

# 論文審査の要旨及び担当者

No.1

報告番号	甲 乙 第 号	氏 名	寺岡 丈博
論文審査担当者	主 査	政策・メディア研究科委員	環境情報学部・教授 今井 むつみ
	副 査	政策・メディア研究科委員	環境情報学部・教授 田中 茂範
	副 査	政策・メディア研究科委員	環境情報学部・教授 萩野 達也
	副 査	慶應義塾大学 名誉教授	石崎 俊
学力確認担当者：			
<p>寺岡丈博君から提出された博士論文は、「動詞連想概念辞書の構築とその応用研究—連想に基づいた意味理解システム—」と題し、6章から成る。</p> <p><b>【論文審査の要旨】</b></p> <p>従来から、コンピュータの言語理解機能を向上させて人間に近付け、真に使いやすくするためには、人間が持つ複雑で膨大な知識を体系化し、利用する必要があると考えられてきている。今日までに様々な辞書や言語資源が構築・応用されてきたが、大規模データへの統計的な手法には限界があり、未だに意味理解・文脈理解に関する精度が不十分であることから、従来から集めてきた知識を見直す必要がある。本研究では、意味理解の中心的な役割を担う動詞に関する知識を連想実験によって抽出して体系化して動詞連想概念辞書を構築し、意味理解システムに応用することでコンピュータの言語理解機能が従来よりも向上することを確認した。</p> <p>連想実験では、小学校の国語の教科書で使用されている基本動詞を刺激語として、動詞との意味的な関係を表す深層格の名称とともに呈示することによって、実験参加者は連想語としてこれらの深層格情報を答える形式となっている。この連想実験から得られたデータと線形計画法を用いて、刺激語と連想語の単語間距離を定量化し、連想のし易さを表す連想距離を定めた。これらの刺激語、連想語、そして連想距離を用いて動詞連想概念辞書を構築し、動詞における基本知識を体系化した。</p> <p>動詞連想概念辞書を評価するため、まず始めに、web 文書の共起情報に基づいて構築された既存辞書と共通するデータに対して比較分析を行った。ある動詞に対して、連想し易い語（刺激語動詞に対して連想距離が短い連想語）と共起し易い語（述語動詞のガ格やヲ格などの表層格に入る共起頻度が高い単語）で共通する語の順位相関などを比較した。その結果、上位で連想し易い語と共起し易い語は必ずしも一致せず、これらの語の内容から、連想情報は日常で身の回りの情報が中心であり、共起情報は新聞などの web 文書の情報が中心であることに起因することが示唆された。さらに、反意語関係の動詞「借りる」、「貸す」に関して、連想情報では共通の連想語が同様の順番であり、動作が行われる状況が「借りる」、「貸す」で共通する傾向があったことに対して、共起情報ではそのような傾向が見られず、これらのような点で両者の性質が異なることを示した。</p> <p>次に、動詞連想概念辞書を「省略語の推定」と「換喩表現の検出」に関する意味理解システムへ応用し、従来手法や既存辞書を応用した手法を参考にしたベースラインを用いて評価実験を行った。まず、「省略語の推定」に関しては、動詞連想概念辞書と既存の名詞連想概念辞書を併用して、名詞と動詞の連想情報に基づいて省略語の推定を行うシステムを構築した。共起情報から成る既存辞書を用いた手法をベースラインとし、評価実験を行ったところ、平均逆順位と N 位正解率、そして平均精度の平均において、本研究の提案手法の性能が高く、人間が実際に推測した省略語の内容に最も近いことを示した。また、「換喩表現の検出」に関しては、動詞連想概念辞書と名詞連想概念辞書の他に、アメリカの心理学者 George A. Miller が構築した概念体系の WordNet の日本語翻訳版である日本語 WordNet を用いた。換喩文における用言が動詞の場合はその連想語を使用し、形容詞などの場合はそれを連想語に持つ刺激語を使用し、「文中の名詞」との間の synset（同義語集合）のノード数による概念間距離を計算するこ</p>			

# 論文審査の要旨及び担当者

No.2

とによって、「文中の名詞」が換喩表現かまたはリテラル（字義通りの表現）かを判別するシステムを構築した。評価実験では、先行研究に基づいた手法をベースラインとし、換喩検出精度を比較したところ、提案手法が最も高い精度を示した。

以上のように、「動詞連想概念辞書の構築」では、動詞に関する基本知識を体系化し、既存辞書と比較分析することで連想情報と共起情報の差異を示した。そして、「動詞連想概念辞書の応用」では、「省略語の推定」と「換喩表現の検出」の意味理解システムにおいて既存辞書を用いた手法や先行研究の手法をベースラインとして用いた各々の評価実験を通して有効性を示すことができた。

以下に、本論文における各章の内容を要約する。

第1章 序論として本研究の背景、目的、概要、そして特色について述べる。

第2章 関連研究として「言語資源の構築」と「連想情報を用いた自然言語処理」について、それぞれ従来の研究を説明する。

第3章 動詞連想概念辞書の構築について説明する。基本動詞を刺激語とした連想実験のデータを用いて刺激語と連想語の連想距離を定量化し、動詞連想概念辞書を構築する過程を述べる。また、動詞連想概念辞書と既存辞書を比較した結果から、連想情報と共起情報の違いについても言及する。

第4章 連想情報を用いた省略語の推定について説明する。動詞連想概念辞書と既存の名詞連想概念辞書を統合して用いて省略語推定システムを構築し、既存辞書を用いたベースラインと比較・評価する。提案手法が平均逆順位や平均精度の平均に関して、各ベースラインよりも統計的に有意に高いことを示した。

第5章 単語間の連想情報に基づいた換喩表現の検出について述べる。動詞連想概念辞書と名詞連想概念辞書の他に、日本語 WordNet を用いて換喩自動検出システムを構築した。先行研究の手法と換喩検出の精度を比較・評価し、提案手法が従来手法のベースラインよりも再現率、適合率、F 値でいずれも高く統計的な有意性を示した。

第6章 結論として本研究の成果をまとめ、今後の課題と展望について述べる。

本研究の新規性と成果は、下記のように要約できる。

## 【新規性】

1. 自然言語処理研究では動詞に関する辞書は従来も作成されているが、人間を実験参加者とした連想実験のデータを利用して、人間が持つ動詞の基本知識をまとめたものは、動詞連想概念辞書の他に類が無い。
2. 連想と共起の特徴について、これまで十分に分析・議論されてきていないため、本研究で動詞連想概念辞書と既存辞書を比較分析して示した「連想し易さ」と「共起し易さ」の差異を明らかにした。
3. 既存の名詞連想概念辞書と本研究で構築した動詞連想概念辞書を同時に用いた応用システムによって、従来の研究よりも精度が高い、人間の名詞と動詞に関する連想情報に基づいた新しい手法を提案した。

## 【成果】

1. 動詞連想概念辞書を構築し、既存辞書と異なる性質の情報（連想情報）を持ち、人間の直感に近い情報を表していることを示した。2013年現在、動詞連想概念辞書は刺激語 440 語に対して連想語が延べ約 176,000 語、異なり語が約 36,000 語の規模となっている。
2. 「省略語の推定」と「換喩表現の検出」における応用システムにおいて、既存辞書を用いた手法や従来研究の手法よりも高い精度を実現し、提案手法の有効性を示した。

本研究は、寺岡丈博君が高度な研究遂行能力と当該分野における豊かな学識を有することを示すものであり、その研究成果は自然言語処理研究の今後の発展に大きく貢献することが期待できることを示している。よって、本学位審査委員会は、寺岡丈博君が博士（学術）の学位を受ける資格があるものと認める。