

自己評価報告書

平成 23 年 4 月 30 日現在

機関番号：17102

研究種目：基盤研究(C)

研究期間：2008～2012

課題番号：20520504

研究課題名(和文) 英文法コーパスの構築とその応用

研究課題名(英文) Constructing School Grammar Corpus of English and its Application

研究代表者

徳見 道夫(TOKUMI MICHIO)

九州大学・大学院言語文化研究院・教授

研究者番号：90099755

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：コーパス, 学校英文法, 機械学習, 自然言語処理, 著作権

1. 研究計画の概要

本研究は、英語を対象として、従来、最も重要な観点の一つであるにもかかわらず見過ごされてきた学校英文法に関する情報を付与した実用的なコーパス(文法コーパス)を整備する(目的1)ものである。ただ単に文法コーパスを整備するのではなく、並行して部分的な構文構造まで考慮した英文からの文法項目の自動検出ルールも蓄積する(目的2)こと、さらに文法コーパスや検出ルールを用いた応用研究(目的3)(文法の統計的分析、教材評価への応用など)を進める。これらの3つのテーマは、互いに相乗的に進展する仕掛けとなっており、コーパス構築の新しい方法論の提案でもある。

2. 研究の進捗状況

本研究は、主に(1)学校英文法コーパスの整備・公開、(2)文法項目の検出ルールの蓄積、(3)これらの応用研究、が軸となる。それぞれの進捗状況を以下に述べる。

(1) 学校英文法コーパスの整備と公開

初年度に学習指導要領・検定教科書・参考書および従来研究を精査し、学校英文法コーパスの基本的な設計(コーパス・デザイン)を与えた。そのうち、主に次の二つの資料をベースに学校英文法コーパスの構築を進めた。

一つは、主に高校～大学生レベルの英語参考書の例文である。特定の文法項目の特徴的構造を見出すには都合が良い。また、参考書中の出現箇所から部分的な文法情報は容易に付与できる、という利点もある。このよう

な例文については、現在、1.5 万文強を電子化し、そのうち完全な文法情報(不正確という意味ではなく、文中の全ての文法情報の使用の是非が付与されている、という意)が付与されているものが4千文弱になる。

もう一つの資料は、代表的な構文解析済み英文コーパスの一つである Penn Treebank の一部(Brown Corpus 部分)で、比較的広い構文情報によって同定される文法項目(たとえば、仮定法や分詞構文など)を約5千文に対して付与した。

これらのコーパスの公開方法について検討し、文法情報部分のみを公開することとした。また、ユーザが本コーパスを容易に再現できるよう、そのためのプログラム群を整備している。

(2) 文法項目の検出ルールの蓄積

機械学習を活用した、文法項目の検出ルールの自動的な記述、形態素レベルで記述される検出ルールの人手による蓄積を進め、それぞれの特徴、文法項目ごとの精度等を見積もった。また、応用研究で求められた n-gram のような部分表現からの文法項目の検出についても検討を開始した。

(3) 応用研究

ESP コーパスからの特徴表現の抽出や、母語話者/非母語話者が書く論文の相違を明らかとするために、(2)のルールの活用を進めた。

3. 現在までの達成度

②おおむね順調に進展している。

学校英文法コーパスの構築は、前項目のとおりで概ね順調であるが、次の3点について

は、やや遅れている。ただし、次項目で述べるように平成 23 年度に比較的容易に達成されるため、上記のように判断した。

- (1) 構築したコーパス全てを公開できていない。
- (2) 文法項目の検出ルール of コーパス構築支援への適用が未着手である。
- (3) 英文章の多様性という点では、まだ不十分である。

4. 今後の研究の推進方策

前項で述べた未達成の 3 点を進める予定であるが、特に(1)は既に一部コーパスを公開し、その形式・方法等は確定しているので大きな困難はない。また、(3)にはフリーの中高英語の副読本を既に確保しており、文法検出ルールを積極的に活用することで、(2)も同時に解消されることが期待される。

なお、応用研究については枚挙すれば限りがないため、本科研関係者がかかわっている研究課題に大胆に絞り、専念する予定である。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 12 件)

- ①徳見道夫, 田中省作: 標準化テストと九州大学における英語教育, 大学教育, 16 (in press), 査読無
- ②徳見道夫, 田中省作: 英文法コーパス構築の有用性, 言語科学, 46, 61-74 (2011), 査読有
- ③Kitao, K. and Tanaka, S.: Characteristics of Japanese Junior High School English Textbooks, 文化情報学, 4, pp.1-14 (2009), 査読有
- ④神谷健一, 田中省作, 北尾謙治: 言語処理技術と教材作成の連携, 自然言語処理, 16(2), 45-58 (2009), 査読有
- ⑤田中省作, 小山由紀江: 構文情報を考慮した ESP コーパスからの特徴表現の抽出, 統計数理研究所共同研究レポート, 239, 13-30 (2009), 査読無

[学会発表] (計 16 件)

- ①田中省作, 冨浦洋一, 安東奈穂子, 柴田雅博: Web を源とした英語科学論文コーパスの構築法 –技術的方法論と法的観点からの検討-, 英語コーパス学会第 34 回大会, 青山学院大学(2009.10.3)
- ②Kitao, K. and Tanaka, S.: Authorized Junior High School English Textbooks in

Japan: a Corpus-based Study of Vocabulary Level and Readability, EuroCALL 2009, スペイン (2009.9.11)

- ③田中省作, 小山由紀江: 日本の英語教科書コーパスを基準とした ESP 特徴表現の抽出, 外国語教育メディア学会第 49 回全国研究大会, 流通科学大学 (2009.8.6)
- ④小林雄一郎, 田中省作, 後藤一章, 徳見道夫, 朝尾幸次郎: 文法情報の自動検出技術を用いたリーディング教材の作成と評価, 語彙研究フォーラム 2008 第 1 回 JACET リーディング研究会・英語語彙研究会合同研究大会, 関西学院大学(2008.12.6)
- ⑤田中省作, 小林雄一郎, 徳見道夫, 朝尾幸次郎: 学校英文法コーパス構築の試み, 人工知能学会第 22 回全国大会,ときわ市民ホール (2008.6.12)

[その他]

朝尾幸次郎: 英語の遊び唄, 月例英語教育講座 (2010)