

Let the movies begin:

A typology of LET-verbs in European languages based on a
parallel corpus of film subtitles

Natalia Levshina

Philipps University of Marburg

WS Parallel Corpora and Linguistic Theory

46th SLE meeting, Split 2013

Outline

- Verbs of letting in European languages: semantic peculiarities
- Quantitative comparison of eight LET-verbs in a parallel corpus of film subtitles
 - why film subtitles?
 - data and method
 - results: language space and exemplar space of LET
- Conclusions

What is LETTING?

- Letting can be defined as 'negative causation'.
- Causation proper: the Antagonist overrides the intrinsic tendency of the Agonist (Talmy 2000).
- Letting: the Antagonist fails, deliberately or not, to override the Agonist's intrinsic tendency.

Antagonist

Agonist

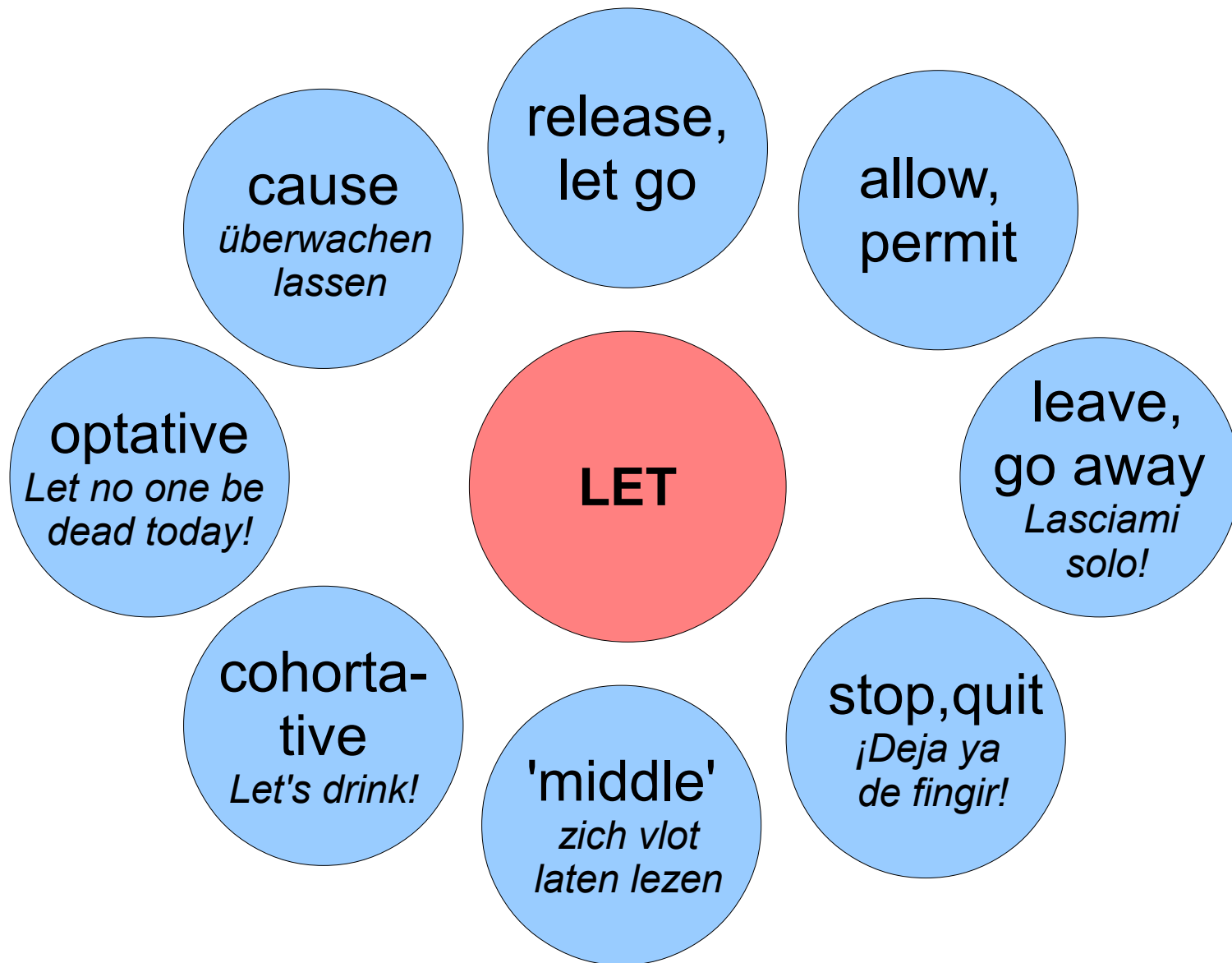
Causation: ***The detective*** made ***the thief*** *confess where he had hidden the stolen diamonds.*

Letting: ***The detective*** let ***the thief*** *escape.*

Semantic variation of LETTING

- letting is a very complex concept, which varies along many dimensions, most importantly:
 - force dynamic patterns:
 - active or passive Permitter (*Let my people go vs. Let it be*)
 - dynamic or static Permittee (*Let my people go vs. Let it be*)
 - physical or psychosocial domain:
 - The plug's coming loose let the water drain from the tank.*
 - The mother doesn't let the child eat junk food.*

A bouquet of functions



The aim of the study

- to explore semantic differences and similarities of etymologically related *let*-verbs in 4 Germanic and 4 Romance languages

let, laten, lassen, låta

laisser, dejar, deixar, lasciare

- historical prototypes (Soares da Silva 2007):

- Romance: Lat. *laxare* 'let go, release'

- Germanic: Goth. *letan* 'leave, let'



from PIE */e-
"to let go, slacken"

- what about now?

A methodological problem

- how to compare highly polysemous constructions in different languages?
 - e.g. German *lassen* and Dutch *laten*. Both can express letting go, leaving, abandonment, indirect causation, hortativity, etc.
- Often the differences are a matter of degree, or 'prototypicality' of a specific sense. These differences, however, may suggest ongoing language change.
 - E.g. French *laisser* is losing its 'let go' function, as a manifestation of its more advanced grammaticalization in comparison with other Romance *let*-verbs (Soares da Silva 2012)

Methodological proposal

- use a multilingual parallel corpus to compare the LET-verbs
- compare the distributions of LET-verbs across many aligned contexts
- compare the semantic scope of each verb by looking at the region occupied by its exemplars in a common semantic space
- compare the prominence of each sense by examining the density of the LET-exemplars in the corresponding region of the common semantic space

Outline

- Verbs of letting in European languages: semantic peculiarities
- **Quantitative comparison of eight LET-verbs in a parallel corpus of film subtitles**
 - why film subtitles?
 - data and method
 - results: language space and exemplar space of LET
- Conclusions

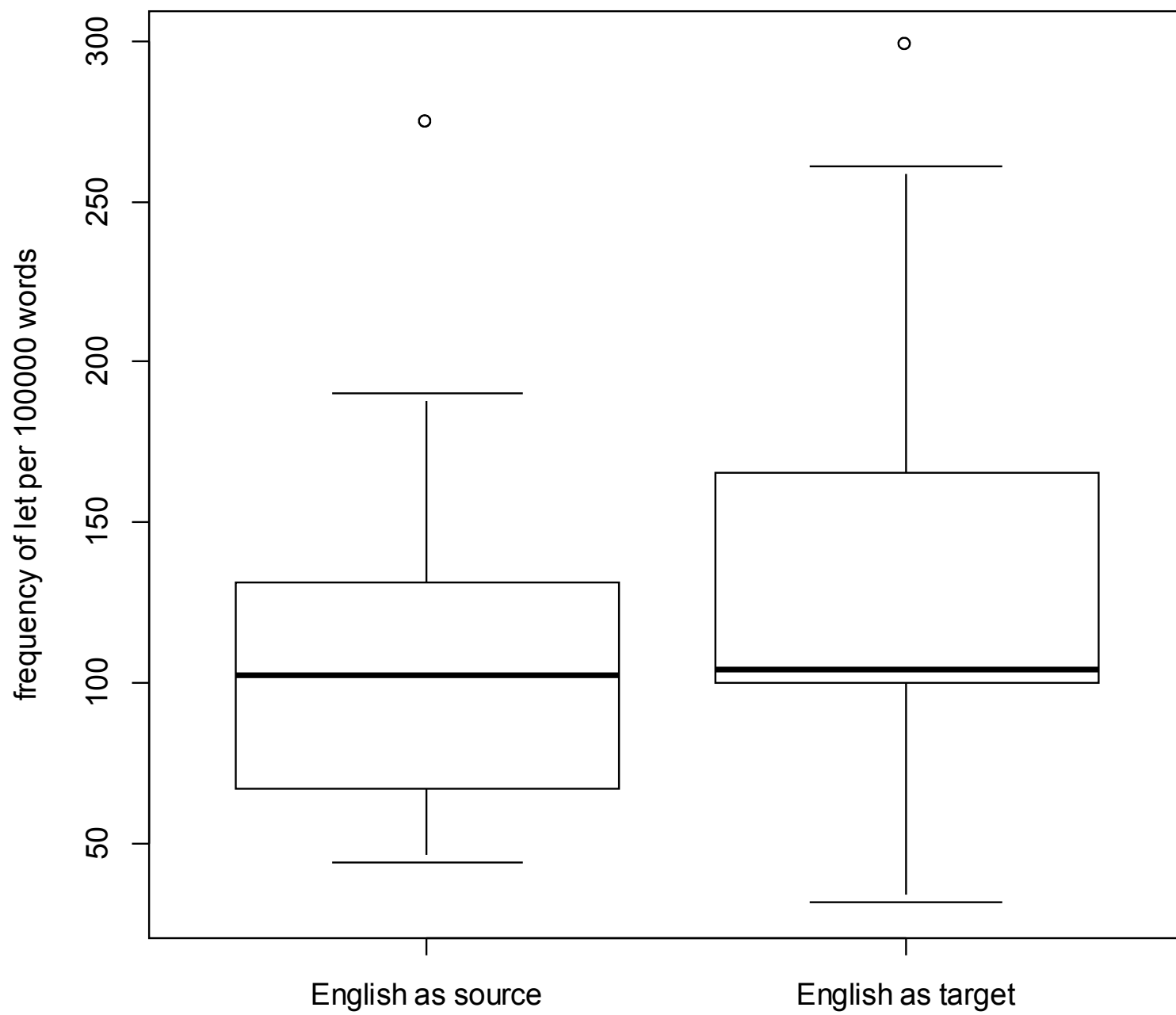
Why film subtitles?

- freely available online for many languages
- timing information helps to align sentences
- free from censorship and normativity, although quality is to some extent controlled by the online community and fans
- closer to 'normal' spoken language than any other existing parallel texts (in comparison to Europarl and Bible translations)

Monolingual subtitles serve as a better approximation of entrenchment in lexical decision experiments than standard 'balanced' corpora!

Is translationese an issue?

- a sample of 15 randomly selected English-source and 15 English-target film subtitles (source languages: German, French, Spanish, Italian, Swedish, Danish, Dutch, Brazilian Portuguese, Russian, Japanese, Chinese), various genres
- search for verb 'let' (per 100 000 words)
- medians are identical, no significant difference in normalized frequencies of *let* (Mann-Whitney test: $p = 0.461$)
- more variation in translated English, although this is not statistically significant (Flinger's test: $p = 0.97$)



Data

- 6 films in different original languages



Data

- subtitles in eight languages from opensubtitles.org
- srt format
- converted to XML
- aligned with the help of timing information (Tiedemann 2011):
trimming, Procrustes procedure
- search for verb stems in each language
- 364 multilingual contexts

srt format

Number

1

Start time

End time

00:00:56,080 --> 00:01:00,798

<i>Sooner or later though,
you always have to wake up.</i>

Text

2

00:01:12,792 --> 00:01:15,476

<i>In cryo you don't dream at all.</i>

3

00:01:15,511 --> 00:01:17,250

<i>Doesn't feel like 6 years.</i>

4

00:01:17,285 --> 00:01:20,688

<i>More like after tequila

and ass kicking.</i>

Avatar.2009.720p.TS.XviD-
ViSiON

SubRip format (.srt)

Matrix

	A	B	C	D	E	F	G	H	I
1	Film	DE_Cx	EN_Cx	ES_Cx	FR_Cx	IT_Cx	NL_Cx	PT_Cx	SV_Cx
2	Tiger	Yes	Yes	No	No	No	Yes	No	No
3	Tiger	Yes	Yes	No	No	Yes	NA	Yes	No
4	Tiger	No	Yes	No	Yes	No	Yes	No	Yes
5	Tiger	No	Yes	No	Yes	No	Yes	No	Yes
6	Tiger	Yes	Yes	No	No	No	No	Yes	No
7	Tiger	No	Yes	No	No	No	Yes	No	Yes
8	Tiger	Yes	Yes	No	No	No	No	No	Yes
9	Tiger	No	Yes	No	No	No	No	Yes	Yes
10	Tiger	No	Yes	No	Yes	Yes	Yes	No	Yes
11	Tiger	Yes	Yes	No	Yes	Yes	Yes	Yes	No
12	Tiger	No	Yes	No	No	NA	No	Yes	Yes
13	Tiger	Yes	Yes	No	No	Yes	Yes	Yes	Yes
14	Tiger	No	Yes	No	No	No	No	Yes	No
15	Tiger	No	Yes	No	Yes	No	Yes	Yes	Yes
16	Tiger	Yes	Yes	No	No	No	Yes	No	No

Quantitative analyses

1. The space of 8 verbs: a matrix of dissimilarities between eight verbs (Gower's distances) represented by MDS.

The Distributional Principle: the **closer** the verbs in the space, the more frequently they are used in the **same contexts**.

2. The space of exemplars:

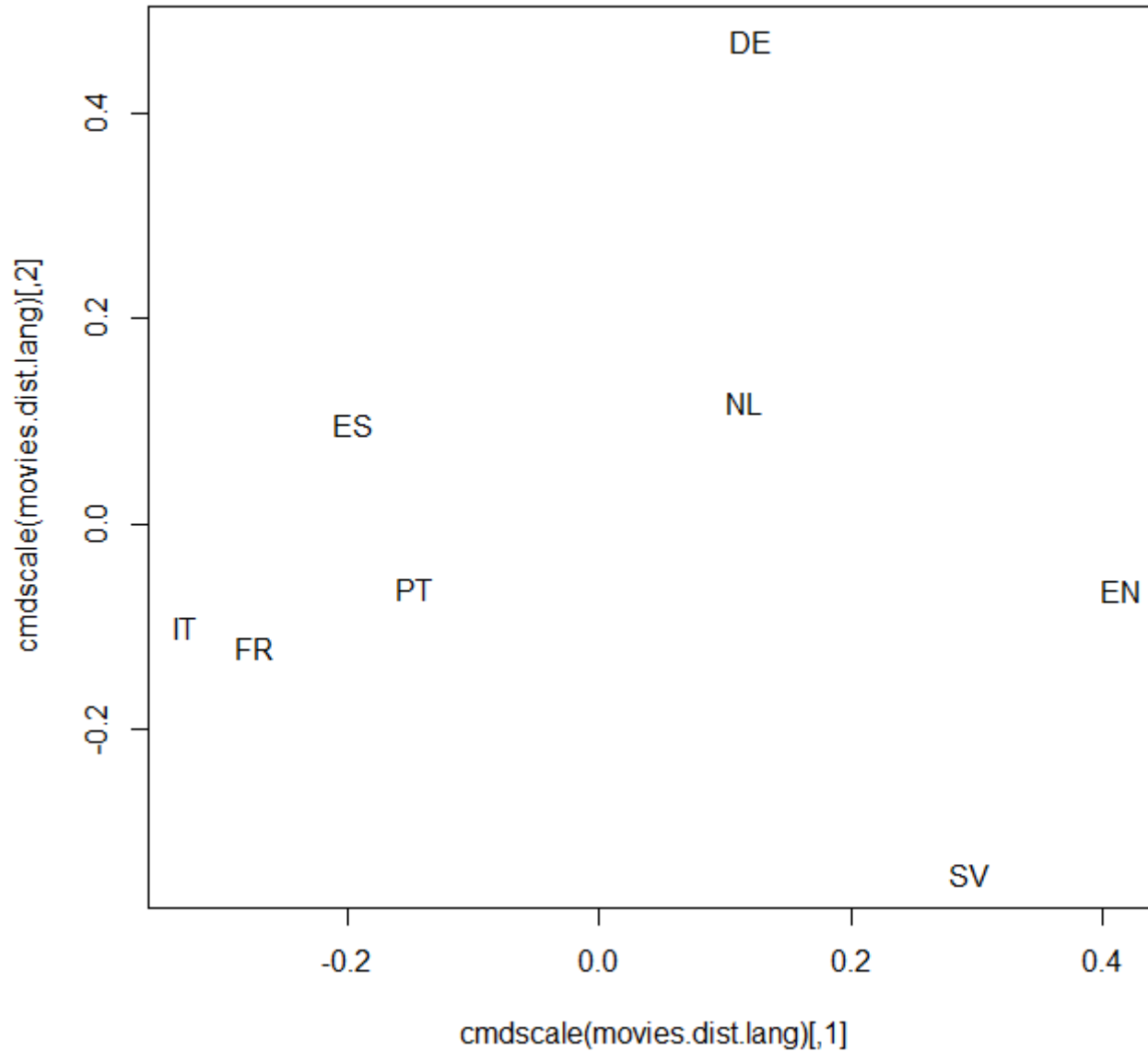
2.1. A matrix of dissimilarities between 364 exemplars represented in a MDS-based semantic space.

The Iconicity Principle: the **closer** the exemplars in the space, the more frequently they are translated with the **same type of construction** (let-verb or not let). In addition, the more central the exemplars in the space, the more frequently they are translated with a let-verb.

2.2. Kriging to obtain central tendencies for each language

Distances between 8 *let*-verbs

Distances between cognates



Quantitative analyses

1. The space of 8 verbs: a matrix of dissimilarities between eight verbs (Gower's distances) represented by MDS.

The Distributional Principle: the **closer** the verbs in the space, the more frequently they are used in the **same contexts**.

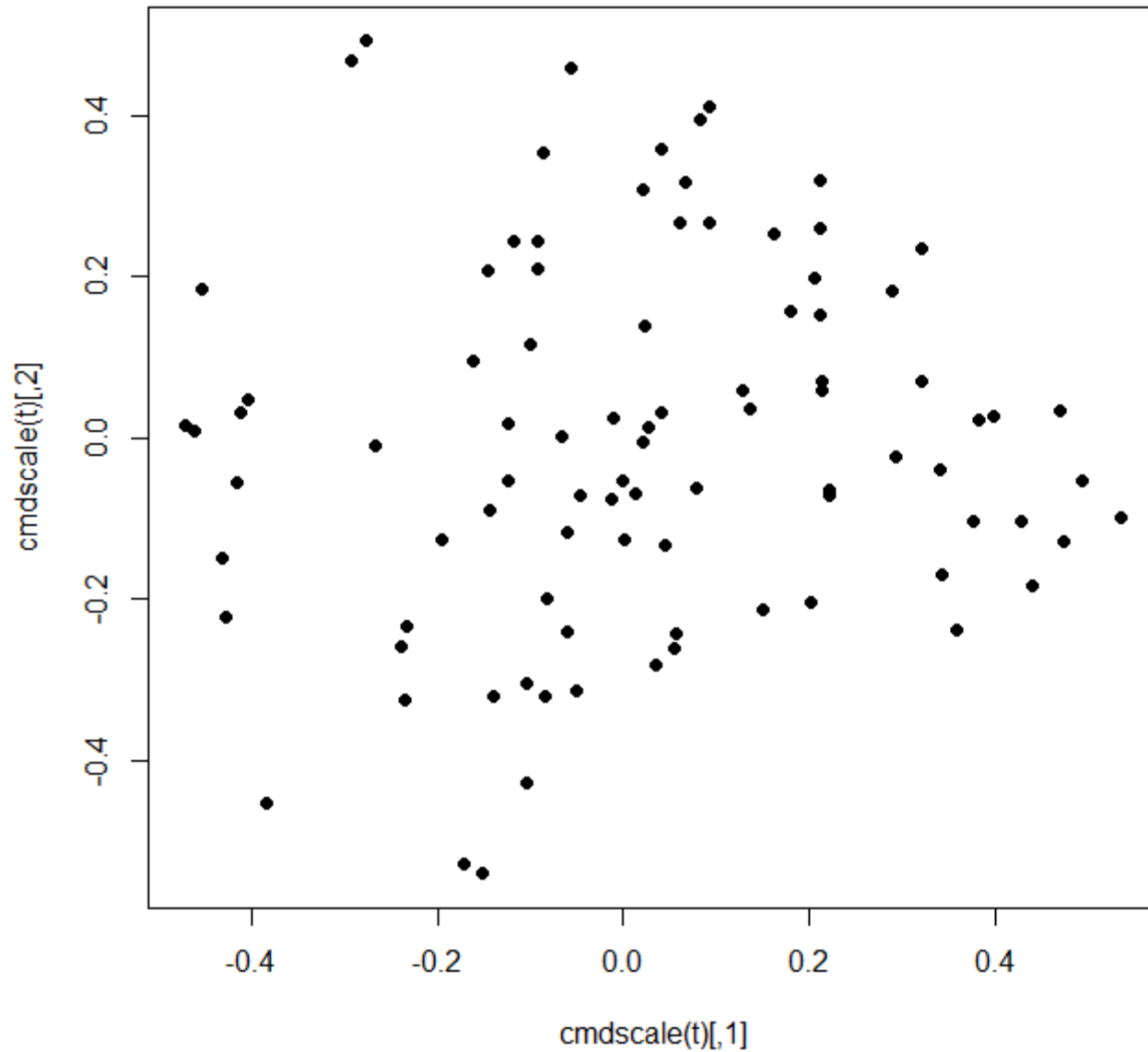
2. The space of exemplars:

- 2.1. A matrix of dissimilarities between 364 exemplars represented in a MDS-based semantic space.

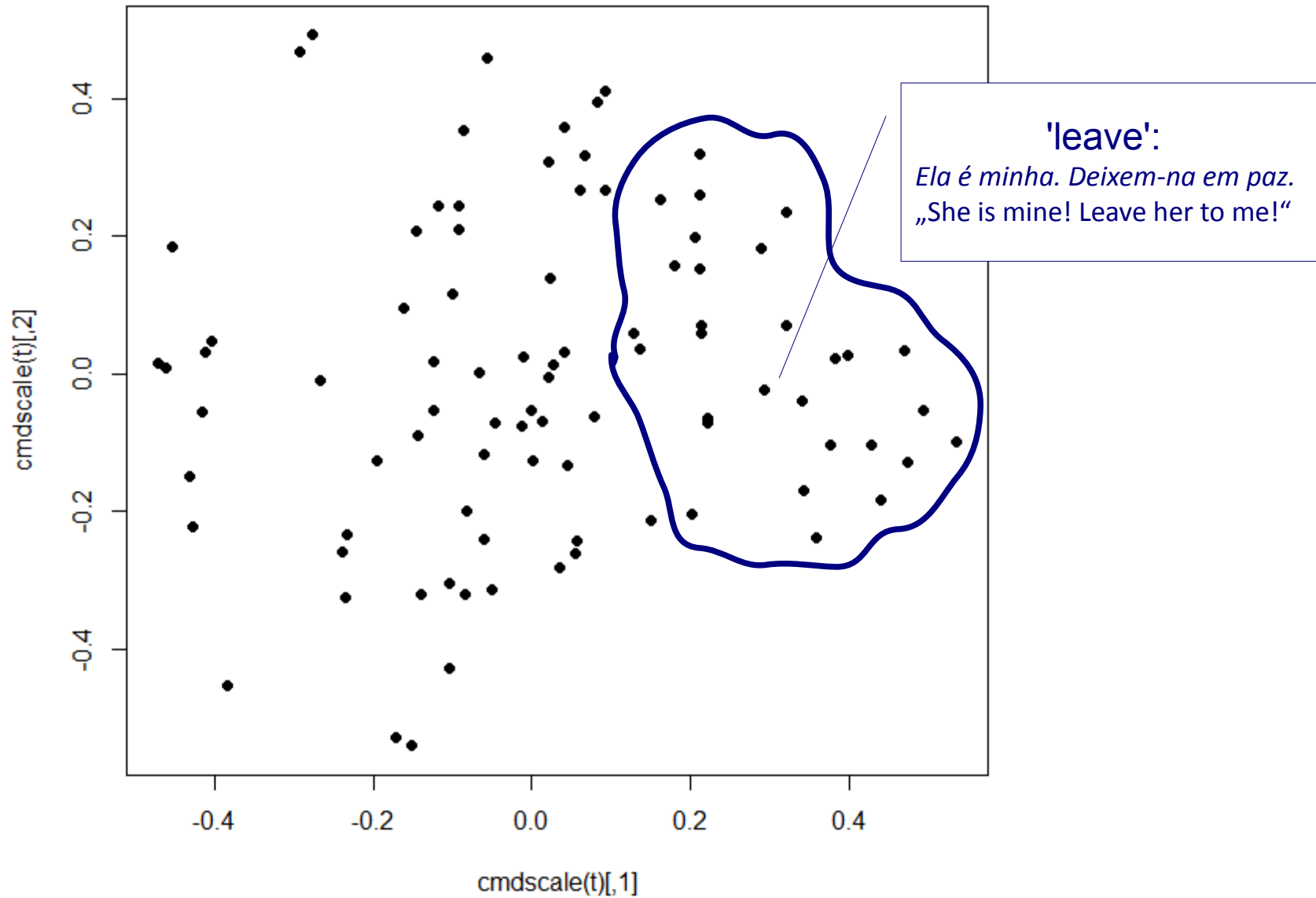
The Iconicity Principle: the **closer** the exemplars in the space, the more frequently they are translated with the **same type of construction** (let-verb or not let). In addition, the more central the exemplars in the space, the more frequently they are translated with a let-verb.

- 2.2. Kriging to obtain central tendencies for each language.

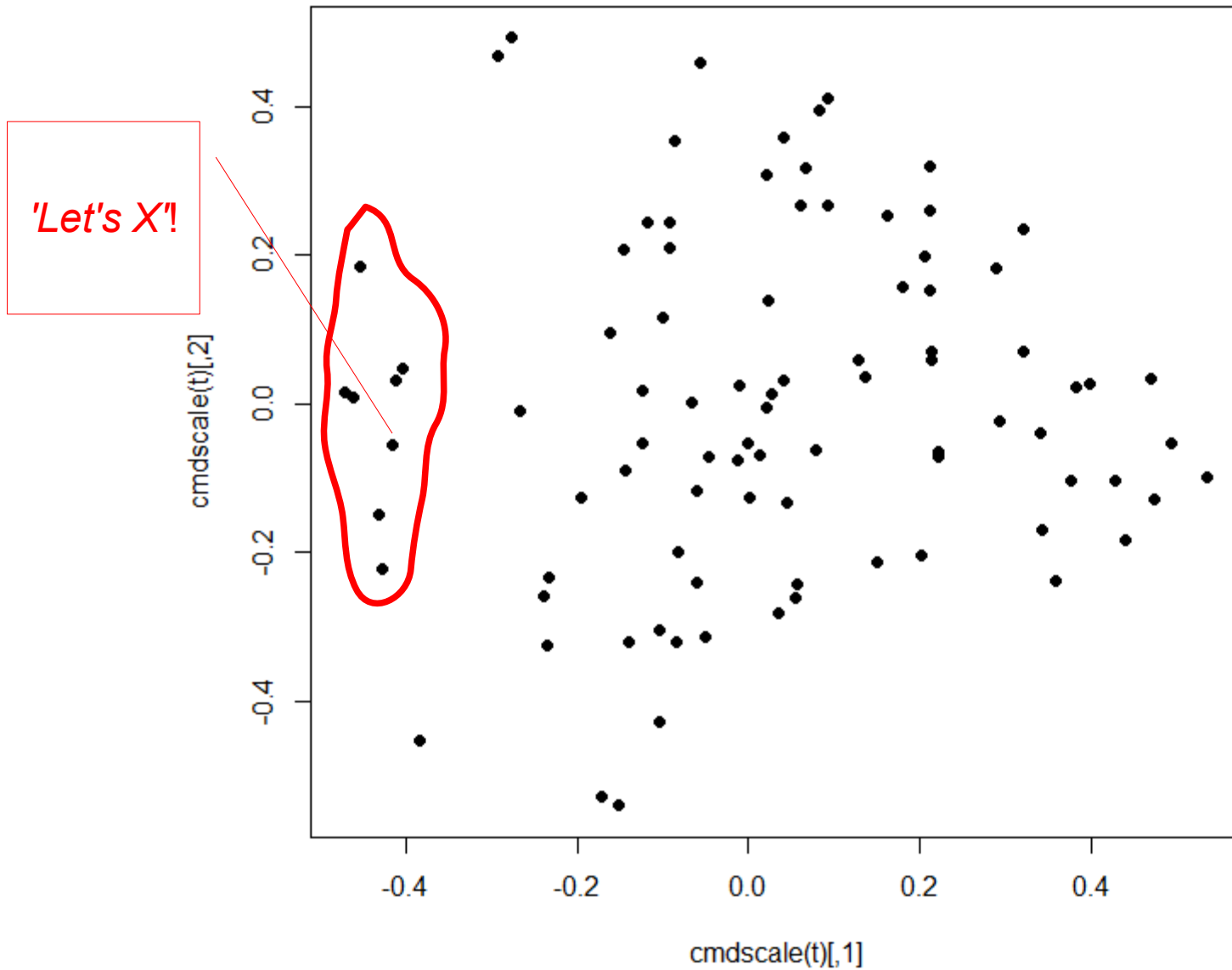
Exemplar space of LET



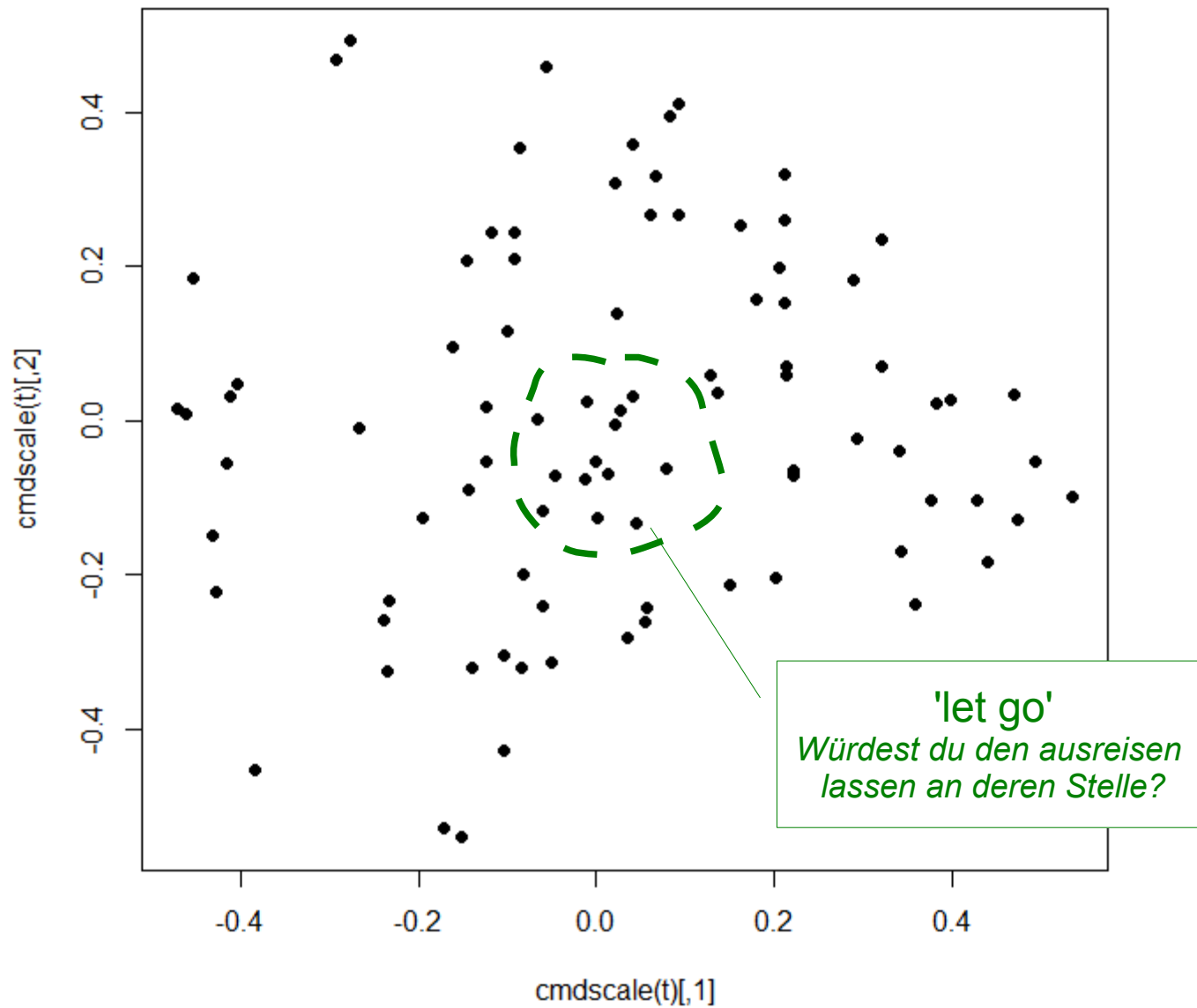
Exemplar space of LET



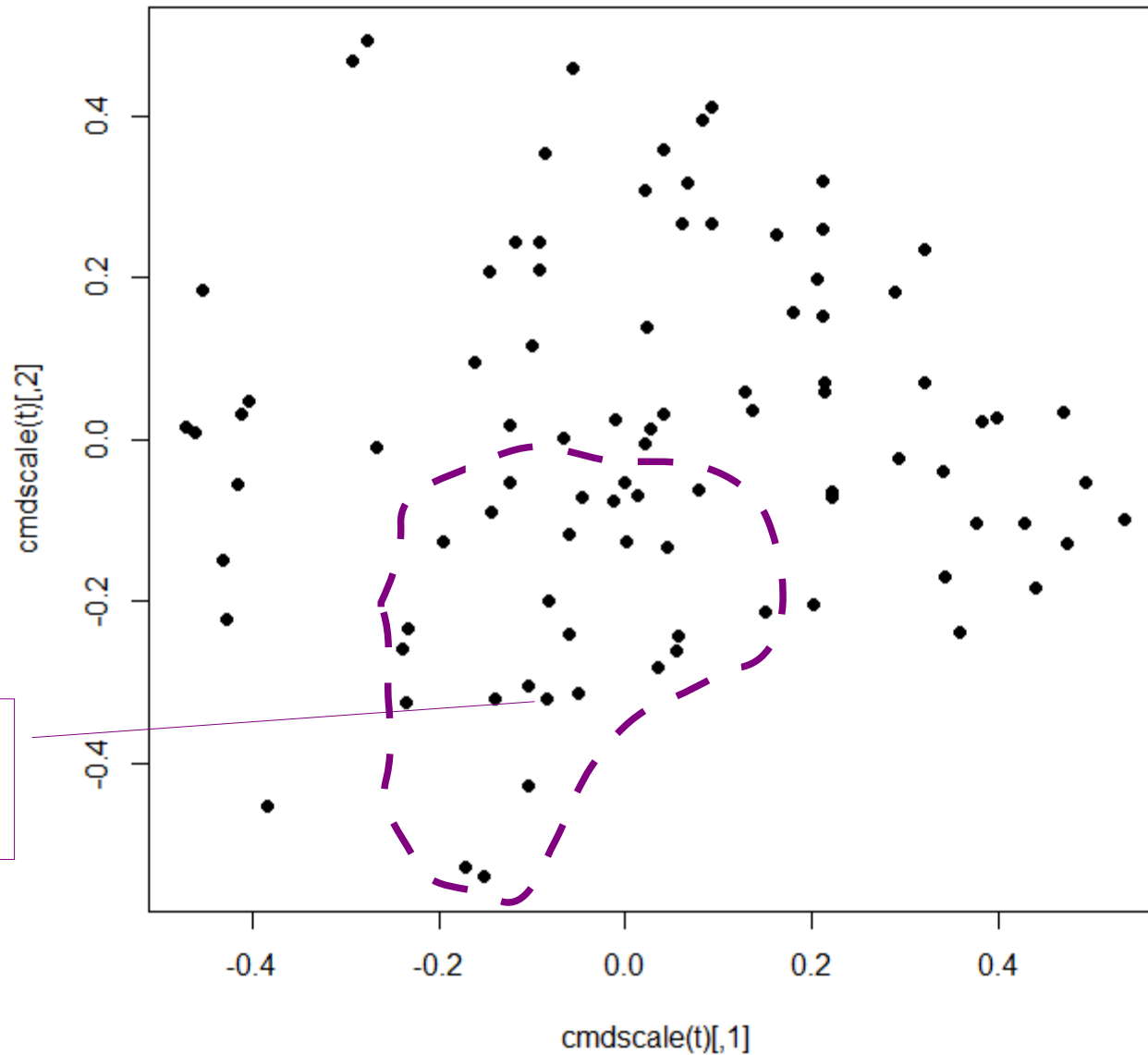
Exemplar space of LET



Exemplar space of LET

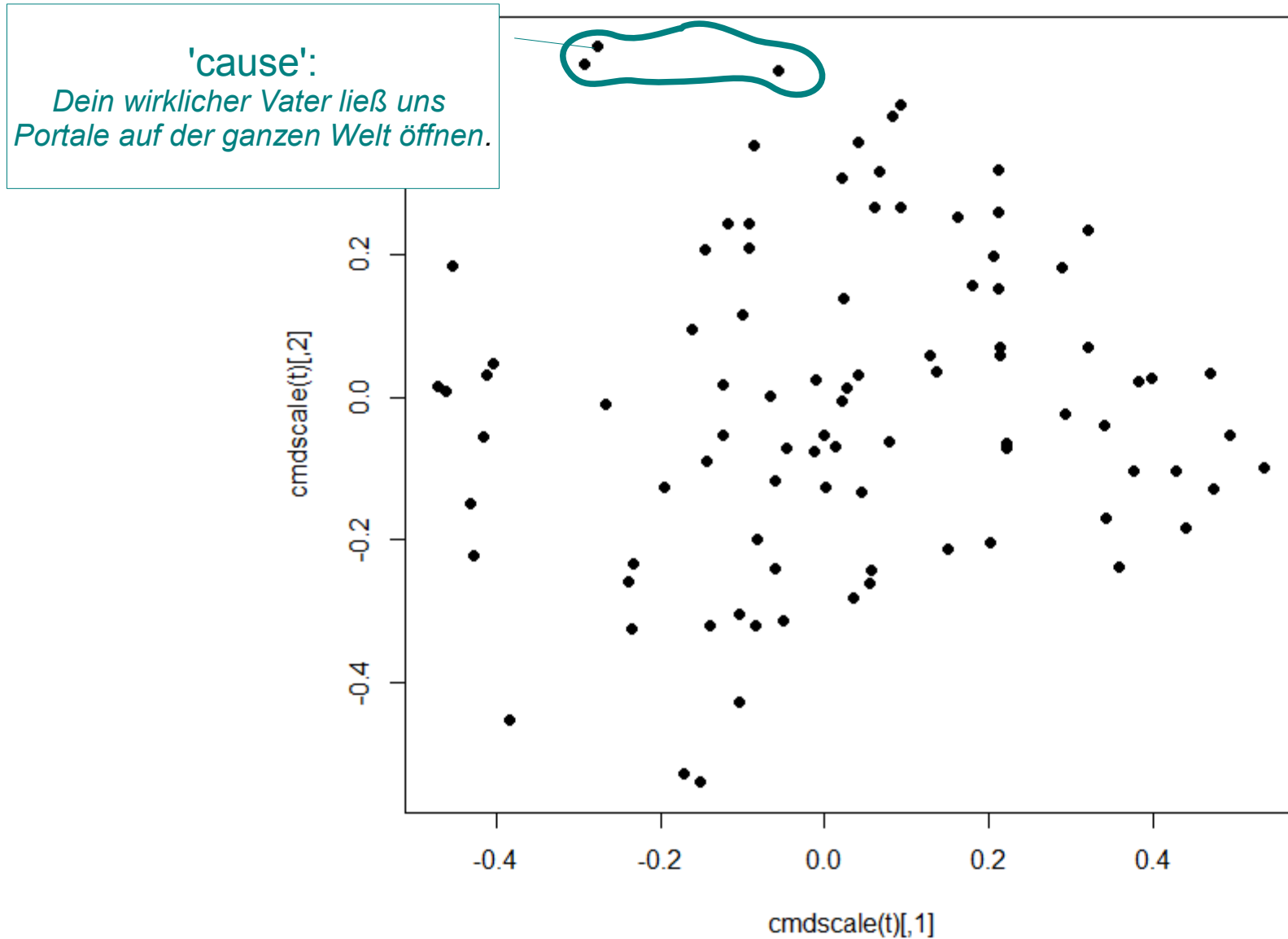


Exemplar space of LET

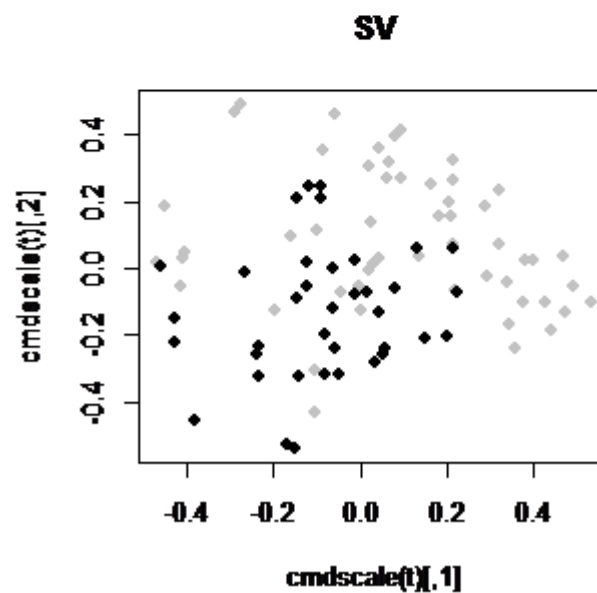
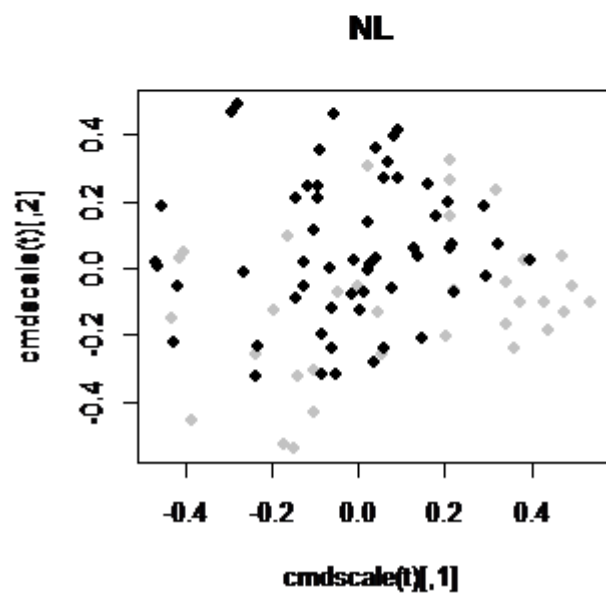
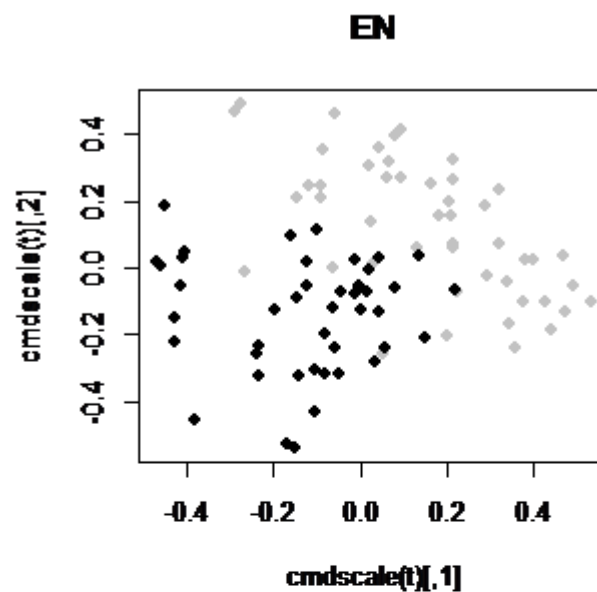
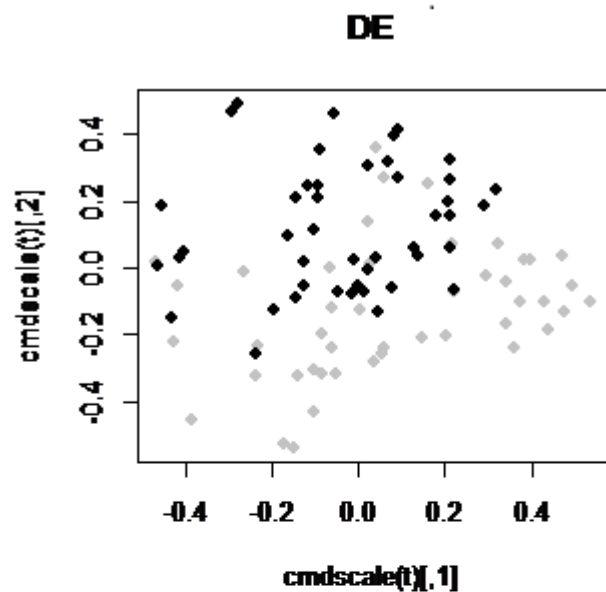


'allow'
*I can't let you
marry.*

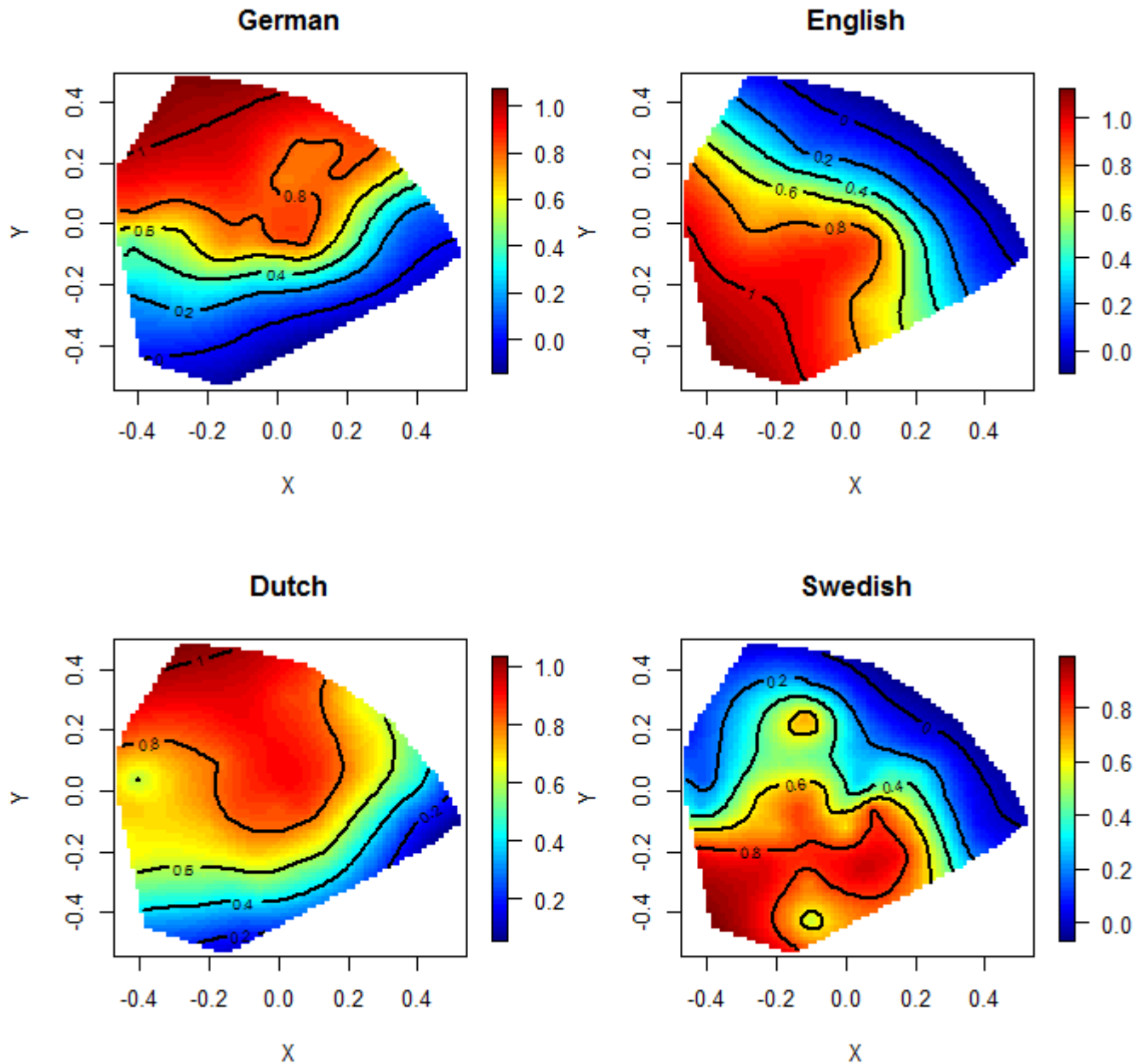
Exemplar space of LET



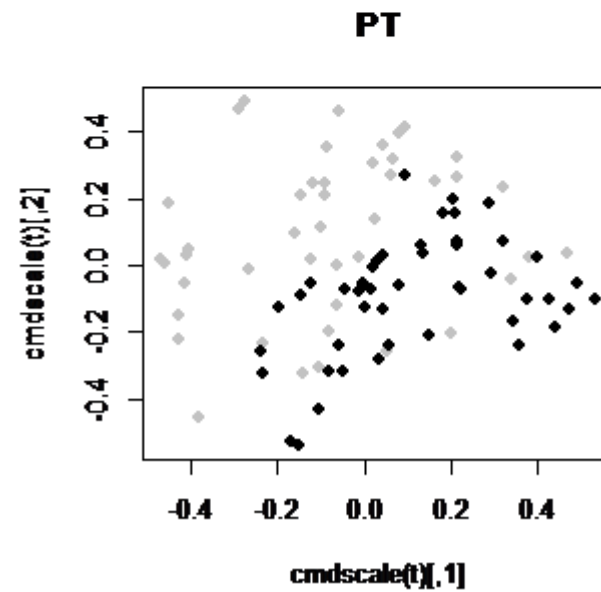
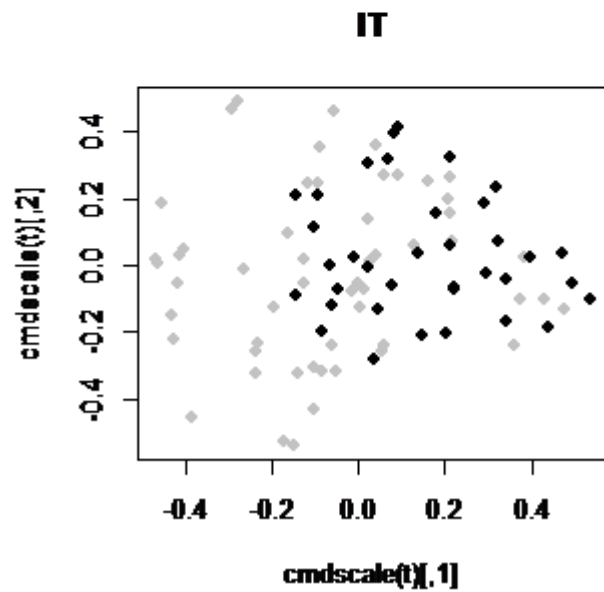
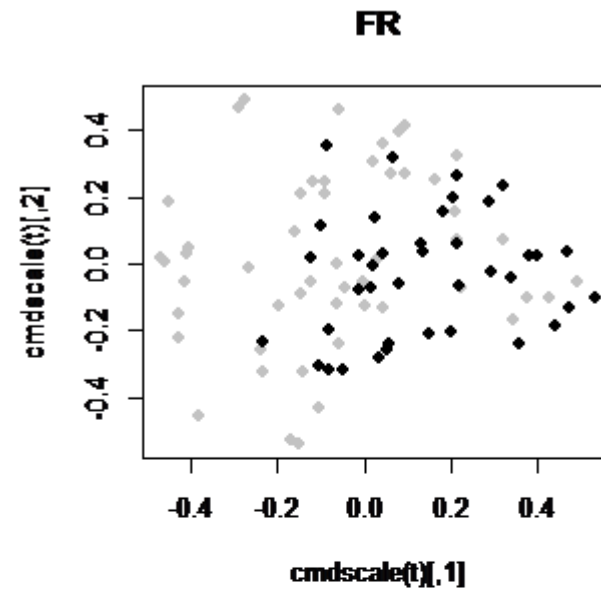
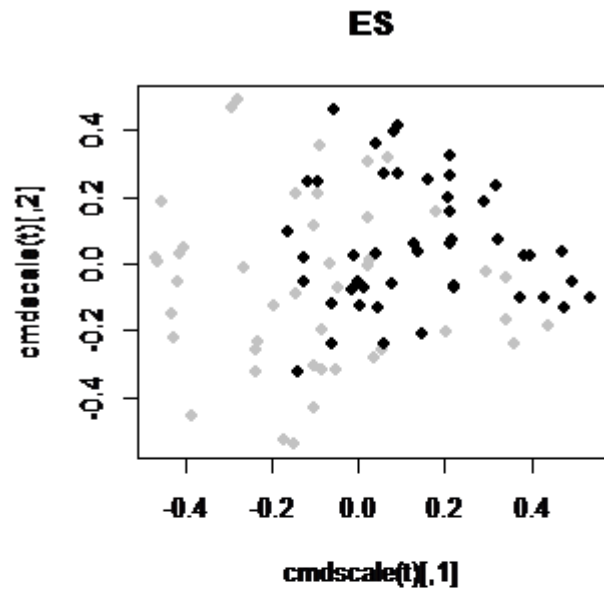
Form-meaning mapping: Germanic



Kriging: Germanic

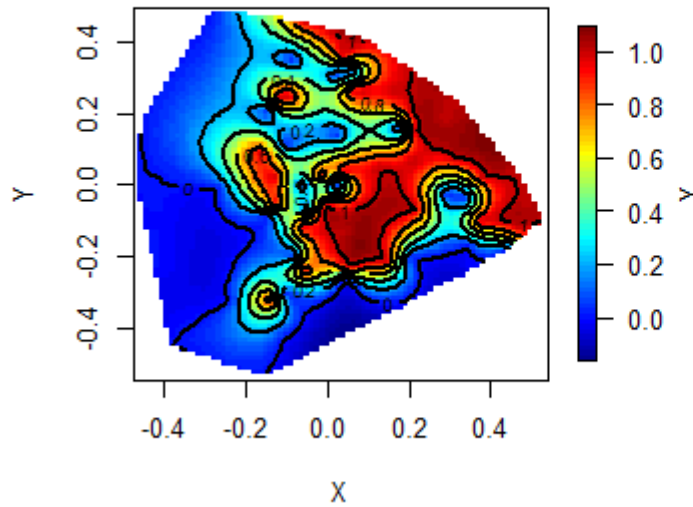


Form-meaning mapping: Romance

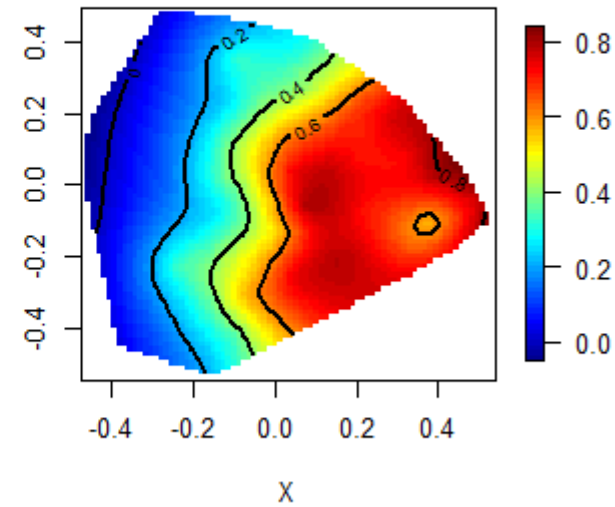


Kriging: Romance

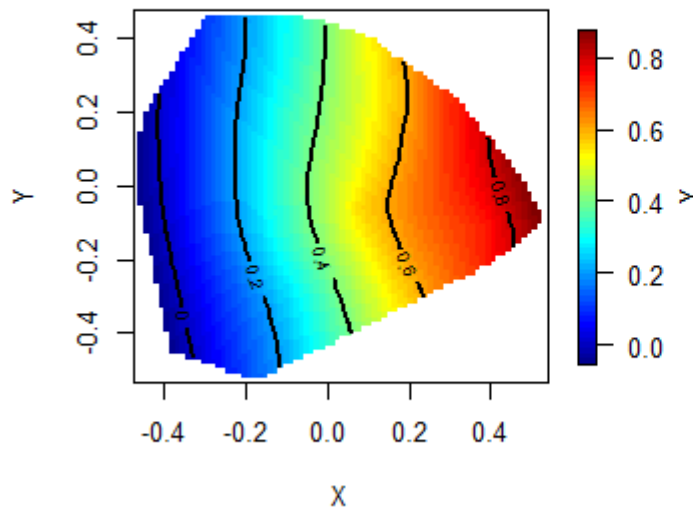
Spanish



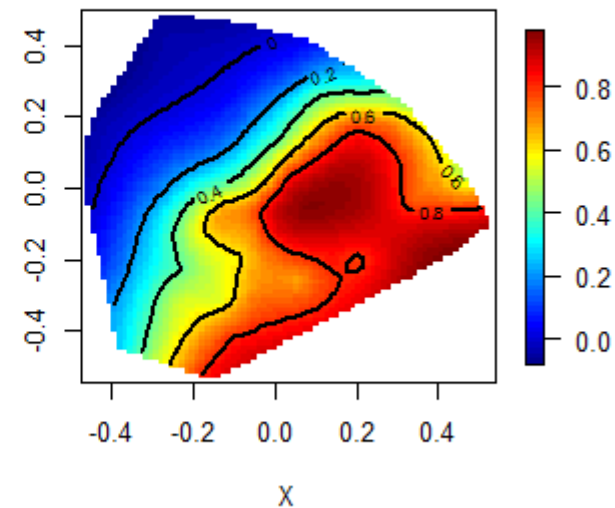
French



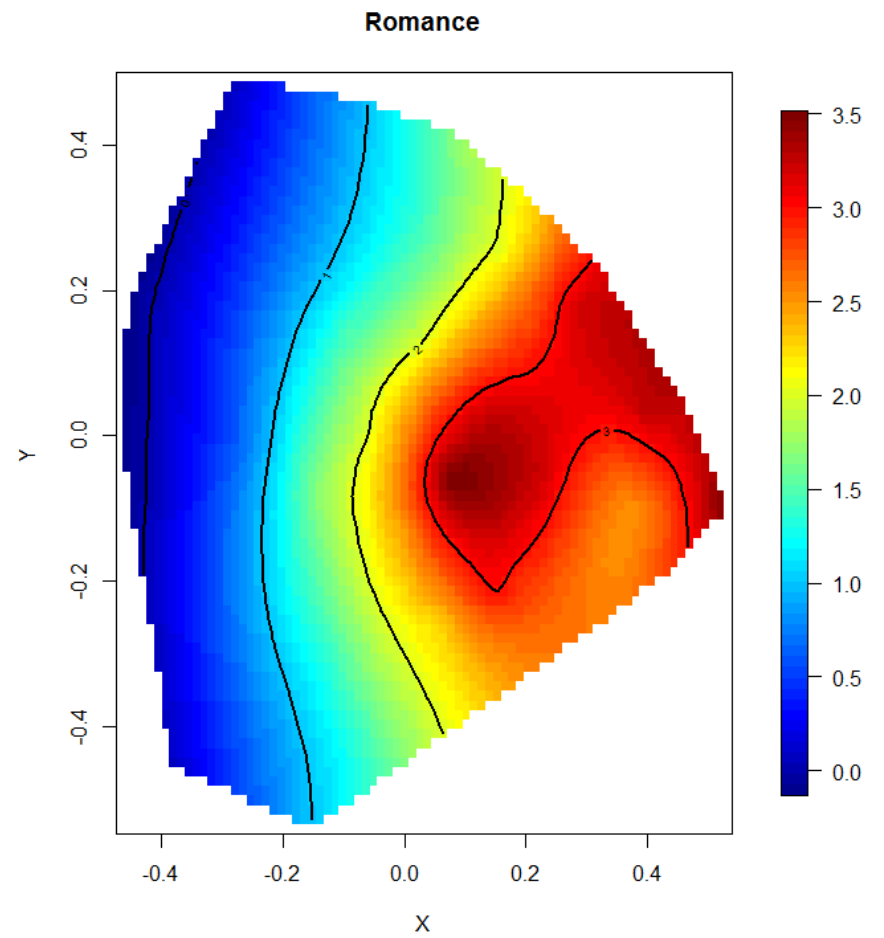
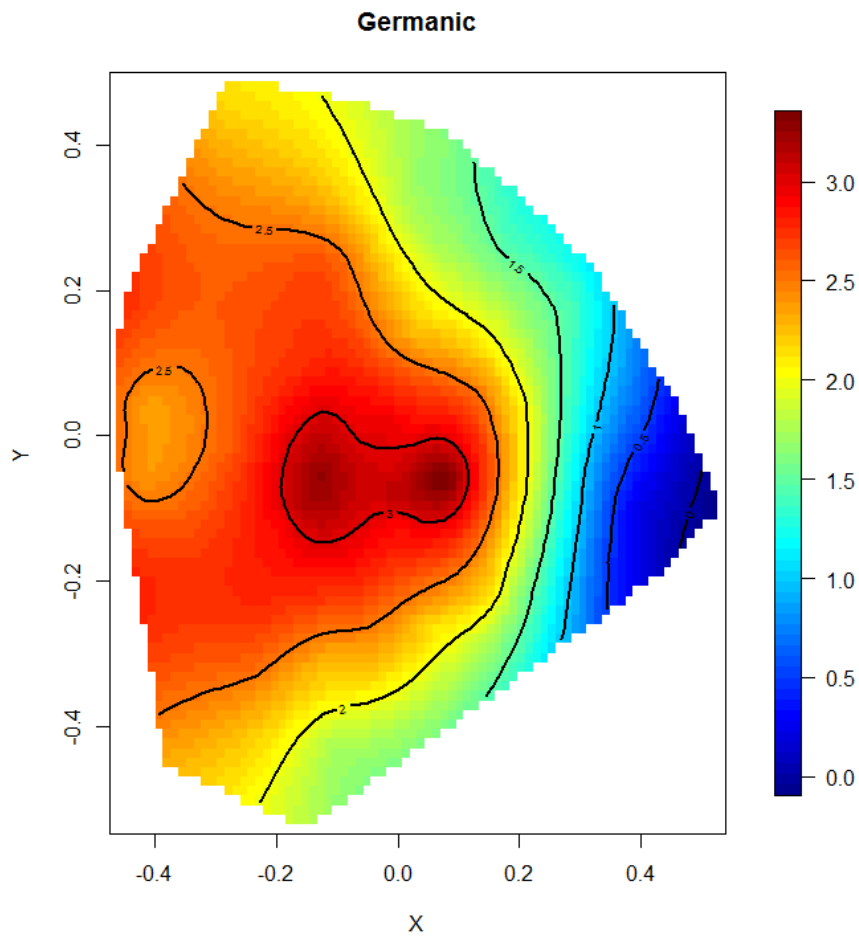
Italian



Portuguese



Germanic vs. Romance



Outline

- Verbs of letting in European languages: semantic peculiarities
- Quantitative comparison of eight LET-verbs in a parallel corpus of film subtitles
 - why film subtitles?
 - data and method
 - results: language space and exemplar space of LET
- **Conclusions**

Conclusions

- Film subtitles are a promising source of data that allows for quantitative comparison of many languages.
- Effects of translationese do not seem to be dramatic, but we need further investigation based on a larger sample, controlled for source languages and genres.
- Germanic and Romance *let*-verbs seem to be developing in the opposite directions, having 'swapped' their most typical meanings in comparison with their historical prototypes. The semantic divergence is higher for the Germanic verbs than for the Romance ones.

