Visualising development in longitudinal first language acquisition data

Danica Pajović, Steven Moran, Jekaterina Mažara, Sabine Stoll
University of Zurich

Background

These visualisations have been devised as part of the *ACQDIV* project („Acquisition processes in maximally diverse languages: min(d)ing the ambient language", P.I. Sabine Stoll). To answer the question how a child can learn any of the world's approximately 7000 languages, ACQDIV analyses data from ten languages taken from five maximally diverse language clusters.

While the inclusion of large collections of data from languages that are understudied in terms of language acquisition is extremely beneficial, the amount of data also introduces new challenges.

Visualizations can be helpful in analyzing large corpora of naturalistic acquisition data more effectively, since tables and traditional figures often cannot cope with the amount of data and the various potentially relevant factors/variables.

Why use visual analytics?

Data exploration
- In visual analytics, visualizations are not restricted to the representation of results
- They allow versatile exploratory approaches to research questions

Guiding principles of Visualization
"Overview first, zoom and filter, then details-on-demand." (Schneiderman et al. 1996)
➤ A visualization should be effective on a global as well as a local level

"Analyse first, show the important, zoom and filter, analyze further, details-on-demand." (Keim et al. 2010)
➤ Focus can be split between data analysis and visualization

Different possibilities
- static figures **(Fig.1)**
- interactive and/or dynamic figures **(Figs. 2 & 3)**
- see also live demonstration of interactive figures

Static example: tree

overview
(excerpt, see computer for full view)

zoom-in
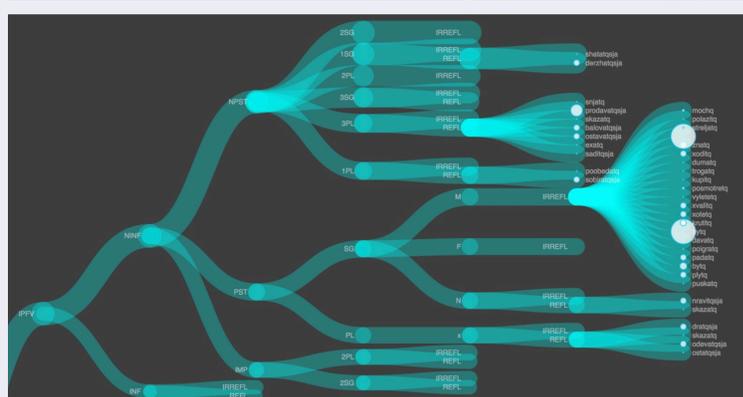
Fig.1: Tree diagram showing all verbs uttered by Russian adults to a target child between the ages of two and three. Thickness of lines and circles indicates frequency of occurrence. The overview shows the distribution of individual verbs, as well as that of aspect and inflectional categories. We can zoom in on particular grammatical categories of the tree, as shown in the lower excerpt (imperfect past singular masculine forms).
These visualizations have been created using mainly the d3.js library (Bostock 2012).

Interactive example I: Sunburst
- From x/y-coordinate space to radial system
- reduces the space used on the screen, still represents the hierarchical structure of our data
- data can be represented fully in one picture or in increments (e.g. month after month) to show the development of verb forms
- if a central layer is clicked, the circle reassembles, showing only the distribution of forms contained in the selected segment
- see demo for fully interactive exploration of the data
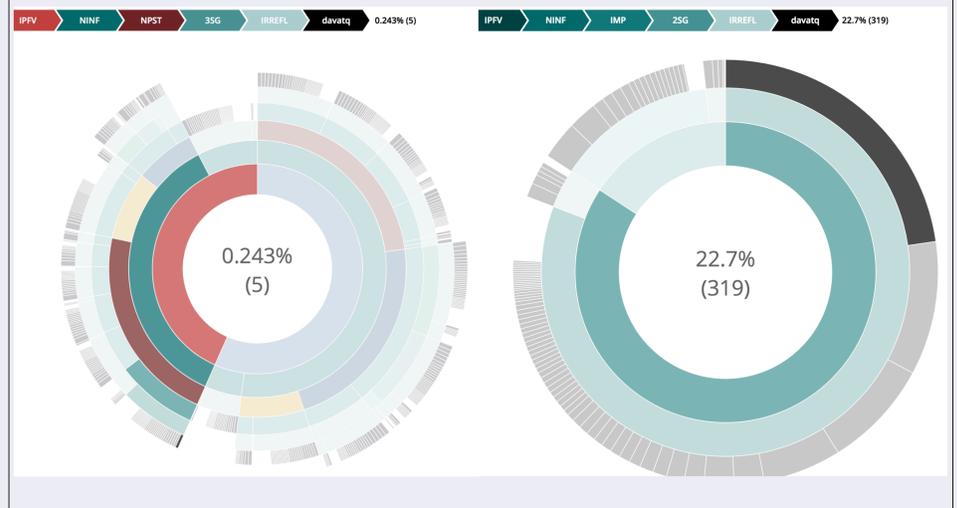
Fig.2a: Sunburst visualization showing the verb forms uttered by a target child in Russian.
Fig.2b: Zoom in on a specific age range on the imperfective portion of the verbs uttered by the child.

Interactive example II: Growing pie charts
- "curve" shape shows overall development over time, we see when a verb form enters a child's inventory
- size of pies shows number of inflectional forms a lemma appears with in a recording period
- summary column at the end shows all forms used with a lemma in all recordings
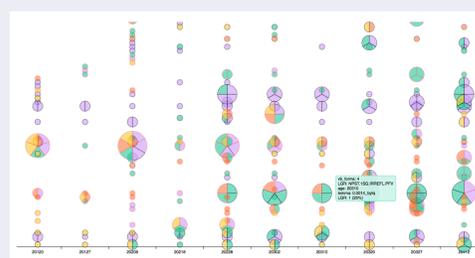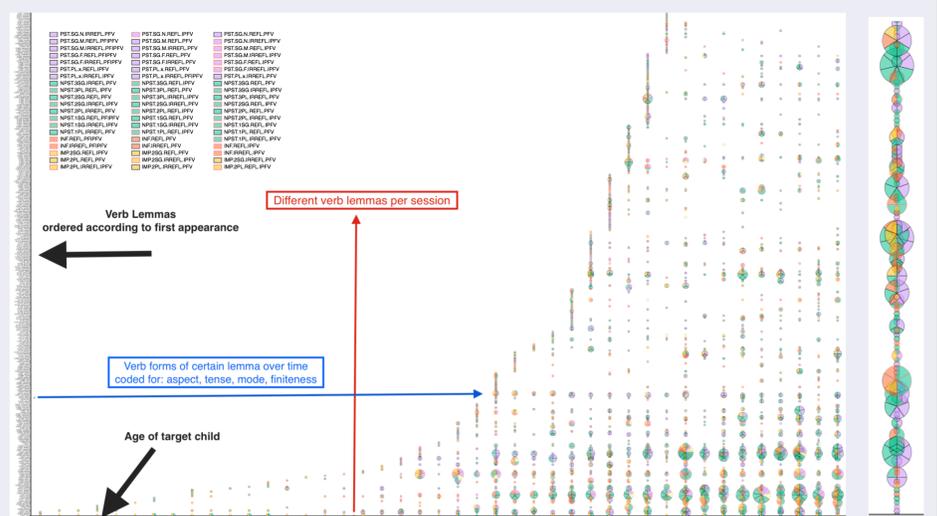
- hovering over parts of the individual pie charts displays details about the form that was used in that particular instance

Contact
corresponding author: Jekaterina Mazara (jekaterina.mazara@uzh.ch)
ACQDIV project home page: http://www.acqdiv.uzh.ch/en.html
www.comparativelinguistics.uzh.ch

References
- M. Bostock. D3. js. Data Driven Documents, 2012.
- D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann. Mastering the information age-solving problems with visual analytics. Florian Mansmann, 2010.
- B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In Visual Languages, 1996. Proceedings., IEEE Symposium on, pages 336–343. IEEE, 1996.