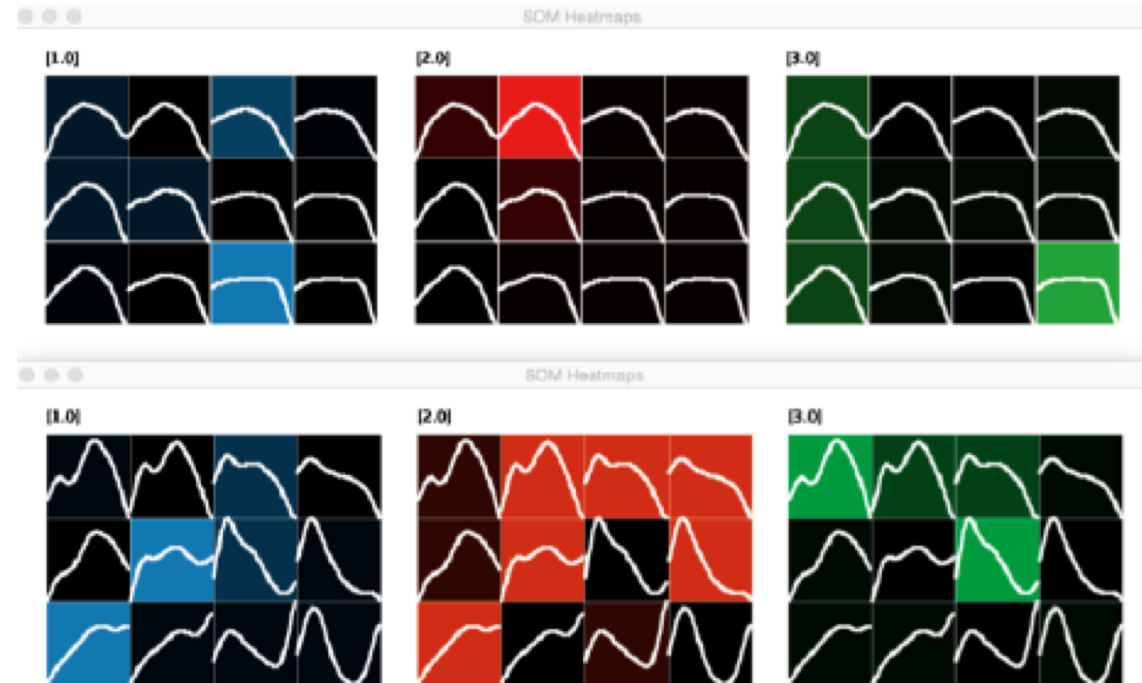


Visualizing Linguistic Structure (LingVis)



Miriam Butt

7th International Conference on Language
and Technology (CLT), February 2020
UET, Lahore

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).
- **But:** as yet only comparatively little use of new **visualization methods**.

Sample Visualizations

EMDialog: Bringing Information Visualization into the Museum

12

display technology carr perspective

data exploration
node
data representation

statement
ring
exhibition

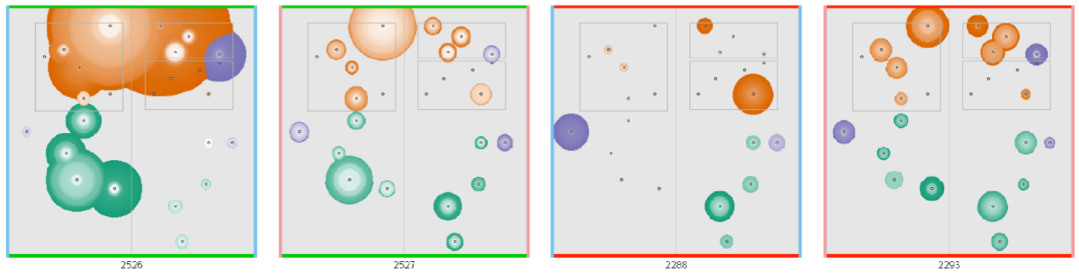
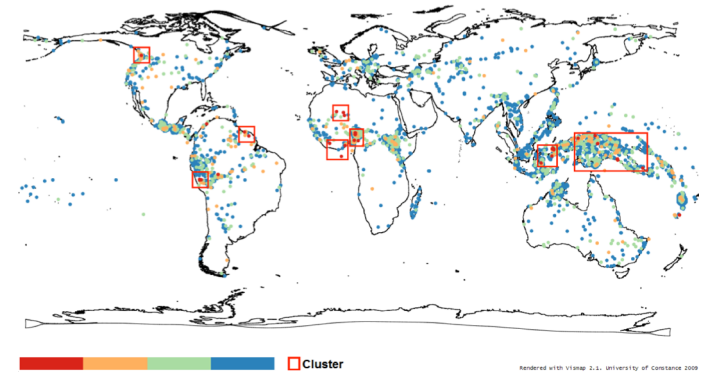
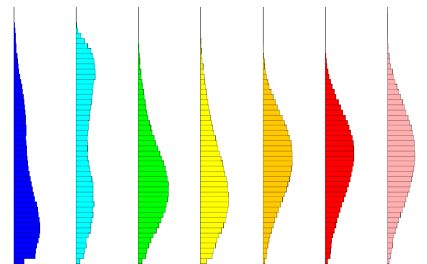
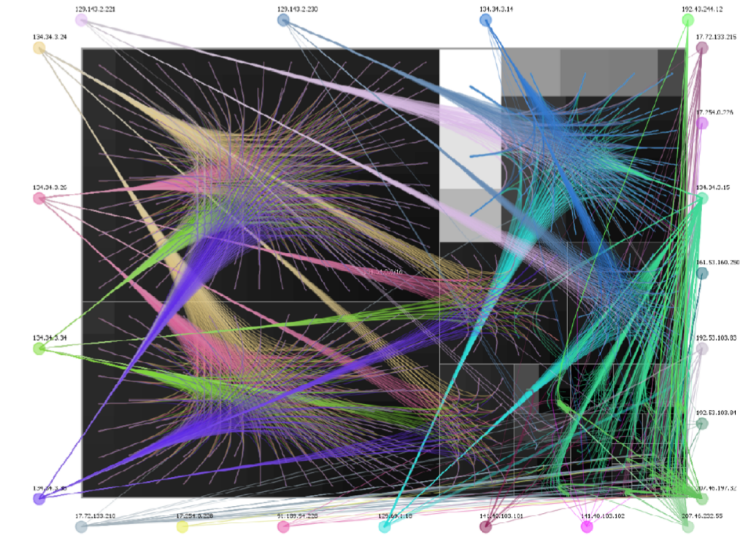
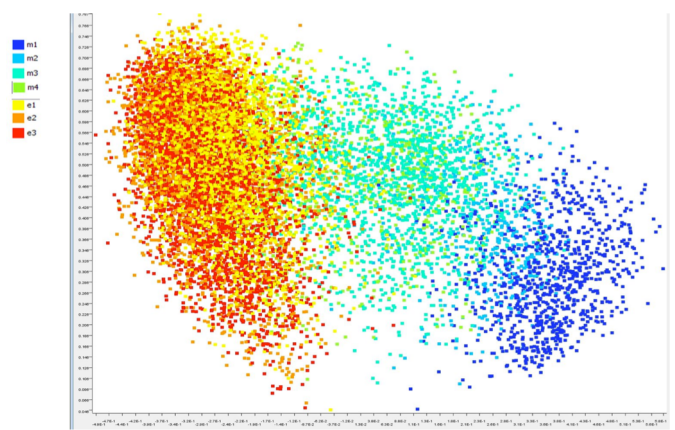
work

installation
tree diagram
information presentation

long-term exploration
public space

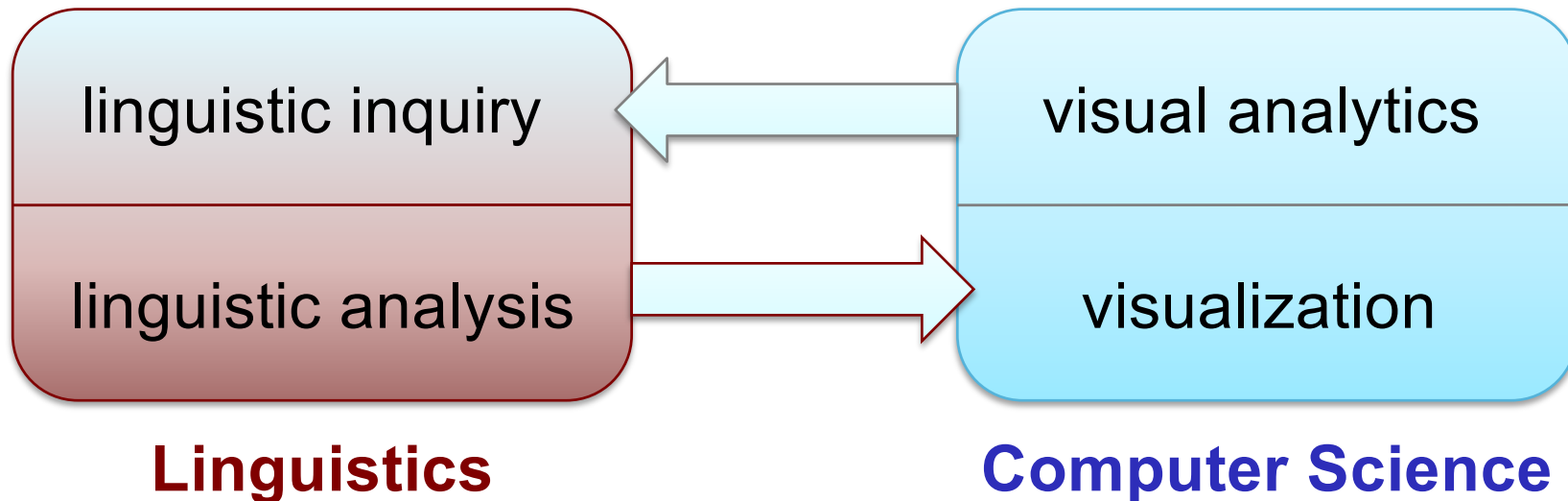
period people

Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale



Overall Interdisciplinary Goal

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

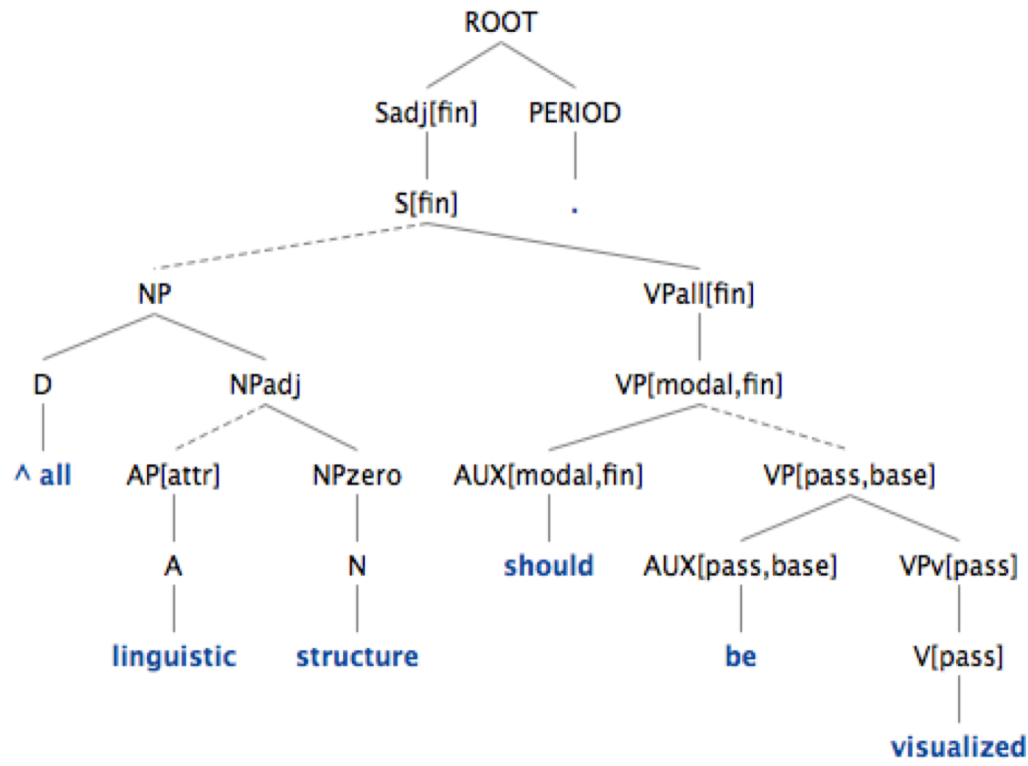


Visual Analytics:

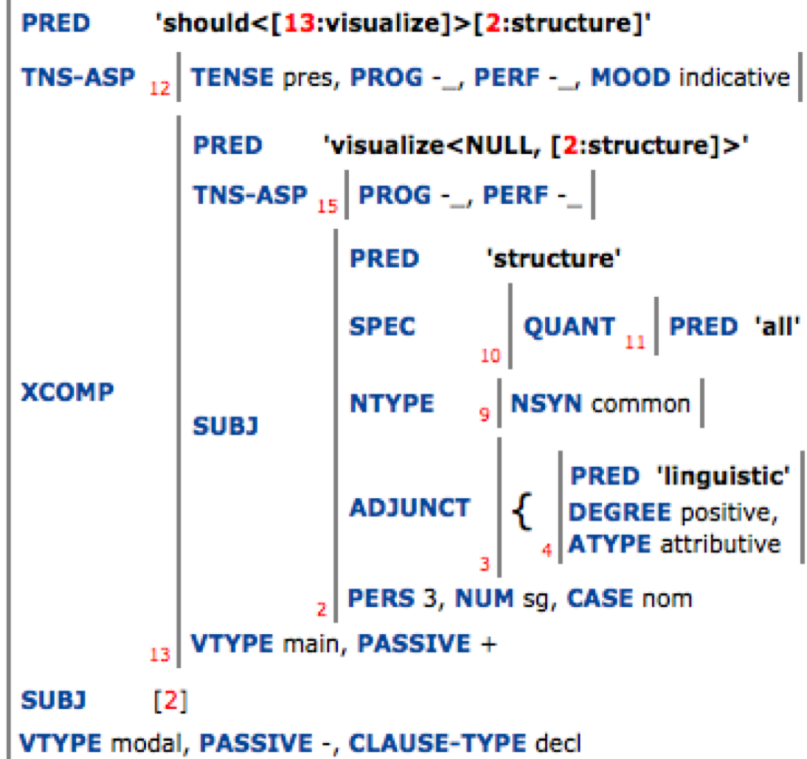
- Interactive exploratory visual access to data.
- Iterations of hypothesis-formation and hypothesis-testing.

Existing Visualization: Syntax

C-structure



F-structure



Syntactic Analysis with Lexical-Functional Grammar (LFG)

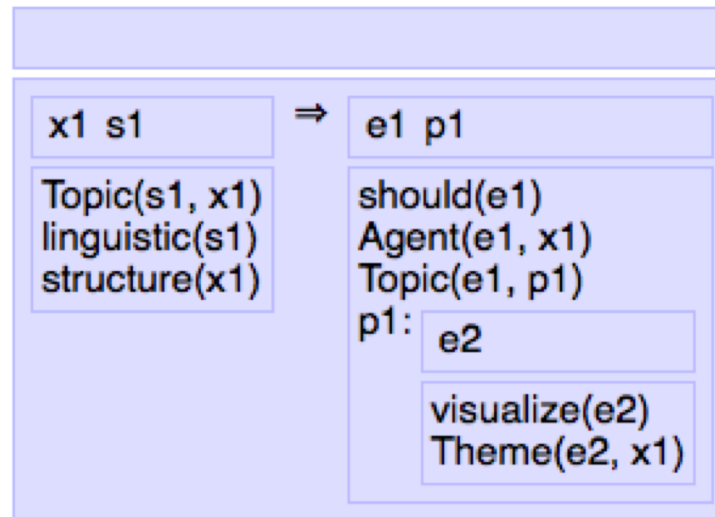
<http://iness.uib.no/iness/xle-web> (Web Interface for LFG Grammars)

Grammar developed at PARC

Existing Visualization: Semantics

DRS

(Discourse Representation Structure)

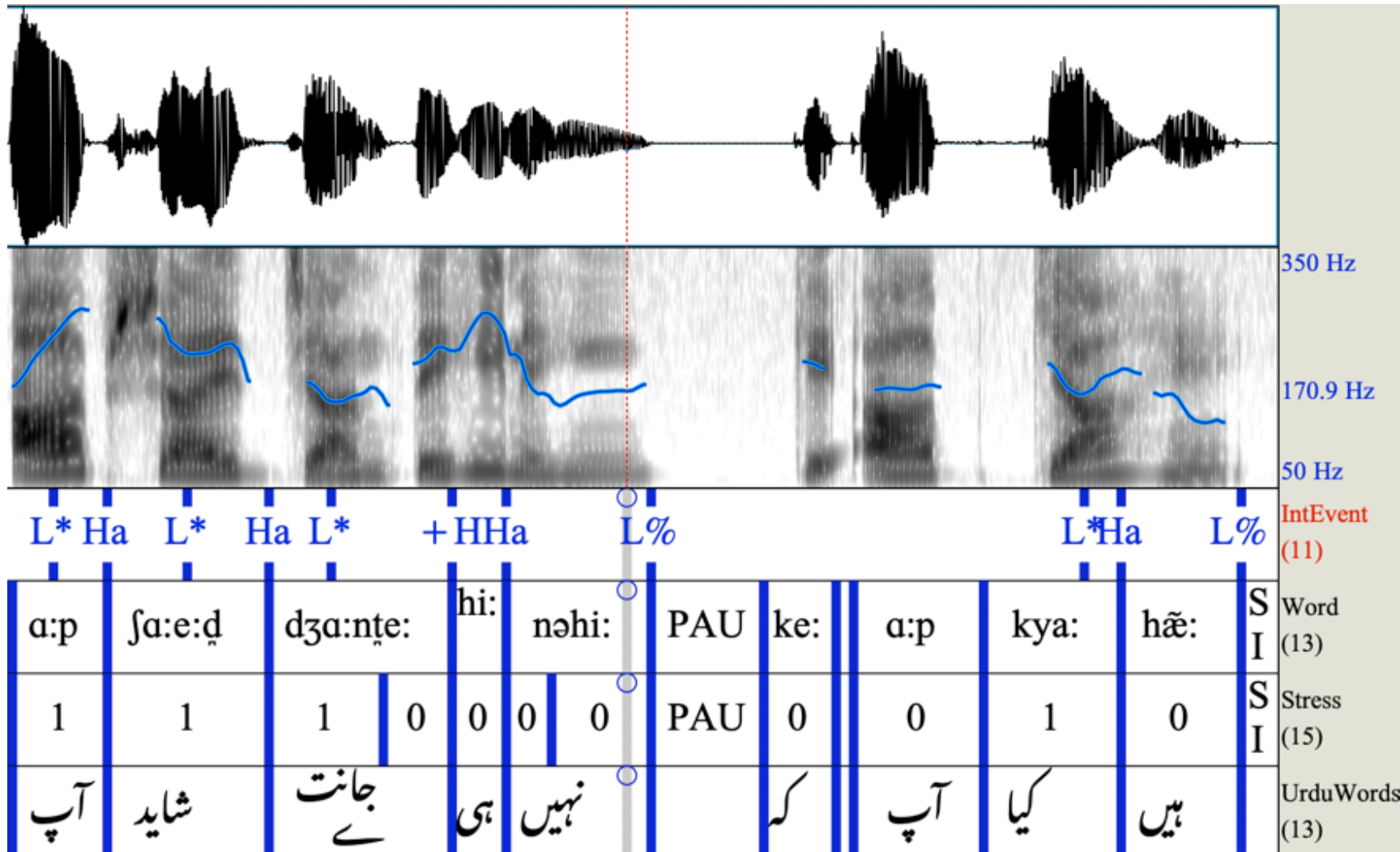


Semantic Analysis with Discourse Representation Theory (Boxer)

<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



(from CLE Corpus of Spoken Urdu)

Les
IV
DECLINAISONS

La I^{re} Déclinaison par le moyen de l'Article **DEFIN.**
Masculin. Singulier la
Feminin. Singulier la
Nom. Accus. de la
Gen. Abl. de la
Dat. au à la
Maj. ou Tom. devant une Vieille.

La II^e Déclinaison par le moyen de l'Article **INDEFINI.**
Masculin. Singulier le
Feminin. Singulier la
Nom. Accus. de
Gen. Abl. de
Dat. à l'
Plurier pour tous.
Nom. Acc. les
Gen. Abl. des
Dat. aux

La III^e Déclinaison par le moyen de l'Article **UNITE**
Masculin. Singulier un
Feminin. Singulier une
Nom. Accus. d'un
Gen. Abl. d'une
Dat. à un à une

La IV^e Déclinaison par le moyen de l'Article **PARTITIF**
Masculin. Singulier de la
Feminin. Singulier de la
Nom. Accus. de
Gen. Abl. de
Dat. à du à de la.
Maj. ou Fem. devant une Vieille.

La I^{re} Déclinaison par le moyen de l'Article **DEFIN.**
Masculin. Singulier un
Feminin. Singulier une
Nom. Accus. d'un
Gen. Abl. d'une
Dat. à un à une

La II^e Déclinaison par le moyen de l'Article **INDEFINI.**
Masculin. Singulier le
Feminin. Singulier la
Nom. Accus. de
Gen. Abl. de
Dat. à l'
Plurier pour l'un et pour l'autre

La III^e Déclinaison par le moyen de l'Article **UNITE**
Masculin. Singulier un
Feminin. Singulier une
Nom. Accus. d'un
Gen. Abl. d'une
Dat. à un à une

La IV^e Déclinaison par le moyen de l'Article **PARTITIF**
Masculin. Singulier de la
Feminin. Singulier de la
Nom. Accus. de
Gen. Abl. de
Dat. à du à de la.
Maj. ou Fem. devant une Vieille.

Il n'a point de Plurier
Exemples.
I. Déd. le Père, la Mère, l'enfant, l'âme, l'homme, l'homme, l'homme.
II. Jean, Adam, Marie, Anne, Martin, moi, toi, soi, mon, ton, son.
III. Déd. un livre, une plume, un enfant, une heure.
IV. du pain, de la viande, de l'eau, des pommes, des poires.

Nouvelle
GRAMMAIRE
à l'usage
des DAMES

CONJUGAISON
de tous
les Verbes réguliers.

Participe actif ant issant evant -ant	INFINITIF. 1. Er, donner, aimer, chercher, 2. Ir, bâtir, finir, punir, agir. 3. Oir, devoir, ve devoir, apercevoir 4. Re, vendre, rendre, attendre	Participe passif e eu
Le Verbe assistant AVOIR. Part. act. ayant Part. pass. eu Imperat. ayez qu'il ait ayons, ayez, qu'ils aient	EXACTEUR ou IMPERATIF. e, e, ons, ez, ent is, isse, issons, issez, issent ois, ois, evous, evous, evous -s, -e, -ons, -ez, -ent	Le Verbe assistant ETRE. Part. act. étant Part. pass. été Imperat. sois qu'il soit soyons, soyez, qu'ils soient
Fai ¹ tu as il a Nous avons Vous avez ils ont	Fais ² tu ayes il ait Nous ayons Vous ayez ils aient	Temps Présent absolu. Temps Présent dépendant e, es, e, ons, ez, ent; e, es, c, ions, iez, ent is, is, it, issons, issez, issent; se, ses, se, sions, siez, sent ois, ois, oit, evous, evous, evouent; ois, ois, ois, evous, evous, evouent -s, -s, -s, -ons, -ez, -ent; -e, -es, -e, -ions, -iez, -ent
J'avois tu avois il avoit Nous avions Vous aviez ils avoient.	J'avois tu avois il avoit Nous avions Vous aviez ils avoient.	Temps passé stable ois, ois, oit, ions, iez, oient.
J'aurais tu aurais il aurait Nous aurions Vous auriez ils auraient.	J'aurais tu aurais il aurait Nous aurions Vous auriez ils auraient.	J. Futur conditionnel crois, rois, roit irois, rions, riez -rois, roient.
J'aurais tu aurais il aurait Nous aurions Vous auriez ils auraient.	J'aurais tu aurais il aurait Nous aurions Vous auriez ils auraient.	J. Futur absolu rai, ras, ra rons, rez, ront.
J'eus tu eus il eut Nous eumes Vous eûtes ils eurent.	J'eusse tu eusses il eût Nous eussions Vous eussiez ils eussent.	J. passé historique absolu ai, as, a, ames, ales, érent is, is, it, imes, ites, irent us, us, ut, umes, ules, urent is, is, it, imes, ites, irent
J'eus tu eus il eut Nous eumes Vous eûtes ils eurent.	J'eusse tu eusses il eût Nous eussions Vous eussiez ils eussent.	J. passé historique dépendant ai, as se, ses, sions, siez sent.
J'eus tu eus il eut Nous eumes Vous eûtes ils eurent.	J'eusse tu eusses il eût Nous eussions Vous eussiez ils eussent.	Je serois tu serois il seroit Nous serions Vous seriez ils seroient.
J'eus tu eus il eut Nous eumes Vous eûtes ils eurent.	J'eusse tu eusses il eût Nous eussions Vous eussiez ils eussent.	Je serois tu serois il seroit Nous serions Vous seriez ils seroient.

Les Verbes Passifs se conjug. avec les Temps du Verbe Etre, tant simples que composés.



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
- **Current project:**
 - Identify good use cases within linguistics
 - Understand evaluation (SFB/TRR 161)

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

1	#this file lists X in X+kar, X+ho, X+hu, X+rakh sequences with corresponding occurrences in the (candidate) CP sequences
2	#X = word occurring directly to the left of LV (LV: kar, ho, hu, rakh)
3	#kar: # of occurrences of X with kar
4	#ho: # of occurrences of X with ho
5	#hu: # of occurrences of X with hu
6	#rakh: # of occurrences of X with rakh
7	X #hu #kar #ho #rakh
8	مفاضلہ 674 466 524 0
9	غور 378 2336 1691 0
10	دولت 366 254 609 0
11	مکامد 359 135 44 0
12	حلم 227 1232 100 0
13	رہنما 183 178 765 0
14	ناصرین 173 0 114 0
15	ایک 172 373 7027 0
16	تباہ 147 394 588 0
17	تقوی 142 105 235 9
18	ادیب 103 754 956 0
19	کالہ 102 1501 3609 0
20	دم‌آرب 80 210 96 0
21	اھکر 74 0 263 0
22	یہ 62 59 1161 0
23	زاغاً 59 315 75 0
24	سی 56 0 2267 0
25	دق‌عزم 54 197 262 0
26	فاش‌کنا 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

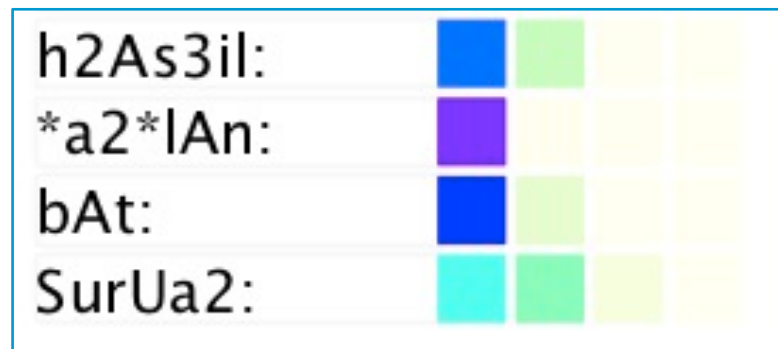
Statistical Data:

ID	Noun	Rel. freq. with <i>kar</i>	Rel. freq. with <i>ho</i>	Rel. freq. with <i>hu</i>	Rel. freq. with <i>rak^h</i>
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

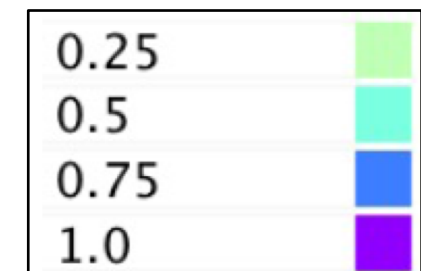
Table 2: Relative frequencies of co-occurrence of nouns with light verbs

do, be, become, put
kar, ho, hu, rakh

'achievement'
'announcement'
'talk'
'beginning'



Color Scale



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

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

h2Asil	■	■	■	■
*a2*IAAn	■	■	: *a2*IAAn: 0.011 (ho)	
bAt	■	■	■	■
SurUa2	■	■	■	■

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

2: *x*rAjAt	■	■	■	■
2: s*p<vao>r*	■	■	■	■
2: uTHA	■	■	■	■
2: k*l<ye>**r	■	■	■	■

Vowel Harmony (VH)

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

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-ın, top-lar-ın

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

1st 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.
- **Results** show that Turkish and Hungarian, for example, pattern similarly. Languages like Spanish or German pattern differently.

Results — Standard Methods: What is the 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

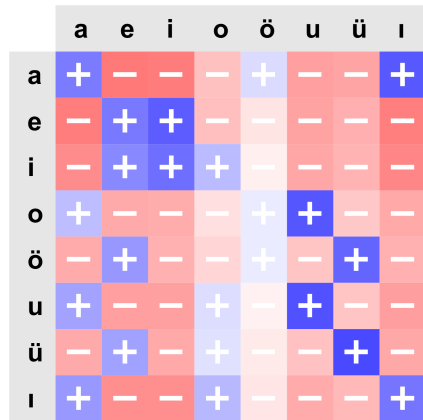
	a	o	i	ü	ö	ä	u	e
a	0.019	0.009	-0.061	-0.034	-0.008	-0.025	0.018	0.035
o	-0.023	-0.004	-0.052	-0.013	-0.020	-0.013	-0.013	0.068
i	-0.069	-0.054	-0.050	-0.039	-0.036	-0.044	-0.003	0.133
ü	-0.067	-0.045	0.070	-0.028	-0.021	-0.033	-0.021	0.050
ö	-0.049	-0.032	0.049	-0.024	-0.013	-0.021	-0.013	0.036
ä	-0.067	-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

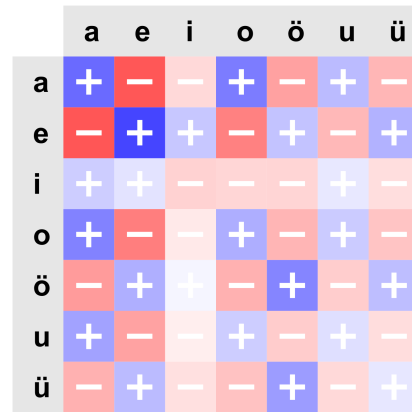
	a	o	u	i	ü	ö	e
a	0.339	0.263	0.070	-0.022	-0.081	-0.136	-0.431
o	0.239	0.099	0.041	-0.007	-0.052	-0.083	-0.253
u	0.132	0.038	0.015	-0.004	-0.017	-0.040	-0.131
i	0.037	-0.026	0.008	-0.030	-0.017	-0.027	0.011
ü	-0.093	-0.056	-0.022	-0.014	0.008	0.148	0.071
ö	-0.152	-0.093	-0.037	0.001	0.065	0.229	0.097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

Hungarian

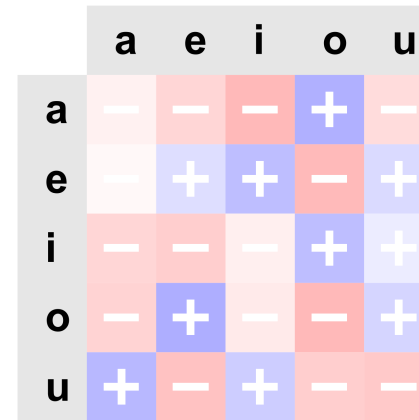
First Visualization: Can you detect a pattern?



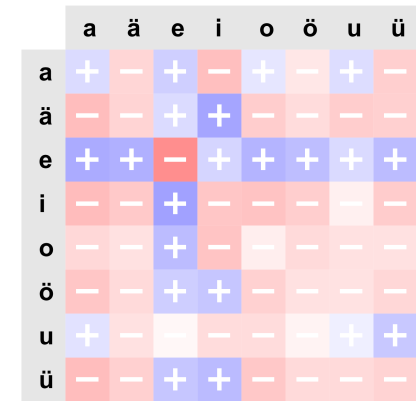
Turkish



Hungarian



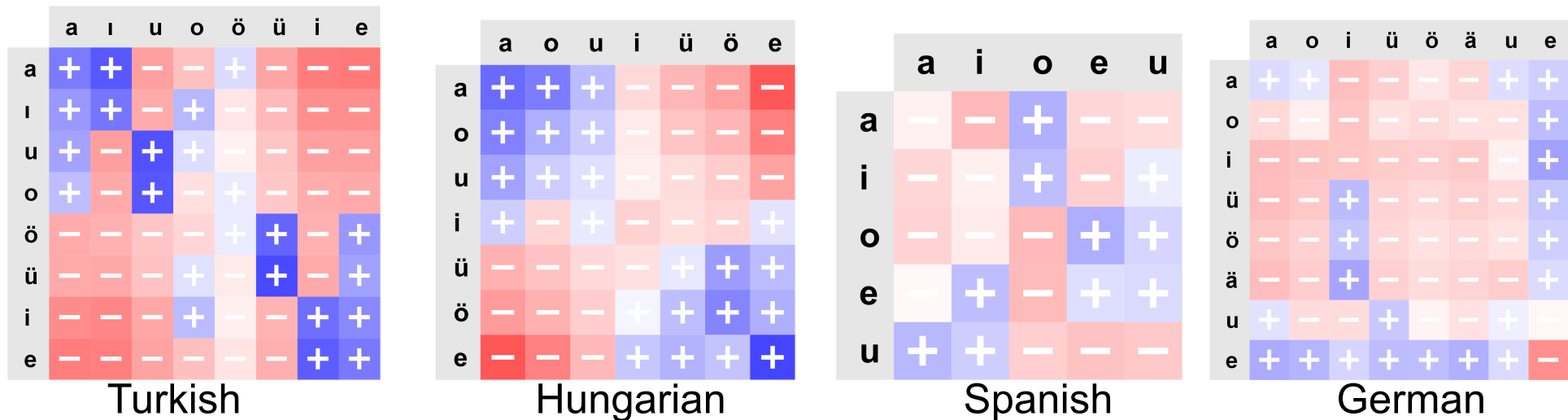
Spanish



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?



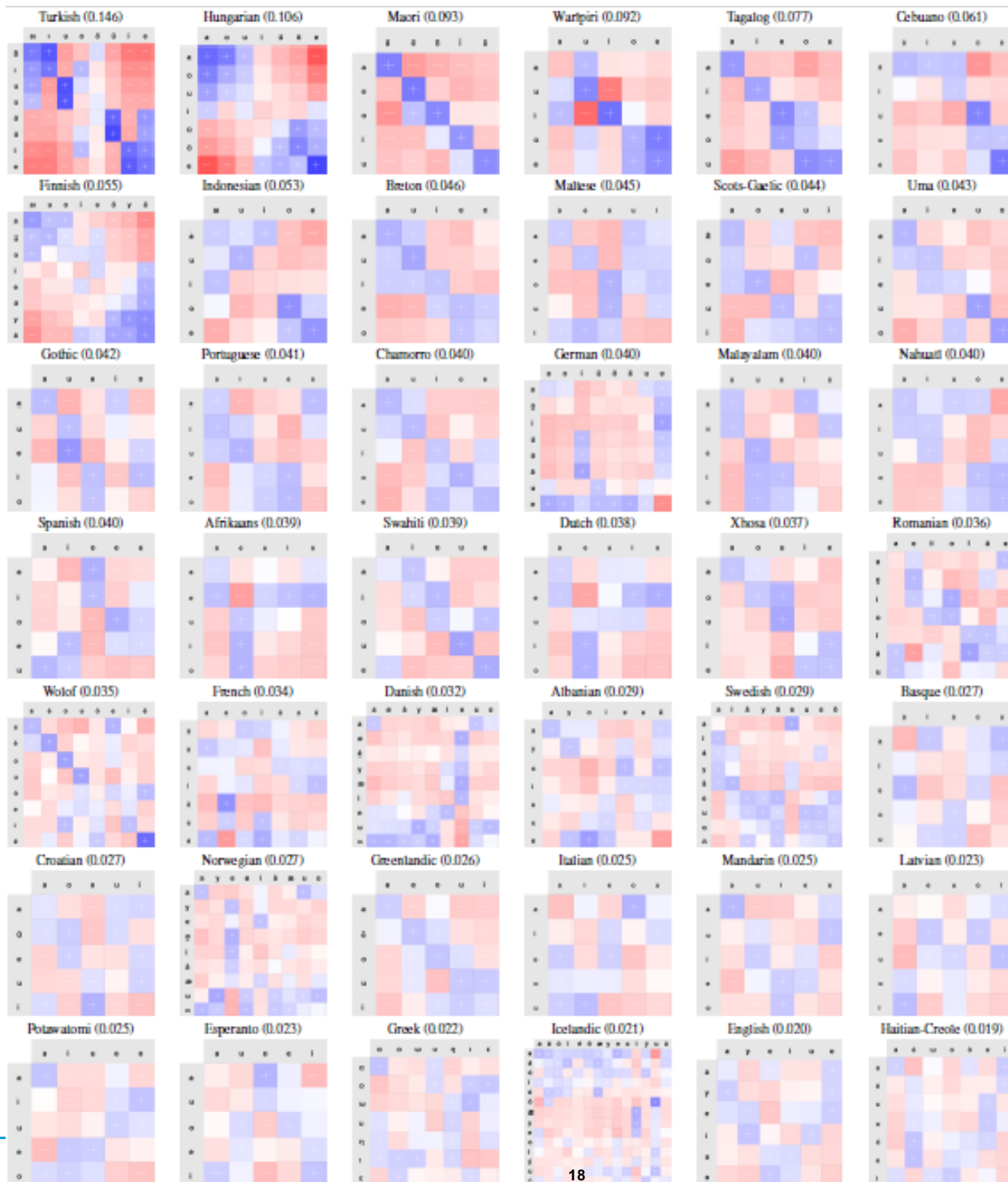
- Vowels **sorted** according to similarity (note: not a trivial process).
- Sorting done according to feature vectors of each of the rows.
- Can even see the **type** of Vowel Harmony involved.

Thomas Mayer, Christian Rohrdantz, Miriam Butt, Frans Plank & Daniel A. Keim. **Visualizing Vowel Harmony**. *Linguistic Issues in Language Technology*, 4(Issue 2):1-33, 2010.

Results – Sorted Visualization:

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

Best viewed on a
Powerwall!

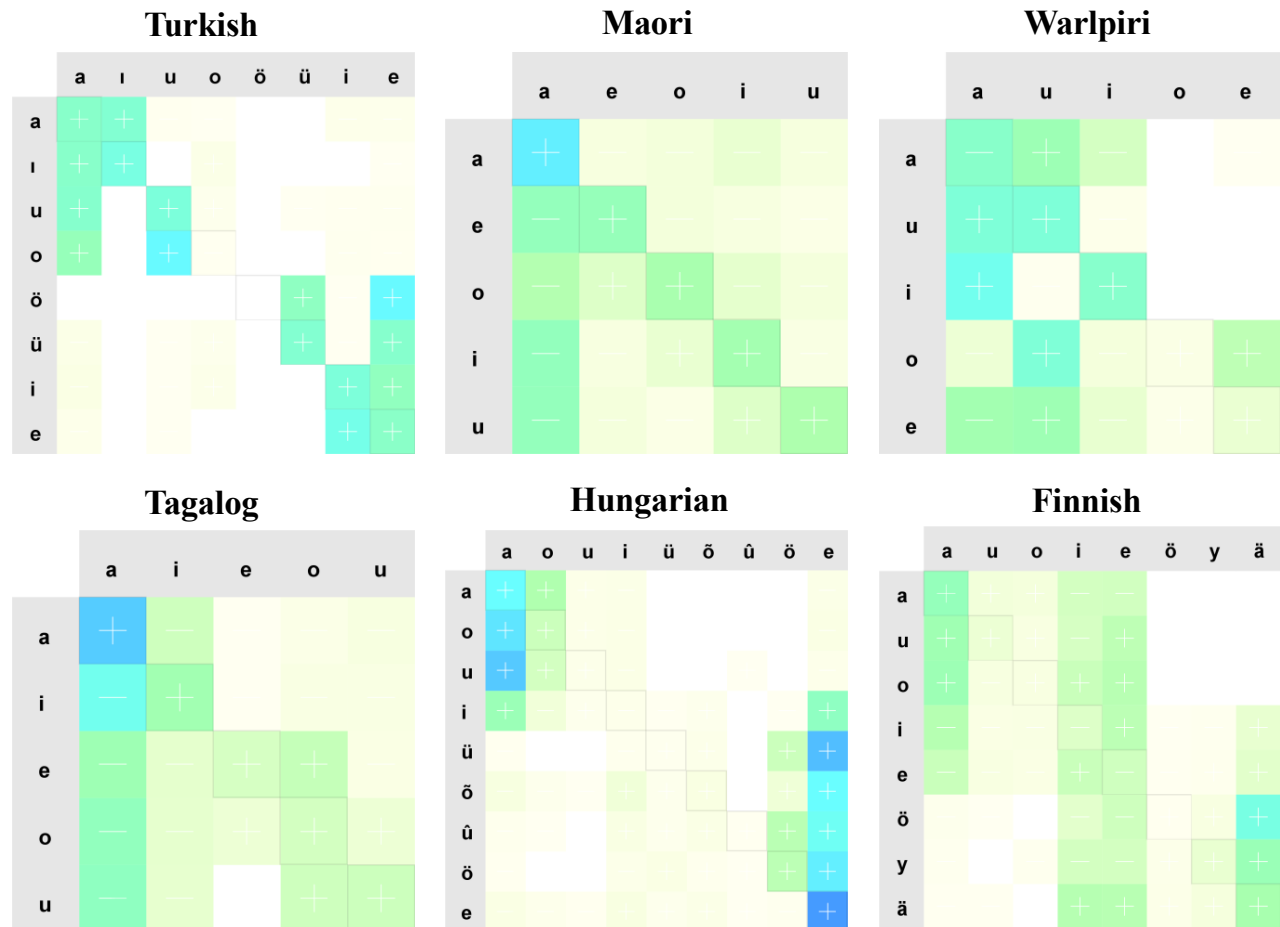


Powerwall, University of Konstanz



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.



Vowel Harmony via Droplets

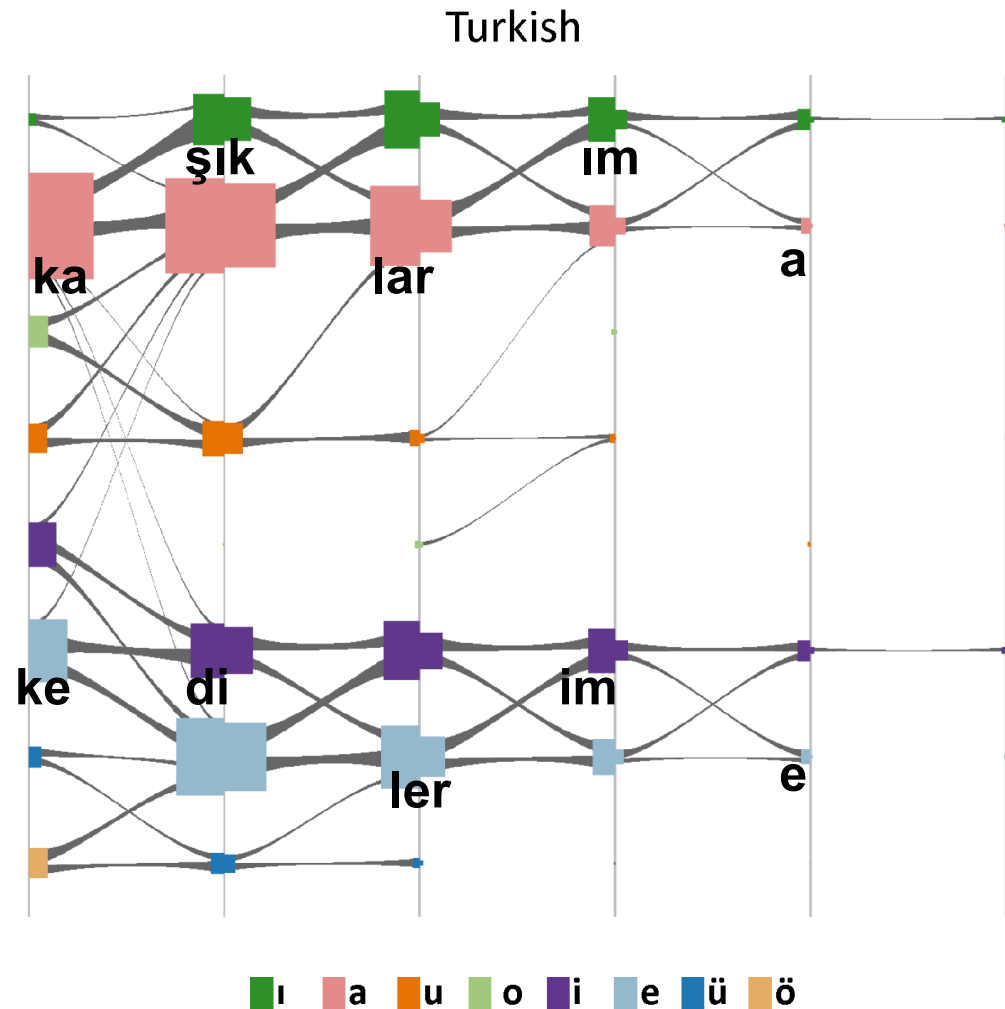
Application of geospatial rendering of places where a picture was taken to Vowel Harmony patterns

kaşık-lar-ım-a

spoon-Pl-1SgPoss-Dat
'my spoons'

kedi-ler-im-e

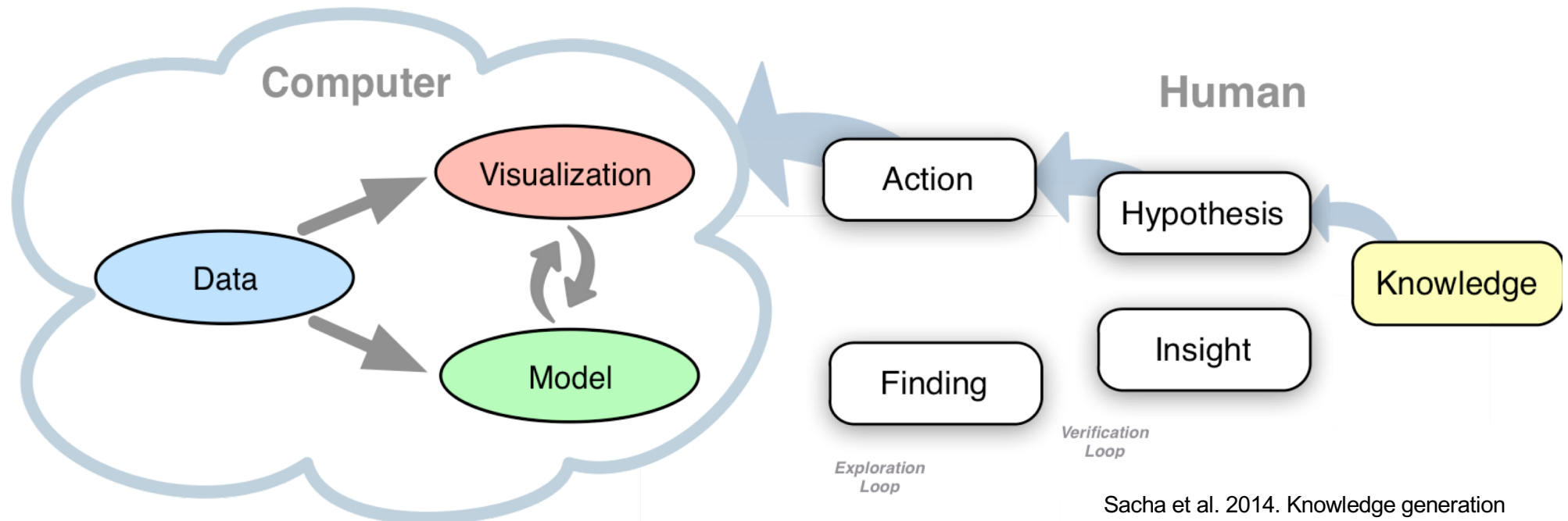
cat-Pl-1SgPoss-Dat
'my cat'



Visual Analytics

Overall Goal: Allow for efficient and insightful data interaction

- “Analyze first, show the important, zoom, filter and analyze further, details on demand” (Keim et al. 2008, based on Shneiderman 1996)
- Compact presentation of large amounts of data
- Different levels of detail on demand (interactivity)
- Exploratory and confirmatory data analysis, Iterative hypothesis testing and generation



Sacha et al. 2014. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20 (12), 1604-1613.

Case Study: Working with Annotated Corpora/Treebanks

- Ever increasing availability of digitized data and annotated corpora for historical linguistic research (e.g., Penn Treebanks, Dependency Treebanks, etc.)
- **For example: IcePaHC (Icelandic Annotated Treebank/Corpus)**
 - syntactically annotated (Penn Treebank style)
 - 60 texts, 12th century CE to 21st century CE

```
(IP-MAT-SPE (NP-SBJ (PRO-D Mér-mér))
  (VBPI finnst-finna)
  (CP-ADV-SPE (WADVP-1 0)
    (C sem-sem)
    (IP-SUB-SPE (ADVP *T*-1)
      (NP-SBJ (PRO-N ég-ég))
      (BEPS sé-vera) (VBN sloppinn-sleppa)
      (PP (P úr-úr) (NP (NP-POS (ONE+Q-G einhvers-einhver)
        (N-G konar-konar)) (N-D fangelsi-fangelsi))))))
  (. .-.))

(ID 1882.TORFHILDUR.NAR-FIC, .603))
```

Annotation sample from IcePaHC (Wallenberg et al. 2011)

Methodological Challenge for Historical Linguistics

- Standard procedure:
 - Calculation of co-occurrence frequencies and statistical significances for a multitude of different linguistic features across different time stages
 - Generation of a multitude of high-dimensional data tables of varying size containing numbers computed for different linguistic features

Texts	Indefinite NPs			Definite NPs			NPs as proper names		
	OV	VO	% OV	OV	VO	% OV	OV	VO	% OV
14th century	28	33	45.9%	11	57	16.2%	3	8	27.3%
15th century	23	30	43.4%	10	25	28.6%	1	3	25.0%
16th century	15	28	34.9%	17	26	39.5%	1	5	16.7%
17th century	28	59	32.2%	18	50	26.5%	0	20	0.0%
18th century	6	28	17.6%	7	31	18.4%	1	7	12.5%
19th century	34	425	7.4%	14	351	3.8%	4	68	5.6%
	134	603	18.2%	77	540	12.5%	10	111	8.3%

Definiteness distribution of NPs across different word orders in Icelandic (Hróarsdóttir 2000, 136)

- **Difficult to read and interpret!**

Case Study

Word Order Change in Icelandic

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

Walked a man into a pub...

Questions

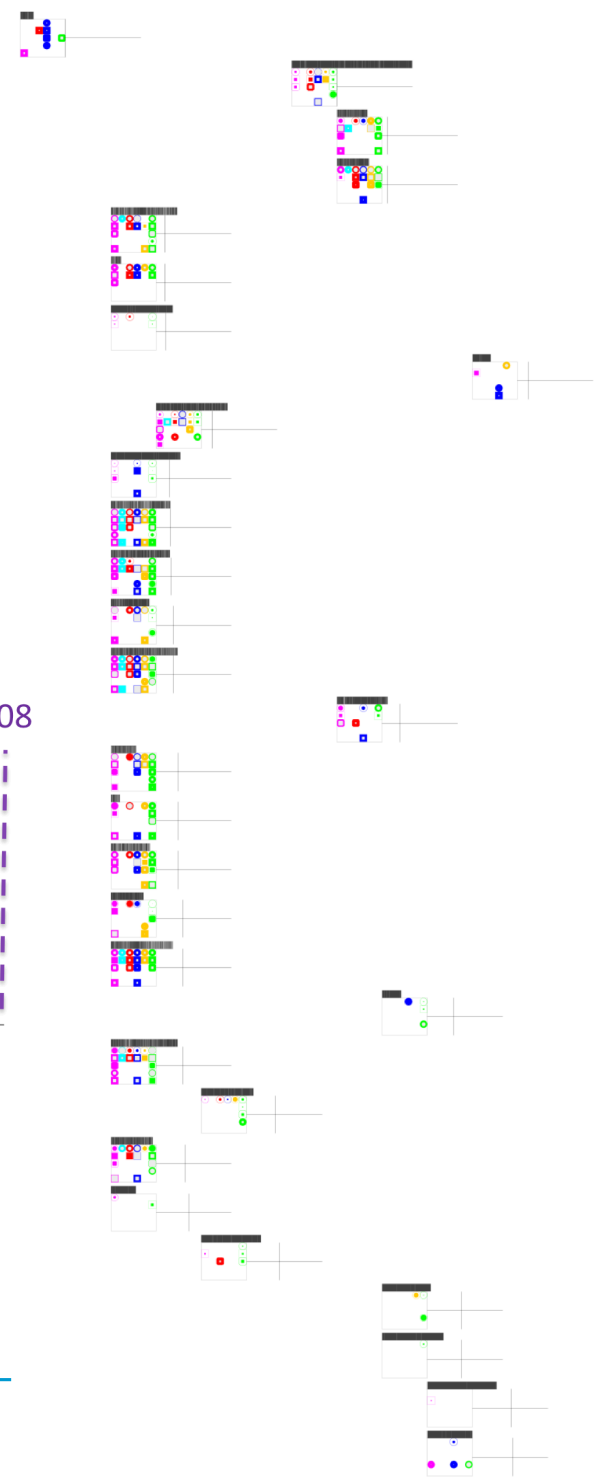
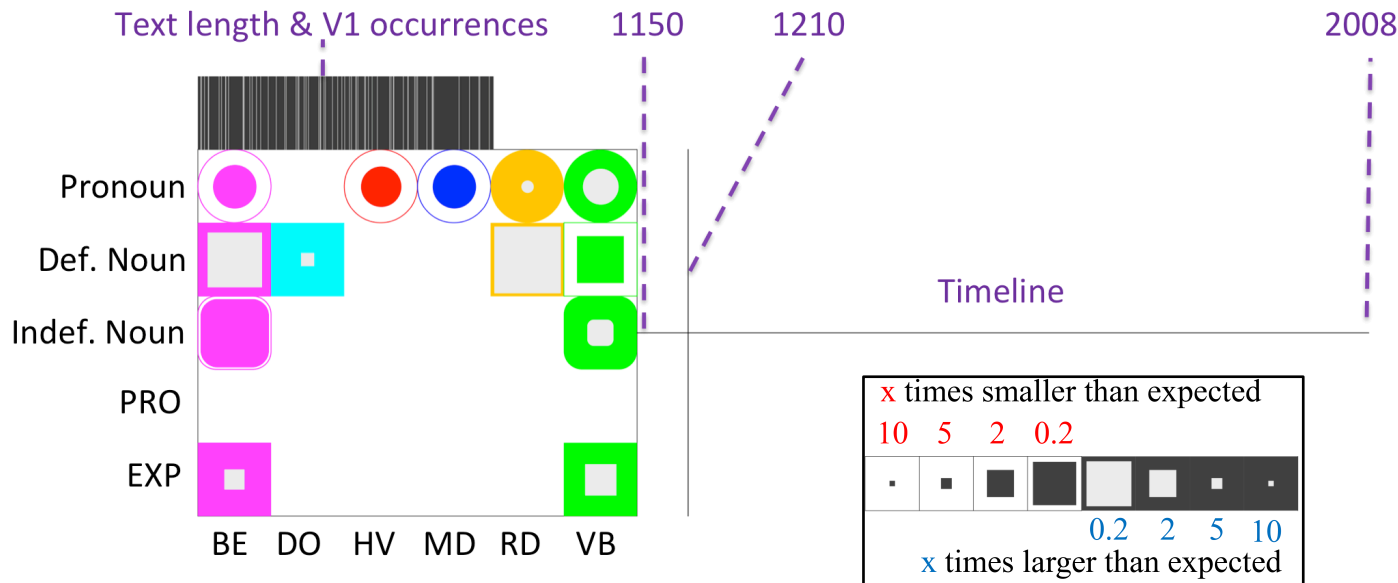
- What determines the (dis)appearance of V1?
- How did this change over the history of Germanic?
- **Case Study:** V1 in the history of Icelandic

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

Example: V1 in Icelandic

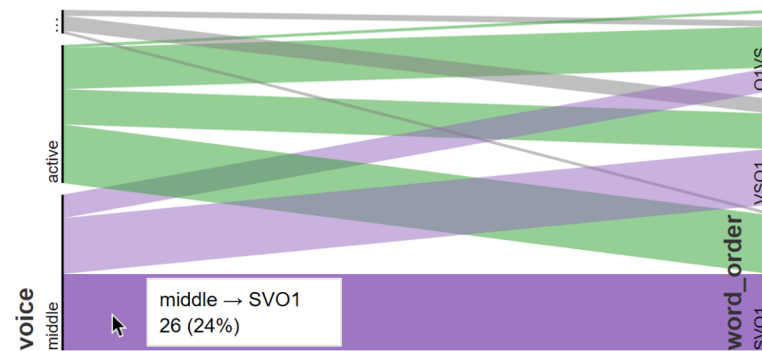
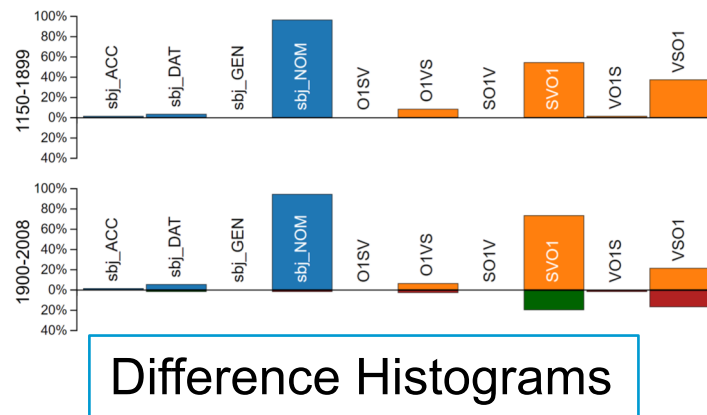
Visual Analytic Access to Data

- Glyph Visualization of likely factors
- Overview of all 60 texts at once
- Genre effect suddenly highly visible
- Can drill down to individual data points interactively
- **Mantra:** Overview First – Details on Demand



Example: V1 in Icelandic

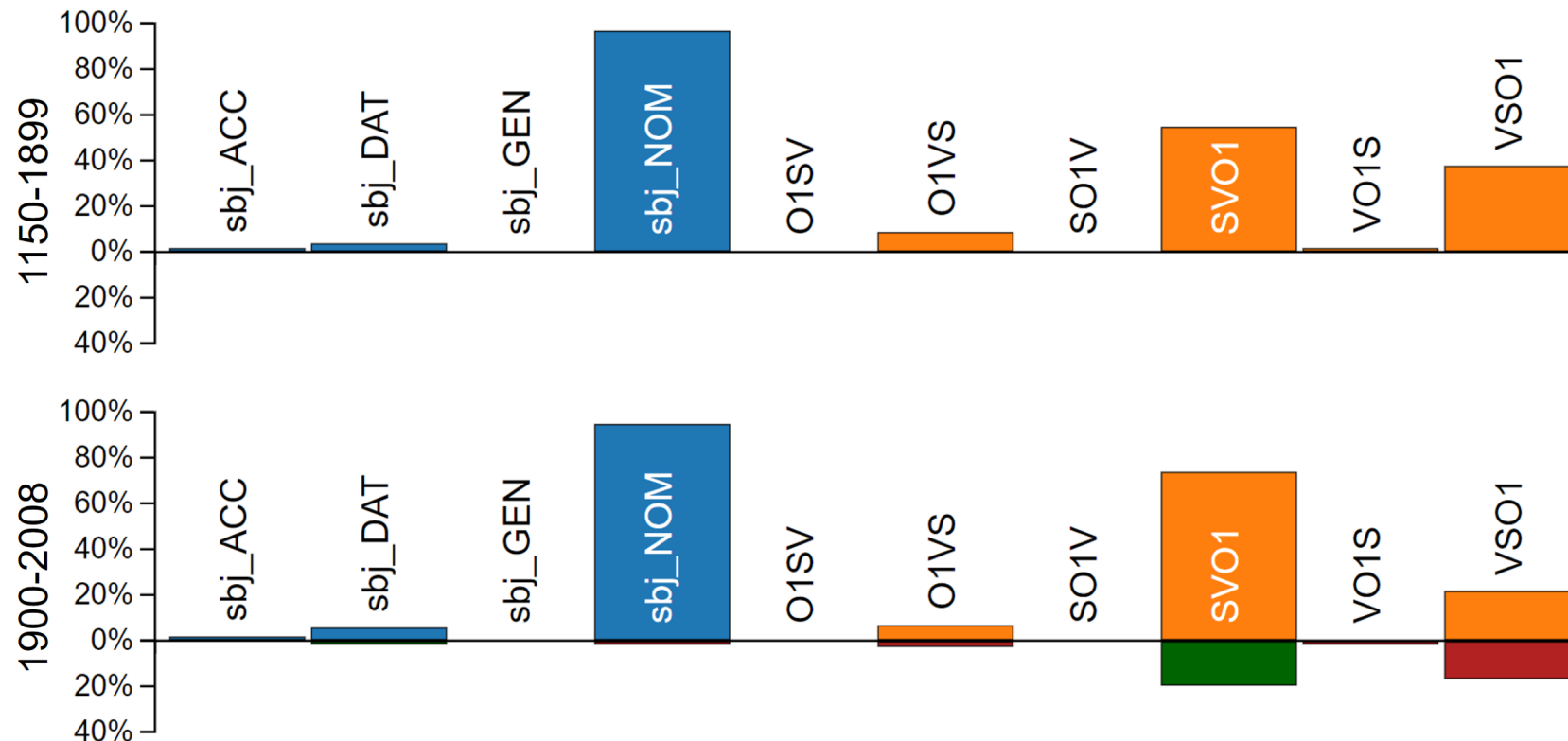
- The Glyph Visualization is complex.
- We have found that for many linguistic purposes, a simpler visualization is more effective.
- Now working with **HistoBankVis**: a multilayer visualization system



Schätzle, Christin, Frederik L. Dennig, Michael Blumenschein, Daniel A. Keim and Miriam Butt. 2019. Visualizing Linguistic Change as Dimension Interactions. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*. Association for Computational Linguistics (ACL), pp. 272--278.

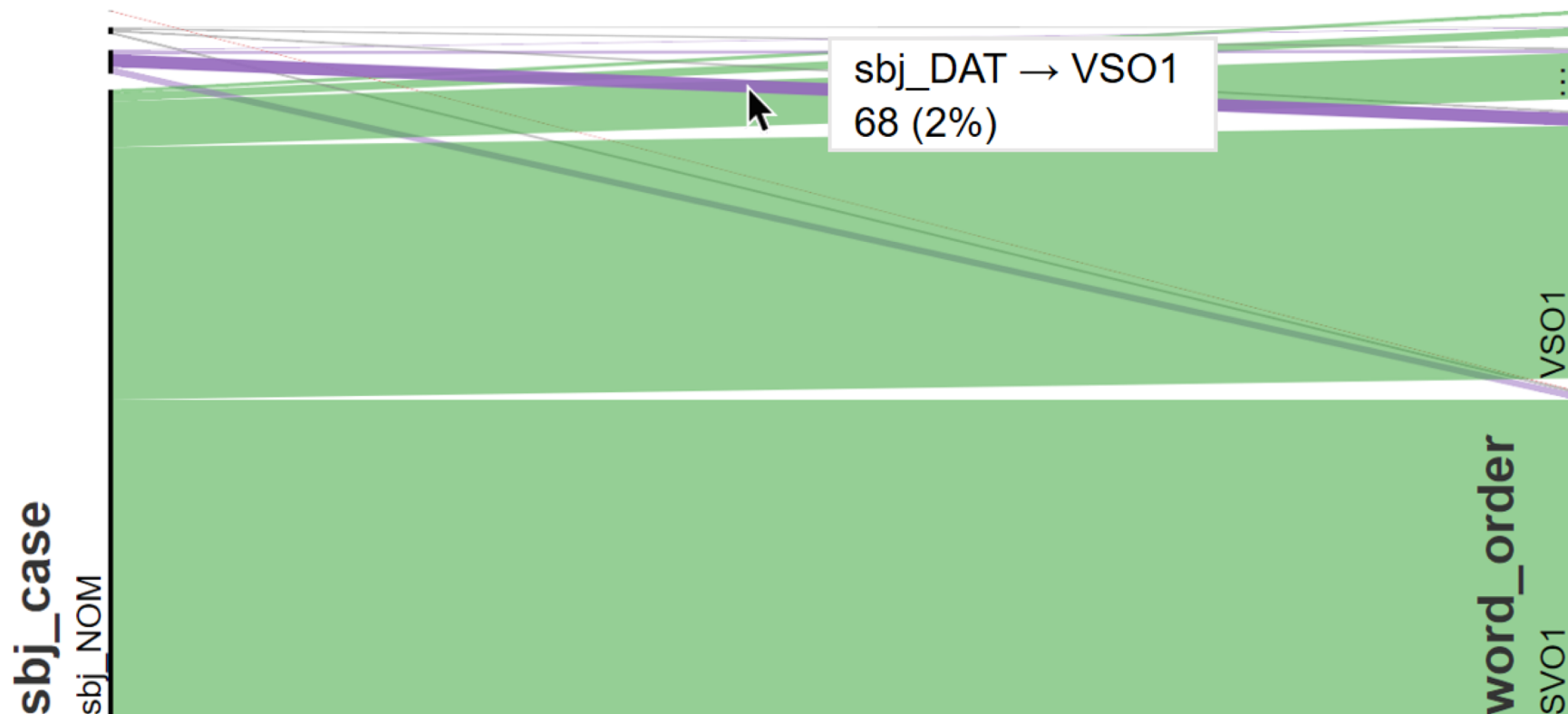
Difference Histograms

- Each time period visualized as one composed bar chart/histogram
 - Dimensions encoded via different colors for parallel inspection
 - Each bar represents an individual feature of a dimension
 - Bar height corresponds to the percentage of sentences containing a feature
- red = feature decrease
green = feature increase



Integration of Parallel Sets Technique

- Models Dimension Interaction
- Feature frequencies are visualized as proportions of equally spaced vertical lines (data dimensions)
- Dimensions are connected by colored ribbons
- Size of a ribbon = share which a feature holds of a feature from another dimension



HistoBankVis

- System available on-line
- Demo use cases
- Instructions on how to format and upload own data

<https://dennig.dbvis.de/histobankvis/#/>

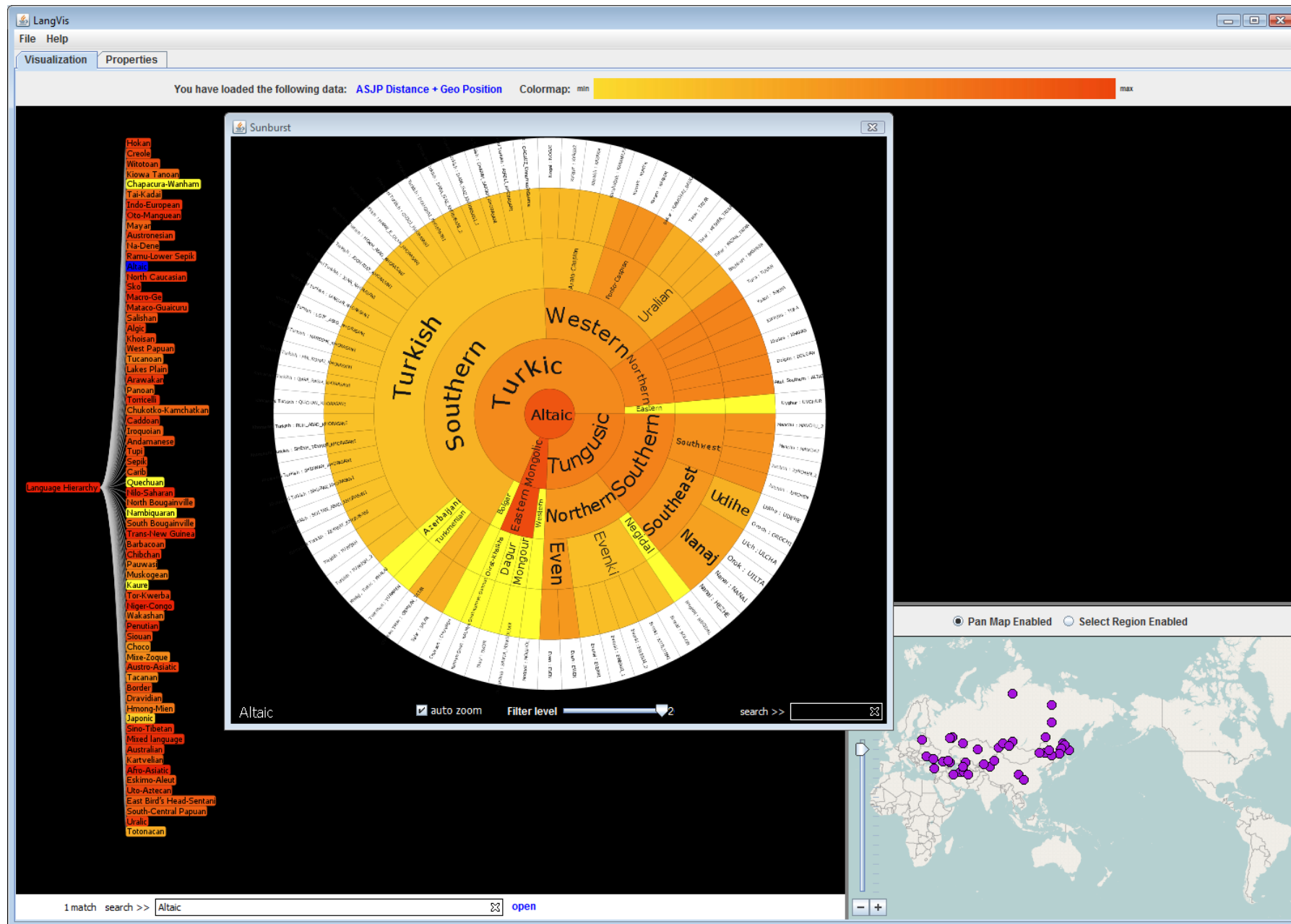
World Languages Explorer

- System that allows comparison of features across languages
- Sunburst Visualization
- Linked with a map
- Data from WALS (World Atlas of Language Structures; <http://wals.info>)
- But can also feed in own data.

<http://th-mayer.de/wals/#30A/>

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

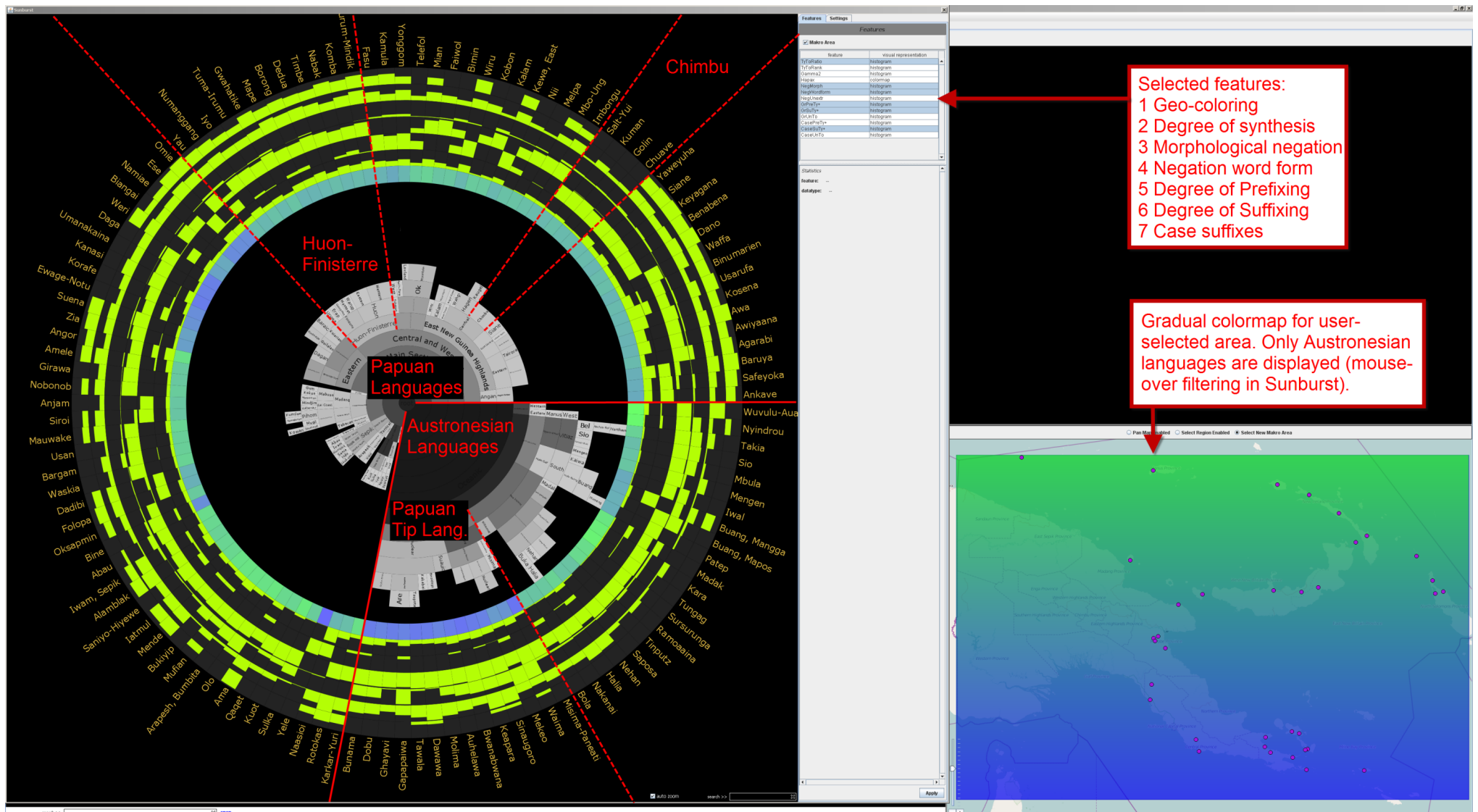
Sunburst and Maps for Language Families



World's Languages 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.



Sorting and Pattern Discovery

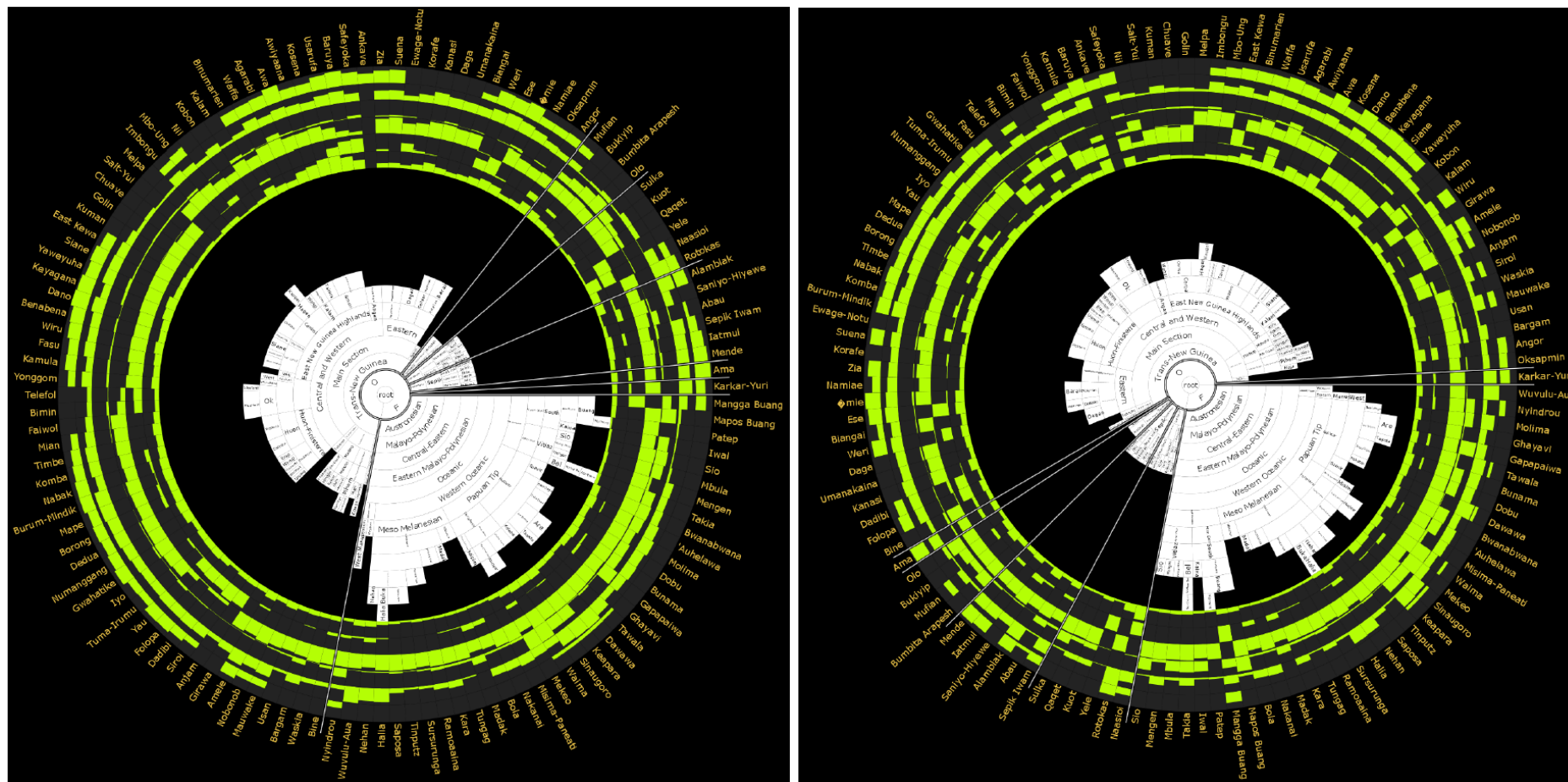
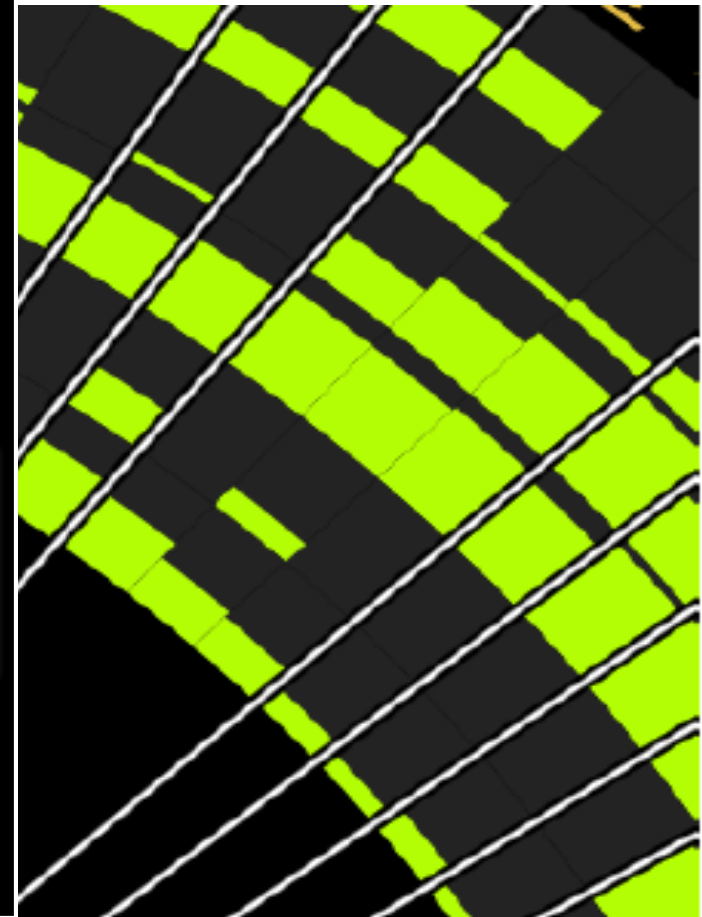
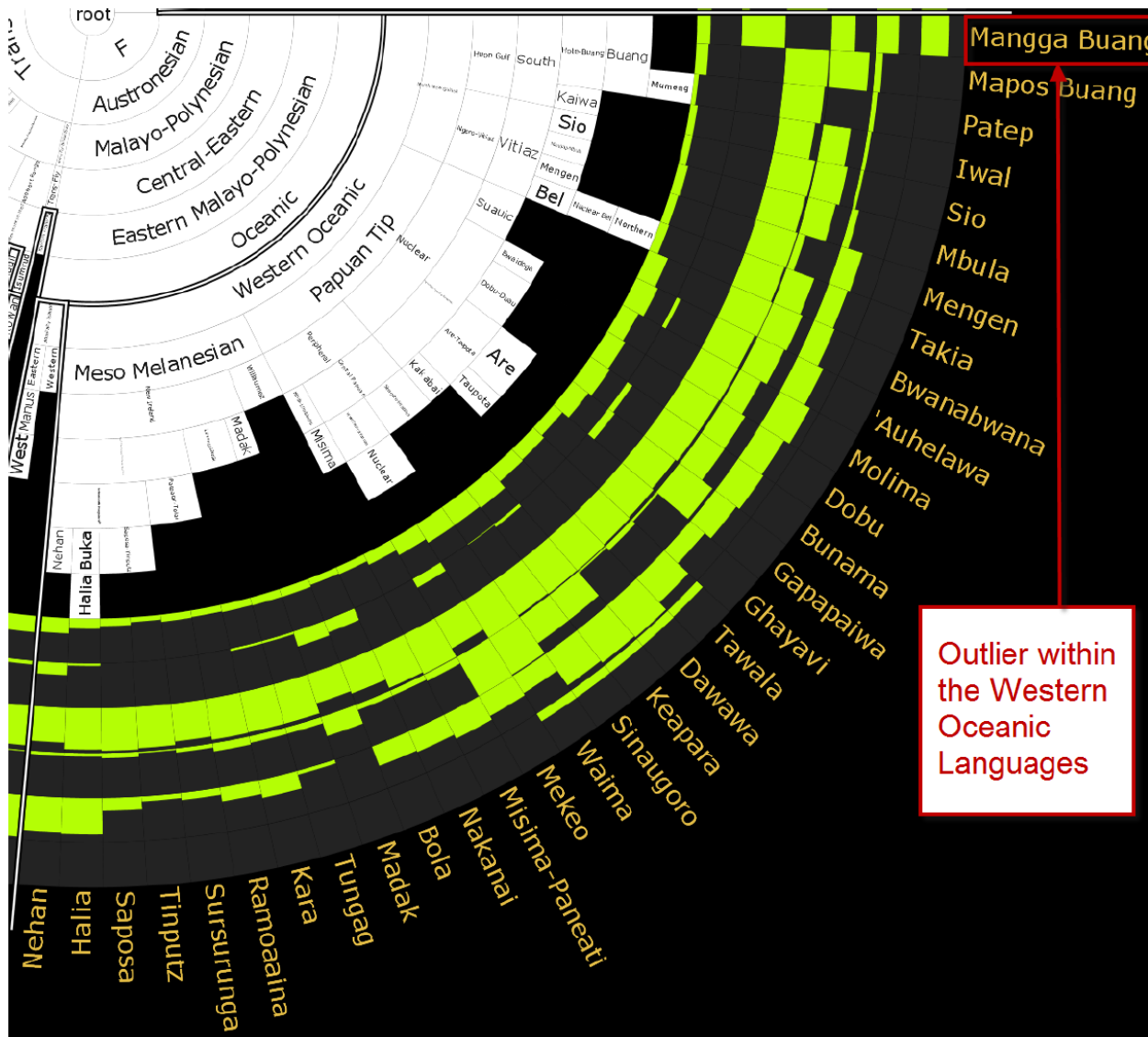


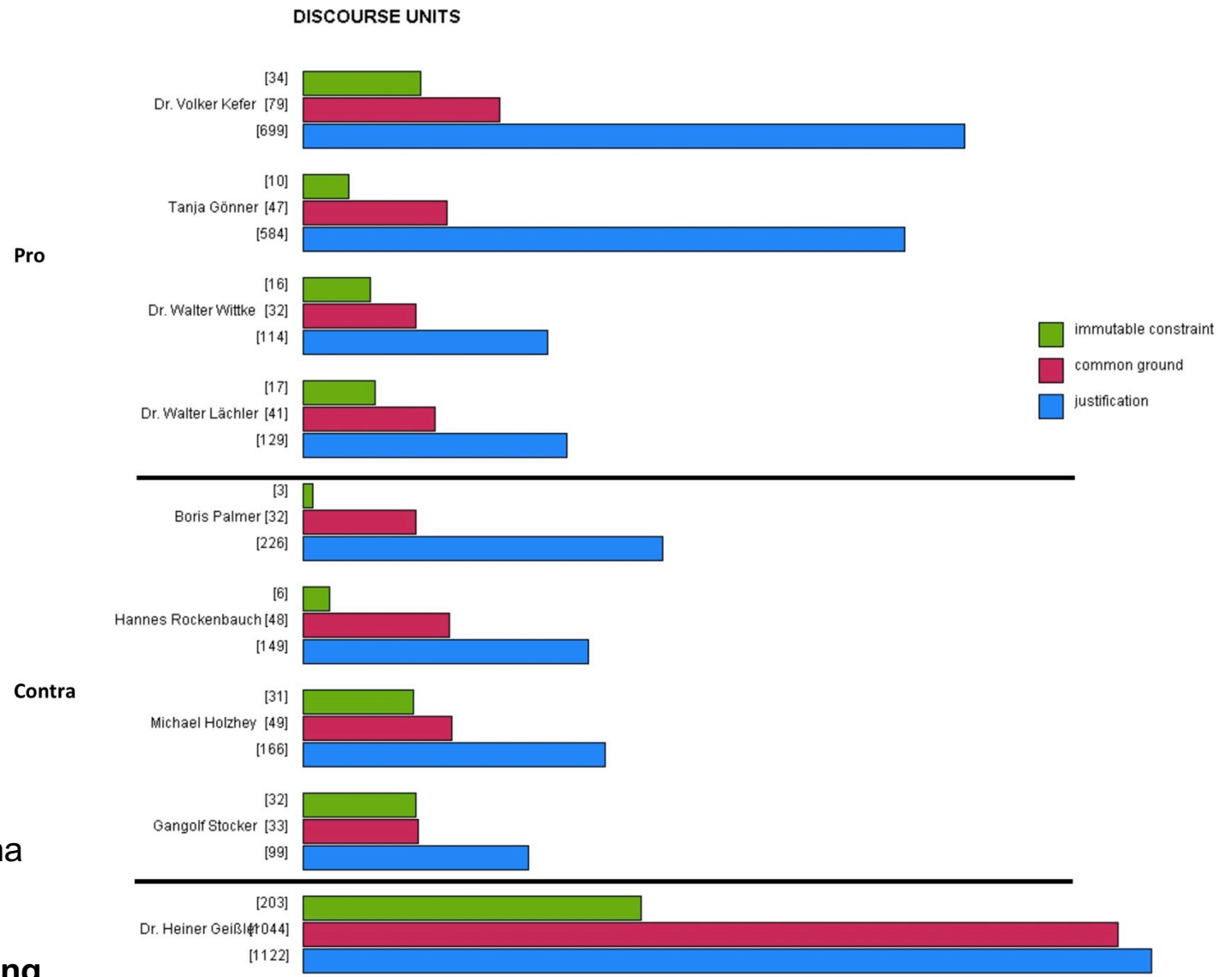
Figure 2.5: High-resolution screenshot showing automatically extracted features for languages from Papua New Guinea with leaves ordered to *maximize* (left) and *minimize* (right) the pairwise leaf similarity for neighbors.

Sorting and Pattern Discovery



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 Tools

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

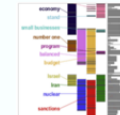
This visualization brings together all measures of deliberation. Using hierarchical glyphs, a fingerprint of the characteristics of utterances can be generated. Interesting sub-dimensions and measures can be interactively selected, depending on the analysis task. For a more holistic interpretation, the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



Lexical Episode Plots

Content Exploration Visualization

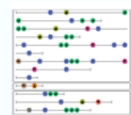
Aligned to the concept of lexical chaining, lexical episode plots allow the analysis of the content of text corpora based on the unexpectedness of the occurrence density of words and n-grams. This view generated an overview of the complete text using the zoomed-out distant-reading view. Using the zooming 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

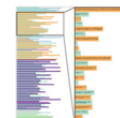
This visualization sets the focus on the exploration of speaker participation in multi-party conversations. By spanning a conversation space using all mentioned topics in the discussion, the interaction of the speakers over time can be traced. This animated visualization enables the analysts to detect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



Named-Entity Relationships

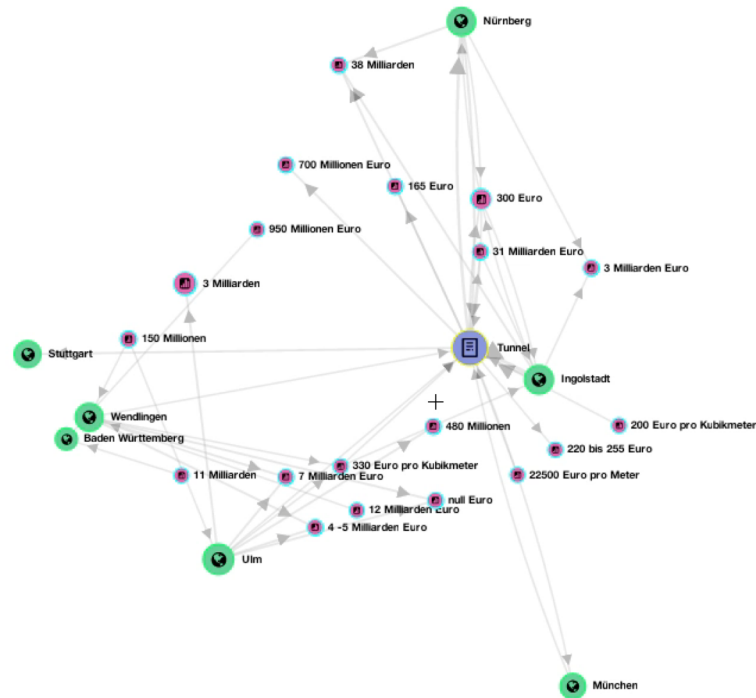
Show positions or speakers

Positions

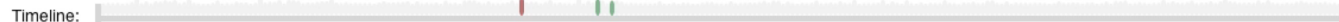
- CONTRA
- NEUTRAL
- PRO

Speakers

- Dr. Martin Viereg
- Prof. Karl-Dieter Bodack
- Werner Wölfle
- Peter Conradi
- Hannes Rockenbauch



Graph over time



Named entity categories

- Person
- Geo-Location
- Date-Time
- Measure
- Measuring-Unit
- Organization

Additional categories

- Context-Keyword
- Politeness-Indicator
- Positive-Emotion-Indicator
- Negative-Emotion-Indicator

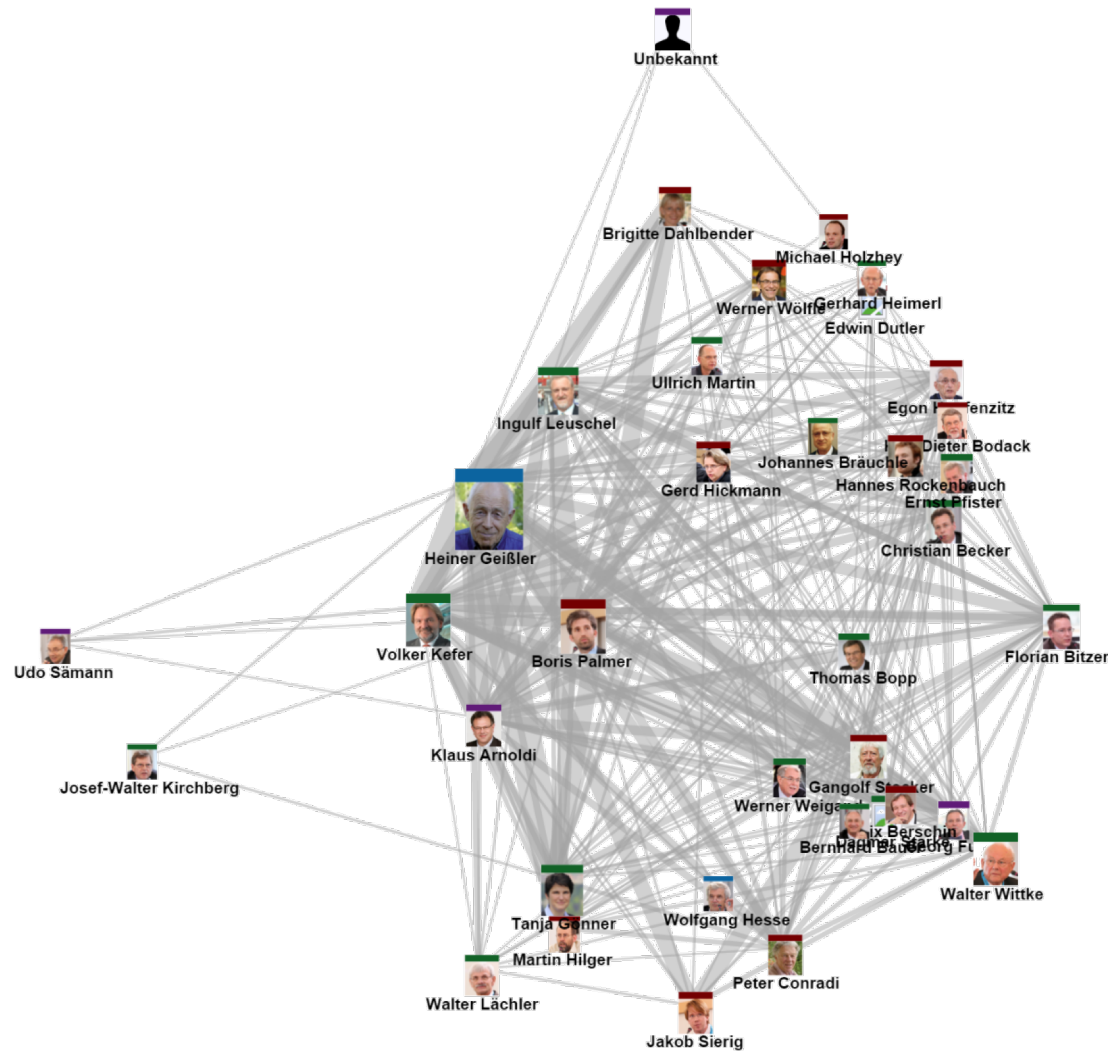
Speaker positions

- NEUTRAL
- UNKNOWN
- PRO
- CONTRA







Used containers

- Tunnel
- Kosten
- Deutschland





Named-Entity Relationships



Named-Entity-Categories

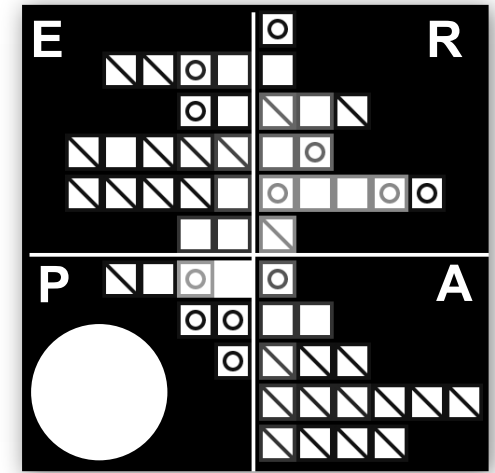
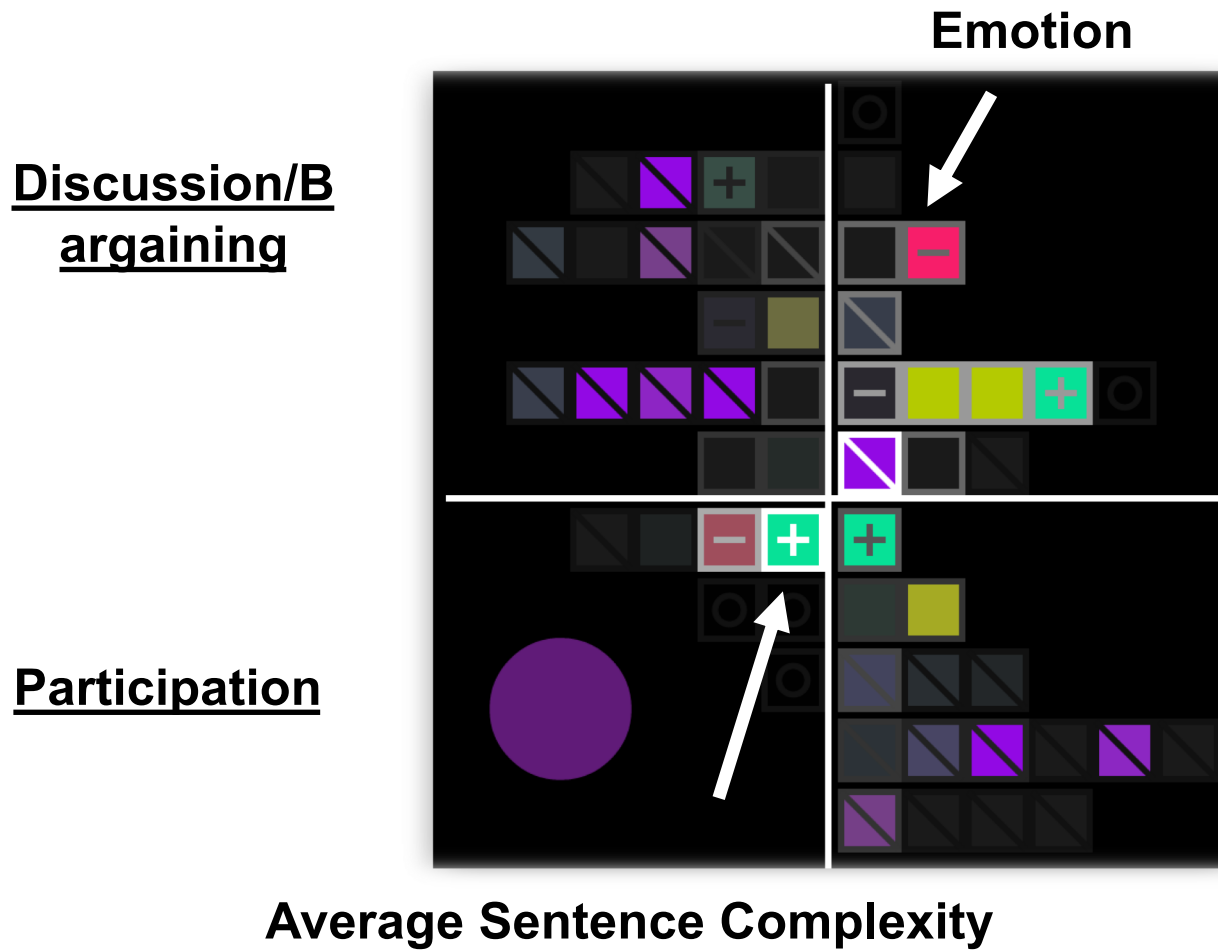
-  Person
-  Geo-Location
-  Date/Time
-  Measure
-  Measuring-Unit
-  Organization

Additional Categories

-  Context-Keyword
-  Politeness-Indicator
-  Positive-Emotion-Indicator
-  Negative-Emotion-Indicator

Glyphs – Deliberative Argumentation

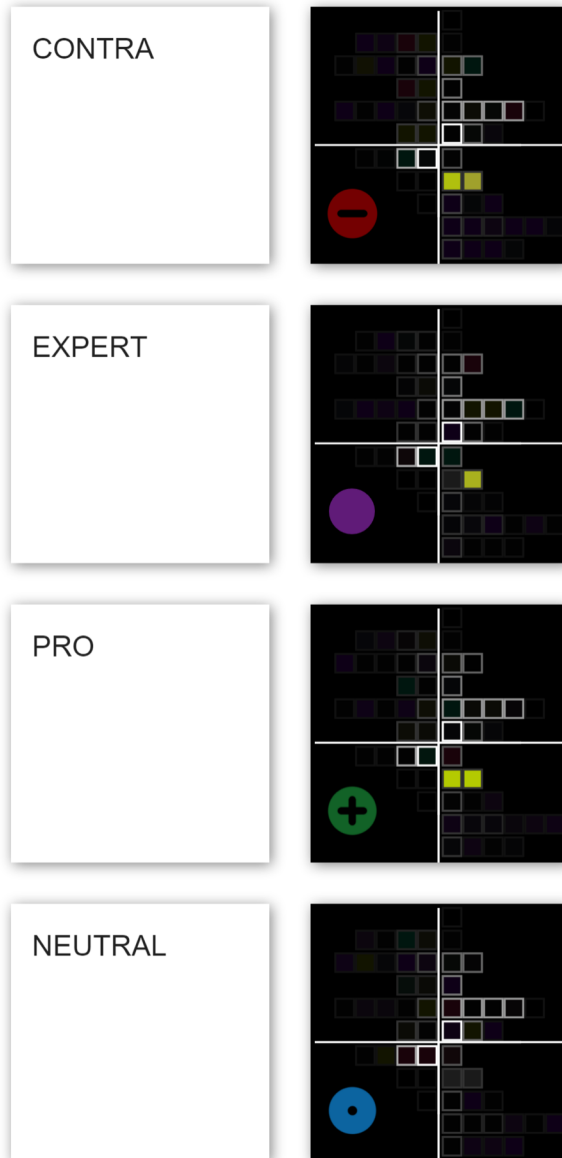
Example



Conversational Atmosphere and Respect

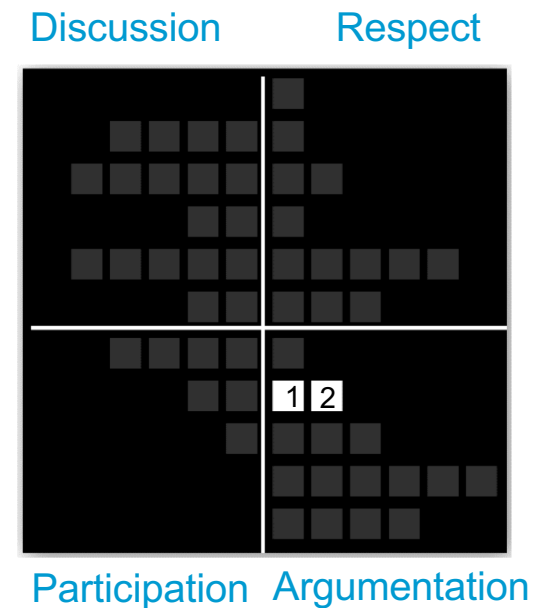
Argumentation and Justification

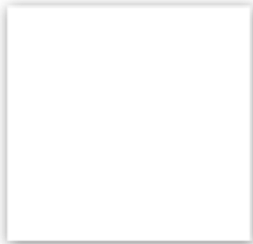
Argumentation



Neutral: Comparatively little Argumentation
Experts: Justifications/Reasons
Pro & Contra: Conclusions and Justifications
→ **Deliberative Argumentation**

- Measures shown:**
1. Conclusion
 2. Justification

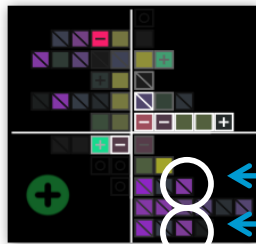




Hauptbahnhof
Gleis Tunnel
Brücke

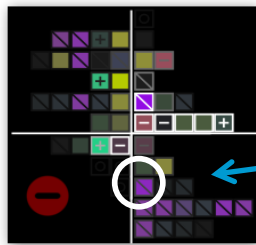
Information mainly from discourse particles and speech act verbs

PRO



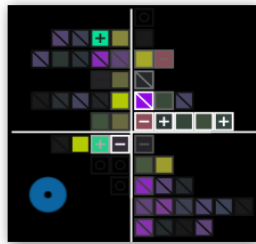
Rejection of "Common" Ground
Refusal to provide information

CONTRA

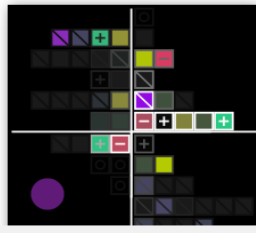


Establishment of Common Ground

NEUTRAL



EXPERT



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

Aussagen Frage Antwort Folio Mittagspause Thema

PRO

CONTRA

NEUTRAL

EXPERT

	Hauptbahnhof Weiche Stellwerk	Hauptbahnhof Weiche Stellwerk	
PRO			Volker Kefer
CONTRA			Boris Palmer
NEUTRAL			Heiner Geißler

The central part of the slide features a 3x2 grid of data visualizations. The columns are labeled 'Hauptbahnhof Weiche Stellwerk' and the rows are labeled 'PRO', 'CONTRA', and 'NEUTRAL'. Each cell contains a grid of small, colorful icons (purple, green, red, blue) with a larger colored circle (+ or -) in the bottom-left corner. To the right of this grid are three white boxes containing the names 'Volker Kefer', 'Boris Palmer', and 'Heiner Geißler'.

Ubertürkeim
Ermschutzwände
Eisensteintunnel

Schichtung
Teilnehmer

Betonwinkel
Unterboden
Verdichten

Visualization of Pitch Contours

- 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.

Data

- Japanese vs. German 'sorry'
- Japanese pitch contour always has a fall
- Germans can vary according to pragmatic intent
- Recorded German and Japanese natives vs. learners of German and Japanese (beginners/advanced)

Dominik Sacha, Yuki Asano, Christian Rohrdantz, Felix Hamborg, Daniel A. Keim, Bettina Braun & Miriam Butt. 2015. Self Organizing Maps for the Visual Analysis of Pitch Contours. *Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa-2015)*, Vilnius, Lithuania, 2015.

Example: Speech Data

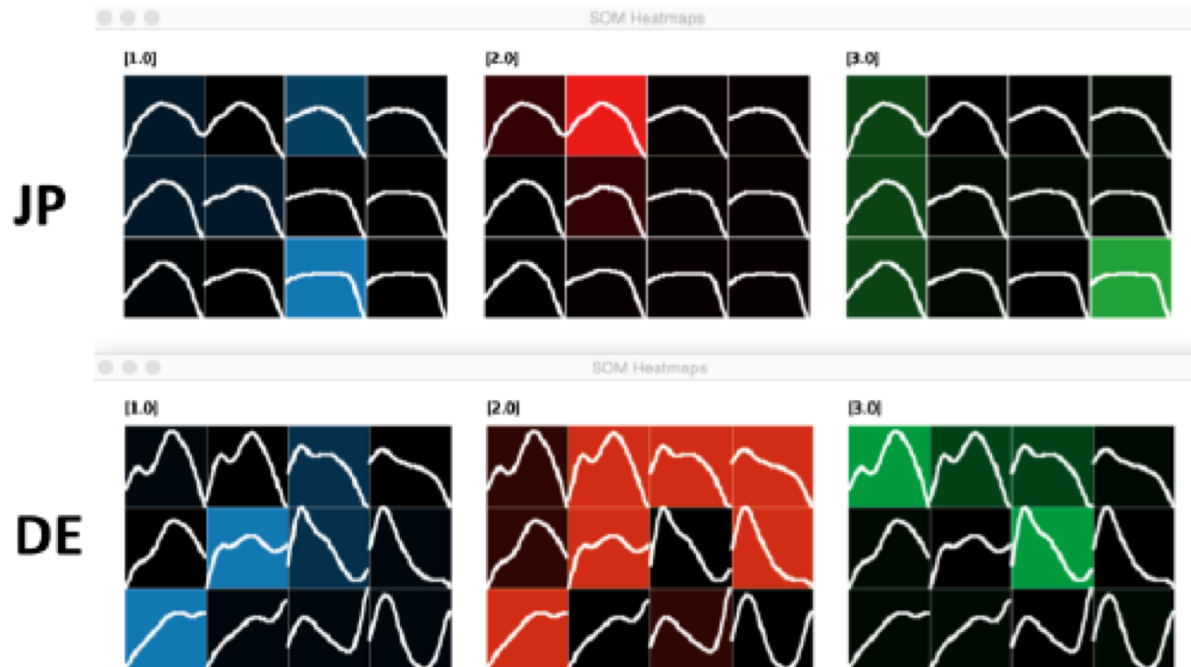
- Japanese Native and German L2 Learner data (pitch contours and meta data)
- F0 contours are smoothed and normalized into pitch vectors
- The pitch vectors are visualized via self-organizing maps (SOM)

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

- Japanese natives do not vary the pitch contour
- German learners do vary the pitch
- German beginner learners do so more

Interactive Exploration:

- individual cells can be merged
- meta data can be inspected
- individual pitch tracks can be examined in context



Example: Speech Data

- Different views on the data and meta data can be explored interactively

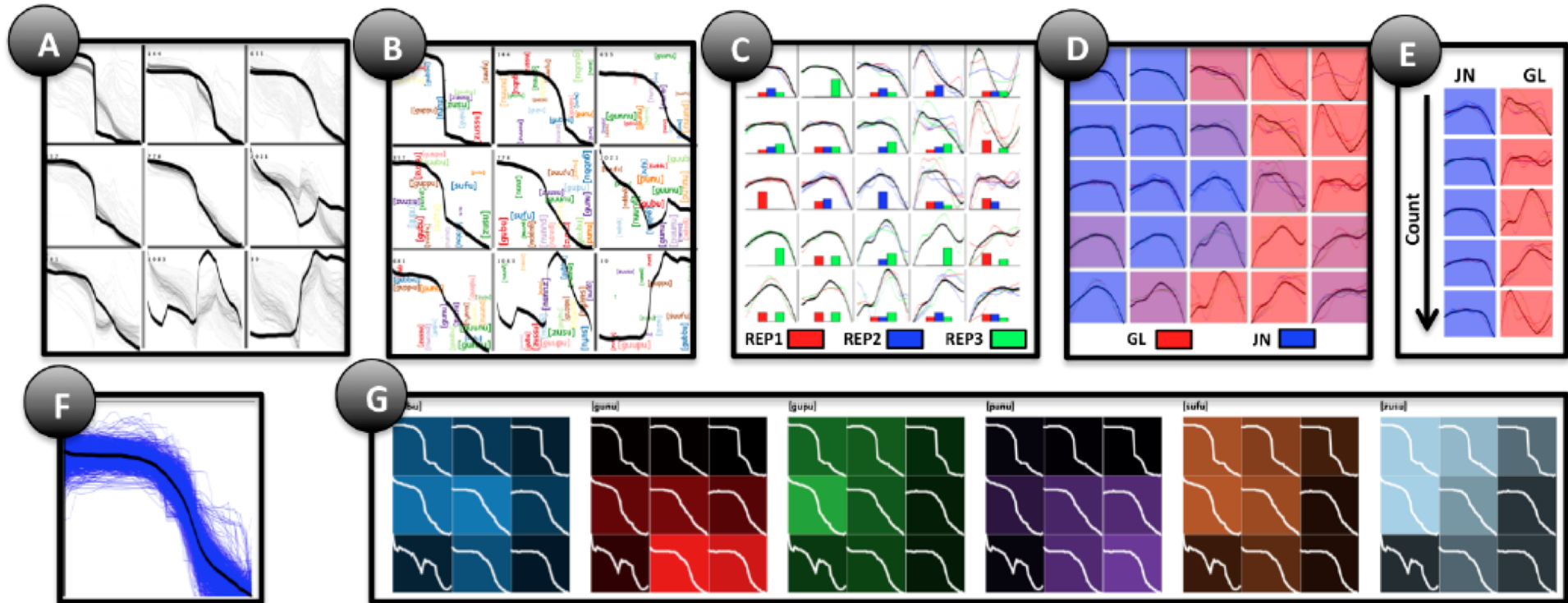
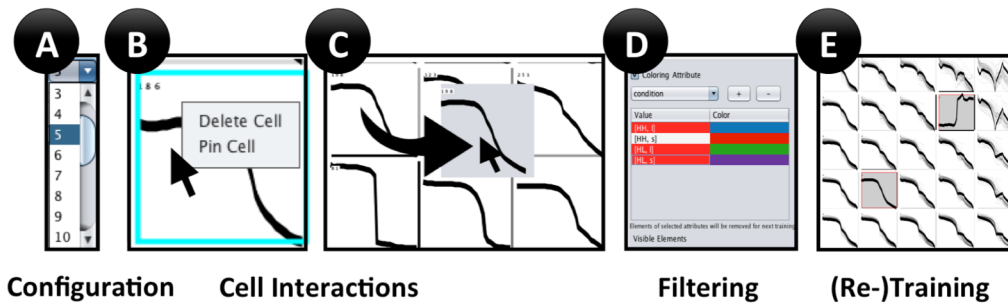


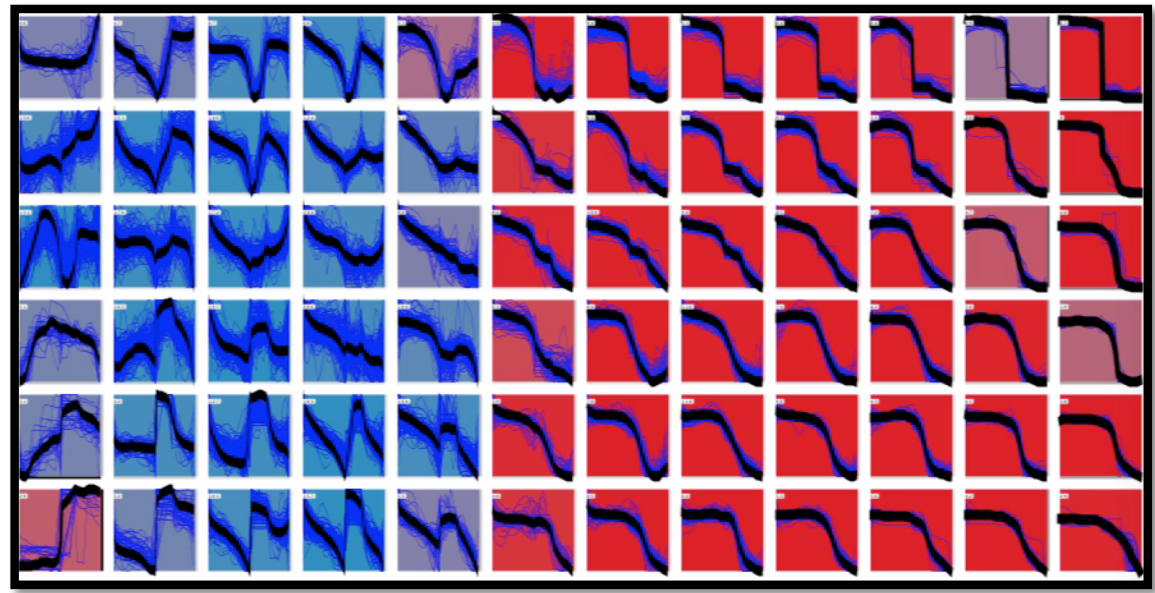
Figure 3: Different approaches to visualize SOM-results according to available meta-data. (A) Grid visualization, (B) word cloud, (C) bar charts, (D) mixed color cells, (E) ranked group clusters, (F) one single cell that visualizes contained vectors and the cluster prototype, (G) separated heatmaps for all values of a categorical attribute.

Interactive Data Exploration: SOMs in Action



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



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.

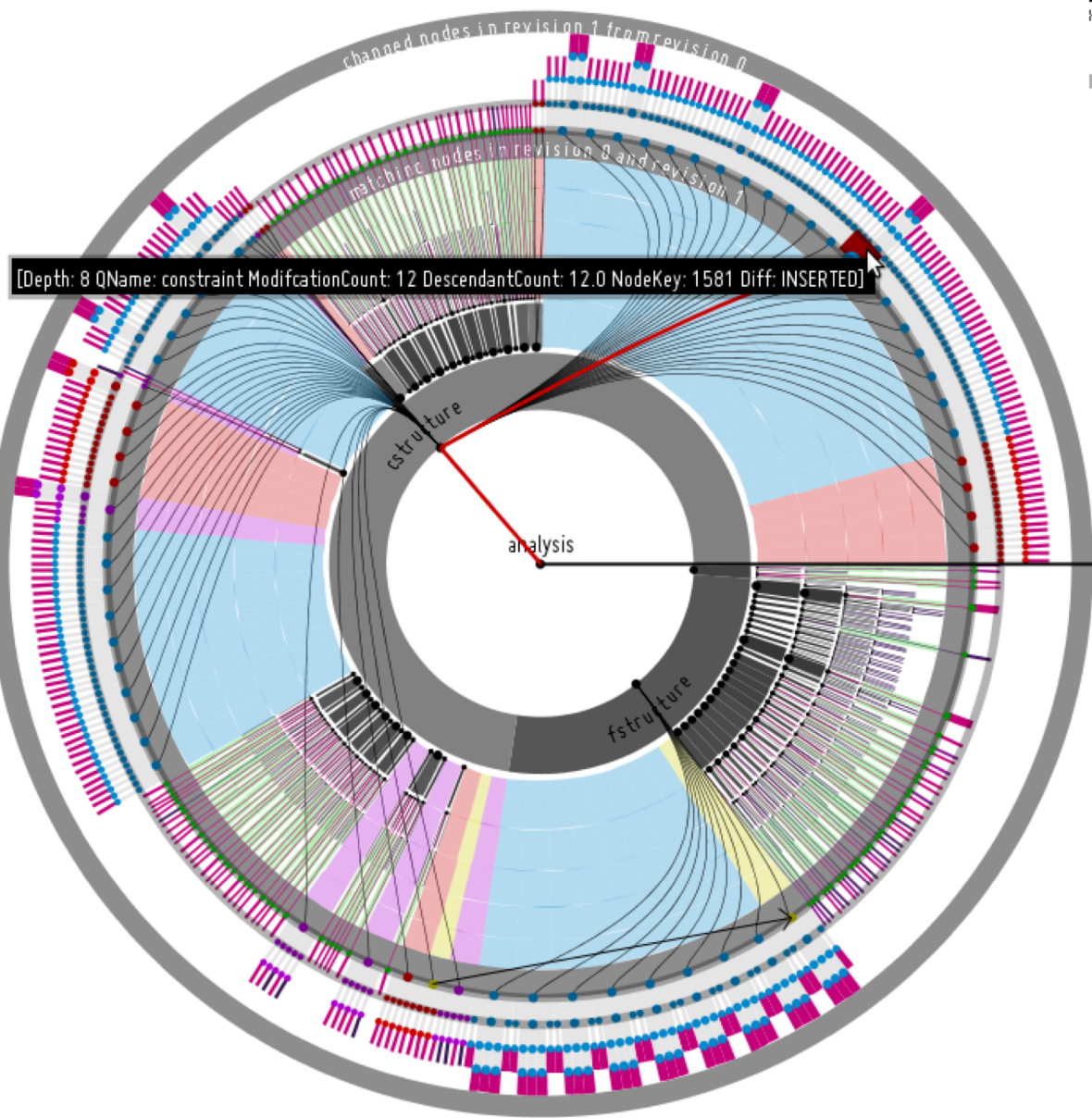
Tree Comparison for Grammar Development

- In grammar development the grammar is routinely updated/changed.
- This necessarily means that the output will differ (relevant for tree banking).
- Would be good to have an automatic visual tree comparison method.
- Proposal by Lichtenberger (2012) – visualization of changes in c-structure/f-structure pairs for LFG grammars.
- The visualizations are not successful.
- Part of the reason is that we are not used to looking at trees this way.

Johannes Lichtenberger. 2012. A Visual Analytics Approach for Comparing Tree Structures. MSc., University of Konstanz. <http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-210419>

HPATH EXPRESSION

REVISION 1



- text node similarity
 - element node similarity

- node moved
- node inserted
- node deleted
- node updated
- node replaced

Outlook

Further Exploration of Possibilities offered by Visual Analytics

- The systems illustrated here are still being experimented with.
- Interactive exploratory linguistic analysis is on-going.

Workflow

- Use cases for Digital Humanities /eHumanities are being developed.
- **Infrastructure** Platforms (mix and match the available tools)

<https://lingvis.io>

Measuring Success

- Development of **Evaluation Metrics** for LingVis.
- Use cases, work flow and result comparison.

Thank You!

Interdisciplinary Cooperation (University of Konstanz)

Linguistics

Tina Bögel, Annette Hautli-Janisz, Thomas Mayer, Frans Plank, Christin Schätzle

Computer and Information Sciences

Daniel Keim, Mennatallah El-Assady, Michael Blumenschein né Hund, Frederick Dennig, Andreas Lamprecht, Christian Rohrdantz, Dominik Sacha, Fabian Sperrle

Political Science

Katharina Holzinger, Valentin Gold