

# Agreement in massively parallel texts

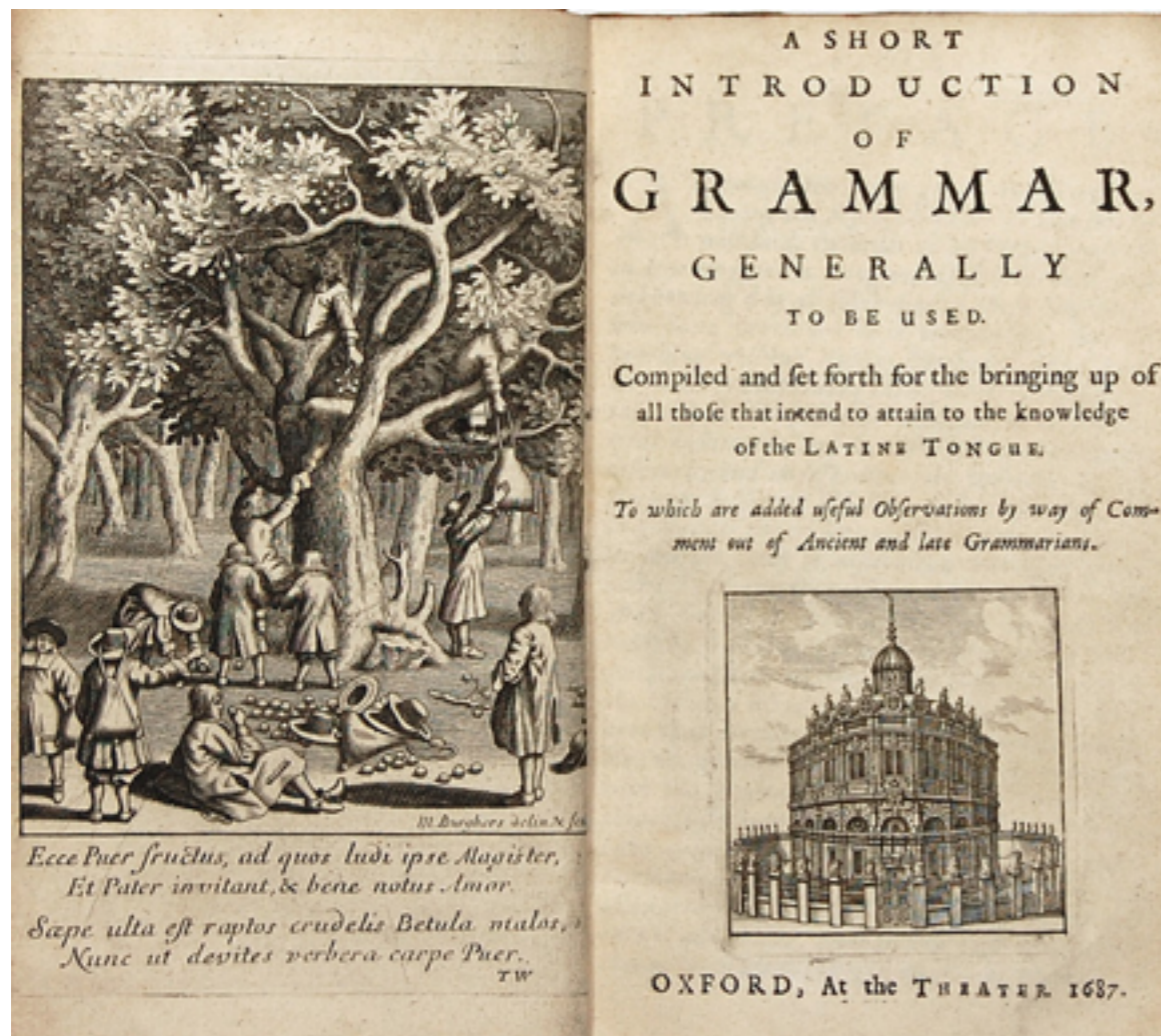
an attempt to induce typology

*Michael Cysouw*  
*Philipps-Universität Marburg*

# A short history of the term “agreement”

# Modistae

- **regere**  
(government)
- **congruitas / concordia**  
(grammaticality)
- **similitudo**  
(subject-verb agreement)



William Lily  
*A short introduction of grammar*  
(1513)

“**Concords** of Latin speech: for the due joyning of words in construction, it is to be understood, that in Latin Speech there be three Concords: The first, between the Nominative case and the Verb : The second between the Substantive and the Adjective : The third between the Antecedent and the Relative.” (Lily 1513: D4r)



## Karl Ferdinand Becker *Deutsche Sprachlehre* (1827)

“Ein Thätigkeitsbegriff z.B. blühen, groß wird auf den Begriff eines Seins, z.B. Baum, Hund dergestalt bezogen, dass beide Begriffe zu einer Einheit des Gedankens, z.B. “der Baum blühet,” “der Hund ist groß,” oder zu einer Einheit des Begriffes z.B. “der blühende Baum,” “der große Hund” verbunden werden. Dieses Verhältniß der Einheit – die **Kongruenz** – von Thätigkeit und Sein wird durch die Kongruenz der Form ausgedrückt, welche sich in der Flexion des bezogenen Wortes blüh-et, blühend-e, groß-e darstellt.” (Becker 1827: 14)

1914

An Introduction  
To The Study Of  
Language  
(1914)



1933



“in a rough way, without real boundaries,  
we can distinguish three general types of  
**agreement**” (1933: 191)

- cross-referring
- congruence
- government

- cross-reference
- concord or congruence
- government

“concord (often called agreement)”  
(Hockett 1958: 214)

“the grammar must contain agreement rules that assign to the Article all of the feature specifications for [Gender], [Number], and [Case] of the Noun it modifies”  
(Chomsky 1965: 174-175)

Steele/Moravcsik/Corbett tradition:  
“agreement” completely replaces “concord”

Chomskyan tradition:  
“agreement” is reduced to subject inflection of the verb

**Massively parallel texts**



# Aim

- Part of larger research objective to automatize linguistics
- This is not geared towards making linguists obsolete !
- Main question: which linguistic tasks are 'easy' and can be automatized
- This would leave only the interesting part of linguistics to us

# Massively parallel texts

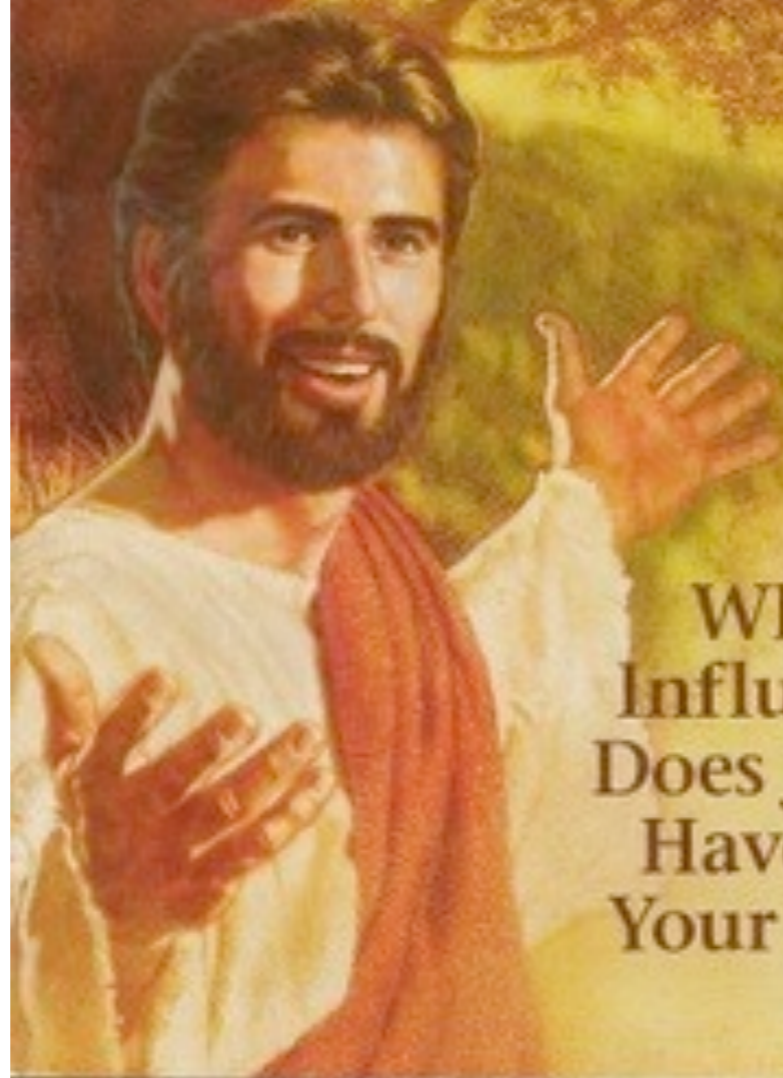
- Same text in very many languages
- Allows for relatively easy comparison, though possible problem of translationese
- Not very many texts, but enough to perform basic typological research
  - ▶ Universal Declaration of Human Rights (300 languages)
  - ▶ Bible (1000 languages)
  - ▶ Jehova's Witnesses Pamphlets (250 languages)
  - ▶ North Wind and the Sun

# Basic ideas

- **Trans co-occurrences**  
**(comparing between languages)**  
forms that significantly co-occur in parallel positions are semantically similar
- **Semantic map**  
**(comparing across languages)**  
parallel positions that recurrently use the same forms are semantically similar

DECEMBER 1, 2011

THE **WATCHTOWER**  
ANNOUNCING JEHOVAH'S KINGDOM



What  
Influence  
Does **Jesus**  
Have on  
Your Life?

- 1 What important information is contained in the Bible?
- 2 Who is the Bible's author?
- 3 Why should you study the Bible?
- 4 The Bible is a precious gift from God.
- 5 The Bible alone tells us what we must do to please God.
- 6 The Bible was written by some 40 different men over a period of 1,600 years, beginning in 1513 B.C.E.
- 7 So God in heaven, not any human on earth, is the Author of the Bible.
- 8 God made sure that the Bible was accurately copied and preserved.
- 9 More Bibles have been printed than any other book.
- 10 Not everyone will be happy to see you studying the Bible, but do not let that stop you.
- 11 But the Bible tells us that there is only one TRUE God.
- 12 But when the Bible was written, the name Jehovah appeared in it some 7,000 times
- 13 God is a Spirit, says the Bible.
- 14 The Bible reveals Jehovah's personality to us.
- 15 The Bible tells us that he is also merciful, kind, forgiving, generous, and patient.
- 16 We learn about God from creation and from the Bible.
- 17 Another way we can learn about God is by studying the Bible.
- 18 By disobeying God's command, the first man, Adam, committed what the Bible calls sin.
- 19 This is what the Bible refers to as the ransom.
- 20 Some of your loved ones may become very angry because you are studying the Bible.
- 21 What is the Bible's view of separation and of divorce?
- 22 The Bible says that a husband is the head of his family.
- 23 Parents need to spend time with their children and study the Bible with them,
- 24 When marriage mates have problems getting along together, they should try to apply Bible counsel.
- 25 The Bible urges us to show love and to be forgiving.
- 26 But God does not approve of them if they come from false religion or are against Bible teachings.
- 27 The only two birthday celebrations spoken of in the Bible were held by persons who did not worship Jehovah.
- 28 The Bible teaches that only a few people are on the narrow road to life.
- 29 The Bible foretold that after the death of the apostles, ...
- 30 True Christians love one another, respect the Bible, and preach about God's Kingdom.
- 31 Another mark of true religion is that its members have a deep respect for the Bible.
- 32 They try to live by the Bible in their everyday life.
- 33 The Bible is the basis for what is taught.
- 34 By now you have learned many good things from the Bible.

# Trans co-occurrences of English word “bible”

---

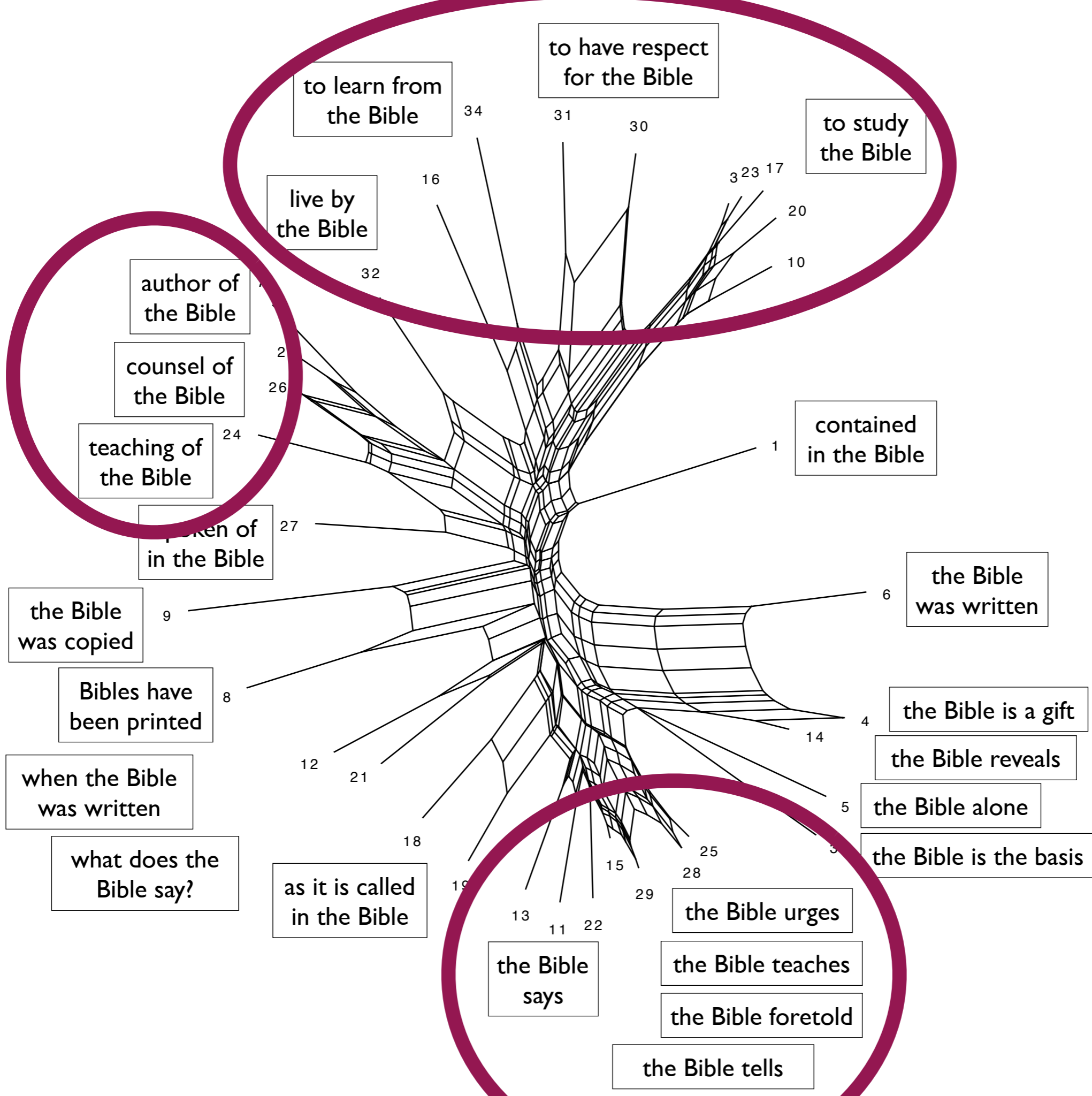
Albanian	Faroese	Estonian	Greenlandic
<i>bibla</i>	<i>biblian</i>	<i>piibel</i>	<i>biibilip</i>
<i>biblën</i>	<i>bibliuna</i>	<i>piiblit</i>	<i>biibli</i>
<i>biblës</i>	<i>bibliunnar</i>	<i>piibli</i>	<i>biibilmik</i>
	<i>bibliuni</i>	<i>piiblis</i>	<i>biibilmi</i>
		<i>piiblist</i>	

---

Context	Albanian	Faroese	Estonian	Greenlandic
1	bibla	bíbliuni	piibel	biibili
2	biblës	bíbliunnar	piibli	biibilimik
3	biblën	bíbliuna	piiblit	biibili
4	bibla	bíblían	piibel	biibili
5	bibla	bíblían	piibel	biibilip
6	bibla	bíbliuna	piibli	biibili
7	biblës	bíbliunnar	piibli	biibilimut
8	bibla	bíblían	piiblit	biibilip
9	bibla	NA	piiblit	biibili
10	biblën	bíbliuna	piiblit	biibilimik
11	bibla	bíblían	piibel	biibilimili
12	bibla	bíblían	piibel	biibilili
13	bibla	bíblían	piibel	biibilimi
14	bibla	bíblían	piibel	biibilimi
15	bibla	bíblían	piibel	biibilimi
16	bibla	bíbliuni	piibli	biibililu
17	biblën	bíbliuna	piiblit	biibilimik
18	bibla	bíblían	piiblis	biibilip
19	bibla	bíblían	piiblis	biibilimi
20	biblën	bíbliuna	piiblit	biibilimik
21	NA	bíblían	piibel	biibilimi
22	bibla	bíbliuni	piibel	biibili
23	biblën	bíbliuna	piiblit	biibilimillu
24	biblike	bíblían	piibli	biibilimi
25	bibla	bíblían	piibel	biibilimi
26	biblës	bíbliunnar	piibli	biibilimi
27	bibla	bíblían	piiblis	biibilimi
28	bibla	bíblían	piibel	biibilimi
29	bibla	bíblían	piibel	biibilimi
30	biblën	bíbliuna	piiblist	biibilimik
31	biblën	bíbliuni	piibli	biibilimik
32	biblës	bíbliuni	piibli	biibili
33	bibla	bíbliuna	piibel	biibilimik
34	bibla	bíbliuni	piiblist	biibilimeersunik

0.00	0.67	0.50	0.75	0.67	0.75	0.69	0.75	0.73	0.56	0.44	0.63	0.53	0.75	0.53	0.56	0.63	0.50	0.63	0.63	0.60	0.50	0.56	0.69	0.75	0.57	0.44	0.69	0.56	0.63	0.63	0.50	0.75	0.63
0.67	0.00	0.73	0.80	0.73	0.67	0.20	0.73	0.64	0.60	0.80	0.67	0.93	0.73	0.57	0.80	0.73	0.73	0.73	0.67	0.50	0.67	0.73	0.40	0.60	0.15	0.53	0.67	0.67	0.60	0.60	0.38	0.73	0.80
0.50	0.73	0.00	0.88	0.80	0.69	0.81	0.75	0.80	0.25	0.75	0.75	0.73	0.88	0.80	0.69	0.13	0.69	0.81	0.19	0.73	0.81	0.06	0.75	0.88	0.64	0.69	0.88	0.88	0.44	0.63	0.64	0.75	0.69
0.75	0.80	0.88	0.00	0.40	0.31	0.81	0.56	0.53	0.88	0.44	0.44	0.40	0.13	0.47	0.75	0.94	0.63	0.63	0.94	0.53	0.44	0.94	0.88	0.25	0.93	0.81	0.31	0.44	1.00	1.00	0.79	0.50	0.94
0.67	0.73	0.80	0.40	0.00	0.60	0.80	0.60	0.79	0.80	0.47	0.47	0.57	0.47	0.50	0.73	0.87	0.53	0.60	0.87	0.57	0.53	0.80	0.80	0.47	0.77	0.67	0.47	0.47	0.93	1.00	0.85	0.53	0.87
0.75	0.67	0.69	0.31	0.60	0.00	0.69	0.63	0.53	0.75	0.63	0.63	0.67	0.44	0.67	0.81	0.75	0.75	0.75	0.88	0.60	0.56	0.75	0.81	0.56	0.79	0.81	0.63	0.63	0.81	0.81	0.71	0.63	0.94
0.69	0.20	0.81	0.81	0.80	0.69	0.00	0.63	0.60	0.75	0.81	0.75	0.93	0.75	0.67	0.81	0.81	0.81	0.81	0.81	0.60	0.75	0.81	0.50	0.75	0.29	0.63	0.69	0.69	0.69	0.69	0.50	0.81	0.81
0.75	0.73	0.75	0.56	0.60	0.63	0.63	0.00	0.47	0.88	0.56	0.50	0.67	0.63	0.53	0.75	0.75	0.63	0.69	0.81	0.80	0.69	0.75	0.75	0.63	0.79	0.75	0.56	0.56	0.75	0.75	0.86	0.81	0.75
0.73	0.64	0.80	0.53	0.79	0.53	0.60	0.47	0.00	0.73	0.80	0.73	0.86	0.67	0.64	0.80	0.87	0.87	0.87	0.80	0.86	0.67	0.87	0.73	0.67	0.69	0.73	0.67	0.73	0.80	0.80	0.62	0.80	0.80
0.56	0.60	0.25	0.88	0.80	0.75	0.75	0.88	0.73	0.00	0.81	0.75	0.80	0.81	0.80	0.69	0.25	0.69	0.81	0.19	0.67	0.88	0.25	0.69	0.81	0.57	0.63	0.81	0.88	0.44	0.63	0.57	0.63	0.81
0.44	0.80	0.75	0.44	0.47	0.63	0.81	0.56	0.80	0.81	0.00	0.56	0.27	0.38	0.27	0.69	0.81	0.44	0.44	0.81	0.53	0.31	0.75	0.81	0.38	0.86	0.63	0.31	0.19	0.88	0.88	0.86	0.56	0.81
0.63	0.67	0.75	0.44	0.47	0.63	0.75	0.50	0.73	0.75	0.56	0.00	0.53	0.50	0.47	0.69	0.81	0.44	0.56	0.81	0.33	0.63	0.75	0.63	0.50	0.64	0.63	0.56	0.56	0.88	0.88	0.71	0.69	0.81
0.53	0.93	0.73	0.40	0.57	0.67	0.93	0.67	0.86	0.80	0.27	0.53	0.00	0.27	0.36	0.60	0.73	0.40	0.33	0.80	0.43	0.40	0.73	0.73	0.33	0.77	0.60	0.33	0.33	0.80	0.80	0.85	0.60	0.73
0.75	0.73	0.88	0.13	0.47	0.44	0.75	0.63	0.67	0.81	0.38	0.50	0.27	0.00	0.33	0.69	0.88	0.56	0.50	0.88	0.40	0.44	0.88	0.75	0.13	0.79	0.69	0.19	0.31	0.94	0.94	0.79	0.44	0.88
0.53	0.57	0.80	0.47	0.50	0.67	0.67	0.53	0.64	0.80	0.27	0.47	0.36	0.33	0.00	0.73	0.87	0.33	0.27	0.80	0.50	0.20	0.80	0.53	0.20	0.57	0.33	0.20	0.07	0.80	0.80	0.64	0.60	0.73
0.56	0.80	0.69	0.75	0.73	0.81	0.81	0.75	0.80	0.69	0.69	0.69	0.60	0.69	0.73	0.00	0.63	0.63	0.75	0.69	0.80	0.75	0.69	0.75	0.75	0.71	0.75	0.75	0.81	0.69	0.56	0.57	0.63	0.44
0.63	0.73	0.13	0.94	0.87	0.75	0.81	0.75	0.87	0.25	0.81	0.81	0.73	0.88	0.87	0.63	0.00	0.75	0.88	0.19	0.80	0.94	0.13	0.81	0.94	0.71	0.75	0.94	0.94	0.31	0.50	0.79	0.63	0.63
0.50	0.73	0.69	0.63	0.53	0.75	0.81	0.63	0.87	0.69	0.44	0.44	0.40	0.56	0.33	0.63	0.75	0.00	0.25	0.75	0.53	0.44	0.69	0.63	0.56	0.64	0.38	0.56	0.44	0.81	0.81	0.64	0.69	0.69
0.63	0.73	0.81	0.63	0.60	0.75	0.81	0.69	0.87	0.81	0.44	0.56	0.33	0.50	0.27	0.75	0.88	0.25	0.00	0.81	0.47	0.31	0.81	0.56	0.38	0.57	0.31	0.38	0.25	0.81	0.81	0.64	0.69	0.75
0.63	0.67	0.19	0.94	0.87	0.88	0.81	0.81	0.80	0.19	0.81	0.81	0.80	0.88	0.80	0.69	0.19	0.75	0.81	0.00	0.80	0.88	0.19	0.75	0.81	0.64	0.69	0.81	0.88	0.44	0.63	0.64	0.69	0.69
0.60	0.50	0.73	0.53	0.57	0.60	0.60	0.80	0.86	0.67	0.53	0.33	0.43	0.40	0.50	0.80	0.80	0.53	0.47	0.80	0.00	0.60	0.73	0.40	0.40	0.46	0.53	0.47	0.47	0.80	0.80	0.62	0.67	0.87
0.50	0.67	0.81	0.44	0.53	0.56	0.75	0.69	0.67	0.88	0.31	0.63	0.40	0.44	0.20	0.75	0.94	0.44	0.31	0.88	0.60	0.00	0.88	0.75	0.31	0.71	0.56	0.31	0.19	0.88	0.81	0.57	0.56	0.75
0.56	0.73	0.06	0.94	0.80	0.75	0.81	0.75	0.87	0.25	0.75	0.75	0.73	0.88	0.80	0.69	0.13	0.69	0.81	0.19	0.73	0.88	0.00	0.75	0.88	0.64	0.69	0.88	0.88	0.44	0.63	0.71	0.75	0.69
0.69	0.40	0.75	0.88	0.80	0.81	0.50	0.75	0.73	0.69	0.81	0.63	0.73	0.75	0.53	0.75	0.81	0.63	0.56	0.75	0.40	0.75	0.75	0.00	0.63	0.21	0.38	0.63	0.63	0.69	0.63	0.43	0.81	0.75
0.75	0.60	0.88	0.25	0.47	0.56	0.75	0.63	0.67	0.81	0.38	0.50	0.33	0.13	0.20	0.75	0.94	0.56	0.38	0.81	0.40	0.31	0.88	0.63	0.00	0.64	0.56	0.06	0.19	0.88	0.88	0.64	0.50	0.81
0.57	0.15	0.64	0.93	0.77	0.79	0.29	0.79	0.69	0.57	0.86	0.64	0.77	0.79	0.57	0.71	0.71	0.64	0.57	0.64	0.46	0.71	0.64	0.21	0.64	0.00	0.36	0.64	0.64	0.57	0.57	0.31	0.86	0.71
0.44	0.53	0.69	0.81	0.67	0.81	0.63	0.75	0.73	0.63	0.63	0.63	0.60	0.69	0.33	0.75	0.75	0.38	0.31	0.69	0.53	0.56	0.69	0.38	0.56	0.36	0.00	0.56	0.44	0.63	0.63	0.43	0.81	0.69
0.69	0.67	0.88	0.31	0.47	0.63	0.69	0.56	0.67	0.81	0.31	0.56	0.33	0.19	0.20	0.75	0.94	0.56	0.38	0.81	0.47	0.31	0.88	0.63	0.06	0.64	0.56	0.00	0.13	0.88	0.88	0.64	0.50	0.81
0.56	0.67	0.88	0.44	0.47	0.63	0.69	0.56	0.73	0.88	0.19	0.56	0.33	0.31	0.07	0.81	0.94	0.44	0.25	0.88	0.47	0.19	0.88	0.63	0.19	0.64	0.44	0.13	0.00	0.88	0.88	0.71	0.56	0.81
0.63	0.60	0.44	1.00	0.93	0.81	0.69	0.75	0.80	0.44	0.88	0.88	0.80	0.94	0.80	0.69	0.31	0.81	0.81	0.44	0.80	0.88	0.44	0.69	0.88	0.57	0.63	0.88	0.88	0.00	0.31	0.57	0.75	0.56
0.63	0.60	0.63	1.00	1.00	0.81	0.69	0.75	0.80	0.63	0.88	0.88	0.80	0.94	0.80	0.56	0.50	0.81	0.81	0.63	0.80	0.81	0.63	0.63	0.88	0.57	0.63	0.88	0.88	0.31	0.00	0.36	0.81	0.56
0.50	0.38	0.64	0.79	0.85	0.71	0.50	0.86	0.62	0.57	0.86	0.71	0.85	0.79	0.64	0.57	0.79	0.64	0.64	0.64	0.62	0.57	0.71	0.43	0.64	0.31	0.43	0.64	0.71	0.57	0.36	0.00	0.79	0.64
0.75	0.73	0.75	0.50	0.53	0.63	0.81	0.81	0.80	0.63	0.56	0.69	0.60	0.44	0.60	0.63	0.63	0.69	0.69	0.69	0.67	0.56	0.75	0.81	0.50	0.86	0.81	0.50	0.56	0.75	0.81	0.79	0.00	0.81
0.63	0.80	0.69	0.94	0.87	0.94	0.81	0.75	0.80	0.81	0.81	0.81	0.73	0.88	0.73	0.44	0.63	0.69	0.75	0.69	0.87	0.75	0.69	0.75	0.81	0.71	0.69	0.81	0.81	0.56	0.56	0.64	0.81	0.00





# Agreement ?

- Strict definition: formal co-variation between forms
- Necessary:
  - some approximation of morphology
  - some approximation of word classes

# Trans co-occurrences of English word “his”

German	Bruder	Vater	Sohn	Hand	Mutter	Weib	Haupt
seine	0	0	0	125	60	0	0
seinen	24	11	34	0	0	0	0
seinem	25	39	24	0	0	2	10
seiner	0	0	0	45	33	0	0
sein	19	63	38	3	4	56	51
seines	0	1	0	0	0	0	0

German	Vater	Vaters	Väter	Vätern
seine	0	0	7	0
seinen	11	0	0	27
seinem	39	0	0	0
seiner	0	0	7	0
sein	63	0	0	0
seines	1	79	0	0

German	Sohn	Sohnes	Söhne	Söhnen
seine	0	0	77	1
seinen	34	0	0	49
seinem	24	0	0	0
seiner	0	0	19	0
sein	38	0	0	0
seines	0	16	0	0

# Trans co-occurrences of English word “his”

French	frère	père	fil	main	mère	frères	mains
sa	0	0	0	122	68	0	0
ses	0	0	104	0	0	44	41
son	33	92	51	0	0	0	0

# Trans co-occurrences of English word “his”

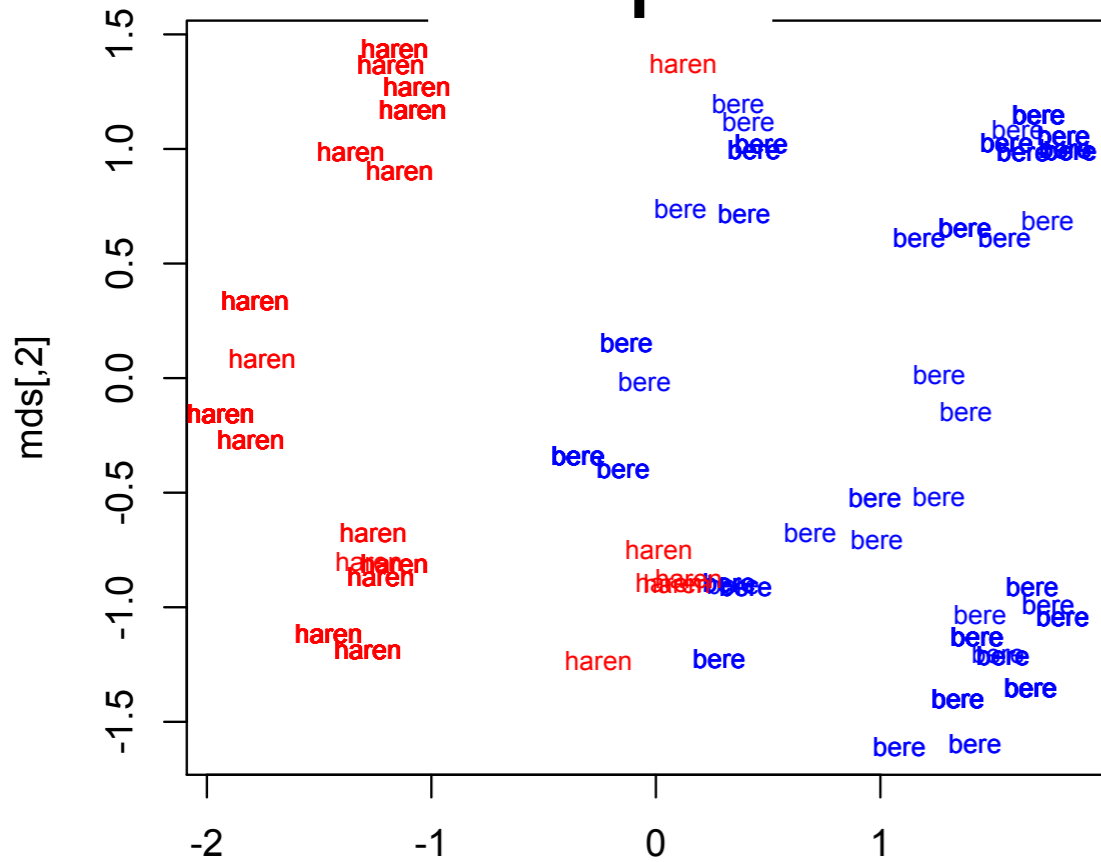
Swedish	broder	fader	son	hand	moder	huvud	land
hans	27	93	126	32	51	30	10
sin	33	89	44	91	20	0	0
sitt	0	0	0	0	0	12	17

# Trans co-occurrences of English word “his”

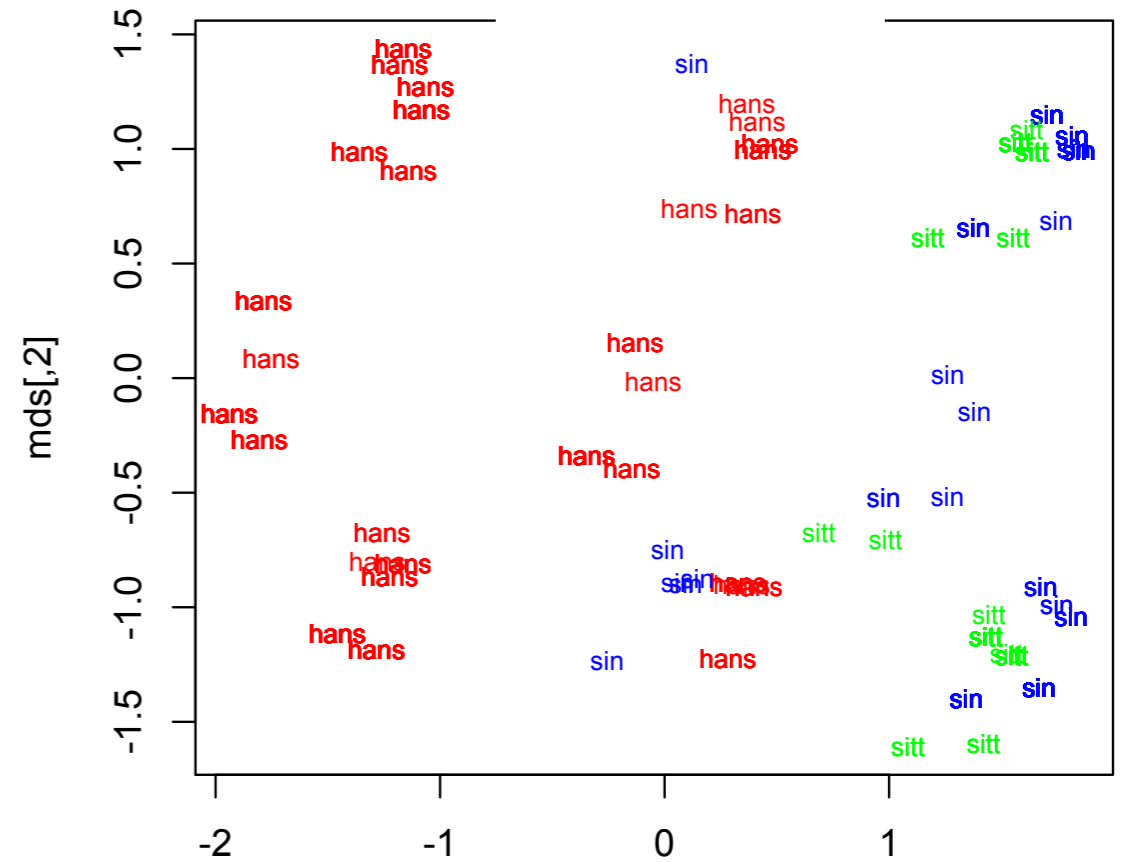
Basque	anayea	aita	semea	escua	ama	buru	lurrera
bere	3	9	6	3	3	6	0
haren	0	5	0	2	7	7	0



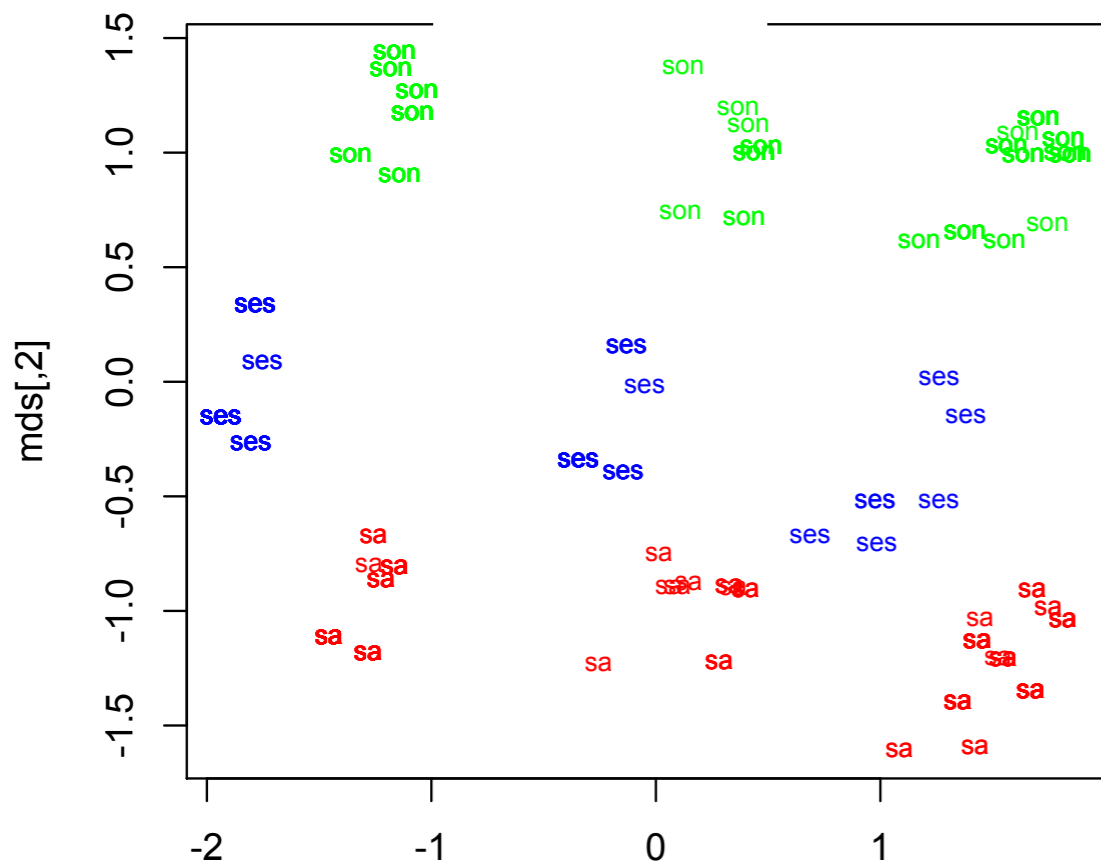
# Basque



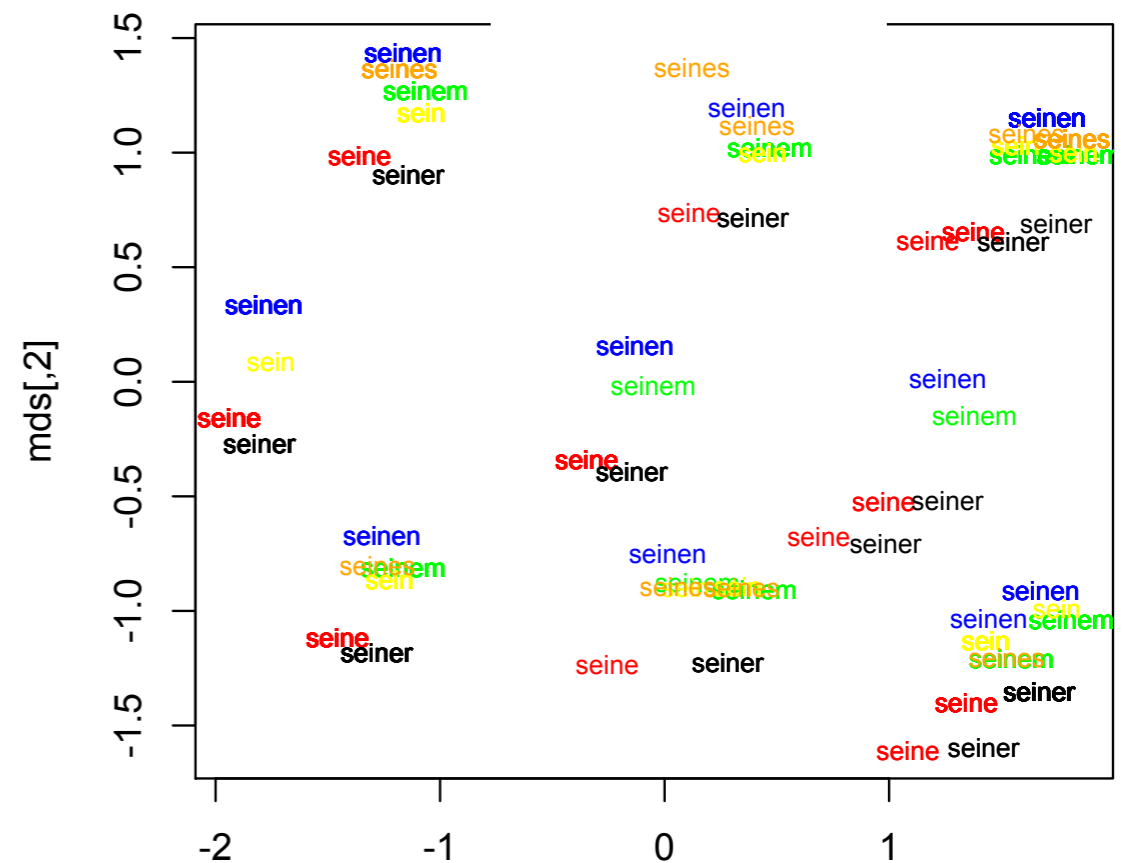
# Swedish



# French



# German



# Conclusions

- Agreement (concord) is an interesting phenomenon for us linguists  
*At least much more interesting than semantic roles and alignment, which are relatively easily extracted automatically*
- For automatic extraction of agreement, we need better inferences of morphology and word classes  
*There are various proposals around in computational linguistics, but it is unclear how well they generalize cross-linguistically*