Identifying Tension in Holocaust Survivors’ Interview: Code-switching/Code-mixing as Cues

Xinyuan Xia, Lu Xiao, Kun Yang, Yueyue Wang

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

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 language varieties (Van Herk, 2012). Since bilinguals or multilinguals tend to use different languages to express different emotions, and the feelings are always mixed (Pavlkenko, 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., place. These cues include hedge words (e.g., booster words (e.g., likely), booster

Methodology & Analysis

Corpus

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

Annotation

• 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), loanswords or borrowings from CS/CM.

Inter-coder Reliability (ICR)

Two annotators (A&B) had annotated all 39 interviews and re-checked 23; B annotated 16 interviews.

• The ICR value between A&B itself is 0.88; the value between A and B is 0.884.
• All identified discrepancies are discussed and reached agreements afterwards.

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.

Analysis

• 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 CS/CM responses.

<table>
<thead>
<tr>
<th>chi square value</th>
<th>Tension responses</th>
<th>Not-Tension responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>232.963, p &lt;0.001</td>
<td>172</td>
<td>103</td>
</tr>
</tbody>
</table>

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

References