

## 格要素間の依存関係を考慮した選択選好モデル

谷塚 太一<sup>†1</sup> 飯田 龍<sup>†1</sup> 徳永 健伸<sup>†1</sup>

選択選好の知識は省略解消や曖昧性解消などに有効な情報であり、その自動獲得の研究には長い歴史がある。これまでに提案された選択選好のモデルの多くでは、格要素と述語の間の1対1の関係しか考慮していない。これに対し、本稿では同時にもう一つの格要素と述語の共起も考慮し、2つの格要素と述語の3項関係からなる選択選好のモデルを提案する。約20年分の新聞記事を使ってモデルを学習し、評価を行う際は正解事例に対し、その事例中に出現する格要素をランダムに他の表現に置き換えた事例を誤り事例とし、その対の正誤を判定する。提案モデルは、述語と各格要素の組み合わせによっては、従来の2項関係のみを用いたモデルに比べ最大約16パーセント正解率が上昇した。

### Selectional preference model using case element dependency

TAICHI YATSUKA,<sup>†1</sup> RYU IIDA<sup>†1</sup>  
and TAKENOBU TOKUNAGA<sup>†1</sup>

There have been many attempts at automatically acquiring selectional preferences, which are useful for a variety of NLP tasks such as reference / ellipsis resolution and word sense disambiguation. Many past models only take into account binary relations between case elements and predicates. This paper proposes a ternary relation model for selectional preference, considering relations among two case elements and a predicate simultaneously. For evaluation we constructed the models using 20 years of news paper articles. Then the model was asked to distinguish a correct sentence from a negative one, where negative sentences were derived from correct sentences by replacing case elements at random. The proposed ternary model was shown to be more effective for certain case element combinations than a binary model by 16% in accuracy in the best case.

### 1. はじめに

自然言語の知識獲得の研究分野では、選択選好の知識は省略解消や曖昧性解消問題などに有効な情報として古くから研究されてきた。初期の頃からよく研究されてきたものとして、語の共起の情報を用いた選択選好知識の獲得が挙げられる。日本語を対象にした選択選好の研究では、EDR辞書<sup>1)</sup>や日本語語彙体系<sup>2)</sup>のような資源に代表されるように、どの格にどのような意味クラスの表現が入りやすいかを人手で記述する試みがなされていた。しかし、例外的な現象の知識も含めて全ての知識を人手で記述し尽くすことは困難なため、網羅性の点で問題がある。このような背景から、コーパス中に出現する共起用例を用いた選択選好の自動獲得の手法を宇津呂らや河原ら、Linらが提案している<sup>3)-5)</sup>。これらの手法では、大規模な用例集合から選択選好のモデルを構築することで、人手作成の資源が記述できていない述語と項の関係を捉えることができる可能性がある。

このような語の共起の情報を利用した研究では、述語とその項の共起用例から名詞  $n$  と動詞  $v$  について条件付き確率  $P(n|v)$  を求め、その値を選択選好の望ましさとして利用している。また、同時確率  $P(n, v)$  が  $n$  と  $v$  が独立に出現する確率の積よりも大きいかなどという情報を捉える相互情報量の値も Linらによって選択選好の望ましさとして利用されており、この相互情報量の値が大きいほど選択選好として望ましいとみなされる<sup>5)</sup>。また、相互情報量の代わりに Dice 係数が利用される場合もある<sup>6)</sup>。

データが疎になるという問題に対して、Daganらは分布類似度に基づき  $v$  に類似する  $v'$  を求め、その共起情報も利用して近似的に  $P(n|v)$  を計算している<sup>7)</sup>。どのような尺度に基づいて類語を導入するかについては、Erkがいくつかの類語計算の尺度を比較しており、この結果、Jaccard 係数と Lin<sup>8)</sup>が提案した情報理論に基づく尺度を用いた場合に最も良い結果を得たことを報告している<sup>9)</sup>。このような分布類似度を用いた研究では、どのような共起尺度を利用するかを吟味し、名詞と動詞の組み合わせが疎な場合に対応するために、分布類似度に基づく類語の共起情報も利用して確率のスージングを行っている。

しかし、共起尺度を利用して選択選好の望ましさを見積もる手法では、典型的にはある格要素と動詞の選択選好の望ましさを他に出現している格要素と独立に見積もるため、他の格要素によって動詞の意味が限定される場合に適切にその選択選好の望ましさを見積もるこ

<sup>†1</sup> 東京工業大学  
Tokyo Institute of Technology

とができない。例えば、「鮭が 証言を する」という事例は、動詞「する」に対して「鮭が」と「証言を」という格要素の組み合わせが不適切であるため選択選好として望ましくないが、従来手法のように動詞とガ格の2項のみで選択選好の望ましさを見積った場合、「鮭が する」という組み合わせは共起情報を算出するコーパス中で観測される可能性が高いため、選択選好のモデルはその頻度情報を利用してこの組み合わせを望ましいものとして出力する可能性がある。この問題を解決するためには、「鮭が」と「する」の整合性の良さを見積もる際に「証言を」という格要素の情報を加えることで動詞の意味を限定した上で、選択選好がどの程度望ましいかを判断することが必要になる。

このような格要素の依存関係の情報を導入するために、本研究では、語の共起ベクトルの和や積によって句の意味を表現する Erk ら<sup>10)</sup> や Mitchell ら<sup>11)</sup> の考え方を採用し、述語と項の組み合わせの共起頻度を推定する。彼らの手法では“catch a ball”と“catch a disease”のような“a ball”や“a disease”という文脈による“catch”の意味の違いを、“catch”の共起ベクトルと文脈となる語の共起ベクトルの和や積といった組み合わせで表現している。Erk らは語の共起ベクトルとして実際にコーパス中に出現している述語とその格要素の共起関係を文法役割ごとに抽出している。また、Mitchell らは、コーパスから得た語の共起ベクトルに対し、ベクトルの和や直積、線形結合などの演算によって、句のようなより大きな単位のまとまりの意味を表現する手法を提案している<sup>11)</sup>。

本研究では、このようなベクトルの演算を用いて語の依存関係を表現する考え方を採用し、格要素間の依存関係を考慮した選択選好モデルの構築を目的とする。本稿ではその第一歩として、従来の動詞と1つの格要素の整合性の良さを見積る選択選好のモデルに他の格要素の情報を統合する手法を提案する。2節では提案手法でのモデル作成の詳細について述べ、3節ではこの提案手法の有効性を調査するために以下の評価実験を行い、従来手法との比較と考察を行う。この際、どのような格要素が動詞の意味を特殊化するのに役立つかについても調査を行う。

- 正誤判定問題：正誤1つずつの選択選好事例のペアを一問とし、各手法で正誤の判定を行う。
  - 省略解消問題：ある動詞の省略された格要素（ゼロ代名詞）に対し、その省略を解消する名詞句を省略以前の文脈から選択選好を利用して選ぶ。
- 4節で本稿の総括と今後の課題の提示を行い、結論とする。

## 2. 格要素間の依存関係を考慮した選択選好モデル

提案する選択選好モデルは、Erk ら<sup>10)</sup> や Mitchell ら<sup>11)</sup> の考え方をもとに、選択選好を評価する対象以外の格要素の情報を語ベクトルの演算によって動詞に統合し、動詞の意味を限定する。以後、選択選好の評価対象となる格要素を評価格、文脈情報として動詞の意味を限定するために参照する格要素を参照格と呼ぶ。例えば、「鮭が 証言を する」の例において評価格をガ格「鮭が」とするとき、参照格はヲ格「証言を」となる。提案モデルでは参照格により意味が限定された動詞と評価格の共起頻度を求め、最終的にいくつかの共起尺度を用いて選択選好の望ましさを見積る。この手順で格要素の依存関係の情報を考慮することで、どのような場合でも参照格の情報を利用することができ、また、従来の選択選好モデルで利用されている共起尺度をそのまま利用することができる。以後、述語と評価格の2項で選択選好を評価する従来手法を2項モデル、さらに参照格を加えて選択選好を評価する提案モデルを3項モデルと呼ぶこととし、それぞれのモデルをどのように作成するかについて説明する。

### 2.1 参照格と述語の共起情報の統合

提案手法では3項モデルを実現するために、参照格の情報を述語の情報と統合することで既存の選択選好モデルで利用されている共起尺度を導入可能にする。この詳細を説明するために、まず2項モデルにおける共起情報とそれに基づく共起確率の計算について説明する。本研究では、述語  $v$  に対し、格要素となる名詞句  $n$  とその表層格  $c$  の組み合わせ  $nc$  の選択選好の望ましさを推定をする問題を扱う。2項モデルでは、述語を  $v_i$ 、評価格を  $nc_j$ 、共起用例中に出現する  $v_i$  と  $nc_j$  の共起頻度を  $q(v_i, nc_j)$  とした場合に、 $v_i$  の出現確率、 $nc_j$  の出現確率、 $v_i$  と  $nc_j$  の同時確率は式1のように最尤推定できる。

$$P(v_i, nc_j) = \frac{q(v_i, nc_j)}{\sum_i \sum_j q(v_i, nc_j)}, P(v_i) = \sum_j P(v_i, nc_j), P(nc_j) = \sum_i P(v_i, nc_j) \quad (1)$$

既存手法では、これらの値をもとに、相互情報量などの共起尺度の値を計算する。これに対し、3項モデルでは参照格  $nc_r$  を述語  $v_i$  に統合した上で評価格  $nc_j$  との共起頻度  $q'(v_i, nc_j)$  を式2によって近似する。つまり、もともとの  $q(v_i, nc_j)$  に参照格と評価格の共起頻度  $q(nc_j, nc_r)$  を掛けることによって  $q'(v_i, nc_j)$  を近似的に求める。この新たに導入した共起頻度を利用して式3のように共起確率を求める。

$$q'(v_i, nc_j) \simeq q(v_i, nc_j) \times q(nc_j, nc_r) \quad (2)$$

ここで、 $q(nc_j, nc_r)$  は  $nc_j$  と  $nc_r$  が共起する頻度を表す。

$$P'(v_i, nc_j) = \frac{q'(v_i, nc_j)}{\sum_i \sum_j q'(v_i, nc_j)}, P'(v_i) = \sum_j P'(v_i, nc_j), P'(nc_j) = \sum_i P'(v_i, nc_j) \quad (3)$$

3 項モデルでは式 3 に示す確率値を利用することで、参照格の情報を参照しつつ、既存手法と同様の共起尺度を利用して評価格に関する選択選好の望ましさを見積ることが可能になる。

このように 2 つの格要素の共起の出現頻度と述語とその項の出現頻度の積を利用して、その 2 つの組合せの情報を得るという枠組みは、Erk ら<sup>10)</sup> や Mitchell ら<sup>11)</sup> がベクトル空間モデルの中で語の共起ベクトルを組合せて求めた句の意味に相当する。彼らはこのような統合の方法を採用することで動詞の意味の特定化が実現でき、その特定化を考慮した意味の類似度を評価した結果、従来の個別の語の共起に基づく手法と比較して良い結果を得たことを報告している。このため、今回取り組む選択選好の知識獲得についても同様の結果が得られる可能性がある。

## 2.2 次元圧縮を用いた確率のスムージング

2.1 で導入した共起確率の積を用いることで参照格の情報を既存手法で利用されている共起尺度の計算時に考慮できるようになったが、評価格と参照格の 2 つの格要素の出現を見る場合には評価格のみを参照する場合に比べて、共起頻度はさらに疎になる。例えば、本稿の 3.3 節で報告する選択選好の評価実験では、評価格のみを参照する場合でも約 2 割の事例が大規模コーパス中に出現しておらず、また 2 つの格要素の組み合わせについては約 7 割がコーパス中に出現していないため、共起確率を求めることができない。

そこで、共起頻度のスパースネスを解消するために一般的に利用される probabilistic latent semantic indexing (pLSI)<sup>12)</sup> を導入し、2 項関係もしくは 3 項関係の頻度のスパースネスの解消を試みる。pLSI では隠れ変数  $z_k$  を導入することで、式 4 のように共起の確率を計算する。

$$P(v_i, nc_j) = \sum_k P(v_i|z_k)P(nc_j|z_k)P(z_k) \quad (4)$$

ここで、この式を 3 項モデルでも利用できるように拡張することを考える。まず、式 5 に示すように参照格の情報を pLSI の式に導入する。

$$P'(v_i, nc_j) = P(\langle v_i, nc_r \rangle, nc_j) = \sum_k P(\langle v_i, nc_r \rangle|z_k)P(nc_j|z_k)P(z_k) \quad (5)$$

この値を計算するためには  $P(\langle v_i, nc_r \rangle|z_k)$  を求める必要があるため、まず式 6 のようにベイズ則で変換した後、 $P(z_k|v_i, nc_r)$  の値を式 7 のように  $P(z_k|v_i)$  と  $P(z_k|nc_r)$  の積で近似する。ここで、 $v_i$  もしくは  $nc_r$  から隠れクラス  $z_k$  を生成する値の積を求めることは 2.1 の式 2 で共起の積を求めることによって参照格の共起情報を述語の共起に統合したことを隠れクラスを媒介させて実現することに相当する。

$$P(v_i, nc_r|z_k) = \frac{P(z_k|v_i, nc_r)P(v_i, nc_r)}{P(z_k)} \quad (6)$$

$$P(z_k|v_i, nc_r) \simeq \frac{P(z_k|v_i)P(z_k|nc_r)}{C} \quad (C \text{ は正規化項}) \quad (7)$$

## 3. 評価実験

本節では、提案する選択選好モデルと従来手法 (2 項モデル) を比較するために 2 つの評価実験を行う。各モデルの選択選好の評価値 (選択選好の望ましさの評価値) は、2 項モデルと 3 項モデルで計算される評価値に加え、2 項モデルと 3 項モデルの評価値の最大値と最小値についても比較対象とする。ここで前者を最大値法、後者を最小値法と呼ぶ。

### 3.1 言語資源

各モデルで使用する確率値の計算には毎日新聞と日経新聞合計約 20 年分の記事から抽出した 47,401,881 事例を用いる。また、評価用データとして、NAIST テキストコーパス<sup>13)</sup> から省略を補完した形で述語項構造を抽出した事例集合を利用する。評価用データ中の事例は、述語の基本形とその必須格となるガ格、ヲ格、ニ格が含まれる。

また、2.2 項で述べたように、これらのデータで計算を行うと、評価用データのうち動詞と評価格の 2 項の組み合わせでは約 2 割、参照格も加えた 3 項の組み合わせでは約 7 割程度がコーパス中に出現しなかった。そこで、本稿では pLSI を用いてスムージングした結果についてのみ掲載する。なお、本実験では pLSI の隠れクラス数を 1000 に設定した。

### 3.2 評価値の計算に用いる共起尺度

本実験で評価値の計算に用いる共起尺度として、Dice 係数、相互情報量、これらの重み付きの 4 種を採用した。それぞれの計算式を表 1 にまとめる。

### 3.3 正誤判定問題

#### 3.3.1 問題設定

評価用データの正例である述語項構造から擬似的に負例を作成し、この正例と負例の正誤判定を行う。実験では、正例と負例の両方の評価値を選択選好モデルを用いて求め、正例の

表 1 共起尺度の計算式

	2 項モデル	3 項モデル
Dice 係数	$\frac{2 \times P(v,nc)}{P(v)+P(nc)}$	$\frac{2 \times P([v,nc'],nc)}{P(v,nc')+P(nc)}$
重み付き Dice 係数	$P(v,nc) \times Dice$	$P([v,nc'],nc) \times Dice$
相互情報量	$\frac{P(v,nc)}{P(v) \times P(nc)}$	$\frac{P([v,nc'],nc)}{P(v,nc') \times P(nc)}$
重み付き相互情報量	$P(v,nc) \times MI$	$P([v,nc'],nc) \times MI$

評価値が負例の評価値を上回れば正解として正解率を求める。また、評価格と参照格の組み合わせごとに評価用データを 6 種に分けて実験を行った (表 2)。

表 2 正誤判定問題の格の組み合わせによる評価用データ分類

評価格	参照格	問題数
ガ	ヲ	26,893
ガ	ニ	14,663
ヲ	ガ	26,893
ヲ	ニ	3,873
ニ	ガ	14,667
ニ	ヲ	3,873

### 3.3.2 負例の作成方法

評価に用いる負例を作成する際には、正例の中の評価格を、名詞の出現分布に基づき他の名詞とランダムに入れ替えて負例を作成する。ただし、大規模コーパス中に出現した述語とその係り受け関係にある格要素の組み合わせには置き換えないこととする。例えば、「首相が 会議に 臨む」という正例では「首相」という表現を「会議に臨む」という文脈では望ましくない「規模」という表現に置き換えた「規模が 会議に 臨む」を負例とする。この問題では、参照格「会議に」と動詞「臨む」の組み合わせに対して、評価格(「首相が」or「規模が」)の選択選好を評価する。

### 3.3.3 実験結果と考察

6 種の格の組み合わせごとに各手法、各共起尺度を用いて問題を解き、その平均正解率を用いて評価を行った。各問題を解いた際の平均正解率を各共起尺度、各手法、格の組み合わせごとに表 3 ~ 6 に示す。また、表中で各格の組み合わせごとの最高正解率の部分を太字で示す。結果として、それぞれの共起尺度で 3 項モデルの正解率が 2 項モデルの正解率を上回る

表 3 正誤判定問題実験結果 : Dice 係数

評価格 / 参照格	2 項	3 項	最大値法	最小値法
ガ / ヲ	.893	.889	<b>.899</b>	.893
ガ / ニ	.868	.836	<b>.877</b>	.835
ヲ / ガ	.929	.885	<b>.930</b>	.885
ヲ / ニ	.918	.897	<b>.924</b>	.897
ニ / ガ	.899	.859	<b>.903</b>	.858
ニ / ヲ*	.879	.889	<b>.897</b>	.884
平均	.898	.876	<b>.905</b>	.875

表 4 正誤判定問題実験結果 : 重み付き Dice 係数

評価格 / 参照格	2 項	3 項	最大値法	最小値法
ガ / ヲ*	.887	.905	.890	<b>.906</b>
ガ / ニ	.864	.850	<b>.866</b>	.849
ヲ / ガ	<b>.920</b>	.873	<b>.920</b>	.873
ヲ / ニ	.912	.892	<b>.913</b>	.893
ニ / ガ	.892	.844	<b>.894</b>	.844
ニ / ヲ*	.871	.881	<b>.883</b>	.878
平均	.891	.874	<b>.894</b>	.874

表 5 正誤判定問題実験結果 : 相互情報量

評価格 / 参照格	2 項	3 項	最大値法	最小値法
ガ / ヲ	.893	.890	.889	<b>.901</b>
ガ / ニ*	.851	.858	<b>.865</b>	.851
ヲ / ガ	<b>.947</b>	.907	.946	.921
ヲ / ニ	.921	.919	<b>.941</b>	.909
ニ / ガ	.929	.898	<b>.932</b>	.909
ニ / ヲ	.917	.908	<b>.925</b>	.914
平均	.909	.897	<b>.916</b>	.901

表 6 正誤判定問題実験結果 : 重み付き相互情報量

評価格 / 参照格	2 項	3 項	最大値法	最小値法
ガ / ヲ*	.637	.800	<b>.828</b>	.629
ガ / ニ*	.641	.761	<b>.804</b>	.627
ヲ / ガ	.905	.814	<b>.925</b>	.877
ヲ / ニ	.852	.811	<b>.892</b>	.835
ニ / ガ	.837	.809	<b>.896</b>	.807
ニ / ヲ*	.785	.813	<b>.879</b>	.767
平均	.776	.801	<b>.871</b>	.757

格の組合わせが存在した。正解率が2項モデルを上回ったのは、表3～6中の“\*”を付けた7つの組合わせである。この格の組合わせは、評価格がガ格である組合わせ2種と評価格が二格で参照格がヲ格である組合わせの3種であった。これは、ガ格の選択選好問題を解くためにはヲ格や二格の情報が、また、二格の選択選好問題を解くためにはヲ格の情報が有用であることを示している。言語学では一般に、内項(ヲ格、二格など)の方が外項(ガ格)よりも依存関係への影響が強いとされている。つまり、依存関係への影響の強い内項を参照格として用いた場合、3項モデルが有効となっているといえる。しかし、その逆の格の組合わせでは3項モデルは有効でない。これらの問題は元々の2項モデルでの正解率が非常に高く、もとより評価格が影響の強いものであったためと考えられる。また、二格に対してヲ格の情報が効果的であるというのは、表3、表4、表6から分かるが、これはヲ格の方が二格よりも依存関係への影響が強いことを示している。以上より、格の種類による依存関係への影響の強さは、経験的に「ヲ格 > 二格 > ガ格」のようになる。このことから、「評価格の強さ < 参照格の強さ」である場合、3項モデルは従来の2項モデルより高い評価能力を持ち、有効であるといえる。

最大値法では、相互情報量の評価格ヲ格/参照格ガ格(表5)以外は2項モデルを上回る正解率であった。正解率が2項モデルを上回らなかった点に関しても、2項モデルとの差は僅かであった。この手法では、2項モデルと3項モデルの評価値の最大値をとるという定義のため、どちらかの評価値が高い、つまり依存関係の考慮がある場合とない場合のどちらか一方のみでも選択選好として望ましい場合に最終的な事例の評価値が高くなるという性質をもっている。つまり、この問題では平均正解率が高くなる場合が多いといえる。本実験で最大値法が最も改善部分が多かったのは、この性質によるものだといえる。

最小値法では、Dice係数の評価格二格/参照格ヲ格、重み付きDice係数の評価格ガ格/参照格ヲ格と評価格二格/参照格ヲ格、相互情報量の評価格ガ格/参照格ヲ格の4つの組合わせで2項モデル以上の正解率であった。これらは2項モデルと3項モデルが同程度の正解率であり、それより良い結果になった。全体的に、これ以外のどちらかのモデルの正解率が低いときは、より低い正解率を記録するなど安定しなかった。この手法では、2項モデルと3項モデルの評価値の最小値をとるという定義のため、両方の評価値が高い、つまり依存関係の考慮がある場合とない場合の両方で選択選好として望ましい場合でなければ最終的な評価値が高くないという性質を持っている。つまり、この問題では平均正解率が高くなる場合が非常に限られている。そのため、この手法が高い性能を持つ場合は少なかったと考えられる。

最大値法、最小値法の2手法の試行により、入力事例ごとのモデルの併用の可能性が示唆された。実際、平均でも最大値法が常に2項モデルを上回っていた。本実験では最大値(最小値)をとるという単純な方法を用いたが、より洗練された方法により事例ごとにモデルを使い分けすることができれば、より高い評価能力をもつ選択選好モデルとなる可能性があるといえる。実際、人手により問題ごとにモデルを使い分けた場合、各格の組合わせ、各共起尺度の正解率は、表7のようになり、非常に高い正解率となる。これは2項モデルと3項モデルを理想的に組み合わせた場合の正解率の上限値となる。

表7 正誤判定問題：正解率の上限値

評価格/参照格	Dice 係数	重み付き Dice 係数	相互情報量	重み付き相互情報量
ガ/ヲ	.965	.959	.957	.900
ガ/ニ	.979	.971	.971	.959
ヲ/ガ	.974	.958	.979	.960
ヲ/ニ	1.00	1.00	1.00	1.00
ニ/ガ	.988	.980	.989	.982
ニ/ヲ	1.00	1.00	1.00	1.00

3.4 省略解消問題

3.4.1 問題設定

ある動詞の格要素が1つだけ省略されており、そのゼロ代名詞の先行詞の候補が複数提示されている状況で正しい先行詞を選択選好の評価値のみを用いて選択する問題を考える。先行詞の選択には選択選好に加えて、出現箇所やその語の顕現性など他の要因が関与するため、選択選好だけは適切には先行詞を選択することができない。そのため、ここでは先行詞を候補集合の中でどのくらい上位にランク付けできるかで選択選好モデルの評価を行う。具体的には、問題ごとに全ての選択肢に対して評価値を求め、その評価値の降順にランク付けを行い、その結果、上位に正解の選択肢が位置することが望ましい。また、本実験でも正誤判定問題と同様に、参照格と評価格の組み合わせごとに評価を行う。各組合わせの出現数を表8にまとめる。

3.4.2 選択肢の作成

NAIST コーパス中のある動詞の省略された格要素(ゼロ代名詞)が存在するとき、その動詞以前の文書中に出現している名詞句を選択肢とする。この選択肢の中でゼロ代名詞と照応関係にある名詞句を正解、それ以外の名詞句を不正解としている。照応関係はNAIST

表 8 省略解消問題の格の組合わせによる評価用データ分類

評価格	参照格	問題数
ガ	ヲ	12,562
ガ	ニ	5,459
ヲ	ガ	2,926
ヲ	ニ	491
ニ	ガ	1,858
ニ	ヲ	746

コーパスのものを用いた。なお、本実験ではゼロ代名詞の部分の評価格としている。

例えば、以下のような記事があったとする。

[大統領が防空壕に避難した。空爆の後、連合側に使者を派遣し、(φが)休戦を提案した。]

この文章には「(φが) 休戦を 提案した」というガ格のゼロ代名詞 φ を含んだ文がある。このゼロ代名詞を含む文の前方文脈には「大統領が 防空壕に 避難した」という文があるので、名詞句「大統領」、「防空壕」、「空爆」、「使者」が出現しているため、この4つの名詞句を先行詞候補とする。さらに、コーパスの情報から、ゼロ代名詞 φ と「大統領」が照応関係にあるため、「大統領」は正解の選択肢とする。ただし、あるゼロ代名詞について前方の複数の表現が先行詞となる可能性があるため、正解は一つとは限らない。また、同じ表現が記事中に複数出現しており、そのうちの1つがゼロ代名詞と照応関係となっている場合がある。この場合、同じ表現が正解と不正解になるという選択選好では扱うことができない問題ができてしまう。選択選好を評価するという観点からは正解となり得る表現を選択することができれば良いため、同じ表現が複数出現する場合はそのいずれかが正解の場合は正解として、それ以外の場合は不正解として問題集合に1つだけ表現を加えることとした。これにより、実際の状況より簡単な問題となることに注意されたい。

### 3.4.3 実験結果と考察

6種の格の組合わせごとに各手法、各評価値を用いて問題を解き、その平均 MRR 値を用いて評価を行う。各問題に対する MRR 値は最も上位だった正解の選択肢の順位の逆数である。例えば前項の例において、実際にランク付けを行った結果、以下のような順位になったとする。

1: 大統領, 2: 防空壕, 3: 連合側, 4: 空爆, 5: 使者, ...

このとき、正解であるものの最高順位は1位であるので、この問題の MRR 値は  $1/1 = 1$

となる。また、正解であるものの最高順位が4位であった場合、この問題の MRR 値は  $1/4 = 0.25$  となる。実験結果を各共起尺度、各手法、格の組合わせごとの平均 MRR 値を表9~12に示す。表中の最大値を示した部分は太字を示す。また、ベースラインとして2項モデルを用いた結果とランダムにランク付けして MRR 値を計算したものをを用いる(表13)。

表 9 省略解消問題実験結果: Dice 係数

評価格/参照格	2項	3項	最大値法	最小値法
ガ/ヲ	<b>.370</b>	.312	.364	.322
ガ/ニ	<b>.282</b>	.248	.279	.251
ヲ/ガ	<b>.320</b>	.299	<b>.320</b>	.299
ヲ/ニ	<b>.213</b>	.200	<b>.213</b>	.200
ニ/ガ	<b>.360</b>	.278	.357	.281
ニ/ヲ	<b>.295</b>	.228	<b>.295</b>	.228
平均	<b>.307</b>	.261	.305	.263

表 10 省略解消問題実験結果: 重み付き Dice 係数

評価格/参照格	2項	3項	最大値法	最小値法
ガ/ヲ	<b>.382</b>	.354	.381	.353
ガ/ニ	<b>.285</b>	.258	.284	.259
ヲ/ガ	<b>.320</b>	.299	<b>.320</b>	.299
ヲ/ニ	<b>.212</b>	.200	<b>.212</b>	.200
ニ/ガ	.371	.325	<b>.372</b>	.325
ニ/ヲ	<b>.299</b>	.228	<b>.299</b>	.228
平均	<b>.311</b>	.277	<b>.311</b>	.277

表 11 省略解消問題実験結果: 相互情報量

評価格/参照格	2項	3項	最大値法	最小値法
ガ/ヲ	.315	.307	.303	<b>.324</b>
ガ/ニ	.258	.242	.250	<b>.261</b>
ヲ/ガ	<b>.295</b>	.283	.292	.293
ヲ/ニ	<b>.213</b>	.198	<b>.213</b>	.208
ニ/ガ	.305	.279	.288	<b>.307</b>
ニ/ヲ	.260	.226	<b>.262</b>	.255
平均	.274	.256	.268	<b>.275</b>

表 12 省略解消問題実験結果：重み付き相互情報量

評価格 / 参照格	2 項	3 項	最大値法	最小値法
ガ / ヲ	.329	.346	<b>.360</b>	.315
ガ / ニ	.262	.255	<b>.269</b>	.257
ヲ / ガ	.315	.292	<b>.320</b>	.291
ヲ / ニ	.213	.198	<b>.214</b>	.206
ニ / ガ	.343	.314	<b>.354</b>	.312
ニ / ヲ	.275	.226	<b>.281</b>	.249
平均	.290	.272	<b>.300</b>	.272

表 13 省略解消問題ベースライン：ランダムにランク付けした場合の MRR 値

評価格 / 参照格	MRR 値
ガ / ヲ	.144
ガ / ニ	.143
ヲ / ガ	.139
ヲ / ニ	.146
ニ / ガ	.147
ニ / ヲ	.159
平均	.146

正誤判定問題と同様に、評価格がガ格である問題において 3 項のモデルを用いた場合に最も向上した部分が多かった。このことから、ガ格の選択選好を解く際には他の格によって文脈情報を考慮すべきであると言える。また、最大値法の結果が他の手法を上回っている部分が多いことも共通している。しかし、正誤判定問題ほど提案手法による改善がみられない。

原因として考えられるのは、省略解消問題は本来、選択選好以外の他の要素も考慮した上で解くべきものであるということである。省略に対する先行詞の選択には、出現箇所やその語の顕現性など他の要因が関与するため、選択選好だけは適切には先行詞を選択することができない。3 項モデルでは、むしろ、より多くの格要素を考慮することによって、格要素がノイズ化してしまった可能性がある。そのため、このように改善された部分が少ない結果になったものと考えられる。

また、人手により問題ごとに MRR 値が大きかった方のモデルを採用した結果を表 14 に示す。これは 2 項モデルと 3 項モデルを理想的に組み合わせた場合の正解率の上限値となる。人手を加えなかった場合と比べて高い MRR 値となっている。これにより、省略解消問

題でも洗練された手法でモデルの併用を行うことは有効であると言える。

表 14 省略解消問題：MRR 上限値

評価格 / 参照格	Dice 係数	重み付き Dice 係数	相互情報量	重み付き相互情報量
ガ / ヲ	.523	.533	.420	.481
ガ / ニ	.476	.485	.367	.417
ヲ / ガ	.367	.358	.328	.357
ヲ / ニ	.386	.381	.381	.380
ニ / ガ	.413	.413	.340	.395
ニ / ヲ	.454	.471	.380	.425

### 3.5 評価実験まとめ

3 項で選択選好を扱うことの有効性 実験の結果より、3 項モデルは格の依存関係への影響力が評価格よりも参照格の方が強い場合において有効であるといえ、このとき格要素間の依存関係を考慮すべきである。逆に、上記以外の場合は参照格の情報を加えない方が良いことが分かった。

格の種類ごとの格要素間の依存関係への影響差 実験の結果から、参照格として動詞と統合した場合の影響の強さが異なることが確認できた。今回用いたガ格、ヲ格、ニ格の 3 つの格では、ヲ格が最も強く、ガ格が最も弱いといえる。これは、外項よりも内項の方が動詞の意味を限定するということになる。

2 つのモデルの併用の可能性 2 つのモデルを併用すればより評価性能の高いモデルが構築できるのではないかという予測のもと、本実験では最大値法、最小値法を用いて試験的に調査を行った。その結果、最大値法でほとんどの問題で 2 項モデルを上回った部分が存在し、2 つのモデルの併用によってさらに性能が向上する可能性が示唆された。

## 4. 結 論

### 4.1 総 括

本研究では、従来の動詞と格要素 1 つ（評価格）から選択選好を評価する手法に対し、さらにもう 1 つ格要素（参照格）を加えて選択選好を評価する 3 項モデルを提案した。

そして、正誤判定問題と省略解消問題の 2 つの問題を通じてモデルの評価を行った。選択選好の評価値の計算に使用する共起尺度は Dice 係数と相互情報量およびそれらの重み付きを選択した。正誤判定問題において、3 項モデルの有効性が示された。また、考慮する格

と選択選好を見積る格の組合わせによってその有効性が異なることも分かった。しかし、省略解消問題においては、問題設定が適切でなかったと思われ、正誤判定問題と比較して3項モデルの有効性が現れなかった。

本研究の結論は、以下の3点である。

- 格の種類ごとに格要素間の依存関係への影響の強さが異なることが実験的に確認できた。
- 評価格に比べて参照格の方が影響が強いならば、格要素間の依存関係を考慮すべきである。
- 格要素間の依存関係を考慮すべき格の組合わせの場合、提案した3項モデルが有効である。

#### 4.2 今後の課題

##### 4.2.1 共起尺度の網羅的な比較

本研究では、計算の利便性および一般的に使用されているという点から Dice 係数と相互情報量の比較を行った。しかし、動詞と格要素の組み合わせについては Erk がさまざまな共起尺度について比較を行っており、Jaccard 係数や Lin の共起尺度の良さを報告している<sup>9)</sup>。しかし、その傾向が今回提案した選択選好のモデルでも同じになるとは限らないため、提案モデルでも同様にさまざまな共起尺度を比較する必要がある。

##### 4.2.2 モデルの併用

本研究では、特定の格の組合わせの場合には3項モデルが有効であるという結論に至った。一方で、従来の2項モデルを使用した方が良い結果になる場合も存在する。そこで、これらのモデルを事例ごとに使い分けることが望ましい。本研究では試験的に、単純に2つのモデルの結果の最大値を採用するものと最小値を採用するものの2つを試行した。結果的に最大値を採用するものが最も良い結果を示したが、より洗練されたモデルの組合わせ手法を用いれば更なる性能向上が見込めると考えられる。

##### 4.2.3 格要素間の依存関係の考慮の方法

本稿では、文脈情報として格要素をもう一つ考慮することによって、従来の文脈情報を無視したモデルからの改善を試みた。しかし、実際には動詞の格要素の数は2つとは限らず、本研究で用いたガ、ヲ、ニ以外の格も存在する。そのため、本研究によって従来に比べ改善はしたものの、未だ選択選好を完全に扱えているとは言えない。完全に扱うためには、ある動詞がある語義である場合にどのような格をとるのか、また、格要素にはどのような選好が存在するのか、ということなども含めた総合的な格要素集合全体の評価が必要である。また、選択選好辞書を用いて入力をフィルタリングすることなども考えられ、今後更なる研

究が必要である。

## 参 考 文 献

- 1) 日本語電子化辞書研究所. EDR 電子化辞書.  
[http://www2.nict.go.jp/r/r312/EDR/J\\_index.html](http://www2.nict.go.jp/r/r312/EDR/J_index.html), 1995.
- 2) 池原 悟, 宮崎 正弘, 白井 諭, 横尾 昭男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦. 日本語語彙大系 CD-ROM 版. 岩波書店. 1999.
- 3) 宇津呂 武仁, 宮田 高志, 松本 裕治. "最大エントロピー法による下位範疇化の確率モデル学習および統語的曖昧性解消による評価", 情報処理学会研究報告, 97-NL-119, pp.69-76. 1997.
- 4) 河原大輔, 黒橋禎夫. 用言と直前の格要素の組を単位とする格フレームの自動構築, 自然言語処理, Vol.9, No.1, pp.3-19. 2002.
- 5) Shane Bergsma, Dekang Lin, Randy Goebel. Discriminative Learning of Selectional Preference from Unlabeled Text. *Empirical Methods in NLP*, pp.59-68, 2008.
- 6) Julie Weeds and David Weir. Co-occurrence Retrieval: A Flexible Framework for Lexical Distributional Similarity. *Association for Computational Linguistics*, pages 439-475. 2006.
- 7) Ido Dagan, Lillian Lee, and Fernando C. N. Pereira. Similarity-based models of word cooccurrence probabilities. *Machine Learning*, 34(1-3):43-46. 1999.
- 8) D. Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL* pp 768-774, 1998.
- 9) Katrin Erk. A simple, similarity-based model for selectional preference. *Association for Computational Linguistics*, pages 216-223. 2007.
- 10) Katrin Erk, Sebastian Padó. A Structured Vector Space Model for Word Meaning in Context. *Proceeding of the 2008 Conference on Empirical Methods in NLP*, pages 897-906. 2008.
- 11) Jeff Mitchell, Mirella Lapata. Vector-based Models of Semantic Composition. *Association for Computational Linguistics Human Language Technology Conference*, pages 236-244. 2008.
- 12) Thomas Hofmann, "Probabilistic Latent Semantic Indexing", *Proceedings of Special Internet Group on Information Retrieval, Association for Computing Machinery Press*, pp.50-57. 1999.
- 13) 飯田龍, 小町守, 乾健太郎, 松本裕治. NAIST テキストコーパス: 述語項構造と共参照関係のアノテーション. 情報処理学会自然言語処理研究会予稿集, NL-177-10, pp.71-78. 2007.