

Visualizing Valency Alternations

Michael Cysouw¹ & Iren Hartmann²

¹Universität Marburg, ²Max-Planck-Institut für evolutionäre Anthropologie

Outline

- i. Data
- ii. Valency Alternations
- iii. Visualizations
- iv. Summary

Data

- counterparts ('translations') from 35 languages for **70 verb meanings** (conceived of as representative of the verbal lexicon)
- one primary valency frame (**coding frame**) and all possible alternative valency frames (**derived coding frames**) for each counterpart
- coding frames include information on argument coding (flags & indexes)
- arguments are assigned to **microroles** ('verb-specific roles') and **coding sets**

Project languages (for basemap)



L's with more detailed alternation data



Alternations and Alternants

Alternation → relation between two alternants

- e.g. Causative:

1 V (*1 runs*) → **3 1 V'** (*3 causes 1 to run*)

1 V 2 (*1 breaks 2*) → **3 1 V' 2** (*3 causes 1 to break 2*)

- **Alternants** (a.k.a Diatheses or Voices)
- Alternants can consist more than one coding frame (a.k.a. subcategorization frame)

Comparative and L specific concepts

Comparative Concepts:

- microroles (EATER, HITTER, LIKER, RUNNER, ...)
- verb meanings (EAT, HIT, LIKE, RUN, ...)

Language Specific Concepts:

- coding set (flagging, cross-referencing)
- coding frames (construction from coding sets)
- alternant (“voice”) (set of coding frames)
- alternation (pair of alternants, typically an ‘unmarked’ and a ‘marked’ one)

Visualization

- How can we visualize linguistic diversity?
 - Show differences/similarities between languages
 - Basic idea: use a **constant “universal” base map**, and show each language as a separate overlay
 - **Optimize base map to highlight variation** among the languages to be compared
 - With sufficient data, suitable base maps approach an average (“universal”) language structure
 - Construction of a base map is similar to the semantic map approach from linguistic typology

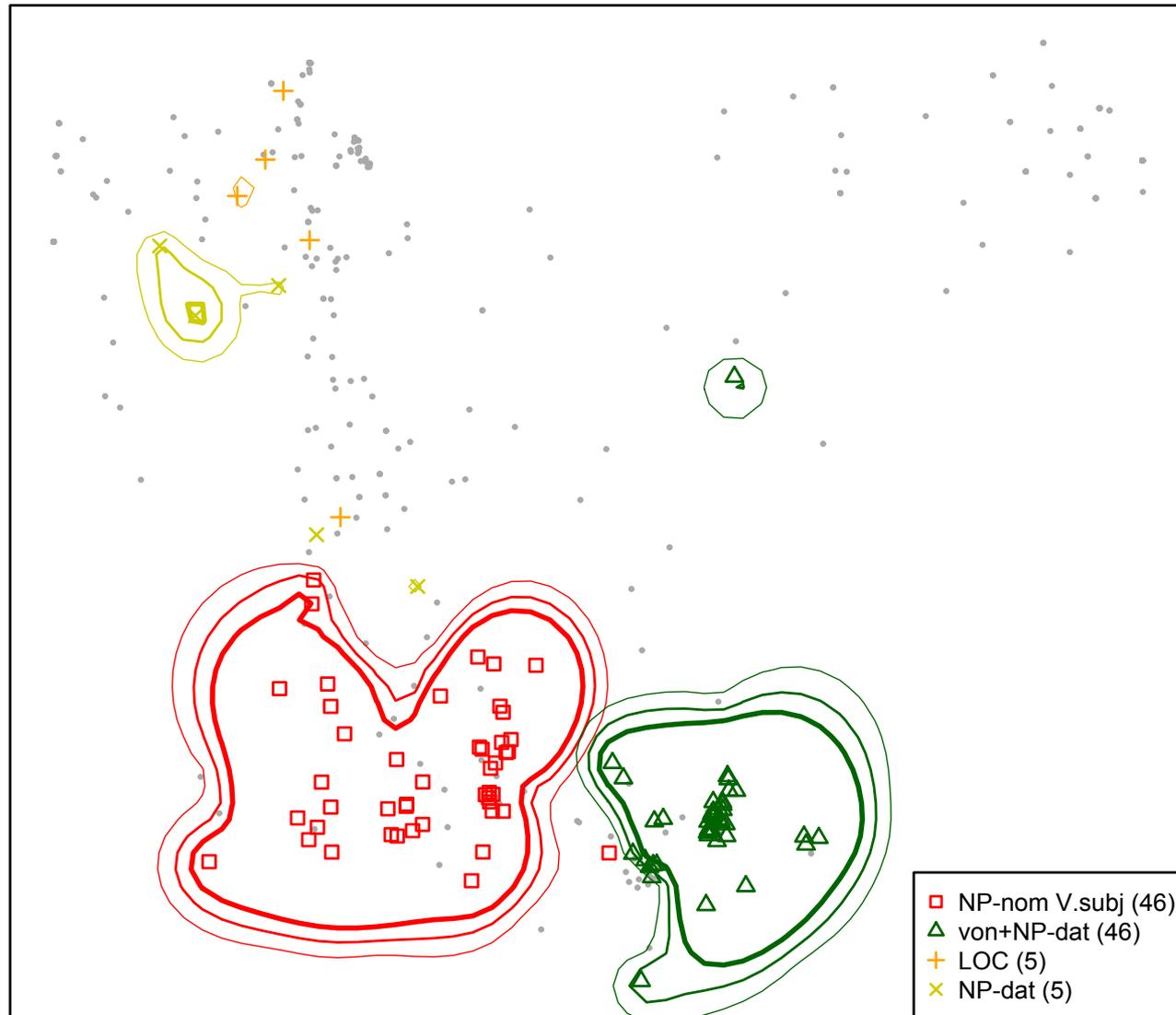
Visualization

- Three aspects to be visualized:
 - Visualize **alternants** (“voices”) individually
 - Visualize **alternations** (“relations between alternants”)
 - Visualize **similarity between alternations** across all languages (“typology of alternations”)

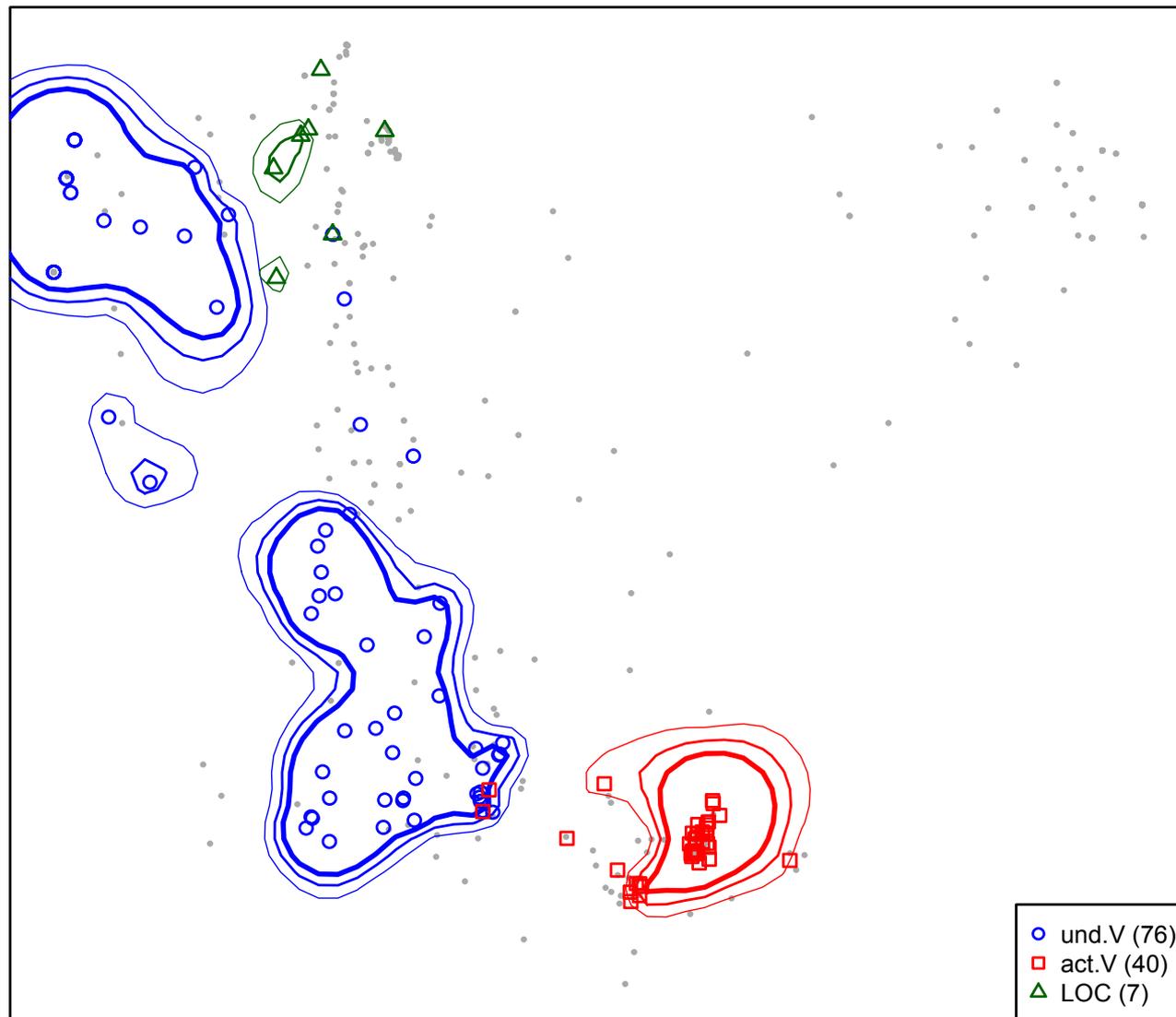
1. Alternants

- An **alternant** (e.g. German Passive marked with *werden+Partizip II*) is characterized by a **mapping between microroles** (“lexical roles”) and **coding sets** (“language specific marking”).
 - HITTER, THROWER, etc (“agents”) are marked with prepositional phrase *von + NP + Dative case*
 - HIT THING, THROWN THING, etc (“patients”) are marked with *Nominative case + Verb inflection*
- For **base map**, use **microroles**, and **position** them **according to average co-expression**

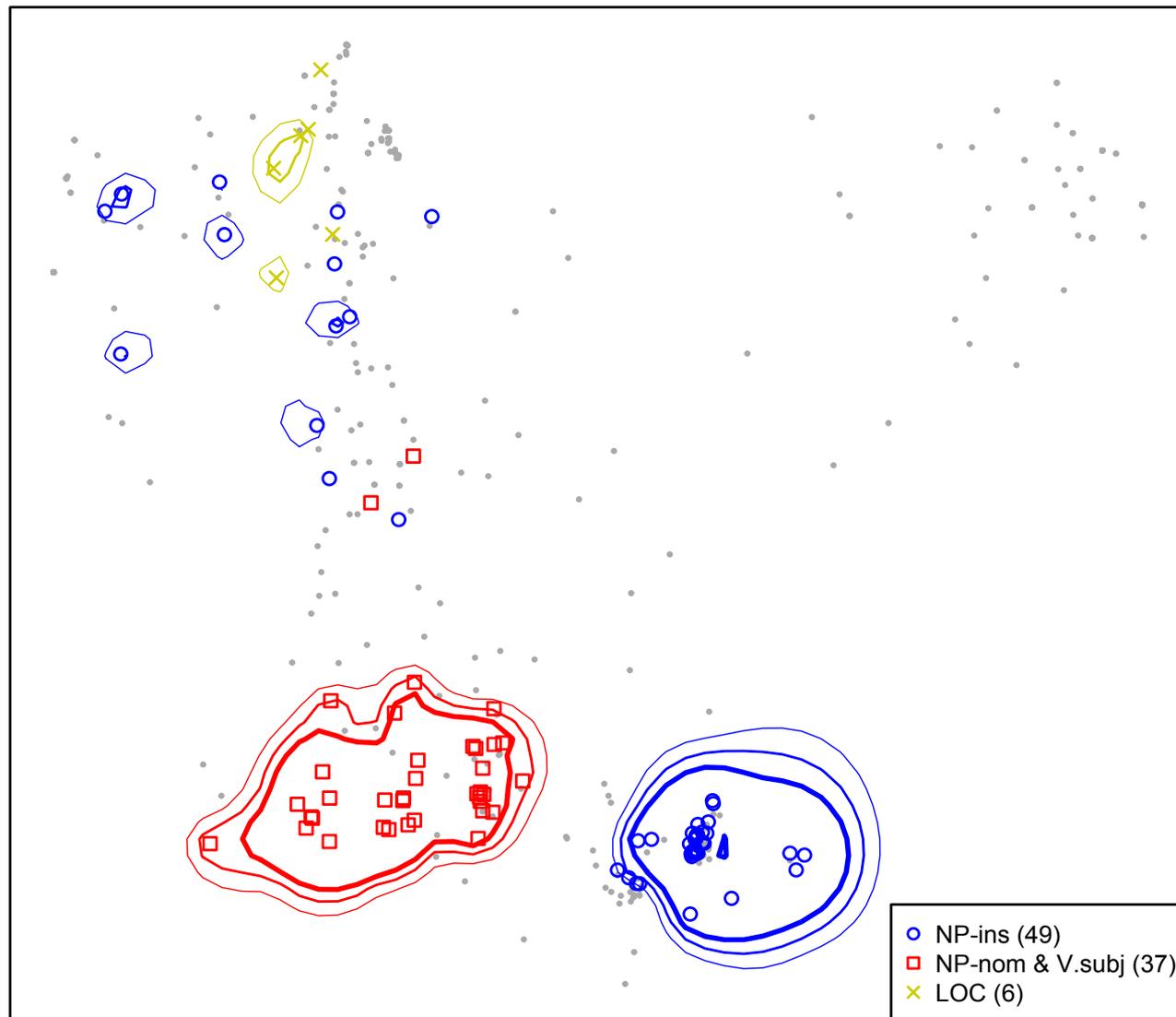
Passive (werden)_German



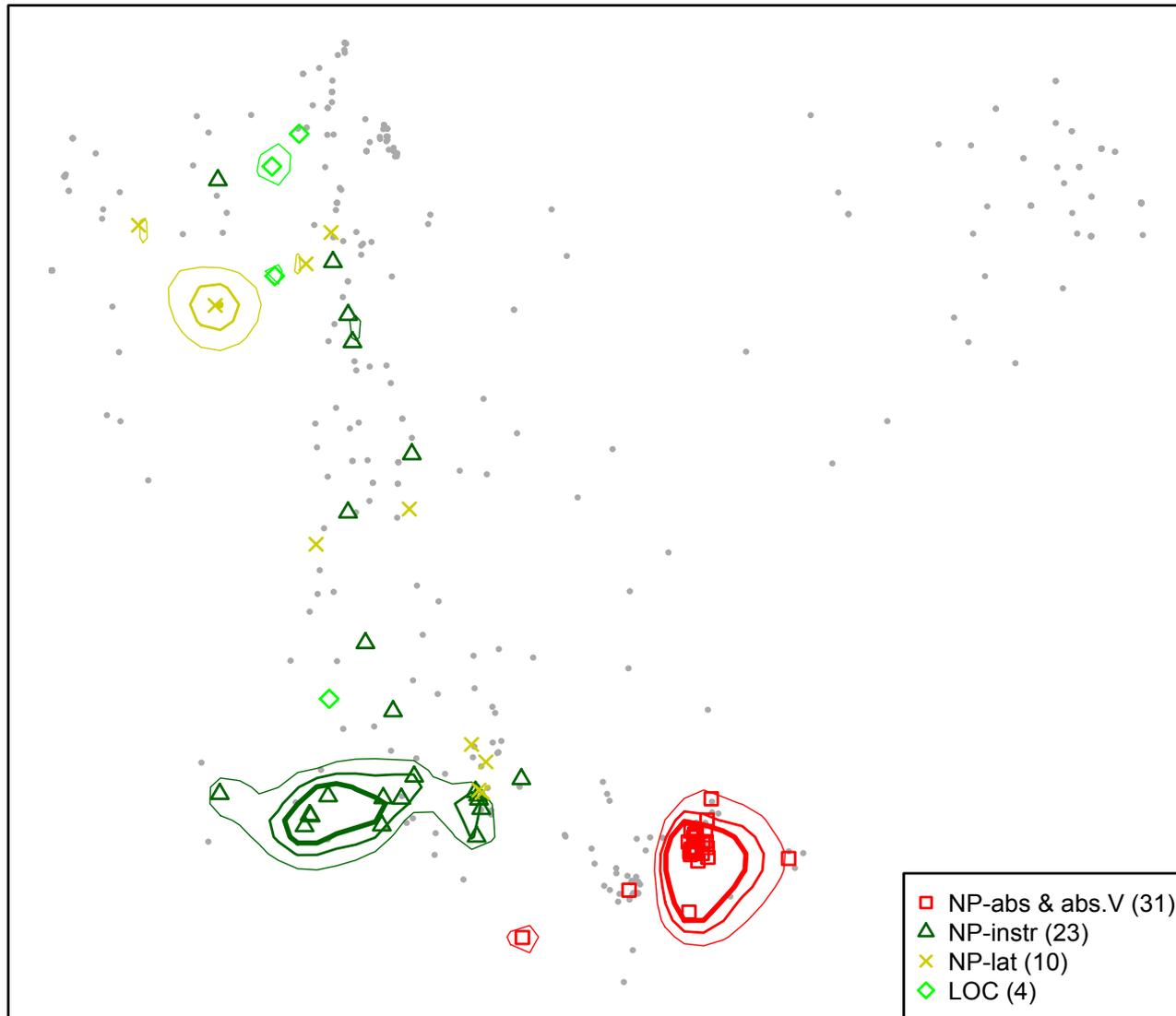
02 benefactive/possession of U (+gi)_Hoocak



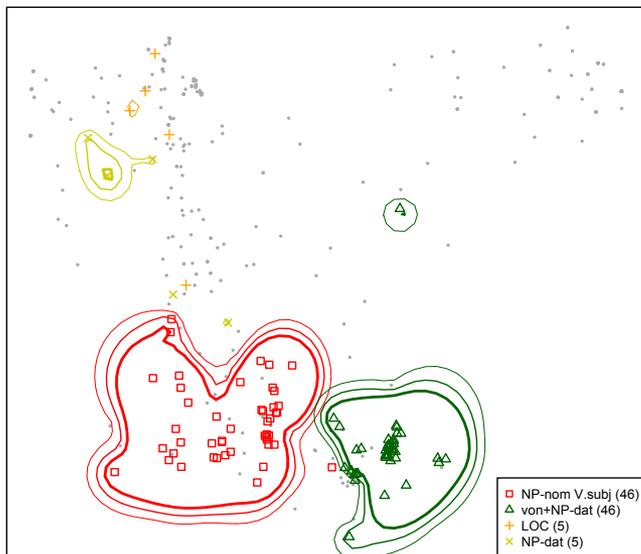
Participial Passive_Russian



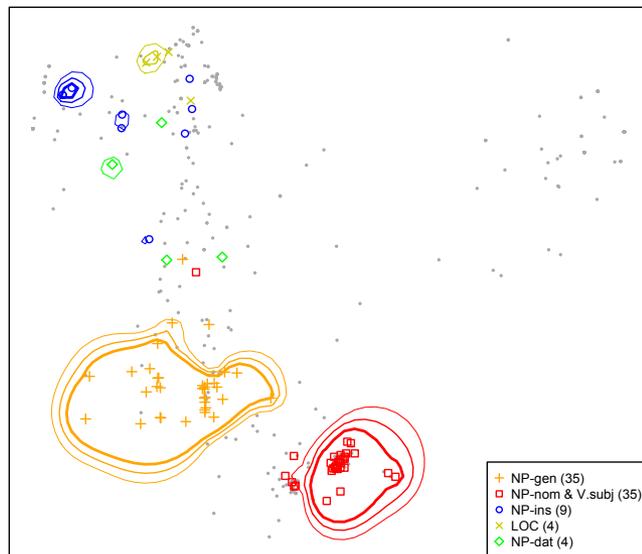
Antipassive Alternation_Bezhta



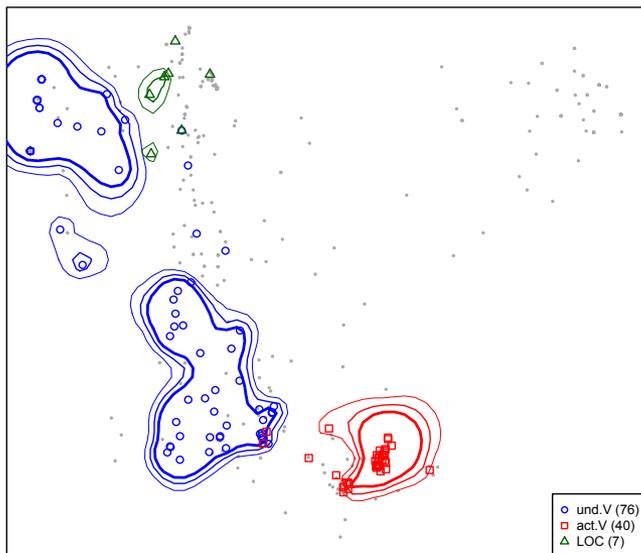
Passive (werden)_German



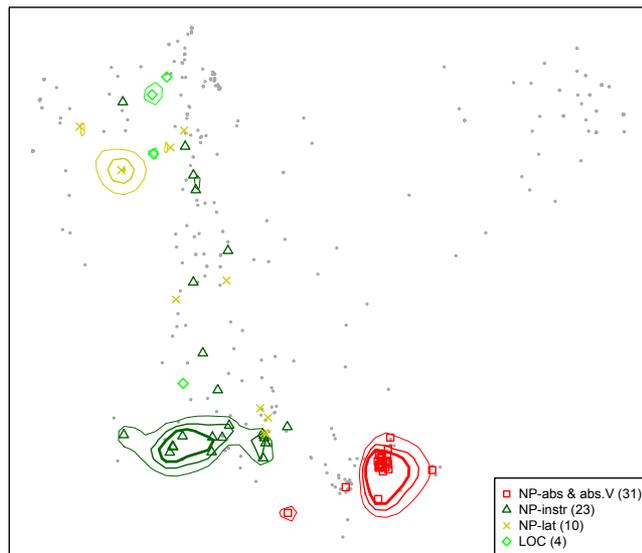
Negative Accusative-Genitive alternation_Russian



02 benefactive/possession of U (+gi)_Hooçak



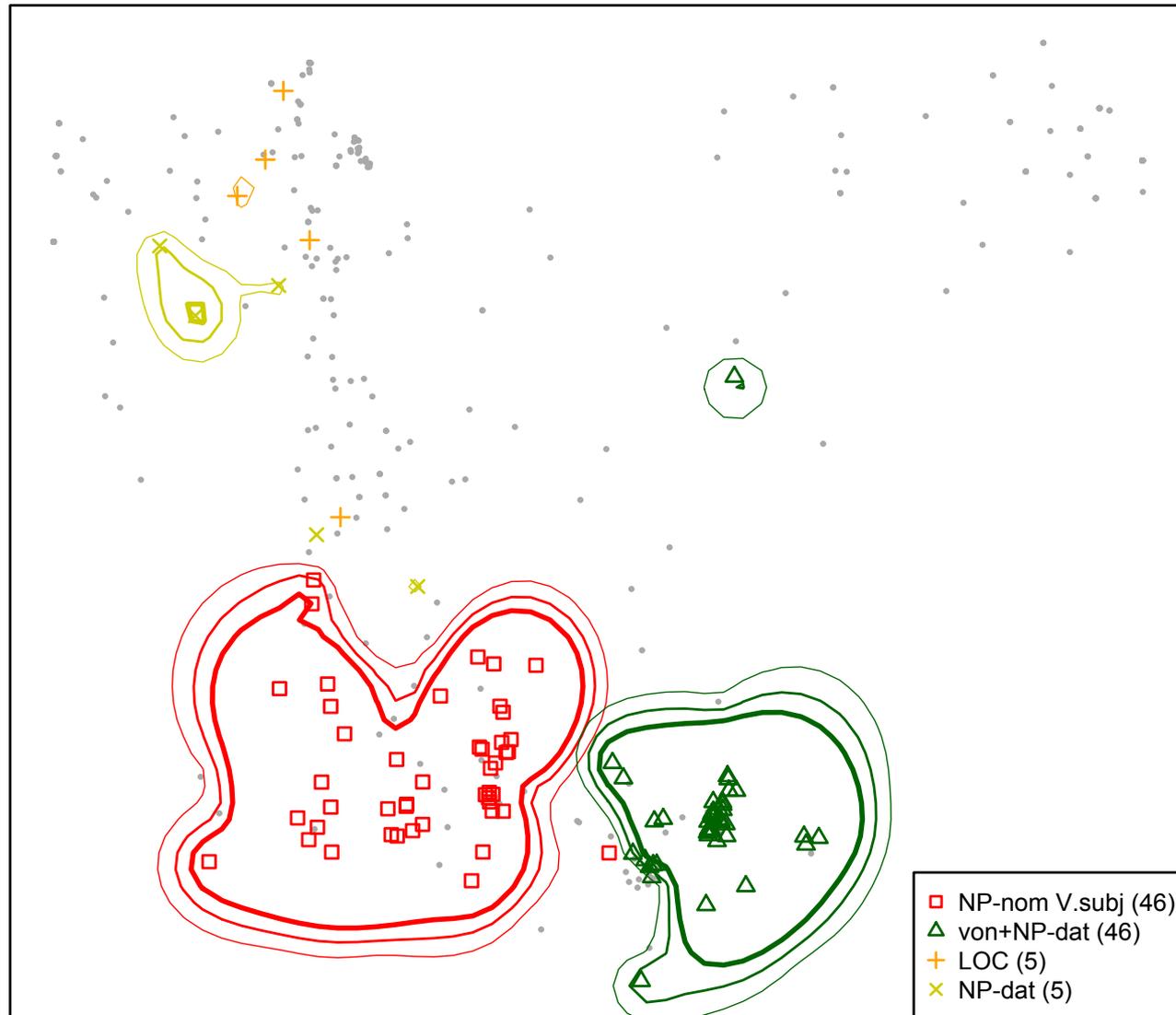
Antipassive Alternation_Bezhta



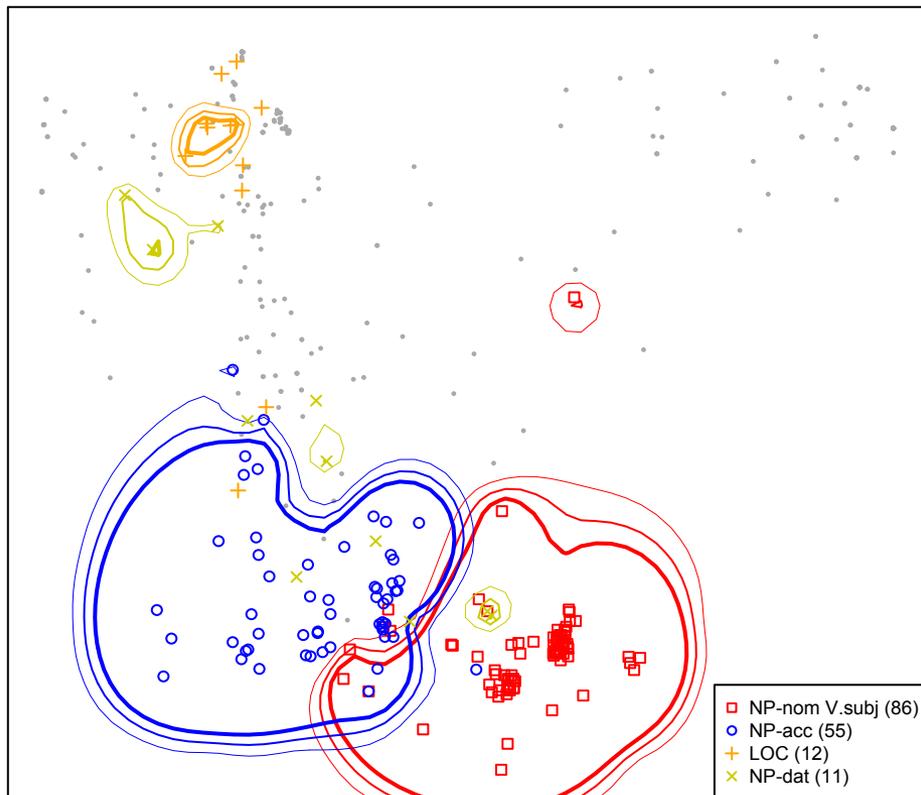
2. Alternations

- The previous visualizations **cannot be directly compared**
 - clusters of microroles are one important notion, but the **actual coding is also important**
 - but: **how to compare language-specific coding?**
e.g. what is the relation between
 - German *von+NP-dat*, Hoocak *act.V*, Russian *NP-ins*, Bezhta *NP-abs & abs.V*
- Our proposal: **compare alternant to ‘unmarked’ alternant of the same language**

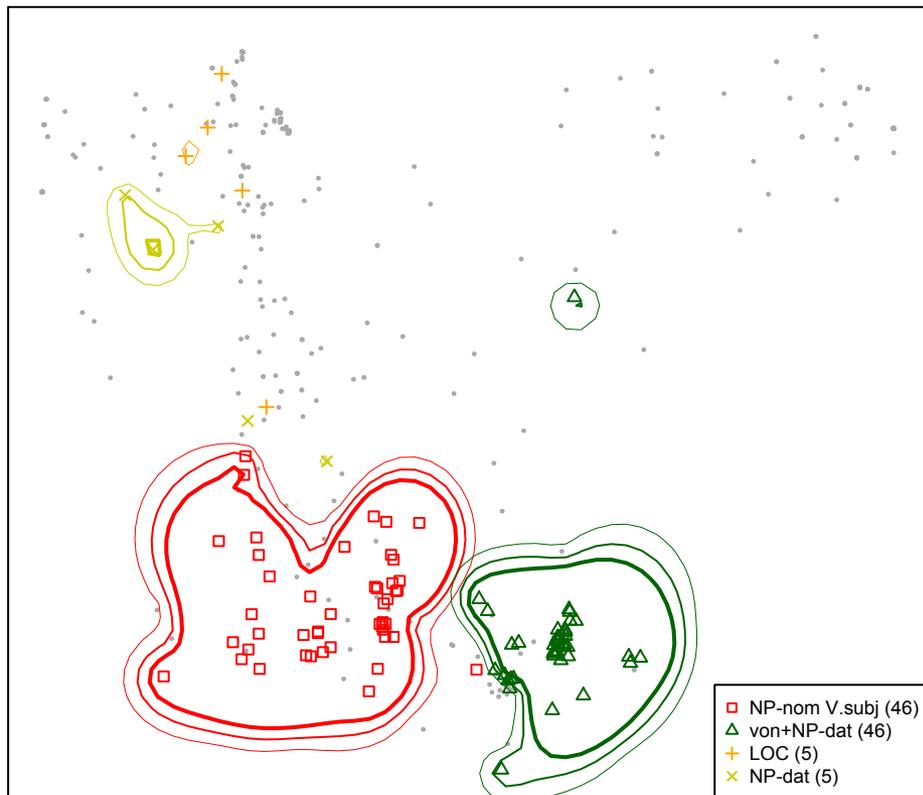
Passive (werden)_German



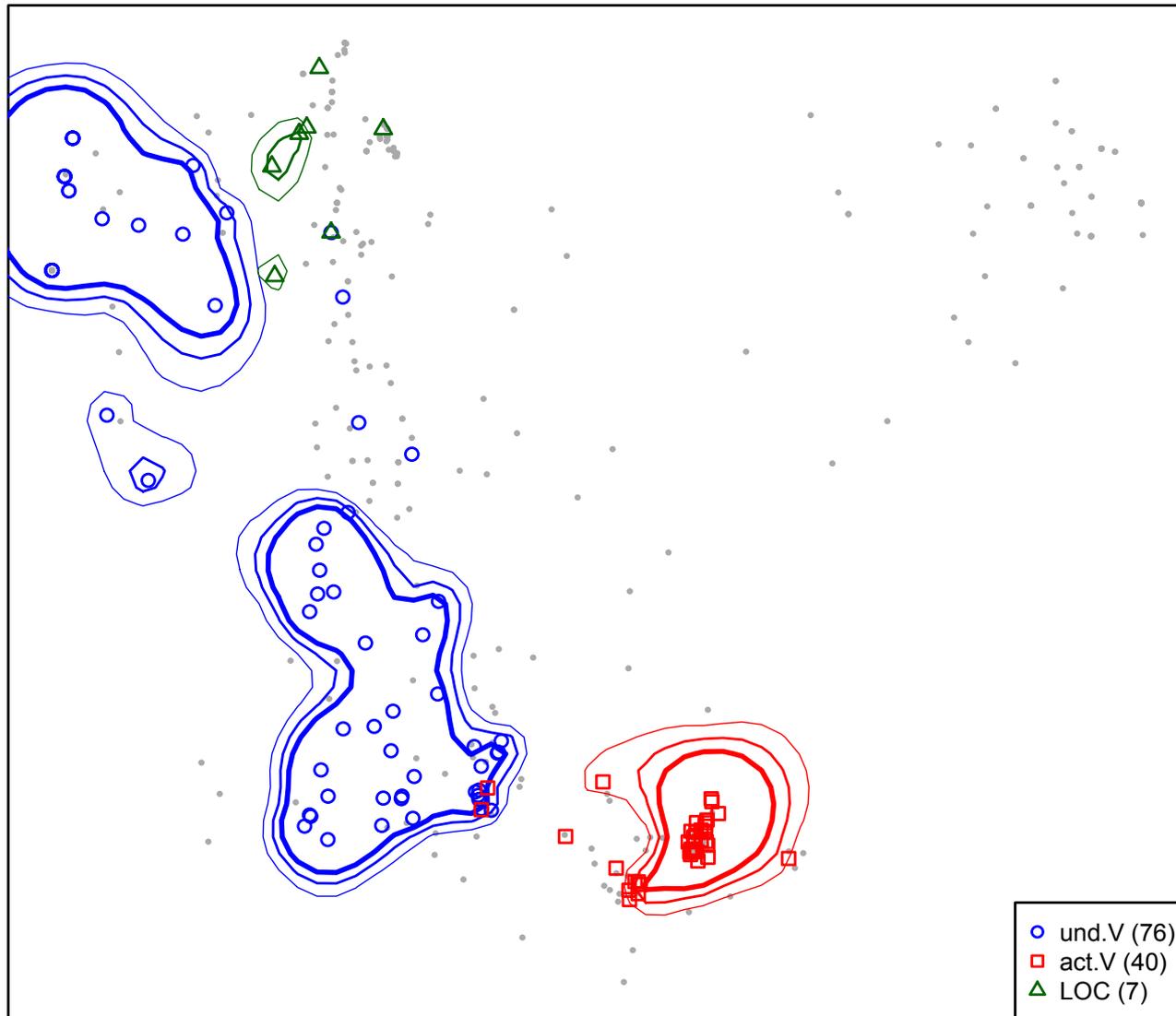
_German



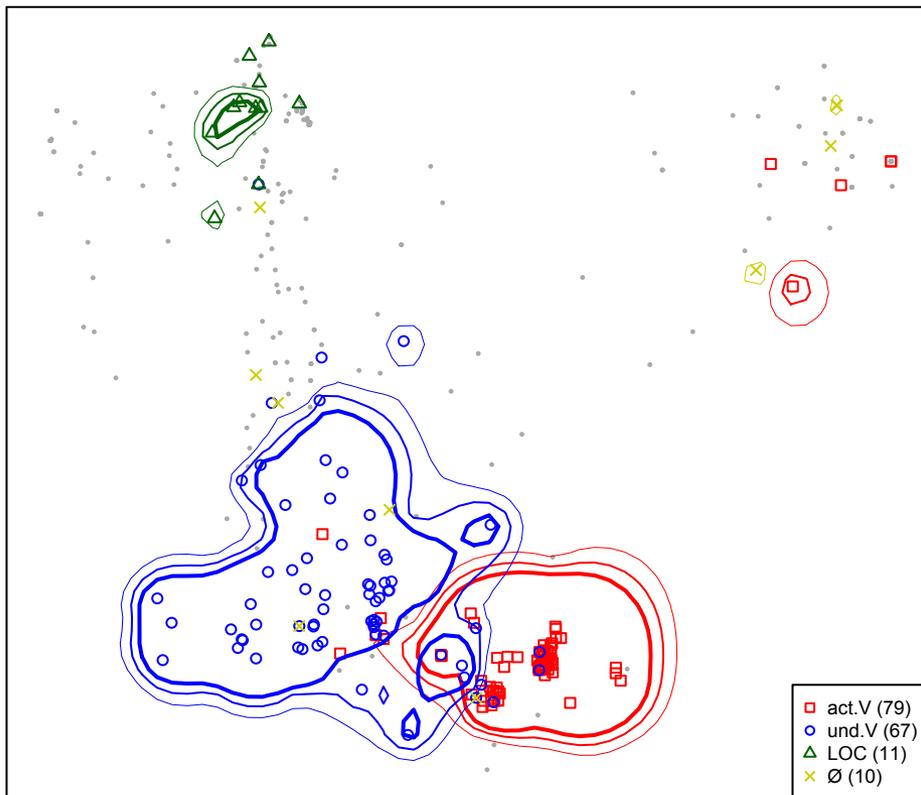
Passive (werden)_German



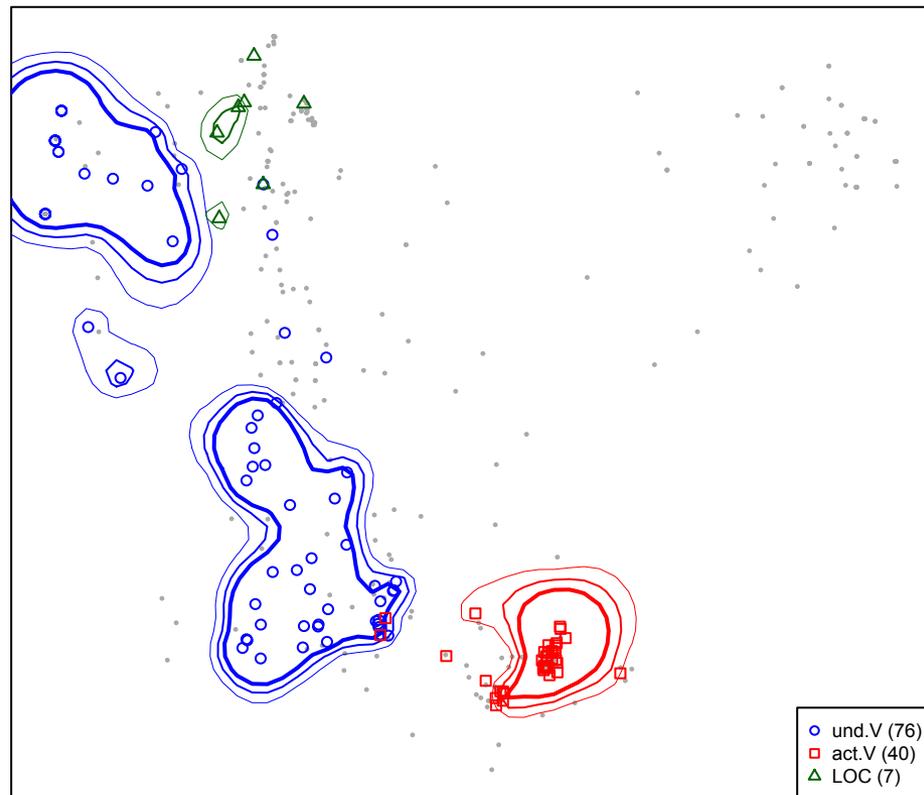
02 benefactive/possession of U (+gi)_Hoocak



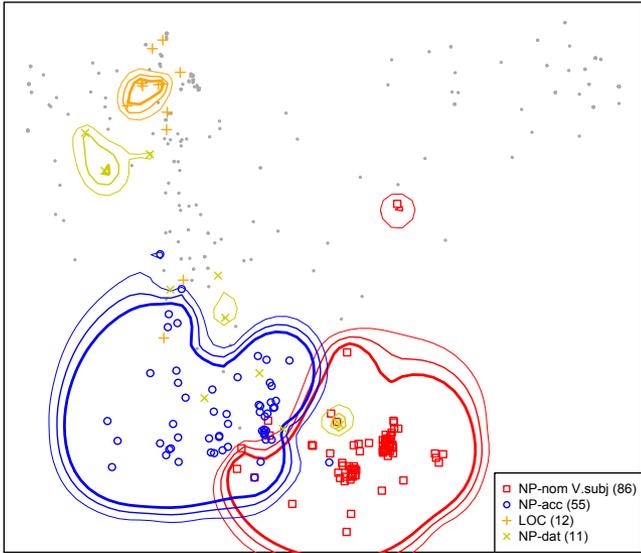
_Hoocak



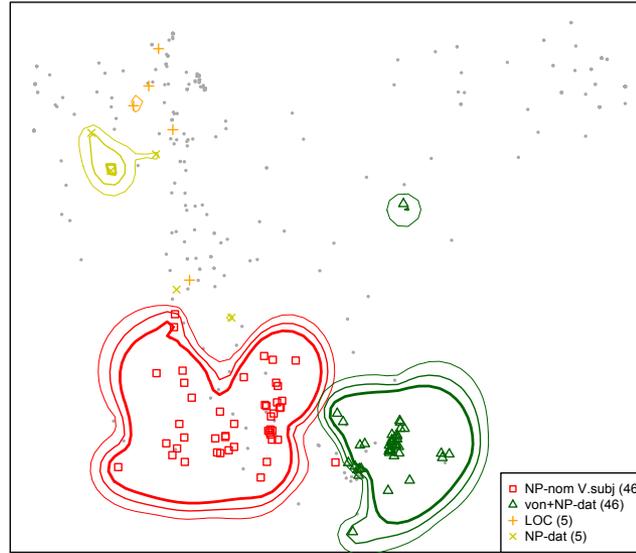
02 benefactive/possession of U (+gi)_Hoocak



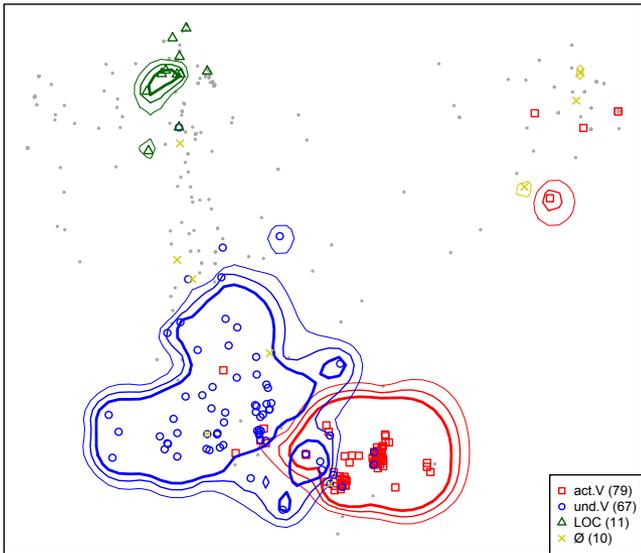
_German



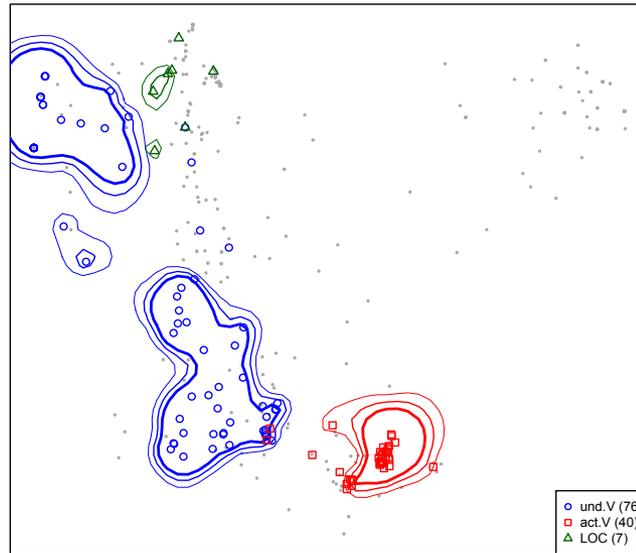
Passive (werden)_German



_Hocak



02 benefactive/possession of U (+gi)_Hocak



Difficult to compare ...

Visualize Alternation

- Another 2-dimensional representation:
 - Use horizontal axis for ‘unmarked’ alternant
 - Use vertical axis for alternant of interest
 - Each axis shows all 431 roles, always in exactly the same order (one-dimensional base map)

**alternant
of interest**

caller
grind causer
hug causer
sing causer
rain maker
meeter
cutter
cooker
leaver
wanter
heard sound
built thing
wanted thing
cover
washing instrument
affliction/disease
carry beneficiary
asked person
look beneficiary
turn instrument
break location
longed for person
absent

**all 431 microroles in the same order
(not all shown here due to space)**

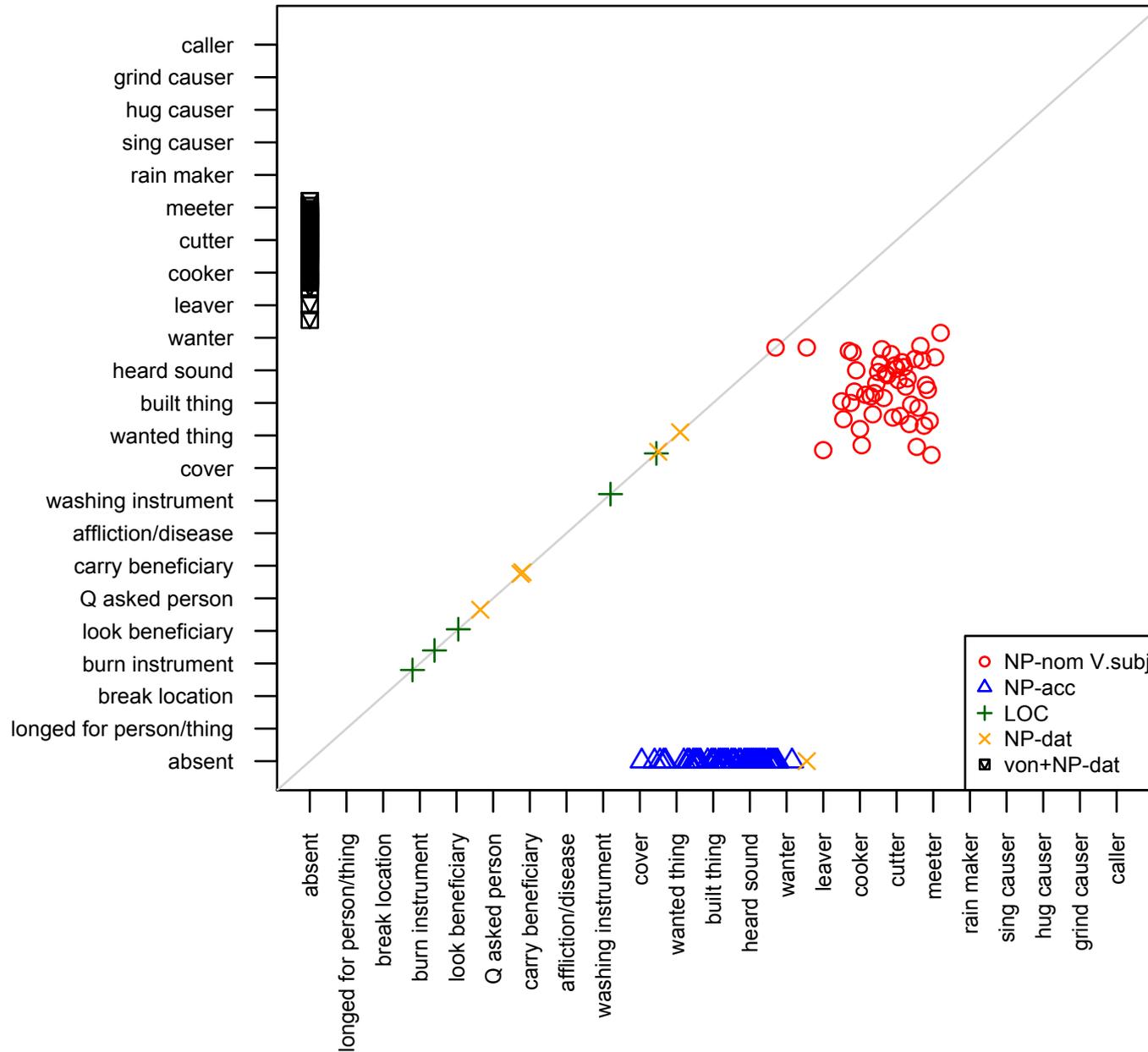
Linear order from Anti-Robinson
seriation by simulated annealing

**added
possibility
of absence**

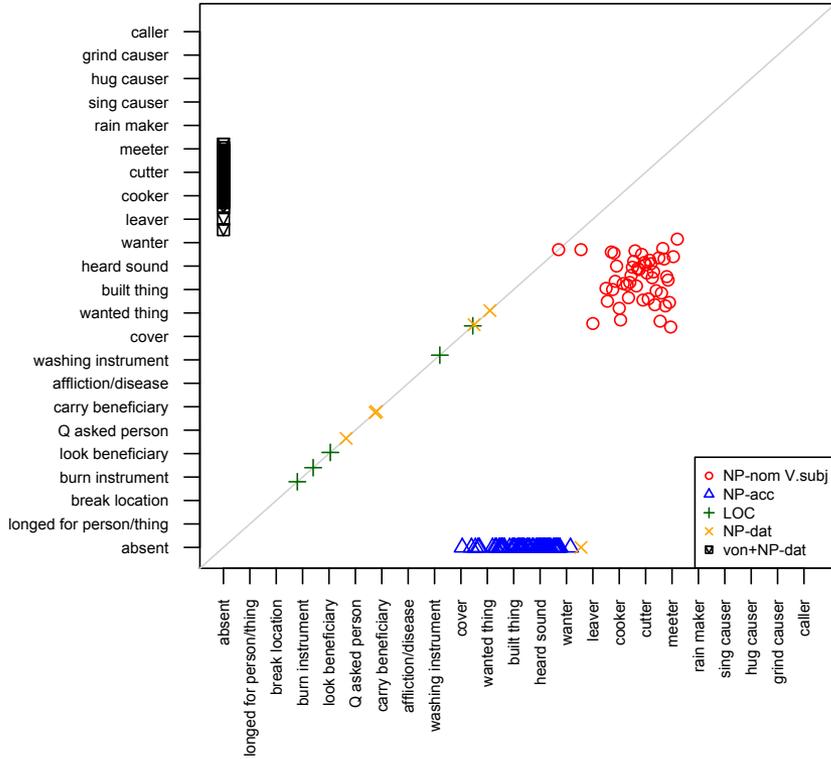
absent
longed for person
break location
turn instrument
look beneficiary
asked person
carry beneficiary
affliction/disease
washing instrument
cover
wanted thing
built thing
heard sound
wanter
leaver
cooker
cutter
meeter
rain maker
sing causer
hug causer
grind causer
caller

**'unmarked'
alternant**

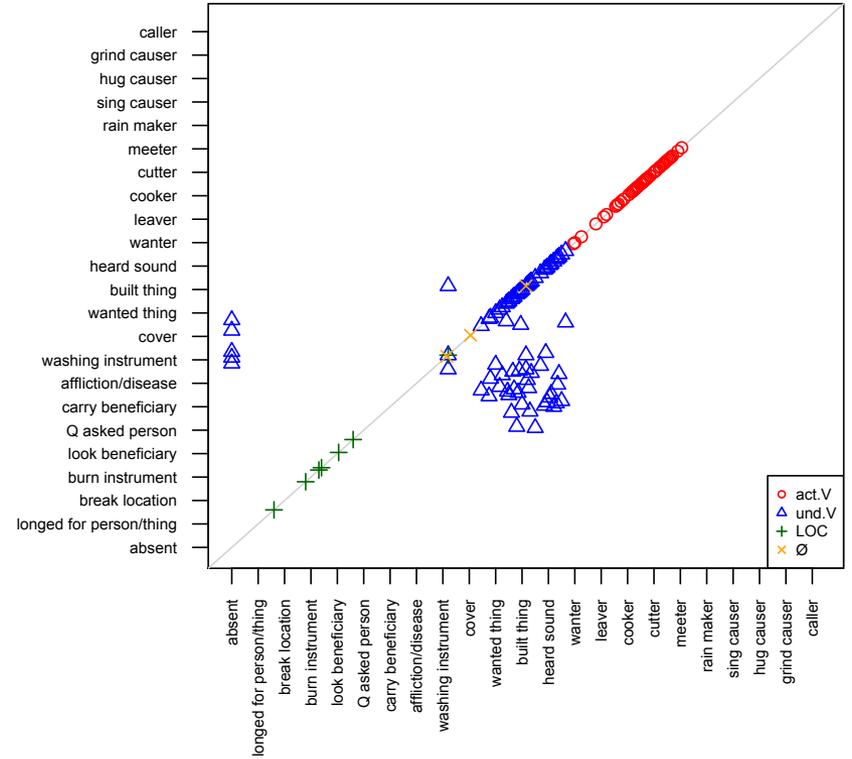
Passive (werden) from German



Passive (werden) from German



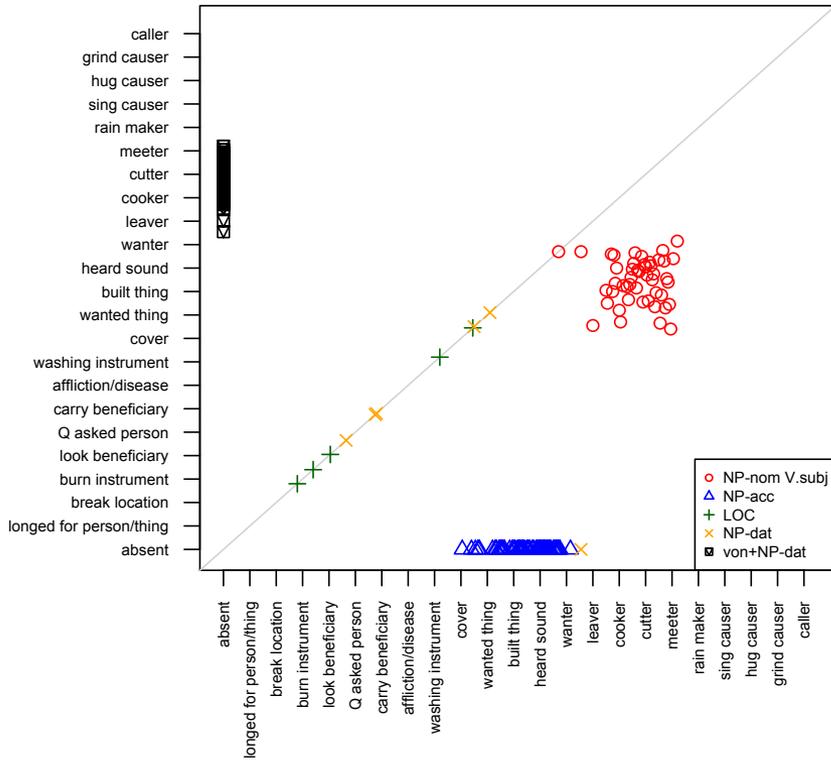
02 benefactive/possession of U (+gi)_Hoocak



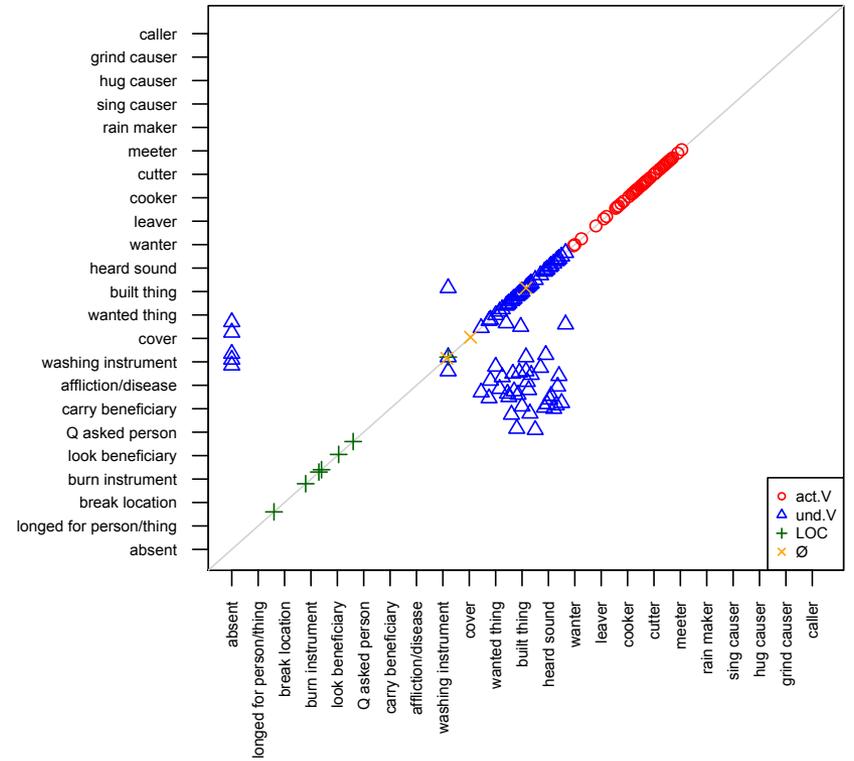
3. Typology of Alternations

- These visual representation of **Alternations** can directly be compared
- Note that the **color is not important** anymore!
- Simple measure of similarity between to alternations is the **number of overlapping dots**
- This similarity can be clustered to show typological structure

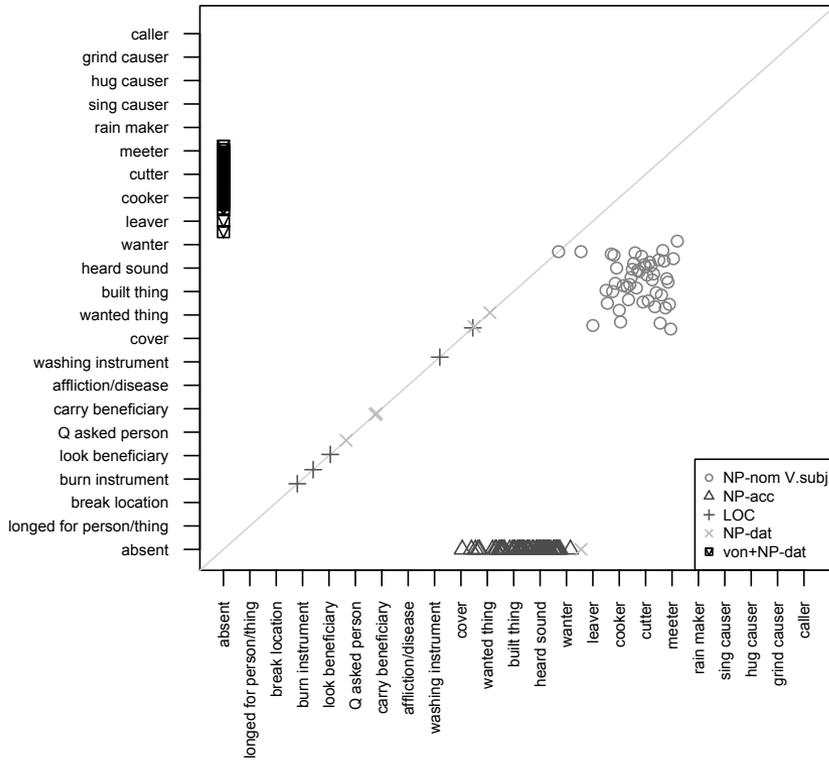
Passive (werden) from German



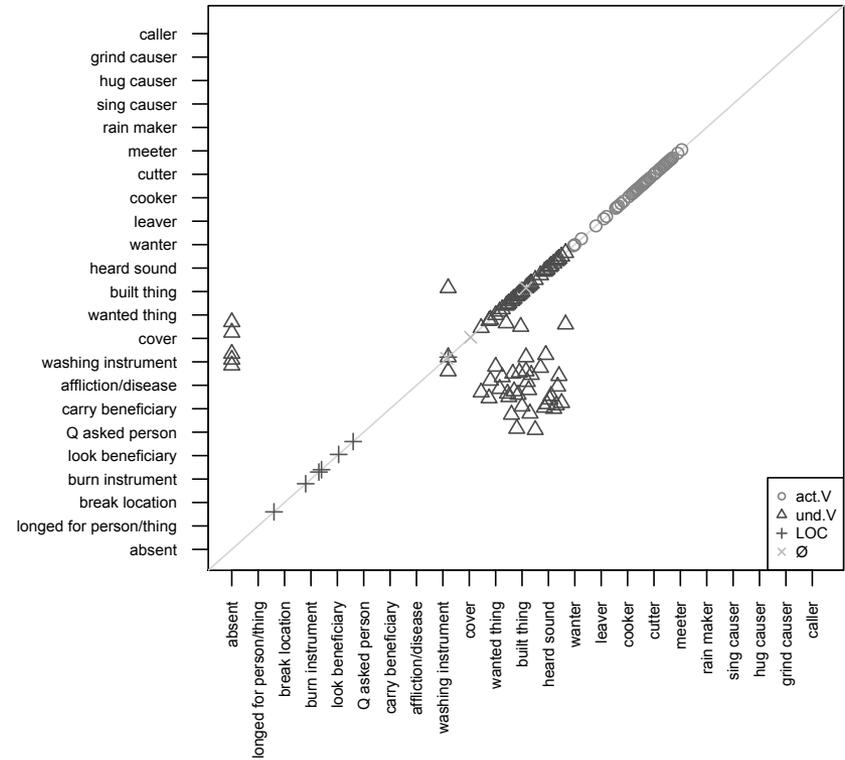
02 benefactive/possession of U (+gi)_Hoocak



Passive (werden) from German

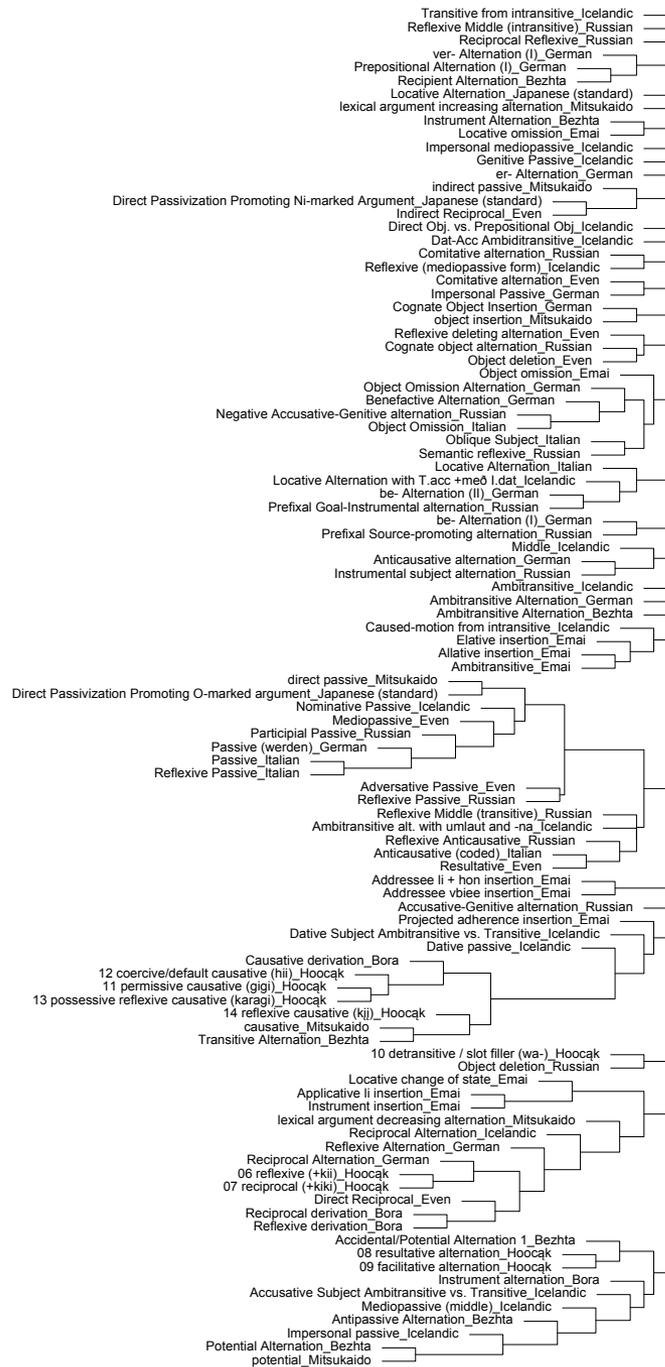


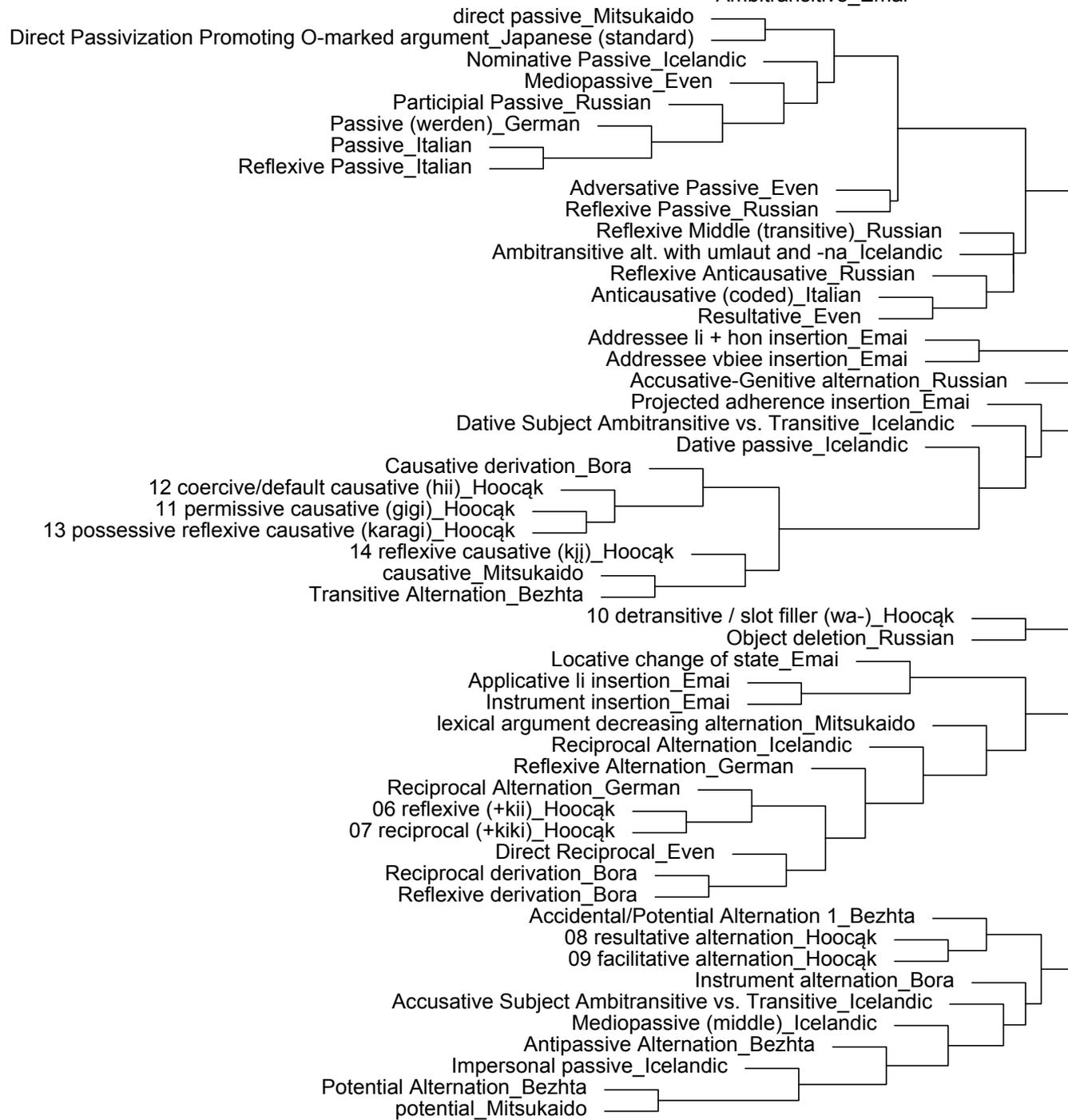
02 benefactive/possession of U (+gi)_Hoocak



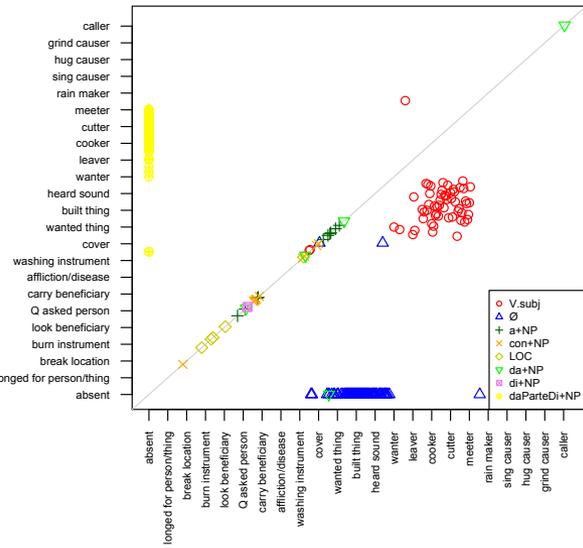
How many points are shared?

That is a combination of how the roles are linked between unmarked and marked alternant, but also of the lexical distribution of an alternation

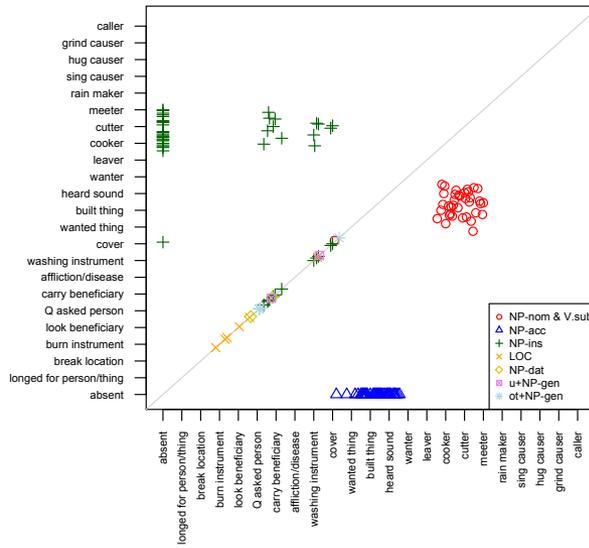




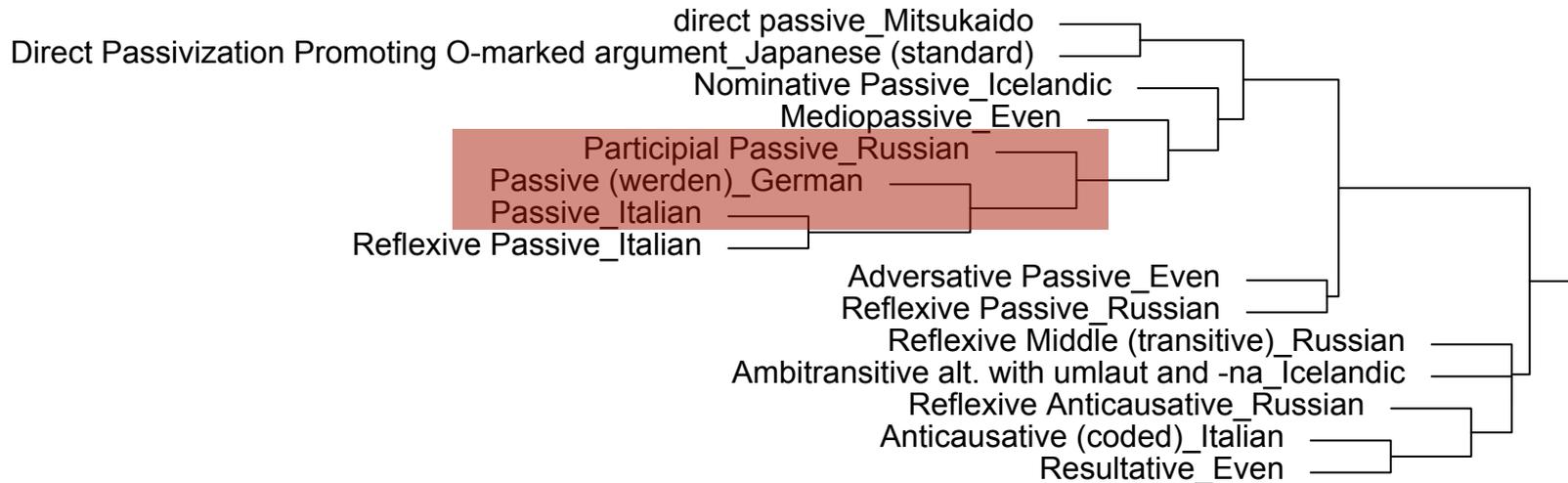
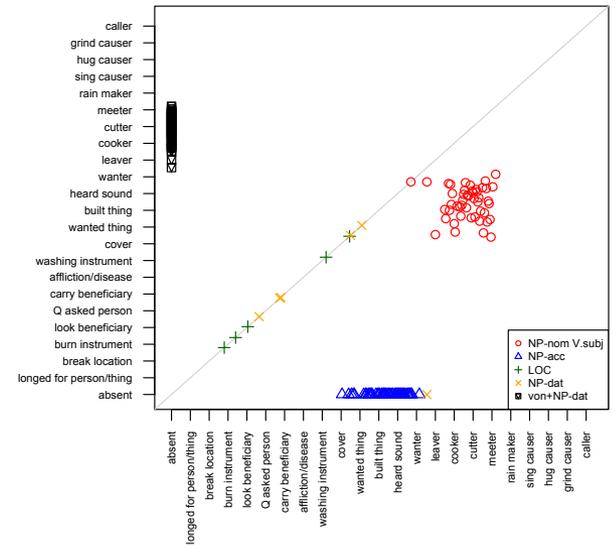
Passive from Italian



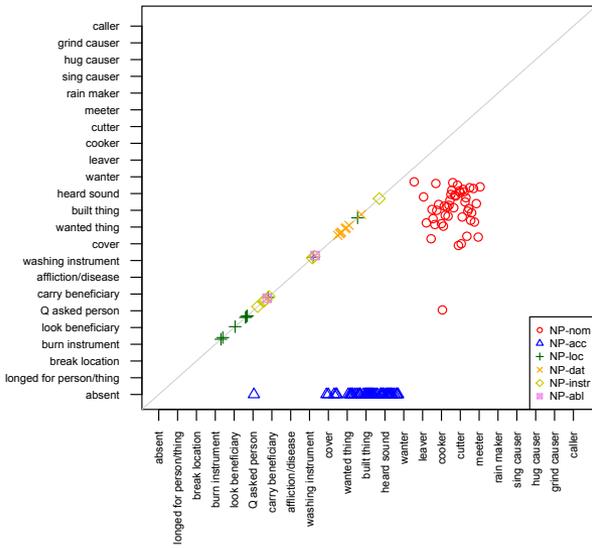
Participial Passive from Russian



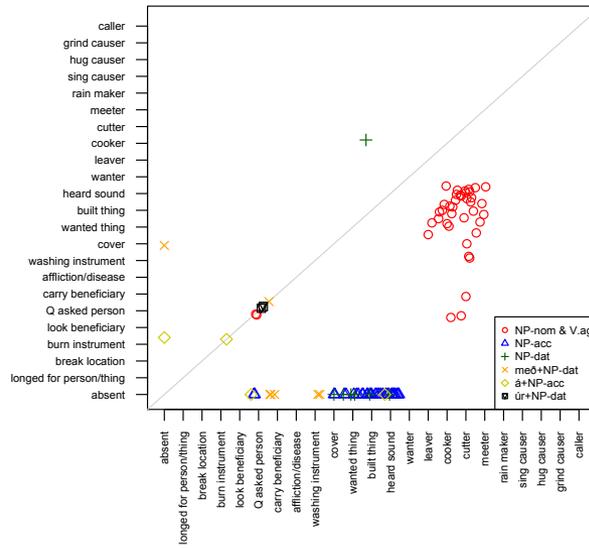
Passive (werden) from German



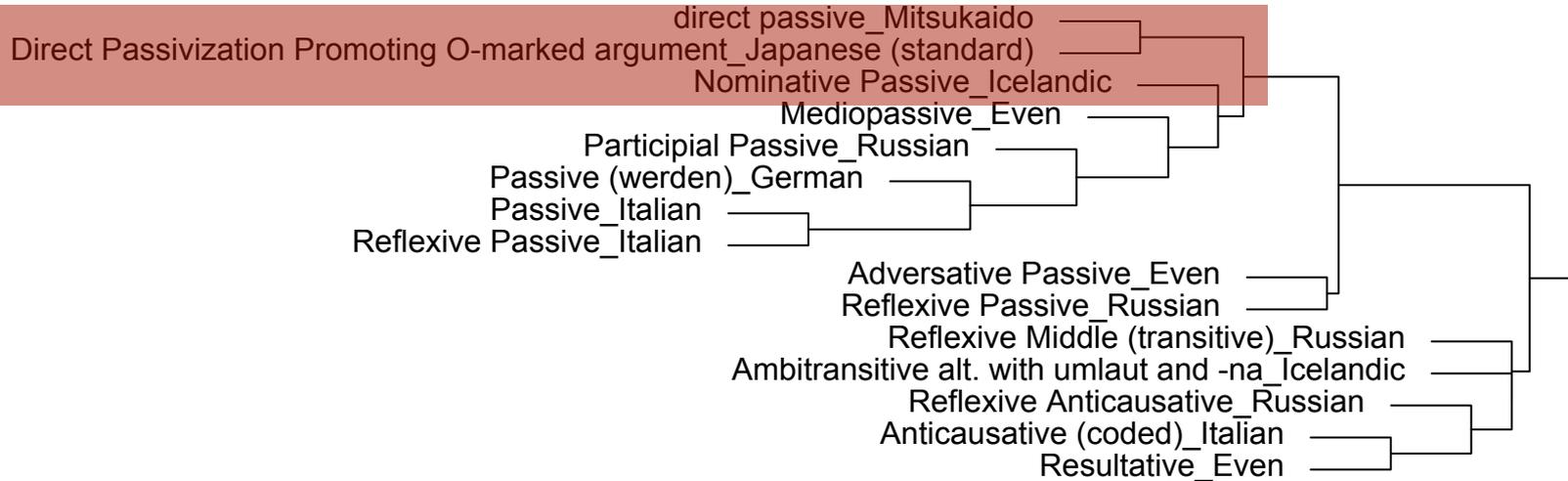
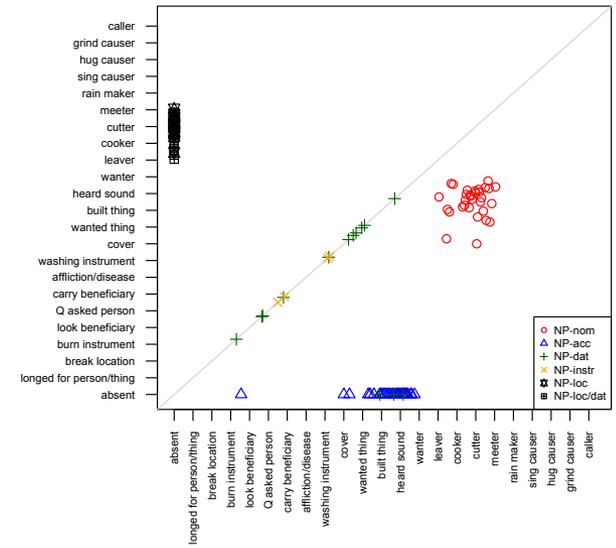
Direct Passivization Promoting O-marked argument from Japanese



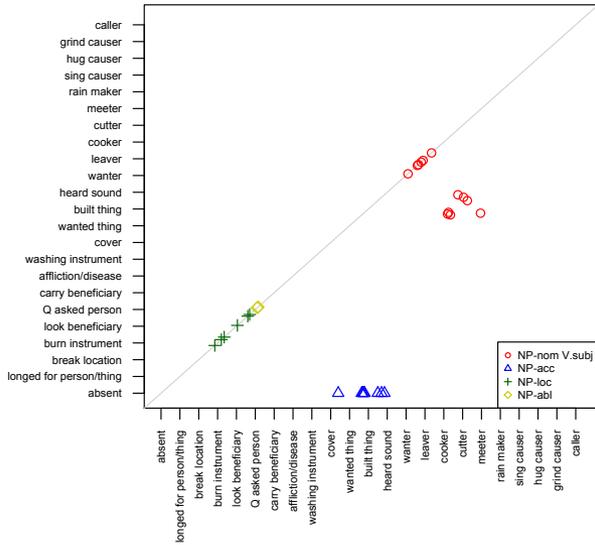
Nominative Passive_Icelandic



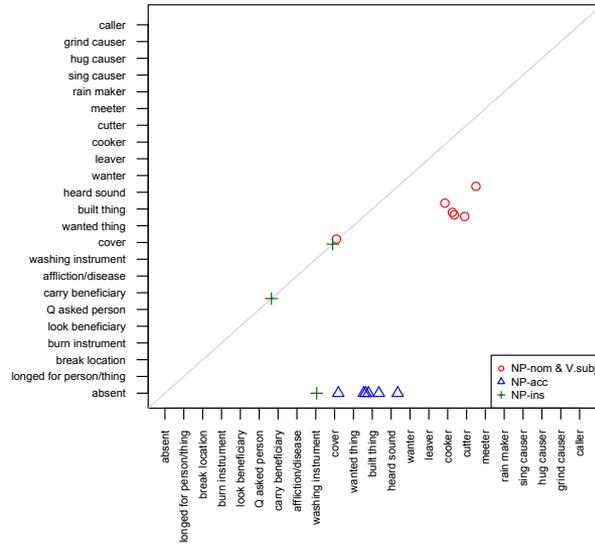
direct passive from Mitsukaido



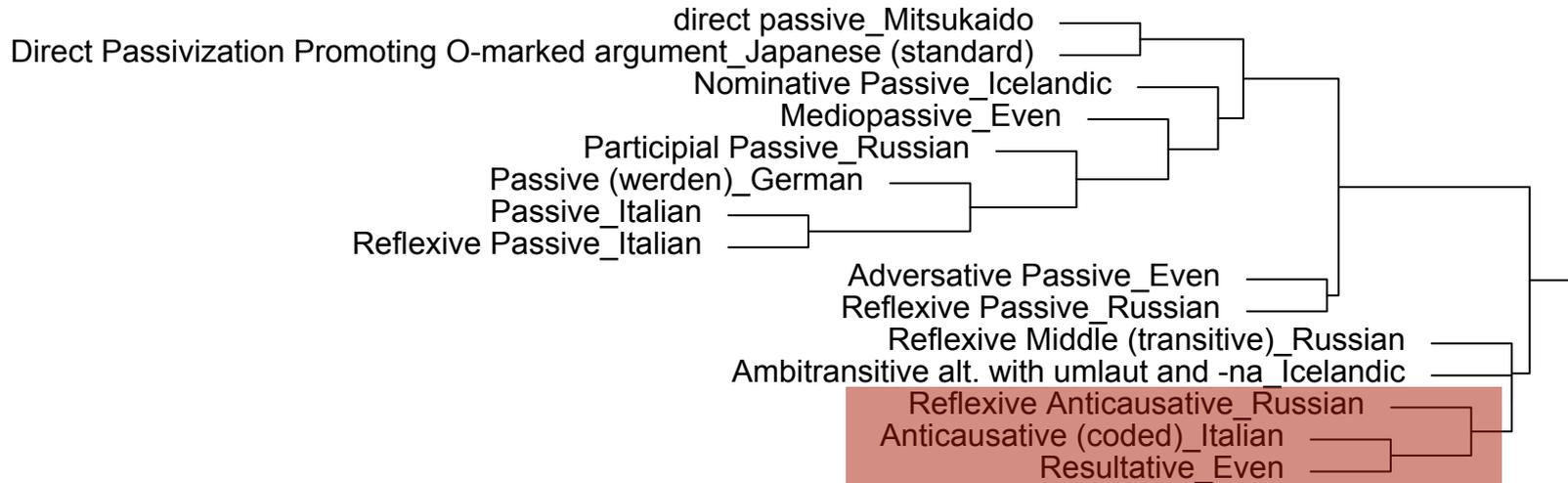
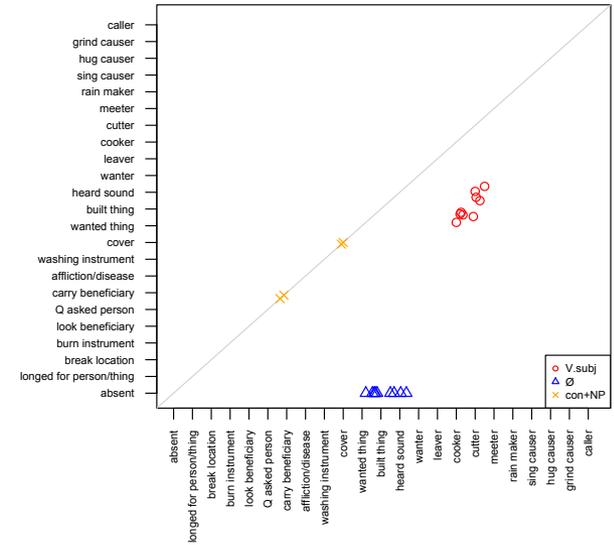
Resultative from Even



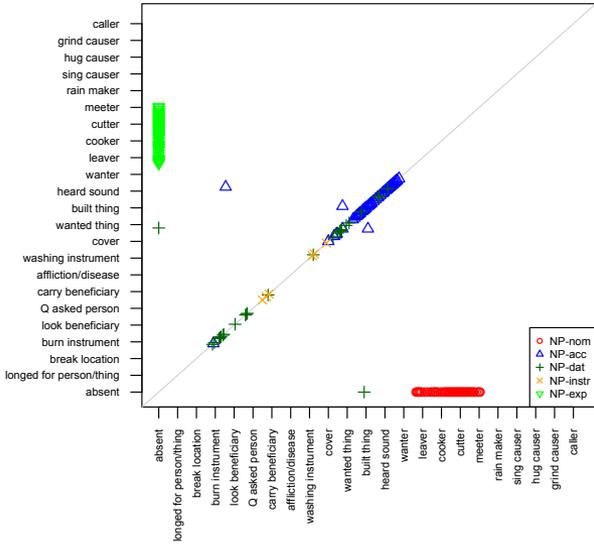
Reflexive Anticausative from Russian



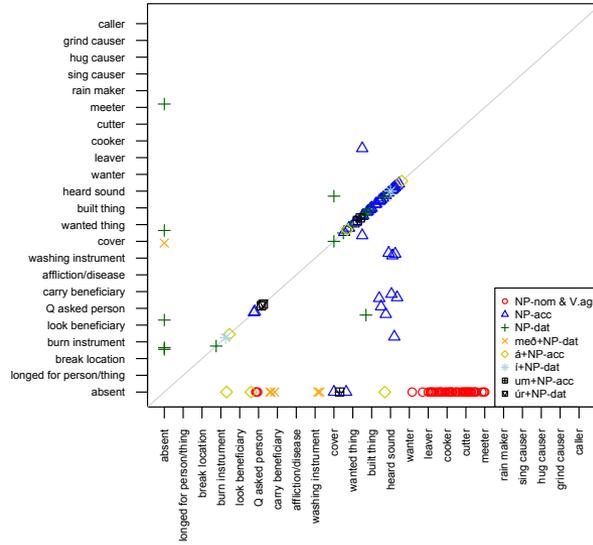
Anticausative (coded) from Italian



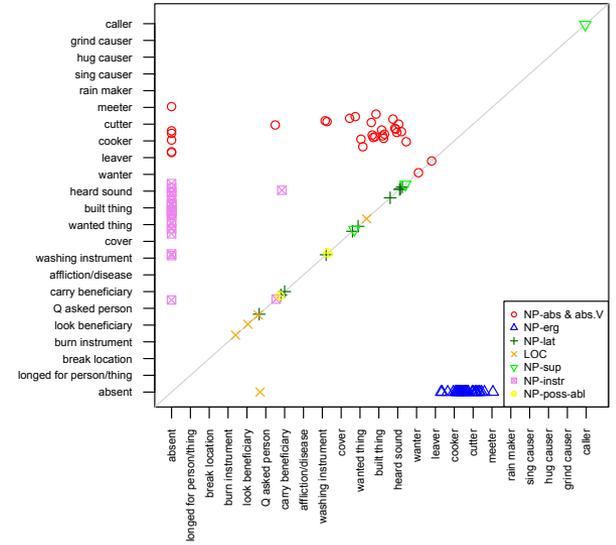
potential from Mitsukaido



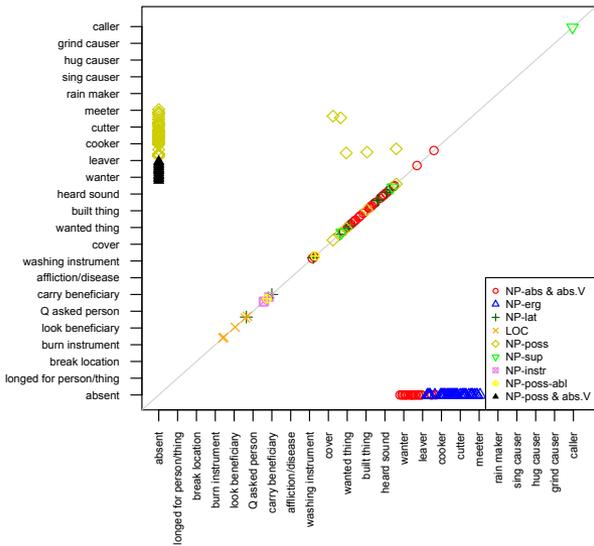
Impersonal passive_Icelandic



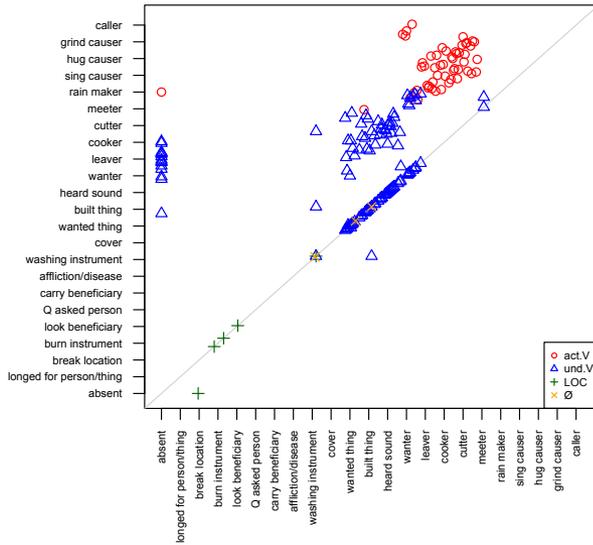
Antipassive Alternation from Bezhta



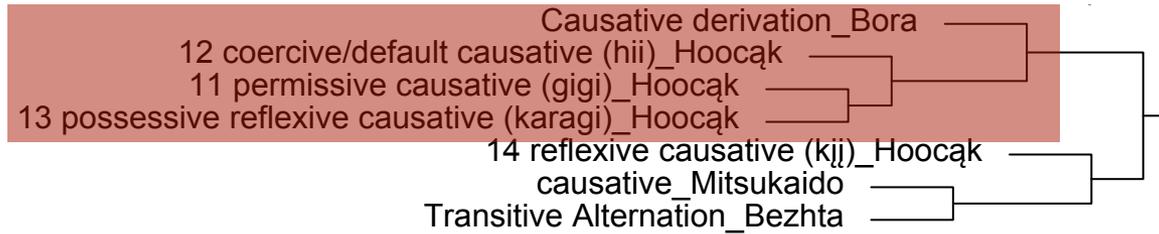
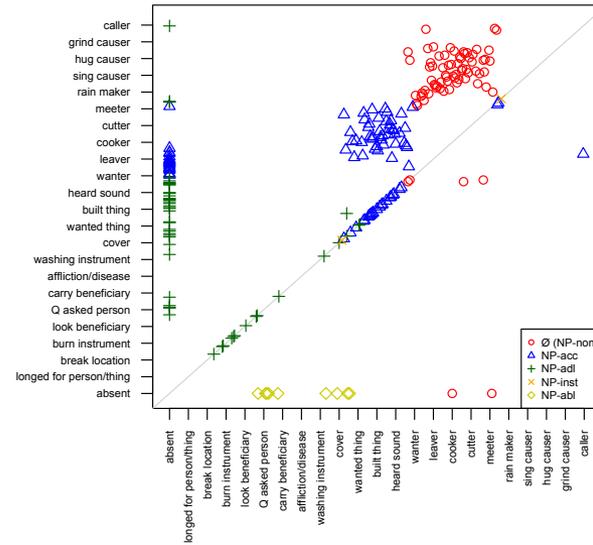
Potential Alternation from Bezhta



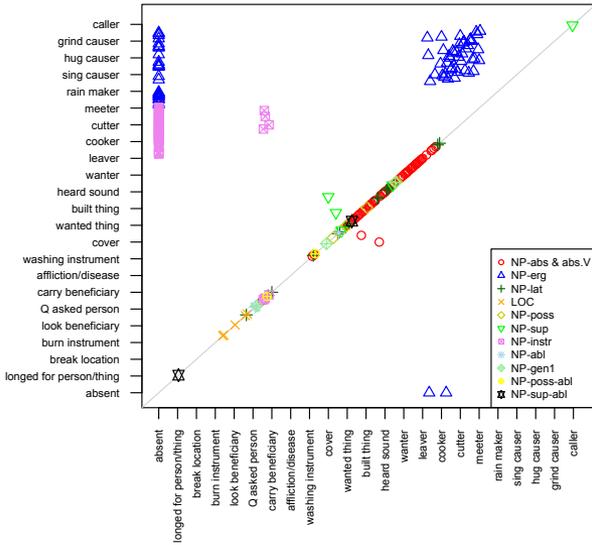
12 coercive/default causative (hii)_Hooçaq



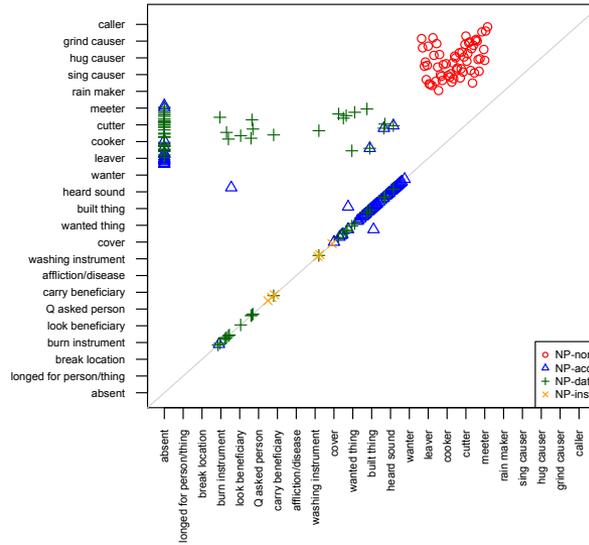
Causative derivation from Bora



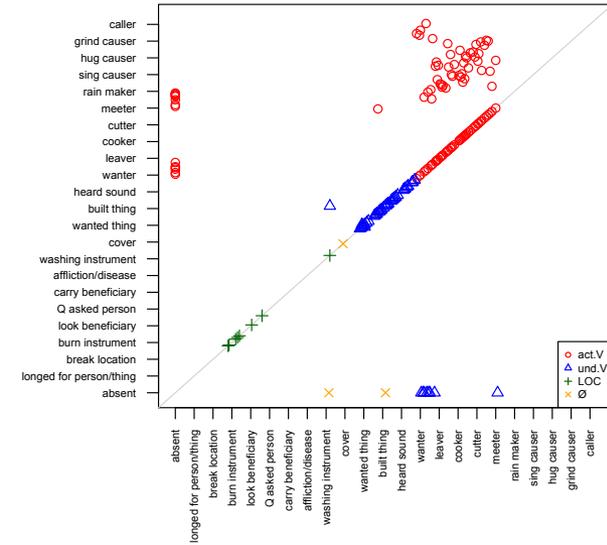
Transitive Alternation from Bezhta



causative from Mitsukaido



14 reflexive causative (kij)_Hooçak



Causative derivation_Bora

12 coercive/default causative (hii)_Hooçak

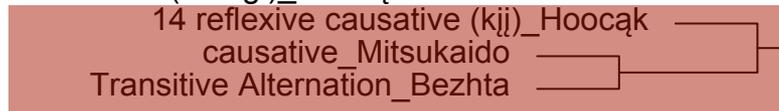
11 permissive causative (gigi)_Hooçak

13 possessive reflexive causative (karagi)_Hooçak

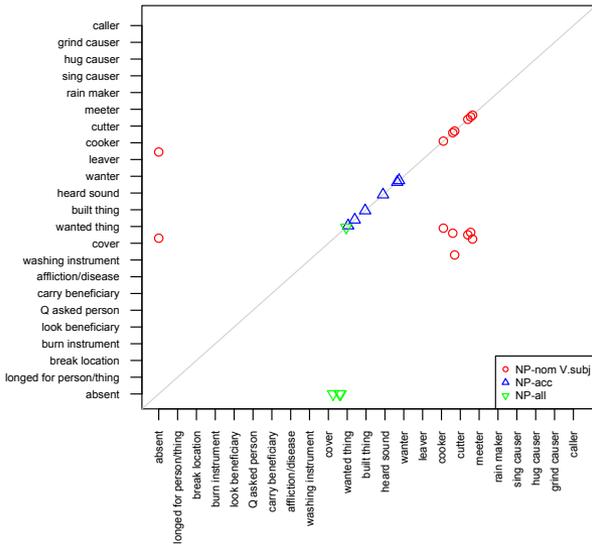
14 reflexive causative (kij)_Hooçak

causative_Mitsukaido

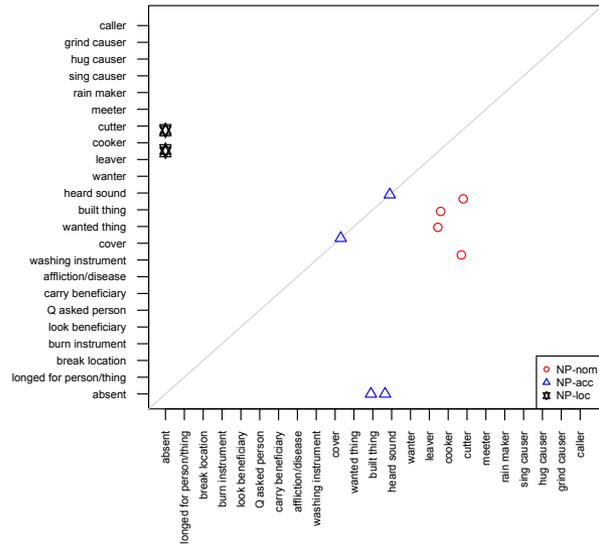
Transitive Alternation_Bezhta



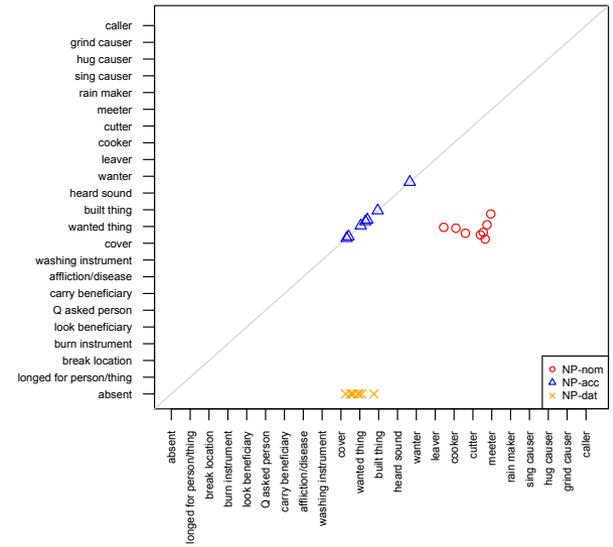
Indirect Reciprocal from Even



indirect passive from Mitsukaido



ssivization Promoting Ni-marked Argument from Japan



Summary

- Objective way of finding **alternation types**
- No typological comparison of language-specific coding necessary
- We would get exactly the same results if **all language-specific constructions were characterized by a random code**
- Likewise, **the actual meaning of the microroles is unimportant for this method**
- Based completely on **mapping between microroles and language-specific coding**