Testing iconicity:
A quantitative study of causative constructions based on a parallel corpus of film subtitles

Natalia Levshina
F.R.S.-FNRS,
Université catholique de Louvain
(Belgium)
Outline

1. Iconicity in causative constructions
2. Data and variables
3. Quantitative analyses
4. Discussion
Causative constructions

- **Lexical** = one predicate
  - e.g. *break, kill, send*

- **Morphological** = a non-causal predicate + productive causative morpheme
  - e.g. Finnish *odotuttaa* “cause to wait” (from *odottaa* “wait”)

- **Analytic** = two predicates
  - e.g. *make X cry, let X go, make X happy*
## Semantic regularities

<table>
<thead>
<tr>
<th>Study</th>
<th>More compact causative</th>
<th>Less compact causative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comrie (1981; 1989)</td>
<td>Direct causation Low control of Causee</td>
<td>Indirect causation High control of Causee</td>
</tr>
<tr>
<td>Haiman (1983; 1985)</td>
<td>Smaller conceptual distance between Cause and Result</td>
<td>Greater conceptual distance between Cause and Result</td>
</tr>
<tr>
<td>Givón (1990)</td>
<td>Inanimate Manipulee</td>
<td>Human-Agentive Manipulee</td>
</tr>
</tbody>
</table>
Iconicity

• All these studies express in different words the same idea: that the degree of formal integration correlates with the degree of semantic integration of the cause and effect.

• An instance of iconic relationship between form and function.

Development of the Chinese character “water”
An extended approach

• Dixon (2000): a tentative list of 9 semantic and syntactic parameters based on a typological survey.
• Not all are directly interpretable in terms of iconicity.
## Dixon’s parameters

### More compact

<table>
<thead>
<tr>
<th>State (or change of state)</th>
<th>Intransitive</th>
<th>No control</th>
<th>Willing (‘let’)</th>
<th>Partially affected</th>
<th>Direct</th>
<th>Intentional</th>
<th>Natural</th>
</tr>
</thead>
</table>

### Less compact

<table>
<thead>
<tr>
<th>Action</th>
<th>(Di)transitive</th>
<th>Control</th>
<th>Unwilling (‘make’)</th>
<th>Fully affected</th>
<th>Indirect</th>
<th>Accidental</th>
</tr>
</thead>
</table>

### Relating to

- **VERB**
- **Causee**
- **Causer**

- **With effort, violence**
Dixon’s parameters

**More compact**

State (or change of state)
- Intransitive
  - No control
  - Willing (‘let’)
  - Partially affected

  - Direct
  - Intentional
  - Natural

**Related to base VERB**

- Control
- Unwilling (‘make’)
- Fully affected

**Action (Di)transitive**

- Indirect
- Accidental
- With effort, violence

**Less compact**

- Related to Causee
- Related to Causer
Dixon’s parameters

State (or change of state)               Action
Intransitive         (Di)transitive
No control      Control
Willing (‘let’)                   Unwilling (‘make’)
Partially affected         Fully affected

Related to
VERB
Related to
Causee
Related to
Causer

More compact

Less compact

Direct
Intentional
Natural

Control
Unwilling (‘make’)
Fully affected

Indirect
Accidental
With effort, violence
Dixon’s parameters

State (or change of state)               Action
Intransitive         (Di)transitive
No control      Control
Willing (‘let’)                Unwilling (‘make’)
Partially affected         Fully affected
Direct              Indirect
Intentional                 Accidental
Natural             With effort, violence

Related to VERB
Related to Causee
Related to Causer

More compact               Less compact
compact

More compact               Less compact
compact
The main question

• Can the formal variation (i.e. degree of compactness) of the causatives be explained by one factor (iconicity-related) or many factors (Dixon)?
• Never investigated quantitatively before!
Outline

1. Iconicity in causative constructions
2. Data and variables
3. Quantitative analyses
4. Discussion
ParTy corpus

• a Parallel corpus for Typology
• subtitles of films and TED talks
• mostly Indo-European languages, but also other major languages (Chinese, Turkish, Finnish, Indonesian, Japanese, Thai, etc.)
• all languages aligned with English
• downloadable files at www.natalialevshina.com/corpus.html
• work in progress...
Why subtitles?

Cluster Dendrogram

Based on the frequencies of 3-grams (Levshina, Accepted)
### Data used in the case study

**Films**

- **Avatar**
- **Black Swan**
- **Inception**
- **Frozen**

**TED talks**

- Ken Robinson: *Do schools kill creativity?*
- Elizabeth Gilbert: *Your elusive creative genius*
- Amy Cuddy: *Your body language shapes who you are*
- Leslie Morgan Steiner: *Why domestic violence victims don’t leave*
- Dan Gilbert: *The psychology of your future self*
- Simon Sinek: *Why good leaders make you feel safe*
## Languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Genus</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>Chinese</td>
<td>Sino-Tibetan</td>
</tr>
<tr>
<td>Finnish</td>
<td>Finnic</td>
<td>Uralic</td>
</tr>
<tr>
<td>French</td>
<td>Romance</td>
<td>Indo-European</td>
</tr>
<tr>
<td>Hebrew</td>
<td>Semitic</td>
<td>Afro-Asiatic</td>
</tr>
<tr>
<td>Indonesian</td>
<td>Malayo-Sumbawan</td>
<td>Austronesian</td>
</tr>
<tr>
<td>Japanese</td>
<td>Japanese</td>
<td>Japanese</td>
</tr>
<tr>
<td>Russian</td>
<td>Slavic</td>
<td>Indo-European</td>
</tr>
<tr>
<td>Thai</td>
<td>Kam-Tai</td>
<td>Tai-Kadai</td>
</tr>
<tr>
<td>Turkish</td>
<td>Turkic</td>
<td>Altaic</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>Viet-Muong</td>
<td>Austro-Asiatic</td>
</tr>
</tbody>
</table>
Data set

- 344 causative situations found in English
- Translations in the 10 languages are found and coded into 3 types of constructions (Analytic, Morphological or Lexical)
Example from *Avatar*

Original

- ENG: *Don't shoot, you'll piss him off.*

Translations

- FRA: *Ne tirez pas. Vous allez l'énérer.* (Lexical)
- TUR: *Ateş etme. Ateş etme. Onu kızdıracaksın.* (Morphological, from *kızmek* ‘become angry’).
- VIE: *Đừng bắn. Câu sẽ làm nó nổi điên đó.* (Analytic)
Data set

• 344 causative situations found in English
• Translations in the 10 languages are found and coded into 3 types of constructions (Analytic, Morphological or Lexical)
• The English sentences are coded for 13 semantic variables (taking into account the context)
Outline

1. Iconicity in causative constructions
2. Data and variables
3. Quantitative analyses
4. Discussion
## Variables (1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
<th>Example(s)</th>
<th>Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CausedEvent</td>
<td>Non-action</td>
<td>John killed Bill.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>Action</td>
<td>I walk my dog.</td>
<td>Longer form</td>
</tr>
<tr>
<td>NoPart (number of participants)</td>
<td>2</td>
<td>John killed Bill.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>I gave him a book.</td>
<td>Longer form</td>
</tr>
<tr>
<td>CeControl (Causee having control)</td>
<td>No</td>
<td>John killed Bill.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Bring your friends!</td>
<td>Longer form</td>
</tr>
<tr>
<td>MakeLet</td>
<td>Let</td>
<td>She let him go.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>Make</td>
<td>John killed Bill.</td>
<td>Longer form</td>
</tr>
<tr>
<td>CeVol (volitional Causee)</td>
<td>No</td>
<td>John caused Bill to die.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>The police let him go.</td>
<td>Longer form</td>
</tr>
</tbody>
</table>
## Variables (2)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
<th>Example(s)</th>
<th>Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrDirect (Causer acting directly)</td>
<td>Yes</td>
<td>He cut his hair.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>He had his hair cut.</td>
<td>Longer form</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrIntent (Causer acting intentionally)</td>
<td>Yes</td>
<td>She wrote a paper.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>It makes me happy.</td>
<td>Longer form</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrForce (Causer acting forcefully)</td>
<td>No</td>
<td>He got him to do it.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>He forced him to do it.</td>
<td>Longer form</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrInvolve (Causer involved in caused event)</td>
<td>No</td>
<td>John killed Bill.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Bring your friends! (and come, too)</td>
<td></td>
</tr>
</tbody>
</table>
## Variables (3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
<th>Example(s)</th>
<th>Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coref (coreferentiality)</td>
<td>Yes</td>
<td>He killed himself.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>He killed Bill.</td>
<td></td>
</tr>
<tr>
<td>Polarity</td>
<td>Pos</td>
<td>She let him do it.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Neg</td>
<td>She didn’t let him do it.</td>
<td></td>
</tr>
<tr>
<td>CrSem (semantics of Causer)</td>
<td>Anim</td>
<td>She made him stay.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Inanim</td>
<td>The rain made him stay.</td>
<td></td>
</tr>
<tr>
<td>CeSem (semantics of Causee)</td>
<td>Anim</td>
<td>John let Mary go.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Inanim</td>
<td>John let it go.</td>
<td></td>
</tr>
</tbody>
</table>
Interrater agreement for semantic variables

Ludivine Crible, UCL

Samantha Laporte, UCL
Light’s kappas

- Min = 0.7 *CrForce* (the Causer acting forcefully)
- Max = 0.93 *CrIntent* (the Causer acting intentionally)
Outline

1. Iconicity in causative constructions
2. Data and variables
3. Quantitative analyses
4. Discussion
A challenge

• The most appropriate method: multiple regression analysis with Cx (Lexical, Morphological and Analytic) as response and the semantic and syntactic variables as predictors.

• But: highly associated semantic variables → danger of multicollinearity

• Solution:
  – Adjusted Multiple Correspondence Analysis of the 13 variables as a dimensionality-reduction technique
  – R packages ca (Nenadić & Greenacre 2007) and FactoMineR (Husson et al. 2015)
MCA: Explained variance (inertia)

Principal inertias (eigenvalues):

<table>
<thead>
<tr>
<th>dim</th>
<th>value</th>
<th>%</th>
<th>cum%</th>
<th>scree plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.034794</td>
<td>65.1</td>
<td>65.1</td>
<td>**********************************</td>
</tr>
<tr>
<td>2</td>
<td>0.004613</td>
<td>8.6</td>
<td>73.8</td>
<td>***</td>
</tr>
<tr>
<td>3</td>
<td>0.002605</td>
<td>4.9</td>
<td>78.6</td>
<td>**</td>
</tr>
<tr>
<td>4</td>
<td>0.000180</td>
<td>0.3</td>
<td>79.0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9e-06000</td>
<td>0.0</td>
<td>79.0</td>
<td></td>
</tr>
</tbody>
</table>
MCA: Dimensions 1 & 2
MCA: Dimensions 1 & 3
## Contributions to dimensions

<table>
<thead>
<tr>
<th>Feature</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>Feature</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrIntent=No</td>
<td>0.00</td>
<td>0.30</td>
<td>0.06</td>
<td>CrIntent=Yes</td>
<td>0.00</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>CrIntent=Yes</td>
<td>0.00</td>
<td>0.09</td>
<td>0.02</td>
<td>CrForce=No</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CrForce=Yes</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>CrInvolve=No</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>CrForce=No</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>CrForce=Yes</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CrInvolve=No</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>CrInvolve=Yes</td>
<td>0.02</td>
<td>0.06</td>
<td>0.27</td>
</tr>
<tr>
<td>CrInvolve=Yes</td>
<td>0.02</td>
<td>0.06</td>
<td>0.27</td>
<td>CrDirect=No</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>CrDirect=No</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>CrDirect=Yes</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>CrDirect=Yes</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>CeControl=No</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CeControl=No</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
<td>CeControl=Yes</td>
<td>0.16</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>CeControl=Yes</td>
<td>0.16</td>
<td>0.00</td>
<td>0.02</td>
<td>CrSem=Anim</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>MakeLet=Let</td>
<td>0.08</td>
<td>0.00</td>
<td>0.05</td>
<td>CrSem=Inanim</td>
<td>0.00</td>
<td>0.37</td>
<td>0.02</td>
</tr>
<tr>
<td>MakeLet=Make</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>CeVol=No</td>
<td>0.05</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CdEvent=Action</td>
<td>0.12</td>
<td>0.00</td>
<td>0.03</td>
<td>CeVol=Yes</td>
<td>0.14</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CdEvent=NAction</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>NoPart=2</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>NoPart=3</td>
<td>0.11</td>
<td>0.01</td>
<td>0.02</td>
<td>NoPart=3</td>
<td>0.11</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Coref=No</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>Coref=Yes</td>
<td>0.02</td>
<td>0.04</td>
<td>0.34</td>
</tr>
<tr>
<td>Coref=Yes</td>
<td>0.02</td>
<td>0.04</td>
<td>0.34</td>
<td>Polarity=Neg</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Polarity=Pos</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Polarity=Pos</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CrSem=Anim</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
<td>CrSem=Inanim</td>
<td>0.00</td>
<td>0.37</td>
<td>0.02</td>
</tr>
<tr>
<td>CeSem=Anim</td>
<td>0.07</td>
<td>0.00</td>
<td>0.02</td>
<td>CeSem=Inanim</td>
<td>0.07</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>CeVol=No</td>
<td>0.05</td>
<td>0.00</td>
<td>0.01</td>
<td>CeVol=Yes</td>
<td>0.14</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>CeVol=Yes</td>
<td>0.14</td>
<td>0.00</td>
<td>0.01</td>
<td>CeVol=Yes</td>
<td>0.14</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Interpretation of dimensions

• Dim1: autonomy (animacy, volitionality, control) of the Causee
• Dim2: non-intentionality (and inanimacy) of the Causer
• Dim3: coreferentiality (and Causer’s involvement)

Coordinates of the 344 causative situations on the dimensions will be predictor variables in regression analysis (Dim1, Dim2 and Dim3). Thus, we have 3 orthogonal variables instead of 13 associated ones!
Regression modelling

• First attempt: ordinal regression with ordinal response (Lexical > Morphological > Analytic), the dimensional coordinates as fixed effects and 344 semantic situations and 10 languages as crossed random effects.
• clmm function in package ordinal
• A nice model, but…
A problem with ordinal model

- Assumption of proportional odds (i.e. the effects of the predictors are the same regardless of the ‘threshold’).
- Separate language-specific fixed-effect models and partial residual plots (package rms) show that this assumption does not hold.
An example: Indonesian
Binary and multinomial logistic models

• Another problem: only in 4 languages all three levels are decently represented.

• Solution: fit 10 separate regression models for each language and compare the coefficients
  – 5 binary models with Lex or Ana (fra, rus, tha, vie, zho)
  – 1 binary model with Lex or Morph (jpn)
  – 4 multinomial models with Lex, Morph or Ana (fin, heb, ind and tur)

• Packages rms (Harrell 2015) and mlogit (Croissant 2013)

• Predictors: Dim1 and Dim2 (Dim3 non-significant)
Dim1: Coefficients and 95% Confidence Intervals

- zho_Ana
- vie_Ana
- tur_Morph, tur_Ana
- tha_Ana
- rus_Ana
- jpn_Morph
- ind_Morph, ind_Ana
- heb_Morph, heb_Ana
- fra_Ana
- fin_Morph, fin_Ana

log-odds

-2  0  2  4
Dim2: Coefficients and 95% Confidence Intervals

- zho_Ana
- vie_Ana
- tur_Morph
- tur_Ana
- tha_Ana
- rus_Ana
- jpn_Morph
- ind_Morph
- ind_Ana
- heb_Morph
- heb_Ana
- fra_Ana
- fin_Morph
- fin_Ana

log-odds
Outline

1. Iconicity in causative constructions
2. Data and variables
3. Quantitative analyses
4. Discussion
Results

- Variation is clearly multifactorial. There are two general semantic factors: autonomy of the Causee (Dim1) and (un)intentionality of the Causer (Dim2).
- On both dimensions, languages mostly ‘agree’ between themselves.
- Overall, Lexical and Morphological causatives are more similar to each other than to Analytic causatives.
- The models demonstrate that multifactorial variation is not only cross-linguistic (Dixon), but is also intra-linguistic.
Discussion

• At the same time, we have found evidence of form-meaning iconicity: the less direct causation (Dim1), the less compact forms.

• Why? The Principle of Iconicity (Haiman 1985) as a form-determining principle?

• But this does not explain why there are differences between the constructions wrt. the second dimension, too.
An alternative view

- A higher-level usage-based explanation:
  - Indirect causation, as well as non-intentional causation, may be less frequent/familiar than the causation type expressed by lexical causatives, very similar to the transitive prototype (Hopper & Thompson 1980)?
Thank you!

The slides will be available at

www.natalialevshina.com/presentations.html

Questions? Suggestions?

natalevs@gmail.com