshi. (2002). The causative continuum. In *The grammar of causation and interpersonal manipulation*, ed. Masayoshi Shibatani, 85–126. Amsterdam: Benjamins.

Talmy, L. (1976). Semantic causative types. In Masayoshi Shibatani (ed.), *Syntax and semantics, vol. 6: The grammar of causative constructions*, 43-116. New York: Academic Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science* 12:49-100.

Van Valin Jr, R. D. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.

Verhagen, A. & S. Kemmer. (1997). Interaction and causation: Causative constructions in modern standard Dutch. *Journal of Pragmatics* 27:61–82.

Wolff, P. 2003. Direct causation in the linguistic coding and individuation of causal events. *Cognition* 88(1): 1–48.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology.* Boston, M.A.: Houghton Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* New York, NY: Hafner.

ありがとう!

Thanks!