Studying Semantic Chain Shifts with Word2Vec: FOOD>MEAT>FLESH

Richard Zimmermann

1st International Workshop
on Computational Approaches
to Historical Language Change, 2019

2 August 2019

Introduction

Chain Shifts

Chain Shifts

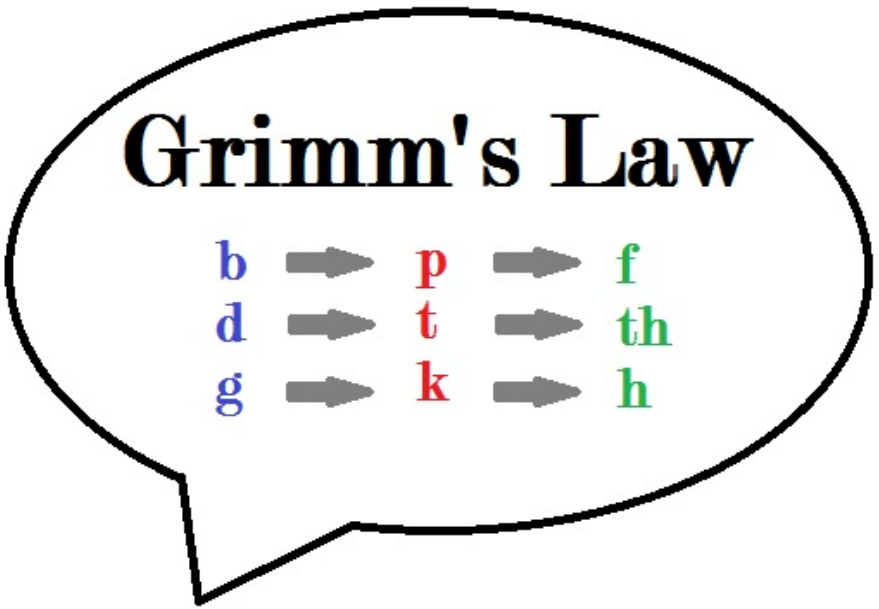


Figure 1 : A simplified representations of Grimm's Law

Chain Shifts

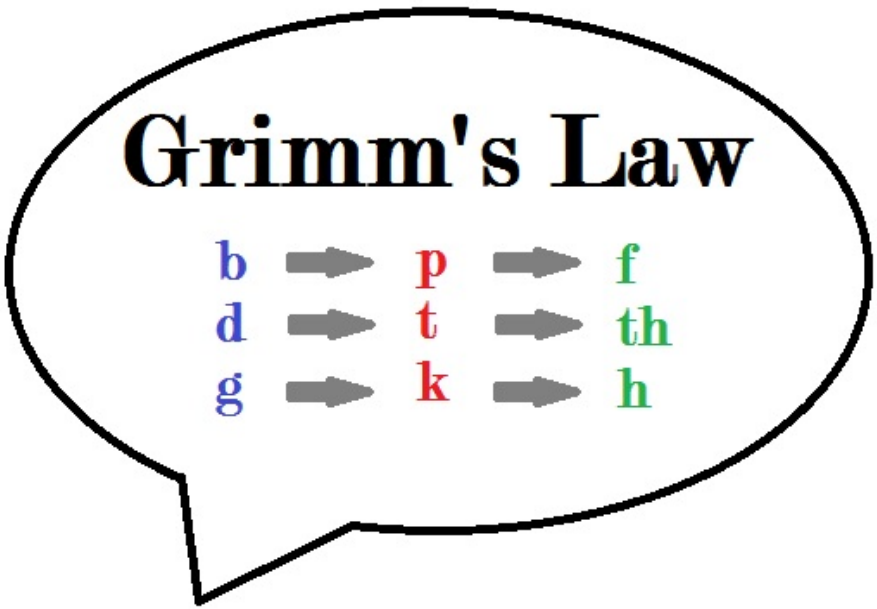


Figure 1 : A simplified representations of Grimm's Law

Lithuanian: *du***b***ùs* ≈ *dee***p**
Latin: *la***b***ium* ≈ *li***p**
Latin: *canna***b***is* ≈ *hem***p**
...

Chain Shifts

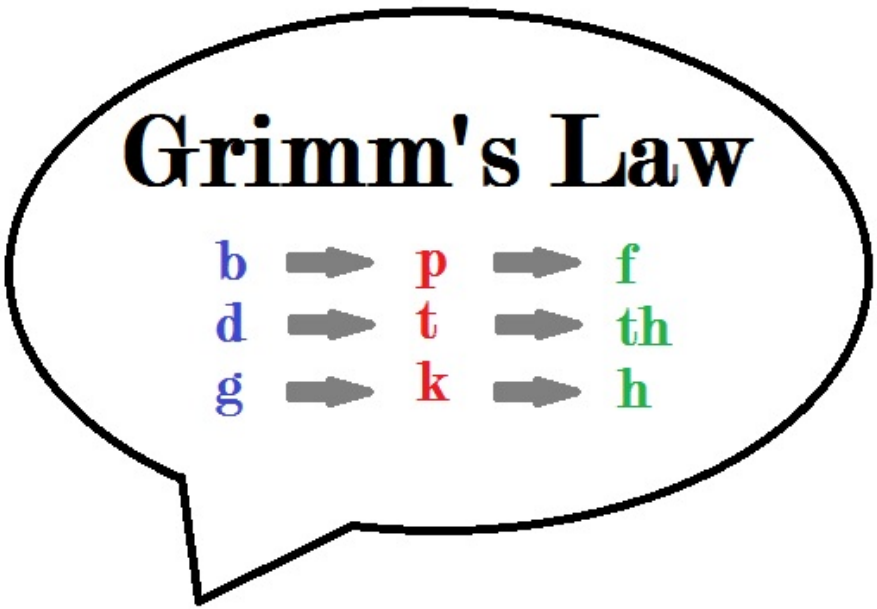


Figure 1 : A simplified representations of Grimm's Law

Lithuanian: *du**b**ùs* ≈ *dee**p***
Latin: *la**b**ium* ≈ *li**p***
Latin: *canna**b**is* ≈ *hem**p***
...

Ancient Greek: **π***ατήρ* ≈ **f***ather*
Latin: **p***isces* ≈ **f***ish*
Italian: **p***iede* ≈ **f***oot*
...

Chain Shifts in Semantics

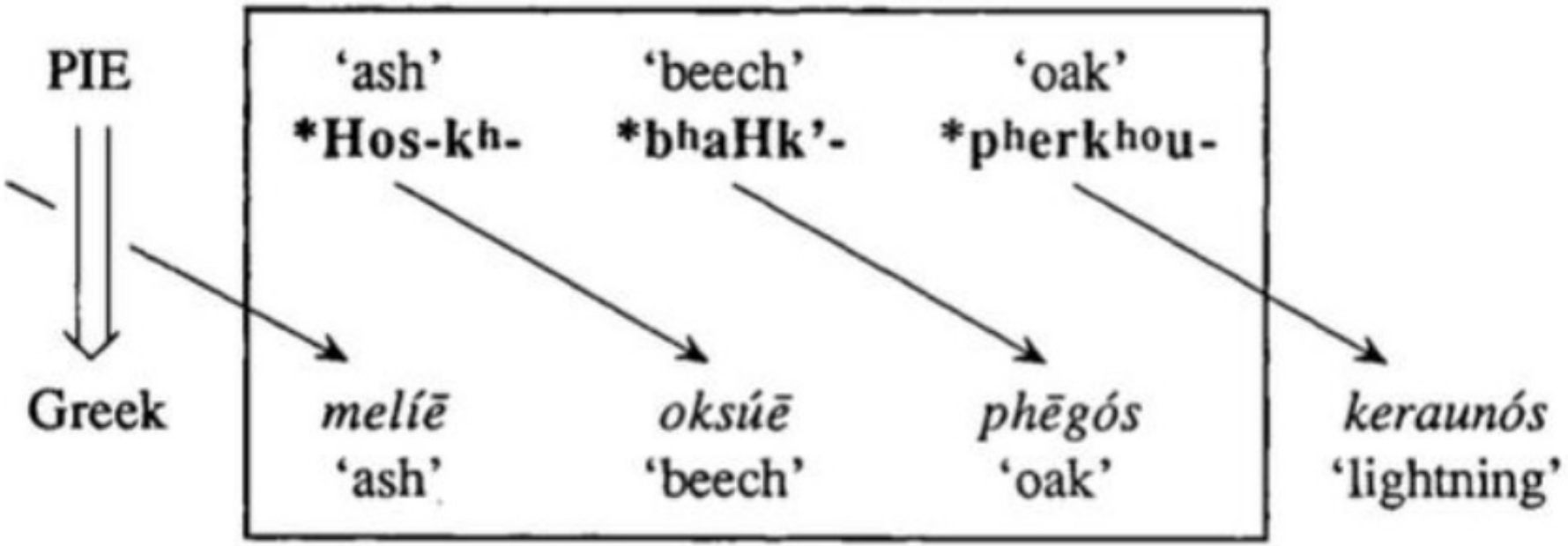


Figure 2 : Semantic chain shift in Ancient Greek tree terminology (from: Gamkrelidze & Ivanov 1995: 538)

FOOD > MEAT > FLESH

FOOD > MEAT > FLESH

(1) a. ic gean þ fen [...] into Ælmham þa preostas to foddan
'I give the fen [...] in Almham to the priests as food' (Ch 1489 27)

FOOD > MEAT > FLESH

(1) a. ic gean þ fen [...] into Ælmham þa preostas to foddan
'I give the fen [...] in Almham to the priests as food' (Ch 1489 27)

b. lar is ðære sawle foda.
'lore is the soul's food' (ÆCHom II, 29 233.110)

FOOD > MEAT > FLESH

(1) a. ic gean þ fen [...] into Ælmham þa preostas to foddan
'I give the fen [...] in Almham to the priests as food' (Ch 1489 27)

b. lar is ðære sawle foda.
'lore is the soul's food' (ÆCHom II, 29 233.110)

(2) a. þæt hi hæfdun to seofon nihtum mete
'so that they had for seven nights meat' (ChronA 82.23)

FOOD > MEAT > FLESH

(1) a. ic gean þ fen [...] into Ælmham þa preostas to foddan
'I give the fen [...] in Almham to the priests as food' (Ch 1489 27)

b. lar is ðære sawle foda.
'lore is the soul's food' (ÆCHom II, 29 233.110)

(2) a. þæt hi hæfdun to seofon nihtum mete
'so that they had for seven nights meat' (ChronA 82.23)

b. morgenmete
'morning meat, breakfast'

FOOD > MEAT > FLESH

(1) a. ic gean þ fen [...] into Ælmham þa preostas to foddan
'I give the fen [...] in Almham to the priests as food' (Ch 1489 27)

b. lar is ðære sawle foda.
'lore is the soul's food' (ÆCHom II, 29 233.110)

(2) a. þæt hi hæfdun to seofon nihtum mete
'so that they had for seven nights meat' (ChronA 82.23)

b. morgenmete
'morning meat, breakfast'

(3) a. þa was corn dære & flec & cæse & butere
'Then was corn dear, and flesh and cheese and butter' (ChronE 1137.39)

FOOD > MEAT > FLESH

(1) a. ic gean þ fen [...] into Ælmham þa preostas to foddan
'I give the fen [...] in Almham to the priests as food' (Ch 1489 27)

b. lar is ðære sawle foda.
'lore is the soul's food' (ÆCHom II, 29 233.110)

(2) a. þæt hi hæfdun to seofon nihtum mete
'so that they had for seven nights meat' (ChronA 82.23)

b. morgenmete
'morning meat, breakfast'

(3) a. þa was corn dære & flec & cæse & butere
'Then was corn dear, and flesh and cheese and butter' (ChronE 1137.39)

b. hi eton swynen flæsc
'they ate swine's flesh' (PPs (prose) 16.14)

Outline

Outline

Goals:

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH
- Determine if the change proceeds as a pull or push chain

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH
- Determine if the change proceeds as a pull or push chain
- Identify the time course of the change

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH
- Determine if the change proceeds as a pull or push chain
- Identify the time course of the change

Roadmap

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH
- Determine if the change proceeds as a pull or push chain
- Identify the time course of the change

Roadmap

- Data

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH
- Determine if the change proceeds as a pull or push chain
- Identify the time course of the change

Roadmap

- Data
- Results

Outline

Goals:

- Exemplify an analysis of semantic chain shifts with Word2Vec: FOOD > MEAT > FLESH
- Determine if the change proceeds as a pull or push chain
- Identify the time course of the change

Roadmap

- Data
- Results
- Conclusion

Data

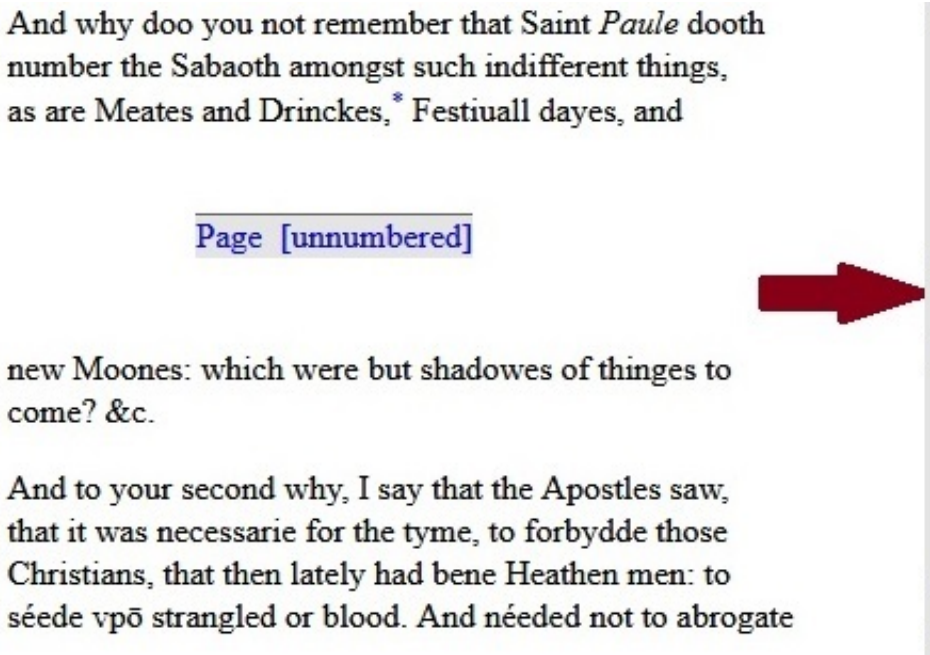
Corpora Used

Period	Corpus	Size
1 1425-1475	Innsbruck*	2.5m
2 1475-1525	EEBO**	9.0m
3 1526-1575	EEBO	35.6m
4 1576-1625	EEBO	149.8m
5 1626-1675	EEBO	330.9m
6 1676-1725	EEBO, ECCO**	230.5m
7 1726-1775	ECCO	31.6m
8 1776-1825	ECCO, CLMET***	38.7m
9 1826-1875	CLMET	10.9m
10 1876-1925	CLMET	8.1m

Table 1 : Periodization of the data, their source corpora, and their size (in million words)

* Markus 2010
** digitzed by the Text Creation Partnership (TCP)
*** Diller et al. 2011

Processing of Textual Material



And why doo you not remember that Saint *Paule* dooth number the Sabaoth amongst such indifferent things, as are Meates and Drinckes,* Festiuall dayes, and

Page [unnumbered]

new Moones: which were but shadowes of thinges to come? &c.

And to your second why, I say that the Apostles saw, that it was necessarie for the tyme, to forbydde those Christians, that then lately had bene Heathen men: to séede vpō strangled or blood. And néeded not to abrogate

```
1079  and whi do iou not remember that
      saint paul do number the sabaoth
      amongst such indifferent thing
      as are MEAT and DRINK festiual
      dai and new moon :
1080  which were but shadowes of
      thinges to come ? &c
1081  and to iour second whi i sai
      that the apostles saw that it
      was necessarie for the time to
      forbid those christian that then
      lateli had bene heathen men :
1082  to seede upon strangled or BLOOD
1083  and needed not to abrogate
```

Figure 3 : Illustration of preparation of the text data

Processing of Textual Material

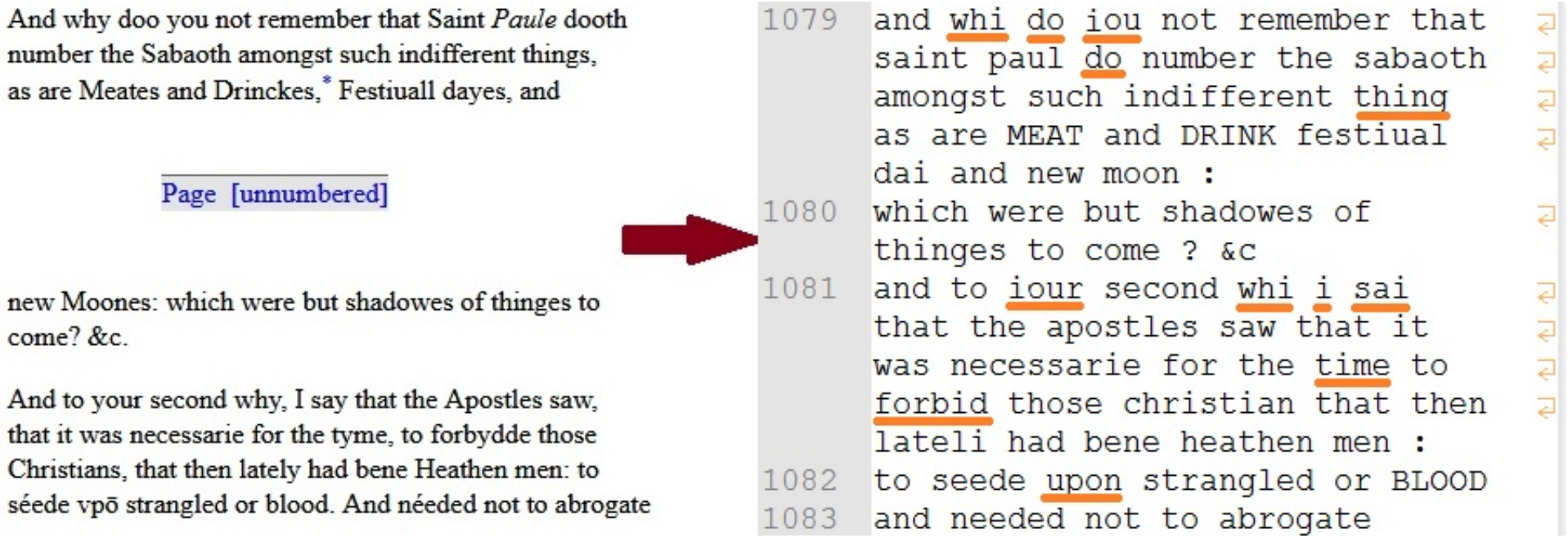


Figure 3 : Illustration of preparation of the text data

- c. 850 RegEx for spelling normalization

Processing of Textual Material

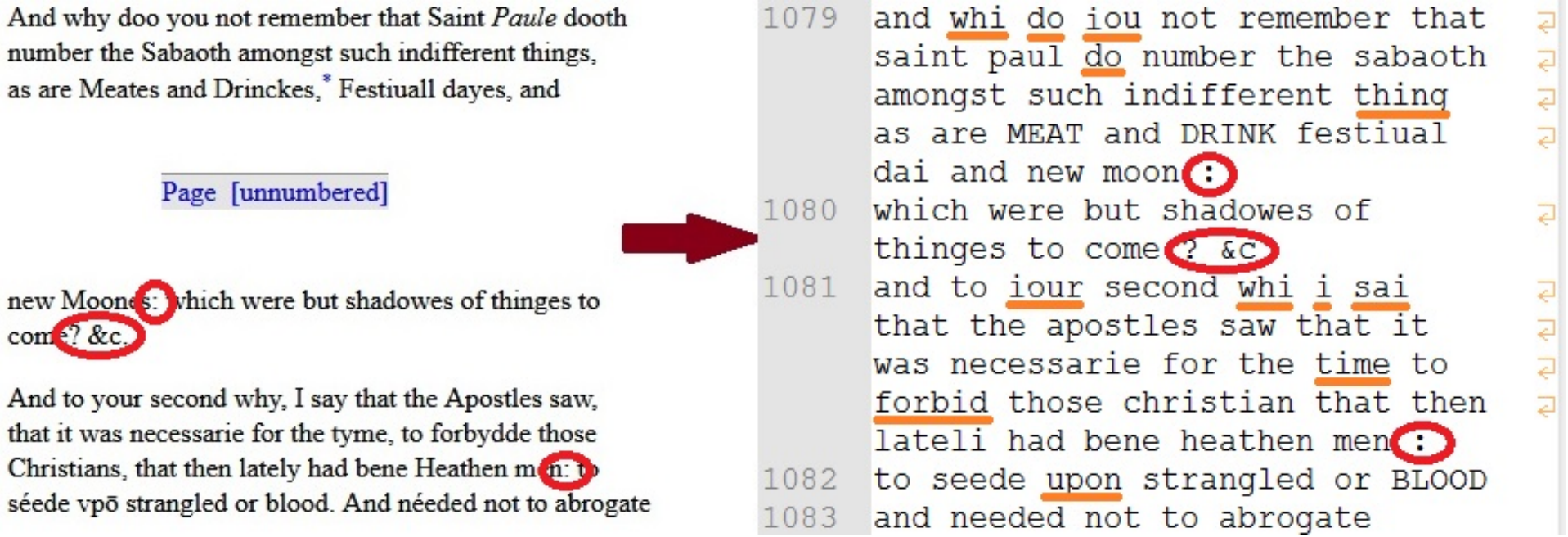


Figure 3 : Illustration of preparation of the text data

- c. 850 RegEx for spelling normalization
- punctuation and paragraphs used for tokenization

Processing of Textual Material

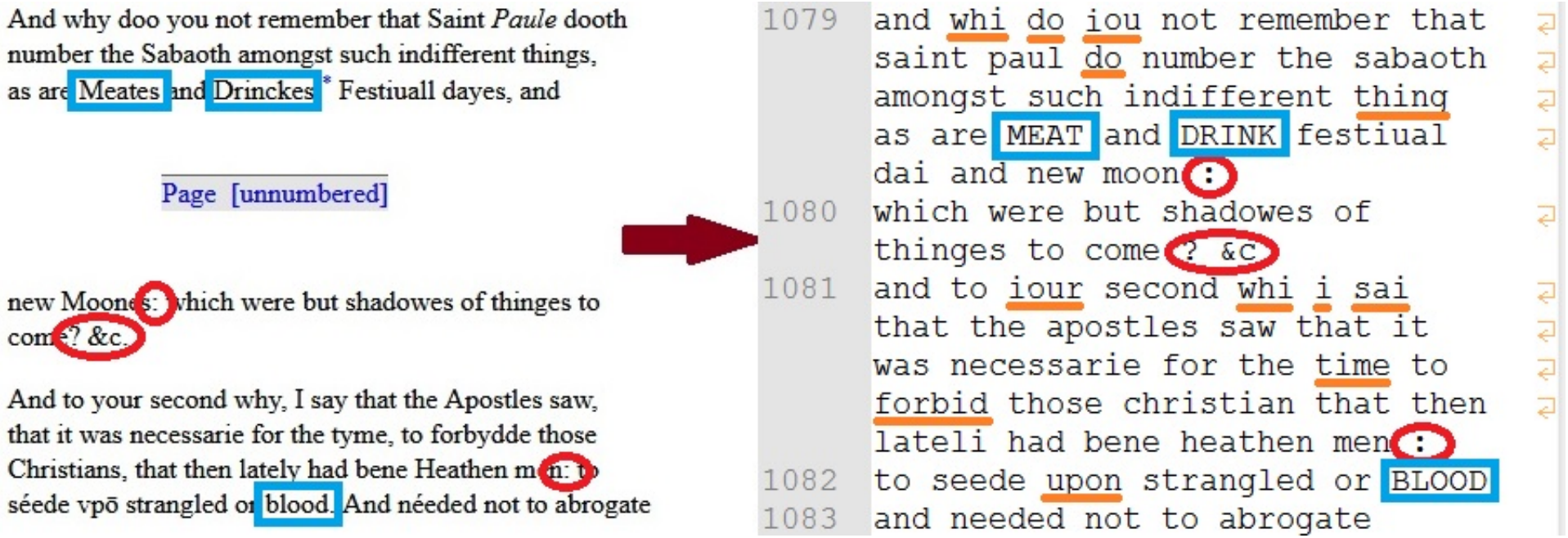


Figure 3 : Illustration of preparation of the text data

- c. 850 RegEx for spelling normalization
- punctuation and paragraphs used for tokenization
- lemmatization of target and central context words

Processing of Textual Material

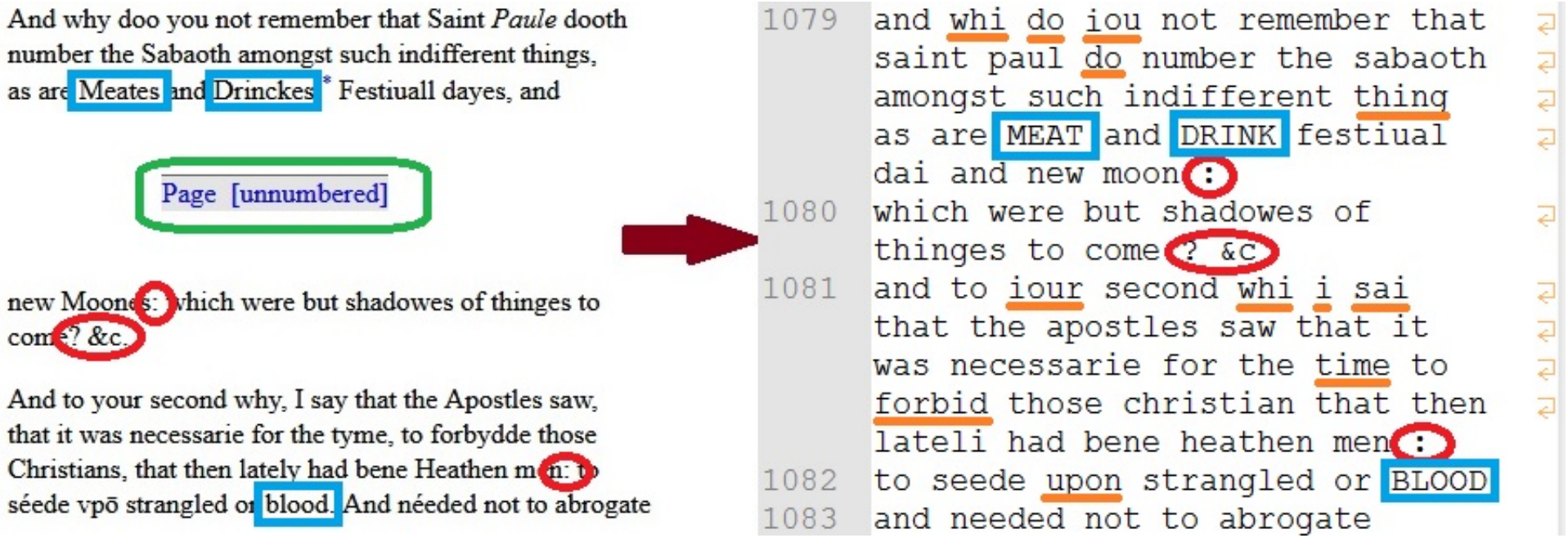


Figure 3 : Illustration of preparation of the text data

- c. 850 RegEx for spelling normalization
- punctuation and paragraphs used for tokenization
- lemmatization of target and central context words
- removal of text mark-up

Word Embedding in *Python*

```
######################################################################
### (1) enter the raw data, a .txt file

file_specified = "no"

while file_specified == "no":
    ## raw = raw_input("Please specify a .txt file you wish to format.\n(e.g. 'C
```

python™

gensim
topic modelling for humans

```
# import modules & set up logging

import gensim
import logging
logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=lo
import time
start = time.time()

# train word2vec on the two sentences
model = gensim.models.Word2Vec(sentences, size=250, window=20, min_count=5, work
```

Word Embedding in *Python*

- Word2Vec model implemented in Python's Gensim library (Řehuřek & Sojka 2010)


```
######################################################################
### (1) enter the raw data, a .txt file

file_specified = "no"

while file_specified == "no":
    ## raw = raw_input("Please specify a .txt file you wish to format.\n(e.g. 'C

# import modules & set up logging

import gensim
import logging
logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=lo
import time
start = time.time()

# train word2vec on the two sentences
model = gensim.models.Word2Vec(sentences, size=250, window=20, min_count=5, work
```

Word Embedding in *Python*

- Word2Vec model implemented in Python's Gensim library (Řehuřek & Sojka 2010)
- Continuous bag of words

Word Embedding in *Python*

- Word2Vec model implemented in Python's Gensim library (Řehuřek & Sojka 2010)
- Continuous bag of words
- 250-dimensional word vectors, 20 words context window, 5 word minimum count

Results

PCA Plot (Li et al. 2019)

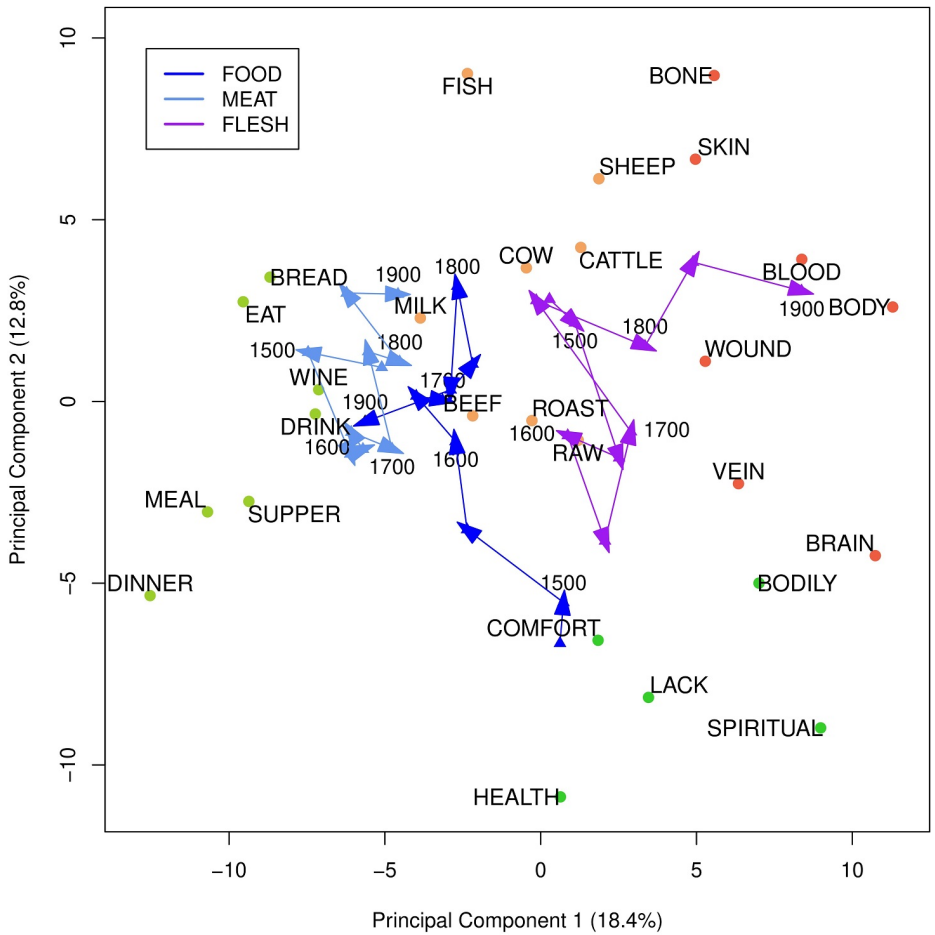


Figure 4 : Representation of semantic chain shift FOOD > MEAT > FLESH

PCA Plot (Li et al. 2019)

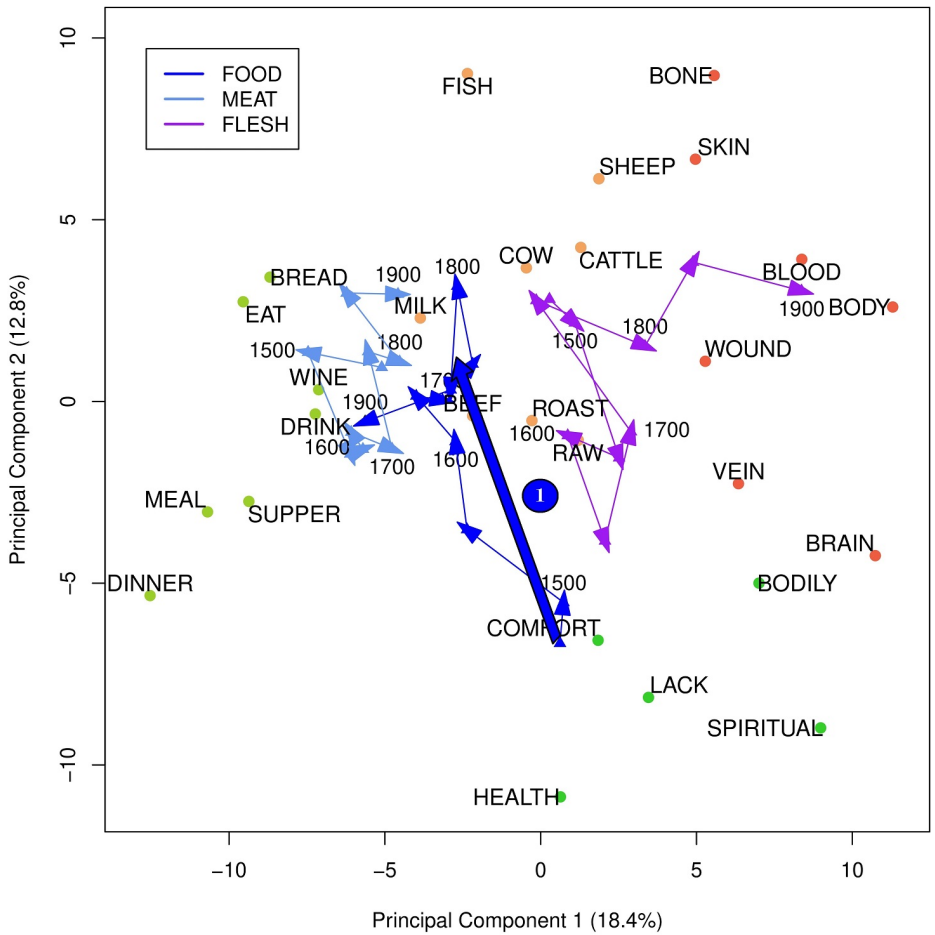


Figure 4 : Representation of semantic chain shift FOOD > MEAT > FLESH

PCA Plot (Li et al. 2019)

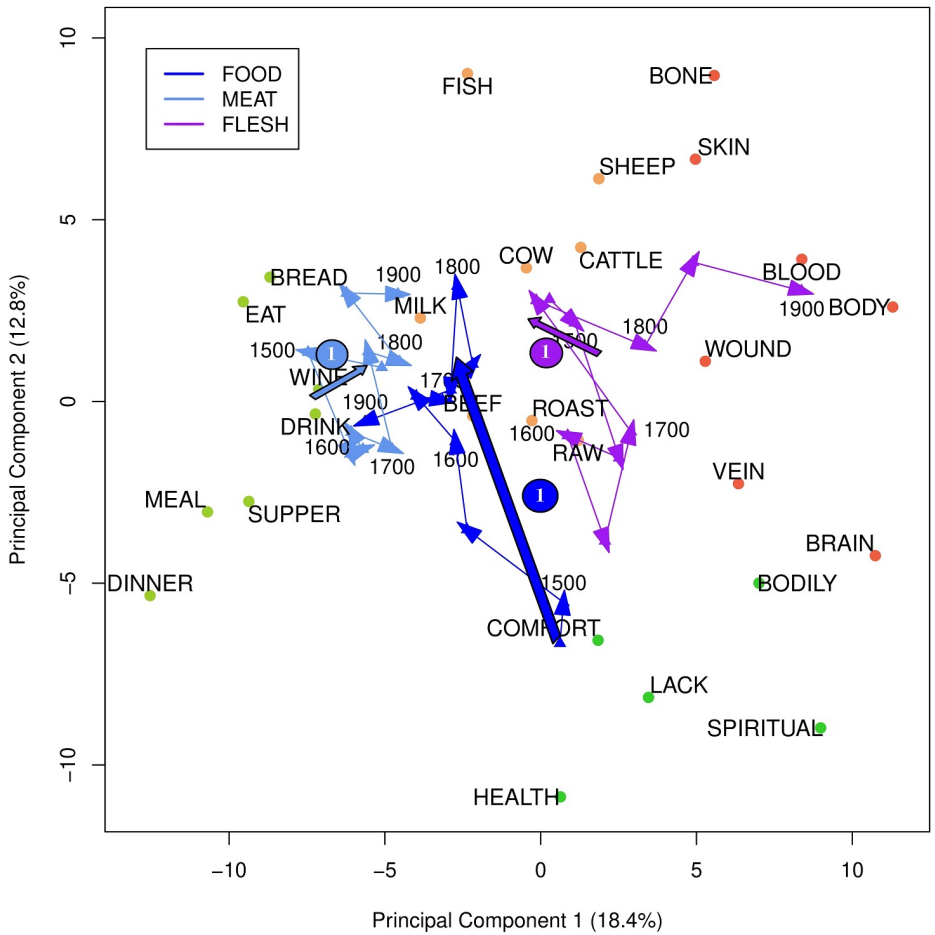


Figure 4 : Representation of semantic chain shift FOOD > MEAT > FLESH

PCA Plot (Li et al. 2019)

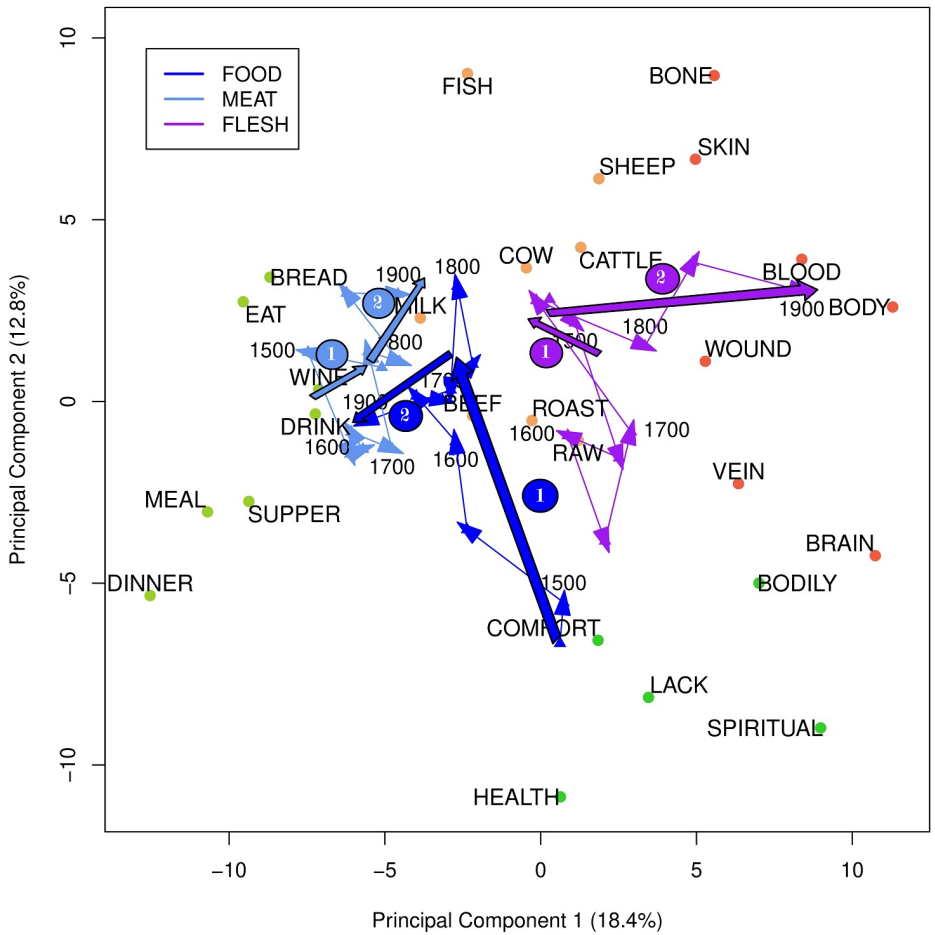


Figure 4 : Representation of semantic chain shift FOOD > MEAT > FLESH

FOOD pushes on MEAT

Pier's Plowman

A-text, c. 1370

And comaundite of his curteisie
in comoun þre þinges;
[...]
þat on is vesture [...]
þat oþer is mete at meel ...

C-text, c. 1385

He comaundid of his cortesye
in comune þre þynges
[...]
The ferst of þo ys fode and
vesture þe secounde

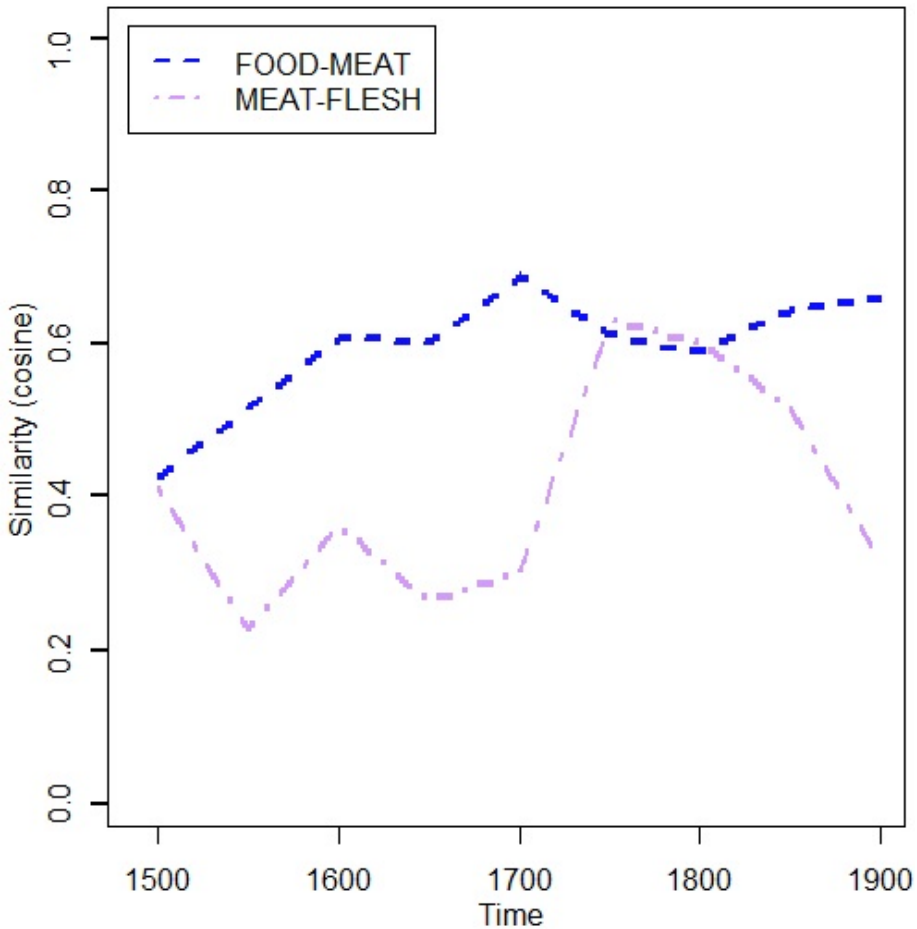


Figure 5 : Cosine similarities

FLESH - High Degree of Polysemy

flĕsh n.

1a. (a) The flesh of the human body; esp., the muscular, gristly, and glandular portions
(b) corresponding portions of an animal's body

1b. *Theol.* The communion bread: **Cristes ~**, bread as transubstantiated

1c. *Med.* and *surg.* (a) A muscle; (b) the flesh of an animal used medicinally;

2a. (a) The flesh of an animal (or fowl) used for human food, meat
(b) *pl.* kinds of meat

2b. (a) **bef ~, calf (calves) ~, camel ~, capoun ~, colver ~, cou(es) ~, deres ~, gos ~, gotes ~, hen ~, hors ~, rotheres ~, shepes ~, swin(es) ~**, etc.; (b) **fat ~**, fat meat; **fresh ~**, fresh meat, unsalted or uncured meat; **gret ~**, coarse meat, plain meat; **ibred ~**, broiled or roasted meat; **rau ~**, raw meat; **rest(i) ~**, rancid meat; **rosted ~**, roast meat; **salt ~**, meat cured with salt; **soden ~**, boiled meat; etc.

...

5a. The physical or sensual nature of man (as opposed to the spiritual)

5b. The sex urge, sexual desire or passion; sexual intercourse

(excerpt from the Middle English Dictionary)

Gradualness and Vagueness of Semantic Change

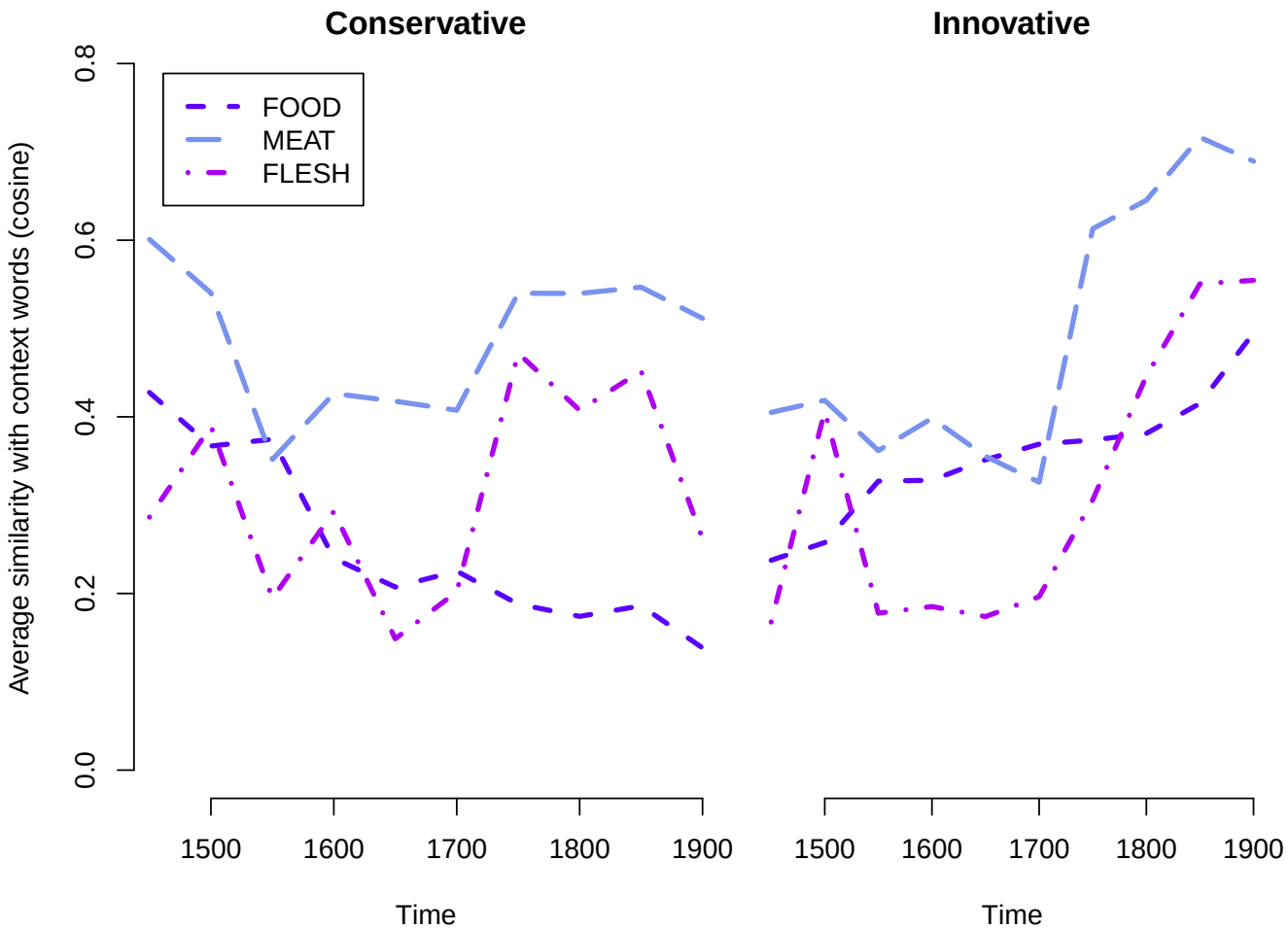


Figure 6 : Average similarity values between a set of conservative and innovative contexts words and each target word

Direction and Time Course of the Chain Shift

Direction and Time Course of the Chain Shift

Central Finding

The semantic shift FOOD > MEAT > FLESH likely proceeded as a **push chain**. It is possible to discern two phases, an initial push **c. 1450-1700** and subsequent changes **c. 1700-1900**.

Direction and Time Course of the Chain Shift

Central Finding

The semantic shift FOOD > MEAT > FLESH likely proceeded as a **push chain**. It is possible to discern two phases, an initial push **c. 1450-1700** and subsequent changes **c. 1700-1900**.

FOOD > MEAT, c. 1450-1700

There is solid evidence that FOOD first encroached on MEAT from the earliest data points on, c. 1450. The change may possibly have begun even earlier (14th century?).

Direction and Time Course of the Chain Shift

Central Finding

The semantic shift FOOD > MEAT > FLESH likely proceeded as a **push chain**.
It is possible to discern two phases, an initial push **c. 1450-1700** and subsequent changes **c. 1700-1900**.

FOOD > MEAT, c. 1450-1700

There is solid evidence that FOOD first encroached on MEAT from the earliest data points on, c. 1450. The change may possibly have begun even earlier (14th century?).

MEAT > FLESH, c. 1700-1900

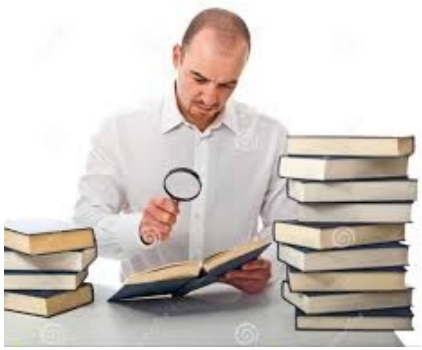
The shifts in MEAT and FLESH are harder to pinpoint. There is some evidence for a secondary push from c. 1700 on (cosine similarity).
However, this finding remains uncertain. (1) FLESH is highly polysemous (genre effect, noise). (2) MEAT does not clearly dissociate from the 'eating' domain even after FOOD has acquired its modern meaning. (3) The MEAT and FLESH measures move roughly in parallel (average similarity).

Conclusion

Alternative Methods

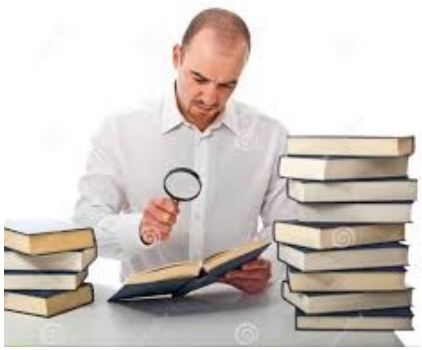
Alternative Methods

Manual analysis

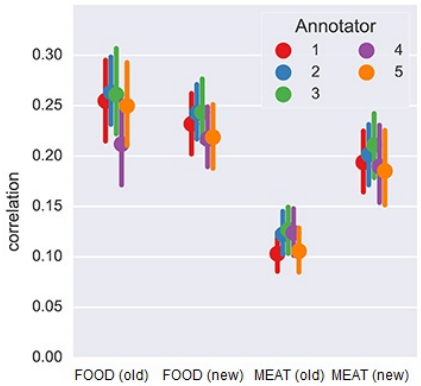


Alternative Methods

Manual analysis

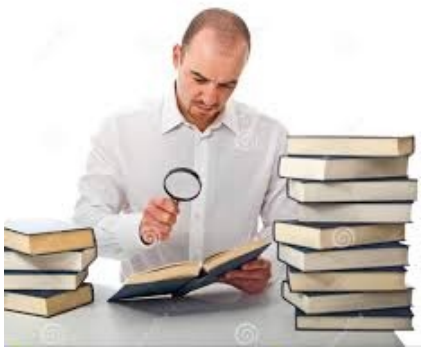


Inter-annotator agreement

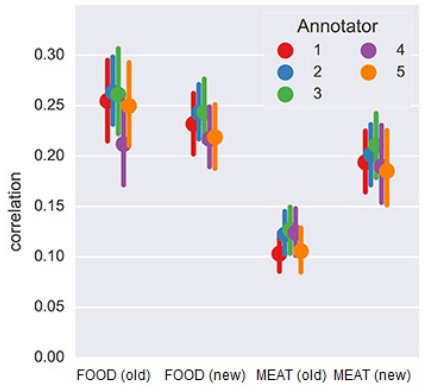


Alternative Methods

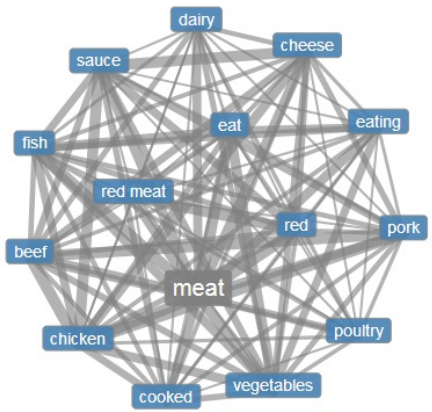
Manual analysis



Inter-annotator agreement

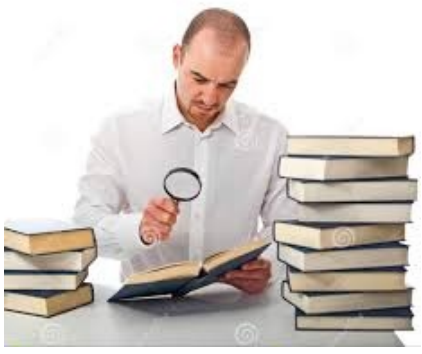


Collocations

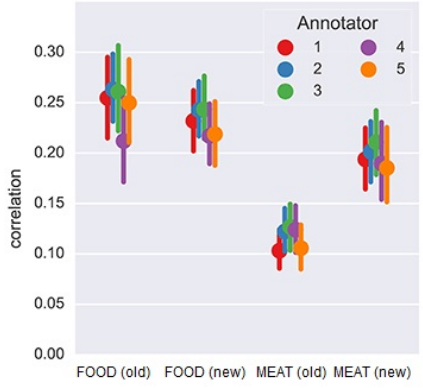


Alternative Methods

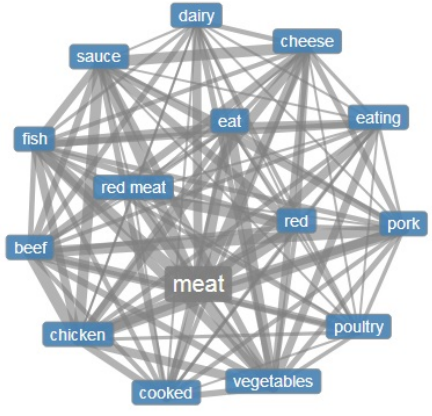
Manual analysis



Inter-annotator agreement



Collocations



Collostructions (Stefanovitsch & Gries 2003)

Collexeme	Collostruction strength	Collexeme	Collostruction strength
trick (92)	2.11E-267	delude (19)	8.83E-49
fool (77)	1.68E-187	talk (62)	2.38E-48
coerce (53)	1.15E-158	goad (18)	1.35E-46
force (101)	6.31E-136	shame (19)	1.28E-45
mislead (57)	9.57E-110	brainwash (13)	2.42E-37
bully (45)	2.53E-109	seduce (17)	2.56E-35
deceive (48)	5.94E-109	push (34)	6.66E-35
con (34)	4.41E-102	tempt (22)	3.37E-32
pressurise (39)	4.8E-101	manipulate (19)	3.3E-31
provoke (48)	4.05E-87	inveigle (10)	1.04E-30
pressure (30)	3.88E-85	hoodwink (10)	1.52E-29
cajole (28)	4.08E-85	panick (15)	7.75E-28
blackmail (25)	3.31E-64	lure (14)	1.23E-27
dupe (19)	7.77E-52	lull (11)	4.62E-26
coax (22)	6E-51	dragoon (8)	1.63E-25

My View on Word2Vec for Semantic Change

My View on Word2Vec for Semantic Change

- + Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec

My View on Word2Vec for Semantic Change

- + Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation

My View on Word2Vec for Semantic Change

- + Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative

My View on Word2Vec for Semantic Change

- + Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative
  3. abstract over the particularities of large amounts of text data

My View on Word2Vec for Semantic Change

- + Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative
  3. abstract over the particularities of large amounts of text data
- - Numerous fine-grained research questions cannot be addressed without further modification

My View on Word2Vec for Semantic Change

- + Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative
  3. abstract over the particularities of large amounts of text data
- - Numerous fine-grained research questions cannot be addressed without further modification
  1. Identification of start and end point of a change

My View on Word2Vec for Semantic Change

- \+ Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative
  3. abstract over the particularities of large amounts of text data
- \- Numerous fine-grained research questions cannot be addressed without further modification
  1. Identification of start and end point of a change
  2. Significance tests and effect size measures to separate real historical effects from possible noise

My View on Word2Vec for Semantic Change

- \+ Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative
  3. abstract over the particularities of large amounts of text data
- \- Numerous fine-grained research questions cannot be addressed without further modification
  1. Identification of start and end point of a change
  2. Significance tests and effect size measures to separate real historical effects from possible noise
  3. Addition of independent variables, such as genre, social class, dialect, etc.

My View on Word2Vec for Semantic Change

- \+ Semantic chain shifts, their direction and general time courses can successfully be investigated with Word2Vec
  1. easy to use for linguists, off-the-shelf libraries and good documentation
  2. objective, replicable, quantitative
  3. abstract over the particularities of large amounts of text data
- \- Numerous fine-grained research questions cannot be addressed without further modification
  1. Identification of start and end point of a change
  2. Significance tests and effect size measures to separate real historical effects from possible noise
  3. Addition of independent variables, such as genre, social class, dialect, etc.
- "Currently, word embedding methods can supplement, but should not supplant, careful linguistic studies on semantic change."

Thank you very much for your attention!

References

Gamkrelidze, Thomas V. & Ivanov, Vjašeslav, V. (1995) *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and a Proto-Culture*. Berlin: De Gruyter.

Li, Ying Engelthaler, Tomas, Siew Cynthia S. Q. & Hills Thomas T. (2019) 'The Macroscope: A Tool for Examining the Historical Structure of Language.' *Behavior Research Methods*.

Řehuřek, Radim & Sojka, Petr (2010) 'Software Framework for Topic Modelling with Large Corpora.' In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta. ELRA. 45-50.

Stefanowitsch, Anatol & Griess, Stefan Th. (2003) 'Collostructions: Investigating the Interaction of Words and Constructions.' *International Journal of Corpus Linguistic* **8.2**, 0943.