Interactive Visualizations for Linguistic Analysis

Verena Lyding and Henrik Dittmann
{verena.lyding/henrik.dittmann}@eurac.edu

Institute for Specialised Communication and Multilingualism, EURAC, Bozen-Bolzano
Information visualization

“The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.” (Card et al., 1999)

Aim: providing a cognitive aid for
• the illustration of data and their structure/organization
• the analysis and manipulation of data

“Good visualizations use graphics to organize information, highlight important information, allow for visual comparisons, and reveal patterns, trends, and outliers in the data.” (Hearst, 2009)
Some examples

Classical ways of visualizing data include: graphs, networks, charts, diagrams, maps, but also text.
More examples

Recently, some visualizations of language data have been introduced, including word clouds, concept galaxies, sparklines, etc.

taken from: http://omnipotent.net/jquery.sparkline/

Mouse speed

Inline line graphs

Bar charts negative values:

Composite inline

Inline with normal range

Composite bar

Discrete

Discrete with threshold

Customize size and colours

Tristate charts (think games won, lost or drawn)

Tristate chart using a colour map

Box Plot

Pre-computed box plot

IN-SPIRE Galaxy visualization: http://infoviz.pnl.gov/tech_inspire.stm

taken from: http://www.dwds.de
How do visualizations work?

Information is transformed into graphics, using ‘visual variables’:

Taken from Carpendale (2003).
Constructing meaningful visualizations

Follow visualization principles.

“Sameness of a visual element implies sameness of what the visual element represents.” (Tufte, 2006)

Follow Gestalt psychology principles of perception, like proximity and similarity.

“Clutter and confusion are failures of design, not attributes of information.” (Tufte, 1999)

Don’t hide information without indicating what is left out.

Present information in context.

“Overview first, zoom and filter, then details-on-demand” (Shneiderman, 1996)
LInfoVis - Linguistic Information Visualization

The application of information visualization principles to display any kind of information concerning language and its use.

LInfoVis is a specialization of InfoVis.

And a particular challenge due to:
• structure and complexity of linguistic data
• textual elements

“The categorical nature of text, and its very high dimensionality, make it very challenging to display graphically.” (Hearst, 2009)
LInfoVis at EURAC

Motivation: **Development and implementation** of visualizations for language data. With focus on the **representation and analysis** of language resources, in particular corpora.

Project **running since the end of 2008**, by the language technologies group at EURAC; initiated by Chris Culy (now University of Tübingen).

By now: Development of several **visualization prototypes**, investigation of **application contexts**, visualization tools for linguistic projects.

www.eurac.edu/linfovis

We are generally interested in collaborations!
Visualizing language data

We can distinguish between visualizations for
a) the **presentation** of data
b) the **analysis** of data

which can be targeted to
1) **language data in context**, e.g. KWIC
2) **information derived from language data**, e.g. frequency lists
LInfoVis for data presentation

Data displays that visually highlight relevant aspects of the data.

Some examples for:
- text
- collocations
- occurrences of words over time
Visualization indicating frequencies of words in their textual context; frequencies are encoded by character size of words.
Graph visualization of the word „Ziel“ and its collocations as calculated based on the DWDS core corpus.
Chart displaying frequencies of the indicated words in newspaper text over a crucial period of time.
LInfoVis for analysis

Visualizations that **highlight data characteristics** and allow for the interaction with the display to explore the data.

Techniques from information visualization:
- search and filter, for focus and context
- overview, zoom-in, details on demand
- multiple views, and brushing and linking
Interactive visualizations

Corpus Clouds (Culy/Lyding, 2009)

- visualization of corpus query results
- multiple panels for different types of information
- interactive features for data exploration
Interactive visualizations

Double Tree (Culy/Lyding, 2010)

Double Tree shows a concordance in a compressed form that allows for interactive exploration.
Interactive visualizations

xLDDs - Extended Linguistic Dependency Diagrams
(Culy/Lyding/Dittmann, 2011b)

xLDDs support the analysis of dependency structures by providing
• a set of visual features (such as color, size and shape) for the presentation of relations
• user controls for focusing on specific information
Visualizing derived information

Linguistic analyses are not only concerned with the linguistic data itself, but also with information about this data.

- e.g. quantitative analysis

Parallel Coordinates (Inselberg, 2009) is a common visualization for high-dimensional data.

For linguistic analyses, we have developed Structured Parallel Coordinates (SPC), (Culy/Lyding/Dittmann, 2011a):

- inherent ordering of the axes
- advanced methods for filtering, selection and highlighting
Parallel Coordinates

Visualization originally invented by d'Ocagne (1885).

In modern Information Visualization developed and popularised by Inselberg in 1959 (cf. Inselberg, 2009).

Taken from Heer et al. (2010).
Interactive visualizations

SPC - Structured Parallel Coordinates (Culy/Lyding/Dittmann, 2011a)

n-grams and frequencies application
Interactive visualizations

SPC - corpus comparisons

evolution of verb constructions over time and register
Interactive visualizations

SPC - ranking comparisons

Live demo:
- using ranking comparisons for sub-corpus analysis
- automatic re-ordering of the axes by similarity
Summing it up

- Information Visualization has a lot to offer for the **description** and exploration of complex data.

- Our LInfoVis work aims at **combining insights and methods from InfoVis and linguistic research settings**.

- Visualizations need to be put to the test in different areas of linguistic analysis. The **users’ needs and experiences** are an important base for:
  - improving existing software and
  - pursuing new directions in LInfoVis
Thank you!

{verena.lyding/henrik.dittmann}@eurac.edu

www.eurac.edu/linfovis
Bibliography


Further links and references


“Visualization as Part of the Linguistic Processing Pipeline” by Culy, C. / Lyding, V., presented at the Linguistic Processing Pipelines workshop at the GSCL conference, 29 September 2009 in Potsdam, Germany.