Using Text Analysis Tools to Understand Autobiographical Memory Retrieval

[1]Shirley Zhang, Jay Sheldon, [2]Matthew Grilli Ph.D, [1]Signy Sheldon Ph.D
[1]McGill University (Department of Psychology), [2]University of Arizona (Department of Psychology)

Introduction

Autobiographical Memory and Aging

- Autobiographical memory is the memory of an individual's past and personal events
- Some studies found aging may lead to changes in the composition of autobiographical memories, with older adults utilizing more semantic details[1]

The "Autobiographical Interview" (AI)[1]

- A widely used tool to measure the characteristics an autobiographical memory
- Involves interviewing individuals about past life events, and annotating the transcripts for episodic and semantic components

Figure 1 Example of AI detail annotation[1]

Limitations

- Annotating transcripts is time consuming
- Annotators must go through significant training
- Human error and subjectivity can occur

Aims

1. To see whether a text analysis tool could demonstrate age differences between younger and older adults.

2. To see whether the memories of older adults may be more similar to each other (based on previous evidence of higher semantic content in older adult memories).

References
[1]Levine, Brian, et al. (2002) Aging and Autobiographical Memory: Dissociating Episodic from Semantic Retrieval. Psychology and Aging, vol. 17, no. 4, pp. 677–689., https://doi.org/10.1037/0882-7974.17.4.677.

Methods

1. Dataset

382 participants, 218 younger (18-34 years) and 164 older (58-92 years).

- Younger adult memories: childhood, teenage, early adulthood
- Older adult memories: childhood, teenage, early adulthood, middle adulthood, late adulthood

2. Text Preprocessing

- Words were tagged as a parts-of-speech (POS) (i.e noun, verb, adjective, adverb)
- Stop words (i.e "a", "in", "the"), punctuation, and contractions were removed

3. Implementing Text Analysis Tool

Latent Dirichlet Allocation (LDA)

- Used to identify hidden topics within a collection of documents
- Output includes scores for each transcript, describing how strongly different topics are associated with the transcript
- Coherence score: higher scores indicate the model found topics close to what may be found by humans

Most recent memory transcript for younger and older adults were used ("early adulthood" and "late adulthood" respectively). Only nouns were included.

4. Analysis of Variance

A repeated measures ANOVA was run with the topic scores for all subjects, with topic # as a repeated effect and age group as a between subjects effect.

Results

LDA model with 5 topics had the highest coherence score

- Coherence score: 0.332

Analysis of Variance

- No main effect of age group ($p=0.559$)
- Significant interaction effect of topic and age group ($p<0.001$)
- Post hoc comparisons showed significant effects of age for:
  - Topic 2 ($p<0.001$), topic 4 ($p<0.001$), and topic 5 ($p=0.028$)

Results (continued)

Topic	Keywords
Topic 1	**house**, friend, day, time, **brother**, **birthday**, **party**, thing, **okay**, **family**
Topic 2	year, people, thing, time, day, lot, **daughter**, **event**, **wife**, **way**
Topic 3	**car**, kind, lot, time, **water**, thing, day, **home**, **something**, **night**
Topic 4	day, **school**, time, **class**, **dad**, **line**, kind, **picture**, people, **mom**
Topic 5	**firstname**, people, kind, **guy**, friend, time, **name**, **bus**, year, thing

Table 1 Top 10 associated keywords for each topic

Keywords per Topic

- Topic 2 appears to capture family dynamics
- Topic 4 appears to capture school-related concepts

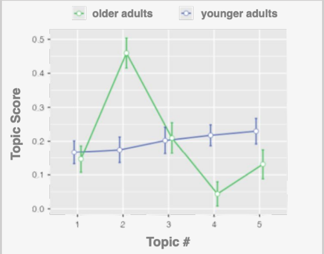

Figure 2 Estimated marginal means for repeated measures ANOVA with all subjects

Aim 1

No main effect of age, although interaction effects show differences for topic 2, 4, and 5.

Aim 2

Older adult memories more likely to be represented by topic 2, whereas younger adult memories are more spread out.

Text analysis tools, such as LDA, may be used to automatically find similarities and differences in autobiographical memory transcripts. This may help speed up research, increase the amount of data able to be analyzed at once, and offer a new perspective.

Future Directions

- Match memory transcripts by life period
- Include all parts-of-speech and more life periods

Acknowledgements

Many thanks to Jay Sheldon, Dr. Matthew Grilli, and Dr. Signy Sheldon for their insightful guidance, continued support, and supervision in this project. Thank you as well to the Office of Science Education at McGill for this opportunity.