

使用者数による語彙制限を用いた日本語学習者のための 文章読解支援

塩田 健人^{1,a)} 梶原 智之^{1,b)} 小町 守^{1,c)}

概要: 語彙平易化とは、難解な語や句を平易な語や句に言い換えることで、子どもや言語学習者などの文章読解を支援する技術である。特に、読者の理解語彙に言い換え対象の語彙を制限することで、文章の可読性を向上させることが期待できる。そこで、本研究では複数の尺度を用いて語彙制限した言い換えを行い、文の難易度について日本語学習者の主観評価を受けた。その結果、Web 日本語 N グラムから得た頻度よりも、Twitter から得た使用者数の指標で行った語彙制限が読解支援に有効であった。

1. はじめに

語彙平易化とは、難解な語や句をより平易な表現に言い換えることにより、子どもや言語学習者の文章読解を支援する技術である。特に、読者の理解語彙に絞って言い換えを行うことで、文章の可読性を向上させることができる。語彙平易化タスクは「言い換え候補の生成」と「言い換え候補の選択」の2つのサブタスクとして定式化できる。言い換え候補の生成とは、難解語とその言い換えとなる表現を抽出するタスクで、言い換え候補の選択とは、言い換えとなる表現の中から、より平易な表現を選択するタスクのことである。本稿では読者の理解語彙になるよう語彙平易化するタスクのことを特に**語彙制限**と呼ぶ。

語彙平易化には大きく分けて2つのアプローチがある。ひとつは辞典の語釈文を使用するアプローチで、もうひとつは統計(頻度)に基づくアプローチである。

梶原ら [1] は国語辞典の見出し語と語釈文を用いた語彙平易化を提案している。これは、国語辞典の語釈文は見出し語よりも平易な表現で見出し語を説明していることを利用して平易な表現を語釈文から獲得し、言い換える手法である。日本語の主要部終端型の特徴を利用して国語辞典の語釈文の末尾の語を利用する梶原ら [2] の手法よりも、広い範囲から換言候補を獲得し、シソーラスに基づく語の類似度が最も高い換言候補にしたことから国語辞典を最大限に利用して高精度な平易化を行っている。また、美野ら [3] は語釈文間の類似度を用いて見出し語の換言対を獲得し、

平易化を行う手法を提案している。

統計に基づくアプローチは、設計がシンプルで、教師データが必要でないという利点がある。例えば、SemEval-2012 English Lexical Simplification Task [4] では、単純頻度のみを用いたベースラインシステムが全12システム中2位の成績を示し、英語の語彙平易化タスクにおける高頻度語への言い換えるの有効性が示された。

1つのスコアに基づく平易化手法は、複雑な機械学習モデルの作成や素性設計を行う必要がない、という利点があるが、語彙平易化タスクにとって単純頻度が最適なスコアであるとは限らない。例えば、高頻度語は平易な語であることが知られている [5] が、Aramaki ら [6] によって提案された Twitter に投稿されたテキストから獲得した「語の使用数」という統計量も語彙平易化の指標に利用できる可能性がある。

この指標は、国語辞典に収載されている語を「自然な日本語」として判定するタスクにおいて、単純頻度よりも優れた指標であることが示されている。これらの尺度はコーパスに基づくためカバー率が高いという利点があるが、日本語学習者に特化したものではなく、日本語学習者の読解支援に有効であるかどうかは分かっていない。したがって、「語の使用数」という指標が語彙平易化をする指標として有効か調査する価値は十分にある。

梶原らや美野らの提案しているシステムは辞書を利用した日本語の語彙平易化に関するシステムを作成しているが、辞書の語釈文は見出し語の意味を解説するために書かれているので、必ずしも語釈文から得た語が語彙として平易になる保証はない。

そこで、本研究では、語の出現頻度および使用者数の指

¹ 首都大学東京

Tokyo Metropolitan University

a) shioda-kent@ed.tmu.ac.jp

b) kajiwara-tomoyuki@ed.tmu.ac.jp

c) komachi@tmu.ac.jp

標を用いた語彙制限を行い、日本語学習者を対象に文章読解を支援する評価実験を行った。本研究も1つの尺度のみを用いた語彙制限による平易化手法を行う。実験の結果、Twitterから得た使用者数の多さがもっとも平易化の尺度として適切であることが分かった。

2. 語彙的換言知識を用いた語彙制限

本研究では、語または句のレベルでの言い換え（語彙的換言）を再帰的に繰り返して、文中の全ての内容語（名詞、動詞、形容詞、副詞）を平易語に制限する。

本研究では語彙平易化のための語彙制限を行うので、より平易な表現が得られるような言い換えを行う。即ち、頻度または使用者数が増加するような表現へ言い換える。

以下に語彙制限の一例を示す。「乗っ取る」という単語が難解語であるとき、これをより平易な「奪う」という単語へ言い換える。言い換えられた語「奪う」が平易語であればここで変換を終えるが、この語がまだ難解語であれば、さらに「取る」へと言い換えを続ける。これを平易語が得られるまで繰り返す。「取る」が難解語で、これ以上言い換え先の語がない場合は探索を打ち切り、言い換えは行わない。また、「乗っ取る」が複数の語へ言い換えることが可能な場合、表1の順に換言知識を探索して出現した際に処理をした。

警察の車を 乗っ取る
↓
警察の車を 奪う
↓
警察の車を 取る

3. 実験

3.1 語彙的換言知識

本研究では、現在利用可能な6種類の日本語の語彙的換言知識（表1）を用いて言い換えを行う。

動詞含意関係データベース^{*1}は、Webから自動的に収集された動詞対に対して、人手で含意関係にあると判定された動詞対である。含意は正確には同義ではないが、本研究では「チンする→加熱する」のような含意関係の変換も扱う。

日本語異表記データベース^{*2}は、編集距離が1である語と句の異表記対である。

基本的意味関係の事例ベース^{*3}は、文脈類似語データベース^{*4}に含まれる各エントリに対して、人手でその名

詞対の関係を分類したものである。本研究ではこの中から、略記、異形同義語、異表記の3つの関係の名詞対を利用する。

PPDB:Japanese [7]^{*5}は、日英対訳コーパスからピボット方式で構築され、KyTeaで形態素解析された句単位の換言知識である。PPDB:Japaneseには各言い換え対に言い換え確率が付与されているが、本研究では言い換え確率が0.5以上の言い換え対のみを使用する。また、本研究ではPPDB:Japaneseから獲得した語は、正規化をせずに表層形を換言知識として使用している。難解表現は語として得られるため、語から語、または語から句への言い換え対のみを扱う。

内容語換言辞書 [8], [9]^{*6}は完全に人手で構築された換言知識である。形態素解析器JUMAN (Ver.7.0)^{*7}の形態素辞書の見出し語に対して、語または内容語3語までの句で言い換えを付与している。

日本語WordNet同義語データベース^{*8}は、日本語WordNet (Ver.1.1)で同じsynsetに登録されている名詞対の中から、人手で同義関係を判定された対を収集した換言知識である。

3.2 実験設定

本研究ではWeb日本語Nグラム^{*9}の7グラムデータから文頭記号と文末記号に囲まれた5単語の文を抽出し、語彙平易化対象文として使用した。内容語を抽出するために、品詞情報はUniDicでMeCab0.996^{*10}を使用して再解析したが、サ変動詞は結合する処理を行わなかった。その他の言語資源は再解析を行わなかった。

抽出した5単語のうち、内容語をさらに抽出して原形にする処理をしてから言い換えを行った。

各上位M語の場合において、難解語を含まない文を除外し、言い換えの実験には各文につき難解語を1語だけ含む文を使用した。これは、言い換える前後で文の難易度を比較するため、及び難解語が複数あるとどの部分を平易化したことによって文が平易になったのかが明らかでないためこのような処理を行った。

この難解語とは、頻度かつ使用者数の上位M語に含まれない語である。例えば、N=5,000の場合、使用者数、頻度Web、頻度Twitterのいずれの上位5,000語にも含まれない語である。本研究では上位5,000語、上位7,500語、上位10,000語への語彙制限を行った。このNの値は、日本語を母語としない日本語学習者の日本語能力を測定する

^{*1} <https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-2>

^{*2} <https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-7>

^{*3} <https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-9>

^{*4} <https://alaginrc.nict.go.jp/resources/nict-resource/>

<li-info/li-outline.html#A-1>

^{*5} <http://isw3.naist.jp/masahiro-mi/jppdb/>

^{*6} <http://www.jnlp.org/SNOW/D2>

^{*7} <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

^{*8} <http://compling.hss.ntu.edu.sg/wnja/>

^{*9} <http://www.gsk.or.jp/catalog/gsk2007-c/>

^{*10} <http://taku910.github.io/mecab/>

表 1 語彙的換言知識

名称	品詞	例	言い換えの向き	換言対
動詞含意関係 DB (Ver.1.3.1)	動詞	チンする→過熱する	左から右	89,784
日本語異表記対 DB (Ver.1.1)	名詞	ゴミ置き場 ⇔ ゴミ置場	両方向	5,513,606
基本的意味関係の事例ベース (Ver.1.4)	名詞	短大 ⇔ 短期大学	両方向	78,260
PPDB Japanese (Ver.0.2.0)	語→句	光速→光の速度	左から右	33,150
内容語換言辞書	語→句	案内→連れて行く	左から右	25,504
日本語 WordNet 同義語 DB (Ver.1.0)	名詞	故障 ⇔ トラブル	両方向	11,753
使用した全言い換え対				11,355,676

表 2 語彙制限による平易化の評価結果

上位 M 語	原文	頻度 Web	頻度 Twitter	使用者数
N = 5,000	85	85 (15)	86 (9)	91 (8)
N = 7,500	88	77 (21)	91 (8)	93 (7)
N = 10,000	81	86 (11)	89 (9)	85 (9)

JLPT*¹¹ の N2 級レベルの人が持つ語彙力が約 6,000 語、N1 級レベルの人が持つ語彙力が約 10,000 語であることから決定した。

言い換えを行った際の意味の保持については著者 1 名が評価を行い、意味が保持できている文のみを平易化の評価に用いた。

平易化に関する評価は日本語能力試験 N1 級を持つ日本語学習者 1 名が評価対象の 100 文に対して行った。

語彙制限を行った後、平易化された語の難易度を調査するために日本語教育語彙表*¹²を使用した。これは語の難易度が 6 段階で表されており、平易化したあとの語がより理解しやすい語に置換されたことを示すために行った。

3.3 評価：理解の可否

被験者が原文および各指標で平易化された平易文のそれぞれを読み、理解の可否を○×で評価した。このとき、被験者はどの文がどの指標で平易化された文なのかを知らない。

表 2 に、原文および各平易文における理解できた文数を示す。また、理解できる原文が言い換えによって理解できなくなってしまった文数を括弧内に示す。

まず、日本語能力試験 N1 級を持つ今回の被験者は、難解文の 80%以上を理解することができた。日本語学習者が難解文を理解できる場合、無理に平易化を行う必要はないので、今回の実験の被験者として日本語能力試験 N1 級保持者は適切ではなかった。今後は日本語学校等に協力を依頼し、日本語能力の異なる様々な被験者から評価を得たい。

次に、理解できた文数に注目すると、上位 5,000 語への厳しい語彙制限の際には使用者数で平易化すると分かりやすくなっており、語彙制限の基準を緩めるにつれて他の指標との差が少なくなっていることが分かる。この結果か

ら、厳しい語彙制限の時、使用者数は語彙平易化の指標に使用できる可能性が十分にあると言える。

さらに、言い換えることにより理解できなくなってしまった文数に注目すると、Twitter から得た頻度や使用者数で平易化した場合に、Web から得た頻度で平易化するよりも、難解な表現に変換してしまう誤りが少ないことが分かる。

3.4 評価：言い換える前後における語の難易度の変化

この節では、言い換える前は理解できなかったが、難解語を言い換えることにより理解できるようになった文について、言い換えを行う前後の文の難易度の変化について日本語教育語彙表を用いて調査した。表 3 に調査の結果を示す。実験と日本語教育語彙表 (ver.1.0) では使用している言語資源が異なるため、単語分割基準が異なり、言い換え前後の語が日本語教育語彙表で級外となってしまうことがある。

表 3 より、N1 級を持つ日本語学習者が平易化によって理解できるようになった文は、各上位 M 位、各指標の列の難易度の数を全て合計すると上位 5,000 語で 24 文、上位 7,500 語で 21 文、上位 10,000 語で 18 文であった。また、日本語教育語彙表に収録されている語で、言い換えをすることによって理解できるようになった文のうち難易度が上がっている例は 1 つもなかった。全ての指標において上級前半の語が主に言い換えられており理解できているという同じような結果が得られたため、日本語学習者は上級以上の難易度の語は言い換えが必要である可能性がある。

この結果から、言い換える前後で語の難易度がどのような分布を示しているかがわかるが、使用者数で言い換えた方が、より難易度の低い語にわずかであるが置換されているため、語彙平易化の指標として使用できる可能性がある。しかし、全指標で非常に類似した分布が得られたので、この分布だけでは適切な指標が見つかったとは言えない。より多くの評価結果が得られれば、現地ではまだ得られていない結果が得られる可能性があるため、今後多くの評価者に評価をしてもらう必要がある。

*¹¹ <http://www.jlpt.jp>

*¹² <http://jhlee.sakura.ne.jp/JEV.html>

表 3 語彙制限による平易化前後で理解できている
文の難易度の変化の数

平易化前	平易化後	5,000 語			7,500 語			10,000 語		
		頻度 Web	頻度 Twitter	使用者数	頻度 Web	頻度 Twitter	使用者数	頻度 Web	頻度 Twitter	使用者数
	初級前半	0 / 1 / 1	0 / 0 / 0	0 / 0 / 0						
	初級後半	1 / 0 / 0	0 / 0 / 1	0 / 0 / 0						
	中級前半	1 / 1 / 1	0 / 0 / 0	0 / 0 / 0						
上級後半	中級後半	0 / 0 / 0	1 / 1 / 0	0 / 0 / 0						
	上級前半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	上級後半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	級外	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	初級前半	0 / 1 / 1	0 / 0 / 0	0 / 0 / 0						
	初級後半	1 / 2 / 1	1 / 0 / 0	1 / 1 / 1						
上級前半	中級前半	4 / 3 / 2	3 / 2 / 5	2 / 2 / 3						
	中級後半	1 / 2 / 4	2 / 5 / 3	3 / 4 / 2						
	上級前半	4 / 2 / 2	2 / 1 / 0	4 / 3 / 4						
	上級後半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	級外	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	初級前半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	初級後半	0 / 1 / 1	0 / 0 / 0	0 / 0 / 0						
中級後半	中級前半	1 / 0 / 0	0 / 0 / 1	1 / 1 / 1						
	中級後半	0 / 0 / 0	1 / 1 / 0	0 / 0 / 0						
	上級前半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	上級後半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	級外	0 / 0 / 0	0 / 0 / 0	0 / 1 / 1						
	初級前半	0 / 1 / 1	0 / 0 / 0	1 / 0 / 1						
	初級後半	1 / 1 / 0	0 / 0 / 0	0 / 0 / 0						
中級前半	中級前半	3 / 2 / 3	1 / 1 / 1	1 / 1 / 0						
	中級後半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	上級前半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	上級後半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	級外	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	初級前半	1 / 1 / 1	2 / 1 / 1	2 / 1 / 0						
	初級後半	0 / 0 / 0	1 / 1 / 1	0 / 0 / 0						
級外	中級前半	3 / 2 / 3	4 / 3 / 4	0 / 1 / 1						
	中級後半	2 / 3 / 2	2 / 3 / 2	2 / 1 / 2						
	上級前半	0 / 1 / 1	0 / 2 / 2	0 / 1 / 2						
	上級後半	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0						
	級外	1 / 0 / 0	1 / 0 / 0	1 / 1 / 0						

表 4 ランキングの評価結果：5,000

	原文	Web	Twitter	使用者数
原文		18	11	9
頻度 Web	39		14	14
頻度 Twitter	37	19		7
使用者数	41	25	9	

表 5 ランキングの評価結果：7,500

	原文	Web	Twitter	使用者数
原文		29	11	10
頻度 Web	36		13	13
頻度 Twitter	39	30		3
使用者数	40	31	4	

表 6 ランキングの評価結果：10,000

	原文	Web	Twitter	使用者数
原文		12	14	12
頻度 Web	48		27	25
頻度 Twitter	39	19		3
使用者数	40	19	6	

3.5 評価：難易度ランキング

原文および各指標で言い換えを行った平易文について、被験者が理解しやすい順に同順を許可して並び替えを行なった。表 4 から表 6 に、3 つの指標ごと難易度の大小関係の分布を示す。縦に書かれている指標で言い換えをする方が横に書かれている指標で言い換えをするよりも理解しやすいことを示している。例を示すと、表 4 において縦：使用者数、横：頻度 Web のセルは、頻度 Web よりも使用者数で言い換えを行った結果、100 文中 25 文理解しやすい文があったことを示している。

各表の 1 列目に注目すると、いずれも平易文は原文より理解しやすくなっていることが分かる。また、頻度 Twitter と使用者数の欄に注目すると、Twitter から得た頻度と使用者数は、同程度の難易度であることが分かる。

上位 5,000 語への語彙制限では、Twitter から得た頻度や使用者数による平易化の効果が大きく、「頻度 Twitter = 使用者数 > 頻度 Web > 原文」という関係で理解しやすいと言える。一方、上位 10,000 語への語彙制限では、Web から得た頻度による平易化の効果が最も高く、「頻度 Web > 頻度 Twitter = 使用者数 > 原文」という関係で理解しやすいと言える。

以上の結果から、より厳しい設定で語彙制限を行う場合（上位 5,000 語）には Twitter から得た頻度や使用者数の指標で平易化することが効果的である。

このことは、Twitter から得た語で上位な語は専門的であり馴染みのない語であることが少ないと考えられるため、Web から獲得した語で言い換えを行うよりも比較的理屈がしやすいからであると言える。

また、より緩やかな設定で語彙制限を行う場合（上位 10,000 語）には Web から得た頻度で平易化することが効果的である。

このことは、本実験の場合、語彙制限を緩くすることによってより難解な語も平易語として扱われる。頻度 Web の上位 5,000 語から上位 10,000 語、かつ上位 7,500 語から上位 10,000 語に存在する語は、使用者数や頻度 Twitter のそれぞれ同じ範囲に存在する語と比較して情報処理や抗菌などといった専門的な語が多かった。したがって、Twitter で獲得した語だけではカバーしきれない意味を持つ語は Web で獲得した頻度を指標として語彙平易化するべきである。

4. おわりに

本研究では、日本語学習者の文章読解支援のために語彙平易化をする際、どの指標で語彙制限すべきかについて検討した。その結果、次の3つのことが明らかになった。

- 上位 5,000 語など、厳しい設定で語彙制限を行う場合は、Twitter から得た使用者数で平易化すると頻度 Web や頻度 Twitter で平易化するよりも平易な表現に言い換えられる
- 上位 10,000 語など、緩い設定で語彙制限を行う場合は、Web から得た頻度で平易化すると、Twitter から獲得した語ではカバーしきれない意味を持つ語に言い換えることが可能である
- Twitter から得た頻度や使用者数の指標で平易化すると、誤って難解に言い換えることが少ない

以上の3点から使用者数は語彙平易化に使用できる可能性があると言える。また、Twitter から獲得した語は、Web から獲得した語よりも一般的な語が獲得できている可能性があり、語彙制限が厳しい時の平易化の指標に有効であることを示した。

今後は日本語学校等に協力を依頼し、日本語能力の異なる様々な被験者から評価を得たい。そして、日本語習熟度に応じた語彙平易化の方法を検討したい。

参考文献

- [1] 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, Vol. 56, No. 3, pp. 983-992 (2015).
- [2] 梶原智之, 山本和英: 小学生の読解支援に向けた複数の換言知識を併用した語彙平易化と評価, 第 19 回年次大会発表論文集, pp. 272-275 (2013).
- [3] 美野秀弥, 田中英輝: 国語辞典を使った放送ニュースの名詞の平易化, 言語処理学会第 16 回年次大会発表論文集, pp. 760-763 (2010).
- [4] Specia, L., Jauhar, S. K. and Mihalcea, R.: Semeval-2012 task 1: English lexical simplification, *Proceedings of the Sixth International Workshop on Semantic Evaluation*, pp. 347-355 (2012).
- [5] 梶原智之, 山本和英: 高頻度語は平易語なのか?, NLP 若手の会第 9 回シンポジウム, 発表 P02, pp. 1-2 (2014).
- [6] Aramaki, E., Maskawa, S., Miyabe, M., Morita, M. and Yasuda, S.: Word in a Dictionary is used by Numerous Users, *Proceedings of International Joint Conference on Natural Language Processing*, pp. 874-877 (2013).
- [7] Mizukami, M., Neubig, G., Sakti, S., Toda, T. and Nakamura, S.: Building a free, general-domain paraphrase database for Japanese, *17th Oriental Chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques*, pp. 1-4 (2014).
- [8] 山本和英, 吉倉孝太郎: 用言等換言辞書を人手で作りました, 言語処理学会 第 19 回年次大会 発表論文集, pp. 276-279 (2013).
- [9] 山形祐輝, 山本和英: 普通名詞換言辞書の構築, 言語処理学会 第 20 回年次大会発表論文集, pp. 7-10 (2014).

表 A.1 実験に用いた指標

指標	データ	異なり語数
頻度	Web 日本語 N グラム	2,565,424
	Twitter	48,324
使用者数	Twitter	48,324

表 A.2 各指標の相関係数

指標	Peason	Spearman	Kendall
頻度 Web × 頻度 Twitter	0.819	0.758	0.783
頻度 Web × 使用者数	0.277	0.732	0.776
頻度 Twitter × 使用者数	0.280	0.986	0.953
親密度 × 頻度 Web	0.027	0.271	0.576
親密度 × 頻度 Twitter	0.028	0.330	0.600
親密度 × 使用者数	0.280	0.353	0.609

付 録

本研究で用いる指標とデータを表 A.1 に示す。頻度は、Web 日本語 N グラム *¹³ から得た頻度と Twitter から得た頻度の 2 つのデータを使用する。使用者数は Twitter から得たデータを使用する。Twitter から得たデータは 2009/11/3 から 2010/3/25 の 143 日間に取得した約 2.5 億ツイート (253,482,784 ツイート) を JUMAN (Ver.7.0) で解析したものを使用した。これらの Twitter 上での語の出現頻度と使用者数に関するデータは、Aramaki ら [6] のものを用いた。

表 A.2 に、データ間の相関係数を示す。我々は、ピアソンの積率相関係数、スピアマンの順位相関係数、ケンドールの順位相関係数、の 3 つの相関係数を調査した。また、各データの散布図を図 A.1 から図 A.3 に示す。いずれのデータ間にも、強い正の相関が見られた。

本研究の実験の初期段階で単語親密度 *¹⁴ を語彙制限の指標に加えていたが、難解語を決定する時に上位 M 語を抽出する際、使用者数や頻度 Web、頻度 Twitter の各指標と重複する語がほとんど単語親密度のデータベースに存在しなかった。図 A.4 から図 A.6 に単語親密度と使用者数、頻度 Web、頻度 Twitter との相関の図を示す。親密度について、親密度が 5.5 以上の語は頻度 Web や頻度 Twitter、使用者数の上位の語に出現するが、4 ポイント台になると急激に頻度 Web や頻度 Twitter、使用者数の上位に出現しない語が増加する。したがって、語彙平易化の手法としては不適切であると判断した。

*¹³ <http://www.gsk.or.jp/catalog/gsk2007-c/>

*¹⁴ http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/brief_3.html

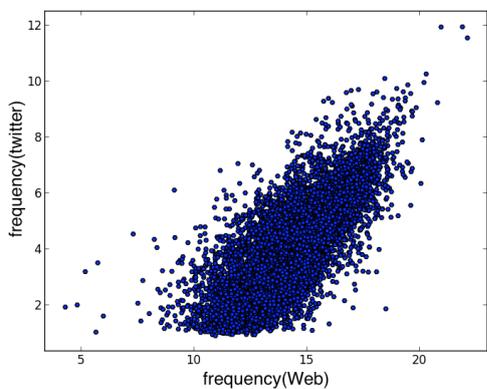


図 A.1 頻度 Web と頻度 Twitter

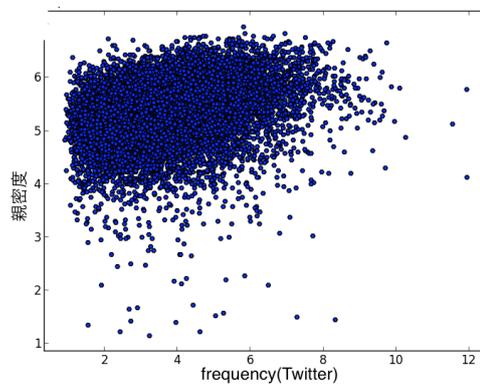


図 A.4 頻度 Twitter と親密度

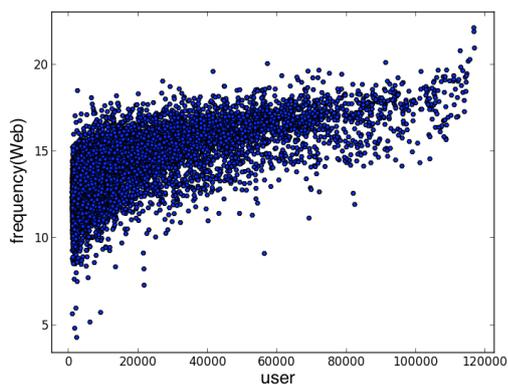


図 A.2 頻度 Web と使用者数

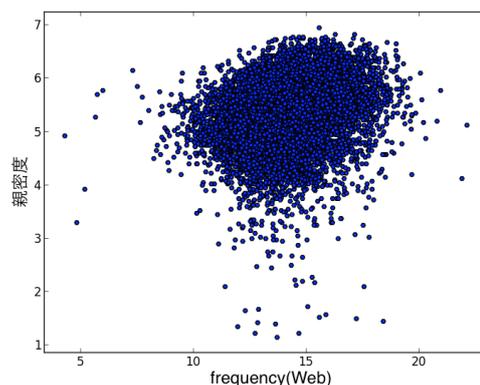


図 A.5 頻度 Web と親密度

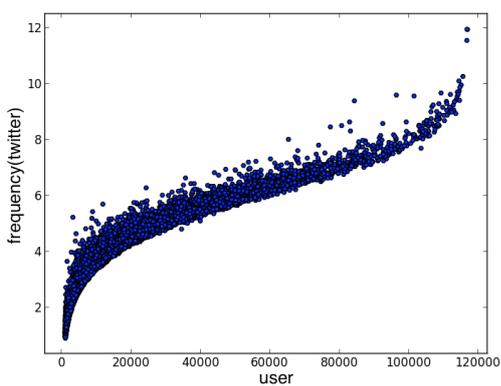


図 A.3 頻度 Twitter と使用者数

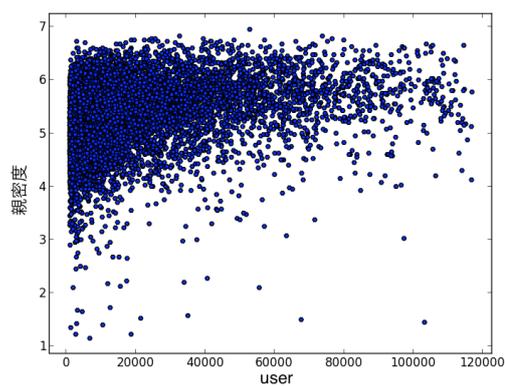


図 A.6 使用者数と親密度