

主観的評価に基づく語間関係の評価尺度の比較

山本英子[†] 乾裕子^{†‡} 井佐原均^{†‡}[†]通信総合研究所 [‡]神戸大学大学院

1 はじめに

コーパスから得られる語と語の関係は書き手の意図を理解することや文末に到達する前の時点で文の内容を予測し理解すること[6]、未知の知識を得ることの手がかりとなる。この語間関係を自動的に推定するために、コーパスにおける語の出現状況の類似性が利用される。類似性を測る尺度はこれまでに多く提案されているが、対象とするコーパスの種類や推定された関係の利用目的によって適する類似尺度は異なると考えられる。例えば、「てほしい」を含む文に言い換え可能な回答は要求とみなせる[7]が、同じ要求意図を持つ表現でも、問題を回答者がどの程度重要と捉えるかによって用いる要求表現に違いが現れる。そこにはなんらかの選択基準があるはずである。そこで、本研究では、この基準を語の共起関係に基づく類似度によって明らかにできないかと考えた。具体的にはアンケートの自由回答コーパス[2]の要求意図表現と問題のテーマを連想できる名詞との関係をさまざまな尺度を用いて推定し、それらの結果と、主観的に要求の強さを測った予備実験結果とを照合する。その結果から、回答者の要求の強さを知ることができる要求表現の選択基準を明らかにする目的に適した尺度を検討する。

2 様々な尺度による語間関係の自動推定

2.1 概要

実験では、アンケートの自由回答コーパスから要求表現を手がかりに回答者の意図を推測することを目的とする。そこで、コーパスから分析者が指定した要求表現と名詞との関係を2.3節に示す尺度を用いて推定する。各尺度において、スコアが高いほど、二つの語は関係があると解釈し、スコアの高い順に分析者に提示する。実際にその関係が正解か不正解であるかは、語を指定した分析者の利用目的に適するかどうかで判断される。

2.2 出現状況を表すパラメータ

本研究では、二つの語の相関はコーパス中の出現状況の類似性を表すものとする。そのため、どの尺度も基本的に出現パターンの重なり具合を測定することでスコア(類似度)を得る[4]。具体的には、コーパスに含まれる文書の総数を次元数 n とし、語が文書 i に出現するなら1、しなければ0を置き、各語の出現パターンを二値ベクトル化し、次に示す四つのパラメータを計数し、各定義式によって類似度を得る。

パラメータ

二つの単語に対する二値パターンをそれぞれ二値 n 次元ベクトル $F = (f_1, f_2, \dots, f_i, \dots, f_n)$, $T = (t_1, t_2, \dots, t_i, \dots, t_n)$ とし、 f_i, t_i は0か1である。パラメータ a, b, c, d は次のように求める。ただし、 $n=a+b+c+d$ である。

$$a = \sum_{i=1}^n f_i t_i, \quad b = \sum_{i=1}^n f_i (1-t_i), \\ c = \sum_{i=1}^n (1-f_i) t_i, \quad d = \sum_{i=1}^n (1-f_i)(1-t_i)$$

それぞれのパラメータは、

- a : 二つの単語がどちらとも出現する文書数。
- b : 一方は出現しないが、他方は出現する文書数。
- c : 他方は出現するが、一方は出現しない文書数。
- d : 二つの語がどちらとも出現しない文書数。

を表す。

2.3 評価尺度

本研究では、八つの尺度を用いて語間関係の推定を試みた。ここでは、その八つの尺度を2.2節に示したパラメータを用いた形式で示す。すべての尺度に与える引数は二つの単語の出現パターンである。ここでは、引数を省略し、左辺は各尺度に対応する関数名とした。

共起頻度[3]

$$Co-oc = a$$

共起と非共起の比[5]

$$BayesRatio = \frac{a+1}{b+c+2}$$

ダイス相関係数[3]

$$Dice = \frac{2a}{2a+b+c}$$

カイ二乗値[1,3]

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

イエーツの補正公式[1]

$$Yates = \frac{n(|ad-bc|-n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

自己相互情報量[3]

$$PMI = \log \frac{an}{(a+b)(a+c)}$$

対数尤度比[1]

$$LLR = a \log \frac{an}{(a+b)(a+c)} + b \log \frac{bn}{(a+b)(b+d)} \\ + c \log \frac{cn}{(a+c)(c+d)} + d \log \frac{dn}{(b+d)(c+d)}$$

補完類似度[4]

$$CSM = \frac{ad-bc}{\sqrt{(a+b)(b+d)}}$$

2.4 推定結果

推定されるのはアンケートの分析者から指定された要求表現と内容語との関係である。この関係を 2.3 節の尺度を用いて推定した。対象コーパスは、将来の道路計画に市民の声を活かす目的で行われたアンケートの自由回答であり、予め設定された交通に関する 12 個のテーマから回答者が自由に選択し回答できる[2]。テーマの内容については図 2 に示す。各テーマから以下に示す高頻度語および「ITS」や「合意形成」のように頻度は少ないが、テーマの性質を表す単語を選んだ。

渋滞 安全 高速道路 料金 歩道 事故 便利
歩行者 景観 環境 自転車 道路 整備 高齢者
合意形成 ITS

これらと、「てほしい」に言い換え可能な要求意図表現のうち助動詞相当語句・動詞型・名詞型のうちから、意味機能の異なる下記の文末表現を選んだ[7]。

しかない たほうがよい (と思う/のではない) たらどうか たらよい (と思います/思う) ていただきたい てほしい なければならない てよい (と思う) なくては いけない べき (だと思ふ/である/ではない) 急務で (ある) 重要だ (と思ふ/と思ふ/思わ) 大切だ (と思ふ) 必要 [がある (と思います/と思ふ/と思ふ/と思ふ) /だと思ふ/である/ではない (か (と思う))] 望ましい 望む する

上記を 2.3 節で挙げた尺度によって、語間関係を推定した結果を表 1 に示す。表 1 は、尺度ごとに類似度の高い上位 5 位を並べたものである。それぞれの尺度によって得られる一般的な傾向については、山本ら[4]や文献[1, 3]が詳しく述べている。表 1 から、頻度に拠らず、要求意図表現と内容語との関係が現れていると考えられるのは、カイ二乗値、イエーツの補正公式、自己相互情報量、対数尤度比、補完類似度である。これらの結果をどのように解釈すべきかについては 4 節で考察する。

表 1 各尺度による要求意図と内容語との類似度

	類似度	語句1	語句2
共起頻度	1137	歩道	てほしい。
	677	安全	てほしい。
	638	歩道	する。
	603	渋滞	てほしい。
	473	渋滞	する。
共起と非共起頻度の比	0.04976	歩道	てほしい。
	0.029885	歩道	する。
	0.029616	安全	てほしい。
	0.024226	渋滞	てほしい。
	0.022813	安全	べきである。
ダイス相関係数	0.090446	歩道	てほしい。
	0.056321	歩道	する。
	0.055847	安全	てほしい。
	0.046143	渋滞	てほしい。
	0.043538	安全	べきである。
カイ二乗値	29.522161	安全	大切である
	27.83922	景観	重要だと思わ
	19.976544	道路整備	必要である。
	18.253464	道路整備	急務であり
	13.44452	道路整備	必要ではないか。

イエーツの補正公式	24.602609	安全	大切である
	19.57055	道路整備	必要である。
	16.713657	望む	急務であり
	15.31751	道路整備	急務であり
	14.950581	景観	重要だと思わ
自己相互情報量	2.773625	高速道路	問題だと思ふ。
	2.74447	景観	重要だと思わ
	2.716335	安全	てもよいと思ふ。
	2.627291	料金	大切だと思ふ。
	2.471481	料金	問題だと思ふ。
対数尤度比	9.000085	道路整備	必要である。
	8.443151	安全	大切である
	5.43216	道路整備	急務であり
	5.403898	道路整備	必要ではないか。
	4.574052	道路整備	急務である
補完類似度	589.245671	安全	大切である
	327.420622	道路整備	必要である。
	312.981375	道路整備	急務であり
	268.60773	道路整備	必要ではないか。
	260.136228	景観	重要だと思わ

3 主観的評価による要求の強さの分析

本節では、表現によって異なる要求意図の強さを主観的評価に基づき調査分析する。

3.1 予備実験の概要

要求の強さの段階性を明らかにするために[8]、予備実験では、要求の強さを構成する要素として、要求の 1) 要求程度、2) 要求内容に関する確信度、3) 要求内容の重要度、を仮定した。要求表現が各観点でどのように順位付けられるか、また総合的にどのような順位になるかを明らかにする。同時に、要求の強さに応じて要求内容の現れ方に異なりが見られるとすれば、この段階性を明らかにするということは要求内容を予測する手がかりともなり得る。したがって、この調査結果をもとに要求意図表現と要求内容との語間関係の傾向を分析する。

3.2 実験方法

被験者は 3 人(男性 1 人女性 2 人)で、要求意図表現を含む刺激文 43 文を用意した。刺激文は「渋滞の解消には公共交通の利用が/渋滞の解消には公共交通を利用」という節に、43 種類の要求意図表現を後続したものである。図 1 にその一部を示す。これらについて、下記の三観点それぞれを基準とした相対的な点数(1~10)付けを行う。ただし、1~10 の各値は少なくとも 1 文以上に与えられる。

観点 1: 要求程度

訴え方の強さを測る。強い要求と読める場合には高い点を付ける。

観点 2: 確信度

要求の内容に対する回答者の確信度を測る。より高い確信を持っていると解釈できる場合には高い点を付ける。

観点 3: 重要度

要求内容が自分にとって重要なものであると共感できる場合には高い点を付ける。

- 1 渋滞の解消には公共交通の利用が一番だ。
- 2 渋滞の解消には公共交通の利用が急務だ。
- 3 渋滞の解消には公共交通の利用が最優先だ。
- 39 渋滞の解消には公共交通を利用するようにしていただきたい。
- 40 渋滞の解消には公共交通を利用するようにしろ。
- 41 渋滞の解消には公共交通を利用する必要がある。

図 1 評価した刺激文の一部

3.3 実験結果と考察

3.2 節の実験から、43 個の要求意図表現に与えられた得点には表 2 に示すような被験者間の相関性が見られた。従って、三つの観点に基づき「強さ」を一般化することは妥当であるとみなし、以下、結果の考察を進める。

表 2 実験結果の相関性

被験者組	被験者 AB	被験者 BC	被験者 AC
要求程度	0.7507	0.8372	0.7435
確信度	0.7674	0.7794	0.6837
重要度	0.7584	0.8529	0.7280

各被験者が観点ごとに与えた得点を合計した各値、および三観念の得点の合計値を降順に並べ、「要求の強さ」に関する順位付けを行った。このうち、各観点について、上位 5 語下位 5 語に現れた語をまとめたものが表 3 である。すなわち、表 3 は各観点から見た場合の、表現としての度合いが強い典型的な表現、弱い典型的な表現を示したものと見える。

この結果を踏まえ、要求の強さが大きい「なければならぬ型」と小さい「てもよい型」をアンケートの 12 個のテーマごとに出現割合で比較したところ、図 2 に示すように「なければならぬ型」が「安全」「事故」「合理形成」など人の命や社会倫理に関わる重要度の高いテーマ(1, 5, 6, 9)に多く現れたことは(ドット部分)、相手に対する制約や義務づけの性質が強いことを示す。一方、「てもよい型」が「渋滞」「便利」「ITS」「道路料金」などの効率向上や利便性を追求するテーマ(2, 4, 7, 10)に多く現れている

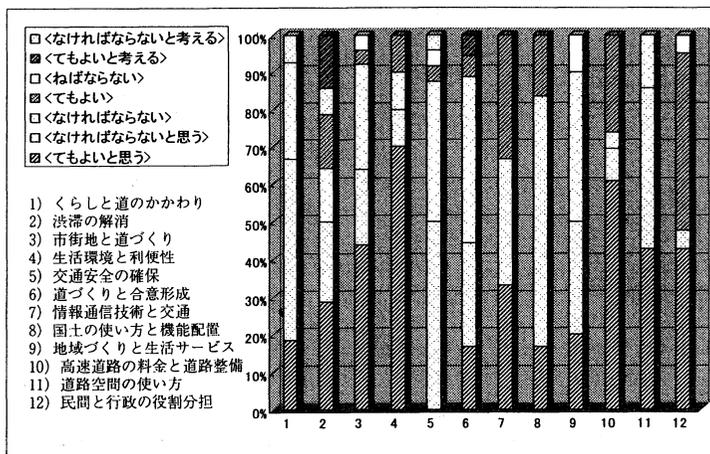


図 2 要求の強さの違いによるテーマ別出現

表 3 主観評価による要求の強さの異なり

順位	「強さ」	程度	確信度	重要度
1	しなければならぬ	しろ	するしかない	が最優先だ
2	するしかない	しなければならぬ/するよ	しなければならぬ	が重要だ/ある
3	しなくてはいけない	しなくてはいい/せよ	しなくてはいけない/	が急務だ/が一番だ
4	するほかに	しかない	するほかに	するほかに
5	するべきだ	するべきだ/	が最優先だ/が急務	するしかない
39	したほうがよいのではな	したほうがよい	したい/したいもので	したいもので
40	したほうがよいのではな	したい/したい	を/したらどうなのか	したらよい/し
41	したらどうか	してはどうか/	したらよいのではな	したらどうか/
42	しては	してはいいか	してよい/ほうがよい	しては
43	してもよい/	してもよい	しては	してもよい/し

ことは(斜線部分)、要求として相手に強い制約を課さない容認型であることを示している。

以上より、主観評価実験から要求の強さと要求内容の重要性の相関性を示すことができた。

4 推定結果と主観評価結果の比較

4.1 概要

比較対象とする語間関係は類似度計算と主観評価ともに、分析者が注目した名詞と各名詞から尺度または評価者によって関連づけられる要求表現の関係である。そして、3 節の実験を踏まえ評価者(分析者)が深い関係にあると考えた順に各名詞事に要求表現を並べたものを正解とする。この正解と表 1 のような推定結果との関連度を測る。

4.2 比較方法

対象とする名詞は評価者によって選択された「事故、歩行者、料金、便利、歩道、高速道路、安全、高齢者、渋滞」の九つとした。これは予備実験において、要求表現の出現割合を調査したところ特徴を示したテーマから連想される名詞である。この名詞ごとに 3

節で示す考察を根拠として正解が作成される。

対象となる尺度は2.3節に示す九つである。各尺度によって上記の名詞と要求表現の関係を推定した結果を正解と比較する。比較にはケンドールの順位相関係数[1]を用いる。この相関係数が1.0であれば、その結果は正解と完全に一致する。すなわち、その結果を得た尺度は正解を完璧に推定できる最適な尺度と考えられる。反対に、-1.0であれば、結果は全く正解と一致しない、すなわち、その尺度は全く適さない尺度と考えられる。表4に「歩行者」に関する各尺度において類似度が高い順に並んだ推定結果を示す。

表4 「歩行者」に関する推定結果

共起頻度	共起と非共起の比	ダイス相関係数	カイ二乗値
べきだと思	べきだと思	べきだと思	なけれはならない
べき	べき	べき	なくてはならない
べきではないか	べきではないか	べきではないか	しかない
たらどうか	たらどうか	たらどうか	べきではないか
しかない	しかない	しかない	べきだと思
なくてはならない	なくてはならない	なくてはならない	てもよいと思
なけれはならない	なけれはならない	なけれはならない	たらどうか
てもよいと思	てもよいと思	てもよいと思	べき

イエーツの補正公式	自己相互情報量	対数尤度比	補完類似度
なけれはならない	てもよいと思	なけれはならない	なけれはならない
なくてはならない	たらどうか	なくてはならない	なくてはならない
しかない	べき	しかない	しかない
べきではないか	しかない	べきではないか	てもよいと思
てもよいと思	べきではないか	べきだと思	べきではないか
べきだと思	なくてはならない	てもよいと思	たらどうか
たらどうか	べきだと思	たらどうか	べきだと思
べき	なけれはならない	べき	べき

4.3 比較結果

表5に「歩行者」について、要求表現を要求の強い順に並べた正解と、各尺度による推定結果との順位相関係数を示す。

表5 「歩行者」に関する正解と推定結果との順位相関係数

正解	共起頻度	共起と非共起の比	ダイス相関係数	カイ二乗値
なけれはならない	.82353	.38875	.90622	.91134
しかない				
なくてはならない				
べきではないか				
べきだと思				
べき				
たらどうか				
てもよい				

イエーツの補正公式	自己相互情報量	対数尤度比	補完類似度
.90963	.87980	.90708	.90793

順位相関係数から、本実験においては「カイ二乗値」がもっとも適し、順に「イエーツの補正公式」「補完類似度」「対数尤度比」「ダイス相関係数」「共起頻度」も高い順位相関を持っている。一方、「共起と非共起の比」は他の尺度に比べ、非常に相関が低かった。このことから、本実験における目的には「共起と非共起の比」は適さないと考える。

「カイ二乗値」は標本分布の代表的なカイ二乗分布の値である。これは独立検定のための尺度なので、

二語に関係がある(独立でない)ならば、高い値を持つ性質を持っている。この性質が本実験における要求表現と内容語の関係を手がかりに、要求意図や強さを把握する目的に合致していると考えられる。次に続く「イエーツの補正公式」はある条件下でカイ二乗検定の精度をあげるため補正した尺度であるので、この結果は予想できる。「共起と非共起の比」は共起頻度が高ければ、高い値を得るが、実験で取り上げる要求表現は内容語に比べ、コーパス中に頻繁に現れるものではない。そのため、頻度の少ない要求表現に関する関係を得ることが難しかったと考える。

5 おわりに

本研究では、アンケートの自由回答コーパスから要求意図表現と要求内容に現れる名詞との関係をさまざまな尺度を用いて推定し、それらの結果と、主観的に要求の強さを測った予備実験結果とを照合する。その結果から、回答者の要求の強さを知ることができる要求表現の選択基準を明らかにするという目的に適した尺度を検討した。

この検討結果を踏まえ、多種多様なコーパスや利用目的に適した評価尺度を分類することは今後の課題である。

謝辞：本稿をまとめるにあたり、富士通研究所の落谷亮氏、柳瀬隆史氏、計量計画研究所の丸元聡子氏に議論に参加していただきました。また、通信総合研究所の内山将夫氏に貴重なご意見を頂きました。ここに感謝いたします。

参考文献

- [1] 池田央, 統計ガイドブック, 新曜社, 1989.
- [2] 道路審議会基本政策部会「21世紀の道を考える委員会」ボイス・レポート, 建設省道路局 建設省都市局, 1996.
- [3] Christopher, D. Manning and Hinrich Schutze, Foundations of Statistical Natural Language Processing, The MIT Press, Cambridge MA, 1999.
- [4] 山本英子 梅村恭司, コーパス中の一対多関係を推定する問題における類似尺度, 自然言語処理 Vol.9 No.2 pp.45-75, 2002.
- [5] 内山将夫 井佐原均, 日英新聞記事の対応付けと精度評価, 情報処理学会 NL-151-3(FI-68-3), pp.15-22, 2002.
- [6] 木田敦子 山本英子 井佐原均, 後続要素を予告する表現の分析, 情報処理学会 NL-152-20, pp.137-143, 2002.
- [7] 乾裕子 内山将夫 井佐原均, 言い換えによる自由記述アンケート回答の要求意図判定基準の作成および検証, 自然言語処理学会 NLP2003, B3-4, 2003.
- [8] 柳瀬隆史 丸元聡子 乾裕子 落谷亮, 「意見の強さ」の分析のための特徴抽出, NLP2003, B7-8, 2003.