Computer-assisted Textual Analysis in Sociology II

Quantitative and qualitative coding

For a large amount of texts, quantitative coding is almost necessity (time, cost, reliability). We can use the inductive and exploratory analysis which combines textual data frequency analysis with non-textual variables like gender, infection rate or economic indicators.

For a little amount of texts (e.g. 10 biographic interviews), the use of quantitative coding is limited for two reasons:
(a) frequencies are usually too low to detect significant patterns,
(b) numbers represent just a fraction of a complex meaning communicated through texts.

Qualitative coding

Qualitative coding proceeds by close reading of data and manual coding.
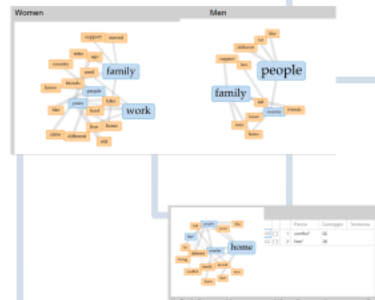It can be deductive with pre-existing code-book or inductive generating codes from the data.

Problem of qualitative coding

Dictionary based vs. inductive coding

family analysis:

Brothers and sisters we were given this book to look around us better.

Discussion of homework

Carola

Emad

Mahmud

Francesca

Computer-assisted way to improve inter-coder agreement in qualitative coding

There is no way to completely avoid disagreement between coders, because every text is open to multiple interpretations (by human readers). However, it is possible - with the help of software - to try to align coders as much as possible or at least make disagreements visible and acknowledge them in the analysis.

Coding - an overview

Coding is an analytical technique by which we assign relevant parts of empirical data to analytical units, called codes.
Coding reduces the inherent complexity of social science data and allows for further systematic analysis.

Brothers and sisters, we were given this book to look around us better.

code A
code B
code C

Concluding remarks

Data Analytics for Students of Social Studies and Humanities. Tuesdays 12:20-13:50
https://ufal.mff.cuni.cz/courses/npfl134

Martin Hájek (martin.hajek@fsv.cuni.cz)

Discussion of homework

Carola

Comparison of male and female migrants stories

Comparison of stories by economic status of country of origin and destination.

Women

Men

movement to countries with lower GDP

Emad

- Most of the migrants have migrated to countries with equal GDP and many others have migrated to countries with higher GDP. But, much less migrants have migrated to countries with lower GDP.

- Migrants who have migrated to countries with equal GDP, have spoken about "better life" more than migrants who have migrated to countries with higher GDP. None of the migrants who have migrated to countries with lower GDP has spoken about "better life". On the other hand, female migrants have had more tendency to migrate to have a "better life" than male migrants have.

- Economic situation of their home countries has been mentioned almost equally by Male and Female migrants as one of the reasons to migrate to another country. Meanwhile, more male migrants have migrated in order to find a new job than female migrants have.

- Based on the narratives of migrants, people who have migrated to countries with higher GDP, "miss" their country of origin, their families and friends, and the past years of their lives in their countries more than migrants who live in the countries with equal GDP. Migrants who live in countries with lower GDP have this sense of missing less than others migrants do. On the other hand, this feeling of "missing" is more in female migrants than male migrants. This applies to migrants in countries with higher, equal and lower GDP.
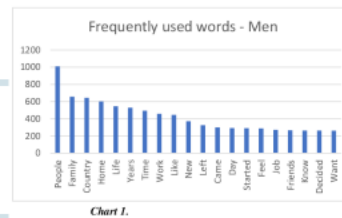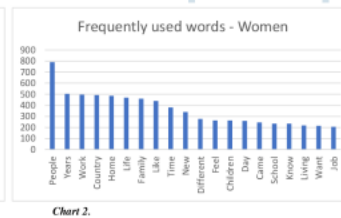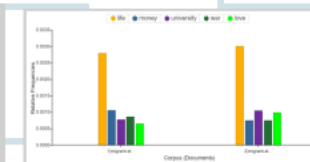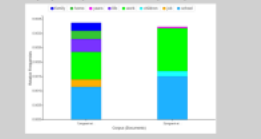
Mahmud

Frequently used words - Men

Chart 1.

Frequently used words - Women

Chart 2.

Francesca

MEN

WOMEN

Moreover, men are more concern of the stability of the whole family unit and this led to a higher apprehension of the quality of their "job", instead women are specifically concern of children.

Coding - an overview

Coding is an analytical technique by which we assign relevant parts of empirical data to analytical units, called codes.
Coding reduces the inherent complexity of social science data and allows for further systematic analysis.

Brothers and sisters, we were given this book to look around us better.

code A

code B

code C

Quantitative and qualitative coding

For a large amount of texts, quantitative coding is almost necessity (time, cost, reliability). We can use the inductive and exploratory analysis which combines textual data frequency analysis with non-textual variables like gender, infection rate or economic indicators.

For a little amount of texts (e.g. 10 biographic interviews), the use of quantitative coding is limited for two reasons:
(a) frequencies are usually too low to detect significant patterns,
(b) numbers represent just a fraction of a complex meaning communicated through texts.

Qualitative coding

Qualitative coding proceeds by close reading of data and manual coding.
It can be deductive with pre-existing code-book or inductive generating codes from the data.

Dictionary based vs. inductive coding

codebook:
kinship
actions
objects

family analysis:

kinship

action

object

Brothers and sisters, we were given this book to look around us better.

family-like address

donation

knowledge

improvement

The problem is that we never know whether the result is valid, whether we have observed the correct elements of the phenomenon under study in our data, because the dictionaries were created on other data.
The problem of agreement between coders.

Most often, researchers combine inductive and deductive strategies of coding.

Validity is not a major issue, but the problem of consistency between coders is critical.

Problem of qualitative coding

Task: code all passages referring to the narrator's discomfort.

"(…) Achol was faced with the hard choice to leave the Protection of Civilians site following the clashes due to her family's ethnicity. Now living in a decrepit church compound with other South Sudanese from the Dinka tribe, Achol and her family are forced to readjust in their temporary accommodations. 'Even though I am out here, there is still a huge need for a nurse but it is difficult to work without supplies,' she stresses. 'I delivered three babies in the past week.' Now her focus is building a future in Malakal town. 'I cant go back to the PoC site, but, honestly, I don't want to go anywhere else. Malakal is my home.'" [extract from the story of Achol]

This disagreement is not surprising because it reflects the varying intensity and type of discomfort that we subjectively experience.

Computer-assisted way to improve inter-coder agreement in qualitative coding

There is no way to completely avoid disagreement between coders, because every text is open to multiple interpretations (by human readers). However, it is possible - with the help of software - to try to align coders as much as possible or at least make disagreements visible and acknowledge them in the analysis.

reQual software

re stands for reproducibility and Qual for qualitative analysis
It is an open-source, free and multipurpose software.
Available as collaborative web app and single user desktop program.
requal.fsv.cuni.cz

It's still work in progress!

https://requal.fsv.cuni.cz/app/

Logging in.
Opening or creating a project.
Adding documents.
Managing users and permissions.
Creating codes and categories.
Workdesk – coding data.
Analysis – filtering and exporting segments
Report – the main tool for enhancing transparency and inter-coder agreement
   • Summary - number of coded segments for documents and coders
   • Agreement – overlap by coder in characters & segments – numeric and graphic
   • Browsing coded documents – discussing highly overlapping and non-overlapping
      segments; a base for improving inter-coder agreement
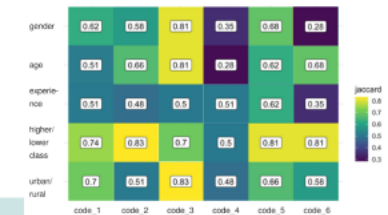User attributes – still unfinished feature.

It is well known to sociologists that members of society see reality according to their social position, including gender, age, ethnicity, class etc.

No CAQDA software takes into account the social positionality of analysts.

The project team can set up relevant personal characteristics such as gender, age, ethnicity, social class or research experience, and each coder will enter these attributes in the application. Other users can not see attributes of others.

After the preliminary coding of the sample data, visualisation of agreement between categories of coders will be displayed.

The purpose of this feedback is to stimulate discussion whenever differences between categories of coders surface to deal with the possible socially conditioned pre-understanding of data.

It is well known to sociologists that members of society see reality according to their social position, including gender, age, ethnicity, class etc.

The project team can set up relevant personal characteristics such as gender, age, ethnicity, social class or research experience, and each coder will enter these attributes in the application. Other users can not see attributes of others.

After the preliminary coding of the sample data, visualisation of agreement between categories of coders will be displayed.

No CAQDA software takes into account the social positionality of analysts.
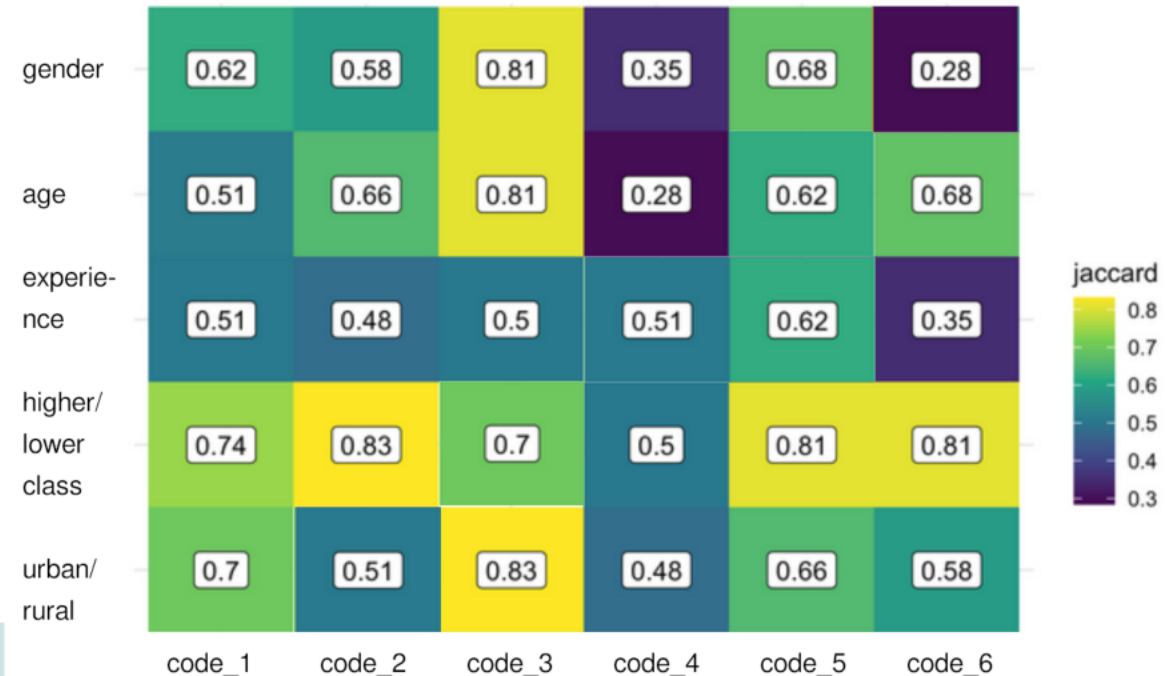
The purpose of this feedback is to stimulate discussion whenever differences between categories of coders surface to deal with the possible socially conditioned pre-understanding of data.

Concluding remarks

- In sociology, software is necessary for the quantitative and useful for the qualitative analysis of textual data.
- The key question is whether to choose a deductive (confirmatory) or inductive (exploratory) approach, or a pragmatic combination of the two.
- It is always beneficial to include non-textual variables in working with textual data.
- In any analysis, ensuring transparency of the analytical process is a requirement.

Thank you for your attention!