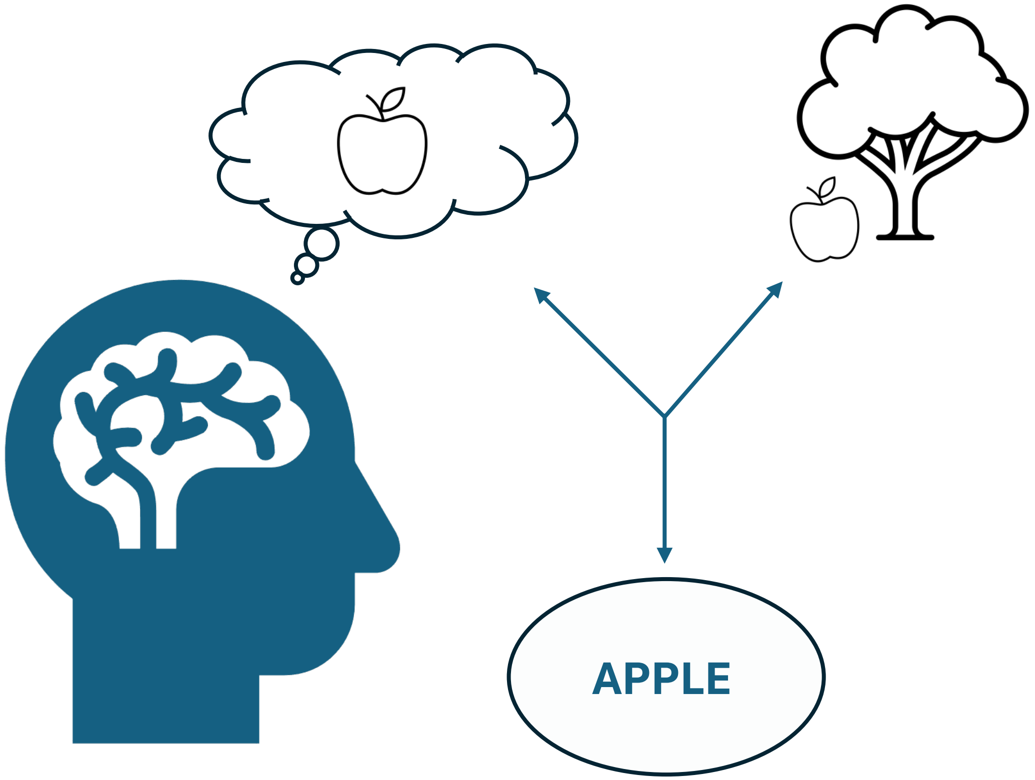
UNVEILING WORD MEANING

A statistical study on primitive semantic features

TOPIC

- Lexical semantics, Psicolinguistics and Neuro-cognitive science
    - **Word meaning**
    - How meaning is constructed
    - **Brain-level organization of meaning**
    - Theories of **semantic composition** and **representation**
    - Find the optimal way to represent meaning



BACKGROUND

- *Classical approaches Vs.* ***Componential approaches***
- *Toward a brain-based componential semantic representation (Binder et al., 2016)*

DATASET *(Binder et al., 2016)*

- Units → **Lexical items** *(i.e., Nouns, Verbs, Adjectives)*
- Features → **attributes** based on aspects of mental experience *(e.g., Visive, Auditive, Spatial, Temporal etc.)*

p =66

Units/Features	**Vision**	**Color**	**Sound**	**Temperature**	**Scene**	....	**Type**
House	*** +	*** +	*** +	*** +	*** +	....	Thing
Reporter						....	Thing
Happy						....	Property
Scream						....	Action
Shy						....	Property
....	....	....	....	....	....	....	
Gorilla						....	Thing
Television						....	Thing

n =534

STEPS OF THE RESEARCH

- **Data preparation**
  - Delete duplicates
  - Missing values
  - Normalization

- **Data analysis**
  - Features analysis
  - Unsupervised learning
    - Clustering
    - PCA
  - Supervised learning
    - Classification task
    - Feature selection

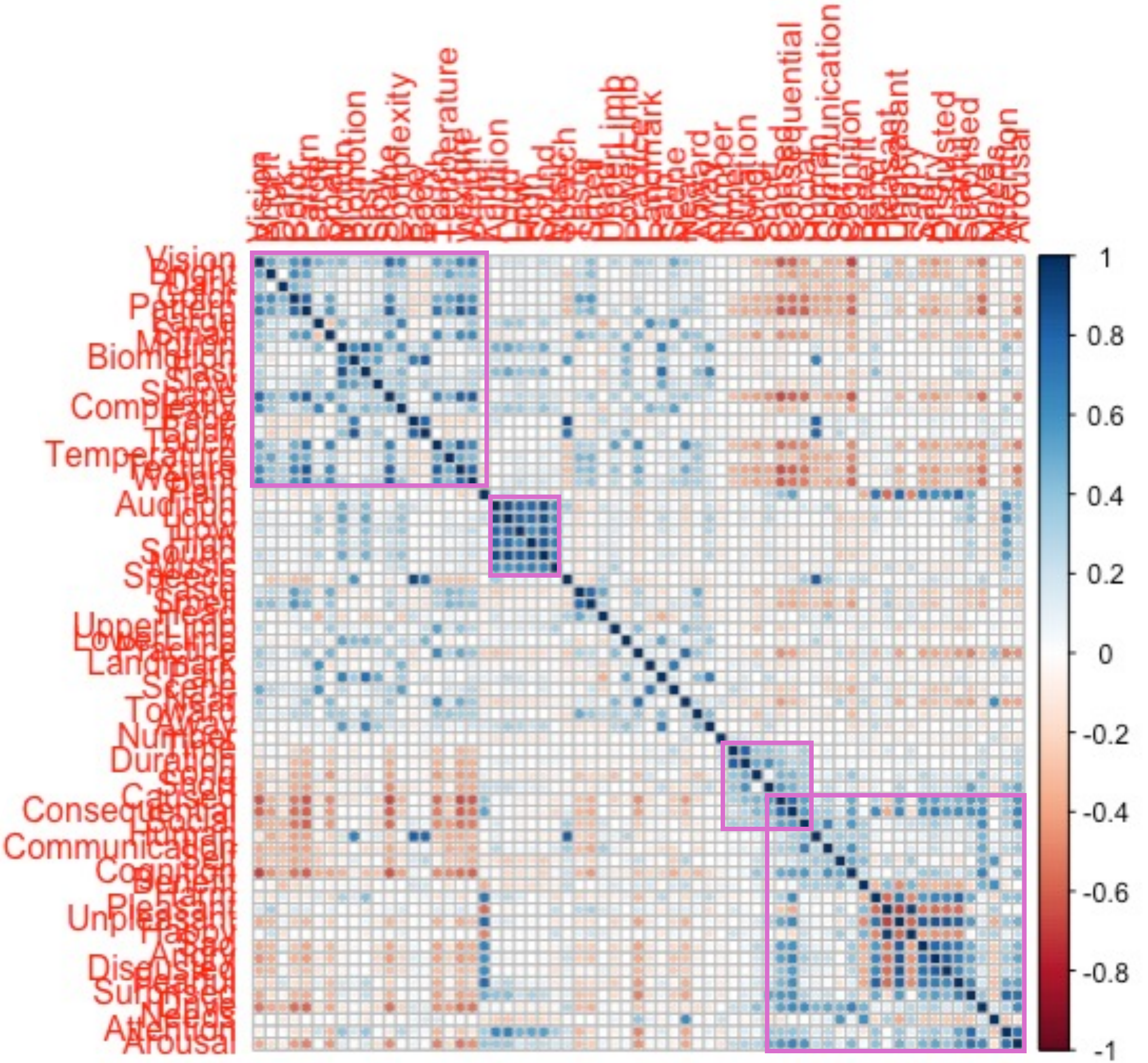
GOALS OF THE RESEARCH

- Prove the validity of such an approach
- Investigate the potential of this componential approach
- Make assumptions on the dynamics of meaning construction

FEATURES ANALYSIS

MULTICOLLINEARITY

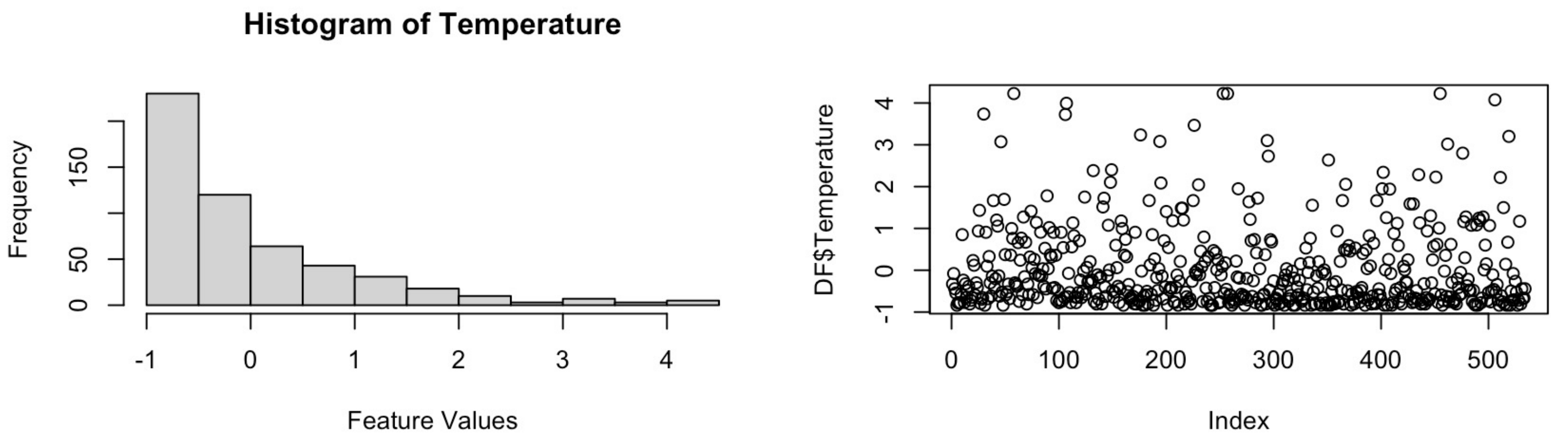
- Linear dependency among some predictors
- 4 main groups:
  - Visive + Somatosensory attributes
  - Auditory attributes
  - Temporal + causal attributes
  - Social + Emotion attributes

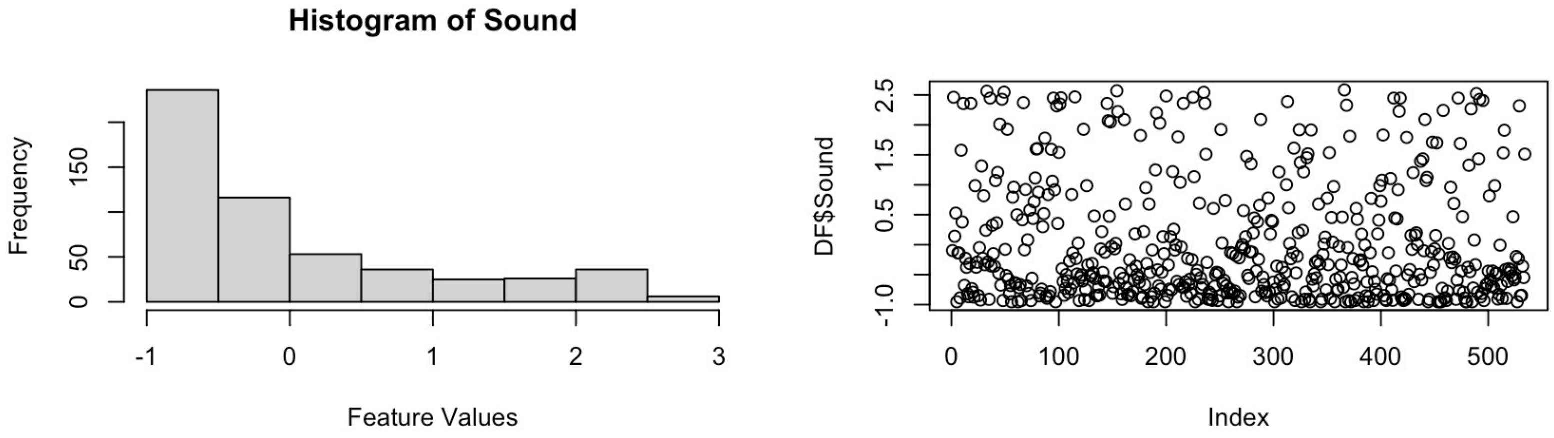


FEATURES ANALYSIS

FEATURES DISTRIBUTION

- Right-skewed behaviour
  - *Skewness_Mean : 1.33952*
- Presence of outliers
- Negative impact on next analysis



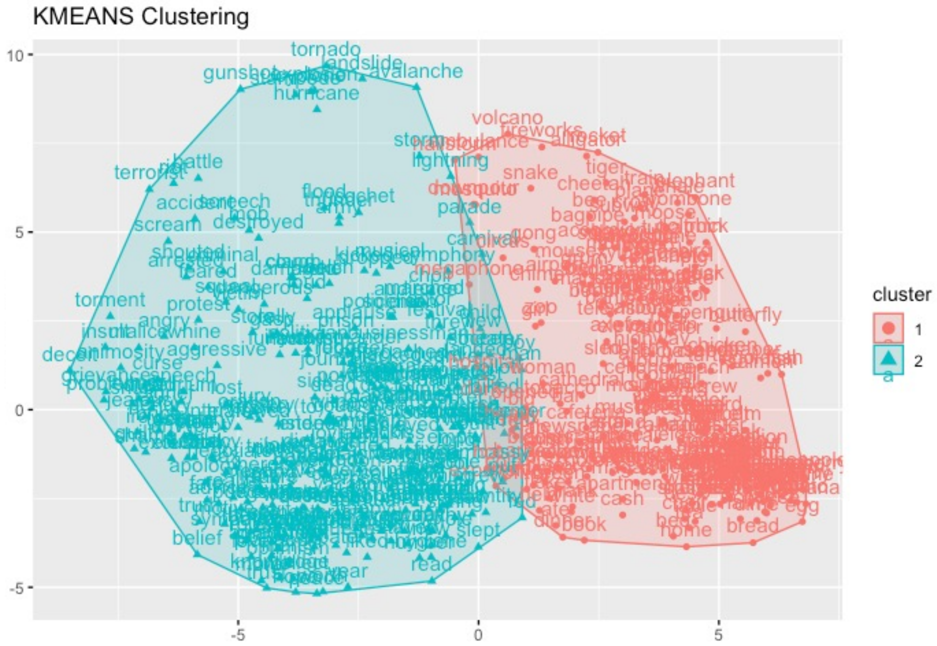


UNSUPERVISED LEARNING

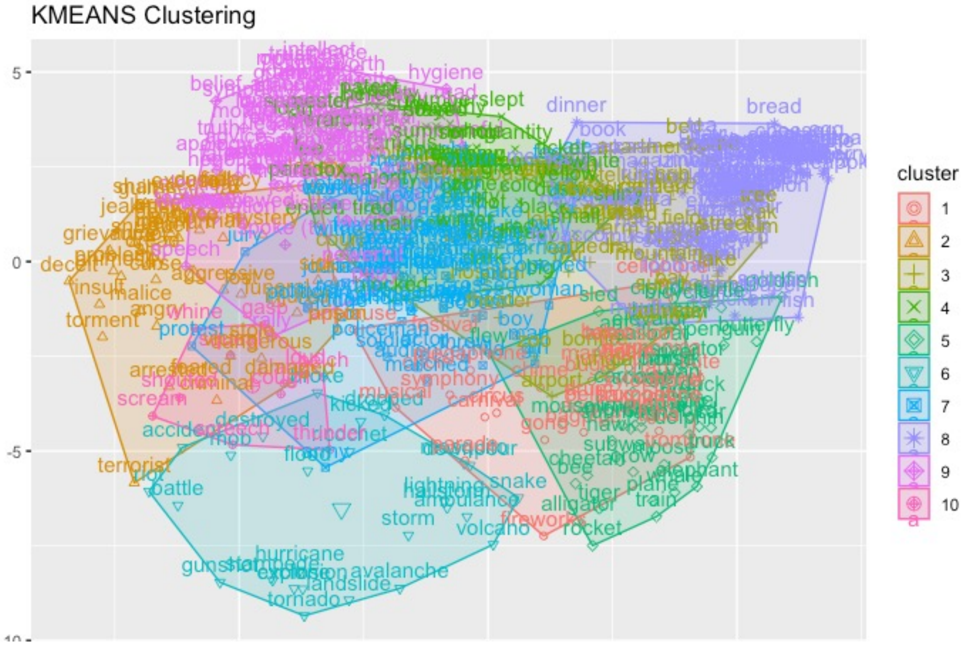
CLUSTERING

Can we identify (semantic) similarity structures?

- Search for optimal number of clusters
  - Elbow method/Hartigan Index : 2 clusters
  - AverageSilhouette : 10 clusters
- Balance between results and Silhouette info
- ***K-means***



- K =2
- Avg. Silhouette = 0.17
- Caotic distinction



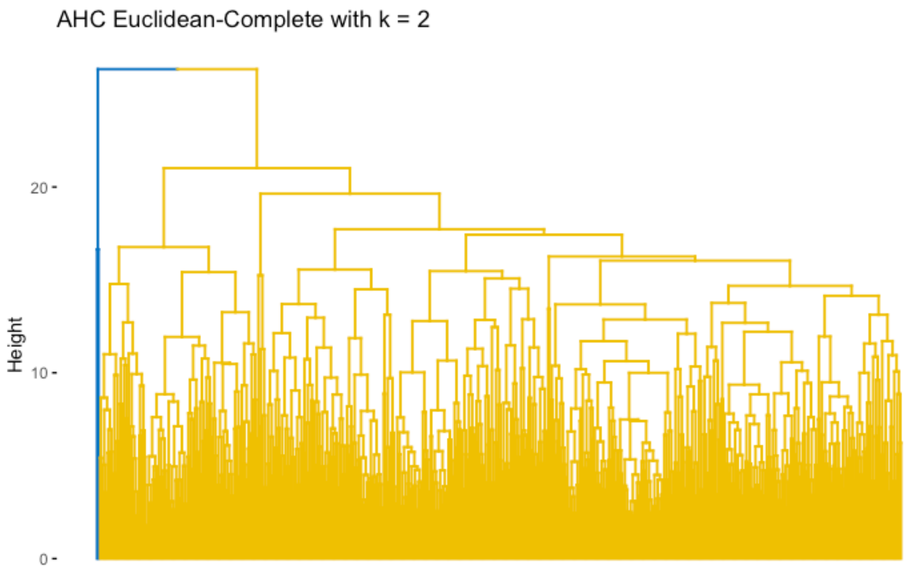
- **K = 10**
- **Avg. Silhouette = 0.20**
- Semantic classes (e.g. Animals, Food, Sentiment)
- Noisy distinction

UNSUPERVISED LEARNING

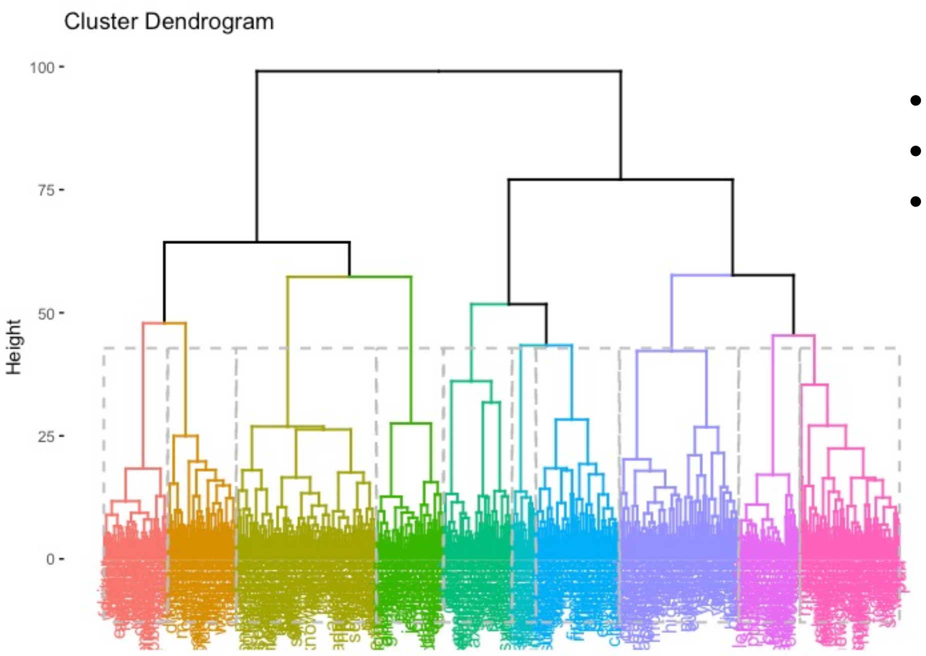
CLUSTERING

Can we identify (semantic) similarity structures?

- Search for optimal number of clusters
  - Elbow method/Hartigan Index : 2 clusters
  - AverageSilhouette : 10 clusters
- Balance between results and Silhouette info
- *K-means*
- ***Agglomerative Hierarchical Clustering***
  - **Complete linkage method**



- K =2
- Avg. Silhouette = 0.16
- Negative Vs. Positive/Neutral entities/events/properties



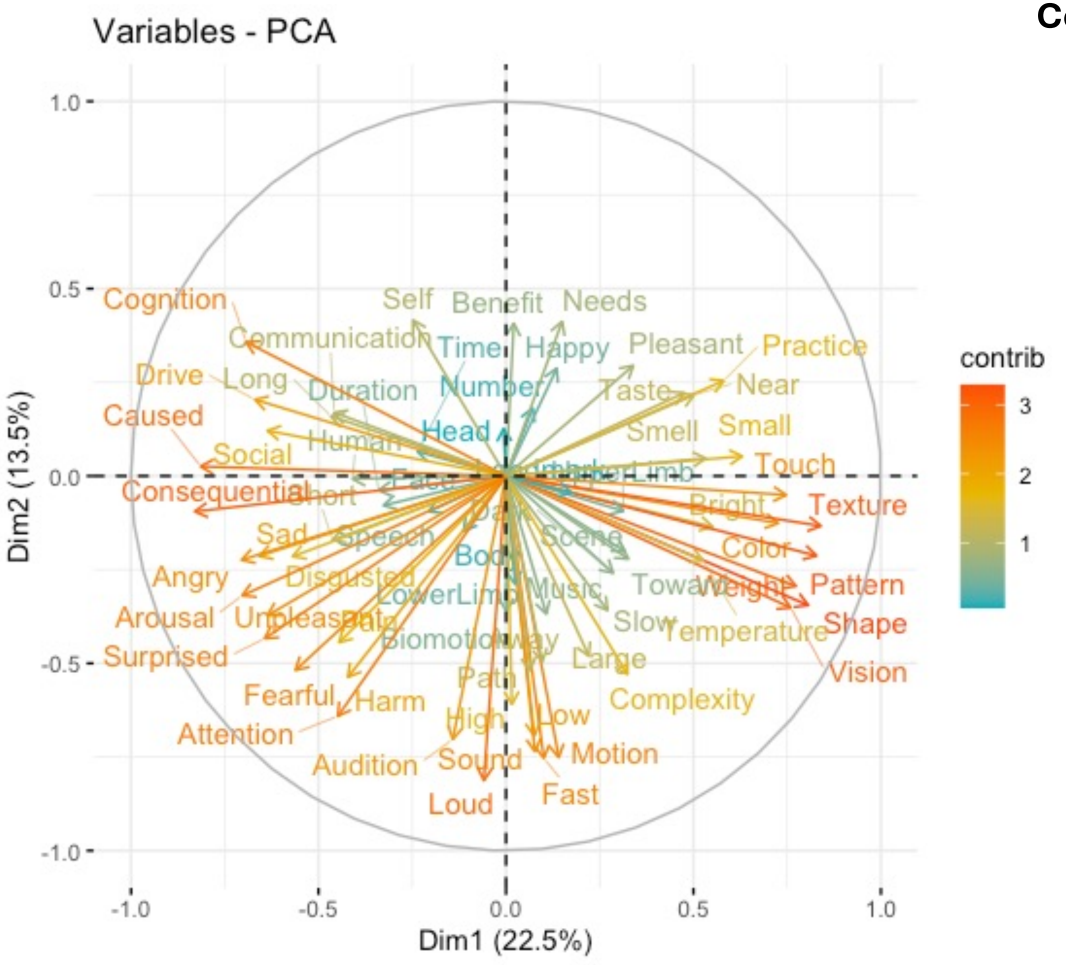
- **K = 10**
- **Avg. Silouette = 0.14**
- Shallow semantic classes
  (Qualitative better than kmeans = 10)
  - Food
  - Animals
  - Musical Instruments
  - Sentiment
  - …

UNSUPERVISED LEARNING

PCA

Can we identify (semantic) similarity structures?

- PCA as **denoising approach**?
  - 75% variance → 10 PCs
- **Took the first 2 PCs (36% cPVE)**
- **Features contribution**



Contribution of variables to

DIM.1	DIM.2
Texture	Fast
Pattern	Sound
Shape	Pain
Vision	Attention
Touch	Loud
Color	Complexity
Bright	Unpleasant
Smell	Fearful
Temperature	Music
....	....

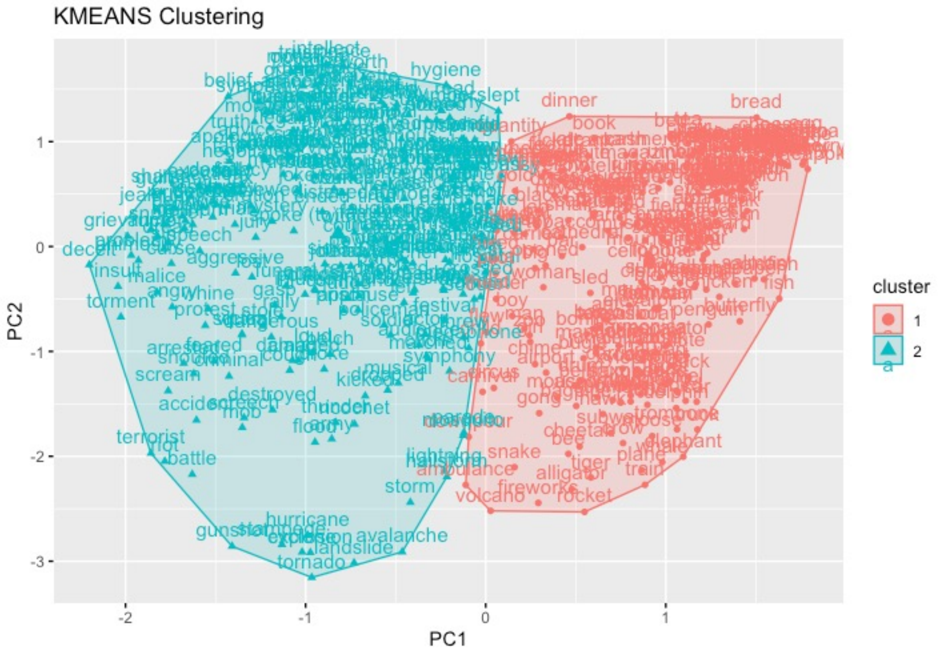
Dim.1 → **Concreteness**
Dim.2 → **Abstractness**

UNSUPERVISED LEARNING

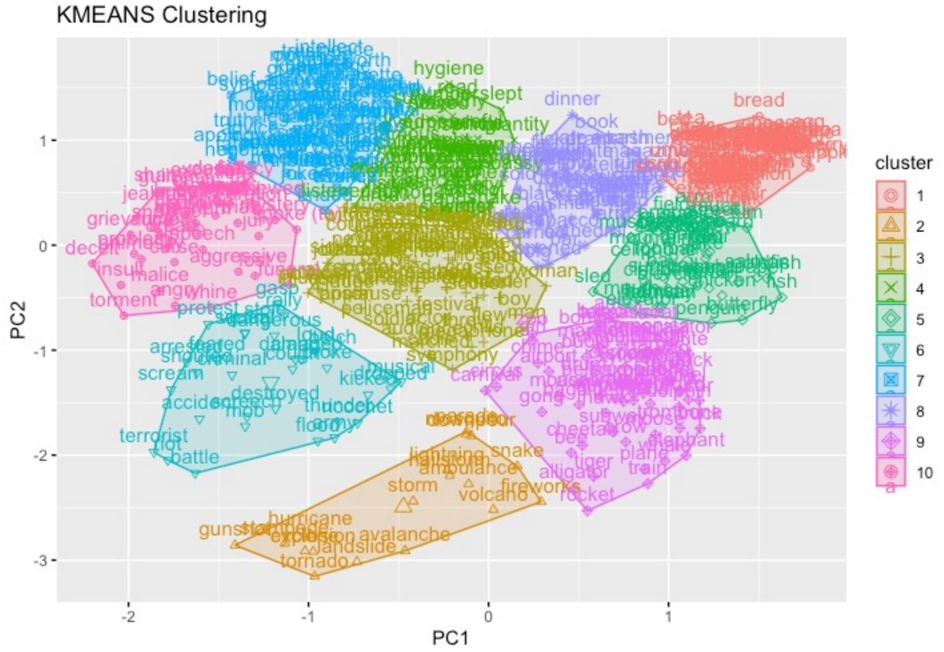
PCA

Can we identify (semantic) similarity structures?

- PCA as denoising approach?
  - 75% variance → 10 PCs
- Take the first 2 PCs (36% cPVE)
- Features contribution
- **Clustering on PCA**



- K = 2
- Avg. Silhouette 0.43
- Abstract Vs. Concrete (Good distinction)



- K = 10
- Avg. Silhouette 0.41
- Exhaustive semantic clusters

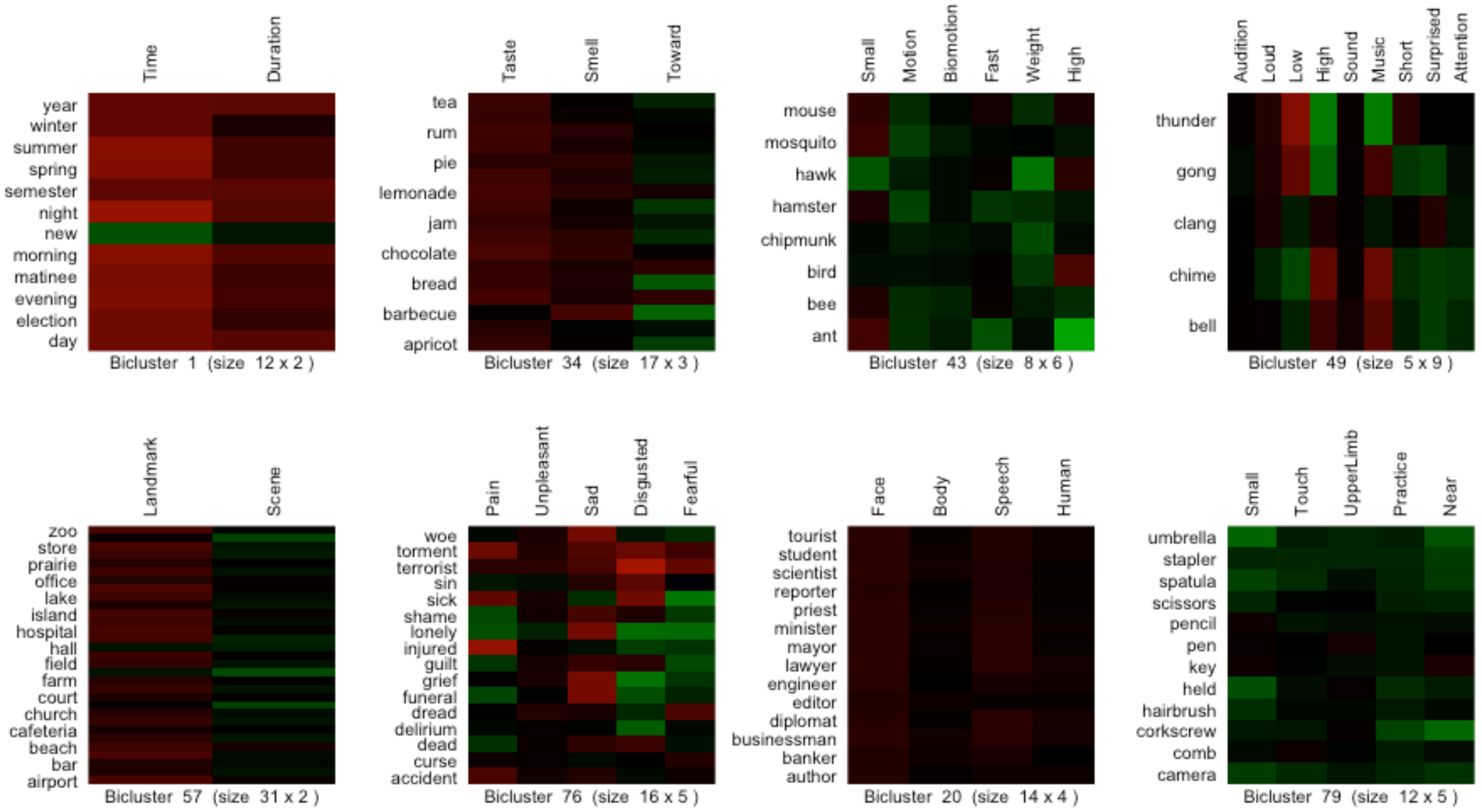
UNSUPERVISED LEARNING

CLUSTERING VS. **BICLUSTERING**

Can we identify (semantic) similarity structures?

- *Iterative Signature Algorithm* (isa2-package)*
- Find biclusters having correlated rows and columns
- **199 semantic clusters**

Time | Food and Drinks | Animals | Sounds



Locations | Negative Events And Feelings | Human Beings (Jobs) | Practical tools

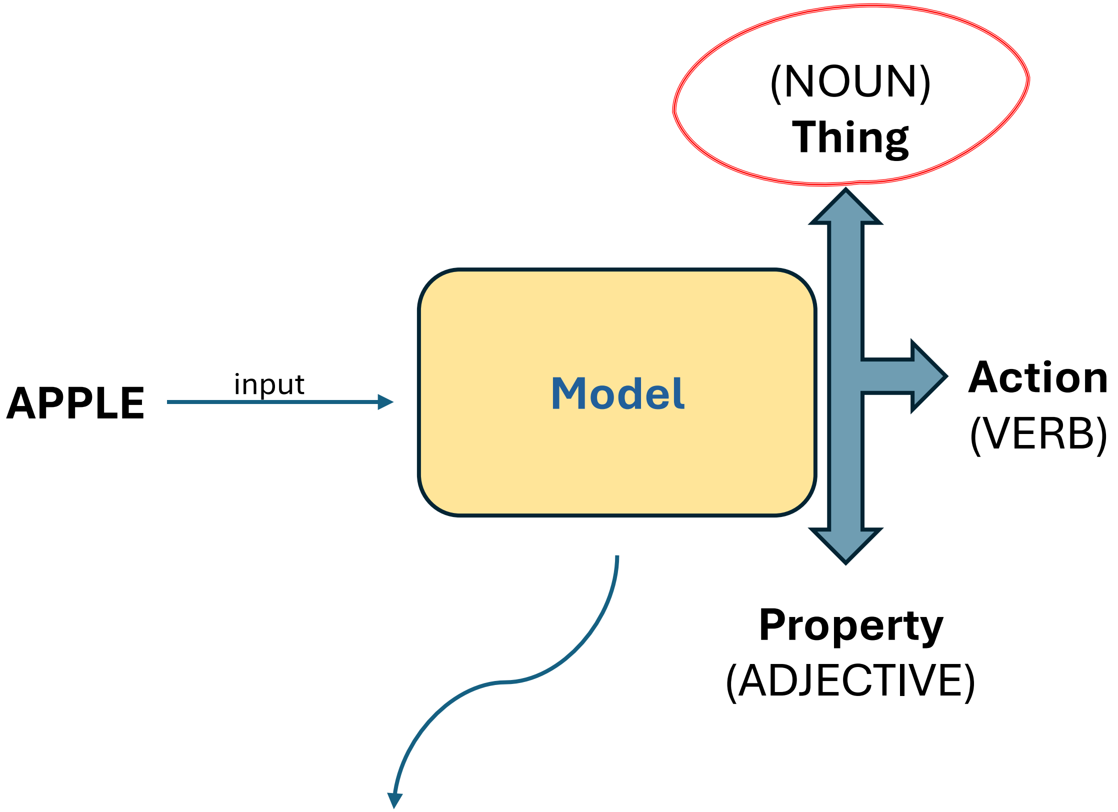
* (Bergman et al., 2003)

SUPERVISED LEARNING

CLASSIFICATION

Can we distinguish morpho-syntactic classes?

- GLMNet package
- **Logistic regression model**



Glmnet()

- `Family:` Multinomial
- `Alpha:` 1 → LASSO Penalty
- `type.multinomial:` Grouped

SUPERVISED LEARNING

CLASSIFICATION

Can we distinguish morpho-syntactic classes?

- GLMNet package
- Logistic regression model
- **First configurations:**
  - **Train/ Test → 70:30**

- **LASSO- 65 Features**
  - Accuracy: 0.84

Prediction	Reference: thing	action	property
thing	130	13	13
action	0	3	0
property	0	0	0

- **LASSO- 2PCs**
  - Accuracy: 0.82

Prediction	Reference: thing	action	property
thing	130	16	13
action	0	0	0
property	0	0	0

- **LASSO- 10PCs**
  - Accuracy: 0.82

Prediction	Reference: thing	action	property
thing	130	16	13
action	0	0	0
property	0	0	0

!

Unbalanced classes :
- Thing: 434
- Action: 56
- Property: 44

SUPERVISED LEARNING

CLASSIFICATION

Can we distinguish morpho-syntactic classes?

- GLMNet package
- Logistic regression model
- First configurations:
  - Train/ Test → 70:30
- **Trial&Error approach to solve classes unbalance**

1. Oversampling → Action/ Property X2
- Accuracy: 0.57
- Confusion Matrix (not good)
- Overfitting

2. CV + optimal λ → *cv_fit$lambda.1se*: 0.0069
- Accuracy: 0.94
- Confusion Matrix (quite better)
- Likely overfitting behaviour

3. a. Stratified CV → k=10
- Accuracy: 0.89
- Confusion Matrixes (good)
- Overfitting

b. Stratified CV – 2 / 10PCs → k =10
- Accuracy: 0.81/0.84
- Confusion Matrixes (good)
- Not overfitting

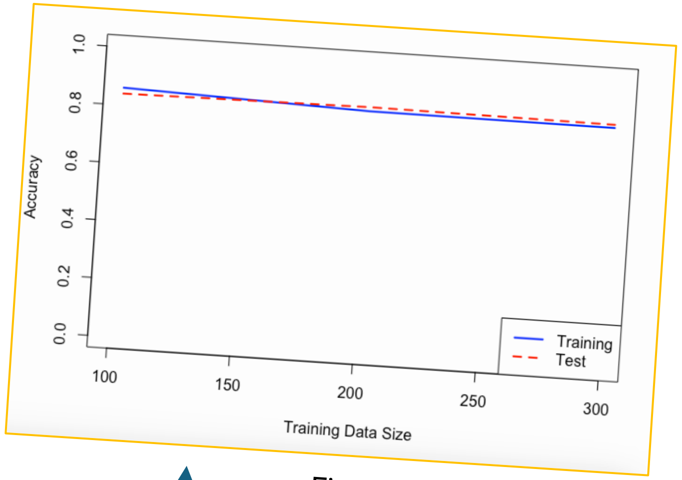


Figure: Accuracy curve

4. Change Data Partition → *Train/Test **60:40*** + optimal λ
- Accuracy: 0.95
- Confusion Matrix (balanced)

SUPERVISED LEARNING

FEATURE SELECTION

Can we distinguish morpho-syntactic classes?

- **(Soft)** Feature Selection
- **LASSO coeff** output
- **Features →** the **most important in PC1/ PC2**
- Fit the model again:
  - Train/ Test → **60:40**
  - **32 Features** (out of 65)
  - Cv + *lambda.1se*

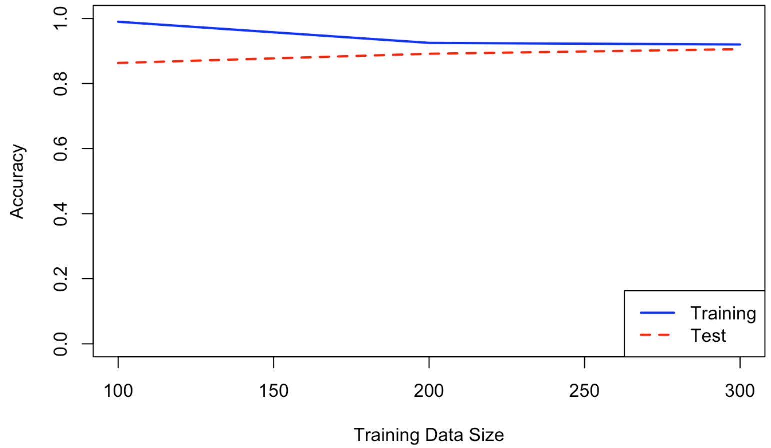
```
"Vision"      "Bright"         "Dark"         "Pattern"   "Large"
"Biomotion"   "Shape"          "Complexity"   "Body"      "Texture"
"Weight"      "Pain"           "Audition"     "Loud"      "Speech"
"UpperLimb"   "LowerLimb"      "Landmark"     "Near"      "Toward"
"Number"      "Time"           "Short"        "Caused"    "Consequential"
"Social"      "Communication"  "Cognition"    "Harm"      "Sad"
"Surprised"   "Needs"
```

Figure: Most important features (LASSO Output)

Model

Accuracy: 0.91

```
            Reference
Prediction  thing action property
  thing       168      7        6
  action        4     15        1
  property      1      0       10
```



CONCLUSION

- Statistical tools **positively confirmed goodness** of this componential (semantic) approach
  - Features **capture** a great quantity of **semantic + relational information**
  - Identification of (macro-)semantic classes = understand word meaning
  - Distinction among lexical classes (nouns vs. verbs vs. adjectives)

- PCA
  - Useful as denoising approach
  - Unveil an important **latent information dicotomy** in the construction of meaning

- **Concrete vs. Abstract** as a **primitive tool for meaning construction (?)** → Future Directions
  - **Improve** complex semantic distinctions (e.g., Food vs. Time vs. Animal)
  - **Contribute** indirectly to lexical class identification (see Feature selection)

Thank you for the attention!