

Semantic maps of causatives : Data types and methods in contrast

Natalia Levshina

Leipzig University

Liège June 27 2018

Introduction

- The temptation to capture such an elusive thing as meaning by representing it as a material object is irresistible.

Introduction

- The temptation to capture such an elusive thing as meaning by representing it as a material object is irresistible.
- This explains, at least partly, the success of semantic maps.

Introduction

- The temptation to capture such an elusive thing as meaning by representing it as a material object is irresistible.
- This explains, at least partly, the success of semantic maps.
- But how do they help us to learn something new about language?

Introduction

- The temptation to capture such an elusive thing as meaning by representing it as a material object is irresistible.
- This explains, at least partly, the success of semantic maps.
- But how do they help us to learn something new about language?
- This talk compares some popular and less well known **statistical** semantic maps based on different data types.

A typology of semantic maps

- What kind of data are used?

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As distances (Multidimensional Scaling or Correspondence Analysis)

Causative constructions

- Formal variation
- Semantic variation

Causative constructions

- Formal variation
- Semantic variation

Formal variation

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

Causative constructions

- Formal variation
- Semantic variation

Control of the Causee

- Does the Causee have control over the caused event?



Control of the Causee

- Does the Causee have control over the caused event?
 - Yes: The teacher **had** the students **read** *War and Peace*.



Control of the Causee

- Does the Causee have control over the caused event?
 - Yes: The teacher **had** the students **read** *War and Peace*.
 - No: The sniper **killed** the terrorist.



Factitive or permissive

- Factitive (making):

That which does not kill us, makes us stronger.

Factitive or permissive

- Factitive (making):

That which does not kill us, makes us stronger.

- Permissive (letting):

Let my people go!

Direct or indirect causation

- Direct:

A Swedish football player **broke** Rudy's nose.

Direct or indirect causation

- Direct:

A Swedish football player **broke** Rudy's nose.

- Indirect:

The politician **had** a rival **poisoned** with Novichok.

Implicative or not

- Are we sure that the caused event happened?

- Implicative:

The secret service **killed** the Kremlin critic (*but he was alive).

- Non-implicative:

She **asked** him **to leave** (but he might have stayed).

Some more types

- Non-intentional:

Oops, I've **broken** your Ming vase!

- Forceful:

You can't **force** anyone **to love** you.

- Assistive:

O God, **help** me **to be pure**, but not now! (St. Augustine)

- Involved/comitative:

Load up your guns and **bring** your friends! (Nirvana)

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As distances (Multidimensional Scaling or Correspondence Analysis)

Database of causatives *TypoCaus*

- Levshina 2013 –
- Over 130 languages analyzed
- R Shiny User Interface
- Here: data from 50 families from all over the world

Database in R Shiny

~/R/MyProjects/CausTyp/CauseR - Shiny

http://127.0.0.1:6516 | Open in Browser

Publish

Database of causative constructions in languages of the world

Search by language

[Semantic maps](#)

Enter the name of a language:

Submit

```
[1] "Basque"
ISO_code Languoid Genus Family Macroarea
eus Basque Basque Basque Eurasia
```

5 causative construction(s) found!

Construction 1 : Lexical

Form: Inchoative/causative alternation

Meaning: NA

Example:

hil 'die/kill'; sartu 'go in, put in'; atera 'go out, take out'; zabaldu 'open'; jantzi 'dress'; galdu 'get lost, lose'

Construction 2 : Morphological

Form: Forms with infix -ra-

Meaning: NA

Example:

erakutsi 'show, make see' < ikusi 'see'; irakatsi 'teach, make learn' < ikasi 'learn'; eragin 'cause to make, affect' from egin 'make'; erabili 'use', from ibili 'walk'; erantzi 'undress' from jantzi 'dress'

Construction 3 : Morphological/Syntactic

Form: Verb/suffix (written separately) (e)raz- added to the participle (Western dialects) or the verbal root (Eastern dialects)

Meaning: More direct causation than eus_01 and eus_02, less direct causation than eus_04

Example:

Berek etzuten nihor hil arazteko bothererik.
they not.AUX anyone die CAUSE.NOM.REL power.PRTT
They did not have anyone die CAUSE.NOM.REL power die

A typology of semantic maps

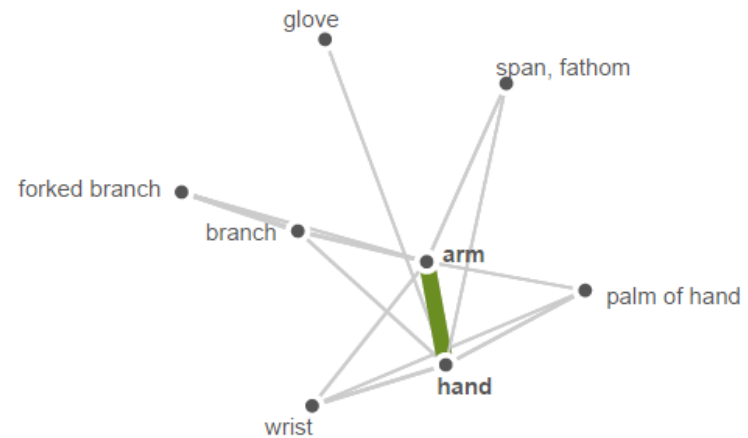
- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As distances (Multidimensional Scaling or Correspondence Analysis)

Networks: Previous work

Line opacity: Line weights: Coloring: Family

85 links for "arm" and "hand":

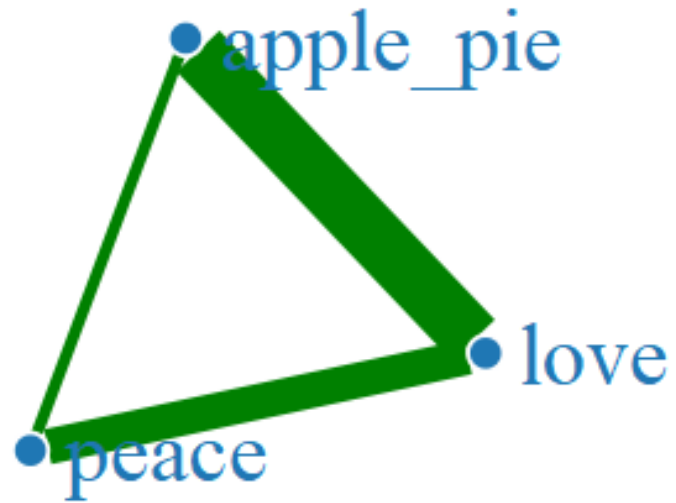
| Language | Family | Form |
|-----------------------|----------------|---------|
| 1. Gawwada | Afro-Asiatic | hargɔ |
| 2. Hausa | Afro-Asiatic | hannu |
| 3. Iraqw | Afro-Asiatic | dawa |
| 4. Polci | Afro-Asiatic | aam |
| 5. Tarifit | Afro-Asiatic | fus |
| 6. Azerbaijani, South | Altaic | əл |
| 7. Kumyk | Altaic | къол |
| 8. Oroqen | Altaic | ŋa:la |
| 9. Gurinji | Australian | wartarn |
| 10. Vietnamese | Austro-Asiatic | tay |



Networks: co-expression data

| SENSE 1 | SENSE 2 | Frequency of co-expression by one form |
|-----------|-----------|--|
| LOVE | PEACE | 3 |
| LOVE | APPLE PIE | 5 |
| APPLE PIE | PEACE | 1 |

Networks: visualization



Networks of causative senses

~/R/MyProjects/CausTyp/CauseR - Shiny

http://127.0.0.1:6516 Open in Browser Publish

Database of causative constructions in languages of the world

Search by language Semantic maps

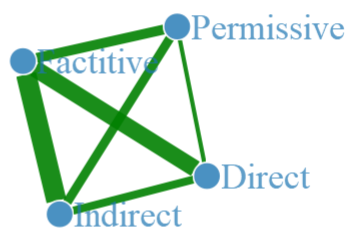
Type of semantic map

- Network of functions
- Distances between functions
- Distances between constructions and between functions

Semantic functions to include:

- Direct
- Indirect
- Control_Causee
- No_Control_Causee
- Forceful
- Not_Forceful
- Factive
- Permissive
- Implicative
- Not_Implicative
- Comitative
- Not_Comitative
- Intentional
- Not_Intentional
- Assistive

Submit



```
graph TD; Factive --- Permissive; Factive --- Direct; Factive --- Indirect; Permissive --- Direct; Permissive --- Indirect; Direct --- Indirect;
```

Networks of functions: evaluation

Advantages

- One can investigate the relationships between individual semantic functions
- No loss of information

Networks of functions: evaluation

Advantages

- One can investigate the relationships between individual semantic functions
- No loss of information

Disadvantages

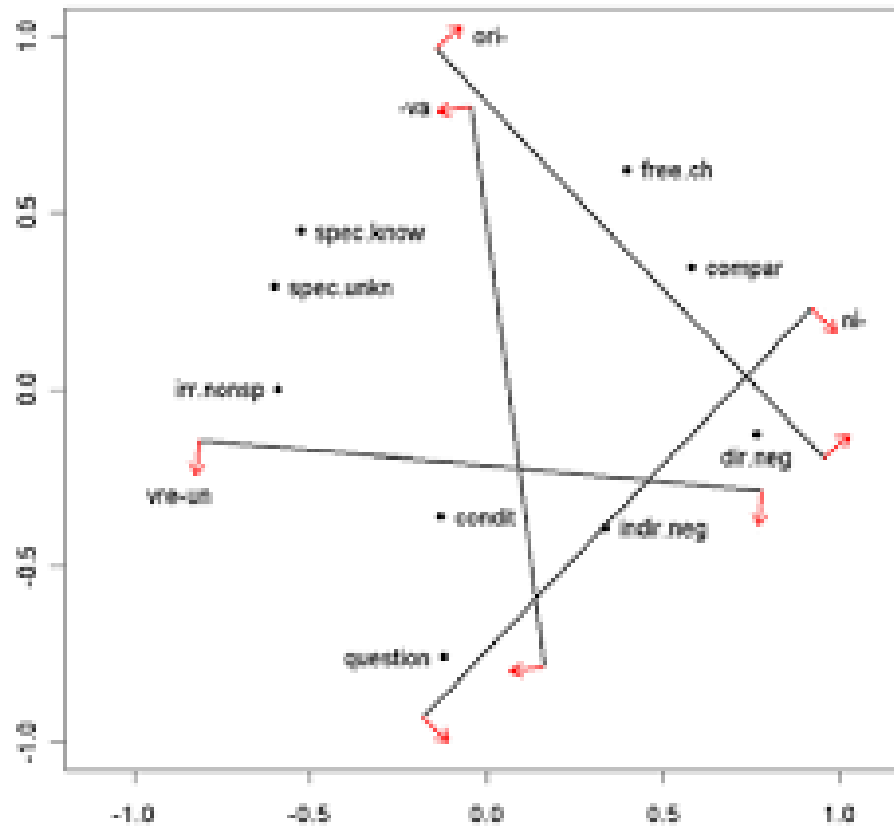
- Very confusing when the number of nodes is large
- No common dimensions of variation

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As distances (Multidimensional Scaling or Correspondence Analysis)

Previous work

- Croft & Poole 2008



Multidimensional Scaling: computing the distances

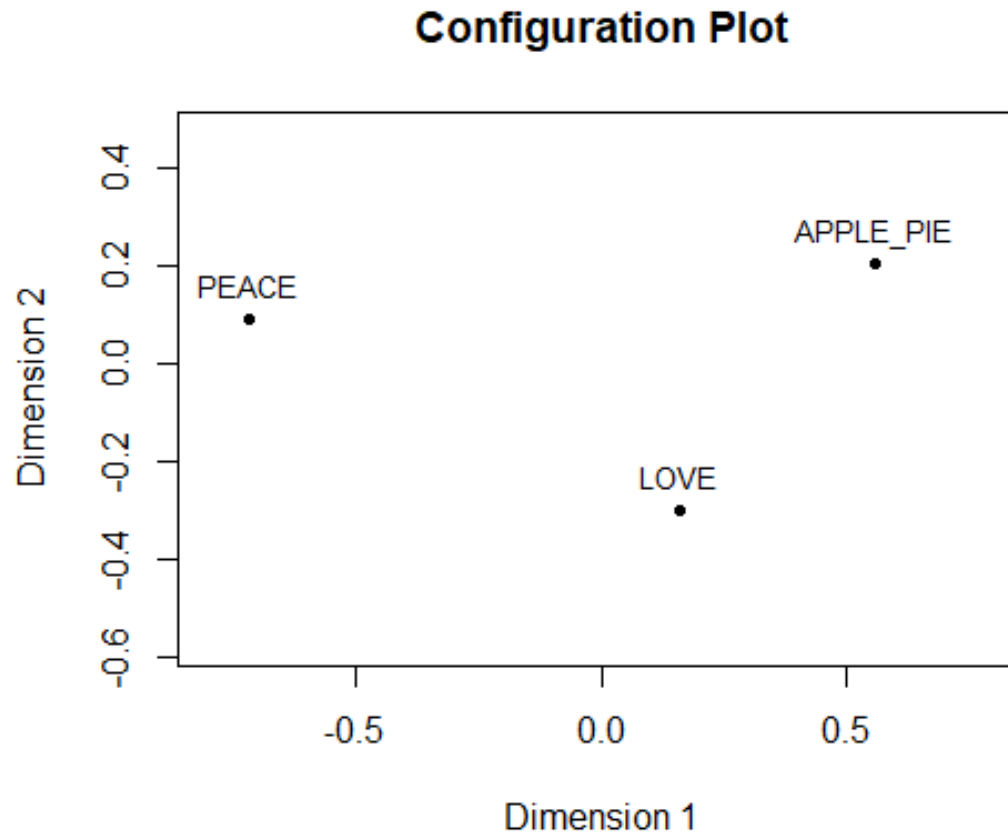
| Sense | Form 1 | Form 2 | Form 3 | Form 4 | Form 5 | Form 6 | Form 7 | Form 8 | Form 9 |
|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| LOVE | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| APPLE PIE | Yes | Yes | Yes | No | Yes | No | No | Yes | Yes |
| PEACE | No | Yes | No | Yes | No | Yes | Yes | No | No |

Distance between LOVE and APPLE PIE: $1 - (5/9) = 0.44$

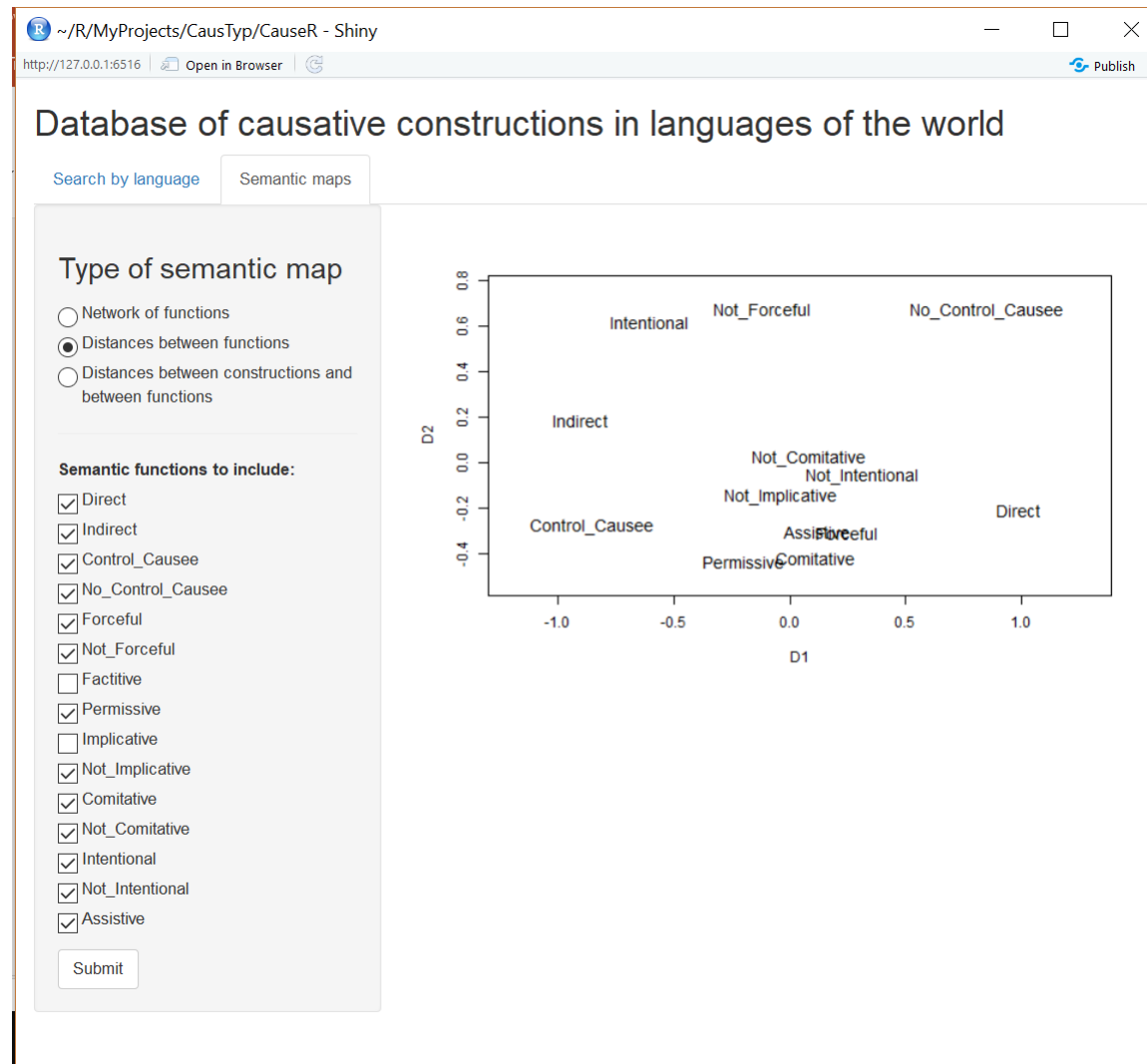
Distance between LOVE and PEACE: $1 - (3/9) = 0.67$

Distance between APPLE PIE and PEACE: $1 - (1/9) = 0.89$

Multidimensional Scaling: visualization of distances



MDS of causative senses



Type-based MDS maps: evaluation

Advantages

- Help to identify dimensions of semantic variation

Type-based MDS maps: evaluation

Advantages

- Help to identify dimensions of semantic variation

Disadvantages

- More difficult to evaluate pairwise relationships
- Loss of information

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - Parallel corpora
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As distances (Multidimensional Scaling or Correspondence Analysis)

The main idea behind CA

- CA is based on comparison of row profiles and column profiles, e.g.

| | M1 | M2 | M3 | Total |
|-------|----|-----|----|-------|
| Cx1 | 20 | 30 | 50 | 100 |
| Cx2 | 10 | 70 | 20 | 100 |
| Total | 30 | 100 | 70 | 200 |

row
profiles
→

| | M1 | M2 | M3 | Total |
|-----|-----|-----|-----|-------|
| Cx1 | 0.2 | 0.3 | 0.5 | 1 |
| Cx2 | 0.1 | 0.7 | 0.2 | 1 |

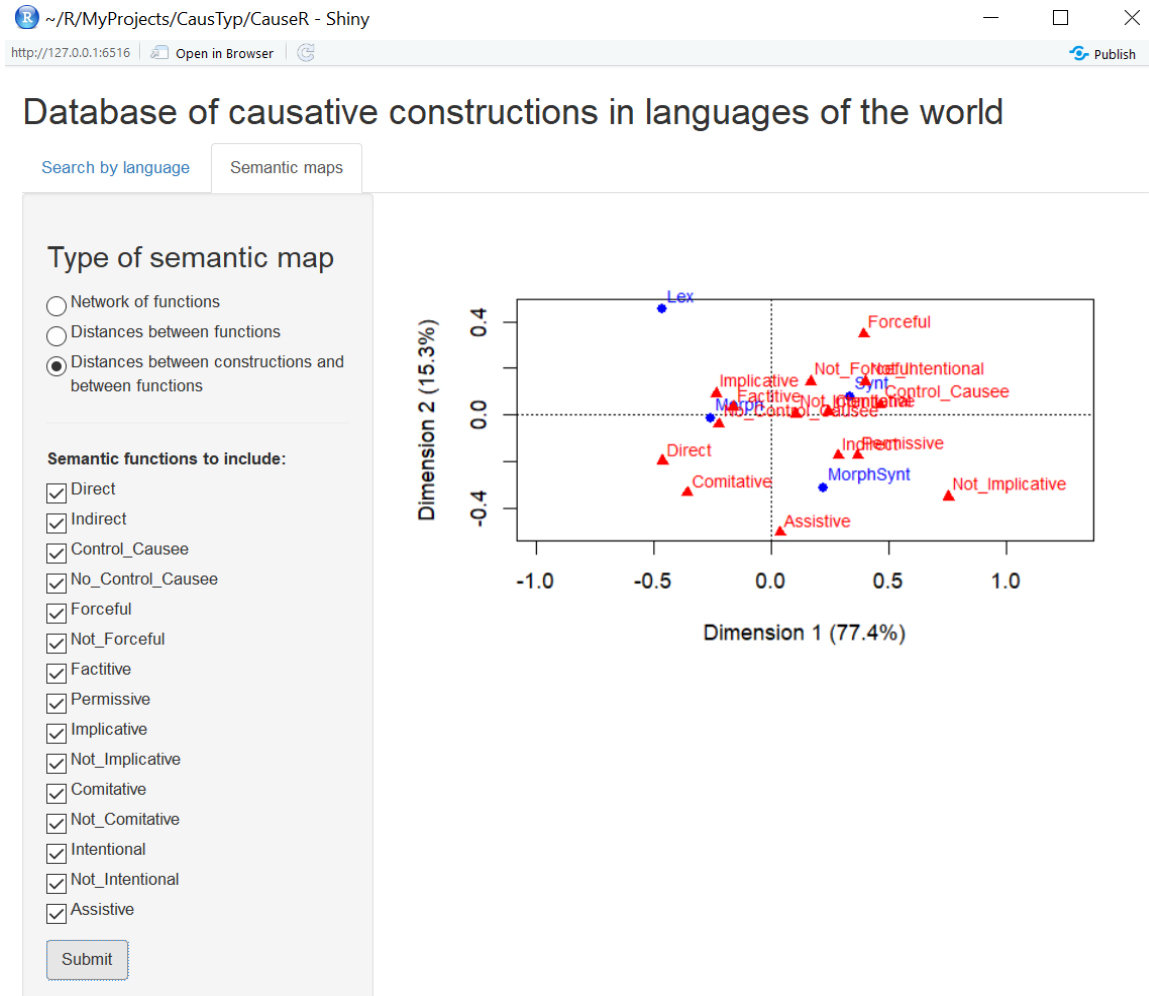
column
profiles
↓

| | M1 | M2 | M3 |
|-------|------|-----|------|
| Cx1 | 0.67 | 0.3 | 0.71 |
| Cx2 | 0.33 | 0.7 | 0.29 |
| Total | 1 | 1 | 1 |

The main idea behind CA

- If two row or column profiles are similar, their labels will be closely located in a semantic map.
- If two row or column profiles are dissimilar, their labels will be located far from each other.

CA of causative formal types and senses



CA maps: evaluation

Advantages

- Easy to investigate form-meaning mapping
- One can explore the semantic dimensions

CA maps: evaluation

Advantages

- Easy to investigate form-meaning mapping
- One can explore the semantic dimensions

Disadvantages

- One cannot interpret the distances between forms and functions directly.
- Loss of information
- The distances are non-Euclidean (chi-squared)
- Outliers are dangerous!

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - **Parallel corpora**
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As distances (Multidimensional Scaling or Correspondence Analysis)

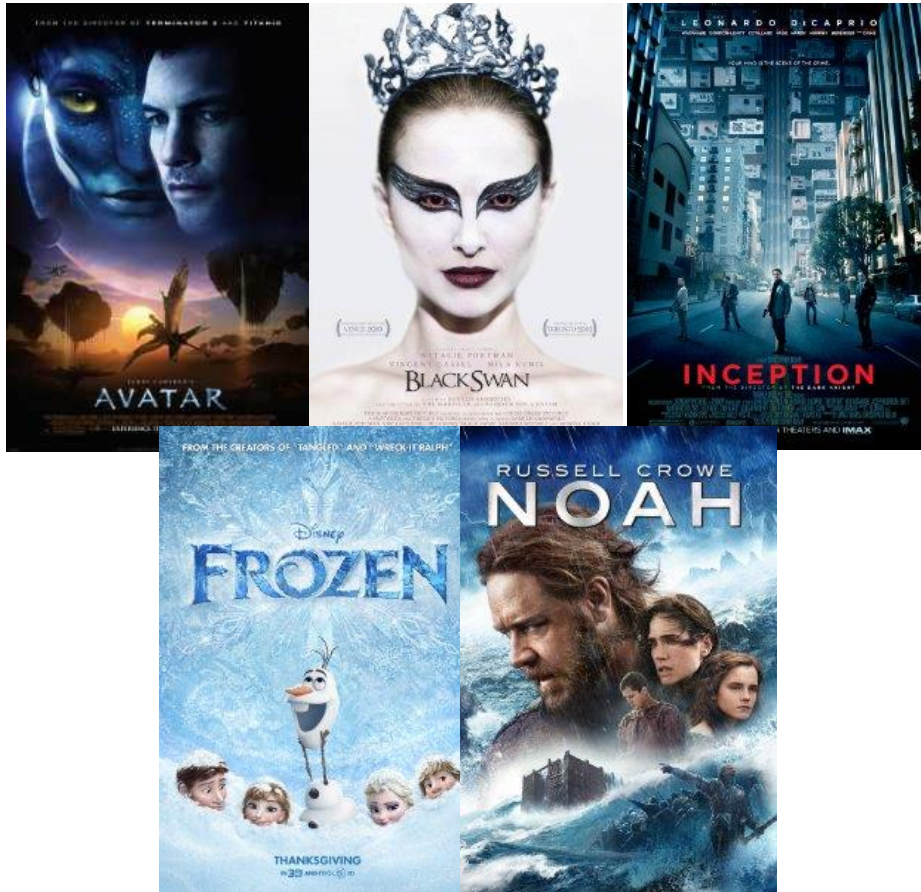
Languages

| 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 |

Subtitles 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*
- Simon Sinek: *Why good leaders make you feel safe*

Data set

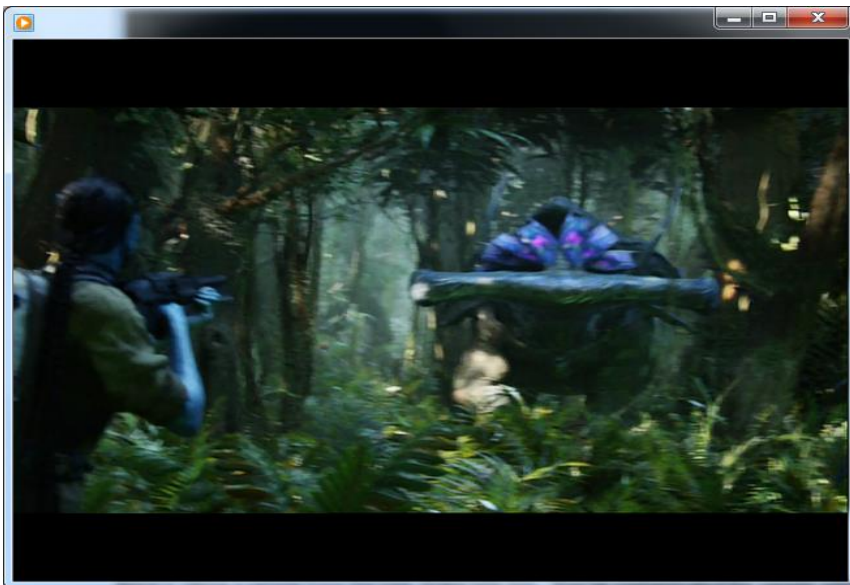
- 344 causative situations found in the English segment of the ParTy corpus*
- Translations in the 10 languages are found and coded into 3 types of constructions (Syntactic, Morphological or Lexical)

*<http://www.natalialevshina.com/corpus.html>

Example from *Avatar*

Original

- ENG: *Don't shoot, you'll piss him off.*



Translations

- FRA: *Ne tirez pas. Vous allez l'énervé.* (Lexical)
- TUR: *Ateş etme. Ateş etme. Onu kızdıracaksın.* (Morphological, from *kızmek* 'become angry').
- VIE: *Đừng bắn. Cậu sẽ làm nó nổi điên đó.* (Syntactic)

Examples of constructions

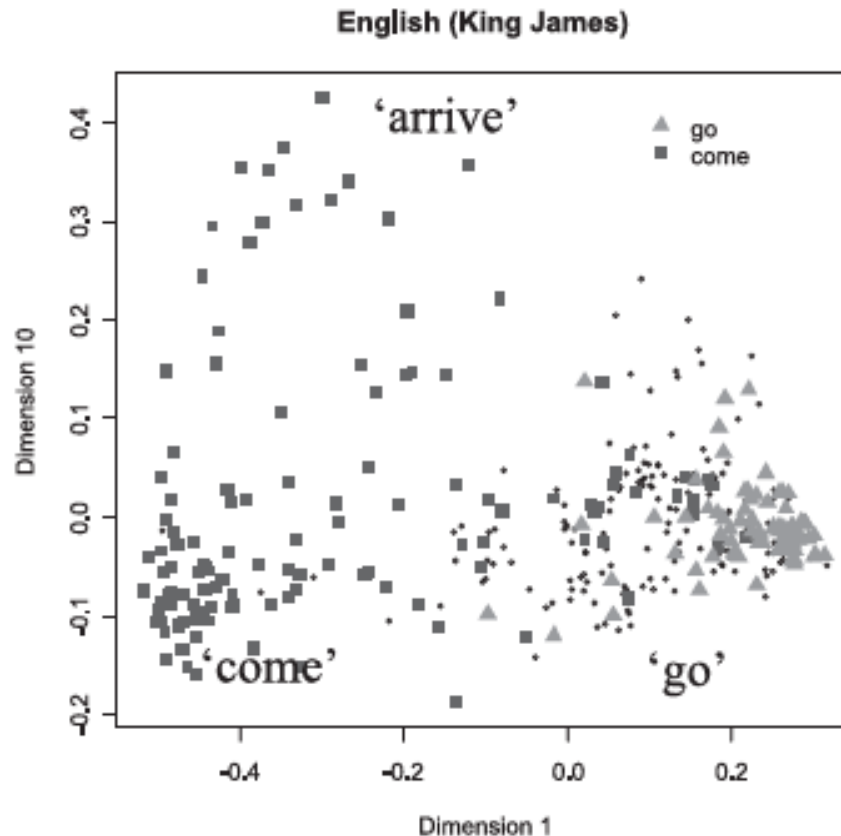
| | Lexical | Morphological | Syntactic |
|------------|---------------------------|---------------------------------|--|
| Chinese | shā sǐ “kill” | - | ràng “let, make” + Pred |
| Finnish | tappaa “kill” | odotu-tt-aa “make wait” | antaa “give” + V1 |
| French | tuer “kill” | - | faire + Vinf |
| Hebrew | harag “kill” <i>pa'al</i> | hotsi “take out” <i>hiph'il</i> | natan “give” + le-Vinf |
| Indonesian | mem-bunuh “kill” | meng-ingat-kan “remind” | membuat “make” + Pred |
| Japanese | korosu “kill” | ikar-ase-ru “make angry” | V _{te} + morau “get” |
| Russian | ubit' “kill” | - | zastavit' + Vinf |
| Thai | kaa “kill” | - | tham hai “do give” + Pred |
| Turkish | açmak “open” | öl-dür- “kill” | V _{mA_DAT} + izin ver- “allow” |
| Vietnamese | giết hại “kill” | - | làm “do” + Pred |

A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - **Parallel corpora**
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - **Semantic situations (exemplars, tokens)**
 - Linguistic forms
- How are the relationships between the objects represented?
 - As links in a network
 - As **distances** (**Multidimensional Scaling** or Correspondence Analysis)

Previous work

- Wälchli & Cysouw (2012): verbs of motion in New Testament



Token-based MDS maps

1. Collect the data (fictitious example)

| | Lang1 | Lang2 | Lang3 | Lang4 | Lang5 |
|-------------|-------|-------|-------|-------|-------|
| Situation 1 | Lex | Morph | Synt | Morph | Lex |
| Situation 2 | Lex | Morph | Synt | Synt | Morph |
| Situation 3 | Morph | Morph | Lex | Morph | Synt |

Token-based MDS maps

2. Compute the distances between the situations (rows)

| | Lang1 | Lang2 | Lang3 | Lang4 | Lang5 |
|-------------|-------|-------|-------|-------|-------|
| Situation 1 | Lex | Morph | Synt | Morph | Lex |
| Situation 2 | Lex | Morph | Synt | Synt | Morph |
| Situation 3 | Morph | Morph | Lex | Morph | Synt |

Overlap 1,2 = 3/5 = 0.6

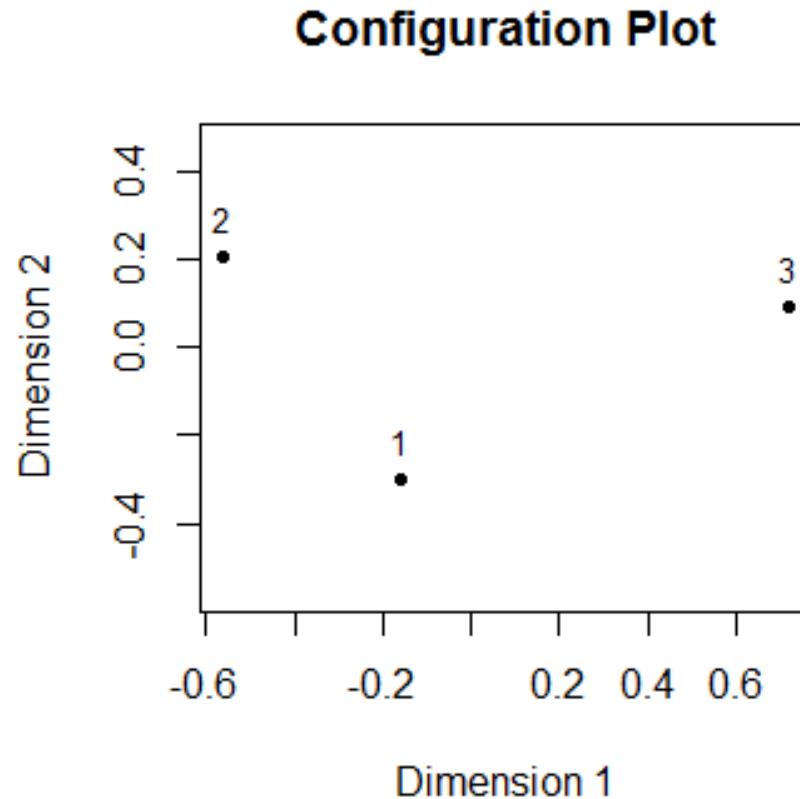
Overlap 1,3 = 2/5 = 0.4

Overlap 2,3 = 1/5 = 0.2

Distance = 1 – overlap

Token-based MDS maps

3. Perform MDS (package smacof)



Interpretation of MDS distances

- The closer two points (i.e. causative situations), the more frequently they are expressed by the same constructions across the languages.

Interactive MDS maps with googleVis

- Exemplars:
 - <http://www.natalialevshina.com/plots/bubblechart1.html>
- Control of the Causee:
 - <http://www.natalialevshina.com/plots/bubblechart2.html>
- Intentionally acting Causer:
 - <http://www.natalialevshina.com/plots/bubblechart3.html>
- Mapping of the constructions: FRA, RUS, FIN, TUR

Token-based MDS maps: evaluation

Advantages

- No need for semantic coding
- Dimensions of semantic variation
- Information about the relative frequencies of meanings

Token-based MDS maps: evaluation

Advantages

- No need for semantic coding
- Dimensions of semantic variation
- Information about the relative frequencies of meanings

Disadvantages

- Often difficult to interpret linguistically
- Loss of information

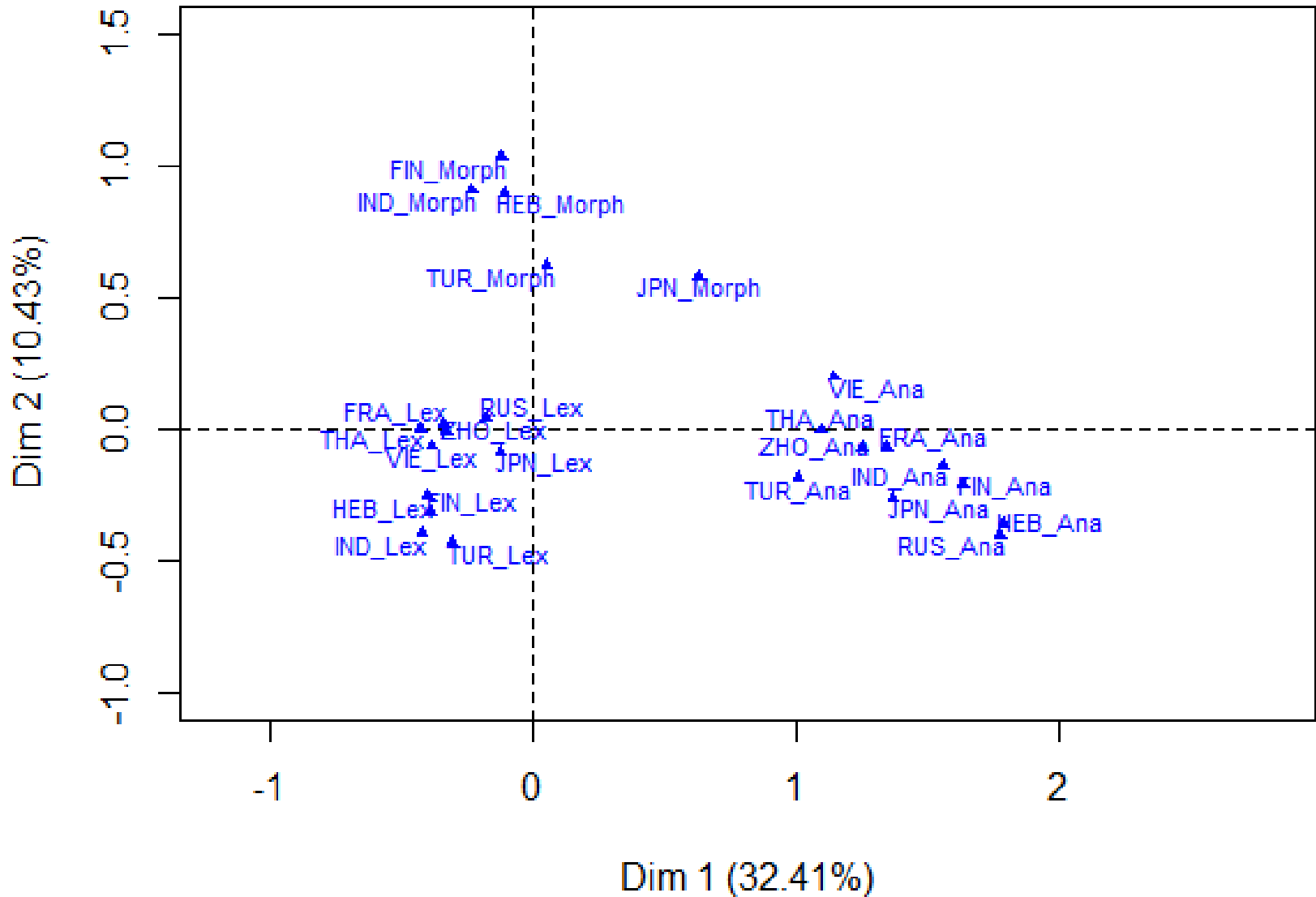
A typology of semantic maps

- What kind of data are used?
 - Grammars or dictionaries
 - **Parallel corpora**
- What kind of objects are shown?
 - Semantic functions (senses, meanings, etc.)
 - Semantic situations (exemplars, tokens)
 - **Linguistic forms**
- How are the relationships between the objects represented?
 - As links in a network
 - As **distances** (Multidimensional Scaling or **Correspondence Analysis**)

Multiple Correspondence Analysis

- Multiple Correspondence Analysis shows how different values of more than two categorical variables are associated.
 - e.g. if Finnish morphological causatives tend to be used in the same contexts as French analytic causatives, they will be located in the same region of the map.
- Package FactoMineR in R

MCA factor map



MCA maps of forms: evaluation

Advantages

- Straightforward cross-linguistic comparison of constructional types

MCA maps of forms: evaluation

Advantages

- Straightforward cross-linguistic comparison of constructional types

Disadvantages

- Loss of information
- Outliers are dangerous!
- Only the average position (no exemplar information)
- What are the underlying semantic features?

(Note: This can be fixed with additional coding and supplementary points, see Levshina 2016)

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography
- We have to rely on the information provided by the author and the few examples

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography
- We have to rely on the information provided by the author and the few examples
- Most frequent types (lexical causatives) are underrepresented

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography
- We have to rely on the information provided by the author and the few examples
- Most frequent types (lexical causatives) are underrepresented

Parallel corpora

- More contextual information (e.g. films)

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography
- We have to rely on the information provided by the author and the few examples
- Most frequent types (lexical causatives) are underrepresented

Parallel corpora

- More contextual information (e.g. films)
- More realistic picture of language use

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography
- We have to rely on the information provided by the author and the few examples
- Most frequent types (lexical causatives) are underrepresented

Parallel corpora

- More contextual information (e.g. films)
- More realistic picture of language use
- Translationese

Grammars vs. parallel corpora

Grammars

- More data for different languages are available, so one can control for genealogy and geography
- We have to rely on the information provided by the author and the few examples
- Most frequent types (lexical causatives) are underrepresented

Parallel corpora

- More contextual information (e.g. films)
- More realistic picture of language use
- Translationese
- Fewer languages available (exception: NT)

Some considerations

- Statistical semantic maps are exploratory methods for generating theoretically interesting hypotheses, not the end goal.

Some considerations

- Statistical semantic maps are exploratory methods for generating theoretically interesting hypotheses, not the end goal.
- If one formulates a cross-linguistic generalization on the basis of a semantic map, one also needs confirmatory methods, which can control for the genealogical and geographical relationships (e.g. mixed-effects models).

Final message

- Semantic maps are almost as diverse as Belgian beers.

Final message

- Semantic maps are almost as diverse as Belgian beers.
- Choose wisely, enjoy responsibly!



<http://www.belgianbeerme.com/why-belgium/>

- The database and the app will very soon be available at

<https://github.com/levshina/TypoCaus>

For questions and suggestions:

natalevs@gmail.com