

コンピュータ科学科
知能システム系
卒業論文

大規模コーパスに基づく
各動詞のメタファー用例率の推定と
言語学的特徴との関係分析

2023 年 2 月

101930007 青野 広太郎

目次

第 1 章	はじめに	5
1.1	本研究の目的	6
1.2	本研究の構成	6
第 2 章	関連研究	7
2.1	静的な単語埋め込み	7
2.2	文脈を考慮した単語埋め込み	8
2.2.1	ELMo	8
2.2.2	BERT	9
2.3	メタファーについての研究	9
2.3.1	認知言語学とメタファー	9
2.3.2	自然言語処理におけるメタファー	11
2.3.3	メタファーと言語学的特徴の関係	12
第 3 章	メタファーの自動判別	13
3.1	自然言語処理におけるメタファーの自動判別	13
3.2	MelBERT	14
第 4 章	単語の特徴量とメタファー用例率との関係性	16
4.1	関係性の分析手法	16
4.2	分析対象とする特徴量	16
4.2.1	認知的な特徴量	17
4.2.2	目的語の多様性に関する特徴量	18
第 5 章	実験	20
5.1	使用するデータセット	20

5.2	実験設定	20
5.3	実験結果	22
	5.3.1 目的語の分布	22
	5.3.2 各特徴との相関	25
5.4	考察	30
第 6 章	おわりに	31
6.1	本研究のまとめ	31
6.2	課題と今後の展望	31

図目次

2.1	BERT の事前学習の構造。Devlin ら [1] より引用	10
3.1	MelBERT のモデル (Choi ら [2] より引用)。対象単語がサブワードで分割されていた場合その二つの埋め込みの平均を利用する	15
4.1	関係分析の全体像	17
4.2	目的語の多様性を数値化する手法	19
5.1	目的語の散布図 左が kill、右が see	23
5.2	目的語の散布図 左が earn、右が look	23
5.3	目的語の散布図 左が attack、右が cause	24
5.4	密集度比のヒストグラム	25

表目次

5.1	恣意的に選んだ 10 動詞とメタファー用例	21
5.2	ランダムに選んだ 50 動詞の一覧	21
5.3	認知的な特徴量との相関係数	26
5.4	目的語の多様性に関する特徴量とメタファー用例率との相関	26
5.5	20 動詞の認知的な特徴量とメタファー用例率	27
5.6	分散、コサイン類似度の平均とメタファー用例率	28
5.7	異なり数、エントロピーとメタファー用例率	29

第 1 章

はじめに

近年、自然言語処理においてメタファー研究は多く行われている。メタファーは隠喩とも言われ、「のような」といった言葉を使わずに行われる比喻表現である。例えば、“He attacked every weakpoint in my argument”といった文が挙げられるが、この文では動詞 attack が文字通りの「攻撃する」という意味ではなく「指摘する」といった意味で使われているのがわかる。このような表現は日常的に頻繁に用いられており、メタファー表現の性質を見極めることはその表現の意味を捉えるうえで言語理解において非常に重要である。また、このような文脈に依存し、文字通りでない意味を捉えることは、機械翻訳や感情分析など、様々な自然言語処理に応用可能である。

自然言語処理においてメタファー研究はメタファー検出、つまり文中のある単語がメタファーかどうか判別するといったタスクを中心に取り組まれてきた。Devlin ら [1] による自然言語処理モデルである BERT による埋め込みを利用したものや、人手で作成された単語の言語学的な特徴量などを用いたものなど、メタファー検出タスクにおける既存研究は数多い。しかし、自然言語処理において大規模なコーパスを用いて収集した大量の文を用いてメタファーの性質を分析するような研究は少ない。

そこで本研究では Web データを収集した大規模なコーパスである Common Crawl と、先ほど述べたメタファー検出の技術を応用し、単語ごとにそれがメタファーか判別するメタファー判別器を使用して、動詞ごとにそれがメタファーとして使用される割合であるメタファー用例率を推定した。そしてそれが各動詞の言語学的特徴とどのような関係にあるのか調査した。また、動詞の目的語の分布がメタファー用例と非メタファー用例でどのように分布するか調べた。その結果、動詞の親密度や目的語の多様性において小さいながらも動詞のメタファー用例率との相関があることが分かった。メタファー用例率に着目することにより、この研究はメタファーの性質の分析やメタファーが用いられる文脈の分析な

どの研究に応用可能であると考ええる。

1.1 本研究の目的

先述の通り、自然言語処理においてメタファー用例率に着目した研究は少ない。また、大規模コーパスと自動メタファー判別器により、人手では分析困難な量のデータを分析し、メタファーの性質や、メタファーが使われる条件を調査するために、本研究では、60種類の動詞を対象に、Common Crawl から各動詞とその目的語を含む文を集め、自動メタファー判別器 (MeLBERT[2]) にかけて各動詞のメタファー用例率を算出する。その後各動詞のメタファー用例率と後述する動詞の親密度や具体性などの言語学的特徴との相関を調べ、動詞メタファーの性質を分析する。

1.2 本研究の構成

ここでは本論文における本章以降の構成について述べる。第2章ではメタファー判別などに利用されている技術である単語埋め込みやメタファーに関する関連研究を紹介し、第3章ではメタファー判別の技術と使用する判別器の概要を説明する。第4章では単語の言語学的特徴を取得するデータセットやそのメタファー研究への応用例を述べ、第5章では実験の具体的な手法とその結果を紹介する。最後に第6章では、本研究のまとめや課題、今後の展望について述べる。

第 2 章

関連研究

本章では本研究に関連する研究を紹介する。2.1 節では静的な単語埋め込み、2.2 節では BERT を用いた文脈化単語埋め込み、2.3 節ではメタファーについての関連研究について説明する。

2.1 静的な単語埋め込み

自然言語処理において埋め込みとは、文や単語など、言語表現の構成要素にある空間におけるベクトルを対応付けることである。多くの単語埋め込みの手法は、「単語の意味は、周囲の単語によって形成される」という分布仮説 [3] と呼ばれる仮説に基づいている。近年ではニューラルネットワークによって埋め込みを作成する手法が一般的であり、同じ単語でも文脈によって埋め込みが変わる文脈化単語埋め込みとある単語に対して一つの対応する単語ベクトルを生成する静的な単語埋め込みが存在する。ここでは静的な埋め込みを紹介し、文脈化埋め込みは次節で紹介する。

静的な単語埋め込みを生成する手法はいくつか存在し、代表的なものとしては word2vec による skip-gram[4]、CBOW[5] や、それ以外にも GloVe[6]、fastText[7] が存在する。word2vec において、skip-gram も CBOW も単語予測タスクをニューラルネットを使って解くことによって単語埋め込みを得るという点では共通しているが、skip-gram では文中の一つの中心語から周辺の単語を予測するタスクを解くのに対して、CBOW は周辺単語から中心語を予測するタスクを解く過程で単語埋め込みを生成する変換行列の重みを得るといった点が異なる。また、skip-gram や CBOW が周辺単語の窓内だけのローカルな情報を使うのに対し、GloVe は語彙全体の大域的な単語間共起行列を考慮することによって、学習効率などの word2vec の弱点を克服した。さらに、2016 年に Facebook が発表し

た fastText では計算が大幅に効率化され、単語をサブワードに分けることによって、単語の活用形や未知語などにも柔軟に対応できるようになった。

静的な埋め込みの問題点としては多義語に対応できない点が挙げられる。例えば、bank という単語は「土手」と「銀行」の二つの意味を持つが、上の手法だとこれらの意味の違いをベクトルに表すことができない。これら文脈によって意味が変わる単語を意味ごとに異なるベクトルで表すことができるのが文脈化単語埋め込みであり、次節で紹介する BERT[1] などの手法が挙げられる。

2.2 文脈を考慮した単語埋め込み

本研究で使ったメタファー判別器では文脈によって変化する埋め込みを特徴量として使用し、判別を行なっている。よってここではその埋め込みを取得する手法として ELMo[8]、BERT[1] について説明する。

2.2.1 ELMo

ELMo とは事前学習により文脈を考慮した単語埋め込みを獲得する手法である。ELMo は二層の LSTM をベースとしたモデルであり、その LSTM モデルを大規模コーパスによって事前学習させる。また、それによって学習された各単語に対応する LSTM の隠れ層の値や Embedding 層の値の平均を取る、もしくは重み付けをして得られるベクトルをその単語の埋め込み表現とする。それを表した式が以下である。

$$R_k = \{x_k^{LM}, h_{k,j}^{\rightarrow LM}, h_{k,j}^{\leftarrow LM} | j = 1, \dots, L\} \quad (2.1)$$

$$= \{h_{k,j}^{LM} | j = 0, \dots, L\} \quad (2.2)$$

式 2.1 において x_k^{LM} は k 番目の単語の Embedding 層の表現であり、 $h_{k,j}^{\rightarrow LM}$ 、 $h_{k,j}^{\leftarrow LM}$ はそれぞれ順方向、逆方向の LSTM による j 層目の隠れ層の表現である。式 2.2 はそれらを連結したものであり、 $h_{k,0}^{LM}$ は x_k^{LM} を表し、 $h_{k,j}^{LM}$ は $[h_{k,j}^{\rightarrow LM}; h_{k,j}^{\leftarrow LM}]$ である。ELMo ではそうして得たベクトルを関数 E に適用することによって一つのベクトルにする。その E はタスクによって以下のようにレイヤーの重みを変えてやることで定義できる。式 2.3 において s_j^{task} は重み付けを変えるパラメータ、 γ^{task} はスケールパラメータである。

$$\mathbf{ELMo}_k^{task} = E(R_k; \Theta^{task}) = \gamma^{task} \sum_{j=0}^L s_j^{task} h_{k,j}^{LM} \quad (2.3)$$

最終的にこうして得た **ELMo** を embedding レイヤーに結合、 $[x_k; \mathbf{ELMo}_k^{task}]$ として学習をすることによって対象となる文章の埋め込み表現の生成モデルを生成する。

2.2.2 BERT

BERT とは、Bidirectional Encoder Representations from Transformers の略である。BERT は複数の Transformers[9] の Encoder で構成されており、MLM(Masked Language Modeling) と NSP(Next Sentence Prediction) の二つの事前学習により、Attention と呼ばれる計算機構を用いて方向性に依存せずに文脈からの単語埋め込みを構成可能である。BERT の学習は事前学習とファインチューニングの2段階に分けられ、事前学習では大量のラベルなしデータから、先述の二つのタスクにおいて学習を行い、ファインチューニングではタスクごとにラベルありデータを用いて学習を行う。

事前学習のうち、MLM は全トークンのうち 15% をマスクし、元のトークンを予測するタスクである。また、NSP は二文を選んでそれが連続する文か当てるタスクである。図 2.1 で示すように元々の文の先頭に [CLS] トークン、末尾に [SEP] トークンを挿入した文から複数層の Transformers を通し、各単語の隠れベクトルを MLM に、[CLS] トークンの隠れベクトルを NSP タスクに使用する。

ファインチューニングではタスクに合わせて比較的少量のラベル付きデータを使い、事前学習で得た重みを初期値として学習を行う。これによって少量のデータからでも高い精度でタスクをこなすことができる。また、BERT 内の Transformers で構成されている各層の出力はモデル内の重みを反映した文脈を考慮した埋め込みと考えることができ、通常出力層に最も近い層の出力を単語埋め込みとして使用する。

これらの学習によって BERT は様々な NLP タスクにおいて SoTA を達成し、現在でも様々なタスクにおいて応用されている。

2.3 メタファーについての研究

本節ではメタファーが自然言語処理、および言語学でどのように扱われてきたかについて述べる。

2.3.1 認知言語学とメタファー

メタファーが最初に議論の場に挙げられたのはもちろん言語学の分野である。しかし、チョムスキーの生成文法が主流だった時代、メタファーはせいぜい言葉の綾であり、文法の中でも周辺的な事象であるとしてあまり重要視されていなかった。しかし、生成文法では扱いきれない、人間の認知構造と言葉の使用を結び付ける認知言語学が台頭してくる

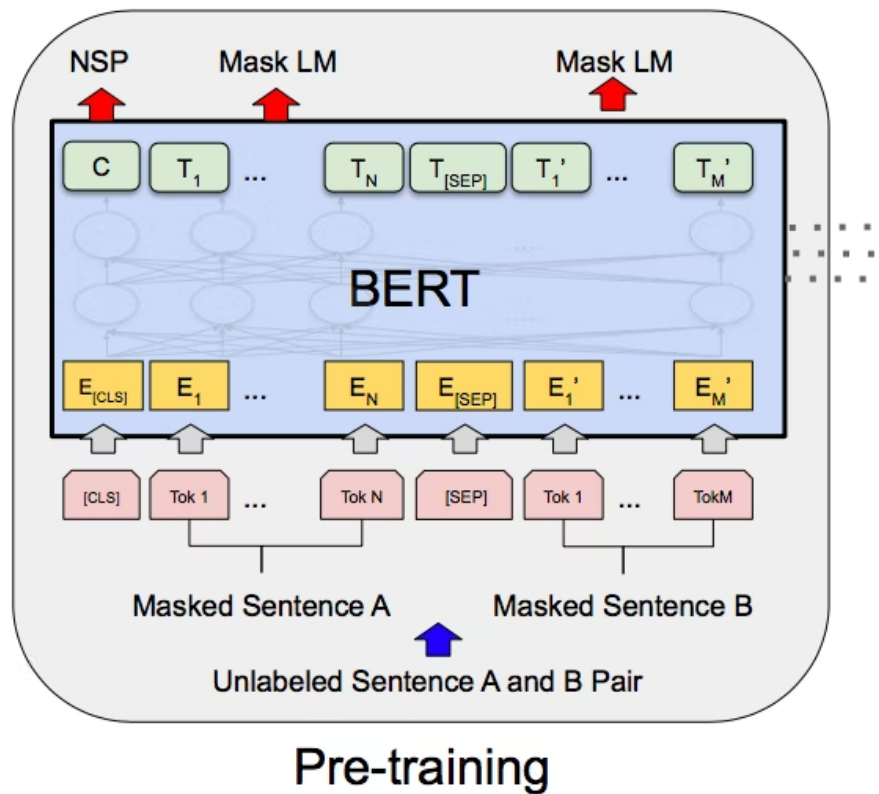


図 2.1 BERT の事前学習の構造。Devlin ら [1] より引用

と、メタファーこそが人間の認知を大きく反映する重要な機能であるとして注目を集め始めた。そのきっかけとして Lakoff らの *Metaphors we live by*[10] が挙げられる。彼らはメタファーを人間の認知の中心的事象と捉え、メタファーを単なるレトリックではなく、対象を概念化するものととらえた。以下は議論にまつわるメタファーの例である。

- (1) He attacked every weakpoint in my argument.
- (2) If you use such a tactics, you can't win this argument.

上の二つの例では、議論に関して attack や tactics、win など戦争に関する用語が使われている。このようなメタファーが使われる地域では、議論を勝ち負けがあり、何らかの戦術を用いて攻めたり守ったりするなどの側面に注目して理解していると考えられる。次に以下のような例も考えられる。

- (3) There is no basis for the argument.
- (4) The debate has already collapsed.

これらの例では、議論を土台や柱があり、組み上げたり崩したりする建物のような側面に焦点を当てて捉えている。これらのような例によって、議論という概念を、戦争や建物などのより理解しやすい概念に写して理解することができる。lakoff らは、このようにメタファーの本質とはある概念を別の概念を通して理解することである、つまりメタファーとは人間の認知の本質であると考えたのである。また、このようなはたらきを概念メタファー ARGUMENT IS WAR(1)(2) や ARGUMENT IS BUILDINGS(3)(4) でとらえた。

この概念メタファー論は言語学に大きな影響を及ぼし、メタファーを研究することで人間の認知の仕方を分析したり、文の意味をよりよく理解できるとして、現在でも自然言語処理などの分野の多くの論文で引用されている。

2.3.2 自然言語処理におけるメタファー

コーパス言語学によると、三文に一回はメタファーが存在しているとされており、隠喩的な表現の正確な理解が自然言語処理に求められているのが現状である。また、現在 NLP コミュニティではメタファー分析をメタファー識別とメタファーの意味理解に分けるのが一般的となっており、そのために多くのデータセットやそれを使った手法などが研究されている。データセットではまず Tsvetkov ら [11] が公開した隠喩的な形容詞-名詞ペアを含むデータセットを発表し、Mohammad ら [12] は WordNet[13] から抽出した文をクラウドソーシングでラベル付けしたデータセットである MOH を公開した。さらに認知言語学者が英国国立コーパスから作成した VU Amsterdam Metaphor Corpus [14] などが存在する。

メタファー識別では最初に登場したニューラルモデルは BiLSTM を使ったものであった。Mykowiecka ら [15] は入力として単語の埋め込みだけでなく後述する MRC データセット [16] による心理言語学的な特徴を使用したモデルを公表し、Gao ら [17] は文脈化単語埋め込みである ELMo[8] を Bi-LSTM エンコーダへの入力として使用し、VUAMC データセットにおいて高い識別精度を達成した。

BERT[1] が登場し始めるとモデルとして BERT を使用したものが多くなり、単語埋め込みに加えて感情の情報も入力として与える手法 [18] や、最近だと Choi ら [2] により、BERT による単語ベクトルのうち、文脈を考慮したものと文字通りのものを比較することによりメタファー検出を行う手法などが提案された。この中でも重視すべきものは感情などの人間の認知などの情報を入力として用いる手法である。比喩処理は、先に述べた概念メタファー論によると、概念が脳／心でどのように組織化されるかに関係し、分類と密接に関係する。したがって、メタファー識別のために範疇の特徴を用いることは理にかなっ

ている。[19] そこで、人間の認知に基づいた特徴量である語の具体性や親密度を使用したアプローチが数多く生み出された。以下ではそれらの例をいくつか紹介する。

2.3.3 メタファーと言語学的特徴の関係

本節では自然言語処理におけるメタファー研究といろいろな人間の認知に基づいたデータセットなどから得られる言語学的特徴がどのような関係にあるか述べる。単語の言語学的な特徴、例えばその単語の品詞情報、具体性やイメージ可能性などは多くのメタファー研究に応用されてきた。Tekiroglu ら [20] は具体性とイメージ可能性などの特徴量を形容詞-名詞ペアのメタファー (例: 甘いメロディー) の判別に利用し、Bulat ら [21] は語の認知科学的な特性をまとめたデータセットを用いてメタファー判別を行なった。また、Shutova ら [22] は視覚埋め込みと単語埋め込みの両方を利用したマルチモーダルモデルを考案し、単一モーダル情報のみを用いた手法に比べ高いスコアを達成した。そして、最近だと Gong ら [23] が語の具体性や WordNet で得た単語の分類などの特徴量を一般的なメタファー判別タスクに利用した。以上の関連研究からは、認知的な特徴量はよくメタファー判別タスクに利用されており、本研究で算出するメタファー用例率にも相関があるという仮説を得た。

また、メタファー判別以外にも、Skalicky ら [24] がメタファーや皮肉の面白さにおける、人間による評価と具体性などの言語学的特徴の関係性について議論しているなど、メタファー研究における言語学的特徴の利用は幅広い。

第 3 章

メタファーの自動判別

本章では、本実験で使用したメタファー判別器に関する技術を説明する。3.1 節で自然言語処理におけるメタファー判別技術の概要を述べ、3.2 節で本実験で使用したメタファー判別器である MelBERT[2] について述べる。

3.1 自然言語処理におけるメタファーの自動判別

まず、自然言語処理においてメタファーの自動判別とは、通常は入力として文を受け取り、それぞれの単語ごとにそれがメタファーかどうかのラベルをつける作業である。ニューラルモデルによるメタファー判別が登場し始めた時はアーキテクチャとして BiLSTM、データセットとして Tsvetkov ら [11] による形容詞-名詞ペアを含むデータセットを用いた判別器が主であったが、近年では BERT と、19 万の字句単位において各単語にメタファー、非メタファーのラベル付けがされた VU Amsterdam Metaphor Corpus (VUAMC) [14] を用いた判別器が多く研究されている。VUAMC は学術テキストや会話、フィクション、ニュースに文がカテゴリ化されており、メタファーの種類も限定されていないため、現在多くのメタファー判別器のデータセットとして使用されている。入力される特徴量としては、トークン (単語) や埋め込み、品詞などが主に見られるが、WordNet や後述する MRC データセット [16] などから取得した単語の言語学的特徴、例えば単語の具体性やイメージ可能性などを入力として使っているものも多く見られる。さらに Shutova ら [22] は各単語について Google Images の画像検索で得た画像を視覚埋め込みに変換し判別に用いたなど、判別の方法は多岐に渡る。

3.2 MeIBERT

ここでは本実験で使用したメタファー判別器である MeIBERT[2] について述べる。MeIBERT はメタファー判別タスクにおける手法のうち、Metaphor Identification Procedure (MIP)[14] と Selectional Preference Violation (SPV)[25] の二つの考えを用いている。前者は一つの単語において、メタファー文脈中での使用例と文字通りの使用例でその単語の意味が異なるという考えを基にしたものであり、文字通りの単語埋め込みと文脈化単語埋め込みの間で生ずる差を用いてメタファー判別を行う。後者は、その単語がメタファーかどうかは、その周囲の単語との意味的な違いによって識別されるという考えに基づいており、文ベクトルと文脈の中での対象単語の埋め込みにおける差を利用してメタファー判別を行う。

ここからはその具体的な手法を述べる。まず、このモデルの構造を図 3.1 で示す。このモデルは文エンコーダ $Enc(S)$ と対象単語エンコーダ $Enc(w_t)$ 、スコアを計算する late interaction mechanism 部の三つで構成されている。まず入力文をトークンに分解し、BERT に見られるようなトークン列の順番を表すための position 埋め込みや、対象単語と局所的な文脈を指定するための target 埋め込み (図では [TAR]、[LOC]) を作成する。また、品詞情報を利用するため最後のトークン [SEP] の後に対象単語の品詞タグを追加する。

これらをもとに、transformer のエンコーダである $Enc(S)$ を用いて文 $S = \{w_1, \dots, w_n\}$ を文脈化埋め込み $\{v_S, v_{S,1}, \dots, v_{S,n}\}$ に変換する。ここで v_S は [CLS] トークンをベクトル化したものであり、 $v_{S,i}$ は文 S の i 番目の単語の埋め込みである。同じように、 $Enc(w_t)$ は対象単語 w_t をその単語のみ、つまり文脈情報なしでベクトル化する (ここでは v_t と表す)。このようにすれば、 $v_{S,t}$ と v_t はそれぞれ対象単語 w_t について文脈化された意味とその単語単体で孤立した意味、すなわち文字通りの意味として表されることがわかる。

このモデルにおいて、MIP、つまり暗喩的な使い方をされる単語を、文脈中の意味と文字通りの意味の間で生じる差において判別する考え方は二つのエンコーダによって得られる $v_{S,t}$ と v_t の二つの埋め込みを比較することによって表現される。この二つは、前者は文脈化埋め込みであり、後者はその単語のみを使った文脈情報のない埋め込みであった。つまり、この二つを比較することによって文脈中の意味と文字通りの意味の差を取得することができる。また、SPV、つまり隠喩的な使用をされる単語は周囲の単語との意味的な差によって判別されるという考え方は、MIP と異なり、一つのエンコーダによるベクトルのみを使って表す。ここで使うベクトルは文ベクトル v_S と文脈化埋め込み $v_{S,t}$ であ

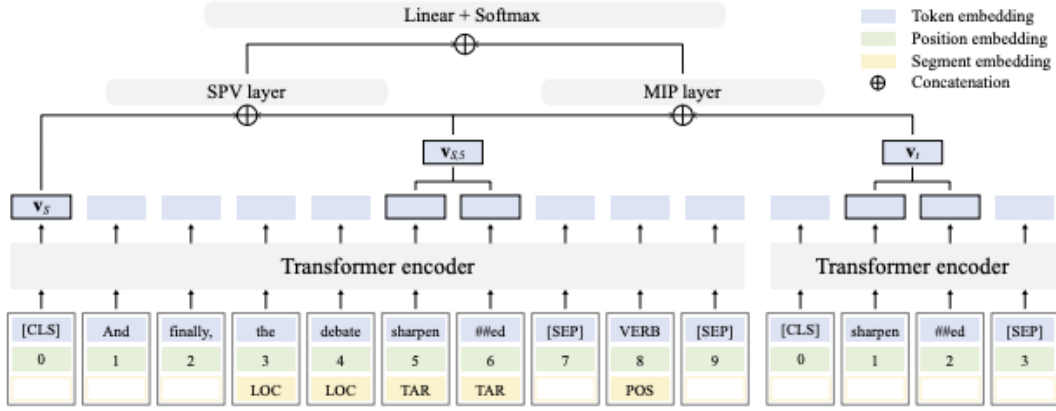


図 3.1 MelBERT のモデル (Choi ら [2] より引用)。対象単語がサブワードで分割されていた場合その二つの埋め込みの平均を利用する

る。ここで、文ベクトル v_S が文 S 中の全ての単語の相互作用を表すのに対して、単語埋め込み $v_{S,t}$ は対象単語 w_t と文中にあるその他の単語の間での相互作用を表す。よって隠喩的な文では $v_{S,t}$ は文ベクトル v_S とより差が生まれる。

これら二つの材料を使って MelBERT はメタファー判別を行う。MIP を表す隠れベクトル h_{MIP} は $v_{S,t}$ と v_t を連結したベクトルから以下の式で求められる。

$$h_{MIP} = f([v_{S,t}; v_t]) \quad (3.1)$$

$f()$ は多層パーセプトロン層において二つのベクトルの差を学習するための関数である。また、SPV も文ベクトル v_S と $v_{S,t}$ を連結したベクトルから求められる。

$$h_{SPV} = g([v_S; v_{S,t}]) \quad (3.2)$$

$g()$ も二つのベクトルの差を学習するための関数である。これら二つの隠れベクトルを連結し、シグモイド関数 σ とパラメータ W 、バイアス b を加えて予測値 \hat{y} を計算する。

$$\hat{y} = \sigma(W^T[h_{MIP}; h_{SPV}] + b) \quad (3.3)$$

最終的に以下の交差エントロピー損失関数によって学習を行う。

$$\mathcal{L} = \sum_{i=1}^N y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (3.4)$$

ここで N はサンプル数、 y_i は正解ラベル、 \hat{y}_i は i 番目のサンプルの予測ラベルである。

第 4 章

単語の特徴量とメタファー用例率との関係性

本章では本研究で動詞のメタファー用例率との相関を取得する手法について述べる。4.1 節で本研究で行う関係分析の手法、4.2 節では本研究で取り扱う特徴量を紹介する。

4.1 関係性の分析手法

本節では本研究で実施する単語の特徴量とメタファー用例率の調査手順を説明する。本研究では動詞のメタファーになりやすさを調査するため、図 4.1 に示すように動詞-目的語のペアを大量に集め、それぞれにメタファー判別器 [2] を使ってメタファー、非メタファーのラベル付けを行った。その後、動詞それぞれについてその動詞の全使用例におけるメタファー用例の割合を算出し、その割合がどのような要因によって決まるのか調査するために、その割合と動詞の各特徴量との間での相関係数を調べた。その特徴量の選び方は以下で述べる。また、機械学習によって動詞ごとにどのような目的語がメタファーと判別されるのか調べるため、目的語の埋め込みを高次元データを 2 次元や 3 次元に落とし込むための次元削減アルゴリズムである t-SNE[26] を使って二次元に圧縮し、散布図を描画した。

4.2 分析対象とする特徴量

本節ではメタファー用例率と相関係数を調べる特徴量と、それを採用した理由について述べる。4.2.1 節ではその中でも単語の親密度や具体性などの認知的な特徴量、4.2.2 節で

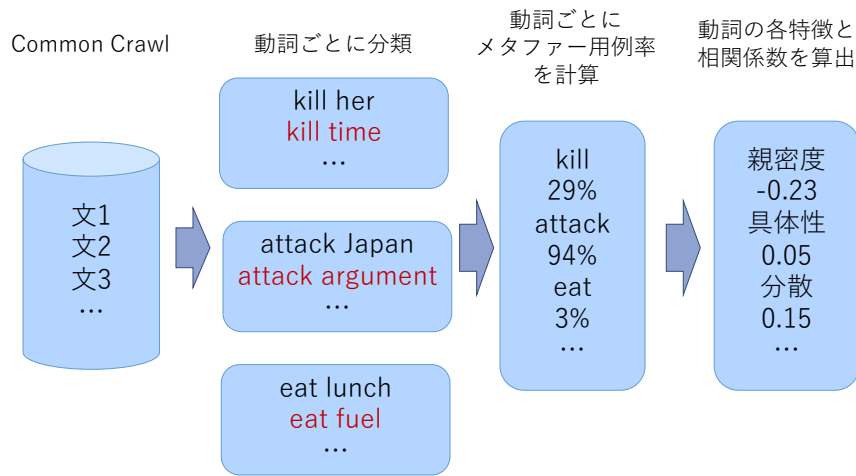


図 4.1 関係分析の全体像

はそれ以外の特徴量について述べる。

4.2.1 認知的な特徴量

まず、認知的な特徴量として、MRC Psycholinguistic database[16](以下 MRC データセット) から得られる動詞の親密度、具体性、イメージ可能性を採用した。MRC データセットは、150837 語にわたって各単語に言語学的、心理言語学的な 26 の特徴が紐づけられている語彙データベースである。その特徴も文字数、音素数、音節数などの言語学的特徴から語の familiarity (親密度 (F)), concreteness (具体性 (C)), imageability (イメージ可能性 (I)) などの心理言語学的な指標まで多岐にわたる。

各属性は 100~700 の値で定められており、例えば yoke(F:441 C:542 I:514) と cause (F:557 C:287 I:282) を見ると、まず yoke は「(家畜に) くびきをかける、(動物を) つなげておく」といった意味の動詞であり、cause は「(事態を) 引き起こす」という意味の動詞である。親密度はそれぞれの単語がどれくらい普段の生活で使うかを示しており、yoke などという動詞は一般的には馴染みが薄く、あまり使わないものと考えられる。逆に cause は日常生活で用いる場面が多い。そういった理由で yoke は familiarity の値は低く、cause は高くなっている。具体性はそれぞれの単語がどれくらい具体的なものやひとと結びついているかを示しており、yoke は現実世界の動物を繋いでおくもの、例えば牛をつなぐくびきや首輪と結びついているが、cause は直接現実世界の何かと結びついているとは思えない。そのため concreteness は yoke の方が高くなっている。イメージ可能

性はその単語を聞いた時に絵や音などの感覚的経験がどれくらい早く、強く思い起こせるかを表す数値であり、yoke という単語を知っていれば動物が繋がれている風景をすぐにイメージできるが、cause と聞いても感覚的な経験をすぐには思い起こせない。そのためイメージ可能性も yoke の方が高くなっている。なお、データベースにおけるこれらの値は心理学実験において人手で作成されている。

2.3.1 節で述べた概念メタファー論によると、メタファーとは概念間の写像であり、より分かりにくく、馴染みの薄い抽象的な概念をより理解しやすい具体的な概念を通して理解する人間の認知のプロセスであるとされていた。そこで動詞メタファーにおいても、メタファー用例となるのは写像先の動詞であるため、メタファー用例の動詞はより分かりやすく、具体的な動詞であると予想する。そこで、MRC データセットで得た親密度や具体性、イメージ可能性を用いてそれを検証する。もしその仮説が一般的に成り立つならメタファー用例率と親密度や具体性、イメージ可能性は正の相関を持つはずである。

4.2.2 目的語の多様性に関する特徴量

本研究では、以上で述べた特徴量以外に、コーパスから収集した文を用いて動詞の目的語の多様性に関する特徴量について相関を取得した。目的語の多様性については、多様性が小さいほど動詞がメタファーになりやすいという仮説を検証するために導入した。例えば動詞 eat は目的語に主に食べ物をとるが、目的語に食べ物以外、例えば人がきて「eat him up」となったら彼を食い尽くす、つまりその人の人間性を味わうといったニュアンスになるし、燃料などの消費物がきて「eat fuel」となったら燃料を消費するといった意味のメタファーになると思われる。それに対して動詞 see は目的語にとる範囲が広い。ものであったらなんでも見ることができるのである程度目的語が何であってもメタファーにならないと推察される。よって目的語の範囲が限定されている eat などの動詞の方が、目的語を広くとる see などの動詞よりメタファー用例率が高くなると予想する。その仮説を実証するために導入した特徴量が目的語の多様性であり、単一の指標では正しく評価できるか信頼性が薄いので複数の指標によって多様性を評価した。

目的語の分散

まずは動詞ごとに目的語がどのくらい意味的に異なる表現で出現するかを示す数値を得るため、いくつかの方法で目的語の多様性を測定した。大まかな手法については図 4.2 で示す。まずは目的語の埋め込みを取得する方法について述べる。本研究ではメタファー用例率を計算するため動詞とその目的語のペアを大量に集める。そこで集まった目的語を word2vec の skip-gram モデル [4] でベクトル化する。そうやった得たベクトル集合を大

きさ 1 に正規化し、次元ごとに分散をとり、その積を目的語の分散とするのが一つ目の方法である。

目的語のコサイン類似度の平均

次に目的語の意味的な類似性をより重視した多様性の測定方法について述べる。集まった目的語集合をベクトル化するまでは先の手法と同様に行い、そこからランダムに選んだ 200 目的語ベクトルにおけるコサイン類似度の平均をここでは多様性とする。このようにすると、目的語の範囲が狭い動詞は目的語が互いに近い意味を持っているので、目的語間のコサイン類似度が大きくなり、この値も大きくなる。

異なり数、エントロピー

次に目的語の意味は考慮しない手法について述べる。ここでは単純に動詞ごとに目的語の種類、つまり異なり数を取得する手法と、目的語の確率分布を作成し、その情報エントロピーを取得する手法の二つを行なった。異なり数については各動詞が使われている文を 10 万文集め、使われている目的語の種類を数えることによって取得した。エントロピーについては異なり数と同様動詞ごとに集めた 10 万文についてそれぞれ目的語の確率分布を作成し、以下の式によって取得した。

$$Entropy = -\sum_{i=1}^n p(i) \log_2 p(i) \quad (4.1)$$

n は集まった動詞-目的語のペアの数 (ここでは 10 万)、 $p(i)$ はその動詞において i 番目の目的語が現れる確率である。

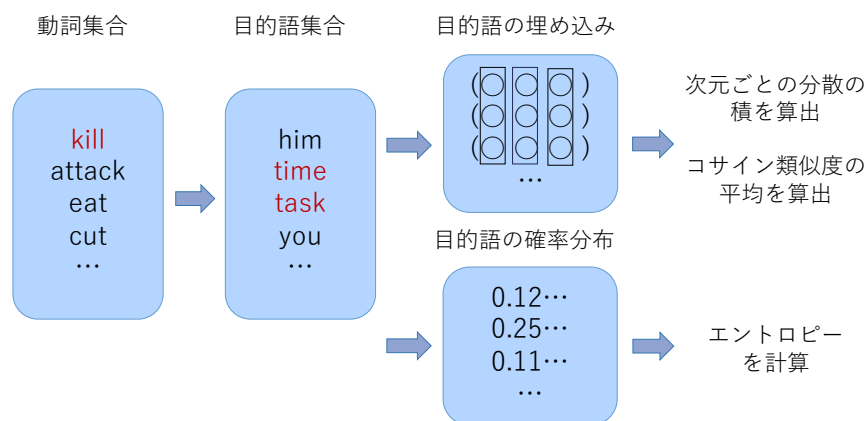


図 4.2 目的語の多様性を数値化する手法

第 5 章

実験

本章では本研究で行なった実験について述べる。5.1 節では本実験で使⽤したデータセットを紹介する。また、5.2 節では実験設定として本研究で行ったタスクの詳細な設定を述べ、5.3 節ではその結果を述べる。最後に 5.4 節ではその結果をもとに考察を行う。

5.1 使用するデータセット

動詞-目的語のペアを集めるための生文として大規模なウェブデータを集めたデータセットである Common Crawl である cc-100 のうち 7.5 億文を⽤し、動詞の特徴量を取得するためのデータセットとして 4.2.1 節で述べた MRC データセット [16] を⽤した。また、本実験で直接⽤したデータセットではないが、メタファー判別器の学習元データセットとして VU Amsterdam metaphor dataset[14] も⽤している。

5.2 実験設定

本実験では 60 動詞に絞って実験を行う。いずれも MRC データセットにデータが存在する動詞であり、そのうち 10 動詞はメタファー⽤例の思いつきやすい、メタファー⽤例と非メタファー⽤例の判別が容易な恣意的に選んだ動詞であり、残りの 50 動詞は cc-100 の中でも頻度の高い 100 動詞からランダムに選んだ。恣意的に選んだ 10 動詞の一覧と、それぞれメタファーとして⽤されている⽂例と⽤されていない⽂例を表 5.1 に示す。また、残りの 50 動詞の一覧を表 5.2 に示す。実験では cc-100 の最初から約 7.5 億文から先述した 60 動詞とその目的語を含む⽂を抽出し、動詞ごとに⼤量の動詞-目的語ペアを集めた。そこからさらに同じ目的語を持つ⽂を集め、その結果同じ動詞-目的語ペアを持

表 5.1 恣意的に選んだ 10 動詞とメタファー用例

動詞	メタファー用例の例	非メタファー用例の例
kill	How did you kill time ?	He killed the enemy.
attack	He attacked my argument .	America attacked Japan
break	He broke a bad habit .	He broke the window.
eat	This machine eats a lot of fuel .	He eats lunch.
read	Reading between lines is important.	He read the book.
cut	The fact cut his heart into two pieces.	He cut a paper with scissors.
heal	He healed our broken relationship .	The wounds is not healed.
build	build a 50-50 relationship .	The house was built.
plant	I planted a bad thought .	I plant some flowers.
ride	He always rides a fad .	He rides a bike.

表 5.2 ランダムに選んだ 50 動詞の一覧

動詞一覧				
come	win	love	ease	view
back	see	search	keep	talk
form	name	know	earn	sit
sign	learn	design	experience	beg
head	end	look	effect	plan
allow	help	land	rent	get
like	own	work	sing	aid
make	want	cause	rate	call
help	gain	count	find	stand
play	mean	lease	point	hand

つ文が 10 文以上あるペアのみここで取り扱った。そして各文をメタファー判別器である MelBERT モデルに入力し、その動詞-目的語ペアがメタファーなのか判別を行なった。

ここで判別における精度を上げるため同じ動詞-目的語ペアをもつ文集合のうち 7 割以上が判別器によってメタファーだと判別されたらそのペアはメタファーであり、3 割以下であればメタファーでないと判別し、残りの動詞-目的語ペアは以下の実験には使用しなかった。最終的に動詞ごとにどのくらいの割合の目的語がメタファーと判別されたかを調

べることで動詞ごとのメタファー用例率を取得した。

その後目的語の多様性を調べたり、散布図を作成するため同じ動詞-目的語ペアが 10 文以上ある目的語を静的な単語埋め込みである skip-gram モデル [4] を用いてベクトル化した。散布図を作成するにはそれらを TSNE を使って二次元に圧縮し、分散を調べる際には 4.2 節で述べた方法で分散を取得した。

最後に動詞ごとのメタファー用例率と目的語の多様性などの自前の特徴量や動詞の具体性などの MRC データセットから得た特徴量などと相関係数を計算し、それらがどのような関係なのか調べた。

5.3 実験結果

5.3.1 目的語の分布

メタファー用例と非メタファー用例で分かれたもの

図 5.1 では、左が kill の目的語の埋め込み、右が see の目的語の埋め込みを表している。kill を見ると、「hour」や「economy」など、生物でない目的語がメタファーと判別されていて、「terrorist」や「mouse」などの生物に関する目的語が非メタファー用例と判別されている。see の場合は「recovery」や「career」など、直接目に見えない抽象的な目的語がメタファーと判別され、それ以外が非メタファーと判別されている。それぞれその意味の違いがベクトル表現に現れ、赤点と黒点がくっきりと分かれている。

メタファー用例と非メタファー用例が混在しているもの

図 5.2 では左が earn の目的語の埋め込み、右が look の目的語の埋め込みの分布を表す。どちらも先に挙げた動詞よりはメタファー用例、非メタファー用例の目的語の埋め込みが混ざり合っているように思われる。

ほとんどどちらかの用例しかないもの

図 5.3 では左が attack の目的語の埋め込み、右が cause の目的語の埋め込みの分布を表す。これを見ると attack はほぼメタファー用例としてしか使われておらず、cause は逆にメタファー用例としては使われていないことがわかる。

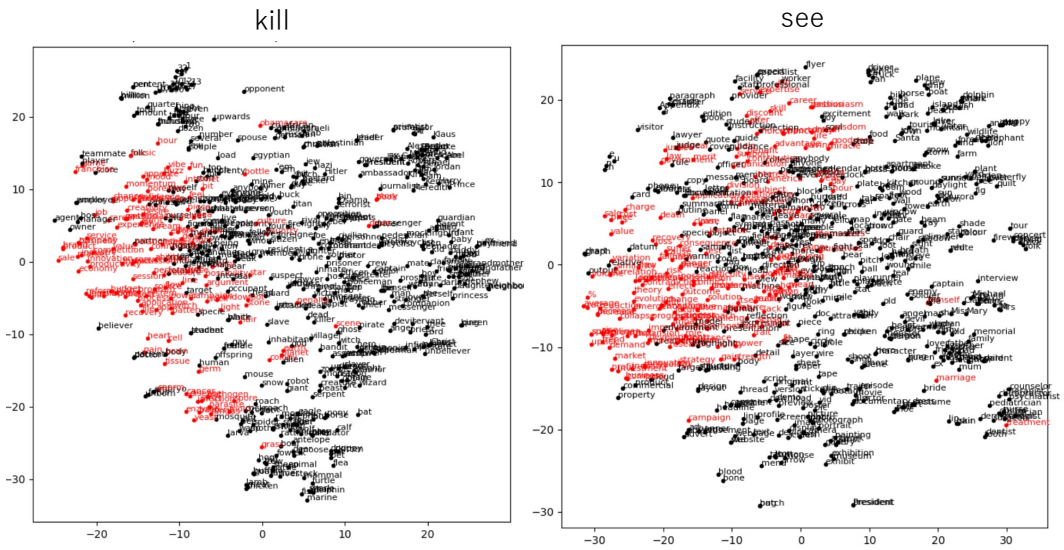


図 5.1 目的語の散布図 左が kill、右が see

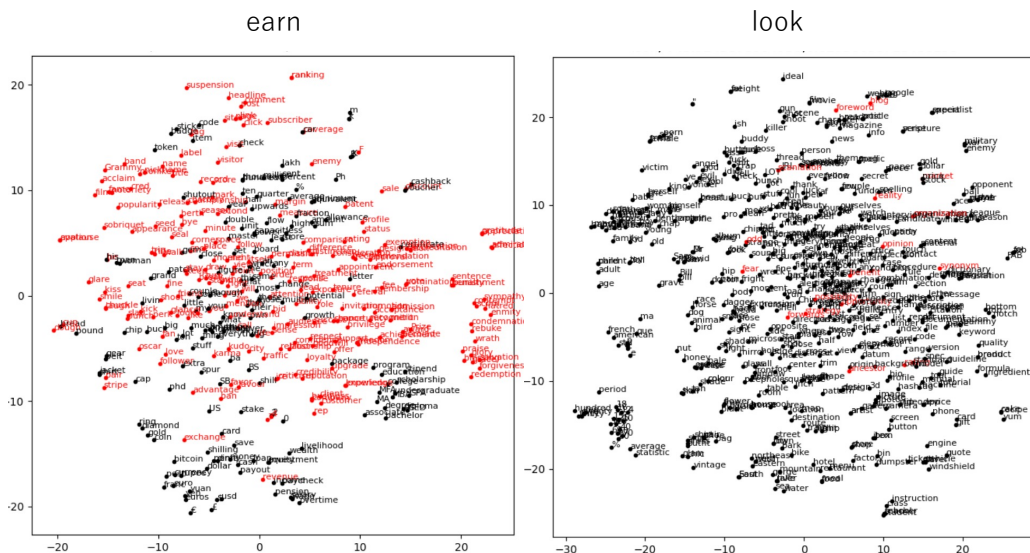


図 5.2 目的語の散布図 左が earn、右が look

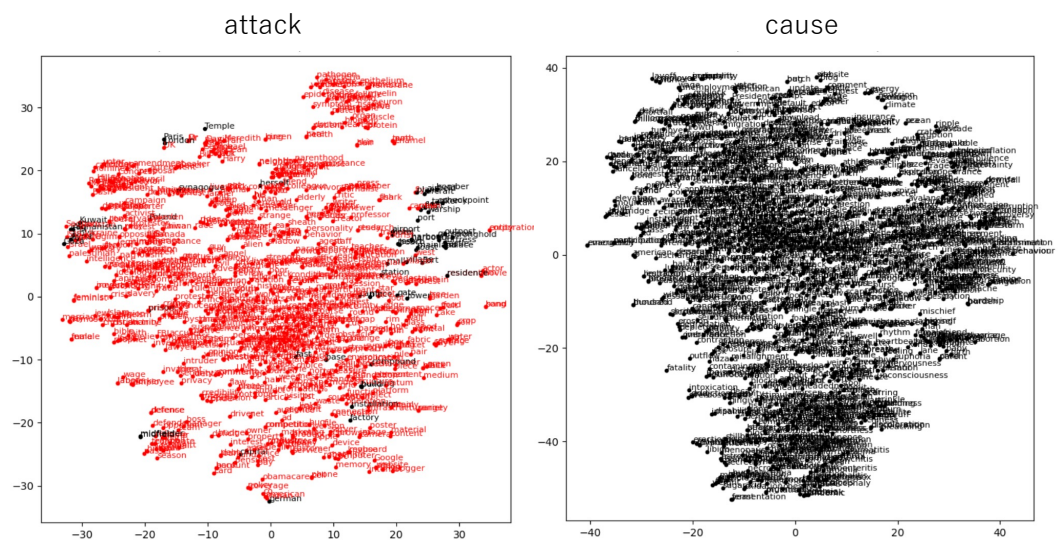


図 5.3 目的語の散布図 左が attack、右が cause

以上 3 パターンの動詞を見てきたが、多くの動詞は 1 つ目のように完全でなくとも、メタファー用例と非メタファー用例で埋め込みの分布が分かれていた。それを検証するために、それぞれのグループの集まり方の評価尺度として、それぞれのグループの集まり方を評価するために以下の密集度比を動詞ごとに計算した。

$$\text{密集度比} = \frac{\text{異色のコサイン類似度の平均}}{\text{同色のコサイン類似度の平均}} \quad (5.1)$$

コサイン類似度はベクトルの角度が近いほど値が大きくなるので上の密集度比が 0 に近いほど同じ色同士で密集していると言える。これを 60 動詞で調べた結果、以下の図 5.4 のような結果となった。

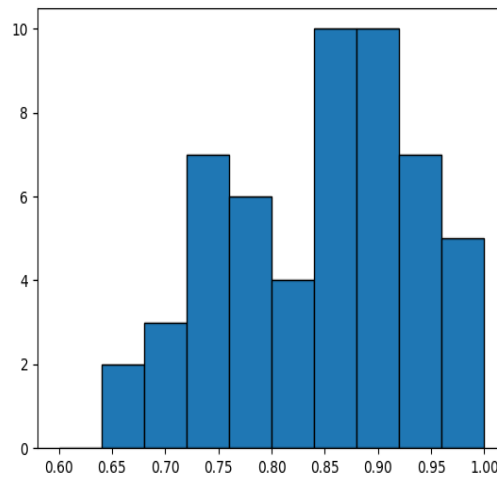


図 5.4 密集度比のヒストグラム

0.6 から 1.0 の範囲に 55 動詞が集まった。なお、5 動詞はメタファー用例率が 0.0 か 1.0 であり、密集度比が定義できないため、分析対象から外した。5.3.1 節でメタファー用例と非メタファー用例で分かれたとして紹介した kill と see はそれぞれ密集度比は 0.76、0.73 と全体から見てもそれぞれの用例で意味的にまとまっていることがわかる。以上より、メタファー用例と非メタファー用例の目的語が多少なりとも意味的なまとまりがあることがわかる。

5.3.2 各特徴との相関

各特徴との相関係数は以下のようになった。動詞の具体性やイメージ可能性などの認知的な特徴量との相関を表 5.3 に、目的語の多様性に関する特徴量を表 5.4 に示す。

表 5.3 認知的な特徴量との相関係数

特徴量	親密度	具体性	イメージ可能性
相関係数	-0.23	0.03	-0.12

表 5.4 目的語の多様性に関する特徴量とメタファー用例率との相関

特徴量	次元ごとの 分散の積	コサイン類似度の 平均	エントロピー	異なり数
相関係数	0.15	-0.15	0.18	-0.02

動詞の親密性に弱い相関が見られ、目的語におけるコサイン類似度の平均と目的語のエントロピーにもさらに小さいながらも相関が見られた。なお、本研究で実験した動詞のうち恣意的に選んだ 10 動詞、さらにランダムで選んだうちから 10 動詞とその全ての特徴量、さらに平均 0、分散 1 に正規化した特徴量を表 5.5 と表 5.6、表 5.7 に示す。

表 5.5 20 動詞の認知的な特徴量とメタファー用例率

	親密度	具体性	イメージ可能性	メタファー用例率
love	619	311	569	0.00
talk	613	422	453	0.01
read	568	420	499	0.02
eat	529	486	563	0.03
win	582	364	454	0.05
see	625	352	379	0.17
ride	576	424	483	0.19
plant	592	594	605	0.20
kill	549	386	481	0.29
search	537	371	402	0.35
build	554	402	399	0.57
cut	581	430	460	0.57
view	523	379	430	0.58
heal	536	417	438	0.90
keep	584	339	284	0.90
attack	542	411	501	0.94
break	529	360	398	0.95
ease	519	305	327	0.97
come	608	355	322	0.97
back	587	540	483	0.99

表 5.6 分散、コサイン類似度の平均とメタファー用例率

	次元ごとの 分散の積	次元ごとの 分散の積 (正規化)	コサイン類似度の 平均	コサイン類似度の 平均 (正規化)	メタファー用例率
love	-1109	0.88	0.21	-0.88	0.00
talk	-1112	0.56	0.22	-0.55	0.01
read	-1108	1.00	0.21	-1.02	0.02
eat	-1124	-0.94	0.26	1.05	0.03
win	-1110	0.75	0.21	-0.74	0.05
see	-1107	1.15	0.20	-1.19	0.17
ride	-1122	-0.81	0.26	0.93	0.19
plant	-1130	-1.71	0.28	1.73	0.20
kill	-1112	0.50	0.22	-0.53	0.29
search	-1114	0.32	0.23	-0.36	0.35
build	-1107	1.15	0.20	-1.21	0.57
cut	-1111	0.66	0.22	-0.70	0.57
view	-1108	1.04	0.20	-1.11	0.58
heal	-1124	-0.92	0.26	0.95	0.90
keep	-1105	1.39	0.19	-1.47	0.90
attack	-1110	0.67	0.22	-0.68	0.94
break	-1109	0.86	0.21	-0.87	0.95
ease	-1123	-0.89	0.26	0.95	0.97
come	-1111	0.61	0.22	-0.61	0.97
back	-1114	0.22	0.23	-0.20	0.99

表 5.7 異なり数、エントロピーとメタファー用例率

	異なり数	異なり数 (正規化)	エントロピー	エントロピー (正規化)	メタファー用例率
love	8490	1.17	8.9	0.71	0.00
talk	8999	1.43	8.7	0.61	0.01
read	5319	-0.45	7.3	-0.96	0.02
eat	4856	-0.69	8.0	-0.17	0.03
win	5212	-0.51	8.3	0.10	0.05
see	9233	1.54	9.3	1.13	0.17
ride	5965	-0.12	7.2	-0.97	0.19
plant	4517	-0.86	7.2	-1.05	0.20
kill	7178	0.50	7.9	-0.24	0.29
search	6491	0.14	8.6	0.43	0.35
build	5726	-0.25	9.1	1.00	0.57
cut	6036	-0.09	8.9	0.79	0.57
view	6587	0.19	9.0	0.86	0.58
heal	4075	-1.09	7.6	-0.63	0.90
keep	6824	0.31	8.6	0.48	0.90
attack	9394	1.63	9.4	1.27	0.94
break	5985	-0.11	8.5	0.37	0.95
ease	3762	-1.25	7.3	-0.91	0.97
come	9130	1.49	8.6	0.44	0.97
back	6973	0.39	8.4	0.22	0.99

5.4 考察

本節では以上の結果を踏まえた考察を行う。これまでの実験により、より親密度が低く、目的語の多様性が大きい動詞ほどメタファーになる傾向があることが分かった。本研究では概念メタファー論に基づいてメタファー用例の動詞はより分かりやすい動詞であり、親密度は高くなると予想していた。しかし、実際には親密度が低い動詞がメタファーになりやすいという結果が得られ、以上のような仮説は実証できなかった。その理由としては、あまりに親密度の高い動詞であっても目的語に決まった語しか取らない動詞はメタファーとして使われにくいことが挙げられる。表 5.5 を見ると、eat や read などのより簡単に親密度の高い動詞でもメタファー用例率が低いことがわかる。これらの動詞は目的語として食べ物や文書などの決まった領域の語をよく目的語とし、メタファー用例となる目的語の出現頻度は低い。このような理由により親密度とメタファー用例率の間に予想と異なる相関が見られたと考える。

また、目的語の多様性に感しては目的語に様々な語をとる動詞ほどメタファーになる傾向があることが分かる。これも先程の考察と同様に、目的語に決まった語しか取らない動詞はメタファーになりにくいことが原因として挙げられる。実際にメタファー用例率の高い break や keep などを見ても、なんでも物質であれば壊すことや保つことはでき、目的語にいろいろなものをとれることがわかる。しかし、相関があまり大きくないことや目的語の多様性が大きいからメタファー用例率が高いのか、それともメタファーとして使われるから目的語の多様性が大きいのかといった因果関係がこの結果からはわからないため、あまり強い結論を導くことができなかった。

第 6 章

おわりに

本章では、本研究のまとめと今後の展望を述べる。6.1 節では本研究のまとめを述べ、6.2 節では課題と今後の展望について述べる。

6.1 本研究のまとめ

本研究では Common Crawl から抽出した文を自動メタファー判別器で判別することにより、動詞ごとにそれがメタファーとして使用される割合であるメタファー用例率を調べた。そしてそれらの割合が動詞のどのような性質と関係するのかを調べるため、動詞の様々な特徴量と動詞のメタファー用例率との間の相関係数を算出した。その結果、動詞の親密度と弱い負の相関が得られ、目的語の多様性と微弱ながらも正の相関を得られた。

それらの結果により概念メタファーなどのメタファーにおける仮説を実証することはできなかったが、より親密度の低い、つまり一般的によく使われる動詞ほどメタファーとして使われにくい、また、目的語にある特定の分野の語しかとらない動詞はメタファーとして使われにくいという知見を得られた。

6.2 課題と今後の展望

本研究では、対象を動詞-目的語のペアに絞って分析を行なった。直接目的語をとるような動詞はその目的語によってメタファー用例かそうでないかが決まると考えたためである。しかし、実際にはそのような動詞ばかりではなく、主語によってメタファーとなる動詞も多い。そのような動詞についてその目的語を用いて相関などの分析を行うのはあまり得策とは言えない。本研究では分析対象となる 60 動詞のうち、50 動詞はランダムに選ん

でいるため、そのような影響を考慮できておらず、主語も考慮に入れたような分析もこれから行いたいと考える。

また、本研究では大量に文を集める必要があるため頻度の高い 60 動詞に絞って分析を行なったが、自動メタファー判別器を使っているという利点を生かしてさらに動詞を増やして実験を行うことによりより信頼性の高い結果を得られると考える。また、動詞の特徴量もデータベースとしては MRC データセットしか使用していないが、他にも単語の特徴量を収集した言語リソースは WordNet や VerbNet など数多く存在する。それらの特徴量も利用することによってもっとよい相関が得られるかもしれない。

謝辞

本研究は、武田・笹野研究室にて、武田浩一教授、笹野遼平准教授のご指導の下で行いました。この場で感謝の意を述べさせていただきます。

指導教員の武田浩一教授は研究のアイデア出しから論文の添削まで様々な部分で丁寧に優しくご指導いただきました。心よりお礼を申し上げます。

笹野遼平准教授は研究のテーマ決めから論文の校正まで、研究全般で細かく的確にご指導いただきました。深く感謝いたします。

研究室の皆様に日頃から刺激をいただいたことで、研究をやり抜くことができました。ありがとうございます。

参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pp. 4171–4186, 2019.
- [2] Minjin Choi, Sunkyung Lee, Eunseong Choi, Heesoo Park, Junhyuk Lee, Dongwon Lee, and Jongwuk Lee. MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pp. 1763–1773, 2021.
- [3] Zellig S. Harris. Distributional structure. *WORD*, Vol. 10, No. 2–3, pp. 146–162, 1954.
- [4] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems (NeurIPS) 26*, pp. 3111–3119, 2013.
- [5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, (ICLR) 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.
- [6] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.

- [7] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146, 2017.
- [8] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pp. 2227–2237, 2018.
- [9] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP)*, pp. 38–45, 2020.
- [10] George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, 1980.
- [11] Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 248–258, 2014.
- [12] Saif Mohammad, Ekaterina Shutova, and Peter Turney. Metaphor as a medium for emotion: An empirical study. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics(*SEM)*, pp. 23–33, 2016.
- [13] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.
- [14] G.J. Steen, A.G. Dorst, J.B. Herrmann, A.A. Kaal, T. Krennmayr, and T. Pasma. *A method for linguistic metaphor identification. From MIP to MIPVU*. John Benjamins, 2010.
- [15] Agnieszka Mykowiecka, Aleksander Wawer, and Malgorzata Marciniak. Detecting figurative word occurrences using recurrent neural networks. In *Proceedings of the Workshop on Figurative Language Processing*, pp. 124–127, 2018.
- [16] Max Coltheart. The mrc psycholinguistic database. *The Quarterly Journal of*

- Experimental Psychology Section A*, Vol. 33, No. 4, pp. 497–505, 1981.
- [17] Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. Neural metaphor detection in context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 607–613, 2018.
 - [18] Verna Dankers, Marek Rei, Martha Lewis, and Ekaterina Shutova. Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2218–2229, 2019.
 - [19] Xiaoyu Tong, Ekaterina Shutova, and Martha Lewis. Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pp. 4673–4686, 2021.
 - [20] Serra Sinem Tekiroğlu, Gözde Özbal, and Carlo Strapparava. Exploring sensorial features for metaphor identification. In *Proceedings of the Third Workshop on Metaphor in NLP*, pp. 31–39, 2015.
 - [21] Luana Bulat, Stephen Clark, and Ekaterina Shutova. Modelling metaphor with attribute-based semantics. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 523–528, 2017.
 - [22] Ekaterina Shutova, Douwe Kiela, and Jean Maillard. Black holes and white rabbits: Metaphor identification with visual features. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pp. 160–170, 2016.
 - [23] Hongyu Gong, Kshitij Gupta, Akriti Jain, and Suma Bhat. IlliniMet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pp. 146–153, 2020.
 - [24] Stephen Skalicky and Scott Crossley. Linguistic features of sarcasm and metaphor production quality. In *Proceedings of the Workshop on Figurative Language Processing*, pp. 7–16, 2018.
 - [25] Yorick Wilks. A preferential, pattern-seeking, semantics for natural language

- inference. *Artificial Intelligence*, Vol. 6, No. 1, pp. 53–74, 1975.
- [26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, Vol. 9, No. 86, pp. 2579–2605, 2008.