

形容詞との共起に基づく概念語の 順序付け

Ordering Concepts by Using Concept-
Adjective Co-occurrences

仁科 俊晴[♥]
鍛冶 伸裕[▲]

Toshiharu NISHINA
Nobuhiro KAJI

吉永 直樹[◆]
豊田 正史^{*}

Naoki YOSHINAGA
Masashi TOYODA

本稿では、複数の概念語を、形容詞で表される性質（「大きさ」や「高さ」など）の程度に基づいて順序付けするタスクを提起し、ウェブテキストから計算した形容詞との共起頻度をもとにこれを解く手法を提案する。提案手法により、ある特定の状況や条件下で人々がどのような手段や物事を好むかといった行動分析や、物事に対して人々が持っている印象分析など、高度な情報収集を行うことが可能になると期待できる。共起頻度を求める際に利用する共起関係としては、同一文内共起、係り受け関係、概念語が主格である係り受け関係の3種類を比較・検討する。実験では、人手による順序付けとの相関を調べることにより、順序付けにおける各共起関係の有用性を検証する。

This article presents a novel task of ordering concepts from viewpoints expressed by adjectives (e.g., “large” or “expensive”), and proposes an ordering method that uses concept-adjective co-occurrences obtained from the Web. Our method will enable advanced information retrieval such as behavior analysis under certain circumstances and impression analysis on common views towards real-world concepts. We explore three kinds of co-occurrence relations to obtain co-occurrences, namely, window-based co-occurrence, dependency, and dependency with nominative concept. In our experiments, we evaluate our method in terms of correlation between the obtained orders and manually-built orders.

[♥] 学生会員 東京大学情報理工学系研究科
nishina@tkl.iis.u-tokyo.ac.jp
[▲] 非会員 東京大学生産技術研究所
ynaga@tkl.iis.u-tokyo.ac.jp
[◆] 正会員 東京大学生産技術研究所
kaji@tkl.iis.u-tokyo.ac.jp
^{*} 正会員 東京大学生産技術研究所
toyoda@tkl.iis.u-tokyo.ac.jp

1. はじめに

我々は日常生活の中で、複数の物事に対して順序関係を与えることによって情報を整理し、意思決定に利用している。例えば、外出時に食事をするときには、「食事代の安さ」を考慮して食べる店を決めることがあるだろう。このように、物事を自身が注目する性質の程度で順序付けすることは、適切な行動を判断する過程でしばしば必要であり、特に、大量の情報が存在する現代において重要であると考えられる。

本研究ではこのような状況を鑑み、複数の概念語をそれらに共通する性質（例えば「大きさ」や「高さ」など）の程度に基づいて順序付けをするタスクを提起し、具体的に大規模ウェブテキストから得た統計量によりこれを解く手法を提案する。提案手法により、ある特定の状況や条件下で人々がどのような手段や物事を好むかといった行動分析や、物事に対して人々が持っている印象分析など、高度な情報収集を行うことが可能になると期待できる。

提案手法では、与えられた概念語と形容詞の共起頻度を概念語の持つ性質の程度の定量化に利用することで、概念語を順序付けする。具体的な共起関係として本研究では、同一文内共起、係り受け関係に基づく共起、概念語が主格である係り受け関係に基づく共起の3種類の共起関係を検討する。形容詞との共起を性質の程度の強さを測るのに利用する理由は、概念語が持つ性質の程度が顕著なほどその性質に関する記述が頻繁に行われると考えられる。

実験では、入力として複数の概念語とそれらに共通する性質を表す形容詞を与え、ブログ記事から得た共起頻度をもとに概念語の順序付けを行った。提案手法の有用性は、4人の被験者による概念語の順序付けとの順位相関係数により評価した。

本稿の流れは以下の通りである。まず2章で本研究で取り組むタスクについて説明する。次に、3章で提案手法について述べ、4章ではその評価を行う。そして、5章で関連研究について述べる。最後に、6章でまとめと今後の課題について述べる。

2. 概念語の順序付けタスク

本章では、本研究で取り組む概念語の順序付けタスクを定義し、具体例を用いて説明する。

本研究では、例えば、「カレー」、「寿司」、「ステーキ」のような複数の概念語と、それらに共通する性質を表す形容詞（例えば「安い」）を入力として与え、入力の概念語を指定された性質の程度（ここでは「安さ」）によって順序付けするタスクを考える。概念語が持つ性質の種類はある程度限られているため、複数の概念語と形容詞のペアは入力として与えられることを仮定した上で、順序付けを行う。なお、概念語には、与えられた形容詞（性質）において相互に順序付け可能な名詞（句）を考える。

概念語を順序付けする際には、(1) 概念語の指示対象が曖昧であること、(2) 程度の判断に主観が含まれることから、絶対的に正しい順序付けを定義することは難しい。(1)の場合では、例えば、食べ物を安さで順序付けしようとするときに

高いカレーもあれば安いカレーもあるように、どの「カレー」を想定するかによって順位が逆転する。しかしながら、我々は通常は物事を典型的事例によって認識しており [1], そのような認識に従えば、例えば寿司がカレーより高いというような順序付けは、異なる評価者の間でもある程度の一致が得られると期待できる。一方で、(2) の場合、例えば食べ物を美味しきで順序付けするような場合については、個人による好みの違いがあるため、順序付け結果が一致しないことが予想される。

本研究の目的は、(1) や (2) の問題が存在する概念の順序関係について、人々の最大公約数的な順序付けを導けるかという問いに答えることにある。したがって、評価においては絶対的な正解を定義することは行わず、複数人の被験者による順序付けとの相関を調べることで順序付けの良さを測ることとする。

なお、順序付けする対象によってどれだけ個人の順序付けに揺れが生じるかについては、各被験者の順序付けの間の相関係数を求めることによって確認する。

3. 提案手法

我々は、各概念語について、その程度が顕著な性質は頻繁に言及される (例えば、「クジラは大きい」という言明は「ネズミは大きい」という言明より、相対的に高頻度で書かれる) という直感に基づき、概念語と注目する性質の程度を表す形容詞の共起頻度によって概念語が持つ性質の程度の強さをスコア化することを提案する。

以下ではまず、概念語と形容詞の共起頻度の算出方法について議論する。その後、得られた共起頻度を用いて概念語を順序付けする方法を説明する。

3.1 順序付けに利用する概念語と形容詞の共起関係

本研究では、順序付けに利用する概念語と形容詞の共起関係として、同一文内共起、係り受け関係、概念語が主格である係り受け関係の3種類を考える。同一文内共起と係り受け関係を比較すると、係り受け関係では形容詞が概念語の性質を記述することが保証されているため、性質の程度の強さのスコア付けに用いる共起としてより適切であると考えられる。しかしながら一方で、係り受け関係は同一文内共起に比べて頻度が少なくなるため、入力によってはスコア付けに十分な統計量が得られない可能性がある。このように、各共起関係は質・量の観点でトレードオフの関係にあるため、本研究では、それら3種類の共起頻度計算方法を比較することにより、その優劣を検証する。

以下では、概念語「カレー」と形容詞「安い」を例として、各共起関係を説明をする。

■**同一文内共起** 同一文内で概念語と形容詞が同時に出現した頻度。例えば、「スーパーで**カレー**の特売がやっていたので、非常に**安い**値段で**カレー**が作れました。」という文の場合は(組み合わせにより)共起が2回、「**カレー**って**安い**けど、**カレー**より**安い**食べ物もたくさんあるよね。」という文の場合は(組み合わせにより)共起が4回観測されたとする。

■**係り受け** 概念語から形容詞への係り受け関係の頻度。例えば、「**カレー**って**安い**よね。」などの文から共起頻度を求める。前述の同一文内共起では、「**カレー**を食べた後行った喫茶店のコーヒーは**安かった**」といった文も共起頻度の集計対象になるが、この文はカレーの安さについて言及しているわけではなく、本論文でスコア付け算出に用いるには適切でない場合がある。係り受け関係を手がかりとすることにより、このような不適切な共起を除外することができる。なお、文中の関係の同定は既存の係り受け解析器を用いて行う。

■**主格係り受け** 概念語から形容詞への係り受け関係で概念語が主格であるものの頻度。主格の判定には、概念語に付属する助詞を用いる。具体的には、概念語と同一文節内に副助詞「は」「も」、格助詞「が」の3つの助詞のうちいずれかが出現している場合、もしくは助詞が全く出現していない場合、その概念語が主格であると判断する。係り受けではあるが、主格係り受けではなく、本論文での共起頻度の集計対象として扱うには不適切な文の例として「**カレー**より**安い**そばを食べた。」のような文がある。

なお、係り受け関係のうち、概念語から形容詞への場合のみを利用する理由は、「いつもより**安いカレー**を買った。」といった限定修飾を共起頻度の集計対象から除外するためである。上記のような文は、ある特定のカレーが安いことに関する言及であり、カレーが安い食べ物であることを述べているわけではない。このような共起を考慮してしまうと、不自然な順序付けを導く可能性がある。

なお、本研究では、概念語と形容詞の共起頻度によって性質の程度を定量化するため、形容詞の直後(同一文節中)に否定形が続く文は共起頻度を計算する対象から除外した。

3.2 共起に基づく概念語のスコア付け

概念語によって出現頻度が異なるため、3.1節で説明した共起頻度を直接スコア付けに利用して概念語を順序付けすることは適切ではない。そこで、本研究では以下のようにして共起頻度を正規化し、順序付けをする際のスコアとした。

$$\text{順序付けのスコア} = \frac{\text{概念語と形容詞の共起頻度}}{\text{概念語の出現頻度}}$$

このようにして得られたスコアをもとに降順で概念語を並べることによって順序付けを行う。

4. 評価実験

本章では、前章で述べた提案手法を用いて複数の概念語の順序付けを行う。さらに、人手で順序付けされた複数の概念語との順位相関係数を求めることにより、順序付けに用いる3種類の共起関係の優劣を検証する。

共起頻度の計算に用いるウェブテキストは、本研究室で継続的に収集しているブログ記事(2006年2月から2012年7月、約2億記事、19億文)を利用した。また、形態素解析にはKajiらの手法[2]、係り受け解析にはYoshinagaらが開発したJ.DepP¹をそれぞれ用いた。

¹ <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>

表1 評価実験に用いた形容詞と概念語集合の一覧
Table 1 List of concepts-adject pairs used in the experiments

形容詞	概念語集合
大きい	ネコ, ウマ, クマ, ウシ, ネズミ, クジラ, キリン, イヌ, ゾウ, サル
速い	自動車, 自転車, 飛行機, 電車, ヘリコプター, スクーター, 新幹線, 船
重い	タンス, じゅうたん, ソファ, 椅子, ベッド, 電気スタンド, 机, カーテン
重い	ストーブ, ホットプレート, コーヒーメーカー, 冷蔵庫, アイロン, 掃除機, 電子レンジ, トースター, 洗濯機
安い	焼きそば, ハンバーガー, チャーハン, パン, 寿司, カレー, ピザ, ステーキ, パスタ
甘い	ナシ, マンゴー, リンゴ, ミカン, モモ, レモン, カキ, イチゴ, グレープフルーツ, パイナップル
可愛い	ウサギ, リス, カメ, ウマ, ネコ, イヌ, ヒツジ, トカゲ, ハムスター, サル
美味しい	焼きそば, ハンバーガー, チャーハン, パン, 寿司, カレー, ピザ, ステーキ, パスタ

4.1 評価用データの作成

実験に先立ち、順序付けをするのに適した組み合わせである概念語集合と形容詞のペアを以下のような手順で生成し、評価データとした。

まず、入力に用いる形容詞をテキスト中での頻度を基準に選択した。具体的には、ウェブテキスト中で頻度上位の形容詞から「ない」「普通だ」「すごい」といった、具体的な性質の程度を表現しない形容詞を除いて選択した。

次に、このようにして得た各形容詞に対して、前述のブログ記事中でその形容詞に高頻度で係っていた名詞(句)の中から概念語を1つを選んだ。なお、概念語を選ぶ際に、その語と比較可能な概念語を想起しにくい場合は、その概念語は候補から除外した。例えば、「大きい」に対する概念語を選択する際には、候補としては「声」「胸」「ネコ」などがあったが、「声」や「胸」は順序付けが可能な概念語集合を想起しづらかったため、候補から除外した。結果、頻度順位が下であっても、「イヌ」や「ウサギ」などの「大きさ」を比較可能な類似概念語が存在する「ネコ」を選ぶこととなった。

このようにして各形容詞に対して得た概念語に対し、言語資源を利用して類似する概念語を収集することで入力形容詞と概念語集合を得た。本研究では、日本語 WordNet²を言語資源として用い、NLTK (Natural Language Toolkit)³の WordNet モジュール [3] を利用することで類似する概念語を最大 10 語列挙した。

得られた評価用データとなる概念語集合と形容詞のペア (8 セット) を表 1 に示す。

4.2 評価尺度

提案手法によって得られた順序と人手によって作成された順序付けの相関はスピアマンの順位相関係数を用いて評価する。本研究におけるスピアマンの順位相関係数 ρ は、順序付けをする概念語の総数を N 、順位相関を求める 2 つの順位を X と Y 、 X と Y における概念語 i の順位差を D_i 、 X と Y において同順位の概念語の個数をそれぞれ n_x と n_y としたときのそれらの順位を t_i と t_j ($i = 1, 2, \dots, n_x, j = 1, 2, \dots, n_y$) と

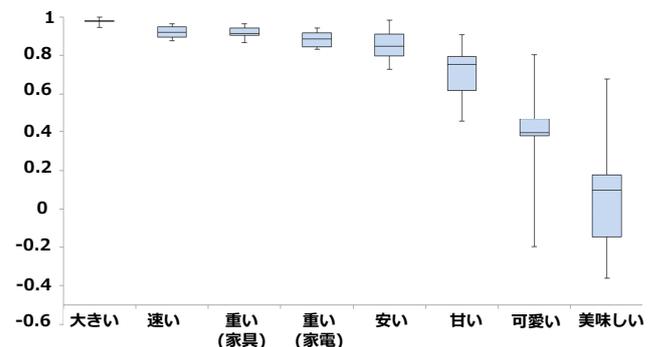


図1 被験者間の順序付けの一致度

Fig. 1 Agreement on concept orderings given by the raters

したとき、以下のように計算する。

$$\rho = \frac{T_x + T_y - \sum_{i=1}^N D_i^2}{2\sqrt{T_x T_y}}$$

$$T_x = \frac{N^3 - N - \sum_{i=1}^{n_x} (t_i^3 - t_i)}{12}$$

$$T_y = \frac{N^3 - N - \sum_{i=1}^{n_y} (t_j^3 - t_j)}{12}$$

順位相関係数は $-1 \leq \rho < 1$ までの値をとる。 $0 \leq |\rho| < 0.2$ はほぼ相関なし、 $0.2 \leq |\rho| < 0.4$ は低い相関あり、 $0.4 \leq |\rho| < 0.7$ は相関あり、 $0.7 \leq |\rho| < 1$ は高い相関あり、と評価される。 ρ が負の場合は負の相関について同様のことがいえる。

なお、スピアマンの順位相関係数は 2 つの順位の相関を求めるものであるため、被験者が与えた順序付けとの相関は、各被験者による順序付けとの順位相関係数の平均を算出することで評価した。

4.3 被験者による順序付けとその一致度

4.1 節で得られた評価用データに対し、4 人の被験者に順序付けを行ってもらった。与えられた各概念語集合の順序付けにおいて、被験者の間でどれだけ順序付けが一致しているかを計るために、被験者同士の順序付けの間の相関係数を求めた。図 1 に、各入力に対する 4 人の被験者の順序付け結果間の順位相関係数を示す。また、表 2 に、被験者の順位相関係数の平均を示す。「可愛い」、「美味しい」に基づく概念語の順

² <http://nlpwww.nict.go.jp/wn-ja/>

³ <http://nltk.org/>

表2 被験者の付けた順位の平均順位相関係数

Table 2 Average ρ between concept orderings given by the rators

形容詞	ρ (平均)
大きい	0.977
速い	0.921
重い (家具)	0.918
重い (家電)	0.884
安い	0.854
甘い	0.709
可愛い	0.377
美味しい	0.163

表3 各形容詞に基づいて順序付けされた相関係数

Table 3 ρ between concept orderings given for each adjective

形容詞	上限値	共起関係		
		同一文内共起	係り受け	主格係り受け
大きい	0.983	0.438	0.462	0.558
速い	0.949	-0.435	-0.164	-0.152
重い (家具)	0.958	0.662	0.217	0.259
重い (家電)	0.939	0.216	0.209	0.209
安い	0.924	-0.468	-0.535	-0.535
甘い	0.866	0.121	0.182	0.103
可愛い	0.722	0.439	0.428	0.418
美味しい	0.598	-0.110	-0.282	-0.282
平均	0.867	0.108	0.0632	0.0753

順序付けに関しては、主観による揺れが大きかったが、それ以外の性質に関する順序付けに関しては、順序付けが高い相関を持っていることが確認された。

4.4 実験結果

表3に被験者が付与した順序付けと提案手法が付与した順序付けとの間の順位相関係数と、全てのありうる順序付けを考慮することで計算した順位相関係数の平均値の上限値を示す。また、表4から表11に、各共起関係によって実際に順序付けされた概念語とその順位を示す。なお、表中のオラクルは相関係数が最大となる場合の順序付けを示す。表3の各共起関係の結果から確認できるように、同一文内共起を共起関係として利用した手法が最も良い結果となった。3.1節では主格係り受けは同一文内共起と比較すると共起関係を厳密に定義しているため、共起頻度こそ少ないものの形容詞が概念語の性質を記述する保証が高くなるためスコア付けに用いる共起としてより適切であると述べた。しかしながら、全ての評価用データの中で主格係り受けの共起頻度が最も多かった「美味しい」に関する順序付けの相関係数をみても同一文内共起よりも低い結果となっていた。このことから、共起関係の質と量を満たせば良い順序付け結果が得られるわけではないということが分かる。次節では概念語を不自然な順位に導く原因について考察する。

4.5 順序付け誤りの考察

前節の実験において、順序付けが不自然と思われる概念語についてその要因を調べた。本章では誤りを、想定される順位よりも高い順位となる(過大評価)と想定される順位よりも低い順位となる(過小評価)の2つの場合に分け、それぞれ原因を説明する。

4.5.1 想定される順位よりも高い順位の場合

この誤りに当てはまる概念語とその性質は、「ネズミ」の「大きさ」(表4)、「スクーター」の「速さ」(表5)、「パイナップル」「イチゴ」「ミカン」の「甘さ」(表9)、「アイロン」「掃除機」「椅子」の「重さ」(表6)、「ステーキ」「寿司」の「安さ」(表8)などが挙げられる。これらは最も厳密な主格係り受けを共起に用いても過大評価となっていたため、主格係り受けに基づく共起が、どのような文に存在していたのかを調べた。以下に、顕著に見られた文を類別して報告する。

■**意外性のある事象を取り上げて報告する文** ブログでは、ユーザは自身が意外な発見をした時にその発見を取り上げて記事にすることがある。例えば、「先日購入したスクーターは想像以上に速かった」のような文がこれに該当する。このような文が共起頻度の計算対象に含まれた結果、スクータは速い乗り物という不自然な結果が得られたと考えられる。

■**限定的な文脈で成立する事象を記述した文** 特定の条件の下でのみ成立する事象を記述した文も多く観測された。例えば、「都会のネズミは大きいんだって」のような文がこれに該当する。この文は「本来のネズミは大きくないが、都会のネズミに限り大きい」という意味であり、ネズミの(普遍的な)大きさを記述した文ではない。このような文が今回の提案手法で負の順位相関係数になった順序付け結果が得られた理由のひとつであると考えられる。

■**形容詞の多義性** 同じ形容詞に複数の意味があることが原因で概念語と形容詞の共起が多く観測された文も多く見られた。例えば、重量に関する性質を記述する形容詞「重い」に関しては「店で見るときにはいい感じだったけど、いざ家に置いてみると椅子は少し重い感じがしました」というように雰囲気に関する性質も記述をする。このような共起が多く観測された結果、椅子は重い家具、という不自然な結果が得られたと考えられる。

4.5.2 想定される順位よりも低い順位の場合

この誤りに当てはまる概念語とその性質は、「ベッド」の「重さ」(表6)、「ヘリコプター」の「速さ」(表5)、「モモ」の「甘さ」(表9)、「カレー」「パン」の「安さ」(表8)などが挙げられる。これらも過大評価と同様、3種類の共起関係全てにおいて過小評価となっていた。過小評価の理由は概念語と形容詞との間に共起関係があまり観測されなかったためである。考えられる原因を以下に列挙する。

■**概念語が持つ性質の多様性** 概念語と形容詞の共起頻度を概念語自体の出現頻度で正規化することでスコア付けを行っているため、概念語が様々な性質と共起する場合、個々の形容詞に対して共起頻度が分散し、(それぞれの性質の程度が顕著であったとしても)スコアが低くなることがある。例えば、「カレー」は「安さ」「美味しさ」の他に「辛さ」などの性質

も話題に上がりやすく、正規化によりスコアが過小評価されたと考えられる。

■**概念語の多義性** 概念語が多義である場合、概念語の頻度による正規化によってスコアが過小評価されることがある。例えば、「モモ」の「甘さ」がこの場合に当てはまる。「モモ」に関する文を見てみると、果物の「モモ」だけでなく、肉の「モモ」や人名の「モモ」などに対する文も含まれていた。これによって、概念語の出現頻度が多くなり、順序付けのスコアが低くなる。その結果、過小評価につながったと考えられる。

■**当然の事象に関する記述不足** 例えば、ベッドが重いものであったり、パンが安いものであったりすることは、一般的には当然のことである。当然の事象は実際に文として記述される機会が少ないのでその結果、順序付けのスコアが低くなり、過小評価につながったと考えられる。

5. 関連研究

著者の知る限り、我々と同種の順序付けタスクに取り組んだ研究はこれまで存在しないが、物事や概念の間に順序付けを行おうという試みはいくつか存在している。以下でこれらの研究について簡単に紹介し、本研究との違いを述べる。

倉島らは具体物を「良さ」によって順序付けするタスクに取り組んでいる [4]。この研究では、比較情報をもとに有向グラフを作成し、各ノードの評価値を求めることにより順序をつけている。彼らが想定する順序付けの尺度は「良さ」だけであり、我々のタスクは彼らのタスクを一般化したものになっている。

形容詞の極性(肯定・否定)や、記述する程度の強さを獲得する研究も行われている。Turney は形容詞が肯定的、あるいは否定的な形容詞かどうかを、その程度も含めてウェブテキストから得られた統計量をもとに計算する手法を提案している [5]。また、de Melo と Bansal は同種の性質を記述する形容詞(例: warm, hot, scorching)を、表現する程度の強さで順序付けする手法を提案している [6]。これらの研究は形容表現自体の程度の強さを順序付けしているものであり、概念語をその性質の程度に基づいて順序付けする本研究とは異なる。

6. おわりに

本研究では、類似する物事の順序付けが情報分析や高度な情報収集を実現をする上で重要であるということに着目し、複数の概念語を共通する性質の程度に基づいて順序付けするタスクを提起した。また、ウェブテキストから得られる統計量を用いて入力概念語と形容詞の共起を算出し、順序付けの指針となるスコアを得る手法を提案した。

本稿では概念語を形容詞の性質に基づいて順序付けを行うというタスクを設定したが、実際には例えば「給料の高さ」といった、より複雑な観点で順序付けをしている機会が多く存在するだろうし、順序付け対象も概念語以外に固有名詞なども考えられる。このような順序付け知識を獲得するためには、本稿で定義した共起関係をより詳細にする必要がある。

また、テキストデータを今回使用したブログ記事からマイクログログに変えて同実験を行い、結果を比較することで情

報源の違いや量による精度の差異についての議論も検討していきたい。

【謝辞】

本研究の一部は JSPS 科研費 25280111 の助成を受けたものです。

【文献】

- [1] John R Taylor. *Linguistic Categorization*. Oxford: Clarendon Press, 1989.
- [2] Nobuhiro Kaji and Masaru Kitsuregawa. Efficient Word Lattice Generation for Joint Word Segmentation and POS tagging in Japanese. *In Proceedings of IJCNLP*, October 2013.
- [3] Dekang Lin. An Information-Theoretic Definition of Similarity. *ICML*, pp. 296–304, 1998.
- [4] 倉島健, 別所克人, 戸田浩之, 内山俊郎, 片岡良治, 奥雅博. 比較情報に基づくランキング手法. *DBSJ*, Vol. 6, No. 1, 2007.
- [5] Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *In Proceedings of ACL*, pp. 417–424, July 2002.
- [6] Gerard de Melo and Mohit Bansal. Good, Great, Excellent: Global Inference of Semantic Intensities. *In Proceedings of ACL*, pp. 279–290, 2013.

仁科 俊晴 Toshiharu NISHINA

2012 明治大・理工・電気電子生命卒。東大大学院情報理工学系研究科修士課程在学中。

吉永 直樹 Naoki YOSHINAGA

2000 東大・理・情報科学卒。2002 同大大学院理学系研究科修士課程了。2005 同大大学院情報理工学系研究科博士課程了。博士(情報理工学)。2002より2008まで日本学術振興会特別研究員(DC1, PD)。2008 東京大学生産技術研究所特任研究員, 特任助教を経て現在, 同大学生産技術研究所特任准教授。計算言語学・機械学習の研究に従事。

鍛冶 伸裕 Nobuhiro KAJI

2005 東京大学大学院情報理工学系研究科博士後期課程了。情報理工学博士。2007 東京大学生産技術研究所特任助教を経て現在, 同大学生産技術研究所特任准教授。自然言語処理の研究に従事。

豊田 正史 Masashi TOYODA

東京大学生産技術研究所准教授。1994 東工大・理・情報科学卒。1996 同大大学院情報理工学研究科修士課程了。1999 同大大学院情報理工学研究科博士後期課程了。博士(理学)。同年, 科学技術振興事業団計算科学技術研究員。ウェブマイニング, ユーザインタフェース, ビジュアルプログラミングに興味をもつ。ACM, IEEE CS, 情報処理学会, 日本ソフトウェア学会各会員。

表4 形容詞「大きい」に基づく順序付け

Table 4 Concept ordering for adjective 'large'

オラクル	同一文内共起	係り受け	主格係り受け
1 クジラ	クジラ	ゾウ	ゾウ
2 キリン	ゾウ	クジラ	クジラ
3 ゾウ	ネズミ	ネズミ	キリン
4 クマ	ウシ	キリン	ネズミ
5 ウシ	キリン	クマ	クマ
6 ウマ	クマ	ネコ	ネコ
7 イヌ	ネコ	ウシ	ウシ
8 サル	サル	イヌ	ウマ
9 ネコ	イヌ	ウマ	サル
10 ネズミ	ウマ	サル	イヌ
ρ	0.983	0.438	0.462
			0.578

表6 形容詞「重い」に基づく順序付け(家具)

Table 6 Concept ordering for adjective 'heavy' (furniture)

オラクル	同一文内共起	係り受け	主格係り受け
1 タンス	タンス	タンス	タンス
2 ベッド	机	椅子	椅子
3 ソファ	椅子	じゅうたん	机
4 机	ベッド	机	じゅうたん
5 じゅうたん	じゅうたん	カーテン	カーテン
6 椅子	ソファ	ソファ	ソファ
7 カーテン	カーテン	ベッド	ベッド
8 電気スタンド	電気スタンド	電気スタンド	電気スタンド
ρ	0.958	0.662	0.217
			0.259

表8 形容詞「安い」に基づく順序付け

Table 8 Concept ordering for adjective 'cheap'

オラクル	同一文内共起	係り受け	主格係り受け
1 ハンバーガー	ステーキ	ステーキ	ステーキ
2 パン	寿司	寿司	寿司
3 焼きそば	ハンバーガー	ハンバーガー	ハンバーガー
4 チャーハン	焼きそば	ピザ	ピザ
5 カレー	ピザ	焼きそば	焼きそば
6 パスタ	チャーハン	チャーハン	チャーハン
7 ピザ	パスタ	パスタ	パスタ
8 ステーキ	パン	パン	パン
9 寿司	カレー	カレー	カレー
ρ	0.924	-0.468	-0.535
			-0.535

表10 形容詞「可愛い」に基づく順序付け

Table 10 Concept ordering for adjective 'cute'

オラクル	同一文内共起	係り受け	主格係り受け
1 ネコ	ウサギ	リス	リス
2 ウサギ	リス	ハムスター	ハムスター
3 ハムスター	ハムスター	ウサギ	ウサギ
4 リス	ネコ	ネコ	ネコ
5 イヌ	ヒツジ	ヒツジ	ヒツジ
6 ヒツジ	カメ	トカゲ	トカゲ
7 カメ	トカゲ	カメ	カメ
8 ウマ	サル	サル	サル
9 トカゲ	イヌ	イヌ	イヌ
10 サル	ウマ	ウマ	ウマ
ρ	0.722	0.439	0.418
			0.418

表5 形容詞「速い」に基づく順序付け

Table 5 Concept ordering for adjective 'fast'

オラクル	同一文内共起	係り受け	主格係り受け
1 飛行機	新幹線	新幹線	新幹線
2 新幹線	スクーター	スクーター	スクーター
3 ヘリコプター	船	飛行機	飛行機
4 電車	自転車	船	船
5 自動車	自動車	自転車	自転車
6 船	飛行機	自動車	電車
7 スクーター	電車	電車	自動車
8 自転車	ヘリコプター	ヘリコプター	ヘリコプター
ρ	0.949	-0.435	-0.164
			-0.152

表7 形容詞「重い」に基づく順序付け(家電)

Table 7 Concept ordering for adjective 'heavy' (electrical goods)

オラクル	同一文内共起	係り受け	主格係り受け
1 冷蔵庫	掃除機	掃除機	掃除機
2 洗濯機	アイロン	アイロン	アイロン
3 電子レンジ	洗濯機	洗濯機	洗濯機
4 ストープ	ホットプレート	ホットプレート	ホットプレート
5 掃除機	冷蔵庫	電子レンジ	電子レンジ
6 ホットプレート	ストープ	冷蔵庫	冷蔵庫
7 トースター	電子レンジ	ストープ	ストープ
8 アイロン	コーヒーマーカー	コーヒーマーカー	コーヒーマーカー
9 コーヒーマーカー	トースター	トースター	トースター
ρ	0.939	0.216	0.209
			0.209

表9 形容詞「甘い」に基づく順序付け

Table 9 Concept ordering for adjective 'sweet'

オラクル	同一文内共起	係り受け	主格係り受け
1 マンゴー	マンゴー	マンゴー	イチゴ
2 モモ	イチゴ	パイナップル	ミカン
3 イチゴ	パイナップル	イチゴ	パイナップル
4 ミカン	グレープフルーツ	リンゴ	マンゴー
5 カキ	リンゴ	ミカン	リンゴ
6 リンゴ	ミカン	グレープフルーツ	グレープフルーツ
7 パイナップル	レモン	レモン	レモン
8 ナシ	ナシ	カキ	カキ
9 グレープフルーツ	カキ	ナシ	ナシ
10 レモン	モモ	モモ	モモ
ρ	0.866	0.121	0.182
			0.103

表11 形容詞「美味しい」に基づく順序付け

Table 11 Concept ordering for adjective 'delicious'

オラクル	同一文内共起	係り受け	主格係り受け
1 寿司	パスタ	ピザ	ピザ
2 ステーキ	ピザ	チャーハン	チャーハン
3 ピザ	パン	パスタ	パスタ
4 カレー	ステーキ	パン	パン
5 ハンバーガー	チャーハン	ステーキ	ステーキ
6 焼きそば	寿司	カレー	カレー
7 パスタ	カレー	ハンバーガー	ハンバーガー
8 チャーハン	ハンバーガー	寿司	寿司
9 パン	焼きそば	焼きそば	焼きそば
ρ	0.598	-0.110	-0.282
			-0.282