

## 単語分散表現を用いた古事記における単語の意味の揺らぎ解析

山本 健人 (愛知大学 経営学部)

岩田 員典 (愛知大学 経営学部)

### 要旨

近年、コンピュータを用いた技術の発展は目覚ましいものがあり、とりわけ人工知能を中心とした画像処理技術や自然言語処理の進歩は顕著である。このように進歩が著しい自然言語処理の研究分野においては、物語や文学に対してコンピュータを用いた分析手法が広まりつつある。その分析手法は主に計量文献学と単語分散表現を用いたアプローチに分類される。ただし、日本の古典文学などにおいては、計量文献学的なアプローチはすでに実施されているが、単語分散表現を用いた分析は行われていない。これは、単語分散表現を用いるには、古典文学に現れる単語が持つ意味の違いを考慮した分析を行う必要があるが、その違いの抽出や検討が困難なためである。そこで、本研究では古事記における単語分散表現を生成し、古事記の頻出単語やそれらの類似単語を抽出する。また、先行研究における重要単語に対する類似単語も抽出する。そして、これらの類似単語について検証することで、古事記における単語の意味の揺らぎに着目し、その特徴と現代語での意味の違いについて検証を行う。

キーワード： 古事記、自然言語処理、単語分散表現、fastText、日本神話

### 1. はじめに

近年、コンピュータを用いた技術の発展は目覚ましいものがあり、とりわけ人工知能を中心とした画像処理技術や自然言語処理の進歩は顕著である。また、それらを応用した生成 AI (Generative Artificial Intelligence) と称されるイラストや文章を生成する人工知能も作成されている。特に ChatGPT (Chat Generative Pre-trained Transformer) <sup>1)</sup> は幅広い分野の質問に詳細な回答を生成するだけで

なく、人間が自然と感じる回答を生成することから大きく注目を集めている。

このように進歩が著しい自然言語処理の研究分野においては、物語や文学に対してコンピュータを用いた分析手法が広まりつつある<sup>2, 3, 4, 5)</sup>。それらの分析手法は主に計量文献学<sup>2)</sup>と単語分散表現<sup>4, 5)</sup>を用いたアプローチに分類することができる。ただし、日本の古典文学などについて、計量文献学的なアプローチはすでに実施されているが<sup>3)</sup>、単語分散表現を用いた分析は行われていない。これ

は、単語分散表現を用いるには、古典文学に現れる単語が持つ意味の違いを考慮した分析を行う必要があるが、その違いの抽出や検討が困難なためである。そこで、本研究では古事記における単語分散表現を生成し、古事記の頻出単語とそれらの類似単語を抽出する。また、先行研究における重要単語に対する類似単語も抽出する。そして、これらの類似単語について検証することで、古事記における単語の意味の揺らぎに着目し、その特徴と現代語での意味の違いについて検証を行う。

なお、この検証を行うにあたり、日本語の自然言語処理において問題となる分かち書きの実施や、分かち書き後の単語分散表現を生成する必要がある。本研究では分かち書きの実施には MeCab<sup>6)</sup> を利用し、単語分散表現を生成する手法として fastText<sup>5)</sup> を採用している。

本論文の構成は以下の通りである。次章で関連研究について説明し、本研究の方向性や日本語の自然言語処理における問題点についても述べる。第3章では本研究で実施する分析手法について解説する。また、第4章では分析を実施したハードウェアとソフトウェア環境について説明する。そして、第5章で分析結果を示し、それらの結果に対する考察を行う。最後に第6章で本研究のまとめと今後の課題について述べる。

## 2. 関連研究

本章では、物語や文学の分析にコンピュータを用いた計量文献学と単語分散表現の概要、先行研究について説明をする。また、これらを行うにあたり生じる日本語独自の問題点と、その解決方法についても述べる。

### 2.1. 計量文献学

計量文献学的研究は著者の推定、真贋判定、成立年の推定、成立順序の推定など多岐に渡って行われている<sup>7)</sup>。真贋の判定についての先行研究には村上らの「真贋の科学：計量文献学入門」<sup>2)</sup>がある。これは日蓮の遺文である「三大秘法稟承事」について分析したものである。この分析において、村上は日蓮の真作 24 編（日蓮本人が著したとされる文献）、日蓮の贋作 16 編（日蓮の作品といわれていたが別人が表した文献）、日蓮門下の作品 5 編、真贋不明作品 5 編の計 50 編に対して、文の長さや品詞の出現率、係り受けの関係などを用いた文の構造に関する変数でのクラスタ分析と、単語自体に着目し、単語の出現率を用いたクラスタ分析を実施した。そして、クラスタ作成においては、分析に用いる変数の組み合わせとクラスタの結合法を変更し最良と考えられるクラスタを選択している。この分析の結果、「三大秘法稟承事」は全て日

蓮の著作のクラスタに分類されており、村上は「三大秘法稟承事」は日蓮の真作であると結論付けている。

また近年、民話や伝説を含む民間伝承のデジタル・アーカイブ化が進んでおり、佐藤ら<sup>8)</sup>は民間伝承の計量分析を行っている。物語・民間伝承の計量的研究としては、文化系統学と呼ばれる進化生物学における系統学の手法を文化に適用する方法が存在する。このアプローチは物語の系統関係や起源の推定などを数理的に検討するものである。佐藤ら<sup>8)</sup>は、怪異や妖怪伝承を対象に特化したデータベース「怪異・妖怪伝承データベース」を計量分析するために、その足掛かりに全体像を把握することを目的として研究を行った。「怪異・妖怪伝承データベース」は民族調査などの結果採取された事例が2021年8月時点で35,414件収録されている。データセットとして地域(都道府県)、オニやムジナなどのタグ情報を用いて分布の分析などを実施している。この研究結果として、資料数としては長野が最も多く2,967件、北海道が最も少なく73件であり地域的偏りについて注意しなければならないことや、クラスタ分析を行いタグのクラスタはランダムに分布しているのではなく、四国地方や東北地方といった地理的に近接した地方区分と部分的に一致することを明らかにしている。

その他にも、物語の構造を統計的に研

究した成果もある。吉田ら<sup>9)</sup>は英雄神話物語の構造化を目標とし、物語分析及び計量的分析から特長を抽出している。この分析においてはギリシア神話、北欧神話、ケルト神話が用いられた。ギリシア神話が対象とされたのは、本神話関する先行研究が多く存在するからである。また、北欧神話とケルト神話は、舞台がギリシア神話の近隣地域であり文化的差異が現れやすく比較研究が容易であることから対象とされている。分析では登場人物の行動を機能として扱う機能分析を用いられている。この分析の結果、既存の神話学研究における特徴を抽出することができ、神話を計量的に分析することに成功したことが示されている。

## 2.2. 単語分散表現による分析

自然言語処理において機械学習を活用するためには、単語の持つ性質や意味を反映したベクトル表現を利用することが求められる。そこで、単語をベクトルで表現することで単語の意味を定量的に把握することを目差しているのが、単語分散表現による分析である。

文字や単語をベクトル化するもっとも単純な手法として「One-hot ベクトル」がある。これは、ベクトルのすべての要素のうちひとつだけが1であり、残りはすべて0であるベクトルを意味する。このように作成したベクトルを各単語に割り

当てる。この手法では、機械的にベクトルを割り当てることができ、同一単語であるかどうかの判定が可能になる。しかし、これでは単語の意味を含んだベクトルにはなっておらず、文章中の全ての単語を対象とすると、ベクトル数が膨大になる（図1）。

文章中	(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
の	(0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
全て	(0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
の	(0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
単語	(0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
を	(0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
対象	(0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
と	(0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)
する	(0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)
と	(0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)
ベクトル数	(0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0)
が	(0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0)
膨大	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0)
に	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)
なる	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0)

図 1: One-hot ベクトルの例

図1のように文章が長くなればなるほどベクトルが多くなり次元数も増えるため、One-hot ベクトルを用いた分析は計算時間が増加してしまうという問題がある。

そこで、考え出されたのが単語を固定長の高次元の実数ベクトルで表現する「単語分散表現」である<sup>10, 11)</sup>。単語分散表現では、近い意味の単語を近いベクトルに対応させるのが基

本となる。また、ベクトルの足し算が単語の意味の足し算に対応する「加法構成性」などを中心に、理論や応用の研究が進んでいる。加法構成性とは「 $\text{vector}(\text{“King”}) - \text{vector}(\text{“Man”}) + \text{vector}(\text{“Woman”}) \rightarrow \text{vector}(\text{“Queen”})$ 」<sup>4)</sup>のようにベクトルの計算により単語の持つ意味を計算できるようにしたものである。

文章からこのような性質を満たす単語分散表現を生成するモデルの一つが Tomas Mikolov らが 2013 に開発した Word2Vec である<sup>4)</sup>。Word2Vec におけるモデルとしては Skip-gram モデルと CBOW (Continuous Bag-Of-Words) モデルがある。Skip-gram モデルは一つの単語を中心として、その周辺にどのような単語が出てくるかということを穴埋め問題のように学習する。一方で、CBOW モデルは周辺の単語からどの単語が用いられるかということを学習する。Skip-gram モデルは処理速度は遅いが精度においては CBOW モデルよりも優れている。反対に、CBOW モデルは精度の点で Skip-gram モデルに劣るが、処理速度の面では優れているという特徴がある（表 1）。

また、Piotr Bojanowski ら<sup>注 1)</sup>によって開発された Word2Vec の後継の fastText<sup>5)</sup>も存在する。fastText においては 1 つ 1 つの単語を学習する際にサブワードに分割して学習を行う。例として、「人工知能」という単語を学習する

表 1: Skip-gram モデルと CBOW モデルの比較

モデル名	Skip-gram モデル	CBOW モデル
学習方法	1つの単語を中心として、その周辺にどのような単語が出てくるかを予測	周辺の単語からどの単語が用いられるかを予測
処理速度	遅い	速い
精度	高い	低い

ことを考える。Word2Vec の場合、文章中の「人工知能」という 1 語についてはそのまま「人工知能」として学習を行う。一方で、fastText の場合には「人工知能」という単語を「人工」、「工知」、「知能」というように単語を分割して学習を行う。そしてこれらのサブワードに対応したベクトルの和の値が「人工知能」という単語のベクトルの値となる。そのため、学習データとして扱うコーパスに存在しない未知の単語に対してもベクトル化することができるという特徴がある。また、このサブワードの作成方法は単純であるため、処理全体に与える影響は小さく Word2Vec と比較して大幅に遅くなるということはない。一方で、サブワードを用いることで学習元となる単語数が少ない場合でも、Word2Vec よりも高精度であるという特徴をもつ。

単語ベクトルの類似度を用いて作者の含意を分析した研究には金子らの「単語ベクトルの類似度を用いた英米文学の通時的含意分析」<sup>12)</sup>がある。この研究は

物語作品において、作家の使う単語と普段我々が使っている単語の意味やイメージの差異を明らかにすることに着目し、同一の作家においても時間の流れで同じ単語でも意味合いが変化するかについて分析している。分析には fastText が利用されており、アメリカの作家のハーマン・メルヴィルの時代の異なる 4 つの作品を選択し、それぞれの作品からモデルを作成している。その際に一般的な言語を表すものとして fastText が提供している Wikipedia のデータを学習したモデルを用いている。なお、分析する単語については、作者の思想や作品研究の中で強い関連を持つ 4 つの単語が採用されている。この分析の結果、作品及年代ごとに言葉の含意が微妙に変化していることを客観的かつ定量的に示すことができたと述べられている。

流行語を単語ベクトルを用いて分析した研究もある。堺らの「単語の分散表現に基づく流行語分析」<sup>13)</sup>では流行語分析の対象として、小説サイトのメタデータ

を用いている。この研究では単に出現頻度の高い単語で流行語を決める方法と、類似単語の出現頻度を考慮した方法での比較を行っている。また、類似単語の意味を考慮するために単語ベクトルが用いられている。分析の結果、流行語の推定においては単語ベクトルを用いて類似単語の出現頻度を考慮する方法が優れていることが示されている。

これらの研究から単語分散表現を用いた研究が文章中の単語の意味を分析するための手段として信頼があるものだといえる。そのため本研究においては単語分散表現を用いて、古事記における単語の分析を行う。

### 2.3. 日本語の自然言語処理

前節まで計量文献学および単語分散表現を用いた先行研究について述べてきたが、本節では日本語の文献の統計解析における問題とその問題を解決するために行われた研究について述べる。自然言語処理を行ううえでの第一歩とされるものが分かち書きである。分かち書きとは文章を形態素という言葉が意味を持つ最小単位に分割することである（図2）。

英文などは「My name is Kento. 」のように意味を成す単語ごとに空白が設けられているため、分かち書きについて考慮する必要がない。しかし、日本語の文は英文などと違い分かち書きがなされて

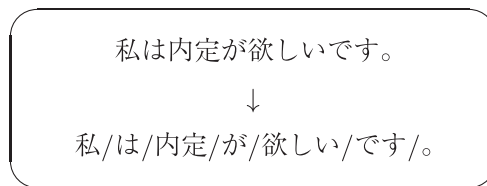


図 2: 分かち書きの例

いない。そのためどこまでが一つの単語であるかという認識が非常に困難であった。例として「東京特許許可局」という要素を分かち書きすると、「東京特許許可局」という一つの単語として扱うのか、もしくは「東京」、「特許」、「許可」、「局」のように扱うべきなのかという問題が生じる。

このような問題に対して、形態素解析と呼ばれるコンピュータを用いた分かち書きの手法が利用されている。代表的な日本語の形態素解析器には MeCab<sup>6)</sup> や JUMAN<sup>14)</sup> などがある。

また、日本における古い文献は漢文で書かれているものや和漢混合文のように漢文を含むものもある。そのため漢文を読み下して和文の状態にするべきなのかという問題も発生する。さらにコンピュータでは扱えない旧漢字もある。このように、ひらがな、カタカナ、漢字などの種類の多さなどが日本語の文献の計量分析を難しくしている。このような問題の解決への研究として小木曾らの「歴史的日本語資料を対象とした形態素解析」<sup>15)</sup>がある。この研究では単語情報がタグ付けされた本格的な通時コーパスの構

築のために、現代語用の UniDic に語彙の追加を行う形で形態素解析辞書「近代文語 UniDic」と「中古和文 UniDic」を作成している。そしてこれらの形態素解析辞書を利用した語彙認定の精度の高さから、通時コーパス構築の基板となる形態素解析システムが整ったと述べられている<sup>注2)</sup>。

### 3. 分析手法

本研究は古事記を分析対象とし、古事記における単語の特徴と現代語での特徴の違いの検証を行う。そのため、古事記において出現頻度の多い単語と先行研究から重要とみなされる単語を基準単語とし、これらに類似すると考えられる単語を古事記から抽出する。また、これら基準単語に類似する現代語の単語を fastText が提供している日本語の学習済みデータを用い抽出する<sup>16)</sup>。そして、抽出された単語について古事記と現代語での違いを考察する。これらを実現するために図3のような処理を実施する。

本章ではこれらの処理について説明をする。

#### 3.1. 古事記テキストデータのダウンロード

古事記には上・中・下巻があり、それらのテキストデータを用意する必要がある。

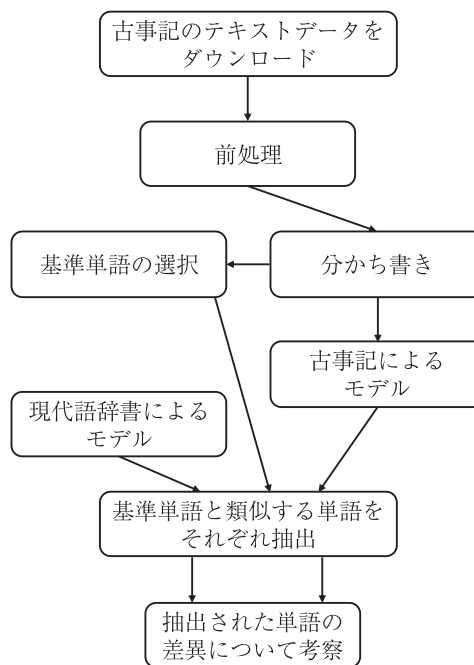


図 3: 処理の流れ

ある。また、現代語との比較を行うことから、古事記の現代語訳されたテキストデータも必要となる。モデルはこれらの現代語訳された古事記のテキストデータから学習させなければならない。そこで、分析に用いるテキストは青空文庫<sup>17)</sup>より提供されている古事記のデータを用いた。ここで提供されているものは上・中・下巻すべてがまとめられている。

#### 3.2. テキストデータの前処理

単語分散表現のためのモデルを作成するにあたって、青空文庫からダウンロードしたテキストファイルでは不要な文字列などが含まれている。そのためダウン

ロードしたファイルをそのまま利用した場合に、単語ベクトルの値に影響を及ぼす可能性がある。そこで、不要な文字列の影響を取り除くために、ダウンロードしたテキストファイルに対して以下の処理を実施した。

- ヘッダー部分の削除
- 「|」の削除
- ルビの削除
- 入力注の削除
- 空行の削除
- 底本の削除

### 3.3. 分かち書き

2.3節で述べたように、日本語を処理するためには分かち書きを実施する必要がある。本研究では分かち書きを行うにあたって MeCab を利用した。また、その際の辞書については MeCab を扱う際に推奨されている IPA 辞書 (ipadic)<sup>18)</sup> を用いた。

### 3.4. 基準単語の選択

古事記での単語の意味の特徴を抽出することを目的としているため、無作為な単語では古事記における特徴の抽出することが難しい。そこで、最初に古事記において頻繁に用いられている名詞の上位10種を抽出する。また、この頻出単語の

分析結果と古事記の先行研究から単語を選択し、これらを基準単語として用いる。

### 3.5. 古事記によるモデルの作成

モデルを生成するにあたって fastText を使い Skip-gram モデルを採用した。なお、ベクトルの次元数に関しては、テキストファイルの容量が 284KB(キロバイト)と小さいため、ファルサイズが 50MB(メガバイト)以下の目安とされている 100 とする。また、名詞・動詞・形容詞において 4 回以上の出現頻度をもつ単語が上位約 25%を占めること、4 回以上の単語の占める割合が約 82%であることから、出現頻度が 3 回未満の単語については重要度が低いと考えモデル作成には利用しない。

### 3.6. 現代語辞書のモデル

現代語辞書のモデルは fastText が提供している日本語の学習済みデータを用いる<sup>16)</sup>。これは Wikipedia などのデータを用いて作成されている。そのため記号などが含まれており、それが結果にも影響を与える可能性がある点には留意しなければならない。



### 3.7. 基準単語と類似単語の抽出

基準として選択した単語と類似度が高い単語をそれぞれのモデルを用い上位10種類抽出する。各単語は fastText によってベクトル化されているため類似度は cos 類似度を用いることで算出できる。ベクトル化された単語  $w_i$  と  $w_j$  の cos 類似度は以下の式で求めることができる。

$$\cos(\vec{w}_i, \vec{w}_j) = \frac{\vec{w}_i \cdot \vec{w}_j}{\|\vec{w}_i\| \|\vec{w}_j\|}$$

ここで、 $\vec{a} \cdot \vec{b}$  はベクトル  $\vec{a}$  と  $\vec{b}$  の内積を表し、 $\|\vec{a}\|$  はベクトル  $\vec{a}$  のノルムを表す。

### 3.8. 抽出された単語について考察

3.7節の手法で基準単語に類似していると判断された単語について、現代語との違いについて考え、古典の観点から意義のあるものであるかどうかについて考察する。

## 4. 分析環境

本研究では表2のような環境を用いて分析を実施した。

なお、3.6節で述べた fastText が提供している日本語の学習済みデータ<sup>16)</sup>は約 1.2 [GB] あり、環境によってはモデルの読み込みに時間がかかってしまう。表2の環境においてモデルを読み込むの

表 2: 実験環境

ハードウェア	
CPU	Intel® Core™ i7-9750H 2.6~4.5 [GHz]
メモリ	16.0 [GB]
ソフトウェア	
OS	Windows 10 64 bit 版
プログラミング環境	
使用言語	Python 3.9.13
パッケージ管理 実行環境管理	Conda 22.11.1
ライブラリ	fastText gensim 4.1.2 mecab-python3 1.0.5 mecab 0.996.3
統合開発環境	JupyterLab 3.4.4

に要した平均時間（試行回数 10 回）は約 381 秒であった。また、分析するために処理した古事記のデータからモデルを生成するのに要した平均時間（試行回数 10 回）は約 1.90 秒であった。

## 5. 分析結果と考察

### 5.1. 基準単語

3.4節で述べたように、古事記において出現頻度の高い名詞 10 種と古事記研究に考慮した重要語を基準単語として採用

する。出現頻度の高い名詞は表3の通りである。なお、古事記を分かち書きした結果における単語の総数は 63,597 語であった。

表3より上位 10 種類「つて」、「神」、「命」、「王」、「天皇」、「國」、「姫」、「なつ」、「時」、「おいで」と先行研究から神に関する用語である「イザナギ」、「イザナミ」、「天照らす」、「大神」、「ヤマツミ」も基準単語として採用する。

「イザナギ」、「イザナミ」は日本という国を創るという「国生み」を行う神である。その内容については世界的に類似しているものが多く国際的な比較研究もなされている<sup>19)</sup>。そのため国生みという我が国を創ったと神話上で語られている二柱に関連した単語の分析は意義がある。

「天照らす」及び「大神」という単語については、日本神話や古来の信仰における主神としての特徴を持っているため採用する。ただし、「天照らす」及び「大神」において注意しておかなければならないことは形態素解析の結果として「天照らす大神」という単語が「天照らす」と「大神」で分割されているという点である。「天照らす大神」については別の表記として「天照大神」や「天照大御神」などがあり、これらの単語で一柱の神と捉えられている。そのためどちらか片方だけの分析では不十分であると考えられる。そのためこれらの単語それぞれの類似単語

を出力する。

表 3: 古事記における名詞の出現回数

順位	単語	出現回数
1	つて	756
2	神	639
3	命	584
4	王	356
5	天皇	320
6	國	311
7	姫	304
8	なつ	285
9	時	282
10	おいで	259
11	御子	248
12	よう	234
13	の	180
14	人	176
15	こと	163
16	結婚	161
17	彦	149
18	女	141
19	天	135
20	方	133

また、「ヤマツミ」は頻出単語の「神」の類似語の分析において抽出した結果である。「ヤマツミ」という単語には山の神という意味を持っており、「ヤマツミ」を信仰の対象として扱っているということが全国で普遍的に認められていること<sup>20)</sup>から分析対象とする。

なお、「イザナギ」、「イザナミ」、「天照らす」、「大神」、「ヤマツミ」については古事記での内容に注目した先行研究を参考にしているため現代語との比較は行わない。

## 5.2. 類似単語の抽出と考察

### 5.2.1. 「つて」の結果と考察

古事記において「つて」と類似していると判断された単語を表4に、現代語において「つて」と類似していると判断された単語を表5にそれぞれ示す。

ここで、古事記において「つて」の出現頻度が高くなったのは形態素解析が適切に行われていないことが原因であると考えられる。今回採用したテキストにおいては「よつて」などの単語の「つ」が古語のため大きな状態である「よつて」という形で記されている。そのため「つて」は「伝手」というような名詞としての扱われ方ではない。しかし、形態素解析の結果において「つて」は名詞として識別されている。このことから「つて」は古事記における出現頻度で最上位ではあったが意義のある分析は不可能であると考えられる。

表 4: 古事記における「つて」の類似単語

順位	類似単語	cos 類似度
1	還	0.854673207
2	追つ	0.763501346
3	向	0.733266175
4	歸	0.731255293
5	上り	0.729565144
6	張	0.724322498
7	腰	0.721265137
8	渡つ	0.720286012
9	到つ	0.718791902
10	越え	0.716301441

表 5: 現代語における「つて」の類似単語

順位	類似単語	cos 類似度
1	つた	0.744962811
2	つたの	0.665537417
3	來上	0.575380147
4	やう	0.573528707
5	といふ	0.570816040
6	例へば	0.570641279
7	代わ	0.558806479
8	なかつ	0.557705820
9	～´{	0.554185092
10	とかいふ	0.553394377

### 5.2.2. 「神」の結果と考察

古事記において「神」と類似していると判断された単語を表6に、現代語におい

て「神」と類似していると判断された単語を表7にそれぞれ示す。

古事記における類似単語で最も類似している単語が「あらわれ」になっているが、そのほかに「出現」という語もあることから「発生する」ものという意味で利用されていると推測できる。「あらわれ」という単語は「現れる」という単語の活用した形として識別されたと考えられるため今回の分析においては自然発生的な意味が含まれた結果となった。いえる。

また、「タカミ」及び「ムスビ」という言葉は古事記において登場する神である「タカミムスビの神」、「カミムスビの神」に関わり抽出されたと思われる。この「タカミムスビの神」は古事記によれば世界ができてからアメノミナカノヌシの神に続き2番目に「出現」した神である。さらに、3番目に「出現」した神は「カムムスビの神」であり、「カミムスビの神」と表される場合もある。「タカミムスビの神」、「カミムスビの神」の二柱の神は「ムスビの神」とされ、天地初発の神であり、古事記においては万物を生成する神と解されることが多いとされている。そしてこの二柱の神は表立った活動は見せていないが神や天皇に対しては天上からの援助や指示を与えるという働きは共通しているため、司令神としての性格があるとされている<sup>21)</sup>。

表 6: 古事記における「神」の類似単語

順位	類似単語	cos 類似度
1	あらわれ	0.721442282
2	女神	0.693964481
3	ムスビ	0.693799138
4	出現	0.666783273
5	タカミ	0.661820412
6	ヤマツミ	0.660791397
7	投げ棄てる	0.643883288
8	ツミ	0.631778717
9	以上	0.630488694
10	左	0.630258560

表 7: 現代語における「神」の類似単語

順位	類似単語	cos 類似度
1	神様	0.621131241
2	神さま	0.575258672
3	バモイドオキ	0.571805537
4	クヌム	0.557660818
5	鷗剣	0.531758189
6	奈中	0.517990768
7	八百万	0.517087102
8	全知全能	0.516898096
9	喚者	0.516126394
10	御子	0.511880875

「ヤマツミ」という単語については「オホヤマツミの神」、「マサカヤマツミの神」、「ハラヤマツミの神」などといった神から出力されたと考えられる。「ヤマツミ」とは山の神という意味であり、「ヤ

マツミ」という単語を含む神は複数存在する<sup>20)</sup>。このそれぞれの神の特徴に着目しても有意であると思われる。今回の「神」という単語の類似単語として山の神に関連する単語を抽出することができた。この結果から、さらに詳細な分析を行う場合にはそれぞれの「ヤマツミ」と名の付く神それぞれの類似単語や「ヤマツミ」自体の類似単語を抽出することで更なる特徴抽出ができると思われる。

今回の分析から古事記における「神」という単語は男性格と女性格から生まれるものではなく自然的に発生するという意味を含むとこの結果から考察できる。しかし古事記内で登場する神は自然発生的なものだけでなく男性的特徴、女性的特徴を持つ神もいることは事実である。そのため「神」という1語のみで古事記に登場する神の特徴を抽出することはできなかった。しかし、これらの考察結果から類似単語を出力することで神としての性格の特徴を抽出できる可能性を示すことができた。

一方、現代語における「神」の単語は「神様」や「神さま」などのよう敬称のつけられている。加えて「全知全能」や「御子」という単語から漠然と人よりも上位の存在として使われていると判断できる。

### 5.2.3. 「命」の結果と考察

古事記において「命」と類似していると判断された単語を表8に、現代語において「命」と類似していると判断された単語を表9にそれぞれ示す。

古事記において「命」という単語は生命を意味する単語だけでなく、神の名称の最後につく「命(ミコト)」のように尊称としての使われ方をしている。そのため最も類似する単語として同じく名前の末尾につく使われ方をしている「彦」が出力されたと考えられる。また、神武天皇の名である「カムヤマトイハレ彦の命」などのように登場した神や天皇などの名において「彦」と「命」の両方が使われている場合もある。そのため、「命」のみ使っているもの、「彦」のみ使っているもの、「命」と「彦」の二つが使われているもので代表的なものの類似語を抽出することでそれぞれの特徴を捉えることに繋がられる可能性はある。このように「命」においては古事記での特徴を適切に抽出することができたといえる。

そして、現代語の「命」との違いは顕著だといえる。「びろいした」、「カラガラ」、「短し」という単語から命びろいや短さが主張されている。そのため名称の末尾の単語としての特徴はなく大切なものであるという意味が含まれているといえる。

表 8: 古事記における「命」の類似単語

順位	類似単語	cos 類似度
1	彦	0.807343662
2	ヤマト	0.772760987
3	フト	0.761658370
4	大	0.758600116
5	オホタラシ	0.757274747
6	ワカタラシ	0.753592670
7	タマ	0.747307360
8	サホ	0.746762097
9	オホヤマト タラシ	0.745998621
10	オホヤ	0.740310013

表 9: 現代語における「命」の類似単語

順位	類似単語	cos 類似度
1	びろいした	0.635943949
2	惜しけりゃ	0.627595127
3	稲飯	0.607133687
4	人命	0.604205191
5	カラガラ	0.595960200
6	びろい	0.588634133
7	モノダネ	0.585827589
8	いのち	0.585082591
9	イノチ	0.578028023
10	みじかし	0.560103178

#### 5.2.4. 「王」の結果と考察

古事記において「王」と類似していると判断された単語を表10に、現代語において「王」と類似していると判断された単語を表11にそれぞれ示す。

「王」という単語については「つて」と同様に有意な結果を示すことは不可能である。類似単語から考察すると「ヤマトラグナの王」などの「ヤマトラグナ」に当たる部分の分かち書き及び形態素解析が適切に行われていないことが原因だといえる。また、「ヤマトラグナの王」など人を指す場合での「王」は採用したテキストにおいては「きみ」と読むことへの対応が必要である。

表 10: 古事記における「王」の類似単語

順位	類似単語	cos 類似度
1	ヲ	0.786937773
2	オシ	0.771272123
3	クマ	0.767167509
4	ワカ	0.761873007
5	ケ	0.761127114
6	オケ	0.760558605
7	サカ	0.759916544
8	ヒコ	0.758204103
9	庶	0.746218443
10	ヒコイマス	0.742589235

現代語においては英語での「king」のような意味ではなく、人の名前の一部で

使われているような使い方などが抽出されている。

表 11: 現代語における「王」の類似単語

順位	類似単語	cos 類似度
1	鬪利	0.594980955
2	薔	0.591907799
3	𠂔/𠂔	0.584982336
4	思婷	0.583225250
5	桓恵	0.570285439
6	溢正	0.567358077
7	漿大	0.562781990
8	濛	0.559408784
9	珞丹	0.554068744
10	闖運	0.553217053

### 5.2.5. 「天皇」の結果と考察

古事記において「天皇」と類似していると判断された単語を表12に、現代語において「天皇」と類似していると判断された単語を表13にそれぞれ示す。

「天皇」という単語の類似単語として古事記では「反正」、「允恭」、「履中」、「崇神」、「顯宗」、「景行」、「繼體」が抽出されており、これらは歴代天皇の名前であるため意味の抽出は成功していると判断できる。「クニオシビト」については考安天皇である「オホヤマトタラシ彦クニオシビトの命」から出力されたと判断できるため「天皇」としての意味の含む特徴を示したと考えられる。また、「カムヤマト

イハレ」は神武天皇である「カムヤマトイハレ彦の命」から出力されたと見える。以上の結果から、天皇としての意味を含む単語を適切に抽出したといえる。しかし「oo 天皇」での「oo」における部分を抽出している場合と「カムイヤマトイハレ」のように天皇の名を抽出している場合があるためそれぞれの単語の比較を行うことでより詳細な分析が行えるはずである。

現代語においてもより丁寧に「天皇陛下」というような表現だけでなく歴代の天皇を出力している。ここで注目すべき点は「明治天皇」においては「明治」と「天皇」というように分けられているのではなく、1つの単語としてまとめられているという点である。これは fastText が提供している公式の学習済みモデルの違いによるものではない。「繼體」、「孝安」においては歴代天皇であるにもかかわらず「天皇」という語を含んでいない。つまり、分かち書きや形態素解析に用いられている辞書において1つの単語で1人の「天皇」という文字を含むものと含まないものがあるということを考慮しなければならないといえる。

### 5.2.6. 「國」の結果と考察

表 12: 古事記における「天皇」の類似単語

順位	類似単語	cos 類似度
1	反正	0.741382003
2	允恭	0.736334026
3	履中	0.734100521
4	崇神	0.732499242
5	クニオシビト	0.726433158
6	顯宗	0.719583154
7	クニ	0.718325675
8	カムヤマト イハレ	0.718304873
9	景行	0.716857016
10	繼體	0.714724004

古事記において「國」と類似していると判断された単語を表14に、現代語において「國」と類似していると判断された単語を表15にそれぞれ示す。

「國」という単語においては古事記内では現代での「国」のような使い方だと推測できる。日本国内の地方の国の呼び方である「近江」、「信濃」、「尾張」、「但馬」の結果からこのことがいえる。また「越え」という単語から国を越える、国をまたいで移動するような表現の中でも使われていることが考察できる。

表 13: 現代語における「天皇」の類似単語

順位	類似単語	cos 類似度
1	天皇陛下	0.67099905
2	繼體	0.63870585
3	明治天皇	0.63729787
4	懿徳	0.63011503
5	ようわてんのう	0.62016016
6	トンデモニセ	0.61946684
7	皇室	0.60639489
8	孝安	0.59636527
9	應神	0.58574194
10	クシロキング	0.57423949

表 14: 古事記における「國」の類似単語

順位	類似単語	cos 類似度
1	越え	0.712679207
2	越	0.696562409
3	野	0.690944850
4	近江	0.686984062
5	信濃	0.664271235
6	造	0.663775980
7	尾張	0.657580972
8	道	0.655382991
9	但馬	0.648845792
10	東	0.647405565



表 15: 現代語における「國」の類似単語

順位	類似単語	cos 類似度
1	鍾建	0.744857728
2	裊	0.738918602
3	蔣經	0.680568635
4	内外	0.680210471
5	阮愛	0.662821531
6	粹会	0.634811044
7	彭偉	0.631997764
8	灣建	0.629675508
9	體論	0.628085315
10	教室	0.621079624

現代語においてはこの「國」の字が旧字体であることから日本やアメリカなどといった「国」としての使い方ではなく、人名などで扱われている結果として出力されたといえる。そのため現代において「國」という言葉は一般的に使用されていないということが推測される。

### 5.2.7. 「姫」の結果と考察

古事記において「姫」と類似していると判断された単語を表16に、現代語において「姫」と類似していると判断された単語を表17にそれぞれ示す。

表 16: 古事記における「姫」の類似単語

順位	類似単語	cos 類似度
1	トヨ	0.813600123
2	庶	0.802409351
3	ミミ	0.797167480
4	サホ	0.796962142
5	タマ	0.791439354
6	マツ	0.789199889
7	オキナガタラシ	0.787300944
8	ヤサカノイリ	0.786278903
9	ヤマト	0.785481393
10	ヌカ	0.782152653

表 17: 現代語における「姫」の類似単語

順位	類似単語	cos 類似度
1	ミミア	0.610977471
2	リアンズ	0.608783543
3	於琴	0.594457746
4	彦鶴	0.591873407
5	スヤリス	0.587195992
6	乃駅	0.584578812
7	逃池	0.581275761
8	獨大	0.578622341
9	かぐ	0.577946007
10	牟宇	0.577938259

「姫」については「王」で述べたことと同様のことがいえる。まず古事記において「トヨタマ姫」という人物が中心となっている物語がある。よって、類似単語に「トヨ」、「タマ」の単語が現れたと推測で

きる。また、「サホ」については「サホ姫」という人物の登場が関係していると考えられる。ただし、この姫に関しては兄に「サホ彦」という人物がいるにもかかわらず、類似単語に「彦」は出てきていない。これらのことから「彦」と「姫」の使い方には大きな違いがあるといえる。また「ヤサカノイリ」という語に関しても「ヤサカノイリ姫」という人物が登場するため出力されたと考えられる。

現代語における「姫」については創作物から出力されたと考えられるものや歴史上の人物が大半を占めているとみられるため違いが顕著に表れている。

### 5.2.8. 「なつ」の結果と考察

古事記において「なつ」と類似していると判断された単語を表18に、現代語において「なつ」と類似していると判断された単語を表19にそれぞれ示す。

「なつ」という語についても「つて」と同様に形態素解析が適切に行われていないことが原因で出現頻度が高まっていることが考えられる。テキストの形式から「～になって」のような表現が「なつて」と表記されていることだけでなく本来動詞として扱われなければならない語であるにもかかわらず名詞と識別されている。よってこの単語では意義のある分析は不可能であると考えられる。

表 18: 古事記における「なつ」の類似単語

順位	類似単語	cos 類似度
1	授け	0.719961822
2	淡路	0.694629848
3	囊	0.691883147
4	なり	0.688225627
5	晝	0.682789981
6	イツモ	0.679984510
7	なろう	0.679590702
8	佩び	0.677898109
9	椅子	0.675706327
10	なる	0.675577581

表 19: 現代語における「なつ」の類似単語

順位	類似単語	cos 類似度
1	うみ	0.712127268
2	搭堂	0.623070359
3	メロブーム	0.569527030
4	あちゅう	0.531142473
5	afoolahoo	0.517229378
6	なつの	0.509951293
7	ふゆ	0.509363472
8	ミライプリズム	0.502020121
9	☆(*^-^)	0.498847067
10	よみかけ	0.497062802

### 5.2.9. 「時」の結果と考察

古事記において「時」と類似していると判断された単語を表20に、現代語にお

いて「時」と類似していると判断された単語を表21にそれぞれ示す。

表 20: 古事記における「時」の類似単語

順位	類似単語	cos 類似度
1	はいり	0.698769629
2	隠	0.675067067
3	遂に	0.664328575
4	處	0.664009929
5	黒姫	0.661461055
6	山口	0.658991873
7	留まり	0.652303517
8	佩び	0.645636320
9	嬢子	0.644227207
10	宴	0.644049227

表 21: 現代語における「時」の類似単語

順位	類似単語	cos 類似度
1	とき	0.754707634
2	頃	0.695726752
3	直後	0.687061310
4	際	0.681109965
5	後	0.677798808
6	直前	0.646012068
7	分	0.623385608
8	時半	0.618538201
9	午前	0.613271832
10	以降	0.602003455

古事記における「時」の類似語は時間に関連する単語は少ない。これは適切に

抽出していないといえる。実際に「はいり」という類似単語については「はいる」という動詞の活用形にもかかわらず名詞という識別になっている。「遂に」という語に関しては何かをしようとしてその時が来たかのような意味を抽出しているといえる。また「處」という語については「何處」や「其處」など旧字の形で表されたものである単語で使われている。そのため適切に形態素解析ができなかったことが考えられる。

現代語においては時間を表す単語が出力されている。そのため現代語と比較することで意義のある結果を得ることは難しい。

#### 5.2.10. 「おいで」の結果と考察

古事記において「おいで」と類似していると判断された単語を表22に、現代語において「おいで」と類似していると判断された単語を表23にそれぞれ示す。

「おいで」における結果についてもこれまでの適切に分析できていない単語と同様な結果だといえる。名詞的な使い方をしている動詞であり、類似単語についても特徴を抽出できたといえるような結果ではないと考えられる。

表 22: 古事記における「おいで」の類似単語

順位	類似単語	cos 類似度
1	いで	0.968661010
2	大宮	0.890182137
3	還り	0.890181124
4	おい	0.888275921
5	熊	0.886641562
6	難波	0.881817162
7	上り	0.880368471
8	役所	0.876894891
9	村	0.876379967
10	日代	0.875208616

表 23: 現代語における「おいで」の類似単語

順位	類似単語	cos 類似度
1	お出で	0.612858415
2	ロート ケプシェン	0.564629853
3	遊ば	0.538924038
4	お越し	0.526203930
5	いらっしゃい	0.525980532
6	来れる	0.518246770
7	手招き	0.514645219
8	ウルトラマン グラフィティ	0.511636078
9	あっち	0.505920470
10	来ん	0.503777862

### 5.2.11. 「イザナギ」の結果と考察

古事記において「イザナギ」と類似していると判断された単語を表24に示す。

表 24: 古事記における「イザナギ」の類似単語

順位	類似単語	cos 類似度
1	イザナミ	0.959555447
2	スサノヲ	0.917612612
3	大國主	0.912840664
4	ヲウス	0.904657364
5	兄弟	0.897306442
6	オホモノヌシ	0.892862201
7	高木	0.889750123
8	天照らす	0.883829176
9	アメノウズメ	0.883770525
10	イスケヨリ	0.878604114

「イザナギ」の類似単語として「イザナミ」が最も類似していると出力されている。これは内容的な面ではなく単語として考察すると fastText の学習方法によるものも考えられる。「イザナギ」と「イザナミ」という単語は1字違いであるため、サブワードとして扱われる単語も類似してくる。そのことから類似度が高いといえることができる。

内容的な面からも考察していく。「イザナギ」と「イザナミ」は5.1節でも述べたように国生みを行った神である。「イザナギ」という単語については濁らずに

「イザナキ」と表されることもあり、「イザナミ」の「ミ」に対して「イザナキ」の「キ」は男性を表す語として考えられている<sup>22)</sup>。

「イザナミ」という語に続いて二番目に類似している単語は「スサノヲ」である。この「スサノヲ」とは古事記で登場する神の1柱であり、「イザナギ」から生まれた神である。古事記内において「スサノヲ」は役目を果たさず災いを引き起こしたり、「天照らす大神」が石屋に閉じこもる原因となっているなど荒々しい神格として表されている。しかし、追放されたのちには八岐大蛇の退治や大国主に試練を与えるなど英雄神としての神格が描かれている<sup>23)</sup>。直接的な親子としての関係だけでなく、「イザナギ」が行った黄泉の国からの帰還などの出来事など神として行ったことからこの結果が出力されたと考えられる。この「スサノヲ」という語の類似語の研究にも意義があるといえる。

「大国主」は出雲大社における主神であり、国づくりという大業を成し遂げた神である<sup>24)</sup>。この国づくりと「イザナギ」と「イザナミ」が行った国生みには共通するものがあり、その結果として「大国主」という単語が出力されたと判断できる。また、「オホモノヌシ」に関しても「大国主」とともに国づくりに関わった神である<sup>25)</sup>。スクナヒコナという神が去った後に登場するが、類似単語にスクナヒコ

ナに関する語は出力されていないため、これらの物語の単語の分析においても有意な結果が得られる可能性がある。

「天照らす」という語は「天照らす大神」から出力されたものだといえる。「イザナギ」と親子関係という点では「スサノヲ」と共通していることや「スサノヲ」の災いの被害を受けている点で共通している<sup>26)</sup>。さらに、「アメノウズメ」という単語は「天照らす大神」が石屋に閉じこもった際に外へ招き出すために神がかりをした女性的神格の神である「アメノウズメの命」から出力されたと思われる。

#### 5.2.12. 「イザナミ」の結果と考察

古事記において「イザナミ」と類似していると判断された単語を表25に示す。

表 25: 古事記における「イザナミ」の類似単語

順位	類似単語	cos 類似度
1	イザナギ	0.959555507
2	兄弟	0.932725489
3	アメノウズメ	0.928923368
4	オホモノヌシ	0.927596033
5	ヲウス	0.926353335
6	オモヒガネ	0.918663621
7	オホタタネコ	0.913374841
8	オシホミミ	0.905309379
9	女神	0.902147353
10	高木	0.901556790

「イザナミ」の類似単語としては第1位に「イザナギ」が出力されているがその考察については「イザナギ」にて示したためここでは省略する。

「兄弟」という語に関しては「イザナギ」と「イザナミ」の関係から出力されたと思われる。「イザナギ」と「イザナミ」は兄と妹という関係であり、近親相姦としての描写が描かれている。そのためこの「兄弟」という単語から物語上の描かれている近親相姦の特徴の一部が抽出することができた。

「アメノウズメ」に関しても「イザナギ」と同様に出力されているが「イザナギ」と比較してみると類似度の値は高い。これは女性的神格としての特徴を抽出できたと考えることができる。

「オモヒガネ」という語に関しては、「タカミムスビの神」の子である「オモヒガネの神」から出力されたと考えられる。この神は「アメノウズメ」と添うように天照大神を石屋からおびき出す際に活躍している。その計画の際に思慮の役割を担う重要な神格である。この神は古事記だけでなく日本書紀も含め、天照大神に関わって登場していることが指摘されていることから、天照大神に重要な相談役のような立場にある神だと考えられている<sup>27)</sup>。

「オホタタネコ」とは「オホモノヌシ」の子である。「オホタタネコ」に関しては「オホモノヌシ」ほど詳細に描かれ

ていないが親子関係などからこの結果が出力されたと推察できる。

「オシホミミ」とは天照大神に連なる神である「アメノオシホミミの命」から出力されたと考えられる。この神は天照大神とスサノヲとの誓約において生まれた神である。この単語において注目すべき点は「アメノ」という文字が分割されているという点である。先述のように「アメノウズメ」に関しては「アメノ」と「ウズメ」の形で分割されていない。「アメノ」という語に関しては美称であり、高天原の系譜を引く神を意味するとされている<sup>28)</sup>。そのため分割される場合とされない場合では分析結果に誤差が出てしまうことは否めない。しかし、「イザナギ」や「アメノウズメ」など類似単語で出力されている神については高天原の系譜を引く神々であるためその特徴を抽出することはできたといえる。

そして「女神」という語が類似単語で出力されている。これは「イザナギ」の類似単語においては出力されていない単語である。そのため女性神格としての特徴を抽出したという「アメノウズメ」と同等の意味の特徴を抽出することができたといえる。

### 5.2.13. 「天照らす」・「大神」の結果と考察

古事記において「天照らす」と類似していると判断された単語を表26に、古事記において「大神」と類似していると判断された単語を表27にそれぞれ示す。

「天照らす」、「大神」については形態素解析で分割されているため、それぞれの分析を行うことには意義がある可能性はある。しかし、今回の分析結果からはこの2つの類似単語においてそれぞれの間に顕著な違いは見られない。そのため今回の分析についてはこの2つの結果の考察をまとめて行うこととする。

「夢」という語に関しては「天照らす大神」が夢に現れお告げをするという描写から出力されたと思われる。「神」の分析における「タカミムスビの神」と同様に「天照らす大神」には司令神としての性格があることが言われている<sup>21)</sup>。夢に現れ、指示を出すというような描写から類似単語として「夢」が出力されたと考えられる。「夢」については日本霊異記における夢告の構造の研究を行った「日本霊異記における夢告譚の構造」<sup>29)</sup>でも示されているように現代と比較して重要な意味を持っており、特色のある構造がある。そのため古事記における「夢」に関する描写について分析すると違った見解が得られる可能性がある。

表 26: 古事記における「天照らす」の類似単語

順位	類似単語	cos 類似度
1	大神	0.944514096
2	夢	0.906404316
3	スサノヲ	0.886729956
4	陛下	0.885306895
5	問わ	0.884806395
6	大國主	0.884797633
7	兎	0.877385676
8	イザナギ	0.874871194
9	病氣	0.873751700
10	イザナミ	0.873235345

表 27: 古事記における「大神」の類似単語

順位	類似単語	cos 類似度
1	天照らす	0.941259921
2	スサノヲ	0.859057665
3	大國主	0.829224646
4	夢	0.818274617
5	イザナギ	0.817764461
6	イザナミ	0.801825821
7	られる	0.798540533
8	降り	0.792240560
9	兎	0.792116702
10	女神	0.774135053

「スサノヲ」に関しては神格としての内容は省略する。まず「天照らす大神」と「スサノヲ」は姉弟関係にある。また、その出自としても「イザナギ」が黄泉の

国から戻った後に左目を清めたときに出現した神が「天照らす大神」であり、「スサノヲ」は鼻を清めた際に出現した神である。そのため出現の仕方からも共通点がある。しかし同じように出現した「ツクヨミ」に関する単語は類似単語には出力されていないことは注目すべき点である。

「陛下」という語においては信仰としての主神という神格や先述のような司令神としての神格から出力されたと考えられる。「天照らす大神」は超越した至高性を有する神という説もある。そして高天原を統治する神であり、天皇の祖として位置づけられている皇祖神であるという特徴を抽出した結果としての「陛下」という類似語が出力されたと考えることができる<sup>26)</sup>。

「大國主」に関しても国づくりの中の主要な神としての特徴と高天原での主要な神としての特徴を抽出した結果だと考えられる。

「兎」については「大國主」に助けられる因幡の白兎の説話の兎から出力されたと思われる。「大國主」の活躍を語る一連の神話において、「兎」は「大國主」の婚姻の達成を予言する働きを担っているという点<sup>30)</sup>から「天照らす大神」の司令神としての神格に通じる特徴を抽出したと考えられる。

「病氣」という語については「天照らす大神」が石屋に閉じこもってしまった

ときに世界に「病氣」が蔓延したということから出力されたと考えられる。しかし、この結果には疑問が残る。それは「天照らす大神」と「病氣」は言うなれば相反するものであるためである。確かに「病氣」の原因となった事象は「天照らす大神」が閉じこもったことにある。ただし「スサノヲ」の暴挙が原因で閉じこもることになったため、この分析結果から「病氣」と「天照らす大神」の関連性を一概に語ることは難しい。

#### 5.2.14. 「ヤマツミ」の結果と考察

古事記において「ヤマツミ」と類似していると判断された単語を表28に示す。

表 28: 古事記における「ヤマツミ」の類似単語

順位	類似単語	cos 類似度
1	ツミ	0.972165942
2	タカミ	0.963885903
3	ヅチ	0.957382917
4	ムスビ	0.952524185
5	女神	0.946900606
6	オカミ	0.945740223
7	オモヒガネ	0.936307788
8	フツ	0.933452368
9	オホモノヌシ	0.930348635
10	ワタツミ	0.925216973

「ヤマツミ」の類似単語としてまず「ツ



ミ」という語がある。この類似性については「イザナギ」と「イザナミ」の類似性と同様に fastText の特性も一つの要因であると考えられる。しかし「神」の結果の考察で述べたように「ヤマツミ」とは山の神という意味であるが、さらに細分化してみると、「ツ」が「の」、「ミ」が「神」の意味を示すこと<sup>20)</sup>から何かを司る神としての単語としての特徴を適切に抽出することができたといえる。

「タカミ」、「ムスビ」の単語の詳細については「神」の結果の考察で述べたためここでは省略する。ここでは神としての特定の権能を意味する単語を類似単語として抽出している。そのため神の中でもそれぞれの役目や権能などについてクラスタ分析等を行うことで更なる研究が見出せる可能性がある。

「ツチ」という語については「タケミカヅチ」や「ノヅチの神」といった神から出力されたと考えられる。「タケミカヅチ」の出自は「イザナギ」と「イザナミ」が関連している。「イザナミ」が「カグツチ」という神を生んだ際に亡くなってしまう。そして「イザナギ」が「カグツチ」の首を切ったとき、刀についた血が湯津石村に走りついて成った三柱のうちの一柱である。「タケミカヅチ」は葦原中国の評定の際には使者としての役割を成功させた神である。刀剣の神として描写が色濃いと指摘されているが、そのことと同時に漢字では「雷」が使われることから、

雷神であるともされている。さらにこの神は外国においても同等のモチーフの類似点があるため国際的な比較によっても有意な結果が得られると思われる<sup>31)</sup>。

また、「ノヅチの神」とは野の霊(チ)、野の精霊と解される<sup>32)</sup>。つまり「ヤマツミ」に通じる自然を司る神に関わる意味のある単語の一部を適切に抽出することができた。

「オカミ」という語とは「オカミの神」から出力されたと考えられる。「オカミの神」とは水を司る龍神と考えられている<sup>33)</sup>。そのためこの単語においても自然に関連するという「ヤマツミ」の類似単語として適切な特徴を抽出することができた。

「フツ」という語は「タケフツの神」や「トヨフツの神」などの語から出力されたと考えられる。この「タケフツの神」や「トヨフツの神」とは「タケミカヅチ」の別名である。また「フツ」という言葉には「すべて」といったような意味を持つこともある。一般的には刀剣の神の呼称として理解されていること<sup>34)</sup>から司る神という意味での単語の特徴を抽出することができたと考えられる。

「ワタツミ」という語に関しては「オホワタツミの神」や「ナカツワタツミの神」といった神から出力されたと思われる。「ワタツミ」における「ワタ」とは海のことを指す。しかし「ワタ」を海洋として捉えて、ウミを湖や池などを含めた

大いなる水の意味の語と捉える見解もある。「ワタツミの神」は自然における海そのものの神格ではなく、海を司る支配者としての存在として考えられている<sup>35)</sup>。よって「ヤマツミ」における類似語として「ワタツミ」は自然を司る支配としての神格の特徴を適切に抽出することができたと考えられる。

## 6. まとめと今後の課題

本論文では古事記を対象として、自然言語処理である単語分散表現を用いた単語間の類似度の分析を行った。その結果として「神」、「天皇」や「命」に関してはそれぞれの意味の特徴を抽出し、類似する語句を検出することができた。特に「神」という単語の分析からは「出現する」といったような自然発生的なニュアンスを含む神の生まれ方に関する特徴を抽出することができた。また、「天皇」という語に関しては、歴代天皇に含まれる単語だけでなく、天皇の即位する前の名前などの特徴を抽出することができた。「命」においても現代語と比較して、古事記内では尊称としての意味が強いことを示した。

しかし、形態素解析を利用し単純な頻度などから単語の分析を行っても有意な結果を得ることが難しいことも判明した。「つて」などの語においては適切な形態素解析が行われず、類似単語の分

析においても評価を下すことは不可能であった。

先行研究を考慮した分析結果では、「イザナギ」や「イザナミ」では兄弟としての特徴や親子関係としての類似単語の抽出することもできた。「イザナミ」においては「イザナギ」には見られなかった女性神格としての特徴も抽出することができた。さらには、「天照らす」、「大神」においては司令神としての特徴として夢告に関する「夢」という単語の出力や兄弟関係にある「スサノヲ」の抽出や、信仰の主神としての特徴や高天原を統治する神、皇祖神としての特徴を抽出することができた。また、今後の研究につながるような相反する単語が類似単語として出力される結果となった。「ヤマツミ」においては自然に関連する神の特徴などを適切に抽出することができた。

今後の課題としては古典表現に対する形態素解析の是正がある。5.2節で述べたように適切に形態素解析がなされていないようなときでは類似する単語を出力しても適切な評価をすることができないという問題がある。この問題を解決するには適切な形態素解析を実施するためのコーパスを作成することが挙げられる。さらにテキストにより表記が定まらないような場合（例：5.2.2項での「カミムスビの神」と「カムムスビの神」など）、その一つ一つを適切に登録する方法についても考える必要がある。

## 謝辞

研究を遂行するにあたり、本学経営学部准教授 横井桃子先生には研究の初期に相談に乗って戴いた。また、本学大学院経営学研究科 修士課程の草開翔太氏には丁寧な添削をして戴き、同大学院 博士後期課程の新村裕太氏には有意なコメントを戴いた。

本稿を執筆するにあたり本学国際コミュニケーション学部教授 梅垣敦紀先生に作成して戴いた p<sub>L</sub>T<sub>E</sub>X のクラスファイル<sup>36)</sup>を利用した。

上記の方々の御助力に厚く感謝申し上げます。

## 注

注 1) Word2Vec の Tomas Mikolov も共同研究者である。

注 2) 通時コーパスとは奈良時代から明治・大正時代までの主要な日本語史資料をコーパス化したものである<sup>37)</sup>。

## 参考文献

- 1) OpenAI. Chat Generative Pre-trained Transformer. <https://chat.openai.com>. (最終閲覧日 2023 年 7 月 9 日) .
- 2) 村上征勝. 真贋の科学: 計量文献学入門. 行動計量学シリーズ 6. 朝倉書店, 1994.
- 3) 村井源. コンピュータを用いた文学研究の近年の動向とその応用に向けて. 情報処理学会研究報告 研究報告人文科学とコンピュータ, Vol. 2016-CH-112, No. 5, pp. 1-5, 2016.
- 4) Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. <https://arxiv.org/abs/1301.3781>, 2013. (最終閲覧日 2023 年 11 月 22 日) .
- 5) Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching Word Vectors with Subword Information. <https://arxiv.org/abs/1607.04606>, 2017. (最終閲覧日 2023 年 11 月 22 日) .
- 6) 京都大学情報学研究科. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. <https://taku910.github.io/mecab/>. (最終閲覧日 2023 年 7 月 10 日) .
- 7) 村上征勝. 計量文献学の歴史と課題. 計算機統計学, Vol. 9, No. 1, pp. 65-74, 1997.
- 8) 佐藤浩輔, 中分遥. 民間伝承の計量分

- 析：「怪異・妖怪伝承データベース」の俯瞰的分析. じんもんこん 2021 論文集, Vol. 2021, pp. 30–37, 2021.
- 9) 吉田拓海, 村井源. 英雄神話物語の構造化に向けての物語特徴の抽出. 情報知識学会誌, Vol. 30, No. 2, pp. 270–275, 2020.
- 10) 岩波データサイエンス刊行委員会. 岩波データサイエンス Vol.2. 岩波書店, 2016.
- 11) 岩波データサイエンス刊行委員会. 分散表現 (単語埋め込み). <https://sites.google.com/site/iwanamidata-science/vol12/%E5%88%86%E6%95%A3%E8%A1%A8%E7%8F%BE%E5%8D%98%E8%AA%9E%E5%9F%8B%E3%82%81%E8%BE%BC%E3%81%BF>. (最終閲覧日 2023 年 11 月 23 日) .
- 12) 金子淳, 大槻恭士, 坂口隆之. 単語ベクトルの類似度を用いた英米文学の通時的含意分析. 人工知能学会全国大会論文集, Vol. JSAI2022, pp. 1K1GS605–1K1GS605, 2022.
- 13) 堺雄之介, 伊東栄典. 単語の分散表現に基づく流行語分析. 情報処理学会研究報告, Vol. 2020-MPS-130, No. 6, pp. 1–8, 2020.
- 14) 京都大学 言語メディア研究室. 日本語形態素解析システム JUMAN. <https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>. (最終閲覧日 2023 年 11 月 23 日) .
- 15) 小木曾智信, 小町守, 松本裕治. 歴史的日本語資料を対象とした形態素解析. 自然言語処理, Vol. 20, No. 5, pp. 727–748, 2013.
- 16) Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhersch, and Armand Joulin. Advances in Pre-Training Distributed Word Representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), 2018.
- 17) 青空文庫. 現代語訳 古事記. <https://www.aozora.gr.jp/cards/001518/card51732.html>. (最終閲覧日 2023 年 6 月 20 日) .
- 18) 京都大学情報学研究科. IPA 辞書. <https://drive.google.com/uc?export=download&id=0B4y35FiV1wh7MWV1SDBCSXZMTXM>. (最終閲覧日 2023 年 11 月 23 日) .

- 19) 尾畑喜一郎. 比較研究における本邦国  
生み神話. 國學院雑誌, Vol. 83,  
No. 11, pp. 26-49, 1982.
- 20) 國學院大學 神名データベース. 大山津  
見神. [https://kojiki.kokugakuin  
.ac.jp/shinmei/oyamatsuminokam  
i/](https://kojiki.kokugakuin.ac.jp/shinmei/oyamatsuminokami/). (最終閲覧日 2023 年 11 月  
25 日) .
- 21) 國學院大學 神名データベース. 高御産  
巢日神. [https://kojiki.kokugaku  
in.ac.jp/shinmei/takamimusuhin  
okami/](https://kojiki.kokugakuin.ac.jp/shinmei/takamimusuhinokami/). (最終閲覧日 2023 年 11 月  
26 日) .
- 22) 國學院大學 神名データベース. 伊耶那  
岐神. [https://kojiki.kokugakuin  
.ac.jp/shinmei/izanakinokami/](https://kojiki.kokugakuin.ac.jp/shinmei/izanakinokami/).  
(最終閲覧日 2023 年 11 月 26 日) .
- 23) 國學院大學 神名データベース. 建速須  
佐之男神. [https://kojiki.kokugak  
uin.ac.jp/shinmei/takehayasusa  
noonomikoto/](https://kojiki.kokugakuin.ac.jp/shinmei/takehayasusa noonomikoto/). (最終閲覧日 2023 年  
11 月 26 日) .
- 24) 出雲大社. 出雲大社と大国主大神. [htt  
ps://izumooyashiro.or.jp/about  
/ookami](https://izumooyashiro.or.jp/about/ookami). (最終閲覧日 2023 年 11 月  
26 日) .
- 25) 坂本勝. 都市の大物主：崇神朝の崇り  
神伝承をめぐって. 日本文学誌要,  
Vol. 83, pp. 35-44, 2011.
- 26) 國學院大學 神名データベース. 天照大  
御神. [https://kojiki.kokugakuin  
.ac.jp/shinmei/amaterasuomikam  
i/](https://kojiki.kokugakuin.ac.jp/shinmei/amaterasuomikami/). (最終閲覧日 2023 年 11 月  
26 日) .
- 27) 國學院大學 神名データベース. 思金神.  
[https://kojiki.kokugakuin.ac.jp  
p/shinmei/omoikanenokami/](https://kojiki.kokugakuin.ac.jp/shinmei/omoikanenokami/). (最  
終閲覧日 2023 年 11 月 26 日) .
- 28) 國學院大學 神名データベース. 天之菩  
卑能命. [https://kojiki.kokugaku  
in.ac.jp/shinmei/amenohohinomi  
koto/](https://kojiki.kokugakuin.ac.jp/shinmei/amenohohinomikoto/). (最終閲覧日 2023 年 11 月  
26 日) .
- 29) 佐原作美. 日本靈異記における夢告譚  
の構造. 駒沢短大國文, Vol. 30, pp.  
1-13, 2000.
- 30) 國學院大學 神名データベース. 菟神. [h  
ttps://kojiki.kokugakuin.ac.jp  
/shinmei/usagigami/](https://kojiki.kokugakuin.ac.jp/shinmei/usagigami/). (最終閲覧  
日 2023 年 11 月 26 日) .
- 31) 國學院大學 神名データベース. 建御雷  
之男神. <https://kojiki.kokugaku>

in.ac.jp/shinmei/takemikazuchi  
noonokami/. (最終閲覧日 2023 年  
11 月 26 日) .

injal.ac.jp/research/cr-projec  
t/project-3/institute/diachron  
ic-corpora/. (最終閲覧日 2023 年  
11 月 23 日) .

32) 國學院大學 神名データベース. 野椎神.  
<https://kojiki.kokugakuin.ac.jp/shinmei/nozuchinokami/>. (最終  
閲覧日 2023 年 11 月 26 日) .

33) 國學院大學 神名データベース. 淤迦美  
神. [https://kojiki.kokugakuin.a  
c.jp/shinmei/okaminokami/](https://kojiki.kokugakuin.ac.jp/shinmei/okaminokami/). (最  
終閲覧日 2023 年 11 月 26 日) .

34) 國學院大學 神名データベース. 佐士布  
都神. [https://kojiki.kokugakuin  
.ac.jp/shinmei/sajifutsunokami  
/](https://kojiki.kokugakuin.ac.jp/shinmei/sajifutsunokami/). (最終閲覧日 2023 年 11 月 26 日) .

35) 國學院大學 神名データベース. 綿津見  
神. [https://kojiki.kokugakuin.a  
c.jp/shinmei/watatsuminokami/](https://kojiki.kokugakuin.ac.jp/shinmei/watatsuminokami/).  
(最終閲覧日 2023 年 11 月 26 日) .

36) 梅垣敦紀. 愛知大学情報メディアセン  
ター紀要「COM」への  $\text{T}_{\text{E}}\text{X}$  投稿の試  
み. 愛知大学情報メディアセンター紀  
要「COM」, Vol. 33, No. 1, pp.  
31–36, 2024.

37) 小木曾智信. 通時コーパスの構築と日  
本語史研究の新展開. <https://www.n>