

# PROBABILISTIC SEMANTIC MAPS OF CAUSATION AND CAUSALITY

---

A study based on a multilingual parallel corpus

Natalia Levshina

F.R.S.-FNRS, Université catholique de Louvain

# Outline

1. Quantitative typological Cognitive Semantics
2. Data: ParTy corpus
3. Method: Probabilistic semantic maps
4. Case studies
  - causative constructions
  - causal connectives
5. Conclusions

# Aims of the study

- Traditional Cognitive Semantics: introspective, qualitative, mostly based on English
- The main aim of this study is to promote a different type of Cognitive Semantics: corpus-based, quantitative AND typological

# Quantitative Typological CogSem

- Which conceptual dimensions and structures are the most common typologically, and which are language-specific (language family- or areally specific)?
  - For a given constructional type (e.g. adpositions of spatial relationships)
  - Across different constructions (e.g. adpositions and verbs)
- These dimensions and structures can emerge automatically from parallel corpora.

# Object of the study

	Causation	Causality
Type of constructions	verbal causative constructions, e.g. X breaks Y, X makes Y do Z	causal connectives, e.g. <i>because</i> , <i>therefore</i>
Construal	Energy transfer from Causer to Causee, typically when the Causer overrides the Causee's intrinsic tendency (Talmy 2000)	Relationship between two propositions, one expressing the cause and the other the effect

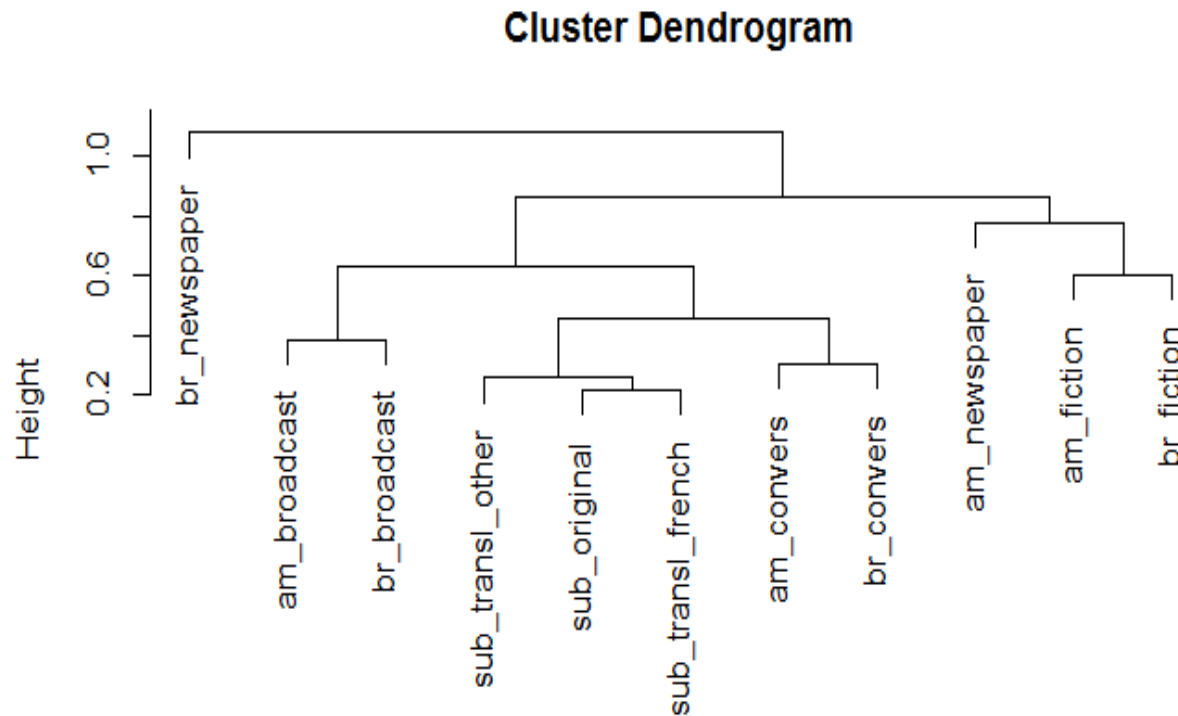
# Outline

1. Quantitative typological Cognitive Semantics
2. Data: ParTy corpus
3. Method: Probabilistic semantic maps
4. Case studies
  - causative constructions
  - causal connectives
5. Conclusions

# ParTy corpus

- a Parallel corpus for Typology
- subtitles of films and TED talks
- mostly European languages, but also other major languages (Chinese, Turkish, Indonesian, etc.)
- all languages aligned with English
- downloadable files at [www.natalialevshina.com/corpus.html](http://www.natalialevshina.com/corpus.html)

# Why subtitles?

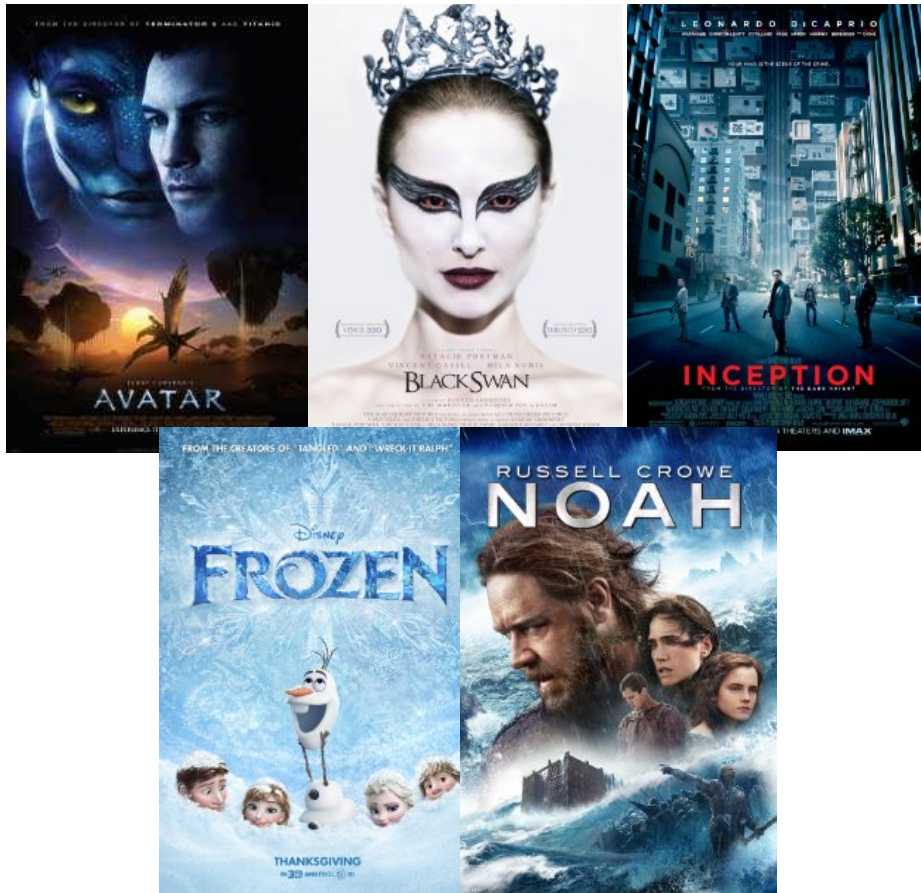


Based on the frequencies of 3-grams (Levshina, Submitted)



# Data used in the case studies

## 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*

# 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

# Outline

1. Quantitative typological Cognitive Semantics
2. Data: ParTy corpus
3. Method: Probabilistic semantic maps
4. Case studies
  - causative constructions
  - causal connectives
5. Conclusions

# Probabilistic semantic maps

- A tool used in typology (Majid et al. 2008; Wälchli 2010; Wälchli & Cysouw 2012; Levshina Forthc.) for different purposes:
  - Induction of cross-linguistic categories (e.g. thematic roles in Hartmann et al. 2014)
  - Cross-linguistic comparison of related constructions (e.g. degrees of grammaticalization of Romance analytic causatives in Levshina, Forthc.)
  - Induction of cross-linguistic semantic dimensions (e.g. categorization of cutting and breaking events in Majid et al. 2008)

# Probabilistic semantic maps

- A tool used in typology (Majid et al. 2008; Wälchli 2010; Wälchli & Cysouw 2012; Levshina Forthc.) for different purposes:
  - Induction of cross-linguistic categories (e.g. thematic roles in Hartmann et al. 2014)
  - Cross-linguistic comparison of related constructions (e.g. degrees of grammaticalization of Romance analytic causatives in Levshina, Forthc.)
  - Induction of cross-linguistic semantic dimensions (e.g. categorization of cutting and breaking events in Majid et al. 2008)

Here: data from a parallel corpus!

# Algorithm for MDS: Step 1

1. Collect the data (fictitious example)

	Lang1	Lang2	Lang3	Lang4	Lang5
Sit1	bla	qu	da	nina	haha
Sit2	bla	qu	da	nana	hihi
Sit3	bla	qa	ta	nina	hehe

# Algorithm for MDS: Step 1

1. Collect the data (fictitious example)

	Lang1	Lang2	Lang3	Lang4	Lang5
Sit1	bla	qu	da	nina	haha
Sit2	bla	qu	da	nana	hihi
Sit3	bla	qa	ta	nina	hehe

Comparative concepts (cf. Haspelmath 2010)

# Algorithm for MDS: Step 2

2. Compute the distances between the situations (rows)

	Lang1	Lang2	Lang3	Lang4	Lang5
Sit1	bla	qu	da	nina	haha
Sit2	bla	qu	da	nana	hihi
Sit3	bla	qa	ta	nina	hehe

Overlap 1,2 =  $3/5 = 0.6$

Overlap 1,3 =  $2/5 = 0.4$

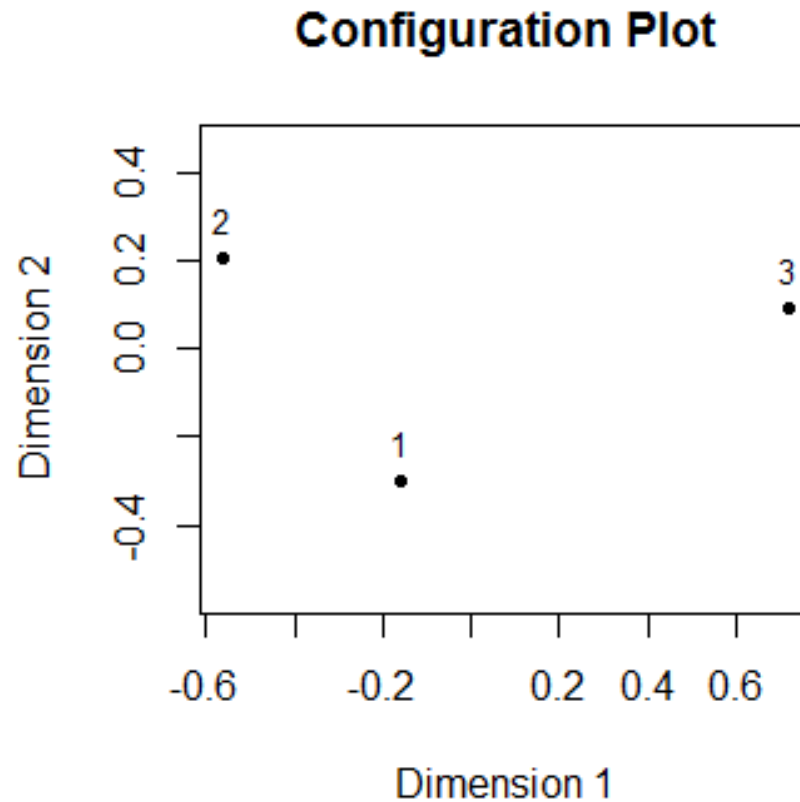
Overlap 2,3 =  $1/5 = 0.2$

Distance =  $1 - \text{overlap}$



# Algorithm for MDS: Step 3

## 3. Perform MDS



# Outline

1. Quantitative typological Cognitive Semantics
2. Data: ParTy corpus
3. Method: Probabilistic semantic maps
4. Case studies
  - causative constructions
  - causal connectives
5. Conclusions

# Causation in CogSem

- Talmy (2000: Ch. 8) makes a number of distinctions:
  - Continuous vs. discontinuous causation chain:
    - *I slid the plate across the table by pushing on it with a stick.*
    - *I made the plate slide across the table by throwing a stick at it.*
  - Extended vs. onset causation
    - *I pushed the box across the ice, going along with it.*
    - *I pushed the box off across the ice, and stayed put.*
  - Effectuating causation vs. enabling causation (i.e. making vs. letting)
  - Presence of self-directedness in mid-causal chain (i.e. the Causee's control)
    - *I threw him downstairs.*
    - *I sent him downstairs.*

# The role of the Causer

- Author causative (non-intended): *I broke the vase in rolling a ball into it.*
- Agent causative (intended): *I broke the vase by rolling a ball into it.*

# The role of the Causer

- Author causative (non-intended): *I broke the vase in rolling a ball into it.*
- Agent causative (intended): *I broke the vase by rolling a ball into it.*

OK, but which distinctions are the most important cross-linguistically in causative constructions?

# Data

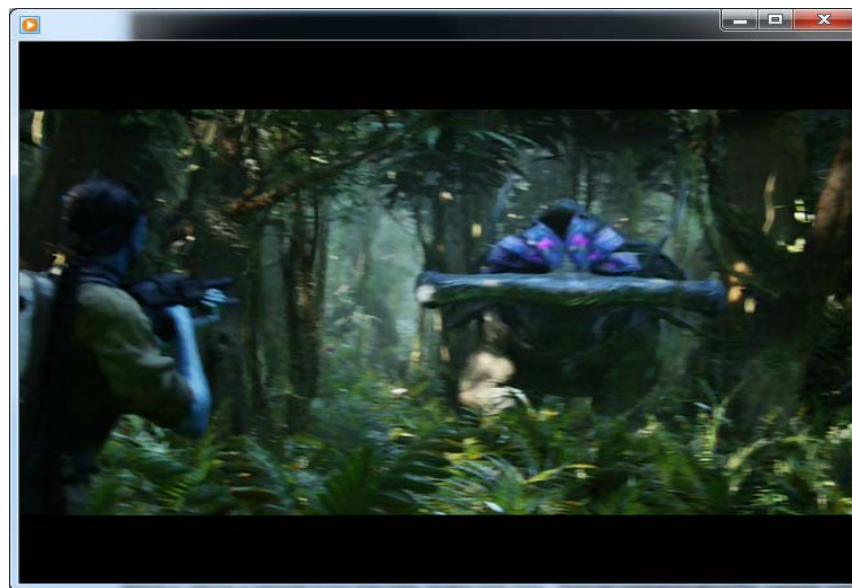
- 344 causative situations from the English subtitles
- Examples:
  - *It is not nice to throw people!*
  - *Now, you made him mad.*
  - *You almost set me on fire!*
  - *You had us worried sick.*
  - *Ah, let me look at you.*
  - *Keep your arms in.*

# Classification of constructions in translations

- Criterion: how the causing and caused events are expressed
  - Analytic causatives (two predicates, e.g. *make X die*)
  - Morphological (the causing event is expressed by a productive causative morpheme, e.g. Turkish *öl-mek* 'to die' - *öl-dür-mek* 'to kill')
  - Lexical (in one word, e.g. *kill, break*)

# Example

- *Don't shoot, you'll piss him off.*
  - French: *Ne tirez pas. Vous allez l'**énerv**er.* (Lexical)
  - Turkish: *Ateş etme. Ateş etme. Onu kız**d**ıracaksın.* (Morphological, from *kızmek* 'become angry').
  - Vietnamese: *Đừng bắn. Cậu sẽ **làm nó nổi** điên đó.* (Analytic)





# Interactive semantic map

<http://www.natalialevshina.com/plots/bubblechart1.html>

# Interpretation

- Dimension 1: (In)directness of causation
  - Left: volitionality, control and animacy of the Causee, letting, longer causation chain, caused action rather than change of state
  - Right: non-volitionality, no control, inanimacy of the Causee, making, shorter causation chain, caused change of state
- Dimension 2: Intentionality of the Causer
  - Top: mostly Causers acting intentionally
  - Bottom: many Causers acting unintentionally (also animate)

# Outline

1. Quantitative typological Cognitive Semantics
2. Data: ParTy corpus
3. Method: Probabilistic semantic maps
4. Case studies
  - causative constructions
  - causal connectives
5. Conclusions

# Variation of causal connectives

- Cause – result – purpose
  - e.g. *because – so, therefore – so that, in order to*
- Degree of subjectivity
  - e.g. Dutch *doordat < omdat < want*
- Register/channel
  - e.g. Fr. *parce que vs. car*
- etc.

e.g. Schiffrin 1987; Sanders & Sweetser 2009; Degand et Fagard 2012;  
Zufferey & Cartoni 2012

# Data

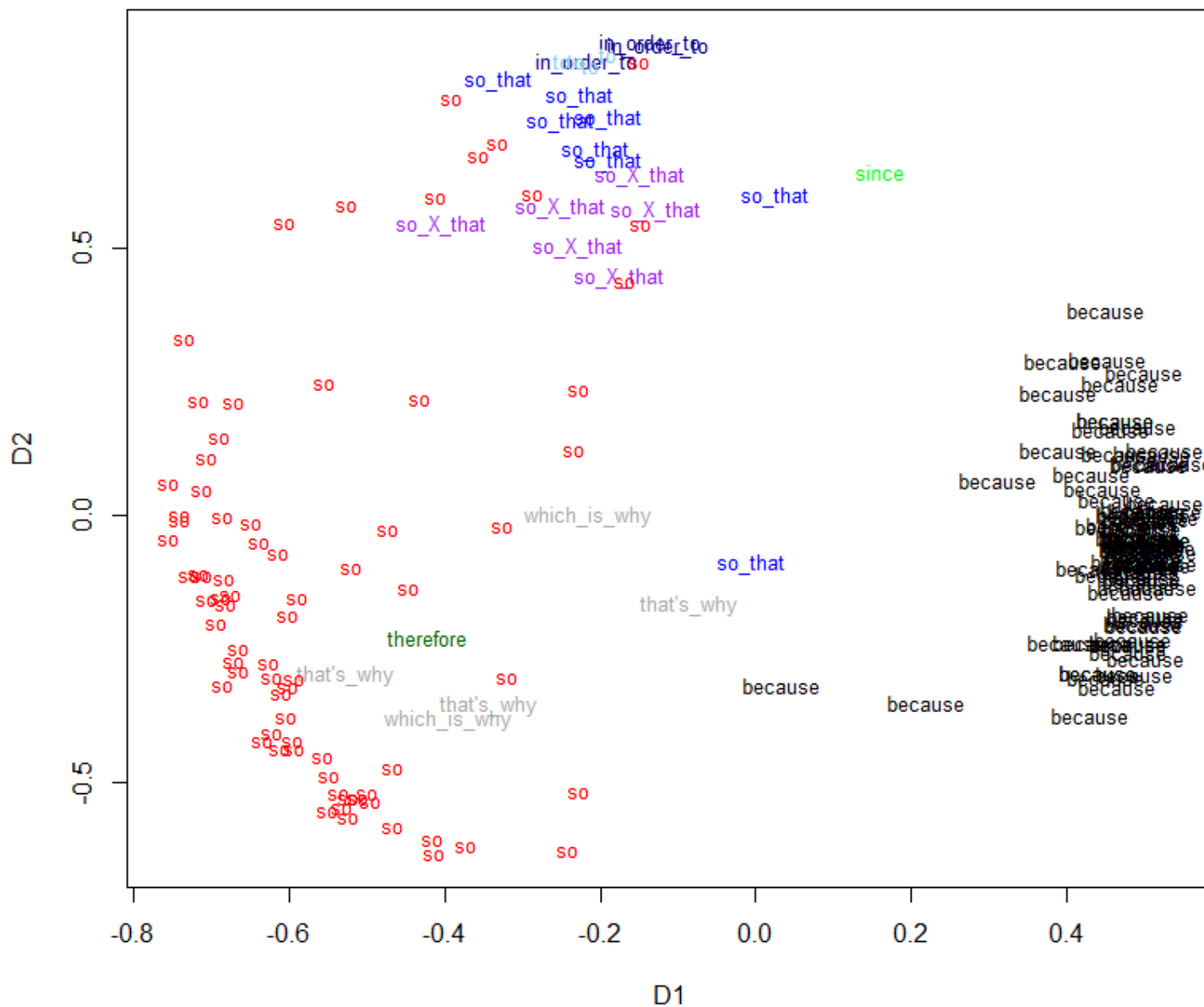
- All causal connectives extracted from the English subtitles (a closed list: *because*, *so*, *therefore*, *since*, etc.)
- In total, 205 instances
- Their correspondences were found in the other languages.

# Example

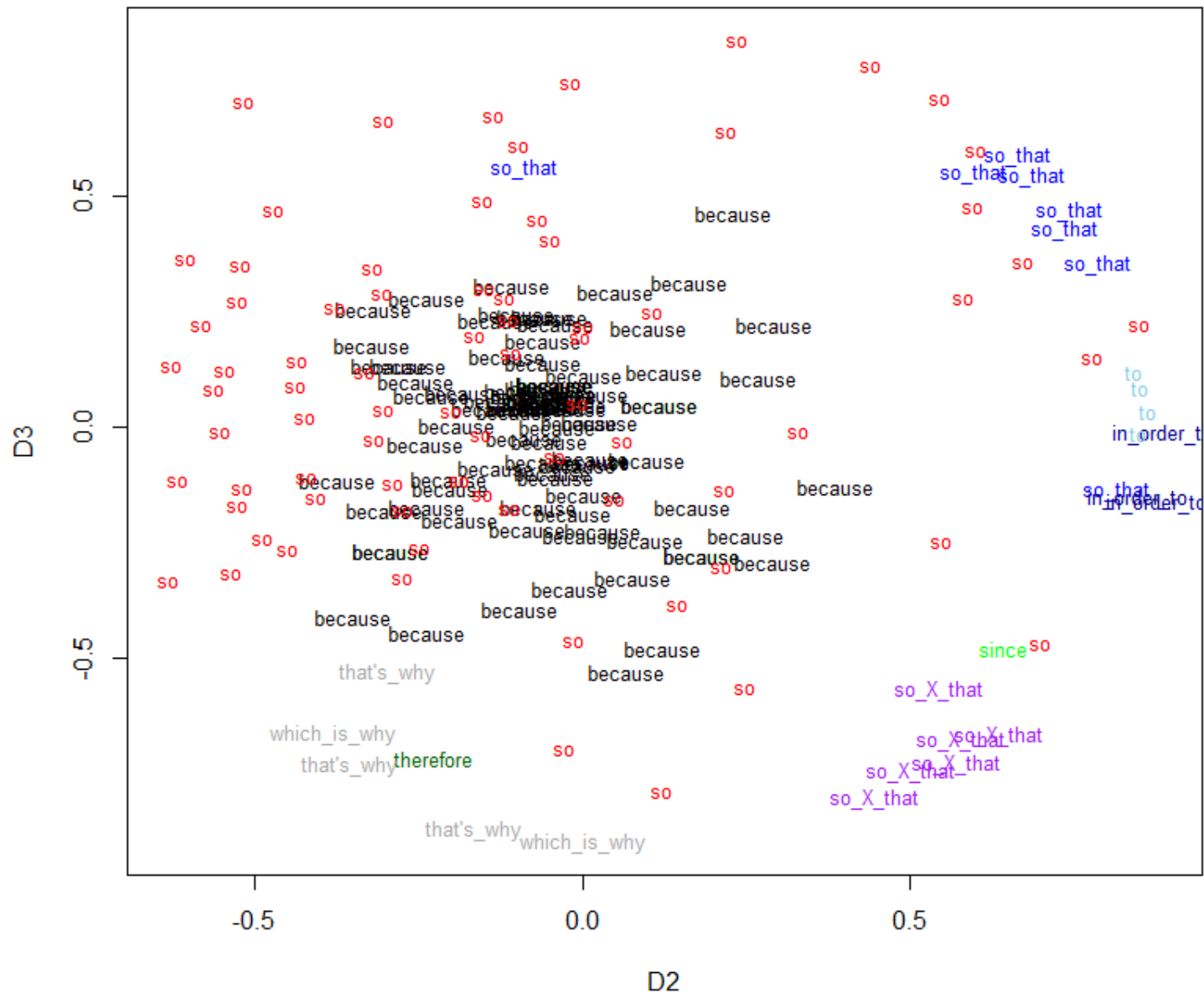
- *I can't stay with her anymore, **because** she doesn't exist.*
  - French: *Je ne peux plus rester, **car** elle n 'existe pas.*
  - Chinese: 我不能再陪着她 **因为**她早已不复存在
  - Finnish: *En voi jäädä Malin luokse, **koska** häntä ei ole olemassa.*
  - Russian: *Я не могу с ней оставаться, **потому что** она не существует.*



## Semantic map of causal connectives



# Semantic map of causal connectives





# Interpretation of semantic map: Dimensions

- Most important: cause vs. result
- Less important: purpose vs. other types
- Even less: emphasis on the causal relationship vs. no emphasis

# Outline

1. Quantitative typological Cognitive Semantics
2. Data: ParTy corpus
3. Method: Probabilistic semantic maps
4. Case studies
  - causative constructions
  - causal connectives
5. Conclusions

# Conclusions

- At both constructional levels, the intentionality/control of the agent(s) represents a key distinction where the languages converge.
  - Language is anthropocentric.
- Hypothesis: we can expect to find this dimension in other constructions across different languages and at different levels of grammatical granularity.
  - Cf. cross-level studies by Degand (2001) and Stukker (2005) on Dutch causal and causative expressions.

Thanks!

The slides are available at

<http://www.natalialevshina.com/presentations.html>

[natalevs@gmail.com](mailto:natalevs@gmail.com)

[natalia.levshina@uclouvain.be](mailto:natalia.levshina@uclouvain.be)