

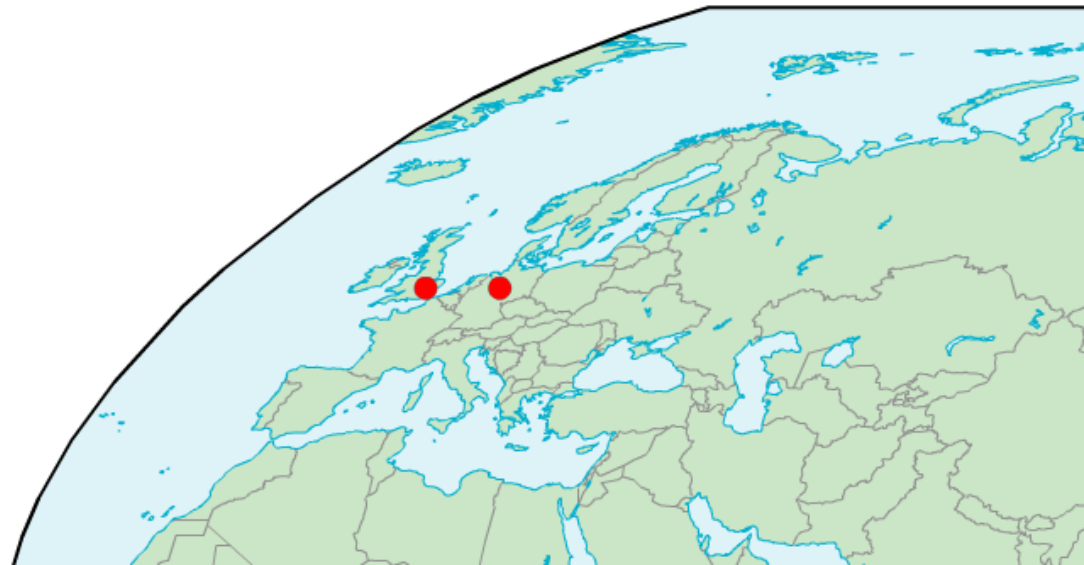


Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

EXPLORING QUANTITATIVE CONTRASTS: THE CASE OF ENGLISH AND GERMAN RELATIVE CLAUSES

DANIEL WIECHMANN



Some disciplines in language comparison:

Scope & Granularity

COARSE-GRAINED

CONTINUUM OF
GRANULARITY OF CONTRAST

Language Typology

- take large sample of languages
- study few investigated properties

(Categorical) Contrastive Linguistics

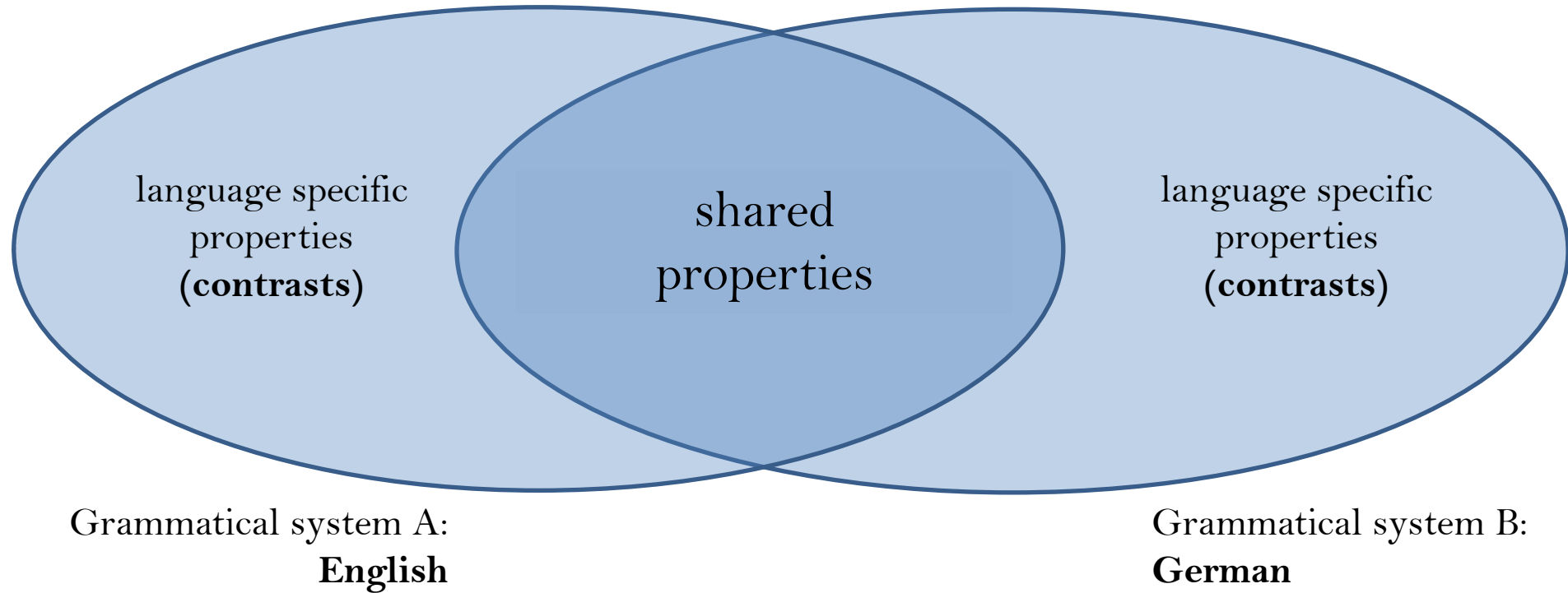
- take few (typically pair of) languages
- study larger number of investigated properties

Quantitative Contrastive Linguistics

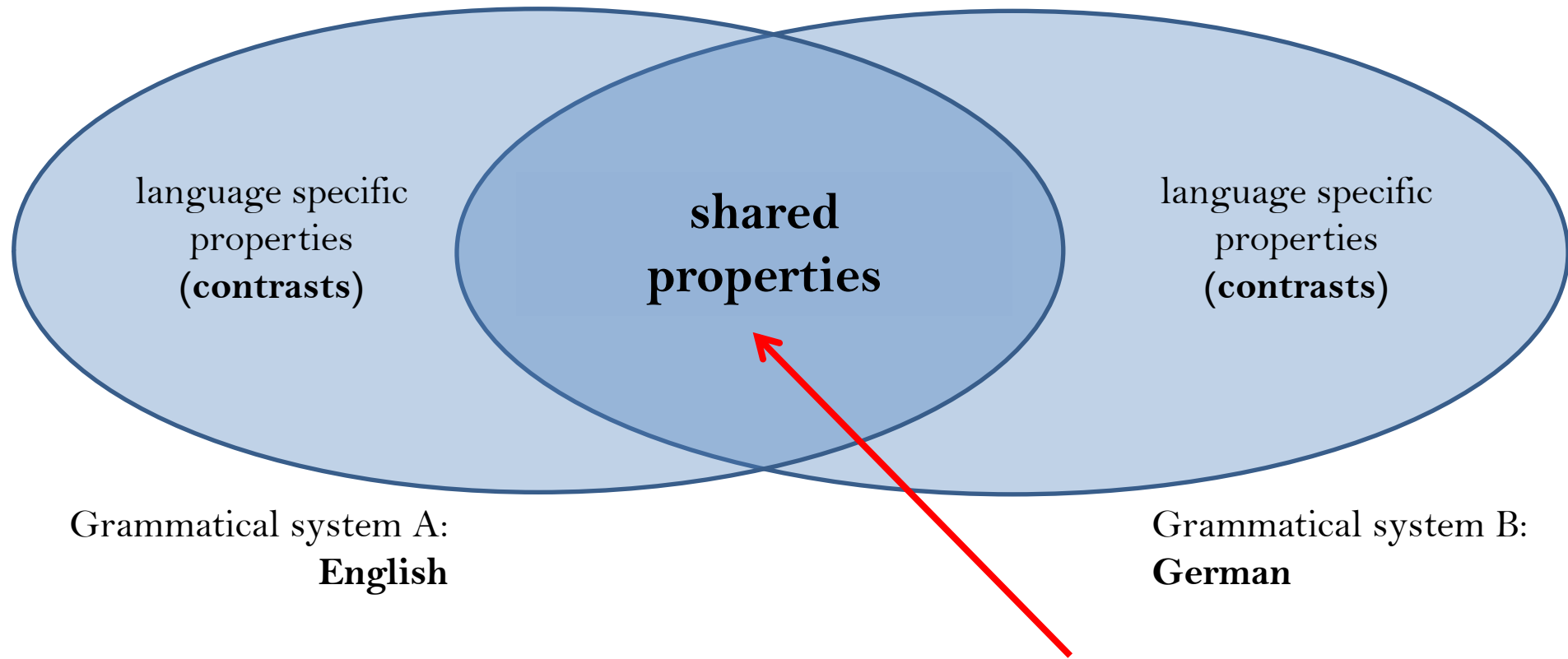
- take few (typically pair of) languages
- study quantitative differences in representative samples

FINE-GRAINED

(Categorical) Contrastive analysis yields:



(Categorical) Contrastive analysis yields:



Grammatical system A:
English

Grammatical system B:
German

domain of
Quantitative Contrastive Linguistics

Goal of this talk

The **goal** of the talk is to sketch a **corpus-based explorative methodology** for the **quantitative contrastive analysis** of (genetically related) languages

Goal of this talk

The **goal** of the talk is to sketch a **corpus-based explorative methodology** for the **quantitative contrastive analysis** of (genetically related) languages

QUESTION 1:

For any given phenomenon of interest, there are usually many potentially relevant contrasts.

► How can we tell which ones are most interesting?

TASK1:

Identify VARIABLES that strongly distinguish language A from language B

Goal of this talk

The **goal** of the talk is to sketch a **corpus-based explorative methodology** for the **quantitative contrastive analysis** of (genetically related) languages

QUESTION 1:

For any given phenomenon of interest, there are usually many potentially relevant contrasts.

► How can we tell which ones are most interesting?

TASK 1:

Identify VARIABLES that strongly distinguish language A from language B

QUESTION 2:

There is a correlation between the frequency of a structure and the way it is cognitively represented (highly frequent patterns are stored and retrieved holistically)

► Are there differences in usage frequency of a given structure across languages?

TASK 2:

Using the variables identified in Task 1,

Investigate INDIVIDUAL PATTERNS and compare their usage frequencies across languages.

An example:

Relative Clause Constructions (RCC)
in English and German

Procedure

I. **Sample from comparable linguistic resources (corpora)**

- GERMAN: 250 tensed RC-constructions IDS corpus compilation via COSMAS II_{web}
- ENGLISH: 250 tensed RC-constructions from written part of ICE-GB R2
 - approx. genre/text-structure of IDS sample

II. Describe corpus data: Code for a wide range of properties

III. Task 1: Identify VARIABLES that strongly distinguish language A from language B

IV. Task 2: Investigate INDIVIDUAL PATTERNS and compare their usage frequencies

Procedure

- I. Sample from comparable linguistic resources (corpora)
 - GERMAN: 250 tensed RC-constructions IDS corpus compilation via COSMAS II_{web}
 - ENGLISH: 250 tensed RC-constructions from written part of ICE-GB R2
 - approx. genre/text-structure of IDS sample
- II. **Describe corpus data: Code for a wide range of properties**
- III. Task 1: Identify VARIABLES that strongly distinguish language A from language B
- IV. Task 2: Investigate INDIVIDUAL PATTERNS and compare their usage frequencies

Procedure

- I. Sample from comparable linguistic resources (corpora)
 - GERMAN: 250 tensed RC-constructions IDS corpus compilation via COSMAS II_{web}
 - ENGLISH: 250 tensed RC-constructions from written part of ICE-GB R2
 - approx. genre/text-structure of IDS sample
- II. Describe corpus data: Code for a wide range of properties
- III. Task 1: Identify VARIABLES that strongly distinguish language A from language B**
- IV. Task 2: Investigate INDIVIDUAL PATTERNS and compare their usage frequencies

Procedure

- I. Sample from comparable linguistic resources (corpora)
 - GERMAN: 250 tensed RC-constructions IDS corpus compilation via COSMAS II_{web}
 - ENGLISH: 250 tensed RC-constructions from written part of ICE-GB R2
 - approx. genre/text-structure of IDS sample
- II. Describe corpus data: Code for a wide range of properties
- III. Task 1: Identify VARIABLES that strongly distinguish language A from language B
- IV. **Task 2: Investigate INDIVIDUAL PATTERNS and compare their usage frequencies**

Describing Relative Clause Constructions (RCCs)

Examples:

- i. *Peter hates **everything** (= head) that **John** (= Subject_{RC}) likes __ .*
- ii. *Peter hasst **alles** (= head), was **Johann** (= Subject_{RC}) mag __ .*

Potentially
interesting
dimensions
of contrast

- **PROPERTIES OF THE RELATIVE CLAUSE**
 - the internal grammatical role of head
 - the external grammatical role of head
 - the type of embedding of the RC
 - the voice of the RC
- **PROPERTIES ENCODED ON THE HEAD NOUN**
 - the animacy (of the referent) of the head
 - the definiteness of the head
 - the type of NP_{head}
- **PROPERTIES ENCODED ON THE SUBJECT OF RC**
 - the animacy (of the referent) of the RC subject
 - the definiteness of the RC subject
 - the type of NP_{RCsubject}

Describing Relative Clause Constructions (RCCs)

Data matrix after description

language	extern.role	intern.role	embed	voice	ani.head	def.head	ms.head	ani.SRC	def.SRC	ms.SRC
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
english	io	do	right	act	ina	def	h.lex	ina	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex

Describing Relative Clause Constructions (RCCs)

Data matrix after description

variables



language	extern.role	intern.role	embed	voice	ani.head	def.head	ms.head	ani.SRC	def.SRC	ms.SRC
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
english	io	do	right	act	ina	def	h.lex	ina	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex

Describing Relative Clause Constructions (RCCs)

Data matrix after description

language	extern.role	intern.role	embed	voice	ani.head	def.head	ms.head	ani.SRC	def.SRC	ms.SRC
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
english	io	do	right	act	ina	def	h.lex	ina	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex

Describing Relative Clause Constructions (RCCs)


Data matrix after description

	variables →									
language	extern.role	intern.role	embed	voice	ani.head	def.head	ms.head	ani.SRC	def.SRC	ms.SRC
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	description	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
english	io	do	right	act	ina	def	h.lex	ina	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex

↓ cases in data set

Task 1:

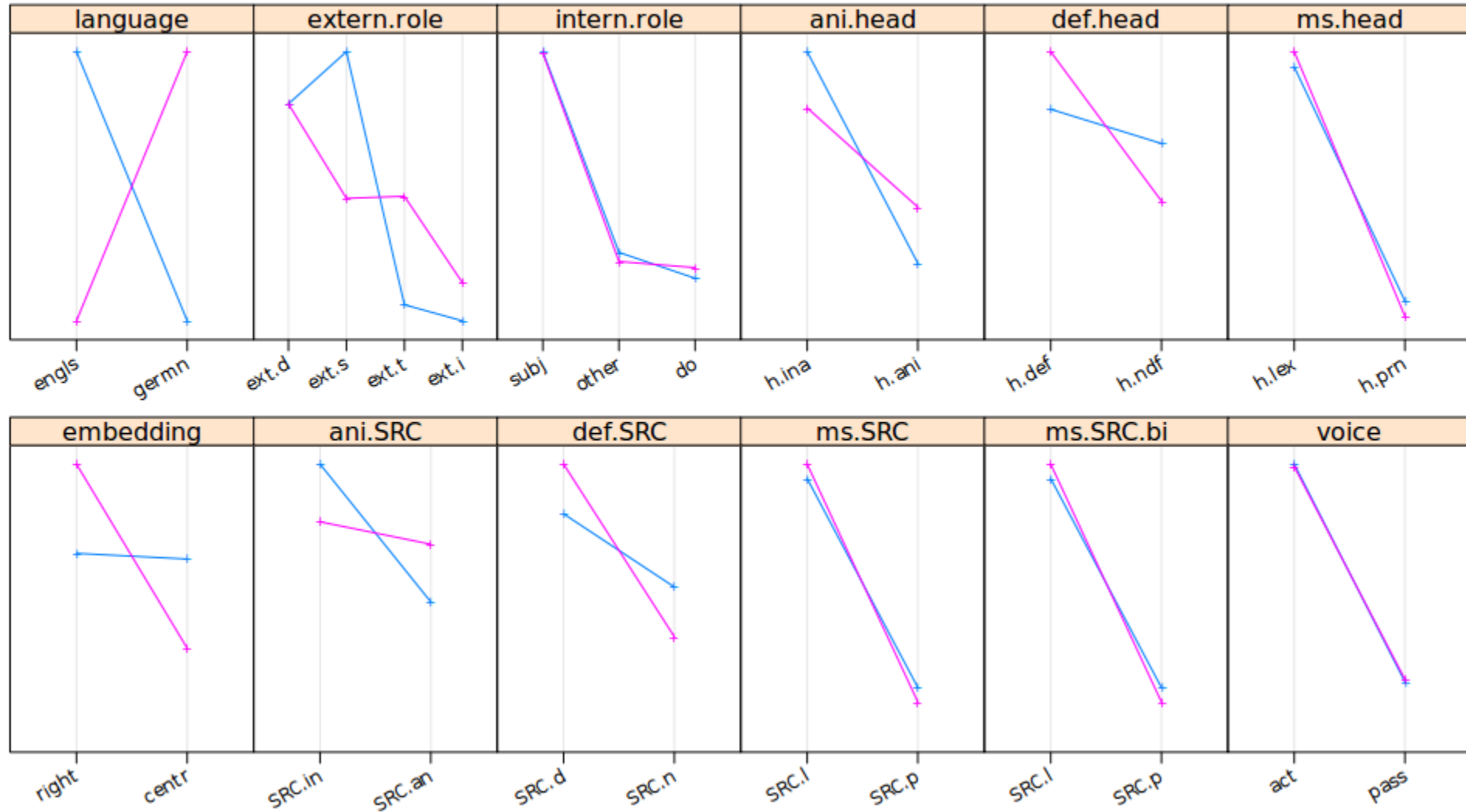
Identify **VARIABLES** that strongly distinguish language A from language B



extern.role	intern.role	embed	voice	ani.head	def.head	ms.head	ani.SRC	def.SRC	ms.SRC
do	do	right	act	ina	def	lex	ani	indef	lex
do	do	right	act	ina	def	lex	ani	indef	lex
do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
io	do	right	act	ina	def	h.lex	ina	def	lex
other	do	right	act	ina	indef	h.lex	ani	def	lex
other	do	right	act	ina	indef	h.lex	ani	def	lex

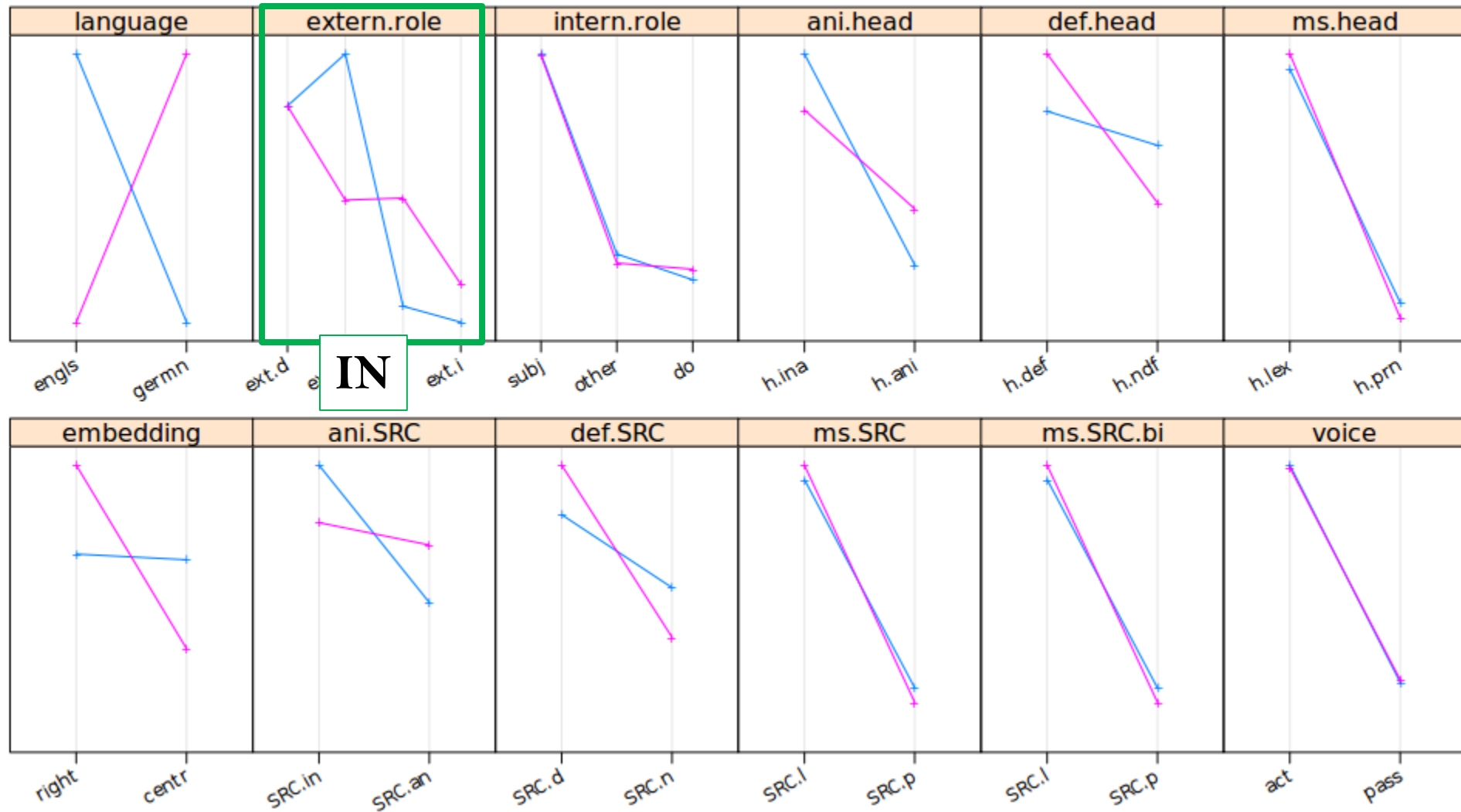
Identify **VARIABLES** that strongly distinguish language A from language B

language
 english ○ — german ○ —



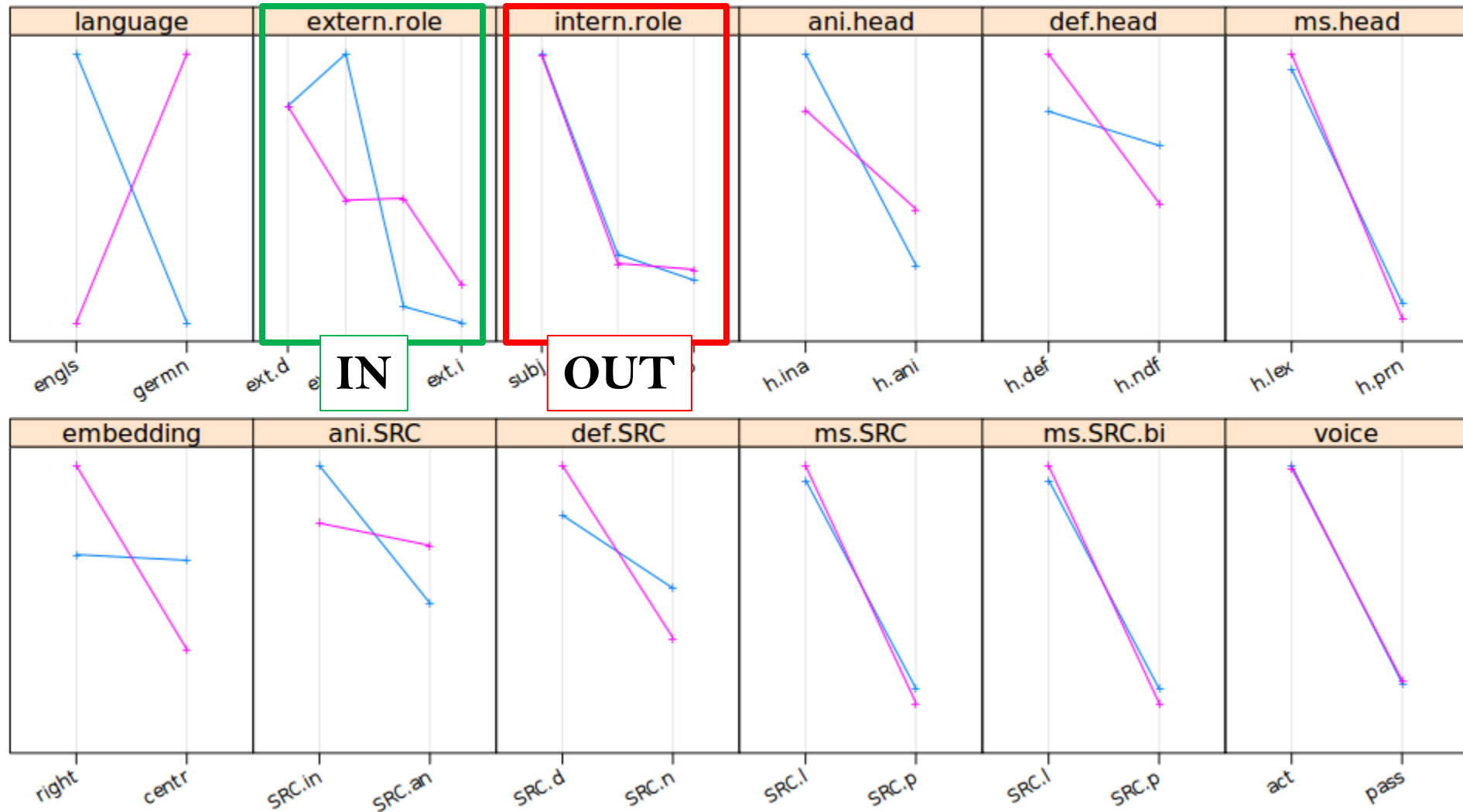
Identify **VARIABLES** that strongly distinguish language A from language B

language
 english ○ — blue line
 german ○ — magenta line



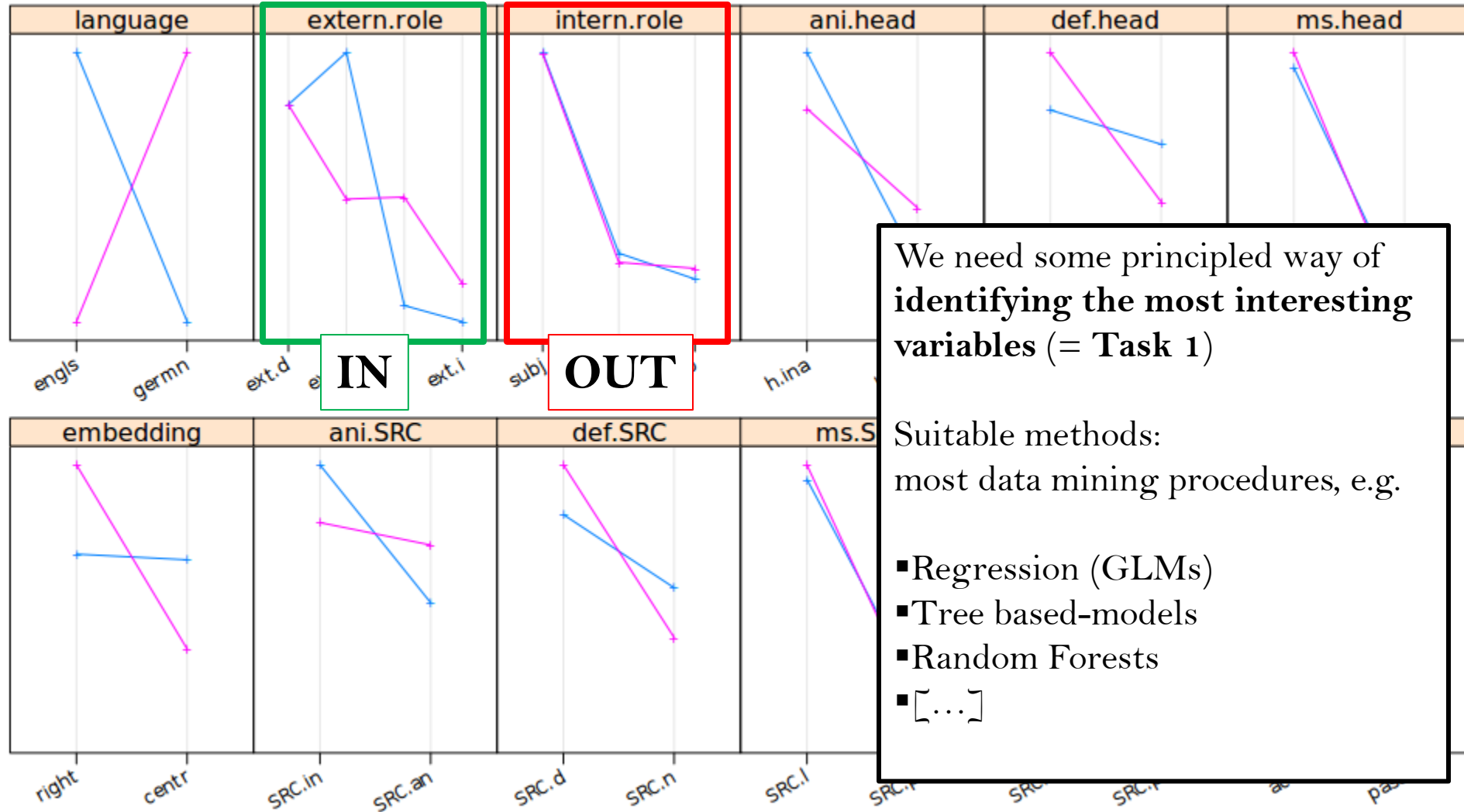
Identify **VARIABLES** that strongly distinguish language A from language B

language
 english ○ — blue — german ○ — magenta



Identify **VARIABLES** that strongly distinguish language A from language B

language
 english ○ — blue — german ○ — pink —



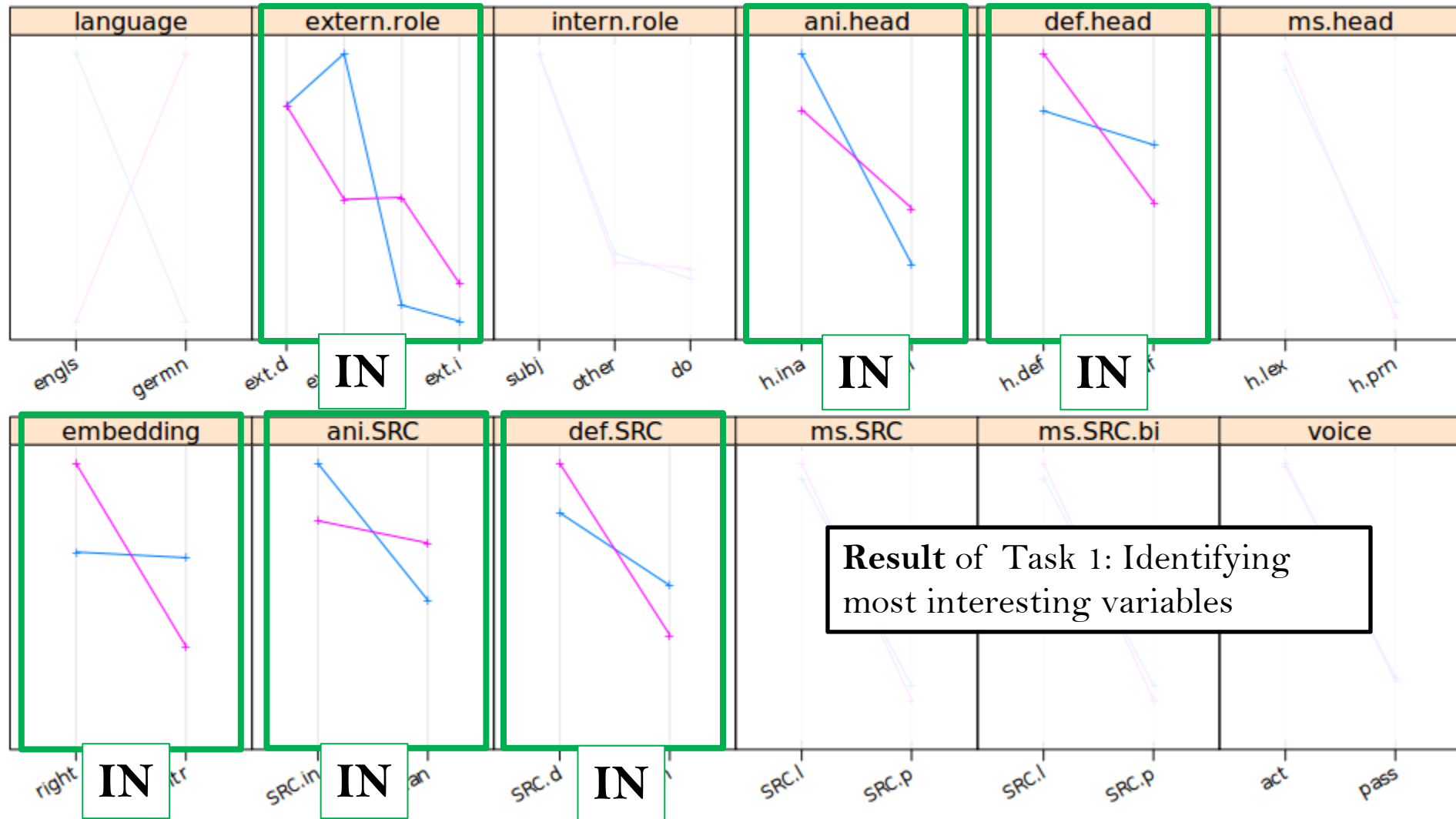
We need some principled way of **identifying the most interesting variables** (= Task 1)

Suitable methods:
 most data mining procedures, e.g.

- Regression (GLMs)
- Tree based-models
- Random Forests
- [...]

Identify **VARIABLES** that strongly distinguish language A from language B

language
 english ○ — blue — german ○ — pink —



Result of Task 1: Identifying most interesting variables

So, we started with 10 descriptor variables...

language	extern.role	intern.role	embed	voice	ani.head	def.head	type.head	ani.SRC	def.SRC	ms.SRC
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
english	io	do	right	act	ina	def	h.lex	ina	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex

So, we started with 10 descriptor variables...

... and **after Task 1** we are left with
only **6 descriptors**

language	extern.role	intern.role	embed	voice	ani.head	def.head	type.head	ani.SRC	def.SRC	ms.SRC
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
german	do	do	right	act	ina	def	lex	ani	indef	lex
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
english	io	do	right	act	ina	def	h.lex	ina	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex
english	other	do	right	act	ina	indef	h.lex	ani	def	lex

Task 2

RCC - PATTERNS and their usage frequencies in English & German

extern.role	intern.role	embed	voice	ani.head	def.head	type.head	ani.SRC	def.SRC
do	do	right	act	ina	def	lex	ani	indef
do	do	right	act	ina	def	lex	ani	indef
do	do	right	act	ina	def	lex	ani	indef
[...]	[...]	[...]	[...]	[...]	description	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
io	do	right	act	ina	def	h.lex	ina	def
other	do	right	act	ina	indef	h.lex	ani	def
other	do	right	act	ina	indef	h.lex	ani	def

Method: Configurational Frequency Analysis (CFA; von Eye, 1990)

- I. Cross all 6 descriptor variables (→ complex contingency table)
- II. List all possible configurations (= factor level combinations)
- III. For each configuration, **compare observed frequency with expected frequency**

How frequent are
combinatorial possibilities
(configurations)

- $\text{Freq}_{\text{OBS}} > \text{Freq}_{\text{EXP}}$
- $\text{Freq}_{\text{OBS}} < \text{Freq}_{\text{EXP}}$
- $\text{Freq}_{\text{OBS}} \sim \text{Freq}_{\text{EXP}}$

Properties of ...					
RC		Head		RC-Subject	
external	role embedding	animate	definite	animate	definite
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>
<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>
<i>subj</i>	<i>center</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>
<i>subj</i>	<i>center</i>	<i>ani</i>	<i>indef</i>	<i>ani</i>	<i>indef</i>
<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>def</i>
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>
<i>subj</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>
<i>do</i>	<i>right</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>indef</i>
<i>other</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>
[...]	[...]	[...]	[...]	[...]	[...]

Results CFA_{English} : Top 15 (by frequency)

Properties of ...											
RC		Head		RC-Subject		Statistics					
external role	embedding	animate	definite	animate	definite	Freq	Exp	Obs-exp	Dec	Q	
1	<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	52	5.54	>	***	0.19
2	<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	48	11.92	>	***	0.152
3	<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	23	9.92	>	*	0.054
⋮	<i>subj</i>	<i>center</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	20	1.24	>	***	0.075
⋮	<i>subj</i>	<i>center</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	14	6.66	>	ns	0.03
⋮	<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>	13	5.82	>	ns	0.029
⋮	<i>subj</i>	<i>center</i>	<i>ani</i>	<i>indef</i>	<i>ani</i>	<i>indef</i>	13	0.69	>	***	0.049
⋮	<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>	8	4.84	>	ns	0.013
⋮	<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>def</i>	8	8.31	<	ns	0.001
⋮	<i>subj</i>	<i>center</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>	6	4.88	>	ns	0.005
⋮	<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>	5	4.06	>	ns	0.004
⋮	<i>subj</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	5	6.88	<	ns	0.008
13	<i>do</i>	<i>right</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	4	1.03	>	ns	0.012
14	<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>indef</i>	4	7.94	<	ns	0.016
15	<i>other</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	3	0.73	>	ns	0.009

Compare most frequent English patterns with their status in German

COMPARE

CFA-results for ENGLISH

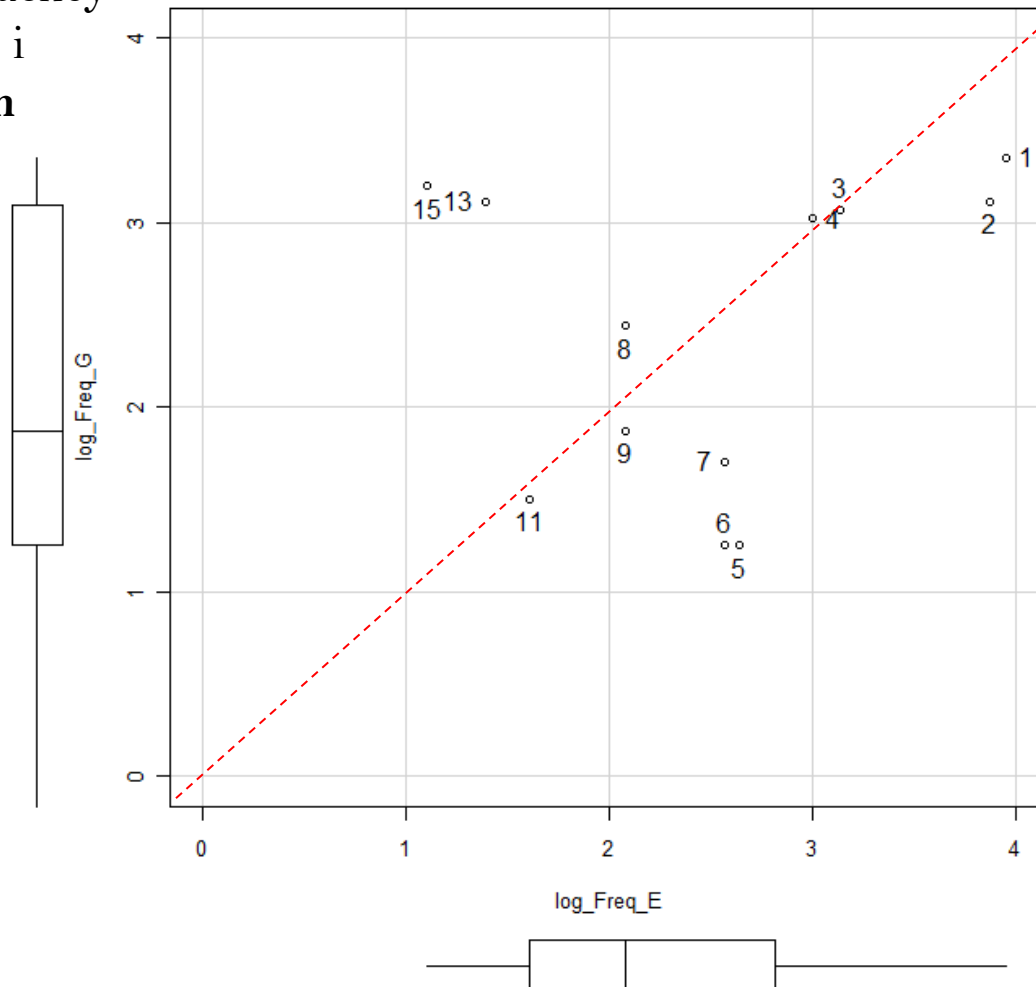
CFA-results for GERMAN

Properties of ...										
RC		Head		RC-Subject		Statistics				
external	role embedding	animate	definite	animate	definite	Freq	Exp	Obs-exp	Dec	Q
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	52	5.54	>	***	0.19
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	48	11.92	>	***	0.152
<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	23	9.92	>	*	0.054
<i>subj</i>	<i>center</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	20	1.24	>	***	0.075
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	14	6.66	>	ns	0.03
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>	13	5.82	>	ns	0.029
<i>subj</i>	<i>center</i>	<i>ani</i>	<i>indef</i>	<i>ani</i>	<i>indef</i>	13	0.69	>	***	0.049
<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>	8	4.84	>	ns	0.013
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>def</i>	8	8.31	<	ns	0.001
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>	6	4.88	>	ns	0.005
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>	5	4.06	>	ns	0.004
<i>subj</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	5	6.88	<	ns	0.008
<i>do</i>	<i>right</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	4	1.03	>	ns	0.012
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>indef</i>	4	7.94	<	ns	0.016
<i>other</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	3	0.73	>	ns	0.009

Properties of ...										
RC		Head		RC-Subject		Statistics				
external	role embedding	animate	definite	animate	definite	Freq	Exp	Obs-exp	Dec	Q
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	28	2.30	>	***	0.104
<i>other</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	24	8.26	>	***	0.065
<i>do</i>	<i>right</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	22	6.87	>	***	0.062
<i>subj</i>	<i>center</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	22	2.57	>	***	0.079
<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	21	14.39	>	ns	0.028
<i>subj</i>	<i>center</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	20	1.22	>	***	0.075
<i>other</i>	<i>right</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	15	3.94	>	**	0.045
<i>do</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ani</i>	<i>def</i>	11	12.87	<	ns	0.008
<i>io</i>	<i>right</i>	<i>ina</i>	<i>def</i>	<i>ina</i>	<i>def</i>	9	2.53	>	ns	0.026
<i>do</i>	<i>right</i>	<i>ani</i>	<i>indef</i>	<i>ani</i>	<i>indef</i>	8	1.10	>	**	0.028
<i>other</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ani</i>	<i>def</i>	6	3.29	>	ns	0.011
<i>do</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>def</i>	6	6.41	<	ns	0.002
<i>other</i>	<i>right</i>	<i>ina</i>	<i>indef</i>	<i>ina</i>	<i>indef</i>	6	1.32	>	ns	0.019
<i>subj</i>	<i>right</i>	<i>ani</i>	<i>def</i>	<i>ani</i>	<i>def</i>	5	3.88	>	ns	0.005
<i>subj</i>	<i>center</i>	<i>ani</i>	<i>indef</i>	<i>ani</i>	<i>indef</i>	5	0.20	>	***	0.019

Compare most frequent English patterns with their status in German

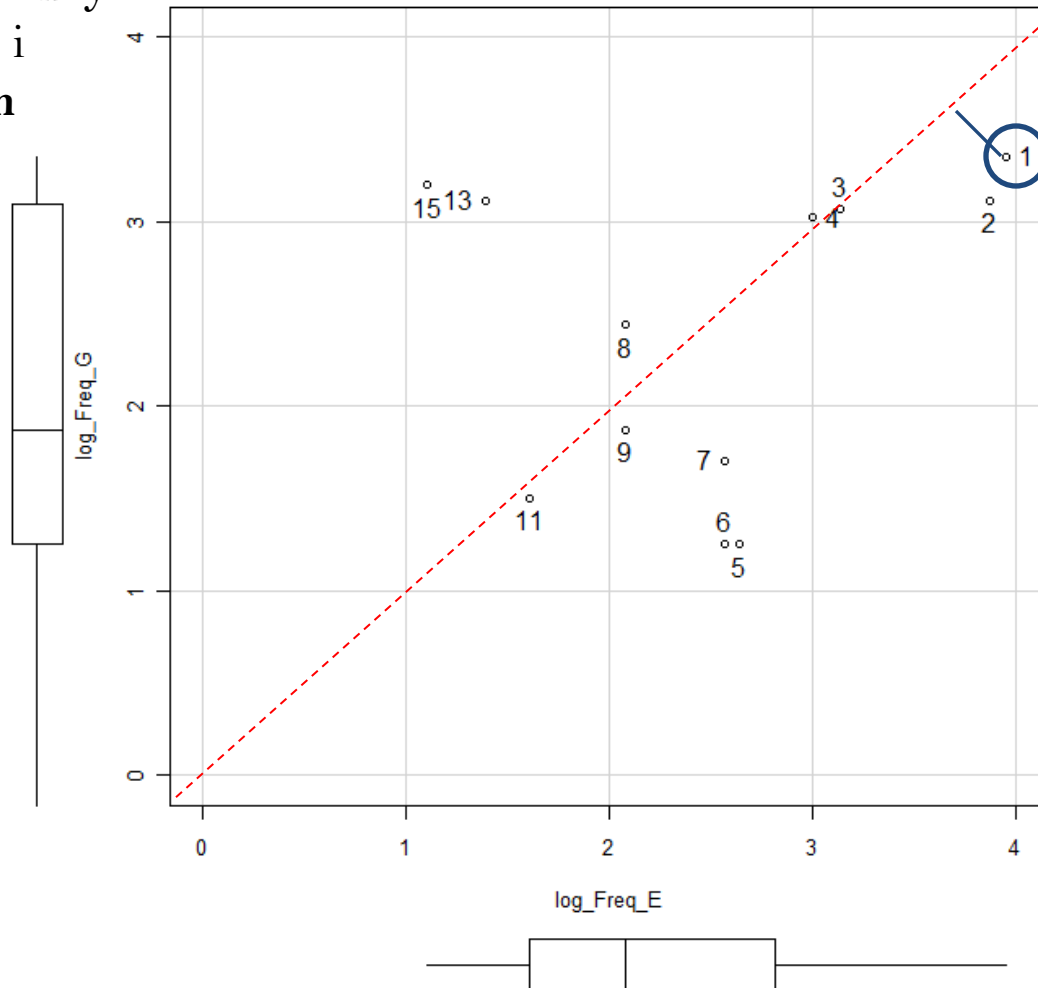
(log) Frequency
of pattern i
in **German**



(log) Frequency
of pattern i
in **English**

Compare most frequent English patterns with their status in German

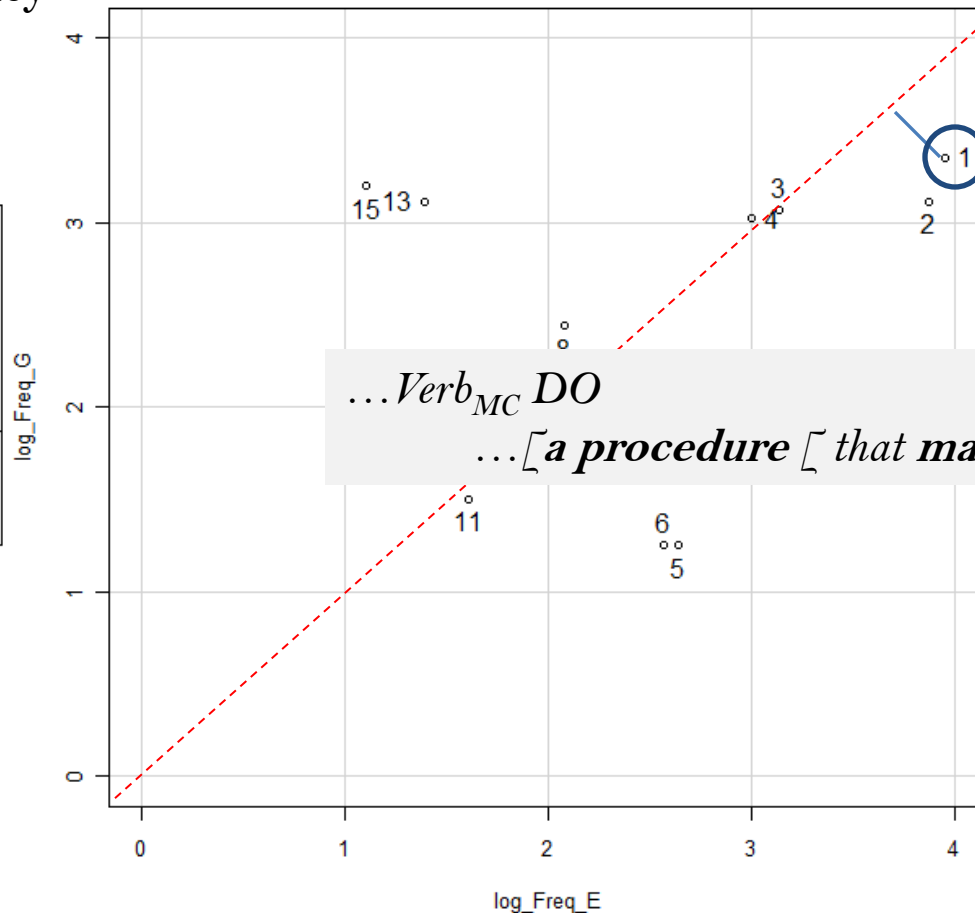
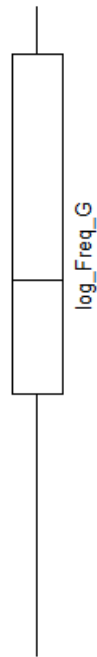
(log) Frequency of pattern i in **German**



(log) Frequency of pattern i in **English**

Compare most frequent English patterns with their status in German

(log) Frequency of pattern *i* in **German**



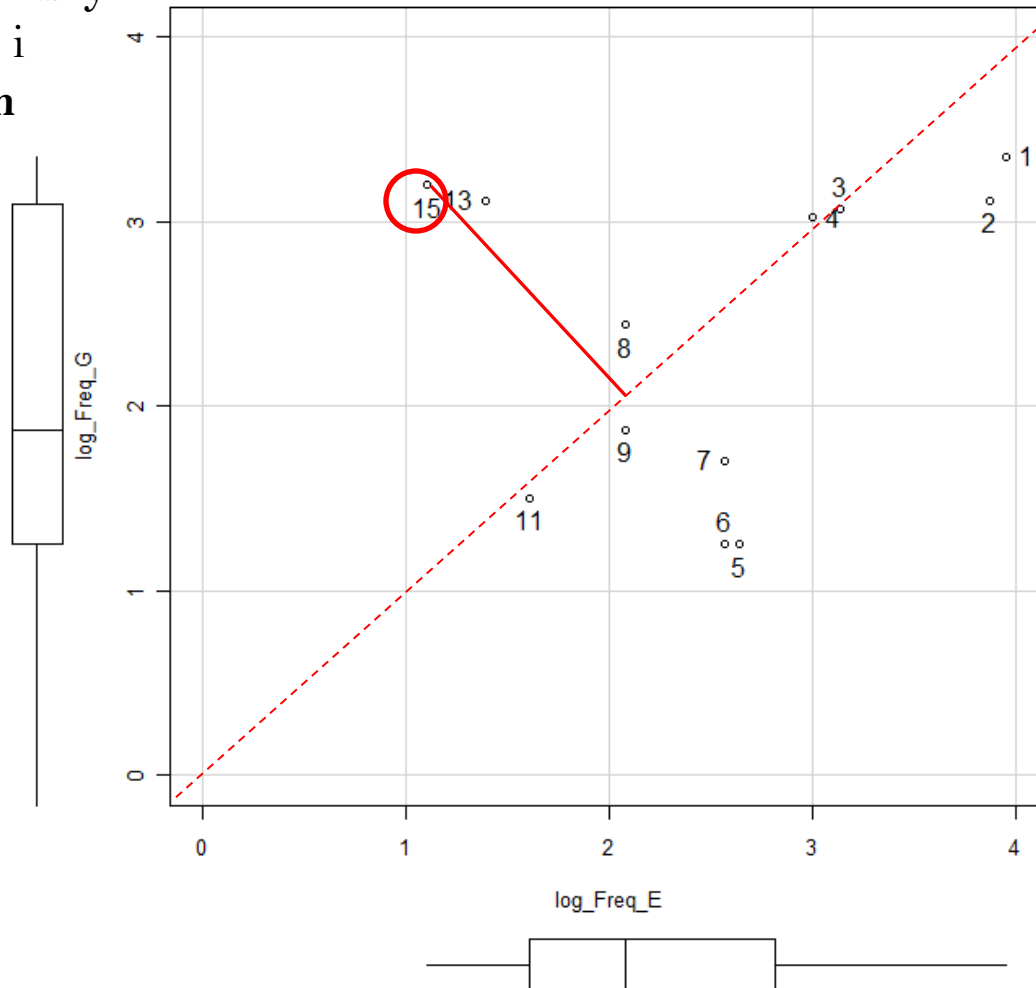
1	EXTERN.ROLE	<i>direct.obj</i>
	EMBEDDING	<i>right</i>
	HEAD.ANI	<i>ina</i>
	HEAD.DEF	<i>def</i>
	RC_SUBJ.ANI	<i>ina</i>
	RC_SUBJ.DEF	<i>indef</i>

(log) Frequency of pattern *i* in **English**



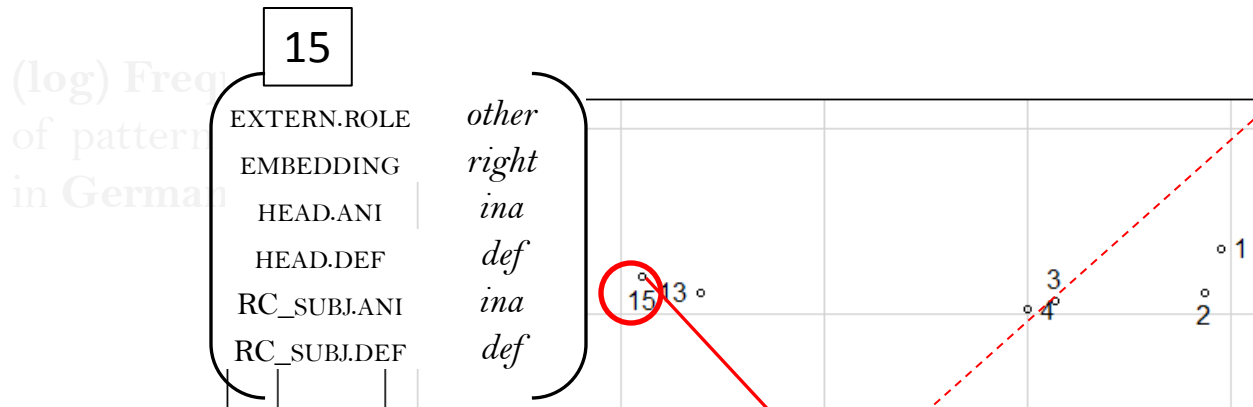
Compare most frequent English patterns with their status in German

(log) Frequency of pattern i in **German**



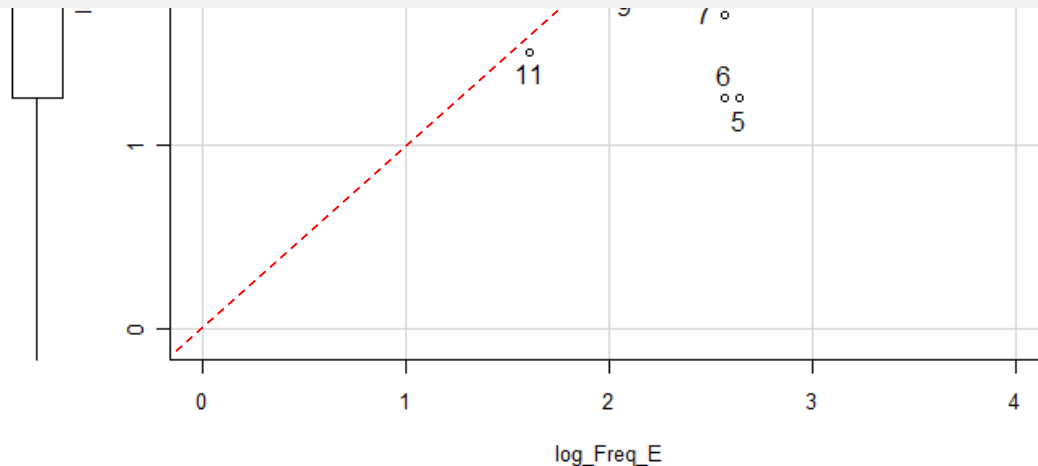
(log) Frequency of pattern i in **English**

Compare most frequent English patterns with their status in German



...*Verb*_{MC} [...] *ADVERBIAL*

... [für [das **Projekt** [, das nicht weiter finanziert werden konnte]].

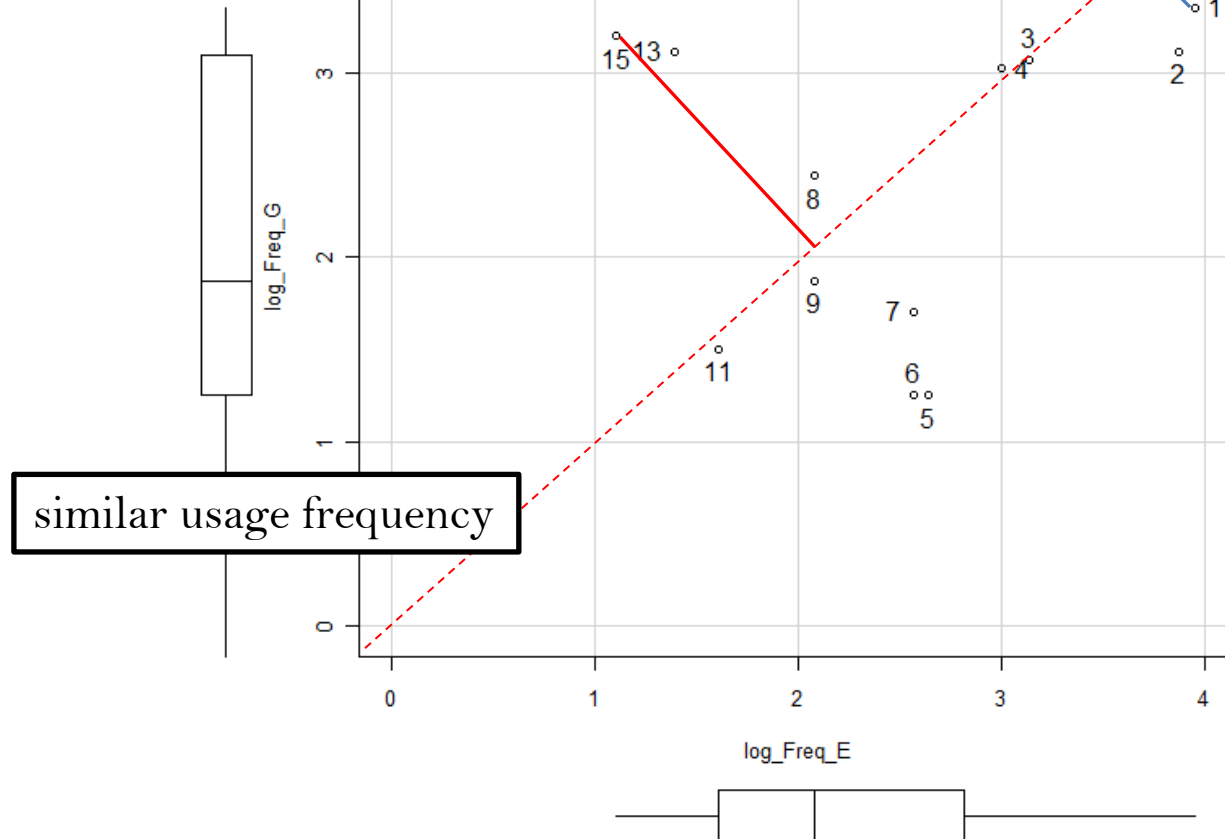


(log) Frequency of pattern i in English



Compare most frequent English patterns with their status in German

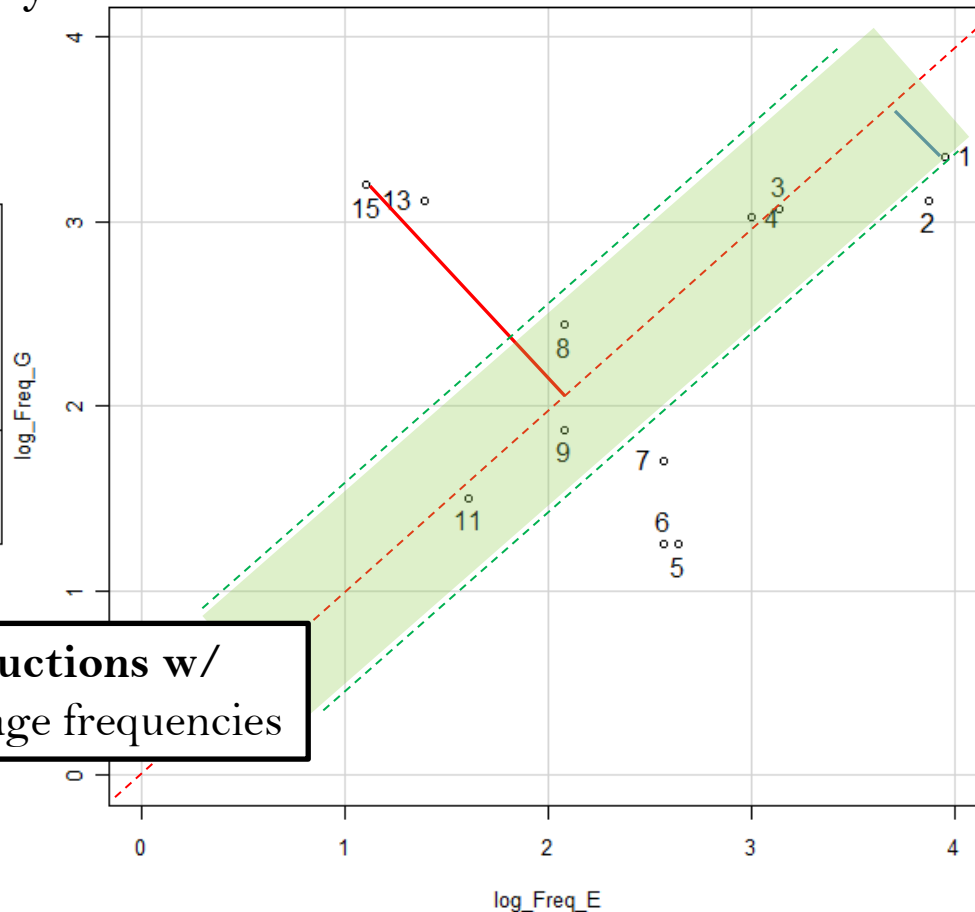
(log) Frequency of pattern i in German



(log) Frequency of pattern i in English

Compare most frequent English patterns with their status in German

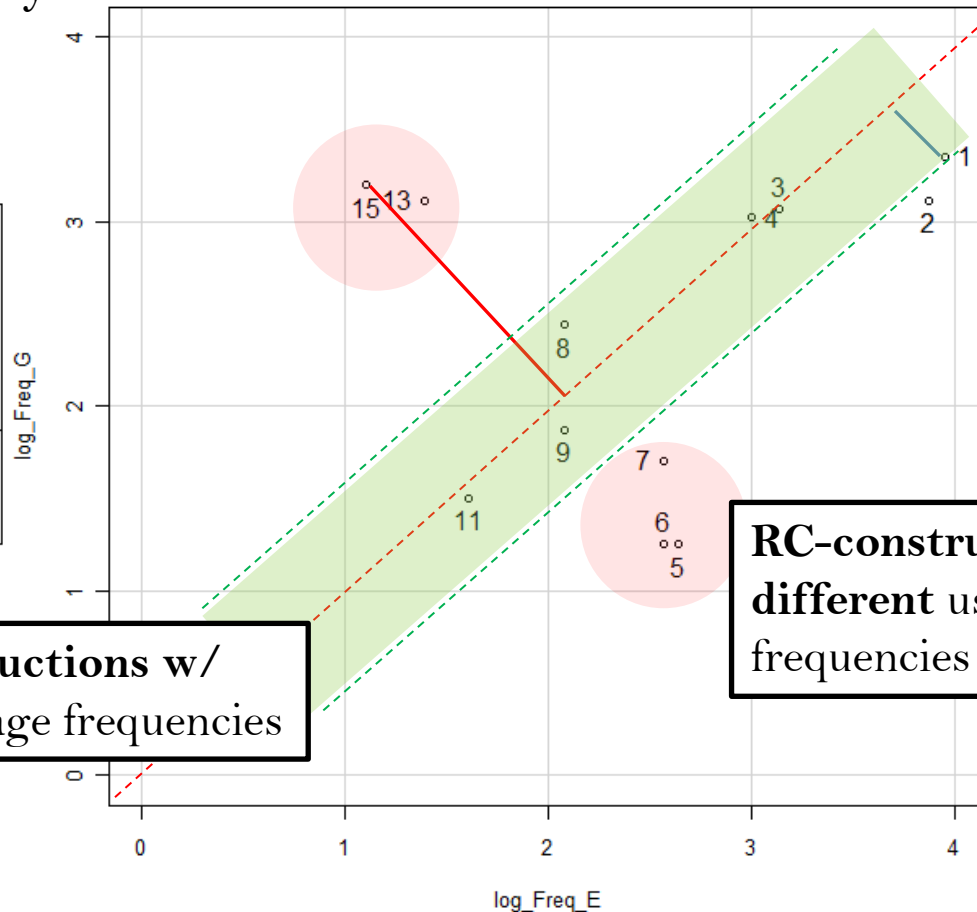
(log) Frequency
of pattern i
in German



(log) Frequency
of pattern i
in English

Compare most frequent English patterns with their status in German

(log) Frequency of pattern i in German



RC-constructions w/ similar usage frequencies

RC-constructions w/ different usage frequencies

(log) Frequency of pattern i in English

Implications for Applied Contexts: **Second Language Acquisition**

Frequency sensitivity pervades all aspects of language processing.
This has profound implications for L2-acquisition
(cf. Ellis 2002 for an overview)

➤ **Proposed methodology opens up new avenues of research into the role of native language/transfer:**

Contrastive Hypothesis (Lado, 1957):

- Learner problems are predictable from similarity between L1/L2
- (similar → easy) & (dissimilar → hard) # empirically inadequate
- salient contrasts are easy; more subtle contrasts are hard

- **frequency acts on salience**
[so, manipulate frequency → manipulate salience]

Conclusion

What?

The goal of the talk was to sketch a **corpus-based explorative methodology** for the **quantitative contrastive analysis** of (genetically related) languages

Step 1: **Identify variables** that strongly **discriminate** between the **languages**

Step 2: **Focus** on individual **patterns** and **compare their usage frequencies**

Why?

The method follows recent trends in linguistic theorizing:

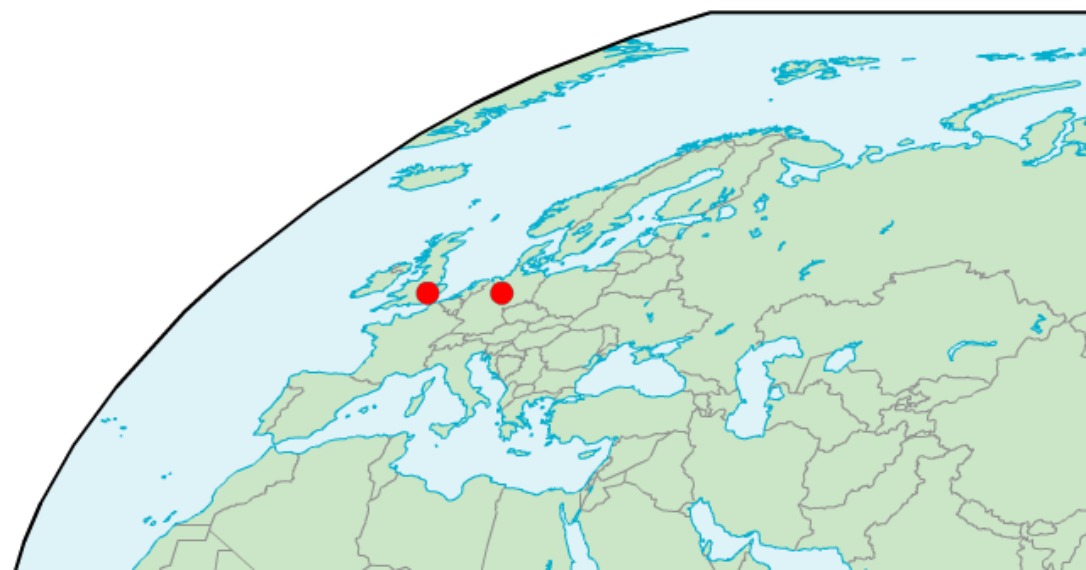
Highly compatible with usage-based constructionist conceptions of language.

The proposed methodology opens up new avenues of research into *Second Language Acquisition* (e.g. role of transfer/saliency):

References

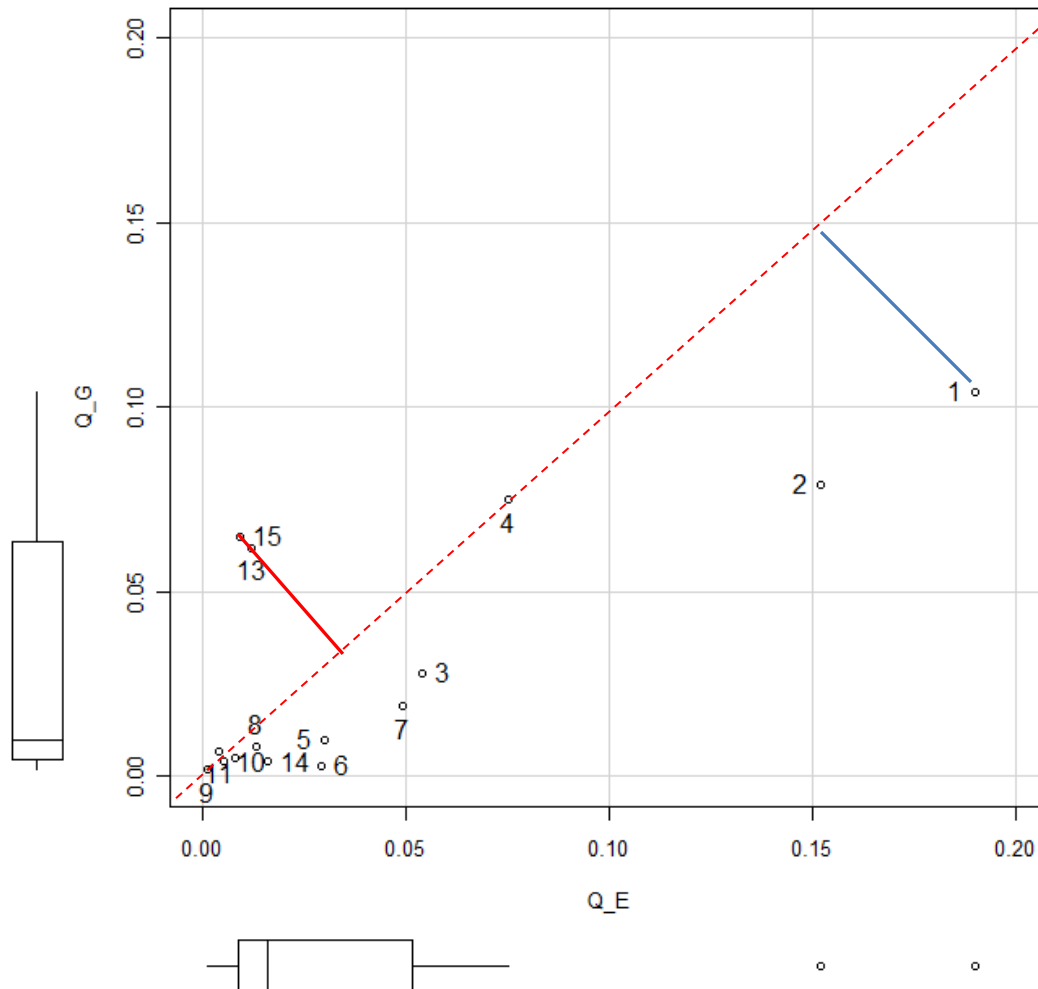
- Ellis, N. 2002. Frequency Effects in Language Processing: A review with implications for theories of
- von Eye, A. 1990. *Introduction to Configural Frequency Analysis: The search for types and antitypes in cross-classifications*. Cambridge, UK: Cambridge University Press.
- Goldberg, A. 2006. *Constructions at Work: the nature of generalization in Language*. Oxford: OUP.
- Lado, R. 1957. *Linguistics across cultures, applied linguistics language teachers*. Ann Arbor, MI: University of Michigan Press.

Thank you very much



Yet to be explored:
Different expressions of frequency information

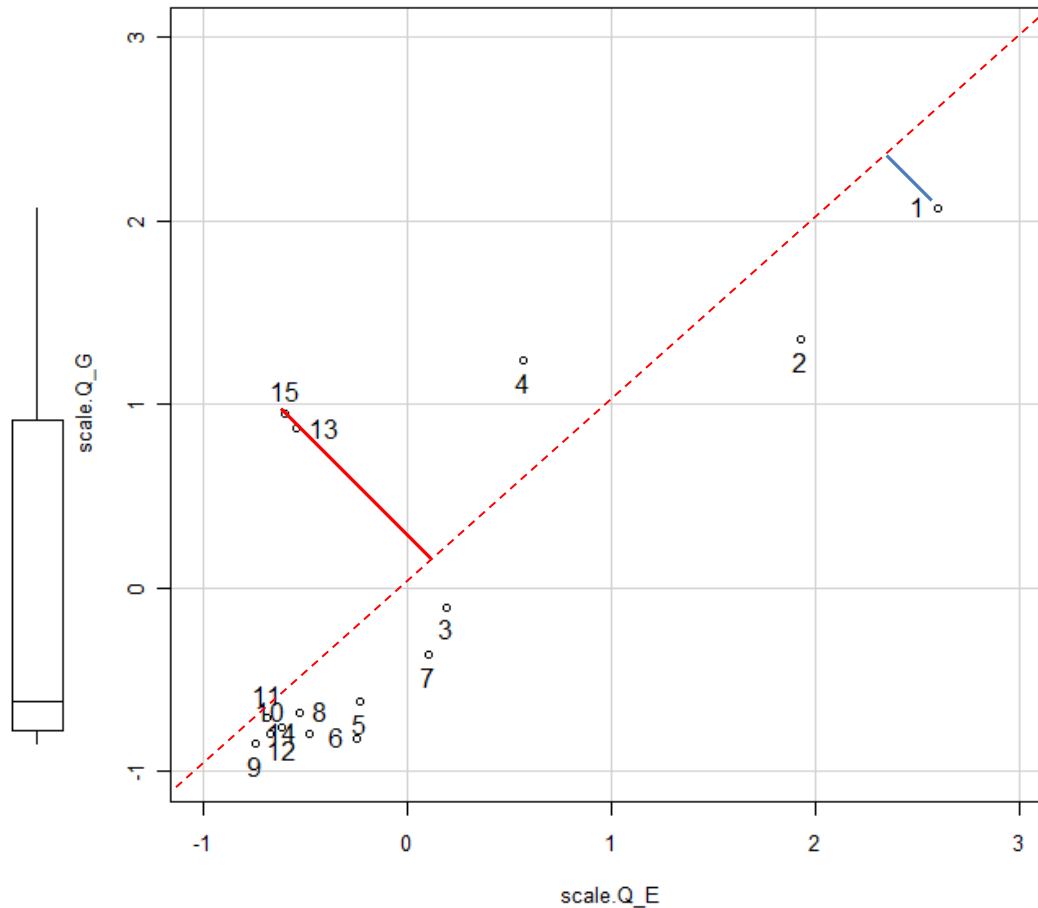
Q (~ association strength)
of pattern i
in **German**



Q
of pattern i
in **English**

Yet to be explored:
Different expressions of frequency information

Q_{centered} (\sim association strength)
of pattern i
in **German**



Q_{centered}
of pattern i
in **English**

