Does beer make Enlil happy?

Semantic explorations in the corpus of Sumerian literary texts from the Old Babylonian Period

(ca. 2000-1600 BCE)

Kamran Vincent Zand – Pseudo-Saffron (Hamburg)

Description of the corpus

- Sumer/Babylonia is located in modern day southern Iraq
- Sumerian literary texts of the corpus were written down between 2000-1600 BCE by Babylonian scribes in cuneiform
- Sumerian is in that period a dead language and is only used for cultic/academic purposes (like Latin in Europe)
- For the purpose of this analysis English translations will be used

Description of the corpus

1. ud šu bal ak-de$_3$ ĝiš-ḫur ḫa-lam-e-de$_3$ (*Cited*
2. ud-de$_3$ mar-uru$_5$-gin$_7$ teš$_2$-bi i$_3$-gu$_7$-e
3. me ki-en-gi-ra šu bal ak-de$_3$
4. bal sag$_9$-ga e$_2$-ba gi$_4$-gi$_4$-de$_3$
5. uru$_2$ gul-gul-lu-de$_3$ e$_2$ gul-gul-lu-de$_3$
6. tur$_3$ gul-gul-lu-de$_3$ amaš tab-tab-be$_2$-de$_3$

1-2. To overturn the appointed times, to obliterate the divine plans, the storms gather to strike like a flood.

3-11. An, Enlil, Enki and {Ninḫursaĝa} {(*2 mss. have instead:*) Ninmaḫ} have decided its fate -- to overturn the divine powers of Sumer, to lock up the favourable reign in its home, to destroy the city, to destroy the house, to destroy the cattle-pen, to level the sheepfold; that the cattle should not stand in the pen, that the sheep should not multiply in the fold, that watercourses should carry brackish water, that weeds should grow in the fertile fields, that mourning plants should grow in the open country,

Description of the corpus

- The Electronic Text Corpus of Sumerian Literature (ETCSL) Oxford University 2003-2006 (https://etcsl.orinst.ox.ac.uk/)
- 377 Compositions/ca. 11600 lines/ca.272000 words

ETCSLcorpus

Catalogue of all available compositions and translations by number

- 0 = ancient literary catalogues (Unicode | Ascii)
- 1 = narrative and mythological compositions (Unicode | Ascii)
- 2 = royal praise poetry and compositions with a historical background (Unicode | Ascii)
- 3 = literary letters and letter-prayers (Unicode | Ascii)
- 4 = hymns and cult songs (Unicode | Ascii)
- 5 = other literature (Unicode | Ascii)
- 6 = proverbs (Unicode | Ascii)

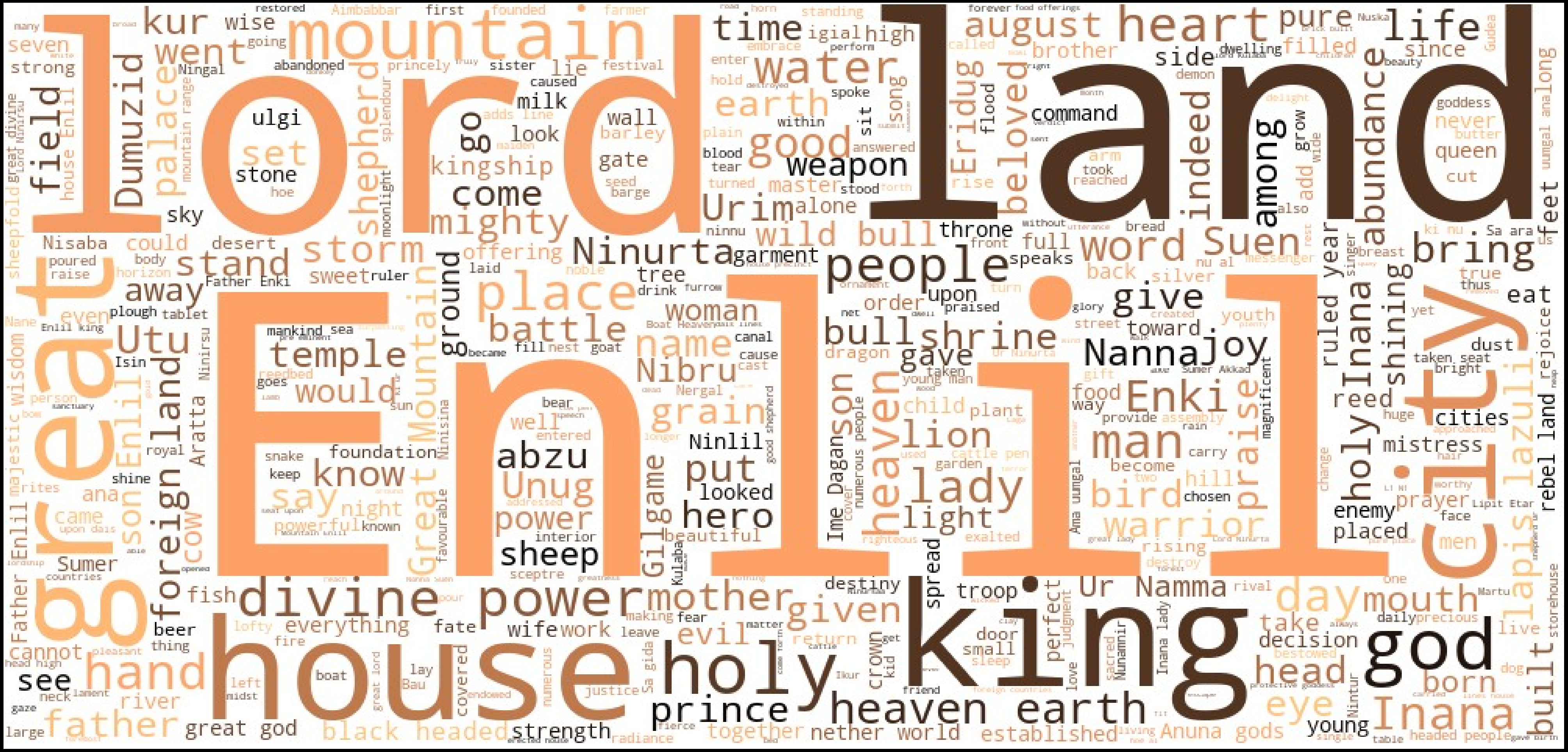
- All compositions (Unicode | Ascii)

Data Pre-processing

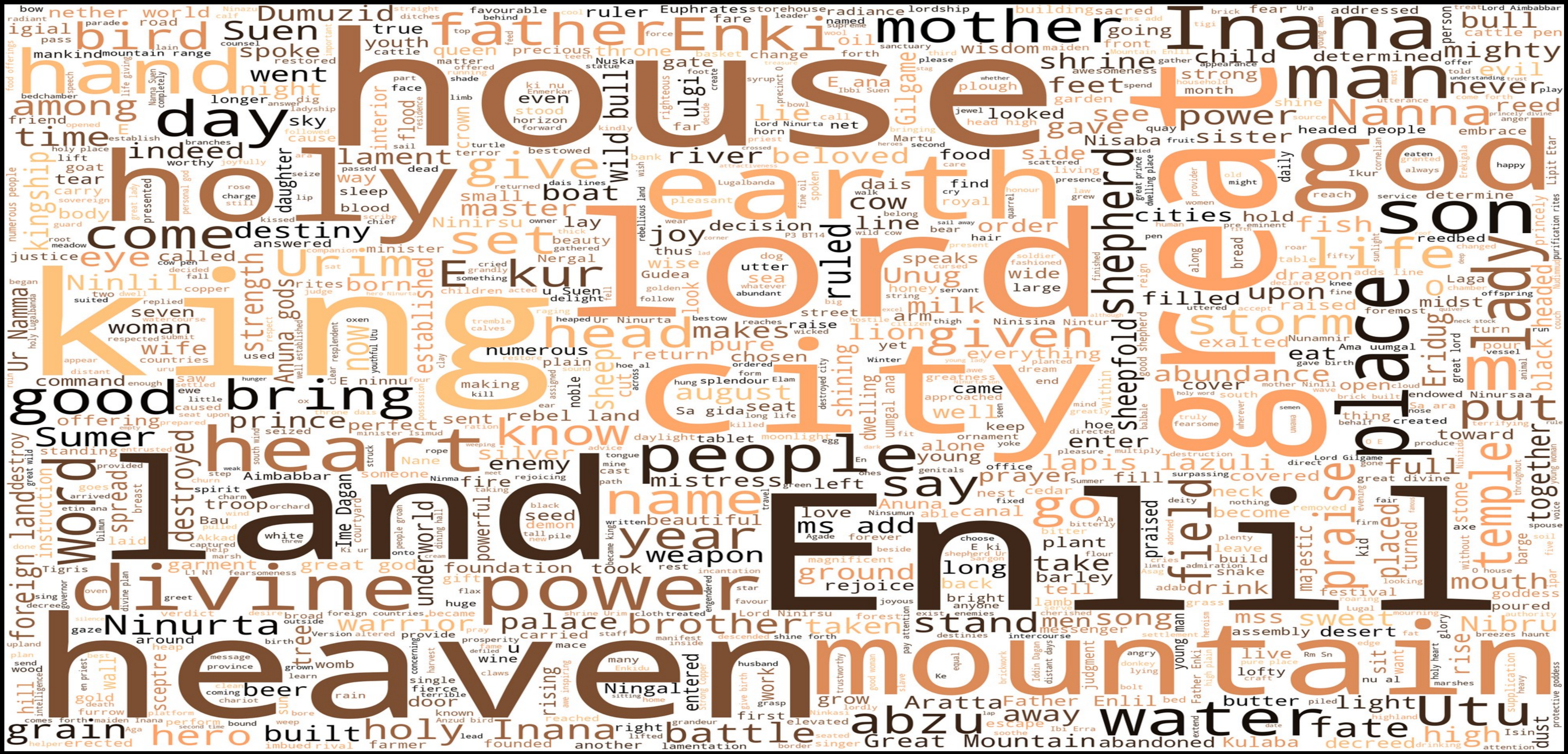
Standard NLP-Workflow:

- Tokenization: Compositions (docs) are split into sentences and words. Lowercase and punctuation-removal
- Stop-word removal
- Lemmatization: removing inflectional endings and returning the base or dictionary form of a word
- Creation of a word cloud: general overview and identification of new stop words

Word cloud (500 most common)

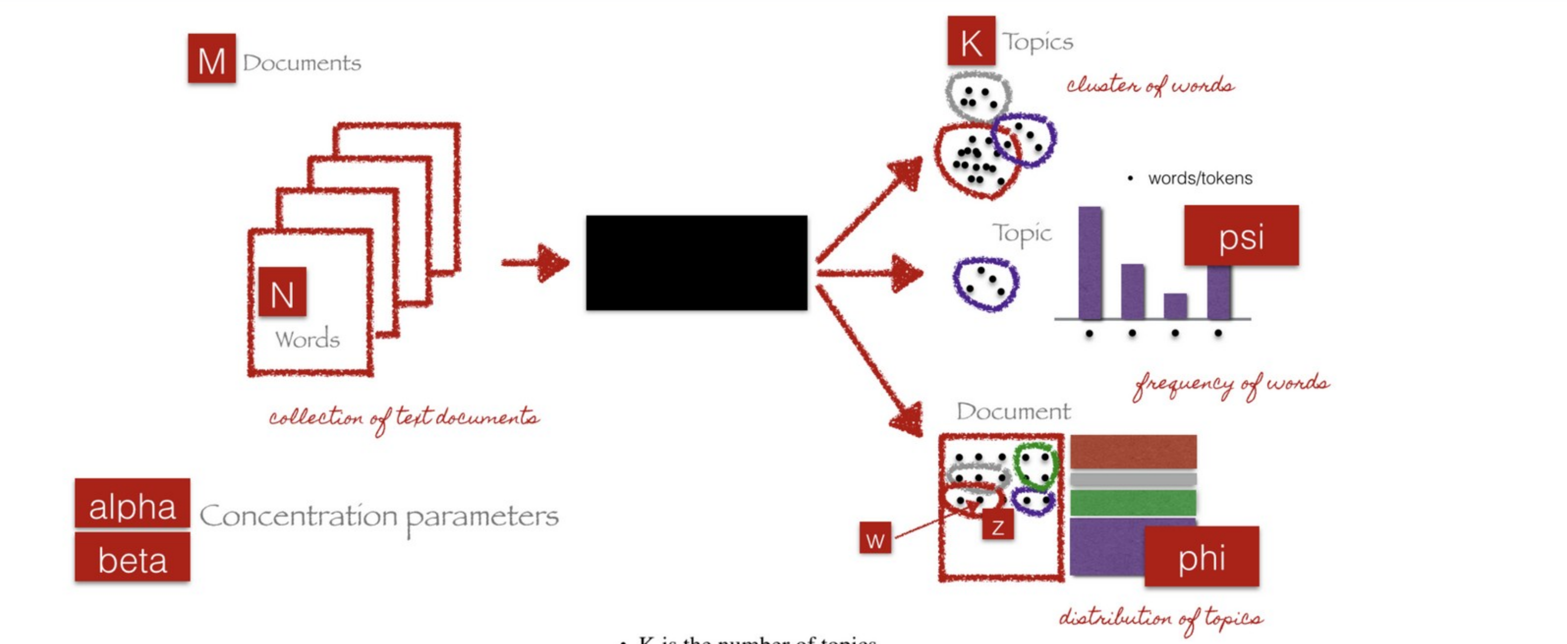


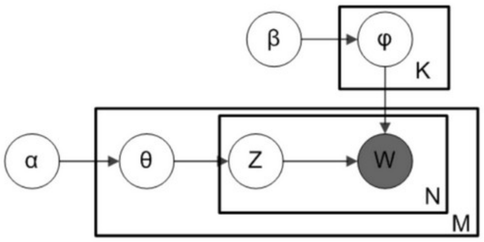
Word cloud (1000 most common)



Choosing topics by LDA

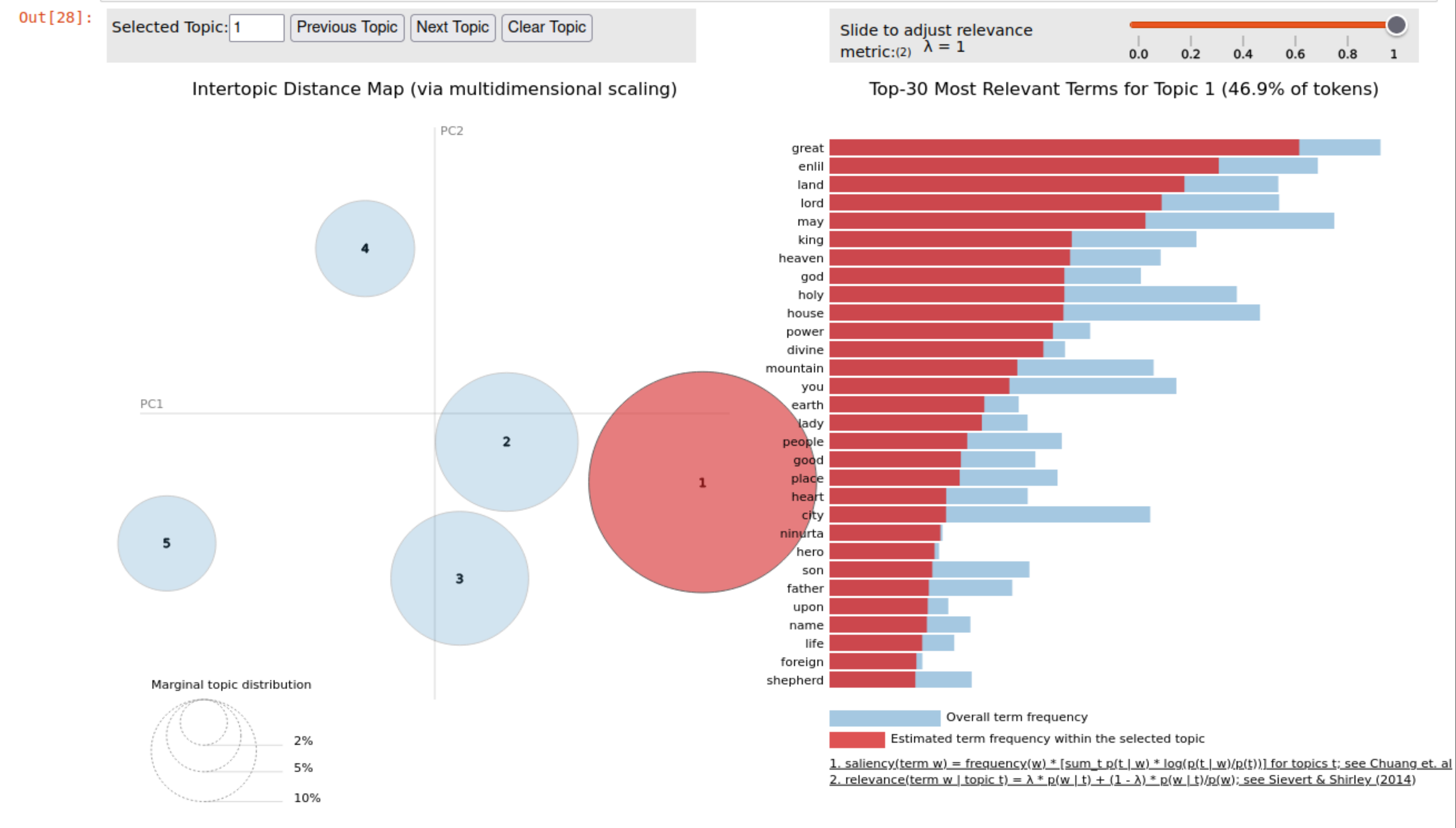
Latent Dirichlet Allocation:



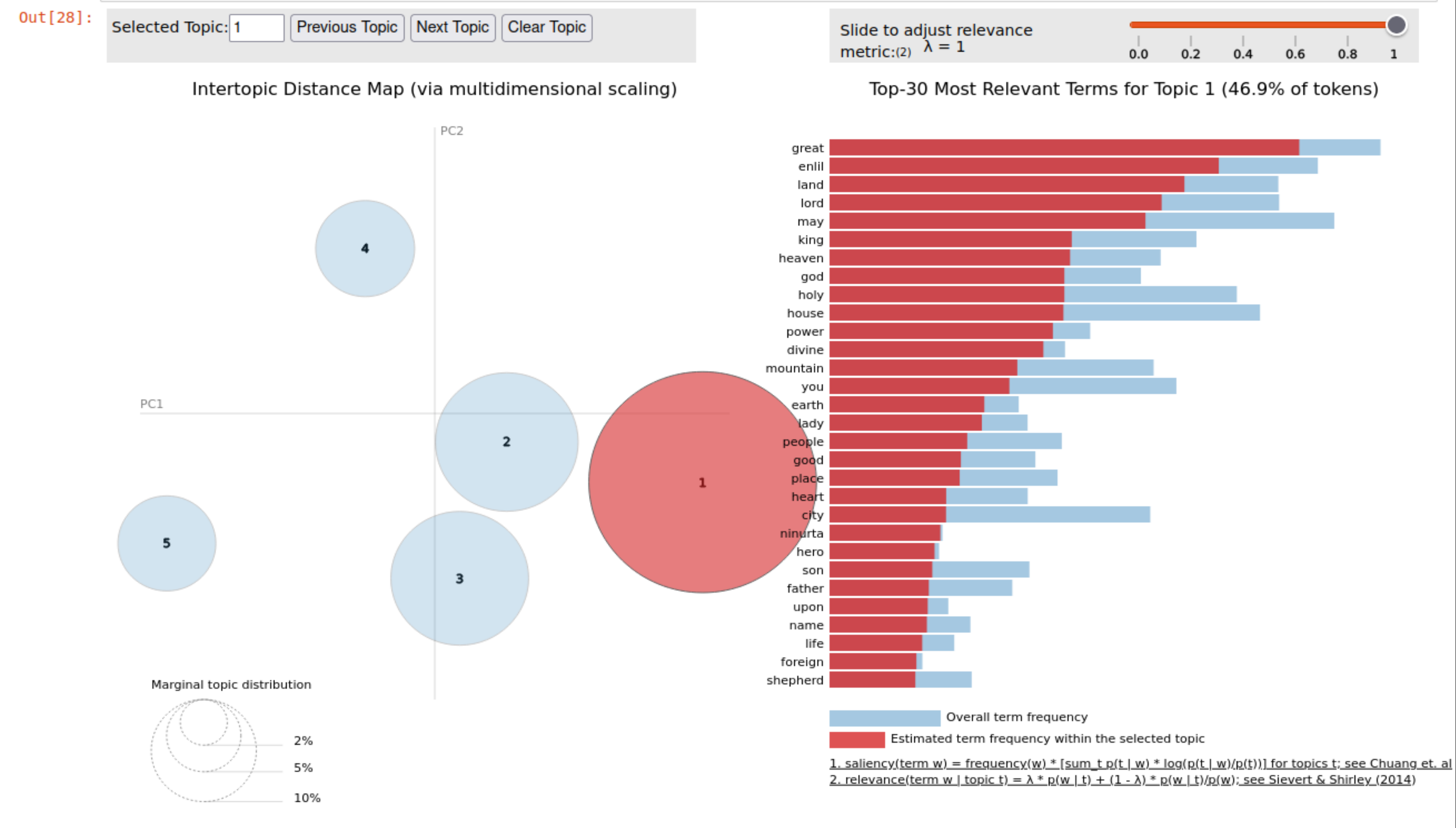


- K is the number of topics
- N is the number of words in the document
- M is the number of documents to analyse
- α is the Dirichlet-prior concentration parameter of the per-document topic distribution
- β is the same parameter of the per-topic word distribution
- $\phi(k)$ is the word distribution for topic k
- $\theta(i)$ is the topic distribution for document i
- $z(i,j)$ is the topic assignment for $w(i,j)$
- $w(i,j)$ is the j-th word in the i-th document
- ϕ and θ are Dirichlet distributions, z and w are multinomials.

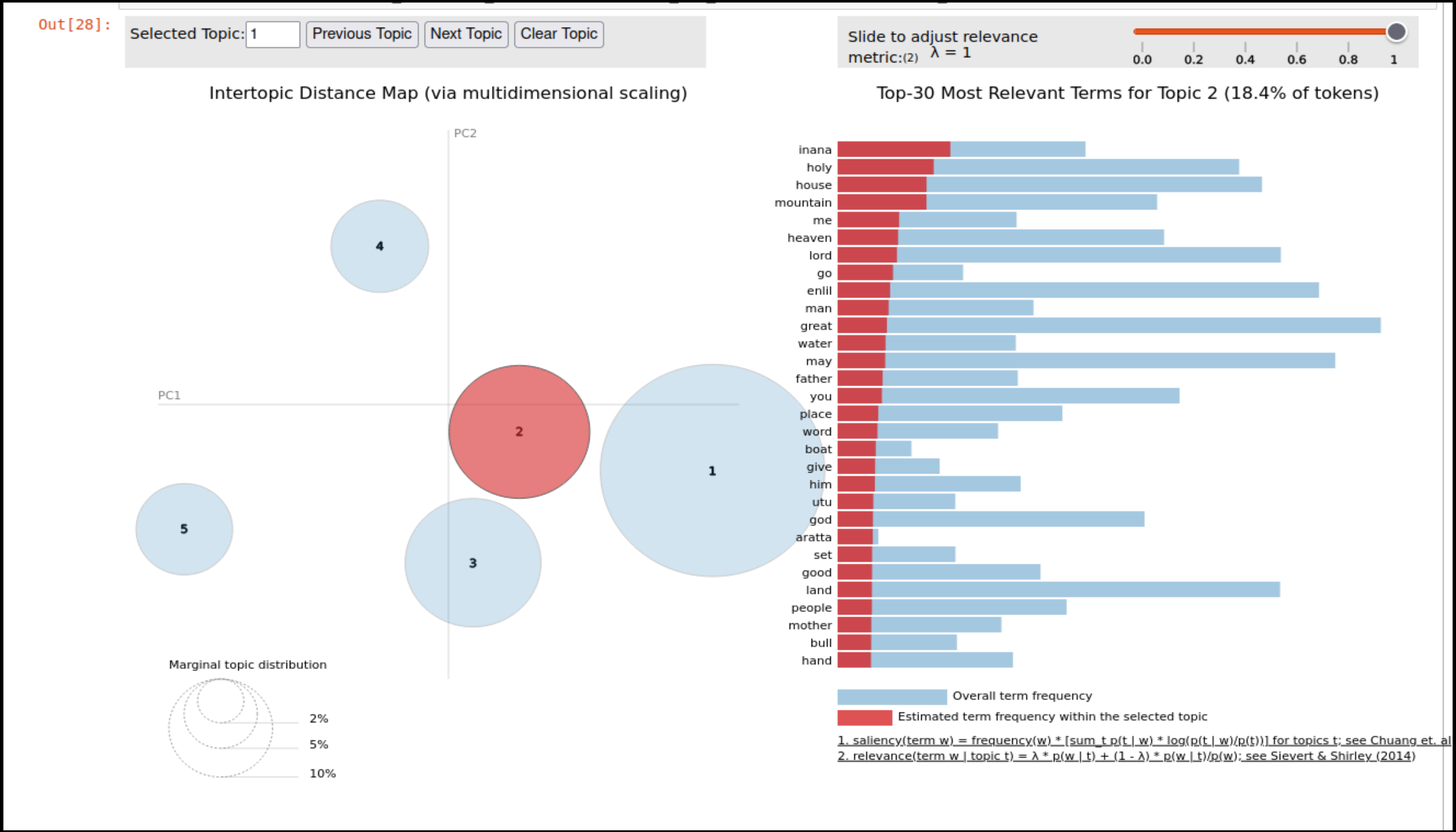
Topics chosen by LDA



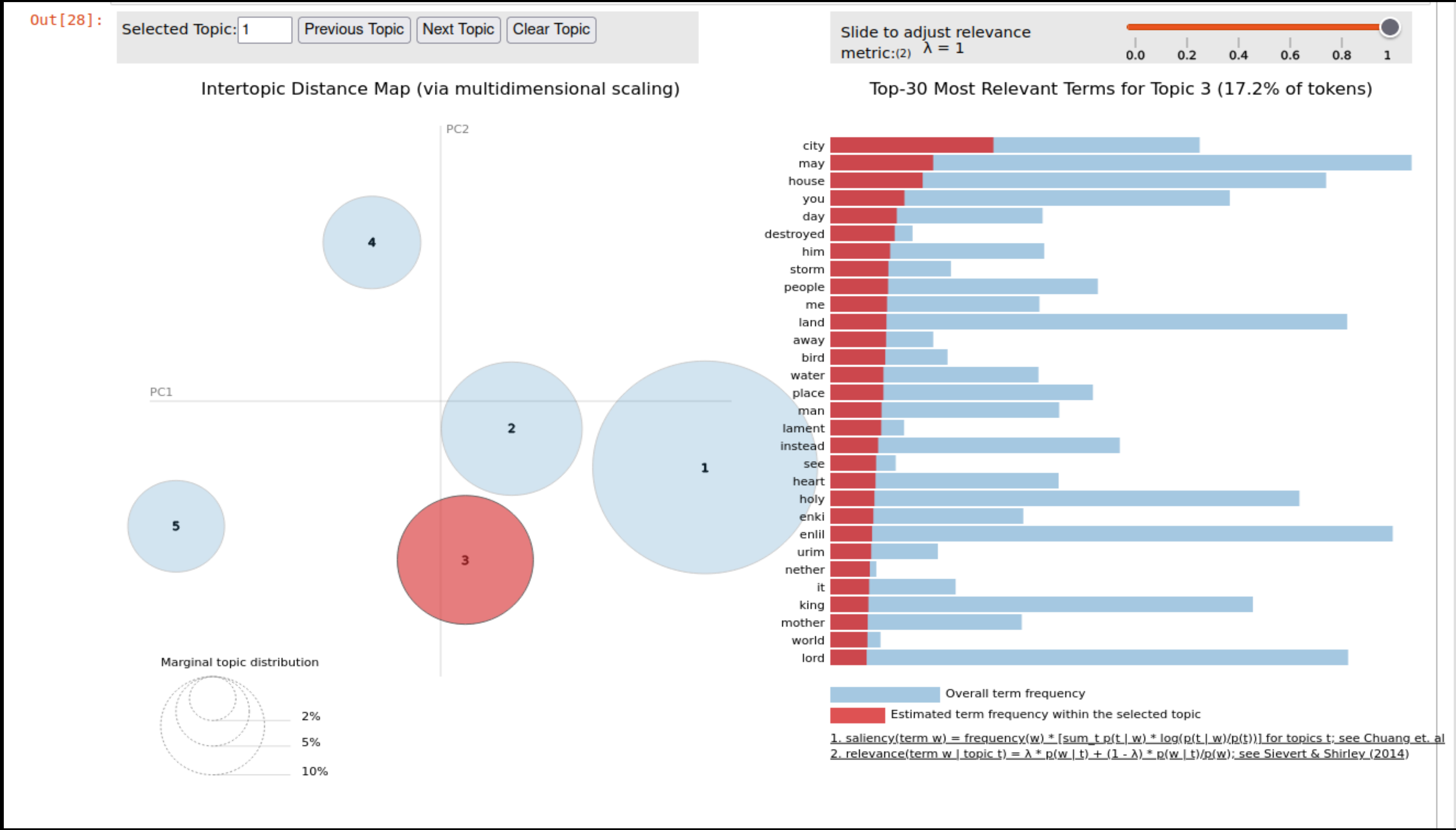
Topics chosen by LDA



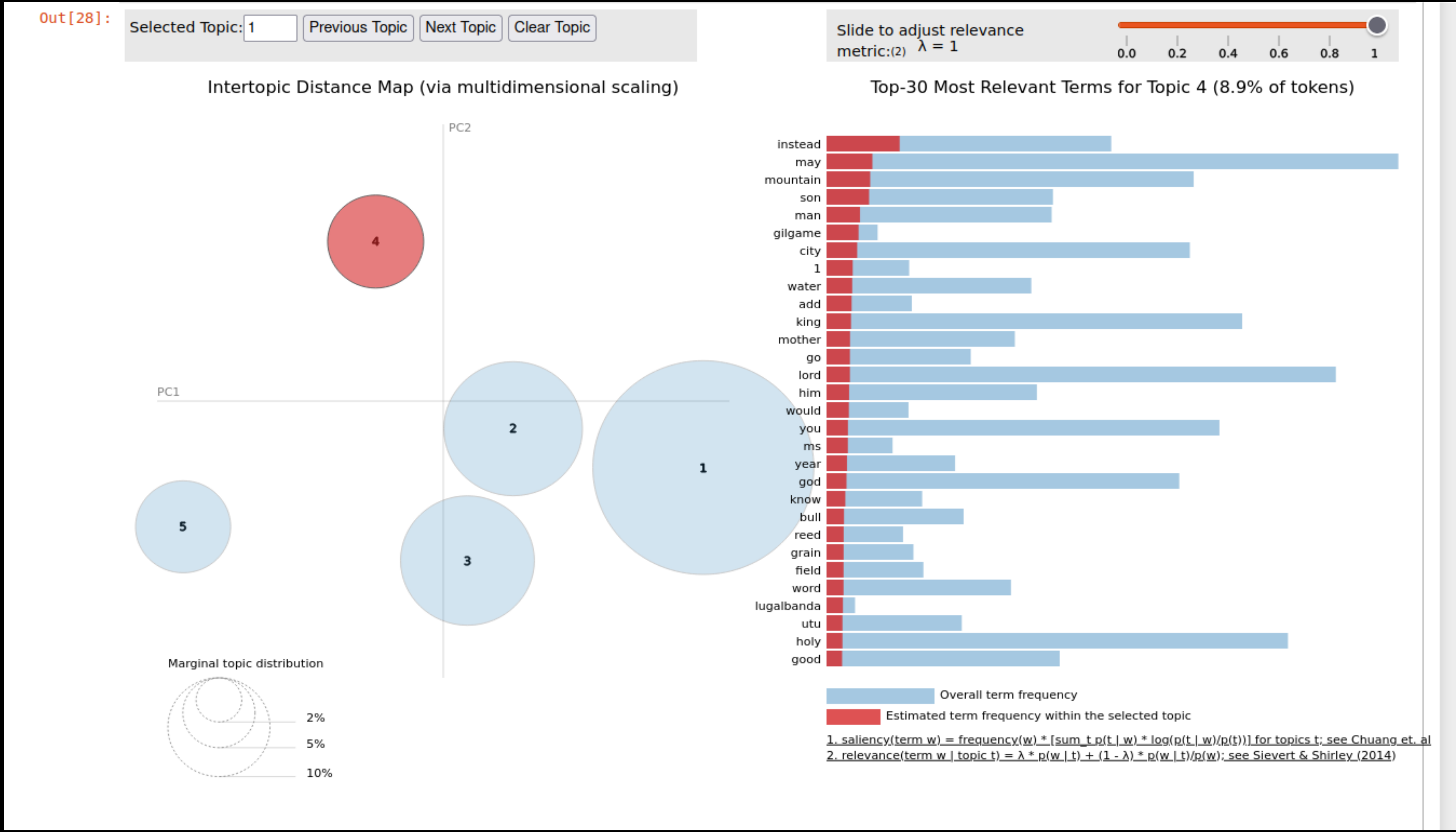
Topics chosen by LDA



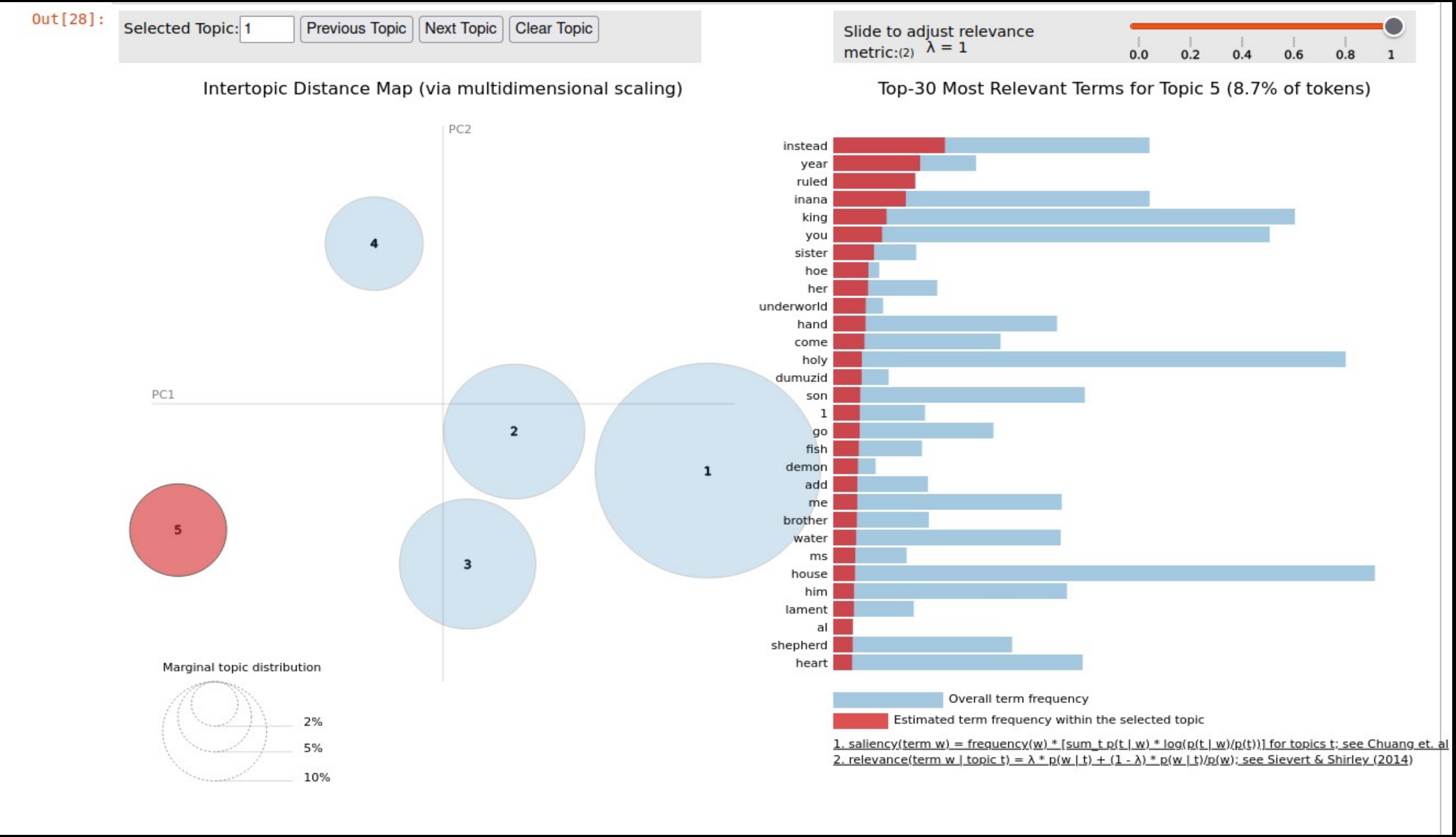
Topics chosen by LDA



Topics chosen by LDA

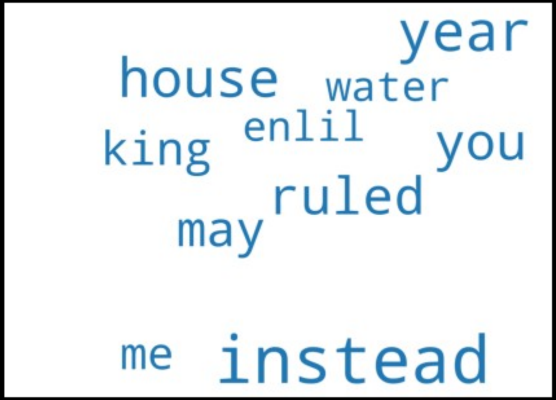


Topics chosen by LDA

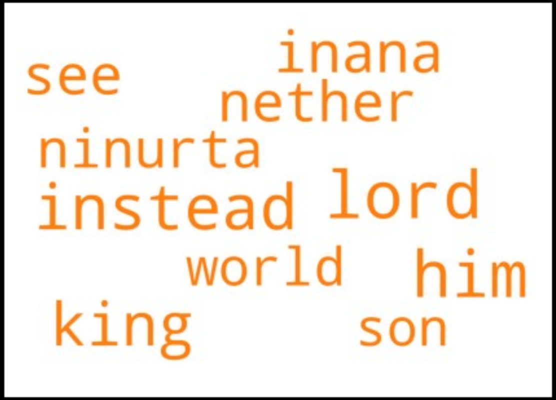


Topics chosen by LDA

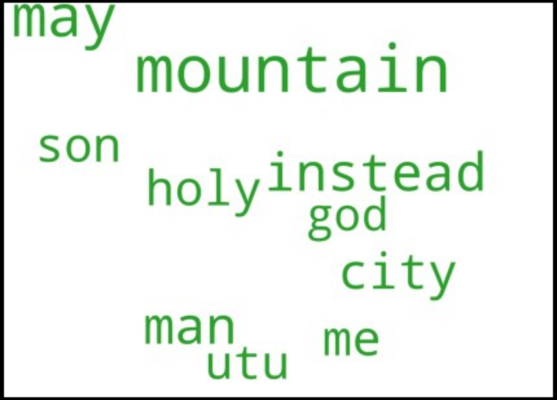
Topic 1



Topic 2



Topic 3

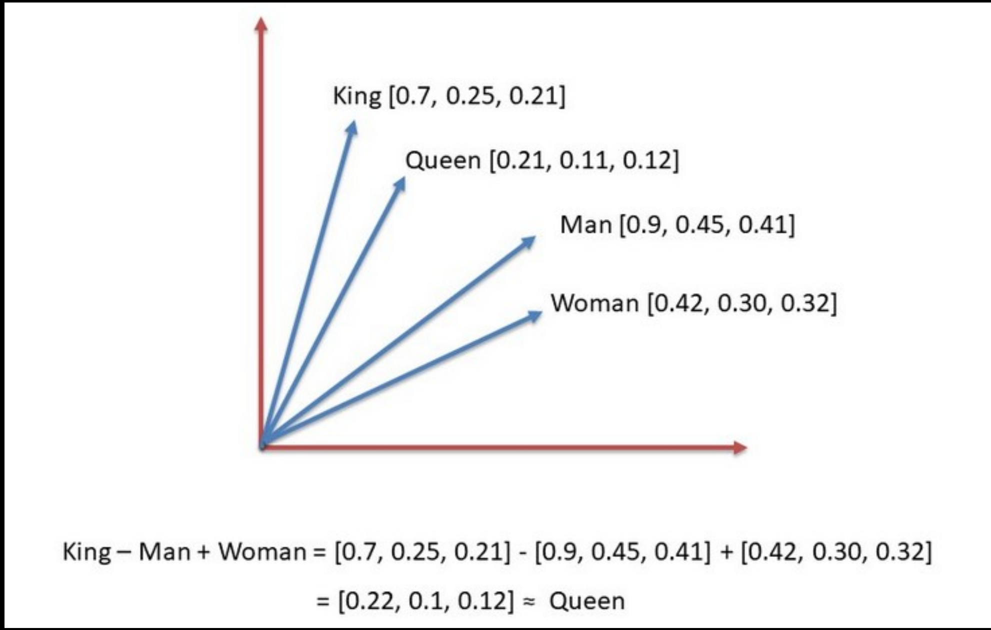


Topic 4

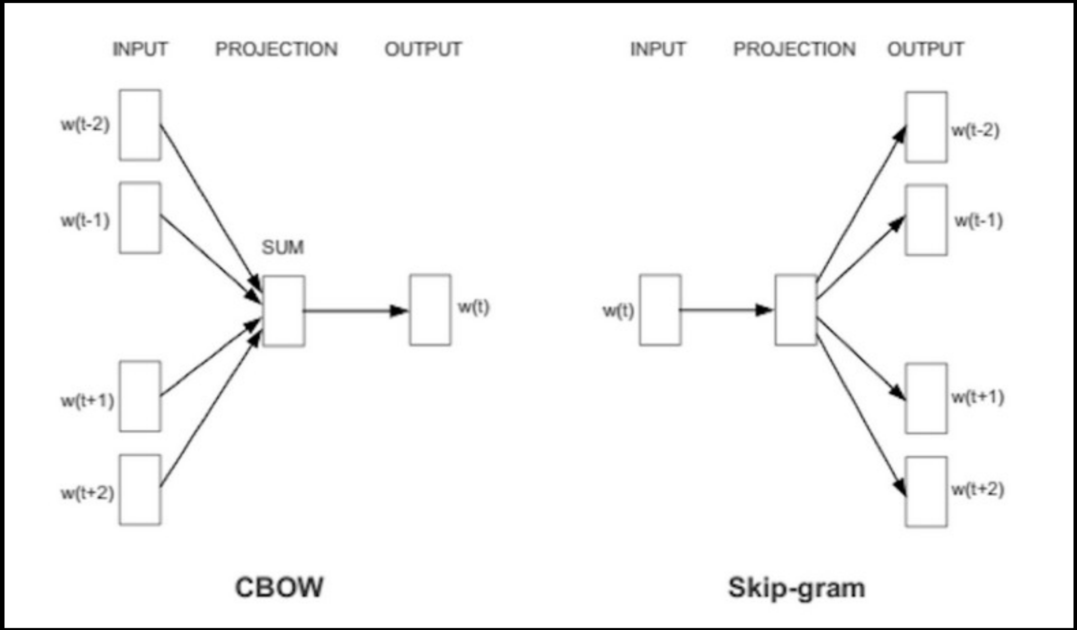
Topic 5

Building a Word2Vec-Model with Gensim

- Word2vec a technique to learn word embeddings (2-layer neural network)
- Input: a text corpus,
- Output : a set of vectors.
- Word embeddings make natural language computer-readable
- further implementation of mathematical operations on words can be used to detect their similarities



Building a Word2Vec-Model with Gensim



- Continuous bag of words(CBOW): using context to predict a target word
- Skip-gram: using a word to predict a target context

Exploring semantics

```
model.wv.most_similar('enlil')
```

```
[('god', 0.9947113990783691),
 ('anuna', 0.9925729036331177),
 ('prince', 0.991898775100708),
 ('name', 0.9901844263076782),
 ('august', 0.9900596141815186),
 ('excel', 0.9877880811691284),
 ('determined', 0.9857959747314453),
 ('lord', 0.9837357997894287),
 ('given', 0.9836603403091431),
 ('ninlil', 0.9835230112075806)]
```

Enlil and his attributes as an indicator for the correctness

Exploring semantics

```python
model.wv.most_similar('god')
```

```
[('anuna', 0.9980718493461609),
 ('august', 0.996114134788513),
 ('enlil', 0.9947112798690796),
 ('prince', 0.993477165699051),
 ('name', 0.9933406114578247),
 ('excel', 0.992594838142395),
 ('an', 0.991404533862305),
 ('inspire', 0.9893007278442383),
 ('determined', 0.9846497178077698),
 ('abzu', 0.9832668900489807)]
```

```python
model.wv.most_similar('nergal')
```

```
[('igial', 0.9994497299194336),
 ('disobedient', 0.9993062019348145),
 ('battlemace', 0.9992561340332031),
 ('support', 0.999177217483505),
 ('highland', 0.9991323351860046),
 ('noble', 0.999116182327270),
 ('country', 0.999106109142303),
 ('ninirsu', 0.999104142189025),
 ('splendour', 0.9990972280502319),
 ('horizon', 0.999084115028381)]
```

Exploring semantics

```
model.wv.most_similar('underworld')
```

```
[('entered', 0.9990392923355103),
 ('garment', 0.9989400506019592),
 ('incantation', 0.9987006187438965),
 ('priestess', 0.9986730813980103),
 ('yours', 0.9986082315444946),
 ('cultic', 0.9985716342926025),
 ('silent', 0.9985690712928772),
 ('mooring', 0.9985321760177612),
 ('charm', 0.9985049962997437),
 ('radiant', 0.9984983205795288)]
```

```
model.wv.most_similar('beer')
```

```
[('pour', 0.9994734525680542),
 ('honey', 0.9992457032203674),
 ('lamb', 0.9988067150115967),
 ('eats', 0.9987311363220215),
 ('poured', 0.9985946416854858),
 ('butter', 0.9985681772232056),
 ('marsh', 0.9985242486000061),
 ('planted', 0.9984870553016663),
 ('little', 0.9984645247459412),
 ('small', 0.9983784556388855)]
```

Exploring semantics

```
model.wv.most_similar('marriage')
```

```
[('ever', 0.9983763694763184),
 ('something', 0.99825119972229),
 ('herald', 0.9982298612594604),
 ('enough', 0.9982125759124756),
 ('table', 0.9981802105903625),
 ('gift', 0.9981703162193298),
 ('performs', 0.9981660842895508),
 ('walk', 0.9981579780578613),
 ('appearance', 0.9981566667556763),
 ('statue', 0.9981500506401062)]
```

```
model.wv.most_similar('child')
```

```
[('answered', 0.9989216327667236),
 ('treat', 0.998796463012695),
 ('talk', 0.99876707792282),
 ('wife', 0.998763978481292),
 ('elder', 0.998733222484588),
 ('speaks', 0.998615443706512),
 ('replied', 0.99861210584640),
 ('lap', 0.998541057109832),
 ('call', 0.998506903648376),
 ('intercourse', 0.998394131660461)]
```

Exploring semantics

King – Man + Woman = Queen

```
model.wv.most_similar(positive = ['king', 'woman'], negative = ['man'])
```

```
[('chose', 0.9678337574005127),
 ('lord', 0.9564954042434692),
 ('ambitious', 0.9556412696838379),
 ('land', 0.9482659697532654),
 ('kingship', 0.9480852484703064),
 ('unchangeable', 0.9447059631347656),
 ('determine', 0.9393962025642395),
 ('determined', 0.9366124868392944),
 ('fifty', 0.934846043586731),
 ('destiny', 0.9346879720687866)]
```

Exploring semantics

Female role of Enlil?

```
model.wv.most_similar(positive = ['enlil', 'god'], negative = ['man'])
```

```
[('earth', 0.9509128928184509),
 ('power', 0.9462100863456726),
 ('divine', 0.9433299899101257),
 ('heaven', 0.9365049600601196),
 ('great', 0.9176678657531738),
 ('an', 0.9054330587387085),
 ('perfecting', 0.9034489989280701),
 ('craved', 0.8756339550018311),
 ('deciding', 0.8756287097930908),
 ('decreeing', 0.8687440752983093)]
```

Exploring semantics

Excluding terms

```
model.wv.doesnt_match('enlil utu inana suen'.split())
```

```
'inana'
```

```
model.wv.doesnt_match('nibru larsam unug urim'.split())
```

```
'nibru'
```

```
model.wv.doesnt_match('fate sun war moon'.split())
```

```
'fate'
```

Exploring semantics

Excluding terms

```
model.wv.doesnt_match('bride bridegroom happy beer'.split())
```

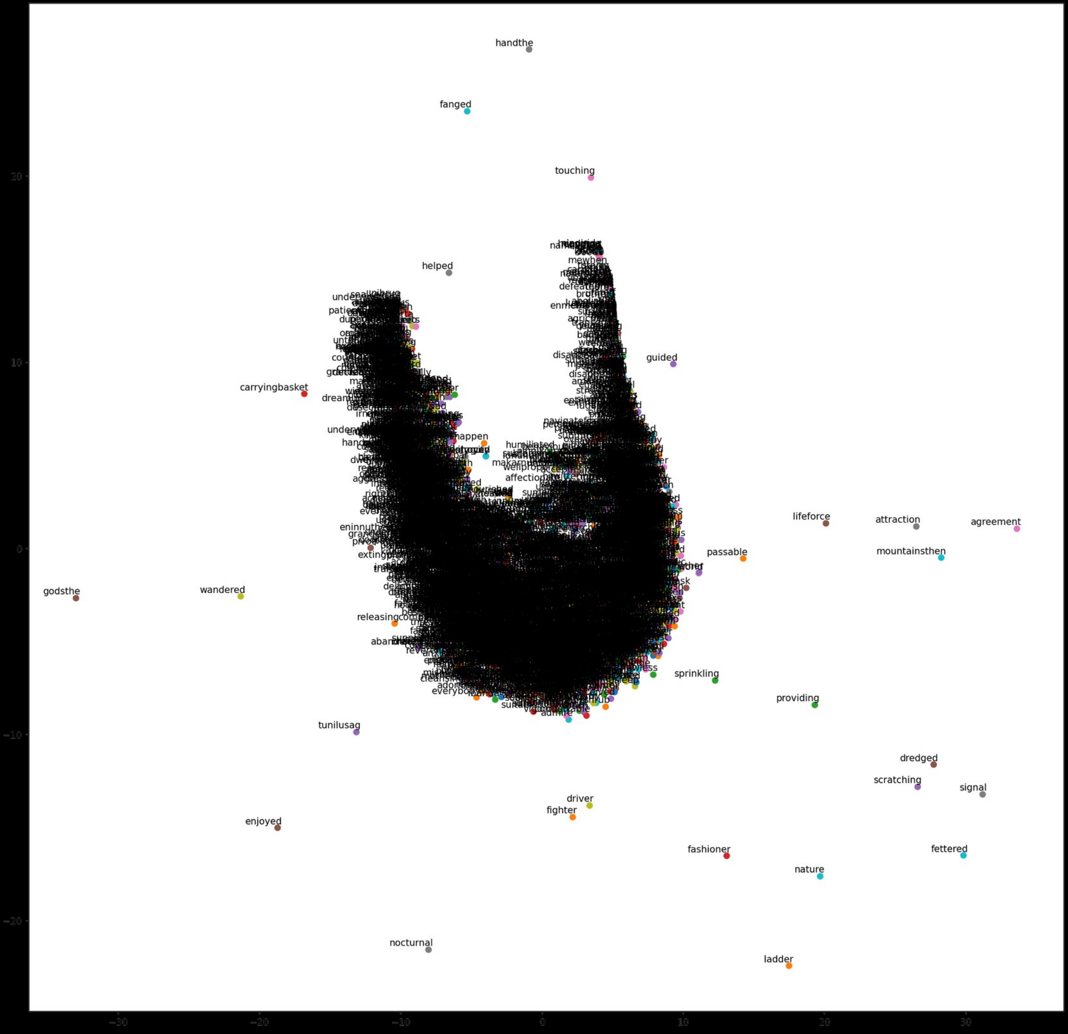
```
'bridegroom'
```

```
model.wv.doesnt_match('wife husband happy beer'.split())
```

```
'beer'
```

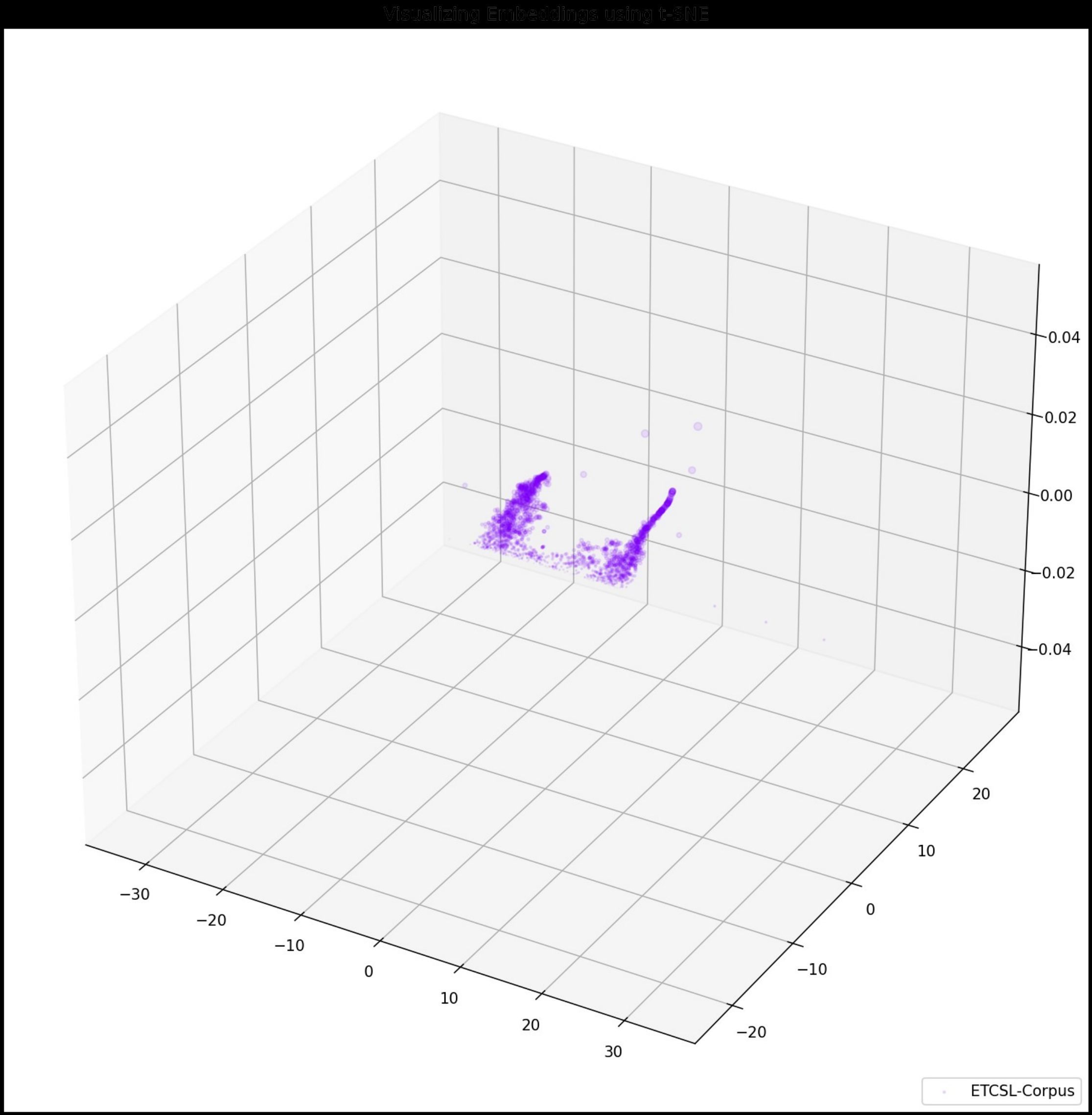
```
model.wv.doesnt_match('love child happy marriage'.split())
```

```
'love'
```



Mapping words

- Few outliers
- Corpus has two separated sections
- These are probably based on different terminology of genres



Mapping words

- Few outliers
- Corpus has two separated sections
- These are probably based on different terminology of genres

```
model.wv.doesnt_match('enlil beer happy'.split())
```

```
'enlil'
```

I thank you for your attention!

Building a Word2Vec-Model with Gensim

Continuous Bag of Words (CBOW) model:

- Dimension of embeddings: 300 vectors
- Minimum word-count: 3
- Window (maximum distance around target word): 10
- Subsampling-rate: 1e-3