Exploratory and confirmatory methods of investigation of causative constructions in a multilingual parallel corpus

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

1. Theoretical background
2. Variables and hypotheses
3. Data and methods
4. Results of quantitative analyses
5. Conclusions
Causative constructions

• Lexical: *His parents sent him to a concert.*
• Morphological: *ik-ase-ta* (Japanese)
  
  go-CAUSE-PAST
  
  “made go”
• Analytic (periphrastic): *His parents made him go to a concert.*
## Functional approaches

<table>
<thead>
<tr>
<th>Study</th>
<th>More compact causative</th>
<th>Less compact causative</th>
</tr>
</thead>
</table>
| Comrie (1981; 1989) | Direct causation  
Low control of Causee | Indirect causation  
High control of Causee |
| Haiman (1983; 1985) | Smaller conceptual distance between Cause and Result | Greater conceptual distance between Cause and Result |
| Givón (1990)      | Inanimate Manipulee                                       | Human-Agentive Manipulee                                   |
Dixon (2001): Parameters of “compactness”

<table>
<thead>
<tr>
<th></th>
<th>More compact causative</th>
<th>Less compact causative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb aspect</td>
<td>State</td>
<td>Action</td>
</tr>
<tr>
<td>(Di)Transitives</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Causee having control</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Volitional Causee</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Causee affected</td>
<td>Partly</td>
<td>Fully</td>
</tr>
<tr>
<td>Causation</td>
<td>Direct</td>
<td>Indirect</td>
</tr>
<tr>
<td>Causer’s intention</td>
<td>Intentional</td>
<td>Accidental</td>
</tr>
<tr>
<td>Naturalness</td>
<td>Yes</td>
<td>No (effort, violence)</td>
</tr>
</tbody>
</table>
Research question

• What are the factors that determine the choice between lexical and periphrastic causatives in European languages?
  - Germanic: English, German, Dutch and Swedish
  - Romance: French, Italian, Spanish and Portuguese
1. Theoretical background
2. Variables and hypotheses
3. Data and methods
4. Results of quantitative analyses
5. Conclusions
Operationalization of factors as variables

- Making or letting ($\approx$ Causee’s willingness)
- Causer’s intention
- Presence of another intentional participant ($\approx$ Causee’s control)
- Number of participants
- Coreferentiality
Making or Letting

• “Let” : *The Stasi released the prisoner.*

• “Make” : *The Stasi arrested the prisoner.*

(see Talmy 1976)
Intentional Causer

• “Yes”: The professor had her students read an essay.

• “No”: My government’s international politics always makes me feel ashamed.
Another controlling participant in causation chain (not Cr)

• “Yes”: They sent their daughter to MIT.
• “No”: I caused the glass to melt.
Number of participants

• “1”: In 10 months, we release... [a prisoner].
• “2”: In 10 months, we release him.
• “3”: In 10 months, we let him leave the prison.

(explicit core arguments: S, DO, IO)
Coreferentiality

• “Yes”: The jealous professor killed herself.
• “No”: The jealous professor killed her husband.
Expectations

Lexical
- Making
- Causer acting intentionally
- No other controlling participants
- Smaller number of participants
- No coreferentiality

Periphrastic
- Letting
- Causer acting non-intentionally
- Another controlling participant
- Larger number of participants
- Coreferentiality
Outline

1. Theoretical background
2. Variables and hypotheses
3. Data and methods
4. Results of quantitative analyses
5. Conclusions
Subtitles of 8 films
Why film subtitles?

- freely available online for many languages
- timing information helps to align sentences (Jörg Tiedemann‘s UPLUG software)
- closer to spontaneous language than any other collection of parallel texts
Data set

- all contexts with periphrastic causatives in every language; random samples to extract lexical causatives
- 360 parallel contexts
- coded for the cx (periphrastic, lexical or other)
- coded for the semantic features (in each language)
## CONSTRUCTIONS

<table>
<thead>
<tr>
<th>ID</th>
<th>EN</th>
<th>DE</th>
<th>NL</th>
<th>SV</th>
<th>FR</th>
<th>ES</th>
<th>IT</th>
<th>PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amelie1</td>
<td>Per</td>
<td>Lex</td>
<td>NA</td>
<td>Lex</td>
<td>Per</td>
<td>Lex</td>
<td>Per</td>
<td>Lex</td>
</tr>
<tr>
<td>Amelie2</td>
<td>Lex</td>
<td>Lex</td>
<td>Lex</td>
<td>Lex</td>
<td>Per</td>
<td>NA</td>
<td>Per</td>
<td>NA</td>
</tr>
<tr>
<td>Amelie3</td>
<td>NA</td>
<td>Per</td>
<td>Per</td>
<td>NA</td>
<td>Per</td>
<td>Per</td>
<td>NA</td>
<td>Per</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

## SEMANTIC FEATURES

<table>
<thead>
<tr>
<th>ID</th>
<th>EN_Make_Let</th>
<th>EN_Intent_CR</th>
<th>EN_Control_CE</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amelie1</td>
<td>Make</td>
<td>Yes</td>
<td>Yes</td>
<td>...</td>
</tr>
<tr>
<td>Amelie2</td>
<td>Make</td>
<td>No</td>
<td>No</td>
<td>...</td>
</tr>
<tr>
<td>Amelie3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

An example
Translationese?

• A manual check of frequencies of *faire* + V and *laisser* + V in 20 films in original French and translated from English
• *make* + V and *let* + V in 20 films in original English and translated from French
• no significant differences in the normalized frequencies of the constructions in original and translated subtitles (the unpaired Wilcoxon test)
Spatiotemporal limitations?

• frequencies of *make* + V and *let* + V in ten films scripts and relate subtitles
• The paired Wilcoxon tests do not yield any significant differences in the frequencies of the causative constructions in scripts and subtitles
Methods

• Exploratory:
  - exemplar-based MDS with 2D Kernel Density Estimation

• Confirmatory:
  - a series of binomial logistic regressions
    (comparison of 95% confidence intervals)
Outline

1. Theoretical background
2. Variables and hypotheses
3. Data and methods
4. Results of quantitative analyses
5. Conclusions
Multidimensional Scaling

• Gower distances for binary & categorical data between all pairs of exemplars:
  small distance = similar cxs in all languages
  large distance = different cxs in all languages
• MDS with stress majorization (smacof package)
Exemplar-based semantic map (MDS)
I always get them when somebody makes me do something I don't want to do.
That’s him, get him!
She likes cracking bones. Ce qu’elle aime, c’est faire craquer les os.
That's how he got caught. C'est là qu'il se fait prendre.
So it may live long...
Laat haar onder ons zijn
based on semantic features of exemplars in the English subtitles
MDS, Causer Acting Intentionally

based on semantic features of exemplars in the English subtitles
MDS, Controlling Participant (Causee)

Based on semantic features of exemplars in the English subtitles.
based on semantic features of exemplars in the English subtitles
Coreferentiality

based on semantic features of exemplars in the English subtitles
MDS, German periphrastic causatives
MDS, Swedish periphrastic causatives
MDS, French periphrastic causatives

![Graph showing MDS, French periphrastic causatives](image-url)
MDS, Italian periphrastic causatives
MDS, Spanish periphrastic causatives
Density: Germanic, periphrastic
Density: Germanic, lexical
Density: Romance, periphrastic
Density: Romance, lexical
Multidimensional Scaling: Summary

• Romance and Germanic PerCx have different range of functions
• the differences within each group seem to correspond to varying degrees of compactness of PerCx:
  French & Italian > Spanish > Portuguese
  German > Dutch > English > Swedish
Logistic regression

• 8 models (one for every language)
• Response: lexical or periphrastic?
• Predictors: Semantic variables
The diagram shows regression estimates for various linguistic features related to lexical and periphrastic expressions. The features include:

- **MAKE (vs. LET)**
- **INTENTIONAL CR**
- **CONTROL CE**
- **NO. PARTICIPANTS**
- **COREFERENTIALITY**

Each feature is represented by horizontal lines, with different colors corresponding to different languages (PT, ES, IT, FR, DE, NL, EN, SV). The x-axis represents the regression estimates, ranging from -4 to 4.
## Predictive power

<table>
<thead>
<tr>
<th>Language</th>
<th>R2</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>0.48</td>
<td>0.83</td>
</tr>
<tr>
<td>German</td>
<td>0.29</td>
<td>0.76</td>
</tr>
<tr>
<td>Dutch</td>
<td>0.41</td>
<td>0.82</td>
</tr>
<tr>
<td>Swedish</td>
<td>0.48</td>
<td>0.85</td>
</tr>
<tr>
<td>French</td>
<td>0.34</td>
<td>0.8</td>
</tr>
<tr>
<td>Italian</td>
<td>0.27</td>
<td>0.76</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.29</td>
<td>0.77</td>
</tr>
<tr>
<td>Portuguese</td>
<td>0.4</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Outline

1. Theoretical background
2. Variables and hypotheses
3. Data and methods
4. Results of quantitative analyses
5. Conclusions
Conclusions

• Effects are observed in accordance with predictions, in most languages, except for coreferentiality (significant only in 2 lang).
• The effect sizes are cross-linguistically similar.
• However, the sizes of semantic areas of periphrastic causatives vary, partly correlating with the degree of compactness.
• A combination of exploratory and confirmatory methods is needed to find this out.
To be continued...

natalevs@gmail.com
natalia.levshina@uclouvain.be