

# GIVE it a go: TAKE parallel corpora, PUT some statistics, and GET cross-linguistic comparisons!

Natalia Levshina, Thomas Mayer  
& Michael Cysouw  
Philipps University of Marburg

# Outline

1. GIVE: theoretical background
2. Parallel corpora as a tertium comparationis for cross-linguistic semantic comparisons
3. An onomasiological study of GIVE in 10 European languages
4. Conclusions and perspectives

# Semantics of GIVE

- **semasiological perspective**
  - semantic extensions (polysemy)
  - paths of semantic change(Newman's 1996, von Waldenfels 2012)
- **onomasiological perspective**
  - how the semantic labour is divided between different expressions of GIVE
  - which conceptual distinctions are cross-linguistically salient

# Outline

1. GIVE: theoretical background
2. Parallel corpora as a tertium comparationis for multilingual semantic comparisons
3. An onomasiological study of GIVE in 10 European languages
4. Conclusions and perspectives

# Multilingual semantic comparisons

- Elicited data (e.g. Levinson & Meira 2003; Majid et al. 2004; Hartmann et al. In press)
  - problems**
    - decontextualized
    - not natural (different from actual use)
    - only physical stimuli in experiments
- Non-elicited data (parallel corpora)
  - problems**
    - register bias
    - translationese

# Corpus-driven semantic maps

- Onomasiological: all instances of items that constitute a semantic field
- Bottom-up: from exemplars to generalizations (relevant dimensions, clusters, prototypes)
- Examples:
  - local phrase markers (Wälchli 2010)
  - motion verbs (Wälchli & Cysouw 2012)
  - letting constructions (Levshina 2012)

# Outline

1. GIVE: theoretical background
2. Parallel corpora as a tertium comparationis for cross-linguistic semantic comparisons
3. An onomasiological study of GIVE in 10 European languages
4. Conclusions and perspectives

# Aims

1. Theoretical: find common dimensions of variation of GIVE in 10 European languages
  - Germanic: EN, DE, NL, SV
  - Romance: ES, FR, IT, PT
  - Slavic: BG, PL
2. Methodological: will the results vary for different registers?
  - Bible translations
  - film subtitles



# Corpora

- Bible translations (New Testament)
  - part of a large-scale DFG project [www.paralleltext.info](http://www.paralleltext.info)
  - word-aligned with the help of GIZA++
- Subtitles for 10 films ([www.opensubtitles.org](http://www.opensubtitles.org), [www.subscene.com](http://www.subscene.com))
  - different original languages and genres, to reduce the translationese effect
  - sentence-aligned with the help of timing information (cf. Tiedemann 2007)

# Data set

- Verbs of giving (FrameNet)  
144 different English forms searched for automatically (*give, bequeath, contribute, donate, fob, hand, pass, leave, etc.* )
- After manual disambiguation and cleaning:

77 sentences in subtitles (all)

+

77 sentences in the Bible (random sample)

---

154 multilingual contexts

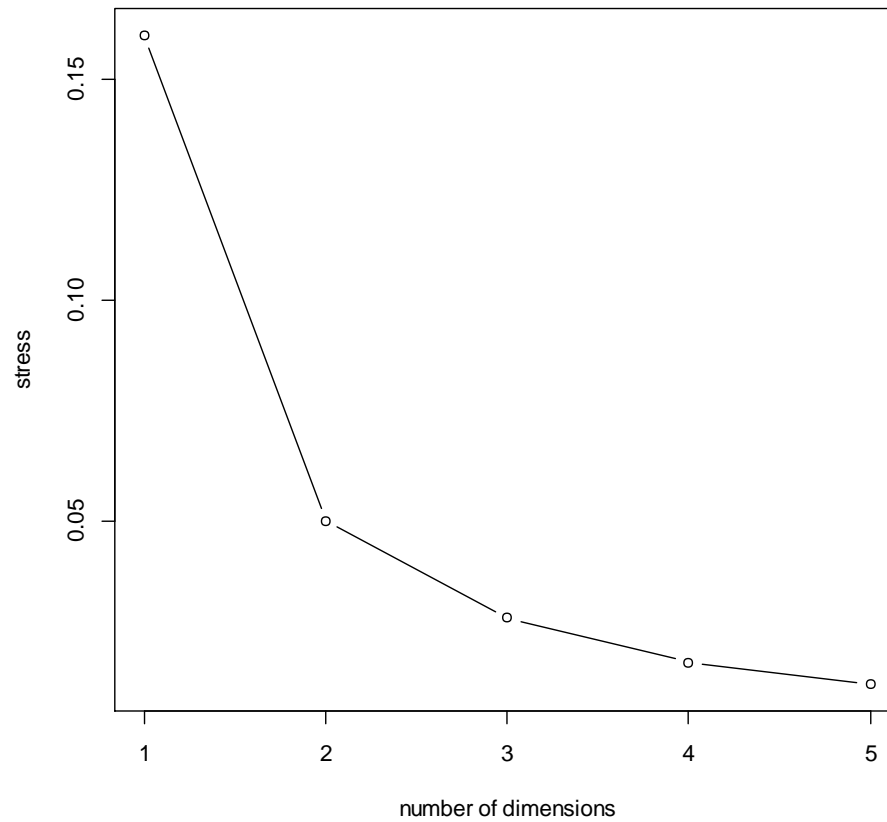
# Data set

	A	B	C	D	E	F	G	H	I	J	K	
1	ID	Source	EN	BG	DE	ES	FR	IT	NL	PL	PT	SV
2	B01	Bible	supply	zapelnjavam	ersetzen	suplir	suppleer	riempire	vervullen	napelnich	suprir	ge
3	B02	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
4	B03	Bible	give	davam	geben	dar	donner	NA	geven	dach	dar	ge
5	B04	Bible	give	davam	geben	NA	donner	dare	geven	dach	dar	ge
6	B05	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
7	B06	Bible	give	davam	geben	dar	donner	dare	zeggen	dach	dar	ge
8	B07	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
9	B08	Bible	give	davam	geben	dar	donner	dare	geven	dach	entregar	ge
10	B09	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
11	B10	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
12	B11	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
13	B12	Bible	give	davam	geben	dar	distribuer	fare	geven	dach	dar	ge
14	B13	Bible	give	davam	geben	dar	etablier	dare	geven	dach	dar	ge
15	B14	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
16	B15	Bible	give	davam	geben	dar	donner	dare	geven	wydach	dar	utge
17	B16	Bible	give	davam	geben	dar	NA	dare	geven	dach	dar	ge
18	B17	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge
19	B18	Bible	give	davam	geben	dar	donner	dare	geven	wydach	ordenar	ge
20	B19	Bible	give	davam	geben	dar	donner	dare	geven	dach	dar	ge

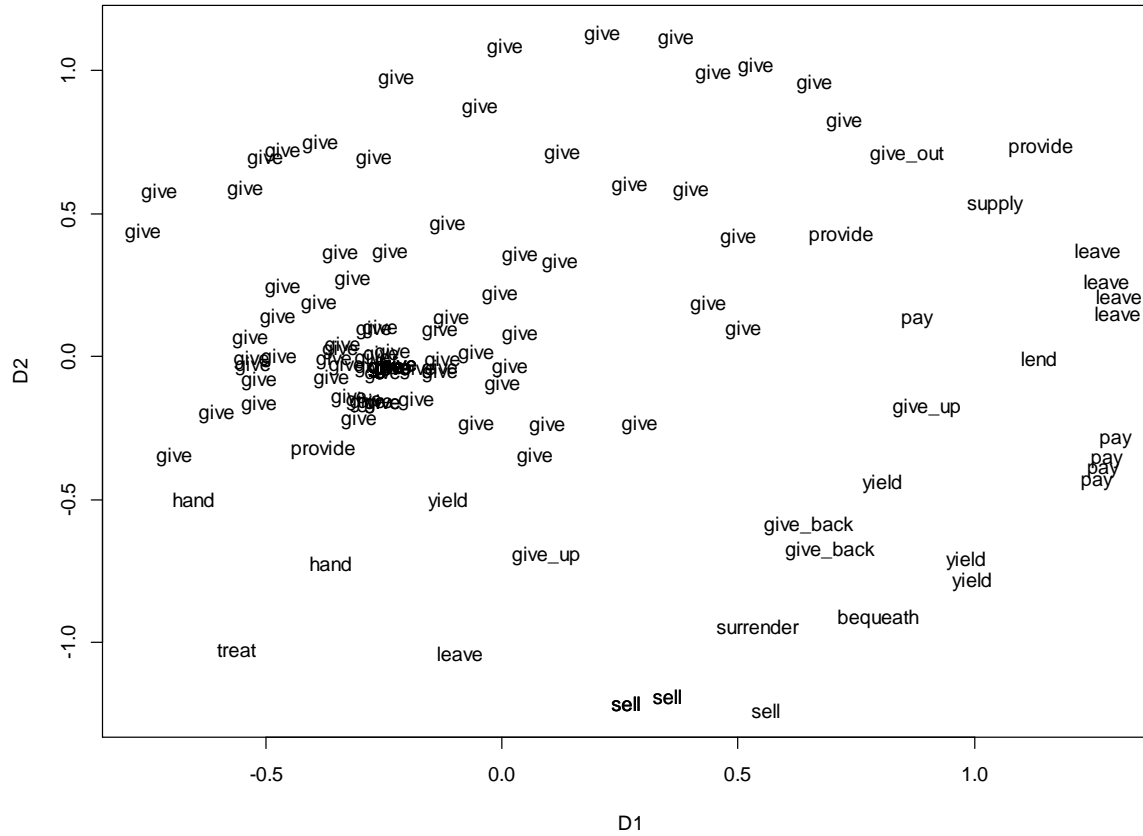
# Matrix of exemplar dissimilarities

- For each pair of exemplars, compute similarity score  $s$ : the sum of all similar translations
- Partial similarity = 0.5
- Dissimilarity score  $d = 1 - s$
- Multidimensional scaling (smacof package, R)

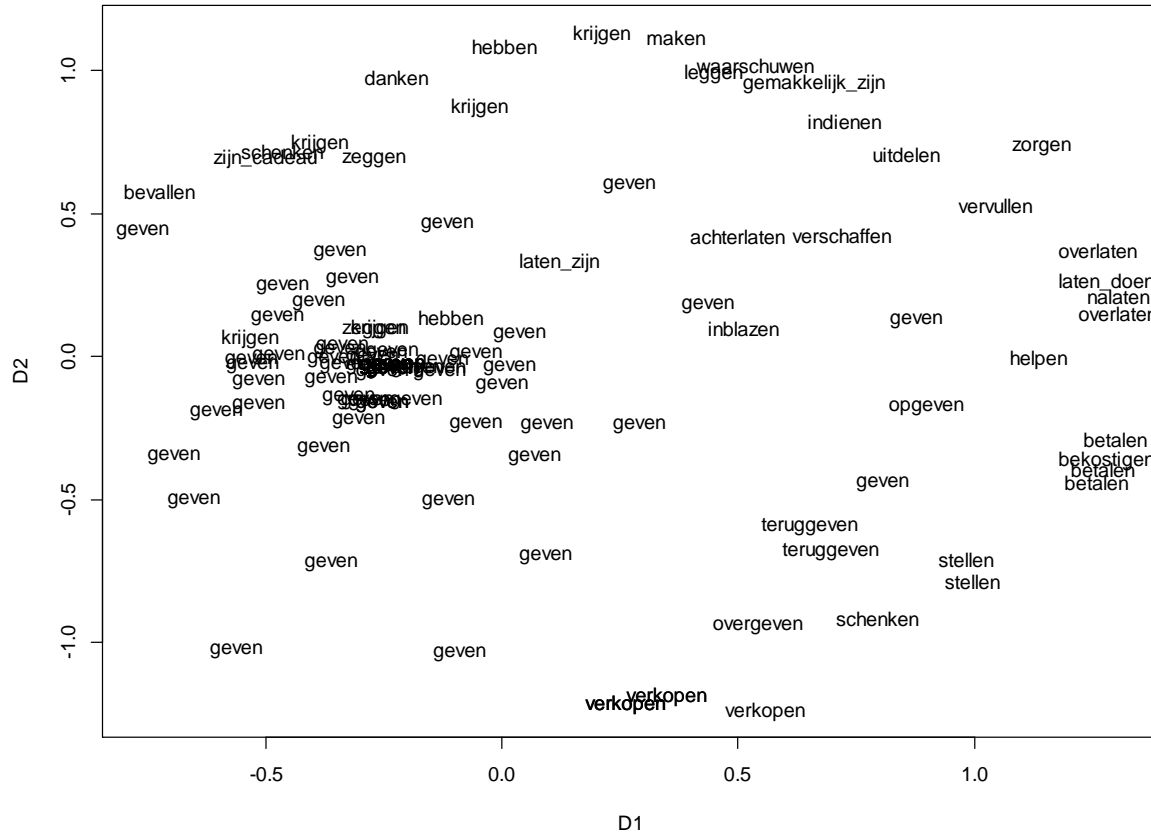
# How many dimensions?



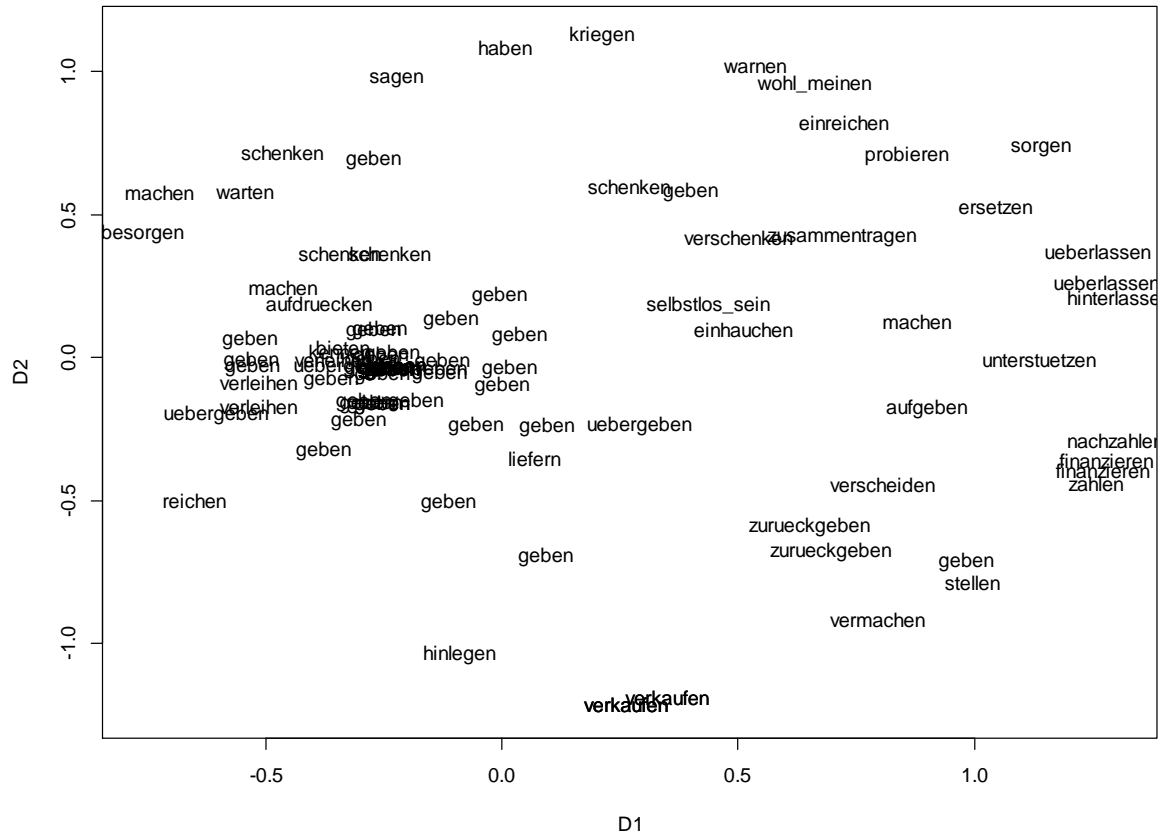
# English



# Dutch

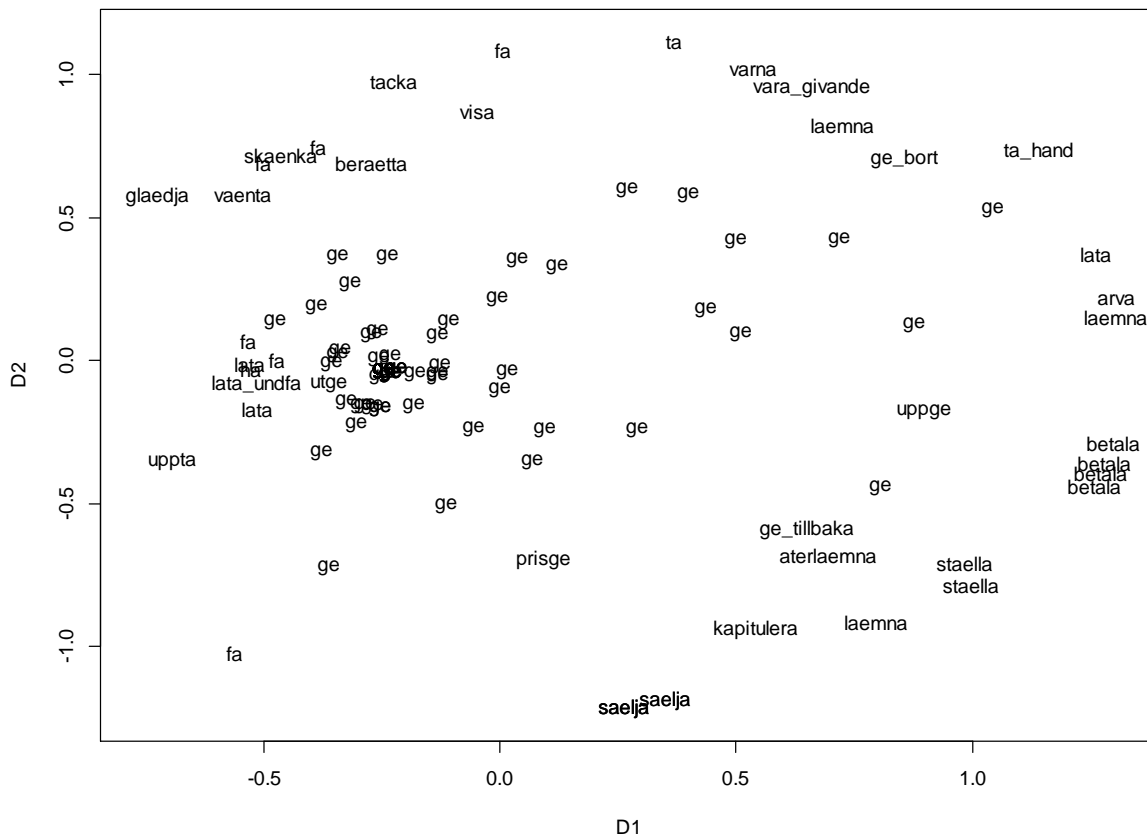


# German

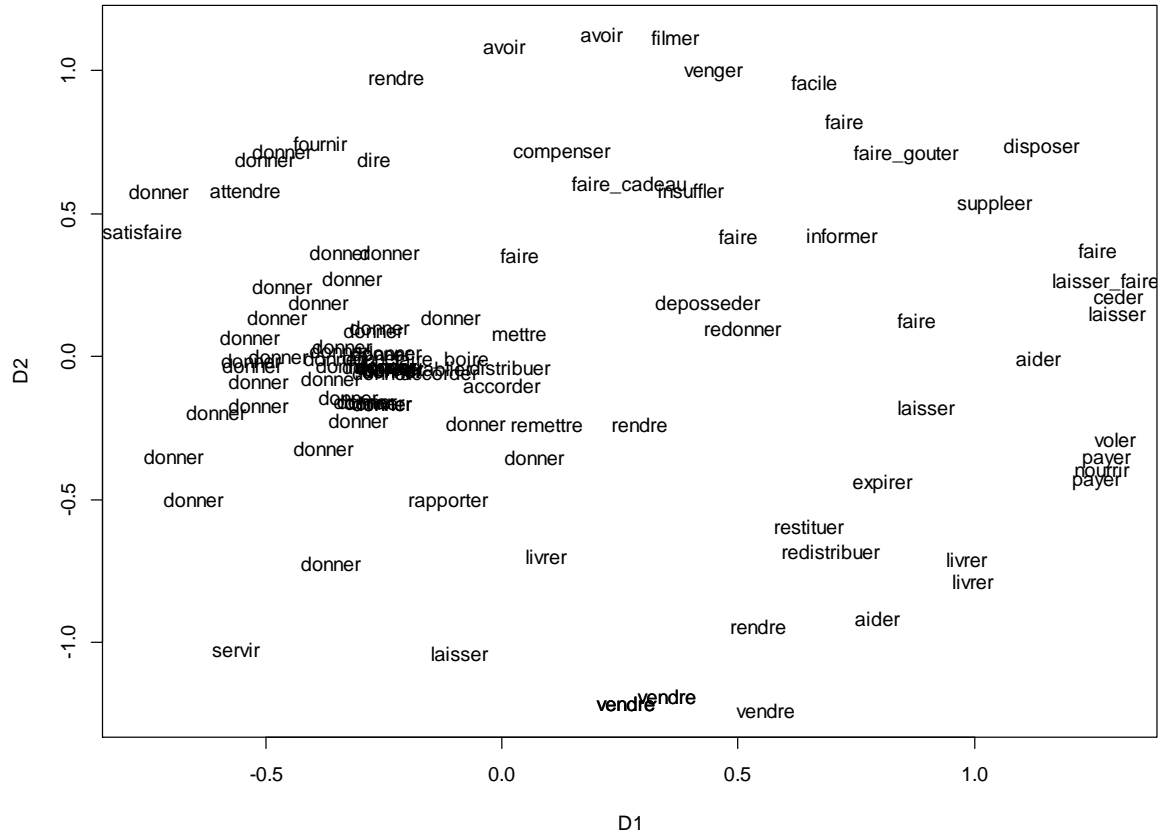




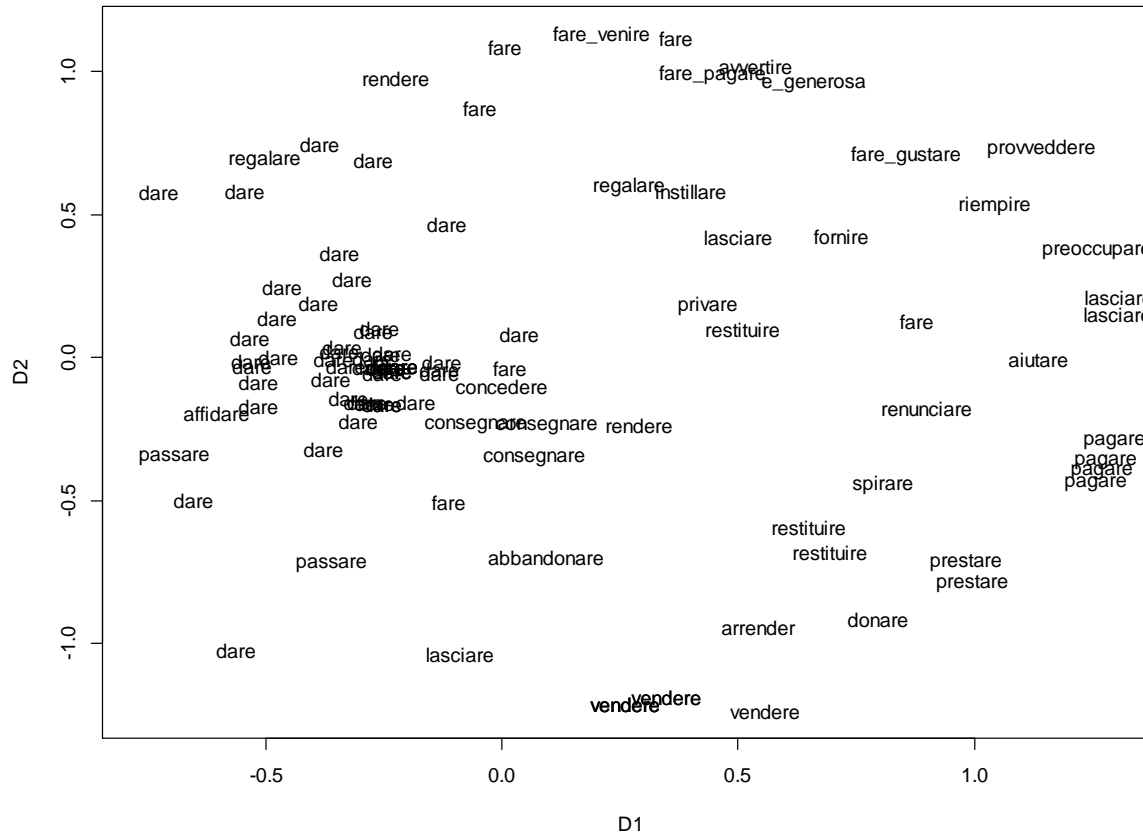
# Swedish



# French

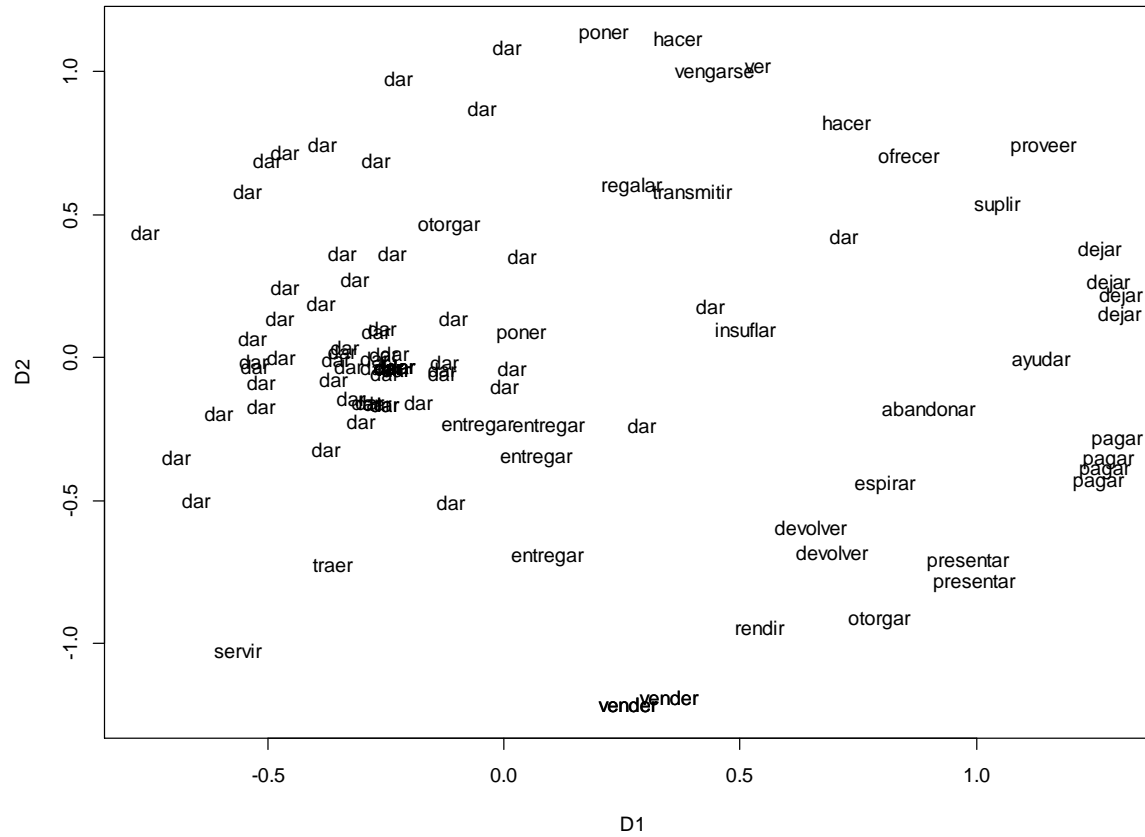


# Italian

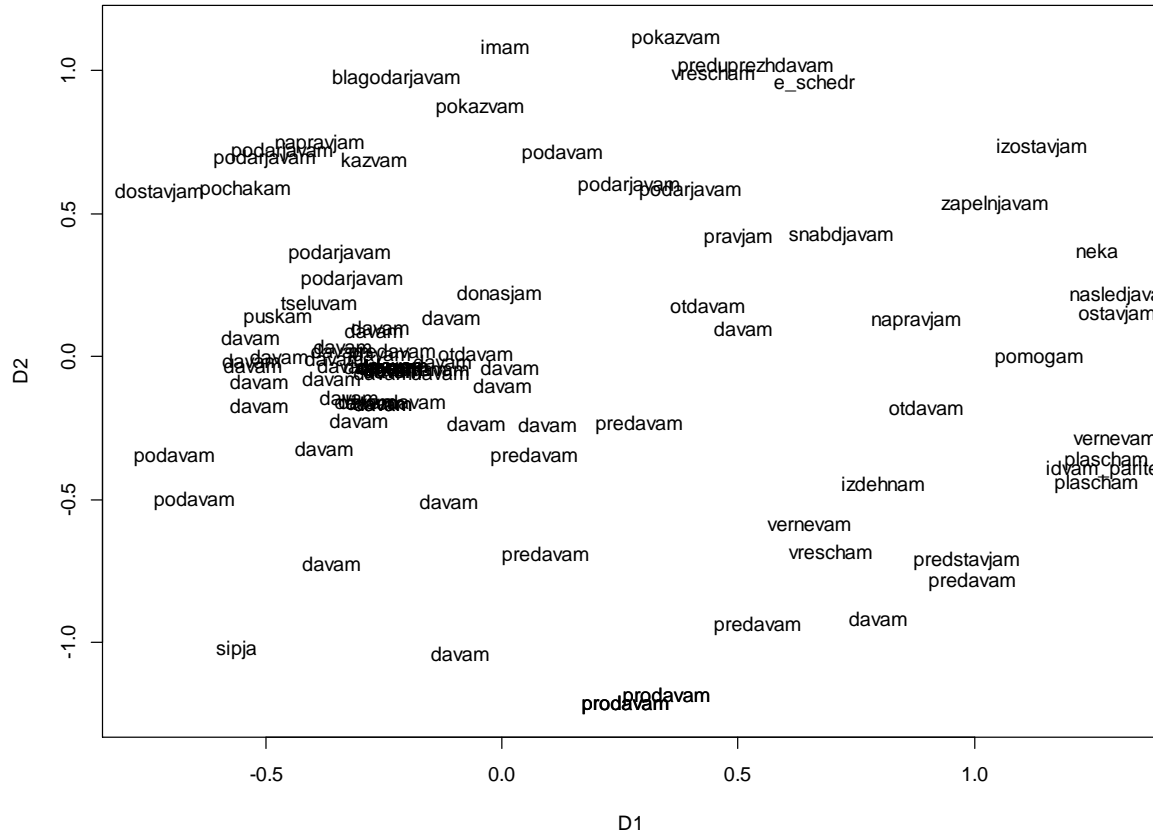




# Spanish



# Bulgarian



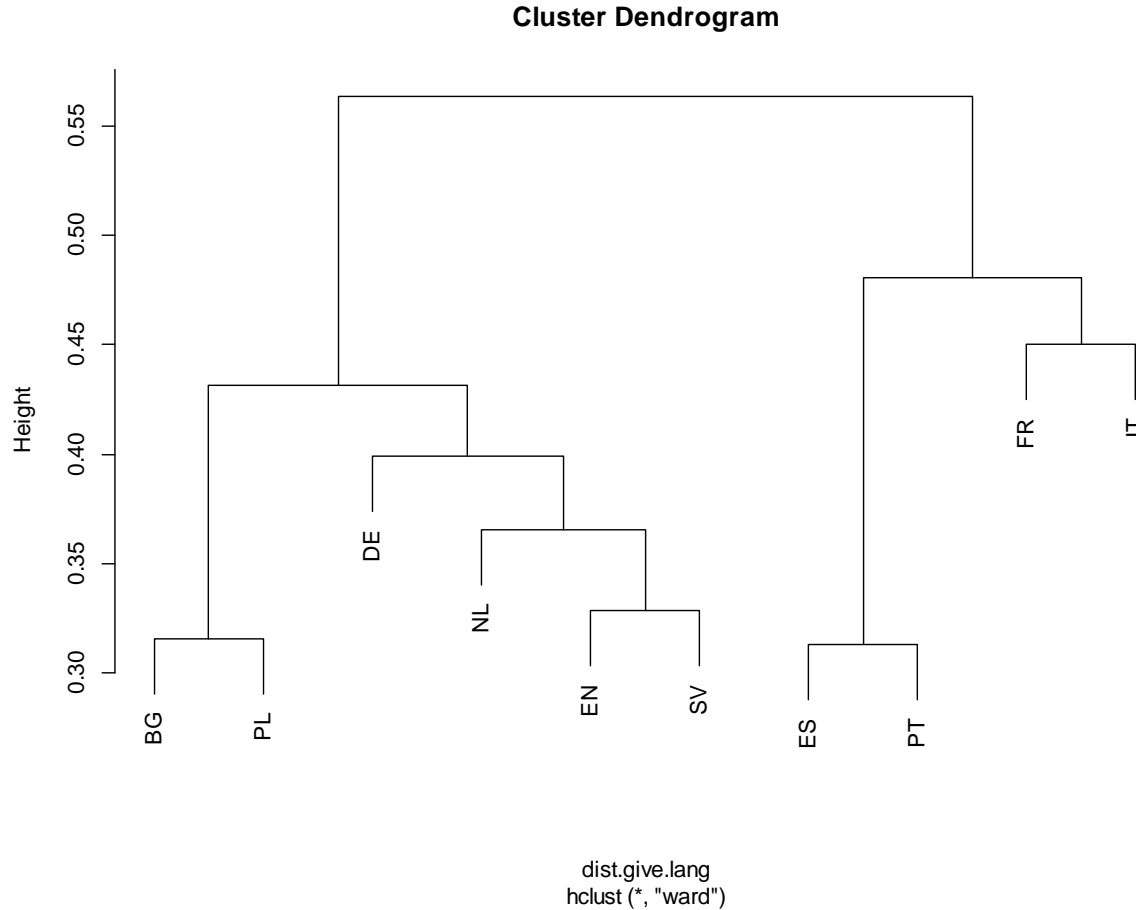


# Dimensions

- Horizontal: distance/contact between the Agent and Recipient
- Vertical: physical vs. non-physical transfer



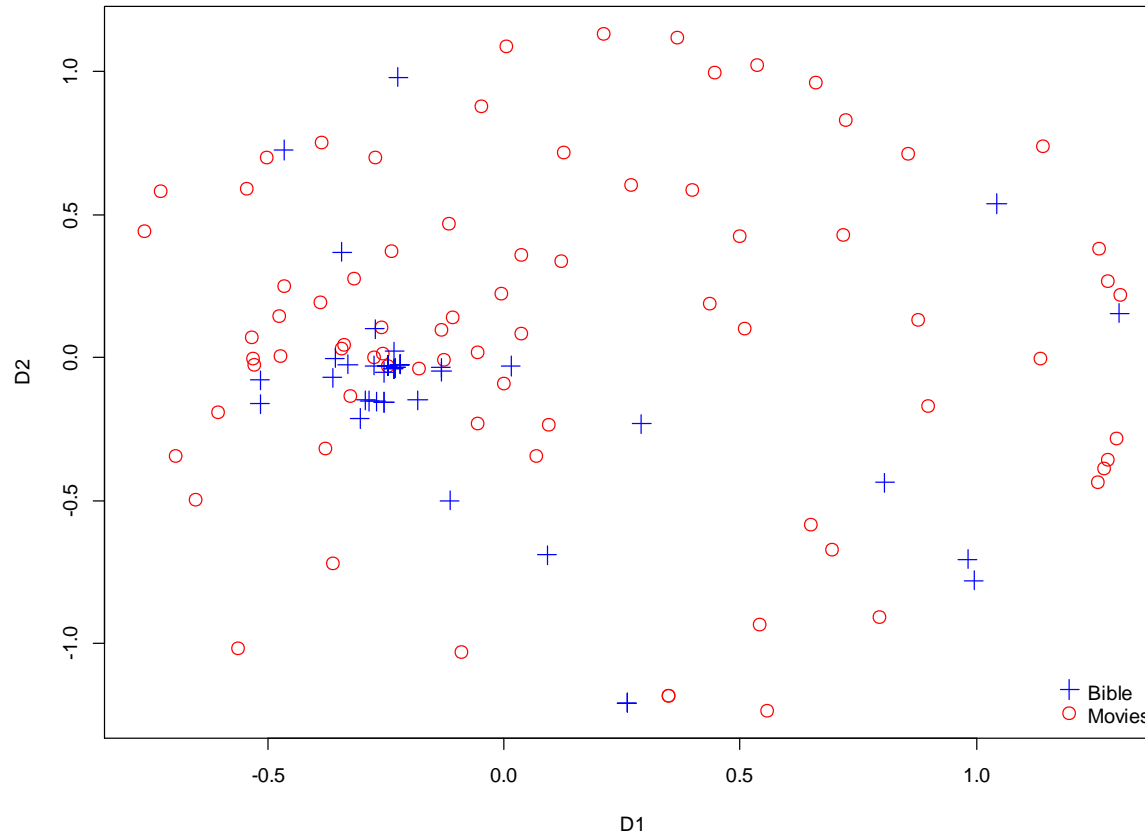
# Clustering the languages



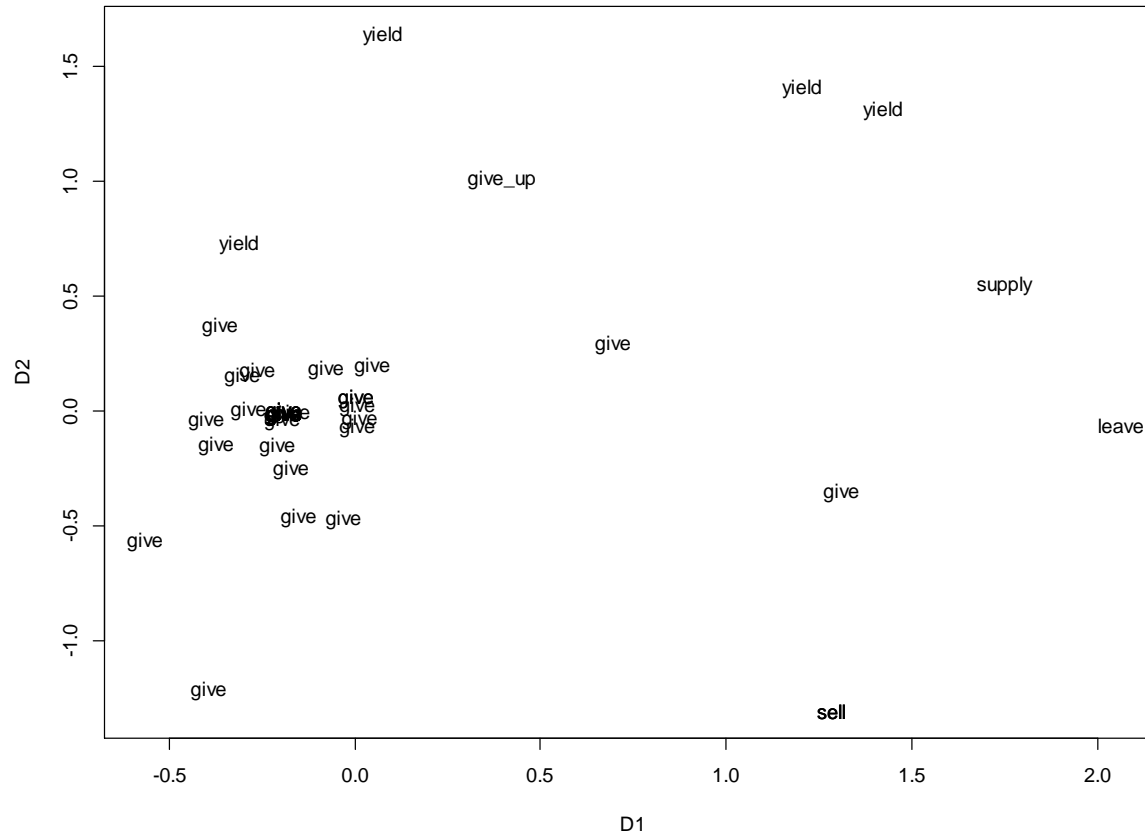
Criterion: Does the root contain the canonical *give* in the language?

Is there register variation?

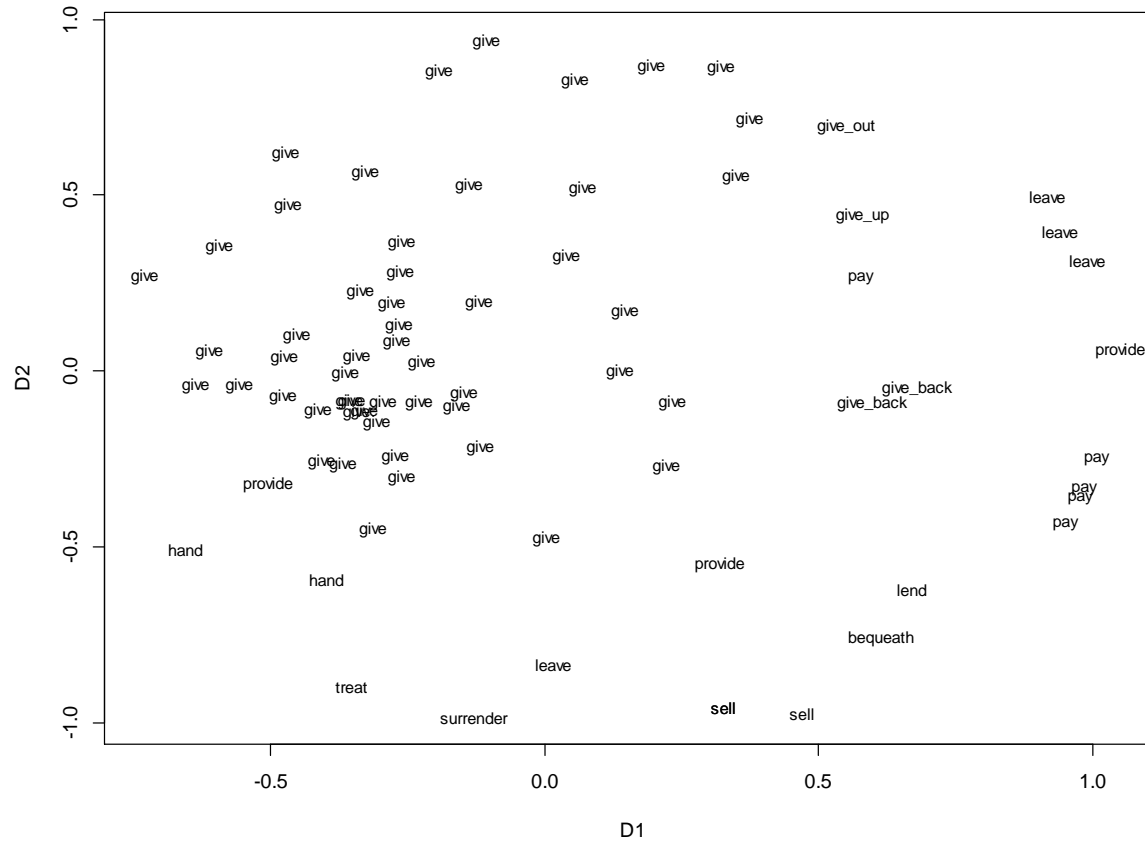
# Bible vs. Movies



# MDS for Bible only



# MDS for movies only



# Register variation

- The Bible GIVE reveals substantially less variability
- The importance of *yielding* in the Bible



Register variation should be taken into account in future studies based on multilingual parallel corpora

# Outline

1. GIVE: theoretical background
2. Parallel corpora as a tertium comparationis for cross-linguistic semantic comparisons
3. An onomasiological study of GIVE in 10 European languages
4. Conclusions and perspectives

# Conclusions

- The main dimensions of GIVE in 10 languages involve contact and (non)physical transfer
- Genealogical relatedness determines the way the languages cut the semantic space (basic vocabulary?)
- Register variation is an important factor in onomasiological corpus-driven typology
- More languages are needed to detect areal phenomena and find new dimensions of variation



Thank you!

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