

アノテーションデータを用いた 実証的計算心理言語学

浅原正幸

理論・対照グループの基幹型共同研究プロジェクト

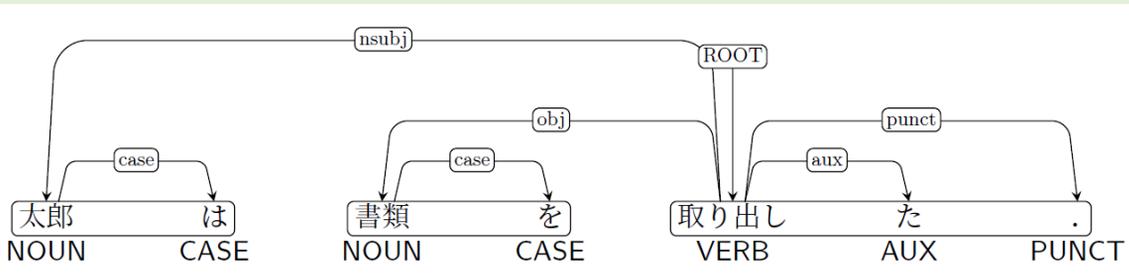
「実証的な理論・対照言語学の推進」のサブプロジェクトとして

「アノテーションデータを用いた実証的計算心理言語学」を進める。

本プロジェクトでは、統語班・意味班・認知班に分かれて以下の研究を行う。

統語班

Universal Dependencies 日本語コーパスの維持・管理



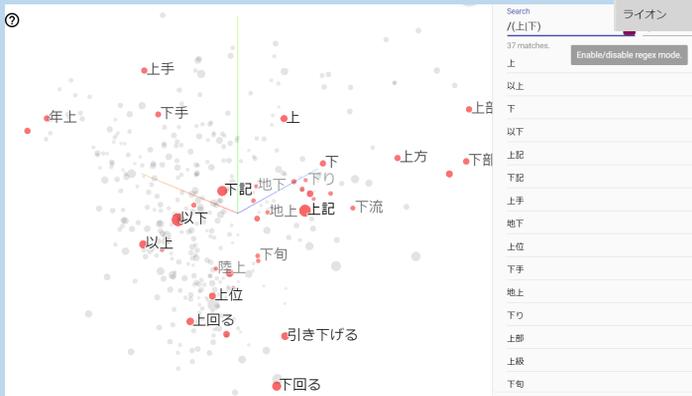
言語横断的に品詞・形態論情報・依存構造を
アノテーションする枠組

Universal Dependencies の

日本語の言語資源を整備するとともに
継続的な公開のための維持作業を行う

意味班

Surface	分類番号	類	部門	中項目	分類項目	知っている (f00)
ゴリラ	1.5501-03-06-04	体	自然	動物	哺乳類	2.19434031554613
コアラ	1.5501-02-04-01	体	自然	動物	哺乳類	2.16988931425696
哺乳類	1.5501-01-03-01	体	自然	動物	哺乳類	2.16628019642134
やぎ	1.5501-11-05-04	体	自然	動物	哺乳類	2.13665095336729
ライオン	1.5501-05-05-03	体	自然	動物	哺乳類	2.1223051693044



分類語彙表に基づく語義情報を
アノテーションするとともに
文脈化単語埋め込み技術を用いて
語義の距離・粒度などを定量的に
評価する

さらに、分類語彙表の語義に基づく
比喩表現データベースの構築を行う

認知班

印象評定・読み時間など
言語受容過程のデータベース
を構築する

心理言語実験サーバを構築し
共同研究者に共用する

