Department of Languages, Literatures, and Linguistics, School of Information Studies, College of Arts and Science, Syracuse University

## Syracuse University College of Arts & Sciences

### Introduction

"Tension," as "an inner striving, unrest, or imbalance often with the physiological indicators of emotion," or "a state of latent hostility or opposition between individuals or groups," always revolves around feelings and emotions and could be affected by psychological, phonological and morphological aspects. Especially in interpersonal communications, people sometimes feel uncomfortable responding to a question or discussing a topic, they may tolerate the situation, and hide their discomfort, or may use various strategies (silence or switching topics) in their responses.

Code-Switching/Code-Mixing (CS/CM) is a linguistic phenomenon, which means that mixed languages are an outcome of switching or mixing between two or more languages (codes); or alternate between at least two languages or languages varieties (Van Herk, 2012). Since bilinguals or multilinguals are tend to use different languages to express different emotions, and the feelings are always mixed (Pavlenko, 2005), therefore code-switching and language choice can influence speakers' expressions of positive and negative feelings.

In this study, we thrive on finding out how CS/CM as a linguistic phenomenon could be a sign of tension in Holocaust survivors' interviews. We created an interview corpus (a total of 39 interviews) that contains manually annotated CS/CM codes (a total of 802 quotations). And then compared our annotations with the tension places in the corpus, which are identified by a computational tool. We found that most of our annotations were captured in the tension places, and its performance implies that CS/CM can be appropriate cues for detecting tension in this communication context.

#### **Tension Detection Tool**

The tension detection tool first used a multi-channel CNN algorithm to recognize emotions in the content of an interviewee's response. Basic negative emotions such as sadness and anger could be the cues for tension. Researchers then examined additional tension cues in a negative emotion place. These cues include hedge words (e.g., probability, likely), booster words (e.g., definitely, never), silence, laughter, the length of an answer, and a list of discourse markers identified by the researchers.

The researchers' examination of these cues was based on the relevant literature. For instance, silence is a critical discursive indicator of various

RESEARCH POSTER PRESENTATION DESIGN © 2019 www.PosterPresentations.com

phenomena in oral history interviews, such as reticent responses or, in turn, unnatural pauses caused by personal trauma. The use of discourse markers such as not really, not that I remember or well, anyway in responses shows how reticence in an interview might be influential. It is also a common strategy embraced by the interviewees to avoid either complete refusal to reply or full disclosure, which can be revealed by the unusual short or long answers given the question.

#### Corpus

Our Holocaust interview corpus is obtained from the United States Holocaust Memorial Museum in Washington, D.C. It contains <u>39 interview</u> data, including the video recordings of the interviews and the transcripts of the interview content. Video and demographics information were also provided (interviewees' age was calculated according to their DOB information and the interview date).

#### Doc

RG-

#### Annotation

Document

RG- 50.030.

### **Inter-coder Reliability (ICR)**

- Two annotators (A&B)

#### **Annotation Results**

The updated annotation results show that there are 802 annotated quotations in 39 transcripts. A total of 22 languages has been detected for CS/CM, and the most frequently occurring languages are German, Polish, Hebrew, and Russian.

Corpus access Link: https://drive.google.com/drive/folders/1GE2DD5Y3KS8lfmPwSsWmYq6gfBy1dbg?usp=sharing

# Identifying Tension in Holocaust Survivors' Interview: Code-switching/Code-mixing as Cues Xinyuan Xia, Lu Xiao, Kun Yang, Yueyue Wang

#### **Methodology & Analysis**

| ument       | Gender | Age | Interview Date | Video Length |
|-------------|--------|-----|----------------|--------------|
| 50.030.0001 | Male   | 66  | 1994 July 28   | 2h58min      |

Use videos as auxiliary instruments and annotated the sentence (clauses) with CS/CM code whenever a CS/CM appeared within the sentence.

Excluded any proper nouns (e.g., names of people and places), loanwords or borrowings from CS/CM.

|       | Quotation Content  | Comment              | Codes           |
|-------|--|----------------------|-----------------|
| .0029 | The Straf company is theeverybody knows what a straf companycompany was, | Straf means punitive | CS/CM<br>German |

<u>A</u> first annotated all 39 interviews and re-checked 23; <u>B</u> annotated 16 interviews. • The ICR value between <u>A</u> itself is 0.889; the value between <u>A</u> and <u>B</u> is 0.884. • All identified discrepancies are discussed and reached agreements afterwards.

#### Analysis

CS/CM responses.

| chi- square value is<br>232.963, p <0.001 | Tension responses | Not-Tension responses |
|---|-------------------|-----------------------|
| CS/CM responses                           | 172               | 103                   |
| No-CS/CM responses                        | 923               | 3309                  |

- and 19 had over 50 percent coverage.

|                                    | -   |
|------------------------------------|-----|
| Translation                        | 186 |
| Glossary                           | 117 |
| Scene Representation               | 22  |
| Glossary & Translation             | 6   |
| Scene Representation & Translation | 11  |

- record may contribute to the detection of tensions.
- further investigation on this topic.

Müller, Beate. (2014). Translating Trauma: David Boder's 1946 Interviews with Holocaust Survivors. Translation and Literature. 23(2), 257-271. Available Online May 2014. Pavlenko, Aneta. (2005). Emotions and bilingualism. Cambridge, UK: Cambridge University Press. Van Herk, Gerard. (2012). What Is Sociolinguistics? 1st Edition. 127-146. Blackwell Publishing Ltd. https://www.merriam-webster.com/dictionary/tension Accessed on June 18th, 2021.

In total, there are 1095 tensions and 3362 not tensions in the interview corpus. Based on our annotation results, there are 275 CS/CM and 4182 no

As for the occurrences of CS/CM captured by the tool, amid 39

transcripts, 11 had all their CS/CM occurrences captured by the model,

There are also 342 CS/CM annotations in the CS/CM responses that have no tension identified by the tension detection tool and is categorized into three main cases according to our interpretation of their occurrences.

> We anticipate that in the third category the interviewees may feel tension, because prior study suggests that while reliving painful or frightening situations, a sudden switch of language at that crucial moment could be viewed as a sign of trauma (Müller, 2014).

#### **Discussion & Future Work**

Our study shows that CS/CM and tension have a strong correlation, and it has potential to be a cue in identifying tensions in Holocaust interviews.

• Also, our analysis of the CS/CM annotations not captured by the tool suggests that the detection of story structure in these communication

It would also be essential to investigate the potential false positives in the dataset to improve and perfect the tool's performance and result.

Furthermore, our annotations also included the silence in the corpus, and we make it as an open accessibility to the research community, calling for

#### **References**