

# 視覚障がい者による仮名漢字変換のための 説明語辞書の作成支援に関する検討

杉浦 知也<sup>†</sup> 西田 昌史<sup>†</sup> 綱川 隆司<sup>†</sup> 西村 雅史<sup>†</sup> 渡辺 哲也<sup>‡</sup>  
静岡大学大学院<sup>†</sup> 新潟大学<sup>‡</sup>

## 1. はじめに

視覚障がい者がパソコンを利用する際は、晴眼者と同じように目で見て操作することは出来ないため、スクリーンリーダーという音声読み上げソフトを用いて操作する。パソコンで文章を入力し変換する場合は、スクリーンリーダーが読み上げた漢字の説明を聞くことで漢字を判断する。しかし、現在説明に広く使われている詳細読みという方法は、渡辺ら[1]により、想起しづらい単語があるという欠点を指摘されている。

この詳細読みの欠点を改善するため、単語変換について西田ら[2]により意味情報による仮名漢字変換方法が提案されている。これは単語の意味を類義語などの意味的に近い言葉で説明する手法であり、詳細読みと比べて使用者の負担が少なく、より速く変換出来る。

この意味的に近い言葉を説明語といい、意味情報による仮名漢字変換を行うためには単語ごとの説明語をまとめた説明語辞書が必要だが、これを作成する時は当該単語を連想しやすい説明語を手探りで探す必要があり、その負担が大きい。そこで、本研究では単語同士の意味の近さを数値化できる Word2Vec や、同義語や上位語など、単語の意味の情報を表す概念辞書の WordNet により、説明語を選ぶ際に用いる情報の追加や整理を行うことで、意味情報による説明語の辞書の作成を支援する手法を提案する。

## 2. 説明語辞書の作成方法

本研究では、意味情報の説明語の作成は ロゴヴィスタ株式会社の日本語大シソーラス - 類語検索大辞典 - を用いて、単語を条件検索して出てきた単語の中から人手で単語を連想しやすいと思った物を選ぶという方法を採用している。

単体の辞書だけではふさわしい単語が見つからない場合にも対応するため、WordNet の同義語と上位語、Word2Vec で算出した類似度の高い単語を候補に追加する手法を提案する。また、

Support Method for Developing Associative Word Dictionary of Kana-Kanji Converter for Visually Impaired People  
Sugiura Tomoya<sup>†</sup> Nishida Masafumi<sup>†</sup> Tsunakawa Takashi<sup>†</sup>  
Nishimura Masafumi<sup>†</sup> Watanabe Tetsuya<sup>‡</sup>  
Shizuoka University<sup>†</sup> Niigata University<sup>‡</sup>

説明語を探す手助けとして、Word2Vec で単語と候補の説明語との類似度を測り、その数値の高い順にソートして説明語を並べることで探しやすくする手法も合わせて提案する。

## 3. 実験

### 3.1. 実験内容

大学生 12 名を対象に説明語辞書を作成する実験を行った。漢字 2 文字の単語 20 個を問題の単語とし、用意した説明語のリストの中から問題の単語を連想しやすいものを選んでもらう。

日本語大シソーラスから取り出した単語を説明語 A、WordNet で出した同義語と上位語の中から説明語 A と被った単語を抜いたものを説明語 B、Word2Vec で出した上位 50 位までの類似度の高い単語の中から説明語 A、B と被った単語を抜いたものを説明語 C とする。Word2Vec のモデルは東北大学の乾・岡崎研究室が作成した日本語 Wikipedia エンティティベクトルのモデル[3]を用いた。

手法①は説明語 A のみ、手法②は説明語 A, B, C を全て、手法③は説明語 A, B, C を Word2Vec によってソートしたものを使う。これらを説明語リストとして問題の単語ごとにまとめた。また、手法②と手法③には、問題の単語について WordNet の定義文を意味例として付け、近い意味の単語を探す際の参考にしてもらった。手法③の説明語リストの例を以下の図 1 に記す。

帰還	
意味例	以前いた場所、状況または活動へ向かうまたは戻る。家に来る、または戻ること。相手の選手へボールを打ち返すテニスのストローク。
説明語 A	生還、帰る、帰念、舞い戻る、復帰、戻る、戻り、帰り、還御(かんぎょ)、還幸(かんこう)、還る、去る、引き取る、回帰、還り、立ち戻る、立ち返る、帰来(きらい)、立ち返り、たぐいま、再帰、引く、還り立つ、帰り立つ、帰り詣(もろず)、帰途につく、お暇(いとま)する、帰(かえり)け、帰って行く、立ち帰る、帰路につく、帰る空、還啓、後にする、帰去来今(かえりなんいざ)、回帰(かいらん)、帰期(きき)、廻歸(かいらん)、帰参(かかなう)、帰り預(わすら)う、走還、駆け戻る、馳せ帰る、飛んで帰る、早帰(そうき)、早帰り。
説明語 B	到着、帰国、帰郷、到達、来着、帰宅、帰省、復讐、返還、返却、参着、行く、復す、参る、復する、再発、着、再現、動く、リターン、帰って来る、還える、復讐する、復讐する、帰還する、還参する、還御する、舞いもどる、舞戻る。
説明語 C	脱出、逃亡、出撃、出立、帰港、入城、引き返し、退還、出航、帰投、出帆、退却、投降、不時着、上陸、入港、逃走、降伏、移乗、進発、敗走、突っ、来襲、撤収、船出、再訪、旅立つ、発つ、一時帰国、亡命、渡海、進攻、進軍、出発、帰任、帰京、帰艦、出征、降艦、出陣、復讐、来襲、進駐、退還、出港、退去、逃げ帰る。

図 1: 説明語リスト例

例題を行った後、本番の 20 個の問題の単語について説明語を 5 つずつ選んでもらい、その時間を測った。手法ごとにこれを繰り返し、実験終了後にどの手法がやりやすかったかの順位付けやその理由などについてアンケートを取った。手法の順番は被験者ごとに変更した。

### 3.2. 実験結果と考察

20 個の単語に説明語を選択するまでにかかった作成時間を以下の表 1 に記載する。

表 1: 手法ごとの作成時間

被験者	実施の順番	手法①	手法②	手法③
A	①→②→③	47分50秒	57分20秒	20分43秒
B	②→③→①	16分31秒	20分57秒	19分22秒
C	③→①→②	18分09秒	20分19秒	26分28秒
D	①→②→③	21分15秒	17分36秒	12分24秒
G	②→③→①	21分32秒	30分20秒	20分46秒
H	③→①→②	16分25秒	15分43秒	16分04秒
I	①→③→②	50分02秒	26分15秒	33分27秒
J	②→①→③	17分41秒	26分59秒	15分56秒
K	③→②→①	15分23秒	12分43秒	23分00秒
L	①→③→②	30分37秒	25分35秒	28分45秒
M	②→①→③	19分38秒	20分26秒	14分31秒
N	③→②→①	19分34秒	24分13秒	31分28秒
平均		24分33秒	24分52秒	21分55秒

平均時間を見ると、手法①と手法②は 20 秒ほどとあまり違いは無いが、手法③が 2 分半ほど速くなっている。

また、どの手法が説明語を選択しやすかったかの順位付けについて、手法ごとに何位がどれだけ選ばれたかについての割合のグラフを以下の図 2 に記載する。

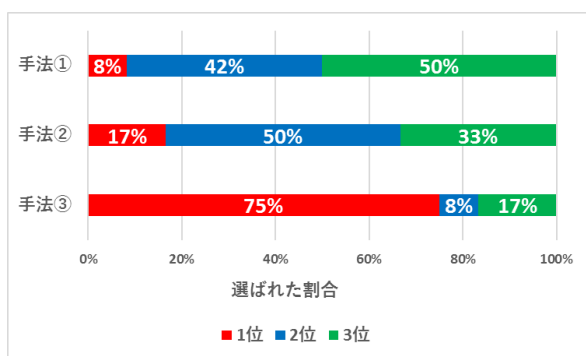


図 2: 説明語の選択しやすさ

グラフから、手法③が 1 位に選ばれることが多く、手法①と手法②はあまり差が無いということが分かった。順位付けの理由を聞いた所、

ソートのおかげという声が多く、ソートが説明語の選択の時に役立っていることが分かった。

手法①と手法②は回答時間も作りやすさ順位もあまり差は無かった。低評価にした理由として、手法①は情報量が少ない、手法②は情報が多すぎて選びにくいという意見が多く、また説明語の候補数が少ない単語では手法②が良かったが候補数が多い単語では手法①がやりやすかったという意見もあり、この 2 つは単語によって一長一短となっていたと考えられる。

選択された説明語を見てみると、手法②では説明語 A, B, C がそれぞれ 66.9%, 26.4%, 6.7% ずつ選択されており、手法③では 74.3%, 20.5%, 5.2% ずつ選択されていた。手法②よりも手法③の方が説明語 A から多く選ばれているが、これはソートにより一番上にある説明語 A から見つけやすくなったためだと考えられる。手法②と手法③共に説明語 C はほとんど選ばれていなかった。

これらの結果から、WordNet や Word2Vec の情報を追加するだけでは全体的な説明語の選択しやすさはあまり変わらないが、Word2Vec の類似度でソートを行えば従来手法より説明語辞書を作成しやすくなることが分かった。

### 4. おわりに

視覚障がい者の仮名漢字変換に使用する意味情報の説明語辞書の作成を支援するため、WordNet と Word2Vec を使って説明語の候補を追加する手法と、候補を Word2Vec で算出した類似度でソートする手法について提案した。

実験の結果、ソートをした手法では作成にかかる時間が短くなり、他の手法よりも作成しやすいことから、この手法に説明語作成の負担を減らす効果があることが分かった。

今後の課題として、より多くの漢字を対象にした実験や、選択された説明語がどれだけ漢字を連想しやすいかについての分析、説明語の数を調整する方法についての検討などを行ってきたい。

#### 参考文献

- [1] 渡辺 哲也, 渡辺 文治, 藤沼 輝好, 大杉 成喜, 澤田 真弓, 鎌田 一雄: スクリーンリーダーの詳細読みの理解に影響する要因の検討-構成の分類と児童を対象とした漢字想起実験-, 電子情報通信学会論文誌, D-I Vol. J88-D-I, No.4, pp.881-899, 2005.
- [2] 西田 昌史, 堀内 靖雄, 黒岩 眞吾, 市川 薫: 視覚障害者支援のための意味情報に基づく仮名漢字変換, 電子情報通信学会論文誌, Vol. J95-D, No.4, pp.960-968, 2012.
- [3] 「日本語 Wikipedia エンティティベクトル」, <[http://www.cf.ecei.tohoku.ac.jp/~m-suzuki/jawiki\\_vector/](http://www.cf.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/)> 2018 年 11 月 25 日アクセス