analysis I: a descriptive look at the data

Figure 4.2. Juncture of the most compact description each participant accepted for each clip
analysis I: a descriptive look at the data (cont.)

- Japanese & Korean: most speakers accepted only clausal junctures for >50% of clips

- Datooga, Sidaama, Yucatec & Zauzou: few speakers required clausal junctures for any clip

- European languages: speakers fell in between

**Figure 4.2. Juncture of the most compact description each participant accepted for each clip**
STUDY II: SEMANTIC TYPOLOGY (CONT.)

- **analysis II: conditional inference trees**
  (Hothorn, Hornik, & Zeileis 2006; Tagliamonte & Baayen 2012)

- Predict the juncture of the most compact description

- Computed top down: which predictor makes the biggest difference?

- e.g **English:**

\[
y = (\text{COMPACT, CORE, CLAUSE})
\]

**Figure 4.3.** Conditional inference tree predicting juncture of the most compact description each English speaker accepted for a given clip.

*IntPart* - Mediation; *CRT* - Causer Type; *CEAFT* - Causee/Affectee Type*
STUDY II: SEMANTIC TYPOLOGY (CONT.)

- **analysis II: conditional inference trees**
  (Hothorn, Hornik, & Zeileis 2006; Tagliamonte & Baayen 2012)

- Compare to **Japanese**: 

---

**Figure 4.4.** Conditional inference tree predicting juncture of the most compact description each Japanese speaker accepted for a given clip.

*IntPart* - Mediation; *CRType* - Causer Type; *CEAFType* - Causee/Afectee Type)
STUDY II: SEMANTIC TYPOLOGY (CONT.)

- analysis III: predictive models - random forests
  (Breiman 2001; Tagliamonte & Baayen 2012)

**Table 4.4.** Conditional variable importance scores from random forest models predicting the juncture of the most compact description each participant accepted for each clip.
preliminary conclusions

- causative structural complexity driven not only by Mediation
- but also by Causer Type and Causee/Affectee Type
- in some languages, those competing variables are more dominant than Mediation

Figure 4.5. A multidimensional continuum model of causation directness
preliminary conclusions (cont.)

Japanese & Korean:

- agentivity, patientivity > mediation
- clause-layer junctures preferred for low-agentivity/low-patientivity scenes
  - i.e., scenes that do not conform to the Transitivity Hypothesis (Hopper & Thompson 1980)
  - exception: communication scenes in Japanese

core junctures - periphrastic causatives:

- Japanese: not available
- Korean: dispreferred for low-agentivity/low-patientivity scenes
SYNOPSIS

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

- Iconicity Principle: 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
- Directness of causation is sensitive not only to mediation, but also to other factors
  - including agentivity and patientivity
languages differ in the primary semantic variable that governs complexity of causatives

- English, Yucatec, Zauzou: **mediation**
  - i.e., presence of an intermediate participant in the causal chain
- Korean & Datooga: **patientivity** is most important
- Russian: **mediation** & **patientivity**
- Swedish: mediation, patientivity, agentivity, participant?
- Japanese: **agentivity** & **patientivity** are more important
  - compact descriptions (incl. morphological causatives) accepted for mediated chains with communication
  - but not with accidental human causers or natural force causers
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


REFERENCES (CONT.)


REFERENCES (CONT.)


ありがとう!
Thanks!
Which variables matter most?

Analysis III: Predictive Models - Random Forests
(Breiman 2001; Tagliamonte & Baayen 2012)

Table 4.4. Conditional variable importance scores from random forest models predicting the juncture of the most compact description each participant accepted for each clip.

<table>
<thead>
<tr>
<th>Population</th>
<th>Causer type</th>
<th>Causee/Afectee type</th>
<th>Mediation</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datooga</td>
<td>0.0156</td>
<td>0.0247</td>
<td>0.0993</td>
<td>0.0054</td>
</tr>
<tr>
<td>English</td>
<td>0.0401</td>
<td>0.0944</td>
<td>0.0147</td>
<td>0.1085</td>
</tr>
<tr>
<td>Japanese</td>
<td>0.1147</td>
<td>0.1405</td>
<td>0.0144</td>
<td>0.0616</td>
</tr>
<tr>
<td>Korean</td>
<td>0.0303</td>
<td>0.1096</td>
<td>0.0336</td>
<td>-0.0338</td>
</tr>
<tr>
<td>Russian</td>
<td>0.0438</td>
<td>0.0788</td>
<td>0.0897</td>
<td>0.0427</td>
</tr>
<tr>
<td>Sidaama</td>
<td>0.0461</td>
<td>0.0018</td>
<td>0.0003</td>
<td>0.1151</td>
</tr>
<tr>
<td>Swedish</td>
<td>0.0739</td>
<td>0.0723</td>
<td>0.0652</td>
<td>0.0944</td>
</tr>
<tr>
<td>Yucatec</td>
<td>0.0420</td>
<td>0.0327</td>
<td>0.1058</td>
<td>0.0653</td>
</tr>
<tr>
<td>Zauzou</td>
<td>0.0138</td>
<td>0.0383</td>
<td>0.0695</td>
<td>0.0309</td>
</tr>
</tbody>
</table>
Compact constructions: did the speaker accept a compact (simplex/nuclear layer juncture) construction for a clip?

Table 4.4. Variable importance scores from random forest models predicting whether a compact description was accepted per participant + clip.

<table>
<thead>
<tr>
<th>Population</th>
<th>Causer type</th>
<th>Causee/ Affectee type</th>
<th>Mediation</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>0.0248</td>
<td>0.0420</td>
<td><strong>0.2196</strong></td>
<td>-0.0088</td>
</tr>
<tr>
<td>Datooga</td>
<td>n/a (not enough compact descriptions tested)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td><strong>0.0444</strong></td>
<td>0.0345</td>
<td>0.0085</td>
<td>0.0319</td>
</tr>
<tr>
<td>Korean</td>
<td>0.0459</td>
<td><strong>0.0810</strong></td>
<td>0.0182</td>
<td>-0.0196</td>
</tr>
<tr>
<td>Russian</td>
<td>0.0360</td>
<td>0.0557</td>
<td><strong>0.1660</strong></td>
<td>0.0166</td>
</tr>
<tr>
<td>Sidaama</td>
<td>0.0014</td>
<td>0.0068</td>
<td>0.0060</td>
<td><strong>0.1302</strong></td>
</tr>
<tr>
<td>Swedish</td>
<td>0.0728</td>
<td><strong>0.1262</strong></td>
<td>0.0975</td>
<td>0.0169</td>
</tr>
<tr>
<td>Yucatec</td>
<td>0.0392</td>
<td>0.0297</td>
<td><strong>0.1661</strong></td>
<td>0.0719</td>
</tr>
<tr>
<td>Zauzou</td>
<td>0.0025</td>
<td>0.0360</td>
<td><strong>0.1359</strong></td>
<td>0.0225</td>
</tr>
</tbody>
</table>