



UNIVERSITÄT
ZU KÖLN

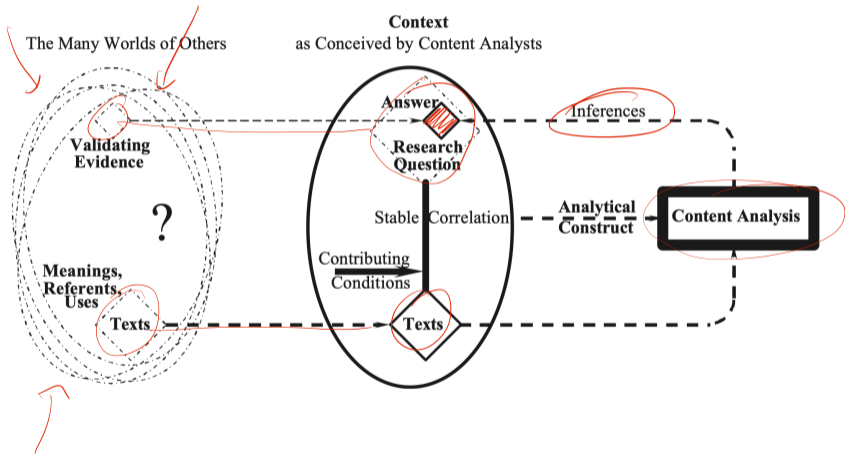
Content Analysis Workflow

HS Anwendungen der Computerlinguistik

Nils Reiter

`nils.reiter@uni-koeln.de`

26. Oktober 2023

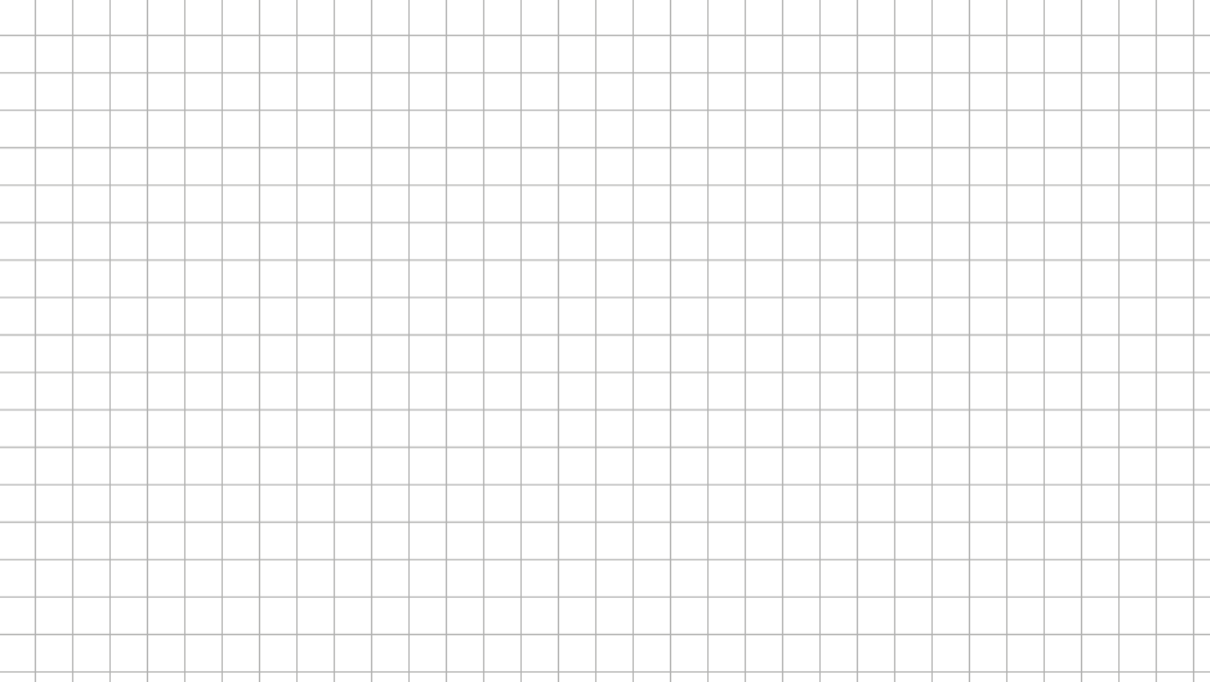


Three Examples

Welcher Text?
 Welcher Kontextfaktoren?
 Indikatoren?

Content-Analytical Research Questions

- ▶ What is the Russian public opinion about the war in Ukraine?
- ▶ How did the portrayal of female characters in German ^{drama} theatre evolve during the 19th century?
- ▶ What are the positions of political parties in Germany on the climate catastrophe?
 (change)

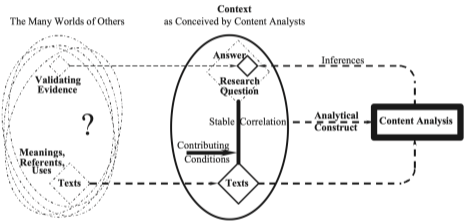


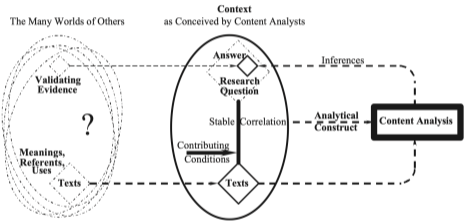
4

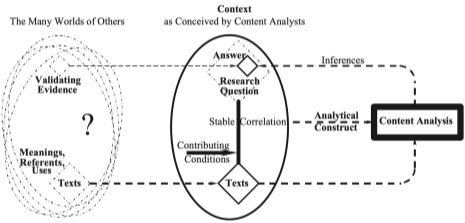
THE LOGIC OF CONTENT ANALYSIS DESIGNS

As a technique, content analysis relies on several specialized procedures for handling text. These can be thought of as tools for designing suitable analyses. This chapter outlines the key components of content analysis and distinguishes among several research designs, especially designs used in the preparation of content analyses and designs for content analyses that collaborate with other research methods to contribute to larger research efforts.

Figure: Klaus Krippendorff (2019). *Content Analysis: An Introduction to its Methodology*. 4th. Sage, Chapter 4







Components

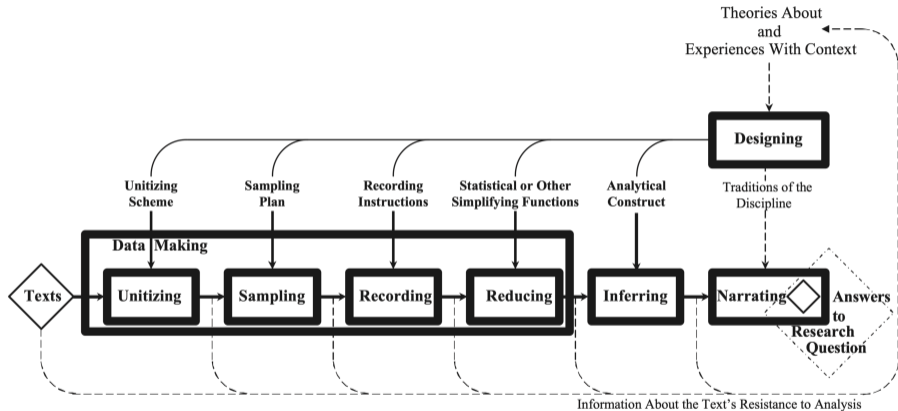


Figure: Components of Content Analysis (Krippendorff, 2019, Fig. 4.2)

Components

- ▶ Not all components are always needed
- ▶ Developing the workflow is different from executing it Krippendorff (2019, 91)
- ▶ “Data making”: There is no raw data, and nothing is purely data-driven

Components

Unitizing

- ▶ Systematically select units out of a stream of undifferentiated text
- ▶ Do we look at documents, chapters, speech turns, time periods, ...?
- ▶ What do we omit?

Components

Sampling

- ▶ Limit observations to a manageable subset of all conceivable units
- ▶ Subset should be representative of the whole
 - ▶ “Ideally, an analysis of a whole population and an analysis of a representative sample of that population should come to the same conclusion.” Krippendorff (2019, 88)

Components

Sampling

- ▶ Limit observations to a manageable subset of all conceivable units
- ▶ Subset should be representative of the whole
 - ▶ “Ideally, an analysis of a whole population and an analysis of a representative sample of that population should come to the same conclusion.” Krippendorff (2019, 88)
- ▶ (Selected) strategies in practice
 - ▶ Random sampling
 - ▶ Systematic Sampling: Looking at every n -th unit
 - ▶ Stratified Sampling: Consider different sub populations explicitly
 - ▶ Convenience Sampling: Use existing selections of texts
 - ▶ “Convenience samples present content analysts with the potential problem of having to undo or compensate for the biases in such data, taking into account the intentions that brought these texts into being and into the analysts’ hands.” Krippendorff (2019, 124)

Components

Recording/Coding (= Annotation)

- ▶ Make (our) reading of a unit explicit
- ▶ Manual through annotators (= coders)
- ▶ Automatic through machine learning models or rule-based systems
- ▶ Categories dependent on research questions and context

Components

Recording/Coding (= Annotation)

- ▶ Make (our) reading of a unit explicit
- ▶ Manual through annotators (= coders)
- ▶ Automatic through machine learning models or rule-based systems
- ▶ Categories dependent on research questions and context
- ▶ Example: “Codebuch zur Inhaltsanalyse der Fernsehdebatte zur Bundestagswahl am 22. September 2013”

Components

Reducing

- ▶ Extract statistics over recorded categories
- ▶ Entails information loss
- ▶ Often also aggregation

Components

Reducing

- ▶ Extract statistics over recorded categories
- ▶ Entails information loss
- ▶ Often also aggregation
- ▶ Summary statistics
 - ▶ Numeric variables: Mean, standard deviation, Median, Quartiles, IQR
 - ▶ Nominal variables: Distribution, majority
 - ▶ Ordered variables: Distribution, majority, distances

Components

Inferring

- ▶ “Analytical Construct”
- ▶ Which inferences do we draw from the results?
- ▶ Why are we allowed to draw them?
- ▶ Arguments taken from context – question of validity most important

Components

Inferring

- ▶ “Analytical Construct”
- ▶ Which inferences do we draw from the results?
- ▶ Why are we allowed to draw them?
- ▶ Arguments taken from context – question of validity most important
- ▶ Justifications
 - ▶ Previous success
 - ▶ Expert knowledge and experience
 - ▶ Established theories

Components

Narrating

- ▶ Making the results comprehensible to others
- ▶ How and to whom depends:
 - ▶ Practical significance to funding agencies / general public
 - ▶ Scholarly value and appropriate methodology to scientific peers
 - ▶ Making recommendations for actions (e.g., legal or political)

Components

Narrating

- ▶ Making the results comprehensible to others
- ▶ How and to whom depends:
 - ▶ Practical significance to funding agencies / general public
 - ▶ Scholarly value and appropriate methodology to scientific peers
 - ▶ Making recommendations for actions (e.g., legal or political)
- ▶ Narrating scientific results is an inseparable part of scientific work
- ▶ If no one knows about your results, it's as if they didn't exist

Components

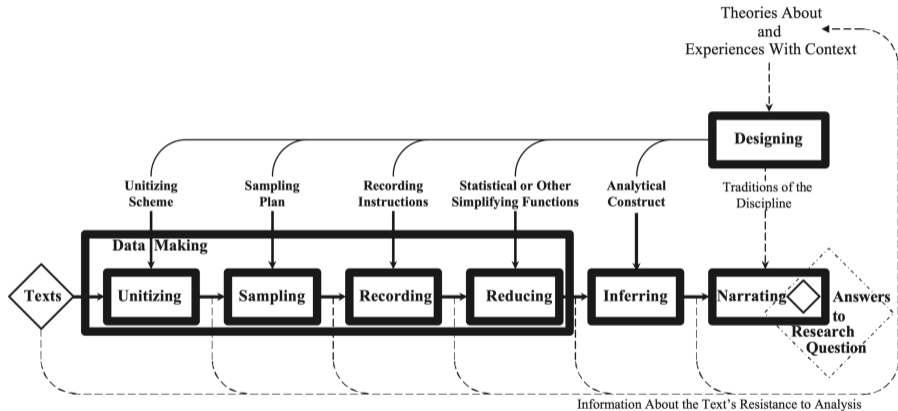


Figure: Components of Content Analysis (Krippendorff, 2019, Fig. 4.2)

References I



Krippendorff, Klaus (2019). *Content Analysis: An Introduction to its Methodology*. 4th. Sage.