

令和 4 年 6 月 10 日現在

機関番号：17102

研究種目：若手研究(B)

研究期間：2017～2021

課題番号：17K17943

研究課題名（和文）メディアと場面の相互連関に基づく英語定型表現の分析とインデックス化

研究課題名（英文）Cross-sectional Analyses and Extraction of English Formulaic Expressions Based on Media and Situation

研究代表者

土屋 智行 (TSUCHIYA, Tomoyuki)

九州大学・言語文化研究院・准教授

研究者番号：80759366

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究課題では、旅行課題遂行データベースおよび大規模テキストコーパスを利用し、定型表現およびその下位カテゴリの表現を、使用場面や使用状況、コミュニケーション形態ごとに抽出・整理し、場面に結びついた定型表現の具体例やコミュニケーションにおける流暢性や創造性の発揮に関わる言語の慣習性的一端を明らかにした。

研究成果の学術的意義や社会的意義

(1) 本研究は、定型表現およびその下位カテゴリの表現の記憶と使用に関して、認知・社会的な観点から分析・考察し、コミュニケーションの流暢性や効率性を向上させるために人が意識的・無意識的におこなっている言語表現の記憶と使用の一端を明らかにしている点で学術的意義がある。

(2) また本研究成果は、定型表現およびその下位カテゴリとなる表現のより客観的な抽出の手法を提案し、英語教育、特に流暢性の向上や場面に適した表現集の開発に資する点で社会的意義がある。

研究成果の概要（英文）：The present research project, using the Database of Travel Planning Task Dialogues and text corpora, extracted and classified formulaic expressions and their subcategories by scene of use, situation of use, and form of communication, and clarified specific examples of formulaic expressions linked to scenes and some of their communicative characteristics.

研究分野：コーパス言語学

キーワード：定型表現 慣習的な表現 構文 Collostruction分析 Multi-word expression n-gram 旅行課題遂行データベース

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 日常的な言語使用から慣習が形成され、その慣習が定型的な表現として個々人の日常的な言語使用に現れる。この定型的な言語表現(定型表現、formulaic expression)はどのような場面でもどのように使用されるのかについて体系的に記述、分析されていない。

(2) コミュニケーションのためのメディア(対面、SNS やメール等)が多様化している現代において、慣習が形成される過程やメディア間の関係性はより複雑化しており、理論的・体系的に論じることが困難となっている。

(3) 言語の慣習と定型性および創造性という観点から、定型表現の下位カテゴリである Multi-word expression、構文(construction)、連語(collocation)、コロストラクション(collostruction)、慣習的な表現等にそれぞれ焦点を絞り、どのような表現が確認され、各表現がどのような文脈・状況で使用されるのかについて記述・分析される必要がある。

(4) 研究代表者による旅行課題遂行データベース(科研費、若手(B)、課題番号 15H06462)を用い、会話の場面や使用するメディア(face-to-face, phone, mail)ごとに分析を進め、定型表現の研究を発展させる必要がある。

2. 研究の目的

(1) 多様なデータ(実際の会話データやテキストデータ)から英語定型表現を抽出し、それらをコミュニケーション形態および場面の観点から整理・分析・考察する。

(2) コミュニケーション形態や場面と定型表現の関係性を認知・社会的な観点から考察し、その理論的な体系化と英語教育への応用に向けた基盤を固める。

3. 研究の方法

(1) 旅行課題遂行データベースは、4名の英語話者(3名英語母語話者、1名非母語話者)を旅行者役(traveler)と代理業者役(agent)に分け、お互いに日本国内の旅程決定課題を対面(face-to-face, f2f)、電話(phone)、メール(mail)でおこなった様子の映像と音声を収録したデータである。このデータベースから抽出したテキストデータから n-gram 列のデータを生成し、複数の分析をおこなった。

パターン束的手法: 抽出した n-gram 列を再帰的に空所化したパターンを集計し、使用メディアごとに傾向を分析した。

他の大規模コーパスの n-gram 列データとの比較: BNC の話し言葉サブコーパスおよび COCA の n-gram 列データの頻度データと結合し、比較した。

会話データの時間情報を用いた慣習的表現の密な区間と疎な区間の分析: 会話の開始時点を 0、終了時点を 1 とし、その間の n-gram 列における前項 b の頻度データの情報から、慣習的な表現が集中している区間とそうではない区間を抽出した。

(2) 新型コロナウイルスにより会話データの収集が困難となったため、現存する大規模英語コーパスのデータを用いて英語談話標識や構文の形成に関する分析と考察をおこなった。

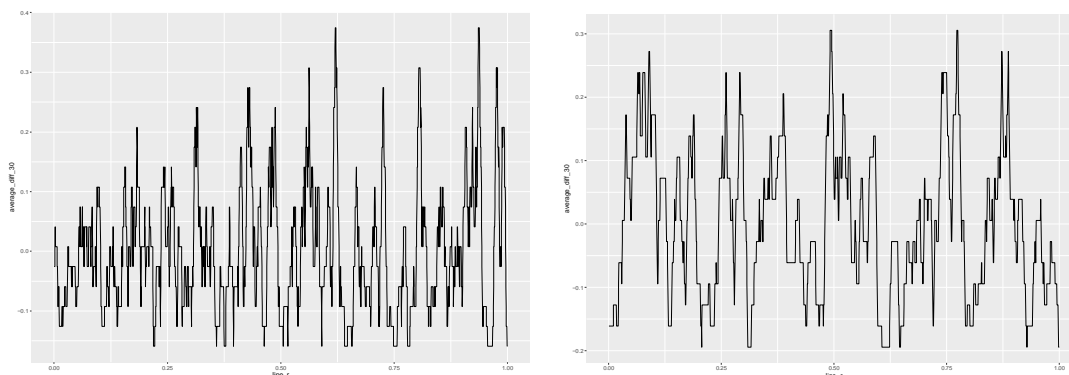
定型性の高い英語談話標識の分析: Sibol corpus of English broadsheet newspapers から文頭に出現する Multi-word expression (複数語がひとまとまりとして慣習化した表現)を抽出し、また近年談話標識化した表現を探索した。

より客観的な構文の抽出方法に関する分析: EnTenTen コーパスから高頻度の動詞(do, say, use, make, include)とその前後の単語を収集し、動詞を中心とした前文脈と後文脈を交差的に分析した。

4. 研究成果

(1) メディアごとに特徴的な定型表現の抽出: 前節(1)において agent のデータを分析した結果、メディアごとに特徴的な n-gram 表現を抽出できた。例として f2f では *if you want to go to...*、phone では *(so) I'll (I will) send you a list of..., would be interested in..., could I take your email address..., would you like to...*、mail では *I look forward to hearing from you..., do our best to..., I would strongly recommend, you and your...* 等の表現が挙げられる。

(2) 慣習的な表現が密な区間と疎な区間の分析: 前節(1) および(1) の分析によって、会話は慣習的な表現が集中する区間と集中しない区間があることを確認した(図1、縦軸は慣習的な表現の集中の度合いを、横軸は会話の経過を表す)。また、会話の冒頭や終了時点で使われる傾向が強い慣習的な表現を抽出した。会話の冒頭では *how can I help you, at this time of the year, at the moment*、等が、会話の終了時点では *is there anything else that..., have a nice day* 等が確認できた。また、慣習的な表現が集中しない区間が必ずしも新奇もしくは創造的な内容を発話しているとは限らないことを示した。



a. agent の発話の例 (f2f_agent_01)

b. traveler の発話の例 (f2f_traveler_01)

図 1: 慣習的な表現の使用状況の推移

(3) 近年使用頻度が上がった談話標識の抽出: 前節(2) の分析から、近年談話標識として使われるようになった表現を抽出することができた。例として *for me, what's more, in truth, that said, and yet, either way* が確認された。

(4) 構文を客観的に抽出する手法の提案と検討: 前節(2) の分析から、慣習的な表現の中でも抽象的なスロットが存在する構文を抽出した。具体的には to do anything/something about..., less/little/lot/more/much to do with, the appropriate/continued/effective/efficient/exclusive/proper/safe/... use of 等が挙げられる。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 0件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Tomoyuki Tsuchiya	4. 巻 27
2. 論文標題 Extracting and Analyzing English Multi-word Expressions with Slots: A Case Study of 'take'	5. 発行年 2021年
3. 雑誌名 言語処理学会第27回年次大会発表予稿集	6. 最初と最後の頁 1134-1137
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 土屋智行	4. 巻 24
2. 論文標題 網羅的なパターン束からみる構文の状況依存性	5. 発行年 2018年
3. 雑誌名 言語処理学会第24回年次大会発表論文集	6. 最初と最後の頁 584-587
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 土屋智行	4. 巻 35
2. 論文標題 言語環境に応じた言語知識の活性化	5. 発行年 2017年
3. 雑誌名 日本英語学会第35回大会 Conference Handbook	6. 最初と最後の頁 185-190
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 土屋智行	4. 巻 69
2. 論文標題 定型表現のコーパス分析からみる構文スキーマの形成プロセス	5. 発行年 2017年
3. 雑誌名 英文学会支部大会 Proceedings	6. 最初と最後の頁 335-336
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計10件（うち招待講演 6件 / うち国際学会 2件）

1. 発表者名 Tomoyuki Tsuchiya
2. 発表標題 Extracting and Analyzing English Multi-word Expressions with Slots: A Case Study of 'take'
3. 学会等名 言語処理学会第27回年次大会
4. 発表年 2021年

1. 発表者名 Tomoyuki Tsuchiya
2. 発表標題 Imaginary Situation in Japanese SNS Quotes: The Case of "A High School Girl in McDonald's" and "Carl Loffler"
3. 学会等名 Workshop on Formulaicity in Discourse
4. 発表年 2019年

1. 発表者名 土屋智行
2. 発表標題 言語の慣習性を中心とした言語研究の手法と展開
3. 学会等名 日本言語教育ICT学会2019年研究大会（招待講演）
4. 発表年 2019年

1. 発表者名 土屋智行
2. 発表標題 構成要素数にとらわれないパターン束分析の手法の検討
3. 学会等名 語用論学会九州山口地区研究会（招待講演）
4. 発表年 2018年

1. 発表者名 Tomoyuki Tsuchiya
2. 発表標題 Euphemistic Use of NPs in Japanese Proverbs
3. 学会等名 Referentiality Workshop (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 土屋智行
2. 発表標題 文頭に出現するMWEの談話標識らしさ
3. 学会等名 日本英文学会九州支部第71回大会 (招待講演)
4. 発表年 2018年

1. 発表者名 Tomoyuki Tsuchiya
2. 発表標題 Processing Conventionalized Sequences
3. 学会等名 No. 150 Shonan Meeting (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 土屋智行
2. 発表標題 網羅的なパターン束からみる構文の状況依存性
3. 学会等名 言語処理学会第24回全国大会
4. 発表年 2018年

1. 発表者名 土屋智行
2. 発表標題 言語環境に応じた言語知識の活性化
3. 学会等名 日本英語学会第35回大会（招待講演）
4. 発表年 2017年

1. 発表者名 土屋智行
2. 発表標題 言語的な慣習の探索手段としてのテキストマイニング
3. 学会等名 第9回異分野融合テキストマイニング研究会
4. 発表年 2017年

〔図書〕 計2件

1. 著者名 土屋 智行	4. 発行年 2020年
2. 出版社 ひつじ書房	5. 総ページ数 152
3. 書名 言語と慣習性	

1. 著者名 大賀哲・内田諭・中藤哲也・石田栄美・土屋智行・他	4. 発行年 2019年
2. 出版社 ひつじ書房	5. 総ページ数 268
3. 書名 異分野融合研究のためのテキストマイニング：基礎と実践	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------