Massively Parallel Text

language comparison using methods from vector-space modeling

Michael Cysouw
Philipps-Universität Marburg
cysouw@uni-marburg.de

- There is a deep methodological affinity between the
  - **Semantic Map** approach from linguistic typology, and
  - **Vector Space** models from corpus linguistics
- Parallel text offer the possibility to link these two fields

Meaning

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

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

A bit of history ...

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.

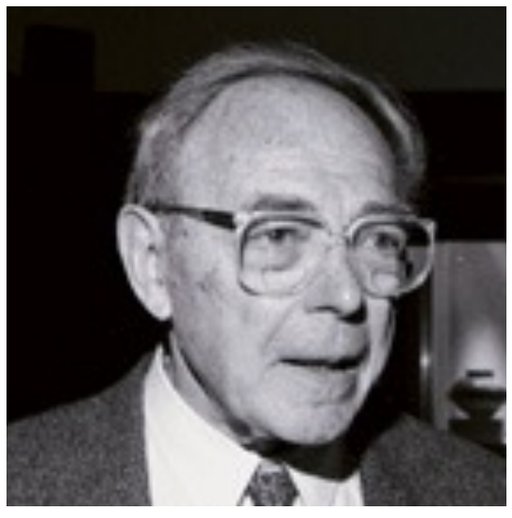
It would hardly be practical to do this by means of a generalized dictionary which contains all possible phrases 2N+1 words long: for the number of such phases is horrifying, even to a modern electronic computer

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)



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

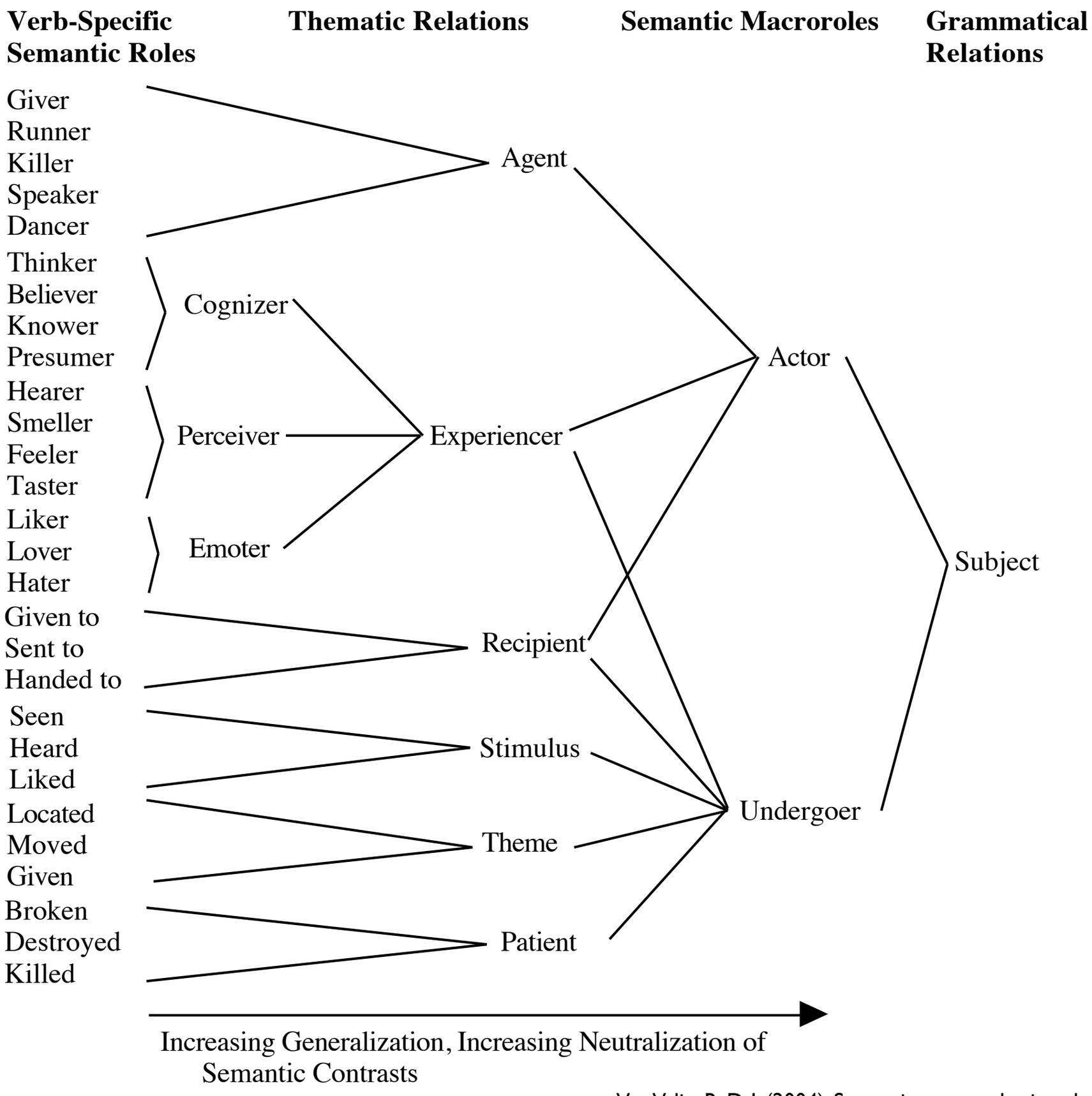
Semantics

- **Exemplar Semantics**
  Meaning of linguistic forms is (re-)instantiated and (re-)produced in individual utterances
- **Similarity Semantics**
  Meaning of linguistic forms is defined relationally to each other

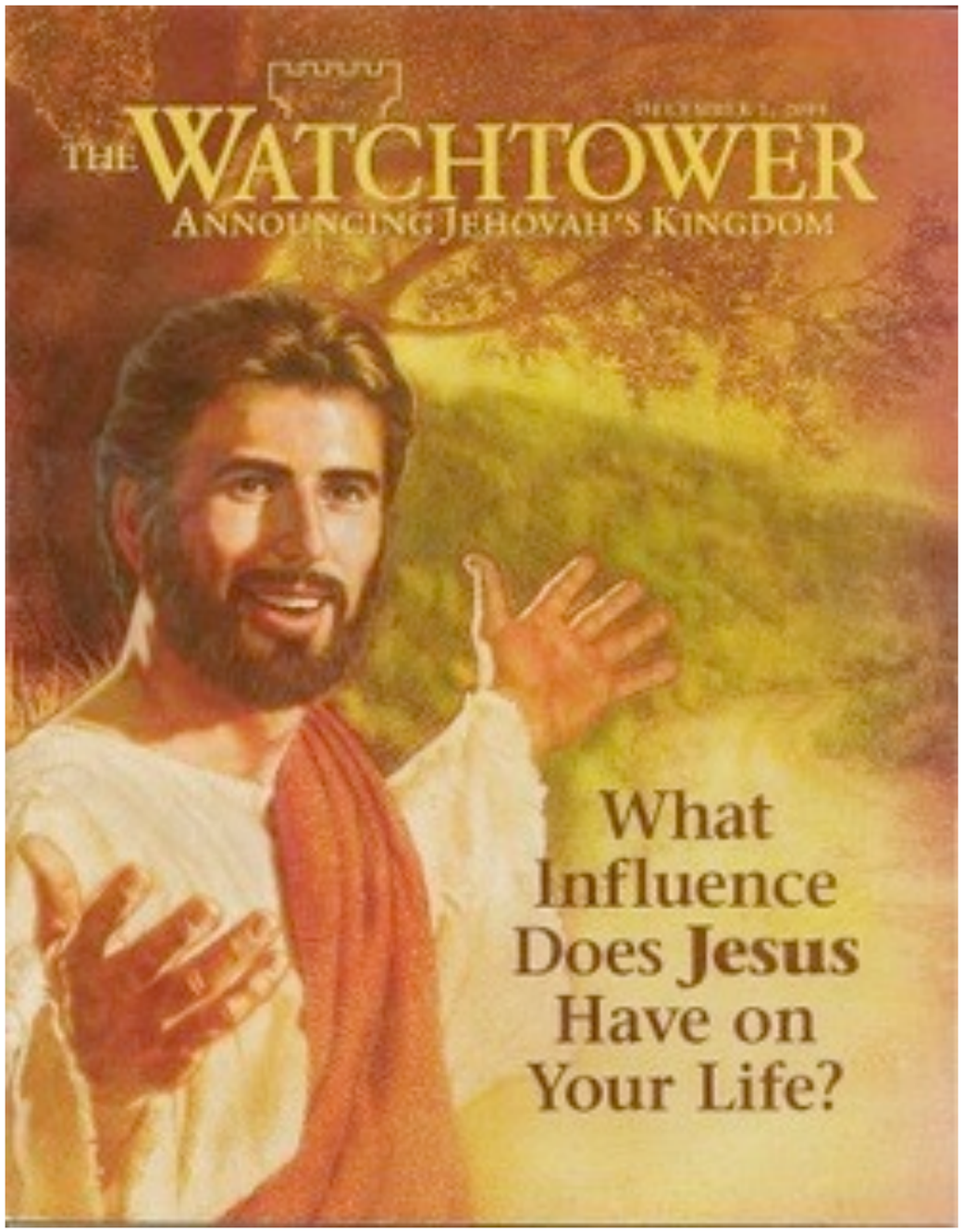
In practice

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

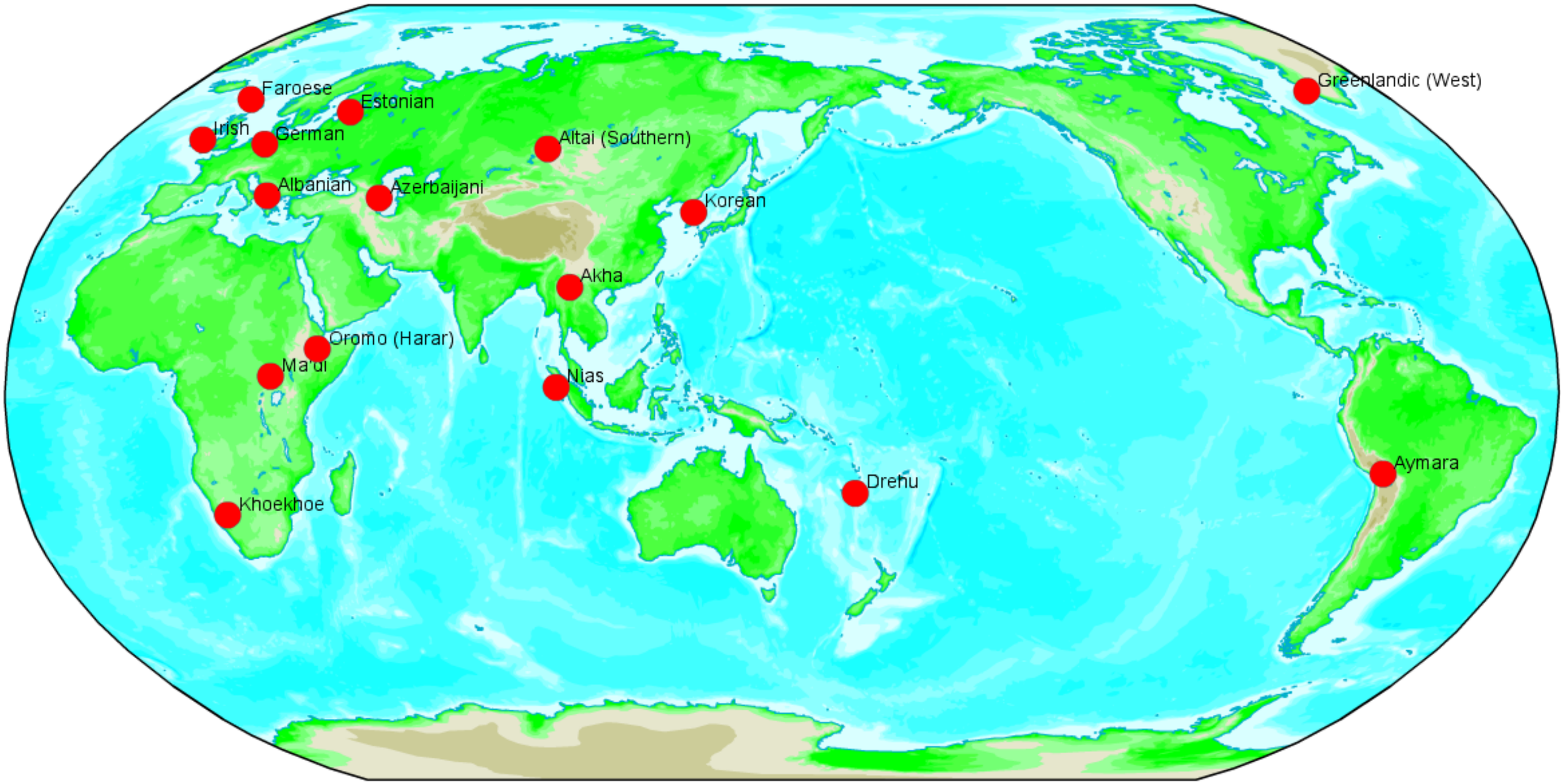
Inducing Semantic Roles



Van Valin, R. D. J. (2004). Semantic macroroles in role and reference grammar. In R. Kailuweit & M. Hummel (Eds.), Semantische rollen. (pp. 62-82). Tübingen: Narr.


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.


Faroese
Estonian
Irish
German
Albanian
Azerbaijani
Altai (Southern)
Korean
Akha
Oromo (Harar)
Ma'di
Nias
Khoekhoe
Dreehu
Greenlandic (West)
Aymara

Language	Genus	Family	Macroarea	Alignment
Oromo	Eastern Cushitic	Afro-Asiatic	Africa	Marked nominative
Khoekhoe	Central Khoisan	Khoisan	Africa	Nominative-accusative
Ma'di	Moru-Ma'di	Nilo-Saharan	Africa	No case marking (?)
Albanian	Albanian	Indo-European	Europe	Nominative-accusative
Irish	Celtic	Indo-European	Europe	No case marking (?)
Faroese	Germanic	Indo-European	Europe	Nominative-accusative
Estonian	Finnic	Uralic	Europe	Nominative-accusative
Altai	Turkic	Altaic	Asia	Nominative-accusative
Azerbaijani	Turkic	Altaic	Asia	Nominative-accusative
Korean	Korean	Korean	Asia	Nominative-accusative
Akha	Burmese-Lolo	Sino-Tibetan	Asia	Ergative-absolutive (?)
Drehu	Oceanic	Austronesian	Pacific	Active-inactive
Nias	Sundic	Austronesian	Pacific	Marked absolutive
Greenlandic	Eskimo-Aleut	Eskimo-Aleut	America	Ergative-absolutive
Aymara	Aymaran	Aymaran	America	Marked Nominative (?)

Albanian	Faroese	Estonian	Greenlandic
bibla Nominative	*biblian* Nominative	*þiibel* Nominative	*biibiliþ* Ergative
biblën Accusative	*bibliuna* Accusative	*þiiblit* Partitive	*biibli* Absolutive
biblës Genitive/Dative	*bibliunnar* Genitive	*þiibli* Genitive	*biibilmik* Instrumental
…	*bibliuni* Dative	*þiiblis* Inessive	*biibilmi* Locative
	…	*þiiblist* Elative	…
		…	

Comparing Roles

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

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		0.53	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			0.60	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			0.47	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																0.81	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																0.44	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																0.80	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																0.88	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																0.44	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																0.69	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																0.88	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

There is no difference between context 5 and itself in any language

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				0.69	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				0.75	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				0.73	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																		.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																		.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																		.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																		.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																		.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																		.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																		.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																		.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																		.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																		.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																		.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																		.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

0.13

On average, there is a small difference between the role in context 4 and 14

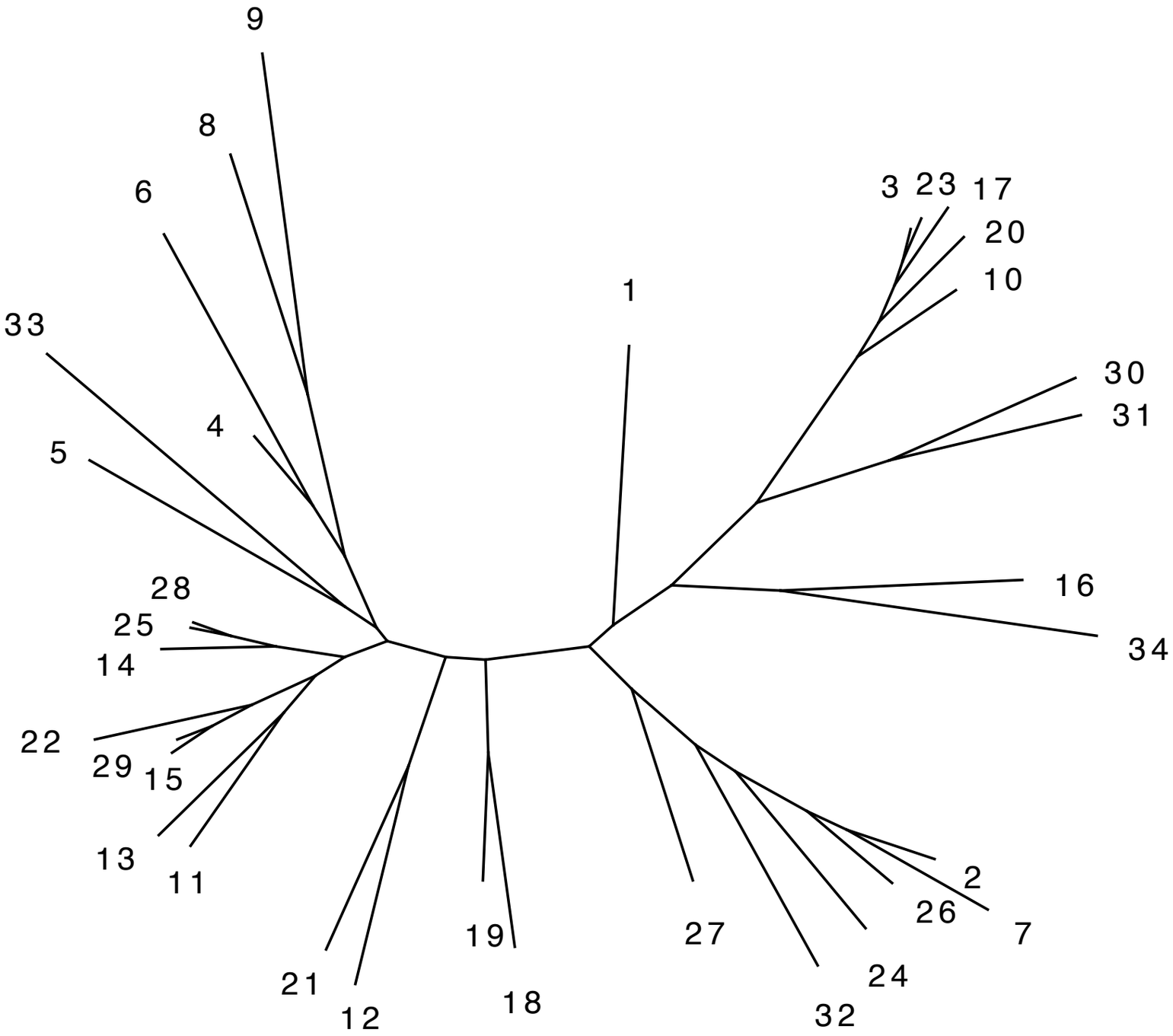
4: The Bible is a precious gift from God.
14: The Bible reveals Jehovah's personality to us.

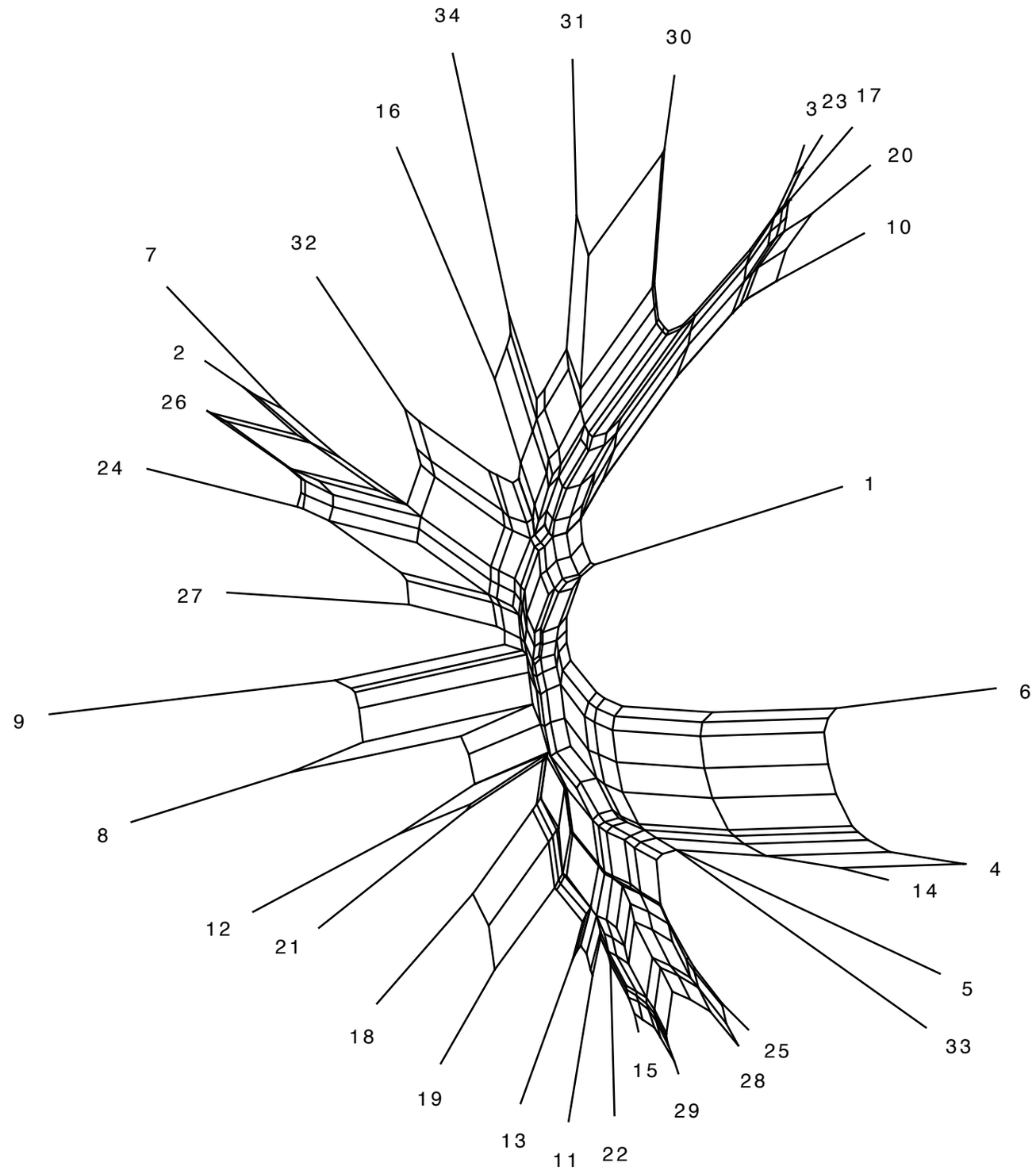
0.88

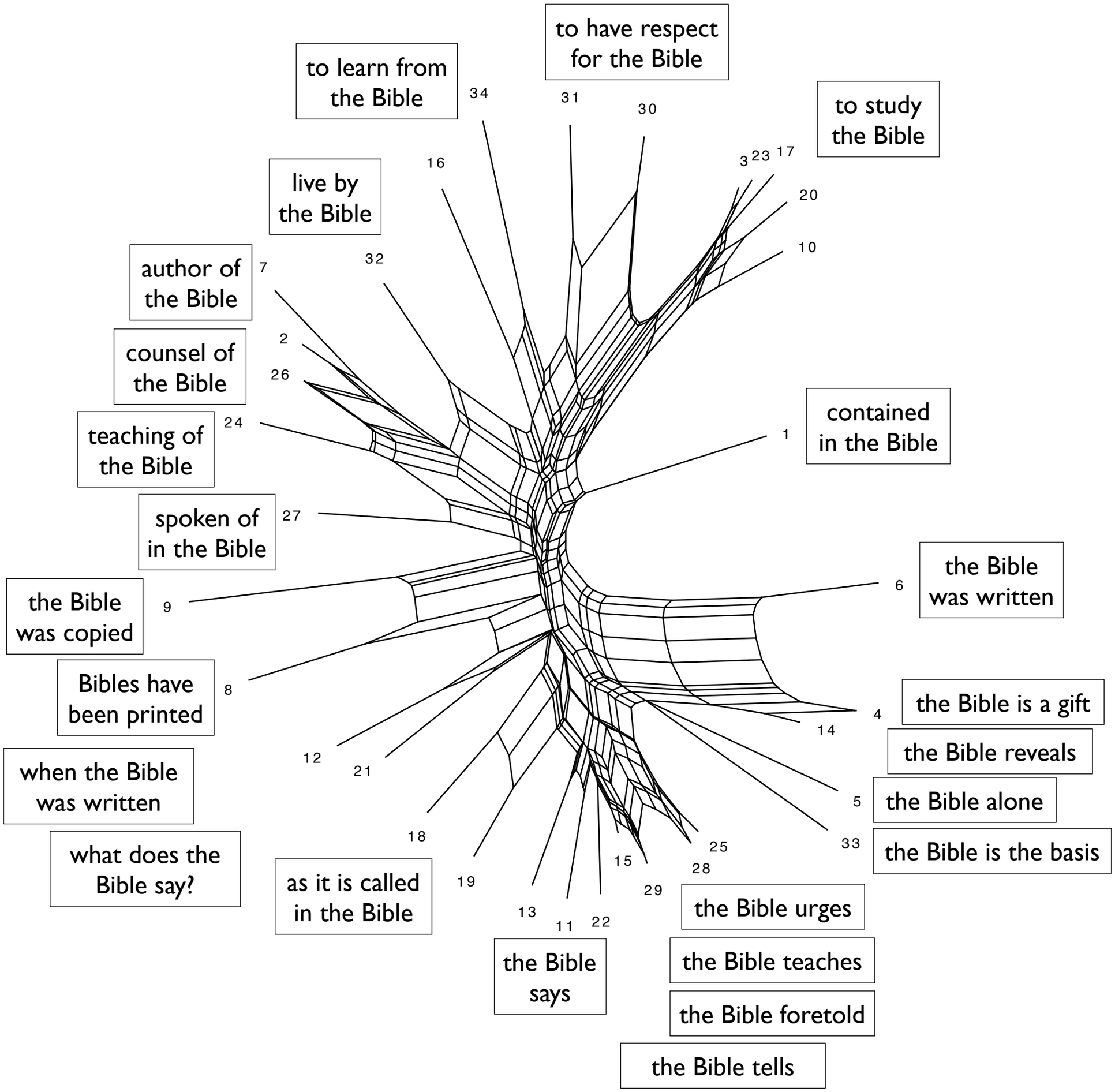
On average, there is a large difference between the role in context 8 and 10

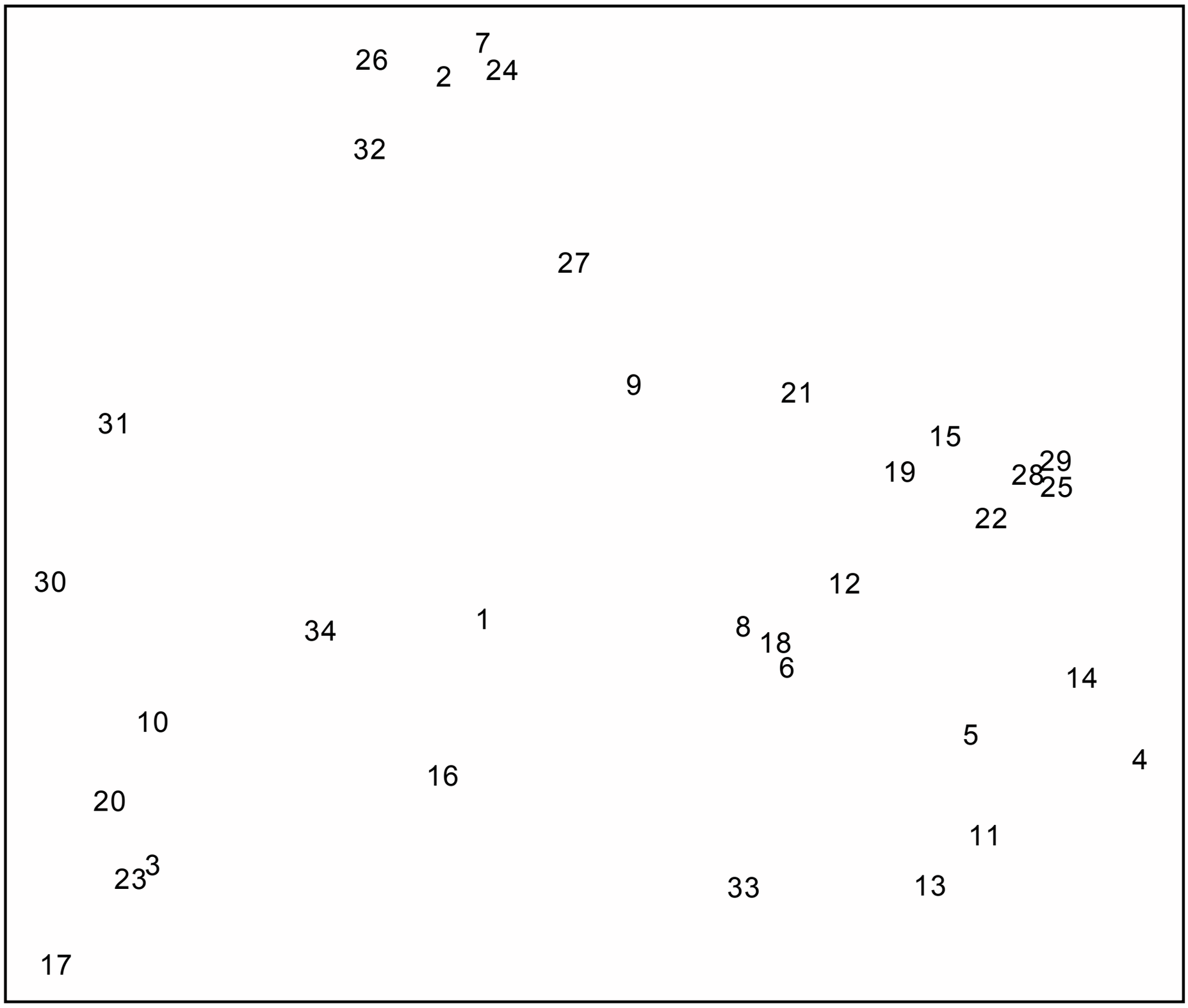
8: God made sure that the Bible was accurately copied and preserved.

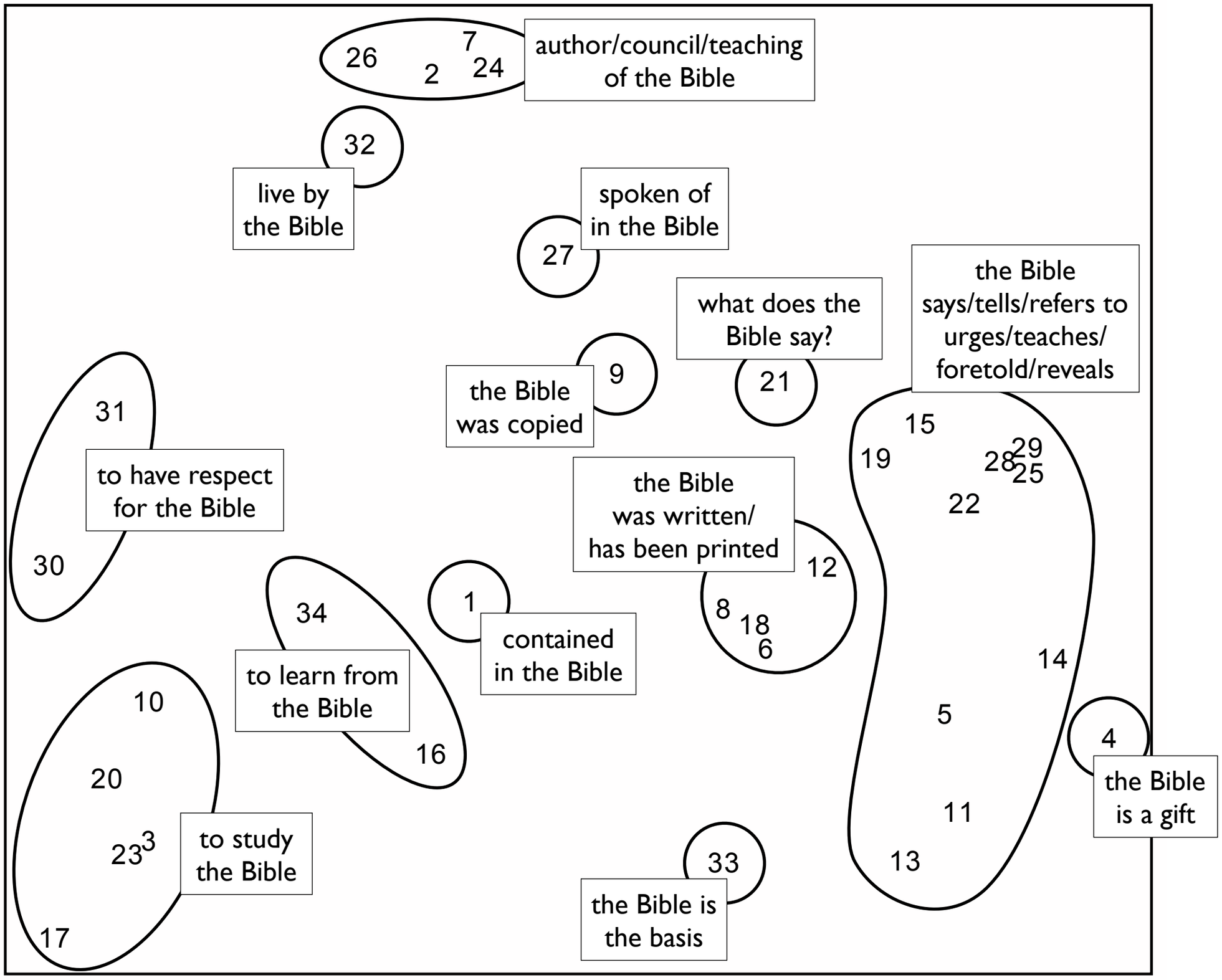
10: Not everyone will be happy to see you studying the Bible.

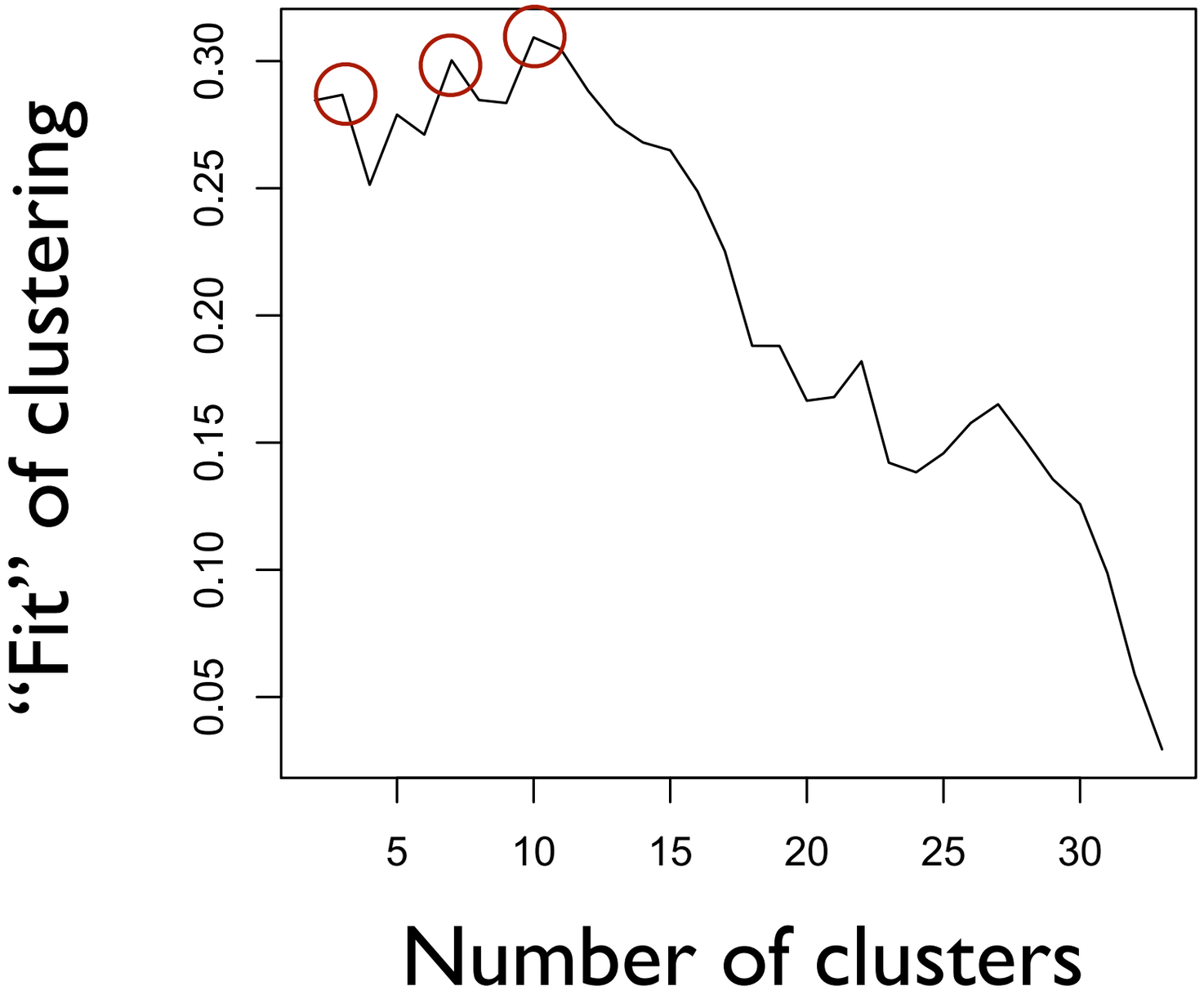

9
8
6
33
4
5
28
25
14
22
29
15
13
11
21
12
19
18
27
32
24
2
26
7
3
23
17
20
10
30
31
16
34
1

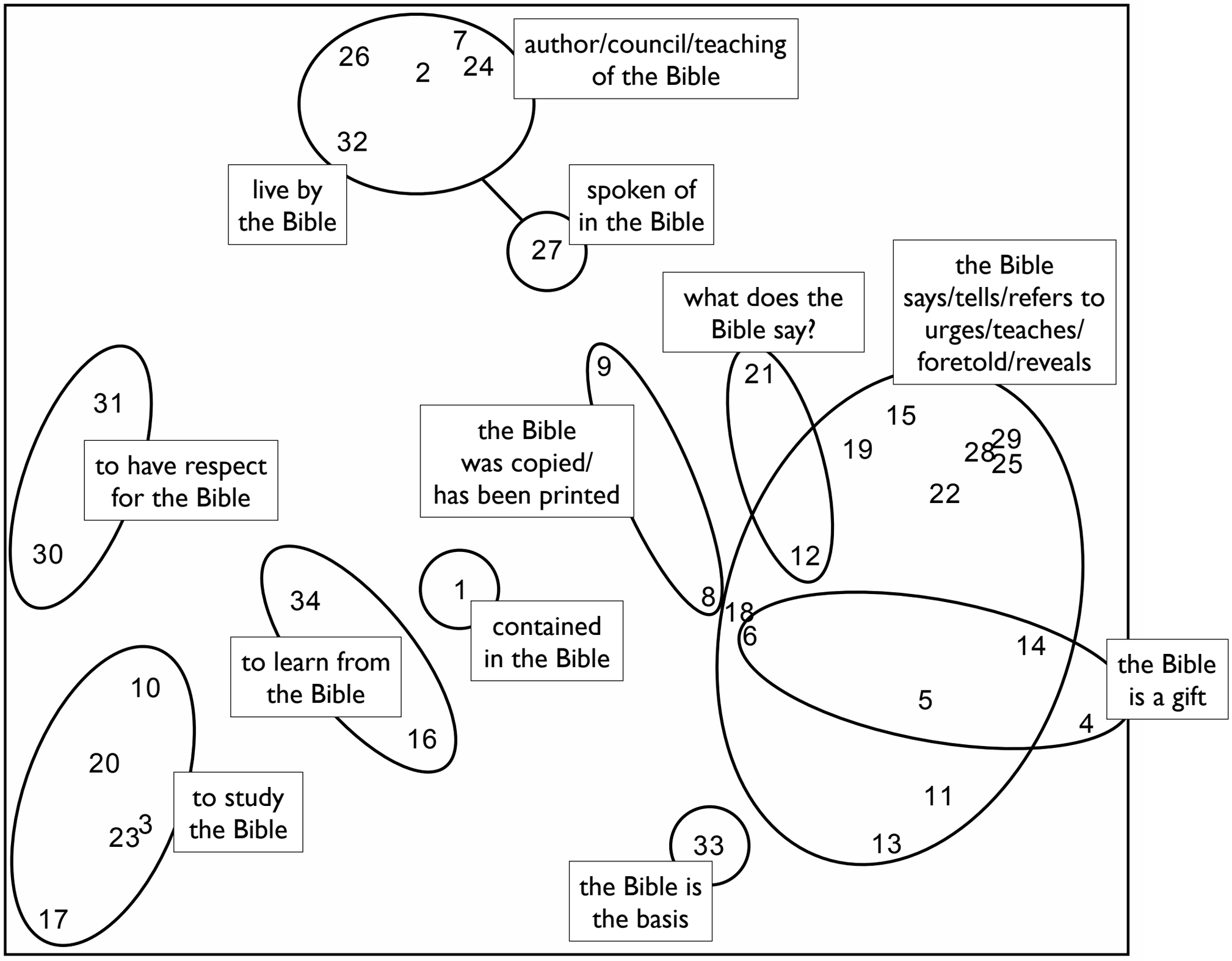

34
31
30
16
3
23
17
20
10
32
7
2
26
24
1
27
9
6
8
4
14
12
21
5
18
33
25
15
28
19
29
13
11
22

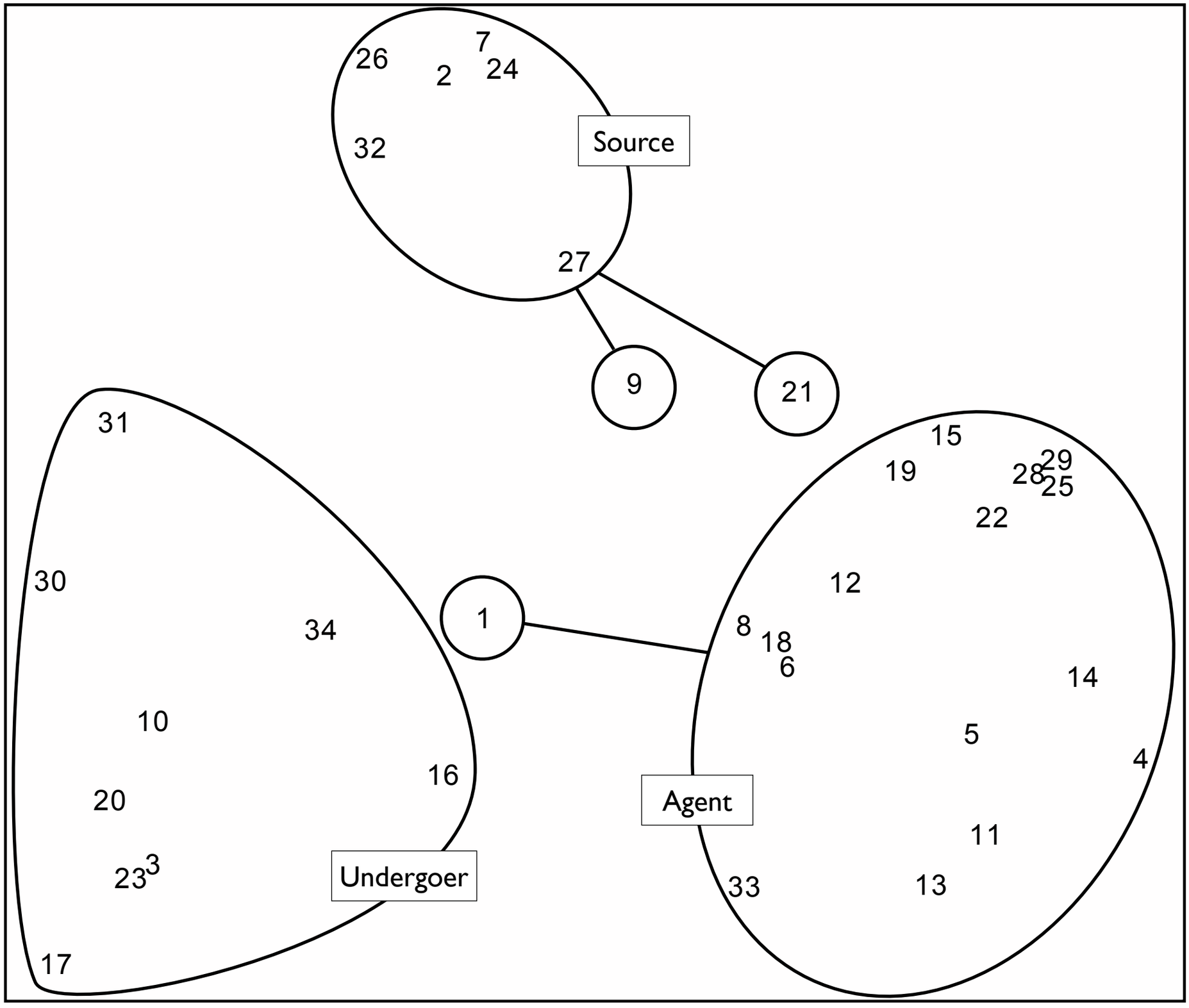

to have respect for the Bible
to learn from the Bible
to study the Bible
live by the Bible
author of the Bible
counsel of the Bible
teaching of the Bible
spoken of in the Bible
contained in the Bible
the Bible was written
the Bible was copied
Bibles have been printed
the Bible is a gift
when the Bible was written
the Bible reveals
what does the Bible say?
the Bible alone
as it is called in the Bible
the Bible is the basis
the Bible says
the Bible urges
the Bible teaches
the Bible foretold
the Bible tells
34
31
30
16
3
23
17
20
10
32
7
2
26
24
1
27
9
6
8
4
14
12
21
5
18
33
19
25
15
28
13
29
11
22


26
7
2
24
32
27
9
21
31
15
19
29
28
25
22
30
12
1
34
8
18
6
14
10
5
4
16
20
11
23
3
33
13
17

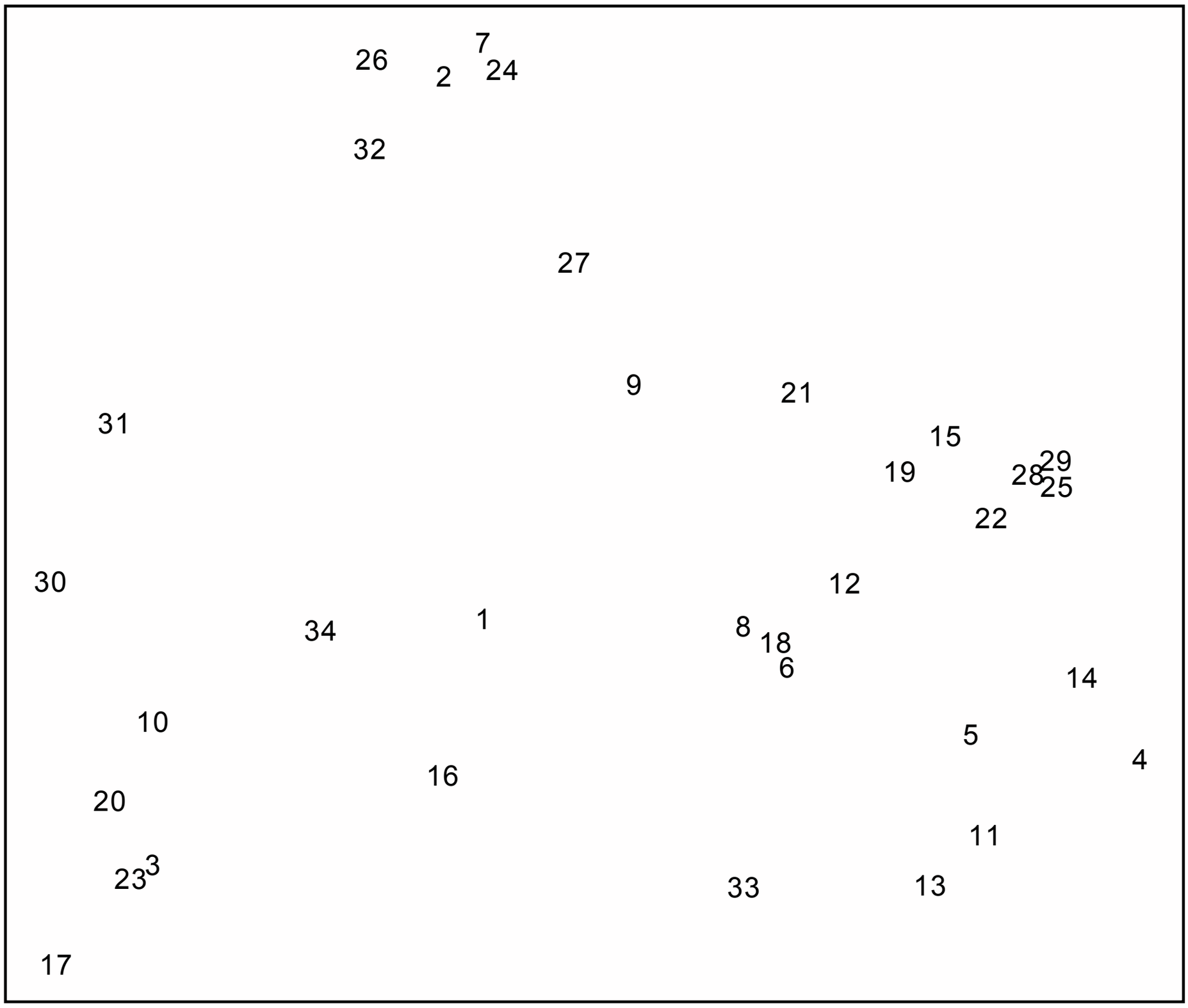

author/council/teaching of the Bible
26
7
24
2
live by the Bible
32
spoken of in the Bible
27
what does the Bible say?
21
the Bible says/tells/refers to urges/teaches/ foretold/reveals
the Bible was copied
9
the Bible was written/ has been printed
to have respect for the Bible
31
30
15
19
29
28
25
22
12
8
18
6
1
contained in the Bible
34
to learn from the Bible
16
14
5
4
the Bible is a gift
10
20
23
3
to study the Bible
17
11
13
33
the Bible is the basis


“Fit” of clustering
0.05
0.10
0.15
0.20
0.25
0.30
5
10
15
20
25
30
Number of clusters

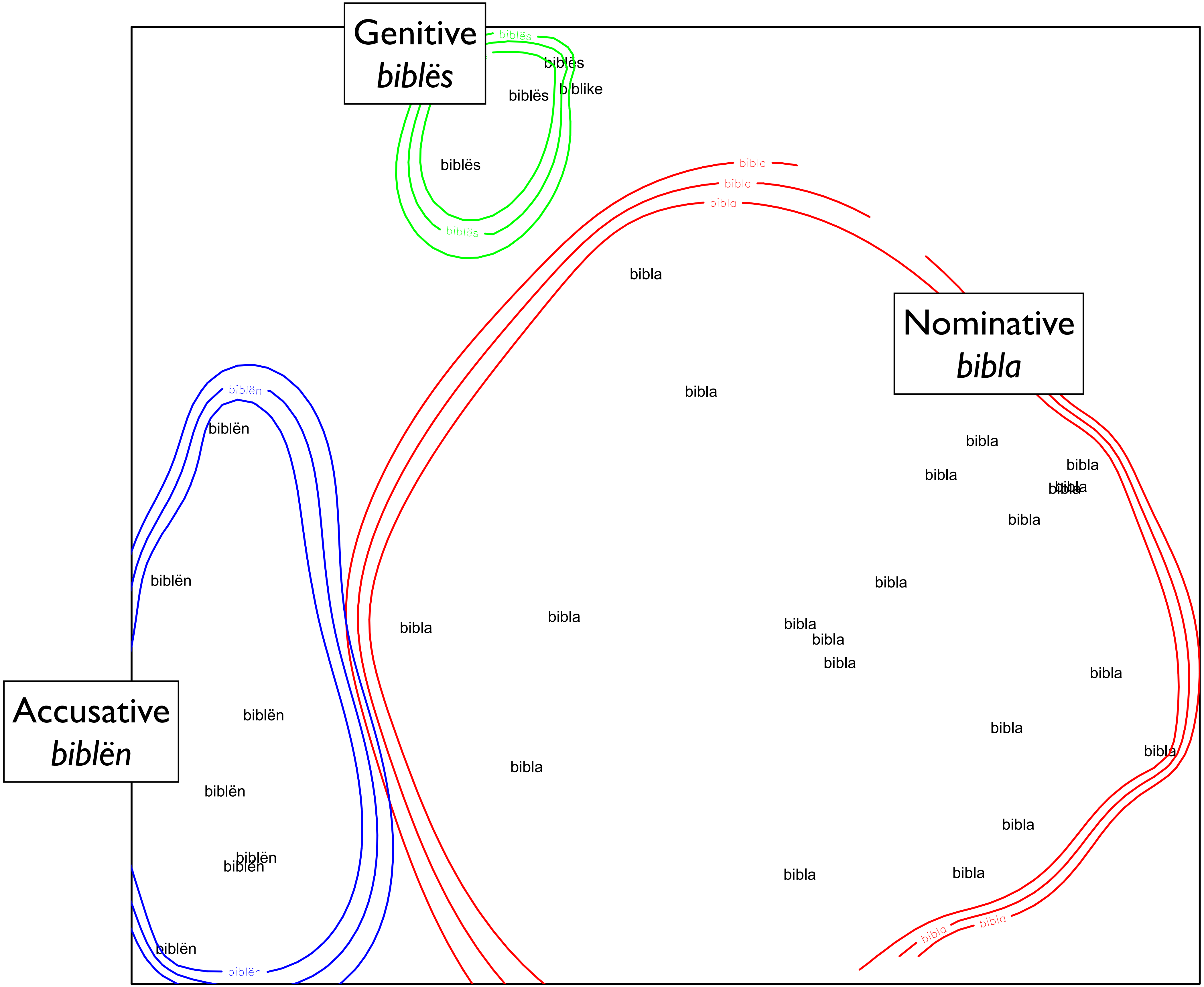

author/council/teaching of the Bible
26
2
7
24
32
live by the Bible
spoken of in the Bible
27
what does the Bible say?
the Bible says/tells/refers to urges/teaches/ foretold/reveals
9
21
15
19
29
28
25
22
12
31
to have respect for the Bible
30
the Bible was copied/ has been printed
1
contained in the Bible
8
18
6
14
5
4
the Bible is a gift
34
to learn from the Bible
16
10
20
23
3
to study the Bible
17
11
13
33
the Bible is the basis


26
7
2
24
32
Source
27
9
21
31
15
19
29
28
25
22
30
12
1
34
8
18
6
14
10
5
4
16
20
Agent
11
23
3
Undergoer
33
13
17

Comparing Languages

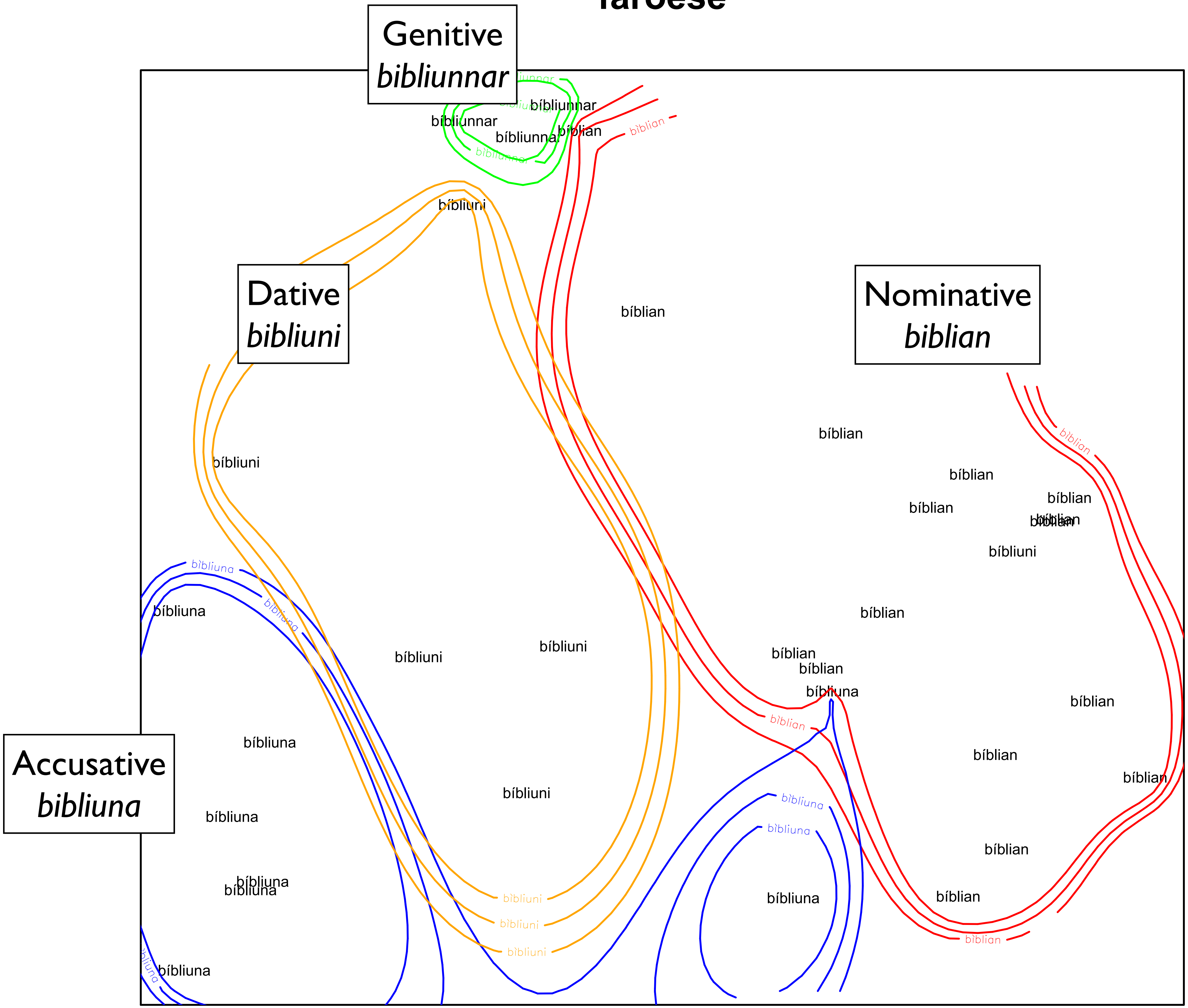

26
7
2
24
32
27
9
21
31
15
19
29
28
25
22
30
12
1
34
8
18
6
14
10
5
4
16
20
11
23
3
33
13
17

albanian

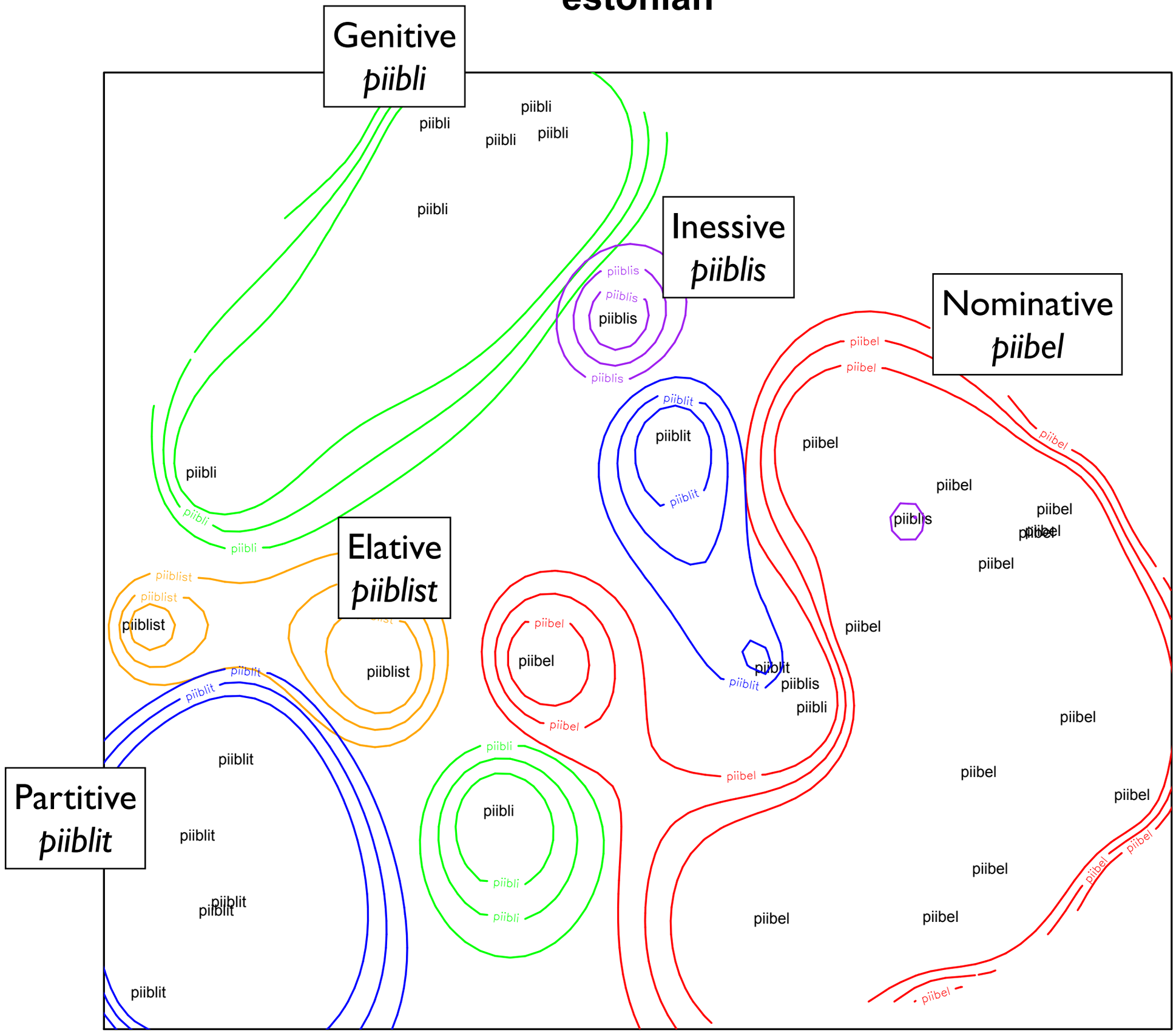


Genitive
biblës
biblës
biblës
biblës
biblike
biblës
biblës
Nominative
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
bibla
Accusative
biblën
biblën
biblën
biblën
biblën
biblën
biblën
biblën
biblën
biblën

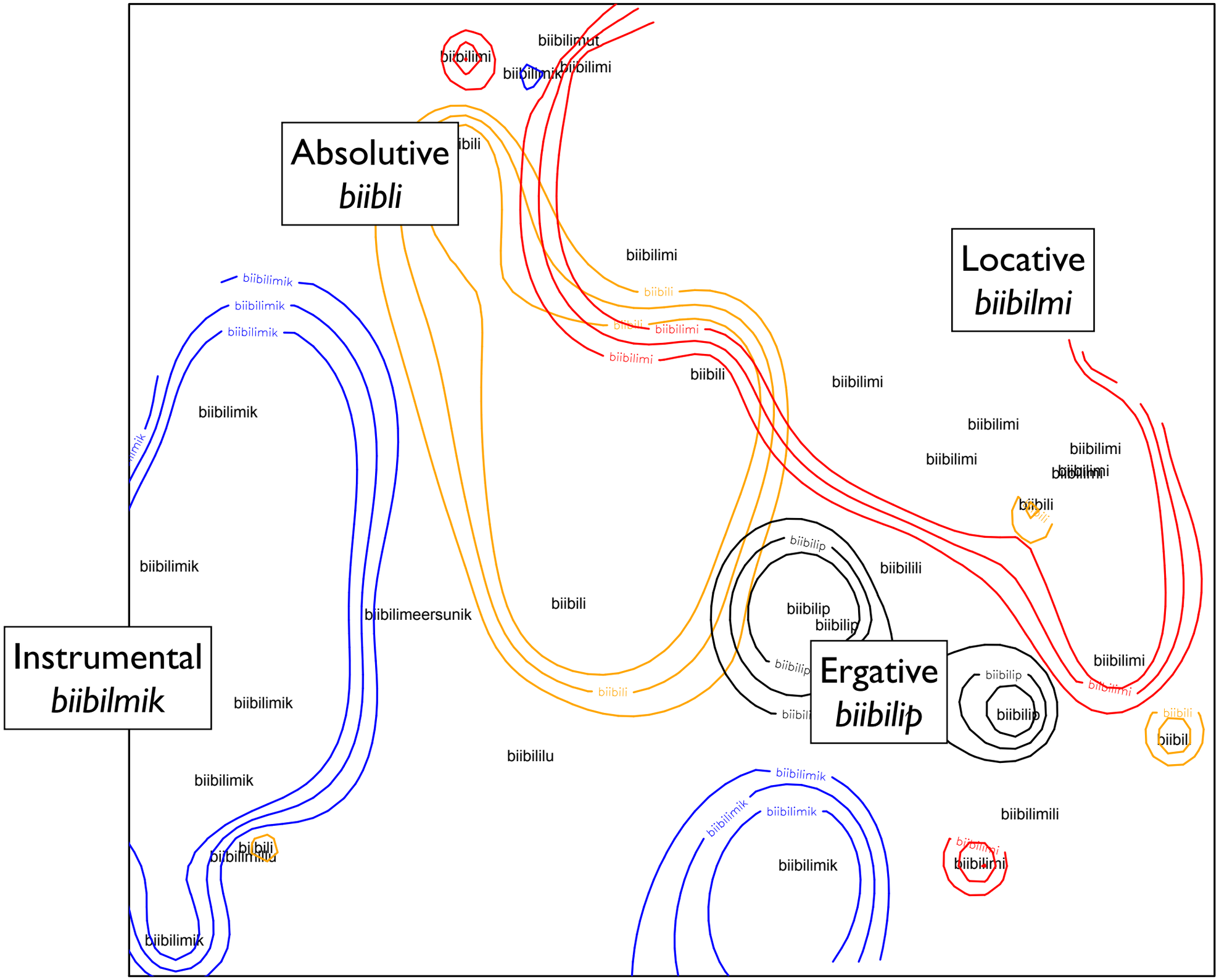
faroese


Genitive
bibliunnar
Dative
bibliuni
Nominative
biblian
Accusative
bibliuna
bíbliunnar
bíbliunnar
bíbliunna
bíblian
bíbliuni
bíblian
bíbliuni
bíblian
bíblian
bíblian
bíblian
bíblian
bíbliuni
bíblian
bíbliuna
bíbliuni
bíbliuni
bíblian
bíblian
bíbliuna
bíblian
bíbliuna
bíblian
bíbliuni
bíbliuna
bíblian
bíbliuna
bíbliuna
bíblian
bíbliuna
bíbliuna

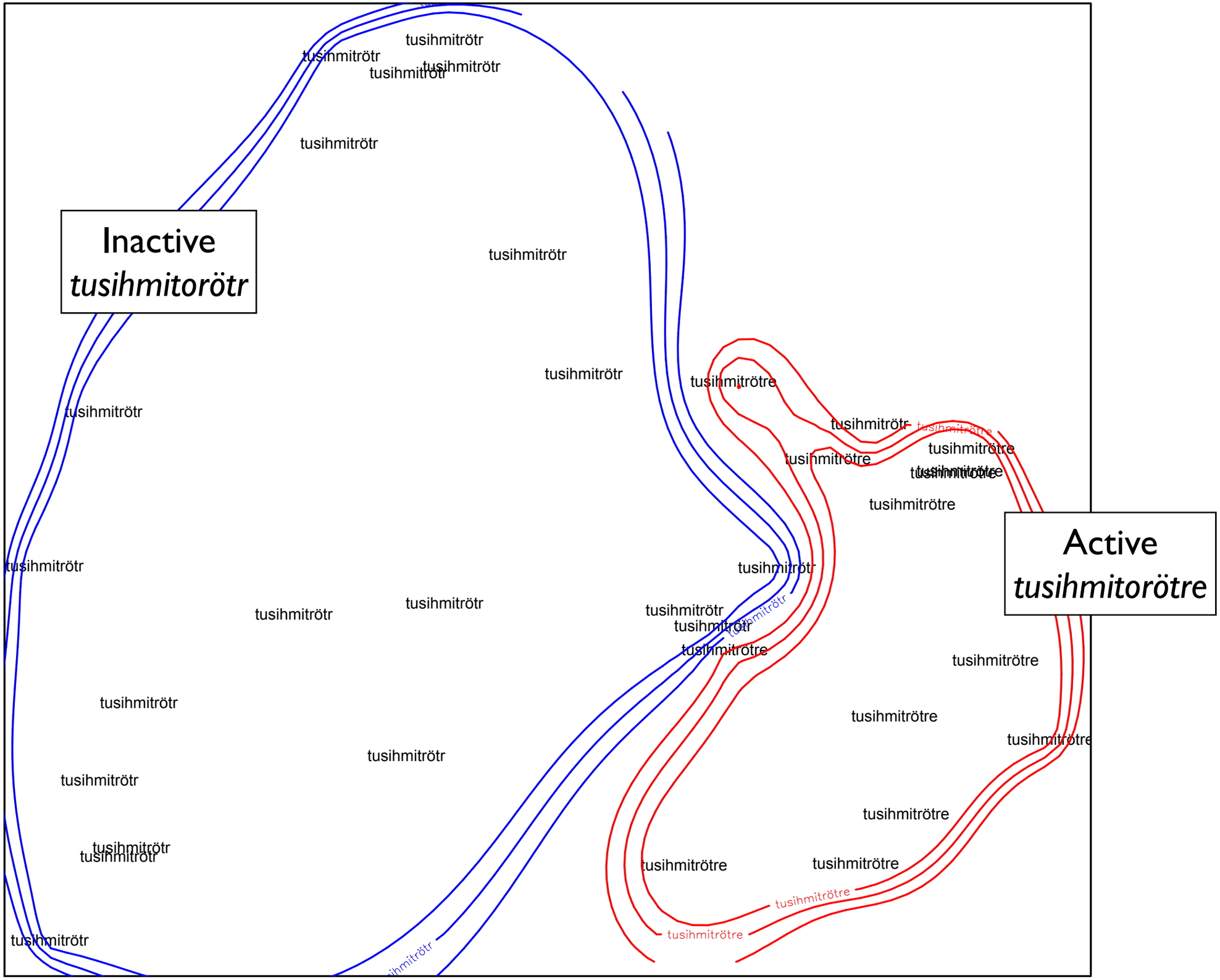
estonian


Genitive
piibli
Inessive
piiblis
Nominative
piibel
Elative
piiblist
Partitive
piiblit
piibli
piiblis
piibel
piiblit
piiblist

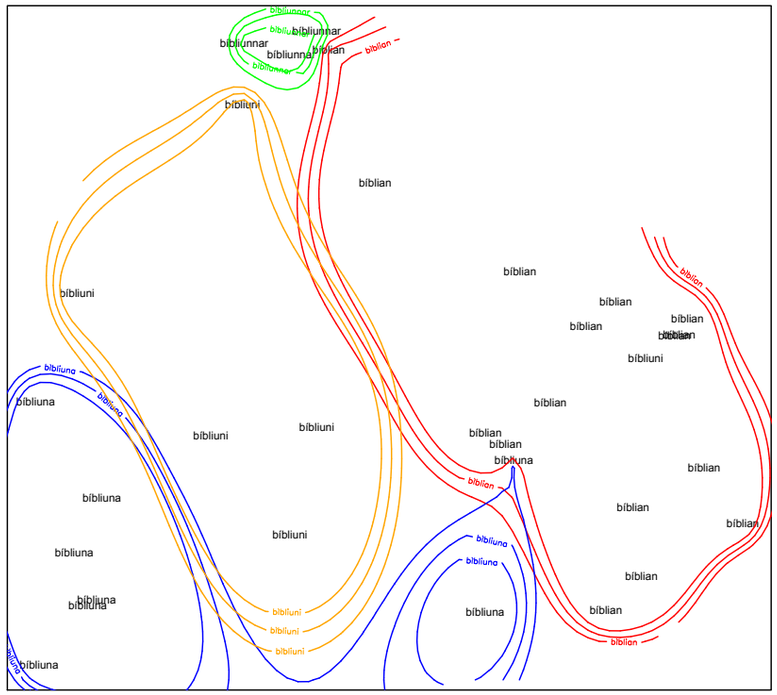
greenlandic



drehu

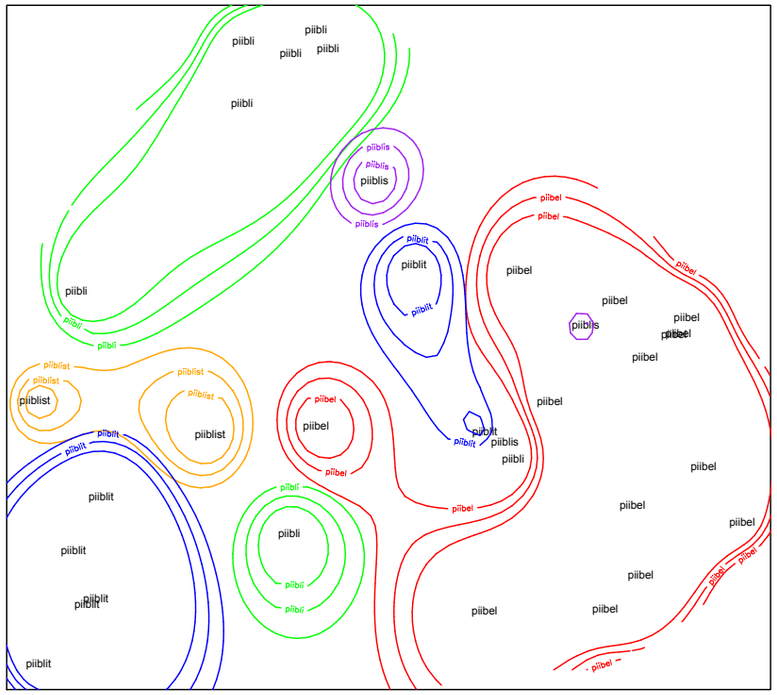

Inactive
tusihmitorötr
Active
tusihmitorötre
tusihmitrötr
tusihmitrötre

faroese

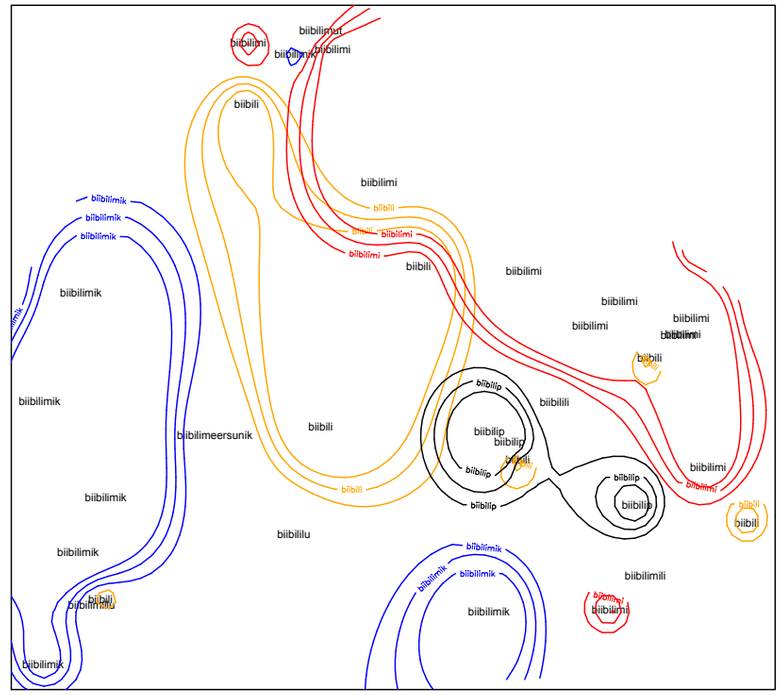


How similar are these languages ?

estonian



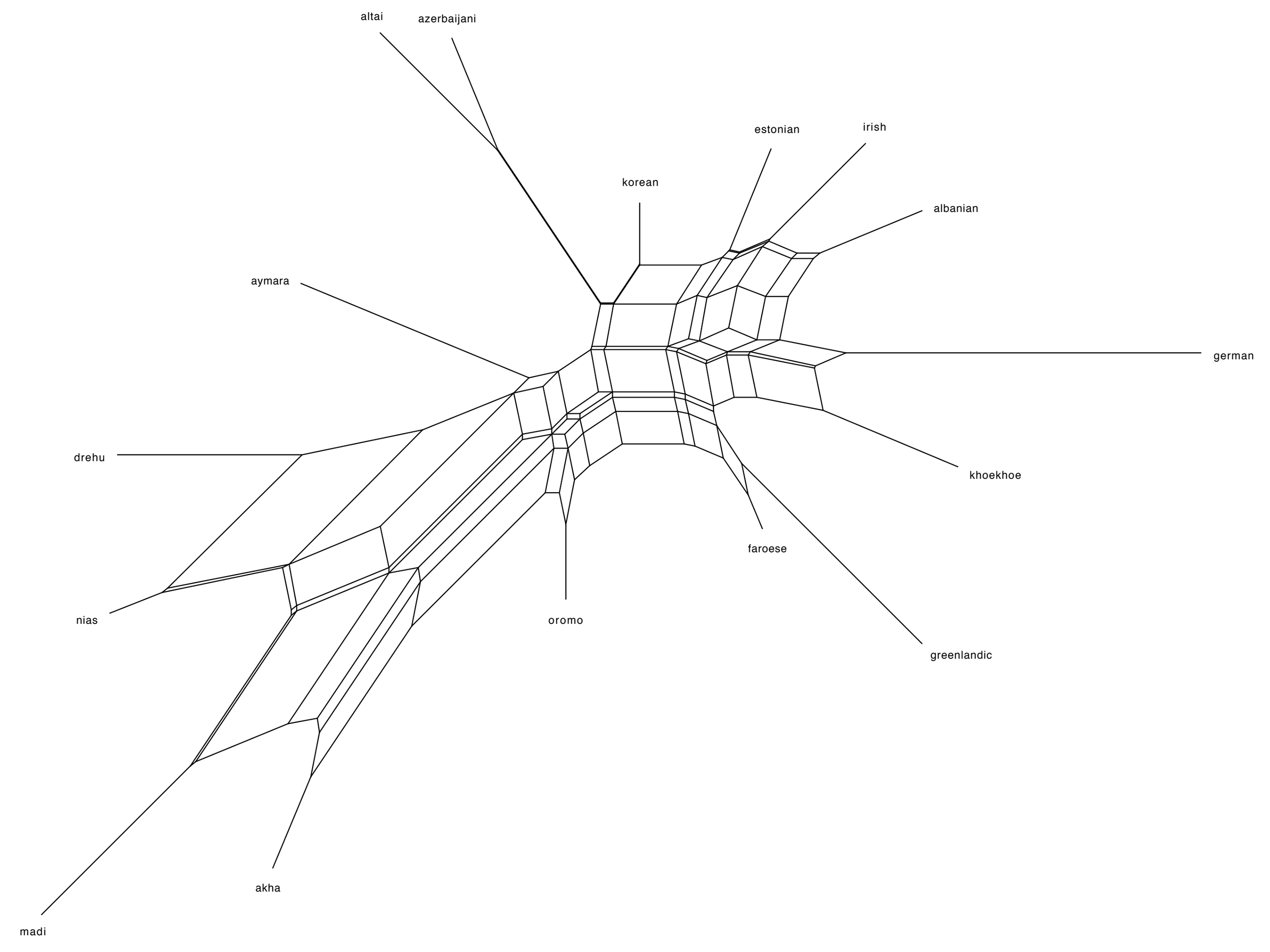
greenlandic

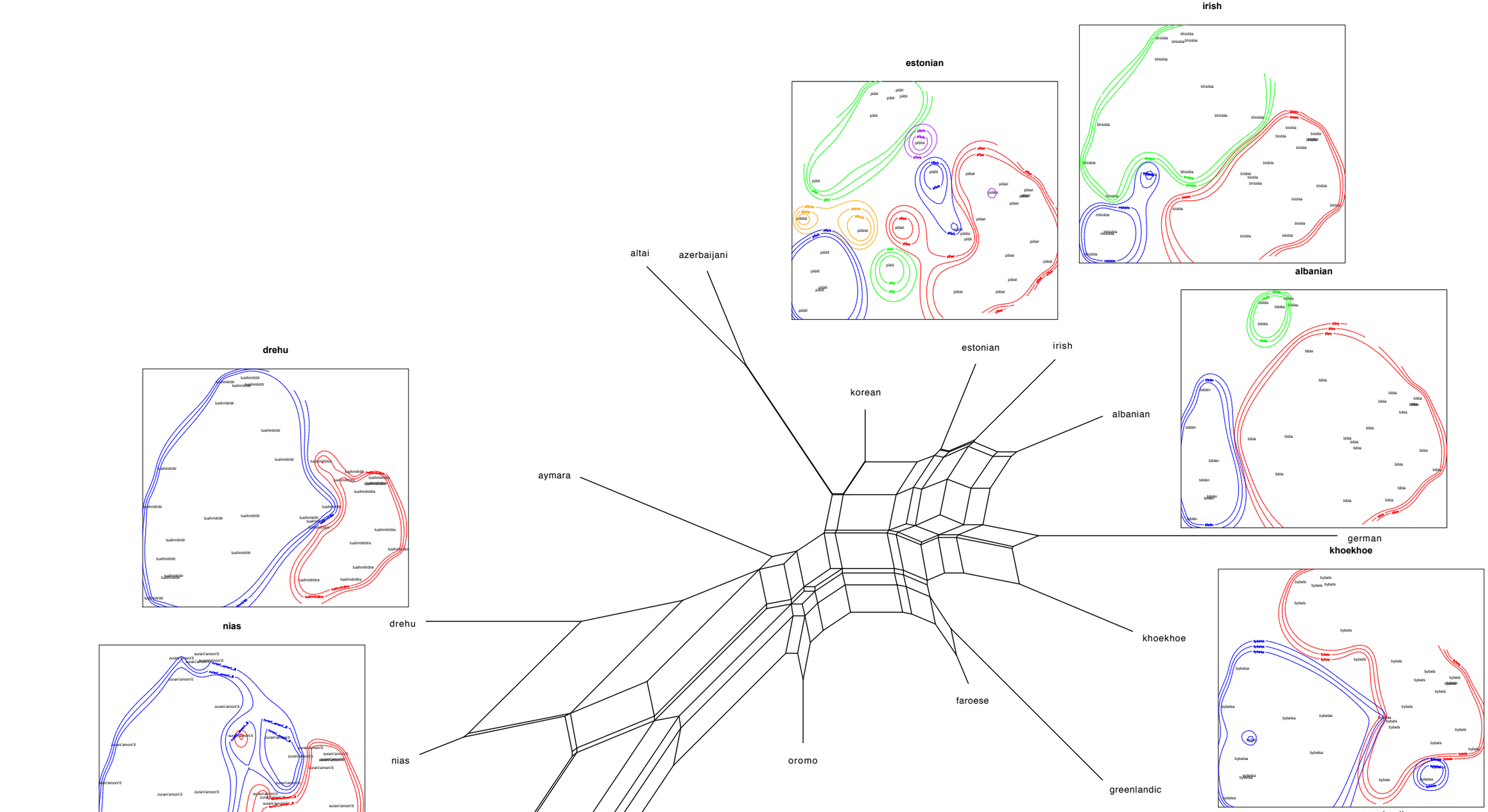


Context	Albanian	Faroese	Estonian	Greenlandic
1	bibla	bíbliuni	piibel	biibili
2	biblës	bíbliunnar	piibli	biibilimik
3	biblën	bíbliuna	piiblit	biibili
4	bibla	bíblian	piibel	biibili
5	bibla	bíblian	piibel	biibilip
		bíbliuna	piibli	
		bíbliunnar	piibli	
		bíblian	piiblit	
		NA	piiblit	
		bíbliuna	piiblit	
		bíblian	piibel	
		bíblian	piibel	
		bíblian	piibel	
		bíblian	piibel	
		bíbliuni	piibli	
		bíbliuna	piiblit	
		bíblian	piiblis	
		bíblian	piiblis	
20	biblën	bíbliuna	piiblit	biibilimik
21	NA	bíblian	piibel	biibilimi
22	bibla	bíbliuni	piibel	biibili
23	biblën	bíbliuna	piiblit	biibilimillu
24	biblike	bíblian	piibli	biibilimi
25	bibla	bíblian	piibel	biibilimi
26	biblës	bíbliunnar	piibli	biibilimi
27	bibla	bíblian	piiblis	biibilimi
28	bibla	bíblian	piibel	biibilimi
29	bibla	bíblian	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

	1	2	3	4	5	…
1	0	1	1	0	0	
2	1	0	1	1	1	
3	1	1	0	1	1	
4	0	1	1	0	0	
5	0	1	1	0	0	
…						

	1	2	3	4	5	…
1	0	1	0	0	1	
2	1	0	1	1	1	
3	0	1	0	0	1	
4	0	1	0	0	1	
5	1	1	1	1	0	
…						


altai
azerbaijani
estonian
irish
korean
albanian
aymara
german
drehu
khoekhoe
faroese
nias
oromo
greenlandic
akha
madi



estonian

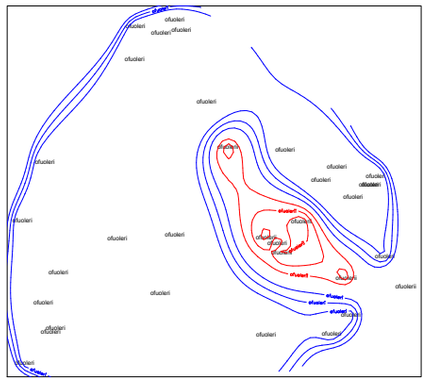
irish

albanian

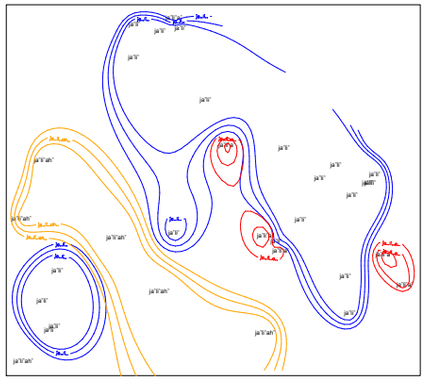
drehu

nias

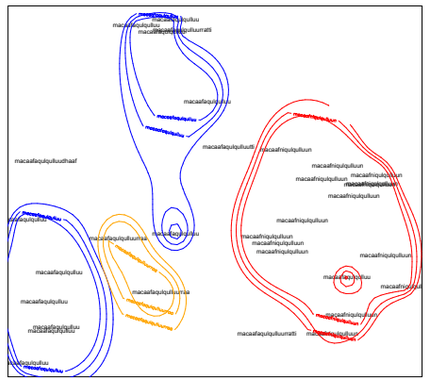
madi



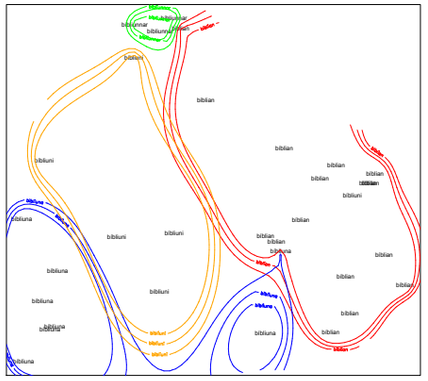
akha



oromo

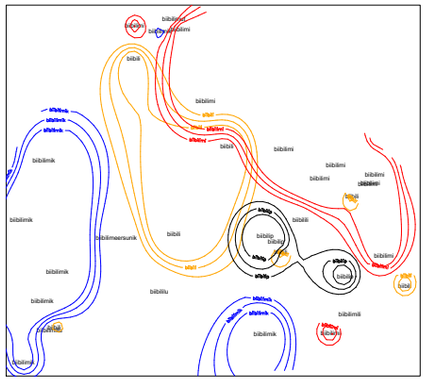


faroese



khoekhoe

greenlandic



How to compare like with like?

- Translations offer a suitable approach to select ***functionally*** parallel data
- Then, form similarity is established ***within each language separately***
- Average ***form similarity across many languages*** is a proxy to general functional similarity
- Distribution of ***form similarity within a language*** gives a language classification

Technicalities

Table III Term-document matrix

Term	*Doc1*	*Doc2*	*Doc3*	*Doc4*	*Doc5*	*Doc6*
Passenger traffic volume	1	1	0	5	2	0
Decrease	1	2	1	0	0	0
Increase	0	2	0	0	0	0
Passengers carried	5	1	0	0	0	0
Personal traffic tools	1	0	0	0	0	0
Grow up	4	1	6	0	0	0
Million	4	1	0	0	0	0
Hundred	0	0	0	0	1	0
FAST rapid transit system	0	2	0	0	0	0
Finished	0	1	0	0	0	0
A1 station	0	0	0	5	4	4
B1 station	0	0	0	1	5	0
C1 station	0	0	0	1	0	0
D1 station	0	0	0	1	0	1
E1 station	0	0	0	1	0	2
Passenger-Kilometers	0	1	7	0	0	0
Columniation	0	0	0	0	2	0
Check the number	0	0	0	0	2	0
Ticket Revenues	0	0	0	0	0	7

random example from the web, taken from Chun-Che Huang, Chia-Ming Kuo, (2003) "The transformation and search of semi-structured knowledge in organizations", Journal of Knowledge Management, Vol. 7 Iss: 4, pp.106 - 123

Word-Parallel Sentence Matrix (WS)

		S1	S2	S3	S4	S5	S6	S7	S8
Language 1	bla	I				I			
	duh	I		I					
	es		I					I	
	...								
Language 2	xfdg			I			I		
	uiso	I			I				
	web		I						
	...								
etc.									

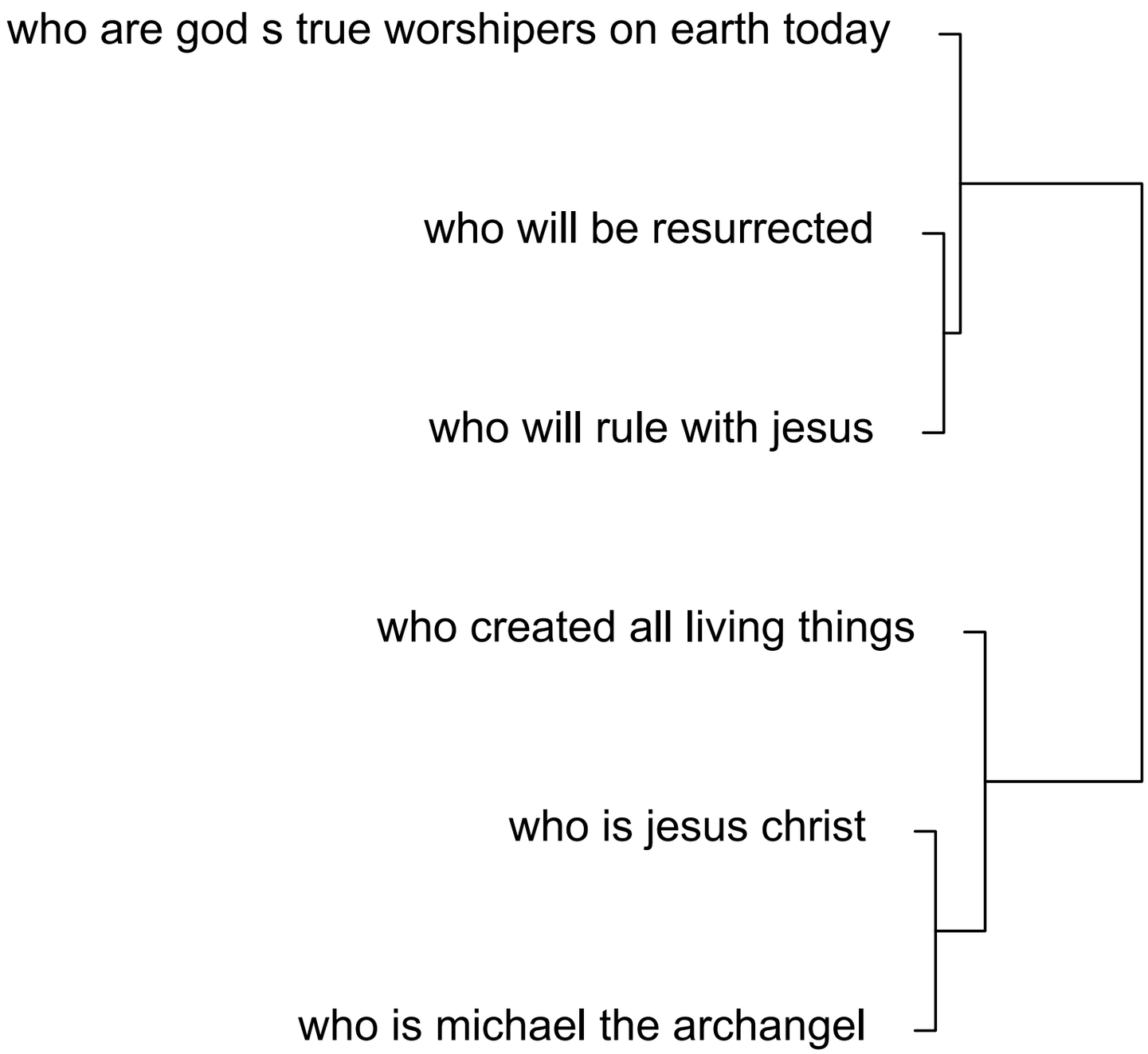
Trans-coocurrences

$$\mathbf{O} = \mathbf{WS} \cdot \mathbf{WS^T}$$

$$\mathbf{E} = \mathbf{WS} \cdot \frac{\mathbf{1_{SS}}}{n} \cdot \mathbf{WS^T}$$

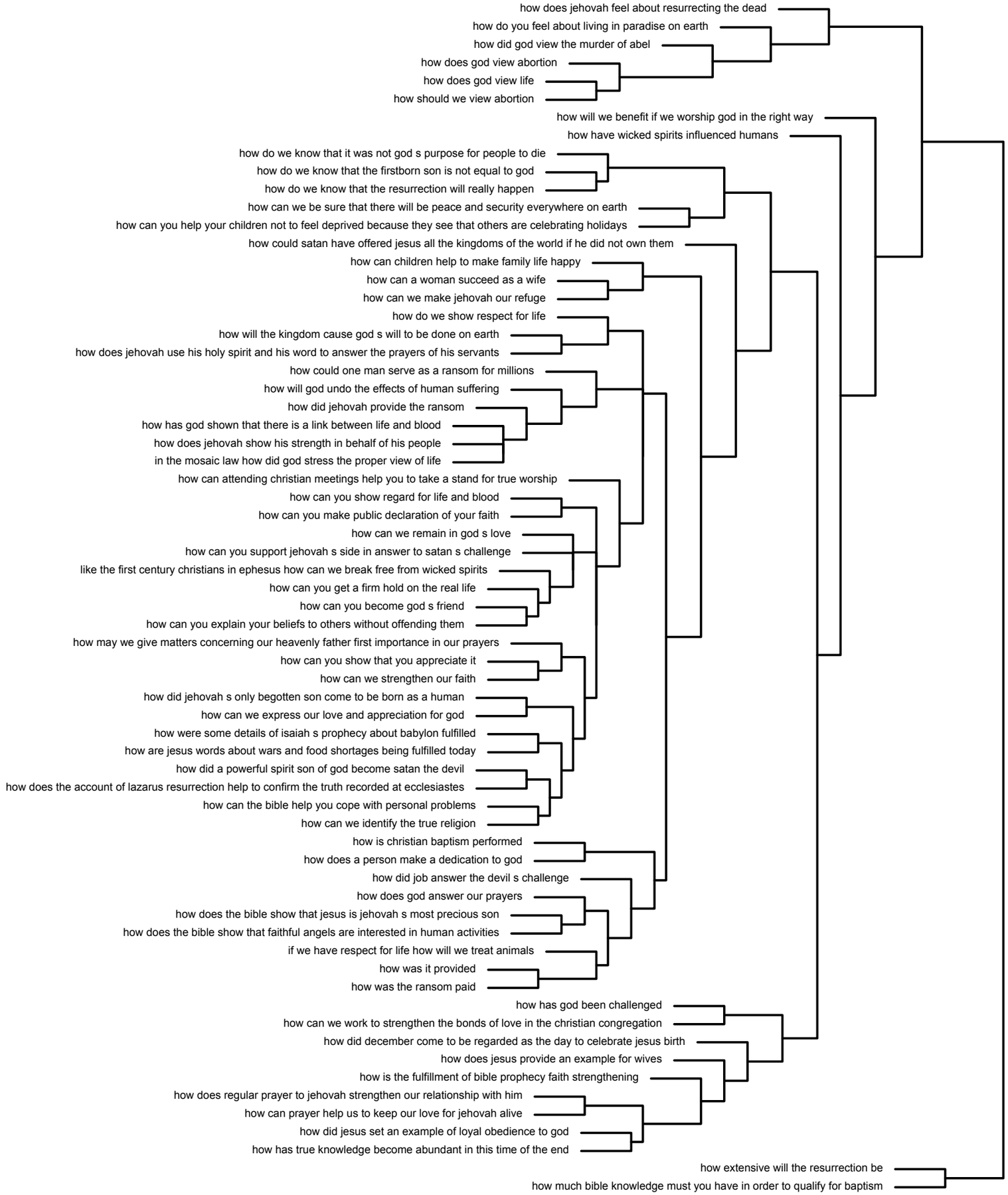
$$\mathbf{WW} = -\log[\frac{\mathbf{E^O}\exp(-\mathbf{E})}{\mathbf{O!}}]$$

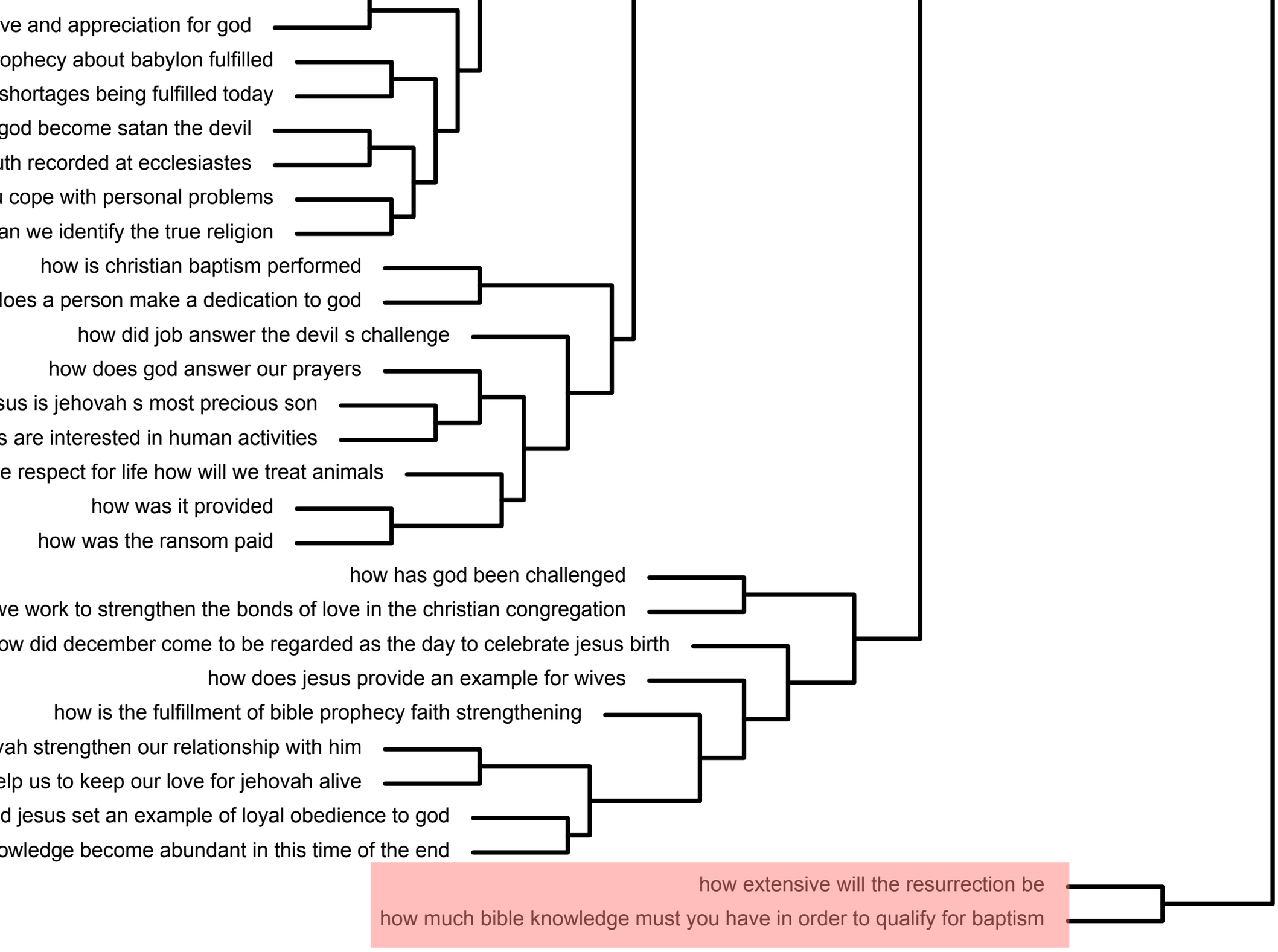
$$= \mathbf{E} + \log \mathbf{O!} - \mathbf{O} \log \mathbf{E}$$



single ‘who’ word

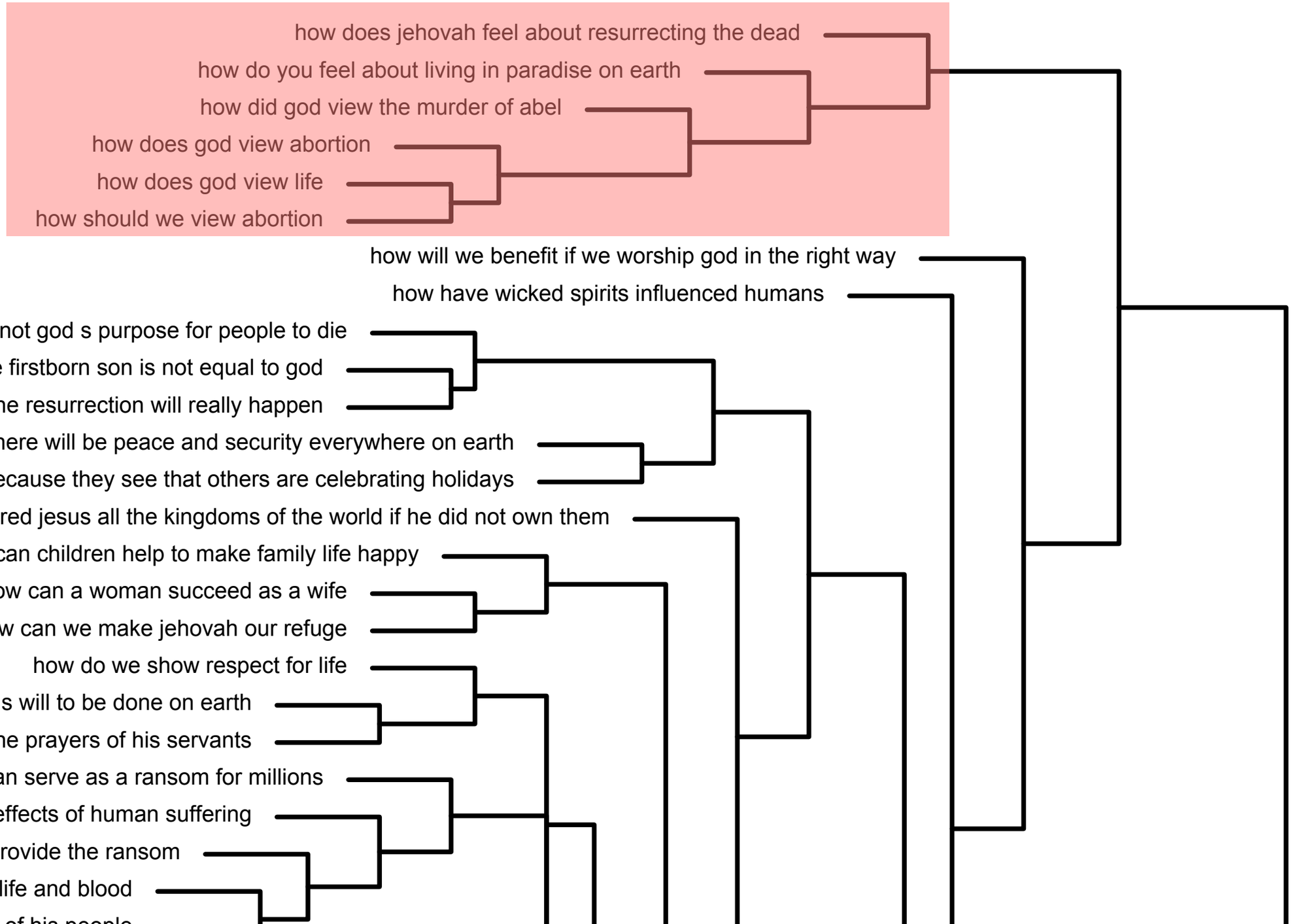
singular vs. plural ‘who’ words


how does jehovah feel about resurrecting the dead
how do you feel about living in paradise on earth
how did god view the murder of abel
how does god view abortion
how does god view life
how should we view abortion
how will we benefit if we worship god in the right way
how have wicked spirits influenced humans
how do we know that it was not god s purpose for people to die
how do we know that the firstborn son is not equal to god
how do we know that the resurrection will really happen
how can we be sure that there will be peace and security everywhere on earth
how can you help your children not to feel deprived because they see that others are celebrating holidays
how could satan have offered jesus all the kingdoms of the world if he did not own them
how can children help to make family life happy
how can a woman succeed as a wife
how can we make jehovah our refuge
how do we show respect for life
how will the kingdom cause god s will to be done on earth
how does jehovah use his holy spirit and his word to answer the prayers of his servants
how could one man serve as a ransom for millions
how will god undo the effects of human suffering
how did jehovah provide the ransom
how has god shown that there is a link between life and blood
how does jehovah show his strength in behalf of his people
in the mosaic law how did god stress the proper view of life
how can attending christian meetings help you to take a stand for true worship
how can you show regard for life and blood
how can you make public declaration of your faith
how can we remain in god s love
how can you support jehovah s side in answer to satan s challenge
like the first century christians in ephesus how can we break free from wicked spirits
how can you get a firm hold on the real life
how can you become god s friend
how can you explain your beliefs to others without offending them
how may we give matters concerning our heavenly father first importance in our prayers
how can you show that you appreciate it
how can we strengthen our faith
how did jehovah s only begotten son come to be born as a human
how can we express our love and appreciation for god
how were some details of isaiah s prophecy about babylon fulfilled
how are jesus words about wars and food shortages being fulfilled today
how did a powerful spirit son of god become satan the devil
how does the account of lazarus resurrection help to confirm the truth recorded at ecclesiastes
how can the bible help you cope with personal problems
how can we identify the true religion
how is christian baptism performed
how does a person make a dedication to god
how did job answer the devil s challenge
how does god answer our prayers
how does the bible show that jesus is jehovah s most precious son
how does the bible show that faithful angels are interested in human activities
if we have respect for life how will we treat animals
how was it provided
how was the ransom paid
how has god been challenged
how can we work to strengthen the bonds of love in the christian congregation
how did december come to be regarded as the day to celebrate jesus birth
how does jesus provide an example for wives
how is the fulfillment of bible prophecy faith strengthening
how does regular prayer to jehovah strengthen our relationship with him
how can prayer help us to keep our love for jehovah alive
how did jesus set an example of loyal obedience to god
how has true knowledge become abundant in this time of the end
how extensive will the resurrection be
how much bible knowledge must you have in order to qualify for baptism



‘to which extent’

‘how do you feel about’, ‘what do you think’




why do we know that food will be abundant in god s new world
why does jehovah choose certain humans to be his friends
why did jesus say that his disciples should pray for god s will to be done in heaven
why can we be sure that the whole earth will become a paradise
why do many people make the mistake of blaming god for the suffering in the world
why do we die
why is the christian ministry urgent today
why can it be said that god uses his angels and his earthly servants to answer our prayers
why should we pray to god
why can you trust the prophecies recorded in the bible
why can we say that satan s challenge involving job applies to all humans
why can we be sure that the resurrection of lazarus really happened
why does the teaching of eternal torment dishonor jehovah
why is it comforting to know the truth about death
why can we be sure that jehovah has a good reason for allowing suffering to continue
why can we say that through jesus we come to know jehovah better
why will the resurrection bring joy to our hearts
in times past why did some people not celebrate christmas
why should we say amen at the end of personal and public prayers
why did jehovah not just destroy the rebels in eden
why is it a loving arrangement for jesus and the to rule over mankind
why is it important for you to know the truth about jesus
why will the destruction of the wicked catch most people by surprise
why does god allow suffering
why should we keep in mind the need to be safety conscious
why is repentance necessary
why do you not need to fear making a dedication to god
why should we worship jehovah publicly in association with true christians
why is it dangerous to try to communicate with the dead
why may we rightly view the kingdom preaching work as a privilege and a treasure
why is it essential to remain in god s love
why is the harmony of the material in the bible particularly noteworthy
why is there a need for a new world
why is it often good to ask questions
why can we trust that god will undo all the effects of wickedness
why do many stop trying to find answers to the questions that matter most
why do we die
why might some hold back from making a dedication to god
why should you think seriously about baptism
why should we take full advantage of the precious privilege of prayer
why should we want to learn about angels
why did an ethiopian court official request baptism
why is it not surprising that the bible is a practical book
why can dedication to jehovah not remain a private matter
why should we be grateful to jehovah for the provision of the ransom
why is it not wrong for a wife to express her opinions respectfully
why should we not hate our fellowman
why is it important to accept jesus christ as god s means of salvation

from Word-Sentence to Word-Alignment

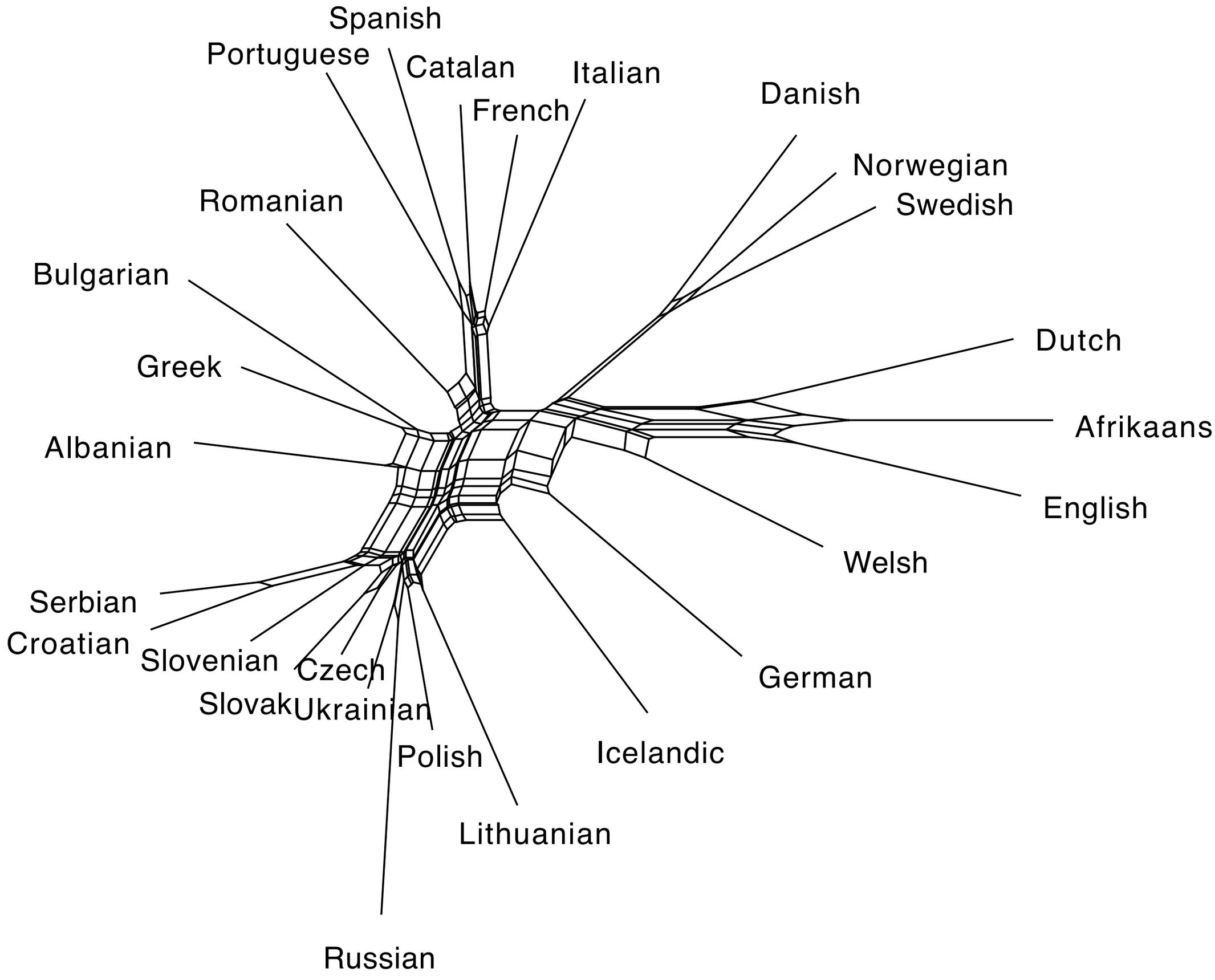
		S1	S2	S3	S4	S5	S6	S7	S8
Language 1	bla	I				I			
	duh	I		I					
	es		I					I	
	...								
Language 2	xfdg			I			I		
	uiso	I			I				
	web		I						
	...								
etc.									

from Word-Sentence to Word-Alignment

		Sentence 1				Sentence 2			
		A1	A2	A3	...	A1	A2	A3	...
Language 1	bla	1							
	duh		1						
	es			1			1		
	...								
Language 2	xfdg								
	uiso	1							
	web		1				1		
	...								
etc.									

Sparse alignment

- Structure of languages are really different
- So: do not force an alignment across all languages
- Only align those words that really form a coherent group, ‘sparse alignment’
  - we used affinity propagation clustering in each sentence
- The more words are aligned between two languages, the more similar are these languages


Spanish
Portuguese
Catalan
Italian
French
Danish
Norwegian
Swedish
Romanian
Bulgarian
Greek
Albanian
Dutch
Afrikaans
English
Welsh
German
Icelandic
Lithuanian
Polish
Russian
Ukrainian
Slovak
Czech
Slovenian
Croatian
Serbian

Prospects

- We are organizing large massively parallel texts for automatic processing
- Main problem: copyright
- Linking to monolingual corpora in cooperation with http://wortschatz.uni-leipzig.de
- Adding information will be open
  - morpheme separation, multi-word recognition, etc.
- **Typology will never be the same again!**