



seit 1558

Friedrich-Schiller-Universität Jena

'Let in Translation': A typological study of the concept of LETTING in a parallel corpus of film subtitles

Natalia Levshina

European Summer School
in Logic, Language and Information
August 6 -17 2012
Opole, Poland

ESSLLI

European Summer School
in Logic, Language and
Information

*Computational Approaches to the Study
of Dialectal and Typological Variation
Workshop*

Outline

1. Constructional typology: Methodological challenges
2. The constructional field of LETTING
3. Data
4. Quantitative analyses
 - conceptual space
 - language space
5. Conclusions

Constructional Typology

- a theoretical and methodological 'update' of lexical typology (see Koptjevskaja-Tamm 2008 about the latter)
- lexical fields as classical examples: kinship terms (Hjelmslev 1957), colour terms (Berlin & Kay 1969), verbs of cooking (Lehrer 1974), aquamotion verbs (Majsak & Rakhilina 2007)
- Constructional Typology: how languages cut the conceptual space with their inventories of constructions – form-meaning pairings at any level of abstraction, including words

Tertia comparationis in semantic typology

- atomistic, or decompositional approaches: comparing elementary semantic components of functionally similar expressions in different languages (e.g. structuralist binary features; Wierzbicka & Goddard's semantic primes)
- holistic approaches: basic frames, or patterns of experience, as input for lexical-typological questionnaires/experiments (e.g. Majid, Boster & Bowerman 2008 – videos with cutting and breaking events). Compare form-meaning mappings in different languages.

Challenges

- the holistic approach looks more attractive (cf. Majsak & Rakhilina 2007)
- however, their inventories of 'frames' are pre-defined and subjective
- we could use non-elicited language data (corpora), as the source of 'frames'
- still, how to compare such contexts from a corpus cross-linguistically?

A proposal

- 'a metric on meaning' (Cysouw 2010):

'The similarity between two meanings can be empirically investigated by looking at their encoding in many different languages. The more similar these encodings, in language after language, the more similar the contexts. So, to investigate the similarity between two contextualized meanings, only judgments about the similarity between expressions within the structure of individual languages are needed.'

- theoretical support: iconicity of form-meaning mapping in usage-based semantic theory
- practical 'carrot': abundance of parallel corpora, which seem to be ideal for this approach

An example

Exemplar 1

EN *...I must be the one to **kill** Harry Potter.*

FR *je dois être celui qui **tuera** Harry Potter.*

NL *ik moet degene zijn die Harry Potter **doodt**.*

Exemplar 2

EN *Ron, **kill** it!*

FR *Ron, **tue-le**!*

NL *Ron, **dood** het.*

An example

	EN	FR	NL
Exemplar 1	<i>kill</i>	<i>tuer</i>	<i>doden</i>
Exemplar 2	<i>kill</i>	<i>tuer</i>	<i>doden</i>

An example

Exemplar 3

EN *the second brother **killed** himself so as to join her*

FR *le deuxième frère se **tua** pour la rejoindre.*

NL *... **berooftde** de tweede broer zich **van het leven**
om bij haar te zijn.*

An example

	EN	FR	NL
Exemplar 1	<i>kill</i>	<i>tuer</i>	<i>doden</i>
Exemplar 2	<i>kill</i>	<i>tuer</i>	<i>doden</i>
Exemplar 3	<i>kill</i> <i>(oneself)</i>	<i>(se) tuer</i>	<i>(zich) van</i> <i>het leven</i> <i>beroven</i>

An example

- if we judged from these 3 examples about the conceptualization of KILLING in these three languages,
 - a) situations 1 and 2 would be more similar than 1 – 3 or 2 – 3, and therefore closer in the **conceptual space**
 - b) EN and FR would be more similar than EN – NL or FR – NL, and therefore closer in the **language space**

Outline

1. Constructional typology: Methodological challenges
2. The constructional field of LETTING
3. Data
4. Quantitative analyses
 - conceptual space
 - language space
5. Conclusions

LET as a Force-Dynamic category

- Causation: the initiator overrides another object's tendency towards rest or motion.

e.g. *The strong wind made the ball roll.*

- Unlike in causation *per se*, in letting the initiator doesn't cause another entity to change its natural tendency. In other words, the initiator refrains to exercise its force-dynamic potential (Talmy 1988).

LET as a Force-Dynamic category

- two main types:
 - cessation of impingement (onset letting)
The plug's coming loose let the water flow from the tank.
 - non-occurrence of impingement (extended letting)
The fan's being broken let the smoke hang still in the chamber.

Wierzbicka 2002 on English *let*

- more specific semantic functions, for instance:
 - *let* of permission (*Please let me go*)
 - *let* of non-prevention (*She let him die*)
 - *let* of tolerance (*Let him do what he wants*)
 - *let* of shared information (*let me know*)
 - *let* of cooperative dialogue and interaction (*Let me conclude by saying...*)
- some functions are less typical of the corresponding verbs in other languages (Russian or German), e.g. the cooperative *let*

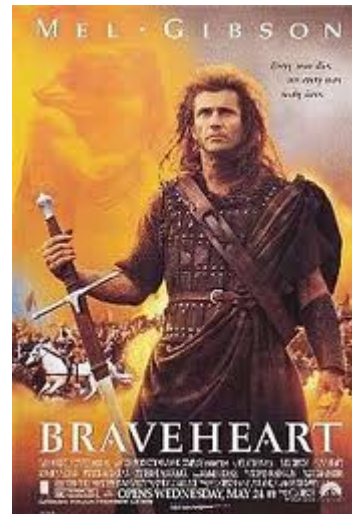
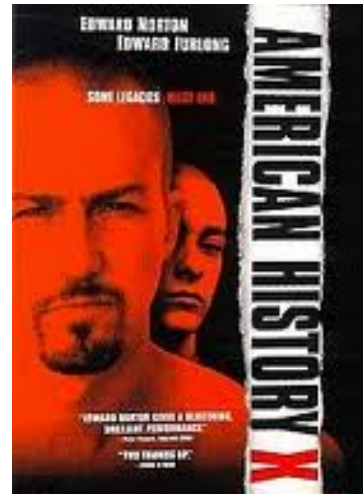
Research questions

- conceptual space:
 - how do languages cut the common conceptual space, which serves as a tertium comparationis (cf. typological semantic maps by Haspelmath 1997; van der Auwera & Plungian 1998; Croft & Poole 2008)?
- language space:
 - which European languages are more similar to one another in their expression of LETTING?

Outline

1. Constructional typology: Methodological challenges
2. The constructional field of LETTING
3. Data
4. Quantitative analyses
 - conceptual space
 - language space
5. Conclusions

Movies



Languages of subtitles



Data sources

- <http://www.subscene.com>
- <http://www.opensubtitles.org>

subscene.com

Inbox (2,156) - natalevs@g x Почта :: Входящие x Mail.Ru: почта, поиск в ин x Грек накатал на своем M x R: Set or Query Graphical P x OPUS - an open source pa x Avatar - Subtitles - Subscer x

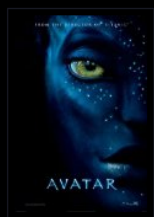
v2.subscene.com/avatar/subtitles-80311.aspx

Sign in | Join

Subscene v.2.3

SUBTITLES: SEARCH UPLOAD ALL SUBTITLES FORUMS

-> Goto the new Subscene! Get subscene news and updates on Twitter!




Avatar
Film Year: 2009

Language filter Upload subtitle for Avatar

Available subtitles - sort by date CD's Author


> Albanian Avatar 2009 TELESYNC H264 AAC-SecretMyth Kingdom-Release	1	Soerensen
> Albanian Avatar.2009.DVDRip [UniVerseHub].cd1.ALB	1	UniVerseHub
> Arabic Avatar 2009PROPER TS XviD-MAXSPEED-1.36 GB	1	gaber
> Arabic Avatar.2009.....480p.TS.XviD-ViSiON 1.45 GB	1	gaber
> Arabic Avatar.DVDScr.xvid-IMAGINE (gaber : ترجمة)	1	BackFire
> Arabic Avatar.2009.1080p.BluRay.X264-AMiABLE-Chan (" (ترجمة احمدchan & تعديل محمد الصادي) ")	1	MohammeD Els
> Arabic Avatar.2009.1080p.BluRay.X264-AMiABLE-gaber (" (ترجمةgaber & تعديل محمد الصادي) ")	1	MohammeD Els
> Arabic Avatar.2009.1080p.BluRay.X264-AMiABLE-Same7 (" (ترجمة SaMeH RaHeeF) محمد الصادي)	1	MohammeD Els
> Arabic Avatar.2009.1080p.BluRay.X264-AMiABLE-Tarek (" (ترجمة طارق عبدالهادي) محمد الصادي)	1	MohammeD Els
> Arabic Avatar.2009.1080p.BluRay.X264-AMiABLE-Don3essa (" (ترجمة أبو عيسى) و خالد الي) محمد الصادي)	1	MohammeD Els
> Arabic Avatar.2009.720p.BluRay.x264-ELITE (By SaMeH RaHeeF, Sync by T.D.D and me :p, also -AMIA	1	foryoux
> Arabic Avatar.2009.720p.BluRay.DTS.x264-HiDt	1	MusaedQ8
> Arabic Avatar 2009 DVDSCR DivXNL-Team ((696MB / Abu Essa ترجمة / Alwahsh50: (التعديل الوقتي))	1	alwahsh50



Schlecker
Unsere Drogeriequalität
Über 4000 Drogerieartikel
24h-Servicecenter
www.schlecker.de

Home Shopping
Günstig online einkaufen im
Schlecker Drogeriemarkt


[SCHLECKER.de](#)



STRATO

Power Hosting
Jetzt 6 Monate lang schon ab
0 € bei STRATO

[www.strato.de](#)



Baby-Markt
Der Baby Fachmarkt für
Babyausstattung zum besten
Preis

[baby-markt.de](#)

EN 12:38 03/08/2012

Example

Number

1

Start time

End time

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

*Sooner or later though,
you always have to wake up.*

Text

2

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

In cryo you don't dream at all.

3

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

Doesn't feel like 6 years.

4

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

More like after tequila

and ass kicking.

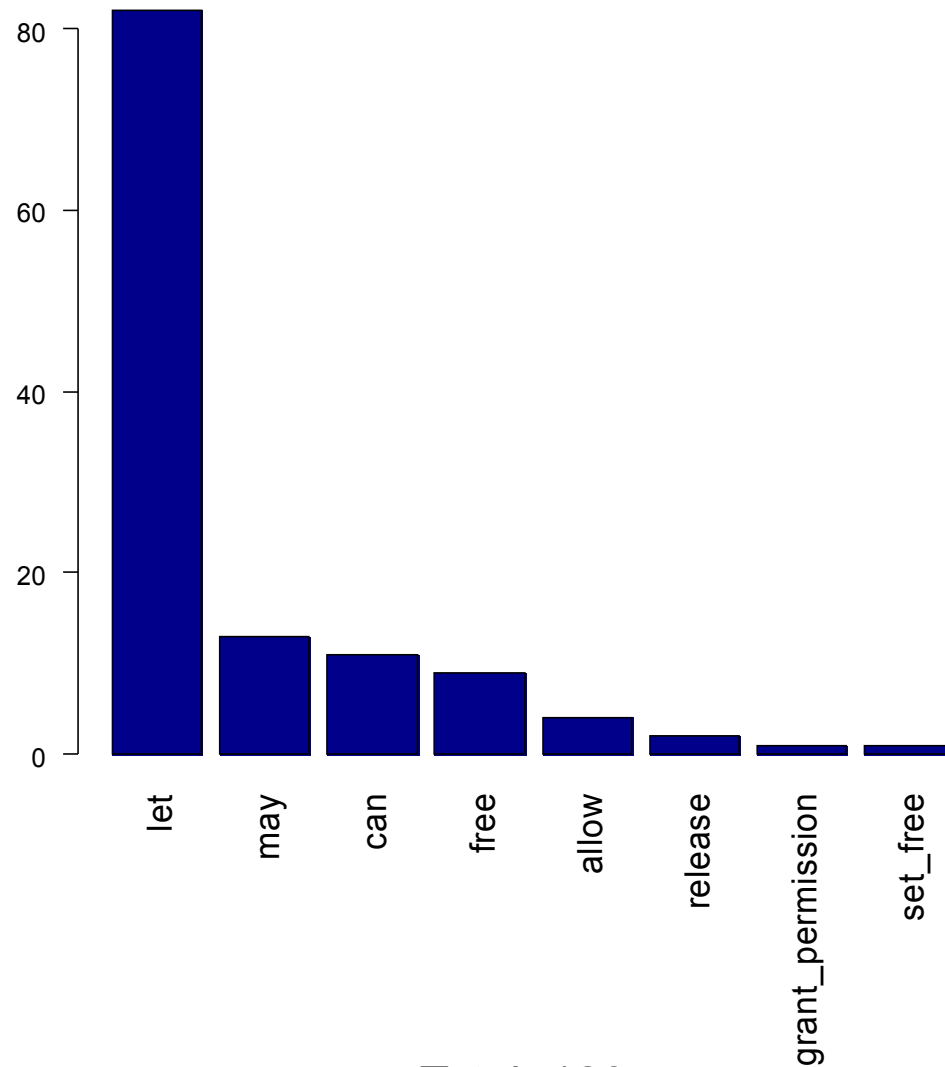
Avatar.2009.720p.TS.XviD-
ViSiON

SubRip format (.srt)

Constructional field of LET

- onomasiological approach (meaning - form)
- all verbs and verb expressions related to LETTING in WordNet and FrameNet
 - WordNet: synset of *let*
 - FrameNet: frames `Make_Possible_to_Do`, `Grant_Permission`, `Permitting`, `Releasing`
e.g. `Permitting` LUs: *accept.v*, *allow.v*, *entitle.v*,
permit.v, *permit.n*, *sanction.v*
- also permissive *can* and *may* (cf. Talmy 2000)

English constructions



Total: 123

Outline

1. Constructional typology: Methodological challenges
2. The constructional field of LETTING
3. Data
4. Quantitative analyses
 - conceptual space
 - language space
5. Conclusions

Distances between the exemplars

- A Python script that computes Gower's similarity metric between pairs of exemplars
- for a pair of exemplars i and j and k languages,

$$s_{ij} = \frac{\sum w_{ijk} \cdot s_{ijk}}{\sum w_{ijk}}$$

where w is 1 or 0 depending on whether the comparison is valid (0 if the translation is missing);

s_{ijk} can be 0 (different constructions), 1 (identical constructions) or 0.5 (partly identical).

- distances: $d_{ij} = 1 - s_{ij}$

An example

EN	DE	NL	SV	ES	FR	IT	PT	BG	PL	RU	SL
<i>let</i>	<i>lassen</i>	<i>laten</i>	<i>låta</i>	<i>dejar</i>	<i>laisser</i>	<i>lasciare</i>	<i>deixar</i>	<i>puskam</i>	<i>pozwalac</i>	<i>pozvolit'</i>	<i>dovoliti</i>
<i>allow</i>	<i>zulassen</i>	<i>laten</i>	<i>tillåta</i>	<i>permitar</i>	<i>permettre</i>	<i>permettere</i>	<i>deixar</i>	<i>pozvoljavam</i>	<i>dopusciac</i>	<i>dopustit'</i>	<i>dovoliti</i>
...

take two first exemplars (rows)

- 3 full overlaps (NL, PT, SL)

- 2 partial overlaps (DE, SV)

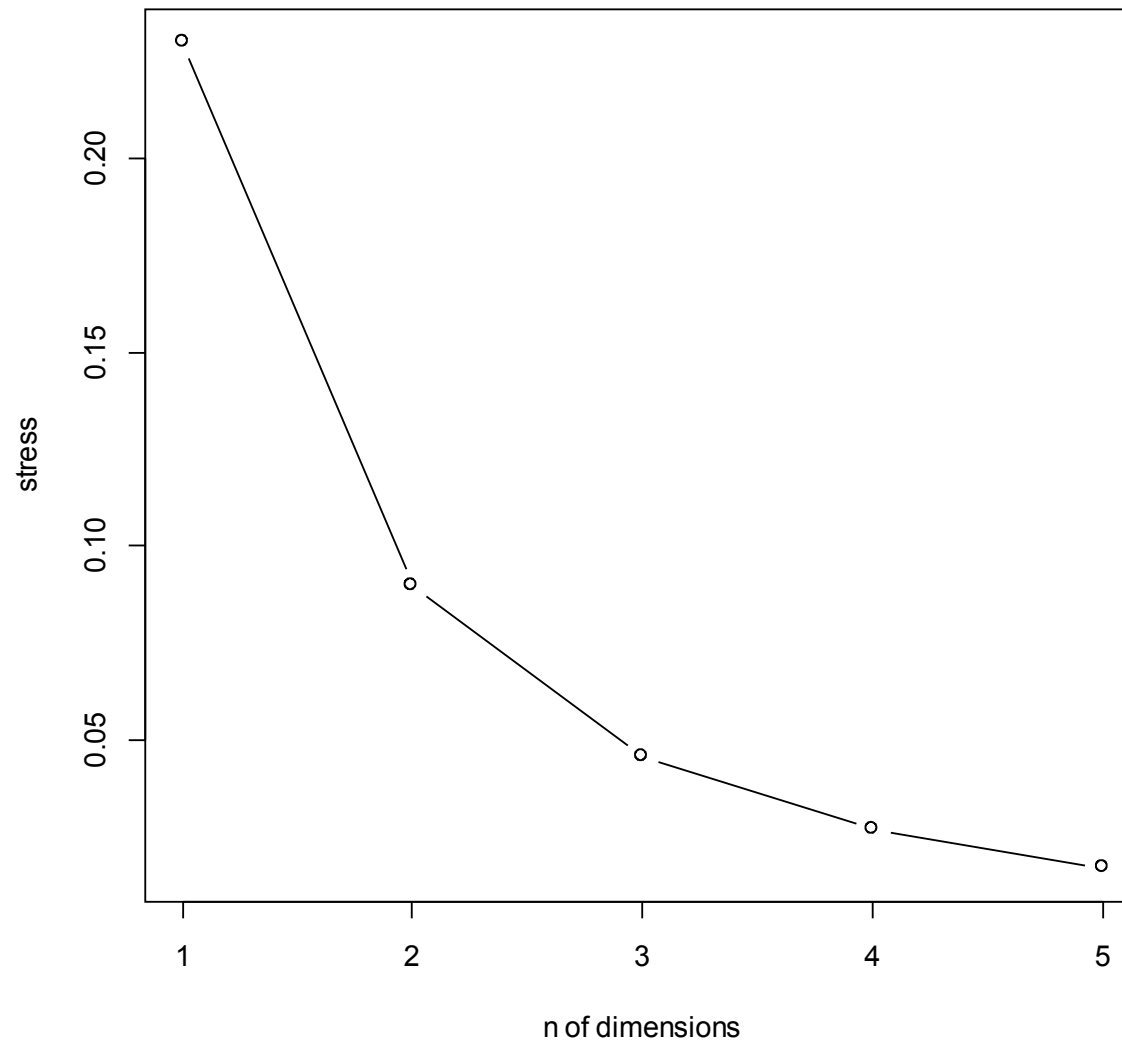
$$d = 1 - (3 + 2 \cdot 0.5) / 12 \approx 0.67$$

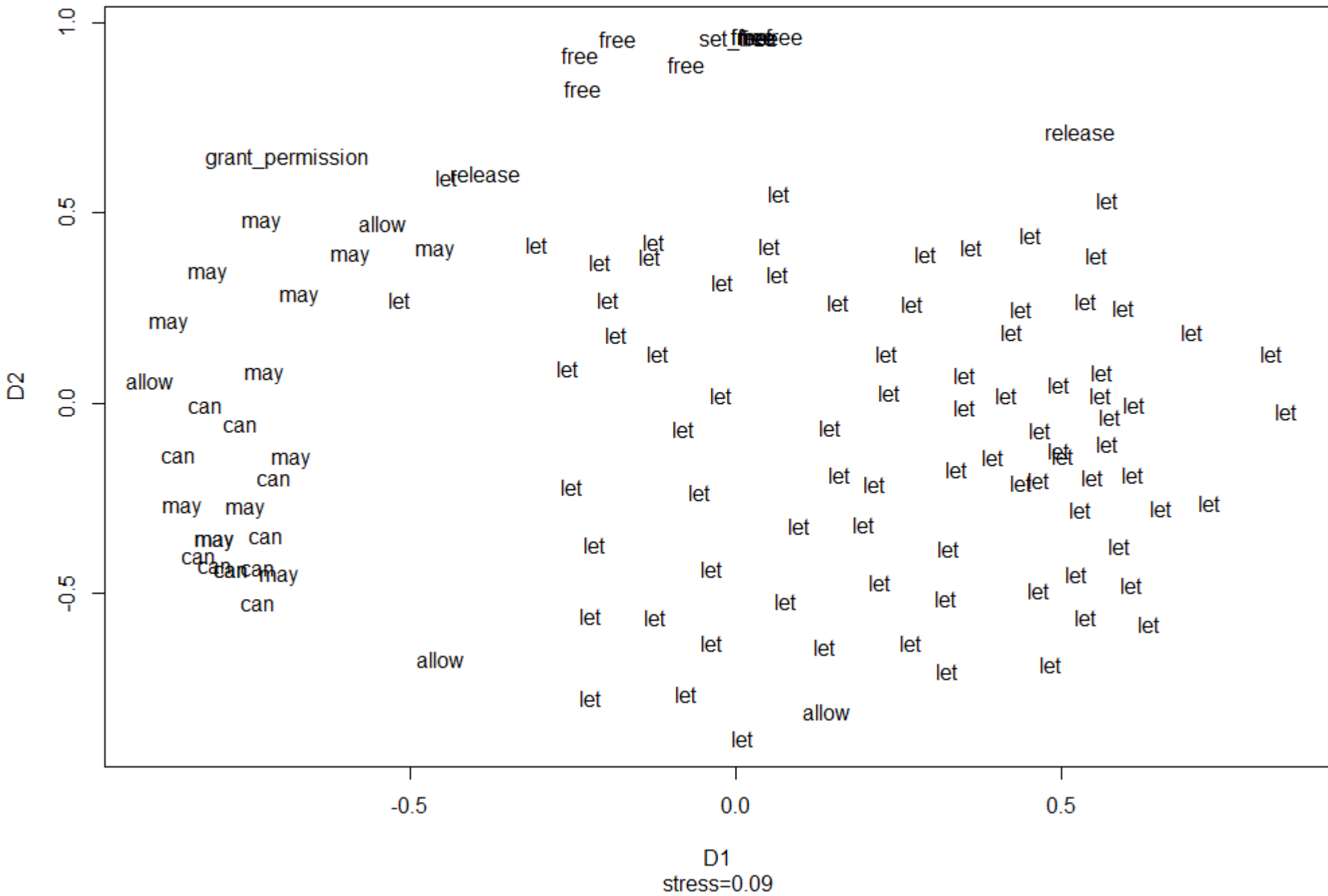
Multidimensional Scaling

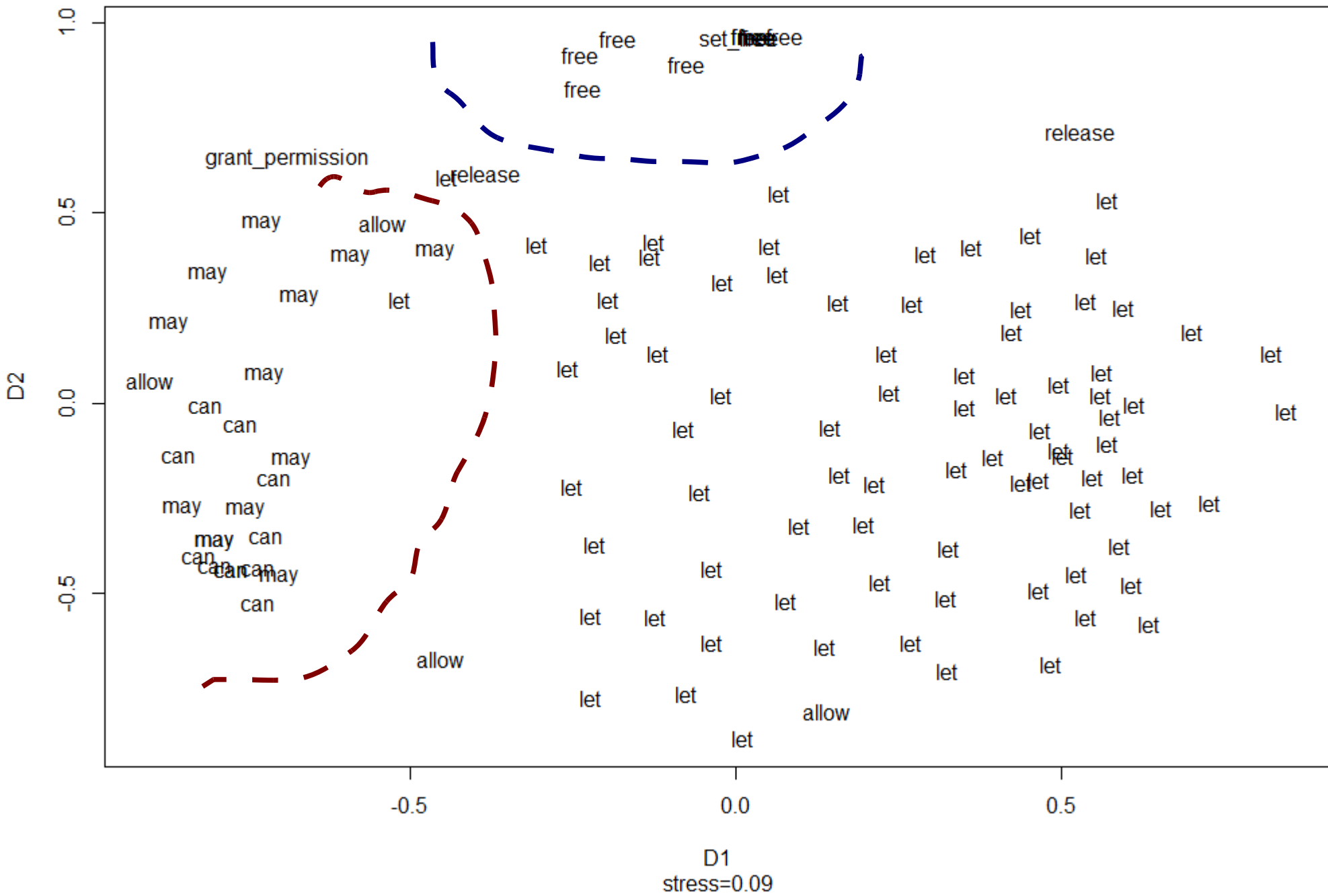
- the matrix of distances serves as input for MDS
- MDS is a dimensionality reduction technique that represents distances between objects in a low-dimensional space
- `smacof` library in R (de Leeuw & Mair 2009)

Choosing the number of dimensions

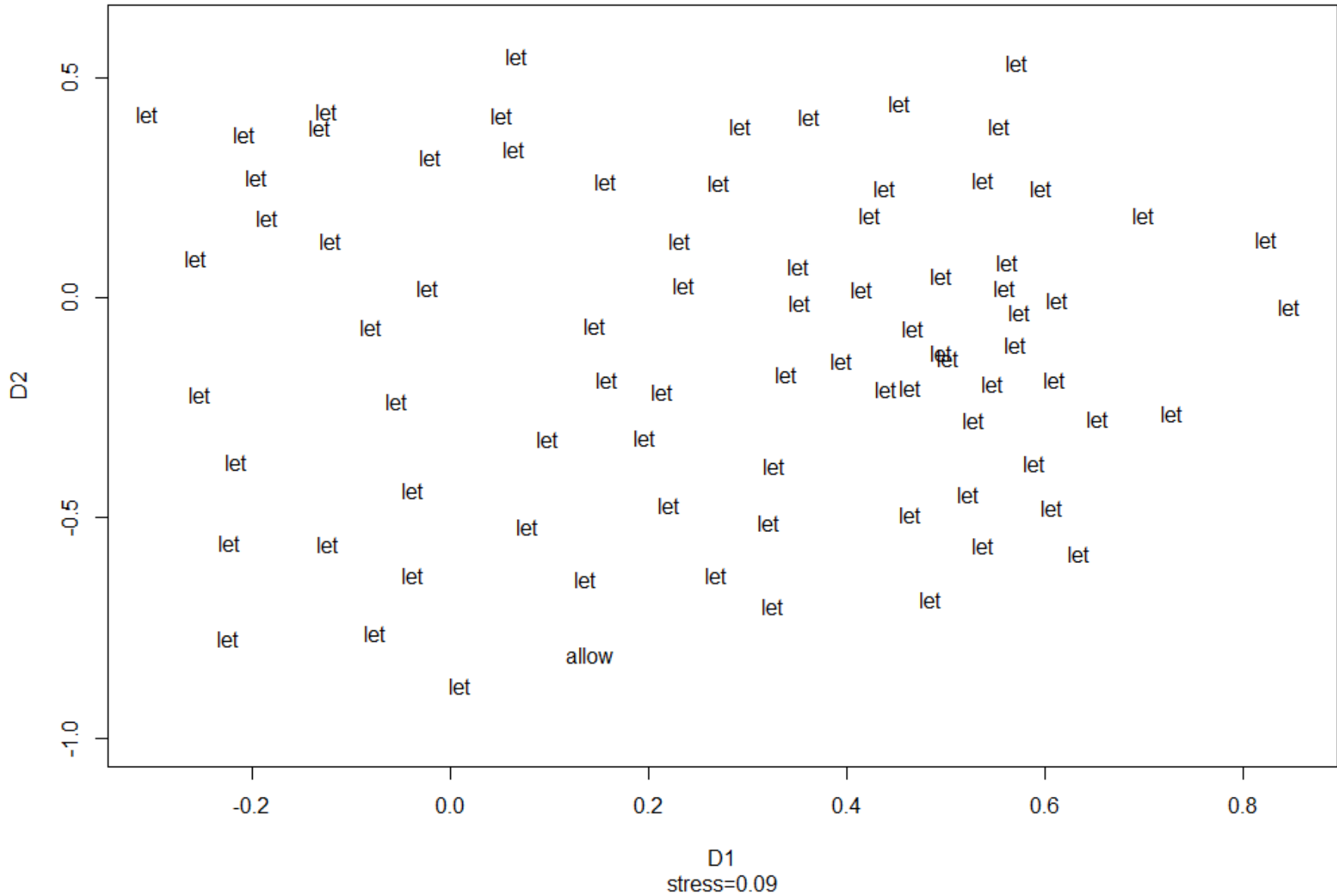
scree plot of n-dimensional MDS solutions



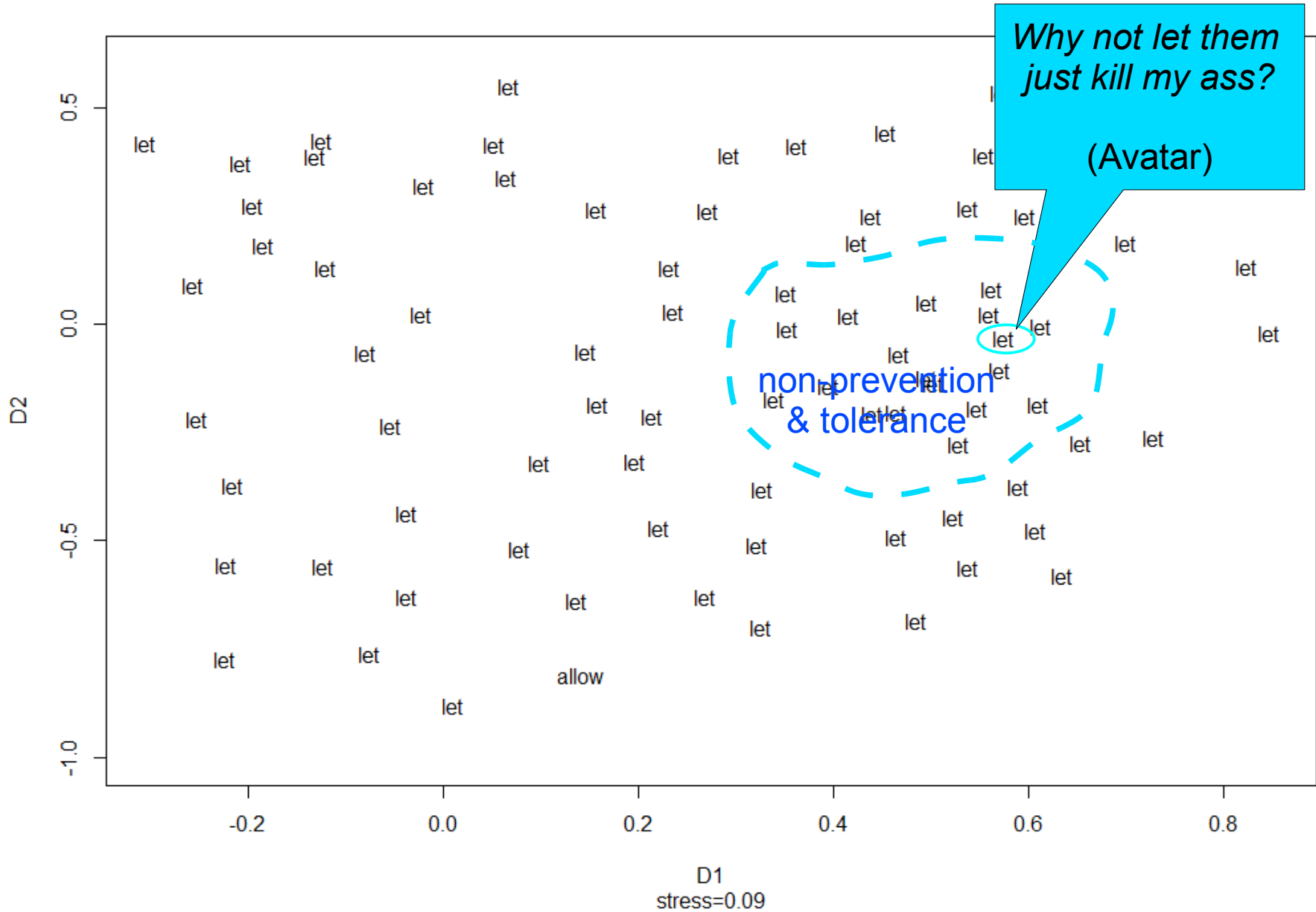




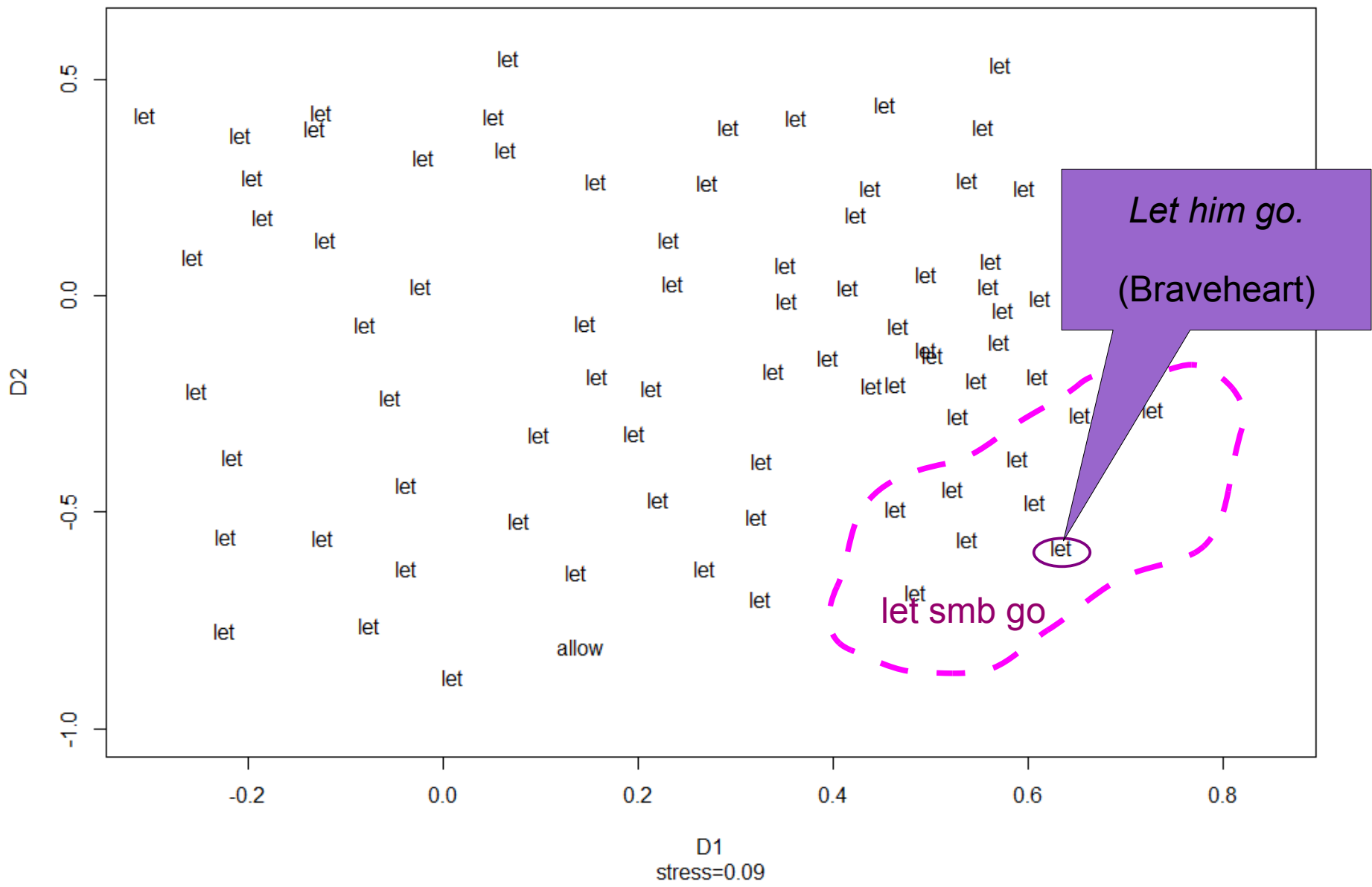
Zoom-in on *let*



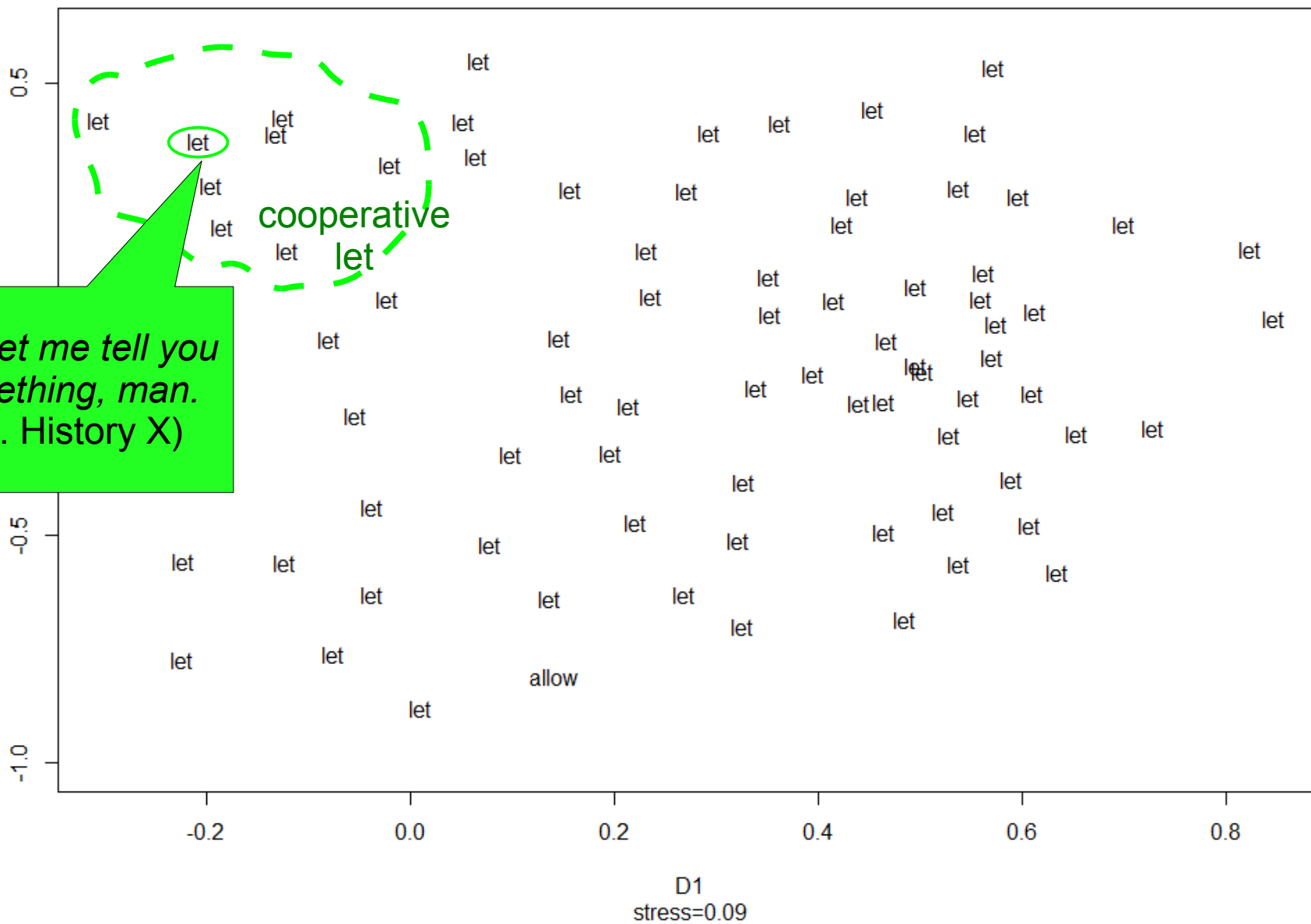
Polysemy of *let*



Polysemy of *let*

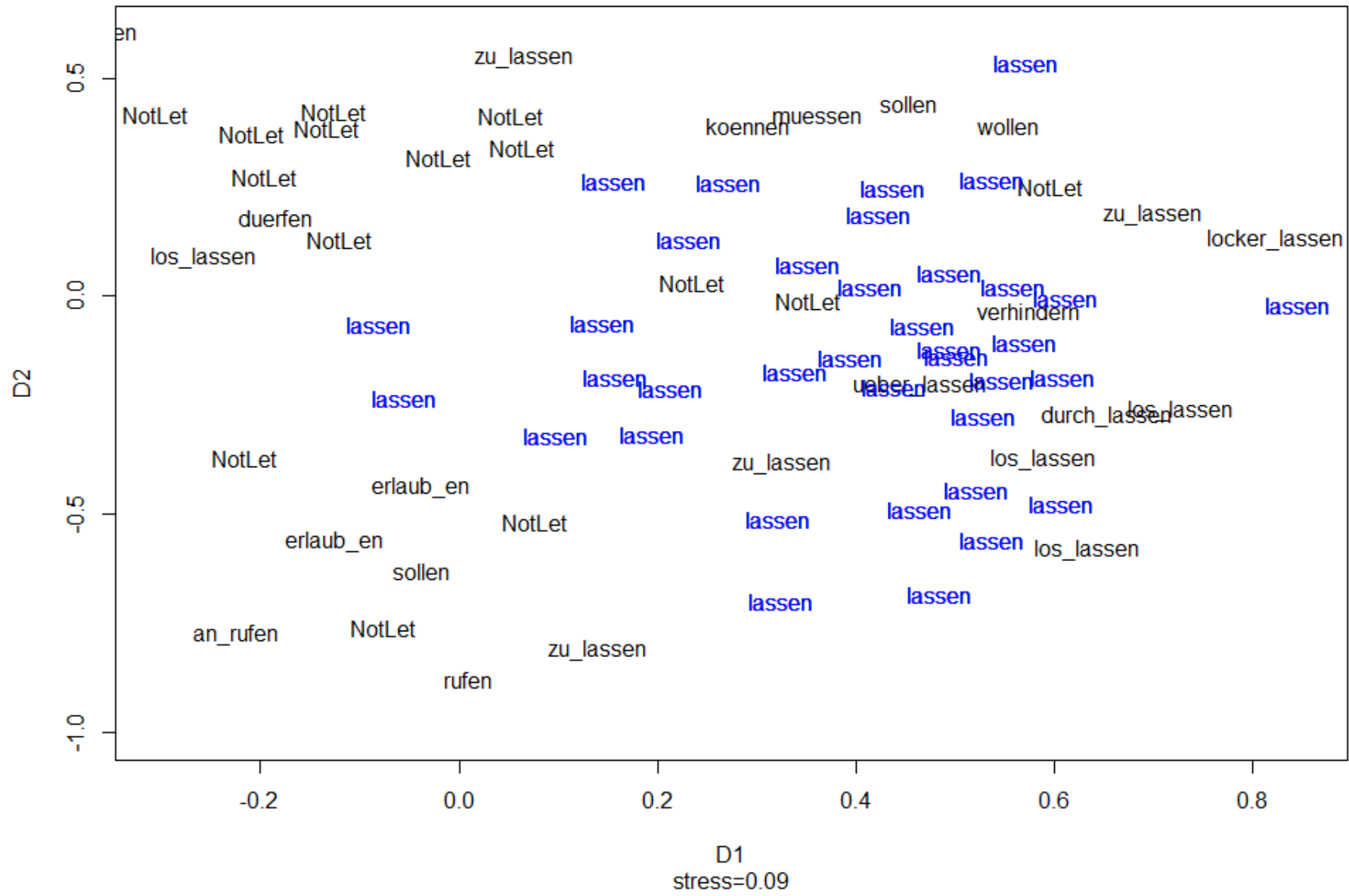


Polysemy of *let*

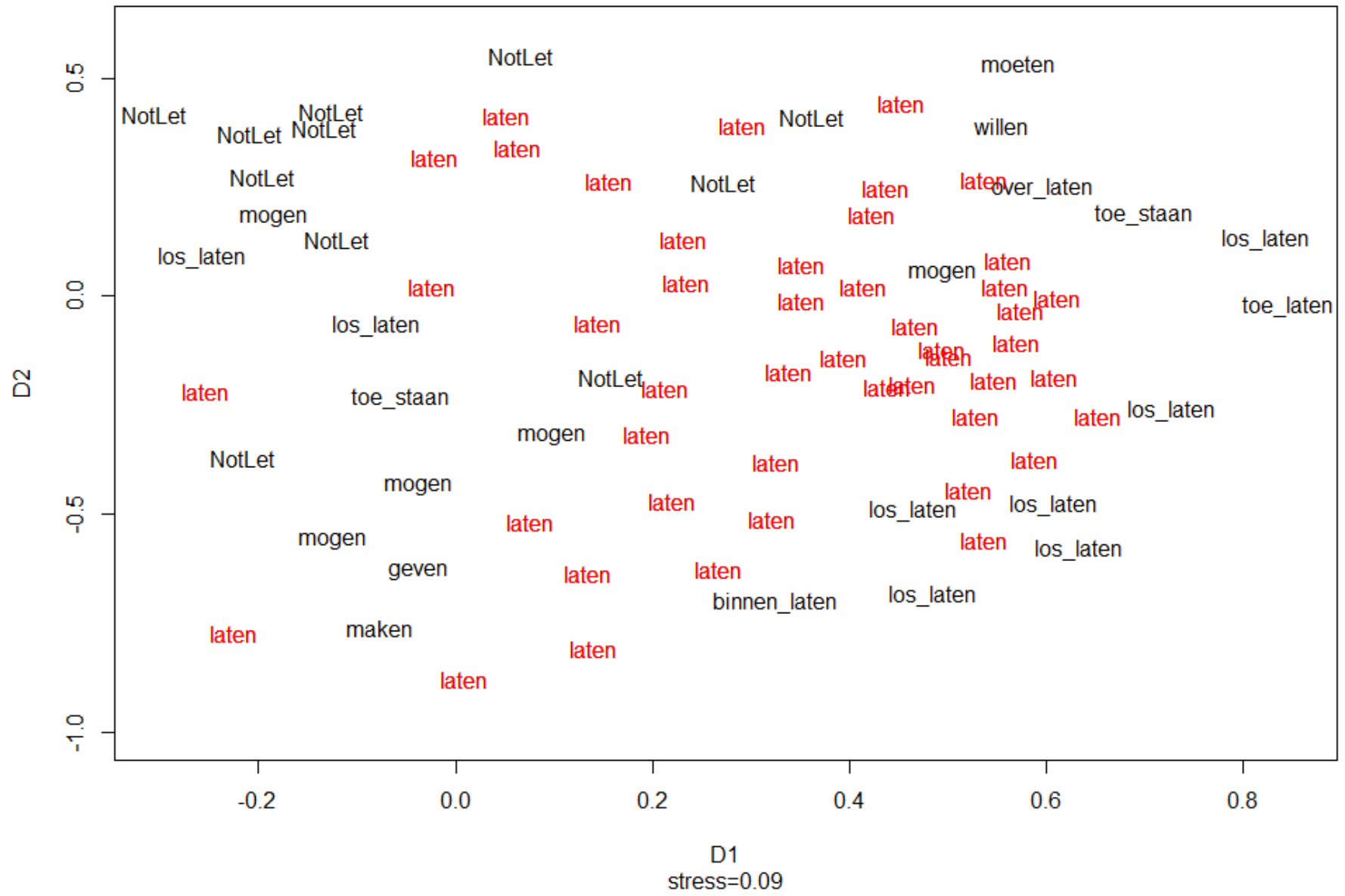


Well, let me tell you something, man. (Am. History X)

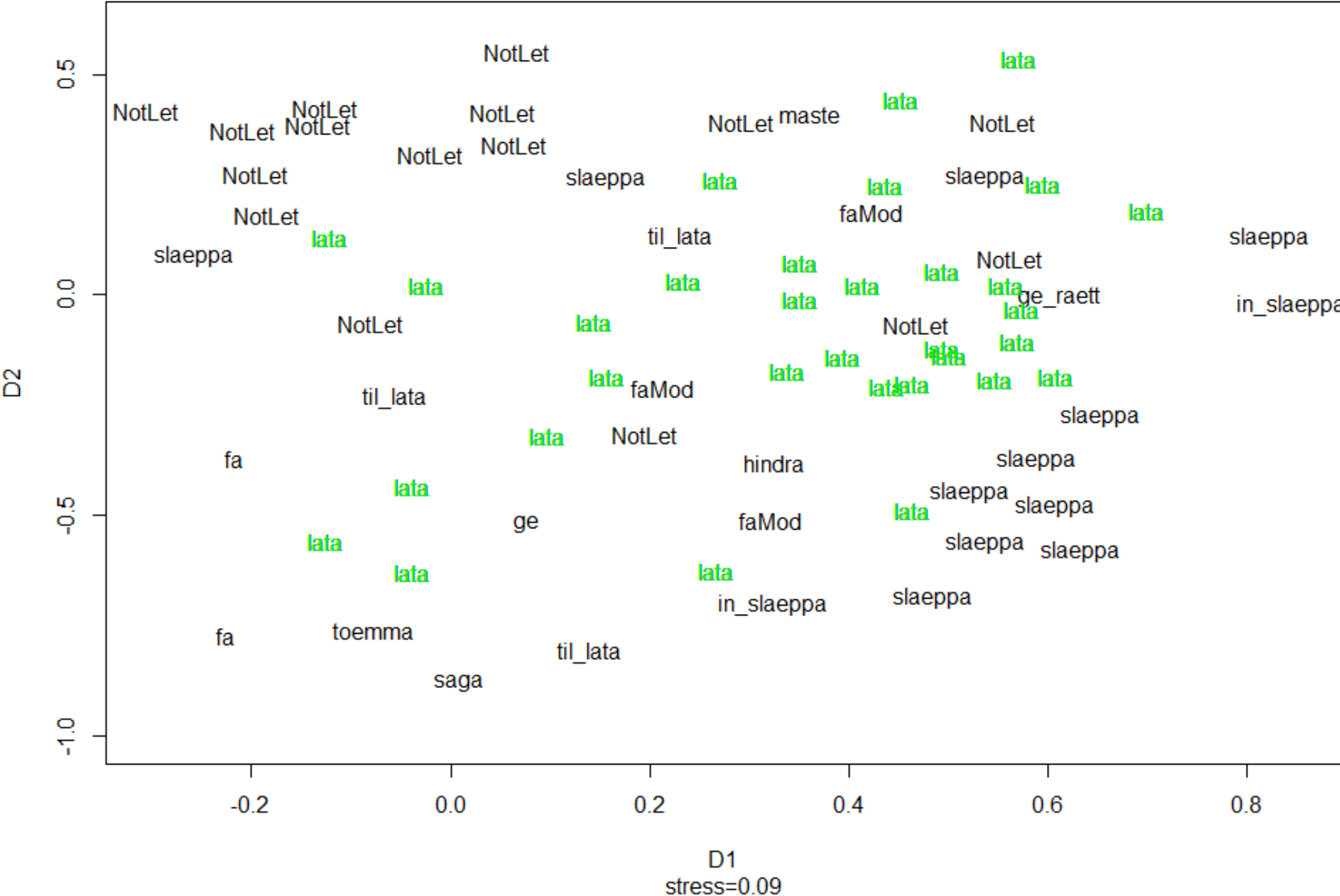
German



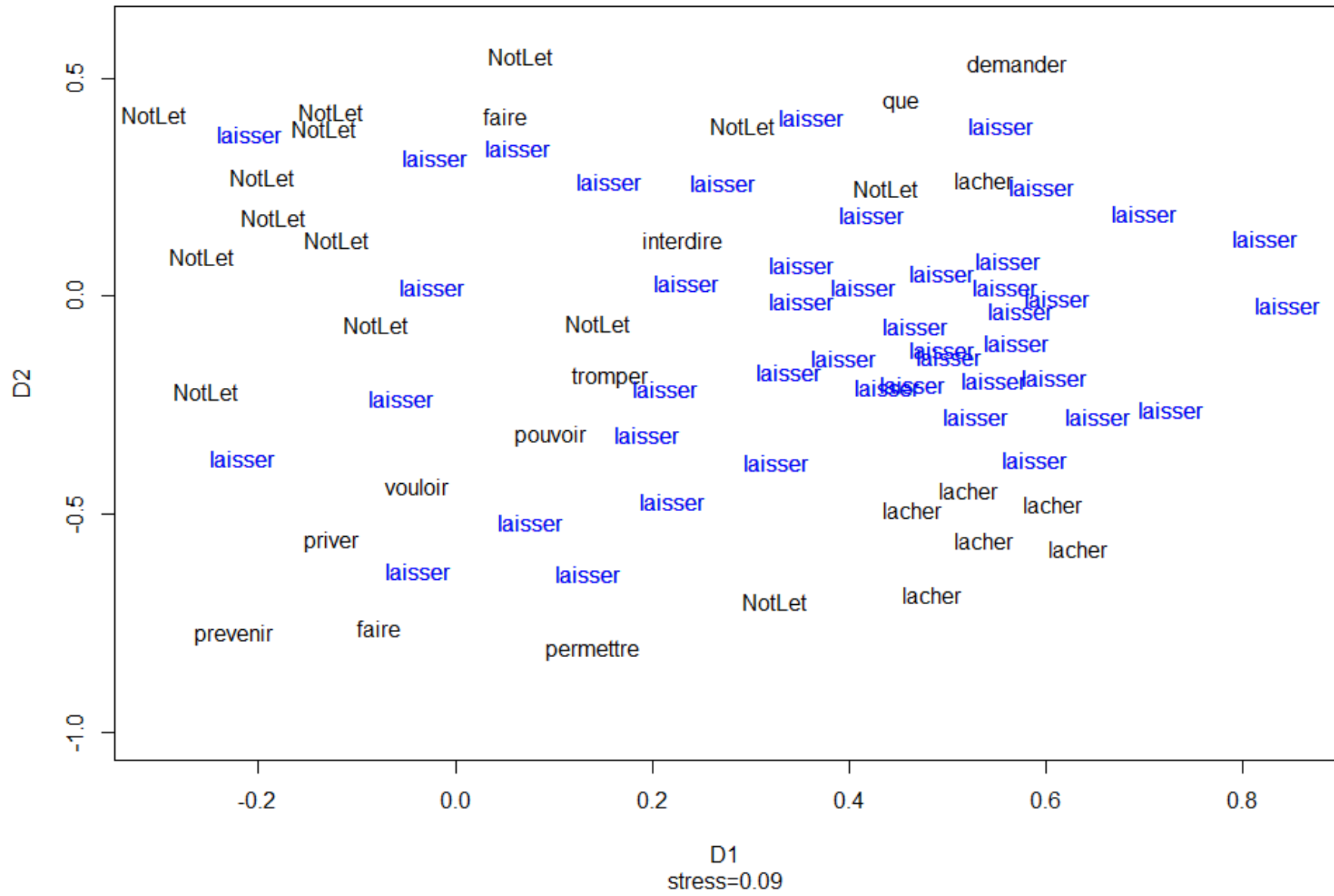
Dutch



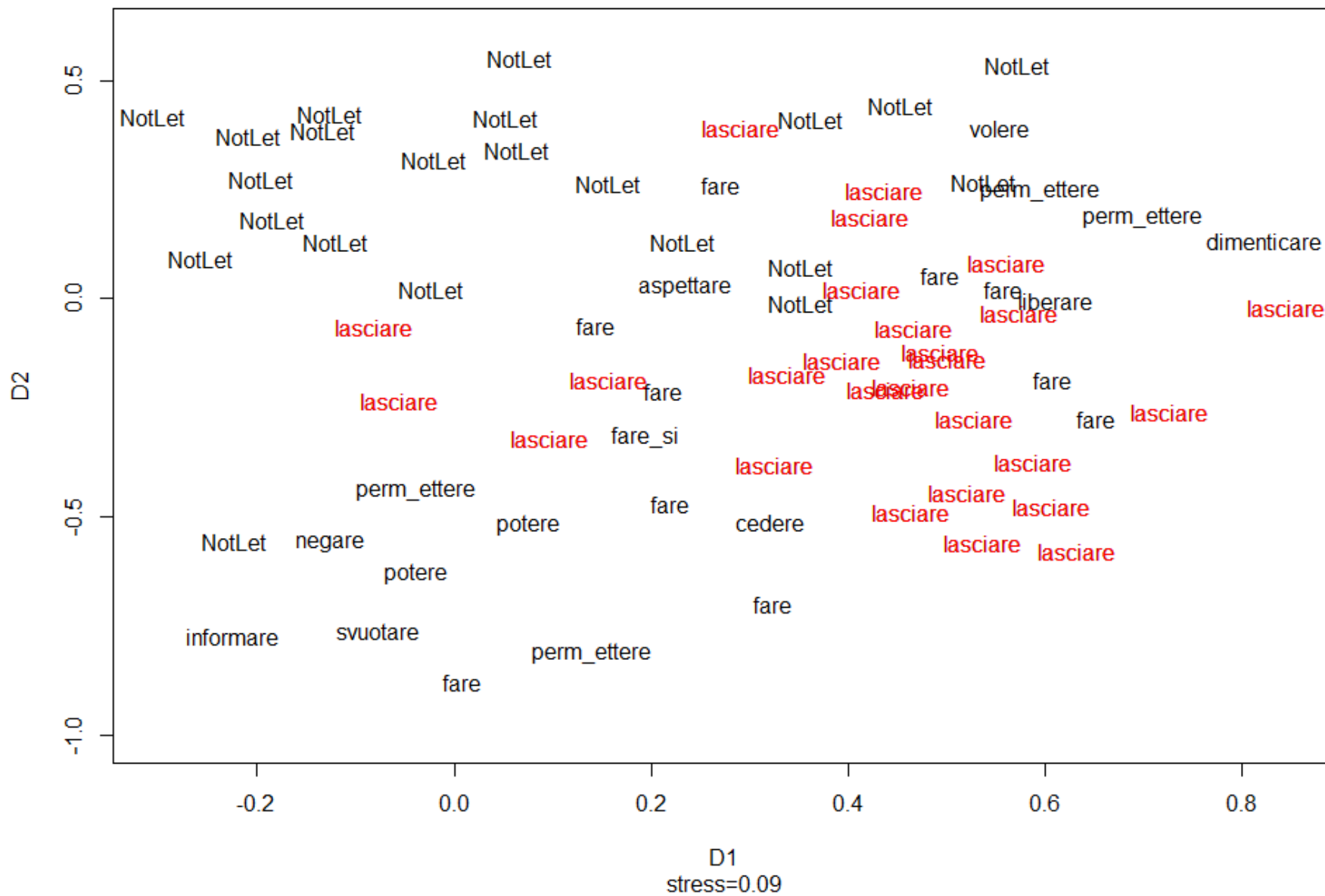
Swedish



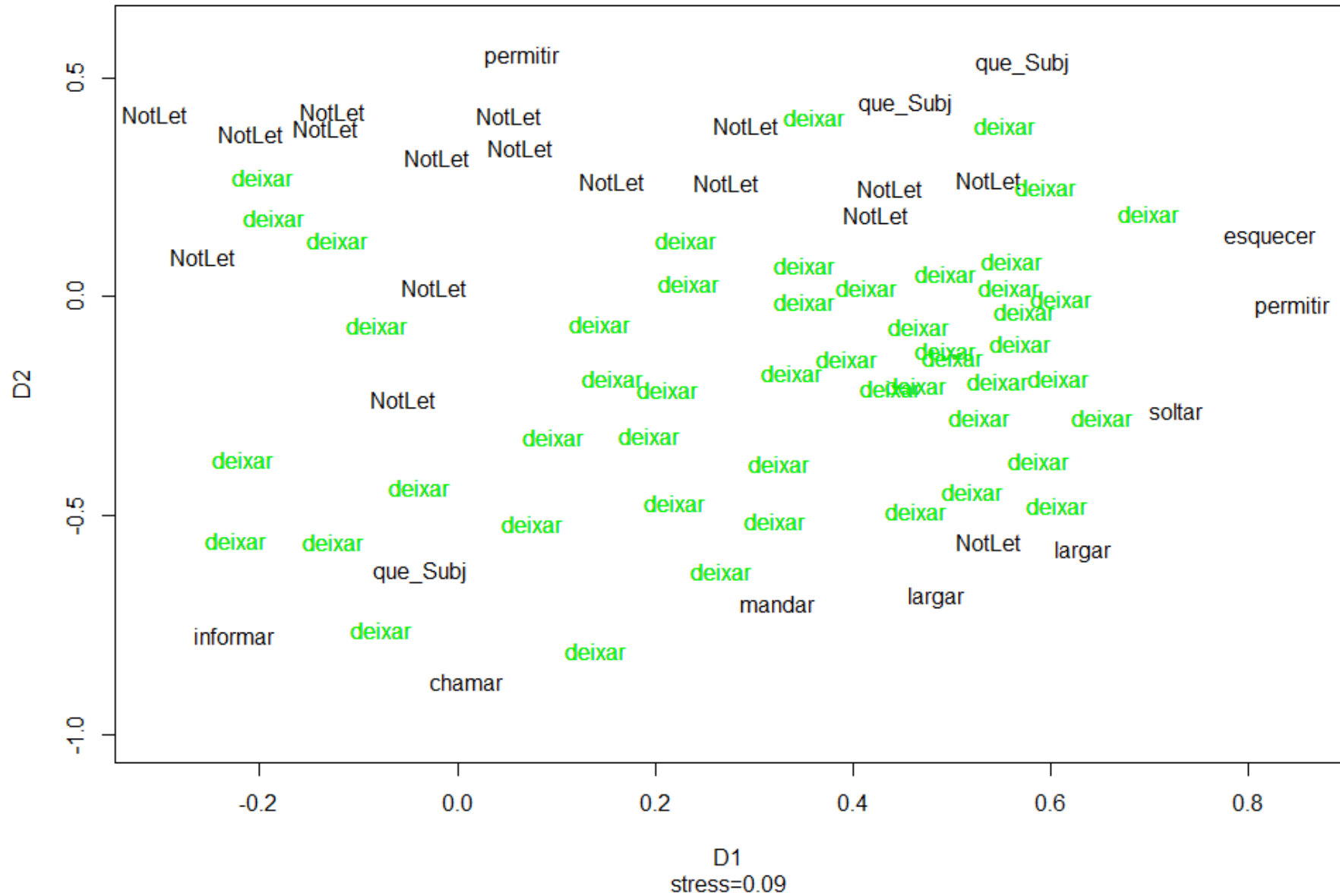
French



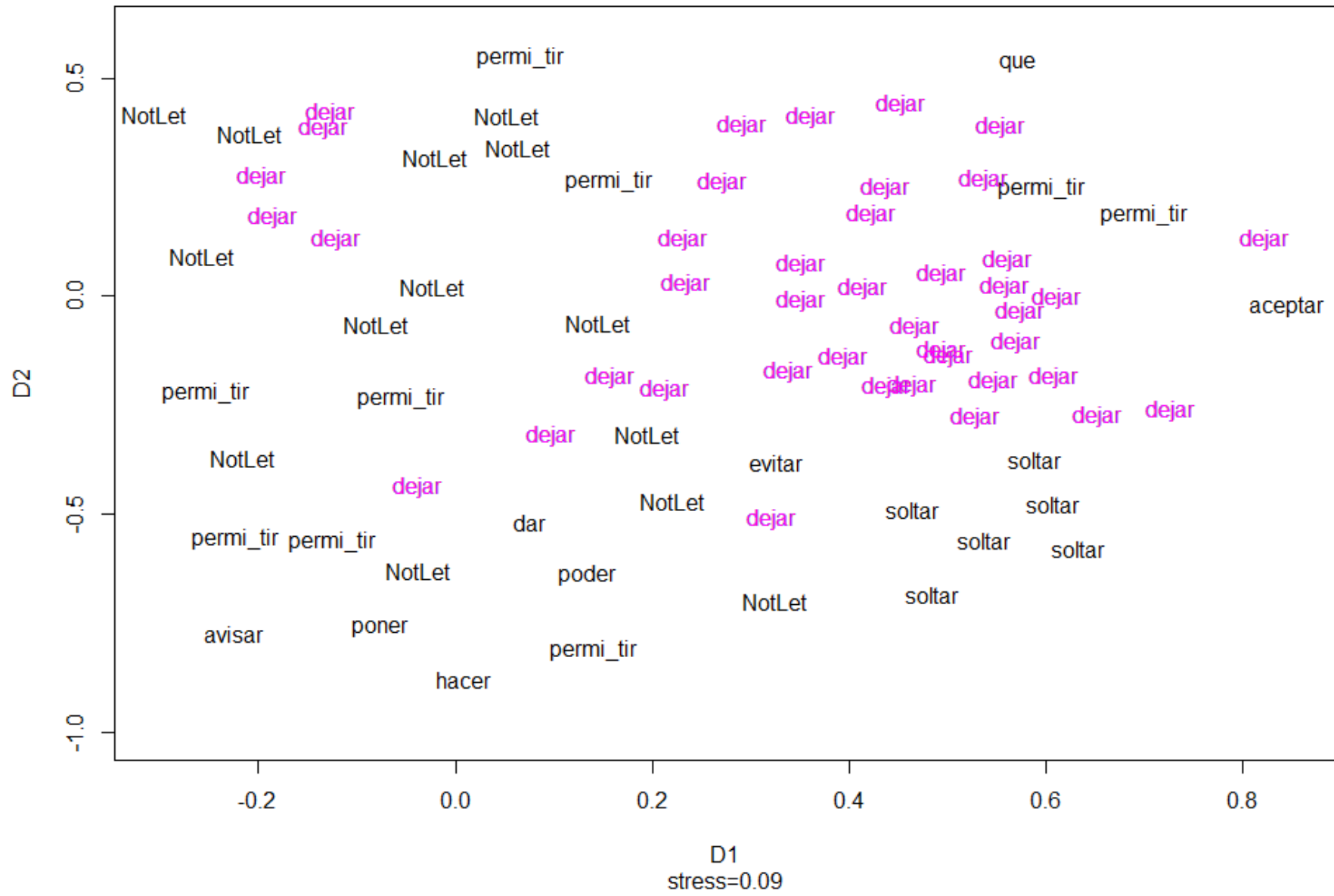
Italian



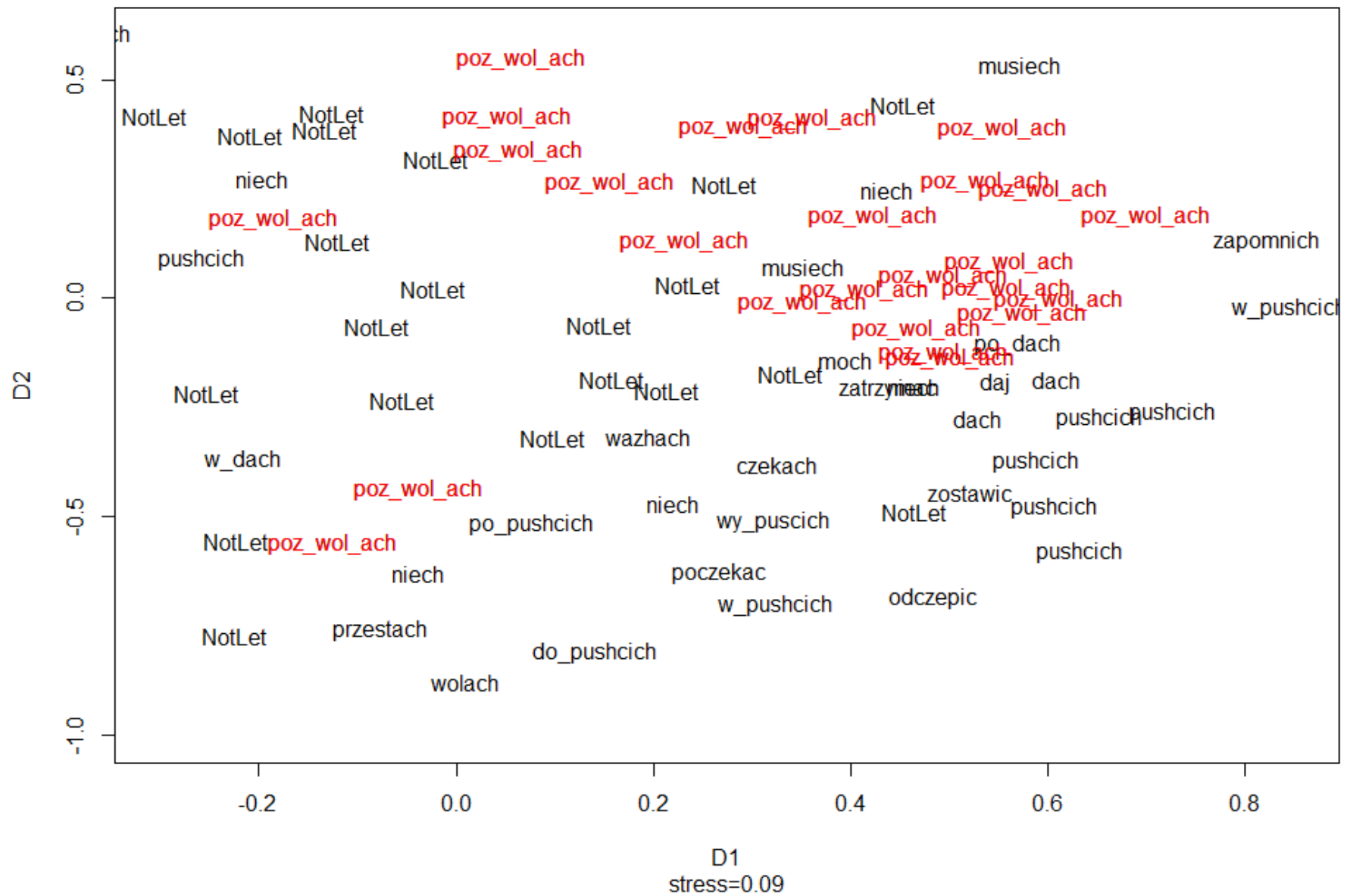
Portuguese



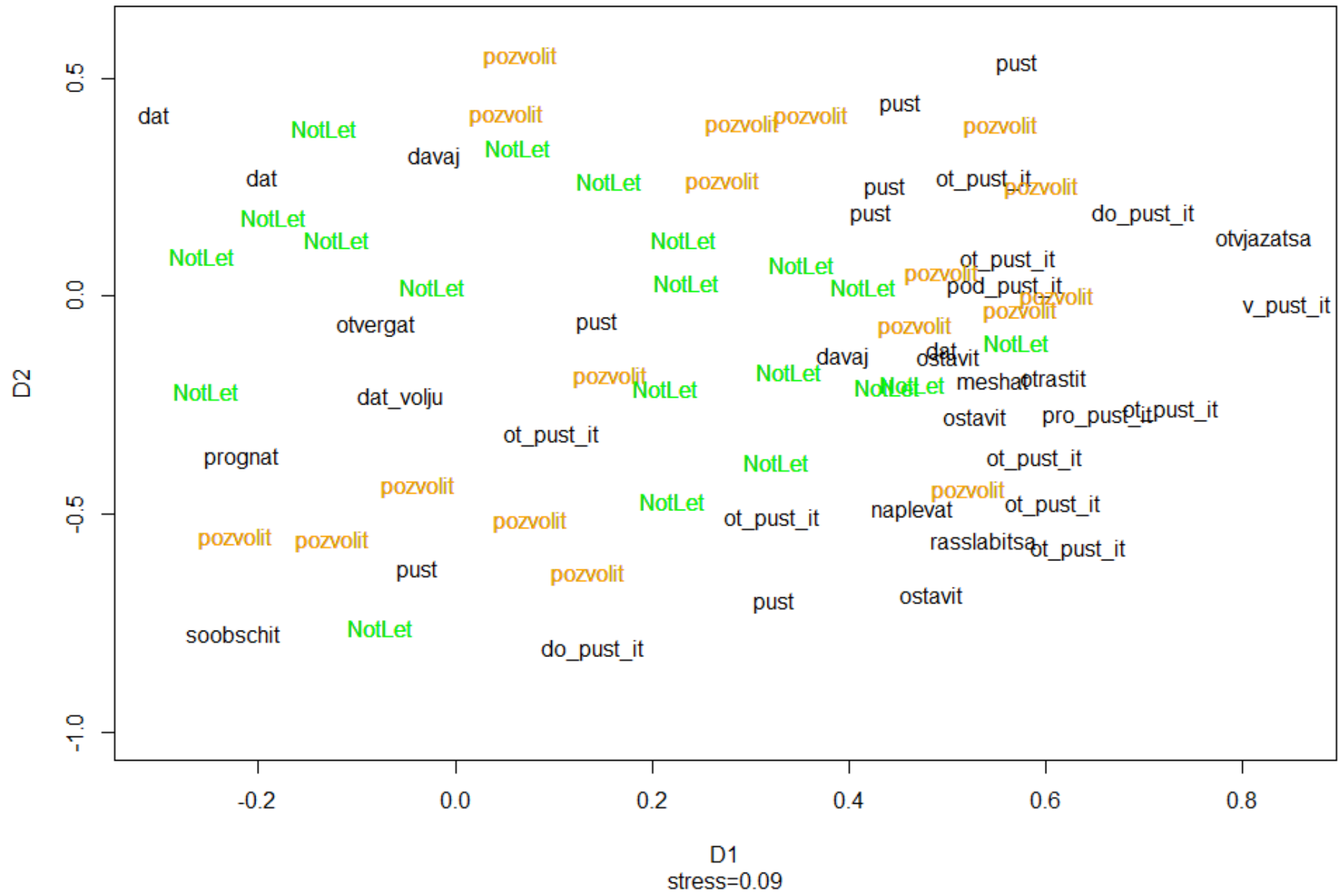
Spanish



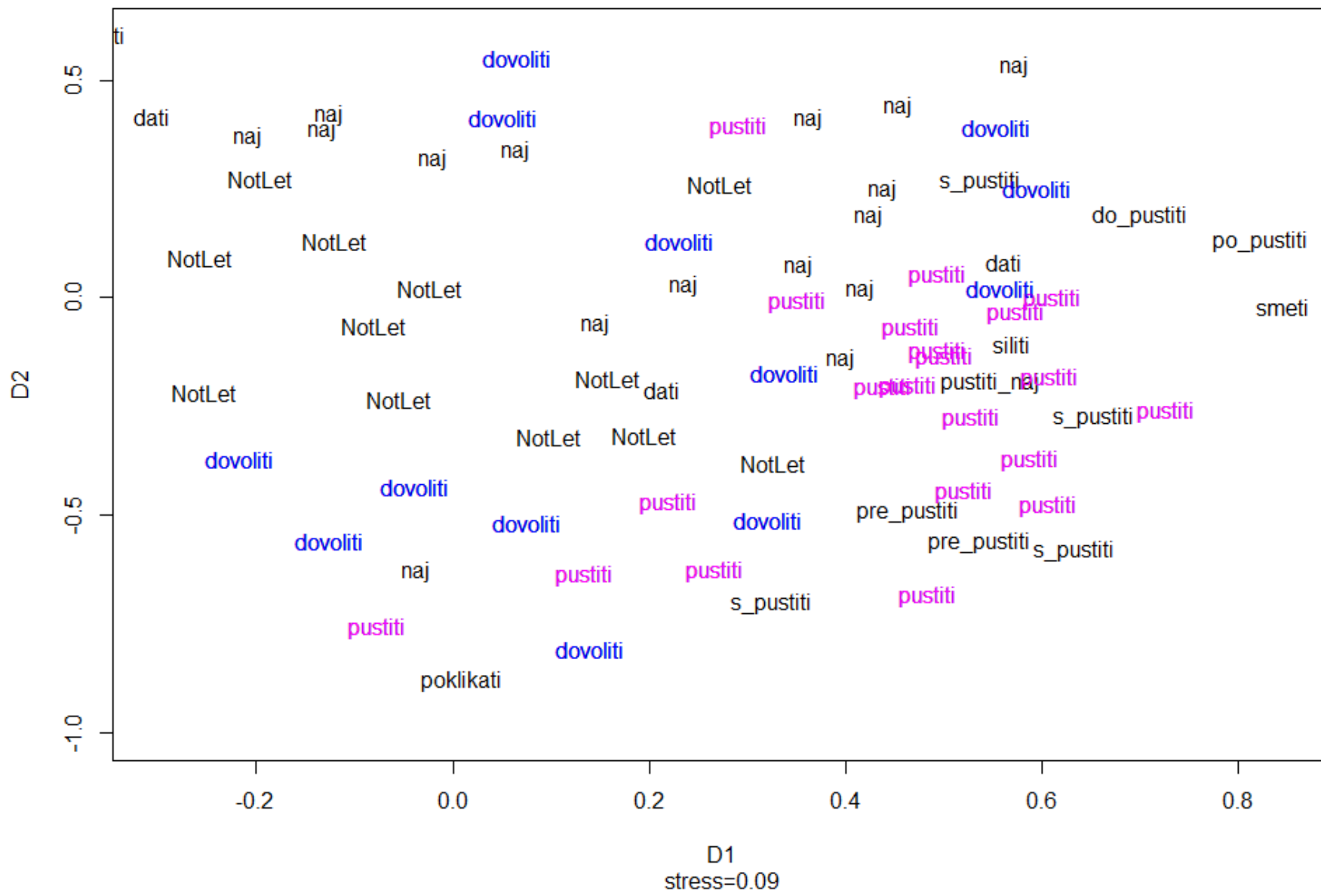
Polish



Russian



Slovene



Interim conclusions (1)

- the most coherently expressed sense is letting as non-interference (the prototype?)
- some senses, e.g. cooperative letting, are rarely expressed cross-linguistically as force-dynamic expressions

Interim conclusions (2)

- the languages cut the space in different ways
- the space fragmentation increases from West to East
- the semantic area of analytic causative expressions (two verbs for cause and effect) also decreases from West to East

cf. Bally – Wierzbicka's 'analytic' –
phenomenological continuum

EN > FR > DE > RU

- these West-East differences seem to override the genealogical relationships

Outline

1. Constructional typology: Methodological challenges
2. The constructional field of LETTING
3. Data
4. Quantitative analyses
 - conceptual space
 - language space
5. Conclusions

Language Space: Procedure

Step 1. For each pair of languages i and j :

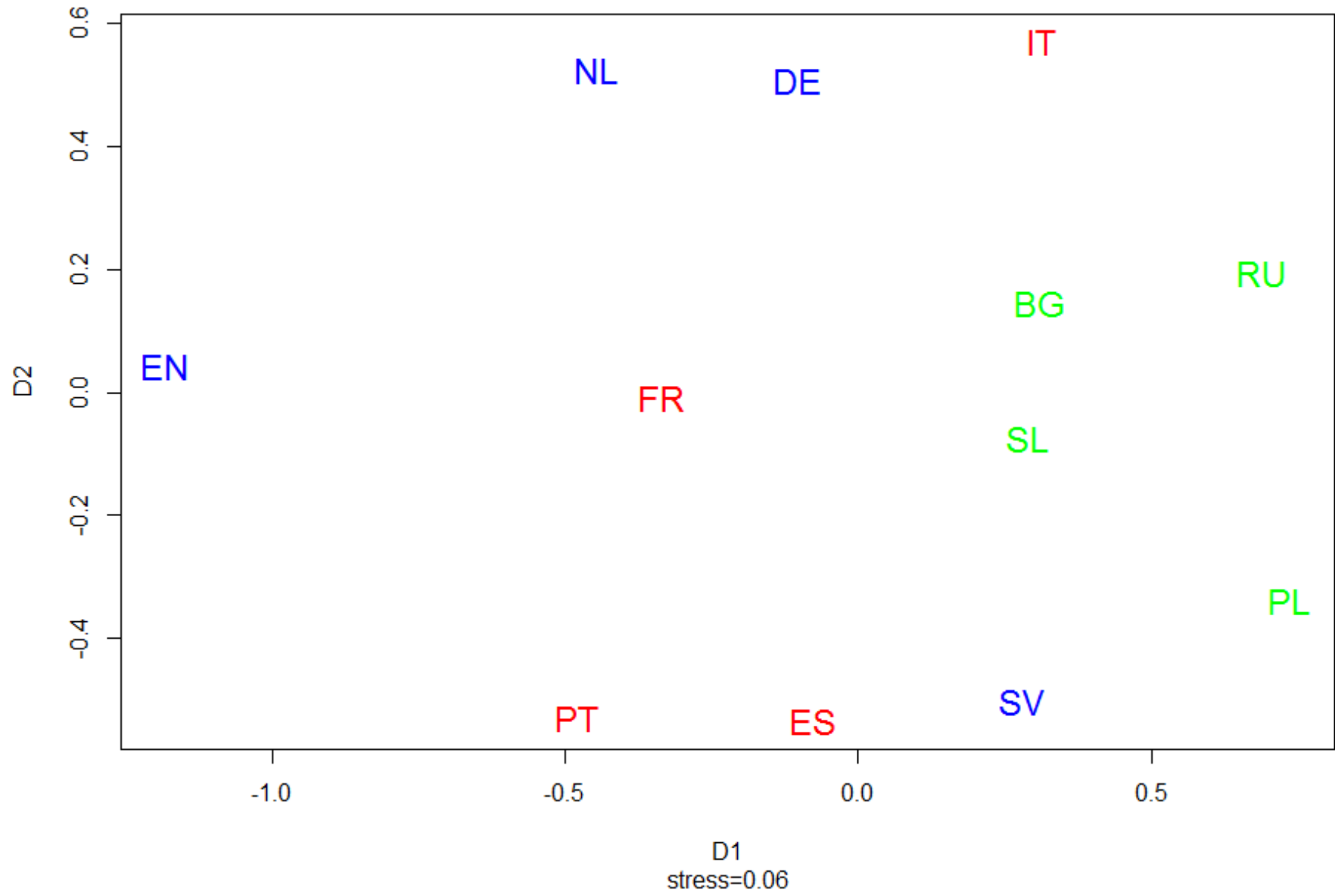
for each pair of exemplars:

- If both languages have the same expressions in both exemplars, similarity score $s_{ij} += 1$.
- If both languages have different expressions, $s_{ij} += 1$.
- If one language has identical expressions, and the other one different ones, $s_{ij} += 0$.

Step 2. Transform the similarities into distances.

Step 3. MDS

MDS



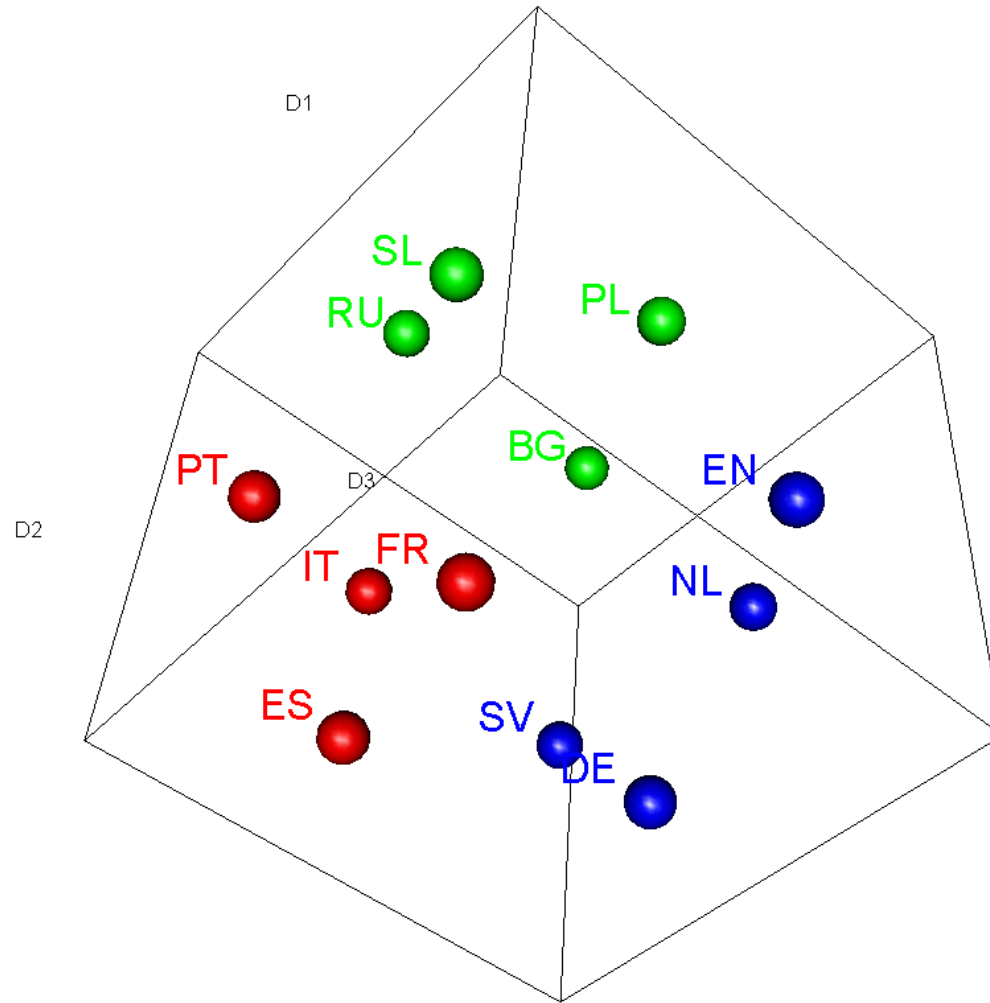
Interpretation

- the genealogical groupings are observed with a few exceptions:
 - Italian is close to German, as well as the Southern Slavic Slovene and Bulgarian
 - Swedish is close to the Slavic languages and, surprisingly, Spanish
- West-East correspondences within each group

A 'lite' version

- Is the force-dynamic meaning expressed at all or not?
(if 'NotLet', $s_{ij} += 1$)
- A 3D MDS solution

3D MDS



Interim conclusions

- the coarser granularity leads to more stable genealogical patterns
- the finer granularity leads to less genealogically transparent results, but reflects the West-East continuum of analyticity to a larger extent

Outline

1. Constructional typology: Methodological challenges
2. The constructional field of LETTING
3. Data
4. Quantitative analyses
 - conceptual space
 - language space
5. Conclusions

Final conclusions

- There is evidence of an 'analyticity' (focus on cause and effect) continuum in the ways the languages cut the conceptual space. There is a decrease of 'analyticity' from West to East.
- The level of granularity of analysis can yield different results. Taking into account the more global strategies leads to a clustering of languages that reflects the genealogical classification to a larger extent.

Perspectives

- More data, also translated into English
- Other registers
- Automatic word alignment (e.g. Tiedemann 2011)
- Dialectometric dimension

