

背景と目的

- 日本語指導が必要な生徒の増加
- 教材開発の必要性が増すのでは？

教材開発の基礎となる語彙のリストが必要

NHK Eテレの教育番組の音声を対象とした語彙調査 (浅井ら, 2020)

仮説「教育番組はその対象学年の日本語母語話者が理解できる単語や表現で構成されている」

NHK Eテレの低学年向け教育番組の音声を書き起こし、生徒にとって重要な単語を抽出する

方法

- 2020年度上半期にNHK for School (www.nhk.or.jp/school/) に公開されていた小学3年生以下を対象とした15番組を書き起こし
- 表記ゆれを修正
- MeCab+UniDic (伝ら, 2007) にて形態素解析
- 記号, 助詞, 助動詞, 接頭辞, 接尾辞, 固有名詞等を除去
- 有用度指標に基づき単語をランキング

表1 対象の番組とその話数, 1話あたりの放送時間

番組名	話数	時間(分)
えいごでがんこちゃん	20	5
おはなしのくに	17	10
おばけの学校たんけんだん	16	10
銀河銭湯パンダくん	40	10
ことぼドリル	20	10
こどもにんぎょう劇場セレクション	10	15
コノマチ☆リサーチ	20	10
さんすう犬ワン	19	10
新ざわざわ森のがんこちゃん	30	10
しぜんとあそぼセレクション	25	15
すたあと	20	5
で~きた	20	10
ふしぎエンドレス理科3年	20	10
理科3年ふしぎだいすき	19	15
ふしぎがいっぱい3年	20	10

有用度指標A

当該単語が出現した番組数を考慮 (浅井ら, 2020)

$$u_i = \frac{d_i}{15} \times \log f_i$$

d_i : 単語*i*が出現した番組数
 f_i : 単語*i*が出現した回数

有用度指標B

番組ごとの時間長の違いを考慮

$$u_i = \sum_{j=1}^{15} \frac{f_{ij}}{t_j}$$

f_{ij} : 単語*i*が番組*j*に出現した回数
 t_j : 番組*j*の時間長

51時間20分

浅井らの約5倍

結果と考察

- 総のべ語数: 108,672語 → 浅井らの5倍以上
- 総異なり語数: 6,172語 → 浅井らの1.6倍以上
- テキストカバー率: 上位2700語で95%

有用度指標Aに基づく語彙表の上位20語

いる, なる, できる, 言う, こと, よい, ない, どう, ある, そう, 行く, する, やる, 何, 皆, 見る, もう, はい, うん, ちょっと

有用度指標Bに基づく語彙表の上位20語

こと, 何, 来る, なる, そう, ため, できる, いる, 言う, どう, 良い, ある, ない, 行く, やる, うん, なる, 皆, もの, 見る

いずれも上位に基本的な単語を抽出できているが, 全体を見ると科目ごとの番組数の差が影響している

まとめ

NHK Eテレの小3年生以下を対象とした教育番組の音声に用いられる単語を抽出し, 有用度指標を用いてランキングした。

付録(1) 語彙リスト全体

Webページにて語彙リスト全体を公開しています。右のQRコードを読み込んでアクセスしてください。



付録(2) 単語レベル判定機能付きエディタ

リーディング・チュウ太のレベルチェッカーの機能を持つエディタを試験公開しています。スマホでもPCでも使えます！

