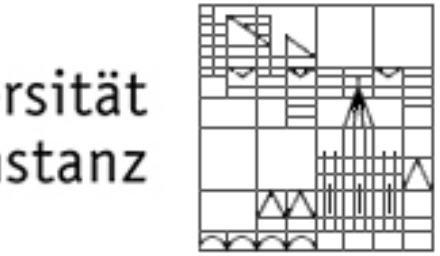
Universität Konstanz



LingVis: Motivation and Use Cases

Miriam Butt & Mennatallah El-Assady

LingVis: Visual Analytics for Linguistics
DGfS 2018 | 6.3.2018

SPONSORED BY THE

Federal Ministry of Education and Research

DFG Deutsche Forschungsgemeinschaft

Tutorial schedule

- 10:00 am – 11:30 am: Motivation and Use Cases
  - Examples of existing visualizations relevant for LingVis.
- 11:45 am – 1:00 pm: Visualization Theory
  - Goals and principles of Visual Analytics
  - Cognitive foundations
- 2:00 pm – 5:00 pm: Hands-on
  - Intro to existing visualization tools/frameworks
  - Work with existing visualizations

Before we start….

Personal Questions

- Who are we?
- Who are you?
  - Programming Background
  - What types of linguistic questions interest you?
  - What is the interest in LingVis?

Motivation and background

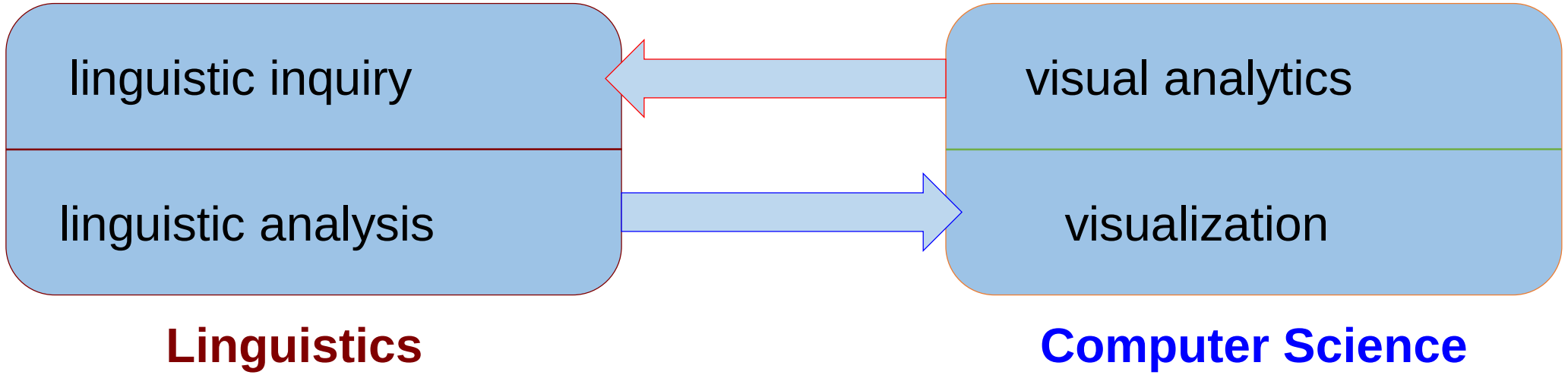
Slides based on the DGfS 2016 and Konvens 2016 Tutorials on LingVis and HistoBankVis 2017 talk.

Contributors: Miriam Butt, Oliver Deussen, Mennatallah El-Assady, Annette Hautli-Janisz, Wolfgang Jentner, Daniel Keim, Thomas Mayer, Frans Plank, Christian Rohrdantz, Dominik Sacha, Christin Schätzle, Rita Sevastjanova

LingVis

Overall Goals:

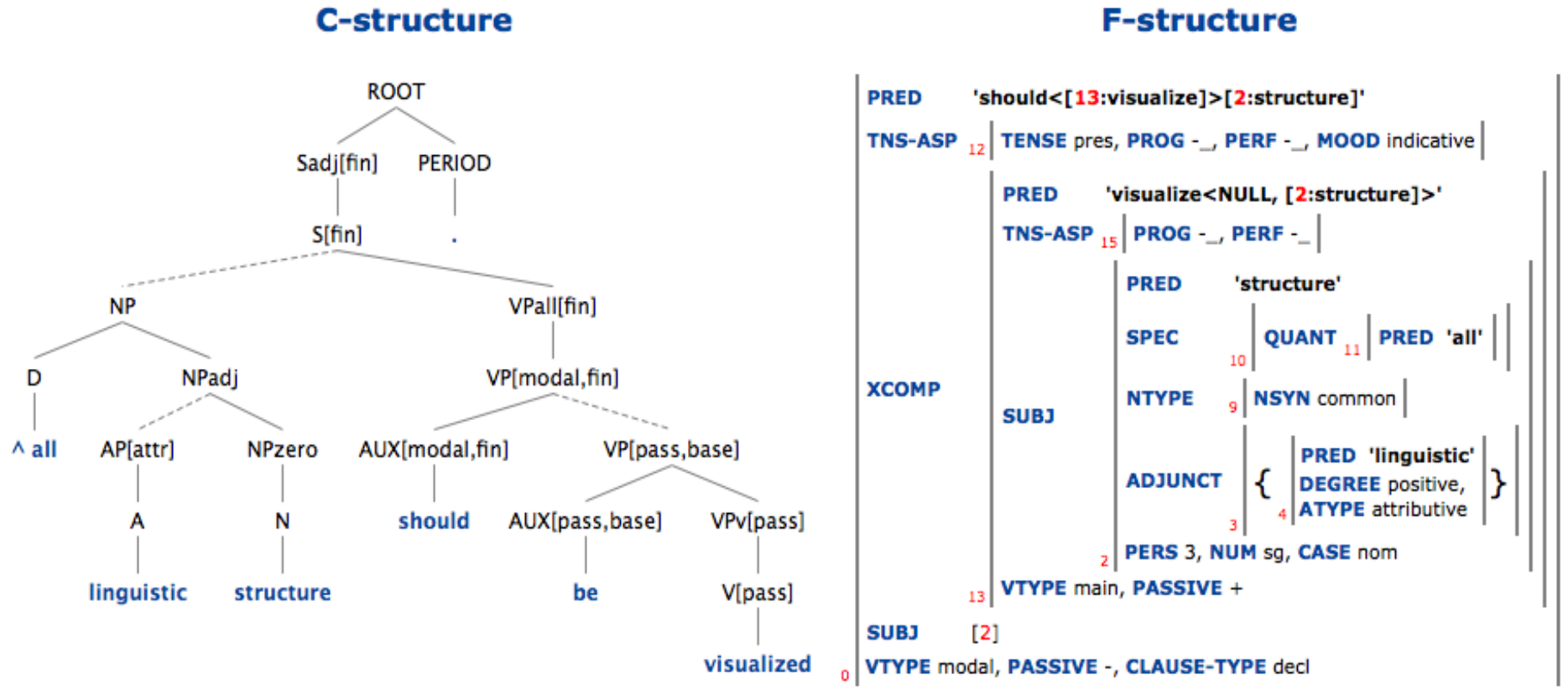
- Integrate methods from **visual analytics** into domains of **linguistic inquiry**.
- Explore challenges based on the needs of **linguistic analysis** for **visualization methods**.



Visual Analytics

- Interactive, exploratory access to data
- Iterations of hypothesis formation and hypothesis testing
- Overview first – details on demand
  - Holistic picture
  - But can drill down to individual data points
  - Abstract representation of multifactorial, multidimensional data.
  - Good for understanding complex interactions in the data.

Existing Visualization: Syntax

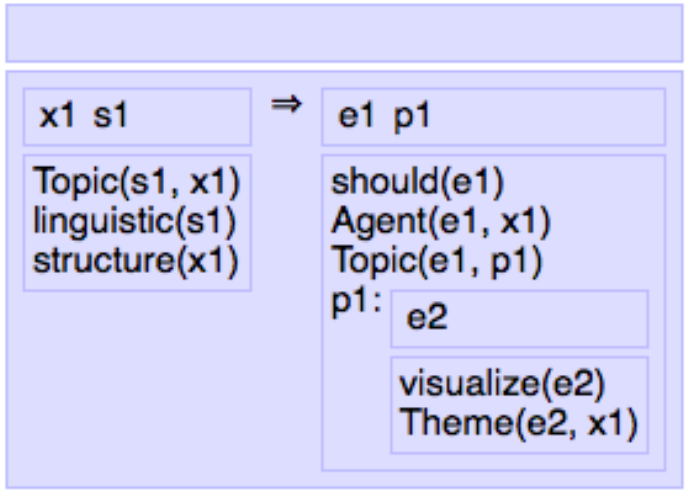


- Syntactic Analysis with Lexical-Functional Grammar (LFG)
- http://clarino.uib.no/iness/xle-web (Web Interface for LFG Grammars)
- Grammar developed at PARC (Palo Alto Research Center)

Existing Visualization: Semantics

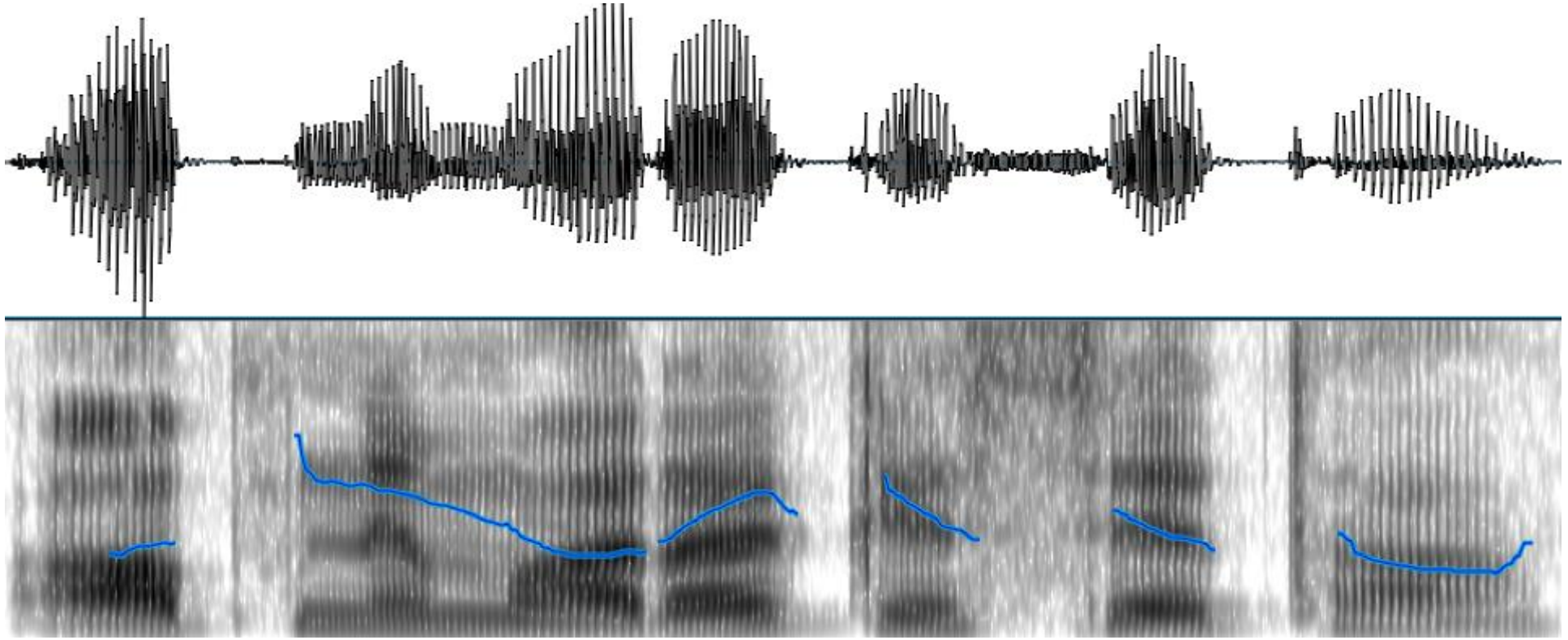
DRS

(Discourse Representation Structure)



- Semantic Analysis with Discourse Representation Theory (Boxer)
- http://gmb.let.rug.nl/webdemo/demo.php (web interface for CCG/DRT)
- Grammar/Semantics developed by Johan Bos and colleagues (Groningen)

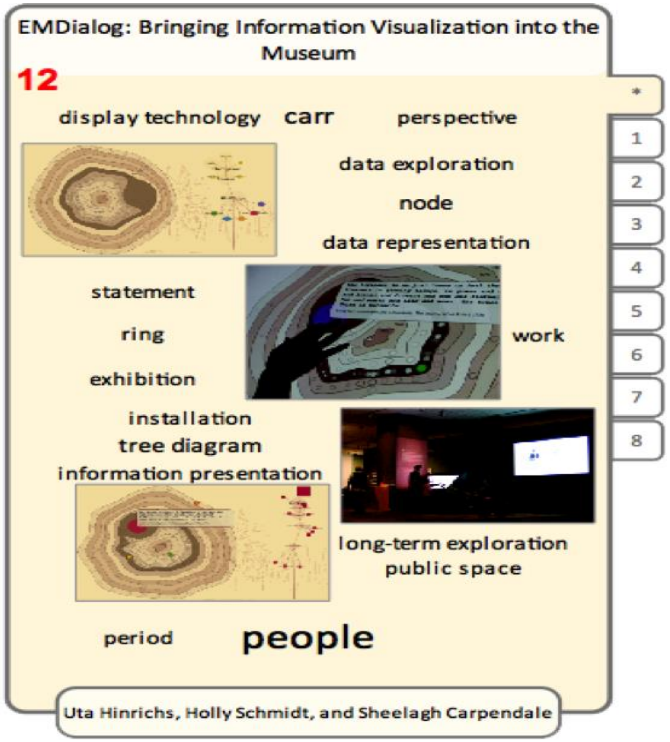
Existing Visualization: Phonetics

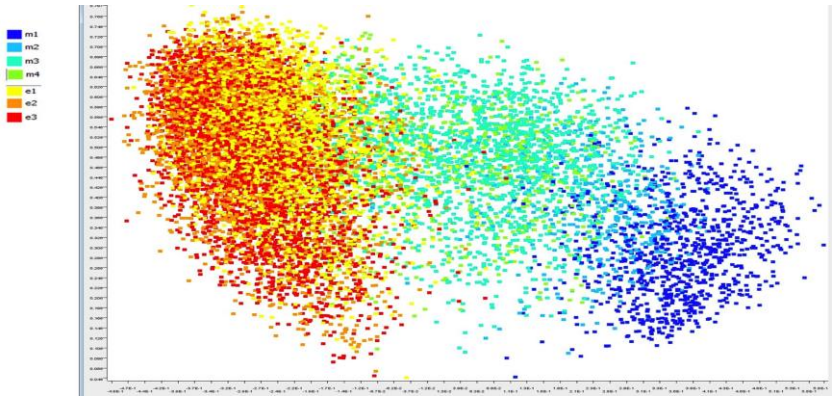


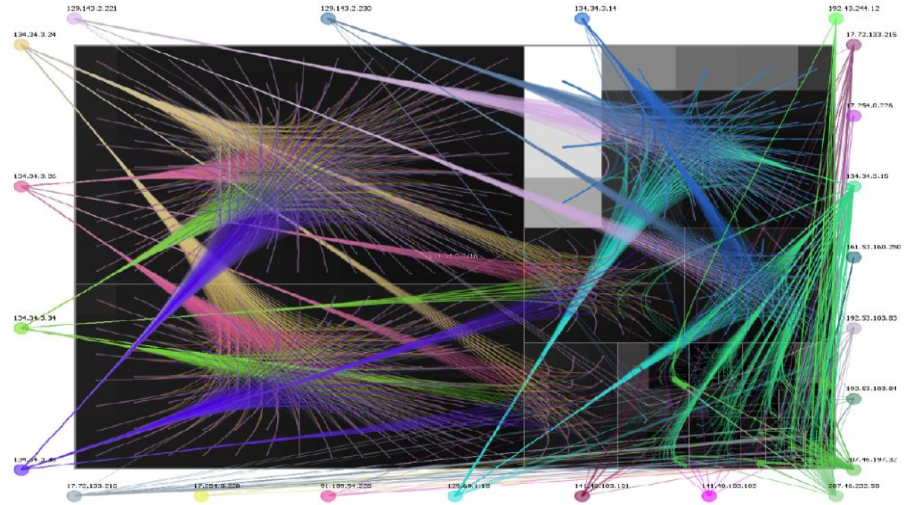
- aap=ne mara kis se thaa

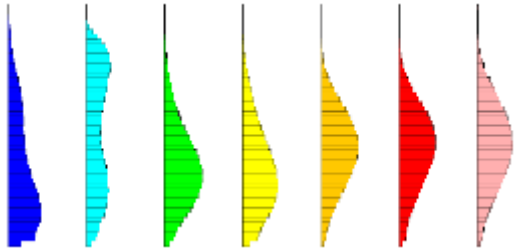

Les IV DECLINAISONS
Nouvelle GRAMMAIRE à l'usage des DAMES
CONJUGAISON de tous les Verbes réguliers
INFINITIF
EXACTEUR ou IMPERATIF
AVOIR
ETRE

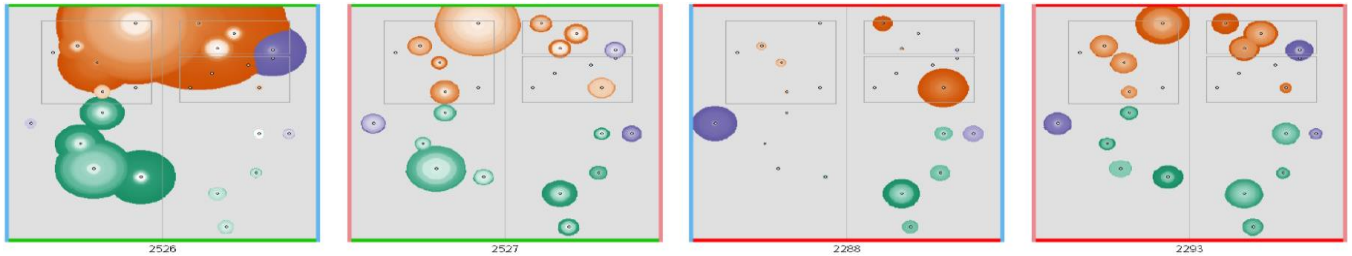
Sample Visualizations

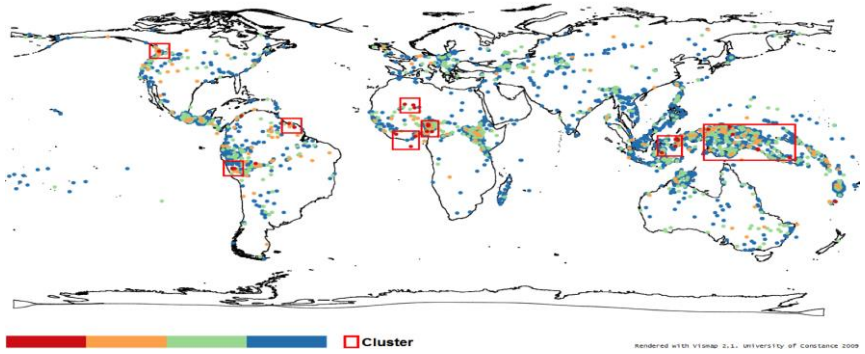












LingVis – Motivation

- Linguists are making more and more use of newly available technology to detect **distributional patterns** in language data.
- Ever increasing availability of **digital corpora** (synchronic and diachronic).
- Increasing interest in language output produced in **social media**.
- Ever better **query** and **search tools** (CQP, COSMAS, DWDS, ANNIS).
- **Programming languages** suitable for text processing, statistical analysis and visualization (e.g., Python, R).

Making Sense of Numbers

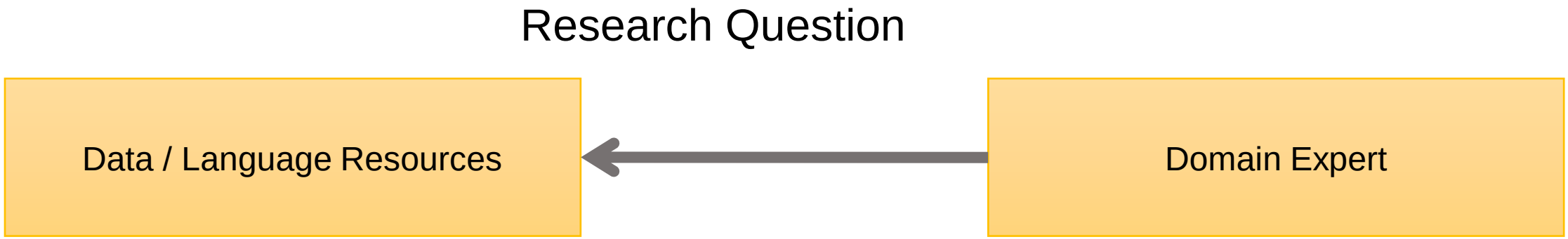
- Current linguistics often includes **corpus work**.
- Linguists try to determine patterns, interactions and usage preferences within a language but also across different languages.
- This work generates a lot of numbers (statistics).
- Numbers are difficult for humans to process.
- Solution: translate **numbers** into **visual properties.**

→ Human visual apparatus can process this easily.

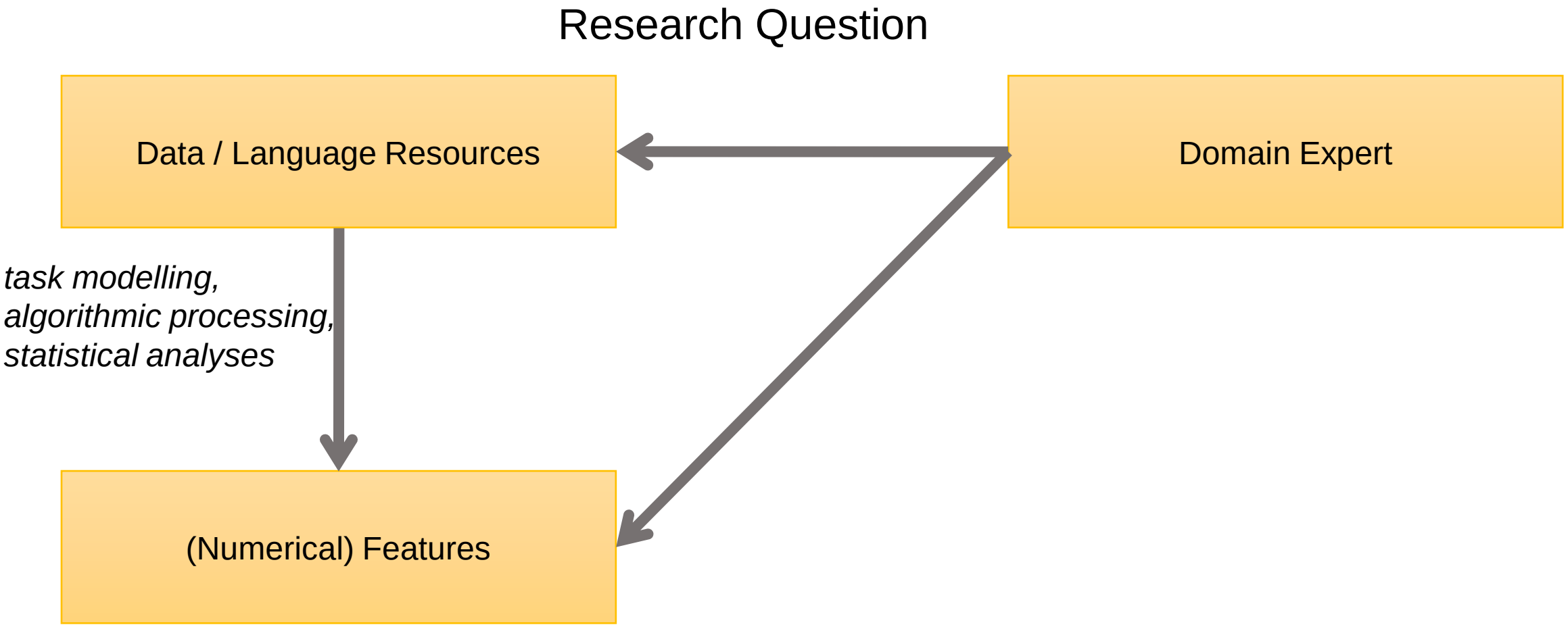
Visualization in Linguistics

- Visualization is part and parcel of linguistics.
- However, very little advantage is taken of new visualization techniques.
- Newest ones borrowed from statistical analysis (bar charts, scatter plots, etc.), but often do not do justice to complexity of linguistic data.
  - Multifactorial/Multidimensional (including temporal)
  - Complex interactions
  - Different data types
  - Metadata
- **Our Work**: Identify and explore good use cases within linguistics

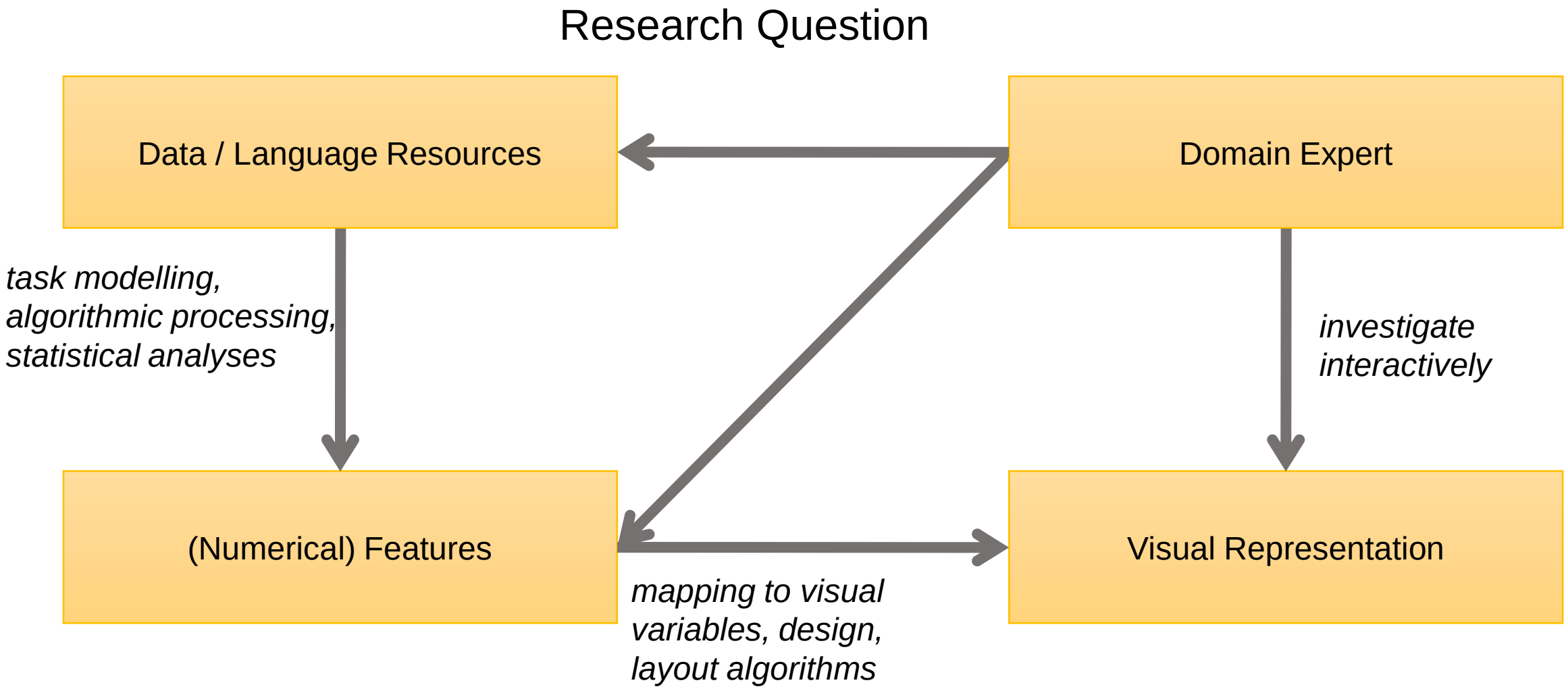
Interdisciplinary Collaboration: LingVis



Interdisciplinary Collaboration: LingVis



Interdisciplinary Collaboration: LingVis



Example: Pixel-Based Visualizations

Two Use Cases

- N-V Complex Predicates
- Vowel Harmony

N-V Complex Predicates

- **N-V complex predicates** occur very frequently in Urdu.
- **Examples:** phone-do, memory-do, memory-become, resolution-do, resolution-be, ...
- **Problem:** would be nice if one knew which nouns were likely to co-occur with which verbs.

Example: N-V Complex Predicates in Urdu

- **Goal:** identify sequences of Noun+Verb for understanding complex predicate patterns
  - *phone-do, use-do, memory-come, begin-do/come*
- **Data:** 7.9 million word raw (unannotated) corpus of Urdu (BBC Urdu)

```
#this file lists X in X+kar, X+ho, X+hu, X+rakh sequences with corr...
esponding occurrences in the (candidate) CP sequences
#X = word occurring directly to the left of LV (LV: kar, ho, hu, rakh)
#kar: # of occurrences of X with kar
#ho: # of occurrences of X with ho
#hu: # of occurrences of X with hu
#rakh: # of occurrences of X with rakh
X          #hu  #kar #ho  #rakh
اضافہ      674  466  524       0
غورش       378  2336 1691      0
معلوم      366  254  609       0
شکایت      359  135  44        0
حملہ       227  1232 100       0
متاثر      183  178  765       0
ناقص       173  0    114       0
کیا        172  373  7027      0
ثابت       147  394  588       0
تقو        142  105  235       9
پیدا       103  754  956       0
کالہ       102  1501 3609      0
آرام      80   210  96        0
اکثر       74   0    263       0
تجزیہ      62   59   1161      0
آغاز       59   315  75        0
ہی         56   0    2267      0
قدغن       54   197  262       0
انکشاف     51   165  13        0
```

Butt, Miriam, Tina Bögel, Annette Hautli, Sebastian Sulger & Tafseer Ahmed. 2012. **Identifying Urdu Complex Predication via Bigram Extraction**. In *Proceedings of the 24th International Conference on Computational Linguistics* (COLING), 409–424. Mumbai, India.

Example: Pixel Visualization

Statistical Data:

ID	Noun	Rel. freq. with *kar*	Rel. freq. with *ho*	Rel. freq. with *hu*	Rel. freq. with *rakh*
1	حاصل	0.771	0.222	0.007	0.000
2	اعلان	0.982	0.011	0.007	0.000
3	بات	0.853	0.147	0.000	0.000
4	شروع	0.530	0.384	0.086	0.000

Table 2: Relative frequencies of co-occurrence of nouns with light verbs

do, be, become, put

kar, ho, hu, rakh

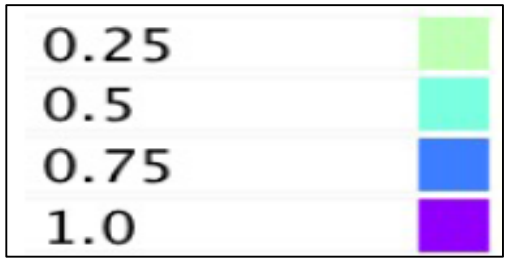
'achievement'

'announcement'

'talk'

'beginning'

Color Scale

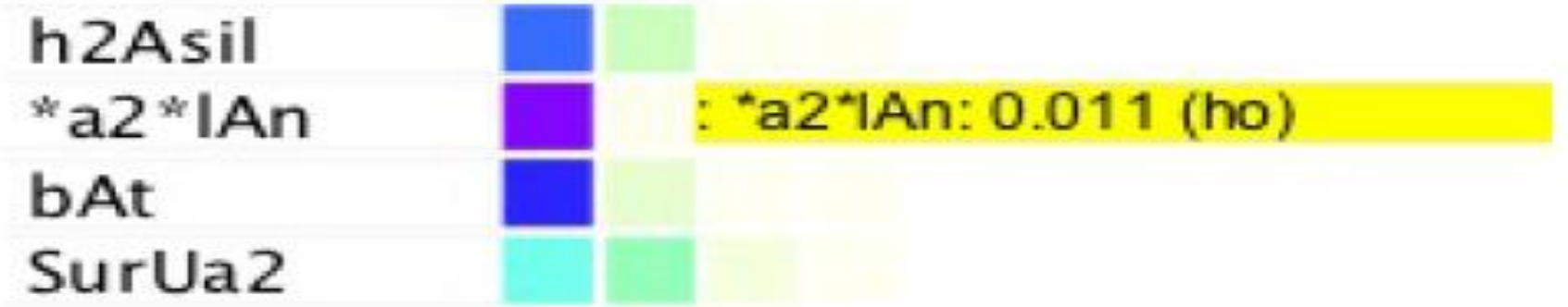


Pixel plus Cluster Visualization

- Performed k-means clustering combined with a pixel visualization.
- Advantages:
  - can inspect clusters visually and detect patterns
  - Outliers spotted easily (mostly errors – “kyA” is not a noun, it is a *wh*-word and was included by mistake).

Example: Identifying N-V complex predicates in Hindi/IUrdu

Tool facilitates zooming and mousing over to see the underlying data set



Outliers/Errors are easily identified (Clustering Algorithm has applied)

Vowel Harmony (VH)

- **Phenomenon (simplified):** Vowels in affixes change according to vowels found in stems.
- **(Famous) Example: Turkish**

	Front		Back	
	Unrounded	**Rounded**	**Unrounded**	**Rounded**
High	i	ü	ı	u
Low	e	ö	a	o

Genitive suffix
deniz-**in**, ev-**in**
tütün-**ün**, çöl-**ün**
kadın-**ın**, adam-**ın**
sabun-**un**, top-**un**

Genitive suffix with plural suffix
deniz-ler-**in**, ev-ler-**in**
tütün-ler-**in**, çöl-ler-**in**
kadın-lar-**ın**, adam-lar-**ın**
sabun-lar-**un**, top-lar-**un**

Vowel Harmony

Goal: Try to determine automatically whether a given language contains patterns indicative of vowel harmony.

Basic Computational Approach:

- Use written corpus (caveat: only approximates actual phonology).
- **Count** which vowels succeed which other vowels in $VC^{+}V$ sequences (within words — again an approximation)
- Through **statistical analysis** find out the association strength between vowels: normalized association strength value ϕ.
- **Results** show that Turkish and Hungarian, for example, pattern similarly. Languages like Spanish or German pattern differently.

Results — Standard Methods: Can you detect a pattern?

	a	ı	u	o	ö	ü	i	e
a	0.266	0.427	-0.141	-0.060	0.019	-0.125	-0.261	-0.275
ı	0.162	0.292	-0.107	0.077	-0.010	-0.075	-0.190	-0.191
u	0.129	-0.143	0.464	0.017	-0.003	-0.051	-0.138	-0.140
o	0.066	-0.112	0.434	-0.015	0.006	-0.045	-0.104	-0.111
ö	-0.107	-0.092	-0.052	-0.026	0.006	0.366	-0.091	0.164
ü	-0.120	-0.114	-0.059	0.014	-0.006	0.507	-0.112	0.134
i	-0.201	-0.224	-0.118	0.071	-0.004	-0.087	0.319	0.211
e	-0.256	-0.251	-0.132	-0.062	-0.010	-0.097	0.400	0.276

Turkish

	a	i	o	e	u
a	-0.003	-0.075	0.094	-0.025	-0.018
i	-0.025	-0.004	0.064	-0.036	0.005
o	-0.028	-0.006	-0.075	0.098	0.026
e	-0,001	0.063	-0.073	0.016	0.021
u	0.077	0.038	-0.036	-0.057	-0.043

Spanish

	a	o	i	ü	ö	ä	u	e
a	0.019	0.009	-0.061	-0.034	-0.008	-0.025	0.018	0.035
o	-0.023	-0.004	-0.052	-0.013	-0.020	-0.013	-0.013	0.068
i	-0.069	-0.054	-0.050	-0.039	-0.036	-0.044	-0.003	0.133
ü	-0.067	-0.045	0.070	-0.028	-0.021	-0.033	-0.021	0.050
ö	-0.049	-0.032	0.049	-0.024	-0.013	-0.021	-0.013	0.036
ä	-0.067	-0.037	0.124	-0.033	-0.018	-0.028	-0.038	0.020
u	0.012	-0.018	-0.019	0.046	-0.002	-0.013	0.004	-0.001
e	0.108	0.084	0.026	0.069	0.063	0.096	0.021	-0.195

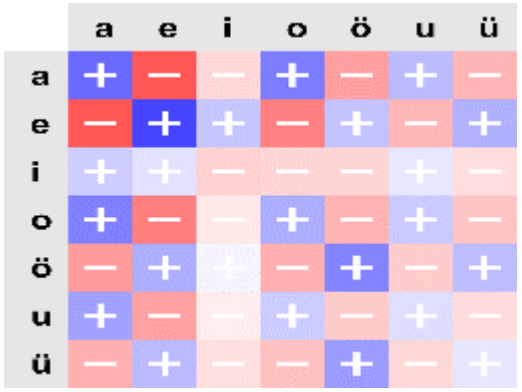
German

	a	o	u	i	ü	ö	e
a	0.339	0.263	0.070	-0.022	-0.081	-0.136	-0.431
o	0.239	0.099	0.041	-0.007	-0.052	-0.083	-0.253
u	0.132	0.038	0.015	-0.004	-0.017	-0.040	-0.131
i	0.037	-0.026	0.008	-0.030	-0.017	-0.027	0.011
ü	-0.093	-0.056	-0.022	-0.014	0.008	0.148	0.071
ö	-0.152	-0.093	-0.037	0.001	0.065	0.229	0.097
e	-0.435	-0.241	-0.076	0.048	0.091	0.054	0.531

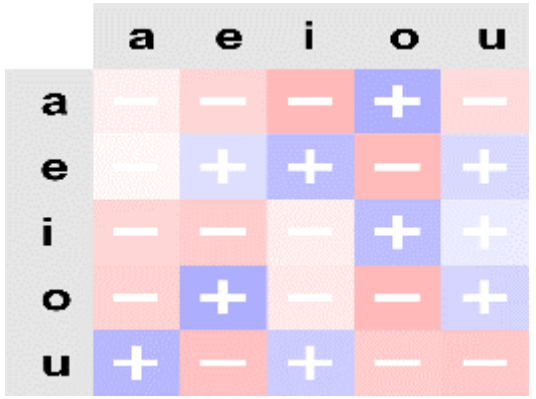
Hungarian

First Simplistic Visualization: Can you detect a pattern?

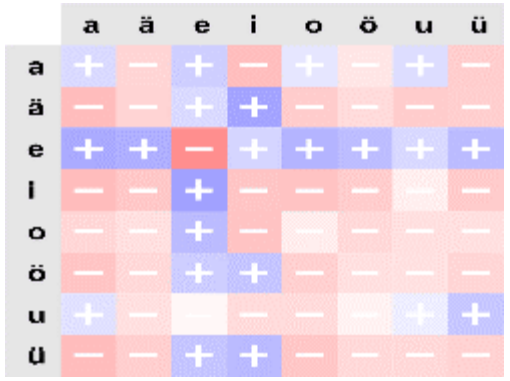
Turkish



Hungarian



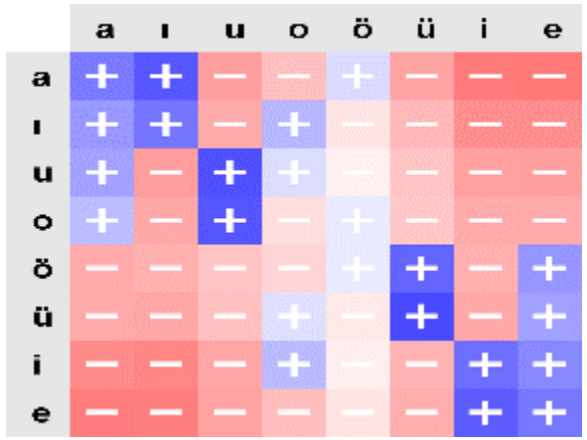
Spanish



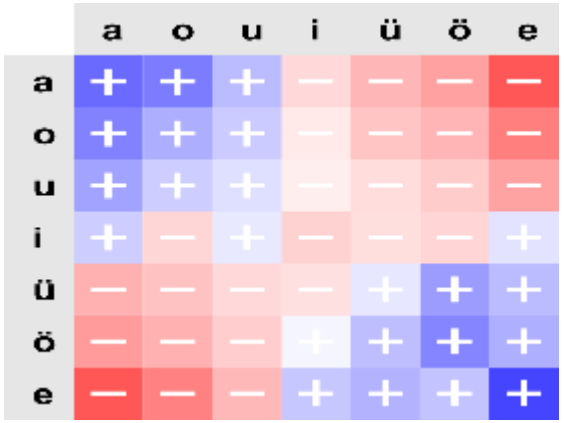
German

- Matrix visualization of association strengths between vowels (deviation from statistical expectation).
- Vowels are sorted alphabetically.
- More saturated colors show greater association strength.
- Blue is for more frequently than expected, red for less.
- The +/– are redundant encodings.

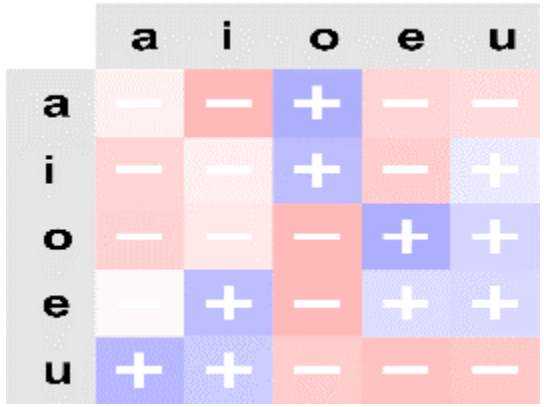
Sorted Visualization: Can you detect a pattern now?



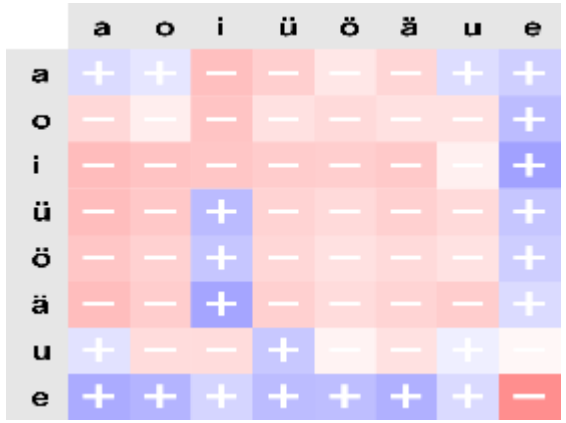
Turkish



Hungarian



Spanish



German

Vowels **sorted** according to similarity (note: not a trivial process)
Can even see the **type** of Vowel Harmony involved.

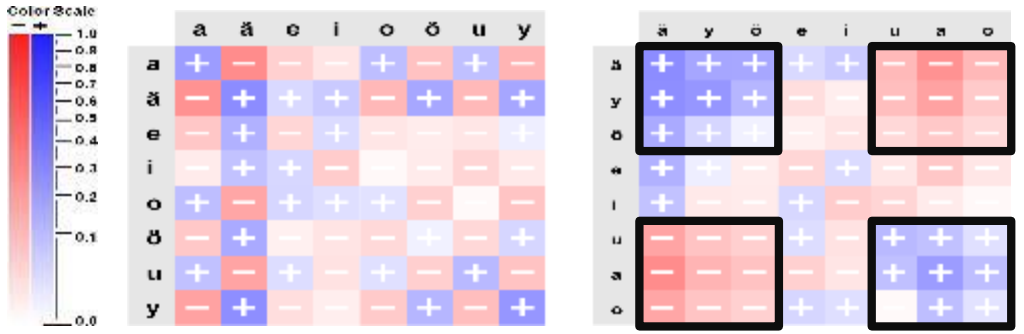
T. Mayer, C. Rohrdantz, M. Butt, F. Plank and D. A. Keim. **Visualizing Vowel Harmony**. *Linguistic Issues in Language Technology*, 4(Issue 2):1-33, 2010.

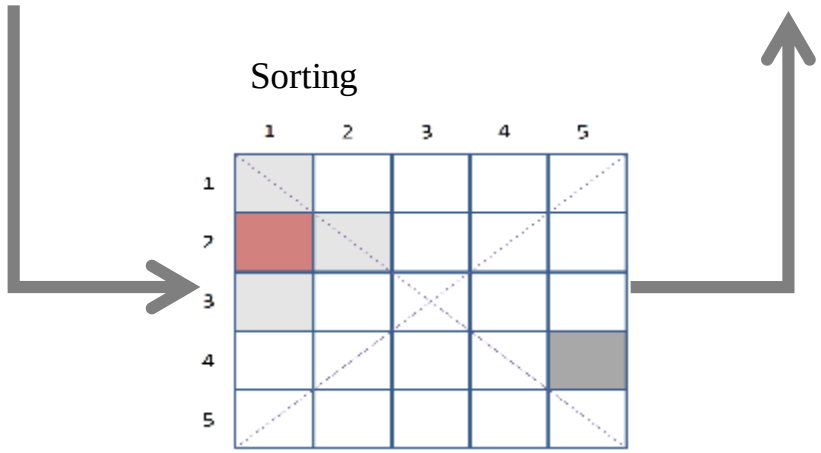
Visualizing Vowel Harmony

Counting Vowel Successions in all Bible Types
Example: Finnish

	a	ä	e	i	o	ö	u	y
a	3548	20	1940	1893	831	0	944	24
ä	35	944	806	820	10	138	33	266
e	1623	1144	1495	1608	419	56	497	187
i	1580	854	1514	1044	376	46	355	135
o	1384	7	1032	902	284	0	294	8
ö	7	125	54	39	0	3	1	18
u	1464	6	1085	850	315	1	547	8
y	39	656	368	368	35	75	4	251

Statistics & Visualization



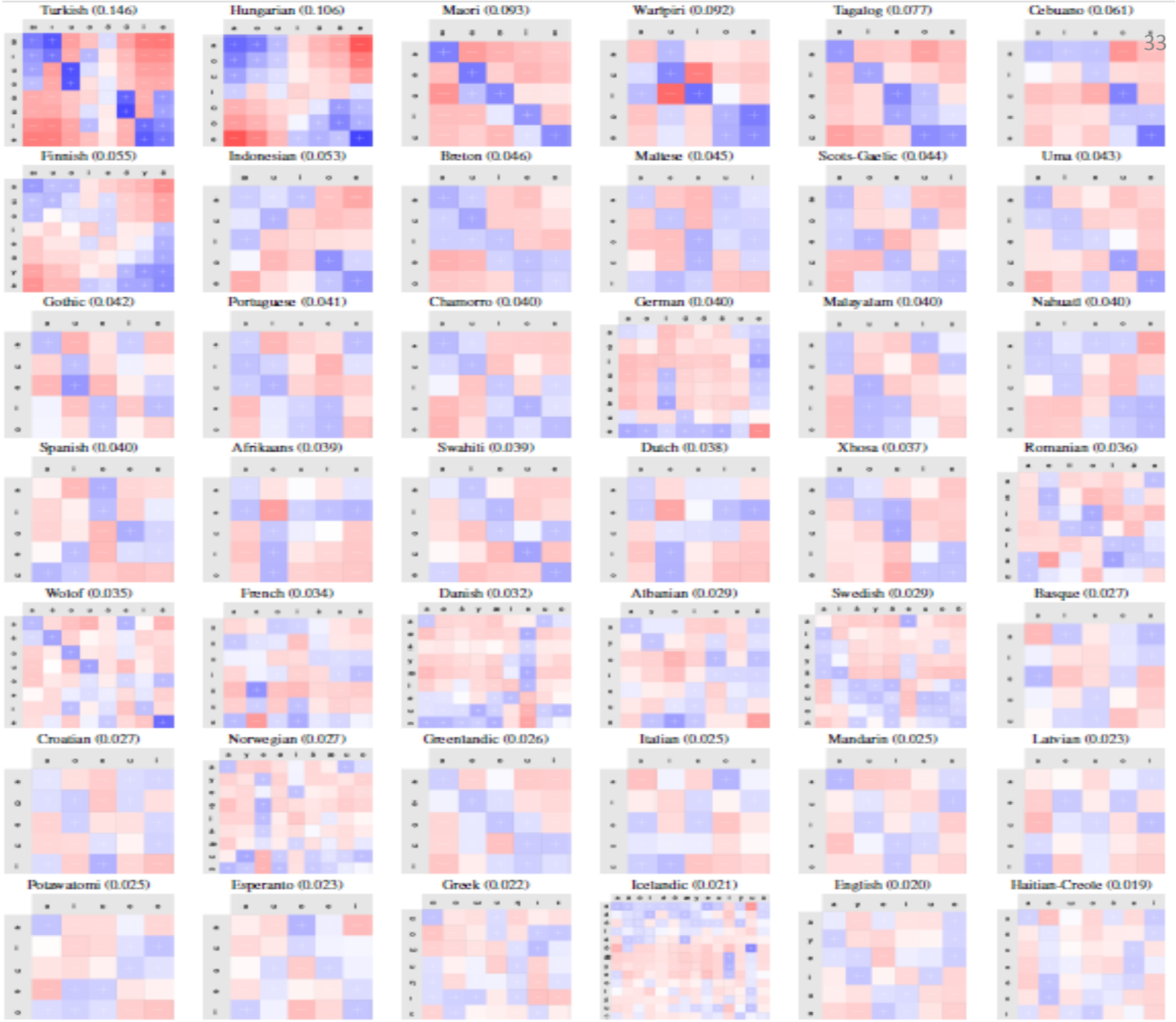


Sorting done according to feature vectors of each of the rows.

[9]

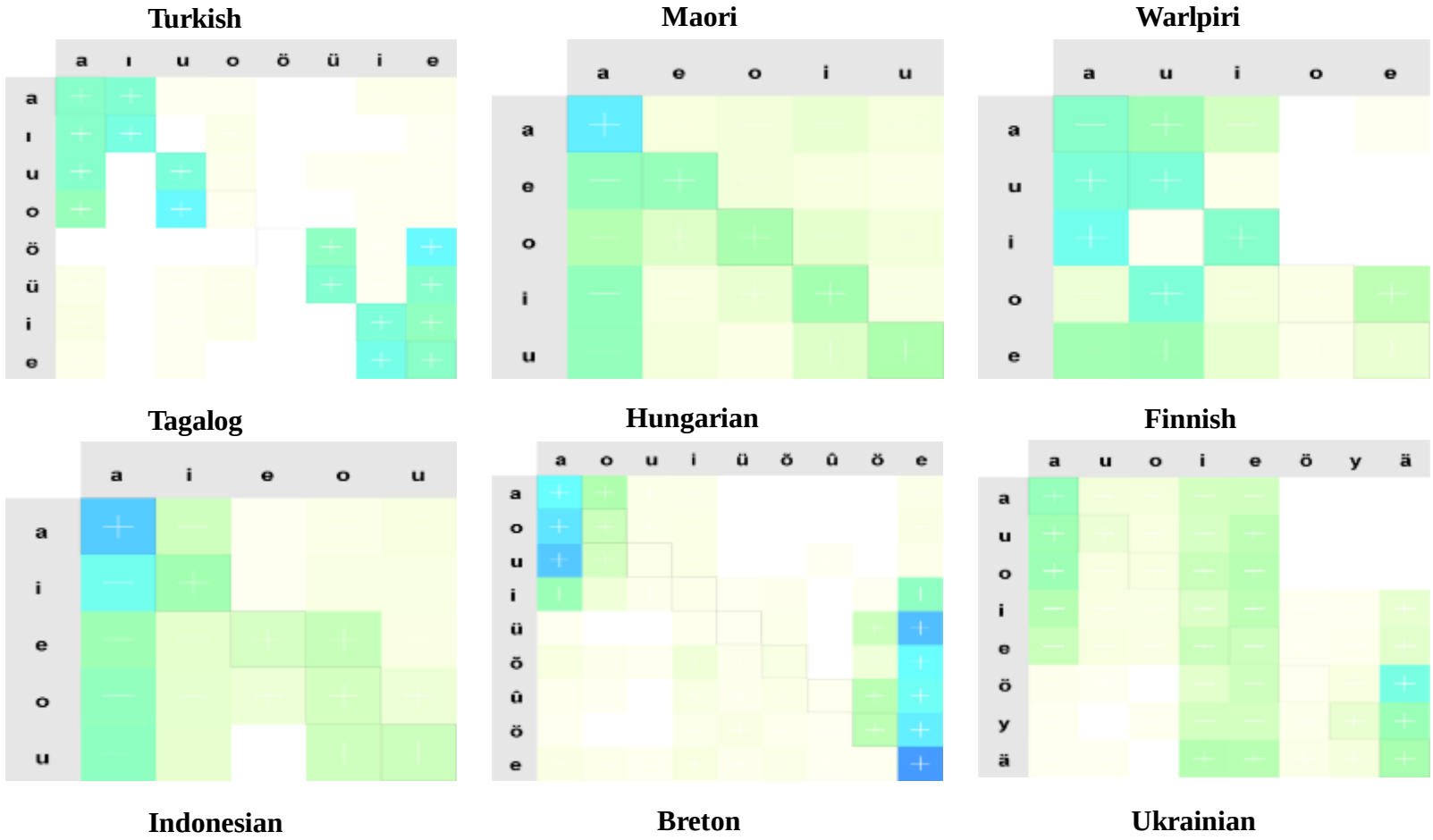
Results –
Sorted Visualization:

- Automatic Visual Analysis of vowel successions
for 42 languages – sorted for effect strength.



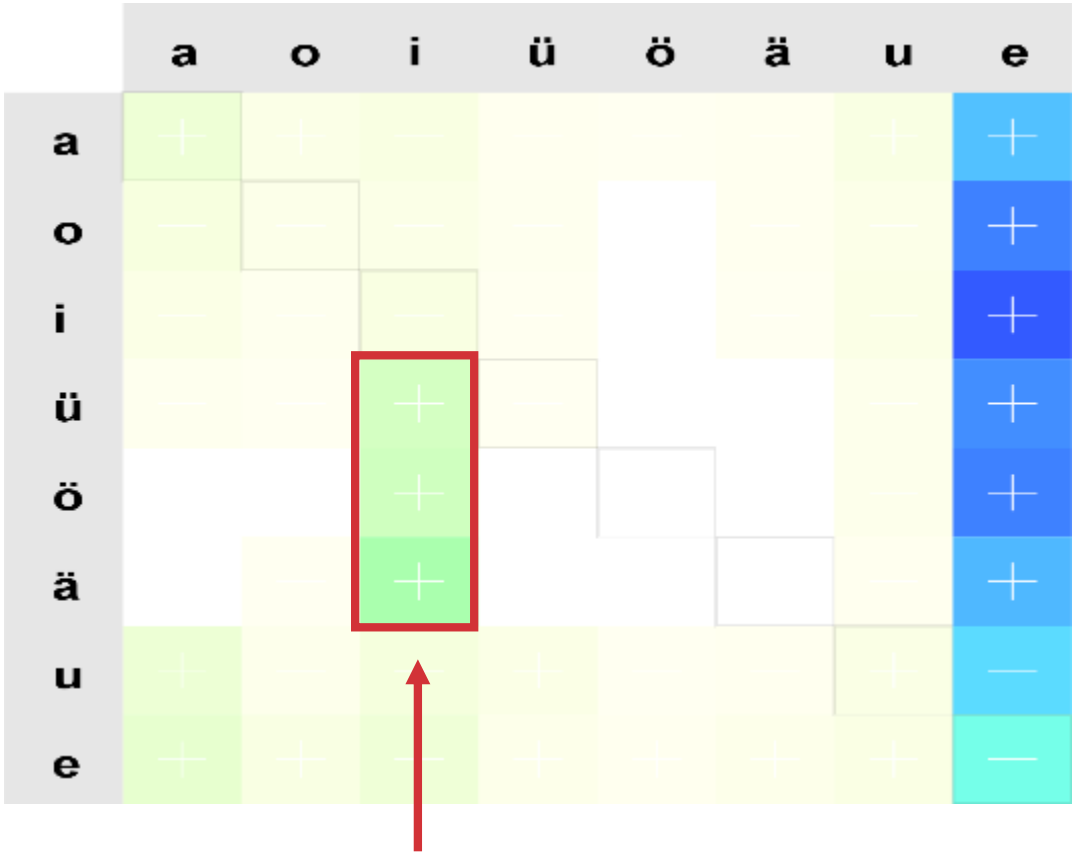
Vowel Harmony vs. Reduplication

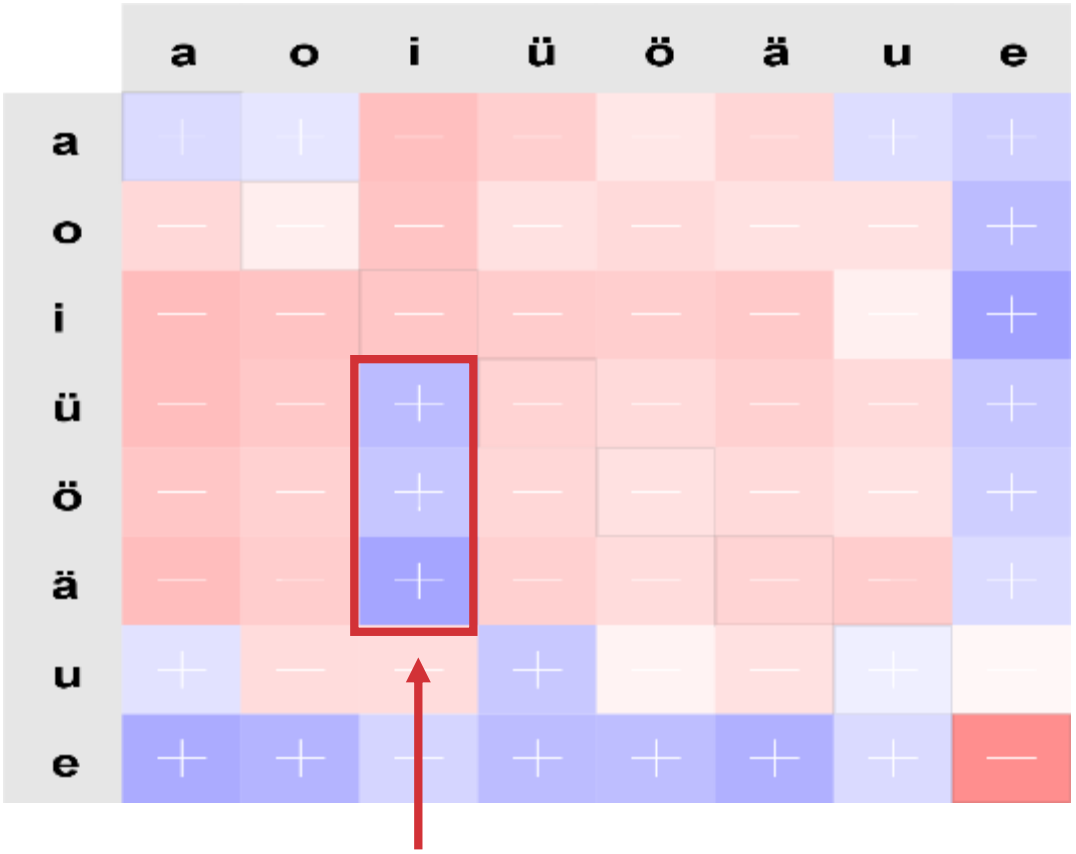
- In VH languages, crucially there are some vowels which never co-occur.
- This can be seen via a calculation of succession probabilities.
- Maori is not a VH language.



Historical Fingerprint: German Umlaut

Even though Umlaut (raising of vowel in stem before high vowel in affix) is no longer a productive process in German, the Umlaut harmony pattern is still visible in the matrices.





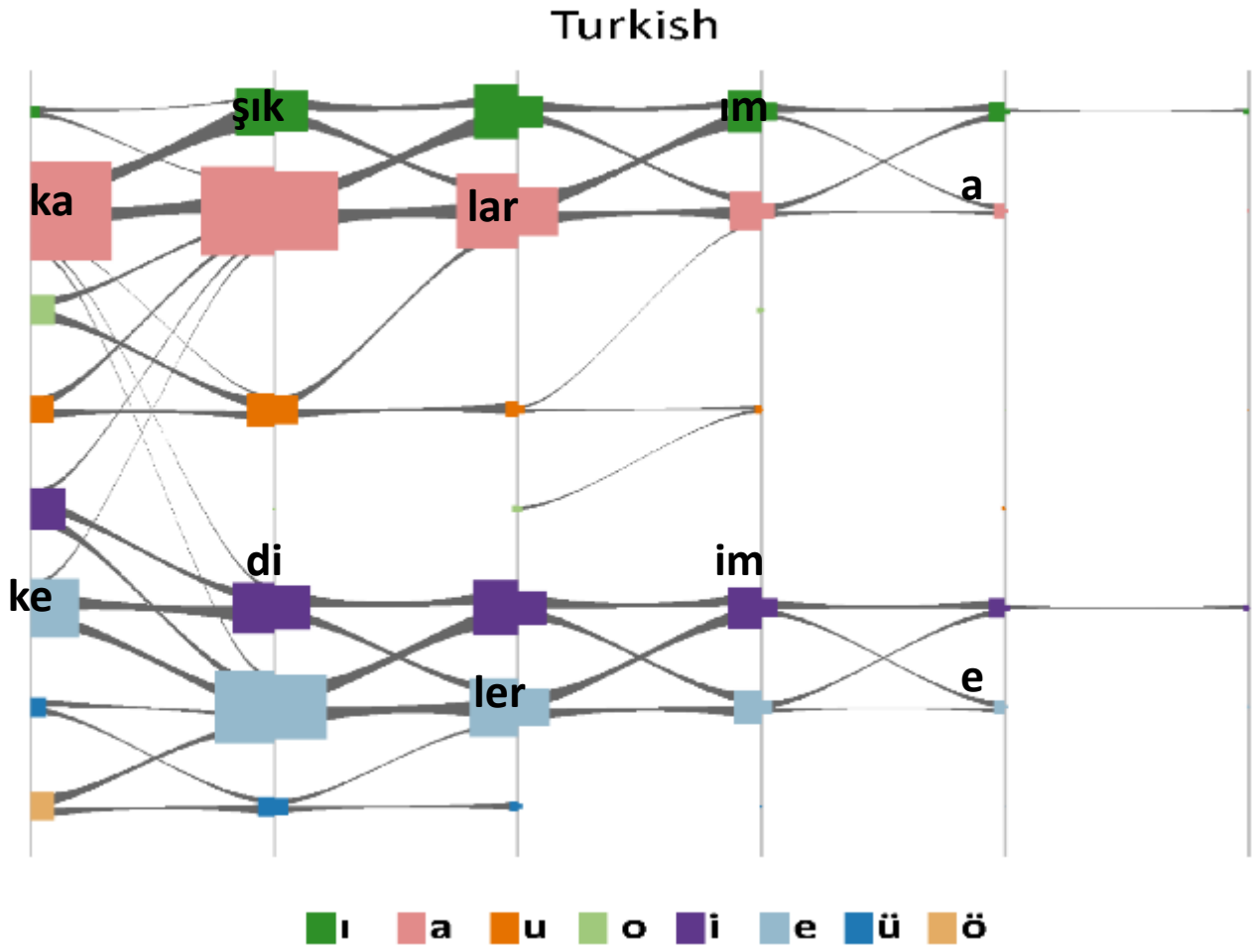
Example: Droplet Visualizations

- Different Types of Visualizations can be used to look at the same data.
- Example: Droplets for Vowel Harmony
- This droplet technique was originally used for rendering geospatial information (an item moving from one place to the next).

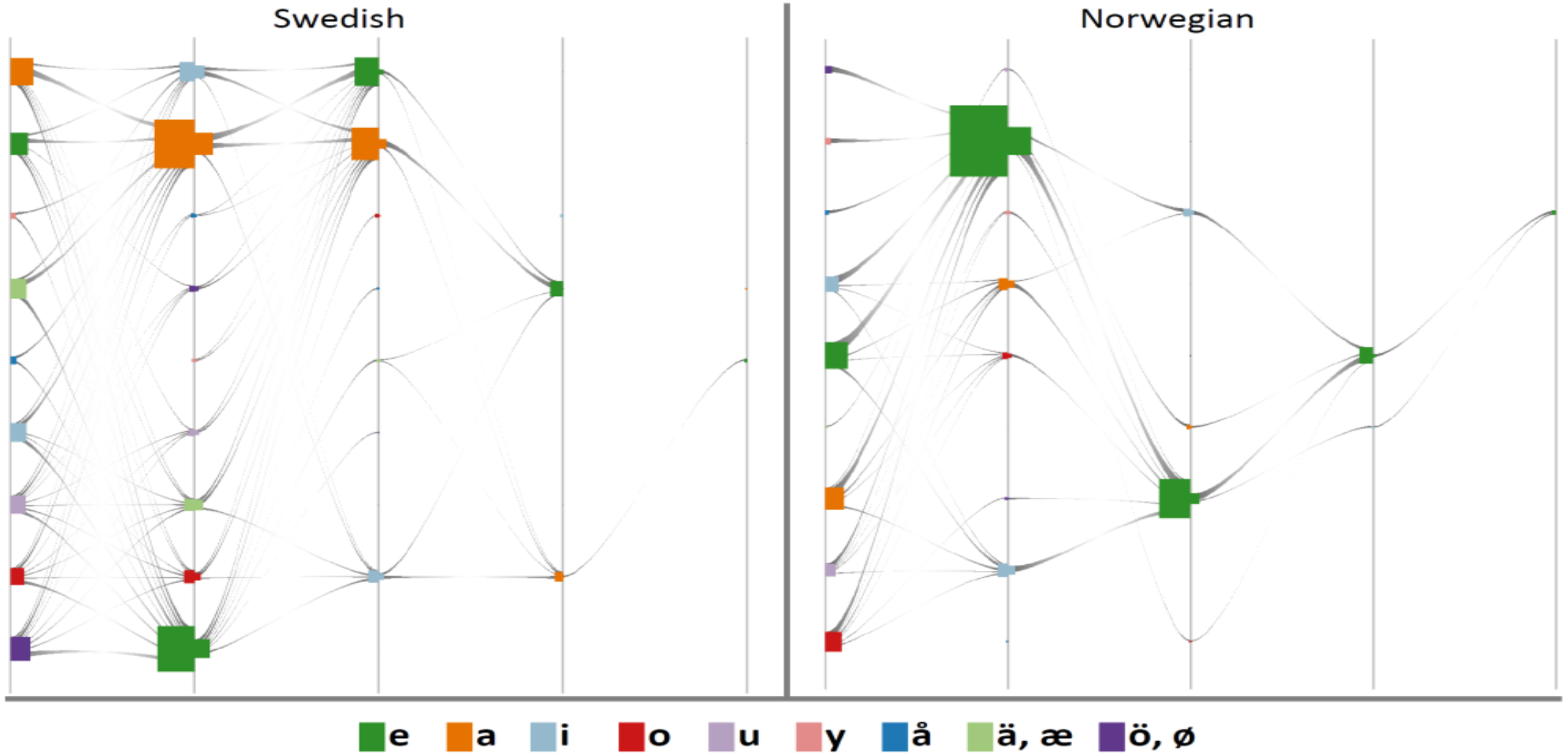
Vowel Harmony via Droplets

kaşık-lar-ım-a
spoon-Pl-1SgPoss-Dat
‘my spoons’

kedi-ler-im-e
cat-Pl-1SgPoss-Dat
‘my cat’



Language Comparison via Droplets



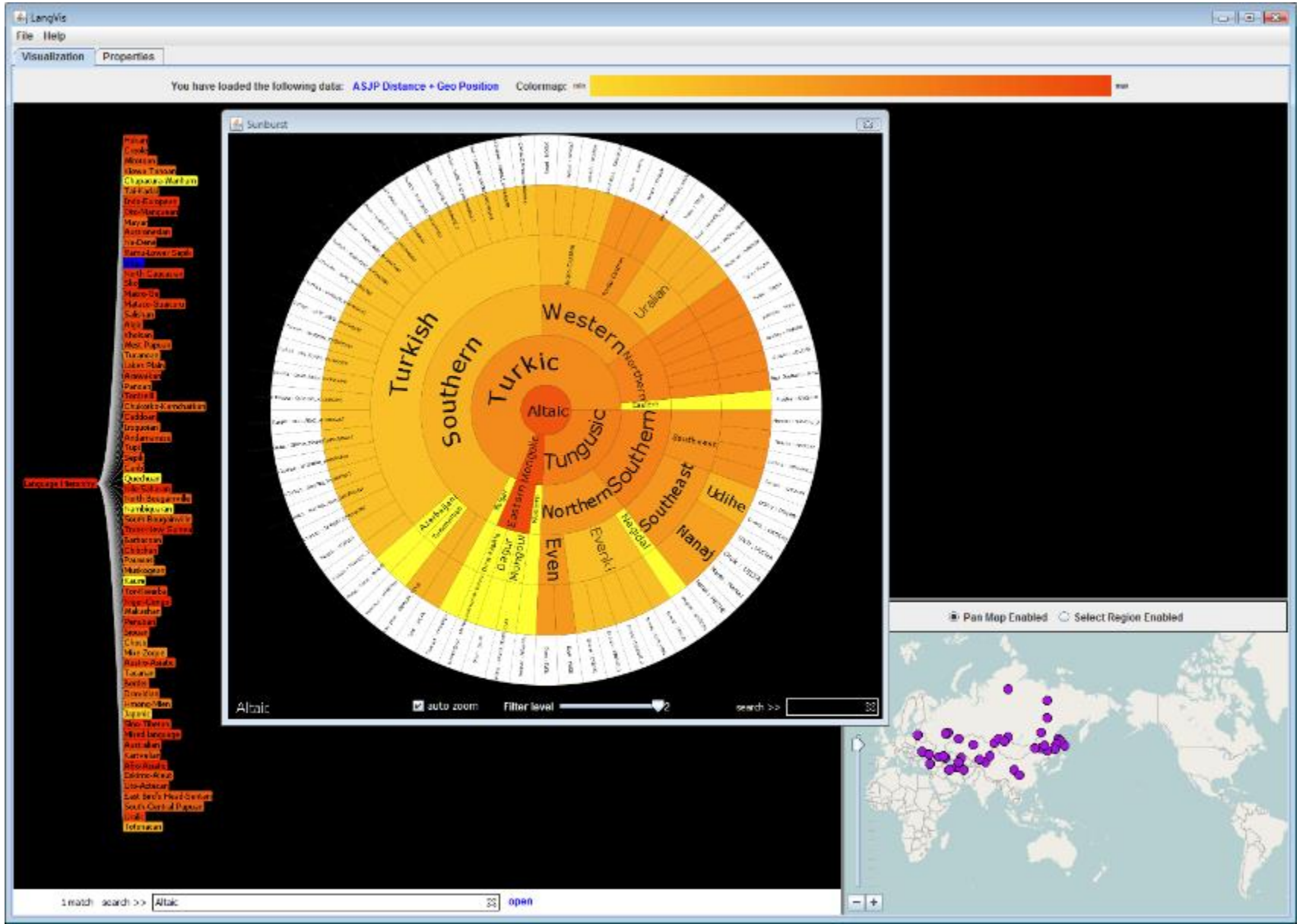
Norwegian shows language change a ➔ e in comparison to Swedish.

Example: Sunburst and Maps

- Another way to compare features across languages is via a sunburst visualization.
- The following visualization combines sunburst with a link to the geographical location of the language.
- The visual analysis is heavily interactive.
  - One can feed in one's own data.
  - One can also use the WALS (World Atlas of Language Structures; http://wals.info).

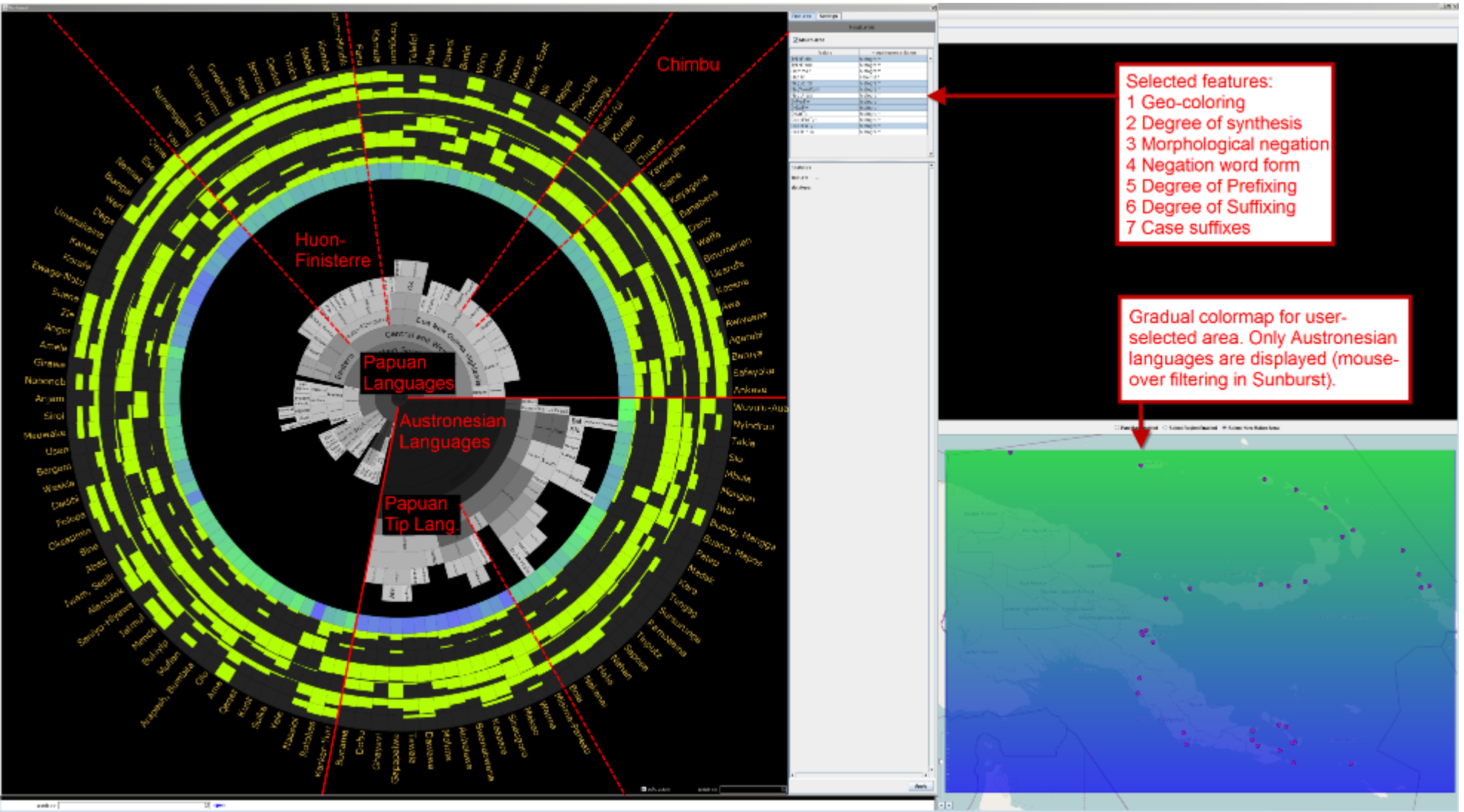
Christian Rohrdantz, Michael Hund, Thomas Mayer, Bernhard Wälchli and Daniel A. Keim. 2012. The World's Languages Explorer: Visual Analysis of Language Features in Geneaologica and Areal contexts. Computer Graphics Forum 31(3), 935-944.

Sunburst and Maps for Language Families



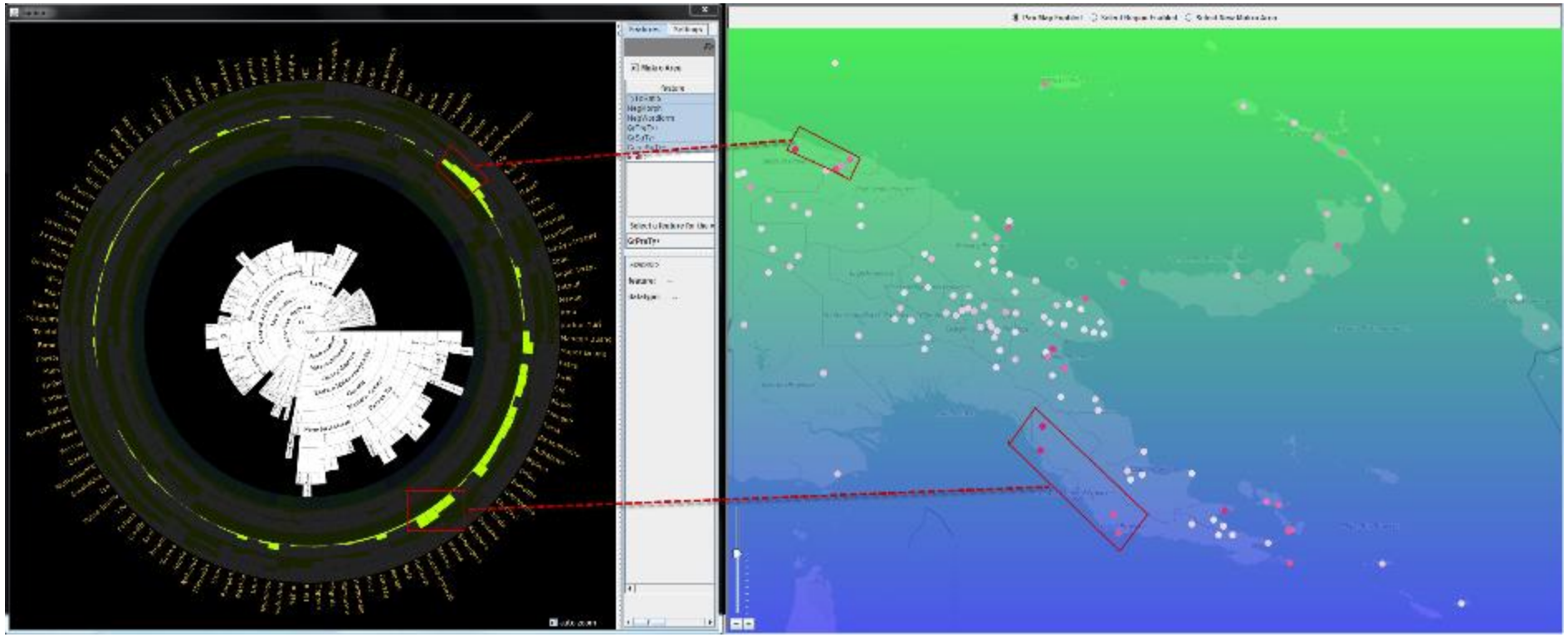
World's Language Explorer

Comparing 126 Languages of Papua New-Guinea based on the New Testament.



Each circle segment represents one language, each ring the values of one feature across all languages.

World's Language Explorer



Bringing genealogy (left) and areal distributions (right) interactively into context: The values of a selected feature ring are color-coded on a map for exploration.

Sorting and Pattern Discovery

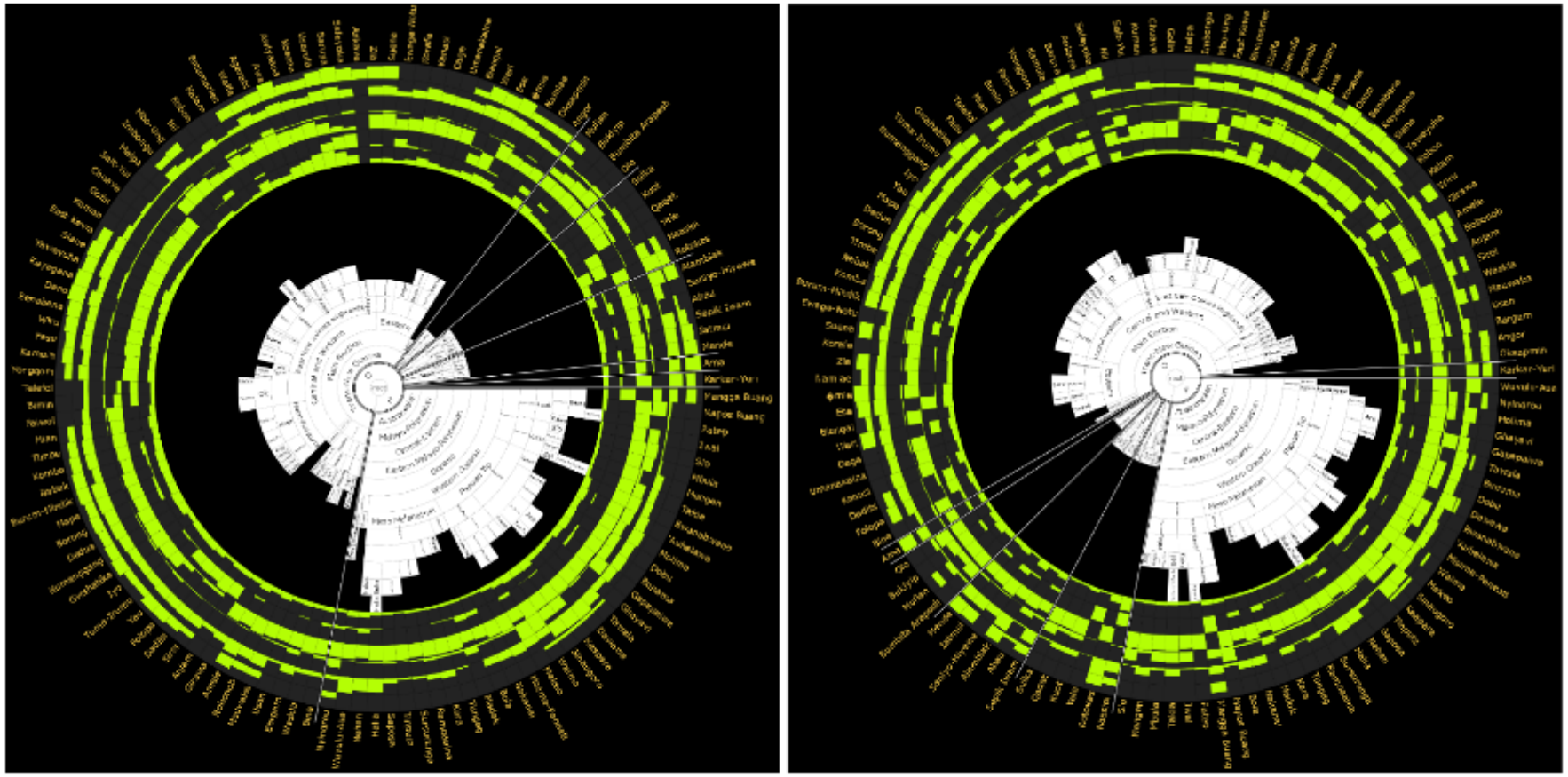
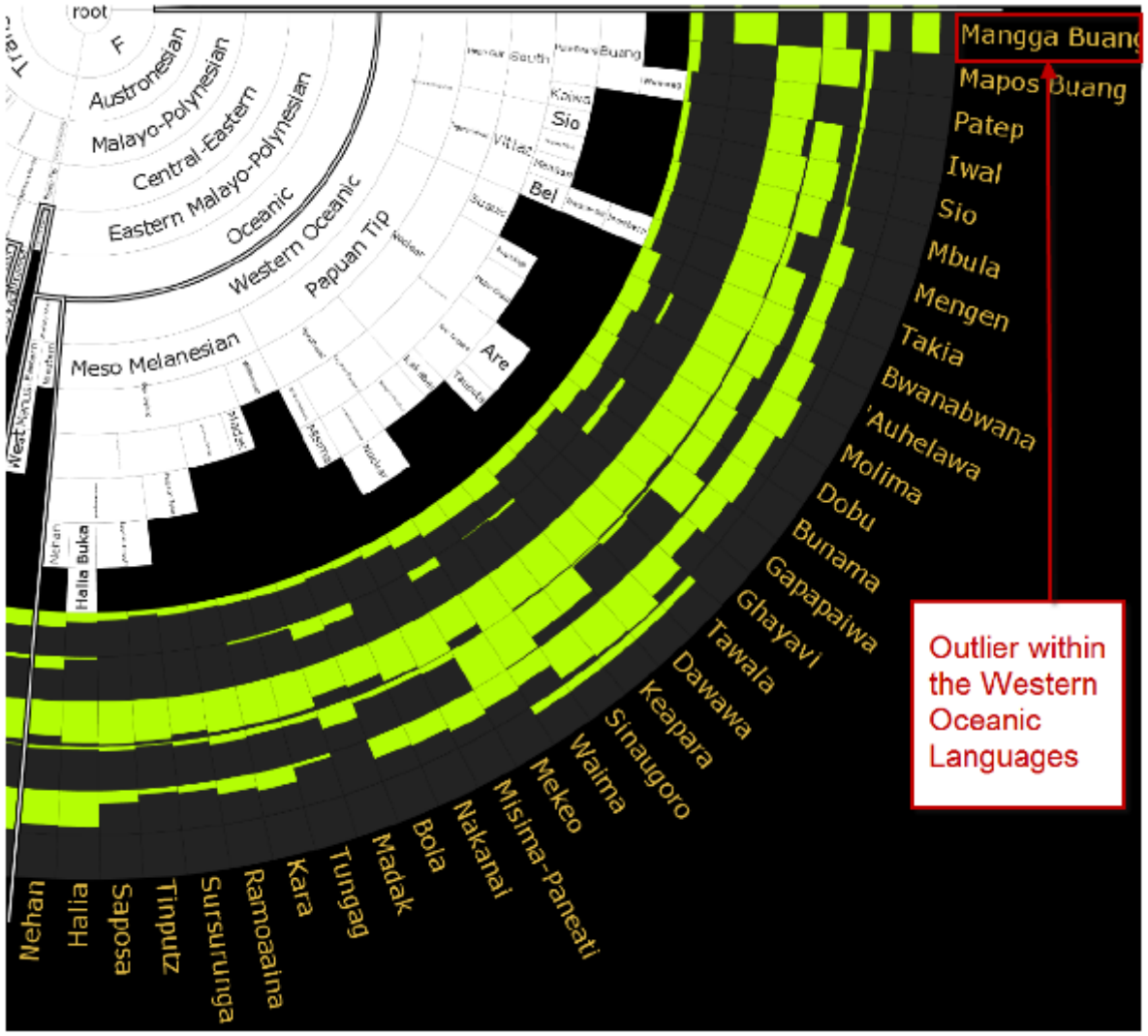


Figure 2.5: High-resolution screenshot showing automatically extracted features for languages from Papua New Guinea with leaves ordered to *maximize* (left) and *minimize* (right) the pairwise leaf similarity for neighbors.

Sorting and Pattern Discovery



WALS Explorer

- A version that is tailored to interact with WALS is available on-line
- http://www.th-mayer.de/wals/

Thomas Mayer, Bernhard Wälchli, Christian Rohrdantz and Michael Hund. 2014. From the extraction of continuous features in parallel texts to visual analytics of heterogeneous areal-typological datasets. In B. Nolan and C. Periñán-Pascual (eds.), Language Processing and Grammars: The role of functionally oriented computational models, 13–38. John Benjamins.

Glyph Visualization for Diachronic Corpora

- Visualization of IcePaHC – Diachronic Corpus of Icelandic
  - Syntactically annotated in Penn Treebank style
  - 60 texts
  - 12th century CE to 21st century CE
- Two case studies so far (on-going)
  - V1 in Icelandic
  - Dative Subjects in Icelandic

Glyph Visualization for Diachronic Corpora

V1 (Verb Initial or Verb First)

- Verb initial structures were common in matrix declaratives in Germanic.
- In German (and English) they mostly survive in narrative/joke contexts

  Walked a man into a pub...

Questions

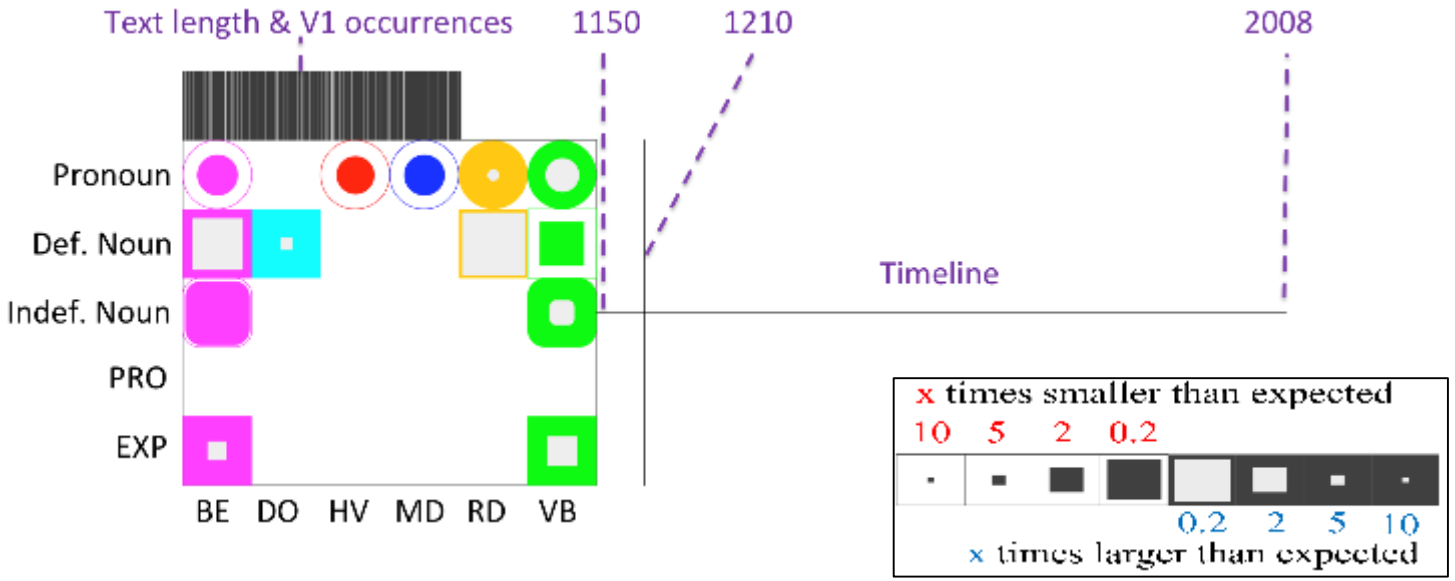
- What determines the appearance of V1?
- How did this change over the history of Germanic?

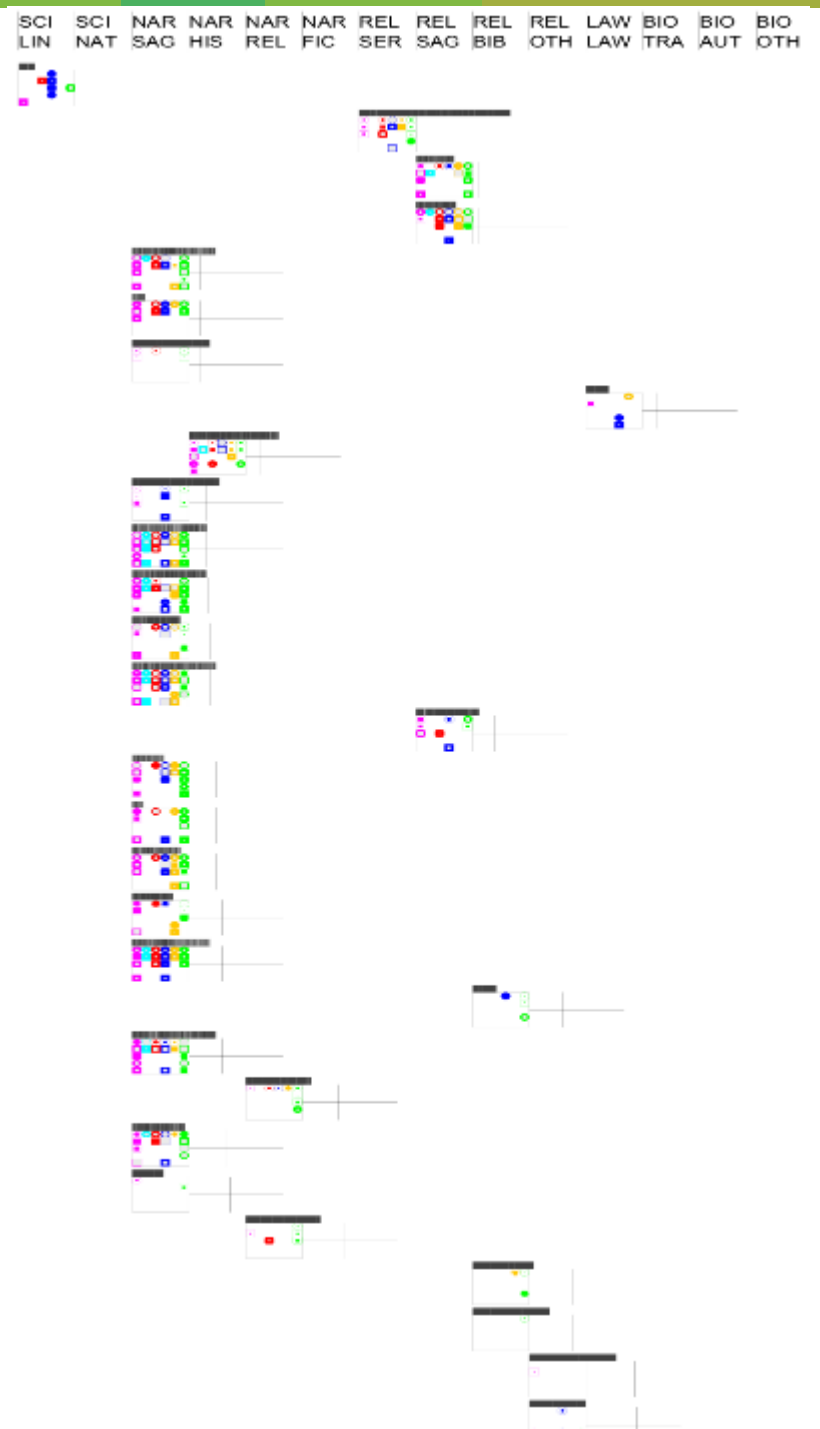
Butt, Miriam, Tina Bögel, Kristina Kotcheva, Christin Schätzle, Christian Rohrdantz, Dominik Sacha, Nicole Dehé & Daniel Keim. 2014. V1 in Icelandic: A Multifactorical Visualization of Historical Data. Proceedings of the LREC 2014 Workshop on Visualization as added value in the development, use and evaluation of LRs (VisLR). Reykjavil Iceland.

Example: V1 in Icelandic

Visual Analytic Access to Data

- Glyph Visualization of likely factors
- Overview of all 60 texts at once
- Can drill down to individual data points interactively
- **Shneiderman's Mantra**: Overview First, Show the Important – Details on Demand





HistoBankVis

- New Visualization for working with diachronic treebanks:
  - HistoBankVis
  - Accessible On-line
  - Interactive and Fast

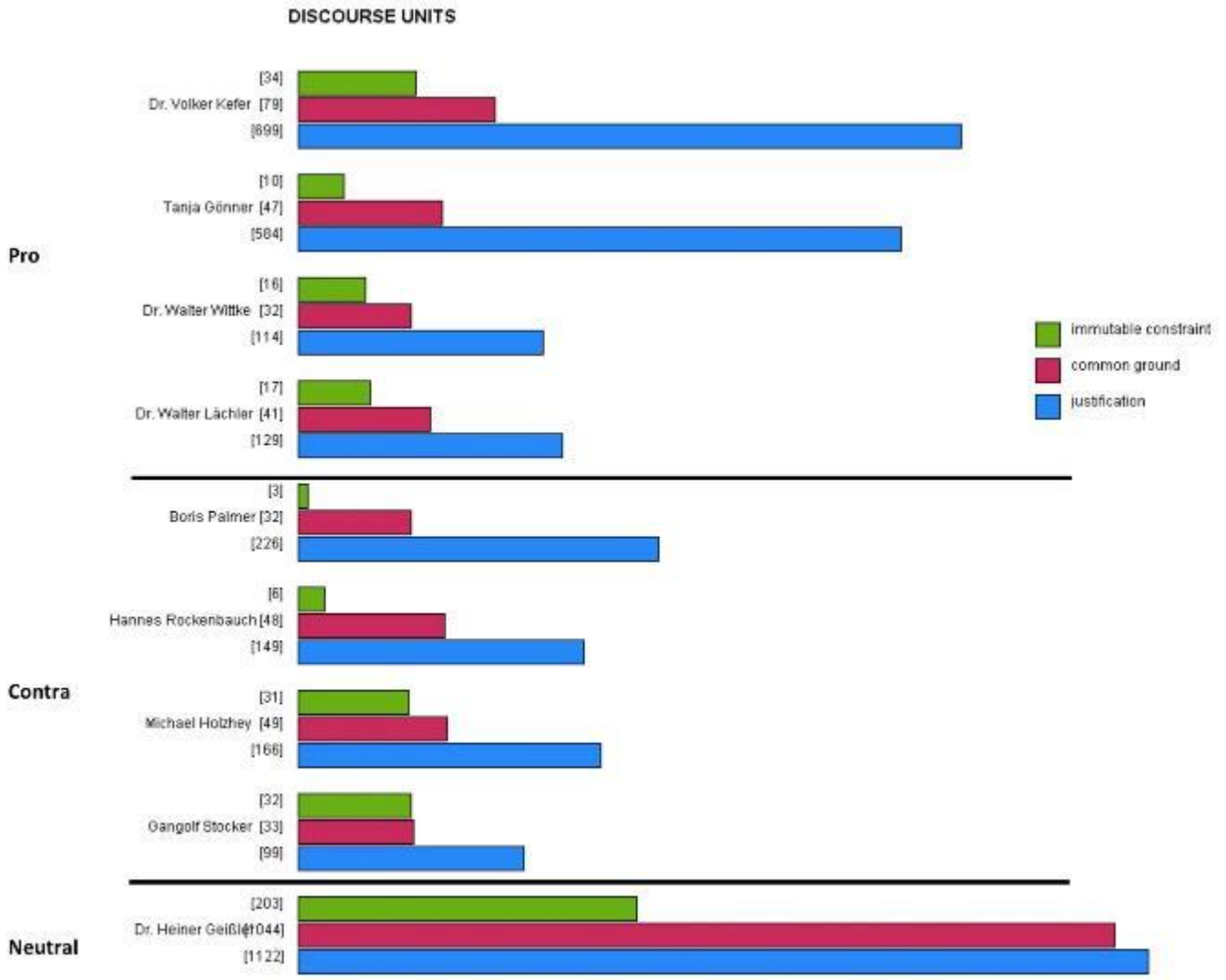
subva.dbvis.de/histobankvis-v1.0

histobankvis.pdf

Schätzle, Christin, Michael Hund, Frederik Dennig, Miriam Butt, Daniel A. Keim. 2017. HistoBankVis: Detecting Language Change via Data Visualization. In G. Bouma and Y. Adesam (eds.): *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, Linköping: Linköping University Electronic Press, pp. 32-39.

Example: Analyzing Political Argumentation (VisArgue)

- Public mediation on S21 (controversy around Stuttgart train station)
- Speakers are either Pro or Contra.
- Mediator is supposed to be neutral
- Data is annotated (rule based)



Gold, Valentin, Mennatallah El-Assady, Tina Bögel, Christian Rohrdantz, Miriam Butt, Katharina Holzinger & Daniel Keim. 2015. **Visual Linguistic Analysis of Political Discussions: Measuring Deliberative Quality.** *Digital Scholarship in the Humanities*, DOI: 10.1093/llc/fqv033

VisArgue Framework (http://visargue.inf.uni-konstanz.de)

Statistical Visualizations

Basic Data Exploration View

To get a feeling of the dataset and explore correlations and high-level patterns that occur in a conversation,simple statistical visualizations can be created dynamically.To explore the data in this view, some data dimensions can be selected interactively and the statistical visualization are generated automatically, according to a heuristic.

☐

Speaker Profile Explorer

Basic Data Exploration View

In order to get an overview over all participants of a conversation, the speaker profile explorer can be used. This view shows metadata and statistics about each speaker, in addition to her picture and party or position in the conversation. Typical speech and argumentation patterns of her utterances are shown in an aggregated argumentation glyph.

☑

Lexical-Unit Annotation Explorer

Basic Data Exploration View

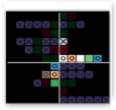
This view presents an interactive component for the exploration of the linguistic annotations in the text. Using interactive selection and highlighting, the annotated corpus can be analyzed.The annotations are subdivided into semantic categories and can be explored on lexeme, as well as, word level.

☑

Argumentation-Glyph Views

Deliberation Analysis Visualization

This visualization brings together all measures of deliberation. Using hierarchical glyphs, a fingerprint of the characteristics of utterances can be generated. Interesting sub-dimensions and measures can be interactively selected, depending on the analysis task. For a more holistic interpretation, the glyphs can be aggregated on multiple levels, e.g. speaker, topic, etc.

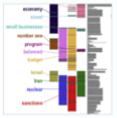


☑

Lexical Episode Plots

Content Exploration Visualization

Aligned to the concept of lexical chaining, lexical episode plots allow the analysis of the content of text corpora based on the unexpectedness of the occurrence density of words and n-grams. This view generated an overview of the complete text using the zoomed-out distant-reading view. Using the zooming and highlighting functionalities, interesting section of the text can be identified and further inspected through close-reading.The significance-level of the episode detection can be varied for multiple levels of detail.

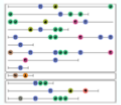


☑

Named-Entity Relationship Visualizations

Content Exploration Visualization

This module provides multiple interlinked visual interfaces that are based on the abstraction of the text into defined named-entity categories.By exploring the distant-restricted named-entity relationship, additional insight can be gained about the semantic structure of the analyzed discussion. Numerous selection and filtering options help the analyst in verifying hypothesis about the content of the corpus.



☑

Topic-Space Views

Content Exploration Visualization

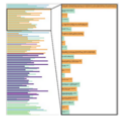
This visualization sets the focus on the exploration of speaker participation in multi-party conversations. By spanning a conversation space using all mentioned topics in the discussion, the interaction of the speakers over time can be traced. This animated visualization enables the analysts to detect interesting interaction patters and form hypothesis about the participating speakers and their alliances.

☐

IHTM Model Visualization

Content Exploration Visualization

In order to understand the underlying structure of the incremental hierarchical topic modeling algorithm, this visualization can be utilized. This visual model representation gives a deep insight into the hidden structure of the topic modeling and opens the black-box of the algorithm to ensure a more trustworthy result and allow a human feedback loop.



☐

Fuzzy Feature Alignment Views

Multiple-File Comparative Analysis

This visualization is based on a fuzzy-feature detection and alignment algorithm that is designed to enhance the comparative analysis of multiple conversation files using their pre-computed measures. The analysis of feature alignments across multiple conversations enables pattern detection and evaluation.

☐

Lexical Episode Plots

User-Steerable Topic Modeling

Content Analysis

Named-Entity Relationship Exploration

Feature Alignment

Discourse and Argumentation

Argumentation Glyphs

Discourse Maps

Speaker Dynamics

Topic Space Views

M. El-Assady, R. Sevastjanova, B. Gipp, D. Keim, and C. Collins.
NEREx: Named-Entity Relationship Exploration in Conversational Text Corpora.
Computer Graphics Forum – EuroVis, 2017.

Moderator
Good evening from Hofstra University in Hempstead , New York .
I am Lester Holt , anchor of " NBC Nightly News . "
I want to welcome you to the first presidential debate .
The participants tonight are Donald Trump and Hillary Clinton .
This debate is sponsored by the Commission on Presidential Debates , a nonpartisan , nonprofit organization .
The commission drafted tonight's format , and the rules have been agreed to by the campaigns .
The 90 minute debate is divided into six segments , each 15 minutes long .
We'll explore three topic areas tonight : Achieving prosperity ; America's direction ; and securing America .
At the start of each segment , I will ask the same lead-off question to both candidates , and they will each have up to two minutes to respond .
From that point until the end of the segment , we'll have an open discussion .
The questions are mine and have not been shared with the commission or the campaigns .
The audience here in the room has agreed to remain silent so that we can focus on what the candidates are saying .
I will invite you to applaud , however , at this moment , as we welcome the candidates : Democratic nominee for president of the United States Hillary Clinton , and Republican nominee for president of the United States Donald J . Trump .
APPLAUSE

Clinton
How are you , Donald ?
APPLAUSE

Moderator
Good luck to you .
APPLAUSE Well , I do not expect us to cover all the issues of this campaign tonight , but I remind everyone , there are two more presidential debates scheduled .
We are going to focus on many of the issues that voters tell us are most important , and we are going to press for specifics .
I am honored to have this role , but this evening belongs to the candidates and , just as important , to the American people .
Candidates , we look forward to hearing you articulate your policies and your positions , as well as your visions and your values .
So , let's begin . we are calling this opening segment " Achieving Prosperity . "
And central to that is jobs .
There are two economic realities in America today .
There's been a record six straight years of job growth , and new census numbers show incomes have increased at a record rate after years of stagnation .
However , income inequality remains significant , and nearly half of Americans are living paycheck to paycheck .
Beginning with you , Secretary Clinton , why are you a better choice than your opponent to create the kinds of jobs that will put more money into the pockets of American works ?

Clinton
Well , thank you , Lester , and thanks to Hofstra for hosting us .
The central question in this election is really what kind of country we want to be and what kind of future we'll build together .
Today is my granddaughter's second birthday , so I think about this a lot .
First , we have to build an economy that works for everyone , not just those at the top .
That means we need new jobs , good jobs , with rising incomes .
I want us to invest in you .
I want us to invest in your future .
That means jobs in infrastructure , in advanced manufacturing , innovation and technology , clean , renewable energy , and small business , because most of the new jobs will come from small business .
We also have to make the economy fairer .
That starts with raising the national minimum wage and also guarantee , finally , equal pay for women's work .

Clinton
I also want to see more companies do profit-sharing .
If you help create the profits , you should be able to share in them , not just the executives at the top .
And I want us to do more to support people who are struggling to balance family and work .
I've heard from so many of you about the difficult choices you face and the stresses that you're under .

Moderator

Good evening from Hofstra University in Hempstead , New York .
I am Lester Holt , anchor of " NBC Nightly News . "
I want to welcome you to the first presidential debate .
The participants tonight are Donald Trump and Hillary Clinton .
This debate is sponsored by the Commission on Presidential Debates , a nonpartisan , nonprofit organization .
The commission drafted tonight's format , and the rules have been agreed to by the campaigns .
The 90 minute debate is divided into six segments , each 15 minutes long .
We'll explore three topic areas tonight : Achieving prosperity ; America's direction ; and securing America .
At the start of each segment , I will ask the same lead-off question to both candidates , and they will each have up to two minutes to respond .
From that point until the end of the segment , we'll have an open discussion .
The questions are mine and have not been shared with the commission or the campaigns .
The audience here in the room has agreed to remain silent so that we can focus on what the candidates are saying .
I will invite you to applaud , however , at this moment , as we welcome the candidates : Democratic nominee for president of the United States Hillary Clinton , and Republican nominee for president of the United States Donald J . Trump .
APPLAUSE

Clinton

How are you , Donald ?
APPLAUSE

Moderator

Good luck to you .
APPLAUSE Well , I do not expect us to cover all the issues of this campaign tonight , but I remind everyone , there are two more presidential debates scheduled .
We are going to focus on many of the issues that voters tell us are most important , and we are going to press for specifics .
I am honored to have this role , but this evening belongs to the candidates and , just as important , to the American people .
Candidates , we look forward to hearing you articulate your policies and your positions , as well as your visions and your values .
So , let's begin . we are calling this opening segment " Achieving Prosperity . "
And central to that is jobs .
There are two economic realities in America today .
There's been a record six straight years of job growth , and new census numbers show incomes have increased at a record rate after years of stagnation .
However , income inequality remains significant , and nearly half of Americans are living paycheck to paycheck .
Beginning with you , Secretary Clinton , why are you a better choice than your opponent to create the kinds of jobs that will put more money into the pockets of American works ?

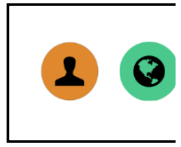
Clinton

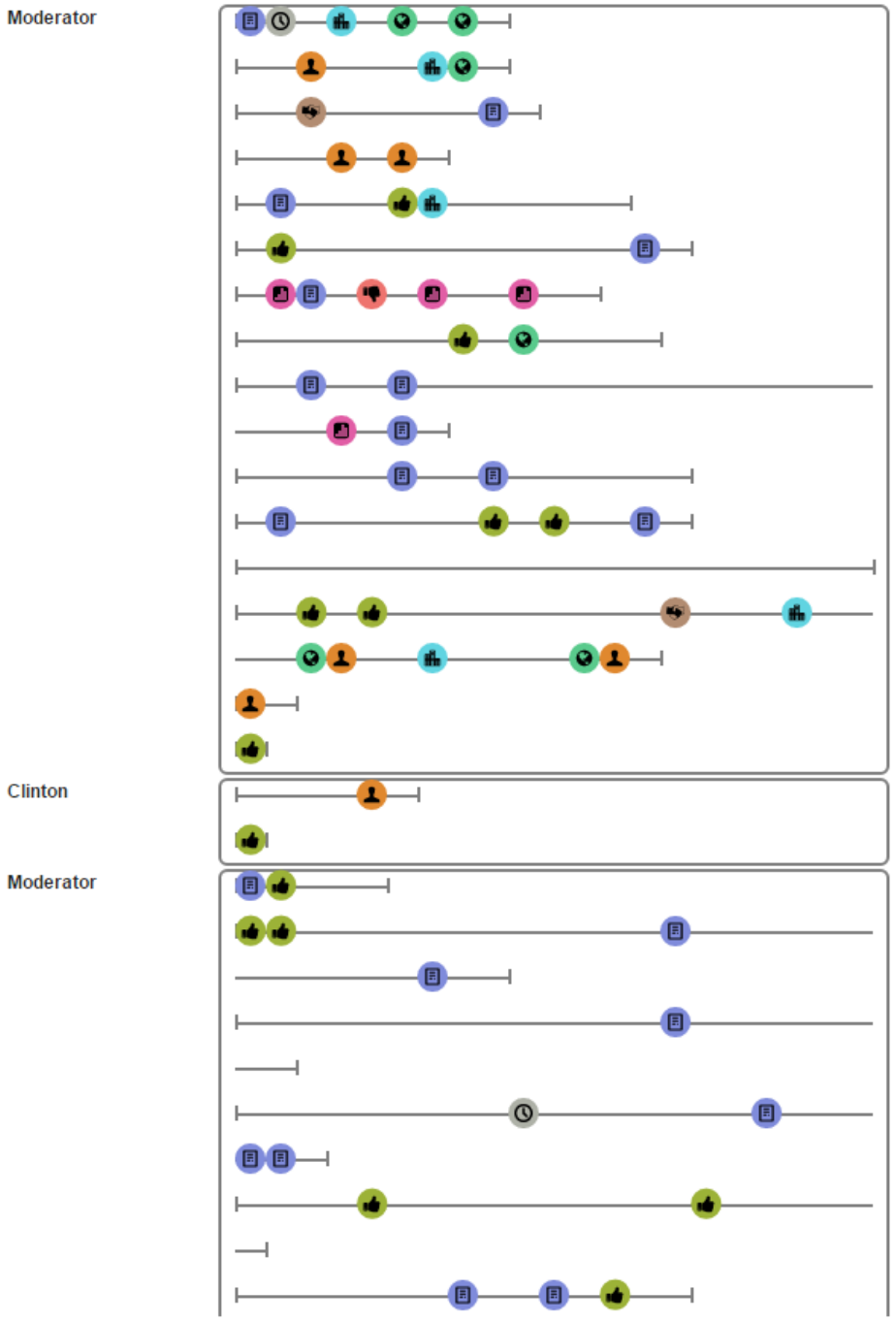
Well , thank you , Lester , and thanks to Hofstra for hosting us .
The central question in this election is really what kind of country we want to be and what kind of future we'll build together .
Today is my granddaughter's second birthday , so I think about this a lot .
First , we have to build an economy that works for everyone , not just those at the top .
That means we need new jobs , good jobs , with rising incomes .
I want us to invest in you .
I want us to invest in your future .
That means jobs in infrastructure , in advanced manufacturing , innovation and technology , clean , renewable energy , and small business , because most of the new jobs will come from small business .
We also have to make the economy fairer .
That starts with raising the national minimum wage and also guarantee , finally , equal pay for women's work .

Clinton

I also want to see more companies do profit-sharing .
If you help create the profits , you should be able to share in them , not just the executives at the top .
And I want us to do more to support people who are struggling to balance family and work .
I've heard from so many of you about the difficult choices you face and the stresses that you're under .

- Person
- Geo-Location
- Organization
- Date-Time
- Measuring Unit
- Measure
- Context-Keyword
- Politeness-Indicator
- Positive-Emotion Indicator
- Negative-Emotion Indicator





Close-Reading

Moderator

Good evening from Hofstra University in Hempstead , New York .
I am Lester Holt , anchor of " NBC Nightly News . "
I want to welcome you to the first presidential debate .
The participants tonight are Donald Trump and Hillary Clinton .
This debate is sponsored by the Commission on Presidential Debates , a nonpartisan , nonprofit organization
The commission drafted tonight's format , and the rules have been agreed to by the campaigns .
The 90 minute debate is divided into six segments , each 15 minutes long .
We'll explore three topic areas tonight : Achieving prosperity ; America's direction ; and securing America .
At the start of each segment , I will ask the same lead-off question to both candidates , and they will each have
From that point until the end of the segment , we'll have an open discussion .
The questions are mine and have not been shared with the commission or the campaigns .
The audience here in the room has agreed to remain silent so that we can focus on what the candidates are
I will invite you to applaud , however , at this moment , as we welcome the candidates : Democratic nominee
Trump .
APPLAUSE

Clinton

How are you , Donald ?
APPLAUSE

Moderator

Good luck to you .
APPLAUSE Well , I do not expect us to cover all the issues of this campaign tonight , but I remind everyone ,
We are going to focus on many of the issues that voters tell us are most important , and we are going to press
I am honored to have this role , but this evening belongs to the candidates and , just as important , to the
Candidates , we look forward to hearing you articulate your policies and your positions , as well as your
So , let's begin . we are calling this opening segment " Achieving Prosperity . "
And central to that is jobs .
There are two economic realities in America today .
There's been a record six straight years of job growth , and new census numbers show incomes have
However , income inequality remains significant , and nearly half of Americans are living paycheck to
Beginning with you , Secretary Clinton , why are you a better choice than your opponent to create the

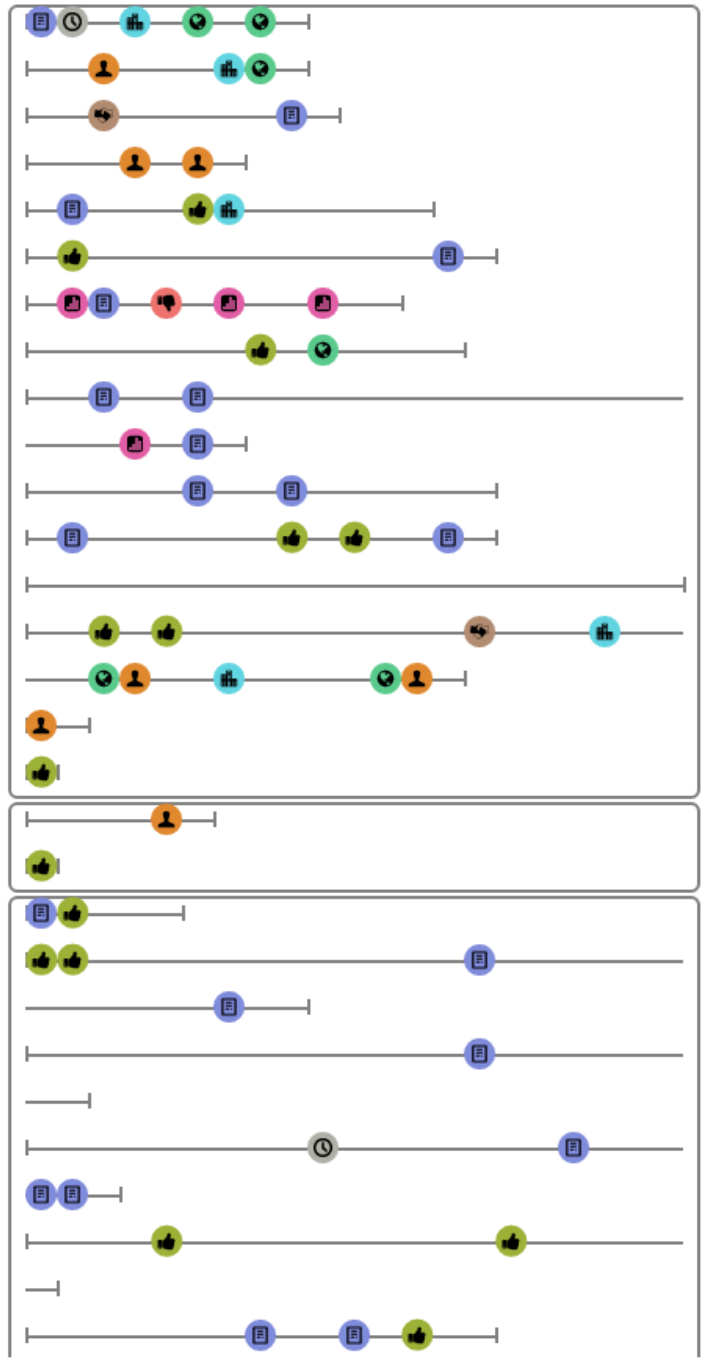
Clinton

Well , thank you , Lester , and thanks to Hofstra for hosting us .
The central question in this election is really what kind of country we want to be and what kind of future we'll
Today is my granddaughter's second birthday , so I think about this a lot .
First , we have to build an economy that works for everyone , not just those at the top .
That means we need new jobs , good jobs , with rising incomes .
I want us to invest in you .
I want us to invest in your future .
That means jobs in infrastructure , in advanced manufacturing , innovation and technology , clean , renewable
We also have to make the economy fairer .
That starts with raising the national minimum wage and also guarantee , finally , equal pay for women's

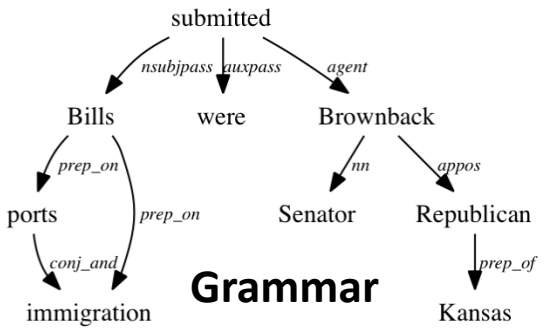
Clinton

I also want to see more companies do profit-sharing .
If you help create the profits , you should be able to share in them , not just the executives at the top .
And I want us to do more to support people who are struggling to balance family and work .
I've heard from so many of you about the difficult choices you face and the stresses that you're under .

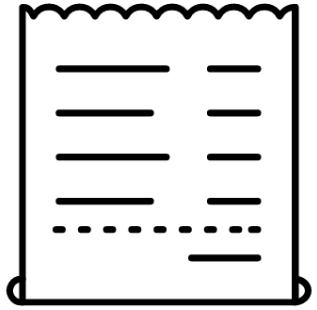
Distant-Reading



Named-Entity Relationship



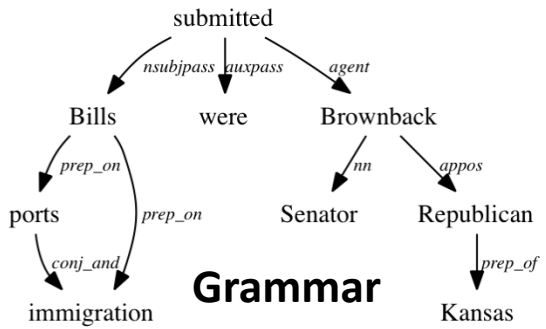
Grammar



Document Structure

Works well for **highly-edited texts** but not for **verbatim conversation transcripts** due to ungrammatical sentences.

Named-Entity Relationship



Grammar

Works well for **highly-edited texts** but not for **verbatim conversation transcripts** due to ungrammatical sentences.

Clinton
Well , Donald , I know you live in your own reality , but that is not the facts .
The facts are -- I did say I hoped it would be a good deal , but when it was negotiated ...

Trump
Not .

Clinton
... which I was not responsible for , I concluded it wasn't .
I wrote about that in my book ...

Trump
So is it President Obama's fault ?

Clinton
... before you even announced .

Trump
Is it President Obama's fault ?

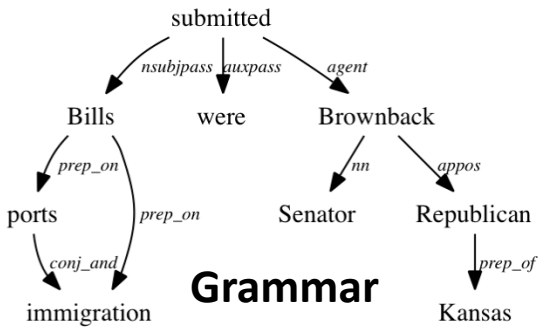
Clinton
Look , there are differences ...

Trump
Secretary , is it President Obama's fault ?

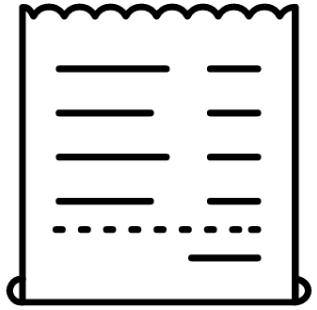
Clinton
There are ...

Trump
Because he is pushing it .

Named-Entity Relationship



Grammar

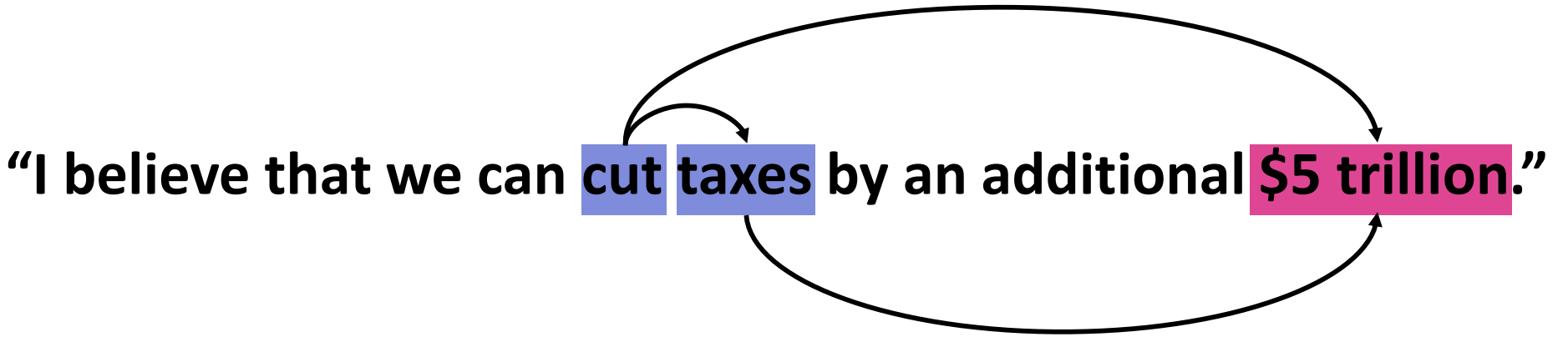


Document Structure

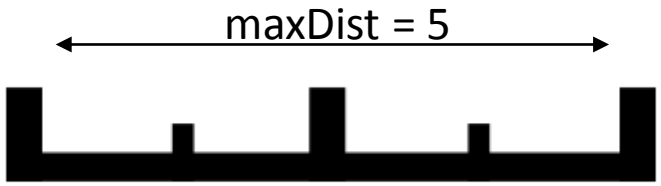
Works well for **highly-edited texts** but not for **verbatim conversation transcripts** due to ungrammatical sentences.

Works well for **text summarization** but constructs **too broad** relations.

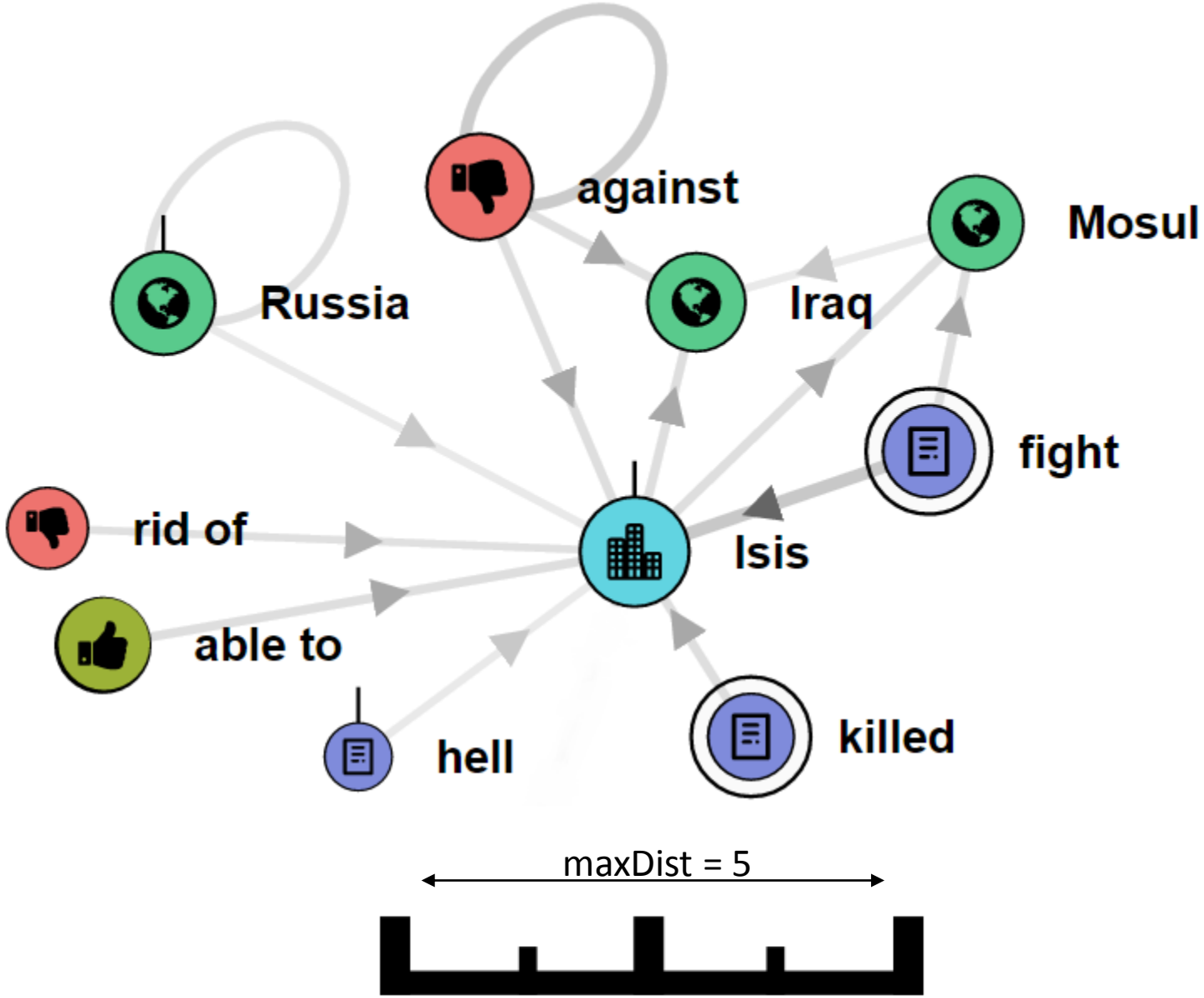
Distance-Restricted Entity-Relationship Model

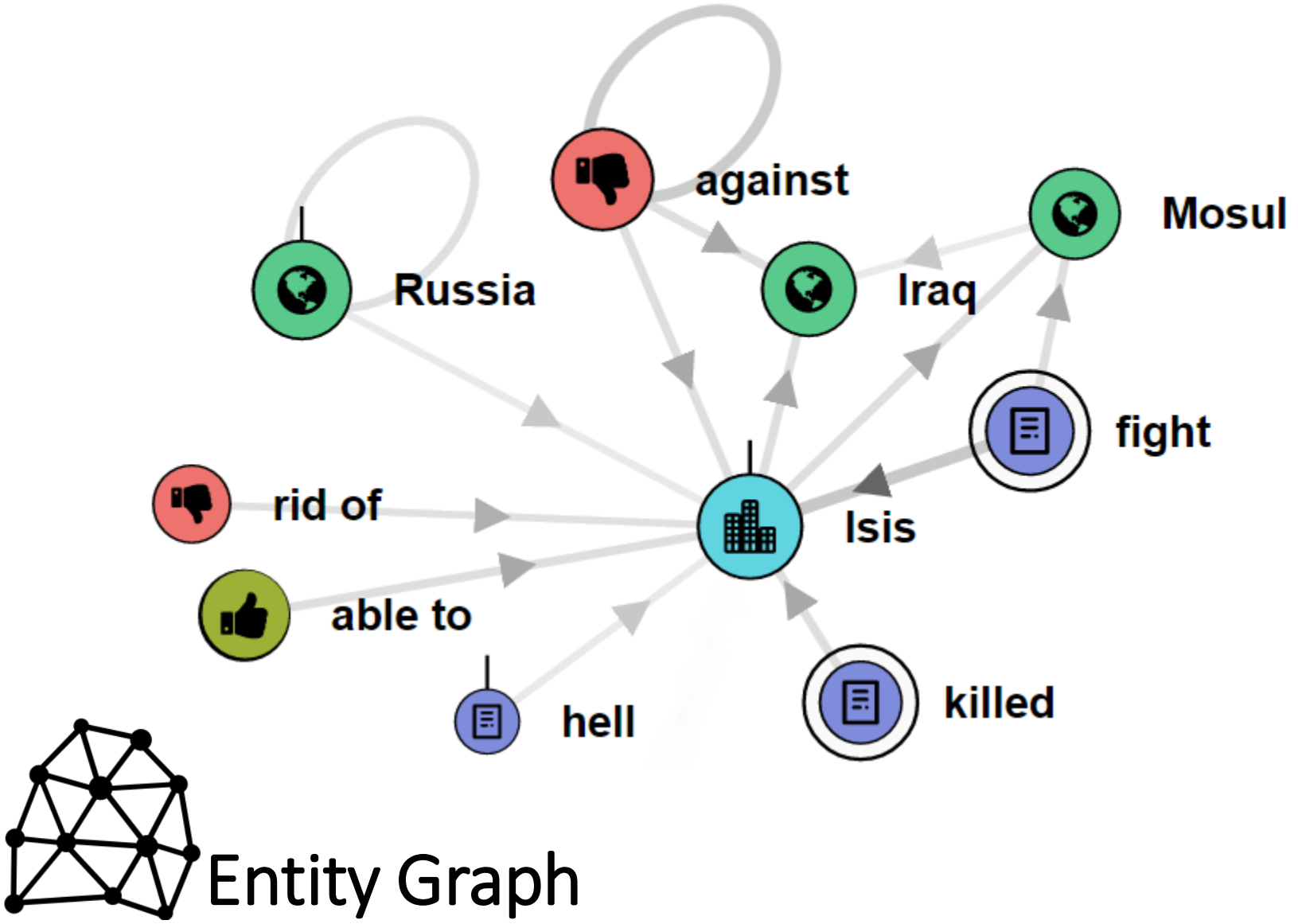


Entity-Pairs: cut taxes | cut $5 trillion | taxes $5 trillion



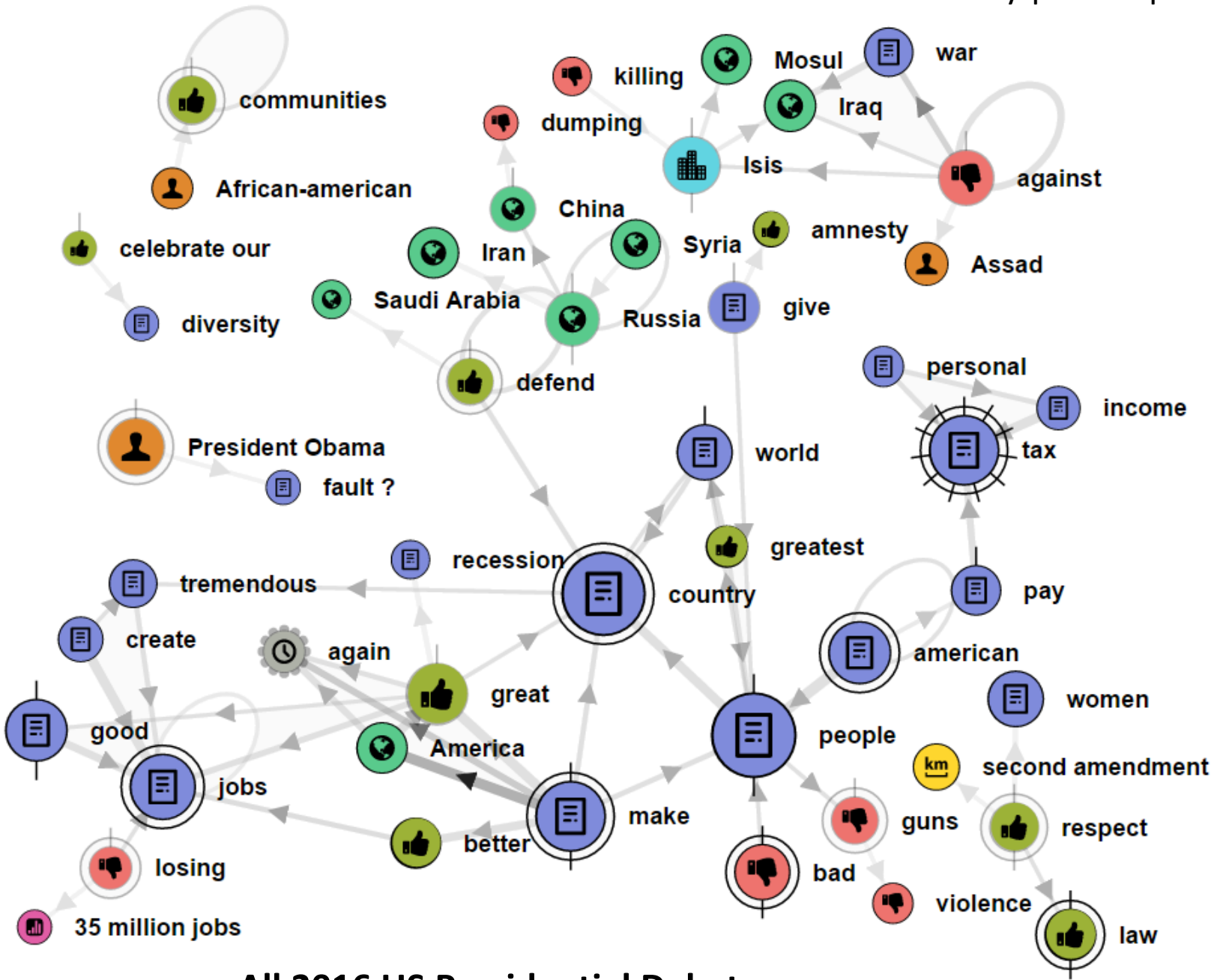
Distance-Restricted Entity-Relationship Model



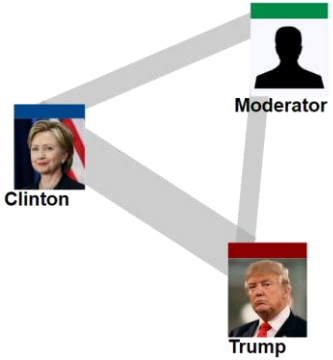


Entity Graph

minimum entity-pair frequency = 3



All 2016 US Presidential Debates

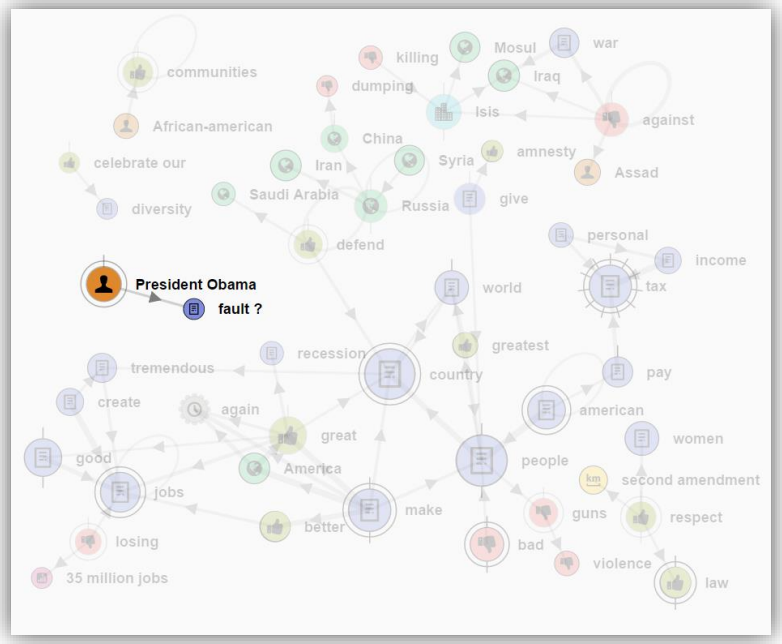


[SG] Speaker Graph

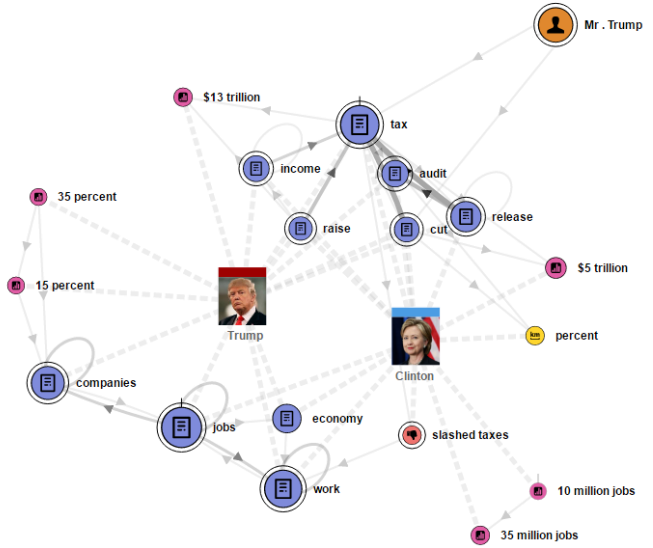
[TLV] Text-Level View

So is it President Obama's fault ?

[ELV] Entity-Level View



[EG] Entity Graph



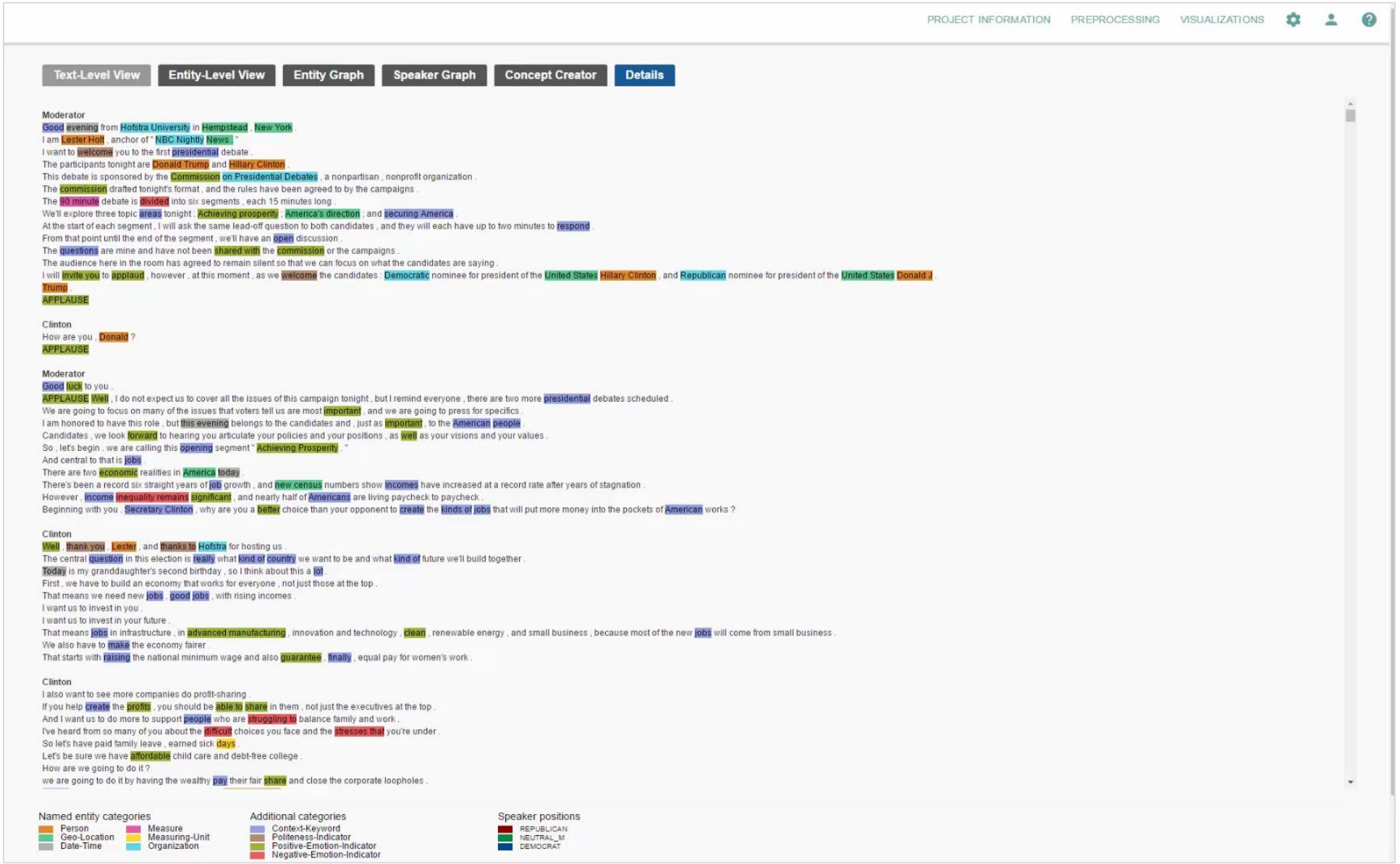
[CG] Concept Graph

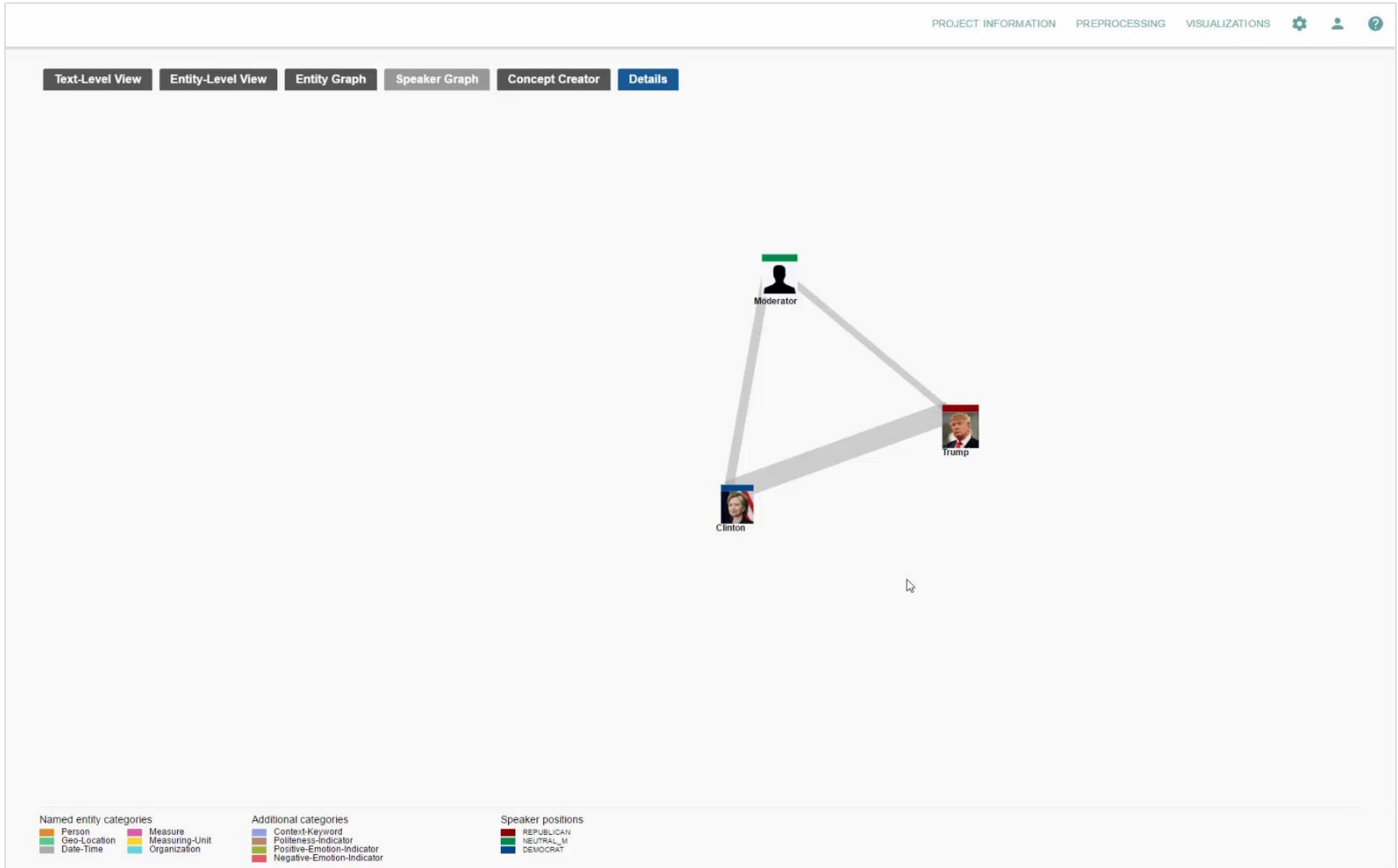
Views

Use Case

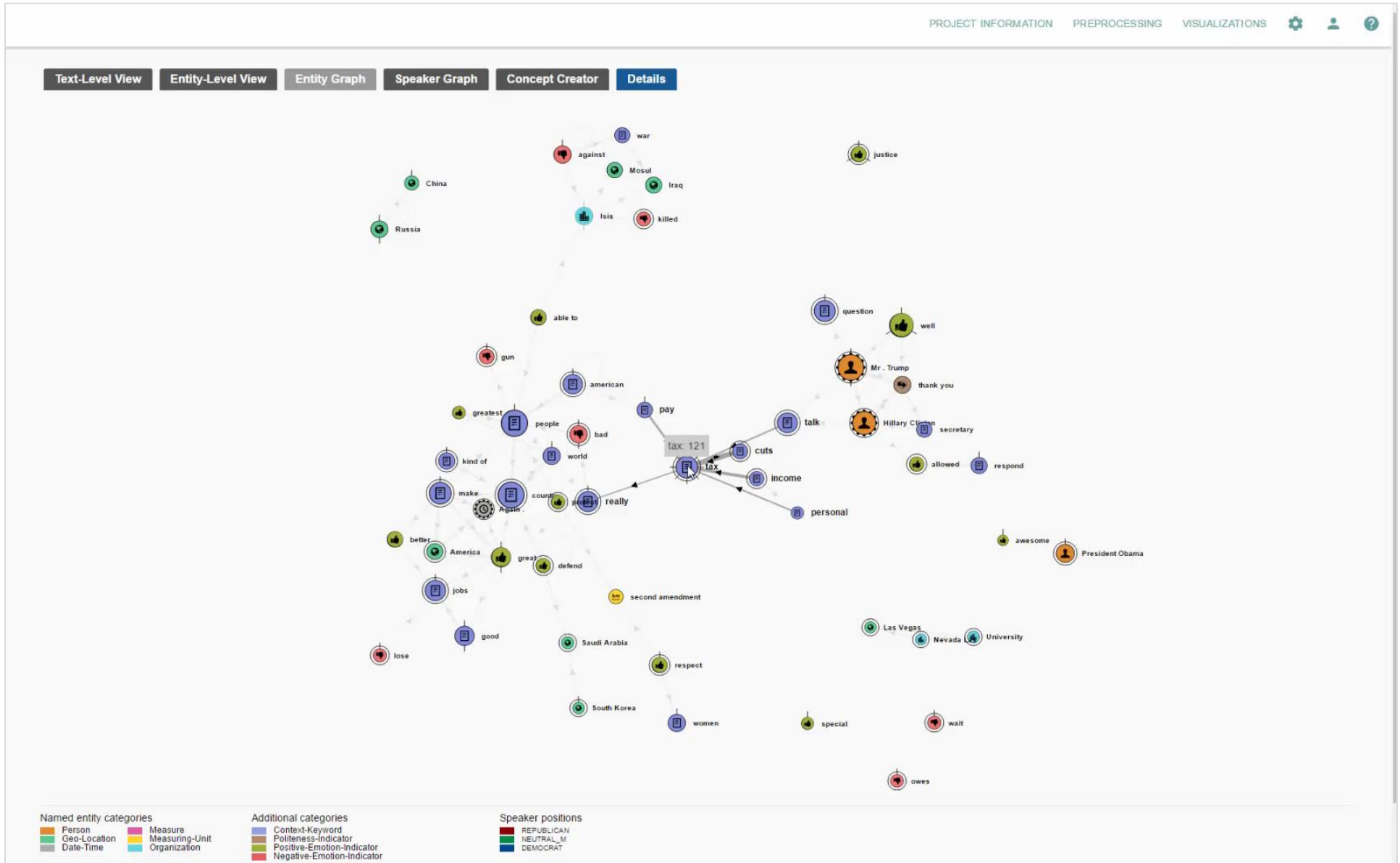
2016 US Presidential Debates

Overall Exploration of the Dataset


PROJECT INFORMATION
PREPROCESSING
VISUALIZATIONS
Text-Level View
Entity-Level View
Entity Graph
Speaker Graph
Concept Creator
Details
Moderator
Good evening from Hofstra University in Hempstead , New York .
I am Lester Holt , anchor of " NBC Nightly News . "
I want to welcome you to the first presidential debate .
The participants tonight are Donald Trump and Hillary Clinton .
This debate is sponsored by the Commission on Presidential Debates , a nonpartisan , nonprofit organization .
The commission drafted tonight's format , and the rules have been agreed to by the campaigns .
The 90 minute debate is divided into six segments , each 15 minutes long .
We'll explore three topic areas tonight : Achieving prosperity ; America's direction ; and securing America .
At the start of each segment , I will ask the same lead-off question to both candidates , and they will each have up to two minutes to respond .
From that point until the end of the segment , we'll have an open discussion .
The questions are mine and have not been shared with the commission or the campaigns .
The audience here in the room has agreed to remain silent so that we can focus on what the candidates are saying .
I will invite you to applaud , however , at this moment , as we welcome the candidates : Democratic nominee for president of the United States Hillary Clinton , and Republican nominee for president of the United States Donald J .
Trump .
APPLAUSE
Clinton
How are you , Donald ?
APPLAUSE
Moderator
Good luck to you .
APPLAUSE Well , I do not expect us to cover all the issues of this campaign tonight , but I remind everyone , there are two more presidential debates scheduled .
We are going to focus on many of the issues that voters tell us are most important , and we are going to press for specifics .
I am honored to have this role , but this evening belongs to the candidates and , just as important , to the American people .
Candidates , we look forward to hearing you articulate your policies and your positions , as well as your visions and your values .
So , let's begin . we are calling this opening segment " Achieving Prosperity . "
And central to that is jobs .
There are two economic realities in America today .
There's been a record six straight years of job growth , and new census numbers show incomes have increased at a record rate after years of stagnation .
However , income inequality remains significant , and nearly half of Americans are living paycheck to paycheck .
Beginning with you , Secretary Clinton , why are you a better choice than your opponent to create the kinds of jobs that will put more money into the pockets of American works ?
Clinton
Well , thank you , Lester , and thanks to Hofstra for hosting us .
The central question in this election is really what kind of country we want to be and what kind of future we'll build together .
Today is my granddaughter's second birthday , so I think about this a lot .
First , we have to build an economy that works for everyone , not just those at the top .
That means we need new jobs , good jobs , with rising incomes .
I want us to invest in you .
I want us to invest in your future .
That means jobs in infrastructure , in advanced manufacturing , innovation and technology , clean , renewable energy , and small business , because most of the new jobs will come from small business .
We also have to make the economy fairer .
That starts with raising the national minimum wage and also guarantee , finally , equal pay for women's work .
Clinton
I also want to see more companies do profit-sharing .
If you help create the profits , you should be able to share in them , not just the executives at the top .
And I want us to do more to support people who are struggling to balance family and work .
I've heard from so many of you about the difficult choices you face and the stresses that you're under .
So let's have paid family leave , earned sick days .
Let's be sure we have affordable child care and debt-free college .
How are we going to do it ?
we are going to do it by having the wealthy pay their fair share and close the corporate loopholes .
Named entity categories
Person
Geo-Location
Date-Time
Measure
Measuring-Unit
Organization
Additional categories
Context-Keyword
Politeness-Indicator
Positive-Emotion-Indicator
Negative-Emotion-Indicator
Speaker positions
REPUBLICAN
NEUTRAL_M
DEMOCRAT


PROJECT INFORMATION
PREPROCESSING
VISUALIZATIONS
Text-Level View
Entity-Level View
Entity Graph
Speaker Graph
Concept Creator
Details
Moderator
Trump
Clinton
Named entity categories
Person
Geo-Location
Date-Time
Measure
Measuring-Unit
Organization
Additional categories
Context-Keyword
Politeness-Indicator
Positive-Emotion-Indicator
Negative-Emotion-Indicator
Speaker positions
REPUBLICAN
NEUTRAL_M
DEMOCRAT

Topic: Taxes


PROJECT INFORMATION
PREPROCESSING
VISUALIZATIONS
Text-Level View
Entity-Level View
Entity Graph
Speaker Graph
Concept Creator
Details
war
against
Mosul
China
Iraq
Isis
killed
Russia
justice
able to
question
well
gun
Mr . Trump
thank you
american
pay
greatest
people
bad
talk
Hillary Clinton
secretary
tax: 121
cuts
world
kind of
tax
allowed
respond
income
make
really
personal
awesome
President Obama
better
America
great
defend
jobs
second amendment
Las Vegas
University
good
Saudi Arabia
lose
respect
South Korea
women
special
wait
owes
Named entity categories
Person
Geo-Location
Date-Time
Measure
Measuring-Unit
Organization
Additional categories
Context-Keyword
Politeness-Indicator
Positive-Emotion-Indicator
Negative-Emotion-Indicator
Speaker positions
REPUBLICAN
NEUTRAL_M
DEMOCRAT

Lexical Episode Plots

User-Steerable Topic Modeling

Content Analysis

Named-Entity Relationship Exploration

Feature Alignment

Discourse and Argumentation

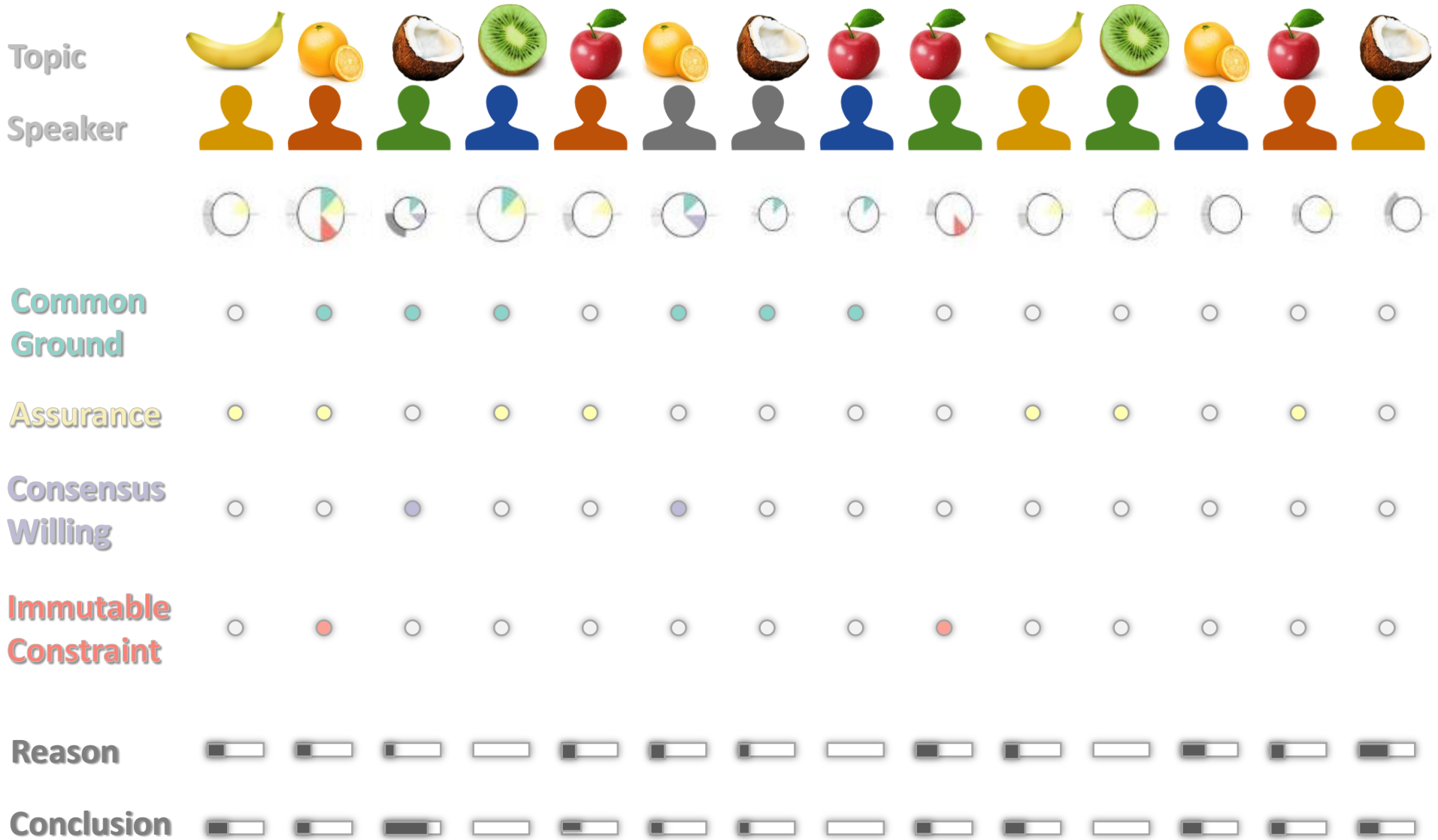
Argumentation Glyphs

Discourse Maps

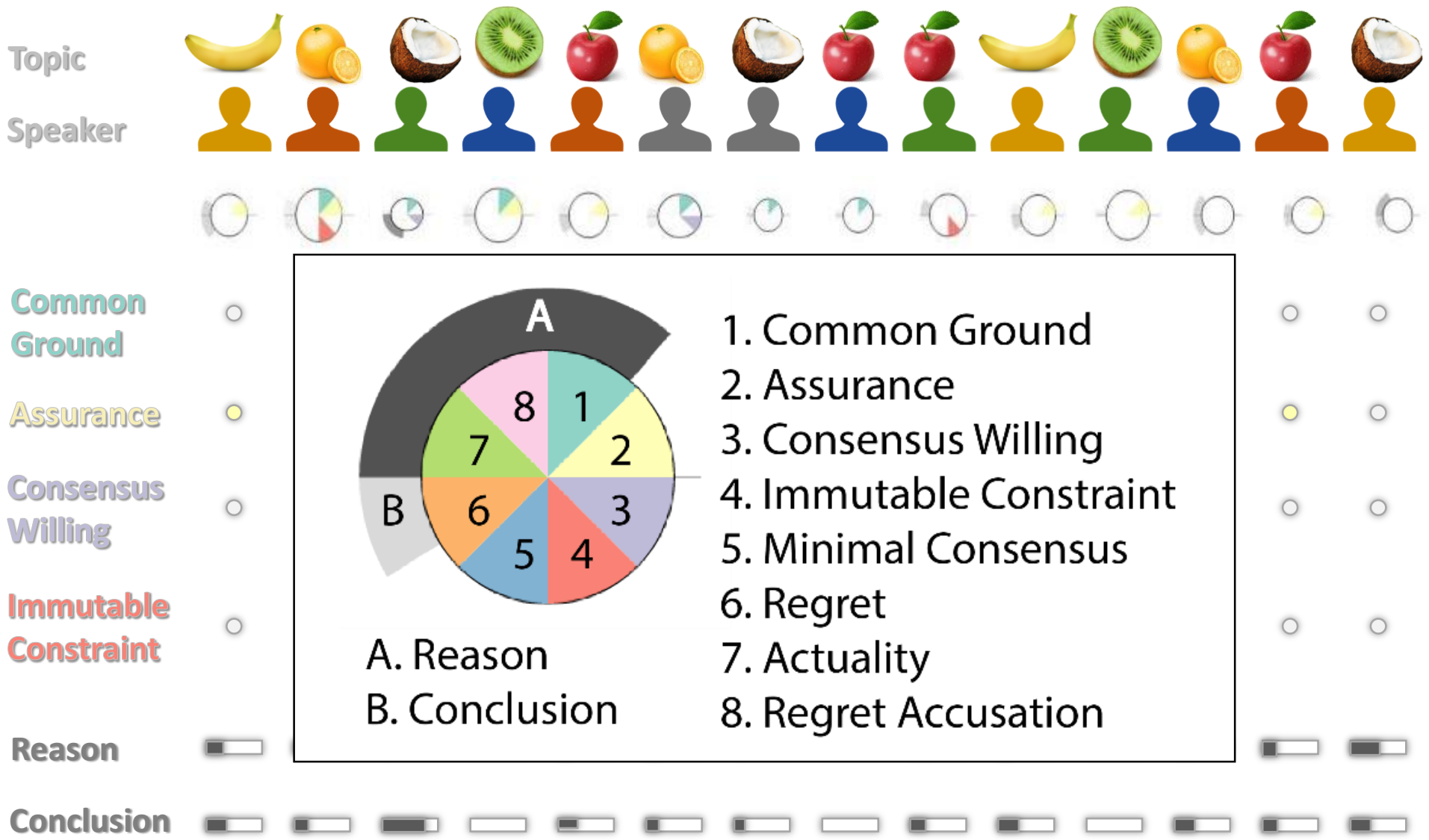
Speaker Dynamics

Topic Space Views

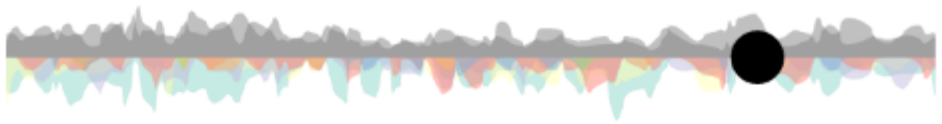
M. El-Assady, V. Gold, C. Acevedo, C. Collins and D. A. Keim.
ConToVi: Multi-Party Conversation Exploration using Topic-Space Views.
Computer Graphics Forum – EuroVis, 2016 .


Topic
Speaker
Common Ground
Assurance
Consensus Willing
Immutable Constraint
Reason
Conclusion

Argumentation Glyphs



Argumentation Glyphs



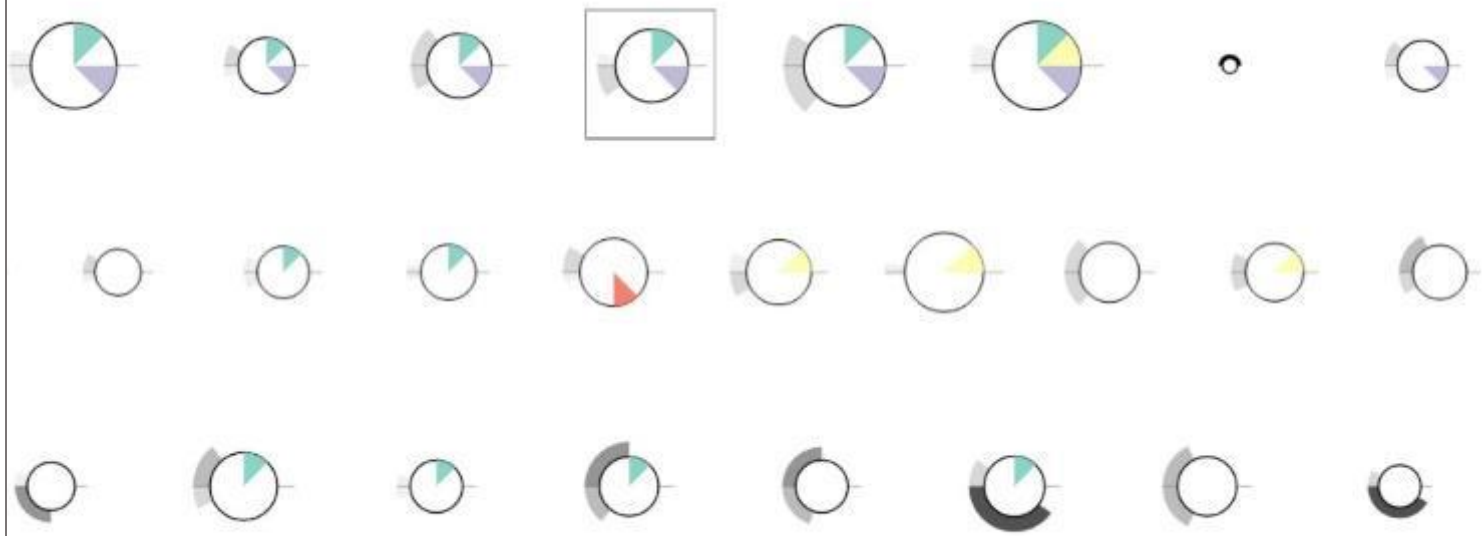
Dr. Heiner Geißler

[Vielen_Dank .]
[Also die Sache haben wir übrigens gestern Abend miteinander beredet nicht und es war klar nicht_wahr hier Gegenstand einer Schlichtung kann nicht sein BAST mit allen Ergänzungen Zusatz Mitteilungen und Verordnungen die von 2002 bis 2010 dann entwickelt worden sind weil das ja rein zufällig ist nicht wahr .]
[Wenn das in der Öffentlichkeit steht dann hat ja jeder die Möglichkeit punktuell situativ irgendwelche Informationen herauszuholen und stellt die dann dar als ob das der aktuelle Stand wäre .]
[Es ist so ein dickes Buch .]
[Wir können das in der Schlichtung einfach so nicht behandeln .]
[Das haben wir gestern Abend miteinander beredet und haben gesagt die Frage die uns interessiert die focussieren wir die konzentrieren wir nicht wahr auf die Frage die Herr_Conradi formuliert hat nämlich wie verhält es sich mit dem Tippfehler mit Euro und D_Mark .]
[So jetzt Frau_Dahlbender .]

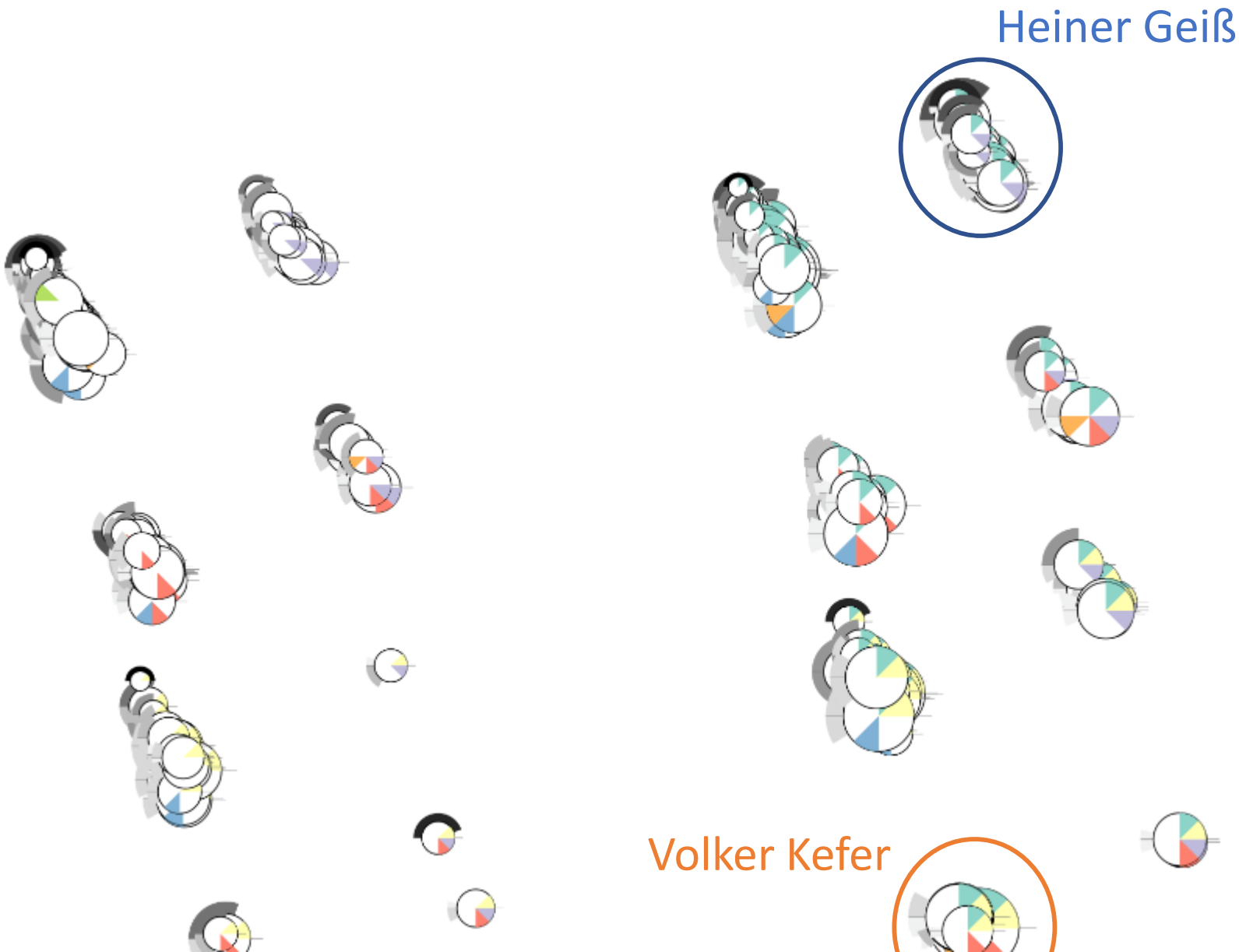
Reason

Conclusion

Common Ground
Assurance
Consensus Willing
Immutable Constraint
Minimal Consensus
Regret
Actuality
Regret Accusation



Argumentation Glyphs – Clustered View



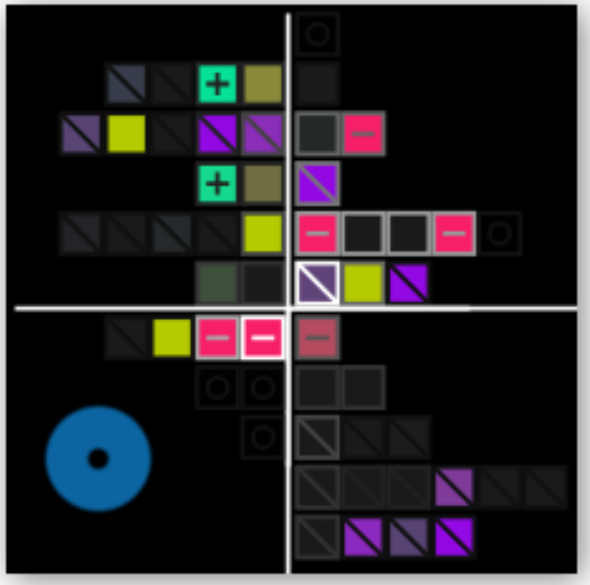
Micro-Linguistic Discourse Features

Dimensions of Deliberative Communication

1. **Participation** (equal and inclusive participation)
2. **Atmosphere & Respect** (equal respect)
3. **Argumentation & Justification** (reason-giving)
4. **Accommodation** (consensus-seeking)

Dimension	Subdimension	Measure	Type	Normalization
Accommondation	Agreement	Agreement	BINARY	BINARY
		Concession	NUMERICAL_CONTINUOUS	NONE
		Consensus	BINARY	BINARY
		Minimal consensus	BINARY	BINARY
		Consensus willing	BINARY	BINARY
	Disagreement	Disagreement	BINARY	BINARY
		Opposition	NUMERICAL_CONTINUOUS	NONE
		Contrast	BINARY	BINARY
		Dissent	BINARY	BINARY
		Activate opposition	BINARY	BINARY
	Agreement vs. Disagreement	Arrangement count	NUMERICAL_CONTINUOUS	NONE
		Arrangement relation	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
	Condition	Condition	NUMERICAL_CONTINUOUS	NONE
		Consequence	NUMERICAL_CONTINUOUS	NONE
	Arguing vs. Bargaining	Negotiation count	NUMERICAL_CONTINUOUS	NONE
		Negotiation relation	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
		Arguing	BINARY	BINARY
		Bargaining	BINARY	BINARY
Argumentation and Justification	Information Exchange	Information giving	BINARY	BINARY
		Information seeking	BINARY	BINARY
		Information refusing	BINARY	BINARY
		Elucidation	BINARY	BINARY
	Information Certainty	Epistemic value	NUMERICAL_BIPOLAR	NONE
	Event Modality	Event modality permission	BINARY	BINARY
		Event modality obligation	BINARY	BINARY
		Event modality alternative	BINARY	BINARY
		Event modality volition	BINARY	BINARY
		Event modality reluctance	BINARY	BINARY
		External constraint	BINARY	BINARY
	Reason Giving	Reason	NUMERICAL_CONTINUOUS	NONE
		Conclusion	NUMERICAL_CONTINUOUS	NONE
	Common Ground	Common ground	BINARY	BINARY
		Activate common ground	BINARY	BINARY
		Reject common ground	BINARY	BINARY
Participation	Equality of Speaker Participation	Expected probability to speak	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
		Moving Gini index	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
	Equality of Speaker Capabilities	Maas index	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
		Average sentence complexity	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
		Number of filler words	NUMERICAL_CONTINUOUS	NONE
		Stalling	BINARY	BINARY
	Topic Comprehensiveness	Network density	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
Atmosphere and Respect	Interruptions	Interruption	BINARY	BINARY
	Sentiments	Sentiment	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
	Conventional Politeness	Politeness	NUMERICAL_CONTINUOUS	NONE
		Impatience	BINARY	BINARY
		Unobtrusiveness	BINARY	BINARY
	Emotions	Emotion count	NUMERICAL_CONTINUOUS	NONE
		Emotion relation	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
	Face Issues	Resignation acceptance	NUMERICAL_CONTINUOUS	NONE
	Responsiveness	Topic shift	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
		Self previous recurrence	NUMERICAL_CONTINUOUS	LINEAR
		Self following recurrence	NUMERICAL_CONTINUOUS	LINEAR
		Self recurrence shift	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR
		Topic persistence	NUMERICAL_BIPOLAR	STANDARD_DEVIATION_LINEAR

Discourse Maps

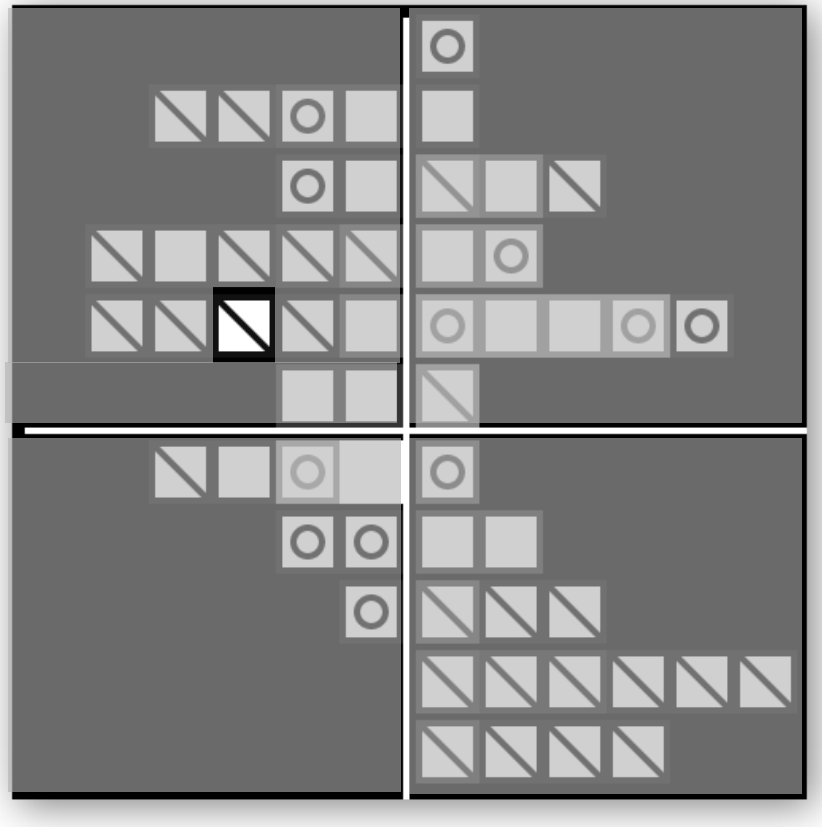


>

A. Reason
B. Conclusion

1. Common Ground
2. Assurance
3. Consensus Willing
4. Immutable Constraint
5. Minimal Consensus
6. Regret
7. Actuality
8. Regret Accusation

Discourse Maps



4 Dimensions

→ Quadrants

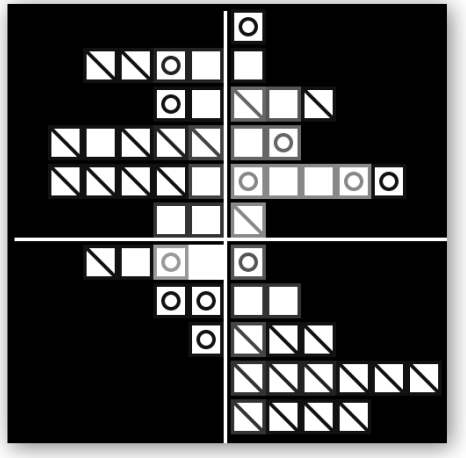
19 Sub-Dimensions

→ Rows

53 Measures / Features

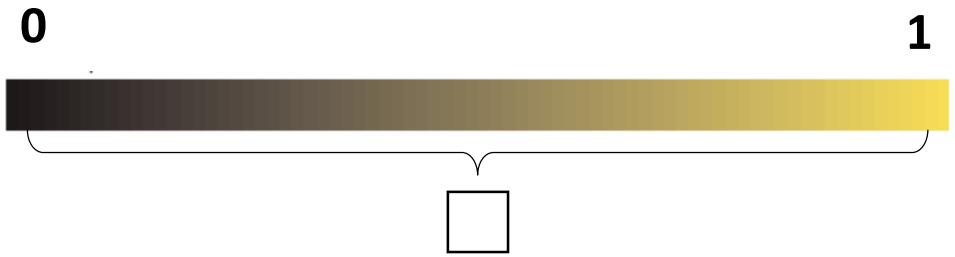
→ Rectangles

Types of measures

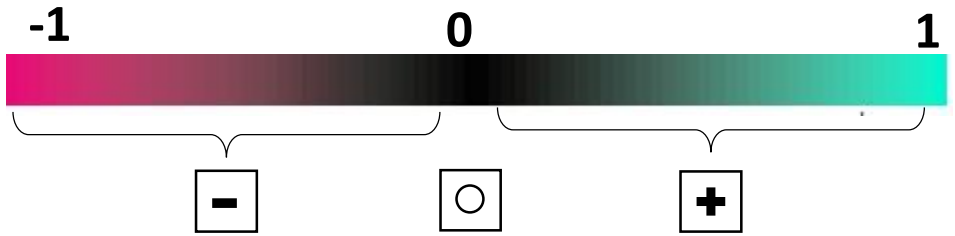


- **Binary** (0/1, e.g. “reason” or not)

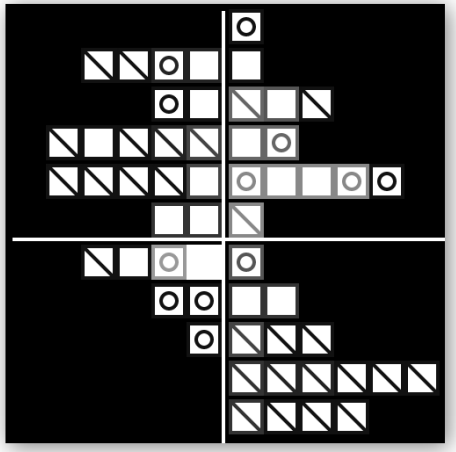
- **Numerical, continuous** (0→1, e.g. epistemic value)

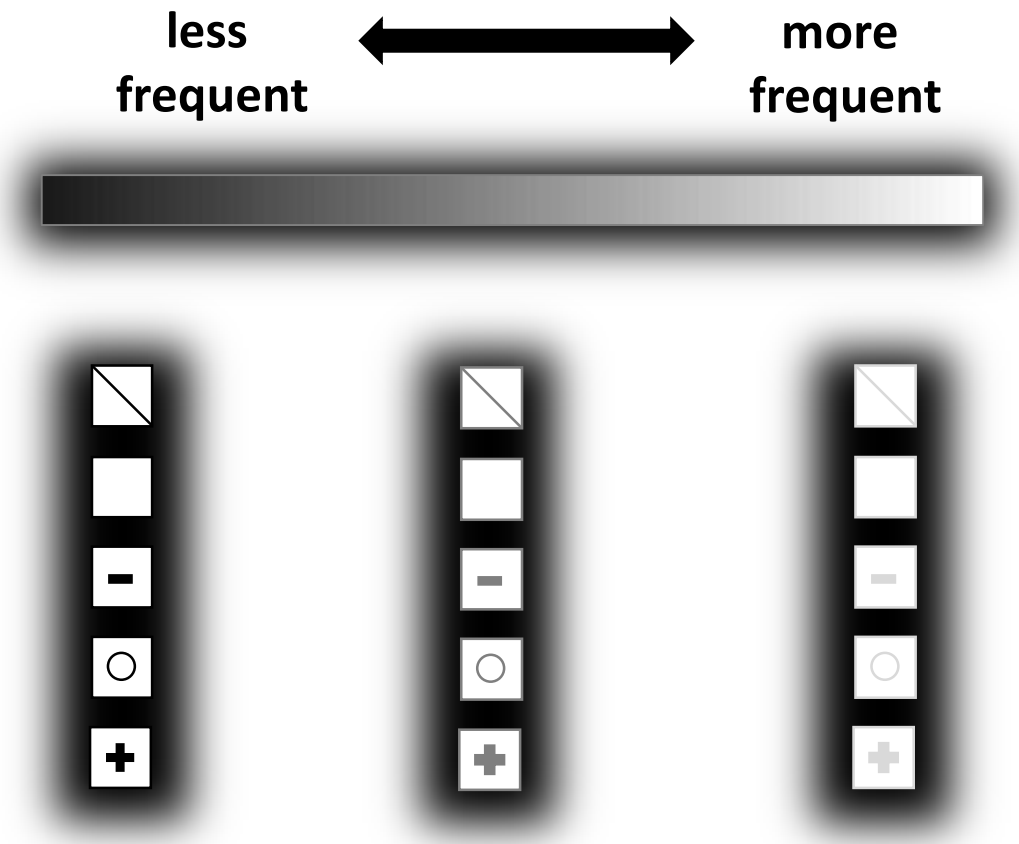


- **Numerical, bipolar** (-1→1, e.g. emotion=“negative/positive”)

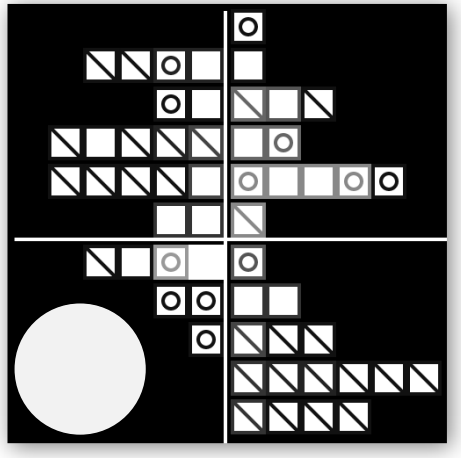


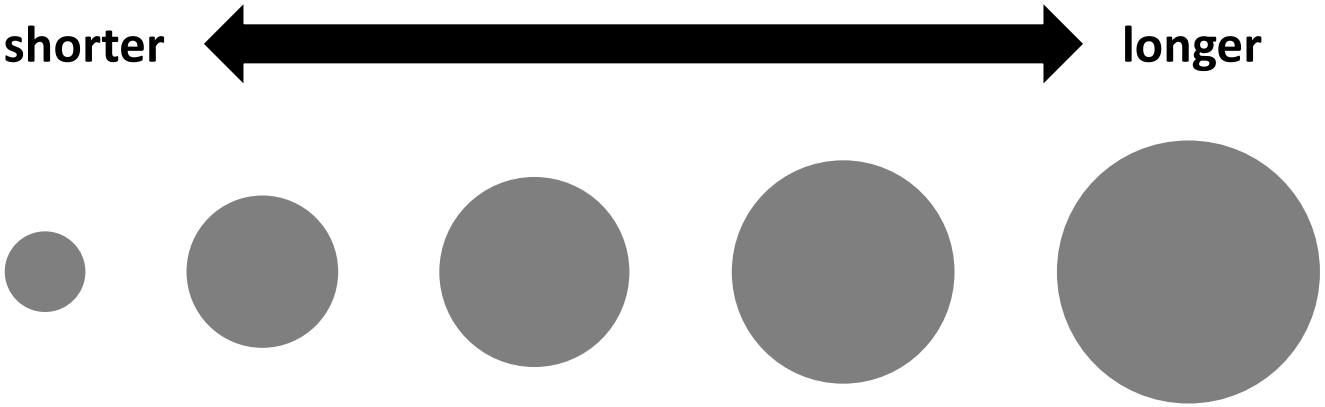
Frequency of the measure



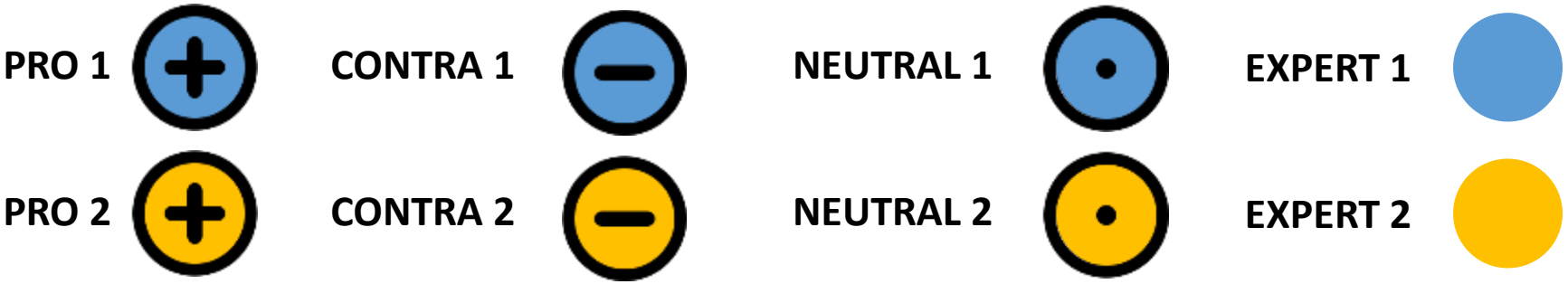


Average utterance length

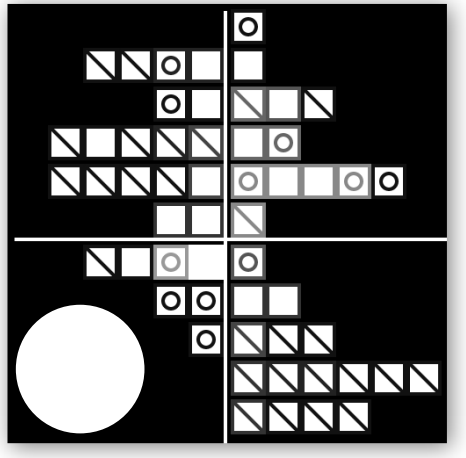




Speakers and speaker parties

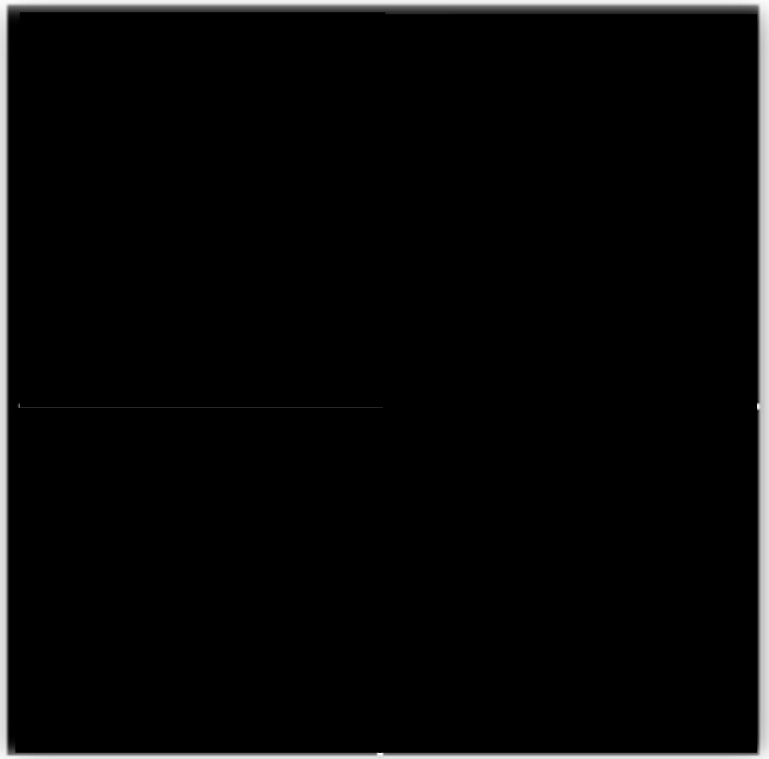


Discourse Maps



Accommodation

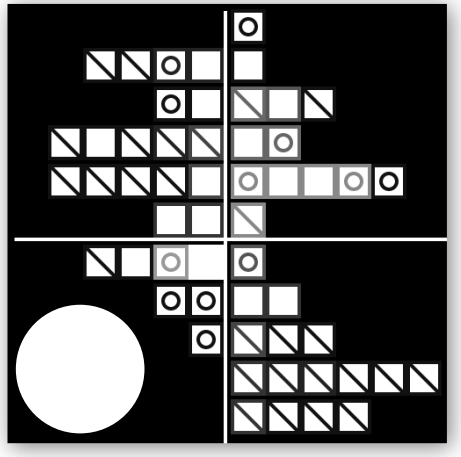
Atmosphere & Respect



Participation

Argumentation & Justification

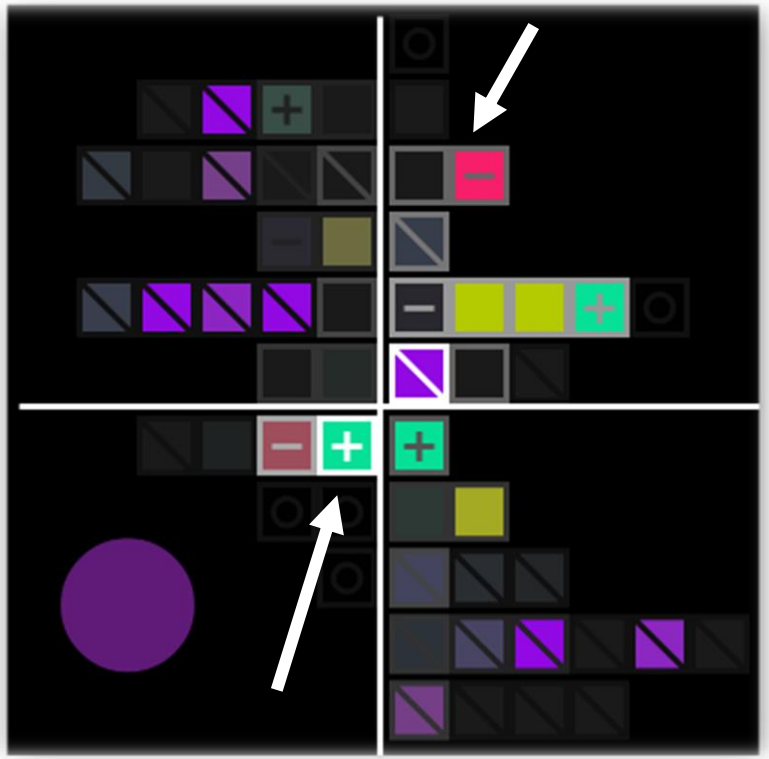
Discourse Maps



Emotion

Accommodation

Atmosphere & Respect



Participation

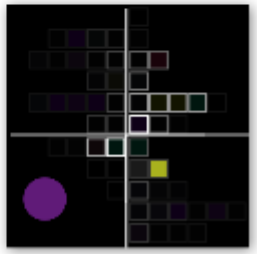
Argumentation & Justification

Average sentence complexity

Argumentation

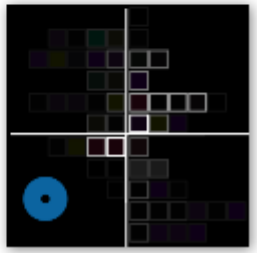
CONTRA

EXPERT



PRO

NEUTRAL



Neutral: Comparatively little Argumentation

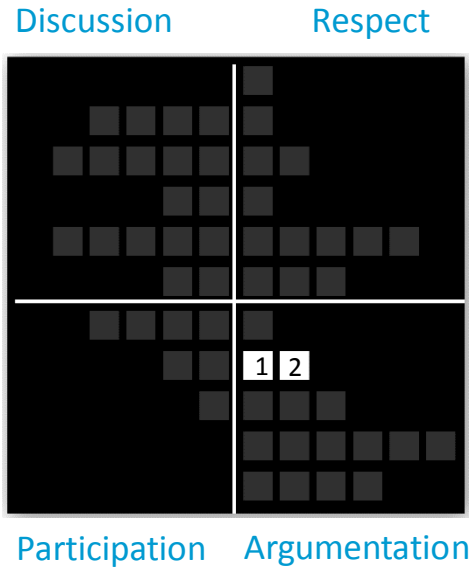
Experts: Justifications/Reasons

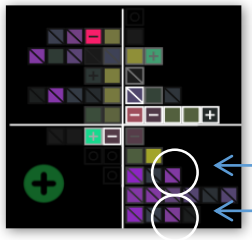
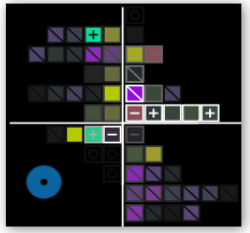
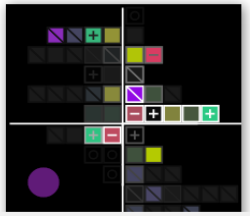
Pro & Contra: Conclusions and Justifications

→ **Deliberative Argumentation**

Measures shown:

1. **Conclusion**
2. **Justification**



	Haptbahnhof Gleis Tunnel Brücke
PRO	
CONTRA	
NEUTRAL	
EXPERT	

Information mainly from discourse particles and speech act verbs

Rejection of "Common" Ground

Refusal to provide information

Establishment of Common Ground

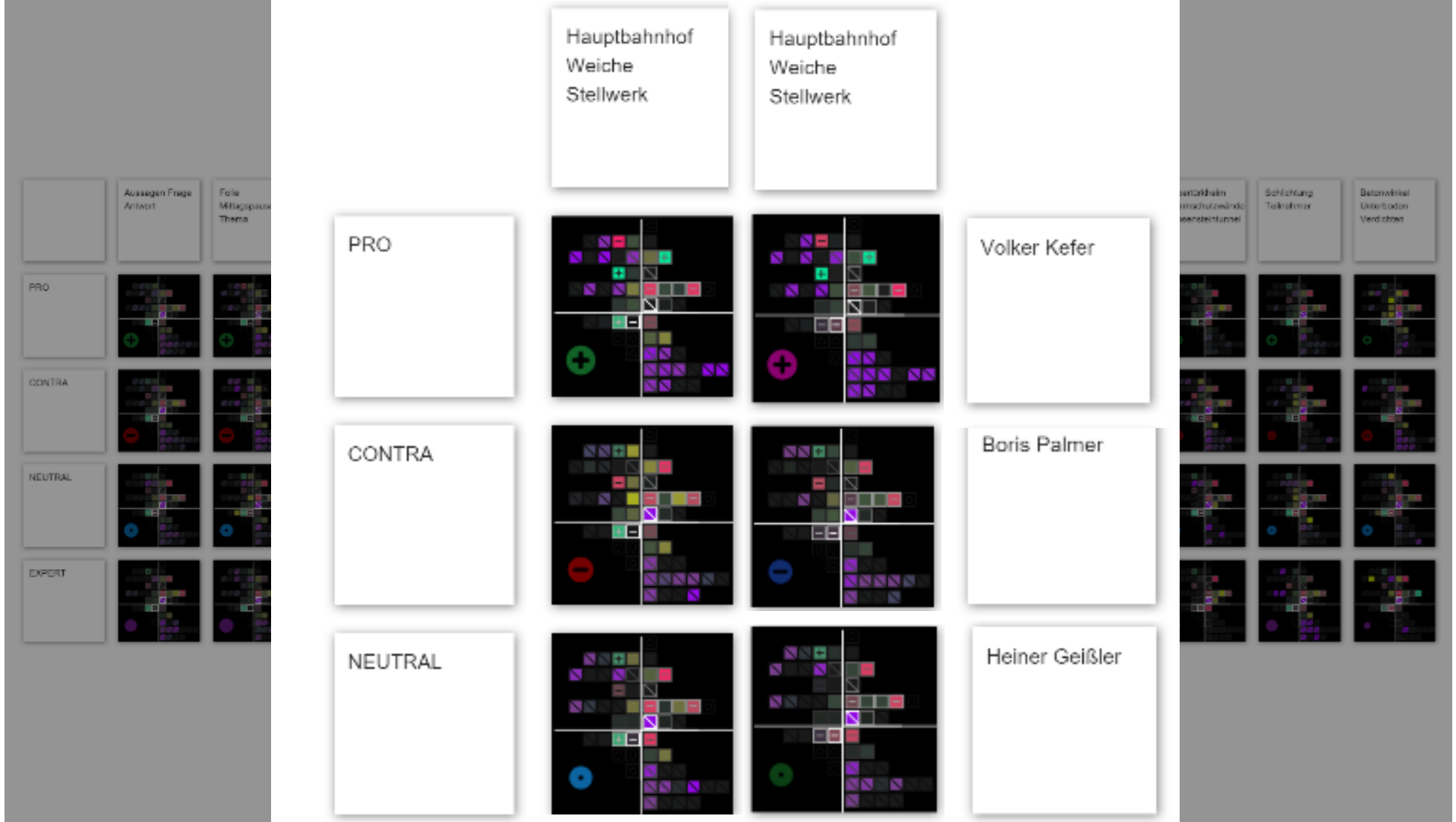
Pro, Contra & Neutral:
Establishment of Common Ground

Pro:
Rejection of "Common" Ground,
Refusal to provide information

→ controversial discussions

Glyphs per Group Type and Topics (Topic-Modelling)

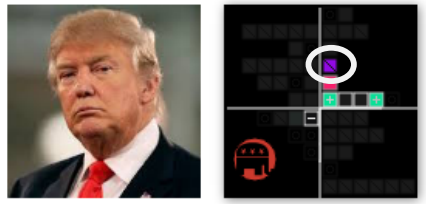
Can aggregate or look at individually



Who was interrupting the debates?

First Debate

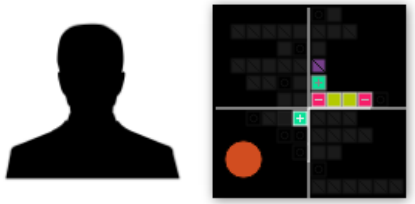
Trump
Position: REPUBLICAN



Utterances: 129

Average words: 148

Moderator
Position: NEUTRAL_M



Utterances: 97

Average words: 43

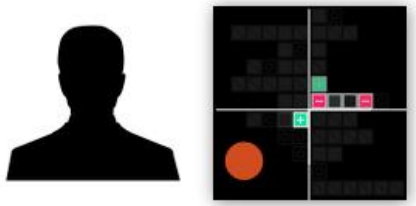
Clinton
Position: DEMOCRAT

Utterances: 94

Average words: 146

Second Debate

Moderator
Position: NEUTRAL_M



Utterances: 114

Average words: 49

Trump
Position: REPUBLICAN

Utterances: 88

Average words: 187

Clinton
Position: DEMOCRAT

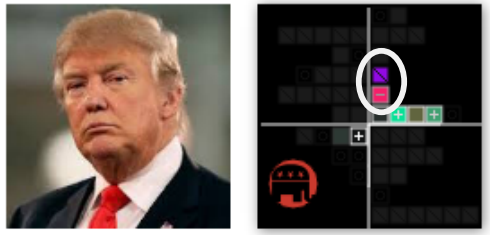
Utterances: 57

Average words: 240

Summary of all three debates

Trump

Position: REPUBLICAN



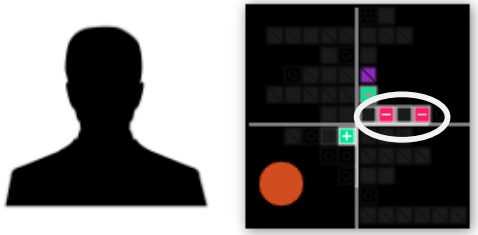
Utterances: 311

Average words: 158

- **High amount of interruptions**
- **Negative sentiment**

Moderator

Position: NEUTRAL_M



Utterances: 294

Average words: 55

- **Multiple topic shifts towards “previous” topics**

Clinton

Position: DEMOCRAT

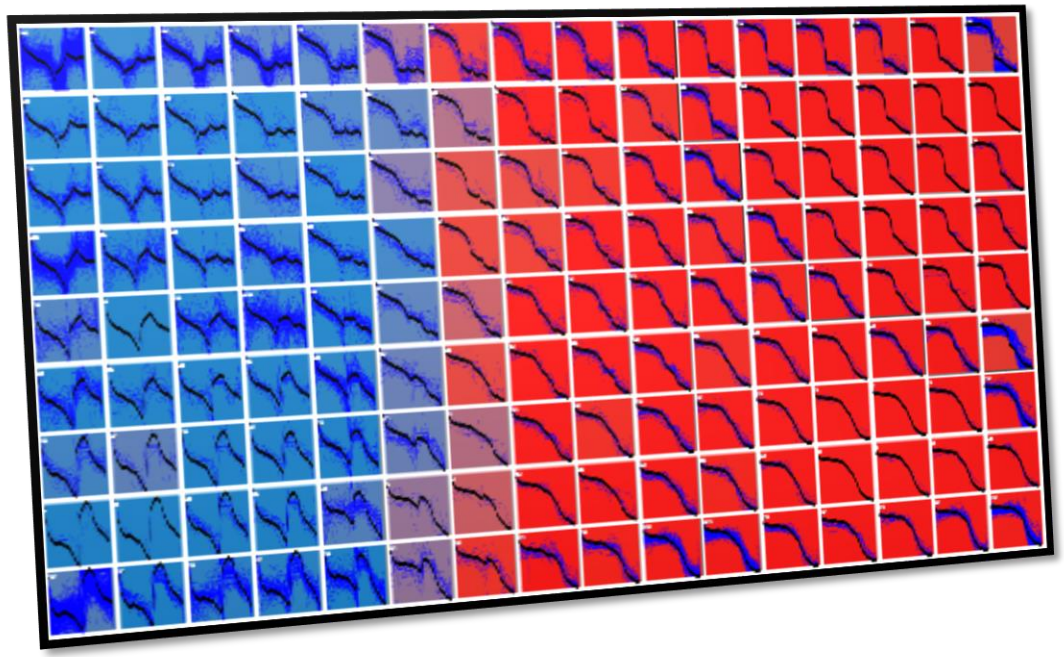
Utterances: 218

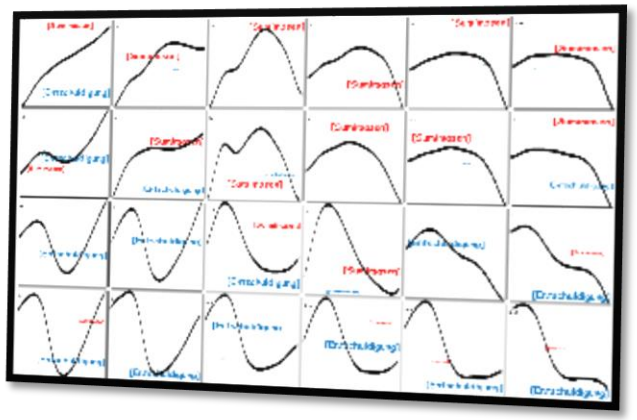
Average words: 191

- **High amount of filler words**
- **Low lexical diversity**

Visualization of spoken language

- So far we have been working with textual data. However, one can also work with spoken data.
- For Visual Analytics, all one needs is to have features (or vectors) that can be computed with



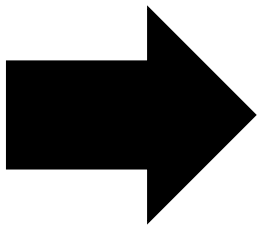


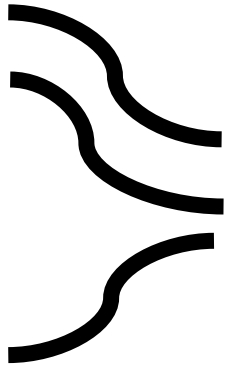
D. Sacha, Y. Asano, C. Rohrdantz, F. Hamborg, D. A. Keim, B. Braun and M. Butt. 2015. Self Organizing Maps for the Visual Analysis of Pitch Contours. Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa-2015), Vilnius, Lithuana, 2015.

Motivation

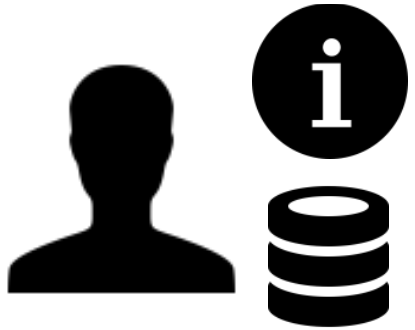
Experiments

Recording





Recording,
Landmarks,
Utterance,
Pitch-Vectors

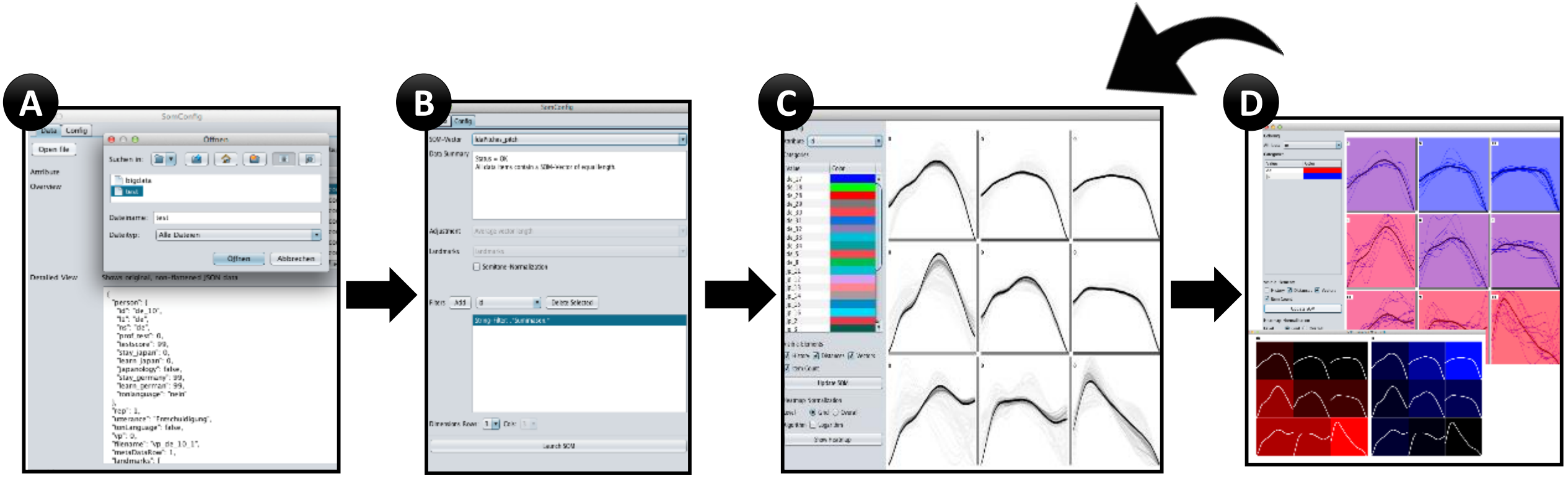


Speaker Information

Analysis

- Manual Inspection
- Statistics
- Functional Data Analysis
- **EXPLORATION**

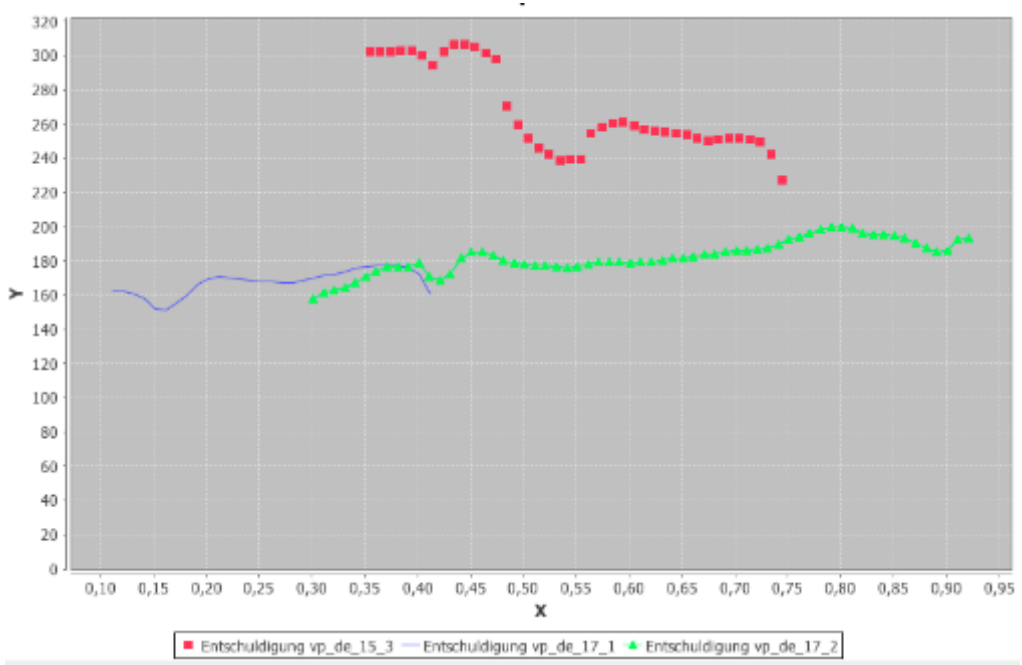
Our Approach - Workflow



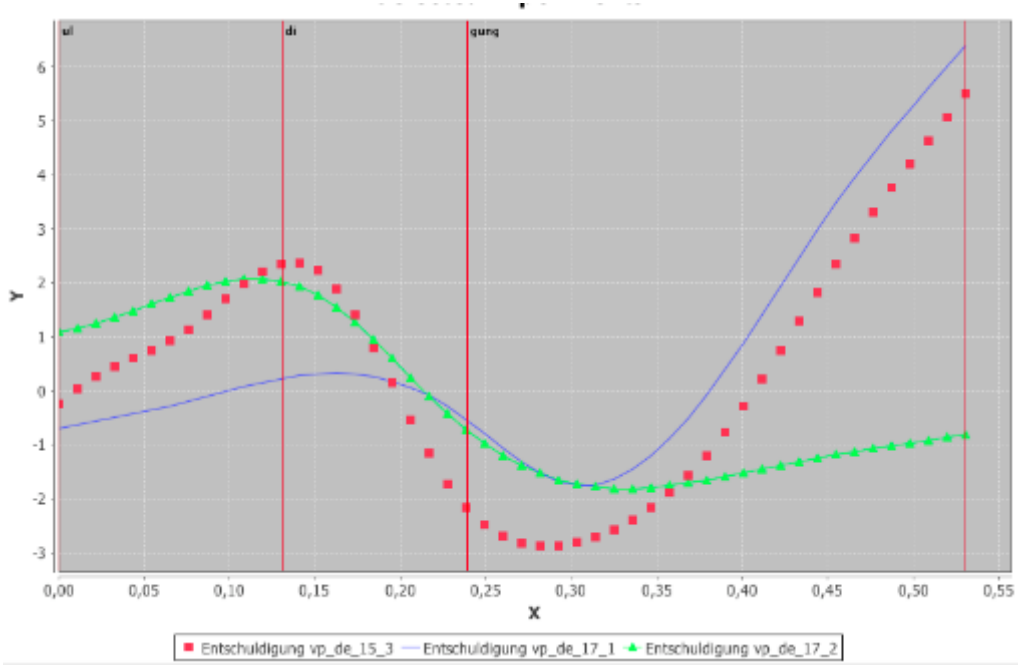
Data Input **Configuration** **Training** **Interactive Visualization**

Configuration and Preprocessing

Essential: Comparable Vectors & Distance Function

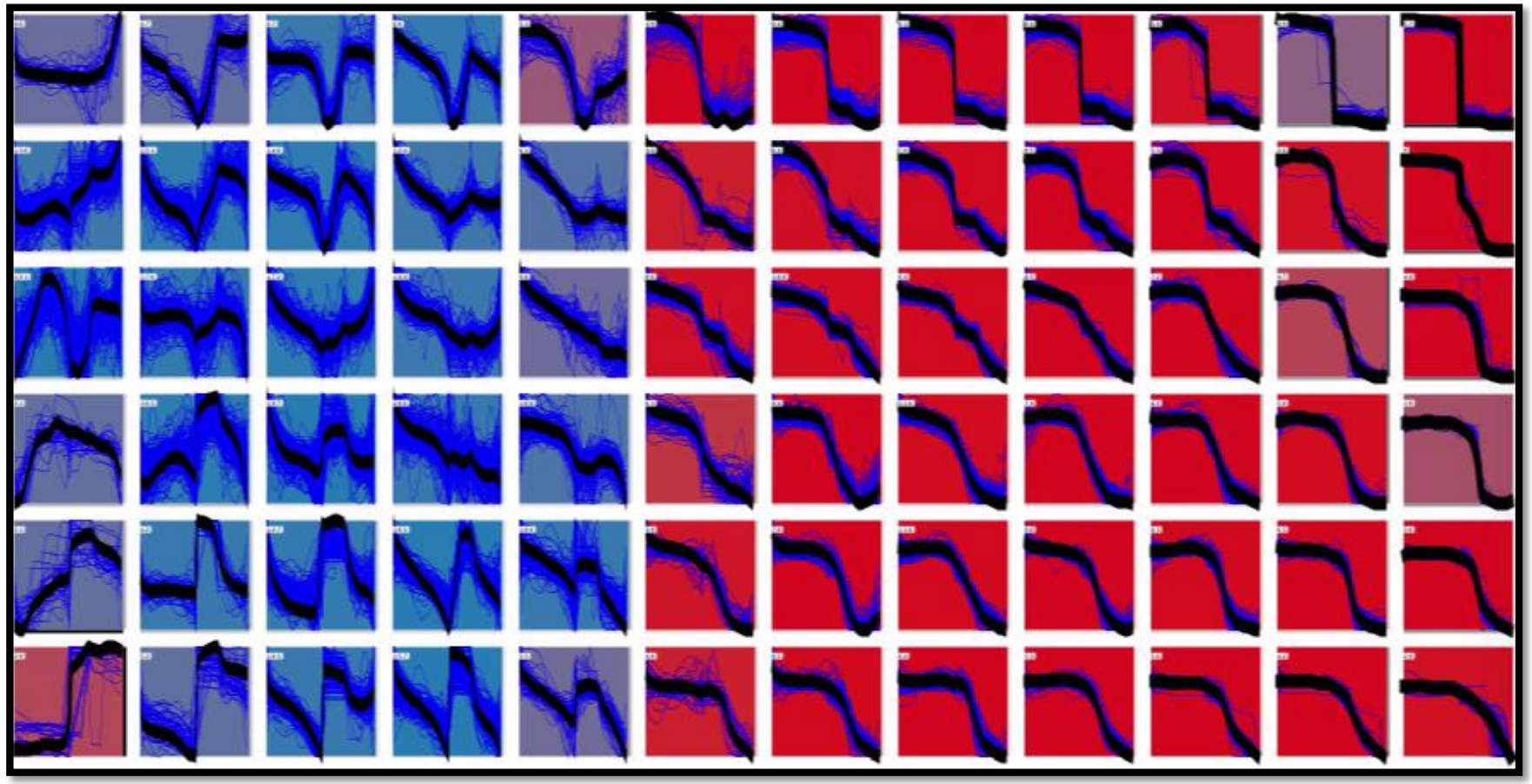


RAW-Pitches



Processed Pitches

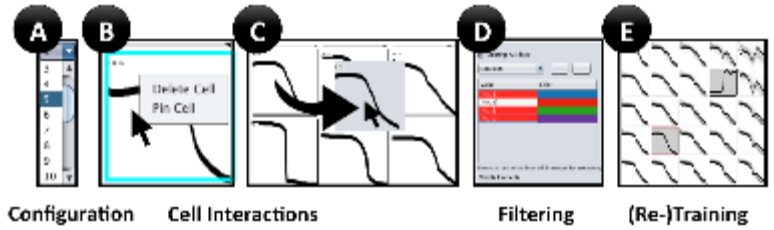
Iterative SOM Training and Interaction



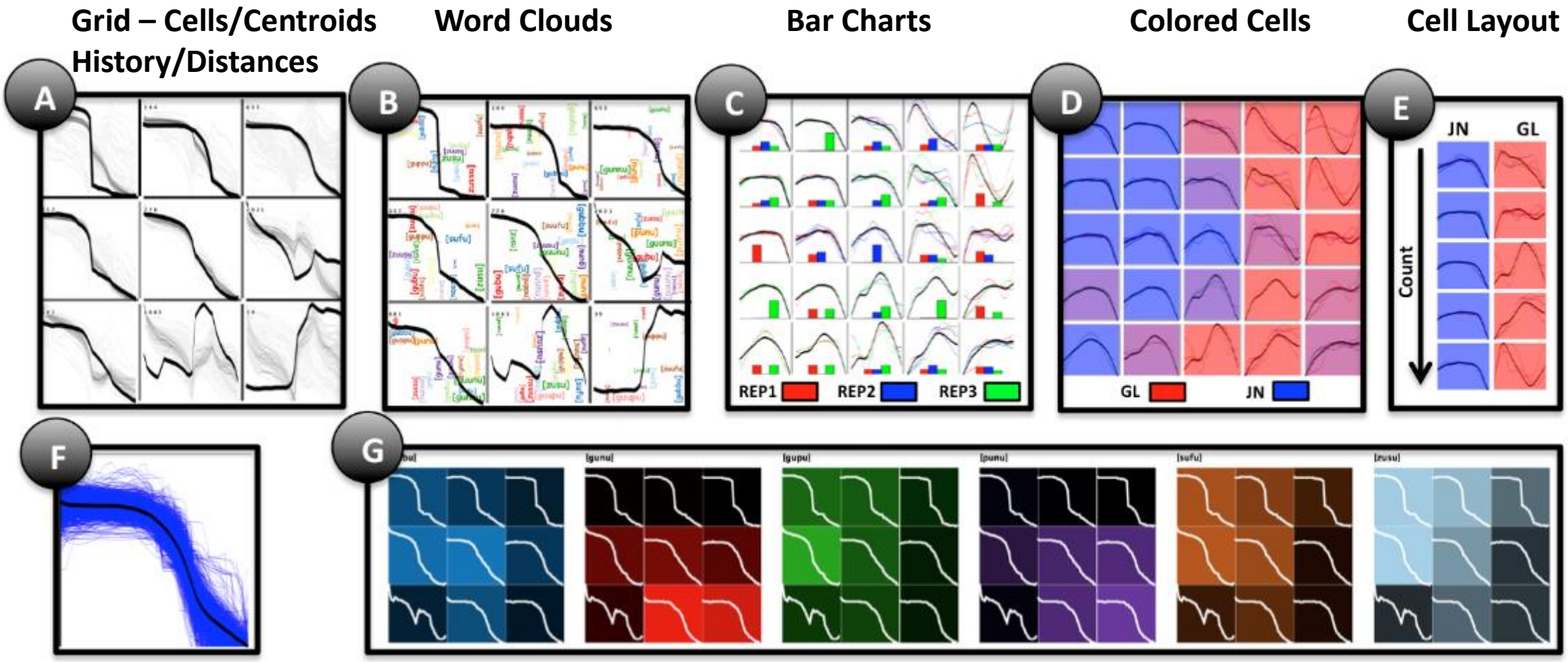
Data Exploration

- SOM learning is fast
- user can switch among different perspectives on the data
- user can interactively delete or pin cells
- and retrain and re-explore

Interactions



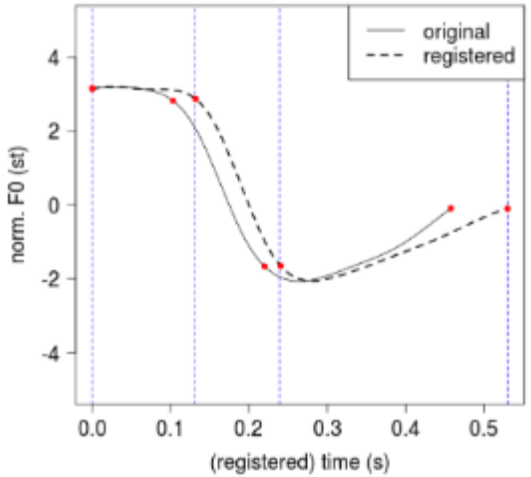
Interactive Visualizations



Use Case

Experiment - Data

FDA



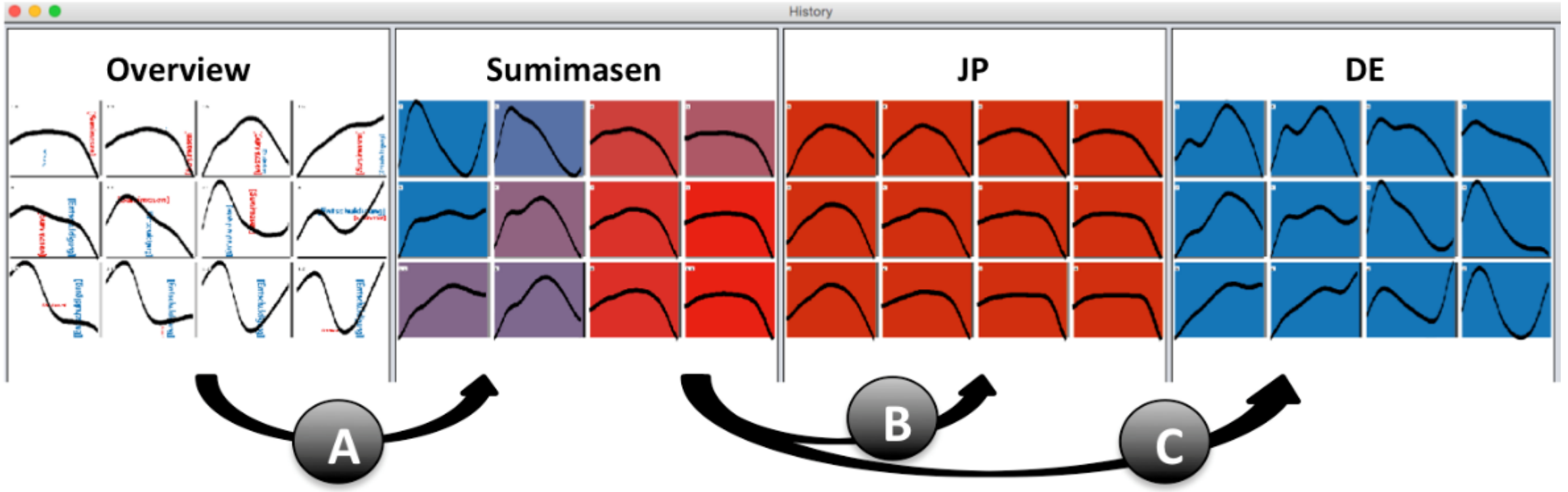
Pre-Processing

Native Speakers vs. Learners

Are there Differences?

Tasks

Use Case – Two Separate SOMs are Trained for Japanese and German Speakers



Speakers pronounced "*sorry/excuse me*" in ever more exasperating circumstances

- Japanese natives do not vary the pitch contour (red SOM - B)
- German learners do vary the pitch (blue SOM - C)

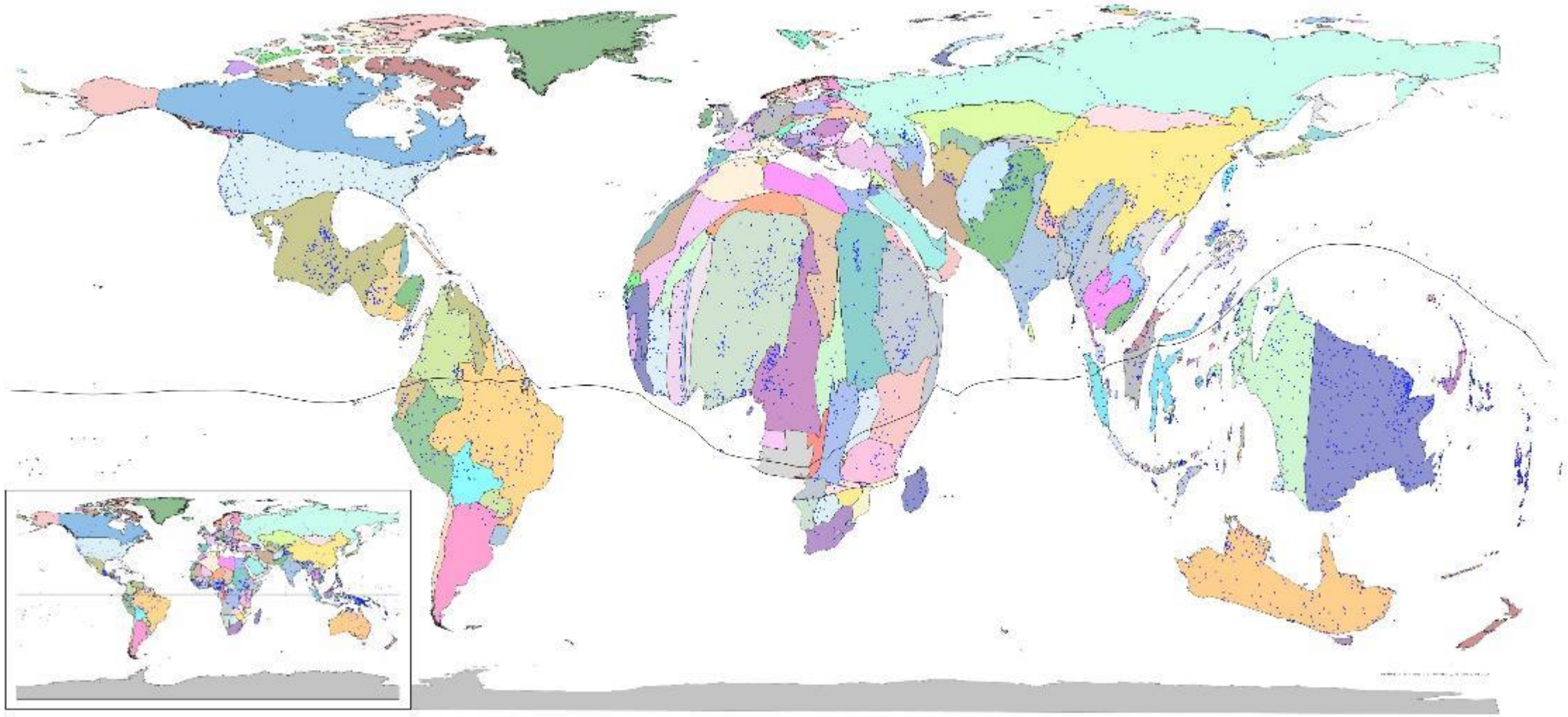
Identifying Optimal Visualizations

- Understanding which visualizations are optimal is not trivial.
- Are the individual dimensions (color, shape, direction, size, etc.) usefully meaningful?
- Does the visualization allow for at-a-glance understanding, or does it confuse the user?

- This also depends on the user's background
  - What is the user used to looking at?
  - How is the user used to understanding the data?
  - How is the user used to interacting with systems?

- Currently evaluations are mainly performed via user studies in Visual Analytics.
- Project (SFB/TRR 161): Establish evaluation metrics.

Distorted Map according to number of languages spoken in area.



Visualization only as good as your data – India massively underrepresented

Outlook

- Further exploration of possibilities offered by Visual Analytics
  - The systems illustrated here are recent
  - Interactive exploratory linguistic analysis is on-going
  - Systems are being fine-tuned
- Workflow
  - Use cases for Digital Humanities /eHumanities are being developed
  - **Infrastructure** Platforms (mix and match the available tools)
- Measuring Success
  - Development of **Evaluation Metrics** for LingVis.
  - Use cases, work flow and result comparison.

What interests Visualizers?

- Need interesting interactions
- Multiple dimensions
- Time depth
- Cross-modular interactions.
- Not just coloring in bits of text that are of interest for linguists.

Summary

- Example of different kinds of visualizations
- These visualizations allow a new approach to linguistic data
- Flexible, interactive, make use of the highly skilled human perceptual system
- More examples to follow.

Now first some design basics!

THANK YOU!

Questions?