

修士論文

記述式小テストの解答群の分類
—類似解答の抽出方法の検討—

平成29年度修了

三重大学大学院工学研究科
博士前期課程 電気電子工学専攻

伊藤 慎治

目次

第1章	はじめに	1
1.1	講義中の講師に求められる行動	1
1.2	理解状況把握のための能動的行動	2
1.3	小テストの種類と計算機による支援	3
1.4	本論文での取り組み	4
第2章	記述式小テスト実施時の支援方法	5
2.1	想定する記述式小テストの内容と実施手順	5
2.2	先行研究としての記述解答支援システム	6
2.3	類義表現の揺らぎの問題点	10
2.4	先行研究の問題点	11
第3章	類義表現の集約	12
3.1	集約手法の検討	12
3.1.1	機械学習を用いた手法	12
3.1.2	概念辞書を用いた手法	14
3.2	類語辞典との比較実験	15
3.2.1	実験内容・評価方法	15
3.2.2	実験結果	17
3.2.3	考察	19
第4章	解答群の分類のための抽出方法	22
4.1	抽出の際に留意すべきこと	22
4.2	提案手法	23
4.2.1	修飾・被修飾の関係を持つ文節対の活用	24
4.2.2	類義表現・助詞の集約	26

4.2.3	抽出方法	28
4.3	実際の解答群を用いた評価実験	29
4.3.1	実験内容・評価方法	29
4.3.2	実験結果	31
4.3.3	考察	31
第5章	まとめ	33
	謝辞	34
	参考文献	35
	発表論文	37

目 次

2.1	フィードバックの流れ	6
2.2	キーワード表示インターフェイス	8
2.3	フレーズ表示インターフェイス	9
2.4	全文表示インターフェイス	9
4.1	係り受け解析の結果 1	25
4.2	係り受け解析の結果 2	25

表 目 次

3.1	3章の評価実験に使用した小テスト	16
3.2	類似度の許容幅に基づくグループ分け	18
3.3	評価結果：名詞「構文」「必要」の比較	18
3.4	評価結果：動詞「用いる」「調べる」の比較	18
3.5	名詞「構文」の類義語と類似度	21
3.6	動詞「調べる」の類義語と類似度	21
4.1	問「機械語について説明せよ」の解答例	23
4.2	学習者X「二進数で 表された」の品詞分解例	27
4.3	学習者Y「二進数で 表されている」の品詞分解例	27
4.4	実験で比較対象とした手法	29
4.5	4章の評価実験に使用した小テスト	30
4.6	評価結果：再現率および適合率	32
4.7	正しく分類できなかった解答一覧	32

第1章 はじめに

本章では、一連の研究背景を述べたうえで、本論文での取り組みを簡単に述べる。

1節では、講義中の講師に求められる行動をまとめ、どのような行動が好ましいか考察する。2節では、理解状況把握のため能動的行動として各々の方法を挙げ、小テストに着目した背景を説明する。3節では、小テストの種類や計算機による支援方法を述べ、特に記述式小テストに着目した理由を述べる。4節では、記述式小テスト支援システムに向けて、本論文でどのようなことを検討していくのか説明する。

1.1 講義中の講師に求められる行動

講師は、自身の講義内容の改善を求められている。理由の一つとして、ファカルティ・ディベロップメント (Faculty Development, FD) による影響が挙げられる。FDとは、「講師が授業内容・方法を改善し向上させるための組織的な取り組みの総称」であり、2008年度以降、大学ごとに実施することを義務づけられている。そのため、大学で講師をする上で、FDは意識せざるを得ない。また、FDは、講師が自発的に取り組むことによって実効性が高まるという意見もあり [1]、積極的に講義内容を改善していくことが望ましい。一方で、大学教育現場では、「学習者のレベルにバラつきがあり、授業を行いにくい」という課題が顕在化している [2]。これは、大学入試において、推薦入試やAO入試が盛んに採用されており、さまざまな学習能力・特性を持った者が混在していることが原因の1つである [3]。そのため多くの講師は、講義改善のために、講義を受けている学習者の理解状況の傾向を把握し、レベルに合わせた指導をする必要がある。

しかし、日本の教育現場においては、学習者からの反応は得づらいという問題がある。そのため、理解状況を把握するには、講師からの能動的行動が必要であ

る。例えば、小林らは、大学などの学習者が多人数で行われる講義において、学習者は講義に集中せず、反応しない傾向があるとしている [4]。また、藤井らは、学習者は他の学習者のことを常に意識しており、講義中の質問行動を取りづらさ可能性があることを指摘している [5]。加えて、祐宗らは、大勢の人の前で自分の意見や考えを発表する訓練がなされていないことが原因で、授業中の学習者の無質問行動が極めて目立つ、としている [6]。これらを踏まえ、講師が受動的な姿勢では、十分な理解状況は把握できないので、能動的な姿勢が求められる。

1.2 理解状況把握のための能動的行動

理解状況を把握するための代表的な能動的行動は、学習者への問いかけ、机間巡視、小テストの出題などが挙げられる。各々のメリット・およびデメリットを検討する。まず、問いかけは、余計な手間は必要なく、質問だけで完結する行動である。一方で、前述したように、特に日本の学習者は反応に乏しい。そのため、問いかけだけでは不十分である。また次に、机間巡視は、個々の学習者の講義中の様子を確認することができる [7]。しかし、講義時間内に学習者全員に対して実施することは難しい。そのため、予め着目する学習者を想定する講師もいる。しかし、この方法は、もし推察が間違っていた場合に学習者の実態がわからず、また、個々の学習者の理解状況を把握できるとは言い難い。最後に、小テストについて説明する。小テストが出題された際、学習者は何らかの解答をする。そのため、解答内容から理解状況を把握することができる。西森の大学の数学基礎教育に関する調査 [8] によると、授業改善の方法として、小テストを多くの講師が利用していることを述べている。また井上らは、小テストは学習者に授業参加を促す一因になることを示唆している [9]。このように、小テストは授業改善のための理解状況の把握方法として有用である。ただし、多人数講義では学習者全員の解答内容を読むことが大変になることも明らかである。これらをふまえ、本研究では、小テストに着目する。

1.3 小テストの種類と計算機による支援

小テストの代表的な種類としては、多肢選択式、穴埋め式、記述式の3種類が考えられる。

多人数講義における小テスト実施の講師の負担を軽減させるため、近年の計算機の発達を理由に、さまざまな支援の方法が検討されている。例えば、Moodle[10]などの学習支援システム(LMS)では、小テストを出題し、学習者の解答を回収し、解答形式によっては自動で採点できる。更に、小テストの結果を自動で分析・集計することで、講師が学習者の理解状況を把握しやすいように支援できる。また、クリッカーと呼ばれる機器では、多数の学習者の反応を即座に集めることができる[11]。しかし、一方で、柳澤らはクリッカーに関して、限られた授業時間の中で、クリッカーの配布、学習者の操作、問題の提示・解答・集計に若干の時間を必要とし、授業運営に支障を与える可能性もあることを述べている[12]。

これらのeラーニングシステムによる支援が容易な小テストは、多肢選択式・穴埋め式に限定される。このような形式の小テストは、学習者は推測で解答できるため、理解状況の把握という観点からは好ましくない。講師はそれを防ぐために入念な準備(選択肢の解答・穴埋め文章の作成)をする必要があり、大きな負担となる。更に、有効な選択肢の作成など、出題時の講師の負担が大きくなる。そのため、講義中に、学生の状況を捉えたいと思った瞬間に、これらの小テストを実施することは難しい。自由記述式の小テストは、多肢選択式とは反対の特徴を持つ。記述式の解答では、自分自身で文章を作成するため、解答文中に理解状況に関する情報が多く含まれる。加えて、講師は問題を準備するだけでよく、講義中にすぐに実施できる。更に、村山は、記述式小テストを実施すると学生自身が理解を深めるように学習することを示している[13]。これらのことから、講師からの能動的行動として、自由記述式の小テストが最も適切だと考える。

このような、講義中に何らかの記述式の作文を出題し、活用されている例としては、リアクションペーパーが挙げられる。リアクションペーパーとは、講師が学習者に「講義に関する感想や質問を記述させる」ツールである。須田は、リアクションペーパーの記述内容から、講義中の学習者の学びの様相を捉える手がかりとして扱うことができることを示唆している[14]。小野田は、リアクションペー

パーは特別な準備が必要なく、また講師の意図に合わせた出題ができ、学習者からの情報発信を得るためのツールとして適切であることを述べている [15]。このことから、広義の意味でのリアクションペーパーとして、自由記述式の小テストを出題することで、講師は学習者の理解状況を把握することができる。

しかし一方で、自由記述であることが原因で、講師は理解状況を把握するために一人ひとりの解答文を読む必要があり、その負担は重い。また、解答を読むことに多くの時間を割いてしまうと、即時的なフィードバックが不可能になってしまう。記述式小テストのメリットを活かすためには、このようなデメリットを解決する必要があるだろう。

1.4 本論文での取り組み

前節までの背景をふまえ、本論文では、講義中に1, 2文の短答自由記述式の小テストを実施した際、学習者の理解状況を把握できるように支援する計算機システムを構築することを最終目的として研究してきた。第2章では、想定される記述式小テスト実施の流れを述べた上で、既存のシステムをいくつか取り上げ、その問題点を述べる。第3章では、自由記述文に散見される類義表現を集約する手法を述べる。具体的には、機械学習 (word2vec) を用いた手法や、概念辞書 (日本語 wordnet) を用いて類似度を計算する手法について比較検討を行う。講師が類似していると感じる語は人それぞれである可能性に対応できる、適切な手法を検討していく。第4章では、講師が解答を閲覧した際、気になった内容の文章を記述している解答の学習者を抽出する手法を述べる。提案手法として、解答文を形態素解析した結果から、解答文に含まれる単語・および文節対を整理することで、気になった内容の文章との類似度を求めた結果を述べる。

第2章 記述式小テスト実施時の支援方法

本章では，記述式小テストの実施において，どのような懸念事項があるのか考え，そのための従来のシステムを簡単に説明したうえで，本研究で取り組む事柄を述べる．

まず，1節でどのような手順で記述式小テストを実施することを想定しているかを述べる．2節では，その記述式小テストを実施したときの支援システムとして，従来のいくつかのシステムを紹介する．また合わせて，3節では，記述式小テストでは生じてしまう可能性の高い類義表現の揺らぎに関する説明もする．最後に，4節では，特に解答分類のシステムについて問題点を挙げ，本論文で目標とするシステムを述べる．

2.1 想定する記述式小テストの内容と実施手順

本論文で想定する記述式小テストの実施手順の概要を以下の図 2.1 に示す．

1. 講師は，講義中に学習者の理解状況を知りたくなったとき，記述式小テストを出題する．
2. 学習者は，講師の出題を基に 1～2 文程度の簡単な解答文を作成し，解答を計算機に提出する．
3. 講師は学習者ごとの解答文を読み，その場でフィードバックの内容を検討する．このとき，計算機側では，近い解答内容の学習者を抽出する．
4. 該当する学習者・または全員へフィードバックを返す．

これらの手順の中で、3.において、講師は学習者一人ひとりの解答文を読む必要があり、負担が多い。また、多人数講義であると多くの時間もかかってしまい、講義後のフィードバックになってしまう恐れがある。しかし、効果的な講義内容の復習としては、フィードバックは講義時間内に行われるべきである。つまり、講師の負担を軽減しつつ、講義時間内に行えるような効率的なフィードバックを実現するには、特に3.の抽出の際に、計算機側から、適切な学習者を抽出できる必要がある。また、講師の気になる内容の解答は1つではないと考えられるので、検索要求は一度ではなく、数回生じる可能性も高い。そのため、さまざまな解答内容に基づき解答群を分類して、記述式小テストを実施した際の講師の支援ができるような計算機システムが求められる。

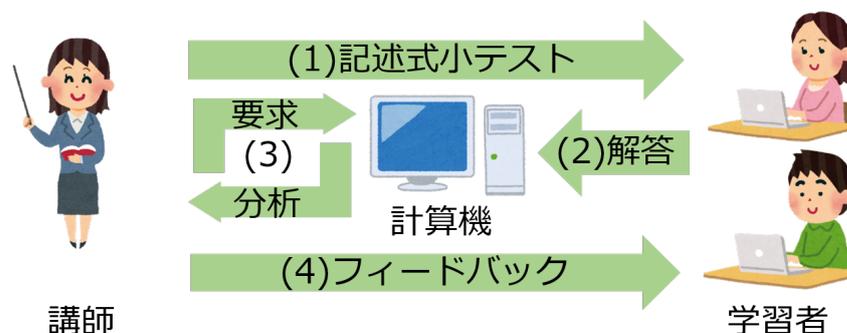


図 2.1: フィードバックの流れ

2.2 先行研究としての記述解答支援システム

記述式の解答に関する支援方法・および関連するシステムは、これまでにいくつか提案されている。例えば、下記のようなシステムが挙げられる。各々について簡単に説明を行う。

1. 文章構造に着目したシステム
2. 解答内容に基づき解答を分類できるシステム
3. 単語の使われ方を確認できるシステム

(1) 文章構造に着目したシステムには、石岡らによって開発された日本語小論文の自動採点システム Jess がある。Jess では、エッセイタイプの解答を採点することができる [16]。また、内容よりは、文章の構造に重みを多くした採点を行っている。ここで、修辞は漢字/カナの割合、ユールの K 特性値、ビクワードの割合で評価し、論理構成は接続詞と指示代名詞の数で評価し、内容は問題文に対して適切な内容かどうかを評価する。しかし、理解状況の把握において、文章の構造ではなく、内容を見て把握する必要がある。そのため、Jess の採点手法では学生がどの部分が理解不足などかはわからない。よって、想定しているような短答記述式小テストには向かない。

(2) 解答を分類するシステムとしては、高瀬・大井らによる記述式小テストの自動分類システムがある [17]。このシステムでは、単体のトピックセンテンスを持つ（複数の内容を持たない）解答群を、解答内容に基づき分類することができる。システムでは、文書中のキーワード（名詞・動詞）・係り受けに基づく重み付けに基づいて各文書を 1 つの特徴ベクトルで表現し、それら特徴ベクトルの類似性に基づき分類を行うことを目指したものである。より多くの文節から修飾を受けている文節ほど、その解答の内容を表していると考え、キーワードの重みを大きくしている。これらをワード法を用いて分類した結果、解答の内容に即した特徴ベクトルを生成できる。インターフェイス化はされていないが、講師の支援を十分にすることができるだろう。

(3) 単語の使われ方を確認できるシステムとしては、大庭による「三段階表示システム」がある。このシステムは、多人数講義での記述式小テストにおいて、講師がその解答群から解答群の概略をとらえやすくし、主要な内容をすばやく把握できるように支援することができる [18]。第一のキーワード表示インターフェイスは、解答群から自動抽出した重要そうな語を表示する。語は重要度に従い色分けされ、使用頻度が高いほど上方に表示される（図 2.2）。第二のフレーズ表示インターフェイスは、文節ごとの三列の表形式となっている（図 2.3）。中央の列はキーワードを含む文節を示し、左右の列の文節は中央の文節との修飾関係をもつ。文節の隣には文節の解答群中での使用回数が見られている。また、使用回数に応じてセルの配色は濃くなり、上位に表示される。第三の全文表示インターフェイスでは、先に指定された文節を含む解答全文を確認することができる（図 2.4）。こ

れら三段階の画面を活用することで、講師は解答群の単語の使い方を確認できる。

本論文では、講師が問題を出題する時点で、学習者からどのような解答が書かれるのか、ある程度想定している場合も多いと考える。そのような場合には、(3)のように、必ずしも単語から使われ方を確認する必要はない。そのため、このような記述式の解答を閲覧するシステムの中で、特に(2)解答内容に基づき解答を分類できるシステムに着目する。解答を内容に沿って分類することで、解答閲覧の負担を軽減することを狙う。

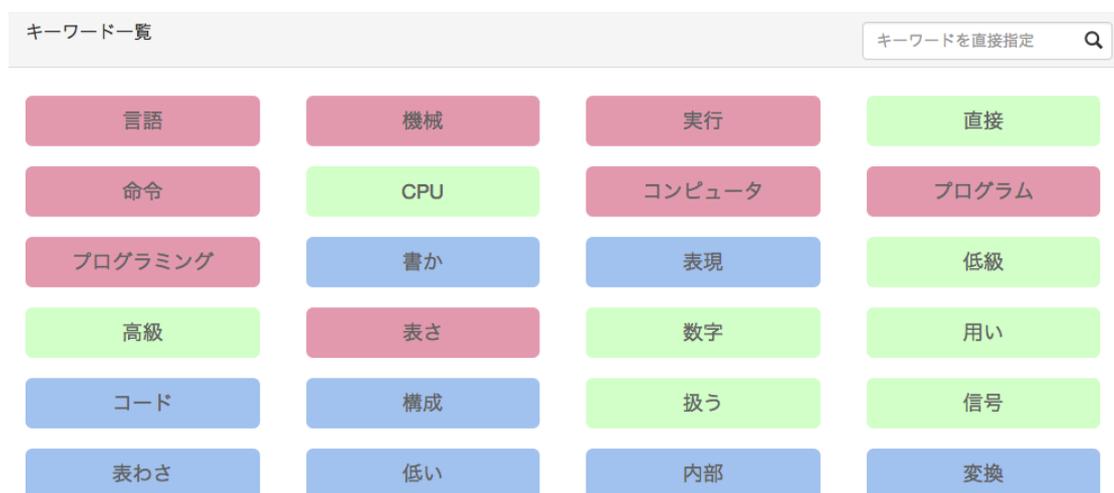


図 2.2: キーワード表示インターフェイス

「CPU」の周辺				上位5個▼
前	検索語	後		
2	コンピュータ内部の	CPUが 63	命令し	0
2	コンピュータ内の		実行できる	0
1	内部である		実行する	0
1	一部である		理解できる	0
1	いい		実行可能な	0
		その他		12
1	コンピュータ内の	CPUに 12	命令する	0
1	一部である		命令したりする	0
1	ある		命令して、	0
			実行、	0
			指示された	0
		その他		0
		CPUによる 0	命令を	0

図 2.3: フレーズ表示インターフェイス

「CPUが」と「命令し」を含む解答 ×

機械語は、マシン語ともいいCPUが命令し直接実行できる言語の総称である。人間が利用する際には二進数、十六進数を用いて表現する。

機械語は、CPUが直接読み取りプログラムを直接実行するために用いられる。また、2つの電気信号で表現されるので、人間が扱う場合は二進数や十六進数を用いている。CPUが命令し解析する。

機械語は、CPUが命令し二進数で表すものであり、プログラムとして直接実行される言語である。また機械語は、低級言語とも呼ばれており、人間は読むことが困難である。

「機械語は、コンピュータ内のCPUが直接実行可能なプログラムである。機械語でCPUが命令しそれをもとにして実行する。機械語は、二進数で表されている。」

機械語は、CPUが命令し直接実行するプログラミング言語の一つである。人間には、数字や一部英字を使い二進数や十六進数で表現される、最も低いレベルの言語である。

機械語は、二進数で構成されているオブジェクトコードのことである。コンピュータの命令できる言語であり、機械語で構成されていないプログラムを直接実行をすることを命令することはできない。またCPUが命令しプログラムを実行することができる。

機械語は、プログラムをコンピュータが実行する際に読み取ることのできるC言語で、その実体は二進数から成り立っている。また、CPUが命令しプログラムを動かす。

機械語は、CPUが読み取ることができ二進数で表記されCPUが命令し実行することができる言語のことを言う。

図 2.4: 全文表示インターフェイス

2.3 類義表現の揺らぎの問題点

解答群の内容に基づき解答を分類するうえで、考慮すべき問題点として、類義表現の揺らぎの問題が挙げられる。例えば、近い意味をもつ単語が計算機内で全く別の単語・語彙として扱われてしまうと、近い意味の解答群を別扱いしてしまう危険性があり、結果として、講師の負担を増加させてしまう。よって、計算機のシステムとして自由記述文章を扱ううえで、類義表現の揺らぎに対して何らかの対策を講じる必要があるといえる。そのため、本論文では、解答内容に基づき、解答を分類を目標とする過程で、類義表現の問題に対しても対策を行う。

記述式小テストの解答には類似表現が散見される可能性が高いことの詳細性を述べる。本論文で想定しているのは小テストの自由記述文であるため、学習者が短時間で記入する。そのため、学習者は語句の細かい意味の違いはあまり意識せず、何気なく使用することが考えられる。例えば、岩佐は、外国人に対する日本語の指導に関して、「学習者が紛らわしく分かりにくいと思う類義語の相違は、最も端的なケースでは『大体同じです』という説明で終わられてしまい、そこまで極端ではなくても、一見無難には見えるものの適切とは言えない語義説明がなされることが多い」と述べている [20]。これは外国人に教える際の問題だが、教える立場であっても、日本人が類義語の相異をあまり意識していないことが伺える。学習者であれば、類義表現は尚更あまり意識しないだろう。加えて、1章で述べたように、大学などの多人数講義の講義において、学習者の能力はさまざまである。これにより、学習者ごとの文章推敲能力にも差が出ることが考えられる。更に、短時間であれば、文章の丁寧な推敲には期待できない。これらを踏まえると、記述式小テストの解答には類似表現が散見される可能性が高い。

類義表現の問題が見受けられるシステムの例として、2節で述べた三段階表示システムが挙げられる。先行研究として、このシステムに関して、講師が解答群から注目すべき記述を探し出せることを評価実験により確認した [3]。しかし、システム使用時の問題点として、キーワード表示インターフェイスでは、類似した意味を持つ語が分かれて表示されている点が挙げられた。キーワード表示インターフェイスの図中の「表現」「表わさ」「表さ」といった語は全て類似した語であり、「表わす」ことを意味するが、別の単語として表示されている。これによ

り、講師は選択した単語と類似した意味を持つ語に対して複数のインターフェースを行き来する必要がある、閲覧の負担が軽減できない。また、このような表現の揺れが分けて計数されると、実質的な解答数をとらえ難い。この原因は、これまでの三段階表示システムでは、解答の文字列に基づいて処理していることに起因しており、語同士の意味の近さを考慮していないためである。

類義表現の揺らぎへの対策の検討に関しては、本論文の3章にて詳細を述べる。

2.4 先行研究の問題点

本節では、2.2節で述べた解答を自動分類するシステムの問題点を挙げ、本論文で求められる支援システムを述べる。

システムの問題として、複数の内容を持たない解答群を対象としていることが考えられる。しかし、実際の記述式小テストでは、1つの解答内に複数の述べるべき事柄があることも多い。また、講師が着目する内容も、同様に1つではない場面があるだろう。そのような状況において、このシステムは対応しきれない可能性が高いといえる。これをふまえ、本論文では、解答を分類する上で、講師の繰り返し生じる検索要求に対応できること、また、複数のフィードバックすべき事柄が含まれていても、分類に問題がないシステムを開発することを目標とする。これらを可能にすれば、講師の解答閲覧の負担は軽減できるだろう。

なお、近年では、新センター試験の実施に向け、短答式記述試験の自動採点および人間による採点を支援する実用可能なシステムを目指して、その試作および実装をした例もある [19]。しかし、このシステムの開発の際、「自然言語におけるテキスト間の同義や含意、推論を公的な試験に採用するほどの精度はこの数年では不可能であるという判断」が行われており、「採点基準」を設け、人間の手を加えることで採点を行っている。この文献を受け、本論文においても、計算機のシステムだけで完全に解答群を分類することは難しいと考える。よって、完全な分類ではなく、提案手法で実現できるような、語の使い方から判断できる範囲での解答群の分類を目指す。

解答群の分類のための類似解答抽出手法の検討に関しては、本論文の4章にて詳細を述べる。

第3章 類義表現の集約

本章では，記述式小テストの類義表現の集約方法に関して，主に機械学習を用いた手法，概念辞書を用いた手法の2種類を比較検討し，評価結果からの考察を行う．

3.1 集約手法の検討

類義語を判定するための簡単な方法は，類義語辞書を使用することであろう．しかし，類義語を集約する際に，講師ごとの考えの違いや，小テストの問題の内容の違いによっては，類義語だと判断する基準は変化する可能性を考慮しなければならない．そのため，似ているか否かのみを判断する類義語辞書では，十分でないと考えられる．そこで，類義語辞典を使うのではなく，単語同士の類似度を算出する手法に着目した．類似度であれば，単語同士がどれだけ似ているかを段階的に判断することができる．本論文では，類義語を判定する手法として，機械学習を用いた手法である，Word2vec[21]，および人の手により作られた概念辞書であるWordnet[22]に着目した．

3.1.1 機械学習を用いた手法

本節では，Word2vec[21]に関して，各々の仕組みの説明を簡単に行う．

Word2vecは，ニューラルネットワークを活用した仕組みであり，Tomas Mikolovにより発明された．文書群を学習することで，その文書群に含まれる各単語の並び順に基づき，各単語をベクトルとして表現することができる．学習により得られたベクトルが，「同じような意味や使われ方を示す単語は，同じような単語の並びで登場する」という前提に基づけば，ベクトルの近い単語同士は，似た意味を

包含すると考えられる。つまり、表記が揺れる単語や、意味の同じ類義語は、ベクトル同士の距離が近くなるといえるだろう。その結果、ベクトルの類似度を判定することで、数値として類似度を得ることができる。また、word2vecにより得られたベクトルでは、「パリ」-「フランス」+「日本」=「東京」というような意味に基づく演算が可能になることも知られている。

Word2vecの学習モデルには、Skip-Gram、CBOWの2つが存在している。今回は、本論文で使用したSkip-Gramについて簡潔に説明する。Skip-Gramはニューラルネットワークのモデルの1つである。Skip-Gramは2層のニューラルネットワークであり、入力層、出力層、隠れ層をそれぞれ1つずつ持つ。また、隣接する層のユニットは全結合している。このモデルを用いて、ある単語が与えられた時、その周辺の単語を予測する。入力層には文章群を単語ごとにone-hotベクトル化した行列を与える。one-hotベクトルとは、“ある”を”1”、“ない”を”0”としてフラグを表現したベクトルのことである。学習過程において得られる重みが単語の意味に基づくベクトルである。これにより、意味が近い言葉は周辺語も似ている可能性が高いことを学習することで、類義語に対し、類似したベクトルを得る。word2vecの詳細な学習の仕組みについては、文献[23]などで述べられているので、本論文では割愛する。

本論文では、単純に着目した二つの単語に対応するそれぞれのベクトルのコサイン類似度（ベクトル同士の角度の差の余弦）により、二つの単語の類似度を求めている。このとき、類似度は-1.0以上1.0以下となる。今回、事前に学習させる文章群（コーパス）としてはWikipediaの日本語版の全文のテキストを使用した。また、日本語の文章群を学習させる場合、分かち書き（単語ごとに区切ること）が必要になる。そのため、この処理にあたり、日本語形態素解析器MeCab[24]を用いて品詞分解を実施した。また、活用語は全て基本形に修正し、「機械語」などの複数形態素が組み合わさった名詞に関しては、日本語係り受け解析器Cabocha[25]により得られる文節内で形態素を連結し、1つの単語として処理を行った。なお、word2vecのパラメータの値は初期値を使用した（次元数200）。

3.1.2 概念辞書を用いた手法

本節では、概念辞書の日本語 Wordnet を用いて類義表現を得る手法について簡単に説明を行う。

日本語 WordNet は Princeton 大学が開発した、WordNet を用いた言語データベースであり、日本語の概念辞書である。元々は、WordNet は英語の概念辞書であり、日本語 Wordnet は英語 wordnet をもとに構築されている。日本語 WordNet では単語が synset と呼ばれる同義語のグループに分類されており、それらが他の synset (同義語のグループ) と意味的に結びついている。これらは上位下位の関係性を持ち、木構造のように存在し、概念の関係を表している。このような synset の仕組みを活かし、概念同士の距離を計算することで、類似度を得ることができる。本論文でも、概念距離を日本語 WordNet により得られる類義語の類似度として使用する。また、synset の木構造により、概念間の距離が複数ある場合、今回は最短距離を概念距離とする。

この日本語 WordNet を活用すると、例えば、名詞「プードル (poodle)」とは「通常刈り込まれる重たくてカールした一色の毛を持つ知的な犬」という単語の説明が得られる。同時に、「犬 (dog)」の下位語であり、「ラージプードル (large poodle)」「トイプードル (toy poodle)」などの上位語、領域 (Semantic Field) が「動物 (animal)」である、などの情報が得られる。しかし、現行の日本語 WordNet は間違いを 5%ほど含んでいると作成者が述べており、誤った類義語を提示してしまう懸念がある。そのため、どの程度適切な類義語が得られるのか、検討する必要があるだろう。類義語であるかどうかの判断は人によって異なるので、今回は、その間違いも含めた結果を使用する。

3.2 類語辞典との比較実験

本節では、前節までに説明してきた、機械学習 (word2vec) を用いた手法、概念辞書 (wordnet) を用いた手法の評価を行う。評価目的は、講師が類似していると感じる語は人それぞれである可能性を考慮した際、word2vec・wordnet で得られる類義語・およびその語の類似度が妥当かどうかを確かめることである。そのため、word2vec, wordnet で得られるそれぞれの類義語に関して、類義語辞典との比較検討を行い、結果の有効性を確かめる。

3.2.1 実験内容・評価方法

今回は、類似度を変更したとき、類語辞典に載っている語がどの程度含まれる、あるいは含まれていないのかを検証する。類義語辞典に載っている類義語は、多くの人が類義語だと感じるはずであり、網羅されている必要があるだろう。これにより、人が類義語だと感じる尺度と、word2vec・wordnet で得られる類義語の判定の妥当性を評価したい。

今回比較対象として用いる類語辞典は、類語新辞典 [26] および weblio 類語辞書 [27] である。それぞれの辞書を選択した理由を述べる。まず、類語新辞典の編集者は、自身の辞書に関して、「単語を意味上の類縁性によって区分し、類似の意味、関係深い言葉、対照的な意味を持つ語を集団化し、人々が心の中で暗々のうちに行っている比較対象を、明確な形でなしうるように眼前に提示するものである。」と述べている。そのため、人が類義語だと感じる語の代表例として適切だと考えられる。次に、weblio 類語辞典は、「日常的な言葉や表現の言い換え方を提案する辞書」だと述べている。web 上で使用できるため、学習者の類語検索をする際に活用していると考えられる。また内容も更新されているので、新語への対応なども期待される。尚、本論文の執筆時点で、weblio は、様々な同義語や同意語の日本語表現を約 40 万語を収録している。これら 2 つの類語辞典のどちらかに記載されていれば、人が類義語だと感じる語だとして問題ないだろうと判断した。

表 3.1 に示すように、実験には、三重大学・工学部・電気電子工学科で実際に行われた講義「計算機基礎及び演習 I」で実施された小テストの解答群を使用した。

小テスト中で用いられている単語の中から、名詞「構文」・「必要」、動詞「調べる」・「用いる」の4単語に関して、評価を行う。

またそれぞれの類似度の許容幅のグループ分けとして、表3.2のように区分した。これらのグループ分けは、講師が類似していると感じる語は人それぞれである可能性を想定した一例である。このような値の区分にした理由として、word2vecによる結果は、類似度1.0の単語はほぼ無く、類似度0.9~0.8代も非常に少ない。よって、講師が類義語処理を行うのであれば、類似度0.7からが妥当だと考えた。wordnetに関しては、synset（同義語のグループ）が同じであれば、その類義語の類似度は1.0である。また、それ以降は概念距離+1の逆数の値を類似度としていたので、概念距離ごとの区分とした。これらの区分は一例であり、最終的には、講師が望むような類似度の幅で調整することができるのが理想形である。

一致率は、以下の式で求める。

$$\text{一致率 } X = \frac{(\text{類語辞典に記載されている類義語}) \cap (\text{求めた類義語})}{\text{求めた類義語}}$$

問題	デバッグとはなにか説明せよ。ただし「ステップ」を用いること。
模範解答例	デバッグはステップ実行などを使用することで、バグを発見できる。また、バグを修正することはできないが、修正の支援をすることができるツールである。
学習者総数（解答数）	102人
出題年度	2008年度
平均字数	75文字

表 3.1: 3章の評価実験に使用した小テスト

3.2.2 実験結果

名詞「構文」「必要」と、動詞「用いる」「調べる」に分けた実験結果を表 3.3, 表 3.4 に示す. 表中の「-」は, その類似度のグループで類義語を抽出できていないことを示している. 具体的には, wordnet では, 構文に関する類義語は類似度 0.33 (類似度の許容幅グループが大, 概念距離 2) まで存在しなかった.

最初に, この表に関して, 実験全体の結果を述べる. word2vec, wordnet ともに, 類似度の許容幅が広くなる (類似度グループが大に近づく) ほど, 類義語辞典から得られる類義語とは異なる語を多く抽出していることがわかる. このことから, どちらの手法でも, 類似度の低い類義語は, 講師の意図する類義語とはかけ離れている可能性が高いといえる.

次に, 名詞「構文」「必要」に関する結果を述べる. 表 3.3 より, word2vec では, どちらの単語の類義語に関しても, 類義語辞典とは合致しない類義語を提示していることがわかる. また一方で, wordnet に関しては, 「構文」では高い類似度を持つ類義語が無いことを示しており, 「命令」の類似度 1.0 (類似度の許容幅グループ小, 概念距離 0) の語を適切に抽出することができている.

最後に, 動詞「用いる」「調べる」に関する結果を述べる. 表 3.4 より, word2vec では, 類似度 1.0~0.70 (類似度の許容幅グループが小) では, 「用いる」の類義語を 66 % (3 単語中 2 単語) 得ることができている. 「調べる」についても, 類似度 1.0~0.70 の類義語の全てが類義語辞典と異なる語ではないので, 名詞よりは良い結果を得ている. 一方, wordnet に関して, 類似度 1.0~0.5 (類似度の許容幅グループが中, 概念距離 1) までは word2vec と同等か, それより良い結果を得ている. 類似度 1.00~0.33 の類義語に関しては, word2vec よりも 3~5 % 悪い結果となった.

以上の傾向を総括すると, 動詞, 名詞ともに, word2vec よりも wordnet のほうが類義語辞典に近い語のみを類義語として抽出できているであろう. ただし, wordnet により得られる類義語でも, 類似度 1.0~0.50 (類似度の許容幅グループ中, 概念距離 1) では, 類義語辞典との一致率が 50 % 以下である. このことから, 講師が類似していると感じる語が人それぞれである可能性を考慮することに関しては, wordnet による手法でも対策しきれていないとはいえない. まとめると, wordnet は, 少なくとも類似度 1.0 (概念距離 0), つまり synset が同じとなる場合, 類義語の集約方法として使うことができる可能性を示したといえる.

類似度の許容幅	word2vec	wordnet
小	1.00~0.70	1.00 (概念距離 0)
中	1.00~0.60	1.00~0.50 (概念距離 1)
大	1.00~0.50	1.00~0.33 (概念距離 2)

表 3.2: 類似度の許容幅に基づくグループ分け

類似度グループ	一致率 [%]			
	構文		必要	
	word2vec	wordnet	word2vec	wordnet
小	0	-	0	100
中	5.0	-	6.0	50
大	3.8	0	2.0	25

表 3.3: 評価結果：名詞「構文」「必要」の比較

類似度グループ	一致率 [%]			
	用いる		調べる	
	word2vec	wordnet	word2vec	wordnet
小	66.6	66.6	25.0	33.0
中	14.0	50	22.0	28.0
大	5.0	3.0	8.0	3.0

表 3.4: 評価結果：動詞「用いる」「調べる」の比較

3.2.3 考察

今回の実験結果に関して、以下の3点について考察を行う。考察のために、4単語の中から「構文」と「調べる」について、word2vec, wordnet を用いて得られる類義語の上位5つ、および類似度を表3-1, 表3-2に示す。

1. 動詞「調べる」について、word2vec, wordnet とともに類語辞典との一致率が低い理由
2. 名詞「構文」について、wordnet では該当する類義語が無いのに、word2vec では類義語があり、かつ類義語辞典との一致率が低い理由
3. 名詞「必要」について、wordnet の一致率が高いこと

第一に、動詞「調べる」に関して検討する。類義語辞典との一致率が低い理由としては、word2vec, wordnet でどちらも類似度が上位5語に入っている「確かめる」という語が、類語辞典では類義語として扱われていないことが挙げられる。また、word2vec では「見つけ出す」という語が出ているが、これに関して、類語辞典には「見出す」という類義語があるため、細かな表現の差が影響した可能性もあるだろう。「探す」と「探し出す」に関しても近いことが言える。このことから、比較対象とした類語辞典の数を増やすなど、評価として細かな現の揺れを許容することも検討すべきだったといえる。これらの影響により、一致率が低くなったと推察される。

第二に、名詞「構文」に関連することについて検討する。表3.5を見ると、wordnet により得られる類義語の類似度は、一番高くても0.33（概念距離2）である。そして、これらの語は類義語辞典による類義語と一致しない。対して、word2vec により得られる類義語の類似度はwordnet と比較して高く、類似度が上位5つの語は全て0.7代である。しかし、「実行時」「文法的」など、wordnet でも近い内容の語が、類似度0.25として得られていることが読み取れる。これらを合わせて考えると、「構文」に関しては、word2vec では判断の難しい語・類似していると考えづらい語の類似度を高く算出してしまった可能性があるだろう。また、「構文」に関する類義語は、weblio 類語辞典では無い、とされている。類語新辞典では、「主語」

など、関連性は高いが類義語とするのは意見の割れそうな語が多い。これらをふまえても、word2vec で上位の類似度の語は、類似度が不適切に高いといえる。

第3に、名詞「必要」について、wordnet の一致率が高い理由を述べる。wordnet により得られた類似度 1.0 の類義語は、「不可分」という語 1 つのみであった。この語が類義語辞典に含まれていたため、一致率 100 % という結果が得られた。また、類似度を 0.5 や 0.3 まで許容した際も、類義語はそれぞれ 2 つ、4 つであったために、全体的に一致率が高くなったといえる。これは、講師が類似していると感じる語が人それぞれであることを、類似度を用いて考慮することができる可能性のある一例だろう。

総括して、語同士が類似しているかどうかの判断は非常に難しいことがわかった。また安直な処理をしてしまうと、講師が解答を読む際の負担が余計に増加してしまう危険性もある。そのため、word2vec や wordnet での細かい類似度を用いて類義語を提示するには、使用方法に更なる検討が必要になるだろう。4 章では、3 章で述べた類義語の処理もふまえて、解答群の分類手法を検討していく。

word2vec		wordnet	
類義語	類似度	類義語	類似度
コンパイラ	0.75	オブジェクト	0.33
実行時	0.72	仕組み	0.33
機械語	0.71	位置	0.25
文法的	0.70	内容	0.25
オブジェクト	0.70	実行	0.25

表 3.5: 名詞「構文」の類義語と類似度

word2vec		wordnet	
類義語	類似度	類義語	類似度
確かめる	0.76	チェック	1.0
見つけ出す	0.73	探す	1.0
見つける	0.72	点検	1.0
探し出す	0.71	眺める	1.0
探す	0.67	確かめる	1.0

表 3.6: 動詞「調べる」の類義語と類似度

第4章 解答群の分類のための抽出方法

本章では、2章で述べたように、複数のフィードバックすべき事柄が含まれている解答群に対しても分類するために、類似解答抽出の提案手法を述べる。具体的には、解答群中の単語の有無だけではなく、単語の使われ方に着目していく。これにより、講師の解答閲覧の負担を軽減することを目的とし、提案手法により適切な類似解答が抽出できることを確かめる。

4.1 抽出の際に留意すべきこと

今回行う小テストは自由記述式であることをふまえると、解答群の抽出の際に生じる問題点は以下の2つである。

1. 1つの解答文中には、複数のフィードバックすべき事柄が含まれており、学習者ごとに語の使い方はさまざまである。
2. 類似した解答内容であっても、学習者ごとに多様な記述の仕方がある。

問題点を確かめるために、具体的な記述式小テストの解答例を基に説明を行う。想定している出題は、「機械語について文字程度で説明せよ。ただし、解答には「CPU」「実行」「二進数」という語をを含むこと。」である。筆者の考えた解答例を表に示す。解答群の記述内容としては、主に以下の2点が考えられるだろう。

- CPU が直接実行できること
- 二進数で表わされる言語であること

問題点1に関して、解答を講師が見たとき、講師によってフィードバックしたい事柄はさまざまであり、また複数あると考えられる。例として、「実行できる」という解答内容に着目したとする。二人の学習者は「実行できる」と述べているが、動作の主体は「CPU」と「機械」で異なる。また他にも「命令する」の使い方など、講師によってフィードバックしたい点はさまざまであり、同じ語を用いている解答であっても、語の使われ方は異なることがわかる。

問題点2に関して、学習者Xは「二進数で表わされた」、学習者Yは「二進数で表わされており」という記述をしている。このことから、両者とも「二進数で表わされる」という意味の解答を行っており、類似した解答内容であるといえる。しかし、自由記述であるため、表現の細部は異なっており、学習者それぞれで多様な記述が想定されるといえる。

これらの解答例を用いて、次節より、類似解答の抽出のための提案手法を説明する。

学習者	解答例
X	機械語は、CPUが直接実行できるプログラミング言語である。CPUへと指示するための言語で、二進数で表される。
Y	機械語は、CPUに命令するための言語で、二進数で表わされており、機械が直接実行できるものである。

表 4.1: 問「機械語について説明せよ」の解答例

4.2 提案手法

本節では、提案手法として、以下の2点の提案に基づく類似解答の抽出を行う。この結果に基づき解答群を分類することで、講師がフィードバックの対象となる学習者を探す際に、学習者を過不足なく特定できるようにし、解答閲覧の負担を軽減することを目標とする。

1. 修飾・被修飾の関係を持つ文節対を活用する

2. 品詞分解に基づき、多様な記述の表記の揺れを削減する。類義語の同一化を行うことで、類似表現の集約を行う

以下の節で、それぞれに関して詳細に説明する。

4.2.1 修飾・被修飾の関係を持つ文節対の活用

提案手法として、修飾・被修飾の関係を持つ文節対を活用する。記述式小テストにおいては、講師の出題した問題に対して解答文を作成するので、解答文中で使用される単語は似通ってくるのが容易に考えられる。また、問題点1で示したように語の使い方は個人によって異なる。よって、語そのものだけではなく、語の使われ方も理解状況を把握する上で重要な情報であるといえる。

2章の2節で述べたように、先行研究では、教育支援システムにおいて語の使われ方を提示する手法として、文節の修飾・被修飾の関係を用いていた[18]。また他にも、キーワードの重みを大きくする方法として、より多くの文節から修飾を受けている文節を活用していた[17]。このように、修飾・被修飾の関係に基づく文節対を用いることは、語の使われ方を得る手法として有効であろう。

本研究では、解答群から文節の修飾・被修飾の関係を得る手法として、形態素解析には日本語形態素解析器 Mecab[24]、係り受け（文節の修飾・被修飾の関係）解析には日本語係り受け解析器 CaboCha[25]を用いる。

具体的な説明のため、先程述べた解答例について、CaboChaを用いて係り受け解析をした結果を図4.1、図4.2に示す。それぞれの図の「実行できる」という語に着目する。どちらの図においても「直接」という文節とは修飾・被修飾の関係を持つので、「直接実行できる」という文節対を得ることができる。同様にして、それぞれ図4.1では「CPUが実行できる」、図4.2では「機械が実行できる」という文節対が得られる。講師が「CPUが直接実行できる」という解答を行った学習者の数が知りたいときは、言い換えると、文節の関係から「CPUが実行できる」「直接実行できる」という内容の文節対を持つ解答を抽出できればよい。この場合は、学習者Xのみが該当し、学習者Yは該当しなくなるので、これにより講師が望む学習者を抽出できる。



図 4.1: 係り受け解析の結果 1

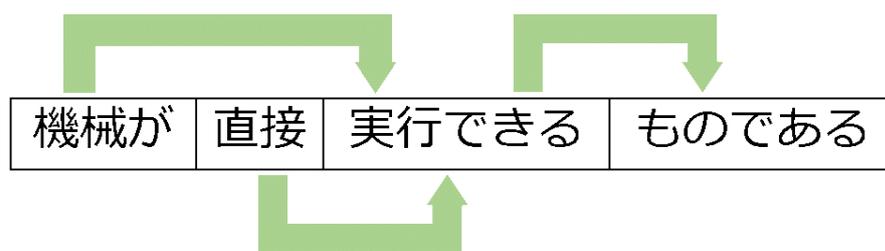


図 4.2: 係り受け解析の結果 2

4.2.2 類義表現・助詞の集約

記述式の解答文から得られる文節対に基づいた手法を提案する際、記述式小テストの解答では、解答の表現の多様性が問題となる。この多様性は、以下の2つの問題に起因するだろう。

1. 言い回しの揺らぎ
2. 類義語の使用

言い回しの揺らぎの対策

まず、1. 言い回しの揺らぎに関しては、提案手法として、品詞分解に基づき、記述の表記の揺れを削減することを狙う。文節の構造に着目すると、1つの文節は、1つの自立語といくつかの付属語から作られる[29]。自立語とは文節内の主要な内容を示す語で、主に名詞、動詞、形容詞である。付属語は自立語を伴って文節を作る語であり、助詞や助動詞のことを指す。更に、文節の中でも修飾節、被修飾節を区別して検討していく。

修飾節に使われる助詞などの付属語は、被修飾節の全文を修飾するわけではなく、その中の一部を修飾することがわかっている[28]。また助詞は全てで4種類あり、格助詞、接続助詞、副助詞、終助詞があるが、この内前半の2つは後ろの語・節との関係を表し、後半2つは話し手の捉え方を表すものである。これらを踏まえて、被修飾節の助詞、助動詞の違い、それに付随して生じる動詞の活用の違いを取り除き、動詞の語幹を抽出することで、学習者の解答内容そのものには手を加えずに、多様な記述に対応できると考える。

修飾文節の助詞は、被修飾節との関係を示すので、特にこのような作業を行わない。ただし、助詞「に」は動作の到達点を示し、「へ」は動作の方向を示すが、これらはほぼ同じ意味を指し示すことがわかっている[30]。また、「へ」に続く「と」は変化や結果を表す格助詞であり、「の」は表現がより細密になった場合に使用する格助詞である。これらに関して、学習者が解答文を推敲する際、微細な意味の使い分けはしないと考えられるので、全て同一視する。

表 4.1 の解答文例より，多様な記述例として，「二進数で表された」「二進数で表されている」という 2 文節対を抜き出す．これらの文節対の主要な内容は，「二進数で表される」であると推察される．よって，これら 2 つは同一視されることが好ましいだろう．これらの文節対に関して，Mecab による形態素解析の結果を表 3，表 4 に示す．この結果から，被修飾節の差異に着目する．すると，必要な部分は「表され」であり，「た」や「ている」の部分は文節中の主要な内容ではない．つまり，「表され」動詞の基本形である「表される」が共通の解答内容であるといえる．以上より，表 3，表 4 の文節対の例は，被修飾節の動詞の語幹を保持し，どちらも「二進数で表される」とすれば，表現の揺らぎの対策ができる．

修飾節		被修飾節	
二進数	名詞	表さ	動詞(未然, 自立)
で	助詞(格)	れ	動詞(連用, 接尾)
		た	助動詞

表 4.2: 学習者 X 「二進数で 表された」の品詞分解例

修飾節		被修飾節	
二進数	名詞	表さ	動詞(未然, 自立)
で	助詞(格)	れ	動詞(連用, 接尾)
		て	助詞(接続)
		いる	動詞(基本, 非自立)

表 4.3: 学習者 Y 「二進数で 表されている」の品詞分解例

類義語の対策

次に、類義語に使用に関しては、3章の評価結果で既に議論したように、日本語 WordNet による概念距離を活用することを提案する。これにより、本来近い意味で使用したはずの語が、別の語の扱いにならずに同一化することができ、多様な記述への対応が可能となるだろう。

表 4.1 の解答文例では、類義語を使用した文節対の例として、「CPU へと命令する」「CPU に指示する」という 2 文節対を抜き出す。これらの文節対の主要な内容は、同じ「CPU が命じる」ことであろう。これに関して、先程述べた「へと」「に」の同一化、および「命令する」「指示する」の類義語の集約を実施すると、どちらも「CPU に命じる」とでき、表現の揺らぎの対策ができる。

4.2.3 抽出方法

前節までで述べた、類義表現の揺らぎを対策した文節対・単語を用いて、類似解答を抽出する。講師が気になった内容の文章も同様の処理を実施し、解答群との文節対・単語の一致数を類似度とする。これは、以下の式で表される。そして、類似度が同じ値になった解答を、講師が気になった内容の観点で、近い意味の解答をしている解答グループとする。これにより、講師が気になる内容の文と類似した内容の解答の学習者の抽出が実現できる。

$$\text{類似度} = \frac{A \cap B}{A}$$

式中の A, B は以下に示すものである。

A 講師の着目した文章の内容を構成する文節対・単語

B 解答文を構成する文節対・単語

4.3 実際の解答群を用いた評価実験

本節では、提案手法が解答群の分類のための類似解答の抽出手法として適切であることを、他の手法と比較検討することで確かめる。

4.3.1 実験内容・評価方法

この実験では、提案手法として使用した、文節対に着目することや、類義表現の有効性を確認することを目標とする。そのための比較対象として、表 4.4 のような 4 つの手法での比較を行う。例えば、手法 A の場合、類似度の計算方法は前節で述べた式を用いる。ただし、提案手法とは異なり、類義語の集約を実施していない単語・文節対に関して、講師が気になった内容の文章と各々の解答文で類似度を計算する。同様に、手法 B では類義語の集約のみを実施し、文節対は考慮しない。手法 C では、単語のみの類似度の計算により、分類する。これら 4 つで比較する意図としては、提案手法では文節対による単語の使われ方への着目と、類義表現への処理の 2 つの要素を含んでいるためである。そのため、各々の効果を確認すること、およびそれを組み合わせた提案手法、これらをせずに単語だけで検索した場合で比較を行うことが最適だと判断した。

分類方法	単語の使用	文節対の使用	類義表現の処理
提案手法	○	○	○
手法 A	○	○	×
手法 B	○	×	○
手法 C	○	×	×

表 4.4: 実験で比較対象とした手法

本実験では、三重大学・工学部・電気電子工学科で実際に行われた講義「計算機工学 I」の中間テストで出題された問題を使用する。具体的な内容を表 4.5 に示す。この解答群を読み、実際に出題者が気になった記述内容は、「マクロ命令をマ

イクロ命令に変換する」であった。そのため、出題者には、筆者が解答群中から白紙ではない解答を無作為に 42 人分抜粋した解答に対して、「マクロ命令をマイクロ命令に変換する」ことが記入できているかどうか、1 人分ずつ解答文を確認し、正誤の判断を行っていただいた。これら 42 人分の解答の中に、正しく記述できていると判断された解答は 8 人分であった。よって、今回の評価実験においては、これら 8 人の解答を適切に抽出できることを確かめる。

尚、このテストは、中間テストであるため、実際には採点を目的としている。しかし、1,2 文で解答できる自由記述文であるため、理解状況を得るために出題する小テスト（採点を意図しない小テスト）との大きな差異はないと考えられる。よって今回は、講師が理解状況を得るために出題した小テストだと仮定して実験を行った。また、本実験において、学習者の推敲した解答文は誤字・脱字を含めて変更せず、提出された文章をそのまま用いている。

問題	マイクロプログラム制御方式とはどのような制御方式かを、40 字以上、80 字以下で説明しなさい。ただし、「メインメモリ」、「マクロ命令」、「マイクロ命令」を説明に用いること。
模範解答	メインメモリに格納されているマクロ命令を CPU が読み込むたびに、マイクロ命令群からなるマイクロプログラムに変換して実行する方法。
出題年度	2017 年度
解答した学習者総数	99 人

表 4.5: 4 章の評価実験に使用した小テスト

評価としては、以下に示す再現率 X (Precision), 適合率 Y (recall) から求める。これらの式を用いて評価する理由を述べる。再現率 X は、講師が気になった内容を解答文中で記述している学習者を多く抽出できるかどうかの評価をすることが目

的である。この値が小さい場合、類似解答の抽出として不十分だといえる。適合率 Y は、解答の抽出が過不足なく行えていることを評価することが目的である。この値が小さい場合、適切な学習者を分類できていても、見なければいけない解答総数は多くなり、講師の負担軽減としての効果が期待できない。また、再現率 X と適合率 Y は相反するものであり、解答を多く抽出すれば、再現率は上昇する一方で、適合率は下降する。そのため、最も適切な分類ができていないか確認する方法として、これら 2 つの指標を合わせた評価としたい。

$$\text{再現率 } X = \frac{(\text{講師が気になった解答の総数}) \cap (\text{適切に分類できた解答の数})}{\text{講師が気になった解答の総数}}$$

$$\text{適合率 } Y = \frac{\text{適切に分類できた解答}}{\text{類似度が最上位に分類された解答の総数}}$$

4.3.2 実験結果

実験のそれぞれの方法ごとの評価結果を以下の表 4.6 に示す。表 4.6 から、実験全体の結果を述べる。提案手法は、他の手法と比べて再現率・適合率ともに一番高く、解答群の分類ができていない。この抽出結果を基に適切な解答群を分類し、講師に提示すれば、講師の解答閲覧の負担が軽減できる可能性を示せたといえる。

次に、再現率・適合率に関して、それぞれの結果を述べる。適合率に関しては、提案手法と手法 A（文節対を用いた抽出）がともに 100% となり、文節対の効果により、講師の気になる内容の解答をしていた学習者を適切に抽出することができている。再現率に関しては、提案手法が 50% と一番高い。しかし、講師が気になった内容の解答をしている学習者の内、半分しか適切に抽出できていないともいえる。そのため、適合率をなるべく下げないように再現率を更に高めていく必要がある。

4.3.3 考察

本節では、今回の実験結果から、再現率が 50% になったことに関する考察をする。理由として、解答群中で使われていた語が、意味が近いと判断するのが難しい

語であり、類義語の処理が上手くできていなかったことが考えられる。表 4.7 は、適切に分類できなかった解答の詳細な内容である。今回、実験の際に出題者が気になった内容としたものは「マクロ命令をマイクロ命令に変換する」ことである。表 4.7 を見ると、学習者 X, Z は「変換する」ことを「分割する」「分ける」と述べており、学習者 Y, W は「変換する」を「取り出す」「する」と述べていることが伺える。つまり、出題者は小テスト内での語の意味として、これらを同義と見なしてもよいと判断しているといえるが、日本語 wordnet による類義語の判断ができていないのが原因だろう。これらの語は使う場面に応じて意味が変化してしまう可能性が高いと考えられ、自動的に処理することは難しい。よって、このような判断の難しい類義語をどのように扱っていくかも今後の課題である。

	提案手法	手法 A (文節対)	手法 B (類義語)	手法 C(単 語のみ)
適合率 [%]	100	100	50	75
再現率 [%]	50	25	37.5	37.5

表 4.6: 評価結果：再現率および適合率

学習者	解答例
X	メインメモリに格納されている命令によってマクロ命令が作られる。マクロ命令には数個のマイクロ命令に分割される。
Y	メインメモリに格納されているマクロ命令を取り出し、そのマクロ命令を執行するのに必要なマイクロ命令を取り出し実行する。よって複雑な操作を行うことができる。
Z	与えられたマクロ命令を、メインメモリ内でさらに多くのマイクロ命令に分けて実行する制御方式のこと。
W	メインメモリに格納されたマクロ命令を制御メモリに格納されるマイクロ命令にしてから実行をする制御方式。

表 4.7: 正しく分類できなかった解答一覧

第5章 まとめ

本研究は、多人数講義において、1,2文の短答記述式小テストを実施した際の講師の負担軽減を目的としたものである。

日本の教育現場には、①学習者の無質問行動が目立つ、②学習者の学力・能力レベルの差がある、など問題がある。そのため、講師は能動的行動により学習者の理解状況を把握する必要があるが生じている。本研究では、理解状況把握のための能動的行動として、1,2文で解答できるような短答記述式小テストを出題する方法に着目した。このような研究背景において、講師を支援するシステムの構築を最終目標としている。本論文では、解答群を分類するために、講師が気になる内容の文と類似した内容の解答の学習者を抽出できるようにすることが目標である。

3章では、解答文中に類似表現が見受けられる問題について検討を行った。具体的には、機械学習（Word2Vec）により得られる単語のベクトルを活用し、類似度を測る方法や、概念辞書（日本語 WordNet）を用いて、概念距離を計算する方法を実践した。類語辞典と各々から得られる類義語を比較検討した結果、本研究の目的においては、概念辞書を活用するほうが適切な可能性があることがわかった。

4章では、解答内容に着目して解答群を分類するために、講師が気になる内容の文と類似した内容の解答の学習者を抽出する手法を検討した。解答文を形態素解析し、各々の解答文内での修飾関係に基づく文節対を整理した。また、類義語の処理を実施した。これらを提案手法とし、評価実験を行った結果、類似した内容の解答の学習者を適切に抽出できていることを示した。

今後の課題は2つ考えられる。1つ目に、類義語の更なる検討が必要である。今回、日本語 wordnet の類似度 1.0 となる語を類義語だと扱ったが、講師ごとに類義語の判断が揺らぐことへの対策ができていない。また、評価実験では、判断の難しい語があることもわかっている。2つ目に、類似内容の解答を抽出した結果を用いて、具体的にどのように講師に解答群を提示するかを検討する必要がある。

謝辞

本論文は，著者が三重大学大学院・工学研究科・博士前期課程に在学中に行った研究をまとめたものである．本研究を進めるにあたり，懇切丁寧なご指導とご督励を賜った三重大学の鶴岡信治教授，高瀬治彦教授，北英彦准教授，川中普晴准教授に深く感謝いたします．また，日頃熱心に討論して頂いた情報処理研究室の皆様方にお礼申し上げます．

最後に，本論文をまとめるにあたり，助言，討論，その他お世話になったすべての方々に感謝いたします．

参考文献

- [1] 岡部光明：「ファカルティ・ディベロップメント（FD）の理念と実践的提案」，
明治学院大学国際学研究，第41号，pp.96-108（2012）
- [2] リベルタス・コンサルティング：「大学教員の教育活動・教育能力の評価の在り方
に関する調査研究」，http://www.mext.go.jp/a_menu/koutou/itaku/_icsFiles/afieldfile/2016/06/02/1371454_01.pdf，2017-12 参照。
- [3] 文部科学省：“平成 29 年度国公立大学入学者選抜について”，
http://www.mext.go.jp/a_menu/koutou/senbatsu/_icsFiles/afieldfile/2016/10/06/1377882_3.pdf，2017-6 参照。
- [4] 小林俊公，柳沢学，池田博一：「呼応度を用いた授業の分析」，工学教育，vol.62，
no.6，pp.56 - 61（2014）
- [5] 藤井利江，山口裕幸：「大学生の質問行動を規定する要因についての研究—学生
はなぜ授業中に質問しないのか？—」，九州大学心理学研究，vol.4，pp.135-148
（2003）
- [6] 祐宗省三，無藤隆，SHWALB David，SHWALB Barbara，仲野好重：「授業
中における大学生の無質問行動をめぐる教育心理学諸問題」，日本教育心理学
総会発表論文集，vol.36，pp.34-35（1994）
- [7] 日本教育工学会『教育工学辞典』実教出版株式会社，pp.137-277（2000）
- [8] 西森敏之：「大学生の授業における態度と数学教師の対策—日本数学会のあ
る調査より—」，高等教育ジャーナル—高等教育と生涯学習—，vol.6，pp.1-31
（1999）

- [9] 井上聡, 平林茂:「授業改善策として学生に行う小テストの効果」, 帝京科学大学紀要, Vol.13, pp.25- 30 (2017)
- [10] Moodle.org: “open-source community-based tools for learning” , <http://moodle.org>, 2014 参照.
- [11] 鈴木久男, 武貞正樹, 引原俊哉, 山田邦雅, 細川敏幸, 小野寺彰:「授業応答システム クリッカー による能動的学習授業: 北大物理教育での一年間の実践報告」, 高等教育ジャーナル, Vol.16, pp.1-17 (2008)
- [12] 柳澤幸江, 鈴木敏和, 藤澤由美子, 金子健彦, 後藤政幸, 櫻井洋一, 中島肇, 古畑公, 湊久美子, 大河原悦子, 杉浦令子, 多賀昌樹, 高橋佳子, 登坂三紀夫, 松井幾子, 本三保子, 難波秀行:「クリッカーシステムを取り入れた管理栄養士国家試験対策および管理栄養士教育向上に関する取り組み」, 和洋女子大学紀要, vol.53, pp.143-150 (2016)
- [13] 村山航:「テスト形式が学習方略に与える影響」, 教育心理学研究, Vol.51, pp.1-12 (2003)
- [14] 須田昂宏:「リアクションペーパーの記述内容をデータとしてどう活用するかー研究動向の検討を中心にー」, 教育論叢, vol.58, pp.19-34 (2015)
- [15] 小野田亮介, 篠ヶ谷圭太:「リアクションペーパーの記述の質を高める働きかけー学生の記述に対する授業者応答の効果とその個人差の検討ー」, 教育心理学研究, vol.62, pp.115-128 (2014)
- [16] 石岡恒憲:「記述式テストにおける自動採点システムの最新動向」, 行動計量学, Vol. 31, No.2, pp.67-87 (2004)
- [17] 高瀬治彦, 大井健太郎, 森田直樹, 北英彦, 林照峯:「記述式小テストの解答を自動分類するための特徴抽出」, CIEC, Computer&Education, Vol.22, pp.54-59 (2007)
- [18] 大庭知也, 高瀬治彦, 川中普晴, 鶴岡信治:「多人数クラスにおける記述式小テストを支援するシステムー学生の理解状況をすばやく把握するためのインターフェイスー」, Computer&Education, vol.39, pp.86-91 (2015)

- [19] 石岡恒憲, 亀田雅之, 劉東岳:「人工知能を利用した短答式記述採点支援システムの開発」, 研究報告自然言語処理, vol.229, No.16, pp.1-6 (2016)
- [20] 岩佐 靖夫:「日本語教育における類義語指導の一考察—系統的な指導原理へ向けての提言—」, 尚美学園大学総合政策研究紀要, vol.20, pp.17-24 (2011)
- [21] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean:”Efficient Estimation of Word Representations in Vector Space”, ICLR, arXiv:1301.3781 (2013)
- [22] Kyoko Kanzaki, Francis Bond, Noriko Tomuro, Hitoshi Isahara :”Extraction of Attribute Concepts from Japanese Adjectives.”, In LREC-2008, Marrakech.
- [23] Xin Rong, ”word2vec Parameter learning explained”, in arXiv preprint, arXiv:1411.2738 (2016)
- [24] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>, 2016年3月8日参照
- [25] 工藤拓, 松本裕治:「チャンキングの段階適用による日本語係り受け解析」, 情報処理学会論文誌, pp.1834-1842 (2002)
- [26] 大野晋, 浜西正人『類語新辞典』角川書店 (1981)
- [27] weblio 類語辞典, <https://thesaurus.weblio.jp/category/wrugj>, 2017-12 参照.
- [28] 山田敏弘『国語教師が知っておきたい日本語文法』くろしお出版, pp.28-29 (2004)
- [29] 町田健, 井上優, 荒木邦起『日本語文法のしくみ』研究社, pp.88-89 (2012)
- [30] 三省堂編修所『何でもわかることばの知識百科』三省堂 (1995)

発表論文

- [1] 伊藤慎治, 他 : 「記述式解答群の主要な内容の把握支援 —三段階表示の効果の検証—」, 2016 PC Conference, pp.43-44
- [2] 伊藤慎治, 他 : 「三段階表示による記述式解答群の主要な内容の把握支援 - 類似表現の集約による改善-」, 2017 PC Conference, pp.45-48
- [3] Shinji Ito, et al.: Supporting system for descriptive quiz in large class — effectiveness of the three-step-view system—, *Procedia Computer Science*, Vol.96, pp.1166-1171,2016
- [4] Shinji Ito, et al.: Discussion on Judging Synonyms for Classification of Descriptive Answers, *The 7th International Symposium for Sustainability by Engineering at Mie University (IS2EMU2017-C)*, A-1, 2017