

対照群付き高齢者コーパスについて

2018/08/07

ソーシャル・コンピューティング研究室

1. 概要

若年対照群付き高齢者コーパスは高齢者群（60～79 歳）60 名と対照群である非高齢者群（20～59 歳）20 名の合計 80 名の実験参加者に対し、インタビューの回答を収集したデータセットである。回答は主に音声データ、一部の質問については音声の書き起こしがテキストファイルとして収録されている。

2. インタビュー・実験内容

80 名の参加者に下記のインタビューを実施し、インタビューの回答を収集した。回答結果は音声ファイルとして収集した。詳細は以下の通りである。

・EP（自然文課題）

参加者は「最近あった楽しい出来事を教えてください」など 10 個の質問について、30 秒から 3 分で回答した。質問の一覧は以下の通り。各質問 ID 末尾のアルファベットは質問の種類を表し、a は主な質問、b は a についていつ起こった出来事かを問う補助的な質問であることを表す。なお、b は EP1, EP6 のみに設定されている。

EP1a: 最近あった悲しい出来事を教えてください

EP1b: それはいつのことですか？

EP2a: 最近不安を覚えた出来事を教えてください

EP3a: 最近怒りを覚えた出来事を教えてください

EP4a: 最近嫌悪感を覚えた出来事を教えてください

EP5a: 最近驚いた出来事を教えてください

EP6a: 最近あった楽しい出来事を教えてください

EP6b: それはいつのことですか？

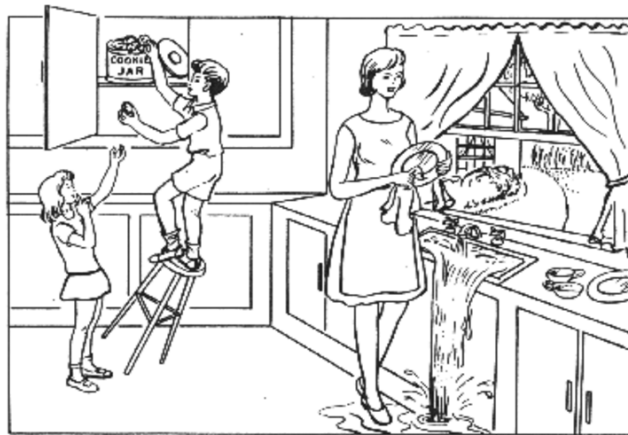
EP7a: あなたの尊敬する人について教えてください

EP8a: あなたが最近夢中になっていることについて教えてください

・ ST (story, イラスト描写課題)

参加者は呈示されたイラストについての「下の絵を見て、何が起きているか説明して下さい」という質問に1分程度で回答した。

提示したイラストは以下。



(Goodglass, H. and Kaplan, E. (1972). “The Assessment of Aphasia and Related Disorders”. Lea & Febiger, Philadelphia.)

・ AM (anime, アニメ描写課題)

参加者は呈示された短いアニメについての「これから再生する動画を見て、起こったことを説明してください」という質問に1分程度で回答した。

呈示したアニメは以下から視聴できる。

<https://www.youtube.com/watch?v=dL-AzBDoFaw>

3. データについて

本データセットのフォルダ構成は以下の通りである。

list: 参加者プロフィールリスト (ID、課題別のファイル一覧を記載)

実験参加者の参加者ID、課題内容、対応するファイル名、音声ファイル長のリストを示した「音声ファイルリスト」と、参加者ID、性別、年代の表である「参加者プロフィール」が含まれている。両リストとも、実験の不備により削除

したデータがあるため、ID が連番になっていない部分がある（他 2 フォルダのデータについても同様）。

voice: 80 名分の音声ファイル

本フォルダには、被験者 ID ごとに EP, ST, AM 各課題の音声ファイル(1 フォルダあたり 12 ファイル)が格納されている。

transcript: 書き起こし

voice フォルダの音声ファイルのうち、EP1a, EP6a, ST, AM についての書き起こしが.docx 形式で収録されている。EP1a, EP6a, ST については言い淀みなどを書き起こさない要約的な書き起こし、AM については言い淀みなどを含めた直接的な書き起こし方式を取っており、テキスト形式で収録されている。

AM については、以下の方針に従って書き起こしを行っている。

- ・フィラー：「あー」「えっと」等、書き起こせるものは書き起こす。
- ・沈黙：会話に沈黙が合った場合、2-3 秒は「…」3 秒以上は「……」とする。
- ・数字，文字：共に全角で表記する。
- ・音声の主，笑い声，聴きとりにくい声，個人情報には以下のタグの付与や加工を行う。

<i>	インタビュアー音声
<e>	回答者音声
<l str=""/>	笑い声
<u str=""/>	発音不明瞭
**	個人情報である氏名，年齢，住所，電話番号，店舗名， 地名(人口 2 万人以下の場合)