What causes causatives?
A corpus-based typological study

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

1. Causative constructions
2. Data and variables
3. Quantitative analyses
4. Results and discussion
The causative continuum

Lexical <-> Morphological <-> Analytic (Periphrastic)

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill*, *break*

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. kill, break     e.g. Turkish öldür- “kill”
                     from öl- “die”

The causative continuum

Lexical <-> Morphological <-> Analytic (Periphrastic)

e.g. *kill, break*  
e.g. Turkish öldür- “kill”  
from öl- “die”  
e.g. *cause X to die, make X disappear*

The causative continuum

Lexical <-> Morphological <-> Analytic (Periphrastic)

e.g. *kill, break*  e.g. Turkish *öldür*- “kill”  e.g. *cause X to die, make X disappear*

from *öl*- “die”

most compact  FORM  least compact

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill*, *break*  
ed.g. Turkish *öldür*- “kill”  
from *öl*- “die”  
ed.g. *cause X to die*,  
*make X disappear*

most compact  FORM  least compact

most direct causation  MEANING  least direct causation

Iconicity

• Most researchers agree that the degree of formal integration correlates with the degree of semantic integration of the cause and effect (e.g. Comrie 1981; Haiman 1983; Givón 1990).

• 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

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

Relating to VERB
Relating to Causee
Relating to Causer

Action
(Di)transitive
Control
Unwilling (‘make’)
Fully affected
Indirect
Accidental
With effort, violence

Less compact
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
Dixon’s parameters

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

- **Related to**
  - base VERB
  - Causee
  - Causer

More compact

Less compact
Dixon’s parameters

More compact

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

Related to VERB
(Di)transitive
Control

Related to Causee
Unwilling (‘make’)
Fully affected

Related to Causer
Indirect
Accidental
With effort, violence

Less 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. Causative constructions
2. Data and variables
3. Quantitative analyses
4. Results and 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
- NOAH

**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. Cậu sẽ làm nó nổi giận đó.* (Analytic)
Proportions of types of causative constructions in 10 languages

Legend:
- Lex
- Morph
- Ana

Countries:
- FIN
- FRA
- HEB
- IND
- JPN
- RUS
- THA
- TUR
- VIE
- ZHO
## Examples of constructions

<table>
<thead>
<tr>
<th>Language</th>
<th>Lexical</th>
<th>Morphological</th>
<th>Analytic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>sha “kill”</td>
<td>-</td>
<td>ràng “let, make“ + Pred</td>
</tr>
<tr>
<td>Finnish</td>
<td>tappaa “kill”</td>
<td>odotu-tt-aa “make wait”</td>
<td>antaa “give“ + V1</td>
</tr>
<tr>
<td>French</td>
<td>tuer “kill”</td>
<td>-</td>
<td>faire + Vinf</td>
</tr>
<tr>
<td>Hebrew</td>
<td>harag “kill” pa’al</td>
<td>hotsi “take out” hiph’il</td>
<td>natan “give“ + le-Vinf</td>
</tr>
<tr>
<td>Indonesian</td>
<td>mem-bunuh “kill”</td>
<td>meng-ingat-kan “remind”</td>
<td>membuat “make“ + Pred</td>
</tr>
<tr>
<td>Japanese</td>
<td>korosu “kill”</td>
<td>ikar-ase-ru “make angry”</td>
<td>V_te + morau “get”</td>
</tr>
<tr>
<td>Russian</td>
<td>ubit’ “kill”</td>
<td>-</td>
<td>zastavit’ + Vinf</td>
</tr>
<tr>
<td>Thai</td>
<td>kaa “kill”</td>
<td>-</td>
<td>tham hai “do give“ + Pred</td>
</tr>
<tr>
<td>Turkish</td>
<td>açmak &quot;open&quot;</td>
<td>öl-dür- “kill”</td>
<td>V_mA_DAT + izin ver- “allow“</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>giết hài “kill”</td>
<td>-</td>
<td>làm “do“ + Pred</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)
• The English sentences are coded for 13 semantic and syntactic variables (taking into account the context), based on Dixon’s parameters...
Outline

1. Iconicity in causative constructions
2. Data and variables
3. Quantitative analyses
4. Results and 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 Action</td>
<td>John killed Bill.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td></td>
<td>He walked the baby upstairs.</td>
<td>Longer form</td>
</tr>
<tr>
<td>NoPart (number of participants)</td>
<td>2 3</td>
<td>John killed Bill.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I gave him a book.</td>
<td>Longer form</td>
</tr>
<tr>
<td>CeControl (Causee having control)</td>
<td>No Yes</td>
<td>John killed Bill.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bring your friends!</td>
<td>Longer form</td>
</tr>
<tr>
<td>MakeLet</td>
<td>Let Make</td>
<td>She let him go.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td></td>
<td>John killed Bill.</td>
<td>Longer form</td>
</tr>
<tr>
<td>CeVol (volitional Causee)</td>
<td>No Yes</td>
<td>John caused Bill to die.</td>
<td>Shorter form</td>
</tr>
<tr>
<td></td>
<td></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 finger. He had his hair cut.</td>
<td>Shorter form, Longer form</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrIntent (Causer acting intentionally)</td>
<td>Yes</td>
<td>She wrote a paper. It makes me happy.</td>
<td>Shorter form, Longer form</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrForce (Causer acting forcefully)</td>
<td>No</td>
<td>He got him to do it. He forced him to do it.</td>
<td>Shorter form, Longer form</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrInvolve (Causer involved in caused event)</td>
<td>No</td>
<td>John killed Bill. Bring your friends! (and come, too)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Values</td>
<td>Example(s)</td>
<td>Expectations</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>------------------------------------</td>
<td>--------------</td>
</tr>
<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 \textit{CrForce} (the Causer acting forcefully)
- Max = 0.93 \textit{CrIntent} (the Causer acting intentionally)
Outline

1. Causative constructions
2. Data and variables
3. Quantitative analyses
4. Results and discussion
Method: Random forests

- Statistical non-parametric method of classification (predicts the choice between causative constructions based on the 13 semantic and syntactic variables)
- Robust in situations of strongly correlated predictors and in situations of many predictors and few observations
- Based on many individual conditional trees
Example: a conditional tree for French (lexical vs. analytic)

Pourquoi ne pas les avoir laissés me botter le cul?

OK, emmène-moi avec toi.

Tu sais ce que ça me force à faire?

Elle a écrit le livre.
Example: random forest for French (based on 1000 trees)
### Var. importance: lex/ana

<table>
<thead>
<tr>
<th></th>
<th>FIN</th>
<th>FRA</th>
<th>HEB</th>
<th>IND</th>
<th>RUS</th>
<th>THA</th>
<th>TUR</th>
<th>VIE</th>
<th>ZHO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CausedEvent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NoPart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CeControl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MakeLet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CeVol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrDirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrIntent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrForce</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrInvolve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrSem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CeSem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Var. importance: lex/morph

<table>
<thead>
<tr>
<th>Feature</th>
<th>FIN</th>
<th>HEB</th>
<th>IND</th>
<th>JPN</th>
<th>TUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CausedEvent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NoPart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CeControl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MakeLet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CeVol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrDirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrIntent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrForce</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrInvolve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrSem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CeSem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Green squares indicate features with high importance in the respective language.
- Red squares indicate features with low importance in the respective language.
Outline

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

• Variation is clearly multivariate, not only cross-linguistically (Dixon), but also intra-linguistically. This may have to do with a variety of different causative constructions per language that have different functions.

• Variables related to semantic integration of cause and effect are the most prominent cross-linguistically, especially in the lex/ana variation:
  • Making or letting
  • CausedEvent: Action or non-action of the Causee
  • NoPart: Length of causation chain
  • Control of the Causee, etc.
Iconicity?

• Thus, in most cases we do observe iconic relationships: more compact forms express more semantically integrated events, less compact forms express less semantically integrated events.

• However, the Causer’s (non)-intentionality is also important, especially in the lex/morph variation in Finnish, Indonesian and Turkish.
A usage-based explanation

• Indirect causation, as well as non-intentional causation, may be less frequent/familiar than direct and intentional causation = the core of the transitive prototype (Hopper & Thompson 1980).
• More complex forms are preferred to express less likely situations. This allows for efficient form-meaning mapping and equilibrium between speaker’s and hearer’s interests (Horn 1984). A result of language evolution based on cooperative social reasoning (Bergen, Goodman & Levi 2012).
• Is iconicity epiphenomenal?
Thank you!

The slides are available at
www.natalialevshina.com/presentations.html

Questions? Suggestions?
  natalevs@gmail.com