

国際的な失語症患者の発話コーパス <http://aphasia.talkbank.org>

課題 (AphasiaBank protocol, Famous People protocol, 他) のトランスクリプト (文法などコーディング済み) ・ビデオ (トランスクリプトとリンク)

⇒ インターネットを通し無料で利用可能 (個人情報 はパスワード管理) オリジナルなデータを作成することも可能

現在英語、独語、広東語など数百人の失語症患者・対照群のデータあり

(日本語データは健常高齢者のみ掲載; 失語症のある人のデータはなし)

コーパス(Corpus)とは

自然言語 (日本語、英語などの書き言葉や話し言葉) をテキスト化し、大量に集めたもののこと。

データに品詞や統語構造などを付与してあるものは「タグ付きコーパス」とよばれコンピュータで解析可能となる。

【コーパスを用いた分析】 例; 頻度分析/コロケーション (共起関係) 分析/用例分析 (特定のパターンを含んだ用例を抽出し、その意味を解釈する)

失語症臨床においては発話を録音→書き起こして分析することは頻繁に行われているが、つどデータを集めてトランスクリプトを作成するのは非常に手間がかかる。コーパスの使用により、多量のデータの中から障害タイプや重症度を選択し、目的とする言語症状を即座に分析できる。

データとその分析システム *AphasiaBank の HP から DL 可

1) データ; 発話のトランスクリプト (文法などコーディング済み)

ビデオ、検査データ (視・聴覚、言語 (呼称、復唱) 等)

2) 表記フォーマット; CHAT MOR (形態素解析プログラム)

TalkBank の形式 (CHILDES などと共通)

失語症に特徴的な誤りも単語・発話レベルでコード

[* p]音韻性錯語 [* n]新造語 [+ jar]ジャルゴン 等

3) 言語解析プログラム; CLAN

FREQ: 特定の単語を出力

KWAL: 特定の単語を含む発話のリストを出力

MLU: 発話数・単語数を出力 ...その他いろいろ

CHAT データファイルの構造

PAR*:otoosan wa tamago o tabeta

* **メインライン** 発話本文

*PAR: 研究協力者の発話

%mor: n|otoo-san=father ptl:foc|wa=TOP n|tamago=egg

ptl:case|o=ACC v:v|tabe-PAST=eat .

%mor: **形態素タグ** 品詞分析

品詞 | 語幹-語尾=英訳

%gra: 1|2|SUBJ 2|5|CASP 3|4|OBJ 4|5|CASP

5|0|ROOT 6|5|PUNCT

%gra: **構文解析タグ** 統語の依存関係

文中の語順|依存先の番号|文法的役割

%ort: お父さんはたまごを食べた.

%ort: **仮名漢字表記**

実施のための材料

* word、excel などのファイルで DL 可

・研究者の発話内容 (台本) ・提示する情景画

・アンケート (利き手、既往歴など)

・失語・知能・気分等の検査類

・研究同意書 ・個人情報入力フォーム

・データベース作成マニュアル など

実施方法

・研究者—参加者の 1 対 1 の対話をビデオ撮影 (参加者の上半身と研究者の後ろ姿が写るアングル)

・研究者が決められた質問をし、談話を収集

課題の流れ (AphasiaBank Protocol)

I. 自発話サンプル

A. 脳卒中の発症と対処

発症時の状況、回復・対処について (非失語対照群は病気・けがの経験)

B. 重要な出来事

人生における重要な出来事

II. 情景画説明 1 枚または複数の絵を提示して、ストーリーを説明してもらう

A. 割れた窓-4 枚の絵の説明

B. 断った傘-6 枚の絵

C. 猫の救出-1 枚の絵

* ほかにも **Famous People protocol**

(有名人の写真を見せて質問する) もあり

III. 物語説明 文字を隠した「シンデレラ」の絵本を見た後、ストーリーを説明

IV. 手続き説明 ピーナツバターとジャムサンドイッチの作り方の語り

日本版プロジェクト

日本語データ掲載に向けて

AphasiaBank 直訳版 (飯塚、2013) 英語版からの直訳 健常高齢者 2 名

AphasiaBank 日本バージョン (飯塚、2014) 一部課題を改変 健常高齢者 2 名

トランスクリプト (コーディング済み)、ビデオが AphasiaBank に掲載

今後の展望

被験者数を増やす/翻訳・課題内容の検討

We are AphasiaBankers!

* 予稿集では幼児の発話コーパスが CHILDS となっていますが、正しくは CHILDES です。お詫びして訂正します。