



Demo Session

Miriam Butt & Mennatallah El-Assady

(based on material from Annette Hautli-Janisz & Dominik Sacha)

LingVis: Visual Analytics for Linguistics

DGfS 2018 | 6.3.2018



Contents

2

- **WALS Explorer**
- **Self Organizing Maps for Pitch Contour Analysis**
- **Diachronic Visualization of Language Properties**
- **VisArgue**
- **DoubleTree (Chris Culy)**

The WALS Sunburst Explorer

Visualization of Typological Patterns

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.



WALS Explorer

- The WALS Explorer can be accessed on-line:
 - <http://th-mayer.de/wals/>
- You cannot use your own data here.
- It is meant for an exploration of the World Atlas of Language Structure (<http://wals.info>).
- Task/Interaction Suggestion:
 - pick a phenomenon you are interested in
 - see what you can find out about it
 - think critically about the visualization and the interactive possibilities

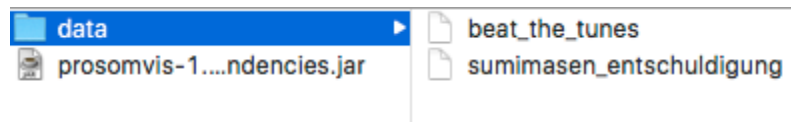
Self Organizing Maps for Pitch Contour Analysis

Analyzing Spoken Data

D. Sacha, Y. Asano, C. Rohrdantz, F. Hamborg, D. A. Keim, B. Braun and M. Butt. Self Organizing Maps for the Visual Analysis of Pitch Contours. Proceedings of the 20th Nordic Conference of Computational Linguistics. Vilnius, Lithuania, ACL Anthology, 23():181-190, 2015.

SOM – Starting the Tool

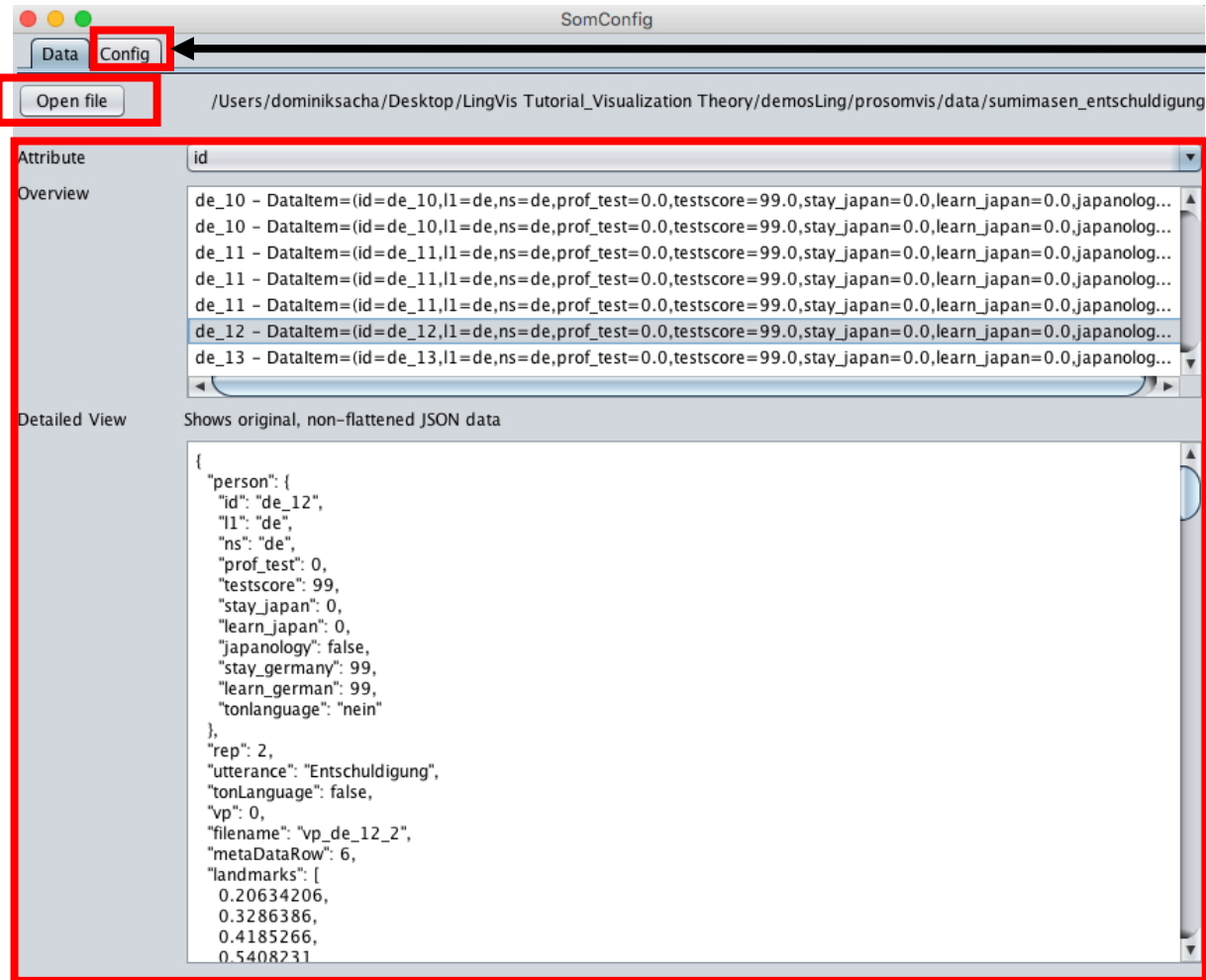
- All material (software and data) is contained in the “prosomvis” folder



- Double-Click the prosomvis-Jar file
 - prosomvis-1.0.0-SNAPSHOT-jar-with-dependencies.jar
 - The App should start

SOM– Loading Data

1. Click on the Button and load a data file



2. You may inspect the loaded data

3. Finally switch to the config tab

SOM– Configuration (Simple)

1. Select the Vector that should be analyzed
(For the sumimasen dataset a preprocessed vector is included)

2. If no preprocessed vector is detected, further preprocessing has to be applied

3. A data filter may be applied
e.g., utterance="sumimasen"
Note, data may be filtered during the analysis process as well

4. Set the SOM dimensions. Feel free to try different ones

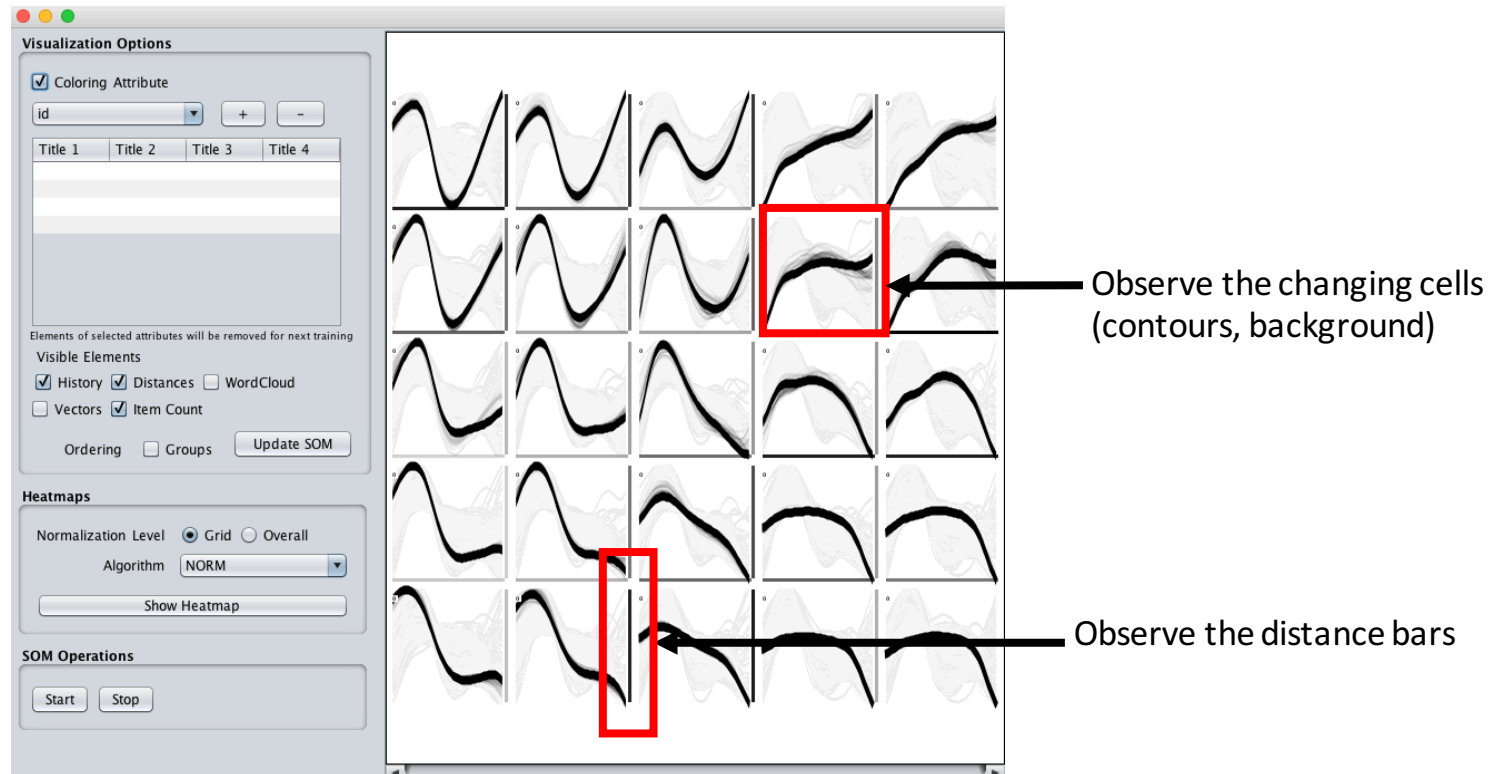
The screenshot shows the 'SomConfig' application window with the 'Config' tab selected. The interface includes several sections: 'Data Summary' (Status = OK, All data items contain a SOM-Vector of equal length.), 'Adjustment' (Average vector length dropdown), 'Landmarks' (Landmark-Vector dropdown set to 'landmarks', Time-Vector dropdown set to 'fdaPitches_time', and a checkbox for 'Semitone-Normalization'), 'Filters' (id dropdown, Add button, Delete Selected button), 'Dimensions' (Rows dropdown set to 3, Cols dropdown set to 3), and 'Training Duration' (dropdown set to 'Short'). A 'Launch SOM' button is at the bottom. Red boxes highlight the 'OM-Vector' dropdown (set to 'fdaPitches_pitch'), the 'Adjustment' section, the 'Filters' section, the 'Dimensions' section, and the 'Training Duration' dropdown. Arrows point from the numbered instructions to these specific elements.

5. The SOM training may be “long” or “short”
Suggestion: Try short first and switch to a longer training to confirm a concrete hypothesis

6. Finally launch the SOM

SOM Training – Just Watch

9

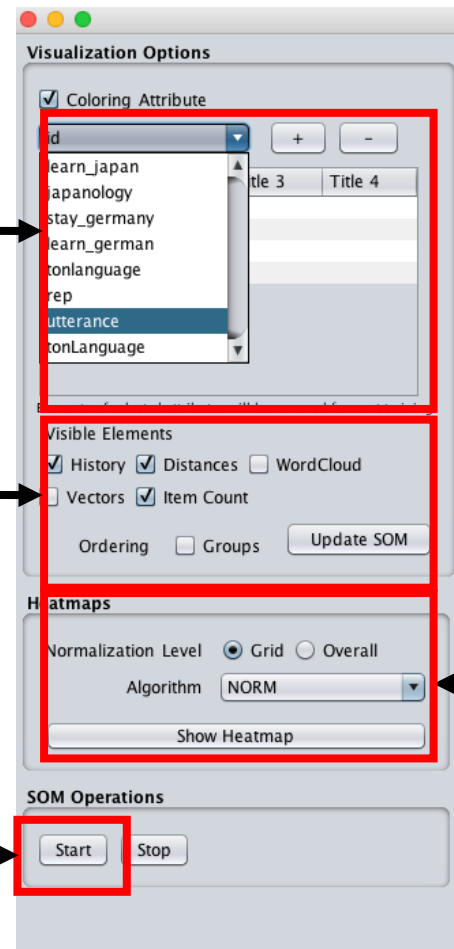


SOM Visualization - Configuration

1.) Add/remove attributes to be visualized. E.g., “utterance” or “ns” (Hint: Start with only one attribute)

2.) Try out different grid and cell visualizations (if the checkboxes do not work use the update button)

4.) A subsequent SOM may be trained with the remaining data items

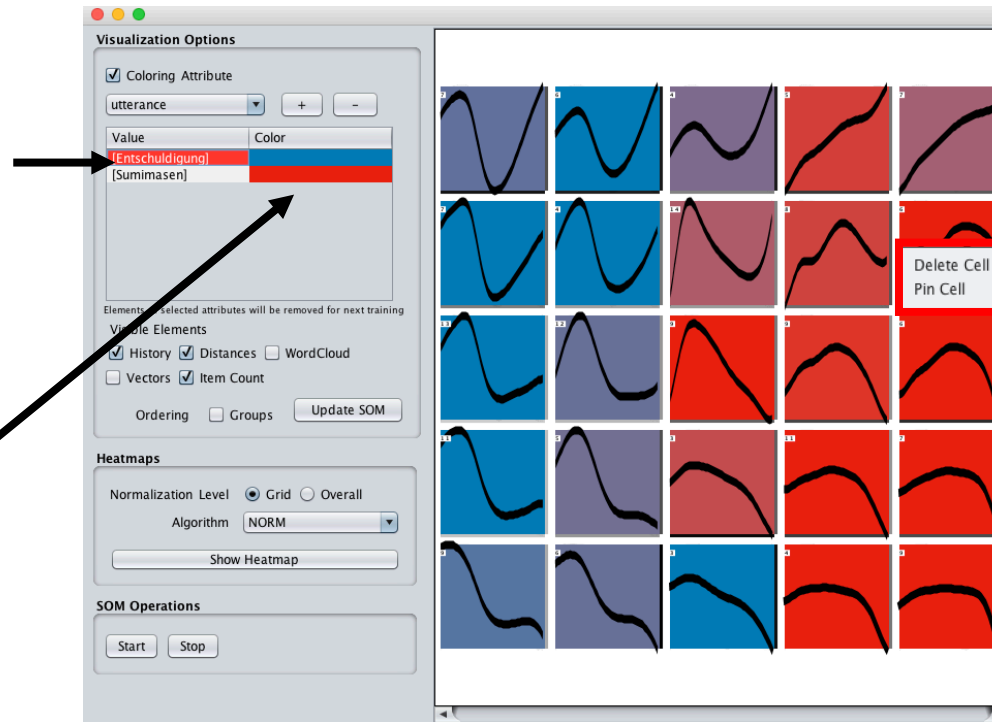
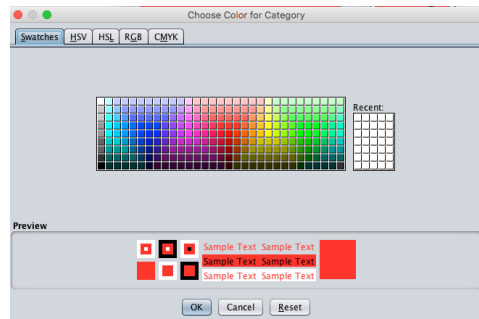


3.) Once an attribute has been selected (1.), a separate heatmap may be opened for each attribute value. Try out different normalization techniques

SOM Visualization – Data Filtering & Grid Interaction

1.) Select an attribute to exclude it for the next SOM training

4.) Choose another color (click on the Color cell)

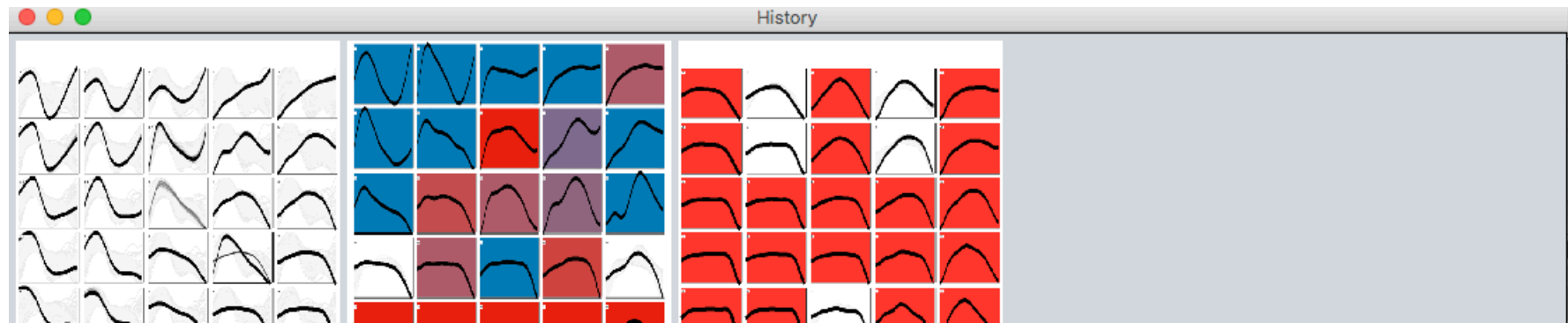


2.) Right click on a cell to open the context menu (cells may be deleted or pinned)

3.) Cells may be drag & dropped (moved) & pinned

SOM - History

12



1.) Click on the SOMs to bring them to the front of the screen



Possible Task I

13

- Sumimasen_Entschuldigung
 - Data set contains recorded utterances “sumimasen” (JP) & “entschuldigung” (GER) [--> Excuse me]
 - German and Japanese speakers and learners (L1, L2)
 - Dataset is well preprocessed and needs no further configuration
- Visualize the whole dataset (select the utterance attribute)
- Filter out the “Entschuldigung” Utterances (in the 1st SOM)
- Visualize the different speaker nationalities (“ns” attribute)

Sumimasen - Example

- **Example:**
 - Analysis of Pitch Contours via Self-Organizing Maps
 - in combination with Visual Analytics
- **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)
 - learners of Japanese were German and vice versa



Possible Task II

15

- Use the “sumimasen” data set
 - Launch a SOM
 - Look at the actual vectors contained in the data set.
 - Pin certain panels and delete certain panels and the relaunch the SOM.

Possible Task III

- Beat the tunes
 - Data set of recorded non-sense words of Japanese and German speakers
 - A lot of meta-data is contained
 - Data set is not preprocessed
- Try out different preprocessings
- Visualize the whole dataset (select the “pitch” attribute)
- Filter out the “HH” Utterances (in the 1st SOM)
- Train another SOM (based on the 1st SOM) that filters out the “HL”
- Analyze the different dialects/segments/speakers (use the heatmaps)

HistoBankVis

Analyzing Language Change Based on Feature Occurrences

Possible Tasks

- Click on the stored examples on the Home Page and interact with and explore the data.
- Look at the interaction between Verb Type and V1 and Voice.
 - Select only the main verbs (VB) and modals (MD).
 - Choose the 5 time stages and compare each range with the previous range.
 - Visualize the interaction.
- Note that modals hardly ever occur and that something seems to be going on with passives and middles.
 - Edit your filter to deselect modals and actives.
 - Revisualize.
 - Can you detect an interesting pattern?

VisArgue

Analyzing Political Argumentation

Possible Tasks

<http://visargue.inf.uni.kn>

- Web-based framework, visit webpage and use demo account
 - Personal accounts available upon request
 - Default parameters can be used and example data is provided
- Analysis Tasks
 - Data Exploration
 - Statistical and Linguistic Analysis based on Micro-Linguistic Features
 - Hypothesis Generation and Verification
- Questions to Answer:
 - WHAT is being said? → Content Analysis
 - HOW is it being said? → Discourse and Argumentation
 - BY WHOM is it said? → Speaker Dynamics

Sample Data Format

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<section name="Conversation Name">
```

```
<topic>
```

```
<utterance name="Speaker Name 0" sprecherId="0"> Text of the first speaker turn.  
Depending on the corpus, alternatively, one document or one paragraph. </utterance>
```

```
<utterance name="Speaker Name 1" sprecherId="1"> Text of the second speaker turn. It's ok  
if two consecutive utterances are from the same speaker. </utterance>
```

```
<utterance name="Speaker Name 2" sprecherId="2"> The text of the third speaker turn.  
</utterance>
```

```
<utterance name="Speaker Name 1" sprecherId="1"> Now you get the pattern ;)  
</utterance>
```

```
</topic>
```

```
</section>
```

DoubleTree

Visualizations by Chris Culy



Exploring Texts

Chris Culy has designed a number of visualization possibilities

- <http://linguistics.chrisculy.net/lx/software/>
- You can either download these or use them on-line

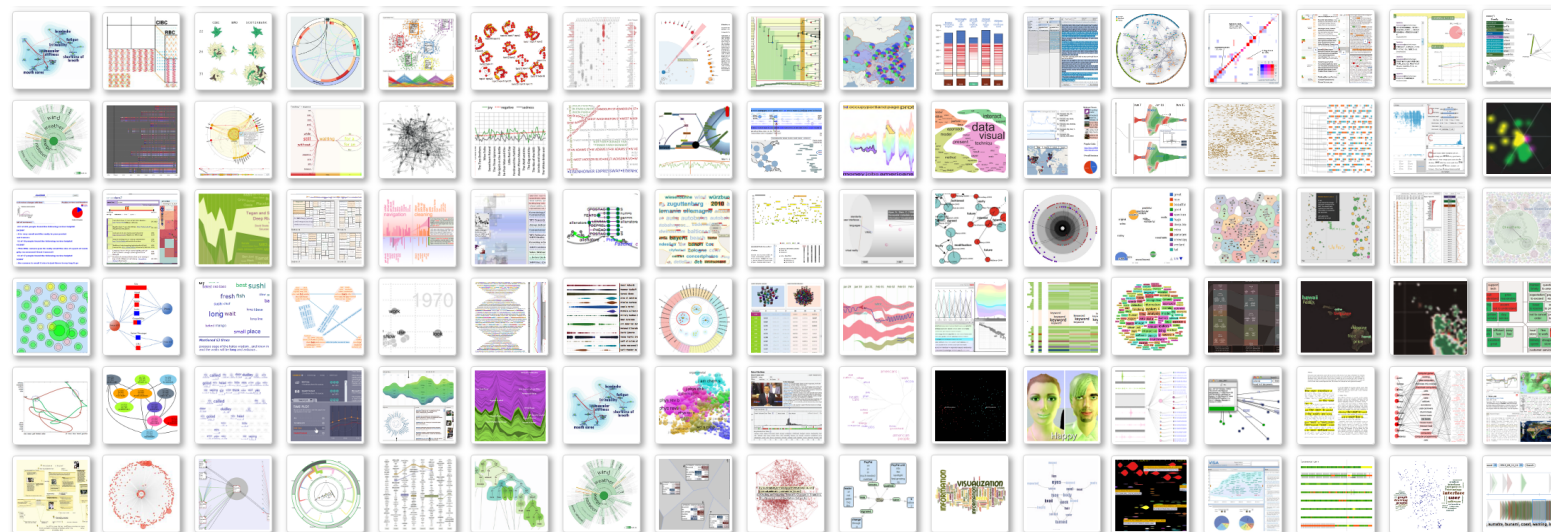
Suggestions:

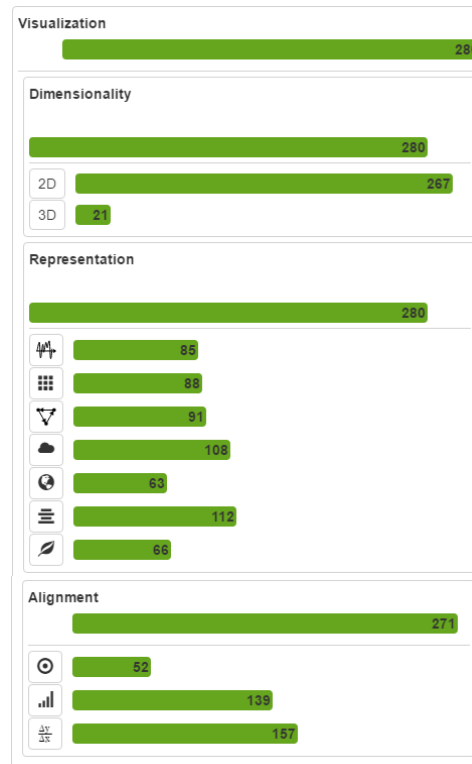
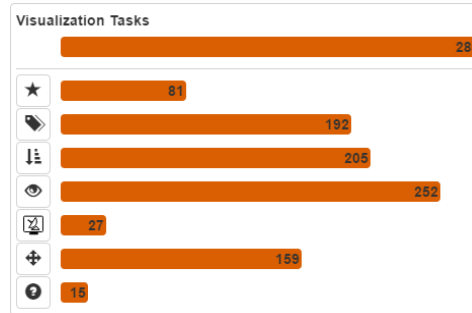
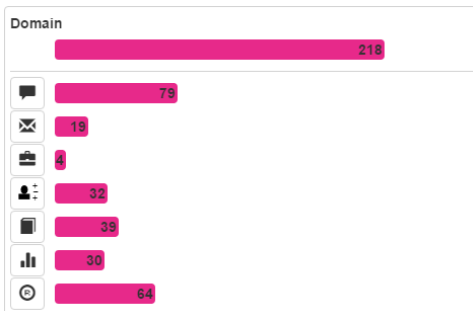
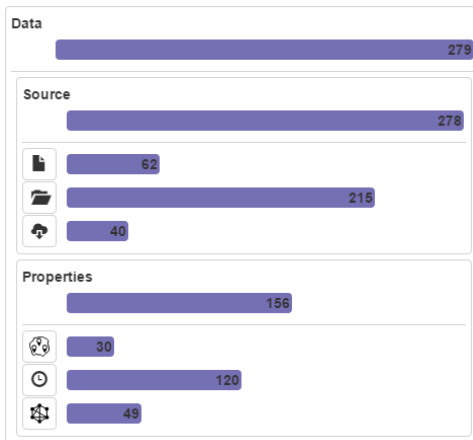
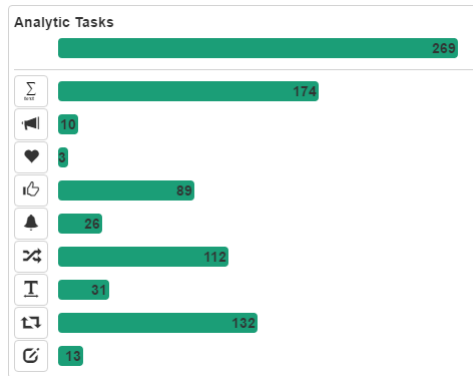
Explore the examples provided with the visualizations

- What are advantages/disadvantages of each?
- What would you like them to do that they can't?
- If you have tagged data available, try your own.



Collection of Linguistic Information Visualization





Text Visualization Techniques

<http://textvis.lnu.se/>

Text Visualization Browser

Developed by Kostiantyn Kucher and Andreas Kerren
ISOVIS group, Linnaeus University, Växjö, Sweden



Some Further Available Visualization Tools/Ideas

- <http://www.knime.org/>