Complex Data Visualization: Visualizing language production dynamics via keystroke analysis

Adam Goodkind
Department of Linguistics

Data Science Research Day
June 25, 2018
Outline

• Complex data visualization overview
• Studying keystrokes (why/challenges)
• Tools for visualization
• TypeShift program
Open Sources

• Data
  • http://www2.latech.edu/~mike/TypingForTenData/

• Software
  • https://angoodkind.shinyapps.io/TypeShift/
  • https://github.com/angoodkind/TypeShift
  • https://shiny.rstudio.com/
Complex Data Visualization:
What is the goal of data visualization?

• Def: Create a visual representation of data
• Clearly communicate relevant comparisons
  • Trends over time
  • Rankings
  • Part-to-whole
Complex Data Visualization:
What options do we have?

• Decide how much audience needs to know
  “Overview first, zoom and filter, then details-on-demand.”
  – Ben Shneiderman
Keystroke Dynamics:
Why is it interesting?

• Language production is a window onto cognition
• Typing is precise and relatively easy to measure as compared to speech
• Every typist has a unique typing pattern
Keystroke Dynamics:
Background

• World War II, “The Fist of the Sender”
• Today, “The Hands of the Typist”
Keystroke Dynamics:
How do we measure it?

• We measure all hold and flight times from when a typist begins a word to when they complete the word.
Our Keystroke Research

• Goals
  • Create a static picture of a dynamic process
  • Capture what kind of typist and typing session are we looking at?

• Methods
  • 1,013 Louisiana Tech students
  • Answered 10-12 different short essay prompts
    • What did you have for breakfast?
    • Design a course curriculum and how you would teach it.

• What we want to measure
  • How a typist compares to other typists
  • How an individual typing session compares to that typist’s other sessions
  • How different linguistic units are produced
Our Solution: TypeShift overview

- Select Typist
- Select Question
- Filter syntactic units (Parts of Speech)
- Filter semantic units
- Add trend lines to compare

Overall typing session
Single word token
Visualizing a typing session

• Overall typing rate
Visualizing a typing session

- Intraword typing rate
Visualizing a typing session

• Gaps in typing
Visualizing a typing session

• Revision behavior
Typing rate comparisons: Speed & Consistency

- Average for user & population vs. Slow for user & fast for population
- Consistent vs. Inconsistent
Typing rate comparisons: Revision behavior
Typing rate comparisons: Single token

![Graph showing overall pause distributions and selected user pauses for single token typing rate comparisons.](image)
Takeaways

• When visualizing complex data:
  • Decide what needs to be emphasized
  • Details will need to be sacrificed
  • Try to create a single objective baseline

• Visualizing language
  • Important to emphasize relevant factors/differences
  • Possible to capture both continuous and discrete nature of process
  • Informative displays of typing information:
    • Can tell a lot about language production
    • Shed light on underlying cognitive processes
Thank You