

Statistisch gestützte Visualisierung von Informationsgliederungen in den Bundestagsreden

Dr. Zakharia Pourtskhvanidze, pourtskhvanidze@em.uni-frankfurt.de
Institut für Empirische Sprachwissenschaft. Goethe-Universität Frankfurt/M

Der handlungsbezogene Aspekt der Sprache wird besonders deutlich in den öffentlichen Auftritten der Politiker. Neben der für die gesprochene Sprache spezifischen Verwendung von lexikalischen Mitteln erscheinen in den Reden zuhörerorientierte Einsetzung von syntaktischen Konstruktionen (Z.B. Anrede, Rhetorische Fragen) und auf die Interaktion abgestimmte Realisierung von pragmatischen kommunikativen Strategien (Z.B. Gesichtserhaltung).

Der linguistisch interpretierte Begriff *Hervorhebung* bezüglich der Gestaltung des informationellen Gehalts einer Äußerung eignet sich besonders als ein Ausgangspunkt für die Verbildlichung (Visualisierung) besonders prominenter Informationseinheit im Vergleich zur informationell neutralen Einheit eines gesprochenen Diskurses.

Die **empirische Basis** der auf dem Poster beschriebenen Analyse stellt das Plenarprotokoll (Stenographischer Bericht) der 2. Sitzung des Deutschen Bundestages am 18. November 2013 zum Thema *die Abhöraktivitäten der NSA und die Auswirkungen auf Deutschland und die transatlantische Beziehungen* dar (ca. 20.000 Token bei 17 Sprechern aus 5 Parteien).

Aufgrund der linguistischen Analyse von Texten wird eine **Ranking-Tabelle** der grammatischen Instrumente der Informationsgliederung erzeugt. In der Tabelle bekommen die einzelnen Instrumente (Partikeln, Vorfeldbesetzung, Spalt-Satzstruktur, Satzarten, Referentielle Bezüge (Anapher bzw. Katapher) etc.) in den Zahlen ausgedruckte Werte (-15 min. ... 0 ... 75 max.). Die fokussierende (rhematisierende) Instrumente bekommen tendenziell hohe Werte (daher stärker hervorgehoben), wogegen die topikalisierte (thematisierende) Instrumente bekommen tendenziell niedrige Werte (daher flachere Visualisierung).

Der *erste Schritt* des Visualisierungsvorgangs sieht die automatische **Ersetzung** des jeden in der Ranking-Tabelle notierten Tokens resp. Skopus im Protokoll-Text mit der entsprechenden Zahl, während der Rest von Tokens im Text automatisch mit dem Wert „0“ ersetzt wird. Die Texte werden in CSV-Dateiformat abgebildet und damit für unterschiedliche Statistik-Analysen endgültig aufbereitet.

Die Visualisierung der Daten erfolgt im *zweiten Schritt* durch eine freie Programmiersprache für statistisches Rechnen und statistische Grafiken **R**.

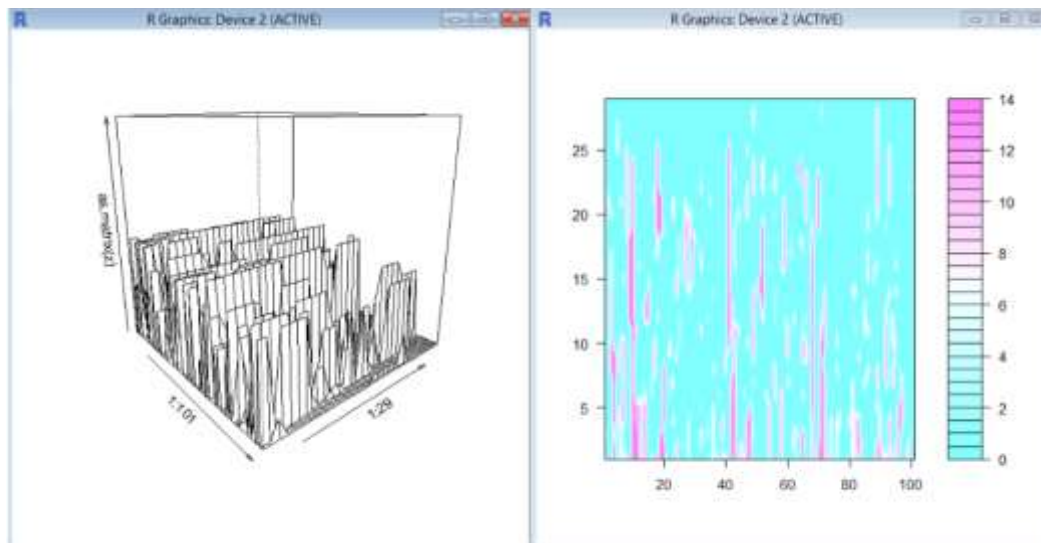


Abb. 1. Visualisierung der Rede des Abgeordneten G. Gysi (Die Linke) mit R-Funktionen: *persp* (links) und *filled.contour* (rechts)

Im Kontext der Auffassung *Make your data tell a story* eröffnet die Visualisierung von Informationsgliederungen die Möglichkeiten die generierten Bilder auf die Zeitschiene zu ordnen und zwar limitiert auf eine Person oder einem Themas.

- Gibt es mögliche Kausalitätstendenzen zwischen den „Informationslandschaften“ der Reden und dem Thema desselben?
- Lässt sich ein Dominanz-Muster für die bestimmten sprachlichen Instrumente der Informationsgliederung abhängig vom Thema etablieren?
- Gibt es individuelle Muster von informationellen Hervorhebungen?
- Lässt sich die Entwicklung des Redestils eines Sprechers in den Visualisierungen von Informationseinheiten über die Zeit dokumentieren?

Eine disziplinübergreifende Analyse mit den Bezügen auf die Gender- und Meinungsforschungen ist denkbar.

Vorgesehen ist die Erweiterung der empirischen Daten in Richtung der Vielfalt der Genres und die Einsetzung von alternativen Visualisierungssoftware (z.B. RStudio).

Das weitere Forschungsvorhaben wird gegenwärtig an der Goethe-Universität für die Aufbau einer integrativen Kooperationsplattform zwischen den Fachbereichen „Sprach- und Kulturwissenschaften“ und „Informatik und Mathematik“ bedacht.