

視覚障がい者用仮名漢字変換プログラムのための 説明語辞書作成支援システム

杉浦知也^{†1} 西田昌史^{†1} 綱川隆司^{†1} 西村雅史^{†1} 渡辺哲也^{†2}

概要： 本研究では、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を支援するシステムとして、単語同士の意味の近さを数値化して計算できる Word2Vec や、同義語や上位語など、単語同士の関係性の情報を持っている概念辞書の WordNet を用いる手法を提案する。評価実験として説明語辞書の作成速度と作成のしやすさ、作成された説明語が変換対象の漢字をどれくらい連想しやすいか、変換対象の漢字と説明語に対しての単語親密度の評価を行った結果、提案手法は従来手法と比べて速く説明語辞書を作成でき、実験後アンケートでは提案手法の方が作りやすく感じられたという結果が得られた。また、提案手法により作成された説明語の連想しやすさは従来手法よりも高かった。さらに、説明語の連想しやすさと親密度の関係を明らかにした。

キーワード： 意味情報, 説明語辞書, Word2Vec, WordNet, 視覚障がい者, 仮名漢字変換

Support System for Developing Associative Word Dictionary of Kana-Kanji Conversion Program for Visually Impaired People

TOMOYA SUGIURA^{†1} MASAFUMI NISHIDA^{†1}
TAKASHI TSUNAKAWA^{†1} MASAFUMI NISHIMURA^{†1}
TETSUYA WATANABE^{†2}

Abstract: In this study, we propose a method using Word2Vec and WordNet for building a system to support development of an associative word dictionary for Kana-Kanji conversion designed for visually impaired people. Word2Vec can quantify semantic distance of words and WordNet is a conceptual dictionary with information on a relation of words. We conducted experiments to examine time and ease of developing the associative word dictionary, how easy it is to associate kanji with the associative words, and word familiarity of kanji and the associative words. As a result, the proposed method is faster and easier to develop the associative word dictionary compared to a conventional method. Moreover, an associativity of associative words created with the proposed method was higher than that of the conventional method. Furthermore, we demonstrated a relation between the ease of association of associative words and word familiarity.

Keywords: semantic information, associative word dictionary, Word2Vec, WordNet, visually impaired people, kana-kanji conversion

1. はじめに

視覚障がい者の方がパソコンを利用する際は、晴眼者と同じように画面を目で見て操作することは難しいため、スクリーンリーダという音声読み上げソフトを用いて文章などの画面の情報を音声で聞く事によって把握して操作している。また、パソコンで文章を入力し仮名漢字変換をする場合は、スクリーンリーダが読み上げた変換候補の漢字の説明を聞くことで今どの漢字に変換しようとしているのかを判断し、仮名漢字変換を行っている。この漢字の説明方法として、現在は詳細読みという方法が広く使われている。詳細読みは漢字一文字一文字に対してその漢字を含む別の単語を使って説明をする方法であり、例として情報という単語を変換する際は「情報のジョウ、なさけ、報告のホウ、

むくいる」といったように説明を行う。しかし、渡辺ら[1][2]により、詳細読みでは想起しにくい単語があるという欠点指摘されている。

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[3]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。これは漢字単体ではなく1つの単語全体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、例として情報という単語なら「インフォメーション、お知らせ」といったように説明する。この意味情報による変換は詳細読みと比べてより直感的に判断できるため使用者の負担が少なく、より短い時間で変換出来る上、詳細読みよりも正しい漢字を選択しやすくなっている。

この意味情報による仮名漢字変換で用いる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どの単語をどの説明語で説明するのかという設定をまとめた説明語辞書が必要不可欠である。しかし、現在この説明語辞書を作成する時は辞書に基づいて当該単語を連想しやすい説明語を全て人手で探す必要があり、その

^{†1} 静岡大学
Shizuoka University

^{†2} 新潟大学
Niigata University

負担がとても大きい。詳細読みに関しては川崎ら[4]による詳細読みの自動生成など、説明用の言葉の準備の手間を省くための研究も行われているが、意味情報による変換についての研究はまだ存在しない。そこで、Word2Vec や WordNet を使用し、説明語を選ぶ際に用いる情報の追加や説明語の提示方法の工夫により、意味情報による仮名漢字変換に必須な説明語辞書の作成を支援する手法を提案する。

提案手法の効果を調べるため、説明語の選択実験や、それによって選ばれた説明語の連想しやすさ、親密度の評価実験を行う。

2. 提案手法

本研究では、意味情報の説明語の作成に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、単語について条件検索して出てきた単語から人手でその単語を連想しやすい物を探して選ぶという方法を採用している。

しかし、この手法では単一の辞書しか使用していないため説明語にふさわしい単語が見つからない場合があるという問題点が存在する。また、検索して出てきた単語は全て同列に並んでいるため、その中でどれがより連想しやすい説明語なのかについて、作成者が自分の感覚だけを頼りに一つ一つ判断しなければならないため時間がかかり、作成の負担がとても大きくなってしまいう問題点が存在する。

そこで、これらの問題点を解決するため、日本語版の WordNet と Word2Vec を使用する。

日本語版の WordNet は国立研究開発法人情報通信研究機構[5]によって開発されている意味辞書である。特徴として語を類義関係のセットでグループ化しており、同義語や上位語、下位語、関連、非構成要素など、単語同士の多様な関係性を持ってグループ同士が繋がっている事が挙げられる。これを用いることによって、日本語大シソーラスだけでは説明語にふさわしい単語が見つからない場合でも対応することが出来る。今回の研究では、同義語と上位語のデータを利用した。同義語や上位語など、関係ごとに区分けされているためより近いものを候補として使える所から WordNet を選んだ。

Word2Vec は Mikolov[6]によって提案されており、単語の意味をベクトルで表現する手法である。ベクトル表現化することによって、単語の意味を演算処理出来るようになり、単語同士の意味の近さを計算することが可能になっている。これを用いることによって、説明語を付与したい単語と説明語候補の単語との間の意味的な近さを数値で出し、その数値を元により意味的に近く連想しやすい単語を優先的に表示することが出来るようになる。また、説明語を作成したい単語と近い類似度の数値の単語を収集することで、

WordNet と同じように他の辞書に載っていない単語を説明語候補に追加し、候補の数を増やす目的でも使用できる。

よって本研究では、説明語の候補を増やすことで候補の中に説明語にふさわしい単語が存在しない場面を減らすため、WordNet の同義語と上位語、Word2Vec で算出した類似度の高い単語を説明語候補に追加する手法を提案する。また、説明語を探す手助けとして、Word2Vec を用いて説明語を付けたい単語と説明語候補の単語との間の類似度を測り、その数値の高い順にソートして説明語候補を並べることでより類似度の高い単語を探しやすくする手法も合わせて提案する。

3. 説明語選択実験

3.1 実験内容

提案手法によって説明語辞書の作成の負担を減らせるか調べるため、大学生 12 名を対象に説明語辞書の作成実験を行った。

帰還、発行などの漢字 2 文字の単語 20 個を問題の単語とし、問題の単語ごとに説明語の候補をまとめたリストを用意した。その中から問題の単語を連想しやすいと思った説明語を単語 1 つにつき 5 つずつ選んでもらう。20 個分の説明語辞書の作成にかかった時間と作りやすさについて評価する。

説明語候補の種類として、日本語大シソーラスから条件検索で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 で出した問題の単語との類似度が高い上位 100 位までの単語の中から説明語 A、説明語 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[7]を用いた。説明語 A は平均で 150 単語、説明語 B は平均 70 単語、説明語 C は平均 60 単語程度の候補数となった。説明語 B、説明語 C は説明語 A だけではふさわしい単語が見つからない場合にそれらを見てそこから探せるようにする事を目的として追加しており、説明語 A に無い単語を探す際に邪魔になる事を避けるために説明語 A と被っている単語は削除している。

手法①は従来手法として、説明語 A のみを使用して説明語辞書の作成をしてもらう。手法②は説明語候補を追加する提案手法として、説明語 A、説明語 B、説明語 C を全て使用して作成してもらう。手法③はソートで探しやすくする提案手法として、説明語 A、説明語 B、説明語 C をそれぞれ Word2Vec の数値を用いて類似度の高い順に並べたものを使用して作成してもらう。また、WordNet を使用する手法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、説明語を探す際の参考にしてもらった。これらの説明語候補を説明語リストとして問題の単

語ごとにまとめ、それを見て説明語を選択してもらった。

問題の単語は手法ごとに変えて実施した。手法③の説明語リストの例として、問題の単語が「帰還」の場合を以下に示す。説明語候補は一部省略した。

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, 帰参, 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), ……	
説明語 B:	
到着, 帰国, 帰郷, 到達, 来着, 帰宅, 帰省, 復職, 返還, 返却, 参着, 行く, 復す, 参る, 復する, 再発, 着, ……	
説明語 C:	
脱出, 逃亡, 出撃, 出立, 帰港, 入城, 引き返し, 退避, 出航, 帰投, 出帆, 退却, 投降, 不時着, 上陸, 入港, ……	

上記のように説明語リストを作成した。手法③の説明語リストの要素は上から順に問題の単語,意味例,説明語 A,説明語 B,説明語 C と並んでいる。手法②は手法③と同じ要素を用いているが,手法①は意味例,説明語 B,説明語 C が無く,問題の単語と説明語 A だけを用いている。手法③はソートをした手法のため,候補は Word2Vec で出した類似度順に並んでいる。手法①と手法②ではソートされていないため説明語 A と説明語 B はそれぞれ検索の出力順で候補が並んでいる。説明語 C は検索の時点でスコア順に並んで出力されるため手法②と手法③で一致している。

手法②と手法③については説明語 A,説明語 B,説明語 C と説明語候補の種類が複数あるが選択する説明語は全ての候補から均等に選ぶ必要は無く,ふさわしい単語が無いと思ったなら全く使わない候補があっても良い事を伝えた。

実験内容や手法の内容について説明をした後,例題として 3 つの単語に対して説明語辞書を作成し,やり方を把握した事を確認した。その後,本番として 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,意味例や説明語 A,説明語 B,説明語 C に関してどう思ったかなどについてアンケートを取った。慣れの影響を減らすため,実施する手法の順番は被験者ごとに変更した。また,説明語を探す際に候補の単語や問題の単語で分からないものがあった場合は goo 国語辞書で調べて良いとした。

3.2 実験結果と考察

手法ごとに 20 個の単語の説明語を選択するまでにかかった辞書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	32分25秒	35分46秒	23分07秒
B	②→③→①	11分06秒	13分32秒	10分53秒
C	③→①→②	21分20秒	21分36秒	20分28秒
D	①→②→③	29分46秒	25分51秒	16分28秒
G	②→③→①	21分20秒	21分36秒	20分28秒
H	③→①→②	21分30秒	22分16秒	27分17秒
I	①→②→③	31分46秒	21分01秒	24分16秒
J	②→③→①	26分24秒	35分24秒	19分28秒
K	③→①→②	38分20秒	43分27秒	37分31秒
L	①→②→③	31分07秒	30分05秒	19分20秒
M	②→③→①	23分36秒	29分29秒	21分13秒
N	③→①→②	19分54秒	20分03秒	16分30秒
平均		25分43秒	26分40秒	21分25秒

手法ごとの平均時間を見ると,短い順に手法③,手法①,手法②となっており,手法①が手法②よりも 1 分ほど,手法③は手法①よりも 4 分 20 秒ほど 20 個の単語の説明語の選択速度が速くなっている事が分かる。また,手法③が 12 人中 11 人で手法①,12 人中 10 人で手法②の作成時間より速く完了している。

手法ごとの作成時間に有意差があるかどうか調べるため t 検定を行った結果,手法③と手法①,手法③と手法②との間の辞書作成時間には有意な差があることが分かった。

意味情報を実用化するために必要な単語の数は膨大であり,また一度作成した後も既存の単語の更新や新しい単語の追加のために何回もチェックを行わなければならないため,その負担を減らすためにはより速く説明語辞書の作成が出来るかは重要であり,その点において提案手法が従来手法よりも効果的であることが分かる。

また,3 つの手法を各被験者が説明語を選択しやすかった順に順位付けした結果について,手法ごとに選ばれた順位の割合のグラフを以下の図 1 に記載する。

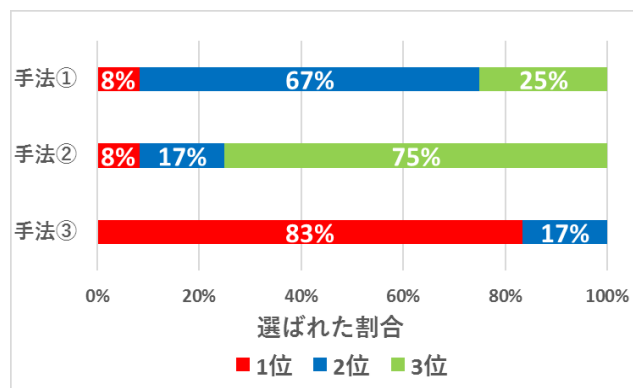


図 1:説明語の選択しやすさ

グラフから、説明語の選択しやすさは手法③が1位、手法①が2位、手法②が3位に選ばれる割合が多いという結果になっていた。

アンケートでこの順位付けの理由について聞いた所、手法③はソートによって単語を探しやすかった、こういう単語がいいかなとイメージしていたものが上の方に来ていてやりやすかったというような声が多かった。同じく説明語A、説明語B、説明語Cを用いているがソートをしていない手法②と比べて選択のしやすさに大きな差が出ている事から、Word2Vecを用いて単語の並び替えを行う事で説明語を選びやすくする効果がある事が分かった。

一方で、WordNetとWord2Vecの情報を使って説明語の候補を増やした手法②よりも手法①の方が説明語を選択しやすいという声が多いという結果となった。この理由については、手法②では情報量が多くなりすぎてしまい、ふさわしい単語を探しにくくなっていたという声が多かった。説明語の選択しやすさだけでは無く、手法②は手法①よりも作成時間が長くなってしまっている事から、説明語候補が多すぎた事によって説明語辞書の作りやすさの順位、説明語辞書の作成時間が共に落ちている事が分かった。ただし、手法①も問題の単語によっては候補が少なすぎて説明語にふさわしい単語が無いという理由で説明語を作成しにくい、説明語Bや説明語Cがそのような場面で役に立ってありがたかったという意見もあり、問題の単語が説明語Aだけで十分な候補があるか、それとも足りないかという性質によって手法①と手法②のどちらが説明語辞書を作成しやすいかは変わると考えられる。よって、この問題を解決するには、説明語A以外の説明語は既に使用している説明語候補の数が一定数以下の時のみ追加するなど、場合によって候補の数を変更する手法や、説明語として選択される可能性が低そうな単語は事前に消去しておいて邪魔にならないようにする手法などが効果的だと考えられる。

また、説明語の候補を追加した手法②と手法③において、実験で選択された説明語が説明語A、説明語B、説明語Cのどの候補からどれだけの割合で選ばれていたかを以下の表2に記載する。

表2：説明語候補の選ばれた割合

	説明語A	説明語B	説明語C
手法②	65%	25%	10%
手法③	66%	27%	7%

表2から、手法②でも手法③でも割合はほとんど同じであり、説明語Aからは65%程度、説明語Bからは25%程度、説明語Cからは10%程度が説明語として選ばれている

事が分かる。

手法③の方が説明語Aと説明語Bの割合がやや多く説明語Cの割合が少ない理由としては、ソートで探しやすくなっている分、説明語Cまで探すまでもなくふさわしい単語を見つけられるケースがあるためだと考えられる。

アンケートでそれぞれの説明語A、説明語B、説明語Cについての感想を聞いた所、説明語Aは基本的にここから探していたという意見、説明語Bは説明語Aだけでは説明語にふさわしい単語が5つ見つからない場合に使ったという意見、説明語Cは連想しやすい単語から外れた意味の単語が多くて時々良い単語はあったがあまり使えなかったというような意見がそれぞれ多かった。上記の選ばれた割合とも一致している感想であり、これらの結果から、説明語Bは説明語Aだけでは足りない場合に使うという追加した目的通りの効果を発揮しており、説明語選択に役立っているが、説明語Cは説明語Bと比べて効果が薄く、前述の情報量が多すぎて逆に探しにくくなってしまう問題も合わせて考えると説明語Cを追加してもデメリットに見合う効果が得られないと考えられる。

また、手法③でソートによってWord2Vecの類似度上位の単語が選ばれやすくなっているのかを調べるため、手法②と手法③で上位に来ている単語がそれぞれどれだけ選ばれたかについて調べた。手法②では選ばれた単語の34%が説明語A、説明語B、説明語Cの上位10番目までの単語だったのに対し、手法③では60%ほどが上位10番目までの単語から選ばれていた。このことから、ソートによって上位に来た単語の多くは連想しやすくて説明語にふさわしい単語と判断されており、またそれらが選ばれやすくなっている事が分かった。ただし、ごく少数であるがソートによって上位に来た単語がピンと来ておらず、あまり役立っていなかったと感じた人もおり、どのような基準でソートをするのかについてはまだ改善の余地があると考えられる。

また、手法②と手法③で追加したWordNetの定義文を使用した意味例について、アンケートでどれくらい使ったかについてや、使った際の感想を尋ねた。全く使わなかったという人もいたが、意味が説明しにくい抽象的な単語やあまり知らない単語について説明語辞書を作成する時に参考にした、ちょくちょく見たという声も多かった。ほとんど見なかった、あまり役に立たなかったという人が4名、参考にしたという人が8名おり、これも説明語辞書を作成しやすくする効果があると言える。

これらの結果から、WordNetやWord2Vecの情報を説明語の候補として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より速く作成でき、より作成しやすく感じられることが分かった。

4. 連想しやすさの評価実験

4.1 実験内容

説明語選択実験の結果、説明語候補を Word2Vec の類似度でソートして並べ直す提案手法が説明語辞書を作成する際にかかる時間を短くし、またより作りやすいと感じられる手法である事が明らかになった。

続いて、提案手法で選択された説明語が従来手法で選択された説明語と比べて単語の連想しやすさに違いがあるかも調べる。そこで、12名の大学生を対象に、説明語選択実験で選ばれた説明語について、それぞれ問題の単語の連想しやすさを評価する実験を行った。

評価は1～5の5段階評価で、1が最も連想しにくい、5が最も連想しやすいとした。また、例外として説明語の意味がそもそも分からなかった場合は0と付けてもらった。連想しやすさの考え方として、実際に意味情報による仮名漢字変換を行う際は説明語と単語の組み合わせで意味を連想して判断するため、問題の単語と説明語を並べて「この単語はこういう意味の単語だ」と連想できるかどうか判断して評価して欲しいと伝えた。

説明語選択実験では、12人の被験者が3つの手法でそれぞれ説明語を選択した。今回の実験ではその被験者ごとにデータを分け、今回の実験の被験者1人につき説明語選択実験での被験者3人分のデータを評価してもらった。これにより、説明語選択実験の1人分のデータを3回ずつ別の人に評価してもらう形で実施した。

4.2 実験結果と考察

手法①、手法②、手法③のそれぞれについて、連想しやすさの評価0～5が選ばれた割合を以下の表3に記載する。

表3:連想しやすさの評価の割合

	0	1	2	3	4	5
手法①	2%	4%	9%	19%	33%	34%
手法②	2%	3%	9%	19%	34%	33%
手法③	2%	3%	7%	16%	38%	34%

表3から、どの手法も連想しやすさは全体的に似た結果になっている事が分かる。連想しやすいと言える評価4以上の割合が7割ほど、連想しにくい評価2以下の割合は1割程度となっており、どの手法でも連想しやすい説明語を選択できていると言える。連想しやすさの評価が4以上になっている割合は手法①が67%、手法②も67%、手法③が72%と手法③が最も高く、手法②は手法①と同じ割合になっていることから、手法②も手法③も従来手法である手法①と同等かそれ以上の連想しやすさの説明語を選ぶ事が出

来ている事が分かった。

手法③は説明語辞書の作りやすさのアンケートでは圧倒的な1位で、手法②は最下位という結果だったが、両者の説明語の連想しやすさの評価にはほとんど差が出ていなかった。また、手法①は他の2手法と比べて説明語の候補が少なかったが、こちらも選択された説明語の連想しやすさに大きな差はなかった。この事から、素早く説明語を選択出来たり、説明語の候補が多くなったりしても選択された説明語の連想しやすさの評価に極端な差は出ない事が分かった。

これらの結果から、どの手法でも一定以上の連想しやすさがある事が分かった。また、説明語の連想しやすさが一番高かった事から、作成された説明語が連想しやすく、速く作りやすいと評価された手法③が3手法の中で一番優れている手法だと考えられる。

5. 親密度の評価実験

5.1 実験内容

説明語選択実験で選択された説明語のデータについて、連想しやすさの他に問題の単語や選ばれた説明語をどれだけ知っているかという親密度の評価実験も行う。

12人の大学生の被験者に、説明語と問題の単語についてどれだけ知っているかについての評価を1～5の5段階で評価してもらった。全く知らない場合は1、良く知っている場合は5で評価してもらう。今回は説明語単体で知っているかどうかの評価であるため、問題の単語とは別々に評価をしてもらった。

連想しやすさとの関連を調べるため、この実験は連想しやすさの評価実験を行った人と同じ被験者で実施した。また、評価してもらう説明語のデータもその被験者が連想しやすさの評価実験で評価したものと同じものを用いて実施した。

5.2 実験結果と考察

手法①、手法②、手法③のそれぞれについて、親密度の評価1～5が選ばれた割合を以下の表4に示す。

表4:親密度の評価の割合

	1	2	3	4	5
手法①	1%	2%	7%	18%	72%
手法②	1%	2%	6%	16%	75%
手法③	1%	1%	4%	16%	77%

表4から、連想しやすさの評価と同じく、どの手法も親密度は全体的に似た結果になっている事が分かる。どの手

法でも9割以上が親密度4以上の評価を付けられており、逆に親密度2以下の評価はほとんどされていない事が分かる。これは知らない単語は調べても良いとしたとはいえ、そもそも説明語を選択する際にあまり知られていない単語は選ばれにくく、良く知っている単語の方が選びやすかったためだと考えられる。

親密度5と評価された割合は手法①よりも手法②、手法③の方が高くなっている。これは手法①よりも候補が増えた事によって知っている単語が増え、あまり知られていない単語から選ばなくてはならなくなる事が少なくなったためという可能性がある。

また、全体の親密度の数値と連想しやすさの数値の組み合わせについて、それぞれ全体の内どれくらいの割合で構成されていたかを以下の表5に示す。

表5: 連想しやすさと親密度の組み合わせの割合

	親密度1	親密度2	親密度3	親密度4	親密度5
連想0	0.50%	0.36%	0.46%	0.26%	0.35%
連想1	0.07%	0.17%	0.33%	0.50%	2.03%
連想2	0.04%	0.31%	0.83%	1.37%	5.72%
連想3	0.16%	0.40%	1.56%	3.85%	11.83%
連想4	0.07%	0.39%	1.96%	7.52%	25.06%
連想5	0.04%	0.08%	0.65%	3.44%	29.67%

表5から、全体の中で親密度の評価が5で連想しやすさの評価も5の場合が一番割合が多い事が分かる。手法①、手法②、手法③ごとに分けた場合でもほぼ同じ割合になっていたため、これについての記載は省略する。

親密度ごとに連想しやすさの割合が最も高くなっている箇所を見ていくと、親密度1では連想しやすさ0の割合が一番高く、親密度2では連想しやすさ3, 4が多く、親密度3, 親密度4では連想しやすさ4が一番高く、親密度5では連想しやすさ5が一番多いという結果になっている。このことから、親密度が高いものは連想しやすさも高くなっている事が分かる。

また、問題の単語の親密度についても調べた。親密度1から親密度5までの割合は順に2%, 5%, 9%, 19%, 65%となっていた。また、問題の単語全体の親密度平均値は4.33となっており、説明語選択実験では問題の単語は全体的に良く知られた上で答えられていた事が分かった。

これらの結果から、親密度は問題の単語についても説明語についても全体的に高くなり、また親密度が高い説明語は連想しやすさも高くなっている場合が多いという事が分かった。

6. おわりに

視覚障がい者のパソコンによる仮名漢字変換を改善出来る意味情報による仮名漢字変換の実用化に必要な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、それらによって出された候補を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、説明語の候補をただ追加するだけの手法では従来手法を上回る結果が出なかったが、類似度で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と感じられていた。また、選択された説明語の連想しやすさが従来手法と比べて高かったことから、提案手法はより漢字を連想しやすい説明語を選択できていた事が分かった。これらの結果から、このソートを行う手法には連想しやすい説明語を従来手法より速く、より容易に選択でき、説明語辞書作成の負担を減らす効果がある事が分かった。

また、親密度についても調べた結果、選択された説明語のほとんどは親密度がとても高いものであり、親密度が高い説明語では連想しやすさも高くなっている場合が多い事が分かった。

今後の課題としては、まだ20単語しか実験が出来ていないため、今後はより多くの漢字を対象にした実験の実施や、実際に音声で聞いて漢字を変換できるかの実験、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いきたい。

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] 渡辺 哲也, 渡辺 文治, 山口 俊光, 大杉 成喜, 澤田 真弓, 岡田 伸一: スクリーンリーダーの詳細読みによる漢字想起実験: 成人を対象とした場合, 電子情報通信学会論文誌, D-I Vol. J89-D, No.3, pp.602-605, 2006.
- [3] 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.
- [4] 川崎 博章, 笹野 遼平, 高村 大也, 奥村 学: テキストコーパスを用いた漢字詳細読みの自動生成, 情報処理学会論文誌, 54(12), 2481-2491, 2013-12-15
- [5] Hazel Mok Shu Wen, Gao Huini Eshley and Francis Bond: Using Wordnet to predict numeral classifiers in Chinese and Japanese, 2012, *GWC*
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, Jeff Dean: Distributed Representations of Words and Phrases and their Compositionality, 2013, *NIPS*
- [7] 日本語 Wikipedia エンティティベクトル”.
http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ (参照 2018-11-25).