

学校英文法コーパスの提案 —デザインと応用可能性—

小林雄一郎* 田中省作** 後藤一章‡ 徳見道夫‡‡ 朝尾幸次郎‡‡
 †法政大学 ‡立命館大学 ‡大阪大学 ‡九州大学
 * kobayashi0721@gmail.com

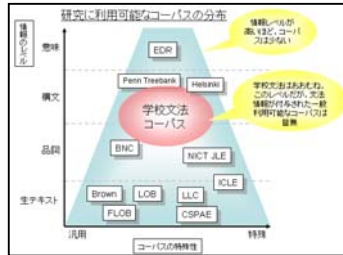
本研究の背景

◆大規模コーパスにおける情報付与

- 品詞情報
- 構文情報
- 意味情報 etc.

◆学校英文法情報

- 日本人にとって重要な理解の観点
- 教育分野での高いニーズ
- 付与された(研究利用可能な)コーパスは皆無



関連研究

◆佐野・猪野 (2000)

- 中学高校の英語教科書や市販の文法書を精査
- それらの難易度に関する順序関係
- 教材の難易度計算の枠組み

◆N-Cube(文法項目別用例集)

- 1320の文法項目
- BNCから用例抽出
- 表層・品詞レベルでの記述
- 正確な精度保証はない

精度保証が可能な
文法情報付き
コーパスの必要性

構築方針

◆使用データ

- Penn TreebankのBrown Corpus部分から約4000文
- 中高の教科書コーパス etc.

今後も
議論・改訂を継続

◆情報付与

- 表層形や品詞・構文情報に基づく automatic tagging
- 複数の作業者による manual tagging

文型	1-5文型	分詞	限定・叙述
文の種類	平叙・疑問・否定・感嘆	動名詞	
文の単複など	単文・重文・複文(・混文)	助動詞	
疑問文の種類	一般・特殊・選択・間接・付加	接続詞	等位・従属
極性	肯定・否定	疑問詞	
否定の種類	全否定・部分否定	疑問詞+to+V	
語法	直接・間接	関係代名詞	主格・目的格・所有格
時制	現在・過去・未来	関係代名詞の種類	制限・継続
時制の一致	一致・不一致	関係副詞	
態	能動・受動	複合関係詞	代名詞・副詞
法	直接・仮定・命令	数量表現	
相	進行・完了	倒置	
to不定詞	名詞的・形容詞的・副詞的	比較級+比較級	
原形不定詞		存在there構文	
独立不定詞		分詞構文	主語の一致・主語の不一致・慣用表現
比較(形容詞)	原級・比較級・最上級	強調構文	
比較(副詞)	原級・比較級・最上級		

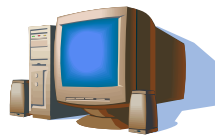
* 個々の項目の定義については、綿貫ほか(2000)を参照

言語処理

◆学校英文法に基づいた英文処理

- 英文に含まれる文法項目を自動検出

Written in a clear hand,
this report is easy to read.



分詞構文 + 受動態

(Being) Written in a clear hand, this report is easy to read.

To不定詞(名詞的用法)

◆田中ほか(2008)

- 機械学習を活用した文法項目の検出
- 広域的情報が必要な項目(仮定法・分詞構文)を一定精度で同定

言語教育

◆辞書・文法書の編集

- 文法レベルの頻度情報の提示
- 文法項目による例文検索



◆教材の開発・評価

- リーダビリティ
- 教材の文法制限

◆学習者言語の評価

- 文法項目の習得研究
- 英作文の採点支援



学習者コーパス研究
(英語教育・第二言語習得との
インターフェイス)

今後の展開

- ◆文法項目の精査とコーパスの拡大
- ◆文法項目の自動検出
- ◆2009年春プロトタイプ公開予定 etc.

主要参考文献

- [1] 佐野洋・猪野真理枝(2000)「英語文法の難易度計測と自動分析」『情報処理学会コンピュータと教育研究会(CE)報告』No. 117 (pp. 5-12).
- [2] 綿貫陽・宮川幸久・須貝猛敏・高松尚弘(2000)『ロイヤル英文法』(改訂新版)旺文社.
- [3] 田中省作・小林雄一郎・徳見道夫・朝尾幸次郎(2008)「学校英文法コーパス構築の試み」『2008年度人工知能学会全国大会(第22回)論文集』CD-ROM.