

複数言語モデルによる  
文章表現の堅さ推定手法の検討

平成29年度 卒業論文

日本大学 文理学部 情報科学科 宮田研究室

中村 仁汰

# 概要

世の中には様々な表現の文章がある。一般の人が文章を書く場合、文章を書く相手や場面によって文章表現を変える必要がある。例えば、友人に送るメールと上司に送るメールや、手紙と論文では文章表現が異なる。しかし、文章作成に熟練していないユーザは、文章にどれくらい堅い表現、または柔らかい表現を使えば良いか、判断するのが難しいと考える。さらに、現在の多くの対話型エージェントは、人間ほどには文章表現を考慮していないせいで、ユーザが堅い表現の質問をしても、柔らかい表現で返答してしまい、場面にふさわしくない会話が行われることが少なくない。そこで我々は、この問題を解決するために、文章表現の堅さを推定する手法を提案する。この手法では、まず、表現が堅い文章群と柔らかい文章群のそれぞれを Word2Vec で分析して、堅い言語モデルと柔らかい言語モデルを構築する。そして、これらの各言語モデルを用いて測定した推定対象文章中の各単語間の概念距離分布を特徴量とした回帰分析により、推定対象文章の表現の堅さレベルを推定する。本稿では文章表現の堅さレベルを推定するプロトタイプシステムの作成、および検証実験を行う。提案システムが推定した文章表現の堅さレベルと、人手で判定した文章表現の堅さレベルの相関係数を測定する実験を行なったところ、0.446 程度という結果が得られ、提案手法に一定の有効性が確認できた。本稿の貢献は次の通りである。

- 複数言語モデルで測定した各単語間の概念距離の頻度分布を利用し、文章表現の堅さを推定する手法を提案した点
- 上記提案方式の推定精度の検証実験を実施した点

# 目次

<b>第1章</b>	<b>序論</b>	<b>1</b>
1.1	研究の背景	2
1.2	研究の目的	2
1.3	本論文の構成	2
<b>第2章</b>	<b>研究事例</b>	<b>4</b>
2.1	現代における日本語表現に関する研究事例	5
2.2	日本語非母語話者を対象とした研究事例	5
2.2.1	日本語非母語話者向けの日本語教育支援についての研究事例	5
2.2.2	日本語非母語話者と日本語母語話者が用いる文章表現の相違についての研究事例	5
2.3	日本語における文章表現の難易度推定に関する研究事例	6
2.4	日本語における言葉の言い換えに関する研究事例	6
<b>第3章</b>	<b>研究課題</b>	<b>8</b>
3.1	問題の定義	9
3.2	研究課題の設定	10
<b>第4章</b>	<b>複数言語モデルによる 文章表現の堅さ推定手法の提案</b>	<b>11</b>
4.1	アプローチ	12
4.2	複数言語モデルによる文章表現の堅さ推定手法の提案	13
<b>第5章</b>	<b>複数言語モデルによる 文章表現の堅さ推定手法の実装</b>	<b>16</b>
5.1	システムの全体像	17
5.2	言語モデル部の実装	17
5.3	文章表現の堅さレベル推定部の実装	17
<b>第6章</b>	<b>評価実験</b>	<b>20</b>
6.1	品詞群パターン別の堅さレベル推定精度の検証	21
6.1.1	実験の目的	21
6.1.2	データセット	21

---

6.1.3	実験の手順	23
6.1.4	実験結果・考察	23
6.2	コーパスサイズ別の堅さレベル推定精度の検証	24
6.2.1	実験の目的	24
6.2.2	実験の手順	24
6.2.3	実験結果・考察	24
<b>第7章</b>	<b>結論</b>	<b>26</b>
	参考文献	28
	付録	31
	研究業績	38

# 目 次

4.1	堅い文章を堅い言語モデルで測定した場合の単語間概念距離の頻度分布 (概念図)	13
4.2	堅い文章を柔らかい言語モデルで測定した場合の2単語間概念距離の頻度分布 (概念図)	14
4.3	柔らかい文章を堅い言語モデルで測定した場合の2単語間概念距離の頻度分布 (概念図)	14
4.4	柔らかい文章を柔らかい言語モデルで測定した場合の2単語間概念距離の頻度分布 (概念図)	15
5.1	堅い文章を堅い言語モデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)	18
5.2	堅い文章を柔らかいモデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)	18
5.3	柔らかい文章を堅いモデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)	18
5.4	柔らかい文章を柔らかいモデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)	19
6.1	相関係数推移	25

# 表 目 次

6.1	データセット内訳 . . . . .	21
6.2	品詞群パターン別相関係数 (上位 10 パターン) . . . . .	23
6.3	品詞群パターン別相関係数 (下位 10 パターン) . . . . .	24
A.1	品詞群パターン別相関係数 . . . . .	32

# 第1章 序論

## 1.1 研究の背景

世の中には様々な表現の文章がある。一般の人が文章を書く場合、文章を書く相手や場面によって文章表現を変える必要がある。例えば、友人に送るメールと上司に送るメールでは文章の表現が異なるように、文章を書く相手によって文章表現が異なる。また、手紙と論文でも使われている表現が異なるように、文章を書く場面によって文章表現が異なる。このように一般の人が文章を書く場合、自分の知識・経験に基づき、相手・場面に適切と思われる堅さの文章表現を用いる。しかし、学生や外国人のような文章作成に熟練していないユーザは、知識・経験が浅いため、どれくらい堅い表現を使えば失礼にあたらぬか、または、どれくらい柔らかい表現を使えば良いか、を判断することは難しいと考える。このような問題を解決するために、自動で文章表現の堅さを判別してくれるシステムが必要であると考えられる。

また、現在の対話型エージェントは、人間ほどには文章表現を考慮していないせいで、ユーザが堅い表現の質問をしても、柔らかい表現で返答してしまい、場面にふさわしくない会話が行われることが少なくない。もし、文章表現の堅さを判別してくれるシステムを対話型エージェントに導入できれば、ユーザが堅い表現の質問をしたら、堅い表現の返答をする、といった文章表現を考慮して場面に合った対話が実現すると考える。さらに、文章表現の堅さを判別してくれるシステムを [1] や [2] のボケる対話型エージェントに導入することで、堅い表現の質問文に対してはボケないで真面目な返答をし、柔らかい表現の質問文の時にのみボケを挟むといった活用法も見いだせる。

## 1.2 研究の目的

日本語における文章表現の堅さ推定システムは数多く存在する。しかし、それらの多くは、教科書コーパスを用いて漢字の出現頻度や文法に基づく具体的なルールを構築し、文章・単語の難易度を推定している。そのため、ルールから外れる文章(例: 学校教育範囲外の文章)は正しく推定されないという問題がある。また、単語の意味を考慮して文章表現の堅さを推定するシステムは少ないため、漢字が簡単でも意味が難しい語(例: 具に)は正しく推定されない問題もある。そこで本研究では、具体的なルールを構築せず、単語の意味を考慮し、文章表現の堅さを正しく推定することを研究目的とする。

## 1.3 本論文の構成

本論文の構成は次のとおりである。

2章では、日本語教育による研究事例や、日本語の文章表現の難易度推定などに関する研究事例について紹介する。



3章では、既存の文章表現の難易度推定システムの問題点について述べ、それらを踏まえ、本研究における課題を設定する。

4章では、言語モデルを用いた文章表現の堅さ推定手法の提案を行う。

5章では、言語モデルの生成法や、文章表現の堅さ推定手法の実装方法について述べる。

6章では、データセットの作成法や実験手順について述べ、提案手法における推定精度を検証する。

最後に7章にて、本論文の結論を述べる。

## 第2章 研究事例

本章では、日本語表現に関する研究事例について述べる。2.1節では、現代における日本語表現について紹介する。2.2節では、日本語非母語話者を対象とした研究事例について紹介する。2.3節では、日本語における文章表現の難易度推定について紹介する。2.4節では、日本語における言葉の言い換えについて紹介する。

## 2.1 現代における日本語表現に関する研究事例

現代における日本語表現に関する研究は数多く行われている。[3]では、動詞「ある」が持つ20の意味を認定し、それぞれ使用される文章の特徴について検討している。日本語の文章構造について[4]では、文章の構成要素の語についての認定のあり方、その分類について述べている。話し言葉の表現に関する研究として[5]では、体験談話の特徴を明らかにするために、日本語母語話者による体験談話のデータ分析を行っている。そして、そこに現れる接続辞・接続詞を対象に、その機能と使用動機について考察している。書き言葉の表現に関する研究として[6]では、ビジネスメール文について日本語述語項構造解析について検討を行なっている。提案手法に曖昧性が見られたが、今後は敬語表現を考慮することで、提案手法改善を試みると述べている。

## 2.2 日本語非母語話者を対象とした研究事例

### 2.2.1 日本語非母語話者向けの日本語教育支援についての研究事例

日本語非母語話者向けの日本語教育支援についての研究として、[7]、[8]がある。[7]では、「変換」・「解釈」・「翻訳」の3つの段階を踏み、文章を敬語から普通の表現に言い換えることで、日本語非母語話者の敬語理解の支援を目指している。[8]では、日本語非母語話者のために、日本語の文法や語彙に制限をかけて、難しい表現を平易化することで、日本語の理解を支援するシステムの試作を行なっている。

### 2.2.2 日本語非母語話者と日本語母語話者が用いる文章表現の相違についての研究事例

日本語非母語話者と日本語母語話者が用いる文章表現の相違について、研究している事例として、[9]、[10]、[11]がある。[9]では、日本語非母語話者と日本語母語話者の誘い場面における、Eメールの談話構造と表現形式を比較、考察している。日本語非母語話者は、親しい友人にメールを送る場合に、失礼な表現を用いてしまうケースが多く見られることを明らかにしている。[10]では、日本語非母語話者と日本語母語話者の発話語彙の相違を分析している。日本語非母語話者の過剰使用語・過小使用語を分析した結果、日本語母語話者がジャンルに適した言語を用いるのに対して、日本語非母語話者にはそのような傾向は見られなかったことを明らかにしている。[11]では、日本語非母語話者のメール文

において、挨拶表現が日本語母語話者とは異なるものが見られることを明らかにしている。日本語母語話者と日本語非母語話者の挨拶表現の違いを比べたところ、日本人は”ありがとう”と述べることが多いと報告している。

## 2.3 日本語における文章表現の難易度推定に関する研究事例

日本語の文章表現の難易度推定を行う研究は数多く行われている。[12], [13], [14], [15], [16], [17] は教科書を利用して、文章表現の難易度推定を行なっている。[12] では、小学校、中学校、高校、大学の教科書からコーパスを作成し、文章難易度レベルを推定している。レベルは13段階(レベル1=小学1年生レベル, レベル12=高校3年生レベル, レベル13=大学生)に分割し、入力された文章に対して、どの学年の文章難易度レベルかを推定している。[13] では、小学校、中学校の国語教科書に収められたテキストの構成要素を分析している。その分析結果から、テキスト全体の平仮名の割合と文の平均述語数を独立変数とした、学年判別式を構築している。[14] では、小・中学生の教科書に出現する語と、文の難しさを反映する特徴量の両方を組み込む手法から、文章難易度を推定する手法を提案している。[15] では、教科書と対象年齢が設定されている絵本を用いることにより、幼児から小学生までの対象年齢を、推定する手法を提案している。[16] では、2つのテキストの難易順序判定器を構成し、これを用いてテキスト集合を難易度順にソートする方法を提案している。この手法では、難易度はスコアではなく、ソートされたテキスト集合の相対順位として表現している。また、日本語だけでなく英語教科書からコーパスも作成し、英語の文章難易度推定も行なっている。[17] では、文字 bigram モデルを用いた日本語の難易度推定を行っており、既存システムの文字 unigram を用いた場合よりも、上回る性能が得られたと報告している。

[18], [19], [20] は教科書を利用しない方法で、文章表現の難易度推定を行なっている。[18] では、学生に講義内容の理解を深めてもらうことを目的として、講義のスライド・講師の発話内容から、日本語表現の難易度推定を行なっている。[19] では、旧日本語能力試験出題の基準に基づいて、単語の難易度によって単語の色分け(5段階)を行うエディタを開発している。[20] では、書き言葉に対する複雑さを測定する手法を提案している。文章中に含まれる内容語の数と、述語を持つ説の数に着目し、語彙密度を計測する方法を示している。

## 2.4 日本語における言葉の言い換えに関する研究事例

言葉の言い換えを行う研究も存在する。[21] では、小学生のための文章読解支援のために、国語辞典を用いた言い換えを行うことで、新聞に出てくる難しい言葉を小学生でも理解可能な表現に言い換えている。[22] では、文章を敬語表現へ言い換えを行うことで、ビジネスの場で文章のやり取りを支援するシステムを構築している。また、係り受け解析により、動詞ごとに主語を取得し、文中に省略されている主語の補完も試みたが、全ての文

章には対応できないと、報告している。[23]では、一般的な人が文章を書く場合、1つの文章においては、一貫して同じ文体を使い続けられることを用いて、文体と難易度を制御しつつ、日本語機能表現(助詞, 助動詞, 接続詞, 形式名詞)を文体に合わせて、適切な表現に言い換える手法を提案している。[24]では、内容の同一性と表現の類似性の観点から、言語生産過程の多様性を評価している。今後の課題の1つとして、内容を捉える尺度と表現を捉える尺度を分離することで、人が“内容が一致していると認知できる表現のゆれ”を捉えることを目標としている。[25]ではメールを送る相手から見たユーザの立場、送り先に与えたいユーザの印象などを考慮した、文章提案システムを構築している。今後は様々な文章を研究対象としたり、機械学習を取り入れることにより、文章に対する印象を自動判別するなど、様々な方向で研究が広げられると考えられると報告している。[26]の研究では、ユーザと対話システムの雑談対話において、ユーザの発話内容を元にユーザの情報を収集し分析している。

## 第3章 研究課題

### 3.1 問題の定義

文章作成に熟練していないユーザは、自分の書いた文章が堅すぎて失敗する、または、柔らかすぎて失敗するということに気づかない。そこで、自分の書いた文章表現の堅さを自動で判別してくれる仕組みが必要であると考え、日本語の文章難易度推定を行う研究例は、2.3節のとおりである。しかし、それぞれの文章難易度推定手法にはいくつか問題がある。

[12], [13], [14], [15], では、文章を1文字単位で分析し、難しい漢字の出現頻度などに基づいて文章表現の難易度を推定している。このため、平易な文字でも意味が難しい単語は、正しく測定できない。例えば、平易な文字でも意味が難しい語「具に」を既存の文章難易度推定システムの帯[12]を用いて分析した。帯に「具に」と入力すると、難易度は”とてもやさしい”となった。しかし、「つぶさに」と入力すると、難易度は”やさしい”となった。さらに、文章で入力した場合、「具に事情を説明」は”高校1年生レベル”、「つぶさに事情を説明」は”高校2年生レベル”となった。この結果から、既存のシステムでは、漢字の難しさを考慮しているが、単語の意味の難しさは考慮されていないことがわかった。また、この文字単位分析のアプローチは、アルファベットなどの少数の文字のみを用いる語(例:英語)には適用できないと思われる。実際、帯にアルファベットからなる単語を入力したところ、正しく推定されなかった。

[18]では、留学生に講義内容の理解を深めてもらうことを目的に、旧日本語能力試験出題の基準に基づいた、講義資料の単語の難易度推定、講師の発話内容における係り受けの複雑さ・助動詞の用法、の観点から講義内容の複雑さを分析している。このため、講師の発話内容において、文章構造の複雑さは分析しているが、意味の難しい単語などを考慮した分析は行われていない。また、研究対象が非母語話者のため、非母語話者にとっての文章難易度推定には一定の効果を発揮するが、母語話者における文章難易度推定には適さない。[19]では、旧日本語能力試験出題の基準に基づいて、単語の難易度によって単語の色分けを行うエディタを開発している。このため、単語の難易度は推定できるが、文章の難易度は推定できない。[20]では、語彙密度(文中に含まれる内容語の割合)を計算し、書き言葉に対する文章の複雑さの判定をしている。この手法では、内容語(名詞・動詞・形容詞・副詞のような独立して意味をなす語)と機能語(助詞・助動詞・接続詞・代名詞のような単独では意味をなさない語)を厳密に区別していない。内容語と機能語を厳密に区別しないことで、形式名詞を正しく分析できない可能性がある。また、語彙密度を用いて文章難易度推定を行う場合、推定する文章の1文の長さによって、難易度の推定結果が大きく左右される。そのため、1文が短い場合は、難易度が”最も難しい”や”最もやさしい”といった偏った推定結果になりやすく、難易度が正しく推定できない可能性がある。さらに、難しい表現が使われている文章でも、代名詞や連体詞を多く含む文を推定すると、難易度が低くなりやすいという問題がある。

また、[12], [13], [14], [15], [16], [17]では教科書コーパスを用いて文章難易度推定を行なっている。教科書コーパスとは、学年別の教科書を元に作成したコーパスである。このため、教科書コーパスを用いて文章難易度推定を行うことで、次のような問題が生じる

と考えられる。学校教育における教科書の範囲を超えた文章の難易度推定が難しい可能性がある。教科書は学習用の教材のため、普通の文章より堅い表現の文章で書かれている。そのため、柔らかい表現の文章の難易度判断が難しい可能性がある。

以上のように、既存研究は、漢字の出現頻度や文法に基づく具体的なルールを構築して文章・単語の難易度を推定している。このため、問題は次のように定義できる。

問題1：ルールから外れる対象（例：学校教育範囲外の文章，外国語）の文章は正しく難易度推定行われぬ可能性がある

問題2：単語の意味の難しさを考慮した文章難易度推定を行う既存研究は少ない

## 3.2 研究課題の設定

3.1節で定義した問題をふまえ、本研究では、具体的なルールを構築しないで、単語の意味の難しさを考慮した文章難易度推定手法を確立することを研究課題に設定する。



## 第4章 複数言語モデルによる 文章表現の堅さ推定手法の提案

## 4.1 アプローチ

3.2節で設定した課題を達成するために、具体的なルールを用いずに、単語の意味を考慮して文章表現の堅さを推定するというアプローチをとる。そのアプローチを実現するためには、言語モデルを用いることで実現できると考えられる。言語モデルとは、文の品詞、単語と単語、文書と文書などの関係性について定式化したもののことである。そこで本研究では、堅い言語モデルと柔らかい言語モデルの複数言語モデルを用いて、文章表現の堅さを推定する手法を提案する。そして、1つの文章において、原則として、一貫して同じ文体が使い続けられることを用いて3つの仮説を立てた。

- 仮説1：1つのテーマについて書かれている文章中の各単語間の概念距離は近い。
- 仮説2：堅い文章中の各単語間の概念距離を堅い言語モデルで分析すると、正しく分析され、概念距離は近くなる。しかし柔らかい言語モデルは、堅い語の概念をうまく捉えることができない。そのため、堅い文章中の各単語間の概念距離を柔らかい言語モデルで分析すると、概念距離は本来よりも遠いと判断してしまう。
- 仮説3：柔らかい文章中の各単語間の概念距離を堅い言語モデルで分析すると、柔らかい語の概念を上手く捉えられない。このため、柔らかい文章中の各単語間の概念距離を柔らかい言語モデルで分析すると、正しく分析され、概念距離はより近いと判断される。

これらの仮説を「お金」をテーマに書かれている文章を例に用いて順を追って説明する。まずは仮説1について説明する。堅い文章には「金銭」、「給与」、「賞与」などの表現が使われており、各単語間の概念距離は近いと考えられる。同様に、柔らかい文章には「おかね」、「おこづかい」、「おとしだま」などの表現が使われており、各単語間の概念距離は近いと考えられる。次に仮説2について説明する。「金銭」、「給与」、「賞与」などの表現が使われている堅い文章中の各単語間の概念距離を、堅い言語モデルで分析すると、正しく分析され、概念距離はより近いと判断されると考えられる。しかし、柔らかい言語モデルでは「金銭」、「給与」、「賞与」などの堅い単語の概念を、堅い言語モデルほどには上手く捉えられることができない。このため、堅い文章中の各単語間の概念距離を柔らかい言語モデルで分析すると、概念距離は本来よりも遠いと判断してしまうと考えられる。最後に仮説3について説明する。これは先程の仮説2とは反対のパターンである。「おかね」、「おこづかい」、「おとしだま」などの表現が使われている柔らかい文章中の各単語間の概念距離を、堅い言語モデルで分析すると、堅い言語モデルは、「おかね」、「おこづかい」、「おとしだま」などの柔らかい語の概念を、柔らかい言語モデルほどには上手く捉えられることができない。このため、柔らかい文章中の各単語間の概念距離を、堅い言語モデルで分析すると、概念距離は本来より遠いと判断されると考えられる。しかし、柔らかい言語モデルで、柔らかい文章中の各単語間の概念距離を分析すると、正しく分析され、概念距離は近いと判断されると考えられる。

## 4.2 複数言語モデルによる文章表現の堅さ推定手法の提案

提案手法について図 4.1～図 4.4 のグラフを用いて説明する。グラフの縦軸は出現頻度、横軸は各言語モデルで測定した2単語間の概念距離を表している。

4.1節で述べた3つの仮説を元にとすると次のように考えられる。堅い文章を、堅い言語モデルで分析すると正しく分析され、各単語間の概念距離の近いペアの語が多く観測される(図 4.1)。しかし、堅い文章を柔らかい言語モデルで分析すると、柔らかい言語モデルは堅い語の概念を上手く捉えることができないので、正しい分析結果にならず、概念距離は本来よりも遠い位置に観測される(図 4.2)。また、反対のパターンでも同じような結果になる。柔らかい文章を堅い言語モデルで分析すると、堅い言語モデルは柔らかい語の概念を上手く捉えることができないので、正しい分析結果にならず、概念距離は本来よりも遠い位置に観測される(図 4.3)。しかし、柔らかい文章を柔らかい言語モデルで分析すると正しく分析され、各単語間の概念距離の近い語のペアの出現頻度が多く観測される(図 4.4)。

これはつまり、分析対象の文章の堅さの違いは、その文章を各言語モデルで分析した際の各単語間の概念距離の頻度分布の違いとして表出する可能性を示唆している。そこで、堅い言語モデルと柔らかい言語モデルの複数言語モデルで分析した、推定対象文章中の各単語間の概念距離の頻度分布を特徴量として機械学習を行い、文章表現の堅さレベルを推定する手法を提案する。

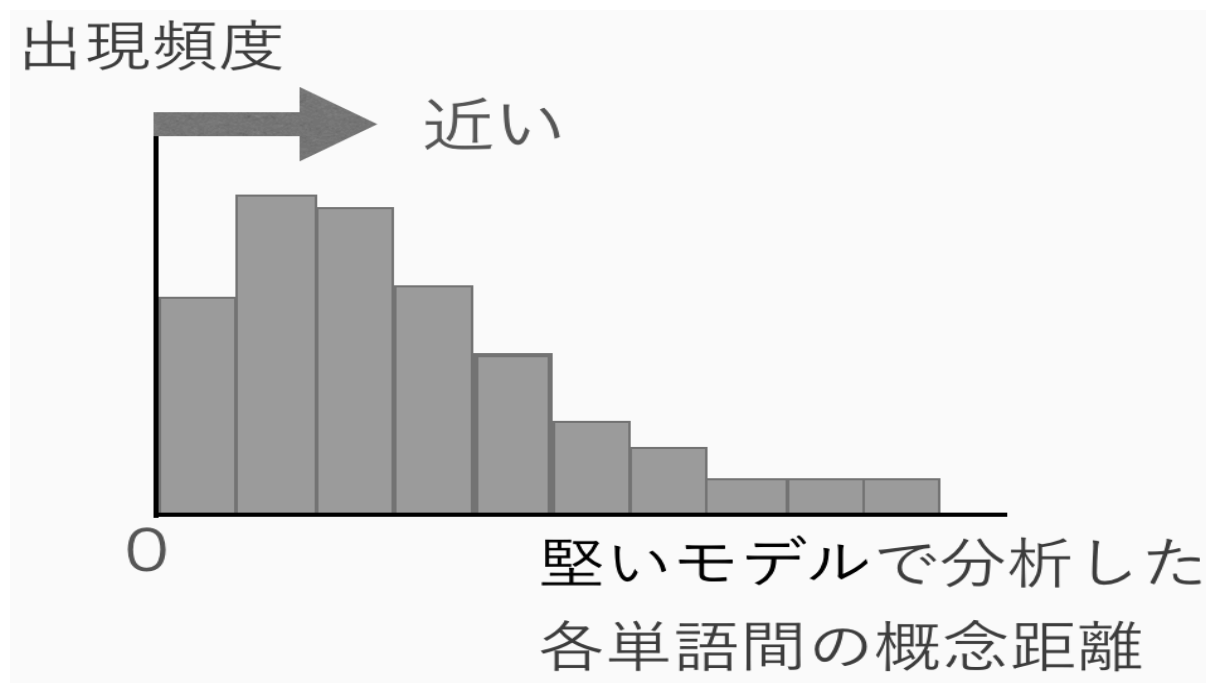


図 4.1: 堅い文章を堅い言語モデルで測定した場合の単語間概念距離の頻度分布 (概念図)

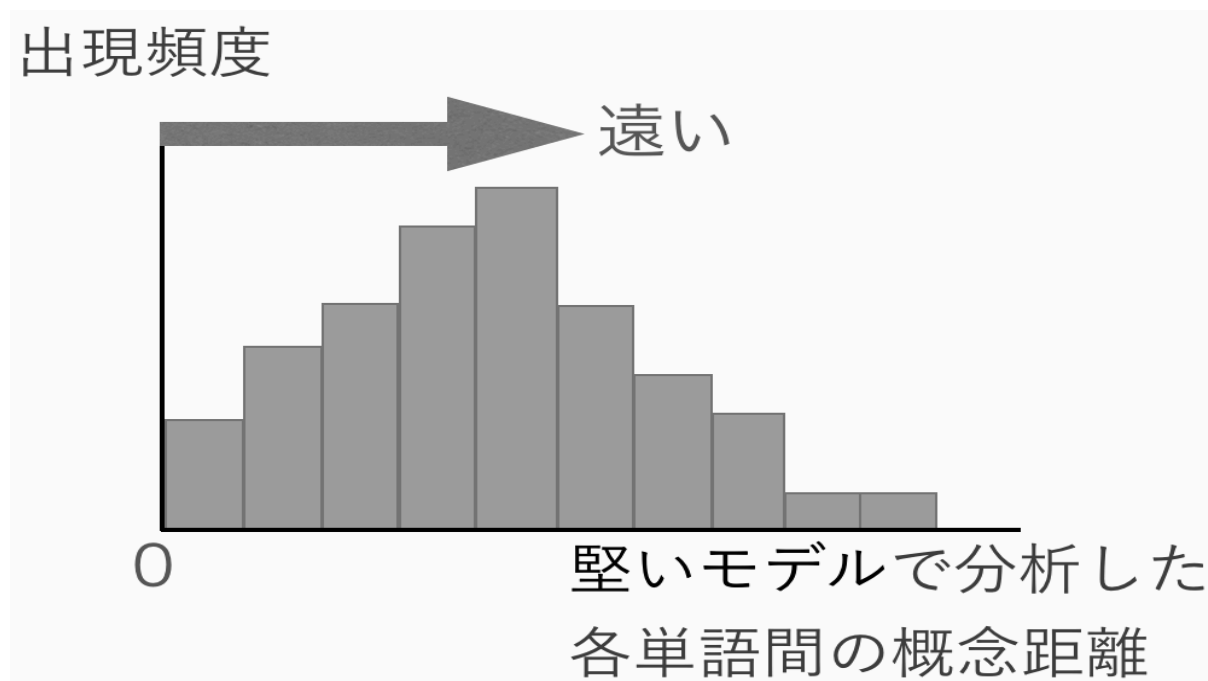


図 4.2: 堅い文章を柔らかい言語モデルで測定した場合の2単語間概念距離の頻度分布 (概念図)

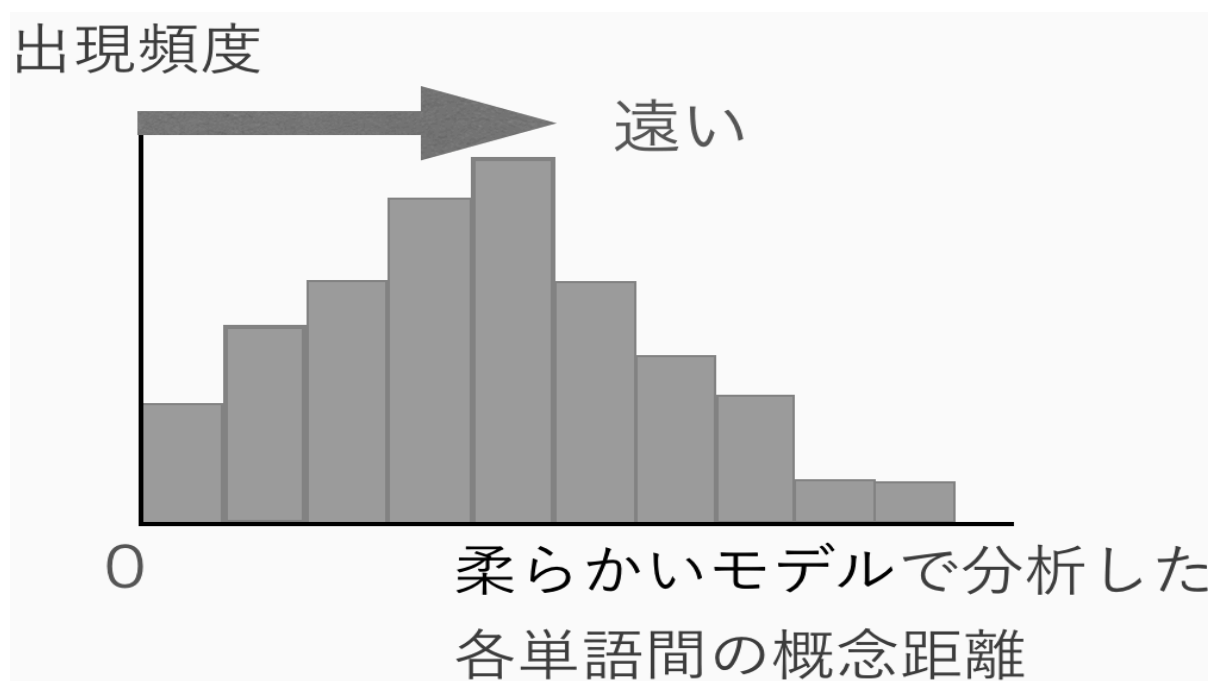


図 4.3: 柔らかい文章を堅い言語モデルで測定した場合の2単語間概念距離の頻度分布 (概念図)

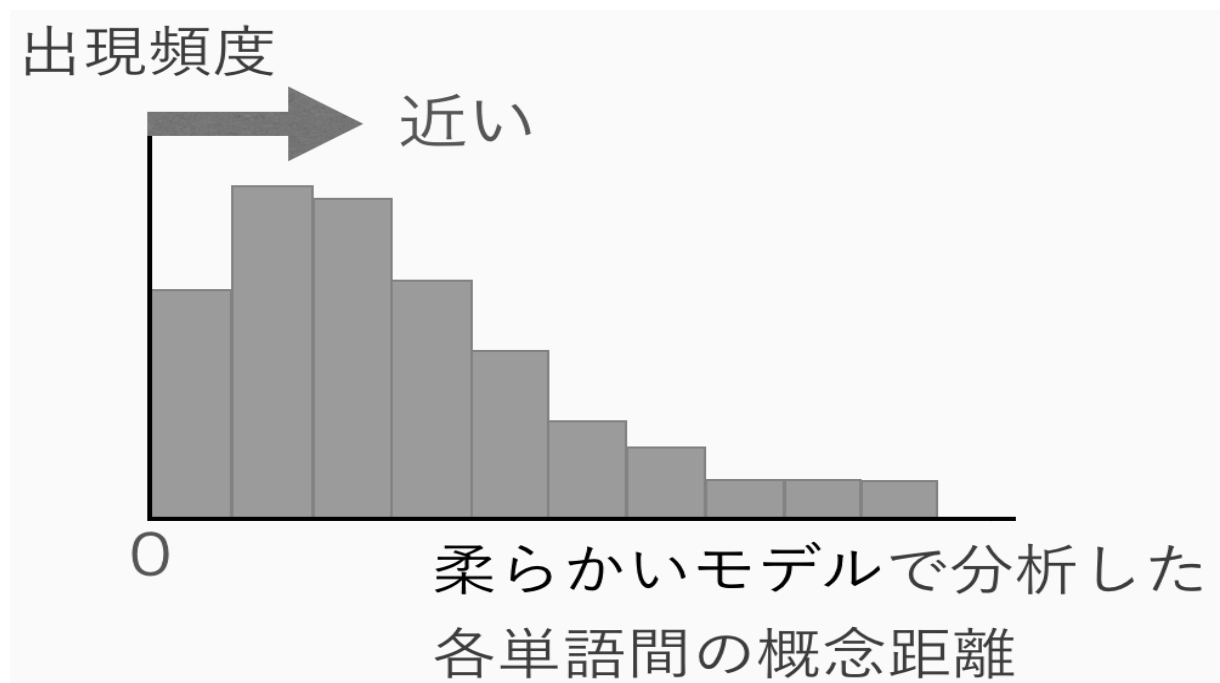


図 4.4: 柔軟い文章を柔軟い言語モデルで測定した場合の2単語間概念距離の頻度分布 (概念図)

## 第5章 複数言語モデルによる 文章表現の堅さ推定手法の実装

## 5.1 システムの全体像

人手で文章表現の堅さレベル付コーパスを作成し、作成したコーパスを MeCab[27] を用いて形態素解析を行い、品詞ごとに分解する。そして、コーパス内の記事1件ごとに、あらかじめ指定した品詞群を抜き出す。どの品詞群が文章表現の堅さレベルの推定精度に影響するかは次節で実験を行う。抜き出した品詞群を合わせたものを1サンプルとし、堅い言語モデルと柔らかい言語モデルで測定する。先ほどの複数言語モデルで測定した、1サンプル中の各単語間の概念距離の頻度分布を特徴量とした回帰分析を行い、文章表現の堅さレベルを推定する。システム全体は Python2.7.11, MeCab のバージョンは 0.996, MeCab の辞書は mecab-ipadic-neologd(最終更新日:2016/6/3) を用いて実装を行う。

## 5.2 言語モデル部の実装

文章表現の堅さを測定するための堅い言語モデル構築に必要なコーパスは、幅広い概念の言葉を含んでいる、かつ、堅い表現で書かれている必要があることから、Wikipedia 日本語記事全文 [28](単語数:960649 語, last visited on 2017/1/31) を選定する。同様に、柔らかい言語モデル構築に必要なコーパスは、幅広い概念の言葉を含んでいる、かつ、柔らかい表現で書かれている必要があることから、幅広い分野(大項目で 97 カテゴリー)の質問文・回答文が投稿されている、教えて goo[29] の質問文 7,290,825 件, 回答文 19,690,004 件 (last visited on 2017/1/31) を選定する。そして Word2Vec[30](次元数:100, ライブラリ:gensim, バージョン:0.12.4, Window サイズ:15) を用いて各言語モデルを作成する。Word2Vec とは、自然言語処理手法の1つで、単語を比較的低次元の空間ベクトル表現することで、単語の意味を捉えられるものである。

## 5.3 文章表現の堅さレベル推定部の実装

文章を 5.2 節で実装した堅い言語モデル・柔らかい言語モデルを用いて測定する。今回の実装において、各言語モデルで測定した 2 単語間の概念距離を合計 30 分割する。そのため、堅い言語モデルで測定する場合の距離を 15 分割、柔らかい言語モデルで測定する場合の距離を 15 分割する。つまり、堅い文章を堅い言語モデルで分析する場合は、概念距離が近いところに多く観測され、図 5.1 のようになる。また堅い文章を柔らかい言語モデルで分析する場合は、本来よりも遠い位置に多く観測され、図 5.2 のようになる。反対に柔らかい文章を堅い言語モデルで分析する場合は、本来よりも遠い位置に多く観測され、図 5.3 のようになる。しかし、柔らかい文章を柔らかい言語モデルで分析する場合は、概念距離が近いところに多く観測され、図 5.4 のようになる。そして、概念距離の頻度分布を特徴量とした線形回帰分析を行い、文章表現の堅さレベルを推定する。回帰式は、サポートベクター回帰 (SVR) を用いる。ライブラリは sklearn でメソッドは svm, メソッドに受け渡すパラメータは svm.SVR(C = .5) となっている。

出現頻度

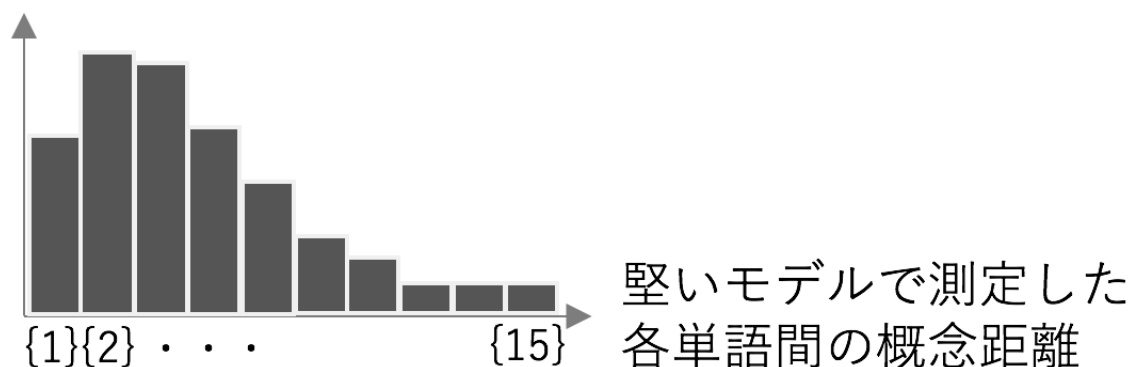


図 5.1: 堅い文章を堅い言語モデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)

出現頻度

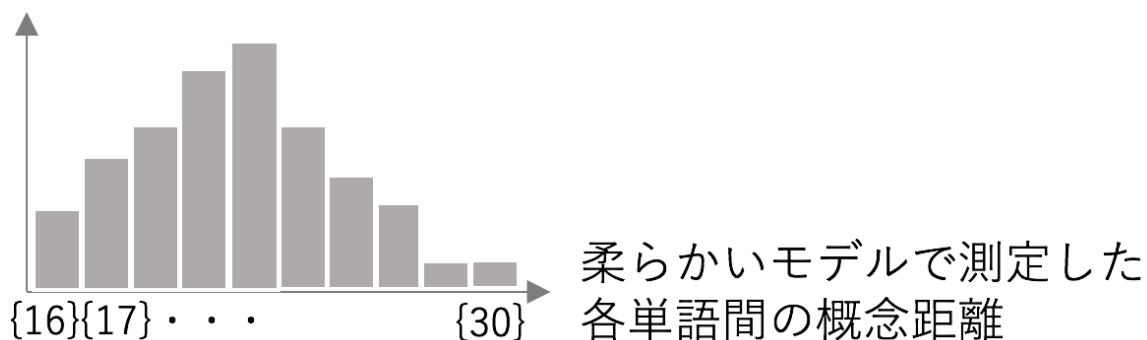


図 5.2: 堅い文章を柔らかいモデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)

出現頻度

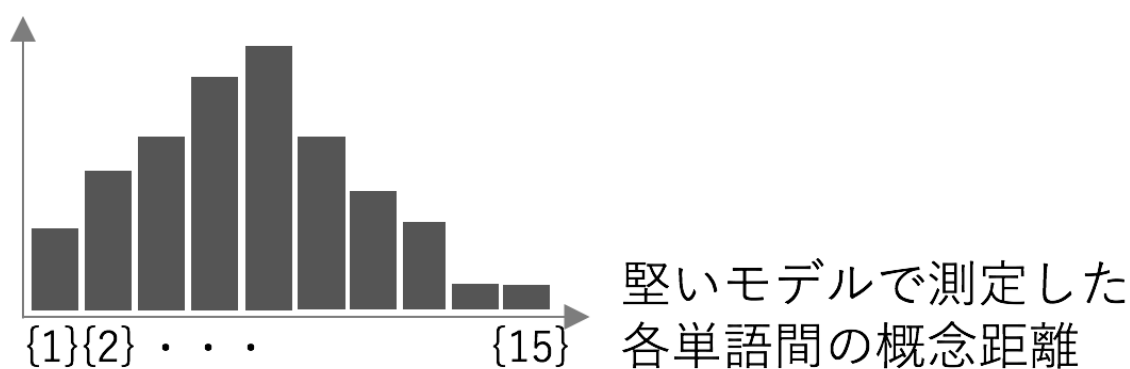


図 5.3: 柔らかい文章を堅いモデルで測定し、2単語間概念距離を15分割した場合の頻度分布 (概念図)



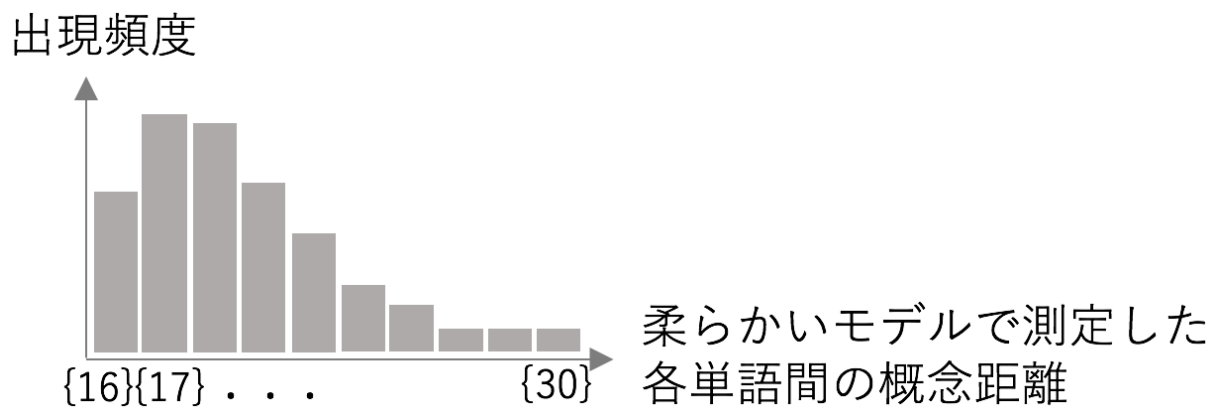


図 5.4: 柔軟い文章を柔軟いモデルで測定し、2 単語間概念距離を 15 分割した場合の頻度分布 (概念図)

## 第6章 評価実験

## 6.1 品詞群パターン別の堅さレベル推定精度の検証

### 6.1.1 実験の目的

提案手法における文章表現の堅さレベルの推定精度を検証する。そのために、どの品詞群パターンが文章表現の堅さレベルの推定精度に影響するかを検証し、分析する。

### 6.1.2 データセット

NHK NEWS WEB[31], Nikkei Style[32], Yahoo知恵袋 [33], LINE Q[34] の4つのWebサイト内から、生活・健康・人生設計などに関する記事を、無作為に3876件抽出して、人手で文章表現の堅さのレベル付けを行ない、コーパスを作成した。コーパス収集したWebサイトごとの記事の内訳は表6.1の通りになっている。レベル付けは、5段階評価を大学生3人が独立に行い、評価平均値をスコアとした。また、記事の例は下記の通りになっている。文章作成者の権利に配慮して固有名詞を排除し、文意・文体を損ねない程度に修正したものを記載する。記事例は文章表現が堅い順に、記事1、記事2、記事3、記事4となっている。

表 6.1: データセット内訳

Web サイト名	記事件数
NHK NEWS WEB	803 件
Nikkei Style	1073 件
Yahoo 知恵袋	1095 件
LINE Q	805 件
合計	<b>3876 件</b>

#### 記事 1(NHK NEWS WEB から引用)

政府内では、大学など高等教育の授業料について、在学中は支払わず卒業後に所得に応じて返還してもらう制度が有力な選択肢として浮上っていて、制度設計に向けた議論が加速する見通しです。この制度は、大学に在学中は授業料を支払わず、卒業後、一定以上の所得が得られるようになった際に、所得に応じて授業料を返還してもらう仕組みで、政府内では「家庭の事情に左右されず希望する教育が受けられる」と評価する意見が出ています。また政府内では、来年度から本格実施される所得が少ない世帯などを対象にした返済の要らない給付型奨学金を充実させる案も検討されていて、今後、財源も含め制度設計に向けた議論が加速する見通しです。

### 記事 2(Nikkei Style から引用)

奨学金は、返還の必要がない「給付型」と、返済義務がある「貸与型」に大別されます。奨学金は従来、「貸与型」だけでした。具体的には、無利子の「第一種」、有利子の「第二種」、そして入学時の一時金にあてるための「特別増額貸与」（有利子）の3タイプで、第一種と第二種の併用も可能です。希望すれば誰でも受給できるわけではなく、学力が一定以上であるとか、収入が一定以下（たとえば世帯3人の給与所得者の場合、第一種で696万円、第二種で1036万円が上限額の目安）といった基準があり、実際に受けられるかどうかは審査によって決まります。また、貸与型奨学金は、保護者が借り入れて返済を行う「教育ローン」とは異なり、学生自身が返済するため、本人がそのことを自覚し、将来の人生設計などを考えたうえで申し込むことも必要です。

### 記事 3(Yahoo 知恵袋から引用)

奨学金について質問です。大学へ進学しようと思っているのですが、下に高2と来年から高校へ入学する妹もおり、経済的には多分厳しいと思っています。なので、奨学金を使って進学しようと思っていますが奨学金は全ての授業料（その他）を支給してくれるのでしょうか？それとも10万（仮）支給し残りの授業料等は自分で払うのでしょうか？入学金は自分でバイトして支払うつもりでいます。なるべく親にも負担をかけさせたくはありません。誰か奨学金に詳しい人ご回答よろしくお願ひします。

### 記事 4(LINE Q から引用)

大学生です。私はバイトをしてません。大学は私学ですが奨学金もあるし親からの仕送りもあるので金銭的には困ってません。親は大学は勉強するところだからバイトはしなくていいよって言ってくれました。行きたいところがあるなら奨学金を使ってでも行きなさい。返済で困ったら助けてあげるからって親が言ってくれました。なので、私はバイト以外のことをしようと決めました。でも、バイトをしている友達が、そんなのあり得ないとか、バイトした方が絶対いいよ！とか、親の金なのによくそんな使えるなって言ってきます。友達の意見が正しいのですかね？

今回は記事ごとの文字数に偏りが生じないように、記事1件の文字数は400～700文字とした。また、記事のジャンルを政治や経済などにすると、文章表現が堅い記事しか収集できない。そのため、今回はジャンルを文章表現が堅い記事と柔らかい記事の両方が収集できる、生活・健康・人生設計などに設定した。

### 6.1.3 実験の手順

コーパスから MeCab を用いて形態素解析を行い、品詞ごとに分解し、あらかじめ指定した品詞群を抜き出し、合わせたものを1サンプルとする。3876件あるサンプルを、学習データ3826件とテストデータ50件にランダムに分割する。そして、堅い言語モデルと柔らかい言語モデルを用いて測定した、1サンプル中の各単語間の概念距離分布を説明変数とした重回帰分析を行い、文章表現の堅さレベルを推定する。品詞パターンごとに重回帰分析による推定する試行を100回実施し、重回帰分析によって推定された文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数を測定する。品詞パターンは「名詞 動詞 形容詞 副詞 連体詞 接続詞 助動詞 助詞」の8つの品詞の組み合わせパターン(8C1 + 8C2 + 8C3 + 8C4 + 8C5 + 8C6 + 8C7 + 8C8 = 255通り)の検証を行う。

### 6.1.4 実験結果・考察

本システムで推定された文章表現の堅さと、人手でレベル付けして文章表現の堅さのレベルの相関係数を測定したところ、最も相関係数が大きかったもので0.446となった。抜き出す品詞群パターンと相関係数の上位10パターンと下位10パターンは表6.2・表6.3となった。(全品詞群パターンは付録の表A.1参照)

表 6.2: 品詞群パターン別相関係数(上位10パターン)

	品詞群パターン	相関係数
1	動詞 形容詞 副詞 助動詞 助詞	0.446
2	動詞 形容詞 副詞 接続詞 助動詞 助詞	0.436
3	動詞 形容詞 助動詞 助詞	0.434
4	形容詞 助動詞 助詞	0.434
5	形容詞 副詞 助動詞 助詞	0.433
6	副詞 連体詞 助動詞 助詞	0.432
7	動詞 連体詞 接続詞 助動詞 助詞	0.432
8	副詞 接続詞 助動詞 助詞	0.431
9	動詞 助動詞 助詞	0.427
10	動詞 副詞 助動詞 助詞	0.421

一番相関係数が大きかったパターンは動詞，形容詞，副詞，助動詞，助詞となり，相関係数は0.446となった。一番相関係数が小さかったパターンは副詞，接続詞となり，相関係数は0.118となった。このことから，品詞群パターンは文章表現の堅さレベルの推定精度に影響すると考えられる。また，上位10パターンに助動詞，助詞が含まれていることから，助動詞，助動詞を含む組み合わせは推定精度が高くなりやすいと考えられる。

表 6.3: 品詞群パターン別相関係数 (下位 10 パターン)

	品詞群パターン	相関係数
246	動詞	0.274
247	動詞 連体詞	0.271
248	動詞 副詞 接続詞	0.269
249	形容詞 接続詞	0.256
250	動詞 接続詞	0.246
251	動詞 連体詞 接続詞	0.241
252	副詞 連体詞	0.214
253	副詞 連体詞 接続詞	0.161
254	副詞	0.151
255	副詞 接続詞	0.118

## 6.2 コーパスサイズ別の堅さレベル推定精度の検証

### 6.2.1 実験の目的

コーパスサイズが文章表現の堅さ推定手法に影響を与えるか検証する。

### 6.2.2 実験の手順

6.1 節の実験で相関係数が一番大きい、動詞、形容詞、副詞、助動詞、助詞の品詞群パターンで検証を行う。また、コーパスは 6.1.2 項で作成したものをを用いる。コーパスサイズは 500 件、1000 件、1500 件、2000 件、2500 件、3000 件、3500 件、3876 件に分けて実験を行う。そして先程と同様に、重回帰分析による推定する試行を 100 回実施し、重回帰分析によって推定された文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数を測定した。

### 6.2.3 実験結果・考察

コーパスサイズ別に相関係数を測定したところ、図 6.1 のようになった。図 6.1 から、コーパスサイズ拡大につれて、相関係数が大きくなっていることがわかる。ここから、コーパスサイズ拡大するにつれて精度向上しているのがわかる。さらに、飽和状態ではないことから、今後も引き続きコーパス収集を行うことで、精度向上に繋がると考えられる。

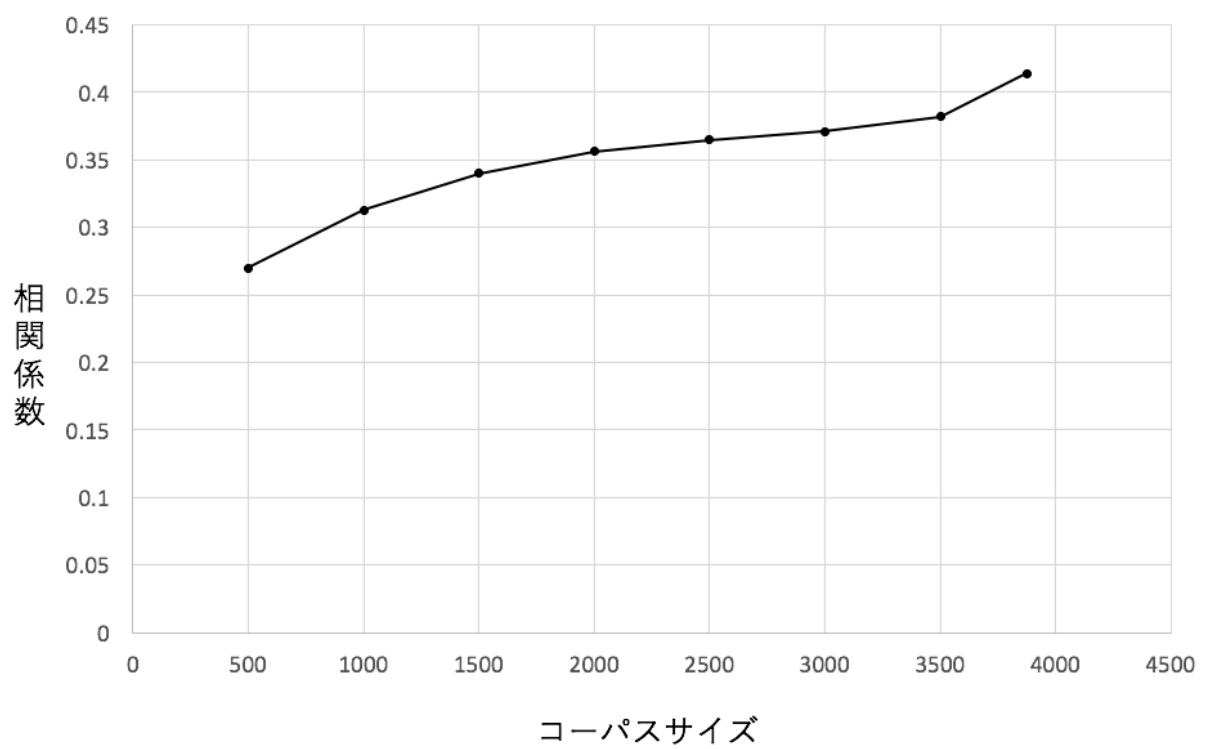


図 6.1: 相関係数推移

## 第7章 結論



学生のように文章作成に熟練していないユーザは、知識・経験が浅いため、どのくらい堅い表現、または柔らかい表現を使えば良いか判断するのが難しいと考えられる。そこで本研究では、文章表現の堅さを推定する手法を提案し、その推定精度を検証した。この手法では、表現が堅い文章群（Wikipedia）と柔らかい文章群（教えてgooの質問文・回答文）のそれぞれを Word2Vec で分析して、堅い言語モデルと柔らかい言語モデルを構築する。そして、これらの各言語モデルを用いて測定した、推定対象文章中の各単語間の概念距離分布を特徴量とした回帰分析により、推定対象文章の表現の堅さレベルを推定する。実験の結果、本システムが推定した文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数は 0.446 程度となり、一定の正の相関が見受けられた。今後も引き続きコーパス収集を行うことで、精度向上に繋がると考えられる。また、品詞群パターン別に相関係数を測定し、どの品詞が文章表現の堅さレベルの推定精度に影響するかを検証したところ、品詞群パターンは推定精度に影響することが分かり、最も推定精度が高くなるのは動詞、形容詞、副詞、助動詞、助詞である。さらに、相関係数が大きい上位 10 パターンには助動詞、助詞が含まれていることから、助動詞、助詞を含む品詞群パターンは相関係数が大きくなりやすいと考えられる。

本研究の期待される活用法として、ユーザと対話型エージェントのコミュニケーションの質の向上が挙げられる。例えば、NTT レゾナント社の教えてgooにはユーザからの質問文を AI が自動で返答する機能がある。このような質問掲示板で、ユーザの質問文が柔らかい表現ならば、AI も柔らかい表現の返答文を選び、反対にユーザの質問文が堅い表現ならば、AI も堅い表現の返答文を選ぶといった利用方法が考えられる。[35]によると、質問者は質問文より文章表現が堅い回答文の方が受け入れやすくなることから、ユーザの質問文より AI 返答文の文章表現を堅くすることで、ユーザが返答文を受け入れやすくなるといった利用方法も考えられる。

## 参考文献

- [1] 鈴木奨, 呉健朗, 瀧田航平, 堀越和, 中辻真, 宮田章裕. ボケて返す対話型エージェントの基礎検討. 情報処理学会研究報告 (GN), 第 117 巻, pp. 11–16, 2017.
- [2] 瀧田航平, 鈴木奨, 呉健朗, 堀越和, 中辻真, 宮田章裕. キャラクター性を持ったボケて返す対話型エージェントの基礎検討. 情報処理学会グループウェアとネットワークサービスワークショップ 2017 論文集, pp. 1–6, 2017.
- [3] 野田大志. 現代日本語における動詞「ある」の多義構造. 国立国語研究所論集, 第 12 巻, pp. 81–100, 2017.
- [4] 中崇, 城田俊. 日本語における語の認定と品詞分類をめぐって—日本語教師のための日本語文法をもとめて—. 就実論叢, 第 46 巻, pp. 46–63, 2017.
- [5] 加藤陽子. 日本語母語話者の体験談の語りについて: 談話に現れる事実的な「タラ」「ソシタラ」の機能と使用動機. 世界の日本語教育: 日本語教育論集, 第 13 巻, pp. 57–74, 2003.
- [6] 平博順, 田中貴秋, 藤田早苗. ビジネスメール文に対する日本語述語構造解析の検討. 言語処理学会第 20 回年次大会発表論文集, 第 20 巻, pp. 1019–1022, 2014.
- [7] 李国慶, 吉野孝. 外国人向け敬語文理解支援システムの開発. 情報処理学会第 77 回全国大会, 第 114 巻, pp. 7–12, 2015.
- [8] 杵真奈見, 山本和英. 「やさしい日本語」変換システムの試作. 言語処理学会第 19 回年次大会, pp. 678–681, 2013.
- [9] 嶋田みのり. 日本語の「誘い」場面における eメールの談話構造と表現形式: 母語話者と中国人学習者の分析を通じて. 創価大学大学院紀要, 第 35 巻, pp. 217–242, 2013.
- [10] 小西円. 日本語学習者と母語話者の産出語彙の相違: I-jas の異なるタスクを用いた比較. 国立国語研究所論集, 第 13 巻, pp. 79–106, 2017.
- [11] 金庭久美子, 金玄珠. メール文における挨拶表現-韓国における日本語学習者のメール文調査から-. 横浜国大言語研究, 第 35 巻, pp. 138–150, 2017.
- [12] 佐藤理史. 均衡コーパスを規範とするテキスト難易度測定. 情報処理学会論文誌, Vol. 52, No. 4, pp. 1777–1789, 2011.

- [13] 柴崎秀子, 玉岡賀津雄. 国語科教科書を基にした小・中学校の文章難易学年判定式の構築. 日本教育工学会論文誌, Vol. 33, No. 4, pp. 449–458, 2010.
- [14] 藤田早苗, 藤野昭典, 小林哲生. 教科書を規準とする難易度推定. 人工知能学会全国大会論文集, 第 29 卷, pp. 1–4, 2015.
- [15] 藤田早苗, 小林哲生, 平博順, 南泰浩, 田中貴秋. 絵本を基にした対象年齢推定方法の検討. 人工知能学会全国大会論文集, 第 28 卷, pp. 1–4, 2014.
- [16] 手塚智史, 寺田博視, 田中久美子. 相対的観点に基づく類似難易度文書検索システムの構築. 言語処理学会第 15 回年次大会, 第 15 卷, pp. 288–291, 2009.
- [17] 小島健輔, 佐藤理史, 藤田篤. 文字 bigram モデルを用いた日本語テキストの難易度推定. 言語処理学会第 15 回年次大会発表論文集, pp. 897–900, 2009.
- [18] 木藤善信, 木村祐介, 椎名広光. 日本語の単語難易度や複雑度を利用した講義の難易度指標作成. 言語処理学会第 19 回年次大会, pp. 624–627, 2013.
- [19] 柴田大介, 村田進, 北村達也, 川村よし子. 単語難易度判定機能を有するエディタの開発. 日本語教育方法研究会誌, 第 20 卷, pp. 70–71, 2013.
- [20] 佐野大樹, 丸山丘彦. システミック文法に基づく書きことばの複雑さ測定. 言語処理学会第 14 回年次大会発表論文集, pp. 1097–1100, 2008.
- [21] 梶原智之, 山本和英. 語釈文を用いた小学生のための語彙平易化. 情報処理学会論文誌, No. 3, pp. 983–992, 2015.
- [22] 明河直樹, 蒲地俊太郎, 堀智允, 韓東力. 敬語表現への自動変換システムの構築. 言語処理学会第 16 回年次大会, 第 15 卷, pp. 764–767, 2010.
- [23] 松吉俊, 佐藤理史. 文体と難易度を制御可能な日本語機能表現の言い換え. 自然言語処理学会誌, Vol. 15, No. 2, pp. 75–99, 2008.
- [24] 浅原正幸, 加藤祥. 文書間類似度について. 自然言語処理学会誌, Vol. 23, No. 5, pp. 463–499, 2016.
- [25] 紺野倭人, 坂本真樹. メール印象を考慮した文章提案システムの構築に関する研究. 人工知能学会第 31 回全国大会, pp. 1–2, 2017.
- [26] 光田航, 東中竜一郎, 牧野俊朗, 松尾義博. 雑談対話における言外の情報を推定するためのデータ収集と分析. 人工知能学会第 30 回全国大会, pp. 1–4, 2016.
- [27] Yet another part-of-speech and morphological analyzer  
<http://taku910.github.io/mecab/> (last visited on 2017/4/1).

- 
- [28] <https://ja.wikipedia.org/wiki/>(last visited on 2017/1/31).
- [29] <https://oshiete.goo.ne.jp> (last visited on 2017/1/31).
- [30] Tomas Mikolov, Kai Chen, Greg Corrad, and Jeffrey Dean. In *In Proceedings of Workshop at ICLR*, 2013.
- [31] <http://www3.nhk.or.jp/news/> (last visited on 2017/9/13).
- [32] <http://style.nikkei.com> (last visited on 2017/9/13).
- [33] <https://chiebukuro.yahoo.co.jp/> (last visited on 2017/9/13).
- [34] <http://lineq.jp/> (last visited on 2017/9/13).
- [35] 栗山和子, 神門典子. Q & A サイトにおける質問と回答の分析 (2) 一文の構造と属性を中心に. 研究報告情報学基礎, 2009.

# 付録

6.1 節の実験結果は表 A.1 である。測定不能の場合はテストデータ内に品詞群パターンの品詞が含まれていない可能性がある。

表 A.1: 品詞群パターン別相関係数

	品詞群パターン	相関係数
1	名詞	0.403
2	動詞	0.274
3	形容詞	0.31
4	副詞	0.152
5	連体詞	測定不能
6	接続詞	測定不能
7	助動詞	0.331
8	助詞	0.389
9	名詞 動詞	0.335
10	名詞 形容詞	0.331
11	名詞 副詞	0.367
12	名詞 連体詞	0.39
13	名詞 接続詞	0.387
14	名詞 助動詞	0.368
15	名詞 助詞	0.367
16	動詞 形容詞	0.293
17	動詞 副詞	0.308
18	動詞 連体詞	0.271
19	動詞 接続詞	0.246
20	動詞 助動詞	0.319
21	動詞 助詞	0.34
22	形容詞 副詞	0.349
23	形容詞 連体詞	0.32
24	形容詞 接続詞	0.256
25	形容詞 助動詞	0.331
26	形容詞 助詞	0.37
27	副詞 連体詞	0.214
28	副詞 接続詞	0.118
29	副詞 助動詞	0.35
30	副詞 助詞	0.399

31	連体詞 接続詞	測定不能
32	連体詞 助動詞	0.316
33	連体詞 助詞	0.338
34	接続詞 助動詞	0.323
35	接続詞 助詞	0.364
36	助動詞 助詞	0.406
37	名詞 動詞 形容詞	0.354
38	名詞 動詞 副詞	0.369
39	名詞 動詞 連体詞	0.309
40	名詞 動詞 接続詞	0.324
41	名詞 動詞 助動詞	0.367
42	名詞 動詞 助詞	0.366
43	名詞 形容詞 副詞	0.34
44	名詞 形容詞 連体詞	0.343
45	名詞 形容詞 接続詞	0.369
46	名詞 形容詞 助動詞	0.345
47	名詞 形容詞 助詞	0.349
48	名詞 副詞 連体詞	0.353
49	名詞 副詞 接続詞	0.392
50	名詞 副詞 助動詞	0.307
51	名詞 副詞 助詞	0.388
52	名詞 連体詞 接続詞	0.355
53	名詞 連体詞 助動詞	0.376
54	名詞 連体詞 助詞	0.338
55	名詞 接続詞 助動詞	0.342
56	名詞 接続詞 助詞	0.364
57	名詞 助動詞 助詞	0.384
58	動詞 形容詞 副詞	0.316
59	動詞 形容詞 連体詞	0.299
60	動詞 形容詞 接続詞	0.275
61	動詞 形容詞 助動詞	0.347
62	動詞 形容詞 助詞	0.362
63	動詞 副詞 連体詞	0.305
64	動詞 副詞 接続詞	0.269
65	動詞 副詞 助動詞	0.333
66	動詞 副詞 助詞	0.39
67	動詞 連体詞 接続詞	0.241
68	動詞 連体詞 助動詞	0.299
69	動詞 連体詞 助詞	0.341
70	動詞 接続詞 助動詞	0.322
71	動詞 接続詞 助詞	0.327
72	動詞 助動詞 助詞	0.427
73	形容詞 副詞 連体詞	0.349
74	形容詞 副詞 接続詞	0.293
75	形容詞 副詞 助動詞	0.345
76	形容詞 副詞 助詞	0.405
77	形容詞 連体詞 接続詞	0.277
78	形容詞 連体詞 助動詞	0.345
79	形容詞 連体詞 助詞	0.394
80	形容詞 接続詞 助動詞	0.364

81	形容詞 接統詞 助詞	0.367
82	形容詞 助動詞 助詞	0.434
83	副詞 連体詞 接統詞	0.161
84	副詞 連体詞 助動詞	0.317
85	副詞 連体詞 助詞	0.389
86	副詞 接統詞 助動詞	0.328
87	副詞 接統詞 助詞	0.369
88	副詞 助動詞 助詞	0.435
89	連体詞 接統詞 助動詞	0.319
90	連体詞 接統詞 助詞	0.351
91	連体詞 助動詞 助詞	0.394
92	接統詞 助動詞 助詞	0.423
93	名詞 動詞 形容詞 副詞	0.362
94	名詞 動詞 形容詞 連体詞	0.324
95	名詞 動詞 形容詞 接統詞	0.32
96	名詞 動詞 形容詞 助動詞	0.334
97	名詞 動詞 形容詞 助詞	0.371
98	名詞 動詞 副詞 連体詞	0.334
99	名詞 動詞 副詞 接統詞	0.337
100	名詞 動詞 副詞 助動詞	0.358
101	名詞 動詞 副詞 助詞	0.363
102	名詞 動詞 連体詞 接統詞	0.338
103	名詞 動詞 連体詞 助動詞	0.34
104	名詞 動詞 連体詞 助詞	0.353
105	名詞 動詞 接統詞 助動詞	0.35
106	名詞 動詞 接統詞 助詞	0.374
107	名詞 動詞 助動詞 助詞	0.363
108	名詞 形容詞 副詞 連体詞	0.369
109	名詞 形容詞 副詞 接統詞	0.356
110	名詞 形容詞 副詞 助動詞	0.36
111	名詞 形容詞 副詞 助詞	0.351
112	名詞 形容詞 連体詞 接統詞	0.326
113	名詞 形容詞 連体詞 助動詞	0.345
114	名詞 形容詞 連体詞 助詞	0.353
115	名詞 形容詞 接統詞 助動詞	0.326
116	名詞 形容詞 接統詞 助詞	0.335
117	名詞 形容詞 助動詞 助詞	0.401
118	名詞 副詞 連体詞 接統詞	0.375
119	名詞 副詞 連体詞 助動詞	0.335
120	名詞 副詞 連体詞 助詞	0.354
121	名詞 副詞 接統詞 助動詞	0.35
122	名詞 副詞 接統詞 助詞	0.356
123	名詞 副詞 助動詞 助詞	0.374
124	名詞 連体詞 接統詞 助動詞	0.327
125	名詞 連体詞 接統詞 助詞	0.353
126	名詞 連体詞 助動詞 助詞	0.397
127	名詞 接統詞 助動詞 助詞	0.389
128	動詞 形容詞 副詞 連体詞	0.33
129	動詞 形容詞 副詞 接統詞	0.315
130	動詞 形容詞 副詞 助動詞	0.36



131	動詞 形容詞 副詞 助詞	0.365
132	動詞 形容詞 連体詞 接統詞	0.295
133	動詞 形容詞 連体詞 助動詞	0.336
134	動詞 形容詞 連体詞 助詞	0.33
135	動詞 形容詞 接統詞 助動詞	0.341
136	動詞 形容詞 接統詞 助詞	0.311
137	動詞 形容詞 助動詞 助詞	0.434
138	動詞 副詞 連体詞 接統詞	0.258
139	動詞 副詞 連体詞 助動詞	0.347
140	動詞 副詞 連体詞 助動詞	0.352
141	動詞 副詞 接統詞 助動詞	0.339
142	動詞 副詞 接統詞 助詞	0.365
143	動詞 副詞 助動詞 助詞	0.421
144	動詞 連体詞 接統詞 助動詞	0.338
145	動詞 連体詞 接統詞 助詞	0.314
146	動詞 連体詞 助動詞 助詞	0.432
147	動詞 接統詞 助動詞 助詞	0.393
148	形容詞 副詞 連体詞 接統詞	0.326
149	形容詞 副詞 連体詞 助動詞	0.374
150	形容詞 副詞 連体詞 助詞	0.418
151	形容詞 副詞 接統詞 助動詞	0.378
152	形容詞 副詞 接統詞 助動詞	0.375
153	形容詞 副詞 助動詞 助詞	0.411
154	形容詞 連体詞 接統詞 助動詞	0.326
155	形容詞 連体詞 接統詞 助詞	0.348
156	形容詞 連体詞 助動詞 助詞	0.408
157	形容詞 接統詞 助動詞 助詞	0.382
158	副詞 連体詞 接統詞 助動詞	0.329
159	副詞 連体詞 接統詞 助詞	0.365
160	副詞 連体詞 助動詞 助詞	0.413
161	副詞 接統詞 助動詞 助詞	0.431
162	連体詞 接統詞 助動詞 助詞	0.401
163	名詞 動詞 形容詞 副詞 連体詞	0.335
164	名詞 動詞 形容詞 副詞 接統詞	0.353
165	名詞 動詞 形容詞 副詞 助動詞	0.356
166	名詞 動詞 形容詞 副詞 助詞	0.363
167	名詞 動詞 形容詞 連体詞 接統詞	0.34
168	名詞 動詞 形容詞 連体詞 助動詞	0.319
169	名詞 動詞 形容詞 連体詞 助詞	0.371
170	名詞 動詞 形容詞 接統詞 助動詞	0.332
171	名詞 動詞 形容詞 接統詞 助詞	0.343
172	名詞 動詞 形容詞 助動詞 助詞	0.375
173	名詞 動詞 副詞 連体詞 接統詞	0.328
174	名詞 動詞 副詞 連体詞 助動詞	0.351
175	名詞 動詞 副詞 連体詞 助詞	0.342
176	名詞 動詞 副詞 接統詞 助動詞	0.339
177	名詞 動詞 副詞 接統詞 助詞	0.364
178	名詞 動詞 副詞 助動詞 助詞	0.365
179	名詞 動詞 連体詞 接統詞 助動詞	0.329
180	名詞 動詞 連体詞 接統詞 助詞	0.345

181	名詞 動詞 連体詞 助動詞 助詞	0365
182	名詞 動詞 接統詞 助動詞 助詞	0.387
183	名詞 形容詞 副詞 連体詞 接統詞	0.329
184	名詞 形容詞 副詞 連体詞 助動詞	0.36
185	名詞 形容詞 副詞 連体詞 助詞	0.332
186	名詞 形容詞 副詞 接統詞 助動詞	0.359
187	名詞 形容詞 副詞 接統詞 助詞	0.349
188	名詞 形容詞 副詞 助動詞 助詞	0.398
189	名詞 形容詞 連体詞 接統詞 助動詞	0.337
190	名詞 形容詞 連体詞 接統詞 助詞	0.344
191	名詞 形容詞 連体詞 助動詞 助詞	0.357
192	名詞 形容詞 接統詞 助動詞 助詞	0.37
193	名詞 副詞 連体詞 接統詞 助動詞	0.333
194	名詞 副詞 連体詞 接統詞 助詞	0.366
195	名詞 副詞 連体詞 助動詞 助詞	0.372
196	名詞 副詞 接統詞 助動詞 助詞	0.353
197	名詞 連体詞 接統詞 助動詞 助詞	0.37
198	動詞 形容詞 副詞 連体詞 接統詞	0.32
199	動詞 形容詞 副詞 連体詞 助動詞	0358
200	動詞 形容詞 副詞 連体詞 助詞	0.343
201	動詞 形容詞 副詞 接統詞 助動詞	0.354
202	動詞 形容詞 副詞 接統詞 助詞	0.367
203	動詞 形容詞 副詞 助動詞 助詞	0.408
204	動詞 形容詞 連体詞 接統詞 助動詞	0.348
205	動詞 形容詞 連体詞 接統詞 助詞	0.355
206	動詞 形容詞 連体詞 助動詞 助詞	0.394
207	動詞 形容詞 接統詞 助動詞 助詞	0.425
208	動詞 副詞 連体詞 接統詞 助動詞	0.344
209	動詞 副詞 連体詞 接統詞 助詞	0.331
210	動詞 副詞 連体詞 助動詞 助詞	0.398
211	動詞 副詞 接統詞 助動詞 助詞	0.407
212	動詞 連体詞 接統詞 助動詞 助詞	0.432
213	形容詞 副詞 連体詞 接統詞 助動詞	0.354
214	形容詞 副詞 連体詞 接統詞 助詞	0.387
215	形容詞 副詞 連体詞 助動詞 助詞	0.421
216	形容詞 副詞 接統詞 助動詞 助詞	0.419
217	形容詞 連体詞 接統詞 助動詞 助詞	0.377
218	副詞 連体詞 接統詞 助動詞 助詞	0.402
219	名詞 動詞 形容詞 副詞 連体詞 接統詞	0.337
220	名詞 動詞 形容詞 副詞 連体詞 助動詞	0.344
221	名詞 動詞 形容詞 副詞 連体詞 助詞	0.358
222	名詞 動詞 形容詞 副詞 連体詞 助詞	0.359
223	名詞 動詞 形容詞 副詞 接統詞 助詞	0.374
224	名詞 動詞 形容詞 副詞 助動詞 助詞	0.373
225	名詞 動詞 形容詞 連体詞 接統詞 助動詞	0.364
226	名詞 動詞 形容詞 連体詞 接統詞 助詞	0.356
227	名詞 動詞 形容詞 連体詞 助動詞 助詞	0.382
228	名詞 動詞 形容詞 接統詞 助動詞 助詞	0.384
229	名詞 動詞 副詞 連体詞 接統詞 助動詞	0.34
230	名詞 動詞 副詞 連体詞 接統詞 助詞	0.341

231	名詞 動詞 副詞 連体詞 助動詞 助詞	0.363
232	名詞 動詞 副詞 接統詞 助動詞 助詞	0.407
233	名詞 動詞 連体詞 接統詞 助動詞 助詞	0.375
234	名詞 形容詞 副詞 連体詞 接統詞 助動詞	0.365
235	名詞 形容詞 副詞 連体詞 接統詞 助詞	0.335
236	名詞 形容詞 副詞 連体詞 助動詞 助詞	0.361
237	名詞 形容詞 副詞 接統詞 助動詞 助詞	0.407
238	名詞 形容詞 連体詞 接統詞 助動詞 助詞	0.367
239	名詞 副詞 連体詞 接統詞 助動詞 助詞	0.395
240	動詞 形容詞 副詞 連体詞 接統詞 助動詞	0.377
241	動詞 形容詞 副詞 連体詞 接統詞 助詞	0.346
242	動詞 形容詞 副詞 連体詞 助動詞 助詞	0.446
243	動詞 形容詞 副詞 接統詞 助動詞 助詞	0.433
244	動詞 形容詞 連体詞 接統詞 助動詞 助詞	0.39
245	動詞 副詞 連体詞 接統詞 助動詞 助詞	0.363
246	動詞 副詞 連体詞 接統詞 助動詞 助詞	0.401
247	名詞 動詞 形容詞 副詞 連体詞 接統詞 助動詞	0.357
248	名詞 動詞 形容詞 副詞 連体詞 接統詞 助詞	0.365
249	名詞 動詞 形容詞 副詞 連体詞 助動詞 助詞	0.364
250	名詞 動詞 形容詞 副詞 接統詞 助動詞 助詞	0.389
251	名詞 動詞 形容詞 連体詞 接統詞 助動詞 助詞	0.408
252	名詞 動詞 副詞 連体詞 接統詞 助動詞 助詞	0.373
253	名詞 形容詞 副詞 連体詞 接統詞 助動詞 助詞	0.394
254	動詞 形容詞 副詞 連体詞 接統詞 助動詞 助詞	0.406
255	名詞 動詞 形容詞 副詞 連体詞 接統詞 助動詞 助詞	0.395

# 研究業績

---

## 査読付き国際会議

- (1) Kenro Go, Nagomu Horikoshi, Shion Tominaga, Jinta Nakamura, Akihiro Miyata: Implementation and Evaluation of a Reminder Registration Interface for Daily Life Objects, Proceedings of the 20th International Conference on Human-Computer Interaction (Presentation planned in July 2018).
- 

## 査読付き国内会議

- (1) 中村仁汰, 玉城和也, 小林舞子, 中辻真, 宮田章裕: 文章表現の堅さ推定手法の基礎検討, 情報処理学会グループウェアとネットワークサービスワークショップ2017 論文集, Vol.2017, pp.1-6 (2017年11月).
  - (2) 呉健朗, 堀越和, 富永詩音, 中村仁汰, 宮田章裕: 実世界オブジェクトへのリマインダ登録インタフェースの実装と評価, 情報処理学会グループウェアとネットワークサービスワークショップ2017 論文集, Vol.2017, pp.1-8 (2017年11月).
- 

## 研究会・シンポジウム

- (1) 玉城和也, 呉健朗, 中村仁汰, 富永詩音, 宮田章裕: 紙をちぎることで電子情報を手渡すインタラクション方式の提案, 情報処理学会インタラクション2018 (2018年3月発表予定).
  - (2) 呉健朗, 玉城和也, 中村仁汰, 宮田章裕: 紙をちぎることで電子情報を手渡すインタラクション方式の提案, マルチメディア、分散、協調とモバイル (DICOMO2017) シンポジウム, pp.1493-1499 (2017年6月).
  - (3) 呉健朗, 中村仁汰, 堀越和, 宮田章裕: InfoClip:実世界オブジェクトへのリマインダ登録インタフェースの基礎検討, 情報処理学会インタラクション2017, インタラクション2017 論文集, pp.783-788 (2017年3月).
-