Iconicity and Usage:

A quantitative corpus-based study of European causatives

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

Outline

1. Iconicity and causation
   - Case study 1: testing the iconicity predictions re. causative constructions in 15 European languages
2. Iconicity and/or usage?
   - Case study 2: interplay of iconicity and frequency in British English
3. Conclusions

Analytic vs. lexical causatives

a. He made his cat come back.

b. She brought her cat back.

Iconicity

Study	Less integrated/compact causative	More integrated/compact causative
Givón (1980)	Lower degree of semantic binding between 2 events	Higher degree of semantic binding between 2 events
Comrie (1981; 1989)	Indirect causation Higher control of Causee	Direct causation Lower control of Causee
Haiman (1983; 1985)	Greater conceptual distance between Cause and Result	Smaller conceptual distance between Cause and Result
Givón (1990)	Human-Agentive Manipulee	Inanimate Manipulee

Analytic vs. lexical causatives

a. He *made* *his cat* *come back.*

b. She *brought* *her cat back.*

Iconicity

Study	Less integrated/compact causative	More integrated/compact causative
Givón (1980)	Lower degree of semantic binding between 2 events	Higher degree of semantic binding between 2 events
Comrie (1981; 1[illegible]	[illegible]	[illegible]ct causation [illegible]ntrol of Causee
Haiman (1983; 1985)	[illegible] between Cause and Result	[illegible]maller conceptual distance between Cause and Result
Givón (1990)	Human-Agentive Manipulee	Inanimate Manipulee

Never been tested empirically, only more or less anecdotal evidence.

Multilingual corpus of film subtitles

.srt format

…

646
00:51:27,880 --> 00:51:32,920
<i>For always evil will look to
find a foothold in this world.</i>

647
00:51:39,440 --> 00:51:42,603
Not good. Not good at all.

648
00:51:50,040 --> 00:51:51,326
Eww.

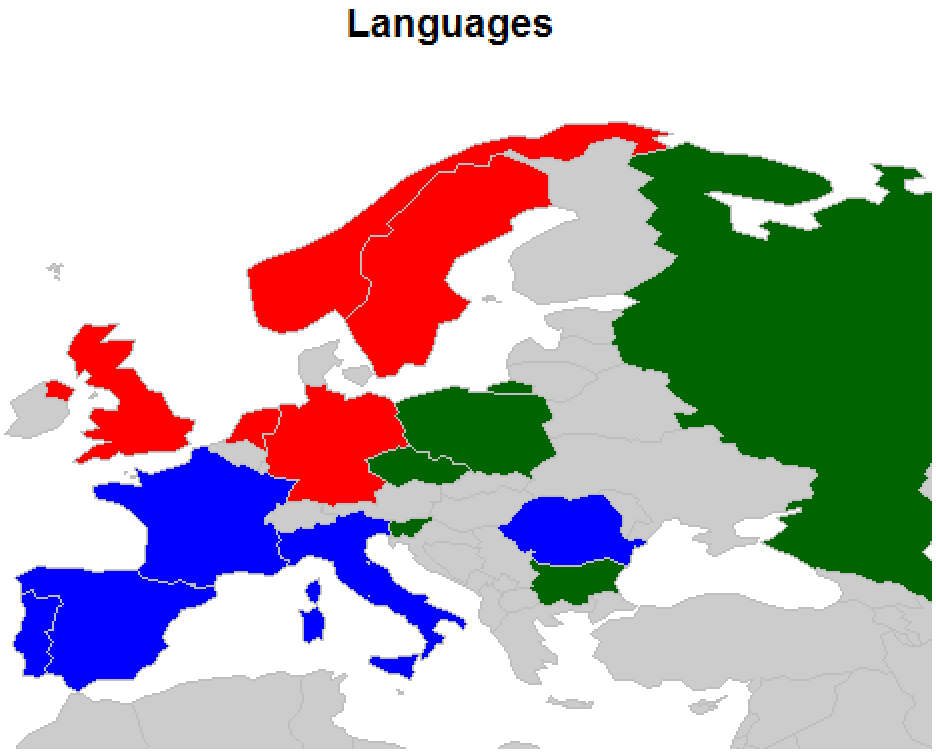
649
00:52:06,760 --> 00:52:09,081
Oh, no. Sebastian.

650
00:52:12,800 --> 00:52:13,847
Good gracious.

651
00:52:34,720 --> 00:52:35,767
Come on.

…

Languages



Data

- Search for ACs
- 363 observations in 15 languages, at least one language contains an AC
- Equivalents in all other languages with the help of sentence alignment

Some examples of ACs

- Eng. *make* + Vinf, *cause* + *to* Vinf, *let* + Vinf, *force* + to Vinf
- Du. *laten* + Vinf, *doen* + Vinf
- Fr. *faire* + Vinf, *laisser* + Vinf
- Ger. *lassen* + Vinf, *bringen* + *zum* Vinf
- Sp. *hacer* + Vinf, *dejar* + Vinf
- Rom. *a face* + *să* Vsubj, *a lasă* + *să* Vsubj
- Cz. *dát* + Vinf, *nechat* + Vinf
- Rus. *zastavljat'* + Vinf, *pozwoljat'* + Vinf
- Bul. *karam + da* Vfin

Matrix

ID	FR	EN	DE	ES	NL	SV	IT	PT	RO
1	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	AC
2	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	AC	<NA>	AC
3	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	AC
4	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
5	AC	Lex	Lex	<NA>	Lex	Lex	AC	<NA>	Lex
6	AC	Lex	<NA>	Lex	Lex	Lex	AC	<NA>	Lex
7	AC	AC	AC	Lex	<NA>	Lex	AC	Lex	Lex
8	AC	<NA>	AC	AC	<NA>	<NA>	AC	Lex	AC
9	AC	AC	AC	AC	AC	AC	AC	Lex	Lex
10	AC	<NA>	<NA>	<NA>	<NA>	<NA>	AC	AC	<NA>
11	AC	<NA>	<NA>	<NA>	<NA>	<NA>	AC	Lex	<NA>
12	AC	<NA>	<NA>	<NA>	<NA>	<NA>	AC	<NA>	<NA>

Matrix

ID	FR	EN	DE	ES	NL	SV	IT	PT	RO
1	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	AC
2	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	AC	<NA>	AC
3	<NA>	<NA>	[illegible]	[illegible]	[illegible]	[illegible]	<NA>	<NA>	AC
4	<NA>	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	<NA>	<NA>
5	AC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Lex
6	AC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Lex
7	AC	AC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Lex	Lex
8	AC	<NA>	AC	[illegible]	[illegible]	[illegible]	AC	Lex	AC
9	AC	AC	AC	AC	AC	AC	AC	Lex	Lex
10	AC	<NA>	<NA>	<NA>	<NA>	<NA>	AC	AC	<NA>
11	AC	<NA>	<NA>	<NA>	<NA>	<NA>	AC	Lex	<NA>
12	AC	<NA>	<NA>	<NA>	<NA>	<NA>	AC	<NA>	<NA>

Can we predict which causative construction is used (AC or Lex) in every particular case?

Predictors

Predictor	Meaning	Examples with AC
CauseeAg	Agentivity of the Causee	'Yes': *I made him leave.* 'No': *My reply made him choke.* 'Both' (undefined)
MakeLet	Making of letting	'Make': *I made him leave.* 'Let': *I let him leave.* 'Both' (undefined)
Aff	If there is another affected entity (Affectee)	'Yes': *I made him eat the soup.* 'No': *I made him eat.*
CrSem	Semantic class of the Causer	'Anim': *I made him leave.* 'Inanim': *The rain made him want to leave.*
CeSem	Semantic class of the Causee	'Anim': *I made him leave.* 'Inanim': *You made my sorrows go away.*
Coref	Coreferentiality of Causer with Causee or Affectee	'Yes': *He let himself go.* 'No': *He let the people go.*

Predictors

Predictor	Meaning	Examples with AC
CauseeAg	Agentivity of the Causee	'Yes': *I made him leave.* 'No': *I made him choke.* 'Both' (undefined)
MakeLet	Making of letting	'Make': *I made him leave.* 'Let': *I let him leave.* 'Both' (undefined)
Aff	If there is another affected entity (Affectee)	'Yes': *I made him eat the soup.* 'No': *I made him eat.*
CrSem	Semantic class of the Causer	'Anim': *I made him leave.* 'Inanim': *The rain made him want to leave.*
CeSem	Semantic class of the Causee	'Anim': *I made him leave.* 'Inanim': *You made my sorrows go away.*
Coref	Coreferentiality of Causer with Causee or Affectee	'Yes': *He let himself go.* 'No': *He let the people go.*

boost AC

Mixed effect logistic regression model

- Regression analysis allows one to measure the effect of each predictor on the response variable while taking into account the other predictors.
- Response: binary, AC or LC (in every situation and every language)
- Mixed effect models enable one to deal with repeated measures correctly (here: situations and languages, treated as random effects)
- glmer (lme4, R)
- C = 0.96

Estimated coefficients

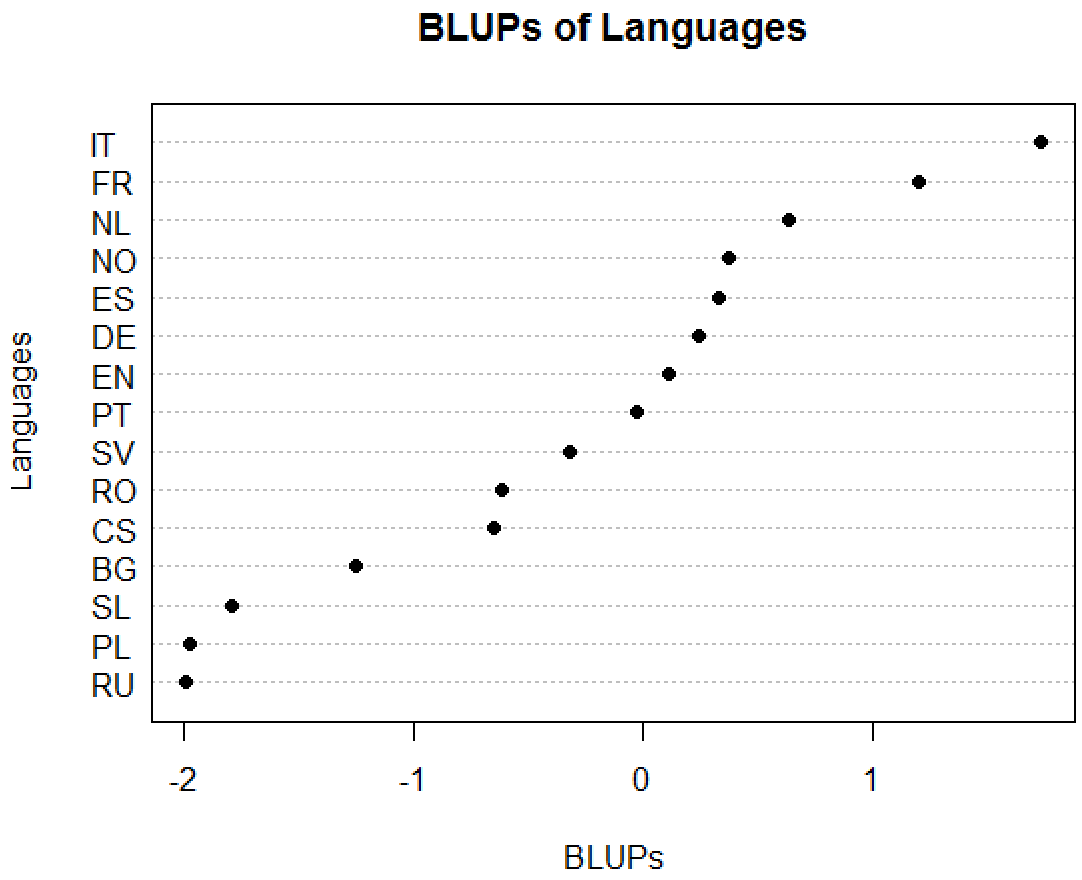
```
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.0042     0.4347  -2.310  0.02088 *
MakeLetBoth   -0.3476     0.5920  -0.587  0.55701
MakeLetLet     2.3020     0.3696   6.228 4.73e-10 ***
CauseeAgBoth  -0.7049     0.6842  -1.030  0.30291
CauseeAgYes    1.4935     0.3704   4.032 5.53e-05 ***
AffYes         1.3942     0.3931   3.547  0.00039 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
1
```

Estimated coefficients

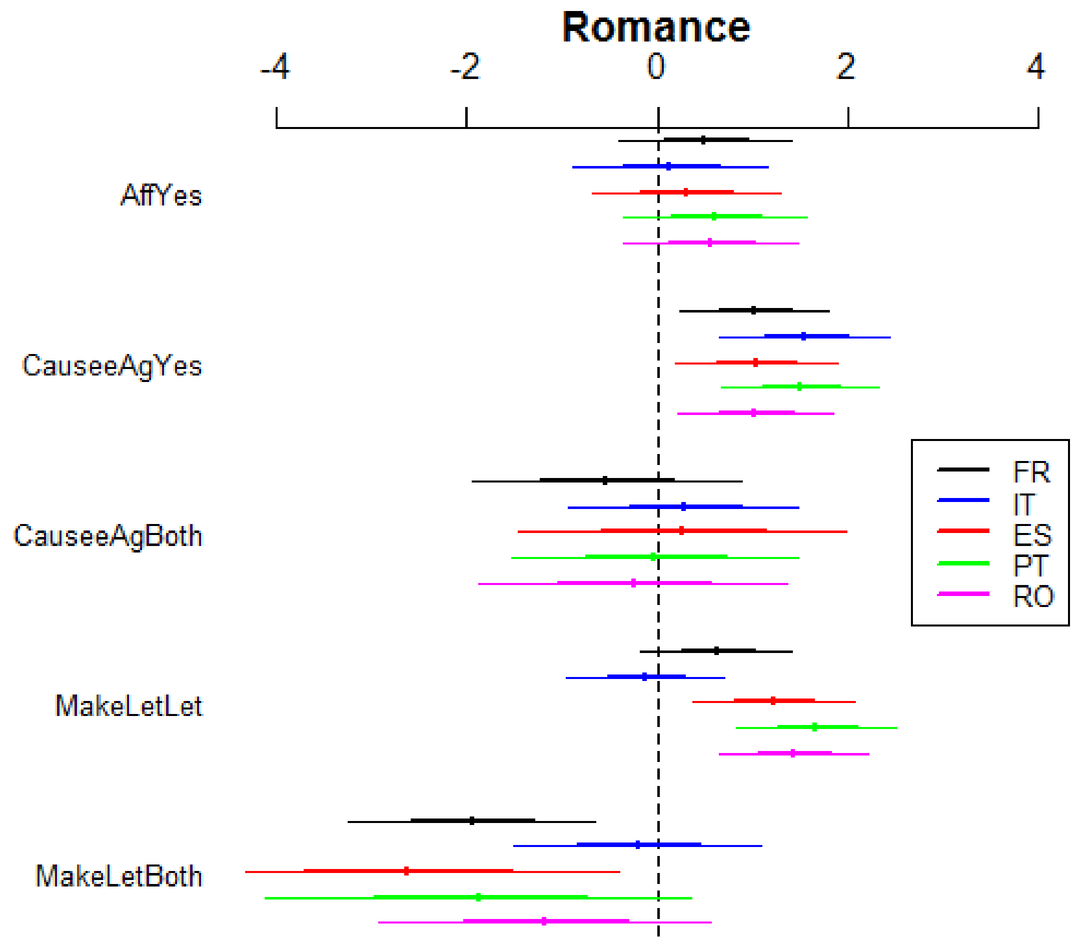
```
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.0042     0.4347  -2.310  0.02088 *
MakeLetBoth   -0.3476     0.5920  -0.587  0.55701
MakeLetLet     2.3020     0.3696   6.228 4.73e-10 ***
CauseeAgBoth  -0.7049     0.6842  -1.030  0.30291
CauseeAgYes    1.4935     0.3704   4.032 5.53e-05 ***
AffYes         1.3942     0.3931   3.547  0.00039 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
1
```

All iconicity-based expectations are borne out!

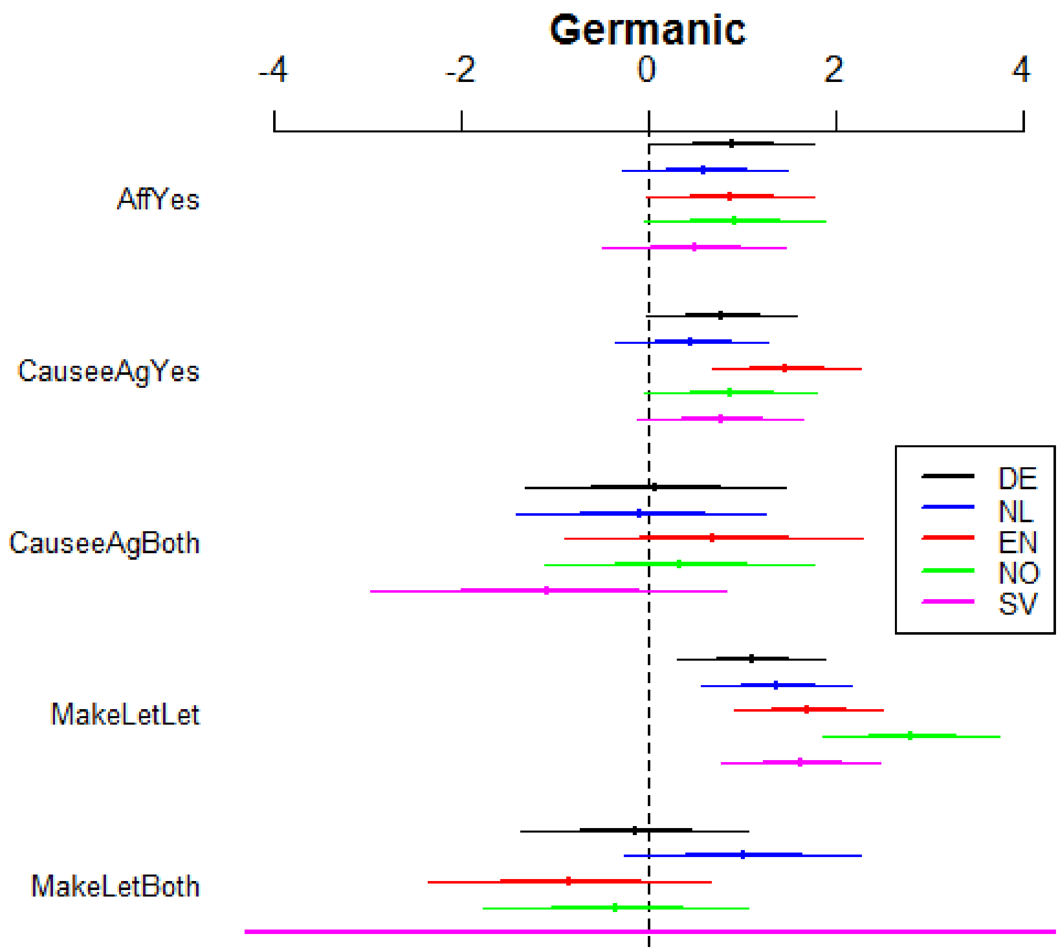
Languages as random effects



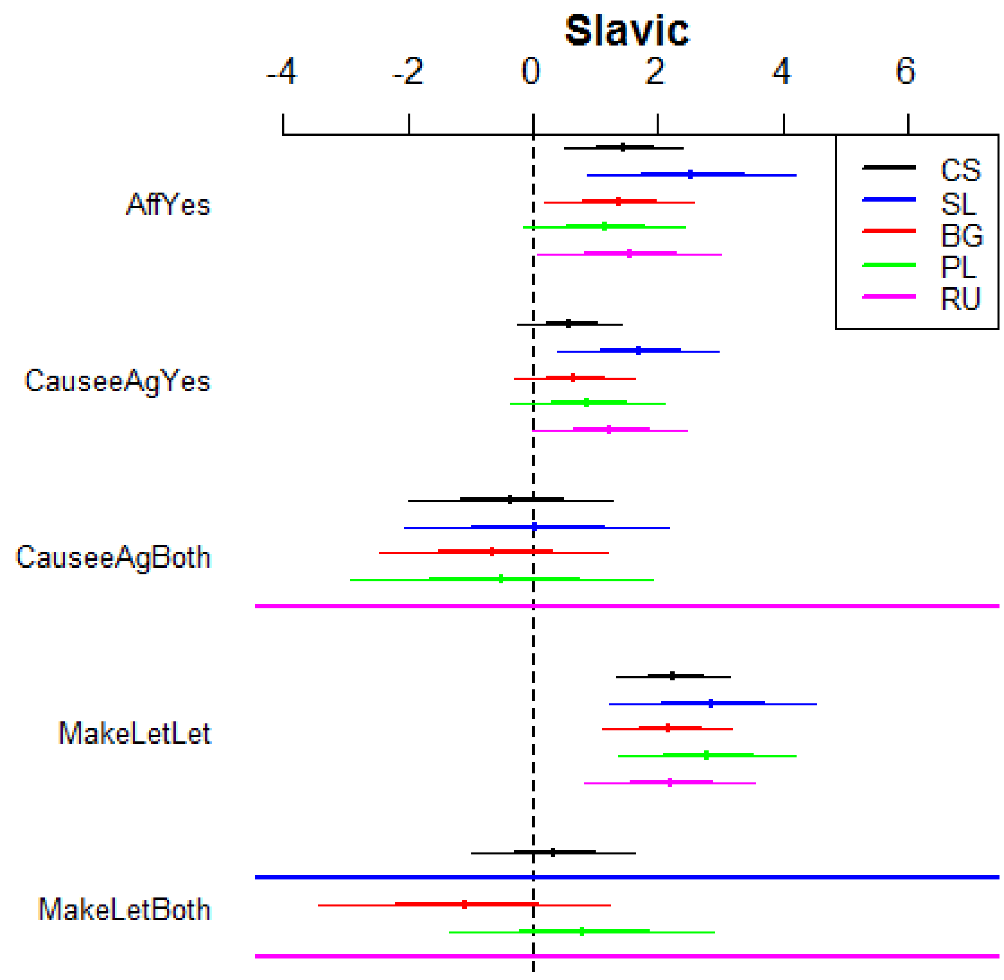
Separate models: Romance



Separate models: Germanic



Separate models: Slavic



Interim conclusions

- In general, weaker integration of the causing and caused events boosts the probability of ACs.
- Romance languages: agentivity of the Causee, letting vs. making is neutralized in Italian and French. The most abstract ACs.
- Germanic languages: letting vs. making (probably, due to the fact that the most frequent ACs are with verbs of letting - Ger. *lassen*, Du. *laten*, etc.).
- Slavic languages: letting vs. making, 2 vs. 3 participants. Cf. Nichols et al. (2004): Russian is a detransitivizing language. Transitives are more formally basic than intransitives.

Outline

1. Iconicity and causation
   - Case study 1: testing the iconicity account in 15 European languages
2. Iconicity and/or usage?
   - Case study 2: interplay of iconicity and frequency in British English
3. Conclusions

But what about usage?

- Does the frequency of the alternative LC have influence on the probability of an AC?

  - E.g. does the frequency of ‘cause to die’ in a corpus depend on the frequency of ‘kill’?

- Statistical pre-emption: People learn to avoid novel formulations by systematically witnessing a competing alternative (e.g. Ambridge et al. 2014).

Data

- BNC XML edition (Stanford parser)
- 63 pairs of lexical and analytical causatives
- CAUS + V = make + V/cause + to V/have + V/get + to V as one abstract Cx

Examples

Analytic	Lexical
CAUS + bend (intr.)	bend (tr.)
CAUS + boil (intr.)	boil (tr.)
CAUS + die	kill
CAUS + rise	raise

Examples

Analytic	Lexical
CAUS + bend (intr.)	bend (tr.)
CAUS + boil (intr.)	boil (tr.)
CAUS + die	kill
CAUS + rise	raise

Can we predict the frequency of ACs in every pair based on the semantic characteristics of the autonomous non-causative event AND the frequency of the lexical counterpart?

Predictors

- *Agent_Prop*: proportion of agentive subjects in a sample of non-causative uses, e.g.
    - 0: *boil, burn*
    - 0.5: *pass, fly*
    - 1: *eat, withdraw*
- *Progress_Prop*: proportion of uses in the Progressive tenses as an indicator of durativity, e.g.
    - min: *vary, believe*
    - max: *shake, prepare*
- *Log(LC)*: Log-transformed frequency of the corresponding lexical causative
- Offset: total frequency of the non-causative verb

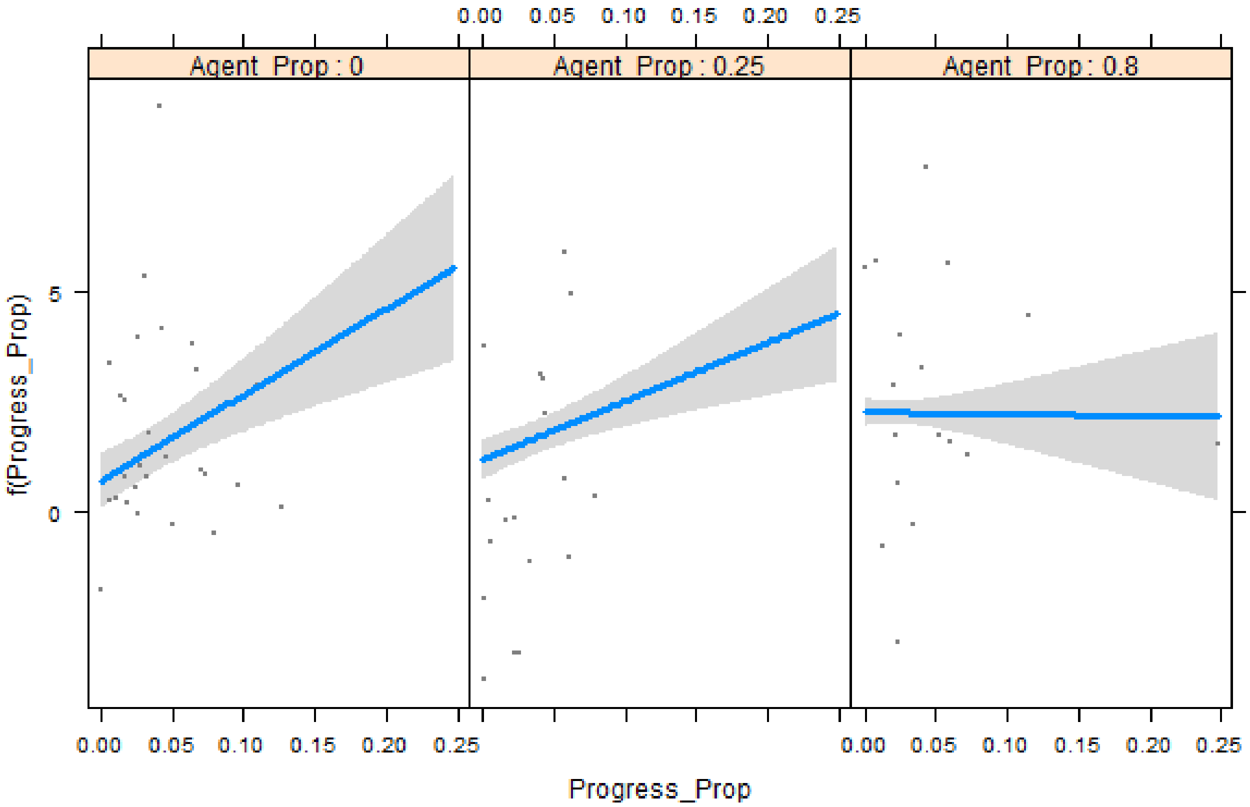
Model

- Quasipoisson glm regression (dispersion parameter = 8.32)
- Pseudo-R^2 = 0.59. A decent result.
- Bootstrap validation due to low *n* (no overfitting).

Estimated coefficients

	Estimate	Std. Error	Pr(>|t|)	
(Intercept)	-4.28119	0.70160	9.78e-08	***
log(LC)	**-0.34656**	**0.07166**	**1.04e-05**	*******
Agent_Prop	**1.94970**	**0.41144**	**1.47e-05**	*******
Progress_Prop	**19.51373**	**4.87467**	**0.000183**	*******
Agent_binYes	-0.55221	0.34748	0.117552	
Agent_Prop:Progress_Prop	**-24.92587**	**8.28368**	**0.003898**	******

Interaction of Agentivity and Durativity



Interim conclusions

- BOTH frequency of the alternative and conceptual factors play a role.
- Verbs that designate more autonomous events, i.e. those which involve an agentive participant or are extended in time, have higher chances to form ACs.
- The frequency of available LCs decreases the probability of ACs, thus supporting the idea of statistical pre-emption.

Outline

1. Iconicity and causation

    - Case study 1: testing the iconicity account in 15 European languages

2. Iconicity and/or usage?

    - Case study 2: interplay of iconicity and frequency in British English

3. Conclusions

Conclusions

- The predictions made by iconicity theory hold cross-linguistically and within one language.
- However, variation of causative constructions is a multifactorial phenomenon depending on both conceptual and usage factors.

To do…

- Semantic similarity between lexical and analytic causatives (might boost the effect of statistical pre-emption).
- Experimental support of iconicity mechanisms.
- Other languages, e.g. Finnish (three types of causative constructions).

Thank you!

natalevs@gmail.com

www.natalialevshina.com