

CEFR Companion Volume
に対応した日本語例文自動分類のアルゴリズム開発
(研究の進捗発表)

メタデータ	言語: jpn 出版者: 公開日: 2020-05-13 キーワード (Ja): キーワード (En): 作成者: 安藤, 聖野, 宮崎, 佳典, 谷, 誠司, 安, 志英 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00027457

CEFR Companion Volume に対応した日本語例文自動分類のアルゴリズム開発

安藤 聖野(情報科学科), 宮崎佳典 (学術院情報学領域) , 谷誠司 (常葉大学外国語学部) , 安志英 (群山大学)

2017 年に CEFR(言語能力の国際基準)を補完する CEFR Companion Volume が公開され, PreA1 レベルが追加された. 一方で日本語教育に向けた CEFR の研究は少なく, 日本語 CEFR 準拠テキストコーパスも管見の限り存在しない. 本研究では, コーパスを作成する際, 例文に CEFR の読解力を表す CDS(能力記述文)を付与する労力を軽減すべく, その自動分類の実装に継続的に取り組んでいる. 分類アプローチには SVM を適用し, 例文の特徴量として文長, 専門性, 文書タイプ, 漢字率を与える. 文書タイプと専門性は fastText を用いて自動推定する. PreA1 レベルは従前のレベル群と難易度に大きな差があり, 全レベルの CDS を一度に分類する先行研究に対し, 2 段階による CDS 分類を試み, 結果の比較や考察を行う.