

Mapping the causative continuum: A corpus-based approach to constructional typology

Natalia Levshina

F.R.S.-FNRS,

Université catholique de Louvain

Outline

1. Theoretical background
 - the causative continuum
 - causatives as comparative concepts
2. Data: the ParTy corpus
3. Mapping the continuum:
 - Token-based maps (MDS)
 - Type-based maps (MCA)
4. Conclusions

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill, break*

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill, break*

e.g. Turkish *öldür*- “kill”
from *öl*- “die”

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill, break*

e.g. Turkish *öldür*- “kill”
from *öl*- “die”

e.g. *cause X to die,*
make X disappear

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill, break*

e.g. Turkish *öldür*- “kill”
from *öl*- “die”

e.g. *cause X to die,*
make X disappear



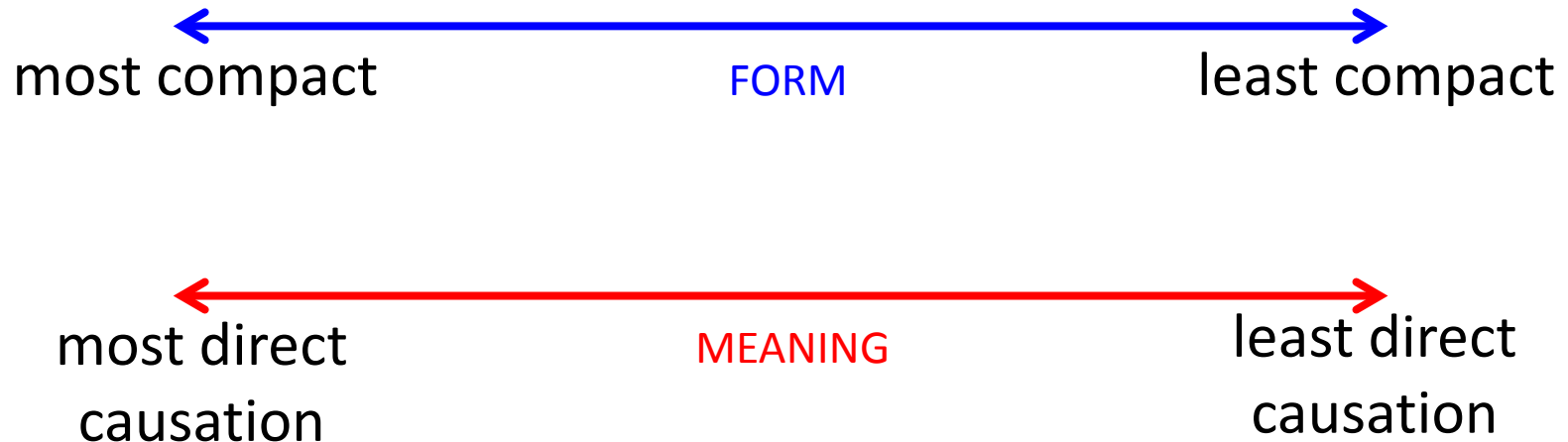
The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

e.g. *kill, break*

e.g. Turkish *öldür*- “kill”
from *öl*- “die”

e.g. *cause X to die,*
make X disappear



Comrie 1981, Dixon 2000, Shibatani & Pardeshi 2002, etc.

Research questions

- Do languages ‘carve up’ the semantic continuum in a similar way or are there large differences?
 - In other words, how similar are analytic, morphological and lexical causatives between themselves?
- Is directness of causation the only semantic dimension?

Comparative concepts

- Haspelmath (2010):

“Comparative concepts are concepts created by comparative linguists for the specific purpose of cross-linguistic comparison. Unlike descriptive categories, they are not part of particular language systems and are not needed by descriptive linguists or by speakers. They are not psychologically real, and they cannot be right or wrong. They can only be more or less well-suited to the task of permitting cross-linguistic comparison.”

Criteria

- Relevant for comparison of form-meaning mappings
- Easy to apply cross-linguistically
- Do not contain problematic concepts (e.g. 'monoclausal' or 'biclausal', cf. Kulikov 2001)
- Do not require tests (corpus-based study)

Lexical causatives as Comparative Concepts

- Such causative constructions where the causing and caused events/states overlap in at least one meaningful unit.
 - Prototype: *break, kill, give*.
 - Less typical: phrasal verbs (e.g. *break off, give away*)

Morphological causatives as Comparative Concepts

- causative constructions with a separate productive element that expresses causation and which cannot be used autonomously.
 - Prototype: Turkish *öldür-* “kill” < *öl-* “die”
 - Less typical: Swahili *chem-k-a* “boil_{INTR}” but *chem-sh-a* “boil_{TR}” (Comrie 1981: 161).

Analytic causatives as Comparative Concepts

- causative constructions that consist of separate predicates. One of them expresses the causing event and is autonomous, and the other expresses the caused event/state.
 - Prototype: *faire* + V, *make* + X + V
 - Less typical: Turkish *izir vermek* ‘give permission’ + V, some serial verb constructions, e.g. Mandarin *huàn-xǐng* “wake up, i.e. call-become awake”

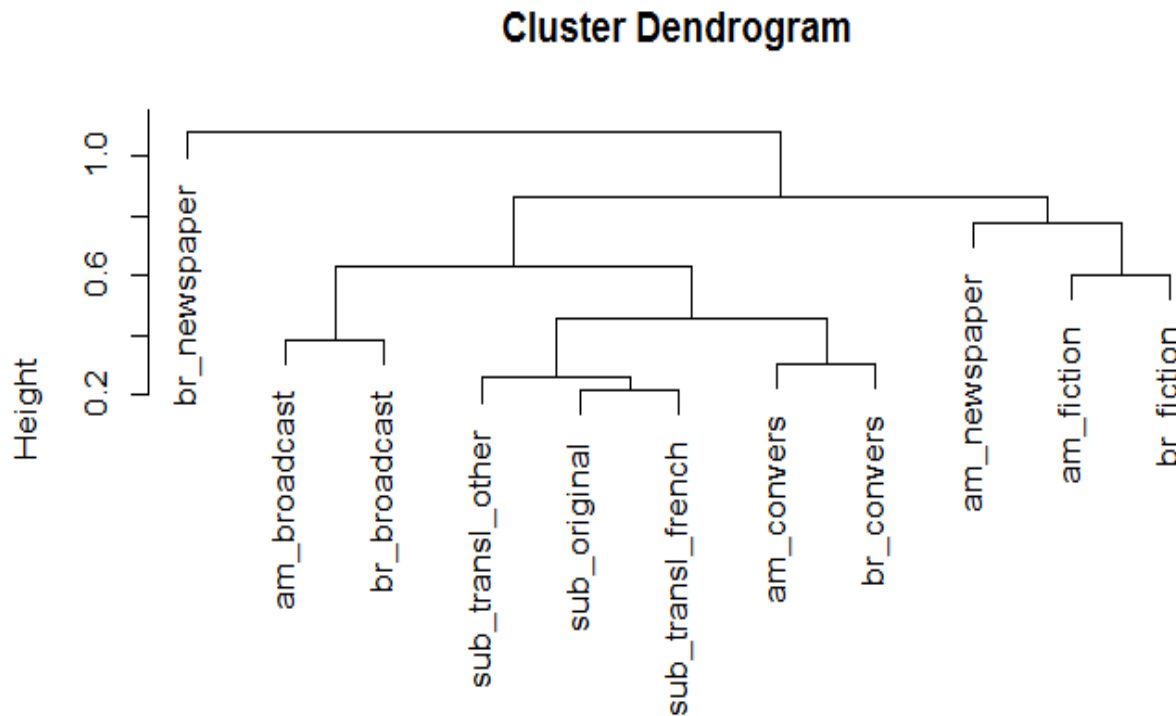
Outline

1. Theoretical background
 - the causative continuum
 - causatives as comparative concepts
2. Data: the ParTy corpus
3. Mapping the continuum:
 - Token-based maps (MDS)
 - Type-based maps (MCA)
4. 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

Why subtitles?



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



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*

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

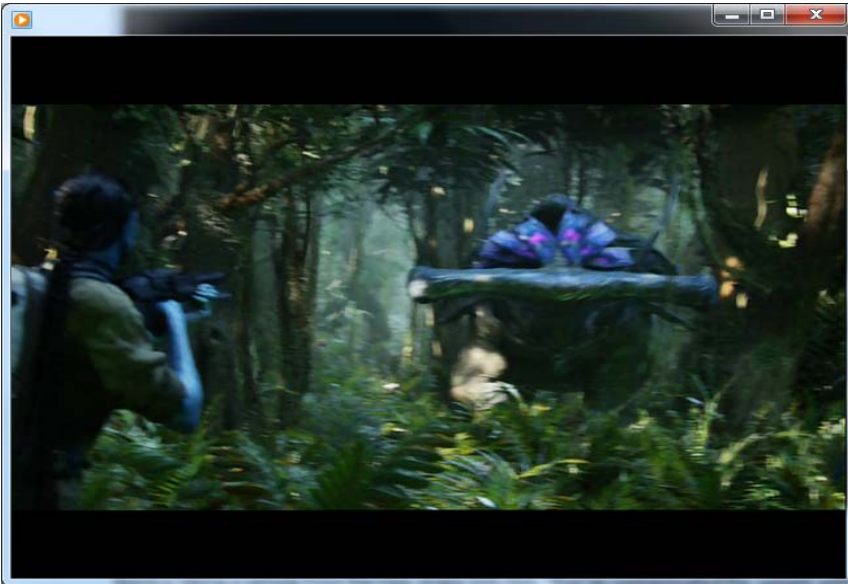
Data set

- 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

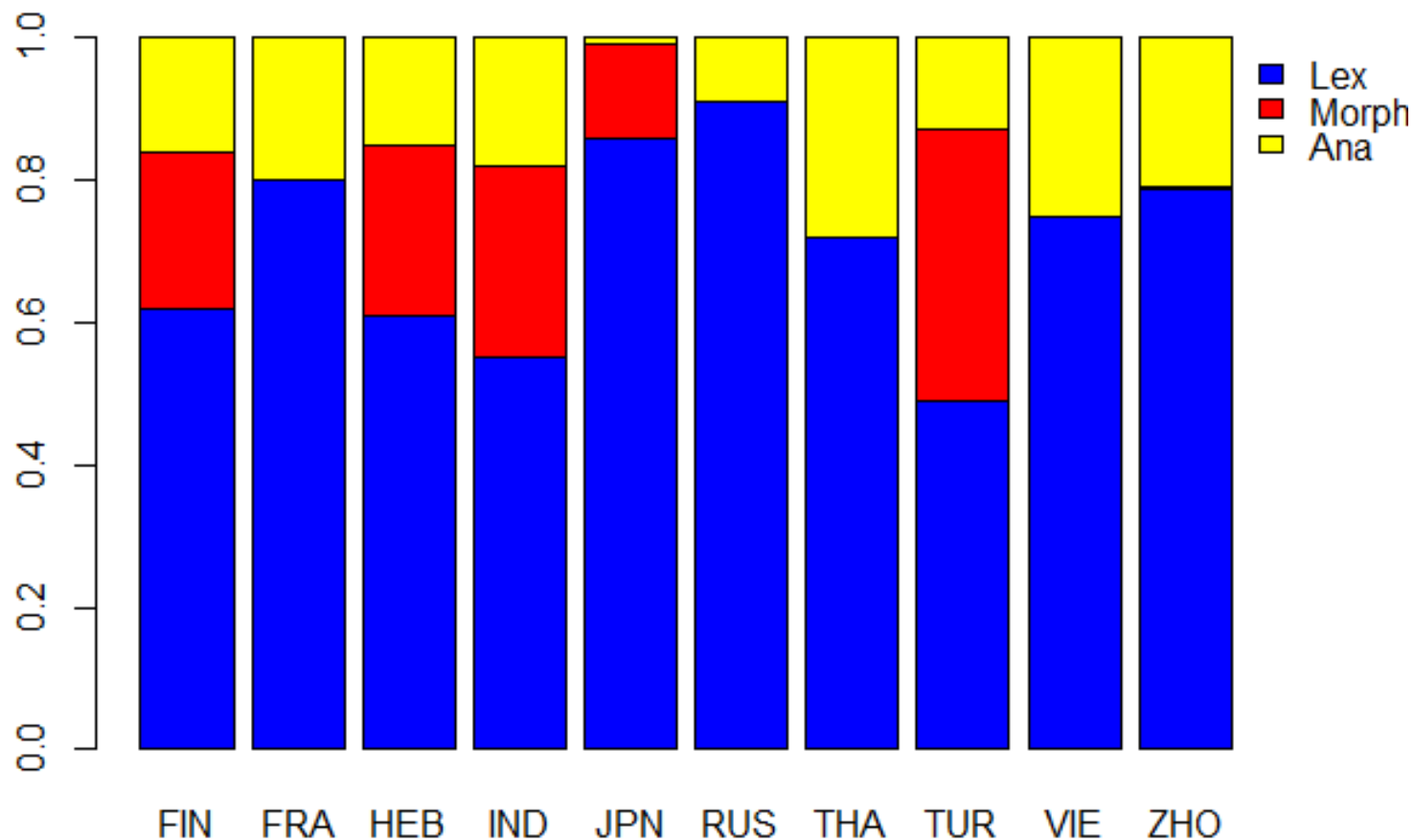
- ENG: *Don't shoot, you'll piss him off.*



Translations

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

Proportions of types of causative constructions in 10 languages



Approach

- The approach is comparable with the one employed by Majid et al. in experimental studies of cutting & breaking events, olfactory categorization, etc.
 - Experimental stimuli → English sentences/situations
 - Subjects' responses → translators' choices
- Advantages:
 - contextualized
 - non-physical, abstract semantics
- Disadvantages:
 - translationese (?)

Outline

1. Theoretical background
 - the causative continuum
 - causatives as comparative concepts
2. Data: the ParTy corpus
3. Mapping the continuum:
 - Token-based maps (MDS)
 - Type-based maps (MCA)
4. Conclusions

Causation maps

- Method 1: Multidimensional Scaling based on tokens (e.g. Wälchli & Cysouw 2012)
 - causative constructions in each language as a cloud of points = constructional instances
- Method 2: Multiple Correspondence Analysis based on types
 - Causative constructions as points in space

Outline

1. Theoretical background
 - the causative continuum
 - causatives as comparative concepts
2. Data: the ParTy corpus
3. Mapping the continuum:
 - Token-based maps (MDS)
 - Type-based maps (MCA)
4. Conclusions

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 at the micro-level!

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 = $3/5 = 0.6$

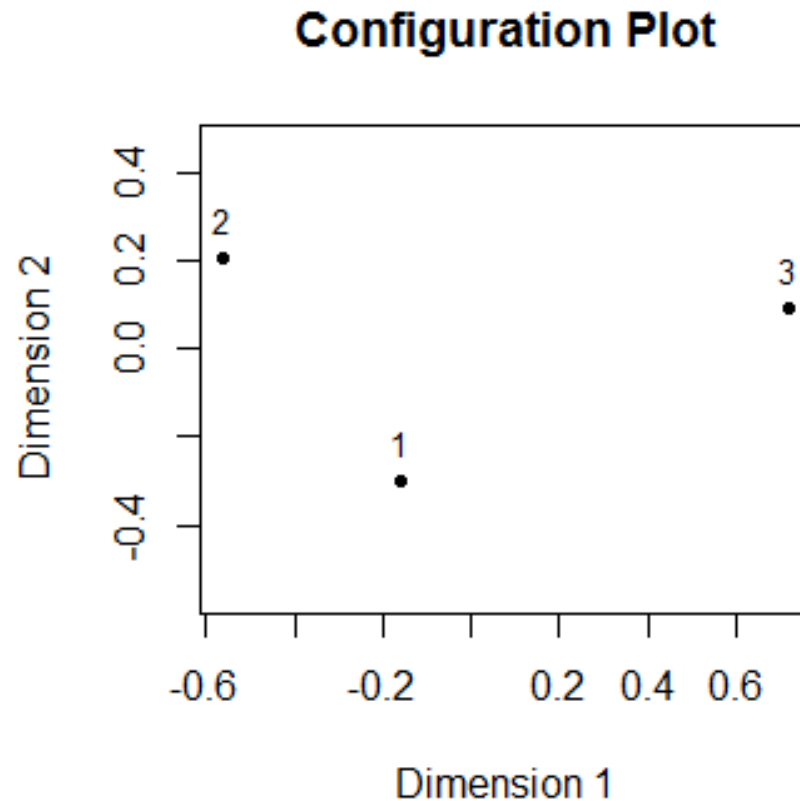
Overlap 1,3 = $2/5 = 0.4$

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

Distance = $1 - \text{overlap}$

Algorithm for MDS: Step 3

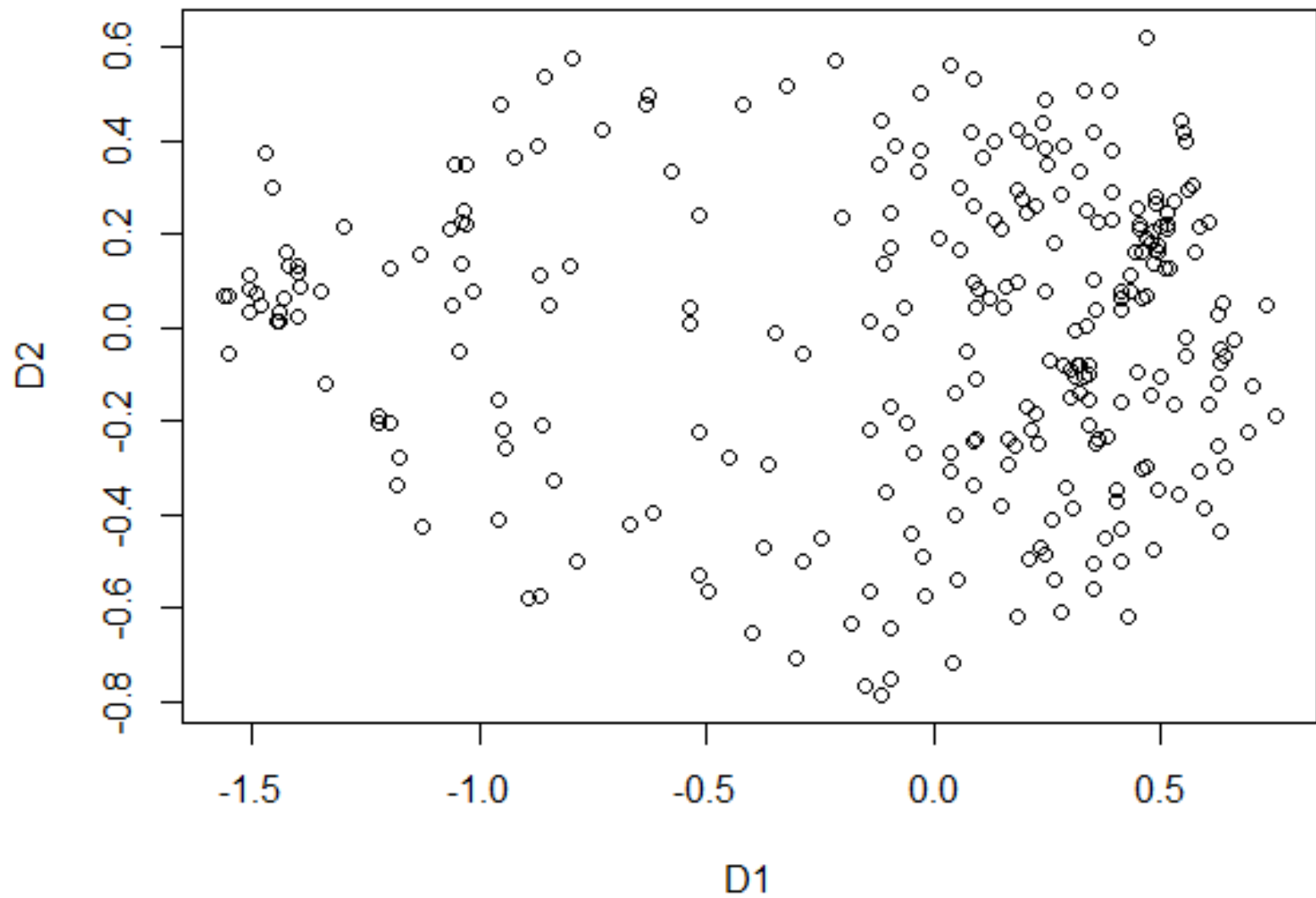
3. Perform MDS (package smacof)



Interpretation of MDS distances

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

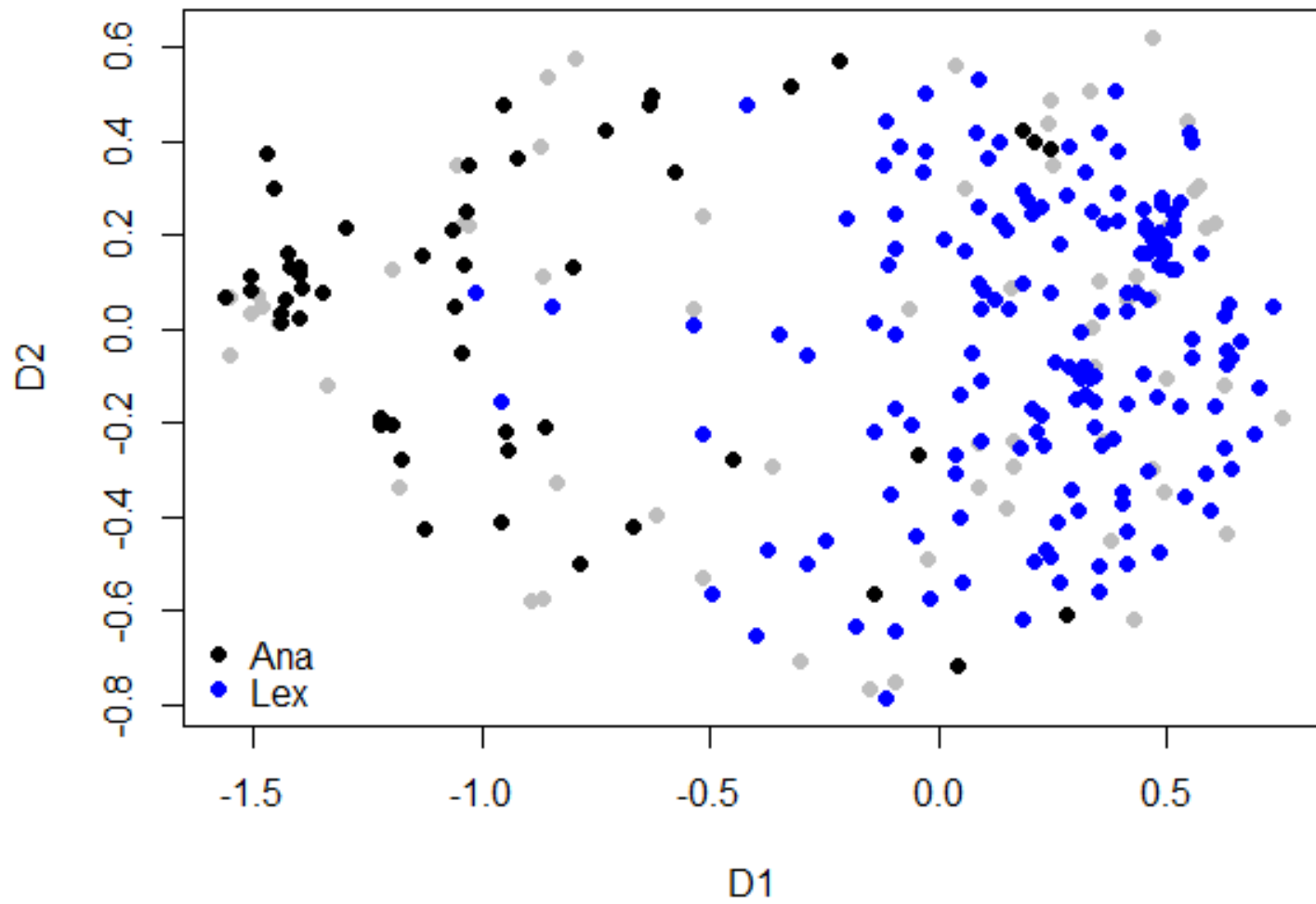
Token-based semantic map of causative constructions



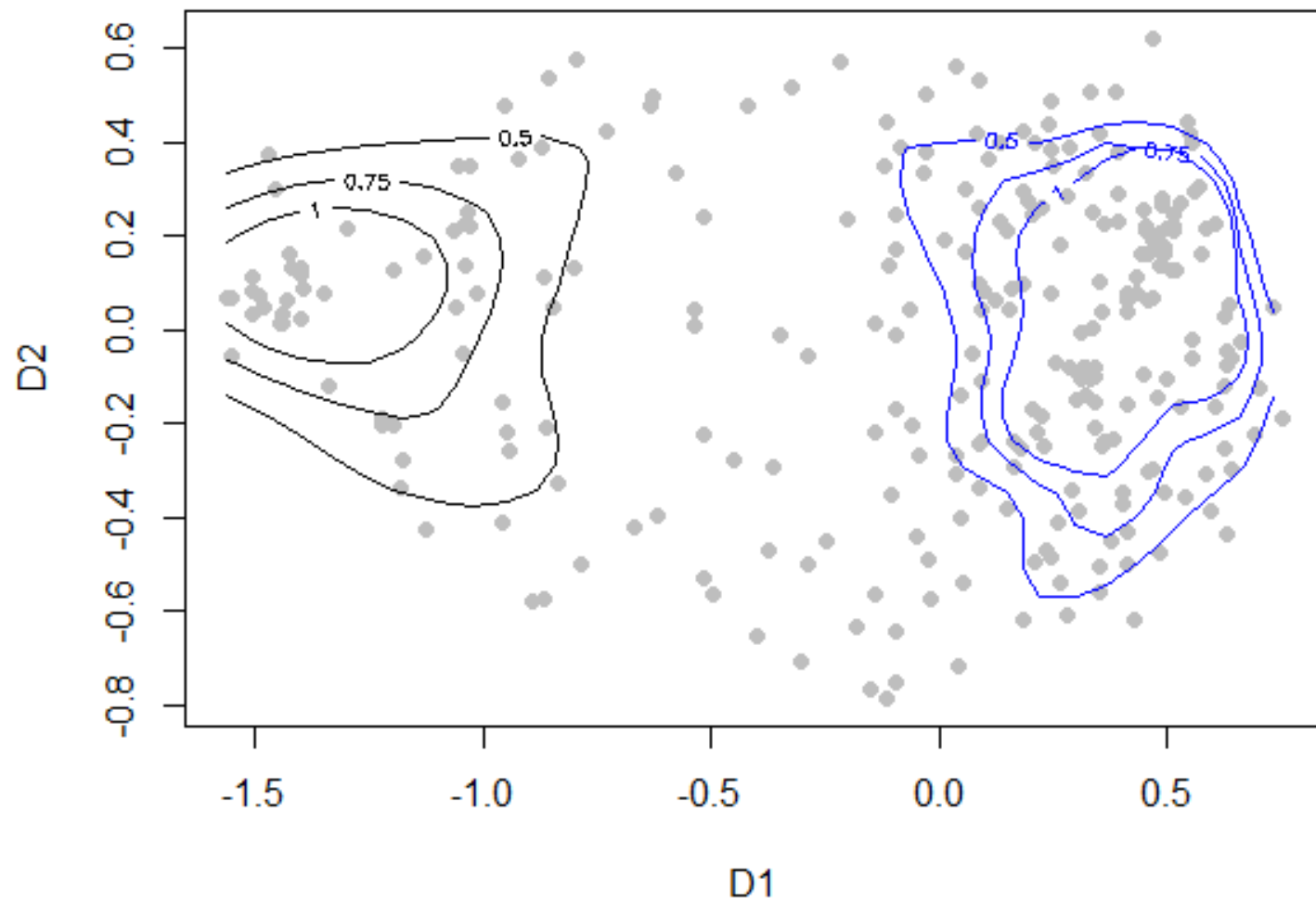
Semantic interpretation of dimensions

- Regression on the dimensional coordinates with 13 semantic variables
- Dim1 (horizontal):
 - letting vs. making: adj. $R^2 = 0.27$
 - Causee having control vs. not having control: adj. $R^2 = 0.20$
 - Caused Event is Action vs. State: adj. $R^2 = 0.19$.
- Dim2 (vertical):
 - non-intentional (bottom) vs. intentional Causer (top): adj. $R^2 = 0.03$.

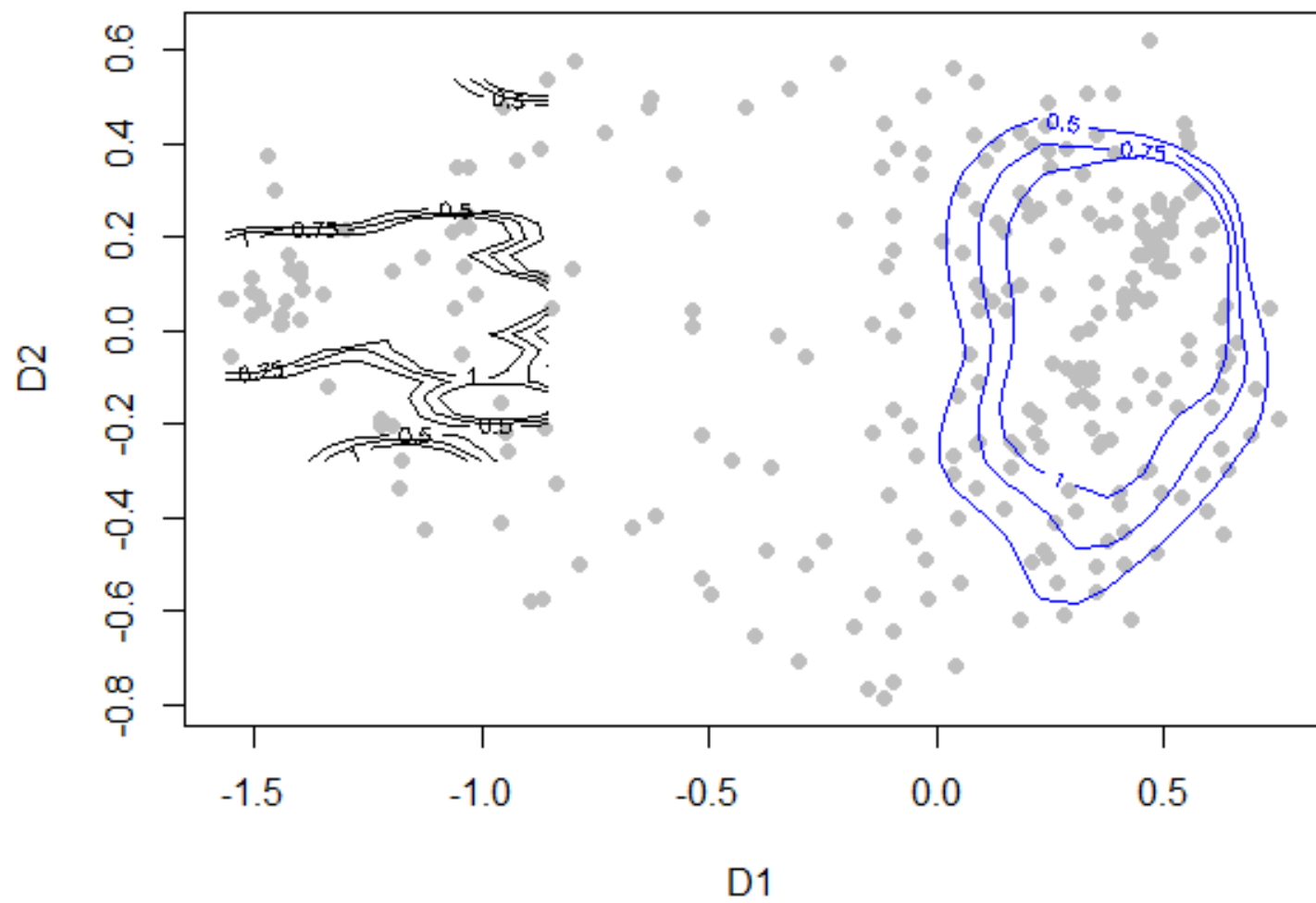
French



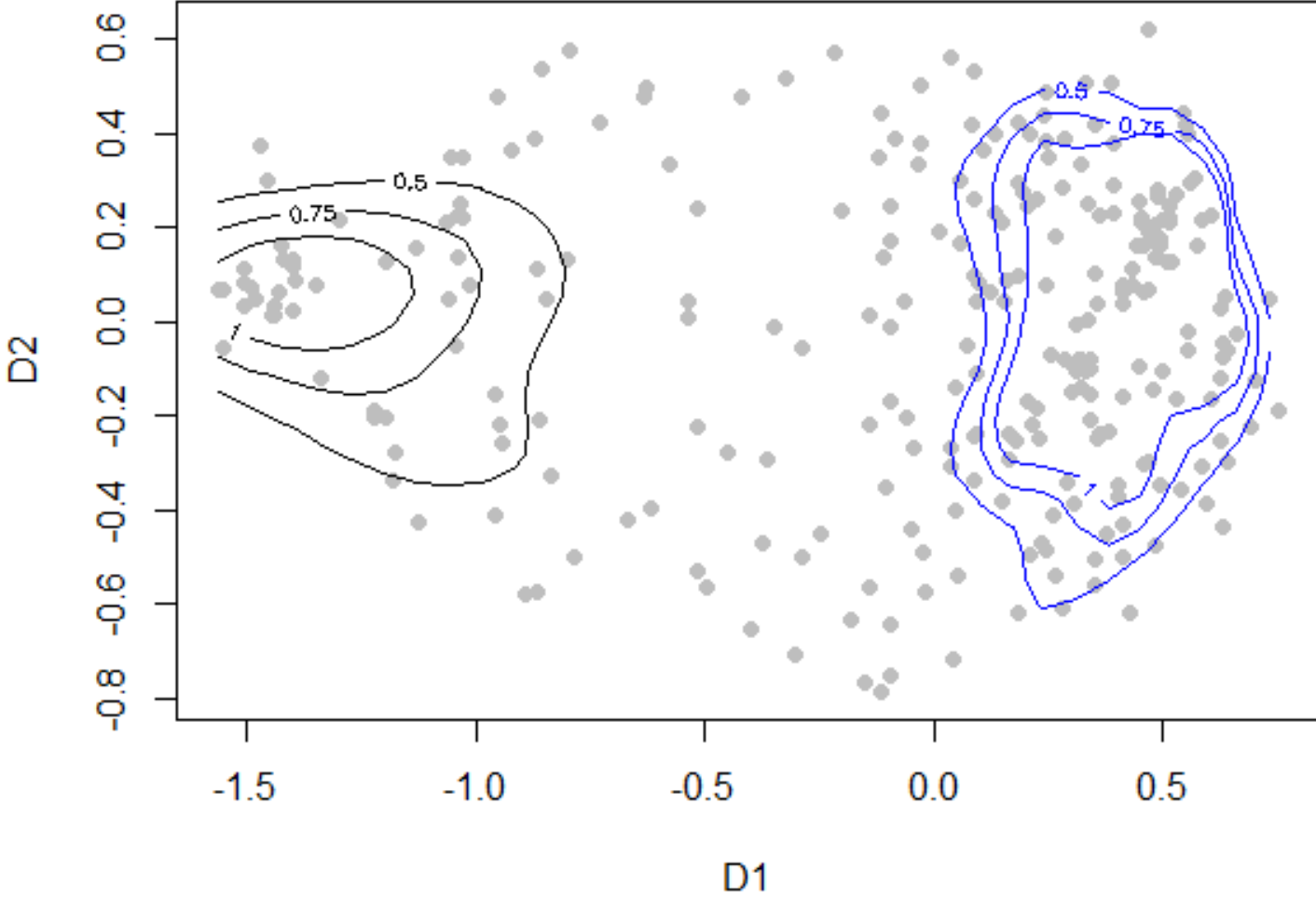
Contour plot, French



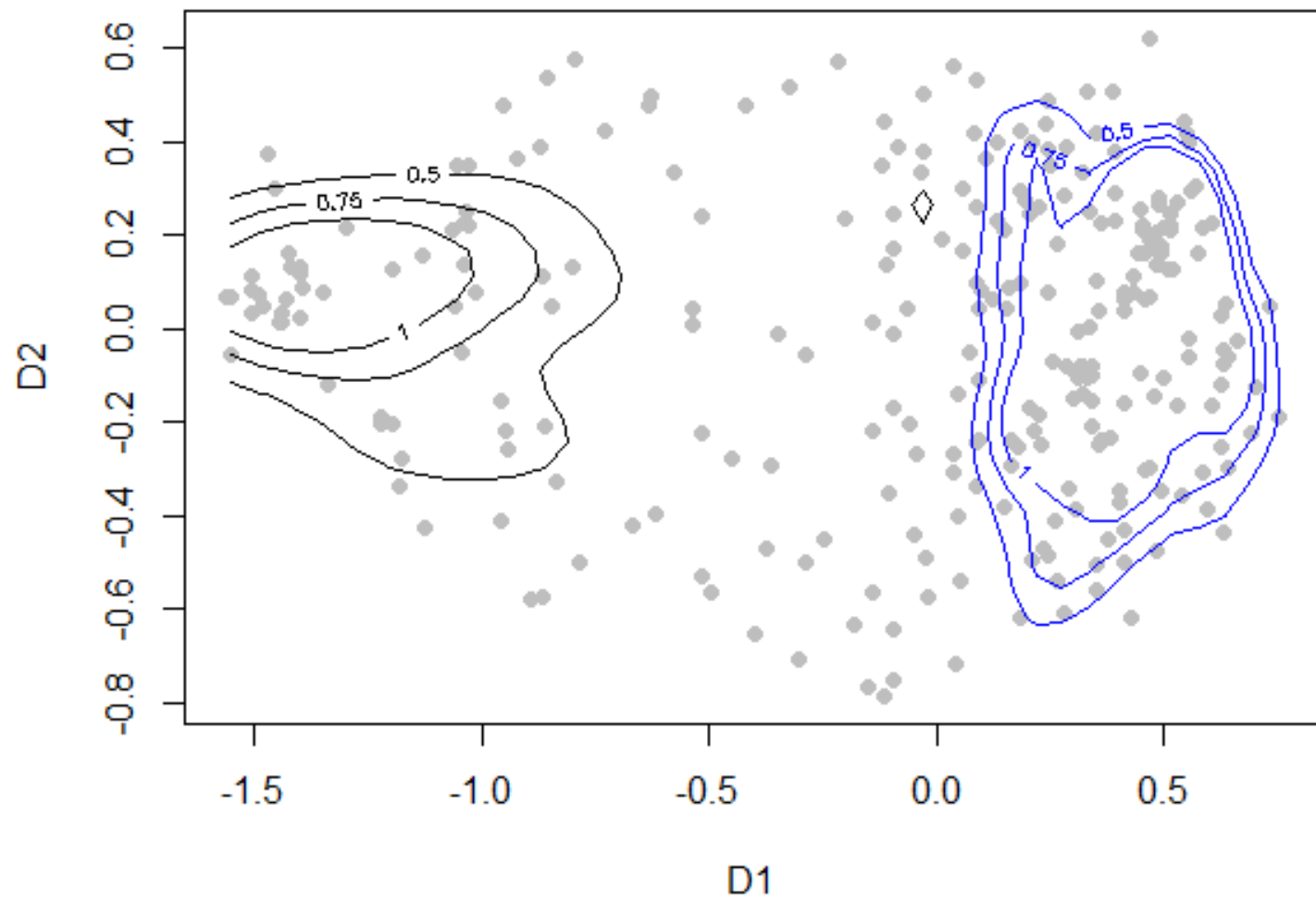
Contour plot, Russian



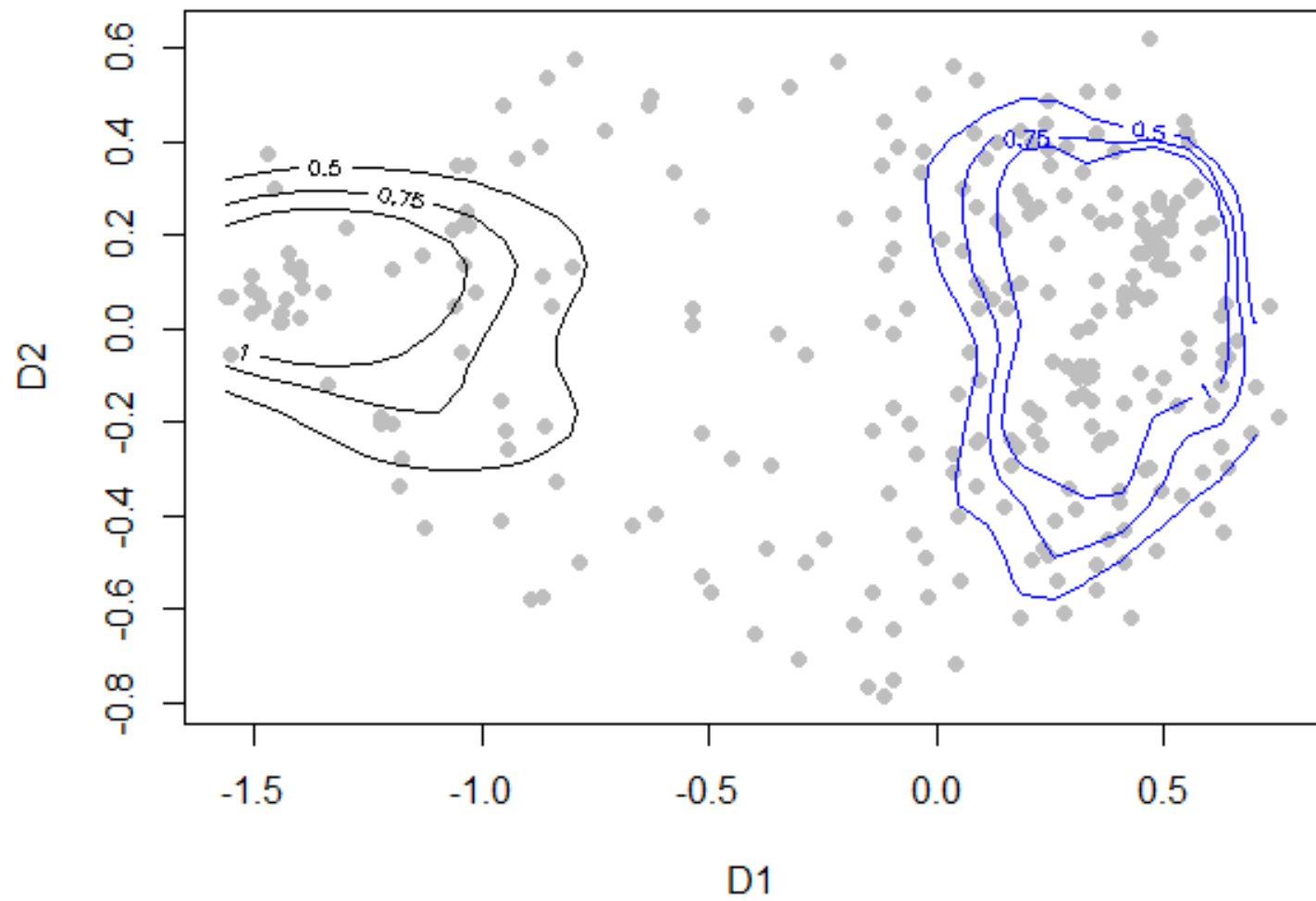
Contour plot, Vietnamese



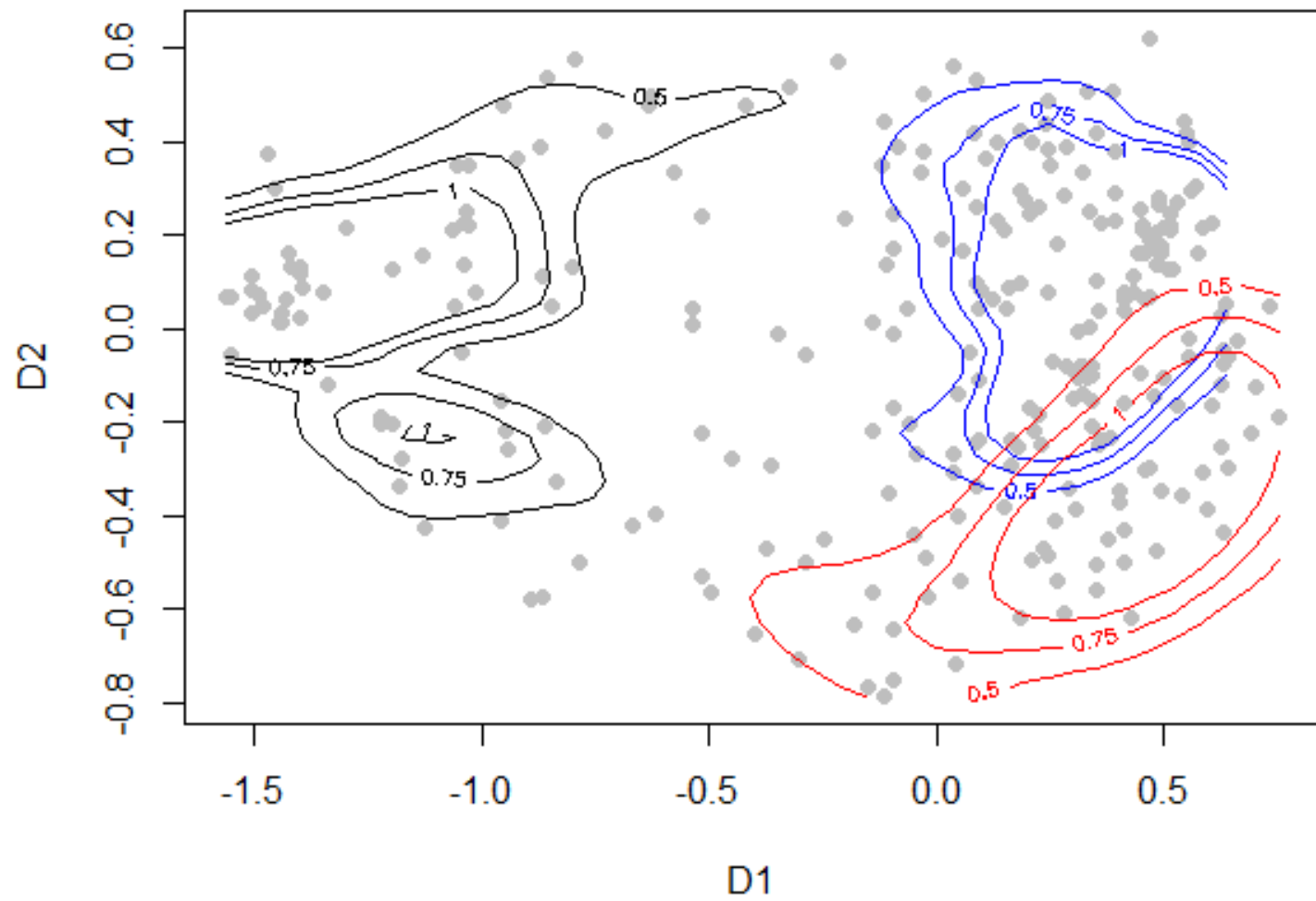
Contour plot, Thai



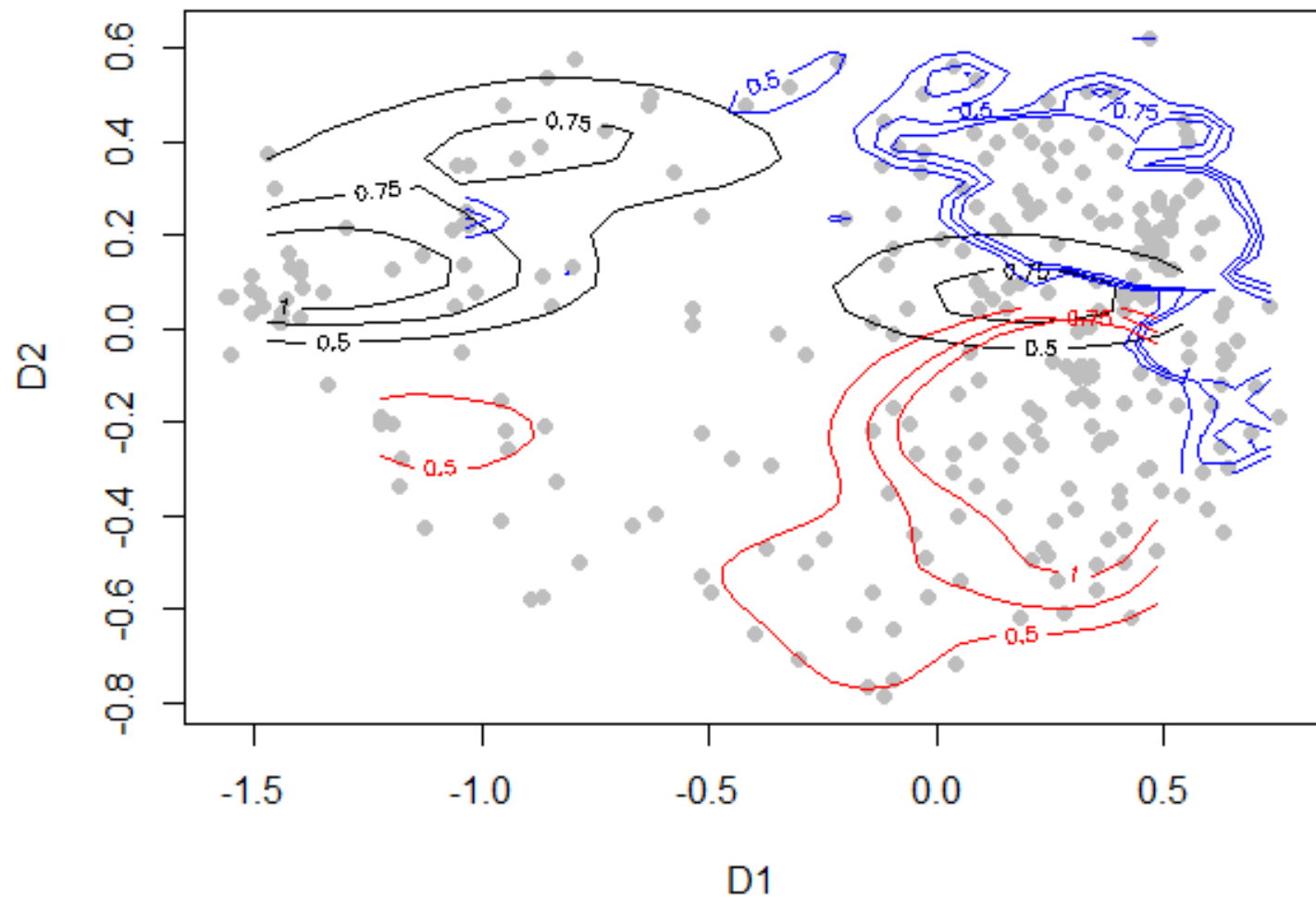
Contour plot, Chinese



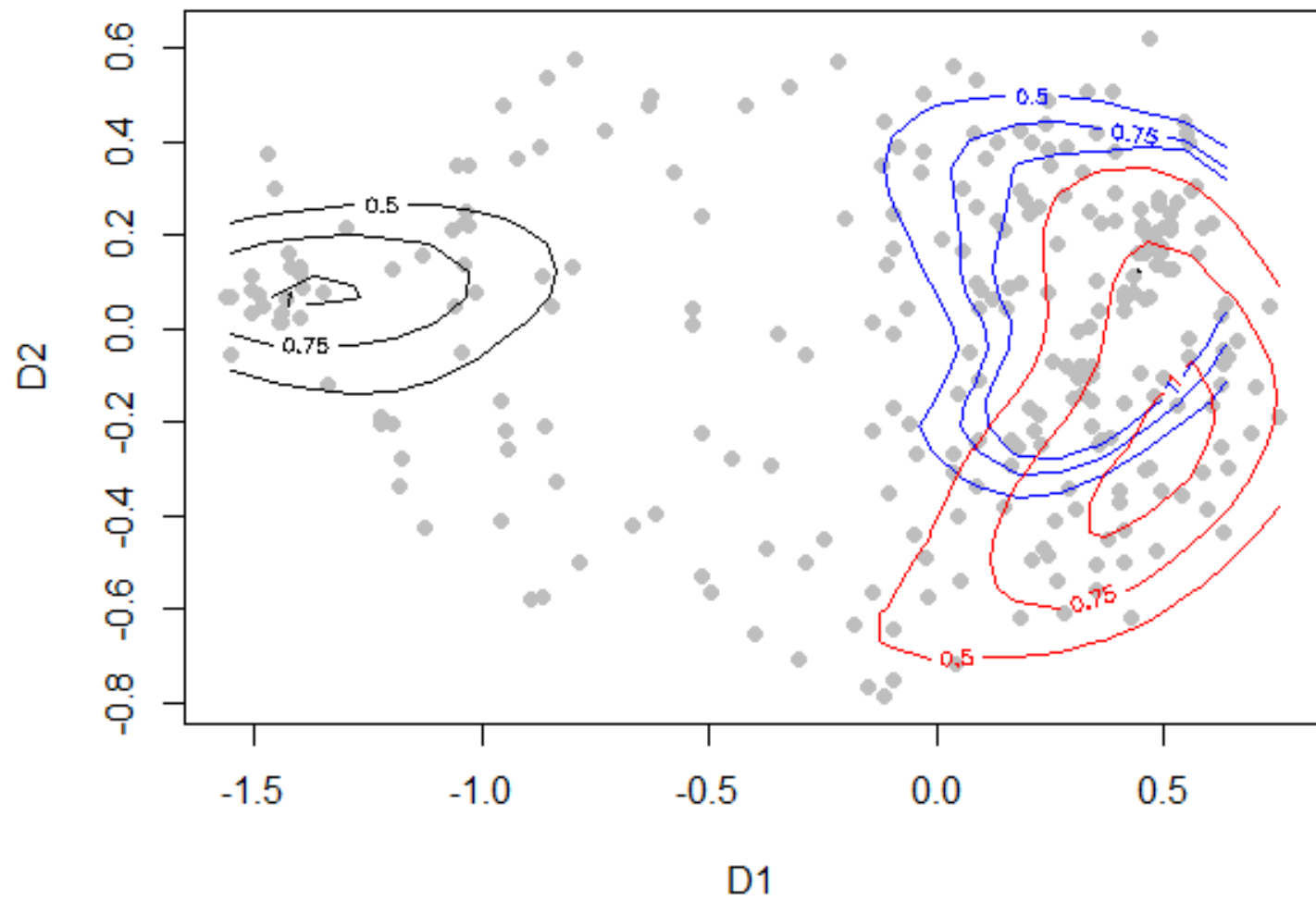
Contour plot, Finnish



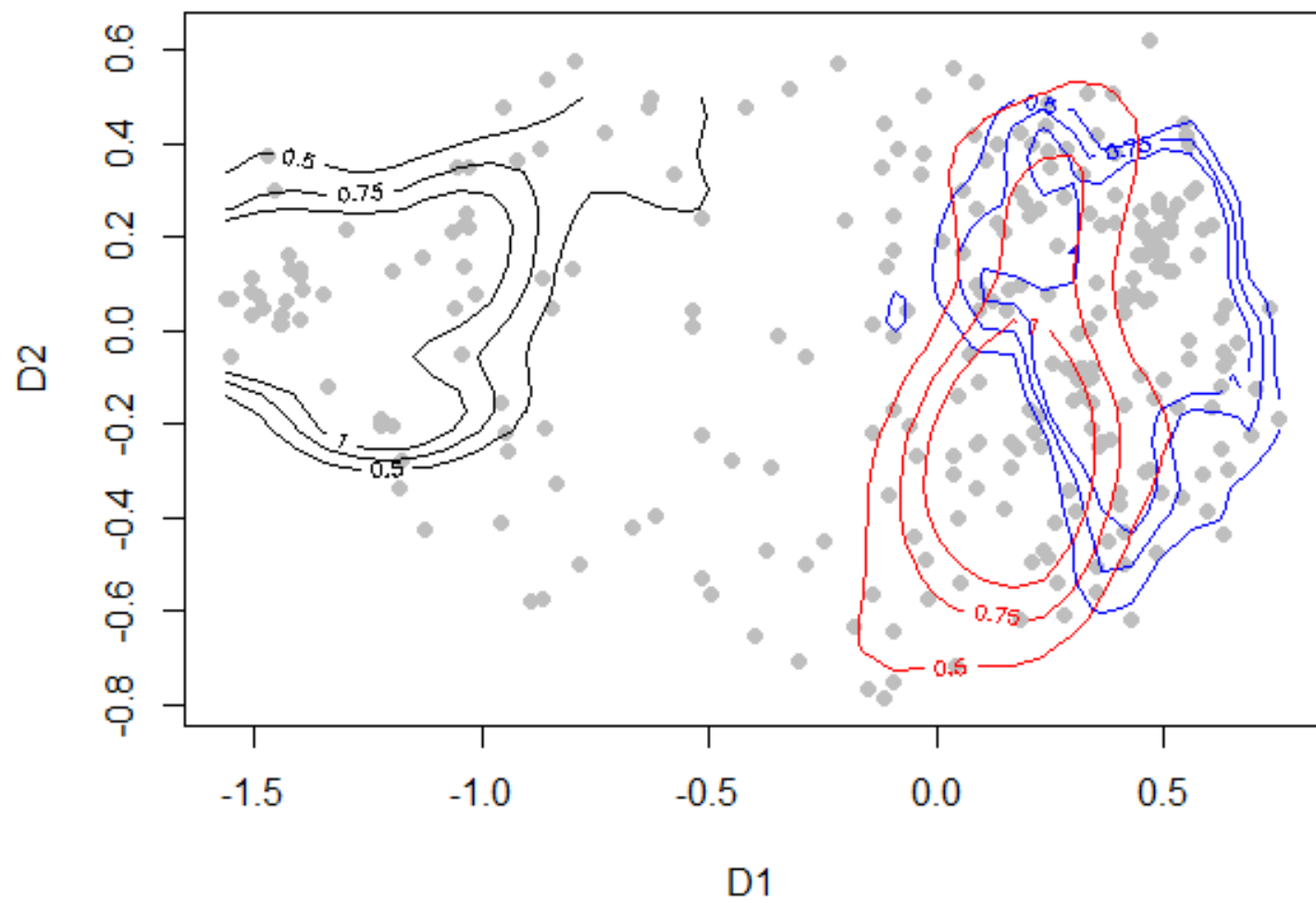
Contour plot, Turkish



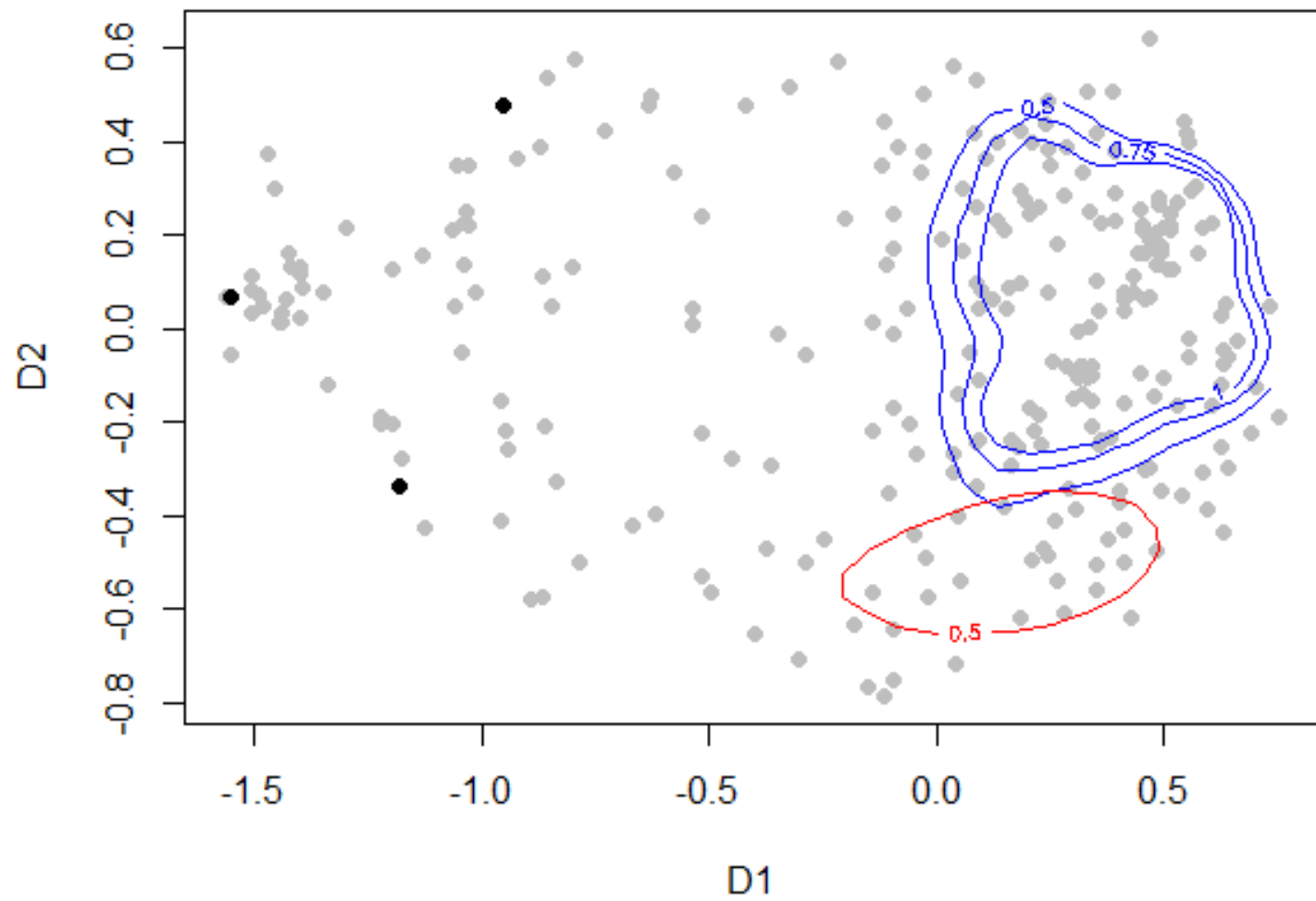
Contour plot, Indonesian



Contour plot, Hebrew



Contour plot, Japanese



Interim conclusions

- There is substantial cross-linguistic similarity in the semantic areas occupied by the lexical, morphological and analytic causatives.
- The lexical and morphological causatives are more similar to each other than to the analytic causatives.
- The variation is at least two-dimensional.

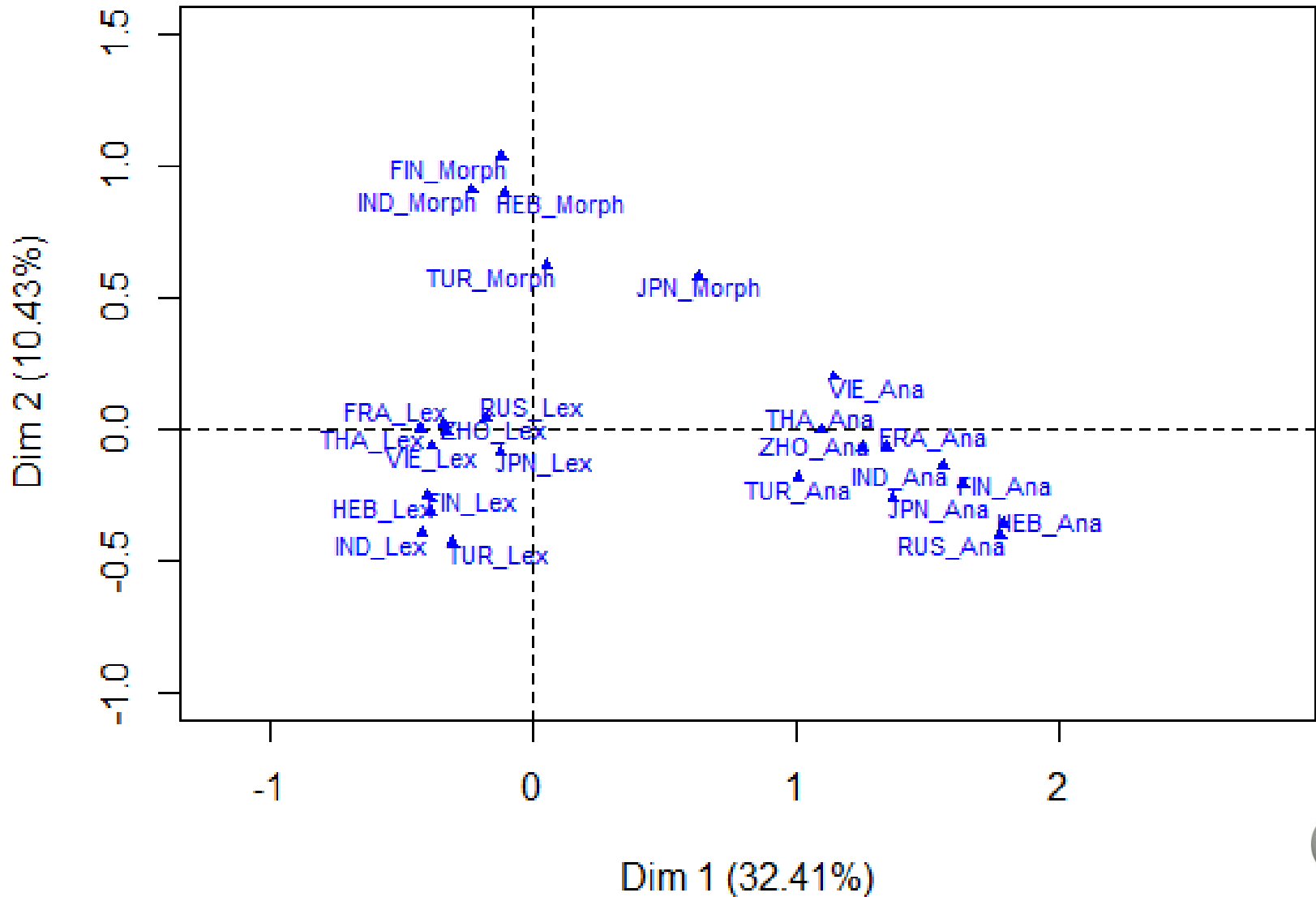
Outline

1. Theoretical background
 - the causative continuum
 - causatives as comparative concepts
2. Data: the ParTy corpus
3. Mapping the continuum:
 - Token-based maps (MDS)
 - Type-based maps (MCA)
4. Conclusions

Type-based maps

- Multiple Correspondence Analysis is a dimensionality-reduction and visualization technique. It shows, among other things, how different values of 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.
- Package FactoMineR in R

MCA factor map



Outline

1. Theoretical background
 - the causative continuum
 - causatives as comparative concepts
2. Data: the ParTy corpus
3. Mapping the continuum:
 - Token-based maps (MDS)
 - Type-based maps (MCA)
4. Conclusions

Results

- Both methods converge:
 - High cross-linguistic distributional similarity for each constructional type (lexical, morphological and analytic)
 - Instead of structuralist arbitrariness we have a probabilistic distributional ‘Universal Grammar’ of causatives
 - No unidimensional causative continuum:
Lexical/morphological vs. analytic = 1nd dimension; lexical vs. morphological = 2st dimension.

The causative continuum

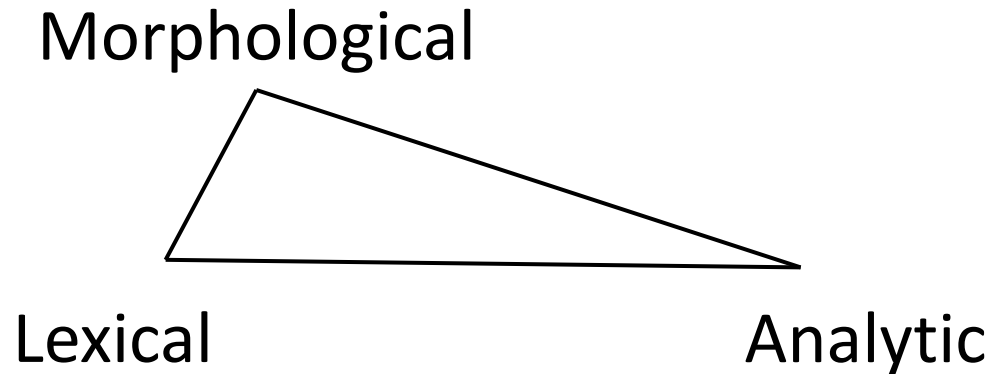
Lexical <> Morphological <> Analytic (Periphrastic)

Comrie 1981, Dixon 2000, Shibatani & Pardeshi 2002, etc.

The causative continuum

Lexical <> Morphological <> Analytic (Periphrastic)

More exactly:



Thank you!

The slides will be available at

www.natalialevshina.com/presentations.html

Questions? Suggestions?

natalevs@gmail.com