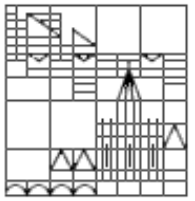
SPONSORED BY THE
Federal Ministry
of Education
and Research



Visual Analytics for Linguists

Miriam Butt & Chris Culy

ESSLII 2014, Introductory Course

Tübingen

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

PHILOSOPHISCHE
FAKULTÄT
Seminar für Sprachwissenschaft

Day 5 – Summary

- Interaction between Visual Analytics and Linguistics
- What kind of data?
- What kinds of hypotheses?
- Insights from working with the software?
- Outlook

What interests Visual Analytics?

- Interesting interactions
  - beyond just basic data types
  - push the boundaries/limitations of visual variables.
- Multiple dimensions (beyond 2)
- Time depth
- Cross-modular interactions.
- Not just coloring in bits of text that are of interest.
- Or drawing lines between pieces of data.
- In short**: Analysis at a Meta-Level**

What is complex in linguistics?

- Complex interactions between different parts of the grammar.
- Complex linguistic representations (trees, AVMs, pitch contours, etc.)
- Comparison of different languages/dialects
  - crosslinguistic (typology/dialectology)
  - diachronic (across stages of time)
- Complex interactions across different texts.

What is good for Linguistics and Visual Analytics?

- Sets of data which have complex interactions that you are interested in exploring.
- Interactions/data that are too difficult for you to eye-ball in its “raw” form.
- Large amounts of data that you want to have an “at-a-glance” overview of.
- Keim’s Mantra: overview first – details on demand

Data

- The Visual Analysis depends on the data.
- It does not make the data, it just helps you understand it.
- So:
  - be clear on what questions you are investigating
  - on what kind of data you are working with
  - whether Visual Analysis will be able to help
  - what kinds of interactive possibilities you might want
- You do not necessarily have to perform a statistical analysis (though that is becoming more and more common) – cf. the Levin Verb Classes Example.

Hypotheses

- Visual Analytics provides a fundamentally explorative approach to data.
- You can just take a bunch of data and go explore.
- However, it is good to have a hypothesis.
- Following cycle:
  - Formulate Hypothesis
  - Gather and Process Data
  - Visualize Data
  - Test Hypothesis and maybe reformulate Hypothesis
  - Reprocess Data
    (e.g., different annotations, focus on different features)
  - Revisualize
  - (Re)Test Hypothesis
  - Start Over

Working with the Software

- Particular low-level issues
  - Make sure the data is in the right format (e.g. UTF-8)
  - For the Cluster Visualization
    - Use the “customized” option for new files (not “quick”)
    - do not specify anything for the bigrams
    - do not try to compute with features that are not numbers

- Otherwise several of you:
  - identified bugs
  - identified features in the software that you would like to have

Very good and Thank You!

Working with the Software

Further reports on experiences/feedback?

The MOTH Manifesto

Ordinary researchers should have access to high level visualizations and analysis tools for their own data.

Culy 2014

Goals for Developing Visualizations

- Make them independent of any particular data set
- When possible, make them (also) independent of a particular application
  - E.g. as components
- Use common/easy file formats for the data
- Give examples!

Research in LingVis?

- Classifying the (higher level) types of data
- Classifying the *kinds* of tasks we want to do
  - The questions we want to answer
- Figuring out how to match data + task with visualizations
  - e.g. the different network visualizations
- Figuring out how visualizations can be connected together in applications

Outlook

- Where do we go from here?
- Where do we see the field as going?