Modeling causative complexity across languages with the Interclausal Relations Hierarchy

Erika Bellingham¹, Pia Järnefelt², Kazuhiro Kawachi³, Yu Li¹, Alice Mitchell⁴, Guillermo Montero-Melis⁵, Sang-Hee Park¹, Anastasia Stepanova¹, Manne Bylund², and Jürgen Bohnemeyer¹
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
INTRODUCING CAL (CONT.)

- 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
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) a. #Le=x-ch’úupal=o’  t-u=nik-ah  le=bàaso-s-o’b=o’
YUC  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:
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
stimuli: the CAL Clips (cont.)

examples (cont.)

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

examples (cont.)

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

examples (cont.)

CR = volitional; CE = volitional/controlled; AF = object; resultant event = process; FD = causation
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 waiting in the wings:

- Ewe (J. Essegbey, UFL)
- Mandarin (J. Du, F. Li, Beihang U)

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

<table>
<thead>
<tr>
<th>Language</th>
<th>Genus</th>
<th>Field site</th>
<th>Participants</th>
<th>Researcher</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datooga</td>
<td>Nilotic</td>
<td>Tanzania</td>
<td>12</td>
<td>A. Mitchell</td>
<td>U of Bristol</td>
</tr>
<tr>
<td>English</td>
<td>Germanic</td>
<td>U.S.A.</td>
<td>13</td>
<td>E. Bellingham, S. Evers</td>
<td>UB</td>
</tr>
<tr>
<td>Japanese</td>
<td>Japonic</td>
<td>Japan</td>
<td>14</td>
<td>K. Kawachi</td>
<td>National Defense Academy of Japan</td>
</tr>
<tr>
<td>Korean</td>
<td>Isolate</td>
<td>R.O.K.</td>
<td>12</td>
<td>S. Park</td>
<td>UB</td>
</tr>
<tr>
<td>Russian</td>
<td>Slavic</td>
<td>Russia</td>
<td>12</td>
<td>A. Stepanova</td>
<td>UB</td>
</tr>
<tr>
<td>Sidaama</td>
<td>Cushitic</td>
<td>Ethiopia</td>
<td>12</td>
<td>K. Kawachi</td>
<td>National Defense Academy of Japan</td>
</tr>
<tr>
<td>Yucatec</td>
<td>Mayan</td>
<td>Mexico</td>
<td>12</td>
<td>J. Bohnemeyer</td>
<td>UB</td>
</tr>
<tr>
<td>Zauzou</td>
<td>Lolo-Burmese</td>
<td>P.R.C.</td>
<td>12</td>
<td>Y. Li</td>
<td>UB</td>
</tr>
</tbody>
</table>
causative coding devices included in the analysis

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

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

<table>
<thead>
<tr>
<th>Language</th>
<th>Juncture level</th>
<th>Field site</th>
<th>Simplex or nuclear-layer</th>
<th>Core-layer</th>
<th>Clause-layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datooga (Nilotic)</td>
<td>Tanzania</td>
<td>MC, TC</td>
<td>AC, PC, SC</td>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>English (Germanic)</td>
<td>United States</td>
<td>RV, TC</td>
<td>PC</td>
<td></td>
<td>AC, CC, SC</td>
</tr>
<tr>
<td>Japanese (Japonic)</td>
<td>Japan</td>
<td>MC, TC</td>
<td>AC</td>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>Korean (isolate)</td>
<td>South Korea</td>
<td>MC, RV, TC</td>
<td>PC</td>
<td></td>
<td>CC, CV</td>
</tr>
<tr>
<td>Russian (Slavic)</td>
<td>Russia</td>
<td>TC</td>
<td>PC</td>
<td></td>
<td>AC, CC, SC</td>
</tr>
<tr>
<td>Sidaama (Cushitic)</td>
<td>Ethiopia</td>
<td>MC, TC</td>
<td>AC, PC</td>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>Swedish (Germanic)</td>
<td>Sweden</td>
<td>RV, TC</td>
<td>PC</td>
<td></td>
<td>CC, SC</td>
</tr>
<tr>
<td>Yucatec (Mayan)</td>
<td>Mexico</td>
<td>MC, TC</td>
<td>PC</td>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>Zauzou (Loloish)</td>
<td>China</td>
<td>RV</td>
<td>CC, CV, PC</td>
<td></td>
<td>CC</td>
</tr>
</tbody>
</table>
- **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; CRTyp - Causer Type; CEAFTyp - 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; CRTType - 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.

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


ありがとう!

Thanks!