

Similarity Semantics

Towards a Formal Calculus for Meaning
without Truth Conditions

Michael Cysouw

Bernhard Wälchli (Stockholm)

Meaning

- We as humans “know” what meaning is
- But: can we operationalize meaning?
- Logic seemed the way to go for centuries
- Vector algebra seems to be a more practical option

Die Bedeutung eines Wortes ist
sein Gebrauch in der Sprache

Ludwig Wittgenstein
Philosophische Untersuchungen, no. 43 (1953)

You shall know a word
by the company it keeps!

John Rupert Firth
Studies in Linguistic Analysis (1957)



... the concept of a translation process in which, in determining meaning for a word, account is taken of the immediate context.

It would hardly be practical to do this by means of a generalized dictionary which contains all possible phrases $2N+1$ words long: for the number of such phrases is horrifying, even to a modern electronic computer

Warren Weaver
Translation (1949)



... if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C.

**In other words, difference of meaning
correlates with difference of distribution**

Zellig Harris

Methods in Structural Linguistics (1951)



and then ...

- Information Retrieval
- Vector Space Model
- Latent Semantic Analysis (1988)
- Google (1998)

Proposal

- Define the meaning of a linguistic form as the set of all contexts in which it occurs
- Under this strongly extensionalistic definition of meaning, variation in meaning becomes readily measurable

Proposal

- Define the **meaning** of a linguistic form as the set of all contexts in which it occurs
- Under this strongly extensionalistic definition of **meaning**, variation in **meaning** becomes readily measurable

Proposal

- Define the **property M** of a linguistic form as the set of all contexts in which it occurs
- Under this strongly extensionalistic definition of **property M**, variation in **property M** becomes readily measurable

Semantics

- **Similarity Semantics**
Meaning of linguistic forms is defined relationally to each other (“context” is taken as purely linguistic)
- **Exemplar Semantics**
Meaning of linguistic forms is (re-)instantiated and (re-)produced in individual utterances,

Usage

- **Within one language**
count co-occurrences of linguistic forms
(within documents, sentences, word-
windows, etc.)
- **Comparing languages**
investigate the co-occurrence of linguistic
forms across languages in parallel texts

Universal Declaration of Human Rights





in, im German

in Dutch

in English

yn Frisian

in Afrikaans

in Scots

אין Yiddish

í Icelandic

i Swedisch

i Nynorsk

i Bokmål

í Faroese

i Danish

Article 10

English: “everyone is entitled in full equality to a fair and public hearing by an independent and impartial tribunal , in the determination of his rights and obligations and of any criminal charge against him .”

Afrikaans: “elkeen het , in volle gelykheid , die reg tot ’n regverdige en openbare verhoor deur ’n onafhanklike en objektiewe tribunaal , in die bepaling van sy regte en verpligtinge en die ondersoek van enige kriminele saak teen hom .”

German: “jeder hat *bei* der feststellung seiner rechte und pflichten sowie bei einer gegen ihn erhobenen strafrechtlichen beschuldigung in voller gleichheit anspruch auf ein gerechtes und öffentliches verfahren vor einem unabhängigen und unparteiischen gericht .”

Danish: “enhver har *under* fuld ligeberettigelse krav på en retfærdig og offentlig behandling ved en uafhængig og upartisk domstol , *når* der skal træffes en afgørelse med hensyn til hans rettigheder og forpligtelser og med hensyn til en hvilken som helst mod ham rettet strafferetslig anklage .”

Article 10

English: “everyone is entitled **in** full equality to a fair and public hearing by an independent and impartial tribunal , **in** the determination of his rights and obligations and of any criminal charge against him .”

Afrikaans: “elkeen het , **in** volle gelykheid , die reg tot ’n regverdige en openbare verhoor deur ’n onafhanklike en objektiewe tribunaal , **in** die bepaling van sy regte en verpligtinge en die ondersoek van enige kriminele saak teen hom .”

German: “jeder hat **bei** der feststellung seiner rechte und pflichten sowie bei einer gegen ihn erhobenen strafrechtlichen beschuldigung **in** voller gleichheit anspruch auf ein gerechtes und öffentliches verfahren vor einem unabhängigen und unparteiischen gericht .”

Danish: “enhver har **under** fuld ligeberettigelse krav på en retfærdig og offentlig behandling ved en uafhængig og upartisk domstol , **når** der skal træffes en afgørelse med hensyn til hans rettigheder og forpligtelser og med hensyn til en hvilken som helst mod ham rettet strafferetslig anklage .”

Article 10

English: “everyone is entitled **in full equality** to a fair and public hearing by an independent and impartial tribunal , in the determination of his rights and obligations and of any criminal charge against him .”

Afrikaans: “elkeen het , **in volle gelykheid** , die reg tot ’n regverdige en openbare verhoor deur ’n onafhanklike en objektiewe tribunaal , in die bepaling van sy regte en verpligtinge en die ondersoek van enige kriminele saak teen hom .”

German: “jeder hat bei der feststellung seiner rechte und pflichten sowie bei einer gegen ihn erhobenen strafrechtlichen beschuldigung **in voller gleichheit** anspruch auf ein gerechtes und öffentliches verfahren vor einem unabhängigen und unparteiischen gericht .”

Danish: “enhver har **under fuld ligeberettigelse** krav på en retfærdig og offentlig behandling ved en uafhængig og upartisk domstol , når der skal træffes en afgørelse med hensyn til hans rettigheder og forpligtelser og med hensyn til en hvilken som helst mod ham rettet strafferetslig anklage .”

Article 10

English: “everyone is entitled in full equality to a fair and public hearing by an independent and impartial tribunal , **in the determination of his rights** and obligations and of any criminal charge against him .”

Afrikaans: “elkeen het , in volle gelykheid , die reg tot ’n regverdige en openbare verhoor deur ’n onafhanklike en objektiewe tribunaal , **in die bepaling van sy regte** en verpligtinge en die ondersoek van enige kriminele saak teen hom .”

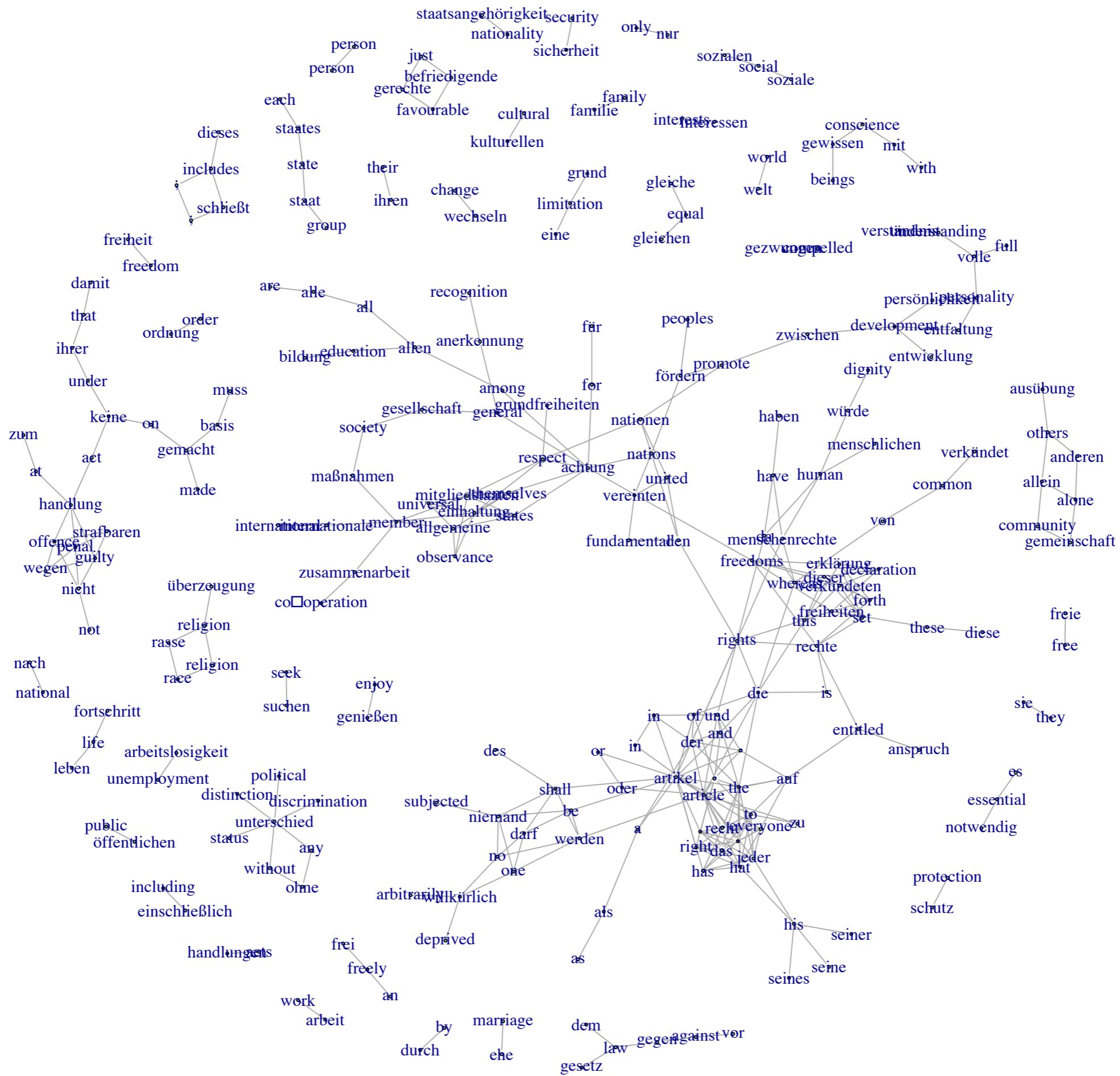
German: “jeder hat **bei der feststellung seiner rechte** und pflichten sowie bei einer gegen ihn erhobenen strafrechtlichen beschuldigung in voller gleichheit anspruch auf ein gerechtes und öffentliches verfahren vor einem unabhängigen und unparteiischen gericht .”

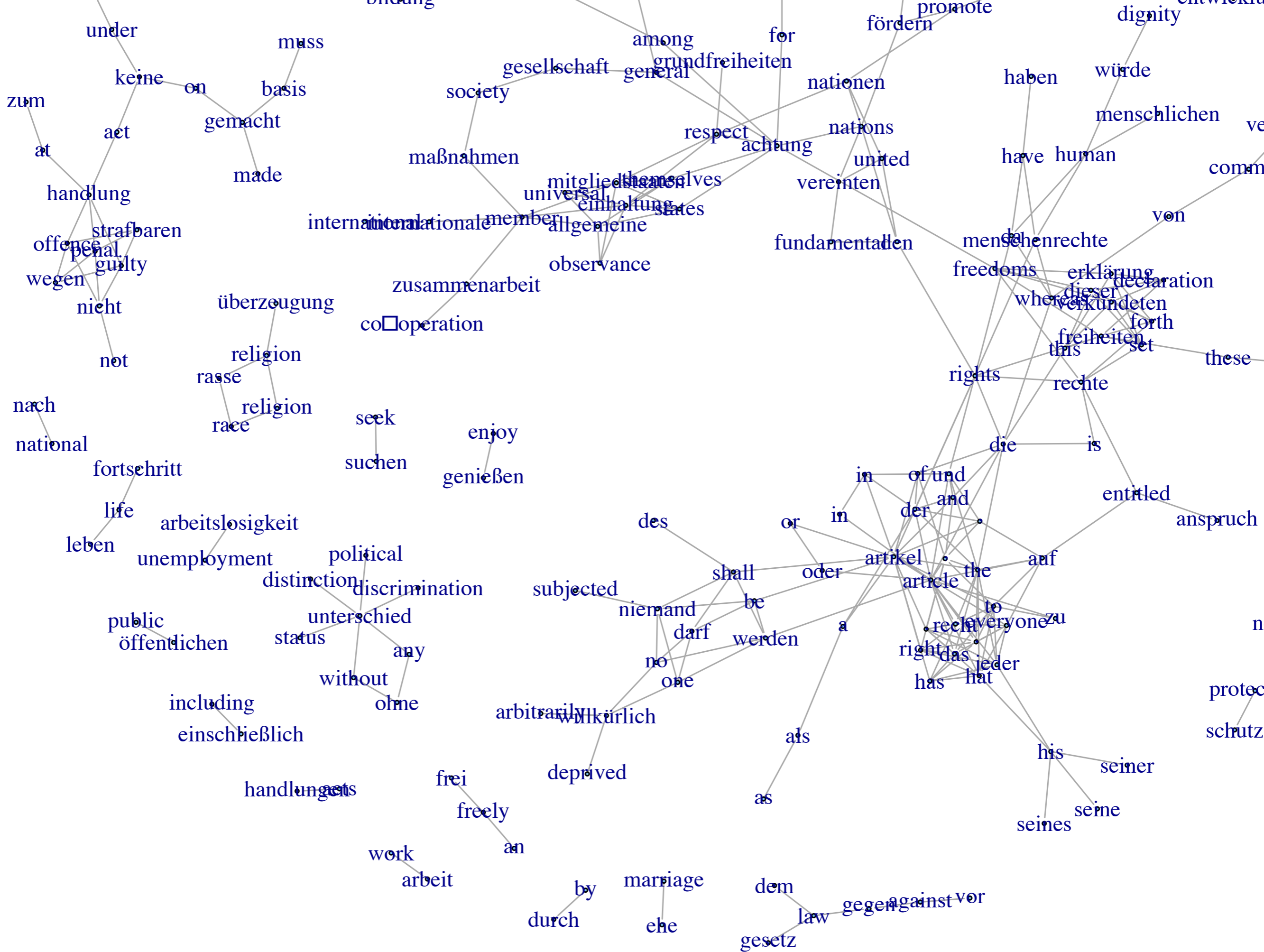
Danish: “enhver har under fuld ligeberettigelse krav på en retfærdig og offentlig behandling ved en uafhængig og upartisk domstol , **når der skal træffes en afgørelse med hensyn til hans rettigheder** og forpligtelser og med hensyn til en hvilken som helst mod ham rettet strafferetslig anklage .”

- For each word from each language, compute how often it co-occurs with each word in each other language
- Only for the 58 paragraphs in the declaration
- significance of co-occurrence of word w and word v based on Poisson distribution:

$$sig(w, v) = \lambda^k \cdot e^{-\lambda} \cdot k!$$

$\lambda = (\text{frequency of word } w) \cdot (\text{frequency of word } v) \div (\text{number of contexts})$
 $k = \text{frequency of co-occurrence of words } w \text{ and } v$





Preamble, sentence 5

English:

“whereas the peoples of the united nations have **in** the charter reaffirmed their faith **in** fundamental human rights , **in** the dignity and worth of the human person and **in** the equal rights of men and women and have determined to promote social progress and better standards of life **in** larger freedom ,”

German:

“da die völker der vereinten nationen **in** der charta ihren glauben an die grundlegenden menschenrechte , an die würde und den wert der menschlichen person und an die gleichberechtigung von mann und frau erneut bekräftigt und beschlossen haben , den sozialen fortschritt und bessere lebensbedingungen **in** größerer freiheit zu fördern ,”

Preamble, sentence 5

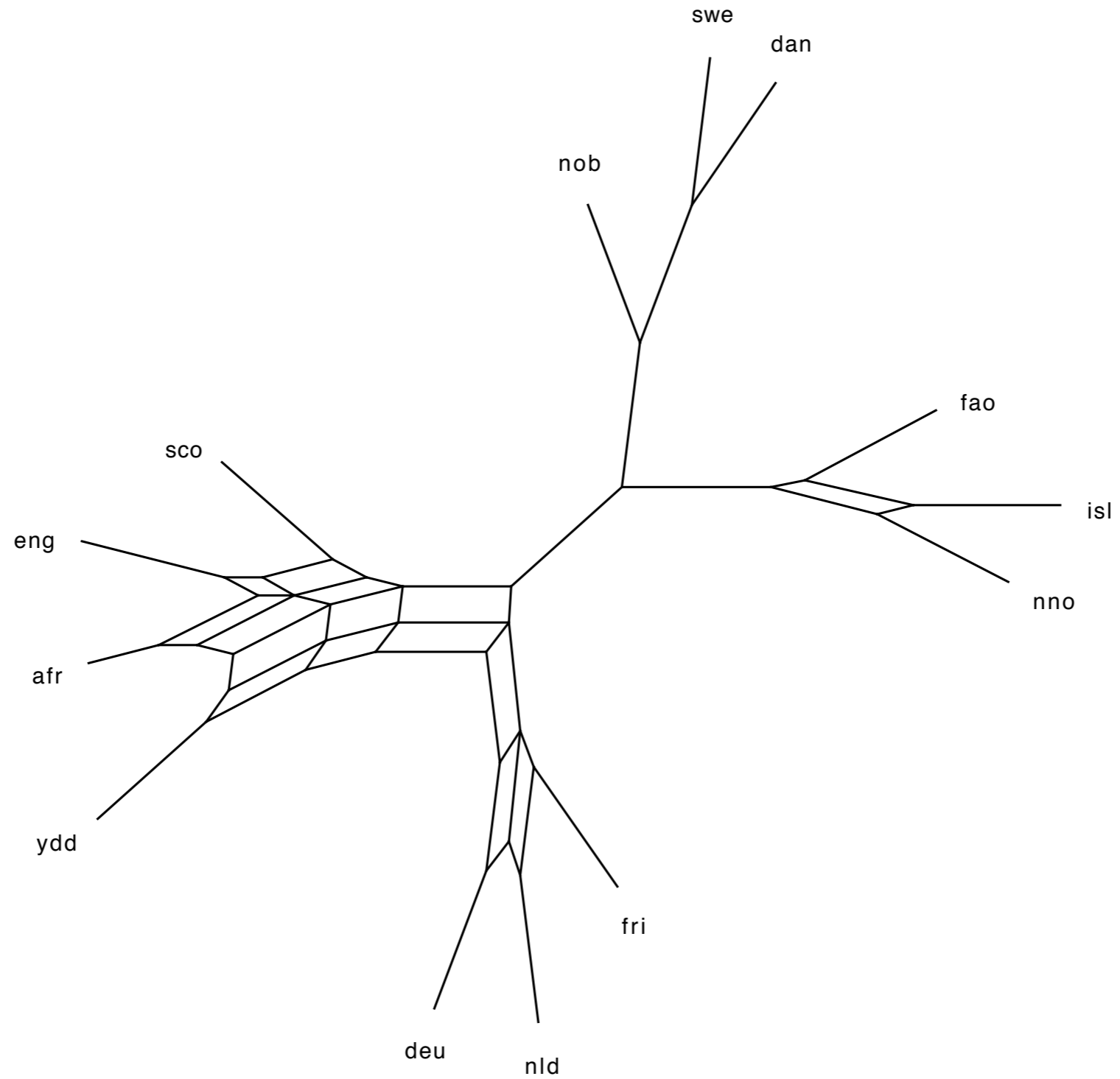
English:

“whereas the peoples of the **united nations** have **in** the **charter** reaffirmed their faith **in** fundamental human rights , **in** the dignity and worth of the human person and **in** the equal rights of men and women and have determined to promote social progress and better standards of life **in larger freedom** ,”

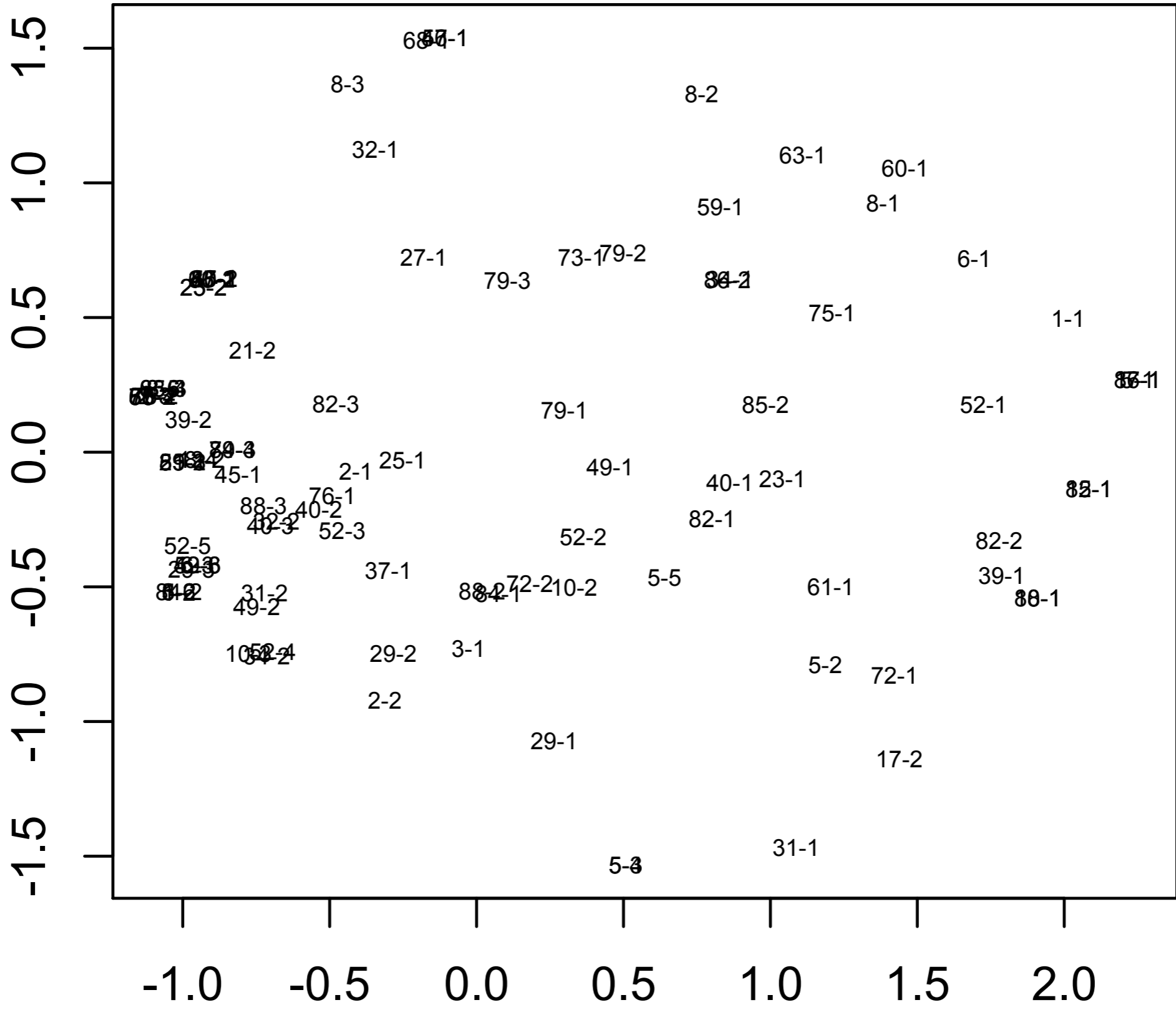
German:

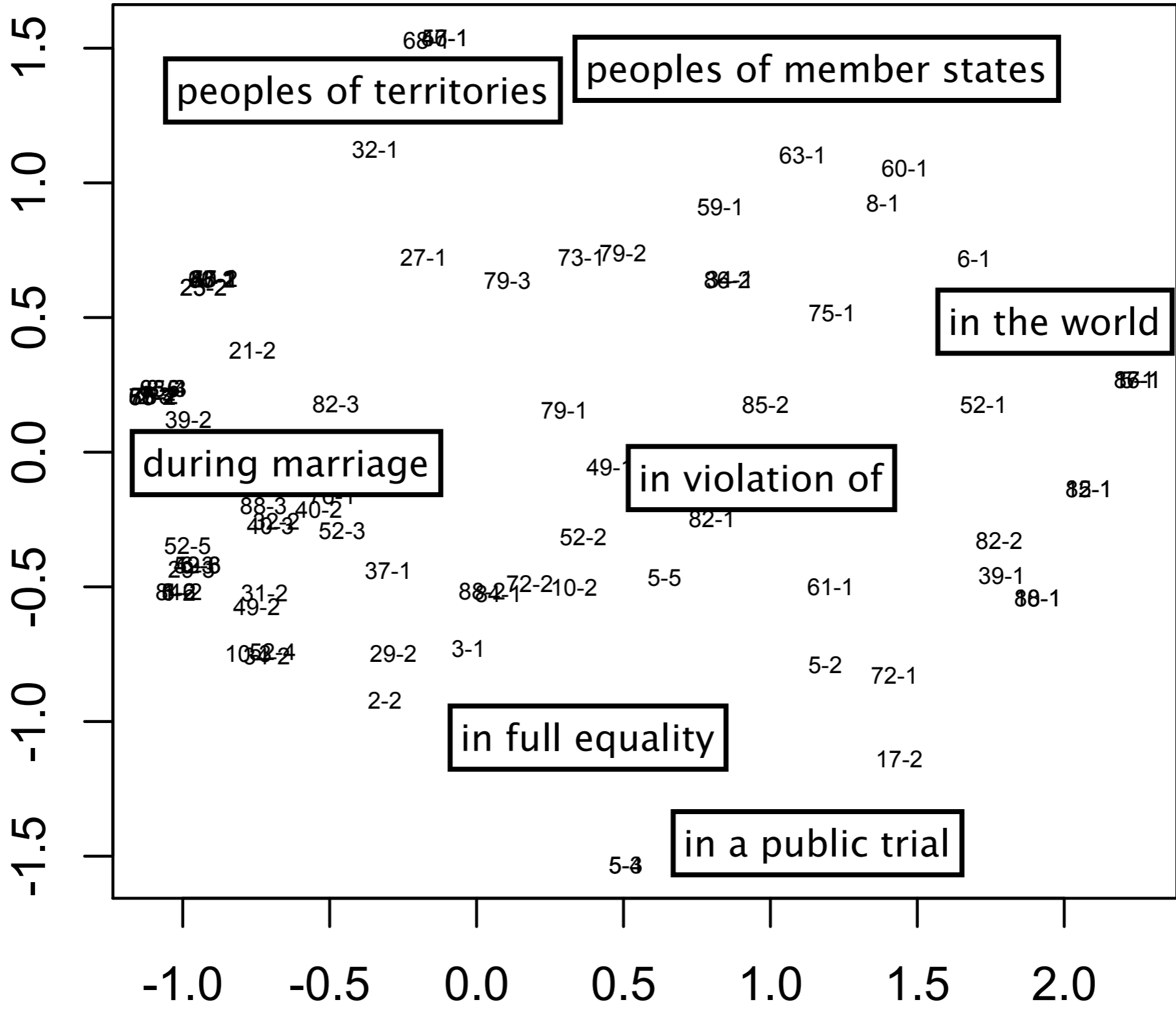
“da die völker der **vereinten nationen in** der **charta** ihren glauben an die grundlegenden menschenrechte , an die würde und den wert der menschlichen person und an die gleichberechtigung von mann und frau erneut bekräftigt und beschlossen haben , den sozialen fortschritt und bessere lebensbedingungen **in größerer freiheit** zu fördern ,”

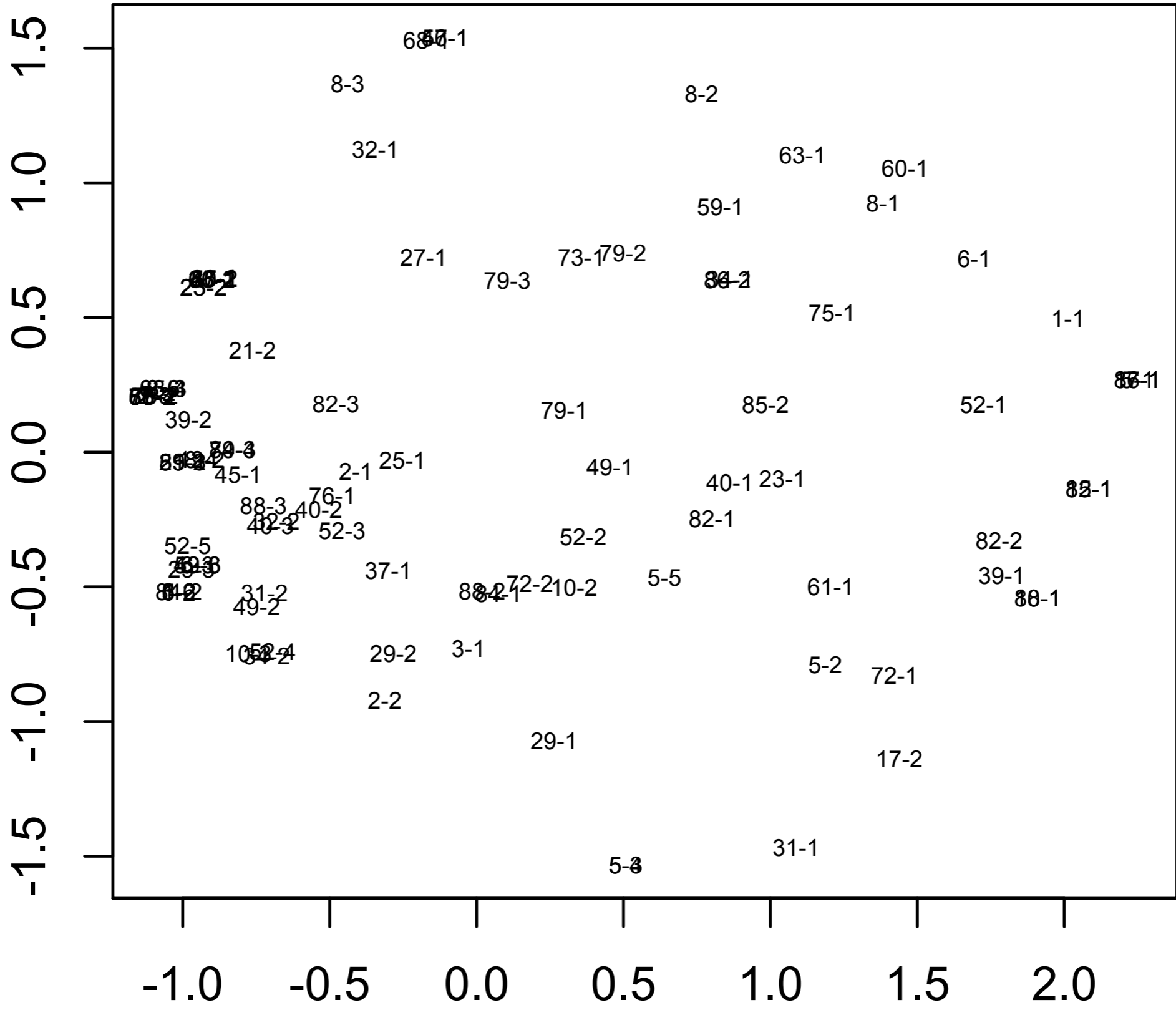
deu	2						
eng	5						
nld	5						
fri	6						
afr	5						
ydd	4						
sco	5						
nno	3						
nob							
swe							
fao	3						
dan							
isl							

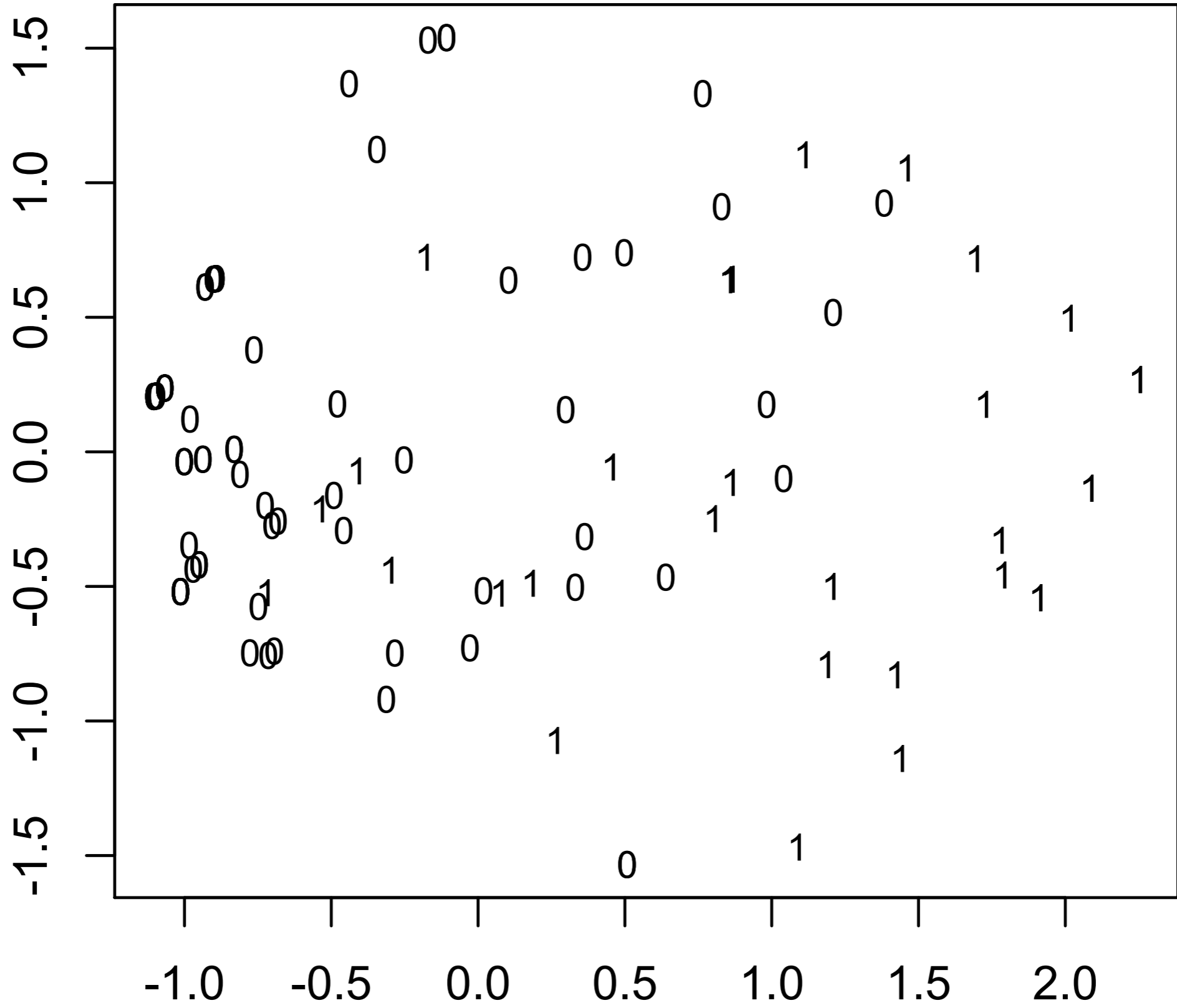


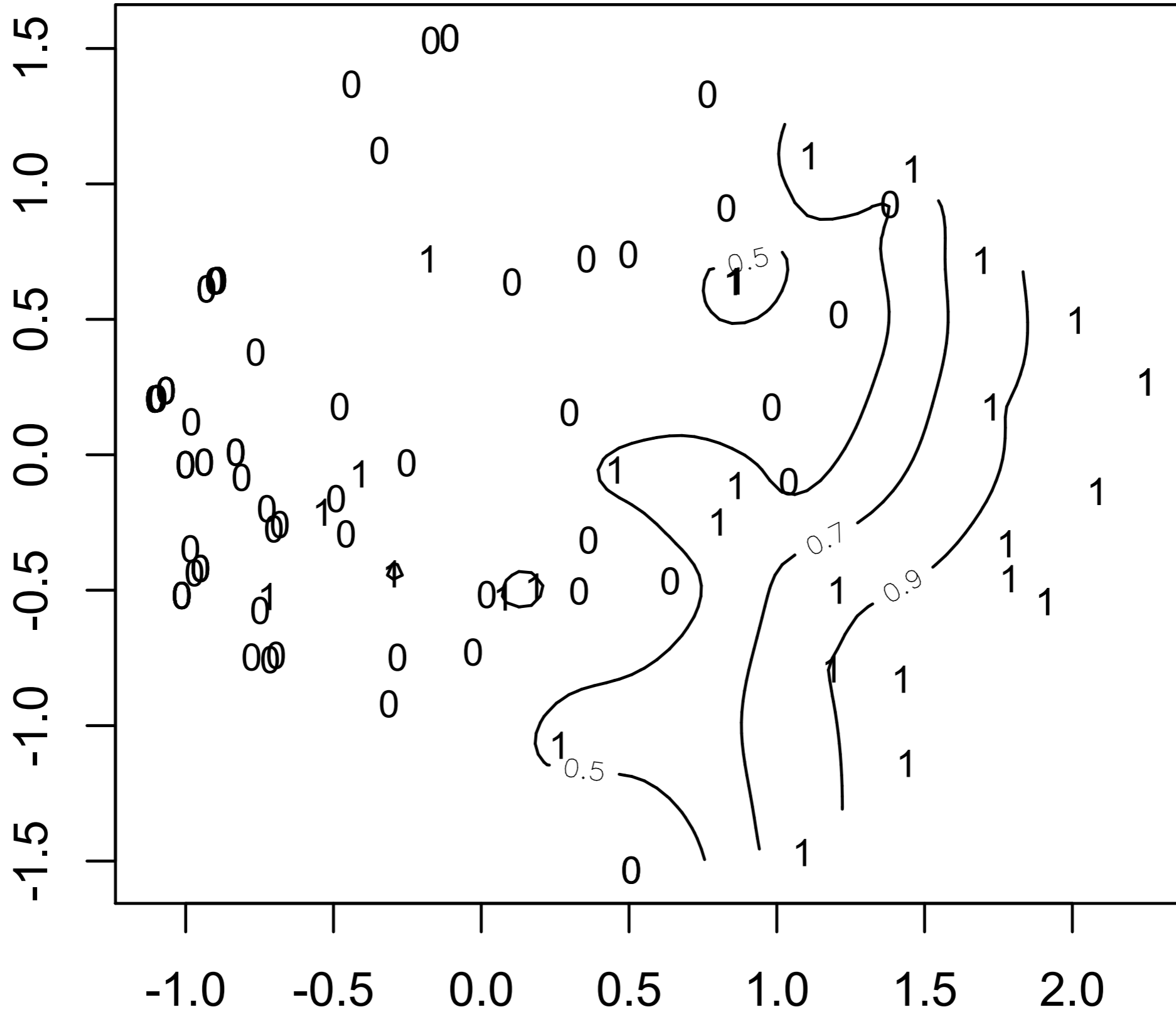
**Consensus network of 4 optimal trees
according to 'dollo' Maximum Parsimony**

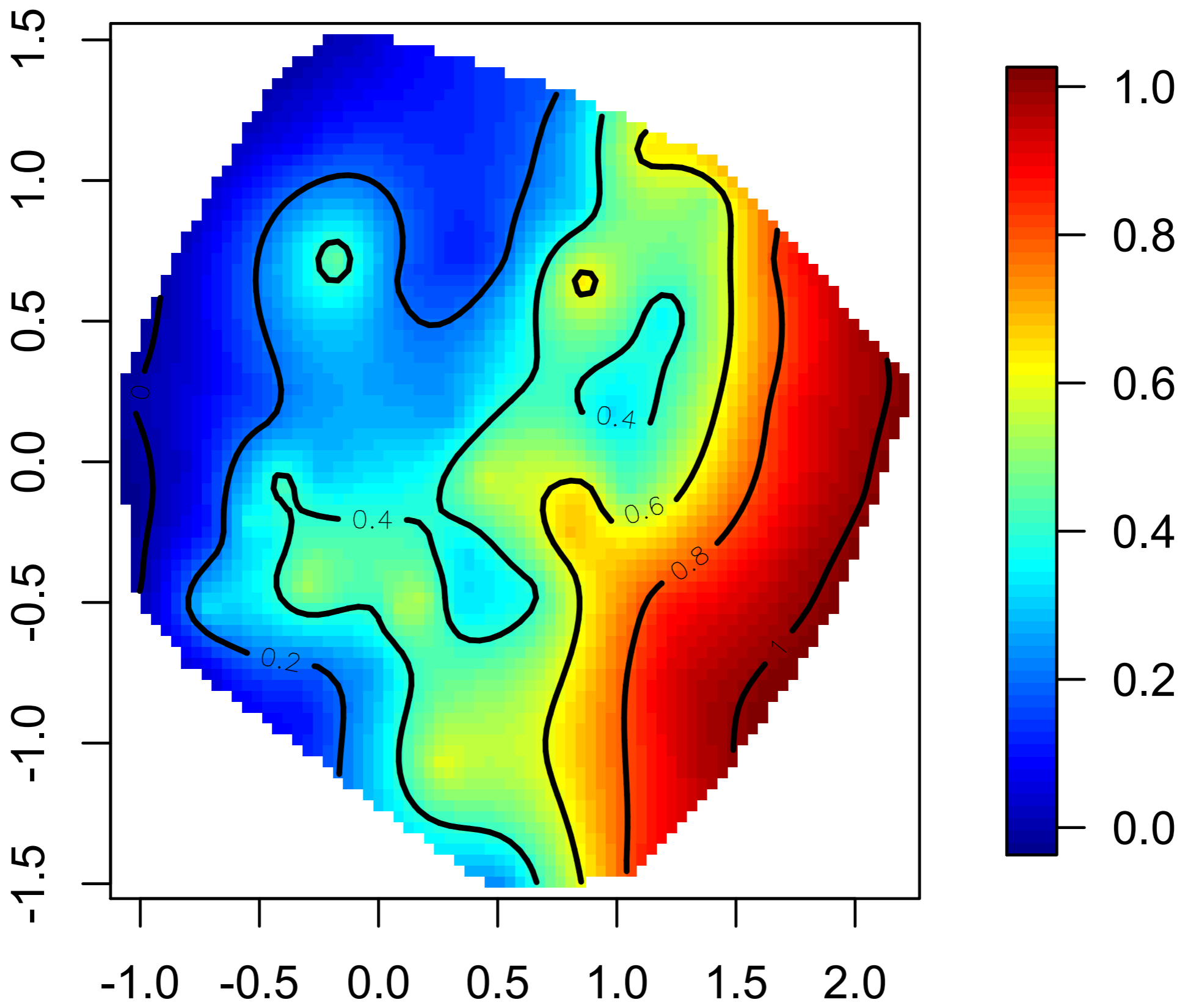


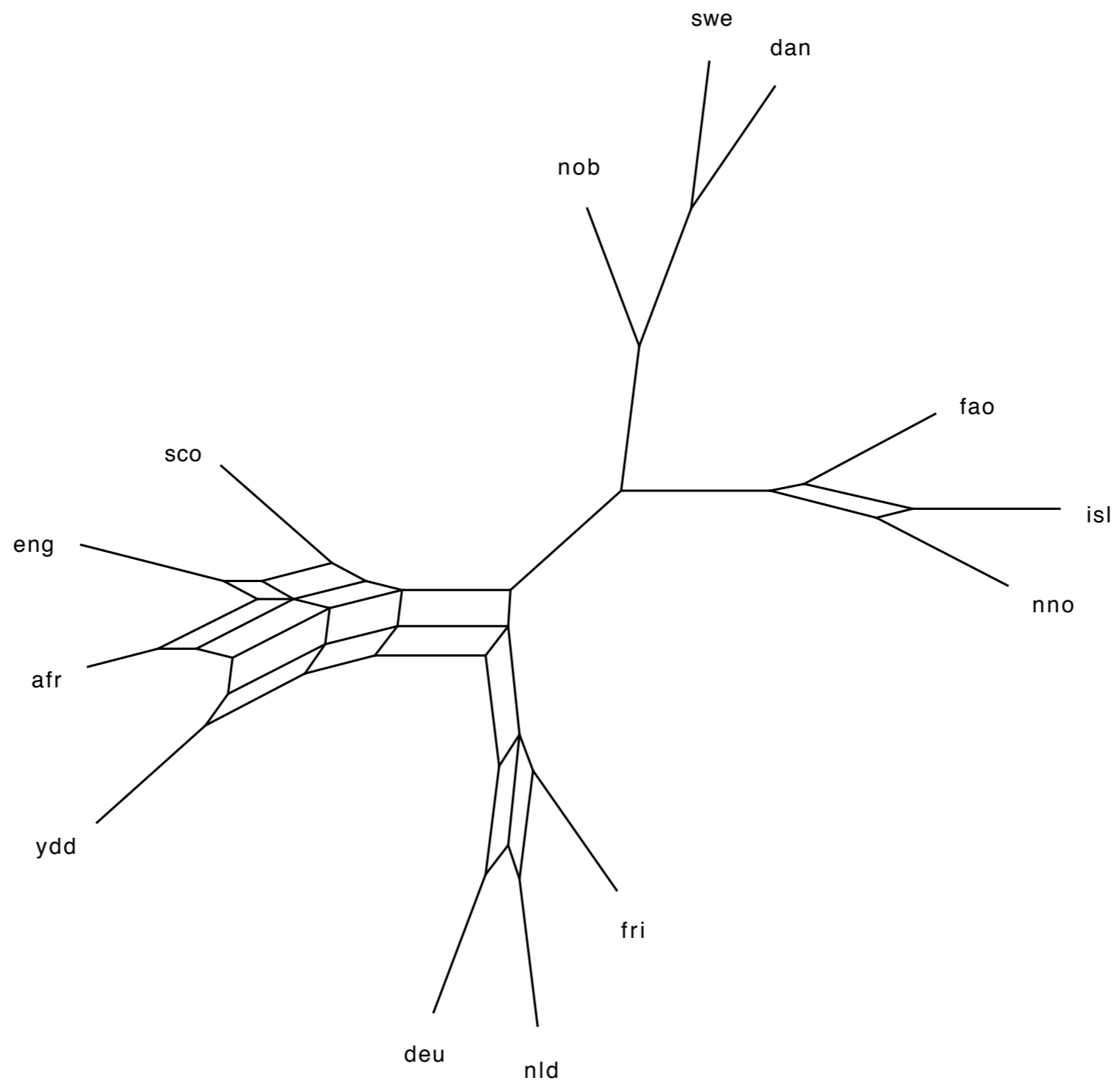


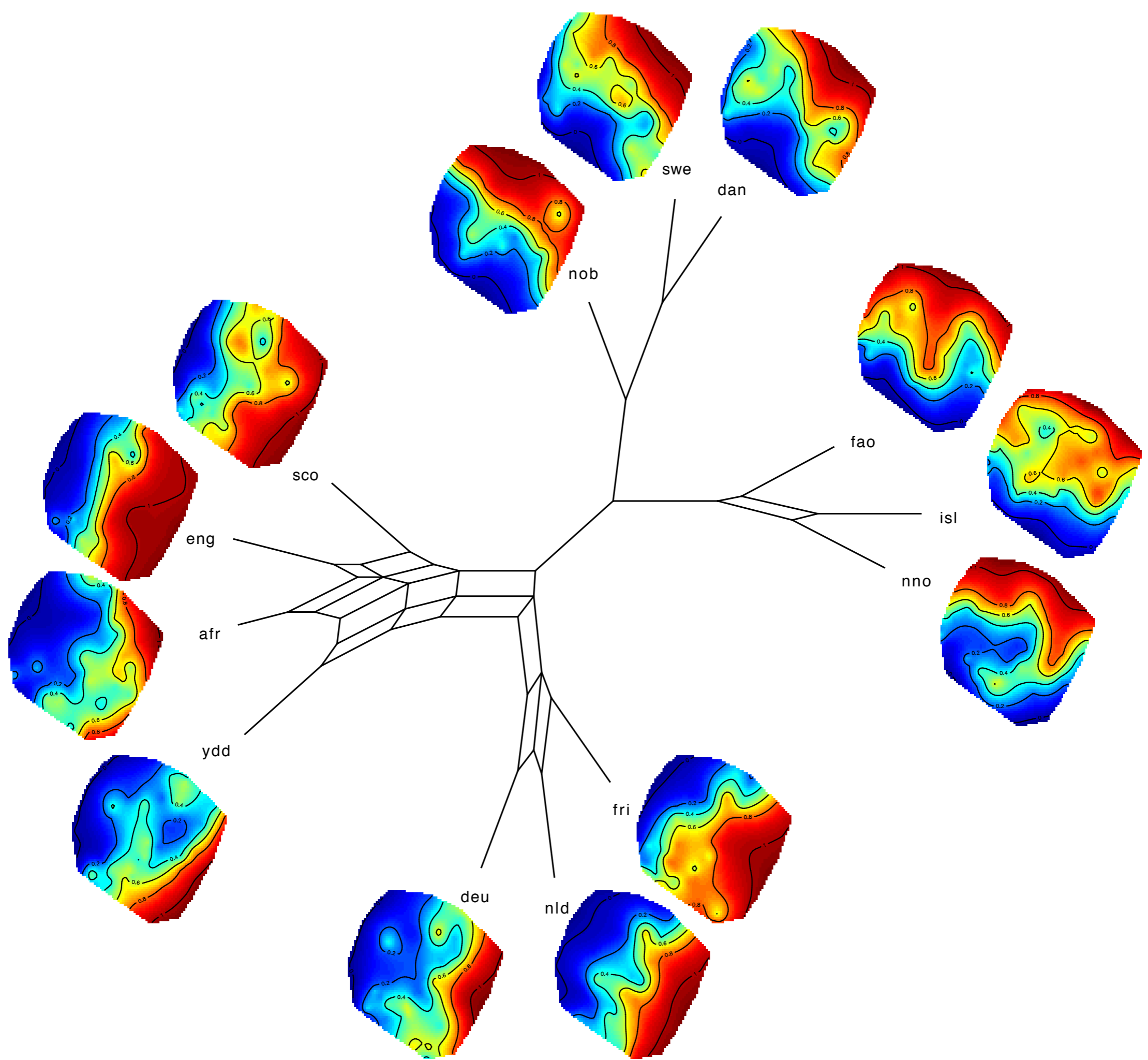












Similarity Semantics

- Define the **meaning** of a linguistic form as the set of all contexts in which it occurs
- **Similarity Semantics**: meaning of linguistic forms is defined relationally to each other (“context” is taken as purely linguistic)
- Investigate the co-occurrence of linguistic forms across languages in parallel texts