Visual History Archive in the Social Scientific Research
Some remarks and experiences from the user's perspective

Annual CLARIN Meeting
Jakub Mlynář
October 22, 2013
Prague

Malach Centre for Visual History
Faculty of Mathematics and Physics
Charles University in Prague
Structure of my talk

(I) What do we have?

(II) Research experience
   (II.1) Theoretical framework
   (II.2) Empirical aspects

(III) Problem areas
   (III.1) Transcription
   (III.2) Search

(IV) Potential improvements and other suggestions
What do we have?

What is the VHA and how is it related to the social sciences?

The Archive:

- **52,000 interviews** with Holocaust and Rwanda genocide witnesses and survivors in 33 languages
- **105,000 hours** of audiovisual material digitalized, indexed, searchable database accessible on-line at 30 access points

The Interview:

- **135 minutes** in average
- **60 %** about WWII, **20 % + 20 %** about pre- and post-war experiences
- **methodology**: oral history / life story / narrative interview / thematic interview

⇒ interdisciplinary relevance
Research experience: Central topics

(1) What is the relationship between memory and identity from the perspective of social theory?

(2) What influence does the social interpretation of the historical events have on individual identity and ways of subjective reflection of (auto)biography?

(3) How is the perception of Jewish identity constituted by the personal recollections and social memory in case of Czech and Slovak Holocaust survivors?
Research experience: Theoretical framework

**sociological view:** mutability and never-ending (re)construction of identity in social interactions (language and agency)

- identity is a process/project/performance, not a fixed „thing“!

**identity:** postulate (Bauman 2004), reflexivity (Neisser 1988), narrativity (Ricoeur 1990)

**memory** has a crucial role in the process of constituting and acquiring the individual, social and collective identities

**identity** as an interplay of **sameness and difference** (+ synchronic & diachronic persp.)

*Think of the individual as embodying a dialectic of sameness and difference. We detect the sameness by seeing what persists within the constant change of our lives. We detect the difference by seeing what has changed against the background of sameness.*

Holland, N. N.: „Reading and Identity“
Research experience: Empirical aspects

almost 52,000 Holocaust survivor testimonies in many languages (over 1,000 interviews in Czech and Slovak) with sophisticated user interface

36 selected interviews ➔ partial transcription ➔ interpretive analysis

focus on narrative/language expressions of collective identities (and membership categorization), multiple individual identities (i.e. „Czech and Jewish“ × „Czech or Jewish“), conflicting identities etc. in the Czech/Slovak testimonies from the VHA

multiple collective identities:
state identity – ethnic identity –
religious identity – political identity

specific features of the (Czech/Slovak) interviews: mostly people who remained in Czechoslovakia (no emigration); filmed in 1990s (particular Zeitgeist); working identity – „Holocaust survivor“ ➔ possible identity exaggeration or restraint
transcription is necessary for the social scientific analytical methods; it is a crucial part of the analysis (Silverman 1993; Cameron 2001: 43).

„It is now widely recognised that it is all but impossible to produce a transcription of a research interview, or any other type of conversation, which completely captures all of the meaning that was communicated in the encounter itself. Any transcription of speech must therefore be understood as a compromise.“ (Elliott 2005: 51)

very different approaches to transcription:

Focus on **content**  Focus on **context**  Focus on **form**

**ANTHROPOLOGY**  **PSYCHOLOGY**  **LINGUISTICS**

**ORAL HISTORY**  **SOCIOLOGY**  **SOCIOLINGUISTICS**
Problem Areas: Transcription

ORAL HISTORY
Focus on content

Focus on form

SOCIOLINGUISTICS

RL: When did your parents marry?

ES: They married in 19…. Oh my goodness me. Well, yes, it must be about 1911, you see, because I know the first child died at one month old. Alice. And with my parents we often went to the Brno cemetery and to her little gravestone, and it stands there today. I had it cleaned up, you know. It’s wonderful. And I found it without any difficulty. My parents have no grave.

RL: and how many children are there in your family?

ES: Ah. My sister Ilse, who was born 1913, and I was born in 1914, 18 months after her. That’s it, and so I grew up with my sister, we’re very close, yes.

RL: What memories do you have of childhood at home, what kind of memories do you have, what’s your earliest memory?

ES: Earliest memory. Before we lived in this house, one of the earliest memories was when I was told off as a child, for being naughty. And they said ‘you must stand in the corner as a punishment’. And I said I don’t mind, it’s nice and warm near the stove there, and they all laughed. But what is not so, what is more serious is I remember when there were, there was unrest, people in the streets, shouting. There was. I think there was either a Communist uprising, or whatever, I don’t know. It was certainly during the war, and I remember my
Problem Areas: Transcription

Focus on content

1. Mom: `huh `Well that’s something else. (0.3) `I don’t think that you should be going to the parties that Beth goes to. She is eighteen years old. An’ you are four:tee:n, da[rlin’.
2. Vir: -> A:LL THE REST OF MY: PEOPLE MY AGE ARE GWAFFS. I promise they are si:[ck.
3. Mom:-> [They’re what?
4. (.)
5. Vir: -> GWAFFS.
6. ???: ( )
7. Pr?: -> What’s a gwaff.
8. (3.1)
9. Vir: -> Gwaff is jus’ someb’dy who’s really (1.1) I just- ehh! `hh
10. -> s- immature.> You don’t wanna hang around people like that.< (1.9)
11. Mom:-> Well, don’tchyou think thet thuh: eighteen year olds, an’ thuh twenty year olds think you’re a gwaff?
12. (0.8)
13. Mom:-> Whatever a gwaff might be.
14. Pru: [ehh huh!
15. Vir: [eWell not if I date ‘em, I mean my go:osh!
16. Pru: ehh!
Problem Areas: Transcription

ORAL HISTORY
Focus on content

SOCIOLINGUISTICS
Focus on form

RL: When did your parents marry?

ES: They married in 19… Oh my goodness me. Well, yes, it must be about 1911, you see, because I know the first child died at one month old. Alice. And with my parents we often went to the Brno cemetery and to her little gravestone, and it stands there today. I had it cleaned up, you know. It’s wonderful. And I found it without any difficulty. My parents have no grave.

RL: and how many children are there in your family?

ES: Ah. My sister Ilse, who was born 1913, and I was born in 1914, 18 months after her. That’s it, and so I grew up with my sister, we’re very close, yes.

RL: What memories do you have of childhood at home, what kind of memories do you have, what’s your earliest memory?

ES: Earliest memory. Before we lived in this house, one of the earliest memories was when I was told off as a child, for being naughty. And they said ‘you must stand in the corner as a punishment’. And I said I don’t mind, it’s nice and warm near the stove there, and they all laughed. But what is not so, what is more serious is I remember when there were, there was unrest, people in the streets, shouting. There was, I think there was either a Communist uprising, or whatever, I don’t know. It was certainly during the war, and I remember my

(4) Virginia, 5:01-23
1  Mom:    hhh "Well that’s something else. (0.3) ’I don’t think that
2     you should be going to the parties that Beth goes to. She is
3     eighteen years old. An’ you are fourteen, darlin’.
4  Vir:   ->   [I KNOW:: BUT
5     ->   A: ALL THE REST OF MY: PEOPLE MY AGE ARE GWAFFS. I promise they
6     are si:cek.
7  Mom:   ->   [They’re what?
8     ()
9  Vir:   ->   GWAFFS.
10   ???:   ()
11  Pr?:   ->   What’s a gwaff.
12   (3.1)
13  Vir:   ->   Gwaff is jus’ someb’dy who’s really (1.1) I just- ehh! ’hh
14     ->   s- immature. > You don’t wanna hang around people like that.<
15   (1.9)
16  Mom:   ->   Well, don’t you think the thuh: eighteen year olds, an’ thuh
17     ->   twenty year olds think you’re a gwaff?
18   (0.8)
19  Mom:   ->   Whatever a gwaff might be,
20  Pru:   [ehh huh!
21  Vir:   [Well not if I date ’em, I mean my gyosh!
22  Pru:   ehh!
Problem Areas: Search

The current VHA thesaurus consists of approx. 55,000 keywords.
Problem Areas: Search

the current VHA thesaurus consists of approx. 55,000 keywords

for some research objectives (e.g. mine) the keywords may be too general or too specific

for other research objectives the keyword index is nearly useless (psychology of emotions, linguistics, study of metaphor etc.)

some language groups are not indexed at all (e.g. Swedish & Danish interviews), some are indexed only partially (e.g. Slovak interviews), some are indexed fully but rather sparsely (e.g. Czech interviews)

certain data were collected but are not searchable (name of the interviewer; religious identity etc.)

„index“ × „fulltext“ search
Potential improvements and other suggestions

no form of transcription will be useful to every social scientist (methodological, theoretical, disciplinary differences)

↓

ideally: allowing users to choose from different transcription „standards“
more likely: allowing users to create their own transcripts in the interface itself

tools for user keyword annotation („coding“)

incorporation of the „fulltext“ search into the VHA interface

multidisciplinary relevance ➔ discipline-structured thesaurus (reflecting different „points of view“)

Web 2.0 aspect: allowing registered users to annotate interviews, comment on the keywords, correct mistakes, discuss, share the transcripts etc.
References


Holland, N. N.: *Reading and Identity*. <http://www.clas.ufl.edu/users/nholland/rdgident.htm>


Thank you!