

# 人間により近いAI開発に向けた

## 自然言語処理と認知科学の研究

### 内海 研究室



内海 彰  
Akira UTSUMI

かつて人工知能(AI)の領域では、人間を理解し、人間を模倣することを目指して研究が進められてきました。しかし、次第にAIはさまざまな分野に分かれ、現在ではコンピュータに言語を理解させる「自然言語処理」と、人間の振る舞いを観察してモデル化する「認知科学」の分野はかなり離れた道を歩みつつあります。

こうした背景において、内海彰教授は、AI発展の観点から、自然言語処理と認知科学の二つの領域を行き来しながら研究に取り組

### 本研究室での研究テーマ

**文章分類・クラスタリング**

様々な観点からの文章のグルーピング

**会話エージェント**

雑談会話  
会話モデル

**感情分析・意見マイニング**

ソーシャルメディアからの情報抽出・発見

**文章生成・要約**

要約や見出しの生成、記事の自動生成

**ことばの工学的処理**

**自然言語処理**

**言語認知科学**

ことばの科学的解明

**小説・物語の言語処理**

創造的な文章の作成支援、選書支援  
小説の紹介文やあらすじの生成

**意味空間モデル**

単語の意味の分散表現の学習・解析手法

**比喩・アイロニー**

比喩やアイロニーの言語理論の提案  
心理実験・認知モデリング等による理解や生成過程の解明

**意味記憶・概念獲得**

意味空間モデル、ニューラルネットワークなどを用いた、人間の意味記憶、語彙・概念獲得、記号接地を支える認知過程の解明

んでいます。というのは、「今後人間自身の研究をもっとAIに取り入れていく必要がある、将来は両分野は再び近づくのではないかと考えているからです。内海教授はとりわけ「ことば」に注目し、工学的なアプローチによって自然言語処理の研究を、科学的なアプローチによって言語認知科学の研究を進めています。

**ディープラーニング+自然言語処理**

自然言語処理とは、AIがことばを操ることを可能にする技術であり、この「ことばを操る能力」は人間の知的な振る舞いの本質といえます。現在のAIブームは、音声認識や画像処理に深層学習

(ディープラーニング)が適用されたことをきっかけに広がりましたが、次にディープラーニングの対象になるのが自然言語処理だといわれているのです。

一方で、音声認識や画像処理は信号とその意味との関係が「規則的」であるのに対して、自然言語処理では、記号と意味との関係が「恣意的」であり、そこに固有の難しさがあります。人間は母国語に翻訳すれば意味を理解できますが、コンピュータにはそもそも母国語がなく、言葉の意味を理解することができません。

**多種多様な文章の分析や生成**

内海教授は自然言語処理の分野では、文章を分類・クラスタリン

#### キーワード

言語情報処理、言語認知科学、Web情報検索支援、情報組織化、認知科学、認知心理学、修辞表現、意味空間

所属	大学院情報理工学研究科 総合情報学専攻
メンバー	内海 彰 教授
所属学会	日本認知科学会、人工知能学会、 情報処理学会、言語処理学会、 日本語用論学会
E-mail	utsumi@uec.ac.jp

ほかに SNS 上の発言から感情を分析したり、意見を発掘(マインニング)したりといった研究や、人間と雑談できる「会話エー

### サッカーのマッチプレビューの自動生成

自動生成されたマッチプレビュー記事  
アウェイで行われた 13 節の新潟戦は、柴崎らの 2 ゴールにより 2-0 で勝利。リーグ戦 2 試合で連続で複数得点を挙げるなど、攻撃陣は好調。14 節は大宮との対戦に臨む。大宮とのリーグ対戦成績は、11 勝 6 分け 3 敗で勝ち越している。高いパフォーマンスを見せつけ、攻撃陣の安定した力を示すことができるか。注目選手は、金崎だ。直前の試合では攻撃を牽引する役割を果たした。大宮戦でも彼のプレーが勝敗を左右するだろう。

実際に配信されたマッチプレビュー記事  
13 節の新潟戦は、前半こそシュートゼロと攻めあぐねたが、後半の 2 得点で勝ち切った。ケガ人が多く厳しい状況は続くものの、2 試合連続の完封勝利と、1st ステージ制覇の原動力となった堅守を取り戻しつつあることは好材料。ただし、流れの中からの得点がないため、14 節の大宮戦で改善が求められる。キープレイヤーは金崎だ。新潟戦では PK で、2 節の広島戦以来となるゴールを挙げた。この一撃を皮切りに不調からの完全復活を目指す。

2016年11月リーグ2ndステージ第14節「鹿島アントラーズ vs. 大宮アルディージャ」

グしたり、生成・要約したりする研究をしています。分類では、例えば参加交流型サイト(SNS)上の発言において、その内容が肯定的か、あるいは否定的かといったように、ある観点から似たような文章を自動的にグループ化します。また、文章生成では、要約や見出しの作成を含む、記事の自動生成などが可能です。一例として、サッカーの試合の見どころを事前に配信するマッチプレビューを自動で生成しました。実際に人間が書いたものと遜色のない記事に仕上がっています。

### 小説の紹介文の自動生成

やりたいこと ...

■背景  
オンライン小説など自作の作品を簡単に公開できるようになった

たくさんの人に読んでもらいたくても  
どうやって宣伝すればいいか...

⇒ 紹介文を自動で作れたら便利かも(紹介文生成システム)



たとえば ...

物語文

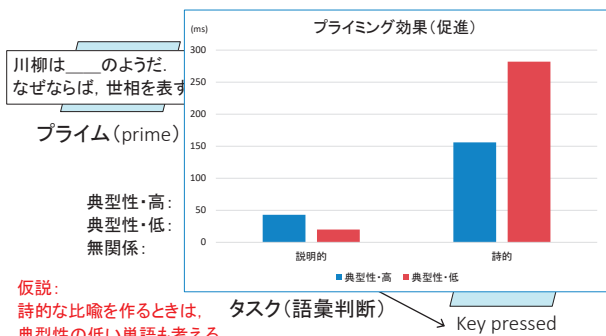
アガサ・クリスティ『オリエント急行の殺人』  
シリアの冬の朝、五時。アレクボ駅のホームに列車が入っていた。鉄道のガイドブックにも出ていない列車で、(タウルス急行)という堂々たる名...

生成システム  
人物、事件の内容、作品の雰囲気... 抽出した情報を組み合わせて紹介文を生成する!

紹介文

真冬の欧州を走る豪華列車オリエント急行には、国籍も身分も様々な乗客が乗り込んでいた。奇妙な雰囲気にも包まれたその車内で、いわずと知れた老富豪が無残な刺殺体で発見される。偶然乗り合わせた名探偵ポアロが捜査に乗り出す。すべての乗客には完璧なアリバイが...ミステリの魅力が詰まった永遠の名作の新訳版。

### 比喩生成過程は目的に依存する



仮説:  
詩的な比喩を作るときは、典型性の低い単語も考える

ジェント」なども開発しています。また特徴的な研究として、小説や物語など創造的な文章の言語処理も手がけています。これは、文庫本の裏表紙にあるような小説のあらすじの自動生成を目指したもので、物語の全体を見てストーリーの展開に関係する重要な部分を抜き出してまとめます。現在はオンライン小説が数多く発表されており、こうした紹介文の自動生成システムは、今後活用されていくかもしれません。

比喩や皮肉の研究  
一方、認知科学とは、人間が知的な処理(認知)を、どのようなメカニズムで行っているかを科学的に探求する学問です。言語を対象とする認知科学を特に言語認知科学と呼び、「自然言語処理の難しさを克服する上でも必要である」ことから、内海教授は言語認知科学の分野でも研究しています。

リングによって人間がそれをどのように生成したり、理解したりしているのかを解明する研究です。比喩の生成では、典型的な単語を選ぶことが多い「説明的な比喩」に対し、「詩的な比喩」を作る際には、典型性の低い単語も併せて考える」という仮説を立て、実際に詩的な比喩の場合は、人間は幅広い単語を考慮していることを実験的に示しました。つまり、比喩の生成過程は目的に依存することが分かったのです。

言葉の意味を数値化する  
また、人間は単語の意味をどのように獲得するのかを明らかにする研究にも着手しました。単語の意味の分散的な表現を学習したり、解析したりするための意味空間モデルやニューラルネットワークなどを用いて、人間の意味記憶や、語彙・概念の獲得、記号接地を支える認知過程の解明を試みています。言葉の意味を「数値化」し、機械学習の手法などで扱えるようにするのが目的です。

内海教授が目指す研究のゴールは、人間のような知的な振る舞いをする人工物の実現であり、それは現在の AI からは想像もつかない姿になるかもしれません。こうした未来の AI の実現に向けて、「認知科学の知見が、これから AI や自然言語処理の分野にどんどん生かされていくことを内海教授は期待しています。」

【取材・文】藤木信穂