修士論文

単語分散表現のなす幾何的配置と
単語共起行列のもつ内部構造の分析

学生番号 2130017
前田晃弘

主指導教員 日髙昇平

北陸先端科学技術大学院大学
先端科学技術研究科
（情報科学）

令和 5 年 3 月

Abstract

It is well known that a word vector is able to solve analogy tests by vector arithmetic, implying that semantic relations are embedded into the vector space and represented geometrically as a parallelogram. Motivated by Harris' structural linguistic view, this study aims to (1) clarify the distributional structure and semantic properties manifested in the distributed representation of words, (2) mathematically formulate the latent structure of the word co-occurrence matrix which is used to learn word vectors, and (3) demonstrate the existence of distributional structures in real corpora.

First, we take a constructive approach and propose a toy model where an artificial corpus replicates word vectors forming geometric shape such as parallelepiped. We mathematically derive the necessary and sufficient conditions for word vectors to form a parallelepiped in high-dimensional vector space. It is revealed that the existence of certain linguistic relations, syntagmatic and paradigmatic, provide the necessary conditions of a parallelogram.

Second, we empirically prove that the word vectors in the semantic relations are, in fact, geometrically arranged in high-dimensional vector spaces. Using a real corpus, we show that there is a significant difference between the distribution of geometric distances between a set of words in analogy relations and randomly selected words.

Third, we propose a mathematical formulation to encode a general sentence into matrix representation and show that the distribution structure of a corpus can be represented as a tensor product. This suggests the regularity and multi-linearity of the language structure.

The contribution of this study is the mathematical formulation to algebraically treat the process of generating the word co-occurrence matrix from sentences and to characterize the obtained word vectors geometrically in the vector space. This study paves the way for future research to construct a computational model of natural language based on distributed representations of words and continued research is expected to be promising.

目次

図目次

表目次

第 1 章

はじめに

1.1　自然言語処理におけるめざましい進化

人工知能における自然言語処理技術の進化は加速度的である．特に，直近 10 年の発達は目覚ましく，2022 年 11 月末に突如出現した対話型言語モデル ChatGPT (OpenAI, 2023) は並外れた言語処理能力を示した．人間からのあらゆる分野の質問に対し適切で筋の通った答えを返し，求めに応じて論文や小説を書くなど知的能力を持つかの様である．

ChatGPI の基盤である GPT-3 (Brown et al., 2020) や，GPT-3 と同じ学習手法である Transformer (Vaswani et al., 2017) を初めて導入した BERT (Devlin et al., 2019) は，いずれも Web 等の膨大なテキストデータから大量のパラメータを学習する大規模言語モデルである．

いずれも公開時に多くの自然言語理解タスクで最先端の性能を達成していたが，その一つ SQuAD（スタンフォード質問回答データセット) と呼ばれる，文章を読んで質問に答える課題に対しても高い精度を示しており (van Aken et al., 2019)，当時より推論する能力を備えているかの様であった．

GPT-3 や BERT を含めた深層学習によるすべての言語モデルに共通するのが，ベクトルによって単語を表現することである．深層学習では情報はニューラルネットワーク上で分散的に表現されるが，言語モデルにおいては単語を表現するベクトルを単語分散表現や単語埋め込みと呼ぶ．

2013 年にコーパス中の単語を予測する言語モデルを用いて学習した単語分散表現 word2vec が，300 次元程度のベクトルにもかかわらず単語の意味を捉えているかのような性質を示した (Mikolov et al., 2013a) ことは，当時驚きを持って受け止められた．

単語分散表現を用いた演算により，単語間の意味的な関係や統語構造を表したり, 文を

構成する単語に対して，文の木構造に沿って分散表現の演算を行うことで文レベルでの意味表現や質問回答のための推論が実現されている．

技術進化によりさまざまな言語タスクにおけるパフォーマンスが改善されてきたが，大規模言語モデルが機能する理由はまだよく理解されておらず現在も進行する研究課題である．例えば，BERT の内部状態を解明する研究は BERTology と呼ばれる独立した研究領域となっている (Rogers et al., 2020)．

ここで次の二つの疑問が自然に湧き上がる．一つ目は，これらの大規模言語モデルは人間の言語を理解しているかという疑問であり，二つ目は，そもそも言語を理解するとはどの様なプロセスかという疑問である．前者は，いわば「機械」を対象とする問であるのに対して，後者は「言語の理解」を対象とする (Gastaldi, 2021)．

この二つの疑問はコインの裏表であり車の両輪のような関係にある．機械により言語が理解され生成されているかのように見える時，その内部で行われている計算プロセスを理解することは人間が言語を理解するプロセスを明らかにする．また，言語が持つ構造を明らかにすることは，言語を機械で扱うための適切なプロセスと計算モデルを明らかにするはずである．

この依然ナイーブな疑問にアプローチするための手がかりとして，次に計算モデルに関する並列分散仮説 parallel distributional hypothesis と言語に関する分布仮説 distributional hypothesis という二つの distribution（それぞれ日本語では分散と分布と訳されている）について概説する．

1.2　認知科学の並列分散処理仮説

2010 年代に入り自然言語処理技術が加速度的に進化した背景には，言語モデルへのニューラルネットワークの導入が挙げられる (Jurafsky & Martin, 2023)．ニューラルネットワークとは，脳の神経細胞網をモデルとして考案された計算モデルあるいは計算手法であり，情報をネットワーク上のノードの活性状態で表現し，活性状態の伝播とその際のノード間の結合の変化により情報処理のための計算を行うものである．特に，ネットワークを多層化したニューラルモデルによる学習は，深層学習 deep learning と呼ばれ (Hinton et al., 2006; Bengio et al., 2006)，BERT や GPT-3 でも用いられた計算技法となっている．

ニューラルネットワークを認知プロセスの計算モデルとして提示したのが Rumelhart et al. (1986) であり，古典的な記号主義に対してコネクショニストモデルと呼ばれた．コネクショニストモデルでは，知識・概念・情報はネットワーク上のノードの活性化パタン

として分散的に表現され，並列的に処理され学習される．このプロセスは，並列分散処理 Parallel Distributed Processing（PDP）(Rumelhart et al., 1986) と呼ばれた．特に，概念を記号として表現する古典な記号主義に対して，PDP では概念を分散表現として捉えることが強調された．

McClelland and Kawamoto (1986) は，言語処理を対象に PDP による文理解のモデルを提示した．そこでは，単語の概念は微細特徴 microfeatrure の組み合わせとして分散的に表現されており，名詞であれば，例えば，人間か非人間の別，大きさや色，感触などの属性が数値化されており，単語の分散表現はそうした数値の集まりとして表された．さらに，こうした特徴は単語により構成される文から学習可能であることが示された．

PDP の流れを汲む Elman (1990) の再帰型ニューラルネットワーク (Recurrent Neural Network; RNN) は現在の言語モデルにも連なるが，RNN により学習した内部表現が言語的特徴を獲得していること示した．

しかしながら，現在では RNN は自然言語処理の言語モデルの発達史の一部として扱われており，自然言語処理研究の中で単語分散表現が何を表象しているのかを直接問われることはないように思われる．その背景には，大規模言語モデルが獲得する大量のパラメータを直接解釈することは難しいことと，表現学習 (Representation learning) という考え方 (Bengio et al., 2013)，すなわち潜在的な構造を持つデータからタスクの目的に応じた最適な特徴を抽出するための学習と捉える立場から，分散表現自体よりも表現を獲得するためのアルゴリズムへ関心が移ってきたことが考えられる．

1.3　構造言語学の分布仮説

現在，自然言語処理分野の論文を見ると，単語分散表現に関する多くの先行研究が，単語分散表現の理論的根拠を Harris (1951, 1954) や Firth (1957) によって提唱された分布仮説に求めている．分布仮説はしばしば Firth の「単語の意味は，周囲の単語によって形成される」という一文により説明されるが，その源流は構造主義までさかのぼる．言語について論じた哲学者 Wittgenstein は「語の意味とは、言語のなかでのその使用である。」(Wittgenstein & 丘澤静也訳, 2013) と記し，記号論の祖で構造言語学を提唱した Saussure は，言語の本質を関係の網として捉え，言語内の単位は構造的で相互依存の関係にあり，差異が意味を生み出すと考えていた (丸山圭三郎, 1983)．

Harris (1951) は，言語学を自律的かつ科学的なものにしようとする立場から，言語には distributional structure（分布構造）が存在していることを主張し，意味論的側面に言及せずに分布構造を記述することを目的とした構造言語学を確立しようとした．Harris

(1954) は，言語を構成する単位を要素（音素，意味素，単語など）として定め，要素間の共起関係に規則性があることを示した．要素間の規則的な共起関係を分布関係と呼び，言語に現れる全ての分布関係を分布構造と呼んだ．Harris は数学における代数構造（集合に備わる演算や作用によって決まる構造）を念頭に置いていたと考えられ (Harris, 1954)，言語は分布構造により記述できるという分布主義を主張した．

Harris は，言語学は言語の持つ自律的構造を解明すべきとのブルームフィールド学派の立場から，意味や歴史的な要因に基づくことなく言語の分布構造を分析し，分布構造を厳密に定式化するための手順を示した (Lyons, 1970)．特に，音素や意味素の同定にあたり単語の意味を判断基準とすることを排除した．

Harris (1954) は一方で，言語と意味は特別な関連を持つが，言語の分布構造と意味の構造は必ずしも一対一には対応しておらず，また言語と意味の区分も不明確であるとした．意味は言語に固有のものではなく，また言語に付随するものではなく，言語とは別に存在する意味の主観世界 subjective world of meaning が存在することを示唆するのみであった．

興味深いのは，意味論に拠らずに厳密に定式化された分析によって発見された分布の規則性が，意味に関する含意を示すことである．Harris (1954) は，*knife* と *knives* の対と，*oculist*（眼科医）と *eye-doctor*（目医者）の対を対称的な例として挙げ，二つの単語がほぼ同じ単語分布を持つが，同時に生起することがあるか否かという観点で二つの対は異なる分布関係にあり，前者は補完的 complementary な意味的関係を，後者は類義関係を含意することを指摘している．分布主義が意味論を排除しようとしたにもかかわらず分布構造が意味構造と対応を持つことを示したことが，分布仮説が単語分散表現の理論的根拠を与えたとされる所以である．

1.4　単語分散表現と言語の分布構造

自然言語処理技術が発達する中で，コーパスを学習して得られた word2vec などの単語分散表現が単語間の意味的な関係を抽出していることはすでに述べた．具体的には，“King is to queen as man is to woman”(*king:queen::man:woman*) のような意味的な類推関係に対して，単語 *word* の単語ベクトルを $\boldsymbol{v}_{word} \in V$ として，４つの単語の単語ベクトルの間に $\boldsymbol{v}_{king} - \boldsymbol{v}_{queen} + \boldsymbol{v}_{woman} \approx \boldsymbol{v}_{man}$ が成り立つことが示された．近似の記号は左辺により計算されるベクトルと右辺のベクトルが同じ方向にあることを表す．

また，BERT などの大規模言語モデルでは，長文の意味を理解して問に答えるために，文章中の文間あるいは文中の単語間の関係などの文脈を分散表現として符号化し，その演

算により推論を行っていると考えられる．例えば，主語と述語に対応した単語ベクトルを組み合わせて文を表すベクトル（文脈化された単語ベクトルと呼ばれる）を生成している．

これらの現象は，単語分散表現と意味理解や意味生成との間に何らかの対応関係が成り立つことを示唆する．コーパスから共起頻度を直接カウントした単語共起行列を行列分解するだけで獲得した単語ベクトルでもニューラルモデルの単語分散表現に近い精度で類推課題を解けることが示されている (Torii et al., 2022)．また大規模言語モデルにおいて学習に用いるのはコーパス（単語列）のみであり，その学習過程で単語や文の意味に関する情報はそれ以外に付与していない．これらのことから単語の共起分布自体に意味関係の抽出につながる規則性が潜在していて単語分散表現の学習はこの分布関係を捉えていると考えられる．

単語間の共起関係を記録して，その分布関係に現れる規則性を抽出するというのが Harris の示した分析手続きであったという観点で，単語分散表現は構造言語学の提唱した分布主義の流れを汲むものといえよう．また，単語分散表現を獲得するための単語共起行列の分解・変換は，Harris が共起関係を分析する手続きの中で用いた表形式の解析と類似していることも大変興味深い．

一方，Harris の構造言語学的視点の内で現在見逃されているのは，Harris が言語には代数構造があると考えていた点である．単語と単語，あるいはその他の言語学的要素の間には，単なる共起関係だけではなく，より多様な関係性が存在しているのであり，要素間に作用するある種の演算を備えた代数構造として言語全体が記述されるというのが Harris の言語観である．

Harris の提唱する分析手続きによれば，単語間の分布関係を複数のクラス（代替的，補完的，類義的，品詞違いなど）に分類することが提案されている一方で，現在の単語分散表現の取り扱いではコサイン類似度などの指標により，単語の関係クラスによらない類似性の評価のみが想定されている．この点で両者は大きく異なる．言語学の立場からは，単語間の意味関係としては類似関係以外にも，Saussure が提示した paradigmatic（連合的）と syntagmatic（連辞的）な関係 (丸山圭三郎, 1983) のほか，類義関係，上下関係，排反関係，反意関係などの関係があることが指摘されている (Lyons, 1968)．

以上の単語分散表現と言語の分布構造に関する考察に着想を得て，研究の指針となる言語観を仮説として列挙する．

1. 単語分散表現を手がかりとして言語の分布構造を数学的に再定義あるいは再定式化できる．

2. 言語はなんらかの代数構造を有しており，単語共起分布に現れる規則性はその構造を反映している．
3. 単語分散表現を基盤とした言語モデルは意味理解や意味生成を実現ないし模倣している．
4. 単語分散表現の内部モデルの解明は言語理解や発話などの人の認知過程の解明につながる．
5. 分布構造と意味構造との関係を明らかにすることができれば新たな言語の計算モデルを構築できる．

これらの仮説的な構想は，著者の今後の長期的な研究に対して動機と指針を与えるものである．その出発地点となる本修士研究では，単語分散表現とその原データとなる単語共起行列を明示的に分析の対象として分布構造の観点から単語間の関係性を解明することを目指した．

1.5 本研究の目的と課題

■本研究の目的 本研究は，(1) 単語分散表現にあらわれる言語の分布構造と意味的な性質を解明すると共に, (2) 単語共起行列の内部構造を数学的に定式化した上で，(3) 自然言語コーパスに潜在する分布構造の存在を実証することを目的とする．

■課題認識 自然言語理解タスクで高い性能を示す大規模言語モデルの内部状態は，その大部分が未解明である．その一つの要因は，それらのモデルが基礎的構成要素にもつ単語分散表現の性質がまだ十分に解明されていないことにある．Harris は単語分散表現の前段階にある単語共起分布に代数的構造が存在すると主張していたが，当時はその代数的構造を定式化するための数学的ツールや言語モデルも確立されておらず，またその存在を実証するための大規模データと計算処理能力も十分ではなかった．

単語分散表現による自然言語の数理モデルを構築するにあたり，言語モデルが示す言語現象を具体的な代数系により記述し定式化して再現することがが本研究の課題であり貢献である．

■本研究の仮説 本研究は言語に代数的構造が備わると仮説を立てた。言語が備える代数構造は、言語の生成系は統語構造（生成ルール）と意味関係（型システム）により構成されると考える．

そのため，生成された文の集積であるコーパスに代数的な分布構造が現れ，さらにその

共起分布が言語モデルを通じて分散表現に写像（変換）される．大規模言語モデルが言語理解と生成のプロセスを模倣しているように見える理由は，言語生成系-単語共起行列-単語分散表現の住むベクトル空間に代数的な同型の構造が生まれるからである（図 1.1）．具体的に言えば、言語の備える代数構造として群や線形代数や束などが想定される．

図 1.1　生成系から分散表現までの同型対応

■問い　　本研究の問いは以下のとおりである．

1. 類推関係にある単語の分散表現が平行四辺形をなすのはなぜか，意味的関係を表象する図形は他にもあるか．
2. 単語共起行列に現われる言語の分布構造と，言語の生成系から規則性を持つ分布関係が生じるプロセスはどのように数学的に定式化できるか．
3. 単語分散表現は高次元ベクトル空間でどのような幾何的配置にあるか，幾何的図形からその単語間の意味関係を識別できるか．

■アプローチ　本研究では，構成論的アプローチとして人工的なコーパスから単語共起の分布構造を構成するトイモデルにより単語分散表現の数学的性質を幾何的に調べる．その際，群や線型代数の数学的道具を用いて行列表現や幾何的性質を考える．さらに簡単なトイモデルからスタートしてそれを徐々に一般化していく．

1.5.1　本論文の構成

第 2 章では，単語分散表現の概要と類型を説明すると共に，単語分散表現の数学的性質とそれが示す言語的な現象に関する先行研究を概観し，それらが示唆する単語分散表現の潜在構造（線形性をもつ構造）について考察する．

第 3 章では，単語分散表現の前処理段階で準備される単語共起行列に焦点を当て，その内部構造を数学的に分析する．類推関係にある単語の分散表現が平行四辺形をなすという現象に着想を得て，言語生成系の持つ対称性の構造が単語の共起分布を経て単語分散表現

へ変換されると，ベクトル空間で幾何的配置を示すことを，トイコーパスを用いて構成論的に示すとともに，その幾何的配置が現れるための言語構造や共起分布の必要十分条件を数学的に導出する．

第 4 章では，実コーパスから生成した単語分散表現を用いて，類推関係にある単語群の単語分散表現が現に平行四辺形の幾何的配置にあることを実証する．そのために高次元ベクトル空間の図形を数学的に定義して，複数の単語に対応するベクトル群がなす配置の図形らしさを評価するためのあらたなメトリックとして図形距離を導入する．

第 5 章では，第 3 章で扱ったトイコーパスを拡張して一般化した言語構造を数学的に扱えるよう，単語の集まりである文と文の集まりであるコーパスを符号化して数学的対象として扱うための定式化の方法を提示する．多次元の線形構造を扱うのに適したテンソルネットワーク活用の可能性に触れる．

第 6 章では，結論と今後の研究に残された課題を述べる．

第 2 章

単語分散表現の先行研究

2.1　単語分散表現の概要

単語分散表現は，言語学・心理学・認知科学・計算言語学・自然言語処理など多くの領域にまたがって進展してきたため，単語埋め込み，単語ベクトルなどさまざまな用語で呼ばれる．またその考え方あるいはモデルも，分散的意味論・ベクトル空間モデル・トピックモデル・潜在意味解析 (LSA) など分野や研究領域により異なる名称で呼ばれる (Lenci, 2018).

単語分散表現は，その理論的根拠とされる分布仮説を実装したものであり，単語の意味は単語の分布によって表現されるという仮定のもと (Firth, 1957)，コーパスを用いて得られた単語の共起分布に基づき，単語の意味を数学的に符号化したベクトルにより表現する.

単語分散表現は，主に単語の意味を実証的に分析するために用いられてきたが，現在では前章で見たように主に言語モデルの計算で用いられる.

初期の単語分散表現に関する研究としては，Harris (1951) の分布仮説に影響を受けて機械翻訳の研究プログラムに単語分散表現を適用した言語学者の Garvin (1962)，ベクトル空間で意味や概念を表現しようとした心理学者の Osgood (1952), 単語分布に基づいて意味の類似性の概念を確立した認知科学者の Miller (1967), さらに認知科学分野で単語分散表現を用いて意味記憶の計算モデルの理論的枠組みを構築した McRae and Jones (2013)，また計算機科学の情報検索領域では Salton et al. (1975) が文書に含まれる単語の分布で文書の内容を表すベクトル空間モデルを構築した.

2.2　単語分散表現の類型

単語分散表現は多分野に跨って開発されてきたため統一的な体系化はなされていないが，ここでは計算機科学の立場から Jurafsky and Martin (2023) と言語学の立場から Lenci (2018) を参考に類型化を試みる．

単語分散表現を生成手法で大きく分けると，コーパス中の単語の共起頻度を用いるカウントタイプと，コーパスを学習データとして単語を予測するモデルを学習する予測タイプに分類される． カウントタイプは，さらに単語共起頻度の数値に対する処理の違いにより，次元数がそのまま維持される顕在カウントタイプ[*1]と，次元数を削減して潜在構造を抽出する潜在カウントタイプに分かれる．また，予測タイプは，後工程に当たる言語タスクでそのまま用いられる静的ベクトルと，文脈を反映して変化していく動的ベクトルに分かれる．

■カウントタイプ（顕在カウントタイプ）　カウントタイプの単語ベクトルの基本的な生成手順は次のとおりである．ここまでの手順は顕在カウントタイプと潜在カウントタイプに共通であり，潜在カウントタイプはさらに追加の工程がある．

1. コーパスに出現する語彙をリストアップする．
2. 語彙ごとにコーパス中で分散表現の対象となる単語（ターゲット単語という）が特定のコンテクストに出現する頻度をカウントする．
3. カウントした出現頻度数を成分とした行列を構成する．これは共起行列と呼ばれ，その行はターゲット単語に，列はコンテクストに対応する．一定範囲内の単語をコンテクストとする場合には，共起行列は，行と列に同じ語彙が対応するので正方行列となる．
4. 行列の成分である出現頻度数を対数化やウェイト付などで変換する．
5. 共起行列の行を抽出して，その行に対応する単語の分散表現とする．
6. 単語ベクトルは単語間の類似度の評価を算出するために用いられる．

これを基本的な生成手順とした上で，次に掲げるパラメータによってさまざまな単語分散表現のバリエーションがある．

まず，コンテクストの選び方についてである．コンテクストとは，ターゲット単語が生

[*1] 顕在の呼び方は，Levy and Goldberg (2014b) と Lenci (2018) による

起する文脈のことであり，多くの場合，文中における語の並びにおいてそのターゲット単語の前後数語の単語を指す．ターゲット単語のコンテクストにある単語はターゲット単語と共起しているといい，その頻度を共起頻度と呼ぶ．このほかに単語ではなく一つの文書全体をコンテクストの単位とする場合もあり，この場合はターゲット単語とコンテクスト文書が共起するという（文書をコンテクストとする単語ベクトルのモデルはトピックモデルと呼ばれる）．

さらに単語単位でコンテクストを考える場合，通常は単語の位置関係に着目して，ターゲット単語の前後数語の範囲をコンテクストとし，その範囲内にある単語とターゲット単語の共起頻度をカウントする．この範囲はウィンドウと呼ばれ，ウィンドウサイズは前後 1 語から 10 語 (Jurafsky & Martin, 2023) であるが 3 から 5 語の場合が多い．ウィンドウサイズが小さいと統語的な関係が単語ベクトルに反映されやすくなり，逆にウィンドウサイズが大きいと意味的な関係が反映されることが経験的に知られている (Lenci, 2018)．また，文の区切り（ピリオド）を跨いでコンテクストを設定するか否かの区別もある．なお，文書をコンテクストとするということは，ウィンドウサイズを文書の語数とすることに実質的には等しい．また，ウィンドウサイズを無限とすると全ての単語の共起頻度は単語単独の分布（ユニグラム分布）に収束していく．

ウィンドウは通常ターゲット単語の前後同数を範囲とするが，片側だけをカウントする場合のほか，両側カウントするがその範囲を前後で非対称とする場合がある．また，ウィンドウサイズが 2 以上の場合，通常は，ターゲット単語までの語数にかかわらず出現頻度をそのままカウントするが，ターゲット単語の位置からの語数分に応じて出現頻度にウェイトをかけて割り引くこともある．

コンテクストの定め方のバリエーションとして，文の統語構造（木構造）に着目して，ノード間にあるエッジに沿ってウィンドウの範囲を定める方法もある (dependency-based と呼ばれる (Padó & Lapata, 2007)．位置関係に基づくウィンドウの場合は，統語構造上偶然隣接した単語ペアも共起としてカウントするためノイズが発生するが，dependency-based の場合にはそのような偶然性を避けて，統語構造上結びつきのある単語の共起のみをカウントする．ただし，実際上はコーパスに対して予め統語構造を割り当てるパーシングが必要となるため計算コストが高く，またパーシングの精度も十分高くないことから，位置関係によってウィンドウを定めることが一般的である．

次の主要なパラメータは共起頻度のウェイト付けの方法である．全コーパスにわたってターゲット単語とコンテクスト単語の共起頻度をカウントした数値を単語共起行列 $C = (C_{ij})$ の成分とした上で，その成分ごとに対数，条件付き確率，自己相互情報量

PMI(Pointwise Mutual Information) (Fano, 1961), 正の自己相互情報量 PPMI(Positive Pointwise Mutual Information) (Church & Hanks, 1990) を求める.

また文書をコンテクストとする場合には, TF-IDF(Term Frequency-Inverse Document Frequency) (Luhn, 1957; Sparck Jones, 1972) が用いられる. それぞれの算出式は次の通りである.

- 対数化： $C_{log_{ij}} = (\log C_{ij})$
- 条件付き確率 $C_{CP_{ij}} = \left(\frac{C_{ij}}{\sum_j C_{ij}} \right)$
- 自己相互情報量 (PMI) $C_{PMI_{ij}} = \log \left(\frac{C_{ij} \sum_{ij} C_{ij}}{\sum_i C_{ij} \sum_j C_{ij}} \right)$
- 正の自己相互情報量 (PPMI) $C_{PPMI_{ij}} = \max \left(\log \left(\frac{C_{ij} \sum_{ij} C_{ij}}{\sum_i C_{ij} \sum_j C_{ij}} \right), 0 \right)$

このようにして得られた単語共起行列の行ベクトルを，その行に対応する単語の分散表現とする.

単語分散表現の主な用途は類似度の評価である. 分散仮説に基づいて二つの単語の意味の類似度はそれぞれの単語分散表現であるベクトルの類似度で評価される. ベクトルの類似度としてはコサイン類似度が用いられることが一般的である. 二つのベクトル $\boldsymbol{u}, \boldsymbol{v} \in V$ のコサイン類似度 $\cos : V^2 \to \mathbb{R}$ は二つのベクトルのなす角度として示され，内積をベクトルのノルムの積で割ることで求められる.

$$\cos(\boldsymbol{u}, \mathbf{v}) = \frac{\langle \boldsymbol{u}, \mathbf{v} \rangle}{\|\boldsymbol{u}\| \|\mathbf{v}\|}.$$

顕在カウントタイプの特徴は，単語の共起頻度カウント（あるいは，ウェイトづけした値）が，そのまま共起行列の成分に現れることである. 行と列にはコーパスの語彙中の全単語が対応するので，各単語の分散表現の次元数は語彙数と同じであり高次元となる. 単語共起行列のほとんどの成分はゼロとなる疎な行列となるので，その後の汎化の際の取り扱いが難しい場合がある. 一方，単語ベクトルの各要素がそれぞれ共起するコンテクスト単語に対応しているので，言語的な解釈が可能となるメリットがある.

■潜在カウントタイプ　疎行列となり汎化が難しいという顕在カウントタイプの課題を解決するために用いられるのが潜在カウントタイプによる分散表現である. 顕在カウントタイプにより得られた単語共起行列に対して行列分解や次元削減を通じて特徴を抽出することで，その潜在構造をより小さい次元数で表現することができる.

潜在カウントタイプの先駆けとなったのが潜在意味解析 (Latent Semantic Analysis;

LSA) 呼ばれるモデル (Landauer & Dumais, 1997) であり，特異値分解 (Singular Value Decomposition; SVD) により単語共起行列を上位 k 個の特異値に対応した特異ベクトルへ写像して k 次元へ次元削減を行うものである．

SVD の他の次元削減の手法としては，主成分分析 (Principal Component Analysis; PCA) や非負値行列因子分解 (Non-negative Matrix Factorization; NMF) (Lee & Seung, 1999)，潜在的ディリクレ配分法 (Latent Dirichlet Allocation; LDA) (Blei et al., 2003)，最小二乗法による回帰 (GloVe)(Pennington et al., 2014) などの潜在タイプの単語分散表表現が提案されている．

潜在カウントタイプの次元数は通常数百次元であり，単語分散表現は要素の多くが非ゼロとなる密なベクトルとなる．この要素は潜在的な特徴を抽出したもので言語的な解釈が難しいことと，大量コーパスから得た単語共起行列の場合次元削減の計算コストが大きいという課題がある．二つのカウントタイプに共通の課題として，新たなデータが追加される都度共起行列を一新して計算し直す必要があることが実用上の欠点となる．

なおカウントタイプに準ずるものとして，ランダムに符号化するモデルがある．BEAGLE (Jones & Mewhort, 2007) はカウントタイプのように共起頻度をカウントして得た行列を分解して密行列を得るのではなく，予め各単語にランダムに生成した密なベクトルを割り当てておいて，コーパスにおける単語共起の回数ごとに共起相手となる単語のランダムベクトルを足し上げていくことで分散表現を獲得する．

■予測タイプ（静的ベクトル）　予測タイプの単語分散表現は，カウントタイプと同様コーパスを用いるが，カウントせずにコーパスを学習データとしたニューラルネットや深層学習により獲得される．

予測タイプの先駆的モデルは Elman (1990) の Simple Recurrent Network とされるが，最も知られている単語分散表現は word2vec (Mikolov et al., 2013a, 2013b) である．

コンテクスト単語からターゲット単語を予測する言語モデル（continuous bag-of-words)，あるいはその反対にターゲット単語からコンテクスト単語を予測するためのモデル (skip-gram) を学習した際に，獲得されたモデルの内部パラメータを単語分散表現として用いたものが word2vec である．Word2vec が “king is to queen as man is to woman”(*king:queen::man:woman*) のような類推関係に対して $\boldsymbol{v}_{king} - \boldsymbol{v}_{queen} \approx \boldsymbol{v}_{man} - \boldsymbol{v}_{woman}$ というベクトル演算が成立することを示し驚きを持って受け止められたことは前述の通りである．

予測タイプの利点は，カウントタイプに比較してバッチ処理や並列処理により高速で効率的に獲得できることにある．またデータを逐次追加しながら継続的に学習できる点もカウントタイプより優れている．

予測タイプの分散表現の本質は，自己教師あり学習 (self-supervised learning) により獲得されることであり，学習データに内在する構造から有用な特徴を抽出する特徴学習 (feature learning) あるいは表現学習 (representation learning)(Bengio et al., 2013) である．予測タイプの単語分散表現は高次元の特徴を低次元のベクトル空間へ写像するので単語埋め込み (word embedding) とも呼ばれる (Landauer & Dumais, 1997)．予測タイプの単語分散表現は，大規模言語モデルへの事前学習済みのインプットして用いられる．

予測タイプが伝統的なカウントタイプと本質的に異なるかどうかは解明にほど遠いとされている (Lenci, 2018)．カウントタイプの分散表現も word2vec と同様に類推関係などの言語的な規則性を捕捉するほか，単語の意味に関するタスクでも予測タイプと同等の性能を示していることが指摘されている (Levy et al., 2015)．Sahlgren and Lenci (2016) は，学習データの量が多い場合のみ予測モデルがカウントモデルに優位となることを報告している．Levy and Goldberg (2014a) は，word2vec の学習モデル skip-gram negative sampling がカウントタイプの一つである単語共起頻度に PPMI によるウェイトづけをしたものと本質的に同等であると主張している．

予測タイプとカウントタイプが本質的に異ならないとする Lenci (2018) の主張は，本研究の観点から重要である．いずれもデータとして用いているのがコーパスであることは共通であることから，予測タイプとカウントタイプはアルゴリズムが異なるにせよ，言語分布に内在する本質的に同じ構造を抽出する可能性がある．仮にそうであるならば，問うべき問題は言語分布に内在する構造とは何かということである．対象としている確率分布により推定量も異なることを踏まえると，予測タイプにせよカウントタイプにせよ，言語分布の構造がわかって初めて適切な特徴抽出が可能になる．

■予測タイプ（動的ベクトル）　予測タイプのうち動的ベクトルは，一般的には文脈付き単語埋め込み (contextual embeddings) と呼ばれるもので，大規模言語モデルである BERT (Devlin et al., 2019) や GPT-3 (Brown et al., 2020) など Transformer (Vaswani et al., 2017) と呼ばれる深層学習モデルの学習過程で生成される．学習の過程で入力した静的ベクトルが文の統語関係や意味関係に沿って相互に結合され，句や文を意味すると思われるベクトルへと変化していくことから動的あるいは文脈を獲得していくという意味で文脈付きと呼ばれる．

BERT はコーパスをデータとして二つの予測を行うよう学習する．具体的にはコーパ

ス中のマスキングをした単語の予測と，二つの文がコーパス中で連続している否かの予測を行うよう膨大な内部パラメータを学習する．文の単語列に対応する事前学習済み静的ベクトルと各単語の位置を入力とすると，注意機構[*2]と呼ばれるメカニズムにより主語と述語や指示代名詞と先行詞の関係など予測に重要と考えられる単語同士を結合していく．BERT の内部には注意機構を含んだニューラルネットワークが 12 層（ベースモデルの場合）があり，それぞれの層で獲得されるパラメータが動的ベクトルとなる．各層にある動的ベクトルが文の統語的・意味的構造の情報を含んでいることは間接的な手法で確認されているが (Clark et al., 2019)，それぞれの動的ベクトルが何を表現しているかについては未解明である．そのため動的ベクトルは BERT のタスク実行に用いられるのみであり，動的ベクトルを BERT から抽出して四項類推課題のような他のタスクに直接利用することは現時点では見当たらない．

前章で触れたように，BERT は質問応答のタスクで高いパフォーマンスを上げているが，タスクを実施する際にその内部では BERT の事前学習で獲得された動的ベクトル間のさらなる結合・変換が生じている．従って，動的ベクトルが表現している対象やタスク実行の際の動的ベクトルの変化を解明することで質問応答タスクに対応する推論のプロセスを特定することができるのではないかと考えられる．静的ベクトルが言語分布における単語間の共起関係を表現しているとすれば，動的ベクトルは階層的な言語の構造にある句や文間の分布関係を何らか抽出している，いわば句ベクトル・文ベクトルと考えられる．これらは現在進行中の研究課題であり本研究のスコープを超えるものであるが，著者の将来的な研究課題としたい．

2.3　単語分散表現の性質

単語分散表現は単語の意味を表現していると言われるが，意味とは何かを直接的に問うことは言語哲学的な意味論に踏み込むことになる．Harris の構造言語学の立場に立てば，正しい問いの立て方は，意味的な関係にある単語は，分布上どのような関係にあるのか，あるいは単語分散表現であるベクトル間にはどのような数学的な関係が成立するのかである．単語分散表現を直接的な研究対象とした先行研究はそれほど多くなく，また，網羅的なサーベイもまだできていないが本節では単語分散表現がもつ言語的な性質を扱った先行研究をアドホックに概括していく．

[*2] 注意機構は系列信号を入力とする予測を行う際に，動的に変更される重要度に応じて入力値を重みづけするメカニズムである (Vaswani et al., 2017).

■類推関係　Word2vec に見られるように単語分散表現が示す代表的な性質は，四項類推関係にある単語ベクトルが平行四辺形をなすことである．平行四辺形モデルは，認知科学の領域で類推（analogical reasoning) を対称的な微細特徴を持つ概念間の関係性を表すモデルとして Rumelhart and Abrahamson (1973) により提案された (図 2.1)．

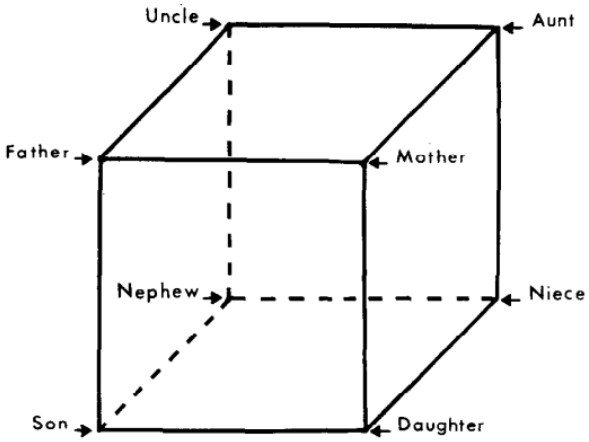

FIG. 2. Three-dimensional representation of relations among eight kinship terms.

図 2.1　Rumelhart&Abrahamson(1973) に提示された類推モデル

コーパスにより獲得された単語分散表現を用いて類推問題に回答することができることは，word2vec 以前に既に Turney and Littman (2005) により示されている．単語分散表現が示す類推関係は言語のどのような分布構造か，またそれが何に由来するのかは第 3 章で検討する．

■類義語・反意語　単語分散表現が，コサイン類似度により 2 つの単語間の類似度を表すことは前述の通りである．コンテクストにある共起単語の分布が類似していることは，2 つの単語が同じように振る舞うことを示す．そして，分布仮説によれば同じように振る舞う単語は類似する意味を持つ．実際に単語分散表現に対してコサイン類似度を計算して上位の単語を見ると，類義語を含めターゲット単語との関連性が深いものばかりが現れる．しかしながらコサイン類似度は，意味的に類似する単語（類義語）だけではなく，統語的に結びつきやすい関連語や反意語に対しても高い類似度を与える (Lenci, 2018)．例えば，hot-cold のような 2 つの単語は類似する文脈で相互に代替可能であるため反対の意味を持っていても類似度が高くなる (Landauer et al., 1998; Landauer, 2002)．類義語と反意語やその他の単語を区別する試みがなされてきているが, コーパスのみから獲得

された単語分散表現では困難な問題とされており (Lin et al., 2003), 統語関係 (Poon & Domingos, 2009) やシソーラスのカテゴリー分類 (Mohammad et al., 2008; Yih et al., 2012) などをコーパスに追加することで類義語と反意語の区別が可能になることが報告されている．

■上位語・下位語・同位語　このほか, 分布上類似しており相互に関連しているが類義語ではないような関係として，上位語・下位語・同位語の区別がある．例えば，*animal-dog-cat* の前二語は上位語・下位語の関係にあり，後二語は同じ上位語を持つ同位語の関係になる．Weeds et al. (2014) は，単語分散表現により上位語・下位語・同位語を判別できることを示している．具体的には，上位・下位などの関係にある単語ペアを教師データとして，サポートベクターマシン (SVM) による教師あり学習をおこなっている．その際，教師データとするペアの単語ベクトルの和・差・アダマール積（要素ごとの積）・結合などの特徴量を追加して，それぞれの識別力を調べた結果ベクトルの結合よりも差分ベクトルが上位語の判定に貢献する一方，同位語にはベクトル和がより高い識別力を示したことを報告している．

■意味カテゴリー　単語分散表現は意味カテゴリーや意味カテゴリー間の階層構造を表している．Rohde et al. (2006) は単語分散表現のベクトル間の距離をメトリックとして階層的クラスタリングを行うと意味カテゴリーが階層構造として出現することを示した図 2.2 の最上部には wrist や ankle など体の部位を表す単語，中程には動物を表す単語，下部には地理的な固有名詞が大きなクラスタを作り，その中で都市名と国名がサブクラスタを作る．

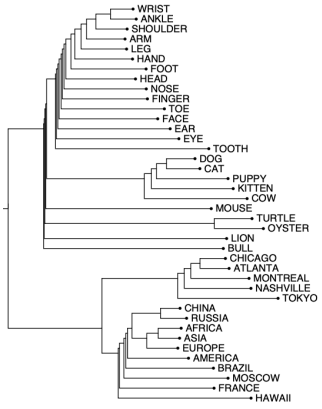

図 2.2　単語分散表現がなす階層的クラスタ; Rohde et al.(2006)

また，Hashimoto et al. (2016) は，同一カテゴリーにある同位語の単語分散表現がクラスタを作るとき，それらの上位語がクラスタのセントロイドとなることを示している．

■全体・部分関係（Meronymy） 上位・下位関係に似た概念として全体・部分の関係がある．例えば，face-nose, car-wheel のような単語の対である．Czachor et al. (2018) は，Fu et al. (2014) の上位語・下位語の分類手法を拡張して，全体・部分関係の単語対を単語分散表現により分類できることを示した．Fu et al. (2014) の手法は，下位語にあたる単語の分散表現を射影すると上位語となるような射影行列が存在すると仮定して，正解データの対からそのような射影行列を線形回帰により学習するものである．そのような射影行列はいくつかの単語ペアのクラスに対応して個別に定められることが報告されている．

■シリーズ 認知科学において提案された言語に関する推論のテストの一つに series compeletion がある (Sternberg & Gardner, 1983). 例えば *penny:nickel:dime:quarter* のような４つ組であり，そのうち３つの単語を与えて残りの１語を回答させるというタスクである．Hashimoto et al. (2016) は，単語分散表現を用いて series completion を回答できることを示している．その際，シリーズの関係にある４つの単語のベクトル間には $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \boldsymbol{v}_2 - \boldsymbol{v}_3 = \boldsymbol{v}_3 - \boldsymbol{v}_4$ の関係があるとの仮定に基づいて 4 つ目の単語を予測している．

■意味の加法構成性 Mikolov et al. (2013b) は単語間の四項類推関係 (*king-queen=man-woman*) の他に単語と単語が組み合わさった語句ベクトルも導入した上で，ベクトル間の加算演算が意味を持つとして例として *Russia + river = Volga river* や *French + actress = Charlotte Ginsberg* などの演算が成り立つことを示している．両式の左辺は二つの単語のそれぞれのベクトルの和を，右辺は二語からなる句に対応する一つの単語ベクトルを示す．そして，このような意味上の演算が成り立つ理由として，単語分散表現がある種の線形構造を備え多くの言語学的規則性やパタンを符号化していると指摘している．

■多義語における線形性 単語分散表現が持つ加法構成性の例として Arora et al. (2018) は，多義語の単語ベクトルが多義語の個々の意味に対応する単語のベクトルの線形結合（線形重ね合わせ）であることを示している．例えば，*tie* の単語ベクトルは，その複数の意味（ネクタイの意，引き分けの意，結びつける意）に対応するそれぞれの単語ベクトルの線形結合として分解される．Arora et al. (2018) は word2vec などのニューラルモデルは出力層に非線形なシグモイド関数を用いているにもかかわらず，その単語埋め込みに線形構造が出現することを Linear Assertion と呼んでいる．

2.4　まとめと考察

本章では，単語分散表現の概要を述べた上で先行研究に沿ってその類型と性質を概観した．単語分散表現を通じて言語の分布構造を解明するとの本研究の観点から重要と考える三点を以下に示す．

第一に，単語分散表現の本質は共起行列であるということである．カウントタイプと予測タイプのいずれもコーパス中のコンテクスト内の単語共起をデータとしていること，予測タイプの word2vec はカウントタイプの PPMI に理論的に帰着できること，評価タスクでの精度に大きな相違がないことから，著者は現時点ではカウントタイプと予測タイプは本質的に異ならないとの Lenci の立場に同意する．もちろん両者には獲得のための計算コストや生成されるベクトルの疎密・次元数の違いはあるが，いずれも単語の共起分布から言語の分布構造と特徴を抽出する機能を果たしている点で本質的に同じであると考える．従って，本研究では単語分散表現より直接的に言語生成系との対応を追跡できる単語共起行列の内部構造を対象として，次章以降，言語の分布構造の解明にアプローチする．なお，単語共起行列の利用に関しては，二語間の共起頻度 (skip bi-gram) が最適な統計量であるとの暗黙の前提があるが，本来は言語の確率分布に応じた統計量を選択すべきである．

第二に，単語分散表現には類義関係に限らない豊かな言語に関する情報が符号化されていることである．コンテクスト単語の分布が類似度だけではなく，二つのベクトルの差ベクトルや射影したベクトルを取ることで類義関係以外の関係性を抽出できることは「単語の意味は周囲の単語によって形成される」という分布仮説以上の構造が単語分散表現に含まれていることを示唆する．一方で，分散表現自体の解釈は難しい，或いはそもそも意味を持たないとされ，先行研究においても教師あり学習による識別に依拠しており，潜在構造へ間接的に接近するにとどまる．Saussure が言語は相対的であり相互依存的であると考えていたことを思いおこすと, 単語分散表現は絶対的な意味を表象しているのではなく，それらの相対的な関係こそが意味関係を示唆する．従って，本研究では単語分散表現の相対的関係を定式化することを目指し，その際，数学的に記述するため幾何的な関係として捉えることとする．

第三に，分布構造は線形的な性質を持っている可能性があることである．類推関係が単

語分散表現間のベクトル演算に対応したり，句の分散表現が句を構成する単語の分散表現の加法で構成されることや多義語の分散表現が個々の意味に対応する分散表現の線形性結合であるとの現象の本質的な意味は明らかではない．おそらく言語の分布構造自体に線形的な性質が内在しており，それは言語生成系に由来する代数的な構造を反映していると予想される．従って，言語の生成段階まで遡り，代数構造がどのように共起行列や分散表現へ反映されていくか数学的な写像・変換として定式化することで解明を試みる．

第 3 章

単語共起行列の数理的分析

3.1　導入

3.1.1　研究目的とアプローチ

第 1 章では本修士研究の目的が，単語分散表現に現れる言語の分布構造と性質の解明，及びそのための数学的定式化と実証であることを述べた．第 2 章では先行研究からの示唆として，(1) 分散表現の本質は単語共起行列にあること，(2) 単語分散表現間の相互関係を捉えるべきこと，(3) 言語には生成系に由来する線形構造が備わっている可能性があることを述べた．

これを踏まえて，本章では単語共起行列を分析の中心的対象として，生成系から分散表現までの変換を数学的に定式化した上で，その数理的構造と性質を明らかにすることを目的とする．

多くの先行研究で実証されている現象として，単語間の類推関係が単語分散表現の間の平行四辺形関係として現れることに着目して，これを数学的に定式化する．言語の生成系までを定式化の対象とするために構成論的アプローチをとり，人工的なコーパスを用いて類推関係を持つ単語分散表現を再現した上で，その数学的な構造と性質を示す (第 2 節)．さらに平行四辺形をなす類推関係を一般化して，単語分散表現間の相互関係を幾何的配置として捉え，平行四辺形以外の図形の類型に拡張する（第 3 節)．

3.1.2　構成論的アプローチについて

構成論的アプローチは，対象を作って動かすことによって理解しようとする科学の一つの方法論である (橋本敬 et al., 2010)．対象を部分に分割して分析する伝統的な科学の手

法では解明が難しい複雑系のようなシステムに対して, 構成論的アプローチは，逆方向の解明，すなわち部分から全体を構築してその挙動を解明することで対象を理解しようとする．特に，その手法は単一のプロセスに焦点を絞り部分と全体の相互作用を理解しようとする点で特徴づけられる (Gerstenmaier & Mandl, 2001).

自然言語には統計物理学的な経験則が成り立つことが知られており，言語の系は複雑系として捉えられる (田中久美子, 2021)．特に，階層的な文法構造から，言語は冪乗則などの自然現象や社会現象が示すような構造的複雑さを持つことが知られている (Kobayashi & Tanaka-Ishii, 2018)．本研究の対象とする単語分散表現に現れる性質と，その背後にある言語の数理構造を解明する手法として構成論的アプローチは有効であると考えられる．

3.2　類推関係のモデル化

3.2.1　方法：平行六面体のモデル概要

具体的な単語共起のモデルとして，3 組の平行関係を持つ六面体（平行六面体）を再現することを想定して，24 文から構成される小規模の人工コーパス（以下，トイコーパスという.）を設計する．Mikolov et al. (2013a) をはじめとする従来研究では，2 組の単語ペア間における類推関係が単語ベクトル空間における 4 つの単語ベクトルがなす平行四辺形関係（2 組の平行関係）として表現される現象が示されてきているが，本研究では文構造との関連性を調べるために，3 組の平行関係（平行六面体関係）へ拡張したモデルを構築して，その挙動を調べるという構成論的アプローチをとる．

トイコーパスは，24 の英文，17 語の語彙から構成されている (図 3.1)．各文は，主語-

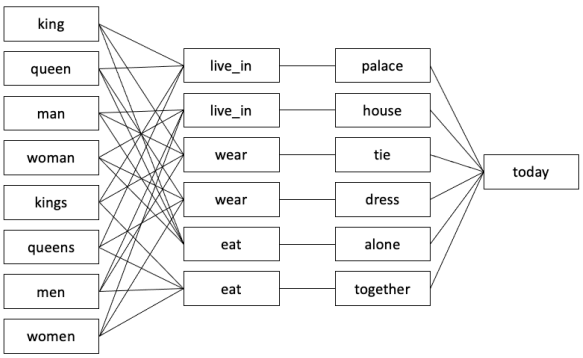

図 3.1　トイコーパスの 24 文の文構造

動詞-目的語-副詞の 4 語からなる同一文型を有しており，品詞ごとに分類された語彙中の

単語を組み合わせて作られている (例:“king live_in palace”)．冠詞や動詞活用を省いた人工的な文法に従う文である．具体的な語彙としては，主語となる名詞が 8 単語（king, queen, man, woman, kings, queens, men, women)，述語となる動詞（句動詞）が 3 単語 (live_in, wear, eat)，動詞句を構成する名詞または副詞が 6 単語 (palace, house, tie, dress, alone, together) ，文全体を修飾する副詞が 1 単語 (today) の計 18 単語である．その際，動詞と目的語，主語と目的語・副詞の間には意味的な対応関係があると想定する．例えば，king は palace と組み合わせることができるが，house とは組み合わないなどである．

文型中のそれぞれの位置にある品詞に対応した単語を自由に組み合わせれば，$8 \times 3 \times 6 \times 1 = 144$ 文となるが，個々の単語間に意味的制約を課すことにより，その部分集合 (24 文）のみがトイコーパスに含まれることになる．これらの意味的対応関係が自由な文の生成を制約する結果として，単語間の共起頻度に一定の規則性が出現することになる．意味的制約の設計にあたっては，主語となる名詞 8 単語間に 3 軸の意味的対称性が生じるよう 3 つの動詞を用い，それぞれの動詞が目的語・副詞としてとれる単語 2 語のうち，主語の属性に応じてどちらかひとつの単語しか対応しないよう制約をした（図 3.1)．すなわち，3 つの動詞に対して 2 つの選択肢があるので，主語となる各単語は 3 ビットで表現されていることになる．

このトイコーパスに対し，同一文内に共起する単語ペアをカウントすることにより共起頻度行列（18 行 18 列）を作成する．各文を構成する単語は，主語・動詞・目的語・副詞の 4 語であるので，共起ペアは一文当たり ${}_4C_2 = 6$ 組生成される．ターゲット単語の前後最大 3 語との共起をカウントするのでウィンドウサイズは 3 語（ただし，文を跨がない）となる．このトイコーパスの各文が 1 回ずつ出現したとすると，共起行列は図 3.2 の通りである．

カウントベースの単語ベクトルのうちもっとも基本的なものが共起頻度行列の行ベクトルを用いるものである（単語ベクトルの生成手法については第 2 章を参照)．ここで，全語彙のうち主語となる名詞単語 8 単語を主語単語とよび，以降の分析の対象とする．主語単語に対応する単語ベクトル（共起行列のうち主語単語に対応する上から 8 行分）を集め

	king	queen	man	woman	kings	queens	men	women	wear	live	eat	tie	dress	palace	house	alone	together
king	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	0
queen	0	0	0	0	0	0	0	0	1	1	1	0	1	1	0	1	0
man	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	0
woman	0	0	0	0	0	0	0	0	1	1	1	0	1	0	1	1	0
kings	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1
queens	0	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	1
men	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	0	1
women	0	0	0	0	0	0	0	0	1	1	1	0	1	0	1	0	1
wear	1	1	1	1	1	1	1	1	0	0	0	4	4	0	0	0	0
live	1	1	1	1	1	1	1	1	0	0	0	0	0	4	4	0	0
eat	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	4	4
tie	1	0	1	0	1	0	1	0	4	0	0	0	0	0	0	0	0
dress	0	1	0	1	0	1	0	1	4	0	0	0	0	0	0	0	0
palace	1	1	0	0	1	1	0	0	0	4	0	0	0	0	0	0	0
house	0	0	1	1	0	0	1	1	0	4	0	0	0	0	0	0	0
alone	1	1	1	1	0	0	0	0	0	0	4	0	0	0	0	0	0
together	0	0	0	0	1	1	1	1	0	0	4	0	0	0	0	0	0

図 3.2　単語共起行列

たものを行列 C とすると次の通りである．

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

全語彙に関する共起行列では，主語単語の単語ベクトルの最初の 8 列の成分は全て 0 であるので，C では省略している．

前節に示した単語共起行列は，トイコーパス中の 24 文が 1 回ずつ含まれていると暗黙に仮定していた．すなわち，各文が一様分布の確率分布（各文の出現確率がすべて $\frac{1}{24}$）から出現するとして共起頻度をカウントしたものであった．次節で見るように，この時 8 つの単語ベクトルは平行六面体を構成するのであるが，平行六面体が出現する条件を検討するために，文の生起確率をパラメータ化して非一様な文の出現確率を考える．24 文のそれぞれの出現確率を p_i $(i = 1, \ldots, 24)$ とする．確率の総和に関して，$\sum_{i=1}^{24} p_i = 1$ が成り立つ．この時，主語単語に関する共起行列 C は次の通りとなり，単語ベクトルは C の各行ベクトルとなる．

$$C = \begin{bmatrix} p_1 & p_9 & p_{17} & p_1 & 0 & p_9 & 0 & p_{17} & 0 \\ p_5 & p_{10} & p_{18} & 0 & p_5 & p_{10} & 0 & p_{18} & 0 \\ p_2 & p_{13} & p_{19} & p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\ p_6 & p_{14} & p_{20} & 0 & p_6 & 0 & p_{14} & p_{20} & 0 \\ p_3 & p_{11} & p_{21} & p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\ p_7 & p_{12} & p_{22} & 0 & p_7 & p_{12} & 0 & 0 & p_{22} \\ p_4 & p_{15} & p_{23} & p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\ p_8 & p_{16} & p_{24} & 0 & p_8 & 0 & p_{16} & 0 & p_{24} \end{bmatrix} \tag{3.1}$$

各行は主語単語に対応しており，1 行目は king, 2 行目は queen に対応するなどである．一方，列は主語単語と共起する単語を表し，1〜3 列目は 3 つの動詞，4〜9 列目は 6 つの目的語に対応する．

トイコーパス中にある 24 文は，主語単語と動詞句（動詞＋目的語）から構成されており，このトイコーパスにおいては，それぞれの主語単語がある動詞句とともに文をなす場合には，その一文でしか共起しないので，その文の生起確率が当該主語単語と動詞句を構成する動詞，目的語との共起確率になる．例えば，king wear tie という文の生起確率を p_1 とした時，king と wear の共起，king と tie の共起はこの文中のみで起きるのでこれらの共起確率はいずれも p_1 となり，行列 C の 1 行目 1 列と 1 行 4 列の成分には p_1 が入る．一様分布の場合は $p_i = \frac{1}{24}$ $(i = 1, \ldots, 24)$ である．

3.2.2 結果１：モデル挙動の分析（一様分布の場合）

■アナロジーテスト (類推課題) トイコーパスより生成した単語ベクトルが類推関係を示すかどうか，アナロジーテストを用いて検証する．アナロジーテストとは，類推関係にある 4 つの単語の単語ベクトルが平行四辺形にあることを利用して，そのうち 3 つの単語から残りの 1 単語を予測する課題である．例えば，$king : queen :: man : woman$ という類推関係に対して，vector offset あるいは 3CosAdd と呼ばれる次のベクトル演算を行う (Levy et al., 2015).

$$\max_{x \in V} \cos\left(v_{king} - v_{queen} + v_{woman}, v_x\right)$$

ここで求めるコサイン類似度を最大化する語彙中の単語として man を選べば正解である．

■一様分布：アナロジーテストの結果 まず一様分布のトイコーパスから生成された単語ベクトルに関して，一例として，$king : queen :: x : woman$ の類推関係にある単語 x を予測するために，$v_{king} - v_{queen} + v_{woman}$ と全ての語彙単語とのコサイン類似度を算出した結果（上位五単語）を表 3.1 に示す.,

表 3.1　3CosAdd によるコサイン類似度上位 5 単語

単語	コサイン類似度
man	1.000
king	0.833
woman	0.833
men	0.833
queen	0.677

最もコサイン類似度の高い単語は man であり，問題にあった x を正しく予測できたことになる．さらにコサイン類似度は 1.000 であった．

8 単語は 12 パタンの類推関係を構成しているが，その全てにおいて四項類推課題を 100% 正解し，またその際の正解に関するコサイン類似度は 1.000 となっている．これは 8 単語が平行六面体をなしていることを意味するのであろうか．

■次元削減による図示化　8 つの単語ベクトルは高次元 (17 次元) ベクトル空間内に住んでいるので，3 次元空間である平行六面体であるかどのように確かめれば良いのであろうか．まず，最初に考えられるのが低次元に次元削減をして図示化することである．任意の 3 つの独立な基底ベクトルを座標軸として座標を求めれば良いので，仮に一様分布における行列 C の上から 3 行をそれぞれ x 軸，y 軸，z 軸の座標軸とする．すると，8 つの単語ベクトルの 3 次元空間における座標を 17 行 3 列の行列 X とすると，$CB = X$ により求められる．ただし，

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

である．

図 3.3 の通り，8 つの単語が，*male-female, royal-common, single-plural* の 3 つの軸に対応した 6 つの平行四辺形を面とする平行六面体をなしていることが確認できる．

一様分布の時の共起行列 C の階数は 4 であるが，3 つの線形独立なベクトル基底 $b_1, b_2, b_3 \in \mathbb{R}^9$ と平行移動 $b_0 \in \mathbb{R}^9$ を使って次のように表現することができる．

$$C = A \begin{bmatrix} b_1^t \\ b_2^t \\ b_3^t \end{bmatrix} + \mathbf{1}_8 b_0^t$$

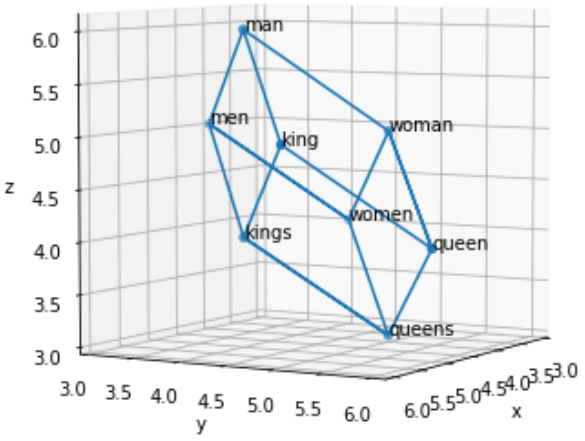

図 3.3　主語 8 単語ベクトルの配置（一様なトイコーパスの場合）

つまり，行列 C は，各行ベクトルが 3 次元のアフィン空間上にあり，$\mathbf{A} \in \mathbb{R}^{3\times 10}$ は，アフィン基底 $\mathbf{B} = (b_0, b_1, b_2, b_3)$ に対応して一意に定まる行列であるような性質にあるのである．この性質より行列 C の行ベクトルを 3 次元空間へ射影すると平行六面体が現れたのであり, 共起頻度行列に 8 単語の平行関係が埋め込まれていることが確認できたのである．

3.2.3　結果２：モデル挙動の分析（非一様分布の場合）

次に非一様な文の出現確率のもとでの主語単語ベクトルを調べる．（n=24 の多項分布により）24 文ごとにランダムに確率を付与して生成されたトイコーパスを用いて，あらためて共起頻度をカウントして共起行列の行ベクトルを取ることで単語ベクトルを生成する．

■アナロジーテスト結果　新たに生成した単語ベクトルを用いて，一様分布の場合と同様に 3CosAdd の計算を行うため $v_{king} - v_{queen} + v_{woman}$ とコサイン類似度が最大となる単語ベクトルを持つ単語を求める．コサイン類似度上位語五単語は表 3.2 の通りであり，正解である man を選ぶことができていない．

表 3.2　3CosAdd によるコサイン類似度上位 5 単語

単語	コサイン類似度
woman	0.794
man	0.614
king	0.426
palace	0.382
house	0.381

■次元削減による図示化　また，非一様分布より生成された 8 つの主語名詞に対応する単語ベクトルの幾何的配置を調べるために，一様分布と同様に任意の独立座標軸として行列 C の最初の 3 行を選び，各ベクトルの座標を求め 3 次元空間に図示化する．図 3.4 に示される通り，歪んだ六面体となり平行六面体を構成していないことは明らかである．

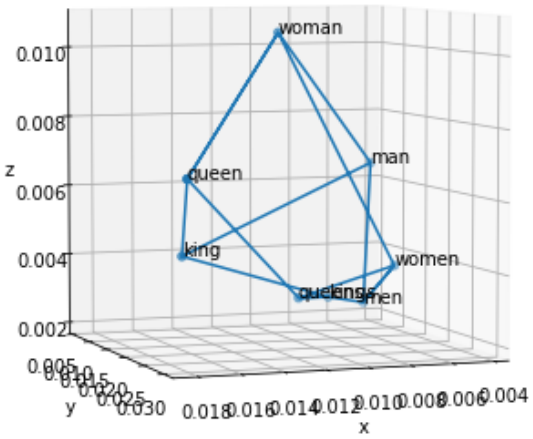

図 3.4　主語 8 単語ベクトルの配置（非一様なトイコーパスの場合）

一様なトイコーパスの場合と比較すると，コーパス中の文の統語構造と各単語間の意味的制約は共通であるので生成される文自体は同一である．それぞれの文中に生起する単語の組み合わせも変わらないので，例えば king は tie, palace, alone と共起するという関係性は一様分布のコーパスでも非一様分布のコーパスでも変らない．これは，単語共起行列の成分のうち, 構文上の制約から一様分布・非一様分布のいずれの場合でもゼロをとる位置があることを意味する．

一方，一様分布と非一様分布で異なるのは共起行列の非ゼロになりうる成分の値ということになる．平行六面体が現れるためには，各文における単語の共起関係のみならず，各文が出現する確率分布上になんらかの対称性が備わっていることが必要であると考えられる．

それでは，単語ベクトルが平行六面体を構成するためには，文の生起確率 p_i $(i = 1, \ldots, 24)$ に関してどのような条件が成り立っていれば良いのであろうか．次に平行六面体をなすコーパスの文生成確率に関する必要十分条件を調べる．

3.2.4 平行六面体の必要十分条件の導出

生成された単語ベクトルが平行六面体をなすための単語共起行列並びにコーパスにおける分布に関する必要十分条件を求める．トイコーパスの 24 文は意図的に単語間に対称性を持たせるよう設計されていたが，その上で各文が同一確率で生起しているとき単語ベクトルは平行六面体をなしており，(1)24 文の単語の組み合わせであることと (2) 文生起確率が一様分布であることが平行六面体の十分条件の一つであった．

この条件が実際のコーパスで成立する可能性は極めて低いと考えられるが，理想化された条件下で明確な条件を求めることで、得られる洞察もあると考えられる．実コーパスから生成された word2vec のような単語ベクトルが四項類推に正答している時にはどのような条件が満たされているのであろうか．それを明らかにするために，平行六面体をなすための必要条件を次のステップにより導出する．

(ステップ 1) 高次元空間における平行六面体を定義する
(ステップ 2) 平行六面体の定義より単語共起行列の必要条件を解く
(ステップ 3) 必要十分条件を満たす解空間を導出する（文の生起確率に関する制約条件となる）
(ステップ 4) 文を構成する単語の組み合わせに関する前提を緩和して一般化した必要十分条件を求める

高次元空間における平行六面体とは，3 次元空間におけるそれと同様に，8 つの頂点からなる幾何的配置であり，辺を 2 つのベクトルが作る差分ベクトルとみて，対辺となる差分ベクトルが等しいものと定義できる．その上で，8 つのベクトルを頂点と見たときに，頂点や辺，面が重複するような特殊なケースを除いて，12 の異なる辺と 6 つの面を持つものと定義する．それぞれの面は平行四辺形である．

また，８つのベクトルを平行六面体のどの頂点に対応させるかについては複数の組み合

わせが存在するので注意が必要であり後述するが，本節では図 3.5 のように 8 つのベクトル $k, q, m, w, ks, qs, ms, ws \in \mathbb{R}^9$ が配置されているとした時に，ここに示された六面体が平行六面体であることは，図中の 12 の辺が作る差分ベクトルがそれぞれ式 (3.2)-(3.10) のように同じであると幾何的に定義できる．さらにベクトルの等式を移行して左辺にまとめる．$\mathbf{0}_9$ は 9 次元ベクトル空間におけるゼロベクトルである．

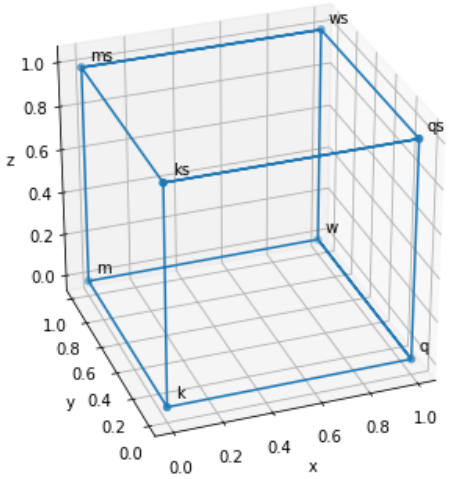

図 3.5　8 頂点の選び方

$$k - q = m - w \Leftrightarrow k - q - m + w = \mathbf{0}_9 \tag{3.2}$$
$$k - q = ks - qs \Leftrightarrow k - q - ks + qs = \mathbf{0}_9 \tag{3.3}$$
$$k - q = ms - ws \Leftrightarrow k - q - ms + ws = \mathbf{0}_9 \tag{3.4}$$
$$k - m = q - w \Leftrightarrow k - m - q + w = \mathbf{0}_9 \tag{3.5}$$
$$k - m = ks - ms \Leftrightarrow k - m - ks + ms = \mathbf{0}_9 \tag{3.6}$$
$$k - m = qs - ws \Leftrightarrow k - m - qs + ws = \mathbf{0}_9 \tag{3.7}$$
$$k - ks = q - qs \Leftrightarrow k - ks - q + qs = \mathbf{0}_9 \tag{3.8}$$
$$k - ks = m - ms \Leftrightarrow k - ks - m + ms = \mathbf{0}_9 \tag{3.9}$$
$$k - ks = w - ws \Leftrightarrow k - ks - w + ws = \mathbf{0}_9 \tag{3.10}$$

式（3.2）と式（3.5），式（3.3）と式（3.8），式（3.6）と式（3.9）は同じ等式であるので，9 つの制約条件は重複を除いた 6 つの等式と同値である．

これらの等式はそれぞれ 9 次元ベクトルであるので行列として次の通り一つの等式で表

現できる．

$$\begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k^t \\ q^t \\ m^t \\ w^t \\ ks^t \\ qs^t \\ ms^t \\ ws^t \end{bmatrix} = \mathbf{0}_{6,9} \tag{3.11}$$

右辺にある $\mathbf{0}_{6,9}$ は 6 行 9 列のゼロ行列である．左辺一つの目の行列を次のように R とし，２つ目の行列は主語単語 8 語の共起分布行列 C であることに留意すると

$$R = \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} k^t \\ q^t \\ m^t \\ w^t \\ ks^t \\ qs^t \\ ms^t \\ ws^t \end{bmatrix}$$

とすれば，8 語に対応するベクトルが平行六面体の頂点に位置する条件は

$$RC = \mathbf{0}_{6,9} \tag{3.12}$$

と同値である．

平行六面体となる共起行列 C の条件を求めるため，式 (3.1) を式 (3.12) に代入すると次のとおりである．まず，R と C の 1 列目と 4,5 列目の積から導出されるを成分を以下

に示す（重複するものは除く）

$$
\begin{aligned}
p_1 - p_5 - p_2 + p_6 &= 0 \\
p_1 - p_5 - p_3 + p_7 &= 0 \\
p_1 - p_5 - p_4 + p_8 &= 0 \\
p_1 - p_2 - p_3 + p_4 &= 0 \\
p_1 - p_2 - p_7 + p_8 &= 0 \\
p_1 - p_6 - p_3 + p_8 &= 0 \\
p_1 - p_2 &= 0 \\
p_1 - p_3 &= 0 \\
p_1 - p_4 &= 0 \\
-p_5 + p_6 &= 0 \\
-p_5 + p_7 &= 0 \\
-p_5 + p_8 &= 0 \\
-p_7 + p_8 &= 0 \\
-p_6 + p_8 &= 0
\end{aligned}
$$

これらの式を整理すると，

$$p_1 = p_2 = p_3 = p_4 \tag{3.13}$$

$$p_5 = p_6 = p_7 = p_8 \tag{3.14}$$

である．同様に，R と C のそれ以外の列の積から導出される成分の等式を整理すると，以下の条件が追加的に得られる．

$$p_9 = p_{10} = p_{11} = p_{12} \tag{3.15}$$

$$p_{13} = p_{14} = p_{15} = p_{16} \tag{3.16}$$

$$p_{17} = p_{18} = p_{19} = p_{20} \tag{3.17}$$

$$p_{21} = p_{22} = p_{23} = p_{24} \tag{3.18}$$

24 文の生起確率の間に 6 つの等式 (3.13)-(3.18) が成り立つことが平行六面体の必要十分条件となる．

ここであらためて共起行列 C を見てみるとつぎのように行列分解できることがわかる．

$$
\begin{aligned}
C &= \begin{bmatrix}
p_1 & p_9 & p_{17} & p_1 & 0 & p_9 & 0 & p_{17} & 0 \\
p_5 & p_{10} & p_{18} & 0 & p_5 & p_{10} & 0 & p_{18} & 0 \\
p_2 & p_{13} & p_{19} & p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\
p_6 & p_{14} & p_{20} & 0 & p_6 & 0 & p_{14} & p_{20} & 0 \\
p_3 & p_{11} & p_{21} & p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\
p_7 & p_{12} & p_{22} & 0 & p_7 & p_{12} & 0 & 0 & p_{22} \\
p_4 & p_{15} & p_{23} & p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\
p_8 & p_{16} & p_{24} & 0 & p_8 & 0 & p_{16} & 0 & p_{24}
\end{bmatrix} \\
&= \begin{bmatrix}
p_1 & 0 & p_9 & 0 & p_{17} & 0 \\
0 & p_5 & p_{10} & 0 & p_{18} & 0 \\
p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\
0 & p_6 & 0 & p_{14} & p_{20} & 0 \\
p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\
0 & p_7 & p_{12} & 0 & 0 & p_{22} \\
p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\
0 & p_8 & 0 & p_{16} & 0 & p_{24}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{aligned}
$$

このとき

$$
P = \begin{bmatrix}
p_1 & 0 & p_9 & 0 & p_{17} & 0 \\
0 & p_5 & p_{10} & 0 & p_{18} & 0 \\
p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\
0 & p_6 & 0 & p_{14} & p_{20} & 0 \\
p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\
0 & p_7 & p_{12} & 0 & 0 & p_{22} \\
p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\
0 & p_8 & 0 & p_{16} & 0 & p_{24}
\end{bmatrix}, \quad
U^T = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

とおくと，上記の行列分解は次のように表現できる．

$$
C = PU^T \tag{3.19}
$$

行列 P の 8 行は主要単語 8 語に対応し，6 列は 6 つの動詞句に対応し，その行列は対応する主語と動詞句を組み合わせた文の共起確率を表すのでこれを文出現確率行列と呼ぶ．また，U^T の 6 つの各行は動詞句に対応して，それぞれの動詞句を構成する合計 9 単語の出現有無を符号化したものとなっているので動詞句符号化行列と呼ぶ．（各行ベクトルは，対応する動詞句に含まれる動詞 1 語，目的語 1 語に対応した成分が非ゼロ “two-hot-vector” となっている点に留意する．（このような文と単語の構成に対応したベクトルや行列の見方は，第 5 章でコーパスを行列を用いて数学的に記述していく際に重要となる．）

式 (3.13)-(3.18) として導出した平行六面体の必要十分条件が成り立つ時に，6 つのパラ

メータ d_i $(i = 1, \ldots, 6)$ を次のように定める．

$$
\begin{aligned}
d_1 &:= p_1 = p_2 = p_3 = p_4 \\
d_2 &:= p_5 = p_6 = p_7 = p_8 \\
d_3 &:= p_9 = p_{10} = p_{11} = p_{12} \\
d_4 &:= p_{13} = p_{14} = p_{15} = p_{16} \\
d_5 &:= p_{17} = p_{18} = p_{19} = p_{20} \\
d_6 &:= p_{21} = p_{22} = p_{23} = p_{24}
\end{aligned}
$$

共起頻度行列 P にこれらのパラメータを代入すると，さらに次のように行列分解できることがわかる．

$$
\begin{aligned}
P &= \begin{bmatrix}
d_1 & 0 & d_3 & 0 & d_5 & 0 \\
0 & d_2 & d_3 & 0 & d_5 & 0 \\
d_1 & 0 & 0 & d_4 & d_5 & 0 \\
0 & d_2 & 0 & d_4 & d_5 & 0 \\
d_1 & 0 & d_3 & 0 & 0 & d_6 \\
0 & d_2 & d_3 & 0 & 0 & d_6 \\
d_1 & 0 & 0 & d_4 & 0 & d_6 \\
0 & d_2 & 0 & d_4 & 0 & d_6
\end{bmatrix} \\
&= \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
d_1 & 0 & 0 & 0 & 0 & 0 \\
0 & d_2 & 0 & 0 & 0 & 0 \\
0 & 0 & d_3 & 0 & 0 & 0 \\
0 & 0 & 0 & d_4 & 0 & 0 \\
0 & 0 & 0 & 0 & d_5 & 0 \\
0 & 0 & 0 & 0 & 0 & d_6
\end{bmatrix}
\end{aligned}
$$

ここで

$$
Q = \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}, \quad D = \mathrm{diag}(d_1, \ldots, d_6)
$$

とおくと，この行列分解は

$$
P = QD \tag{3.20}
$$

と表現される．ただし $\mathrm{diag} : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ は，任意の n 次元ベクトル $(d_1, \ldots, d_n) \in \mathbb{R}^n$ に対し，そのベクトルの成分 d_i を対角成分 (i, i) に持つ対角行列を返す関数である．

以上より，式 (3.19) と併せると，平行六面体の必要十分条件が成り立つとき，単語共起行列 C は

$$C = PU^T = QDU^T \tag{3.21}$$

と行列分解される．

行列 Q の行は 8 つの単語に対応し，列は 2 列ずつ 3 組で合計 6 列が 3 軸のバイナリー関係に対応していると解釈でき，それぞれ平行六面体の 8 つの頂点，3 つの軸の構造をなしていると考えることができるので，Q を平行六面体の構造を定める「構造行列」と呼ぶことにする．

また，行列 D は，6 つの動詞句の共起頻度を表して平行六面体の自由なパラメーター (確率であるので総和 1 という制約を考慮すると自由度は 5) を表していると考えられる．トイコーパスの 24 文の生起確率が等しい均一分布の場合は，$d_1 = d_2 = \ldots = d_6$ であり，$P = \frac{1}{24}Q$ の特別なケースであることがわかる．

ここで再び平行六面体の定義として定式化した式 (3.12) に着目して，$RC = \mathbf{0}_{6,9}$ を満たす共起行列 C のある部分空間を導出する．厳密に言えば，C を 8 行 9 列の行列ではなく，$8 \times 9 = 72$ 次元のベクトル空間内の 1 点と見た上で，そのような全ての点のうち平行六面体条件を満たす点がなす部分空間を行列 R のカーネル $\ker R$ を用いて求めるのである．

$$\ker R := \{X \in \mathbb{R}^{8\times 9} | RX = \mathbf{0}_{6,9}\}$$

ここで行列 R はランク 4 であり[*1]，その行空間の基底は 4 つあるので，カーネル R の基底も 4 つあることに注意すると, 以下の行列 B の 4 つの列ベクトル $\mathbf{b}_i \in \mathbb{R}^8$ $(i = 1, \ldots, 4)$ は, それぞれ $R\mathbf{b}_i = \mathbf{0}_6$ を満たし，かつ相互に独立であるので，カーネル R の基底であることがわかる．

$$B = [\mathbf{b_1} \quad \mathbf{b_2} \quad \mathbf{b_3} \quad \mathbf{b_4}] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{3.22}$$

カーネル $\ker R$ にあるベクトルは基底 $\mathbf{b}_i$ $(i = 1, 2, 3, 4)$ の線形結合で表すことができ, 逆にこれらの基底の線形結合で得られるベクトルは平行六面体条件を満たす．

[*1] 行列 R に対してガウスの消去法を適用すると先頭がゼロでない行は 4 つとなる

なお，前節で導出した構造行列 Q は，$\ker R$ の基底を集めた行列 B を用いて，以下のように分解ができる．

$$Q = BA$$

ただし，行列 A は

$$A := \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

である．

以上をまとめると，トイコーパスより生成された単語共起行列 C において，その行ベクトルである 8 つの単語ベクトルが平行六面体をなすための必要十分条件を充足しているとき，C はカーネル R の解空間にあるので，式 3.19 はさらに次のように行列分解できる．

$$C = PU^T = QDU^T = BADU^T \tag{3.23}$$

参考のために，この行列分解に現れる各行列のノーテーションの一覧を示しておく．

- C　単語共起行列
- P　主語・動詞句の共起頻度行列（文の生起確率）
- U^T 動詞句符号化行列
- Q　構造行列（平行六面体の構造を表す）
- D　パラメーター d_i の対角行列
- B　解空間の基底の行列
- A　座標行列

3.2.5　平行六面体の一般条件

ここで，構造行列 Q がカーネル $\ker R$ の基底による特定の線形結合であるということは平行六面体をなす条件を満たす特殊な解であることを考えると，さらに条件を一般化することができることがわかる．前節までは，トイコーパス中に現れる文を 24 文による単語の組み合わせに限定して，平行六面体をなす条件を導出していた．すなわち，コーパス中には 6 つの動詞句のみが現れ，8 つの主語単語のそれぞれは，そのうち 3 つのみと共起するという言語構造上の制約下で文の生起確率に関する条件として導出されたのであった．

トイコーパスに出現する文の制約を外し，単語共起行列 C に関する平行六面体を構成するための必要十分条件をあらためて考える．平行六面体の定義は

$$RC = \mathbf{0}_{6,9}$$

であった．両辺を列ベクトルとして見ると，行列 C の全ての列ベクトル $\mathbf{c}_j$ $(j = 1, \ldots, 9)$ が $\ker R$ にあるということであった．すなわち，$\mathbf{c}_j$ が，カーネルの基底 $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ の線形結合で表現できるということである．

$$\mathbf{c}_j = \sum_{i=1}^{4} w_{i,j}\mathbf{b_i} \quad (j = 1, \ldots, 9)$$

これは，次のように分解できるような行列 $W := (w_{i,j})$ が存在するということと同値である．

$$C = BW \quad (W : 4 \times 9) \tag{3.24}$$

すなわち，単語共起行列 C が基底行列 B とその線形結合のウェイトを表す行列 W に行列分解できることが平行六面体の一般条件である．

3.2.6 考察：平行六面体条件の言語学的含意

8 つの単語ベクトルが平行六面体をなす必要十分条件は，それらをを列方向に並べた単語共起行列 C がカーネル $\ker R$ にあることであったが，そのカーネルの基底，すなわち行列 B の各列ベクトルにはどのような意味があるのだろうか．

ある空間の基底は無数に選ぶことができるが，$Z_2 = \{0, 1\}$ を成分とするベクトルに限定すると，式 (3.22) として選んだベクトル $\mathbf{b}_i$ $(i = 1, 2, 3, 4)$ が $\ker R$ の唯一の基底となる．

一方，単語共起行列 C が $C = BW$ と分解できるということは，その列ベクトル $\mathbf{c}_j$ が基底の線形結合であることを意味する．この列ベクトルは，C の行に対応する主語単語 8 語 (taget word) に対して，ある単語 (context word) が共起する確率を成分として持つベクトルを表すことに留意する．

すなわち j 列に対応する context word と 8 つの単語それぞれとの共起確率 $\mathbf{c}_j$ は次の

ように表される．ただし，行列 $W = (w_{ij})_{4\times 9}$ であるとする．

$$\begin{aligned}\mathbf{c}_j &= \begin{bmatrix}1 & 1 & 1 & 1\\ 1 & 1 & 1 & 0\\ 1 & 1 & 0 & 1\\ 1 & 1 & 0 & 0\\ 1 & 0 & 1 & 1\\ 1 & 0 & 1 & 0\\ 1 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\end{bmatrix} \begin{bmatrix}w_{1j}\\ w_{2j}\\ w_{3j}\\ w_{4j}\end{bmatrix} \\ &= w_{1j}\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} + w_{2j}\begin{bmatrix}1\\1\\1\\1\\0\\0\\0\\0\end{bmatrix} + w_{3j}\begin{bmatrix}1\\1\\0\\0\\1\\1\\0\\0\end{bmatrix} + w_{4j}\begin{bmatrix}1\\0\\1\\0\\1\\0\\1\\0\end{bmatrix}\end{aligned} \tag{3.25}$$

式 (3.25) の各基底ベクトルを線形結合するウェイト係数 $(w_{1j}, w_{2j}, w_{3j}, w_{4j})^t$ は，共起行列 C の j 列に対応する context word が，8 主語単語に対してどのような共起パタンを示しているかを表すと言える．例えば，ある context word が Royal の特性を表象しているのであれば w_{3j} は正となり，逆に Common の特性を持つのであれば負となる．この意味するところは，ある context word が与えられた時に，平行六面体をなす 8 つの単語が共起するかどうかの挙動は次のいずれかであるということである．

(共起パタン 1) 8 単語が context word に対して同じように共起する
(共起パタン 2) 単数単語 (king, queen, man, woman) と複数単語 (kings, queens, men, women) でそれぞれ同じように共起する
(共起パタン 3) Royal を表す単語 (king, queen, kings, queens) と Common を表す単語 (man, woman, men, women) でそれぞれ同じように共起する
(共起パタン 4) Male を表す単語 (king, man, kings, men) と Female を表す単語 (queen, woman, queens, women) でそれぞれ同じように共起する
(共起パタン 5) 上記４つの組み合わせ

これは言い換えれば，潜在的な属性が４種類あり，それぞれの属性が特定の動詞句を惹起する．主語単語となる 8 つの単語はそれらの潜在的な属性を合わせ持つため，結果的に共起関係に規則性 regularity が生じてくると解釈できる．また，主語単語８語の集合は，

次にように３通りの方法で同じ挙動をする部分集合に分割される．

$$
\begin{aligned}
S_{11} &= \{king, queen, man, woman\} \\
S_{12} &= \{kings, queens, men, women\} \\
S_{21} &= \{king, queen, king, queens\} \\
S_{22} &= \{man, woman, men, women\} \\
S_{31} &= \{king, man, kings, men\} \\
S_{32} &= \{queen, woman, queens, women\} \\
S &= S_{11} \sqcup S_{12} \\
S &= S_{21} \sqcup S_{22} \\
S &= S_{31} \sqcup S_{32}
\end{aligned}
$$

共起パタン１，２，３における tatget word と context word の共起関係は，図 3.6 のように二部グラフを用いて表現することができる．二部グラフとは，ノードの集合を２つの集合に分割した，それぞれの部分集合に属するノードがお互いにエッジで結ばれていないグラフのことを言う．この場合，主語となる単語の集合と目的語となる単語の集合に重なりがないために文中での共起関係を二部グラフで記述することが可能である．

図 3.6 中のそれぞれのグラフの左横に，共起パタンに対応する基底ベクトルを示す．グ

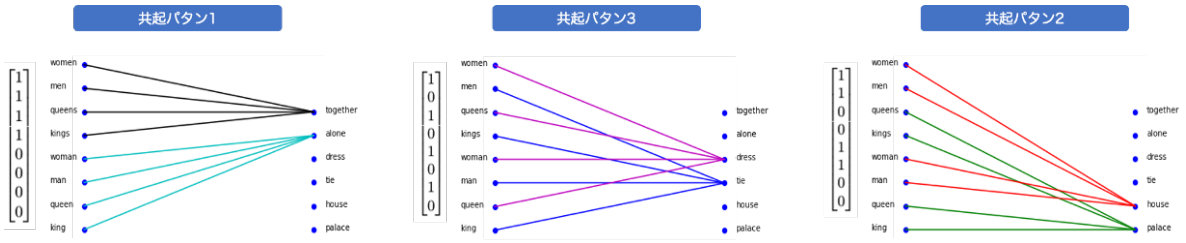

図 3.6　基底に対応する共起パタン

ラフ理論的に言えば，二部グラフで記述される主語単語集合（各グラフの左側のノード集合）に属する単語と目的語単語集合（右側のノード集合）に属する単語との間の隣接関係は図に示すように３つの部分二部グラフに分解できる．

また，各部分グラフにおいては主語単語集合の８つの単語が，目的語単語集合の２つのノードのいずれかと隣接するという関係がある．例えば，共起パタン 1 においては，主語単語は together か alone のいずれか一つと共起している．このとき目的語単語集合の同じノード（例えば alone) を共有する主語単語集合の元 (king, queen, man, woman) は何らかの性質（単数性と名づけることができる）を共有していることを意味する．一方，

他方のノード（例えば together) を共有する主語単語の元 (kings, queens, men, women) は，別の性質（複数性と名づけることができる）を共有しており，これら性質（単数性と複数性）は，相補的であるといえる．すなわち，単数性か複数性かいずれかの性質をもち，両方の性質を持つことはない．さらに，同じ性質を共有する主語単語は，単数単語というグループを構成することになる．

ここまでの考察が示唆するのは，言語における分布構造にある共起関係は二部グラフで表される数学的構造であり，意味関係や意味構造に関連があるということである．

そして言語の意味関係という観点から二部グラフを見ると，Saussure の導入した paradigmatic relation（連合関係）と syntagmatic relation（連辞関係）という概念 (丸山圭三郎, 1983) が想起されるのである．ここではそれらの言語学的定義の検討は避けるが，一般的な説明として前者は入れ替えることが可能な単語同士の関係を指し，後者は同じ文脈の中に共起する単語同士の関係を指すと言われている．これを二部グラフを用いて考えると，syntagmatic relation とは隣接関係を示すエッジにあたり，paradigmatic relation は同じノードと繋がるエッジを持つもう一方のノード間の関係ということになる (同じノードの部分集合にあるためそれらの間にはエッジは存在しない)．

また Harris (1954) は，単語間の共起関係を類型化した一つとして補完的関係を取り上げ，「ほぼ同一の共起要素を持つが，一方はある文脈で生起するが他方は起きない」と定義した上で，補完的関係にある単語のペアとして knife-knives の例を挙げている．これはまさに共起パタン 1 における単数単語と複数単語の関係に対応すると考えられる．

本研究ではこれ以上言語的含意を考察することはできないが，これら考察から言えるのは，言語の分布構造（の部分）には，二部グラフで記述できるような関係があり，グラフ理論を用いると言語学研究において検討された意味関係に係るさまざまな概念を数学的に厳密に定義することができる可能性があるということである．これは，今後の研究課題としたい．

■二部グラフと形式概念分析　言語の分布構造を分析するための数学的ツールとして二部グラフが有望であるとの点に関して，二部グラフと形式概念分析 Formal Concept Analysis (Ganter & Wille, 1999; Ganter, Stumme, & Wille, 2005; Davey & Priestley, 1990) について付言する．

形式概念分析は，応用数学の一つの研究領域であり，順序関係と束論を用いて対象 (objects) と属性 (properties) と名付けられた 2 つの集合の要素間の隣接関係から形式概念を数学的に定義した上で，さらに形式概念の間の順序関係が構成できることを定式化し

たもので，オントロジーなどで用いられている．形式概念分析の前提にある対象と属性の集合間の要素の関係は二部グラフで表現できる．

形式概念分析において定義される形式概念は，二部グラフの中の完全部分グラフと同値である (Bradley, 2020).

完全二部グラフの定義　$\mathcal{G} = (A, B, E)$ を二部グラフとする．ただし,A, B は互いに交わりを持たない集合でその要素はノードを表し，E は直積 $A \times B$ の部分集合で，両集合間のエッジを表す．$\mathcal{G}$ が完全 (complete) であるとは,

$$A \times B = E$$

が成り立つことをいう.

図 3.6 の共起パタン 1 には，部分二部グラフとして，2 つの disjoint な完全二部グラフが存在しており，それぞれが形式概念を構成していることがわかる．これらの形式概念は，それぞれ名詞の単数と複数にあたる．また，8 つの単語を一つのグループとして見るときに，これらの互いに交わりを持たない集合は Harris の言うところの complementary な関係と言うことができる．このように，束論と形式概念分析により，単語の共起関係とそこに現れる意味構造を数学的に取り扱うことができるのである．

さらに形式概念分析は概念の間に自然に順序関係が生じることを示している．単語が品詞などの統語的カテゴリーや意味的なカテゴリーに分類され，階層的なカテゴリーが現れるという言語現象の背景には，二部グラフと形式概念分析が取り扱う数学的な構造があることが予想される．

■代数的構造と統計的特性　単語ベクトルが平行六面体をなすための条件は，トイコーパスとして文（単語の組み合わせ）が固定して与えられた時の条件としては式 (3.23)，文の制約がない場合の単語共起行列の条件として式 (3.24) であることが明らかになったが，これらは言語の構造的要件と出現頻度に関する要件の両方が関わっていることを示している．

前者の場合について，トイコーパスに採用した文には 8 つの主語単語と 6 つの目的語単語の間に言語構造上の対称性があったが，それだけでは平行六面体を構成するには至らず，文の出現確率に一定の制約が求められた．この言語構造的要件は，一様分布のトイコーパスと非一様分布のトイコーパスのそれぞれから生成された共起頻度行列において，ゼロと非ゼロの要素が同じ位置にあることに表れている．仮に出現頻度に関する条件が満たされていなかった場合（ノイズを含む実コーパスではその可能性は大であるが），何ら

かのバイアス補正を行うことで，共起分布から言語構造の規則性を抽出することができるのはないかと考えられる．word2vec などの予測タイプの単語ベクトルがそのような言語構造的規則性を抽出しているのか否か今後の研究課題となる．

一方，単語の共起行列の列ベクトルに関しては，カーネルの基底の線形結合である必要があったが，これは基底が意味関係の基本的な構造を与えた上で，線形結合に現れるウェイトが確率分布を定めているとみなすこともできる．カーネルの基底は，二元体のベクトルとしては平行六面体を構成するものはユニークに定まることを既に指摘したが，これもゼロと非ゼロの構造が先に与えられていることを示しているように思われる．

これらの考察から言えるのは，言語には代数的構造と統計的特性の二つの側面があり，これが regularity（規則性）を生み出していると言うことである．言語は符号化・復号の記号体系であり，そこに何らかの最適化・効率化が作用しているとすれば，代数的構造と統計的特性がそれぞれどのように寄与しているのであろうか．本研究では，その解明のための数学的準備はまだできていない．第 5 章で，代数的構造と統計的特性の双方を一体として扱える数学的フレームワークを構築するための探求を行う．

3.2.7 補論 1：平行六面体における頂点の選び方 (群論)

8 つのベクトルが平行六面体を構成するか否かを調べる際にベクトルと平行六面体の頂点の対応を所与として考えていた．すなわちベクトル $\mathbf{x}_i$ $(i = 1, \ldots, 8)$ が与えられとき，$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_3$ が半時計回りに平行六面体の底面の平行四辺形をなし，$\mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_8, \mathbf{x}_7$ が同様に上面の平行四辺形をなし，その際，$\mathbf{x}_i, \mathbf{x}_{i+4}$ $(i = 1, \ldots, 4)$ がそれぞれ隣接していると考えていた．

しかしながら，この場合，上面の平行四辺形の 4 頂点に割り当てられたベクトルと，底面に割り当てられたベクトルを入れ替えてもまた別の平行六面体を構成することができる．これらを異なる立体図形としてそれぞれ必要十分を求めるのではなく，頂点を入れ替えて合同となる立体図形を同一視した平行六面体の必要十分条件を導出したい．そのためには頂点をどのように選べか良いのであろうか．すなわち置換によって不変となる立体図形は平行六面体をどのように扱うのか，これを頂点の選び方問題と呼ぶ．

3 次元空間にある多面体を自分自身に写す変換は群をなすことを踏まえ，平行六面体をそれ自身の上に写す変換全体のなす群を考える．頂点とベクトルはそれぞれ 8 個あるので，その対応のさせ方は 8! 通り存在する．これを頂点の置換と見れば，この置換は 8 次対称群 S_8 である．

従って，単語ベクトルを行ベクトルとして列方向に積み上げた単語共起行列 C の各行

ベクトルを置換をするためには，置換行列 P として，PC を考えればよい．P は 8 行 8 列で，全ての行と列のそれぞれにおいて，一つの成分が 1 があり残りの成分は 0 であるような行列である．この時，平行六面体の一般化された必要十分条件式 (3.24) は，

$$(\exists P) \quad PC = BW$$

となる．

しかしながら，実証上の問題点として，8 つの単語ベクトルが平行六面体をなすかどうか確認するために全ての置換を調べようとすると膨大な組み合わせを計算しなければならなくなる．8! の組み合わせのうち回転や鏡映により同じ平行六面体となる頂点の選び方はいくつあるのであろうか．

ここでは簡単化のために，平面上の平行四辺形の頂点の選び方をまず考えてみる．平行四辺形の頂点は 4 つであるので，その置換は 4 次対称群 S_4 であり，その位数は $|S_4| = 4! = 24$ である．仮に頂点に $1, 2, 3, 4$ と番号をつけて，1 と 4，2 と 3 がそれぞれ平行四辺形の同じ対角線上にあるとする．

平行四辺形を自分自身に写す変換には，(1) 対角線を軸とする鏡映が 2 つ，(2) 重心を原点とする 180 度回転が 1 つ，(3) 恒等変換，(4) それらの組み合わせがある．

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}$$
$$\sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$$
$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$
$$\epsilon = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

であり，

$$H = \{\epsilon, \sigma_1, \sigma_2, \tau, \sigma_1\sigma_2, \sigma_1\tau, \sigma_2\tau, \sigma_1\sigma_2\tau\}$$

は群をなしている．なお，$\sigma_1^2 = \sigma_2^2 = \tau^2 = \epsilon, \sigma_2\sigma_1 = \sigma_1\sigma_2, \tau\sigma_1 = \sigma_2\tau, \tau\sigma_2 = \sigma_1\tau$ である．H は S_4 の部分群であり，その位数は 8 である．

S_4 は部分群 H により 3 つの剰余類に類別されるので，4 つのベクトルが平行四辺形をなすかどうか調べる場合には，その 3 つの剰余類の代表元について調べれば良い．このことは第 4 章で平行四辺形までの距離を定義する際に必要となり，具体的には 3 つの不変式を考えることになる．

平行六面体に関しては，これを回転と鏡映によって自分自身に写すような変換を考えれば良いが，その数え上げについてはここでは省略する．

3.2.8　補論 2: 条件付き確率化した行列

第 2 章で触れたように共起頻度から単語ベクトルを生成する際に，単語共起行列の成分にウェイト付の変換を行っているものがある．このような単語ベクトルにおいて平行四辺形が構成される条件はどのように考えられるだろうか．条件付き確率による単語共起行列を $\bar{C}$ とすると

$$\bar{C} = \begin{bmatrix} p_1 & p_9 & p_{17} & p_1 & 0 & p_9 & 0 & p_{17} & 0 \\ p_5 & p_{10} & p_{18} & 0 & p_5 & p_{10} & 0 & p_{18} & 0 \\ p_2 & p_{13} & p_{19} & p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\ p_6 & p_{14} & p_{20} & 0 & p_6 & 0 & p_{14} & p_{20} & 0 \\ p_3 & p_{11} & p_{21} & p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\ p_7 & p_{12} & p_{22} & 0 & p_7 & p_{12} & 0 & 0 & p_{22} \\ p_4 & p_{15} & p_{23} & p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\ p_8 & p_{16} & p_{24} & 0 & p_8 & 0 & p_{16} & 0 & p_{24} \end{bmatrix}$$

条件付確率なので $\bar{C}$ の各行ごとの和は 1 に等しい．すなわち以下が成り立っている．

$$\begin{aligned} 2(p_1 + p_9 + p_{17}) &= 1 \\ 2(p_5 + p_{10} + p_{18}) &= 1 \\ 2(p_2 + p_{13} + p_{19}) &= 1 \\ 2(p_6 + p_{14} + p_{20}) &= 1 \\ 2(p_3 + p_{11} + p_{21}) &= 1 \\ 2(p_7 + p_{12} + p_{22}) &= 1 \\ 2(p_4 + p_{15} + p_{23}) &= 1 \\ 2(p_8 + p_{16} + p_{24}) &= 1 \end{aligned}$$

これと式 3.13-3.18 を合わせると平行六面体の必要十分条件は以下の通りとなる．

$$\begin{aligned} &p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 \\ &p_9 = p_{10} = p_{11} = p_{12} = p_{13} = p_{14} = p_{15} = p_{16} \\ &p_{17} = p_{18} = p_{19} = p_{20} = p_{21} = p_{22} = p_{23} = p_{24} \end{aligned} \tag{3.26}$$

式 (3.26) は，各主語単語の生起を条件とした時同じ動詞の動詞句全ての条件付き確率が等しいことを意味しており，一つの式は一つの動詞に対応している．

3.3　幾何的配置の類型

3.3.1　平行四辺形以外の図形への拡張

前節では，2 組の単語ペアが類推関係にあるとき，それらの 4 つの単語ベクトル群が四項類推課題を解くという現象に着想して，8 つの単語ベクトルが平行六面体を構成する時のコーパスや共起行列にかかる必要十分条件を導出した．

自然言語処理では，通常，高次元ベクトル空間にある単語ベクトルを調べる際に 2 つのベクトル間のコサイン類似度やユークリッド距離を用いるのが一般的で四項類推課題もコサイン類似度を用いた演算を使っていた．また，高次元ベクトルを可視化する手法として次元削減をおこなって，多くの場合 2 次元平面で単語ベクトルの位置関係を見ることも行われている．

しかしながら，単語間の分布構造が二項関係だけではなく，単語群として豊かな関係を持っていることが予想されるし，また，高次元空間におけるベクトルの相対的な関係を特徴づけるのは，類似度の評価や次元削減による可視化では不十分である．例えば，上位語・下位語の関係にある単語群は，animal などのカテゴリーを表す上位語の下に dog, cat, monkey などの下位語があり，またこれらの下位語動詞も co-hyponym（同位語）と呼ばれる関係を形成している．

このような単語群（特に 3 以上の単語を含む単語群）に含まれる単語の関係性には，類推関係を平行四辺形として捉えることができるように，その他の図形が対応しているのではないかと考えられるのである．図形は幾何学的に取り扱えるのみならず，群の変換における不変性の概念を用いて数学的に厳密に特徴づけられるメリットがある．

なお図形は一般的には 2 次元平面上の平面図形かあるいは 3 次元空間内の立体図形を指すが，単語ベクトルの関係性を考えるためには高次元空間内の複数ベクトル間の幾何的関係を取り扱うので，図形ではなく幾何的配置と呼ぶこととする．

ここで以降の分析で活用していく数学的概念として「単語ベクトル集合」を定義しておく．

定義　m 次元単語ベクトル（列ベクトル）を $x \in \mathbb{R}^m$ で表すとして，n 個の単語ベクトルを集めて行方向に並べて行列としたものを単語ベクトル集合と言い，行列

$$X = [x_1 \quad x_2 \quad \dots \quad x_n] \in \mathbb{R}^{m \times n}$$

で表すとする．

単語ベクトル集合はベクトルの単なる可算部分集合であるが，そのうち幾何的性質を備えたものを単語ベクトル群として群論的に特徴づけること意図している．物理学や化学で扱われる「点群」を念頭においたものである．数学的には，点群とはある図形の形を保ったまま行う移動操作のうち、少なくとも 1 つの不動点を持つものを元とする群のことであり，結晶や分子対称性を数学的に記述することができる．

以下のいくつかの単語ベクトルがなす図形（幾何的配置）を数学的に特徴づける際に，方程式を用いてその解空間として示すという手法を用いるが，それが頂点の置換に対して不変となることから群との関係も現れてくる．また，複数の単語ベクトルを行列により表現をすると行列演算による取り扱いが可能となるという利点もある．このような数学的な取り扱いは計算機科学の先行研究でも例を見ないものと考えられる．

なお，本節では，単語ベクトルが列ベクトルで表現されていることに留意する．データを行単位で取り扱うのは情報科学の流儀であり，単語ベクトルを行ベクトルとするのが自然言語処理では一般的であるが，数学特に線形代数の慣習に沿ってベクトルは列ベクトルとする．

類推関係は平行四辺形であったが，その他の単語間の統語的関係や意味関係はどのような図形で表せるであろうか．出発点として，まず関係を構成する単語の数が図形の頂点に対応することは前提とできると考えられる．幾何学で取り扱われる図形としては離散的な頂点をもつものは，多角形や多面体である．2 つの頂点がなす幾何的配置は線分または直線であり，それらの直線の関係として平行や直交であることが考えられる．3 つの頂点がなす図形は三角形であり，その中でも正三角形，2 等辺三角形，直角三角形，鋭角・鈍角三角形が考えられる．

以下の各節では，例示的に，4 頂点のケースとして正四面体，5 頂点のケースとして正五角形，6 頂点のケースとして三角柱，さらに 3 点が同一直線上にある幾何的配置，4 点が同一平面上にある幾何的配置を取り上げ，それらの図形をなす単語ベクトル群を数学的に特徴づけていく．

3.3.2 正四面体

前節で見た平行六面体と同様に，正四面体を作るトイコーパスを試作し，その分布構造をモデル化して分析する．ここで考えるコーパスを図 3.7 に示す．正四面体の頂点に対応する 4 つの単語は, north, east, south, west である．コーパスから単語共起頻度をカウントして単語共起行列を作り，4 つの単語に対応するブロック行列を抜き出してこれを C とする．なお，前節とは異なり，単語ベクトルを列ベクトルとしていることに留意する．

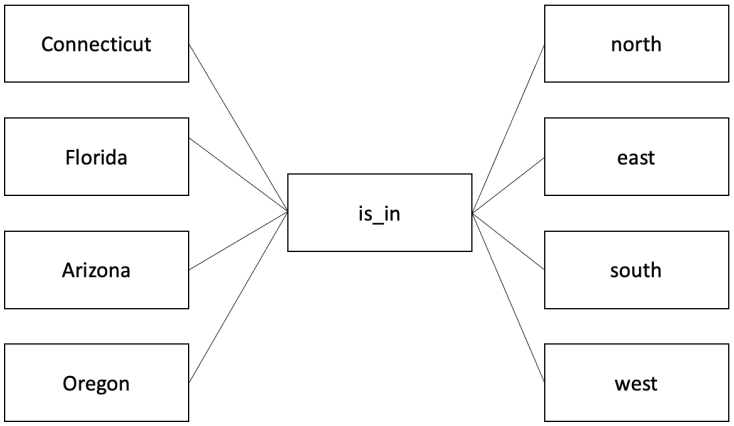

図 3.7　正四面体のコーパス

第 1 列より順に,north, east, south, west に対応する.

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

行列 C の成分を列間で比較すれば明らかであるが，4 つの単語ベクトルいずれの 2 つを選んでも，それらのベクトル間の距離は等しい．4 つの頂点を持つ図形で，そのいずれの頂点間も等距離であるような図形は平面図形ではありえず，立体図形の正四面体である．従って，この 4 単語の単語ベクトルは正四面体を構成している．

行列 C の列ベクトルは 5 次元であるので，任意の座標軸を用いて 3 次元空間へ射影してみると，図 3.8 の通り正四面体が現れる．厳密に言えば，行列 C を特異値分解してその第 2-4 特異ベクトルを座標系としている．特異ベクトルは相互に直交しているので，射影前の空間の距離を保存するので，射影後の空間においても正四面体における頂点間の等距離関係が保存されている．(下記の系を参照)

高次元ベクトル空間において正四面体をなす 4 つの単語ベクトルの関係性を数学的に定義することを試みる．単語群は 4 語からなり，単語ベクトル群を表す行列 X を

$$X = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4] \in \mathbb{R}^{m \times 4}$$

とする（単語ベクトルは m 次元とする)．高次元空間における正四面体の幾何的配置を，4 ベクトル間が等距離であると定義する．4 つのベクトルから 2 つ選ぶ方法は ${}_4C_2 = 6$ 通りであり，これらの 6 組の単語ベクトルの差分ベクトルは次の行列 D で表すことができ

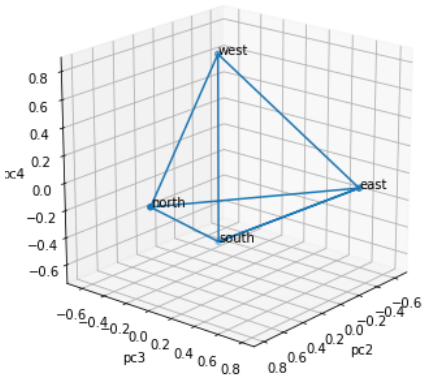

図 3.8　正四面体

る．ただし，二行目に現れる右側の行列を T とする．

$$
\begin{aligned}
D &= [\mathbf{x}_1 - \mathbf{x}_2 \quad \mathbf{x}_1 - \mathbf{x}_3 \quad \mathbf{x}_1 - \mathbf{x}_4 \quad \mathbf{x}_2 - \mathbf{x}_3 \quad \mathbf{x}_2 - \mathbf{x}_4 \quad \mathbf{x}_3 - \mathbf{x}_4 \quad] \\
&= [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4] \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{bmatrix} \\
&= XT
\end{aligned}
$$

頂点 4 点間の距離がすべて等しいということは，6 つの差分ベクトルのノルム（の二乗）がすべて等しいことと同値であるが，これは行列 D^TD の対角成分がすべて等しいということである．

このことは $\mathbf{1}_6 = (1, 1, 1, 1, 1, 1)^T$ として，次のように表現することができる．

$$diagex(D^TD) = k\mathbf{1}_6 \quad (k \in \mathbb{R})$$

ただし，$diagex : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ は，任意の n 行 n 列の行列 $A = (a_{ij})$ に対して，その対角成分をベクトルの成分とする n 次元ベクトル $(a_{11}, \ldots, a_{ii}, \ldots, a_{nn}) \in \mathbb{R}^n$ を返す関数である．

また，D は頂点の置換に関する不変式となっていることに留意する．

$D = XT$ を持ちいれば

$$D^TD = (XT)^TXT = T^T(X^TX)T$$

と展開できるので，単語ベクトル群が平行四面体をなすための必要十分条件は

$$diagex(T^T(X^TX)T) - k\mathbf{1}_6 = \mathbf{0}_6$$

を満たすような k が存在することである．

4 つの単語ベクトルが与えられたときに，その幾何的配置が正四面体にどれだけ近いのか，あるいは遠いのかを「正四面体までの距離 d」として定義することを考える．正四面体の必要十分条件より，4 つのベクトルの作る 6 つの差分ベクトルのノルムを成分とする 6 次元ベクトル空間で特徴付けられ，X が $diagex(T^T(X^TX)T) - \mathbf{1}_6 = \mathbf{0}_6$ のカーネルの中にあるときに正四面体をなしていると捉えることができる．

従って，任意の配置にある 4 つのベクトル群 X の正四面体までの距離は，カーネルまでの最短距離として次のように定義することができる．

$$d = \min_{k\in\mathbb{R}} ||\mathrm{diagex}(T^T(X^TX)T) - k\mathbf{1}_6||$$

$$\mathbf{r} = diagex(T^T(X^TX)T) \in \mathbb{R}^6$$

と置くと

$$\begin{aligned} d &= \min_{k\in\mathbb{R}} ||\mathbf{r} - k\mathbf{1}_6|| \\ &= \min_{k\in\mathbb{R}} \sqrt{\sum_{i=1}^{6}(r_i - k)^2} \\ &= \min_{k\in\mathbb{R}} \sqrt{6k^2 - 2\sum_{i=1}^{6} r_i + \sum_{i=1}^{6} r_i^2} \\ &= \min_{k\in\mathbb{R}} \sqrt{6\Big(k^2 - 2\bar{r}\Big) + \sum_{i=1}^{6} r_i^2} \end{aligned}$$

より d は，$k = \bar{r}$ の時に最小となる．よって，

$$d = \sqrt{\sum_{i=1}^{6}(r_i - \bar{r})^2}$$

■系：特異値分解による座標変換における距離不変　行列 X は特異値分解により $X = U\Sigma V^T$ と表すことができるが，V^T を座標軸としてこれを変換して得られる新たな座標を Y とすると，

$$Y = U^TX = U^T(U\Sigma V^T) = \Sigma V^T$$

となる．このとき，Y の正四面体までの距離を求めるために次を求める．

$$\begin{aligned} Y^TY &= (\Sigma V^T)^T(\Sigma V^T) \\ &= V\Sigma^T\Sigma V^T \\ &= V\Sigma^T(U^TU)\Sigma V^T \\ &= (V\Sigma^TU^T)(U\Sigma V^T) \\ &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= X^TX \end{aligned}$$

従って，

$$diagex(T^T(X^TX)T) = diagex(T^T(Y^TY)T)$$

であり，正四面体までの距離を図るための 6 次元ベクトル空間においては，特異値分解による座標変換は不変であることが示された．

正四面体をなす単語群として考えられるものとしては，同じカテゴリーに属する単語で，そのカテゴリーの中では４つの単語間の関係はいずれも等しい（対称的である）ようなものが予想される．例えば，トイコーパスとして用いた方向を表す単語カテゴリーに属する north, east, south, west のほかに，季節を表す単語カテゴリーに属する spring, summer, autumn, winter や，ものを表す代名詞に属する something, anything, everything, nothing などが正四面体を成している可能性がある．

言語学上，このような単語意味関係に特別な名称は著者の知る限り与えられていないと思われるが，認知的には人間は一連のまとまりとして捉えていることが容易に想像できるので，あえて名付ければ「シリーズ」である．Hashimoto et al. (2016) は，単語ベクトルを用いてこのような４つの単語間の関係の予測を試みており，そのタスクを Series completion と呼んでいる．

3.3.3 正五角形

次に正五角形を作るトイコーパスを示す（図 3.9).

正五角形の頂点に対応する５つの単語は，{monday, tuesday, wednesday, thursday, friday} である．コーパスから単語共起をカウントして単語共起行列を作り，５つの単語に対応するブロック行列を抜き出してこれを C とする．単語ベクトルは列ベクトルであ

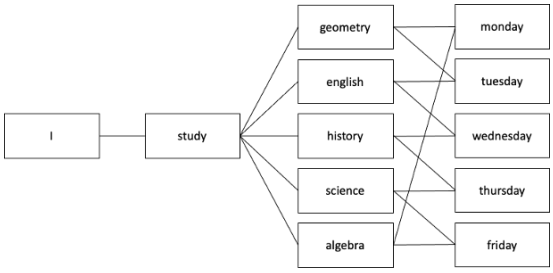

図 3.9　正五角形を作るトイコーパス

り，第 1 列は monday に対応する．C を次に示す．

$$C = \begin{bmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

行列 C の列ベクトルは 7 次元であるので，任意の座標軸を用いて 2 次元平面へ射影すると，図 3.10 のように正五角形が現れる．[*2] 高次元ベクトル空間において正五角形をなす

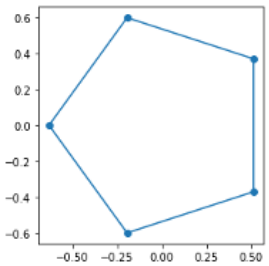

図 3.10　正五角形をなす単語ベクトル群

5 つの単語ベクトルの関係性を数学的に定義する． 単語群は 5 語からなり，単語ベクトル群を表す行列 X とする (単語ベクトルは m 次元)．

$$X \in \mathbb{R}^{m \times 5}$$

$\mathbf{x}_1 \to \mathbf{x}_2 \to \ldots \to \mathbf{x}_n \to \mathbf{x}_1$ の順番で，正五角形の隣り合う頂点であるとする．

平面における正五角形は，重心を中心として $\frac{2\pi}{5}$ 回転すると，それぞれの頂点は隣接する頂点へ移り元の図形に重なる．またそのような回転を 5 回繰り返すと最初の頂点を持つ

[*2] 行列 C を特異値分解した時の第 2 特異ベクトルと第 3 特異ベクトルを座標軸としている．

正五角形に戻る（恒等変換)．この性質を高次元空間の幾何的配置に拡張して，次のように正五角形の定義とする．

すなわち，m 次元ベクトルを線形変換するある行列 R（ある部分空間における回転行列）が存在して，これは 5 回合成すると恒等変換になるとする．I_m を m 行 m 列の単位行列とすると

$$R^5 = I_m \quad (R \in \mathbb{R}^{m \times m})$$

を満たす．

単語ベクトル群のベクトルを順に隣り合ったベクトルへ置換する行列を置換行列 P で表すと次の通りである．

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

行列 P は右から行列 X に作用して，X の 2 列目の列ベクトルを第 1 列へ，3 列目の列ベクトルを第 2 列へと置換して新たな行列 XP を作る．P は 5 行 5 列であるが，一般化して正 n 角形を考える場合には，n 行 n 列の置換行列として次のように定義できる．ただし，$\mathbf{e}_{n,i}$ は n 次元ベクトルでその i 番目の成分が 1 で，それ以外の成分が 0 のものである．

$$P = \sum_{i=1}^{n} \mathbf{e}_{n,i} \otimes \mathbf{e}_{n,i-1(\mathrm{mod}\, n)}^{t}$$

なお, 演算記号 $\otimes$ は，クロネッカー積を表す．

クロネッカー積の定義　$m \times n$ の行列 $A = (a_{ij})$ と $p \times q$ の行列 $B = (b_{ij})$ のクロネッカー積 $A \otimes B$ は次の式で与えられる $mp \times nq$ の行列である．n 次元のベクトルは，$n \times 1$ の行列とする．

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}$$

□

このように定めた回転行列 R と置換行列 P を用いると，5 つの単語ベクトル群 X が正五角形をなすための条件は

$$RX = XP \quad (R^5 = I_m) \tag{3.27}$$

として表すことができる．なお，行列 R は，5 次巡回群の表現であり，行列 P は置換群の表現となっており，正五角形の対称性が直交群（回転群）が作用した時の安定化群として定式化されることを示している．

最後に一般化して，正 n 角形をなすベクトルのある空間を求めるために，カーネルを導出する．正五角形条件を表す式 (3.27) を正 n 角形の条件に一般化をしてさらにベクトル化すると

$$\begin{aligned} RX &= XP \quad (R^n = I_m) \\ vec(RX) &= vec(XP) \\ (I_n \otimes R)vec(X) &= (P^t \otimes I_m)vec(X) \end{aligned}$$

なお，演算記号 vec は，行列をベクトル化する演算子を表し，次のように定義される．

vec 演算子の定義　A を n 列の行列とし，$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_1 & \dots & \boldsymbol{a}_n \end{bmatrix}$ のとおり，$\boldsymbol{a}_i$ $(i = 1, \dots, n)$ を A の列ベクトルとするとき，vec 演算子は次により定義される．

$$vec(A) = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix}$$

□

これは，n 個の m 次元ベクトルを表す m 行 n 列行列 X を，mn 次元ベクトル空間における 1 点として見做して，正 n 角形をなすような点が占める mn 次元空間の部分空間をカーネルとして求めていることになる．カーネルは，以下の式を満たすような $vec(X) \in \mathbb{R}^{mn}$ の集合である．

$$\Big((I_n \otimes R) - (P^t \otimes I_m)\Big)vec(X) = \mathbf{0}_{mn}$$

3.3.4　三角柱

三角柱は 6 つの頂点を持つ立体図形で，合同な 2 つの三角形を上面と底面としてもち 3 つの側面が平行四辺形であるような図形である．従って，6 つの単語の単語ベクトル群を考える．ここで考えるコーパスを図 3.11 に示す．三角柱の頂点に対応する 6 つの単語は，go, goes, went, come, comes, came である． 2 つの動詞のそれぞれ現在形，三単現，過去形の 3 単語が上面・底面をなす三角形を形成する．

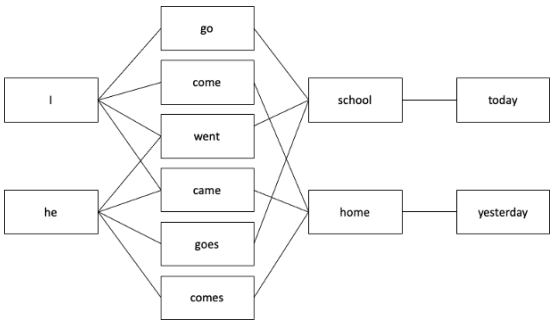

図 3.11　三角柱を作るトイコーパス

コーパスから単語共起頻度をカウントして単語共起行列を作り，6 つの単語に対応するブロック行列を抜き出してこれを C とする．単語ベクトルは列ベクトルに対応し，第 1 列より順に go, goes, went, come, comes, came に対応する．

$$C = \begin{bmatrix} 2 & 0 & 1 & 2 & 0 & 1 \\ 0 & 2 & 1 & 0 & 2 & 1 \\ 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 2 \\ 2 & 2 & 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 & 2 \end{bmatrix}$$

行列 C の列ベクトルは 6 次元であるので，任意の座標軸を用いて 3 次元空間へ射影してみると，図 3.12 の通り三角柱が現れる．厳密に言えば，行列 C を特異値分解してその第 2-4 特異ベクトルを座標系として座標を求めたものを図示している．

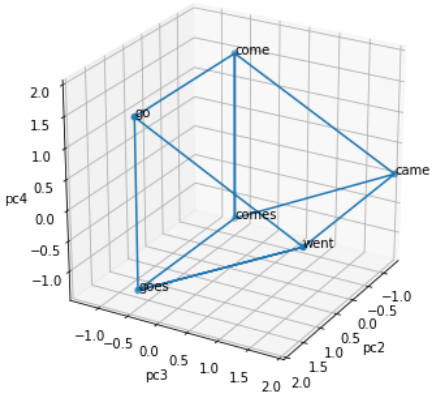

図 3.12　三角柱をなす単語ベクトル群

なお，上面と底面の三角形は二等辺三角形をなしている．これは，現在形と三単現が単語 today と共起し，過去形が yesterday と共起する一方，現在形が I，三単現は he と共

起するが，過去形はIとheの両方と共起する関係があることを反映して，過去形のベクトルに対して現在形と三単現が対称的な位置（線対称）にあることに由来する．他の図形と同様に三角柱をなす単語ベクトル群も数学的に定式化することができる．単語群は6語からなるので，単語ベクトル群を表す行列 X を

$$X \in \mathbb{R}^{m\times 6}$$

とする．単語ベクトルは m 次元であり，$X = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 & \mathbf{x}_6\end{bmatrix}$ において，$\mathbf{x}_1 \to \mathbf{x}_2 \to \mathbf{x}_3$ が反時計回りの順に一つの底面である三角形をなし，残りの $\mathbf{x}_4 \to \mathbf{x}_5 \to \mathbf{x}_6$ が別の底面をなし側面の辺は $\mathbf{x}_1$ と $\mathbf{x}_4$,$\mathbf{x}_2$ と $\mathbf{x}_5$,$\mathbf{x}_3$ と $\mathbf{x}_6$ をそれぞれ結んだベクトルとする．

この時，このベクトル群 X が三角柱をなす条件は3つの側面が平行四辺形を成していることと定義すると，それぞれの平行四辺形の条件は，次の3つの等式で表される．

$$\begin{aligned}\mathbf{x}_1 - \mathbf{x}_4 &= \mathbf{x}_2 - \mathbf{x}_5\\ \mathbf{x}_1 - \mathbf{x}_4 &= \mathbf{x}_3 - \mathbf{x}_6\\ \mathbf{x}_2 - \mathbf{x}_5 &= \mathbf{x}_3 - \mathbf{x}_6\end{aligned}$$

3つ目の式は，1つ目と2つ目の式が成り立てば成り立つので，三角柱の条件を行列表現すれば

$$\begin{bmatrix}\mathbf{x}_1 - \mathbf{x}_4 - \mathbf{x}_2 + \mathbf{x}_5\\ \mathbf{x}_1 - \mathbf{x}_4 - \mathbf{x}_3 + \mathbf{x}_6\end{bmatrix} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 & \mathbf{x}_6\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 0\\ 0 & -1\\ -1 & -1\\ 1 & 0\\ 0 & 1\end{bmatrix} = \mathbf{0}_{m,2}$$

従って，

$$S = \begin{bmatrix}1 & 1\\ -1 & 0\\ 0 & -1\\ -1 & -1\\ 1 & 0\\ 0 & 1\end{bmatrix}$$

とおけば，三角柱の条件は

$$XS = \mathbf{0}_{m,2}$$

あるいはベクトル化して，$6m$ 次元空間を考えると

$$(S^T \otimes I_m)\mathrm{vec}(X) = \mathbf{0}_{2m}$$

であり，これを満たすような点の集まりがカーネルとなる．

3.3.5　アフィン従属

ここまで平面図形または立体図形を拡張して，高次元空間におけるベクトル群の幾何的配置を定式化してきたが，図形以外の幾何的配置は存在するのだろうか．図形としては認知されないが，単語ベクトル間の何らかの規則的な関係性を網羅的に記述する方法は存在するであろうか．

F.Kelin が幾何学を定義した 1872 年の Elrangen プログラム (河田敬義, 1976) を踏まえ，次のように幾何的配置を考える．

- 空間 Ω を考える．単なる集合と考えてよく，これを基礎集合と呼ぶ．
- Ω 上の図形 F を考える．
  - 本研究での用語の使い方に従えば Ω はベクトル空間であり，ここで Klein が図形と呼ぶものは「ベクトル群」に対応する．
  - すなわち，集合の元を二つ以上集めた部分集合がある性質を持つとき図形となる．
  - 集合の元ただ一つは点となるが，これを図形と呼ぶか否かは場合による．
- 図形 F の持つのある種の性質を P とする．
  - 図形の持つ性質とは実はトートロジーであり，ある種の性質を共有して持つものが図形である
  - 本研究では，ベクトル群が図形をなすことを，「ベクトル群が幾何的配置にある」と表現してきた．
- 集合 Ω から集合 Ω への全単射 f で，次の性質を持つものを考える．
  - f は図形 F を同種の図形 $f(F)$ に写す
  - F のもつ性質 P は，$f(F)$ においても不変に保たれる．
- このような性質を持つ写像 f の全体を $\mathbf{G}$ とする．
- この時，$\mathbf{G}$ は写像の合成を積とする群（変換群）を作る．
- 逆に，空間 Ω と，Ω 上の変換群 $\mathbf{G}$ が与えられたとき，Ω 上の図形 F の性質で任意の $f \in \mathbf{G}$ によって不変に保たれるものを Ω 上の幾何学的性質という．

この立場からは，幾何には，それに固有の一つの変換群 $\mathbf{G}$ が定められていて，幾何学的性質と変換群 $\mathbf{G}$ に属する各変換 f で不変に保たれる性質が同値であることになる．例えば，線分の距離を保つ合同変換群は合同な図形を同一視する幾何=合同幾何であり，線分の比（または角度）を保存する相似変換群は相似幾何（ユークリッド幾何）に対応する．

このように整理すると，ここまで考えてきた平面図形や立体図形は相似幾何における幾何的性質であり，本節の冒頭の問いは，相似変換群より大きい変換群にはどのようなものがあるのかという問いに言い換えられる．幾何学で知られているものとして，ユークリッド空間における変換群で相似変換群より大きい変換群としてアフィン変換群がある．

ユークリッド空間 Ω のアフィン変換 $f : \Omega \to \Omega$ は全単射であって，(i) 任意の成分 AB の像は，$f(A), f(B)$ を結ぶ線分であることと,(ii) 同一直線上の 3 点 A, B, C に対して $\frac{\overline{AB}}{\overline{AC}} = \frac{\overline{f(A)f(B)}}{\overline{f(A)f(C)}}$ が成り立つような写像として定義される．アフィン幾何における幾何的性質として，アフィン従属を考える．アフィン変換によって，同一直線上にある 3 点の像は，再び同一直線上をなすし，同一平面上にある 4 点の像は再び同一平面上にある．すなわち，単語ベクトル群が同一直線上や同一平面上にあることをアフィン従属として幾何的性質を特徴づけることができるのである (Gordon & McNulty, 2012).

これを踏まえてアフィン従属を次のように数学的に特徴づけることができる．

■同一直線上 (collinear) にある３つの単語群　単語ベクトル群を表す行列は

$$X \in \mathbb{R}^{m \times 3}$$

X が collinear にある条件は,

$$\exists \mathbf{p} \in \mathbb{R}^3 \quad \text{s.t} \quad X\mathbf{p} = \mathbf{0}_m, \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \quad \sum_{i=1}^{3} p_i = 1$$

これを満たす単語共起行列を例示すれば

$$X = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 0 & 3 \\ 3 & 3 & 3 \end{bmatrix}$$

であり，$p_1 = 1, p_2 = -\frac{1}{3}, p_3 = -\frac{2}{3}$ で条件を充足する．

■同一平面上 (coplanar) にある 4 つの単語群　単語ベクトル群を表す行列は

$$X \in \mathbb{R}^{m \times 4}$$

X が coplanar にある条件は,

$$\exists \mathbf{p} \in \mathbb{R}^4 \quad \text{s.t} \quad X\mathbf{p} = \mathbf{0}_m, \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}, \quad \mathbf{1}_4^t \mathbf{p} = 0$$

$\mathbf{p} = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix}^t$ も $\mathbf{1}_4^t \mathbf{p} = 0$ を満たすので，この場合 $X\mathbf{p} = \mathbf{0}_m$ は X が平行四辺形をなすことの必要十分条件であることに留意すれば，平行四辺形にあるベクトル群は自動的に同一平面上にあることになり，これは幾何学的直観に合致する．

■n 個の単語間のアフィン従属　アフィン従属による幾何的配置の特徴づけは容易に一般化することができる．n 個の単語からなる単語ベクトル群を表す行列を

$$X \in \mathbb{R}^{m \times n}$$

とすると，X がアフィン従属である条件は，

$$\exists t \in \mathbb{R} \quad \text{s.t} \quad X\mathbf{p} = \mathbf{0}_m, \quad \mathbf{1}_n^t \mathbf{p} = 0$$

である．

アフィン従属にある単語群がどのような意味関係を形成しているのかは，現時点では不明である．アフィン変換において不変な性質として他にどのようなものが存在しているのか網羅的に調べ，それらの言語的な含意の有無を検討することは今後の研究課題である．

3.4　まとめと展望

本章では，単語共起行列を分析の対象として，生成系から分散表現までの変換を数学的に定式化し，その数理的構造と性質を明らかにした．構成論的アプローチをとり，人工的なトイコーパスにより単語間の類推関係を表現する単語分散表現を再現した．単語共起行列の中に類推関係に対応した潜在構造が存在することを示し，その必要十分条件が文の生起確率と言語生成系における単語間の共起パタンにあることを数学的に導出した．また，単語分散表現の間の相互関係を幾何的配置として捉える手法を提案し，類推関係以外の単語間の意味関係を表すための図形の類型とその数学的定式化を示した．

本章の研究の最大の成果は，高次元の単語ベクトル間の関係性を幾何学的に特徴づけた上で，言語の生成系からベクトル空間における表象までを線形代数・行列や群論などの数学により取り扱うことを可能にしたことにある．Saussure が述べたように言語の本質は関係の網であることを踏まえると，言語の構造と性質を単語分散表現に現れる分布構造を幾何的な性質として捉えることは重要であると考える．また，類推関係を示す分布構造を単語間の関係を示す二部グラフが特徴づけていたことは，二部グラフと形式概念分析との数学的な関連を鑑みると，言語の意味関係や意味構造の本質的な解明につながる可能性があると期待できる．

次に取り組むべき課題としては，本章で示した数学的定式化のアプローチが現実の言語に適用可能かどうか，適用できるようにするためにはどのように拡張・改変をしていく必要があるかということである．実証的な課題として，本章で提示した幾何的配置が実際の言語構造の中に存在することを実コーパスにより実証することである．この課題は次章で

扱う．理論的な課題としては，言語が持つ階層的な木構造を取り扱うために，本章で示した数学的枠組みをどのように拡張するかである．言語構造を行列表現として符号化するための一般的な枠組みづくりを第 5 章で試みる．

第 4 章

図形距離による類推関係の識別

4.1　背景と目的

第 3 章において検討した理論的モデルにおいては，トイコーパスにある文中に現れる単語間の共起関係が，単語ベクトルの幾何的配置を出現させていた．例えば，意味的統語的関係に基づいてある共通の context 単語に対して，対称的に共起する単語群は平行四辺形や平行六面体，三角柱を構成する可能性があることが明らかとなった．

こうした幾何学的配置は実際に生じているのであろうか． word2vec は，類推関係にある単語の４つ組に対してコサイン類似度を用いて四項類推課題を解くことができていた．その前提として，4 つの単語ベクトルが平行四辺形をなしていることが想定されていたが，実際に高次元のベクトル空間上で平行四辺形をなしているのであろうか．また，四項類推課題の 4 つ組以外にも平行四辺形をなしている単語群が存在するのであろうか．これらの疑問を検証するのが本章の目的である．

本章では実際の単語ベクトルは幾何的配置をなしているのかどうか，また，幾何的配置にある単語群は何らかの意味的統語的関係にあるのかどうかを調べる．具体的には，4 つの単語が平行四辺形をなしているかどうかを判別する指標を用いて，類推関係にある単語群が実際に平行四辺形を形成しているかどうかを評価することを目的とする．特に，平行四辺形距離により類推関係を識別できる可能性を評価する．

複数の単語ベクトルの集まりを単語ベクトル集合と呼ぶ時，そのれらがある図形をなす単語ベクトル群である，すなわち，ある幾何的配置[*1]にあるかどうかを判定したい．さらに，厳密にはそのような幾何的配置にない時も，どれくらい幾何的配置（図形）に近いの

[*1] 3 点が同一直線上にあるような配置など，通常図形とは呼ばない点の配置も対象とするもので幾何的配置と呼ぶ．幾何的配置は数学的に定義される．

かどうか「図形らしさ」を定量化したい．そのため，単語ベクトル集合から考慮している図形までの距離（図形距離）を定義する．

4.2　図形距離の導入

4.2.1　平行四辺形までの距離

単語ベクトルが d 次元ベクトル空間 $V = \mathbb{R}^d$ の元であるとして，考慮している幾何的配置が n 個の単語ベクトルから構成されるとする．この時，nd 次元のベクトル空間 $W = \mathbb{R}^{nd}$ を考えると，任意の n 単語ベクトル群はこのベクトル空間 W の 1 点 $\mathbf{x} \in W$ として考えることができる．その上で，ある幾何的配置をなしているベクトル群（これは W 上では点である）を W 内の部分空間 $U \subset W$ として表すことができれば，点 $\mathbf{x}$ が U に含まれていれば単語ベクトル群であることになる．また，点 $\mathbf{x}$ から部分空間 U までの距離（ユークリッド距離）を求めることができるので，これを図形までの距離（図形距離）と定義することができる．この場合，図形距離がゼロであることが，単語ベクトル群が幾何的配置にあることを示し，図形距離が小さいほど幾何的配置に近いということになる．以下で明らかになるように，単語ベクトル群が幾何的配置にあるための条件を方程式として数学的に定義することができれば，この方程式を満たす点の集合を核（カーネル）として，W の部分空間を構成することができる．

前節の目的で述べたように，本章では 4 つの単語ベクトルがなす平面図形として平行四辺形を考える．

4 つの単語 w_1, w_2, w_3, w_4 を表す単語ベクトルをそれぞれ $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbb{R}^d$ とする．d は単語ベクトルの次元数である．単語群を単語ベクトルを行方向に沿って並べた行列

$$X = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4]$$

で表す．X は d 行 4 列の行列である．

前章において平行六面体を定義したのと同様に，平行四辺形は 4 点のうち 2 点ずつを組みにした対辺ベクトルが等しい図形と定義すると，対辺の組み合わせ方は 3 通りあることから，次の 3 つの方程式のいずれかが成り立つことに等しい．

- $\overrightarrow{\mathbf{x}_1\mathbf{x}_4}, \overrightarrow{\mathbf{x}_2\mathbf{x}_3}$ をそれぞれ対角線とする平行四辺形の場合は，$X\mathbf{p}_1 = \mathbf{0}_d$
- $\overrightarrow{\mathbf{x}_1\mathbf{x}_3}, \overrightarrow{\mathbf{x}_2\mathbf{x}_4}$ をそれぞれ対角線とする平行四辺形の場合は，$X\mathbf{p}_2 = \mathbf{0}_d$,
- $\overrightarrow{\mathbf{x}_1\mathbf{x}_2}, \overrightarrow{\mathbf{x}_3\mathbf{x}_4}$ をそれぞれ対角線とする平行四辺形の場合は，$X\mathbf{p}_3 = \mathbf{0}_d$

ただし，

$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{p}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \mathbf{p}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \tag{4.1}$$

3 つの方程式のいずれかを満たすような行列 X の集合は，それぞれの方程式のカーネルとして次のように表すことができる．

$$K_i = \{X \in \mathbb{R}^{4\times d} | X\mathbf{p}_i = \mathbf{0}_d\} \quad (i = 1, 2, 3)$$

$\mathbb{R}^{4\times d} \cong \mathbb{R}^{4d}$ であり，行列 X を $4d$ 次元ベクトル空間内の点 $\mathbf{x}$ と同一視して，平行四辺形の方程式を次のようにベクトル化する．なお，vec は行列をベクトル化する演算（行列の列ベクトルを 1 列目から順に縦に並べて一つの列ベクトルとする演算）を表し，$\otimes$ はクロネッカー積を表す（定義は，小節 3.3.3 を参照）．

$$\begin{aligned} Xp_i = \mathbf{0}_d \quad &\Leftrightarrow \quad \mathrm{vec}(X\mathbf{p}_i) = \mathrm{vec}(\mathbf{0}_d) \\ &\Leftrightarrow \quad (\mathbf{p}_i^t \otimes I_d)\mathrm{vec}(X) = \mathbf{0}_d \\ &\Leftrightarrow \quad (\mathbf{p}_i^t \otimes I_d)\mathbf{x} = \mathbf{0}_d \\ &\Leftrightarrow \quad A_i\mathbf{x} = \mathbf{0}_d \end{aligned}$$

ただし，$A_i := \mathbf{p}_i^t \otimes I_d \in \mathbb{R}^{d\times 4d}$ であり，I_d は d 行 d 列の単位行列を表わし，また $\mathrm{vec}(X) = \mathbf{x} \in \mathbb{R}^{4d}$ である．

なお，この方程式の中身を確認するためにブロック行列を用いて表すと ($i = 1$ の場合)，

$$\begin{aligned} A_1\mathbf{x} &= (\mathbf{p}_1^t \otimes I_d)\mathrm{vec}(X) \\ &= [I_d \quad -I_d \quad -I_d \quad I_d] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{bmatrix} \\ &= I_d\mathbf{x}_1 - I_d\mathbf{x}_2 - I_d\mathbf{x}_3 + I_d\mathbf{x}_4 \\ &= \mathbf{x}_1 - \mathbf{x}_2 - \mathbf{x}_3 + \mathbf{x}_4 \\ &= X\mathbf{p}_1 = \mathbf{0}_d \end{aligned}$$

単語ベクトル群を行列 X で表す表現とベクトル $\mathbf{x}$ で表す表現が等しいことを確認できる．

これより，先に求めた K_i は行列 A_i のカーネルとして次のように表現できる．

$$K_i = \{\mathbf{y} \in \mathbb{R}^{4d} | A_i\mathbf{y} = \mathbf{0}_d\}$$

ベクトル空間内の任意の点 $\mathbf{x}$ から部分空間であるカーネル K_i までの距離を求めるためには，$\mathbf{x}$ から K_i 内にある点で最も $\mathbf{x}$ に近い点を求め，両点の間の距離を出せばよい．最小距離は点 $\mathbf{x}$ とそれをカーネルへ直交射影した点により与えられるので，次のように直交射影行列を定める．

点 $\mathbf{x}$ を最小距離の位置にある点 $\mathbf{y}_i \in K_i$ に対応づける行列を P_i とする．

$$P_i\mathbf{x} = \mathbf{y}_i \tag{4.2}$$

P_i は K_i への直交射影行列である．

次に，A_i の行空間を $C(A_i)$ とする．行空間は，行列の全ての行ベクトルが張る空間をいう．この時，行列 A_i のカーネル K_i は，$C(A_i)$ の補空間であり，$W = K_i \oplus C(A_i)$ である．従って，$C(A_i)$ への直交射影行列 Π_i を考えると，

$$P_i + \Pi_i = I_{4d} \tag{4.3}$$

である．ただし，I_{4d} は恒等写像を表す $4d \times 4d$ の単位行列である．

$\mathbf{x}$ を $C(A_i)$ へ直交射影した点を $\mathbf{z}_i$ とする．すなわち，

$$\Pi_i\mathbf{x} = \mathbf{z}_i \tag{4.4}$$

$\mathbf{z}_i \in C(A_i)$ であるので，$A_i^t\mathbf{w} = \mathbf{z}_i$ なる $\mathbf{w} \in \mathbb{R}^d$ が存在する．行空間 $C(A_i)$ にあるベクトルは A_i の行ベクトルの線形結合で表されるので，転置行列 A_i^t を用いてこのように表される．

点 $\mathbf{x}$ から点への直線は $\mathbf{z}_i$ は部分空間 $C(A_i)$ への垂線となるので，次が成り立つ．

$$A_i(\mathbf{x} - \mathbf{z}_i) = \mathbf{0} \tag{4.5}$$

式 (4.5) を次のように展開する．

$$\begin{aligned}
A_i(\mathbf{x} - \mathbf{z}_i) &= \mathbf{0} \\
A_i(\mathbf{x} - A_i^t\mathbf{w}) &= \mathbf{0} \\
A_i\mathbf{x} - A_iA_i^t\mathbf{w} &= \mathbf{0} \\
A_i\mathbf{x} &= A_iA_i^t\mathbf{w} \\
(A_iA_i^t)^{-1}A_i\mathbf{x} &= \mathbf{w}
\end{aligned}$$

ここで，求めた $\mathbf{w}$ を式 (4.4) に代入して展開する．

$$\begin{aligned}\Pi_i\mathbf{x} &= \mathbf{z}_i \\ &= A_i^t\mathbf{w} \\ &= A_i^t(A_iA_i^t)^{-1}A_i\mathbf{x}\end{aligned}$$

この式は任意の $\mathbf{x}$ について成り立つので

$$\Pi_i = A_i^t(A_iA_i^t)^{-1}A_i \tag{4.6}$$

である．

$A_i = \mathbf{p}_i^t \otimes I_d$ を用いて，さらに式 (4.6) を展開する．

$$\begin{aligned}\Pi_i &= A_i^t(A_iA_i^t)^{-1}A_i \\ &= (\mathbf{p}_i^t \otimes I_d)^t\big((\mathbf{p}_i^t \otimes I_d)(\mathbf{p}_i^t \otimes I_d)^t\big)^{-1}(\mathbf{p}_i^t \otimes I_d) \\ &= (\mathbf{p}_i^t \otimes I_d)^t\big((\mathbf{p}_i^t \otimes I_d)(\mathbf{p}_i \otimes I_d^t)\big)^{-1}(\mathbf{p}_i^t \otimes I_d) \\ &= (\mathbf{p}_i^t \otimes I_d)^t\big((\mathbf{p}_i^t\mathbf{p}_i) \otimes (I_dI_d^t)\big)^{-1}(\mathbf{p}_i^t \otimes I_d) \\ &= (\mathbf{p}_i^t \otimes I_d)^t\big((\mathbf{p}_i^t\mathbf{p}_i)^{-1} \otimes I_d^{-1}\big)(\mathbf{p}_i^t \otimes I_d) \\ &= (\mathbf{p}_i \otimes I_d^t)\big((\mathbf{p}_i^t\mathbf{p}_i)^{-1} \otimes I_d\big)(\mathbf{p}_i^t \otimes I_d) \\ &= \big(\mathbf{p}_i(\mathbf{p}_i^t\mathbf{p}_i)^{-1}\big) \otimes I_d^tI_d\big)(\mathbf{p}_i^t \otimes I_d) \\ &= \big(\mathbf{p}_i(4)^{-1}\mathbf{p}_i^t\big) \otimes \big(I_d^tI_dI_d\big) \\ &= \frac{1}{4}(\mathbf{p}_i\mathbf{p}_i^t) \otimes I_d\end{aligned}$$

K_i と $C(A_i)$ が直交補空間であることから式 (4.3) で示したように，カーネル K_i への射影行列 P_i は，行空間 $C(A_i)$ への射影行列 Π_i より，$P_i = I_{4d} - \Pi_i$ で与えられる．

従って，点 $\mathbf{x}$ から点 $\mathbf{y}_i$ までの距離 d_i は次のように求められる．

$$\begin{aligned}d_i &= ||\mathbf{x} - \mathbf{y}_i|| \\ &= ||\mathbf{x} - P_i\mathbf{x}|| \\ &= ||(I_{4d} - P_i)\mathbf{x}|| \\ &= ||(I_{4d} - (I_{4d} - \Pi_i))\mathbf{x}|| \\ &= ||\Pi_i\mathbf{x}||\end{aligned}$$

これを $i=1$ の場合に関して具体的に求めれば,

$$\begin{aligned}\Pi_1\mathbf{x} &= \Big(\frac{1}{4}(\mathbf{p}_1\,\mathbf{p}_1^t)\otimes I_d\Big)\mathbf{x}\\ &= \frac{1}{4}\begin{bmatrix} I_d & -I_d & -I_d & I_d\\ -I_d & I_d & I_d & -I_d\\ -I_d & I_d & I_d & -I_d\\ I_d & -I_d & -I_d & I_d\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \mathbf{x}_3\\ \mathbf{x}_4\end{bmatrix}\\ &= \frac{1}{4}\begin{bmatrix}\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3+\mathbf{x}_4\\ -\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3-\mathbf{x}_4\\ -\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3-\mathbf{x}_4\\ \mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3+\mathbf{x}_4\end{bmatrix}\end{aligned}$$

従って

$$||\Pi_1\mathbf{x}||^2 = \frac{4}{16}||\mathbf{x}_1-\mathbf{x}_2-\mathbf{x}_3+\mathbf{x}_4||^2 = \frac{1}{4}||X\mathbf{p}_1||^2$$

であり,

$$d_1 = ||\Pi_1\mathbf{x}|| = \frac{1}{2}||X\mathbf{p}_1|| = \Big|\Big|\frac{\mathbf{x}_1+\mathbf{x}_4}{2}-\frac{\mathbf{x}_2+\mathbf{x}_3}{2}\Big|\Big| \tag{4.7}$$

となる．$\frac{\mathbf{x}_1+\mathbf{x}_4}{2}$ が点 $\mathbf{x}_1$ と点 $\mathbf{x}_4$ の中点であり，$\frac{\mathbf{x}_2+\mathbf{x}_3}{2}$ が点 $\mathbf{x}_2$ と点 $\mathbf{x}_3$ の中点であることに留意すれば，ここで導出された距離の定義の幾何的な解釈としては，四角形の 2 つの対角線の中点間の距離が平行四辺形までの距離ということになる．2 つの対角線の中点が一致する場合には，中点間の距離はゼロになるが，まさにそれが平行四辺形の幾何学的な定義の一つである．

$i=2,3$ の場合にも同様に求めることができる．すなわち，4 つの頂点から対角線をなす頂点として異なる組み合わせを考えた場合の平行四辺形をなす単語ベクトル群の部分空間としてカーネルを求めて，その部分空間までの図形距離を考えることができる．

$$d_2 = \Big|\Big|\frac{\mathbf{x}_1+\mathbf{x}_3}{2}-\frac{\mathbf{x}_2+\mathbf{x}_4}{2}\Big|\Big| \tag{4.8}$$

$$d_3 = \Big|\Big|\frac{\mathbf{x}_1+\mathbf{x}_2}{2}-\frac{\mathbf{x}_3+\mathbf{x}_4}{2}\Big|\Big| \tag{4.9}$$

これらの 3 つの部分空間までの距離のうち，最小のものを平行四辺形の図形距離 d とする．

$$d = \min\{d_1, d_2, d_3\} \tag{4.10}$$

4.2.2 図形距離の正規化

単語ベクトル群が平行四辺形をなしている場合その図形距離は $d=0$ であるが，平行四辺形でない場合には，最も近い平行四辺形までの距離 $d>0$ をユークリッド距離によって

測定することになる．いわば平行四辺形からの歪み度を評価しているわけであるが，その場合，現在の定義では，距離を対角線の中点間の距離として測定しているので，四角形の大きさの影響を受けてしまう．すなわち，お互いに相似な平行四辺形でない四角形があった場合，大きい四角形は小さい四角形より平行四辺形までの距離が遠くなる．ユークリッド幾何において合同・相似な図形を同一視するのと同様に，合同・相似な図形は同じ図形距離を持つように距離を定義するためには，前節で求めた距離を正規化する必要がある．

平行四辺形距離は，単語ベクトル群を表す $4d$ 次元ベクトル空間の点 $\mathbf{x}$ から 3 つのカーネル（部分空間）K_i までの距離として定義したことを想起すれば，原点 O から $\mathbf{x}$ までの距離により正規化することが考えられる．これにより，大きさは異なるが相似であるような 2 つの図形は，正規化することで原点からの距離にかかわらず同じ距離（正規化距離）が付与されることとなる．

しかしながら，この場合同じ図形であったとしても，4 点とも原点から遠い位置にあるような場合には，より大きな距離で正規化されることとなり，合同の図形に対して異なる図形距離を割り当てることになってしまう．従って，平行移動に対しても中立となるような点を原点としてそこからの距離により正規化する必要がある．

具体的には，カーネルにあるいずれの点も平行四辺形をなすベクトル群を示しているが，それらのうち平行移動によって互いに移り合うような点を同一視することとして，平行移動に対して不変となるような原点からの距離により正規化することを考える．その際のトリックとしては，ユークリッド幾何における図形の合同は図形自体を平行移動して重なるかどうかを調べるが，$4d$ 次元ベクトル空間においては，（カーネル内にある）平行四辺形を示している点ではなく，原点自体を平行移動して平行四辺形から最も近くなる点を原点として選ぶのである．

ここで 2 つの空間を考えていることに留意する．すなわち，4 つの単語ベクトルが住んでいる d 次元ベクトル空間 V と，それらの単語ベクトル群を 1 点として取り扱っている $4d$ 次元ベクトル空間 W の 2 つである．

V において 4 つのベクトル $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ がなす図形と合同な図形は，それぞれのベクトルを同じベクトル $\mathbf{b}$ で平行移動して新たに得られるベクトル群であり，$\mathbf{x}_1 + \mathbf{b}, \mathbf{x}_2 + \mathbf{b}, \mathbf{x}_3 + \mathbf{b}, \mathbf{x}_4 + \mathbf{b}$ として表すことができる．すなわち，同じ $\mathbf{b}$ により平行移動される図形を同一視するということである．

合同な図形は，W においてはどのように取り扱えば良いであろうか．これらの 4 つの

単語ベクトルは空間 W において点 $\mathbf{x}$ で次のように表される.

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{bmatrix}$$

そして，これらを $\mathbf{b}$ により平行移動した図形を表す点は

$$\begin{bmatrix} \mathbf{x}_1 + \mathbf{b} \\ \mathbf{x}_2 + \mathbf{b} \\ \mathbf{x}_3 + \mathbf{b} \\ \mathbf{x}_4 + \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \mathbf{b} \\ \mathbf{b} \\ \mathbf{b} \end{bmatrix} = \mathbf{x} + \mathbf{1}_4 \otimes \mathbf{b} = \mathbf{x} + \mathbf{b}'$$

ただし,

$$\mathbf{b}' = \mathbf{1}_4 \otimes \mathbf{b}$$

つまり，いかなる $\mathbf{b}$ に対して，$\mathbf{x}$ と $\mathbf{x} + \mathbf{b}'$ を同一視するということであり，そのような $\mathbf{b}'$ のうち最も近いものを原点として選ぶということである.

ここで空間 W において任意の点 $\mathbf{x}$ とすると，最短距離を与える原点（の平行移動)$\mathbf{b}$ を次のように求める.

$$\begin{aligned} \min_{\mathbf{b}'} \delta &= \min_{\mathbf{b}'} \left\| \mathbf{x} - \mathbf{b}' \right\| \\ &= \min_{\mathbf{b}} \left\| \mathbf{x} - \mathbf{1}_4 \otimes \mathbf{b} \right\| \\ &= \min_{\mathbf{b}} \Big(\sum_{i=1}^{4} (\mathbf{x}_i - \mathbf{b})^2 \Big)^{1/2} \end{aligned}$$

δ の代わりに δ^2 を最小化しても $\mathbf{b}'$ は不変であるため，δ^2 を $\mathbf{b}$ で偏微分する.

$$\frac{\partial \delta^2}{\partial \mathbf{b}} = -2 \sum_{i=1}^{4} \mathbf{x}_i + 8\mathbf{b} = 0$$

従って,

$$\mathbf{b} = \frac{1}{4} \sum_{i=1}^{4} \mathbf{x}_i := \mathbf{m}$$

すなわち，空間 V において 4 つの単語ベクトルの中点 $\mathbf{m}$ として，空間 W においてこれに対応する $\mathbf{1}_4 \otimes \mathbf{m}$ が $\mathbf{x}$ に最も近い原点（平行移動により同一視される点) ということになる.

以上より，図形距離の正規化に用いる正規化項 Z は，$\mathbf{1}_4 \otimes \mathbf{m}$ までの距離ということになる.

$$
\begin{aligned}
Z &= \|\mathbf{x} - \mathbf{1}_4 \otimes \mathbf{m}\| \\
&= \left\| \begin{bmatrix} \mathbf{x_1} - \mathbf{m} \\ \mathbf{x_2} - \mathbf{m} \\ \mathbf{x_3} - \mathbf{m} \\ \mathbf{x_4} - \mathbf{m} \end{bmatrix} \right\| \\
&= \Big(\sum_{i=1}^{4} \|\mathbf{x}_i - \mathbf{m}\|^2 \Big)^{1/2} \\
&= \Big(\sum_{i=1}^{4} \|\overline{\mathbf{x}_i}\|^2 \Big)^{1/2}
\end{aligned}
$$

ここで $\overline{\mathbf{x}_i}$ は，4 点の重心 $\mathbf{m}$ で中心化したベクトルと考えることができるので，これは結局，4 つのベクトルを重心の分だけマイナスに平行移動していることであり，合同となる図形を同一視して，その正規化項を導出していることになる．

4.3　平行四辺形距離による実データの分析

4.3.1　目的

定義した平行四辺形距離を踏まえ，本実験では実コーパスから生成された単語ベクトルを用いて，類推関係にある単語群が実際に平行四辺形を成しているかどうか評価することを目的とする．特に，平行四辺形距離により類推関係を識別できる可能性を評価する．

4.3.2　データ

実コーパスより単語共起行列をカウントして単語ベクトルを生成する．実コーパスは，English wikipedia dump(20171001 時点）であり，コーパス全体のトークン数は約 79 億語，語彙数は約 260 万語である．共起頻度をカウントするコンテクストは窓枠を 5 として, ターゲット単語の前後 5 語以内に生起した単語を共起単語としてカウントする．全語彙のうち出現頻度上位 1000 語をとった上で，上位 1000 語には含まれないが，正解データとして用いるテストセット（後述）に含まれる単語を追加した合計 1487 語に関する共起行列 (1487 行 1487 列）を `freq1000` とする．同様に，出現頻度上位 10000 語をとった上で，上位 1000 語には含まれないが，正解データとして用いるテストセット（後述）に含まれる単語を追加した合計 10072 語に関する共起行列 (10072 行 10072 列）を `freq10000`

とする．それぞれの行列の成分は，行と列に対応した 2 つの単語がコーパス中で共起した頻度（カウント数）である．また，これらの行列は対称行列である（行列を入れ替えて転置しても同一の行列となる）

その上で，これら 2 つの共起行列について，その成分を対数化したものを，それぞれ `logfreq1000`，`logfreq10000` とする．単語に対応する行ベクトルをその単語の単語ベクトルとして用いる．対数化する直接的な理由は，対数化していない共起行列を用いた四項類推課題の精度は悪い一方，対数化した共起行列では一定の精度が確保されることが実証的に明らかとなっているからである (Torii, Maeda, & Hidaka, 2022). [*2]

さらに，対数化した共起行列に対して，特異値分解 (SVD) を行い，その特異値の上位 300 個に対応した特異ベクトルに射影して座標を得たもの，すなわち 300 次元に次元削減したものを `svd1000` と `svd10000` とする．

4.3.3 正解データ

四項類推課題のためのテストデータとして準備された Google Analogy Test Set (Mikolov et al., 2013a) を用いる．これには，19,544 組の単語 4 つ組 (例：king, queen, man, woman) が含まれていて，14 のカテゴリー（うち 9 つは統語的な類推関係，5 つは意味的な類推関係）に分類されている．これらの単語 4 つ組は，2 つの単語ペアの組み合わせ (例：(king, queen) と (man,woman)) からなっており，ユニークな単語ペアとしては 573 組である．Google Analogy Test Set では，全ての単語ペアの組み合わせを網羅していないため，本論文では，あらためて各カテゴリーごとに全ての単語ペアを組み合わせて 30,494 組の 4 つ組を生成した．なお，順番を問わずに 2 つの単語ペアの組み合わせをカウントすると，この半分の 15,247 組である（以下，これを「正解テストセット」という)[*3]

4.3.4 手法

正解テストセットを 1 群として，無作為抽出した 4 つ組を 2 群として，それぞれについて算出された正規化平行四辺形距離の分布を比較する．2 群は，正解テストセットと同数

[*2] 単語の頻度分布が冪乗分布様を示すことが知られている（Zipf の法則) ことを踏まえると，対数化により単語分布に内在する線形構造がより良く捉えられると考えられるが，あくまで経験則による処理であり理論的な根拠づけは今後の課題である

[*3] Google Analogy Test Set には，(king, queen, man, woman) と (man, woman, king, queen) が別の 4 つ組として含まれている．

統計量	logfreq10000		logfreq1000		svd10000		svd1000	
	Analogy	Random	Analogy	Random	Analogy	Random	Analogy	Random
平均	.331	.436	.297	.353	.332	.435	.302	.357
分散	.0066	.0058	.0069	.0095	.0066	.0059	.0068	.0093

表 4.1　正規化平行四辺形距離の計算結果

統計量	logfreq10000	logfreq1000	svd10000	svd1000
t(15,246) 値	116.22	53.94	113.39	53.51
p 値	<.001	<.001	<.001	<.001

表 4.2　t 検定結果

を抽出する．この比較を 4 つの単語ベクトルすなわち，logfreq1000，logfreq10000，svd1000，svd10000 のそれぞれの共起行列から生成された単語ベクトルについて行う．比較は t 検定により行う．

4.4　実験結果

4.4.1　二群間の比較

1 群と 2 群のそれぞれに含まれる単語 4 つ組に対して計算された正規化平行四辺形距離の平均と分散を表 4.1 に示す．表中，1 群（正解テストセット) は Analogy または analogy_testset，2 群は Random また random_sampling と表記している．また，それぞれの分布を図 4.1 に示す．サンプル数は，1 群,2 群とも 15,247 である．

平行四辺形距離は正規化しているので，0 から 1 の間の値を取り，平行四辺形を成している場合には 0 となる．対角線をなす頂点の選び方は 3 通りあり，それぞれに対して計算された距離のうち最小のものを選ぶ（最も近い平行四辺形までの距離を測っている）ので，平行四辺形距離が 1 となることはなく最大値は 1 未満である．計算結果において，それぞれの単語ベクトルにおける平行四辺形距離の最大値は 0.574 ～ 0.577 であった．

2 群の平行四辺形距離の分布について有意水準を 0.001 と定めて t 検定を行った結果，いずれの単語ベクトルにおいても正解テストセットとランダムサンプリングの間に有意な差が得られた（表 4.2).

これより定義された平行四辺形距離が，類推関係のある単語群を捉えていることが示された．

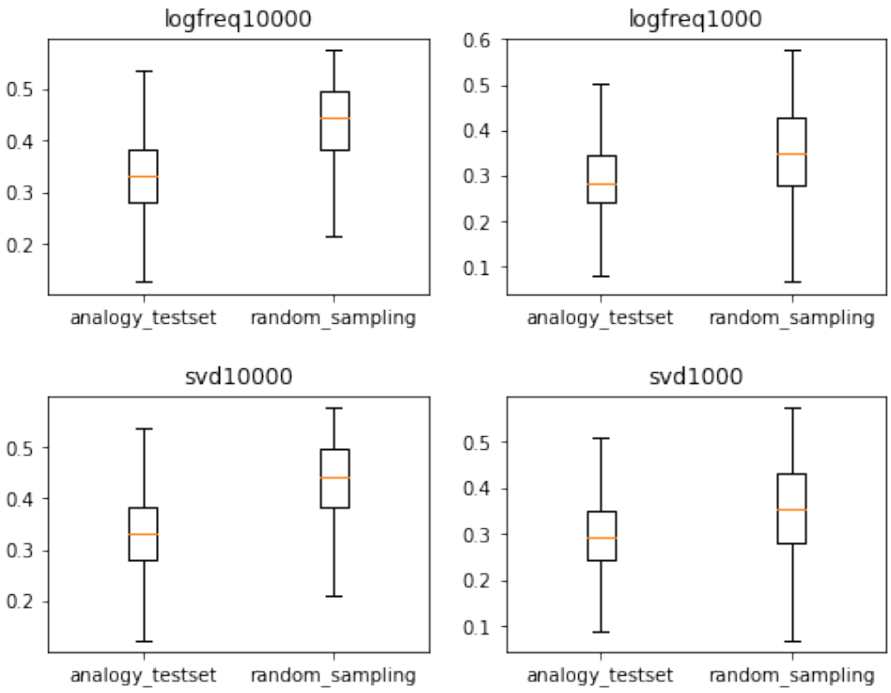

図 4.1　平行四辺形距離の分布

4.4.2　カテゴリー間の比較

カテゴリーごとに含まれている単語群の平行四辺形距離の平均を比較すると，最も距離が小さい（平行四辺形に近い）カテゴリーは family カテゴリーで平均 0.257，次に小さいカテゴリーは gram4-superlative カテゴリーで平均 0.277 である一方，距離が最大のカテゴリー (平行四辺形から遠いカテゴリー）は，gram1-adjective-to-adverb カテゴリーで平均 0.369，次に大きいカテゴリーは opposite カテゴリーで平均 0.358 であった (表 4.3).

このうち距離が最も小さい（平行四辺形に近い）family カテゴリーについて，カテゴリー内の上位 5 つの単語群と下位 5 つの単語群を表 4.4 に示す.

最も平行四辺形に近い単語群は (his, her, nephew, niece) の 4 つ組であり，その値 0.095 は，全てのカテゴリーの中でも最も小さい平行四辺形距離となっている．その他の上位に上がっている単語群の平行四辺形距離も極めて小さいが，それらの単語群の内容を見てみると，(his, her) など出現頻度の高い単語同士のペア（大きいノルムを持つ）と，(nephew, niece) や (stepbrother, stepsister) などの出現頻度の低い単語同士のペア（小さいノルムを持つ）である.

カテゴリー	サンプル数	平行四辺形距離の平均
capital-common-countries	253	0.327
capital-world	6670	0.338
city-in-state	2278	0.341
currency	435	0.296
family	253	0.257
gram1-adjective-to-adverb	496	0.369
gram2-opposite	406	0.358
gram3-comparative	666	0.289
gram4-superlative	561	0.277
gram5-present-participle	528	0.336
gram6-nationality-adjective	820	0.331
gram7-past-tense	780	0.331
gram8-plural	666	0.325
gram9-plural-verbs	435	0.326
Random sampling	15,247	0.436

表 4.3　カテゴリーごとの平行四辺形距離の平均 (`logfreq10000`)

反対に，最も平行四辺形から遠い単語群は，(policeman, policewoman, stepbrother, stepsister) の 4 つ組であり，その値は，0.549 である．ランダムサンプリングを含めた平行四辺形距離の最大値に近い値である．下位単語群の内容をみると，(policeman, policewoman）の policewoman の出現頻度が小さく対称性が低いと考えられるほか，(stepbrother,stepsister) などのようにペアのいずれも出現頻度が低くノイズが大きいことが影響していると思われる．

表中に現れない単語群の例として，(king, queen, man, woman) の平行四辺形距離は 0.320 で順位は 253 組中 176 位である．

4.5　分析と考察

4.5.1　実験結果の考察

類推関係にあると思われる正解テストセットの単語群の平行四辺形距離は，類推関係のないランダムサンプリングに比べて有意に小さい距離を示していた．しかしながら，正解テストセットの中で見ると平行四辺形距離の最大値に近い値を示すような単語群もあり，また主観的には類推関係が高いと思われる単語群の平行四辺形距離はそこまで小さくない

順位		平行四辺形距離	単語群
上位 5 位	1	0.095	his, her, nephew, niece
	2	0.095	father, mother, stepbrother, stepsister
	3	0.096	his, her, stepbrother, stepsister
	4	0.097	grandson, granddaughter,his, her
	5	0.097	his, her, stepson, stepdaughter
下位 5 位	249	0.499	boy, girl, brothers, sisters
	250	0.500	grandpa, grandma, policeman, policewoman
	251	0.503	dad, mom, groom, bride
	252	0.539	policeman, policewoman, stepson, stepdaughter
	253	0.549	policeman, policewoman, stepbrother, stepsister

表 4.4　family カテゴリー平行四辺形距離の上位下位の単語群 (`logfreq10000`)

一方，平行四辺形距離の上位の単語群はノルムの小さい単語ペアを含むものであった．

これらの観察から示唆されるのは，平行四辺形距離と類推関係の対応は十全ではなく，改善の余地があるということである．大きく分けて 2 つの課題が考えられる．一つ目は，正解テストセット自体の質に関することである．正解テストセットの元となった Google Analogy Test Set の作成方法は Mikolov et al. (2013a) には詳細には記述されていないが，それぞれのカテゴリーに対応した単語間の関係性のタイプを想定して人間がマニュアルで単語ペアを抽出したものと想定される．従って，そこにどの程度厳密な類推関係が存在しているのか，すなわちどのような対称性がどの程度存在しているのかは実は何ら保証されていない．最も顕著な問題として考えられるのは多義性を持つ単語の取り扱いである．例えば，family カテゴリーに含まれる単語ペアの（man, woman) において，man は男性の語義のみならず人全般の語義でも用いられるので，そのペアの関係性は (father, mother) のような対称関係とは異なる可能性が考えられる．正解テストセット内における類推関係の非均質性ついては，次小節の分析 1 で検討する．

二つ目は，本研究で定義した平行四辺形距離自体の課題についてである．より正確に言えば，正規化の方法に関する課題である．family カテゴリー内のランキング上位に上がる単語群に見られる傾向として，出現頻度の高い（ノルムの大きい）単語ペアと出現頻度の低い（ノルムの小さい）単語ペアの組み合わせからなる単語群となっている．図形的な直観に即して述べれば，4 つの単語が構成する四角形が細長い図形となっているということである．すなわち，四角形の 4 つの辺のうち，一方の対辺はいずれも短く，他方の対辺はいずれも長いというような図形を成している．4 つの辺の長さが似通っている（菱形に近い）図形と比べると，平行四辺形からのユークリッド距離が同じであっても（同じノル

ムだけ平行四辺形より歪んでいるとしても），細長い図形の場合は正規化に用いる正規化項が大きくなる（分母が大きくなる）ため，結果的に正規化後の平行四辺形距離が大幅に小さくなる効果がある．正規化項が出現頻度の異なるペアからなる単語群を過大評価する一方，縦横のバランスの取れた四角形を過小評価している可能性が考えられる．例えば，family カテゴリーにおいては，(his, her, stepbrother, stepsister) の平行四辺形距離が 0.096 である一方，(father, mother, brother, sister) のような直観的には対称性が高いと思われる単語群の距離は 0.345 である．平行四辺形距離の正規化項の改善について，次次小節の分析 2 で検討する．

4.5.2 分析 1：正解テストセット内の頂点配置

小節 4.2.1 で示したように，任意の 4 点が与えられたとして，平行四辺形をなす頂点の選び方は 3 通りあるため，3 通りの頂点の選び方に対応した距離の中から最小のものを平行四辺形距離と定義してきた（式 4.10）．すなわち 4 つの d 次元ベクトルを $4d$ 次元ベクトル空間の 1 点として表し，その点から平行四辺形をなす点の集まりである 3 つの部分空間のいずれかまでの最短距離を平行四辺形距離としたのであった．この時, 式 (4.10) の d_1 が選ばれた時の頂点配置をパタン 1 として，d_2 が選ばれた時の頂点配置をパタン 2，d_3 が選ばれた時の頂点配置をパタン 3 とすることとする．パタン 1，2，3 は，それぞれ式 (4.1) の $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ による平行四辺形の定義を適用していることに対応する．

正解データセットに含まれている単語ペアは，ある関係性のもとでその関係の方向を反映した並びとなっている．例えば，family カテゴリーでは，(father, mother) のように男性を表す単語・女性を表す単語の並びとなっており，このような単語ペアを 2 つ組み合わせて 4 つ組を作ると単語ベクトルはそれぞれ $\mathbf{x}_1$ がペア 1 の男性単語，$\mathbf{x}_2$ が同じくペア 1 の女性単語，$\mathbf{x}_3$ がペア 2 の男性単語，$\mathbf{x}_4$ が同じくペア 2 の女性単語という並びとなる．従って，仮にこの 4 単語が平行四辺形に最も近い図形を構成するのであれば，

$$\mathbf{x}_1 - \mathbf{x}_2 \simeq \mathbf{x}_3 - \mathbf{x}_4$$

が成り立って，$d_1 \simeq 0$ が最短距離となる，すなわちパタン 1 が選ばれ，次式が成り立つ．

$$d_1 = \frac{1}{2}\|X\mathbf{p}_1\| = \| \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x} & \mathbf{x}_4\end{bmatrix} \begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}^t \| \simeq 0$$

なお，図形距離 d_1 はペア 1 とペア 2 の交換に対して不変である．

しかるに，パタン 2 の頂点配置が最短距離を与える場合には，式（4.8) が適用される

カテゴリー	パタン 1	パタン 2	パタン 3
capital-common-countries	93.28%	3.95%	2.77%
capital-world	96.48%	1.53%	1.99%
city-in-state	54.92%	2.90%	42.19%
currency	69.43%	11.72%	18.85%
family	72.73%	26.09%	1.19%
gram1-adjective-to-adverb	53.63%	39.72%	6.65%
gram2-opposite	72.41%	9.11%	18.47%
gram3-comparative	91.89%	0.00%	8.11%
gram4-superlative	89.66%	5.53%	4.81%
gram5-present-participle	60.42%	34.66%	4.92%
gram6-nationality-adjective	95.61%	2.56%	1.83%
gram7-past-tense	55.13%	43.85%	1.03%
gram8-plural	87.69%	12.16%	0.15%
gram9-plural-verbs	81.38%	14.48%	4.14%

表 4.5　頂点の配置パタン

カテゴリー	family	adj-to-adv	present-participle
パタン 1	he-she	free-freely	generate-generating
	father-mother	calm-calmly	think-thinking
パタン 2	husband-wife	immediate-immediately	sing-singing
	groom-bride	quick-quickly	swim-swimming

表 4.6　順方向と逆方向の差分ベクトルを持つ単語ペアの事例

ので

$$\mathbf{x}_1 - \mathbf{x}_2 \simeq \mathbf{x}_4 - \mathbf{x}_3$$

が成り立っていることになる．

本実験において計算を行った際の頂点配置の選び方の占率を，各カテゴリーごとに示したのが表 4.5 である．多くのカテゴリーにおいては，パタン 1 の配置が最短距離を与えている一方で，パタン 2 が 20% 以上占めているのが family, gram1-adjective-to-adverb, gram5-present-participle, gram7-past-tense の 4 つのカテゴリーである．また，パタン 3 が 20% 以上占めているのは city-in-state カテゴリーである．

示唆されるのは，男性を表す単語と女性を表す単語の間の差分ベクトルの向きが,2 つの組みにおいて逆転する現象が起きているということである．表 4.6 に差分ベクトルが順方向に向いているペアをパタン 1, 逆方向に向いているペアをパタン 2 として例示する．

例えば，いずれも順方向のペアを組み合わせた 4 つ組 (he, she, father, mother) の場合には差分ベクトルの向きがいずれも順方向なので次が成り立つ．

$$\mathbf{v}_{he} - \mathbf{v}_{she} = \mathbf{v}_{father} - \mathbf{v}_{mother}$$

従って，以下の四項類推演算により he,she,mother の 3 単語から father を推定することになる．

$$\mathbf{v}_{he} - \mathbf{v}_{she} + \mathbf{v}_{mother} = \mathbf{v}_{father}$$

しかし，逆転が起きているペアの場合，例えば，(he,she,husband,wife) の 4 つ組では

$$\mathbf{v}_{he} - \mathbf{v}_{she} = \mathbf{v}_{wife} - \mathbf{v}_{husband}$$

が成立しているので，四項類推演算を行う場合には

$$\mathbf{v}_{he} - \mathbf{v}_{she} + \mathbf{v}_{wife} = \mathbf{v}_{husband}$$

ではなく

$$\mathbf{v}_{he} - \mathbf{v}_{she} + \mathbf{v}_{husband} = \mathbf{v}_{wife}$$

を用いて四項類推演算を行わなければならない[*4]．

4.5.3 分析１の考察：反転現象の原因

単語ペアの差分ベクトルが反転する原因については，一義的には単語ベクトルのノルムの逆転が起きているからである。単語共起行列から生成された単語ベクトルでは，その成分は共起頻度を対数化したものであるため，そのノルムの大きさはその単語の出現頻度と正の相関がある．従って，例えば family カテゴリーにおいては多くの単語ペアにおいて男性単語が女性単語よりも出現頻度が多く，男性単語の単語ベクトルのノルムは女性単語のそれよりも大きくなっている．一方，反転しているペア (husband, wife) や (groom, bride) の場合には女性単語のノルムが男性単語のそれよりも大きくなっている．その他のカテゴリーにおいても，同じ語幹を持つ形容詞と副詞では形容詞の頻度が多い単語ペアが多数である一方，表 4.6 に示された (immediate, immediately) などでは副詞の方が頻出している．動詞の現在形と進行形の間でも同様の現象が生じている．

family カテゴリーの多くのペアにおいて男性単語のノルムが女性単語のそれよりも大きいのは，コーパスに含まれる文中で女性よりも男性への言及が一般に多いからである．

[*4] 各式の wife, husband の単語ベクトルの符号が，father, mother を含む式の場合と正負反転していることに注意する．

しかし，出現頻度の逆転しているペア，例えば (husband, wife) のペアで，husband よりも wife が頻出するのは his wife と共起するように wife が生起する文脈は男性単語と同じ文脈である一方，husband は女性の文脈で生起していることが考えられる．同じ family カテゴリーに属するペアであっても，(king,queen) と (husband,wife) のそれぞれの単語間には微妙に異なる関係性が存在していると考えられる．

また，形容詞-副詞や動詞の現在形-現在進行形の関係においても品詞や文法上の区分としては同じ対応関係にあっても，単語ペアによっては用法上の差が存在しているのであって，例えば，present-participle のカテゴリーでは，逆方向となっている単語ペアの現在進行形 (例：singing, swimming) は形容詞として名詞を修飾に使われている頻度も高いと考えられ，そのような異なる用法を持つ単語のペアがサブカテゴリーを成している可能性もある．

以上の分析が示唆するのは，正解テストセットに含まれている類推関係は必ずしも均質ではないということである．すなわち，正解テストセットは人間がマニュアルで各カテゴリーのペアをリストアップすることで作成されているが，その際想定していた意味的関係性以外のタイプの関係性が含まれていることが考えられる．例えば，family カテゴリーであれば男性-女性，adj-to-adv カテゴリーであれば形容詞-副詞，present-participle カテゴリーでは現在形-進行形の対称的関係が想定されていたのに対して，実際の単語分布が示唆する関係は，それと異なる意味的対称性を持っていたり，追加的な意味要素を併せ持つなどしていることが想定される．

逆に言えば，本研究で導入した平行四辺形距離を用いることで，正解テストセット内に潜在している単語間の関係性のタイプの違いを評価して発見することができるということである．

4.5.4 分析２: 正規化項の妥当性

これまでに定義した平行四辺形距離が図形的には何を表しているのか，特に，平行四辺形でない四角形の場合には何を定量化しているのか．前節においては，平行四辺形距離は，四角形の 2 つの対角線の中点間の距離であることを述べたが，図形の性質に関しては何を意味しているのか．さらに，適切な正規化項は何かを明らかにするために，定義した平行四辺形を 4 つ組の単語群に含まれる各単語ペアの差分ノルムに着目して分析をする．

4 つの単語ベクトルを $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ としたときに，4 点（4 つのベクトル）の間には

${}_4C_2 = 6$ つの辺（差分ベクトル）が存在するので，それらを次の差分ベクトルとして表す．

$$\begin{aligned}
\mathbf{u} &:= \mathbf{x}_1 - \mathbf{x}_2 \\
\mathbf{v} &:= \mathbf{x}_3 - \mathbf{x}_4 \\
\mathbf{s} &:= \mathbf{x}_1 - \mathbf{x}_3 \\
\mathbf{t} &:= \mathbf{x}_2 - \mathbf{x}_4 \\
\mathbf{p} &:= \mathbf{x}_1 - \mathbf{x}_4 \\
\mathbf{q} &:= \mathbf{x}_2 - \mathbf{x}_3
\end{aligned}$$

$\mathbf{u}, \mathbf{v}$ は，$\mathbf{x}_1$ と，$\mathbf{x}_4$ が対角線をなすと見るときの平行四辺形 (四角形）の対辺をなし，$\mathbf{s}, \mathbf{t}$ と $\mathbf{p}, \mathbf{q}$ もそれぞれ対辺をなすことに留意する．さらに，$\mathbf{u}, \mathbf{v}$ が四角形の横の辺である場合には，$\mathbf{s}, \mathbf{t}$ は縦の辺となる．

このとき，頂点配置のパタン 1 に対応した平行四辺形距離は，次のように差分ベクトルで表現することができる．

$$\begin{aligned}
d_1 &= \Big\| \frac{\mathbf{x}_1 + \mathbf{x}_4}{2} - \frac{\mathbf{x}_2 + \mathbf{x}_3}{2} \Big\| \\
&= \frac{1}{2} \big\| (\mathbf{x}_1 - \mathbf{x}_2) - (\mathbf{x}_3 - \mathbf{x}_4) \big\| = \frac{1}{2} \big\| \mathbf{u} - \mathbf{v} \big\| \\
&= \frac{1}{2} \big\| (\mathbf{x}_1 - \mathbf{x}_3) - (\mathbf{x}_2 - \mathbf{x}_4) \big\| = \frac{1}{2} \big\| \mathbf{s} - \mathbf{t} \big\|
\end{aligned}$$

すなわち，平行四辺形距離は，横の対辺または縦の対辺で見た場合の差分ベクトルのさらに差分のノルムの半分ということになる．

さらに，差分ベクトルの差分（の二乗）は次のように分解することができる．

$$\begin{aligned}
\|\mathbf{u} - \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\
&= \|\mathbf{u}\| \|\mathbf{v}\| \Big(\frac{\|\mathbf{u}\|}{\|\mathbf{v}\|} - 2 \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} + \frac{\|\mathbf{v}\|}{\|\mathbf{u}\|} \Big) \\
&= \|\mathbf{u}\| \|\mathbf{v}\| \big(r - 2 \cos(\mathbf{u}, \mathbf{v}) + \frac{1}{r} \big) \\
&= \|\mathbf{u}\| \|\mathbf{v}\| \big((r + \frac{1}{r}) - 2 \cos(\mathbf{u}, \mathbf{v}) \big)
\end{aligned}$$

ただし，$r := \frac{\|\mathbf{u}\|}{\|\mathbf{v}\|}$ であり，cos は 2 つのベクトル間のコサイン類似度を表す．この時，同様に以下が成り立っている．

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{s} - \mathbf{t}\|^2 = \|\mathbf{s}\| \|\mathbf{t}\| \big((q + \frac{1}{q}) - 2 \cos(\mathbf{s}, \mathbf{t}) \big)$$

ただし，$q := \frac{\|\mathbf{s}\|}{\|\mathbf{t}\|}$ である．

他の 2 つの頂点配置の場合も含めて，整理すると 3 つの平行四辺形距離と 6 つの差分ベクトルの関係は次の通りである．

$$
\begin{aligned}
4d_1^2 &= \|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{s}-\mathbf{t}\|^2 \\
4d_2^2 &= \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{p}-\mathbf{q}\|^2 \\
4d_3^2 &= \|\mathbf{s}+\mathbf{t}\|^2 = \|\mathbf{p}+\mathbf{q}\|^2
\end{aligned}
$$

以上の分解より，正規化する前の平行四辺形距離の二乗は，対辺を成している差分ベクトルのノルム，並びにそれらのノルム比率とコサイン類似度によって決まることが明らかとなった．

2 つの向き合う差分ベクトル $(\mathbf{u},\mathbf{v})$ のノルム比率が $1(r=1)$ すなわち対辺が同じ長さであり，かつコサイン類似度が 1 すなわち同じ向きで平行であるとき，$r+\frac{1}{r}-2\cos(\mathbf{u},\mathbf{v})=0$ となるが，これは向き合う対辺のベクトルが等しいことを意味する．この 2 つの条件が成り立つとき，また，その時に限り平行四辺形をなす．さらに，これが $\mathbf{u},\mathbf{v}$ において成り立つ時には，$\mathbf{s},\mathbf{t}$ においても成り立つ．すなわち，横の対辺ベクトルが同一である場合，縦の対辺ベクトルも同時に同一でなければならない．

一方，平行四辺形でない場合には，正規化前の平行四辺形距離はノルム比率が 1 より異なるほど大きくなり，またコサイン類似度が 1 より小さくなるほど大きくなる．

これらの洞察より，$r+\frac{1}{r}-2\cos(\mathbf{u},\mathbf{v})$ は，正規化前の平行四辺形距離の二乗を差分ベクトルのノルムの積で正規化したものであり，横の対辺ベクトルの観点から見た平行四辺形からの相対的な歪みと解釈することができる．つまり，平行四辺形からの距離を対辺のノルムの比率とベクトルの向きのずれとして見ることができるのである．

なお，留意すべきは，これは縦の対辺ベクトルの観点から見た平行四辺形の相対的な歪み $(q+\frac{1}{q})-2\cos(\mathbf{s},\mathbf{t})$ とは必ずしも一致しないということである．（平行四辺形の場合にはいずれもゼロとなり一致する）

従って，これらの 2 つの値の平均を平行四辺形からの相対的な歪みとして指標化することが考えられる．

$$
\overline{d_1}^2 := \frac{1}{2}\Big((r_1+\frac{1}{r_1}-2\cos(\mathbf{u},\mathbf{v}))+(r_2+\frac{1}{r_2}-2\cos(\mathbf{s},\mathbf{t}))\Big) \tag{4.11}
$$

d_2, d_3 についても同様に定義される．差分ベクトルに付される符号に留意して，表記すれば以下の通りである．

$$
\overline{d_2}^2 := \frac{1}{2}\Big((r_1+\frac{1}{r_1}+2\cos(\mathbf{u},\mathbf{v}))+(r_3+\frac{1}{r_3}-2\cos(\mathbf{p},\mathbf{q}))\Big) \tag{4.12}
$$

順位		平行四辺形距離（修正版)	単語群
上位 5 位	1	0.532	he, she, his, her
	2	0.535	his, her, prince, princess
	3	0.535	he,she, prince, princess
	4	0.539	his, her, son, daughter
	5	0.541	he, she, king, queen
下位 5 位	249	1.394	boy, girl, man, woman
	250	1.397	boy,girl, prince, princess
	251	1.446	boy, girl, brothers, sisters
	252	1.514	policeman, policewoman, stepson, stepdaughter
	253	1.793	policeman, policewoman, stepbrother, stepsister

表 4.7　family カテゴリー平行四辺形距離（修正後）の上位下位の単語群 (`logfreq10000`)

$$\overline{d_3}^2 := \frac{1}{2}\Big((r_2 + \frac{1}{r_2} + 2\cos(\mathbf{s},\mathbf{t})) + (r_3 + \frac{1}{r_3} + 2\cos(\mathbf{p},\mathbf{q}))\Big) \tag{4.13}$$

ただし，

$$r_1 = \frac{\|\mathbf{u}\|}{\|\mathbf{v}\|}, r_2 = \frac{\|\mathbf{s}\|}{\|\mathbf{t}\|}, r_3 = \frac{\|\mathbf{p}\|}{\|\mathbf{q}\|}$$

と再定義している．平行四辺形距離は，これらの 3 つのうち最小で与えられる．

この代替的な正規化の手法により修正した平行四辺形距離を再計算し，その単語群の順位の変化を調べてみると，ノルムの大きい単語ペアとノルムの小さい単語ペアの順位が大幅に減退し，最初の正規化項による過大評価の効果が緩和されている．表 4.7 に，family カテゴリーの上位下位の単語群を示しているが，上位には直観的に対称性が高いと思われる単語群が並ぶように改善されている．

4.6　結論

本章は，単語分散表現が実際に幾何的配置にあることを実証することを目的として，図形距離による類推関係の識別を試みた．単語の類推関係を平行四辺形として捉えて，実コーパスから生成された単語ベクトルが実際に平行四辺形を成しているかどうか定量化するため，高次元ベクトル空間における平行四辺形までの距離を点からカーネル（部分空間）までの距離として定義した．その際，合同・相似な図形の距離を同じとするための正規化の手法を提案し実コーパスから作成された単語ベクトルを用いて，類推関係にある単語群と無作為抽出した単語群の平行四辺形距離の分布を比較した結果，類推関係にある単語群の平行四辺形距離は有意に小さいことを示した．

しかしながら，類推関係にある単語群の中においても平行四辺形距離のばらつきが大きく，また，正規化された平行四辺形距離が類推関係における直観的な対称性を正しく反映していないとの課題が見られた．正解テストセットの均質性に関する課題として，そもそも 1 つのタイプとして見られる類推関係の中にも関係性の種類の違いが存在している可能性があり，平行四辺形距離の測定はそのような関係性の違いを抽出できることが示唆された．一方，正規化の際のノルムの影響，特に，ノルムの異なる単語ペアの組み合わせの場合過大評価される傾向があり，代替となる正規化手法の提案を行った

本研究の実験結果と分析から見えてきた課題として 4 点を指摘する．

課題 1: 単語間の関係性定義の厳密化　Google Analogy Test Set に含まれる単語群の類推関係は人によりマニュアルで作成したものであり，その性質と程度において均質ではないことが明らかになった．14 のカテゴリーのそれぞれのカテゴリーの中にも異なる関係性の類推関係が混在している可能性である．また，14 カテゴリー以外にも類推関係を持つ単語群が存在していることは容易に想像できる．平行四辺形距離は単語ベクトル間の相対的な位置関係つまり幾何的配置を定量化することを可能にするものであり，これを用いて，類推関係の分類と特定の厳密化を行う必要がある．すなわち，Google Analogy Test Set のカテゴリー分類をさらに細分化したり，まだ含まれていない単語ペアを追加したりすることで，より精度の高い類推関係のテストセットへ改善することができる．

また，Google Analogy Test Set 以外の類推関係のテストセットとしては，SEMEVAL 2012 Task 2 dataset(Jurgens et al., 2012) が存在しており，これは類推関係の関係性の種類やその当てはまり度合いについて人間による評価を経ているものであるので，このテストセットに対する平行四辺形距離を算出して比較することで，人間の直観的な類推関係の分類との対応づけを行うことも考えられる．

課題 2: 距離空間の不均一性　平行四辺形距離の正規化にあたり取り扱いが困難となったのはノルムの大きさの違いをどのように反映するかである．直観的には，単語ベクトルのあるベクトル空間は均質なユークリッド空間ではなく，単語ベクトルが不均一に分布して，局所的に疎密な領域があるような不均一な位相空間であると思われる．例えば，Zipf 則に見られるように単語の出現頻度は冪乗分布の様であり，対数化をおこなったのちでも頻出する単語とレアな単語のノルムの差は大きく，異なるスケールのノルムを持つ単語ベクトル間の距離をユークリッド距離で測定してきたことには問題があるように思われる．

また，ベクトル空間内において単語ベクトルが均一に分布しているわけではなく，あるローカルでは稠密に分布している一方で，別のローカルでは疎となっていることが経験的に予想されている．この観点からも，均質なユークリッド距離を適用することは不適切で

あるように思われる．

単語ベクトル空間のトポロジーをより適切に捉えるためには，双曲空間（ポワンカレ埋め込みと呼ばれる単語ベクトルのアイデア）の他，射影空間の可能性も考えられ，またこれ以外にも非均質な距離を考えるような位相空間の可能性も検討すべきと考える．

課題 3: 多義性・ノイズの取り扱い　距離空間の不均一性とも関連する課題として，単語ベクトルの元データとなる単語共起頻度に関する確率分布をどのように仮定するかという問題がある．コーパスによって単語共起頻度の実現値は異なってくるので，確率的に取り扱うよう平行四辺形距離を拡張していくことが望ましい．その際，異なるローカル領域にある単語ベクトル間のノイズをどの様に取り扱うのか，すなわちノイズを織り込んだ平行四辺形距離の評価を行う必要がある．また，ノイズに含めるべきかどうか現時点では不明であるが，多義性を持つ単語を含む単語群の平行四辺形をどのように扱うかも明らかにする．理想的には，それぞれの語義ごとに単語ベクトルを分解できれば，類推関係に関与する語義の単語ベクトル（の部分）のみを用いて，平行四辺形距離を計算することが考えられる．[*5]

課題 4: 部分空間の抽出　使用頻度の高い単語においては，むしろ多義性を持つ単語の方が多数であることを踏まえると，単語ベクトルの部分のみにおいて単語間の平行四辺形関係が成立していることも考えられる．SVD や PPMI などの前処理や Skip-gram negative sampling などのニューラルネットによる学習を得て得られる単語ベクトルが類推課題でより高い精度を達成するのは，単語共起行列の部分空間にある規則性を抽出していると考えることができる．これらの前処理や学習モデルはあくまで経験的あるいは結果的に言語構造を取り出しているに過ぎず，より意図的・合目的的に言語構造を分解するために単語ベクトルの部分空間に存在する幾何的構造を抽出するような手法を検討することができるのではないだろうか．

今後の研究の展望について述べる．図形距離によりこれまでに見つかっていない単語間の関係性の共通性（類推関係）を同定したり，またより厳密な関係性の分類が可能となると考えられる．また，図形距離は，三以上の単語間の相対的な位置関係を幾何的性質として位置付けることを可能にする新たな研究ツールとなるものであり有望である．さらには，幾何的性質の本質は群における不変性であり，言語の代数構造を特徴づける道筋となる．

[*5] 多義語の単語ベクトルは，各語義に対応する単語ベクトルの線形結合として分解できることが示されている (Arora et al., 2018).

第 5 章

言語構造の行列表現

5.1　文構造の数学表現の一般化

これまでに，トイコーパスから平行六面体が出現するための必要十分条件を求めてきているが，トイコーパス中の文の生成系には，強い言語構造（意味に基づく制約関係）を課してきていた．

具体的には，3 つの動詞 (*live,wear, eat*) が取る目的語や副詞をそれぞれ 2 つに限定（すなわち 6 つの動詞句 VP に限定）した上で，主語となる 8 単語 (*king, queen, man, woman, kings, queens, men, women*) が結びつく VP を，それぞれ 3 つに限定していた（各主語単語は，3 つの動詞のそれぞれに対応する動詞句を一つずつ選ぶ）．

これらの制約はどこまで本質的であるのか，すなわち，あらかじめ与えていた言語構造の意味的制約がどこまで単語ベクトル間に平行六面体関係が出現する条件を定めているのか，どこまで条件を緩和することができるのかを調べる．第 3 章では，平行六面体の必要十分条件に関わる前提を次のように緩和し一般化していた．

1. まず，文の単語構成と文の出現頻度を均一分布に従うとした．
2. 次に出現頻度は非均一な多項分布に従うよう緩和した．
3. 主語と動詞句の組み合わせパタンの制約を除いた．
4. 動詞句を構成する単語の組み合わせの制約（意味制約）を緩和した．
5. 残された前提は，単語は特定の統語範疇に属することと，文が定められた統語範疇の組み合わせ型に従うことのみである（いずれも言語の構造に関わる前提となる）．

意味制約まで緩和した段階で，単語共起行列 C が平行六面体のカーネルの基底行列 B に対して，$C = BX$ と行列分解できることが平行六面体の必要十分条件として導出されて

いた．

しかしながら，ここで条件の対象となったのは単語共起関係を示す単語共起行列のみであり，文の構造や文の出現頻度を直接取り扱うことはできなかった．文の構造がもたらす単語共起への構造的制約を明らかにするためにコーパスを直接数学的に定式化することを試みるのが本章の狙いである．

第 3 章においては，トイコーパスから生成された単語共起行列 C が平行六面体を構成する必要十分条件である式 (3.12) $RC = \mathbf{0}_{6,9}$ に関して，式 (3.19) $C = PU^T$ と行列分解したうえで，そのうち文の生起確率を表す行列 P に着目した上で P の満たすべき条件を導出していた．R は，平行六面体の定義として対辺ベクトルが同じであることを表す行列であり，P は文生起確率を表す行列である．

一方，U^T は特定のトイコーパス中の文がもつ構造に依拠した具体的な行列であり，平行六面体の導出にあたり所与の条件として取り扱っていた．

それでは U^T は，平行六面体条件に対してどのような役割を果たしているのだろうか．U^T の持つ構造はどのようなもので，どこまで緩和された構造とすることができるのであろうか．まず，U^T のもつ意味を考察して，特定のコーパスに寄らない，より一般的な平行六面体条件の導出を考える．

行列 U^T を分析すると，以下のようなブロック行列に分解することができることがわかる．

$$
\begin{aligned}
U^T &= \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \\
&= \begin{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} &
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{bmatrix} \\
&= \begin{bmatrix} I_3 \otimes \mathbf{1}_2 & I_6 \end{bmatrix} \\
&= I_3 \otimes \begin{bmatrix} \mathbf{1}_2 & I_2 \end{bmatrix}
\end{aligned}
$$

但し，

$$\mathbf{1}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

また，演算記号 $\otimes$ はクロネッカー積である (定義は小節 3.3.3 を参照).

U^T の行は主語単語 8 語に対応しており，列はブロック行列ごとに見ると，最初の 3 列が 3 つの動詞，次の 6 列が 6 つの動詞の補語（目的語）に対応したものと考えることができる．そして，U^T の各行 u_i^t $(i = 1, \ldots, 6)$ を見ると, いずれも最初の 3 列に対応した要素のうちひとつの要素のみが 1 を取り残りの要素は 0 である．また，次の 6 列も同様にひとつの要素のみが 1 となっていることから，u_i^T が動詞と目的語をひとつずつ組み合わせた動詞句を表す “Two-hot vector” として符号化されていることがわかる．

“Two-hot vector” とは，語彙中の動詞と目的語あわせた 9 語を次の順番のように並べた時,

$$(\textit{wear, live, eat, tie, dress, palace, house, alone, together})$$

符号化の対象としている動詞句のなかに出現した動詞と目的語のそれぞれについて，ベクトル中の対応する場所にある要素を 1 として，それ以外を 0 とするものにあたる．U^T の各行ベクトルを符号化する前の動詞句に対応させると以下のとおりである．

$$\begin{aligned}
u_1^T &= (1,0,0,1,0,0,0,0,0) && \textit{wear_tie} \\
u_2^T &= (1,0,0,0,1,0,0,0,0) && \textit{wear_dress} \\
u_3^T &= (0,1,0,0,0,1,0,0,0) && \textit{live_palace} \\
u_4^T &= (0,1,0,0,0,0,1,0,0) && \textit{live_house} \\
u_5^T &= (0,0,1,0,0,0,0,1,0) && \textit{eat_alone} \\
u_6^T &= (0,0,1,0,0,0,0,0,1) && \textit{eat_together}
\end{aligned}$$

U^T は，コーパス中に現れるすべての動詞句を集約したリストと解釈できる．

U^T の「因子」として，二つの単位行列 I_3, I_6 が現れている．前者は，動詞カテゴリの 3 語から一語が選ばれて，それら 3 語が相互に背反な関係（いずれか一つのみが生起し，同時には生起しない関係）にあることを示す．後者も同様に目的語カテゴリーから一語が選ばれて，カテゴリーに属する 6 語が相互に背反な関係にあるということ示す．

特定の動詞には特定の目的語のみが結びつくので，この結合関係は直積の部分集合である．つまり，3 つの動詞が，6 つの目的語すべてと結びつくのあれあば $3 \times 6 = 18$ の動詞句ができるが，各動詞は意味的な制約のもとでそれぞれ異なる 2 つの目的語としか結び

つかないので動詞句の組み合わせは 6 組である．例えば，*live* の対象になる (*livable* である）のでその目的語となりうるのは *palace* と *house* であり，それ以外の目的語は取らないという制約を反映している．各動詞は 2 つの目的語しかとれないので $\mathbf{1}_2$ という行列が現れている．

さらに $I_3 \otimes \begin{bmatrix} \mathbf{1}_2 & I_2 \end{bmatrix}$ は，3 つの動詞のそれぞれに対応する目的語が二つずつサブカテゴリをなす（2 行 2 列の単位行列で表現）構造を表していると解釈できる．

こうした意味関係の制約を反映した構造を一般化して，コーパスや文の構造に制約を設けずに一般的な条件を調べるために，言語構造を行列形式で表現して，行列の操作，行列分解，そしてカーネル空間を用いて調べる．

以下では，まずこれまでの平行六面体条件導出において用いてきたベクトルや行列の操作を振り返りながら，コーパスから単語ベクトルを生成し，それらの幾何的配置関係を調べるプロセスをすべて数学的に定式化して，分析のためのフレームワークとすることを目指す．

5.2　コーパスの符号化：集合と直積の導入

トイコーパスに現れる文は，いずれも主語・動詞・目的語すなわち $S-V-O$ の構文を持つ．各統語範疇から，それに属する単語一語を選び，$S-V-O$ の順序で結合することで文が構成される．

これを踏まえて，統語範疇をそれに属する単語の集合として数学的に定式化していく．具体的には，主語単語の集合 S，動詞単語の集合 V，目的語単語の集合 O として，次のように外延的に定めることができる．

$$
\begin{aligned}
S &= \{king, queen, man, woman, kings, queens, men, women\} \\
V &= \{wear, live, eat\} \\
O &= \{tie, dress, palace, house, alone, together\}
\end{aligned}
$$

単語は，属する統語範疇の集合ごとに，その集合のサイズを次元数とする one-hot ベクトルで表す．例えば，主語単語の集合 S の位数は $|S| = 8$ であるので，各単語はベクトル $\mathbf{e}_i \in \mathbb{R}^8$ $(i = 1, \ldots, 8)$ である．ただし，$\mathbf{e}_i$ は i 番目の要素が 1 で，それ以外の要素として 0 を持つ行ベクトルである．この時，*king* は $\mathbf{e}_1$，*queen* は $\mathbf{e}_2$ で表される．さらに，S に属する 8 つの単語に対応するベクトル $\mathbf{e}_i$ $(i = 1, \ldots, 8)$ を列方向に並べると 8×8 の単位行列 I_8 となるので，これを S と同一視する．

一般に，それぞれの統語範疇に対応する集合の要素数を $|S| = m, |V| = n, |O| = l$ とす

るとき，各集合を次の単位行列で表現する．

$$\begin{aligned} S &:= I_m \\ V &:= I_n \\ O &:= I_l \end{aligned} \tag{5.1}$$

トイコーパスに現れる文は，統語範疇を表す 3 つの集合の直積と考えることができる．例えば, *king live-in palace* という (人工の) 文は, 順序付き 3 つ組み $(king, live\_in, palace)$ で表すことができる．数学的に表記すれば次のとおりである．

$$sentence = (subject, verb, object) \in S \times V \times O$$

この直積集合の要素数は mnl であり，トイコーパスの例では $8 \times 3 \times 6 = 144$ 文が，潜在的には存在する，すなわち文の構成を制約する規則が $S - V - O$ の構文に従うことのみである時には 144 個の文法的には正しい文が生成される．

3 章で分析したトイコーパスに 144 文ではなく 24 文のみが現れるのは，構文の制約に加えて特定の動詞がとれる目的語は制約されていることと，特定の主語単語は特定の目的語としか共起しないという意味関係上の制約が追加的に課されているからである．

この観点は以下の定式化において特に重要であり，統語的な構造のもとで生成される文は直積の要素である一方，非文でない文（意味をなす文）は直積の部分集合であり，直積を構成する直積因子間に制約（おそらく意味関係上の制約）があることに由来していることを強調しておく．

次に，動詞 $v \in V$ のと目的語 $o \in O$ を組み合わせて動詞句 $(v, o) \in V \times O$ を数学的に構成するために，2 つの集合の直積に対応する表現として，動詞と目的語の統語範疇に対応する単位行列 I_n, I_l に対して次の操作により新たな行列 VP を生成する．

$V \times O$ は動詞の集合と目的語の集合の直積であるのに対して，VP の各行は，ある動詞句を two-hot vector で符号化した行ベクトルを表現しており，どの動詞句にもただ 1 つの行ベクトルが 1 対 1 対応しているという意味で同型である．

$$VP = \begin{bmatrix} \mathbf{1}_l \otimes I_n & I_l \otimes \mathbf{1}_n \end{bmatrix} \tag{5.2}$$

VP は集合 V の元と集合 O を並べ合わせているという点で直和である $VP = V \oplus O$ と表される．[*1]

[*1] 圏論的には，アーベル群の圏では，積と余積が自然な同型をなすことが証明できる (Awodey, 2010).

トイコーパスの例では，$n = 3, l = 6$ であるので，VP の 1,2,4 行目は次のように表現される．また各行に対応する動詞句を示す．

$$VP_1 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad wear\ tie$$
$$VP_2 = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad live\_in\ tie$$
$$VP_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad wear\ dress$$

ただし，2 つ目の VP_2 は意味をなさないので出現確率がゼロとなる動詞句である．このよう統語的にはありうる文（句）であるが意味的には成立しない文も，この定式化によりゼロの出現確率をもつ文として表現することができる[*2]．直積 $V \times O$ の要素数が $3 \times 6 = 18$ であるのに対応して，VP は 18 行の行列となる．一方，その列数は，直積の 2 つの因子集合の要素の和となり $3 + 6 = 9$ 列である．

なお，$VP = \begin{bmatrix} \mathbf{1}_l \otimes I_n & I_l \otimes \mathbf{1}_n \end{bmatrix}$ のように生成することは，直積の要素を辞書式順序で並べることを意味する．この定め方は，2 つ目の集合の要素を固定した上で，1 つ目の集合の要素を順番に並べる方式を意味しており “Little-endian convention” と呼ばれる (Cichocki, 2013)．反対に 1 つ目の集合の要素を固定して 2 つ目の集合の要素を順番に並べる方式は “Big-endian convention” と呼ばれる．

動詞句をこのよう符号化することの利点は，符号化した行列 VP と，文の生起確率を表す行列の積をとることで，次節に述べるように共起行列を行列の演算として導出することができることである．

例えば，トイコーパス中に king を含む文は 3 文あるが，これを king と各動詞句の順序付き組みとして表し，それぞれの組が表す文の生起確率を次のようにおくとする．

$$Pr\{(king, wear\_tie)\} = p_i$$
$$Pr\{(king, live\_palace)\} = p_j$$
$$Pr\{(king, eat\_alone)\} = p_k$$
$$Pr\{(king, otherwise)\} = 0$$
$$i, j, k \in \{1, \ldots, 144\}$$

さらに 3 つの動詞句 *wear_tie, live_palace, eat_alone* に対応する VP の行をそれぞれ VP_p, VP_q, VP_r とすると，次に示す演算により，king に対応する単語共起行列の行ベク

[*2] チョムスキーの非文の例としてよく知られている “Colorless green ideas sleep furiously” は文法的には正しいが意味をなさない．

トル v_{king} が導出がされる．

$$
\begin{aligned}
& p_i VP_p + p_j VP_q + p_k VP_r \\
=& \begin{bmatrix} p_i & p_j & p_k \end{bmatrix} \begin{bmatrix} VP_p \\ VP_q \\ VP_r \end{bmatrix} \\
=& \begin{bmatrix} p_i & p_j & p_k & p_i & 0 & p_j & 0 & p_k & 0 \end{bmatrix} \\
=& v_{king}
\end{aligned}
$$

ただし，動詞句は次のように符号化されている．

$$
\begin{aligned}
VP_p &= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
VP_q &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\
VP_r &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}
\end{aligned}
$$

共起行列の行ベクトルを単語ベクトルするならば，その単語と文を構成する各動詞句に対応するベクトルと，それらの共起確率（すなわち文の生起確率）との積の和が単語ベクトルとして算出されるのであり，模式的に表せば次のような演算を行っていることになる．

$$
\begin{aligned}
king =& Pr(king, wear\_tie) \times wear\_tie \\
&+ Pr(king, live\_palace) \times live\_palace \\
&+ Pr(king, eat\_alone) \times eat\_alone
\end{aligned}
$$

5.3 分布確率の導入

統語範疇の 3 つの集合の直積が文を表すことから，文（潜在的な文を含む），すなわち直積に属する要素の数は mnl 個であり，その各文の出現確率を p_i $(i = 1, \ldots, mnl)$ とする．確率の総和は $\sum_{i=1}^{mnl} p_i = 1$ である．

p_i を “Little-ending convention” に従って並べて，3 階のテンソル $\underline{\mathbf{P}} \in \mathbb{R}^{m \times n \times l}$ とする．テンソルとはベクトルを 1 次のテンソル、行列を 2 次のテンソルとして考え，多次元配列へ一般化したものである．この 3 階のテンソル $\underline{\mathbf{P}}$ は 3 つの添え字で表される要素 $q_{\alpha,\beta,\gamma}$ $(\alpha \in \{1, \ldots, m\}, \beta \in \{1, \ldots, n\}, \gamma \in \{1, \ldots, l\})$ を持ち、$q_{\alpha,\beta,\gamma} = q_{i \pmod m, i\%m \pmod n + 1, i\%mn+1} = p_i$ である．この式は，i 番目の成分 p_i を，3 階のテンソルの縦方向 α 行，列方向 β 列，奥行方向 γ 段の成分 $q_{\alpha,\beta,\gamma}$ とすることを意味しており，添字 i の順で並んだ要素を手前の行列の縦方向 1 列目，2 列目，$\cdots$，次に奥行き方向

へと並べていることに相当する．演算子 $(\mathrm{mod}\ m)$ は m を法とする剰余演算の結果を返し，演算子 % は整数商を返す．

トイコーパスの例 $(m = 8, n = 1, l = 6)$ を使って，3 階テンソルの各要素 $p_i (i = 1, \ldots, 144)$ の位置を行列スライスで示すと次のとおりである．例えば，$p_{124} = q_{4,13,6}$ である．各行列スライスは，m 行 n 列の行列であり直積 $S \times V$ に対応し，それぞれの行列スライスの添字 i_l は，l 個ある O の要素に対応する．直方体として例えるならば，高さが主語範疇，幅が動詞範疇，奥行きが目的語範疇に対応し，ここで示す行列スライスは正面から奥行きに向かって順にスライスして，2 次元化したものを示していると考えることができる．

$$\underline{\mathbf{P}}_{i_l=1} = \begin{bmatrix} p_1 & p_9 & p_{17} \\ p_2 & p_{10} & p_{18} \\ p_3 & p_{11} & p_{19} \\ p_4 & p_{12} & p_{20} \\ p_5 & p_{13} & p_{21} \\ p_6 & p_{14} & p_{22} \\ p_7 & p_{15} & p_{23} \\ p_8 & p_{16} & p_{24} \end{bmatrix}, \underline{\mathbf{P}}_{i_l=2} = \begin{bmatrix} p_{25} & p_{33} & p_{41} \\ p_{26} & p_{34} & p_{42} \\ p_{27} & p_{35} & p_{43} \\ p_{28} & p_{36} & p_{44} \\ p_{29} & p_{37} & p_{45} \\ p_{30} & p_{38} & p_{46} \\ p_{31} & p_{39} & p_{47} \\ p_{32} & p_{40} & p_{48} \end{bmatrix}, \cdots, \underline{\mathbf{P}}_{i_l l=6} = \begin{bmatrix} p_{121} & p_{129} & p_{137} \\ p_{122} & p_{130} & p_{138} \\ p_{123} & p_{131} & p_{139} \\ p_{124} & p_{132} & p_{140} \\ p_{125} & p_{133} & p_{141} \\ p_{126} & p_{134} & p_{142} \\ p_{127} & p_{135} & p_{143} \\ p_{128} & p_{136} & p_{144} \end{bmatrix}$$

$\underline{\mathbf{P}}$ を文生起確率テンソルと呼ぶ．次節で示すように，前節で導入した統語範疇を符号化した行列と文生起確率テンソルを用いた行列計算により単語共起行列を導出できるようになる．その際に必要となるテンソル演算として **mode-n unfolding** を導入する (Cichocki, 2013)．

■mode-n unfolding の定義 $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_N}$ を N 階のテンソルとする[*3]．$\underline{\mathbf{X}}$ の **mode-n unfolding** $\mathbf{X}_{(n)}$ は，次のように定義される．

$\mathbf{X}_{(n)}$ は，I_n 行 J 列の行列であり，その i_n 行 j 列の成分は次の通りである．$i_n \in \{1, \ldots, I_n\}$ に対して，

$$\mathbf{X}_{(\mathbf{n})_{i_n, j(i_1, \ldots, i_N)}} = \underline{\mathbf{X}}_{i_1, i_2, \ldots, i_N}$$

[*3] 本定義内においてのみ，I_n を添字集合を表す記号として用いている．本定義以外の箇所では I_n は n 行 n 列の単位行列として用いられている．

但し，

$$J := \prod_{k \in \{1,\ldots,N\} \setminus \{n\}} I_k$$
$$j(i_1, \ldots, i_N) := 1 + \sum_{k \in \{1,\ldots,N\} \setminus \{n\}} (i_k - 1) J_k$$
$$J_k := \prod_{m \in \{1,\ldots,k-1\} \setminus \{n\}} I_m$$

mode-n unfolding は N 階のテンソルを，2 階のテンソルである行列に並べ替える操作であり，n は演算結果となる行列の行に対応する．具体例を示す．第 3 章で扱ったトイコーパスにおいて文の生起確率を表すテンソル $\underline{\mathbf{P}}$ は 3 階のテンソルであるので，前項の定義に従えば，次のように 3 つの **mode-n unfolding** を求めることができる．

$$\mathbf{P}_{(1)} = \begin{bmatrix} p_1 & p_9 & p_{17} & p_{25} & p_{33} & p_{41} & \cdots & p_{121} & p_{129} & p_{137} \\ p_2 & p_{10} & p_{18} & p_{26} & p_{34} & p_{42} & \cdots & p_{122} & p_{130} & p_{138} \\ p_3 & p_{11} & p_{19} & p_{27} & p_{35} & p_{43} & \cdots & p_{123} & p_{131} & p_{139} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ p_8 & p_{16} & p_{24} & p_{32} & p_{40} & p_{48} & \cdots & p_{128} & p_{136} & p_{144} \end{bmatrix} \tag{5.3}$$

$$\mathbf{P}_{(2)} = \begin{bmatrix} p_1 & p_2 & \cdots & p_8 & p_{25} & p_{26} \cdots & p_{32} & \cdots & p_{127} & p_{128} \\ p_9 & p_{10} & \cdots & p_{16} & p_{33} & p_{34} \cdots & p_{40} & \cdots & p_{135} & p_{136} \\ p_{17} & p_{18} & \cdots & p_{24} & p_{41} & p_{42} \cdots & p_{48} & \cdots & p_{143} & p_{144} \end{bmatrix}$$

$$\mathbf{P}_{(3)} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 & \cdots & p_{21} & p_{22} & p_{23} & p_{24} \\ p_{25} & p_{26} & p_{27} & p_{28} & \cdots & p_{45} & p_{46} & p_{47} & p_{48} \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \vdots \\ p_{121} & p_{122} & p_{123} & p_{124} & \cdots & p_{141} & p_{142} & p_{143} & p_{144} \end{bmatrix}$$

$\mathbf{P}_{(1)}$ は主語範疇を行側の軸として展開したものであるので８行１８列の行列となり，$\mathbf{P}_{(2)}$ は動詞範疇を行側の軸として展開したもので３行４８列，$\mathbf{P}_{(3)}$ は目的語範疇を行側の軸として展開したもので６行２４列となる．いずれの行列も成分の数は 144 で同じであるが，添字を展開していく並び順はそれぞれ列方向，3×8 のブロック単位，行方向と異なっている点に留意する．

より正確に述べれば，**mode-n unfolding** とは，N 階テンソルに対して，その第 n 番目のモード（次元）を縦軸とし，それ以外の N-1 個の次元については辞書順序形式で 1 次元に平坦化して横軸とした行列を生成する操作である．

トイコーパスの例では 3 つの統語範疇 S, V, O があったので，3 階テンソルである $\underline{\mathbf{P}}$ に対する **mode-n unfolding** は，$n = 1$ であれば，縦軸は主語単語に対応するので 8 行，

横軸には動詞と目的語を組にした動詞句に対応した 18 列を持つ行列を導出する．つまり，この行列　$\mathbf{P}_{(1)}$ の要素はその行にあたる主語単語と列に対応する動詞句が共起する確率を表していることになる．$n = 2, 3$ の場合も同様であり，それぞれ動詞を縦軸とした共起確率を表す行列，目的語を縦軸とした共起確率を表す行列を意味する．

- n=1 の場合：$\mathbf{P}_{(1)}$ は，主語と動詞句の共起確率を表す行列に対応し，8 行 (3×6) 列=8 行 18 列
- n=2 の場合：$\mathbf{P}_{(2)}$ は，動詞と主語・目的語の組の共起確率を表す行列に対応し，3 行 (8×6) 列=3 行 48 列
- n=3 の場合：$\mathbf{P}_{(3)}$ は，目的語と　主語・動詞の組の共起確率を表す行列に対応し，6 行 (8×3) 列=6 行 24 列

5.4　単語共起行列の生成

第 2 節で各統語範疇の単語を組み合わせた文や句を符号化した行列を定め，第 3 節では文の生起確率を統語範疇の直積とするテンソルから, テンソルの演算により単語と句の共起確率を表す行列を導出した．

実はこれらの 2 つの行列の積を取ると，以下に示すように求める単語共起行列を生成することができる．前者の行列は，動詞句などの句がどの単語から構成されているかを表しており，後者の行列は，特定の単語とその他の単語から構成される句の共起確率を表している．これらの行列計算を行うことで単語の共起確率を求めることができる．

単語共起行列 C のうち，主語単語に関する共起行列（共起行列全体のうち主語単語に対応する行ベクトルを集めたブロック行列）を C_S とする．また，すべての動詞句を符号化した行列を VP としていたが，これは動詞集合 V と目的語集合 O の直和であることを踏まえ $V \oplus O := VP$ と表現することとする．式 (5.1) と式 (5.2) を用いて, 動詞句を生成する直和演算は次の通り定義される．

定義：動詞句生成の直和演算　動詞の集合 V と目的語の集合 O のそれぞれの要素数が $|V| = n, |O| = l$ であり，$V := I_n, O := I_l$ と符号化するとき，動詞句を生成するために二つの集合の直和をとる操作は，符号化した行列に対する演算 $\oplus$ として次の通り定義する

$$V \oplus O := \begin{bmatrix} \mathbf{1}_l \otimes V & O \otimes \mathbf{1}_n \end{bmatrix} = \begin{bmatrix} \mathbf{1}_l \otimes I_n & I_l \otimes \mathbf{1}_n \end{bmatrix} \tag{5.4}$$

直和演算を用いると，主語単語の共起ブロック行列 C_S は，文生起確率のテンソル $\underline{\mathbf{P}}$ の mode-1 unfolding と $V \oplus O$ の積で求められる．

$$
\begin{aligned}
C_S &= \mathbf{P}_{(1)}(V \oplus O) \\
&= \mathbf{P}_{(1)} \begin{bmatrix} \mathbf{1}_6 \otimes I_3 & I_6 \otimes \mathbf{1}_3 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_{(1)}\mathbf{1}_6 \otimes I_3 & \mathbf{P}_{(1)} I_6 \otimes \mathbf{1}_3 \end{bmatrix}
\end{aligned}
$$

同様に，動詞単語に関する共起行列ブロックを C_V, 目的語単語に関する共起行列を C_O として，主語目的語の対を符号化した行列を $S \oplus O$, 主語動詞の対を符号化した行列を $S \oplus V$ とする．

$$
\begin{aligned}
C_V &= \mathbf{P}_{(2)}(S \oplus O) \\
&= \mathbf{P}_{(2)} \begin{bmatrix} \mathbf{1}_6 \otimes I_8 & I_6 \otimes \mathbf{1}_8 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_{(2)}\mathbf{1}_6 \otimes I_8 & \mathbf{P}_{(2)} I_6 \otimes \mathbf{1}_8 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
C_O &= \mathbf{P}_{(3)}(S \oplus V) \\
&= \mathbf{P}_{(3)} \begin{bmatrix} \mathbf{1}_3 \otimes I_8 & I_3 \otimes \mathbf{1}_8 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{P}_{(3)}\mathbf{1}_3 \otimes I_8 & \mathbf{P}_{(3)} I_3 \otimes \mathbf{1}_8 \end{bmatrix}
\end{aligned}
$$

この時，単語共起行列全体は次のように表現される．

$$
C = \begin{bmatrix} C_S \\ C_V \\ C_O \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{8,8} & \mathbf{P}_{(1)}(\mathbf{1}_6 \otimes I_3) & \mathbf{P}_{(1)}(I_6 \otimes \mathbf{1}_3) \\ \mathbf{P}_{(2)}(\mathbf{1}_6 \otimes I_8) & \mathbf{0}_{3,3} & \mathbf{P}_{(2)}(I_6 \otimes \mathbf{1}_8) \\ \mathbf{P}_{(3)}(\mathbf{1}_3 \otimes I_8) & \mathbf{P}_{(3)}(I_3 \otimes \mathbf{1}_8) & \mathbf{0}_{6,6} \end{bmatrix}
$$

なお，単語共起行列は対称行列であるので，それぞれ対称の位置にあるブロック行列は次のように互いの倒置行列となっている．

$$
\begin{aligned}
\big(\mathbf{P}_{(1)}(\mathbf{1}_6 \otimes I_3)\big)^t &= \mathbf{P}_{(2)}(\mathbf{1}_6 \otimes I_8) \\
\big(\mathbf{P}_{(1)}(I_6 \otimes \mathbf{1}_3)\big)^t &= \mathbf{P}_{(3)}(\mathbf{1}_3 \otimes I_8) \\
\big(\mathbf{P}_{(2)}(I_6 \otimes \mathbf{1}_8)\big)^t &= \mathbf{P}_{(3)}(I_3 \otimes \mathbf{1}_8)
\end{aligned}
$$

単語共起行列 C のうち，主語単語の単語ベクトルに対応するブロック行列 C_S について数値計算例を示す．

動詞句を符号化した行列 $V \oplus O$ は次の 2 つのブロック行列で示される.

$$
\begin{aligned}
V \oplus O &= \begin{bmatrix} \mathbf{1}_6 \otimes V & O \otimes \mathbf{1}_3 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{1}_6 \otimes I_3 & I_6 \otimes \mathbf{1}_3 \end{bmatrix} \\
&= \begin{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}
\end{bmatrix}
\end{aligned}
$$

一方，第 3 章で扱ったトイコーパスの 24 文の生起確率を $p_1, \ldots, p_{24}$ として，3 階テンソル $\underline{\mathbf{P}}$ を正面からスライスした行列で表す.（添字 1 は *tie*，添字 2 は *dress*，添字 3 は *palace*，添字 4 は *house*，添字 5 は *alone*,，添字 6 は *together* に対応しており，各行列の行は主語，列は動詞に対応する.)

$$
\underline{\mathbf{P}}_1 = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & 0 & 0 \\ p_2 & 0 & 0 \\ 0 & 0 & 0 \\ p_3 & 0 & 0 \\ 0 & 0 & 0 \\ p_4 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \underline{\mathbf{P}}_2 = \begin{bmatrix} 0 & 0 & 0 \\ p_5 & 0 & 0 \\ 0 & 0 & 0 \\ p_6 & 0 & 0 \\ 0 & 0 & 0 \\ p_7 & 0 & 0 \\ 0 & 0 & 0 \\ p_8 & 0 & 0 \end{bmatrix}, \underline{\mathbf{P}}_3 = \begin{bmatrix} 0 & p_9 & 0 \\ 0 & p_{10} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & p_{11} & 0 \\ 0 & p_{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
$$

$$
\underline{\mathbf{P}}_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & p_{13} & 0 \\ 0 & p_{14} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & p_{15} & 0 \\ 0 & p_{16} & 0 \end{bmatrix}, \underline{\mathbf{P}}_5 = \begin{bmatrix} 0 & 0 & p_{17} \\ 0 & 0 & p_{18} \\ 0 & 0 & p_{19} \\ 0 & 0 & p_{20} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \underline{\mathbf{P}}_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & p_{21} \\ 0 & 0 & p_{22} \\ 0 & 0 & p_{23} \\ 0 & 0 & p_{24} \end{bmatrix},
$$

ブロック行列 C_S のサブブロック行列ごとに計算過程を示す．一つ目のサブブロック行

列は次のように導出できる.

$$\mathbf{P}_{(1)}(\mathbf{1}_6 \otimes I_3) = \begin{bmatrix}\underline{\mathbf{P}}_1 & \underline{\mathbf{P}}_2 & \underline{\mathbf{P}}_3 & \underline{\mathbf{P}}_4 & \underline{\mathbf{P}}_5 & \underline{\mathbf{P}}_6\end{bmatrix} \begin{bmatrix} I_3 \\ I_3 \\ I_3 \\ I_3 \\ I_3 \\ I_3 \end{bmatrix}$$

$$= \sum_{i=1}^{6} \underline{\mathbf{P}}_{I_l=i} = \begin{bmatrix} p_1 & p_9 & p_{17} \\ p_5 & p_{10} & p_{18} \\ p_2 & p_3 & p_{19} \\ p_6 & p_{14} & p_{20} \\ p_3 & p_{11} & p_{21} \\ p_7 & p_{12} & p_{22} \\ p_4 & p_{15} & p_{23} \\ p_8 & p_{16} & p_{24} \end{bmatrix}$$

次に，2 つ目のサブブロック行列を求める前に，$\mathbf{E}_i = \begin{bmatrix} \mathbf{e}_i^{\mathrm{t}} \\ \mathbf{e}_i^{\mathrm{t}} \\ \mathbf{e}_i^{\mathrm{t}} \end{bmatrix}$ と置く．但し，$\mathbf{e}_i \in \mathbb{R}^6$ は，
i 番目の要素が 1 で，それ以外は 0 の要素を持つ列ベクトルとする. 例えば,

$$\mathbf{E}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

である．この時 C_S の 2 つ目のサブブロブロック行列は，次のように導出される.

$$\mathbf{P}_{(1)}(I_6 \otimes \mathbf{1}_3) = \begin{bmatrix}\underline{\mathbf{P}}_{I_l=1} & \underline{\mathbf{P}}_{I_l=2} & \underline{\mathbf{P}}_{I_l=3} & \underline{\mathbf{P}}_{I_l=4} & \underline{\mathbf{P}}_{I_l=5} & \underline{\mathbf{P}}_{I_l=6}\end{bmatrix} \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \mathbf{E}_3 \\ \mathbf{E}_4 \\ \mathbf{E}_5 \\ \mathbf{E}_6 \end{bmatrix}$$

$$= \sum_{i=1}^{6} \underline{\mathbf{P}}_{I_l=i}\mathbf{E}_i = \begin{bmatrix} p_1 & 0 & p_9 & 0 & p_{17} & 0 \\ 0 & p_5 & p_{10} & 0 & p_{18} & 0 \\ p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\ 0 & P_6 & 0 & p_{14} & p_{20} & 0 \\ p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\ 0 & p_7 & p_{12} & 0 & 0 & p_{22} \\ p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\ 0 & P_8 & 0 & p_{16} & 0 & p_{24} \end{bmatrix}$$

$C_S = \begin{bmatrix} \mathbf{0}_{8,8} & \mathbf{P}_{(1)}(\mathbf{1}_6 \otimes I_3) & \mathbf{P}_{(1)}(I_6 \otimes \mathbf{1}_3) \end{bmatrix}$ に代入して，次の計算結果を得る．

$$C_S = \begin{bmatrix} 0\cdots0 & p_1 & p_9 & p_{17} & p_1 & 0 & p_9 & 0 & p_{17} & 0 \\ 0\cdots0 & p_5 & p_{10} & p_{18} & 0 & p_5 & p_{10} & 0 & p_{18} & 0 \\ 0\cdots0 & p_2 & p_3 & p_{19} & p_2 & 0 & 0 & p_{13} & p_{19} & 0 \\ 0\cdots0 & p_6 & p_{14} & p_{20} & 0 & P_6 & 0 & p_{14} & p_{20} & 0 \\ 0\cdots0 & p_3 & p_{11} & p_{21} & p_3 & 0 & p_{11} & 0 & 0 & p_{21} \\ 0\cdots0 & p_7 & p_{12} & p_{22} & 0 & p_7 & p_{12} & 0 & 0 & p_{22} \\ 0\cdots0 & p_4 & p_{15} & p_{23} & p_4 & 0 & 0 & p_{15} & 0 & p_{23} \\ 0\cdots0 & p_8 & p_{16} & p_{24} & 0 & P_8 & 0 & p_{16} & 0 & p_{24} \end{bmatrix}$$

以上より，文の出現確率を表すテンソル積に対してテンソル演算を行なって，第 3 章で示したトイコーパスの主語単語に関する単語共起行列を再構成することができた．

8 つの単語ベクトルが平行六面体を構成する必要十分条件は（置換を考えない場合），単語ベクトルを集めた共起行列 C が，式 (3.12) $RC = \mathbf{0}$ を満たすことであった．(行列 R は平行六面体を対辺ベクトルの同一条件として定義する．）

潜在的な文も含めた 144 文の生起確率を表す 3 階テンソルは，平行六面体を構成する時にはどのような条件を充足しなければならないのであろうか．主語単語に対応する単語共起行列 C_S を式 (3.12) に代入すると $RC_S = \mathbf{0}$ であり，行列の積をブロック行列ごとに見ると，次の 3 つの式がすべて成り立つ必要がある．

$$R\mathbf{0}_{8,8} = \mathbf{0}_{6,8} \tag{5.5}$$
$$R\mathbf{P}_{(1)}(\mathbf{1}_6 \otimes I_3) = \mathbf{0}_{6,3} \tag{5.6}$$
$$R\mathbf{P}_{(1)}(I_6 \otimes \mathbf{1}_3) = \mathbf{0}_{6,6} \tag{5.7}$$

式 (5.5) は恒等的に成り立ち式 (5.6) と式 (5.7) が成り立つための必要十分条件は以下の補題により $\mathbf{P}_{(1)}$ の各列ベクトルがカーネル $\ker R$ に含まれることである．

式 (5.3) を再掲する．

$$\mathbf{P}_{(1)} = \begin{bmatrix} p_1 & p_9 & p_{17} & p_{25} & p_{33} & p_{41} & \cdots & p_{121} & p_{129} & p_{137} \\ p_2 & p_{10} & p_{18} & p_{26} & p_{34} & p_{42} & \cdots & p_{122} & p_{130} & p_{138} \\ p_3 & p_{11} & p_{19} & p_{27} & p_{35} & p_{43} & \cdots & p_{123} & p_{131} & p_{139} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ p_8 & p_{16} & p_{24} & p_{32} & p_{40} & p_{48} & \cdots & p_{128} & p_{136} & p_{144} \end{bmatrix}$$

また，カーネルの基底を示す式 (3.22) を再掲する．

$$B = \begin{bmatrix} \mathbf{b_1} & \mathbf{b_2} & \mathbf{b_3} & \mathbf{b_4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

すなわち平行六面体の必要十分条件は，$\mathbf{P}_{(1)}$ の各列ベクトルが，カーネルの基底 $\mathbf{b_1}, \mathbf{b_2}, \mathbf{b_3}, \mathbf{b_4}$ の線形結合で表されることである．

3 章で取り上げたトイコーパスを当てはめていえば，$p_1, \ldots, p_{144}$ のうち非ゼロとなるのは，コーパスの 24 文に対応する 24 個の成分のみである．列ベクトルで言えば，$\mathbf{P}_{(1)}$ は 18 列の列ベクトルを持つが，そのうちゼロベクトルは 12 列，残り 6 個の列ベクトルがそれぞれ 4 個の非ゼロ成分を持つので全体では 24 個の非ゼロ成分がある計算となる．これらの 6 個の非ゼロ成分を含む列ベクトルが 4 つの基底ベクトルの線形結合で表されるということは，コーパスが 4 次の部分空間上にあることを意味する．

3 階のテンソル積を用いることで，SVO という文型の制約はあるものの，文構造的に許される全ての文の生起確率をモデル化できるまで一般化することができたのであり，自然言語の共起構造を幅広く表せる枠組みを得たことになる．そして平行六面体などの幾何学的配置の条件がかなり特殊な構造を反映していることが明らかとなった．そのような特殊な構造は，言語に内在する部分構造を反映していると考えられる．

■補題：行列の積の各列ベクトルがカーネルにある条件

補題　行列 B の各列ベクトルがあるカーネル $\ker R$ に含まれていない時に，行列の積 AB の各列ベクトルがカーネル $\ker R$ に含まれるための必要十分条件は A の各列ベクトルがが $\ker R$ に含まれることである．(単位行列を因子とするクロネッカー積は R のカーネルにないことは計算により確認できる)

証明　行列 AB がカーネルにあることと，AB の各列ベクトルがカーネルの基底の線形結合であることは同値である．一方，AB の各列ベクトルは，行列 A 列ベクトルの線形結合である．従って，これらがカーネルにないとすると AB の列ベクトルでカーネルにないものが存在することになり矛盾である．従って，行列 A の各列ベクトルはカーネルにあるので A はカーネルにある．(証明終)

5.5　考察と展望

前節までの展開を要約すると，我々はコーパスにある言語の分布構造をテンソル積により数学的に定式化できることを示した．そして，単語分散表現の生成に用いられる単語共起行列をテンソル積の演算により導出できることを示した．

この定式化において，文は統語範疇の組み合わせに対応した集合の直積として扱われる．そして，n 階のテンソル積は全ての文=単語の組み合わせの出現確率を表す．

テンソル積の階数にあたる統語範疇の組み合わせは，言語の生成ルール（文法あるいは文型）にあたる．テンソル積の成分がゼロとは，文法的には正文であるが出現確率がゼロの文を意味する．文法的に非文な組み合わせはテンソル積の中に対応する成分を持たない．

反対にテンソル積の成分が非ゼロとは，その組み合わせに対応する文が文法的にも意味的も生起することを意味する．さらに類推関係のような共起分布の規則性は，テンソル積の成分に示される出現確率の間に規則性があることを示す．この規則性は，類推関係を表す平行六面体の必要十分条件のような数学的条件を満たすものとして定義される．

言語の分布構造がテンソル積であるとの示唆は重要であり，ここからいくつかの論理的な帰結が導き出される．まず単語共起行列と単語分散表現に現れる規則性は，テンソル積として表される言語の分布構造に由来することである．テンソル積を共起行列に変換するテンソル演算は，テンソルである言語構造を 2 次元の共起行列へ射影しており，変換で保存される規則性が単語分散表現に現れていると考えられる．

そしてテンソル積が多重線形性を持つことが，単語分散表現が第 2 章で見たように線形性を示していることの背景にあることが考えられる．言語が，生成ルールのもとで単語の組み合わせから構成されることが，必然的に言語の構造に線形性をもたらしていると予想される．

また複数の組み合わせの間に構造的な制約がある場合，群における半直積のような構造的な性質をもたらしている可能性も考えられる．

上に述べた点がテンソル積の構造自体に由来する性質であるとすると，これとは別にテンソル積の成分に表される文の出現確率の間に規則性があることはテンソル積の内部にさらに部分構造があることを示唆する．

ここから多次元配列である n 階のテンソル積の中にある部分構造を抽出する手法とし

てテンソルネットワーク (Cichocki, 2013) を考えるのが有効ではないかと考える．テンソルネットワークは，量子論における多体問題を扱うための手法であり，統計力学や量子科学の理論分野のほか機械学習との関連においても近年有用性が認識され研究が進んでいる．第 3 章で単語共起行列の行列分解を行なって平行六面体の必要十分条件を表す規則性があることを見てきたが，テンソルネットワークも高階のテンソルをより低階のテンソルのネットワークへ分解してその部分構造を取り出すことを目的としており，その間には関連があることが予想される．例えば，テンソルネットワークでも特異値分解によるテンソルネットワークの分解が有効な手法であり，共起行列への射影を行わず，コーパスを表すテンソルをダイレクトにテンソルネットワーク分解することもアイデアとして考えられる．

最後に，テンソルネットワークを用いた今後の研究への課題を述べる．

第一に，実コーパスをテンソルネットワークにより表現するために，統語範疇をどのように定義するかである．与えられたコーパスから，bi-gram に代わる統計量をどのように決めるかということに等しい．その点では，dependency-based の共起行列に類すると考えられるので，これを 2 次元の行列ではなく，n 階のテンソルとして扱うアプローチが考えられる．

第二に，テンソルネットワーク分解のためのアルゴリズムがあるかという課題である．言語現象は物理現象とは異なり階層構造が複雑に組み合わさっていることが予想される．階層構造を持つテンソル積を分解するための手法がどのようなものかを明らかにするのが今後の研究課題である．特に統語範疇以外の範疇として意味範疇をどのように定義するかがポイントと考えられる．

第 6 章

まとめ

コーパスを学習して得られる単語分散表現が単語間の意味的な関係を抽出していることに着目して，本研究は，(1) 単語共起行列の持つ構造と性質を解明すると共に, (2) 言語の構造を数学的に定式化した上で，(3) 実コーパスにおける分布構造の存在を実証することを目的としてきた.

まず，第 2 章では単語分散表現に関する先行研究に沿ってその類型と性質を概観した結果，(1) 単語分散表現の本質は共起行列であること, (2) 単語分散表現には類義関係に限らない言語の分布関係が符号化できること, 並びに (3) 分布構造は線形的な性質 (Arora et al., 2018) を持っている可能性があるとの洞察を得た.

次に，第 3 章で単語共起行列の内部構造を生成段階まで遡り，分散表現間の幾何的な関係が立ち現れる過程の例を数学的に分析した．構成論的アプローチをとり人工的なコーパスにより単語間の類推関係を表現する単語分散表現が平行六面体を構成する状況をトイモデルにより再現した．そして単語共起行列の中に意味関係に対応した潜在構造が存在することを示し，類推関係を持つ単語ベクトルの必要十分条件が文の生起確率と言語生成系における単語間の共起パタンにおける対称性であることを示した．また，この対称性が言語学上の syntagama と paradigma の概念に対応する意味関係と関連づけられ，数学的に完全二部グラフにより記述できることを示した．さらに，トイモデルを拡張して，ベクトル空間における単語分散表現の間の相互の位置関係を，平行四辺形以外の図形など一般の幾何的配置として捉える手法を提示し，単語の分布構造が 4 単語間の類推関係以外の意味関係をもつ可能性を示した.

第 3 章の単語共起行列に関する数理研究の結果を踏まえ，第 4 章では平行四辺形などの幾何的配置が実際の共起構造の中に存在することを実コーパスにより実証するため，高次元ベクトル空間における単語ベクトル群の特定の幾何的配置までの近さを定量的に評価す

るためのカーネル（部分空間）までの図形距離を定義し，合わせて，合同・相似な図形を同一視できるよう図形距離の正規化の手法を提案した．その上で，実コーパスから作成された単語ベクトルを用いて，類推関係にある単語群と無作為抽出した単語群の平行四辺形距離の分布の間に有意な差があることを示した．

第 5 章では，第 3 章のトイコーパスで措定した単語生成過程に関する強い仮定を緩和して文の構造がもたらす単語共起への構造的制約を明らかにするため一般的な文構造を行列表現として符号化する手法を提案し，コーパスの分布構造がテンソル積や直和などのテンソル操作で表現できることを示した．単語分散表現の生成に用いられる単語共起行列をテンソル積の演算により導出できることを示した．また，テンソルネットワークを用いて言語の分布構造の内部にある部分構造を抽出できる可能性を指摘した．

本研究の問い (1.5 節) に即して結論すれば，第一に，類推関係にある単語の分散表現が平行四辺形や平行六面体の幾何的配置にあるための単語の共起関係の構造と生起確率の数学的条件を導出した．また類推関係以外にも同位語関係など対称性を持つ意味関係が，平行四辺形以外の単語分散表現の幾何的配置をなすことを数理的に示した．第二に，文を統語範疇の直積と考え言語の分布を n 階のテンソル積で表すことで，単語共起行列の前段階にある文の生成系（文単位での単語の共起関係）を反映した言語の分布を数学的に記述できた．さらに単語共起行列は 2 階テンソルへの射影であり，言語の分布における線形性や規則的な性質が共起行列へ保存されている可能性を指摘した．第三に，高次元ベクトル空間における単語分散表現の配置を定量的に評価するための図形距離を定義したが，精度は不十分であり, ベクトル空間の位相の特徴を踏まえたノルムの定義とノイズの取り扱いに課題を残した．

本研究の最大の成果は，高次元の単語ベクトル間の関係性を幾何学的に特徴づけた上で，言語の生成系からベクトル空間における表象までを線形代数・行列や群論などの数学により取り扱うための道筋を示したことにある．

また，言語の分布構造がテンソル積で表現可能であるとの示唆は重要であり，仮に自然言語がテンソル積の構造を持つのであれば，単語共起行列と単語分散表現に現れる規則性は，テンソル積として表される分布構造に由来する可能性を指摘した．

単語分散表現に現れる分布関係の規則性は，テンソル積から 2 階の共起行列への射影写像を通じて保存されたものであるとの新たな認識に立てば，コーパスの分布構造を表す高階のテンソル積をテンソルネットワーク分解の手法により，複数の低階なテンソルの組み合わせに分解することで，言語の分布構造の内部に潜在する部分構造，すなわち単語分

散表現に現れる分布関係の潜在的な規則性を特定することができるようになると考えられる．

単語分散表現が単語の意味関係を表象するような既存研究で観察された言語現象を具体的な代数系により記述し定式化して再現できたことで単語分散表現による自然言語の数理モデルを構築する今後の研究への道筋をつけられた．

図形距離は，これまでに見つかっていない単語間の関係性の共通性（類推関係）を同定したり，より厳密な関係性の分類が可能となると考えられる．また，図形距離は，単語間の相対的な位置関係を幾何的性質として位置付けることを可能にする新たな研究ツールとなるものであり有望である．

今後の研究に残された課題としては，第一に，理論面での課題として，自然言語の持つ複雑な代数構造をテンソルネットワークとしてモデル化することである．本研究成果の対象は，単語の統語範疇とその組み合わせ規則としての構文タイプが単純なものに限定されていたが，自然言語は再帰的な生成ルールに由来する複雑系である．特に言語の持つ階層的な木構造や複雑な構文構造を代数構造の中で明示的に取り扱うことが必要である．そのためには，統語範疇の中にサブカテゴリがあるような階層構造や，多義語のように複数の統語範疇で用いられる単語を扱えるようテンソルによる定式化を拡張する必要がある．また，実コーパスを対象にした学習という観点で，他分野で研究の進むテンソルネットワーク分解の手法を自然言語に応用する可能性もあり有望な研究テーマとなる．

第二に，根源的な課題として，なぜテンソルや群などの代数が自然言語に現れるのか、そもそも単語列は何を符号化しているのかという問いがある．言い換えれば，言語という記号体系で表現される意味とはなにかを問うことである．Harris は，言語の分布構造は意味の構造と 1 対 1 ではない対応があると考えていたが，意味とは何かを問う意味論を構造言語学の研究対象から除外していた．一方，Saussure は言語の意味＝価値は，関係の網で相互依存的に決定されると考えていた．本研究では，平行六面体の数学的条件の言語学的含意を考察した際に (第 3 章)，主語単語間の類推関係の根底には二部グラフの完全部分グラフとして見做せる数学的関係性があることを示した．また，Paradigma や Syntagma などの言語学上の概念が数学的に特徴づけできる可能性もすでに示唆した (小節 3.2.6)．今後の具体的な研究課題としては，単語の単純な共起関係から代数構造が現れるメカニズムを二部グラフ構造と形式概念分析を用いてモデル化することが考えられる．

第三に，実コーパスから生成される単語分散表現の空間がユークリッド空間ではないことが示唆されており，実証的な課題として，射影空間などの非ユークリッド空間の可能性を研究し単語共起行列から生成される単語ベクトルのすむ空間の位相構造を特定する必要

がある．

これは，Zipf の法則にみられる言語の冪乗分布・指数分布的な構造や，単語ベクトルにおけるコサイン類似度や単語出現頻度の対数化が有効に機能していることの解明にもつながると考えられる．

第四に，認知科学的な課題として，BERT などの大規模言語モデルにみられるように単語ベクトルの演算は推論などの認知プロセスを実行している可能性があるので，その計算機序を解明することが考えられる．BERT の内部処理において行われている文脈付き単語ベクトルの演算を分析することで，質問応答などの推論を行う演算処理を特定・定式化する．さらに，認知科学のうち脳科学に関わる研究領域において，幾何的・空間的な認知機能の存在が明らかになってきており，単語分散表現を用いて脳内における言語処理プロセスを実験する可能性も考えられる．

これらの課題解決は，単語分散表現による自然言語の計算モデルの構築につながるものであり，そのための継続的な研究は有望であると期待される．

謝辞

修士研究を進めるに際し，終始適切な助言を賜り、また丁寧に指導して下さった本学の日髙昇平先生に深く感謝します。既成観念に囚われることなく自由に思考する機会を与えていただき，また数学という武器を手に自由闊達に研究対象に切り込む楽しさを教えていただきました.

日髙研究室に所属されていた東京電機大学鳥居拓馬先生には，研究上のアドバイスのみならず，研究データの準備や参照すべき先行研究の特定，数学上の問題解決など細やかにサポートいただきました.

社会人を卒業して大学院で研究に専念することを許してくれた妻と二人の娘に感謝します.

参考文献

Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2018). Linear Algebraic Structure of Word Senses, with Applications to Polysemy. *Transactions of the Association for Computational Linguistics*, *6*, 483–495.

Awodey, S. (2010). *Category Theory, Second Edition*. Oxford University Press.

Bengio, Y., Courville, A. C., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, *35*(8), 1798–1828.

Bengio, Y., Schwenk, H., Senécal, J.-S., Morin, F., & Gauvain, J.-L. (2006). *Neural Probabilistic Language Models*, pp. 137–186. Springer Berlin Heidelberg, Berlin, Heidelberg.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, *3*(null), 993–1022.

Bradley, T.-D. (2020). *At the Interface of Algebra and Statistics*. Ph.D. thesis.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language Models are Few-Shot Learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., & Lin, H. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 33, pp. 1877–1901. Curran Associates, Inc.

Church, K. W., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, *16*(1), 22–29.

Cichocki, A. (2013). Era of Big Data Processing: A New Approach via Tensor Net-

works and Tensor Decompositions..

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What Does BERT Look at? An Analysis of BERT's Attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 276–286 Florence, Italy. Association for Computational Linguistics.

Czachor, G., Piasecki, M., & Janz, A. (2018). Recognition of Hyponymy and Meronymy Relations in Word Embeddings for Polish. In *Proceedings of the 9th Global Wordnet Conference*, pp. 251–258 Nanyang Technological University (NTU), Singapore. Global Wordnet Association.

Davey, B. A., & Priestley, H. A. (1990). *Introduction to lattices and order*. Cambridge University Press, Cambridge.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Burstein, J., Doran, C., & Solorio, T. (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Fano, R. M. (1961). Transmission of Information: A Statistical Theory of Communications. *American Journal of Physics*, *29*(11), 793–794.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-55... *1952-59*, 1–32.

Fu, R., Guo, J., Qin, B., Che, W., Wang, H., & Liu, T. (2014). Learning Semantic Hierarchies via Word Embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1199–1209 Baltimore, Maryland. Association for Computational Linguistics.

Ganter, B., Stumme, G., & Wille, R. (Eds.). (2005). *Formal Concept Analysis: Foundations and Applications*. Springer-Verlag, Berlin, Heidelberg.

Ganter, B., & Wille, R. (1999). *Formal concept analysis : mathematical foundations*. Springer, Berlin; New York.

Garvin, P. L. (1962). Computer Participation in Linguistic Research. *Language*, *38*(4), 385–389.

Gastaldi, J. L. (2021). Why Can Computers Understand Natural Language?. *Philos-*

ophy and Technology, *34*(1), 149–214.

Gerstenmaier, J., & Mandl, H. (2001). Constructivism in Cognitive Psychology. In Smelser, N. J., & Baltes, P. B. (Eds.), *International Encyclopedia of the Social Behavioral Sciences*, pp. 2654–2659. Pergamon, Oxford.

Gordon, G., & McNulty, J. (2012). *Matroids A Geometric Introduction*, chap. 5 Finite Geometry. Cambridge University Press.

Harris, Z. (1951). *Methods in Structural Linguistics*. University of Chicago Press, Chicago.

Harris, Z. (1954). Distributional structure. *Word*, *10*(2-3), 146–162.

Hashimoto, T. B., Alvarez-Melis, D., & Jaakkola, T. S. (2016). Word Embeddings as Metric Recovery in Semantic Spaces. *Transactions of the Association for Computational Linguistics*, *4*, 273–286.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.*, *18*(7), 1527–1554.

Jones, M., & Mewhort, D. (2007). Representing word meaning and order information composite holographic lexicon. *Psychological review*, *114*, 1–37.

Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing (3rd Edition Draft)*. Prentice-Hall, Inc., USA.

Jurgens, D., Mohammad, S., Turney, P., & Holyoak, K. (2012). SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pp. 356–364 Montréal, Canada. Association for Computational Linguistics.

Kobayashi, T., & Tanaka-Ishii, K. (2018). Taylor's law for Human Linguistic Sequences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1138–1148 Melbourne, Australia. Association for Computational Linguistics.

Landauer, T., Foltz, P., & Laham, D. (1998). An Introduction to Latent Semantic Analysis. *Discourse Processes*, *25*, 259–284.

Landauer, T. K. (2002). On the computational basis of learning and cognition: Arguments from LSA.. Vol. 41 of *Psychology of Learning and Motivation*, pp. 43–84. Academic Press.

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge.. *Psychological Review*, *104*, 211–240.

Lee, D., & Seung, H. (1999). Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature*, *401*, 788–91.

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, *4*.

Levy, O., & Goldberg, Y. (2014a). Linguistic Regularities in Sparse and Explicit Word Representations. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pp. 171–180 Ann Arbor, Michigan. Association for Computational Linguistics.

Levy, O., & Goldberg, Y. (2014b). Neural Word Embedding as Implicit Matrix Factorization. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, p. 2177–2185 Cambridge, MA, USA. MIT Press.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225.

Lin, D., Zhao, S., Qin, L., & Zhou, M. (2003). Identifying Synonyms among Distributionally Similar Words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, p. 1492–1493 San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development*, *1*(4), 309–317.

Lyons, J. (1968). *Introduction to theoretical linguistics / John Lyons*. Cambridge University Press London.

Lyons, J. (1970). *Chomsky*. Penguin books.

McClelland, J., & Kawamoto, A. (1986). Mechanisms of Sentence Processing: Assigning Roles to Constituents. In *Parallel Distributed Processing. Volume 1: Psychological and Biological Models*. MIT Press.

McRae, K., & Jones, M. (2013). Semantic Memory..

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013a). Efficient Estimation of

Word Representations in Vector Space. In *International Conference on Learning Representations*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013b). Distributed Representations of Words and Phrases and their Compositionality. In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., & Weinberger, K. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc.

Miller, G. A. (1967). Empirical methods in the study of semantics..

Mohammad, S., Dorr, B., & Hirst, G. (2008). 2008) Computing Word-Pair Antonymy.. pp. 982–991.

OpenAI (2023). ChatGPT: Optimizing Language Models for Dialogue.. `https://openai.com/blog/chatgpt/`.

Osgood, C. E. (1952). The nature and measurement of meaning.. *Psychological bulletin*, *49 3*, 197–237.

Padó, S., & Lapata, M. (2007). Dependency-Based Construction of Semantic Space Models. *Comput. Linguist.*, *33*(2), 161–199.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543 Doha, Qatar. Association for Computational Linguistics.

Poon, H., & Domingos, P. (2009). Unsupervised Semantic Parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1–10 Singapore. Association for Computational Linguistics.

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What we know about how BERT works..

Rohde, D., Gonnerman, L., & Plaut, D. (2006). An Improved Model of Semantic Similarity Based on Lexical Co-Occurrence. *Cognitive Science - COGSCI*, *8*.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group (Eds.). (1986). *Parallel Distributed Processing. Volume 1: Foundations*. MIT Press, Cambridge, MA.

Rumelhart, D. E., & Abrahamson, A. A. (1973). A model for analogical reasoning.. *Cognitive Psychology*, *5*, 1–28.

Sahlgren, M., & Lenci, A. (2016). The Effects of Data Size and Frequency Range on Distributional Semantic Models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 975–980 Austin, Texas.

Association for Computational Linguistics.

Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. *Commun. ACM*, *18*(11), 613–620.

Sparck Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*.

Sternberg, R. J., & Gardner, M. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General*, *112*, 80–116.

Torii, T., Maeda, A., & Hidaka, S. (2022). Embedding parallelepiped in co-occurrence matrix: simulation and empirical evidence. In *Joint Conference on Language Evolution (JCoLE2022)*.

Turney, P., & Littman, M. (2005). Corpus-based Learning of Analogies and Semantic Relations. *Machine Learning*, *60*.

van Aken, B., Winter, B., Löser, A., & Gers, F. A. (2019). How Does BERT Answer Questions? A Layer-Wise Analysis of Transformer Representations. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, p. 1823–1832 New York, NY, USA. Association for Computing Machinery.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., & Polosukhin, I. (2017). Attention is All you Need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., & Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.

Weeds, J., Clarke, D., Reffin, J., Weir, D., & Keller, B. (2014). Learning to Distinguish Hypernyms and Co-Hyponyms.. pp. 2249–2259.

Wittgenstein, L., & 丘澤静也訳 (2013). **哲学探求**. 岩波書店.

Yih, S., Zweig, G., & Platt, J. (2012). Polarity Inducing Latent Semantic Analysis..

河田敬義 (1976). **アフィン幾何・射影幾何**. 岩波書店.

丸山圭三郎 (1983). **ソシュールを読む**. 岩波書店.

橋本敬, 稲邑哲也, 柴田智広, & 瀬名秀明 (2010). 社会的知能発生学における構成論的シミュレーションの役割と SIGVerse の開発. **日本ロボット学会誌**, *28*(4), 407–412.

田中久美子 (2021). **言語とフラクタル**. 東京大学出版会.