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 *odotu**tt**aa* “cause to wait” (from *odottaa* “wait”)

- Analytic = two predicates

  e.g. *make X cry, let X go, make X happy*

High formal integration, most compact

Low formal integration, least compact

Semantic regularities

Study	More compact causative	Less compact causative
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

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		Less compact
State (or change of state) Intransitive No control	**Relating to VERB**	Action (Di)transitive Control
Willing ('let') Partially affected	**Relating to Causee**	Unwilling ('make') Fully affected
Direct Intentional Natural	**Relating to Causer**	Indirect Accidental With effort, violence

Dixon’s parameters

More compact ⟷ Less compact

State (or change of state)
Intransitive

Related to base VERB

Action
(Di)transitive

No control
Willing (‘let’)
Partially affected

Related to Causee

Control
Unwilling (‘make’)
Fully affected

Direct
Intentional
Natural

Related to Causer

Indirect
Accidental
With effort, violence

Dixon’s parameters

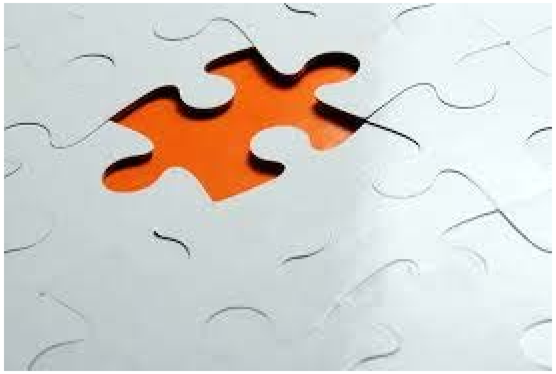
More compact		Less compact
State (or change of state) Intransitive	**Related to VERB**	Action (Di)transitive
No control Willing (‘let’) Partially affected	**Related to Causee**	Control Unwilling (‘make’) Fully affected
Direct Intentional Natural	**Related to Causer**	Indirect Accidental With effort, violence

Dixon's parameters

More compact		Less compact
State (or change of state) Intransitive No control	**Related to VERB**	Action (Di)transitive Control
Willing ('let') Partially affected	**Related to Causee**	Unwilling ('make') Fully affected
Direct Intentional Natural	**Related to Causer**	Indirect Accidental With effort, violence

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!



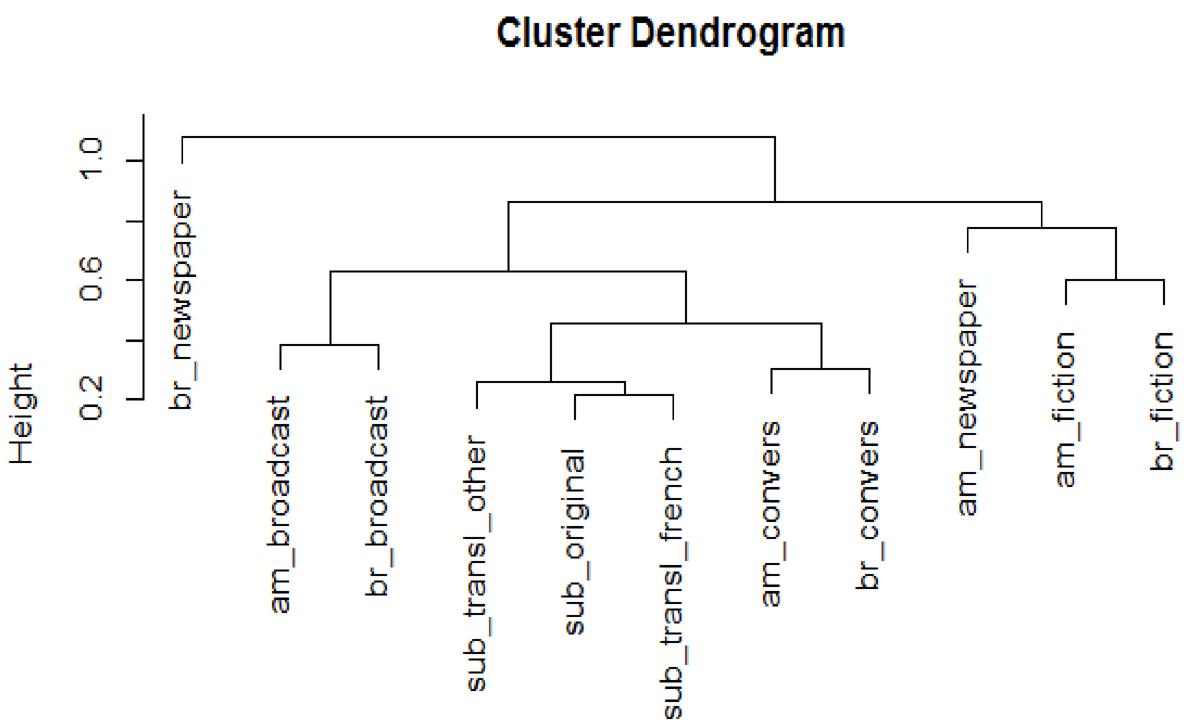
Outline

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…

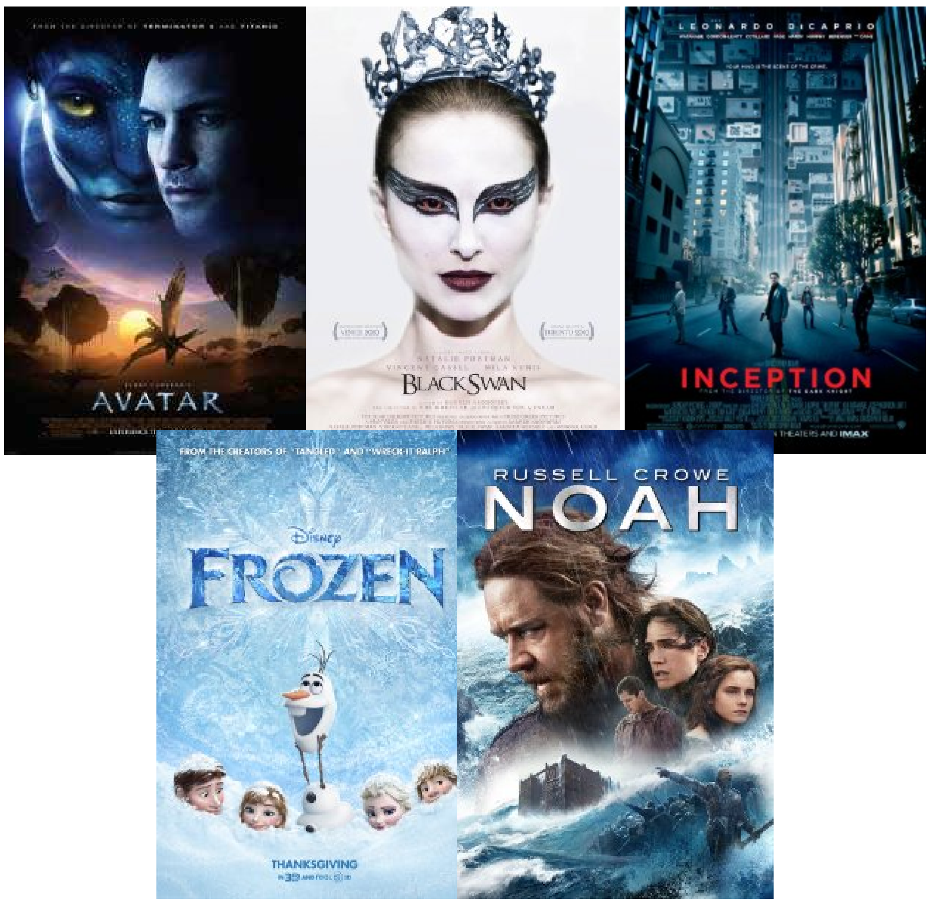
Why subtitles?



Based on the frequencies of 3-grams (Levshina, Accepted)

Data used in the case study

Films



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

Language	Genus	Family
Chinese	Chinese	Sino-Tibetan
Finnish	Finnic	Uralic
French	Romance	Indo-European
Hebrew	Semitic	Afro-Asiatic
Indonesian	Malayo-Sumbawan	Austronesian
Japanese	Japanese	Japanese
Russian	Slavic	Indo-European
Thai	Kam-Tai	Tai-Kadai
Turkish	Turkic	Altaic
Vietnamese	Viet-Muong	Austro-Asiatic

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'énerver*. (Lexical)
- TUR: *Ateş etme. Ateş etme. Onu kızdıracaksın*. (Morphological, from *kızmek* ‘become angry’).
- VIE: *Đừng bắn. Cậu 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)

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

Variables (2)

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

Variables (3)

Variable	Values	Example(s)	Expectations
Coref (coreferentiality)	Yes No	He killed himself. He killed Bill.	None
Polarity	Pos Neg	She let him do it. She didn't let him do it.	None
CrSem (semantics of Causer)	Anim Inanim	She made him stay. The rain made him stay.	None
CeSem (semantics of Causee)	Anim Inanim	John let Mary go. John let it go.	None

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

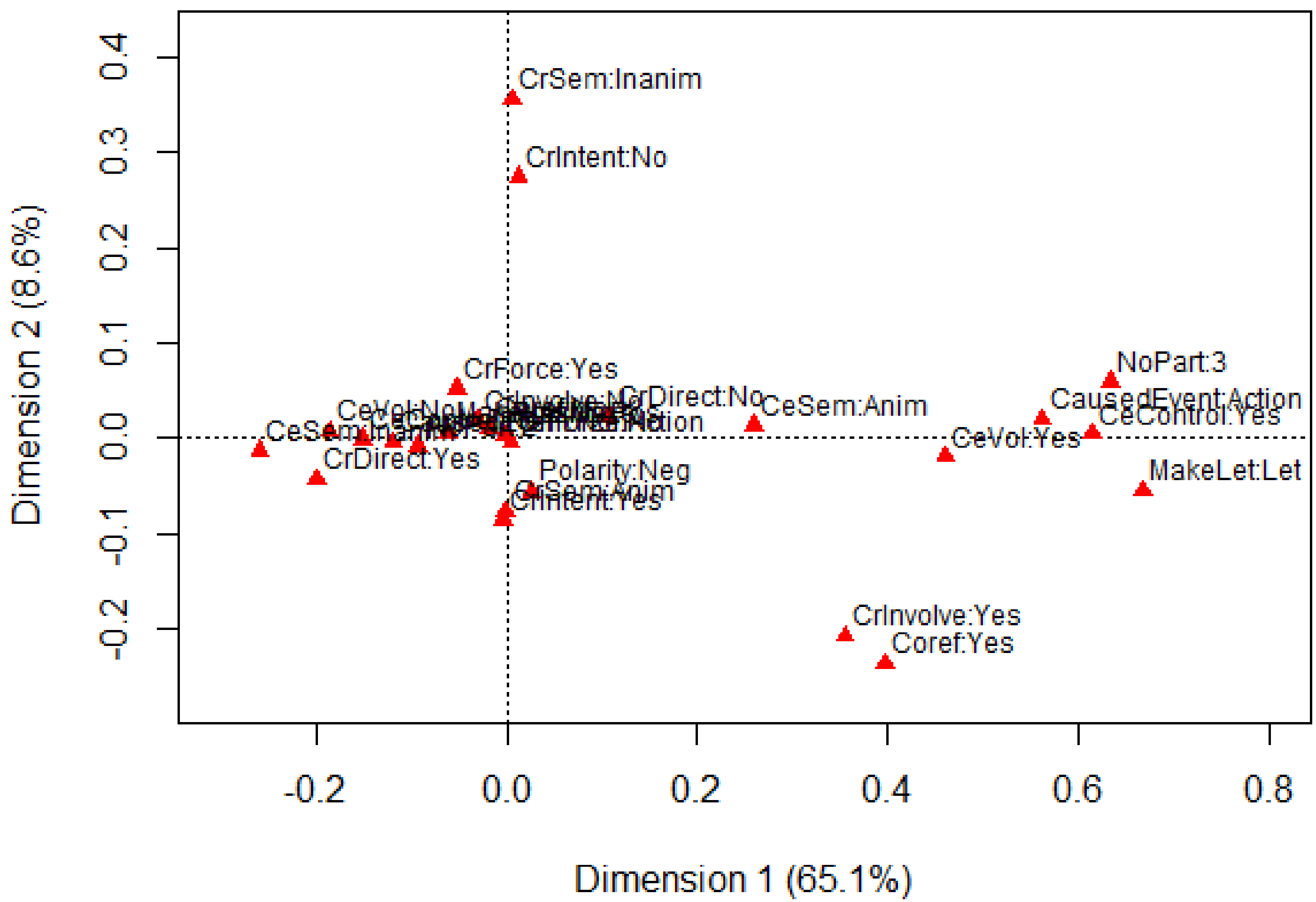
- The most appropriate method: multiple regression analysis with Cx (Lexical, Morphological and Analytic) as response and the semantic and syntactic variables as predictors.
- But: highly associated semantic variables → danger of multicollinearity
- Solution:
  - Adjusted Multiple Correspondence Analysis of the 13 variables as a dimensionality-reduction technique
  - R packages `ca` (Nenadič & Greenacre 2007) and `FactoMineR` (Husson et al. 2015)

MCA: Explained variance (inertia)

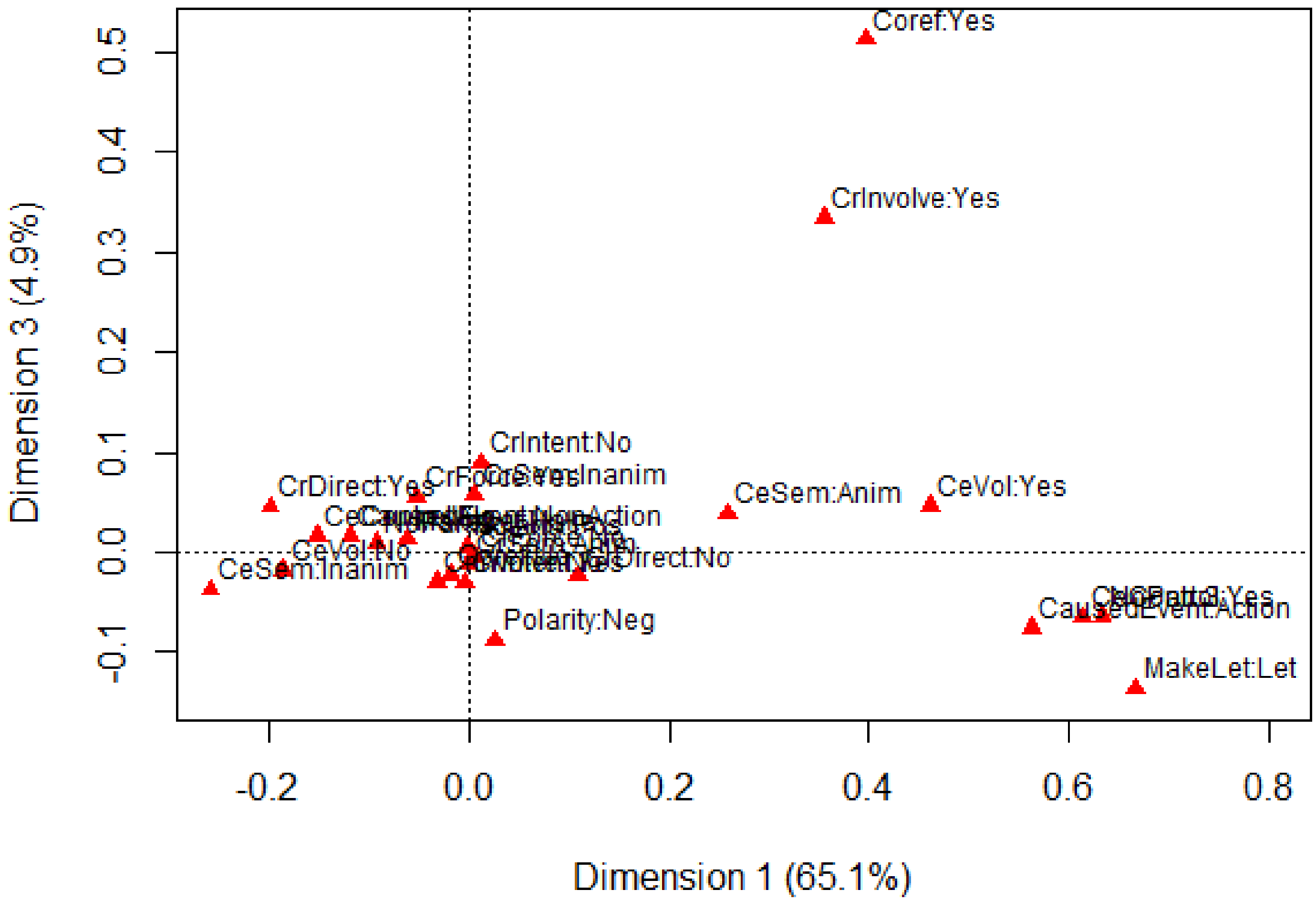
Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.034794	65.1	65.1	*********************
2	0.004613	8.6	73.8	***
3	0.002605	4.9	78.6	**
4	0.000180	0.3	79.0	
5	9e-06000	0.0	79.0	

MCA: Dimensions 1 & 2



MCA: Dimensions 1 & 3



Contributions to dimensions

Feature	D1	D2	D3
CrIntent=No	0.00	0.30	0.06
CrIntent=Yes	0.00	0.09	0.02
CrForce=No	0.00	0.00	0.00
CrForce=Yes	0.00	0.00	0.01
CrInvolve=No	0.00	0.01	0.02
CrInvolve=Yes	0.02	0.06	0.27
CrDirect=No	0.02	0.01	0.01
CrDirect=Yes	0.03	0.01	0.02
CeControl=No	0.04	0.00	0.01
CeControl=Yes	0.16	0.00	0.02
MakeLet=Let	0.08	0.00	0.05
MakeLet=Make	0.01	0.00	0.00
CdEvent =Action	0.12	0.00	0.03

Feature	D1	D2	D3
CdEvent=NAction	0.03	0.00	0.01
NoPart=2	0.02	0.00	0.00
NoPart=3	0.11	0.01	0.02
Coref=No	0.00	0.00	0.02
Coref=Yes	0.02	0.04	0.34
Polarity=Neg	0.00	0.00	0.01
Polarity=Pos	0.00	0.00	0.00
CrSem=Anim	0.00	0.08	0.00
CrSem=Inanim	0.00	0.37	0.02
CeSem=Anim	0.07	0.00	0.02
CeSem=Inanim	0.07	0.00	0.02
CeVol=No	0.05	0.00	0.01
CeVol=Yes	0.14	0.00	0.0

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!

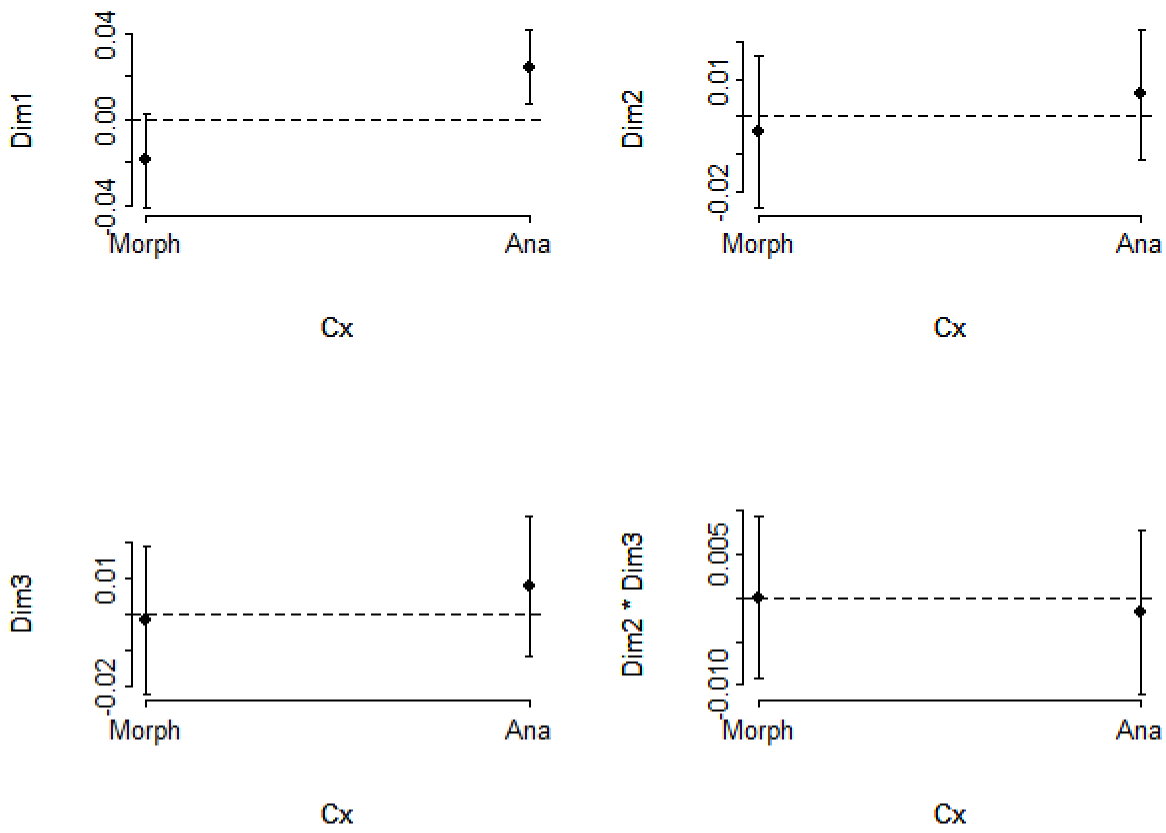
Regression modelling

- 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

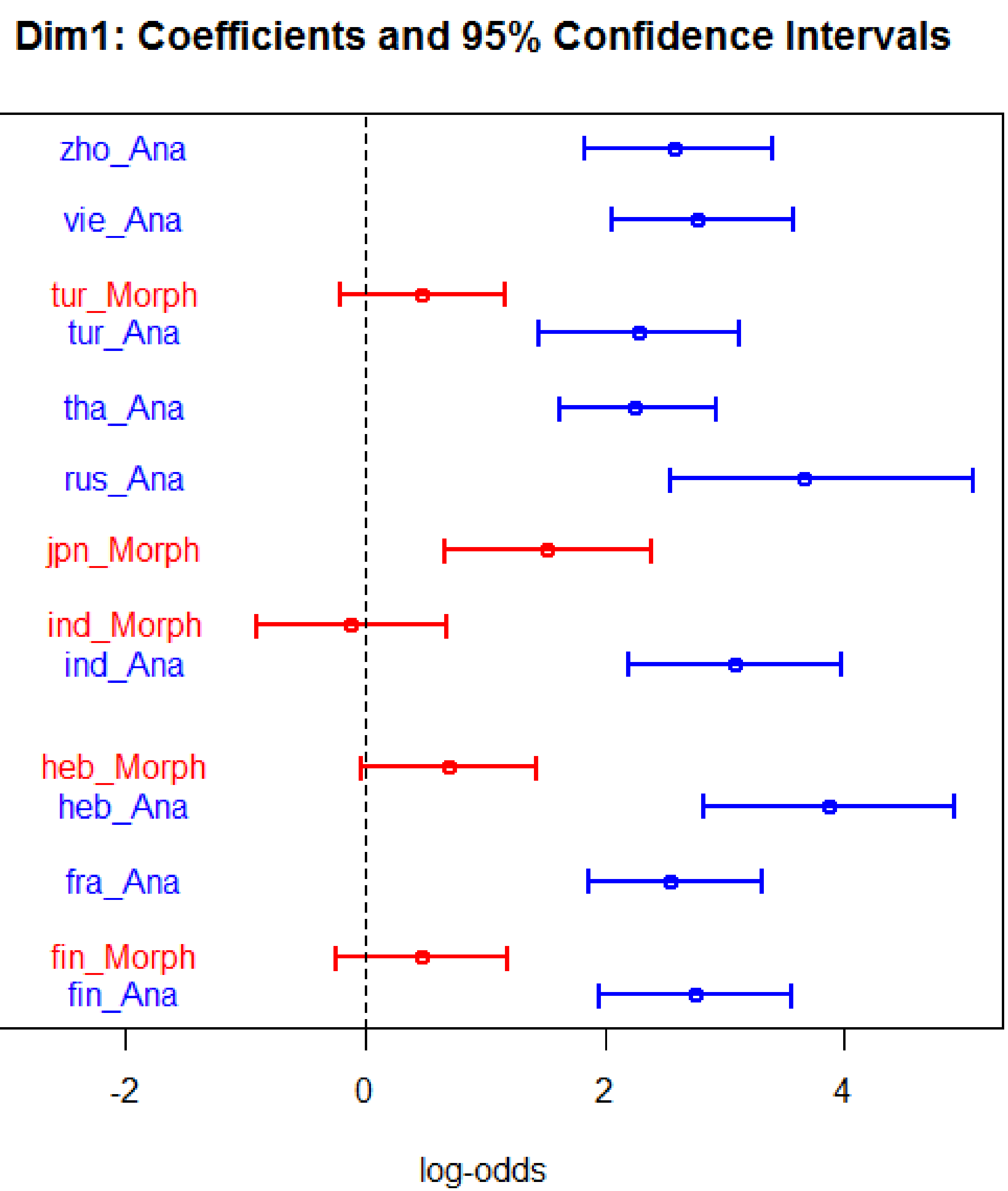
- Assumption of proportional odds (i.e. the effects of the predictors are the same regardless of the 'threshold').
- Separate language-specific fixed-effect models and partial residual plots (package rms) show that this assumption does not hold.

An example: Indonesian

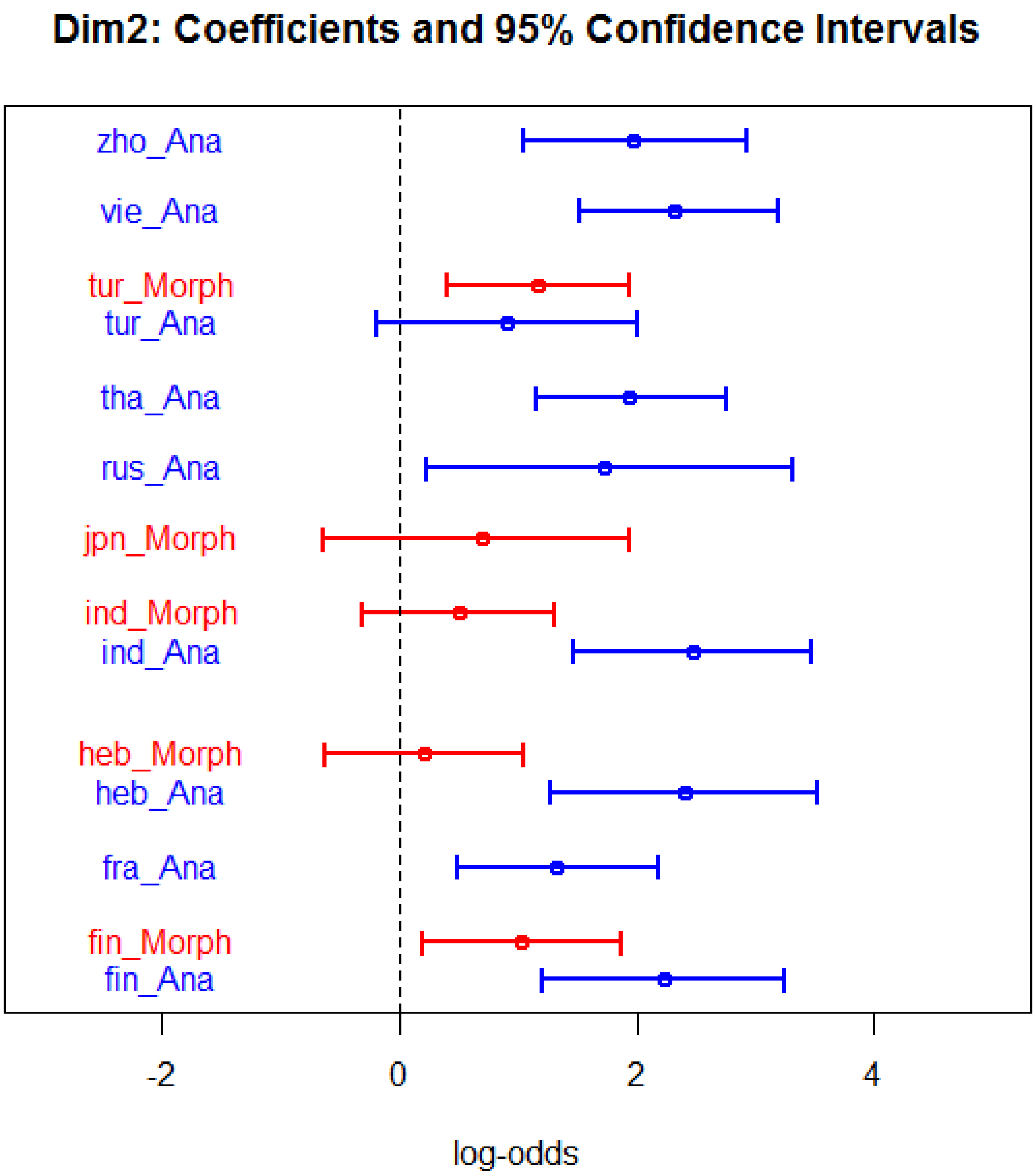


Binary and multinomial logistic models

- 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
-2
0
2
4
log-odds


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
-2
0
2
4
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:
  - Haspelmath 2008; Haspelmath et al. 2014: Less frequent/familiar situations tend to be expressed by longer forms (Principle of Economy).
  - 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