

## 情報科学発展演習：自然言語解析器を用いた実験

### ◆ 演習概要

本科目は、「自然言語解析」の演習版です。したがって履修者が同期開講の自然言語解析を履修していることを前提としています(履修していないと以下の説明の内容を理解することは困難です)。

本演習では、当研究室で研究・開発中の以下の日本語自然言語解析器((1)～(5))の出力結果を用いた演習を行います。

(1)形態素解析器、(2)形態素結合器(chunker)、(3)係り受け解析器、(4)構文解析器、(5)照応解析器

以下の講義計画に記したように、それぞれの解析機構について詳細な説明をします。理工学分野の文書中にテキスト(例えば 1000 文)を対象に、当研究室設置の PC で(1)～(3)を実行し、出力結果(ファイル)を提示します。

現時点では、(1)～(2)の解析精度は 90%±ですので、それぞれの出力結果は 100±の誤解析を含むことになります。(1)～(3)それぞれについて、これらの誤解析文のうち、50～100 文(／1 人)について受講者が誤解析結果を分析し、その要因の検討、改善案の提示が課題です。分析に際しては、同一のテキストに対してこの分野の研究で広く利用されている解析ツール(Mecab(1)、Cabocha(3)、等)でも実行して、その結果を対比することを期待します。この要因分析-改善案提案によって、日本語文の特性、日本語文解析器の問題点について理解を深めることが本演習の主眼です。

### ◆ 講義計画(2023 年度)

期日 回数 講義内容

5.25 第 1 回 言語の形式的モデルⅡ

6. 1 第 2 回 演習：形態素解析器  
演習課題

6. 8 第 3 回 演習：形態素結合器  
演習課題

6.15 第 4 回 演習：係り受け解析器  
演習課題

6.22 第 5 回 演習：構文解析器

6.29 第 6 回 演習：照応解析器

7. 6 第 7 回 演習課題の発表