

網羅的なパターン束分析からみる構文の状況依存性

土屋 智行
九州大学

tsuchiya@flc.kyushu-u.ac.jp

1 はじめに

Iwasaki (2015) は、話し手が異なる状況に応じて言語表現パターン（構文、または文法）を使い分けている実態を指摘し、そのパターンの使い分けを理論的に考察している。この理論を言語教育に応用する場合、それぞれの状況に応じて良く使用される表現パターンを抽出・整理する必要があるが、コミュニケーションの状況の分類や表現パターンの抽出方法について一貫した見解は提示されていないのが現状である。

本論文では、英語の会話内容から作成した 4gram リストに対し、パターンマッチング処理を用い、抽象的な表現パターンへと網羅的に変換し、それらのパターンを全体的、およびコミュニケーション形態別に集計し、分析をおこなう。

2 分析

本論文では、土屋 (2017) で作成した「旅行課題遂行会話データベース」を利用し、旅程を決定する英語話者の会話からコミュニケーション形態 (ex. 対面会話、電話、メール。以降「メディア」) に特有の表現パターンの抽出を試みる。

2.1 分析データ

本論文では、分析用のデータとして「旅行課題遂行会話データベース」の一部を利用した。

「旅行課題遂行会話データベース」は、著者の研究プロジェクトの一環¹として収録・構築したものであり、英語話者 4 名による、異なるメディアを利用した会話を収録したものである土屋 (2017)。データベースの基本構成は、以下のとおりである。

- **協力者:** 英語母語話者 3 名 (Ted, Mitch, Finn) と英語非母語話者 1 名 (Jon) の 4 名
- **時間:** 合計約 10 時間 (自然発話約 40 分, 課題遂行会話約 9 時間)
- **メディア:** 対面 (合計約 2 時間), 電話 (合計約 2 時間), メール (5 時間)
- **収録回数:** 協力者ごとに、対面・電話・メールでのやりとりを各 3 回, 合計 9 回 (全 18 回)
- **内容:** 「印象に残った旅」に関する自然発話 (予備収録), 日本国内の旅行計画決定に関する課題遂行会話 (本収録)
- **課題内容:** 協力者を代理業者役 (Mitch, Finn) と旅行者役 (Ted, Jon) に分け、日本国内の旅行計画を相談し、決定

本論文では、代理業者役を担当した 2 名によるすべてのやり取りの書き起こしデータを使用した。データに関する情報は表 1, 表 2 の通りである。

表 1: 課題遂行会話の収録時間

メディア	収録回数	各回の収録時間	合計時間
対面	6 回	20~25 分	約 2 時間
電話	6 回	15~25 分	約 2 時間
メール	12 回	20~25 分	約 5 時間
(送・返信各 6 回)			

表 2: 各データの語数

	旅行者	代理業者	合計
対面	6,221 (14.5%)	14,616 (34.1%)	20,837 (48.6%)
電話	4,263 (9.9%)	12,328 (28.7%)	16,591 (38.7%)
メール	1,953 (4.6%)	3,536 (8.2%)	5,489 (12.8%)
合計	12,437 (29%)	30,480 (71%)	42,917 (100%)

2.2 分析手法

本論文では、代理業者役のやり取りのみを分析の対象とした。したがって、全体の約 7 割を分析すること

¹JSPS 科研費 (研究活動スタート支援, 「言語環境の違いによる英語定型表現の選択と調整に関する研究」)

となる。データの分析にあたって、以下の手続きをおこなった。

- (1) コーパスから 4gram のリストを作成
- (2) リストの各表現の各語を再帰的に空所化（空所化の結果=パターン）
- (3) すべてのパターンの継承関係を記述
- (4) より多くの表現と継承関係をもつパターンを z スコアにもとづいて抽出

このうち、(2)-(3) の手続きは、PL 分析用ツールである `rubyp1b` を利用した。PL 分析のの要点を以下に記す。

■ PL 分析

Kuroda (2009) の Pattern Lattice Model は、きわめて具体的な言語表現の記憶を想定した言語理論であり、発話の産出に有効な構文をボトムアップ的に抽出していく分析手法をとる。本論文の PL 分析は、この分析手法を援用したものであり、特定の発話・発信者が特定の文脈で多く利用している表現のうち、一部が抽象化されているものの抽出を可能とする。

たとえば、(5) の 3 つの発話の PL 分析をおこなうと、それぞれの発話の一部が再帰的に抽象化され、共通するパターン同士でまとめあげられることで、図 2.2 のようなラティス図が生成される。

- (5) he calls me mom/he makes me laugh/she calls me babe

ボックス下部の数値は、出現頻度と z スコア（標準偏差）を示し、値が高いほどボックス内のパターンが構文として説明力の高いものであることを示している。また、各パターンの z スコアの高さは、図中のボックスの色温度によって視覚的に示される（色が濃いほうが高い）。

本分析では、発話全体の 4gram リストを全て PL 分析にかけ、生成された dot ファイルからテキスト情報のみを抽出し、集計をおこなった。なお、本論文では、生成された n -gram リストの処理の規模の関係上、4gram の PL 分析のみをおこなった。

2.3 抽出されたパターンの検討

表 3 をみると、すべてのコミュニケーション形態で共通して高い z 値を示している表現パターンを構成する語の多くは、前置詞 (ex. *and, of, to, in, etc.*) と代名

表 3: 代理業者役の発話全体における構文の出現頻度 (括弧内は z 値, $z \geq 5$)

exp	f2f · agent	phone · agent	mail · agent
and ... you ...	54 (10.57)	34 (7.07)	20 (9.47)
to ... you ...	32 (6)	22 (4.32)	12 (5.14)
you ... have ...	28 (5.16)	26 (5.24)	14 (6.23)
you ... to ...	116 (23.48)	106 (23.56)	12 (5.14)
a ... of ...	98 (19.73)	74 (16.23)	28 (13.79)
the ... of ...	84 (16.82)	84 (18.52)	24 (11.63)
in ... of ...	36 (6.83)	50 (10.73)	12 (5.14)
to ... and ...	32 (6)	30 (6.15)	12 (5.14)
for ... and ...	32 (6)	24 (4.78)	14 (6.23)
and ... can ...	28 (5.16)	36 (7.53)	12 (5.14)
the ... of the	13 (18.85)	6 (8.62)	3 (5.83)
we can ... you	10 (14.09)	8 (12.15)	3 (5.83)
there is a ...	14 (20.44)	5 (6.86)	3 (5.83)
as well as ...	8 (10.92)	5 (6.86)	5 (11.82)
you ... have to	6 (7.75)	4 (5.09)	5 (11.82)
for you and ...	7 (9.33)	7 (10.39)	7 (17.81)
in terms of ...	17 (25.2)	16 (26.27)	5 (11.82)
you and your ...	7 (9.33)	6 (8.62)	11 (29.79)

詞 *you* であることが確認される。一部、*in terms of* のように、具体的かつ慣習的な表現パターンが頻出していることが確認できる。それにたいして、本論文の最後に記載した各コミュニケーション形態特有の表現パターンは、より具体的な語彙が出現していることが確認できる。

また、各コミュニケーション形態特有の表現パターンを参照し、リストの中で共通した語を含むパターンを観察すると、4 語を超えるより長い表現の頻繁な使い回しが存在することが確認できる。たとえば、*f2f · agent* 特有の頻出表現パターンでは、*if you want to go to...*, *the good thing about...*, *phone · agent* 特有の頻出パターンでは、*i'll send you a list of...*, *could i take your email address...*, *mail · agent* 特有の頻出パターンでは、*i look forward to hearing...*, *i would strongly recommend...*, *if you are interested in...* が多用されていることが確認できる。

3 考察

本論文では、異なるコミュニケーション形態という状況の違いの中で人が言語表現をどれだけ使い分けているのかという問題を提起し、英語話者による実際のやり取りの内容の網羅的な分析をおこなった。分析の結果、全てのコミュニケーション形態に共通して使われやすい表現のパターンが確認したが、パターンを構成する語のタイプとしては、前置詞、代名詞などの抽象的な語彙が多く見られた。それに対し、コミュニケー

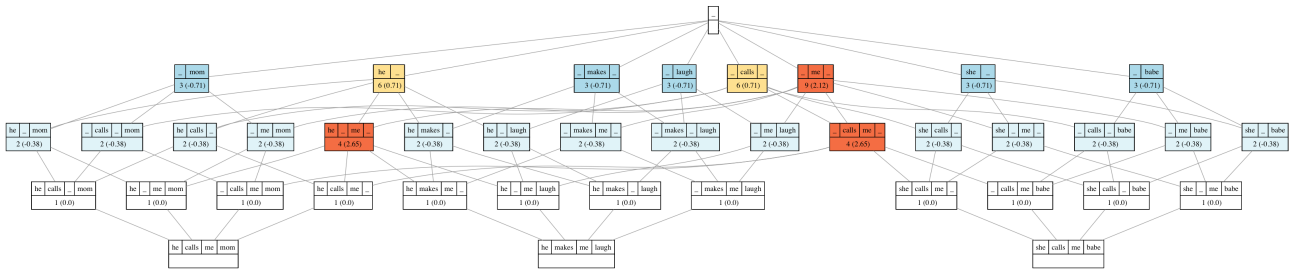


図 1: (5) の例から作成されたラティス図

ション形態に特有の表現パターンはより具体的な語彙を使用している傾向が確認できた。

人はコミュニケーションにおいて、的確な状況に適した定型な言い回しを記憶し、使用しているという点が理論言語学において言われている (cf. Fillmore, 1979) が、本論文でも、具体的な状況であるほど、それに適した具体的な表現パターンを使用していることが推測できる。

複数の単語からなる言語表現パターンを、より多くの状況に適応させるためには、必然的にパターンを構成する語彙の抽象性を上げなければならないことは容易に予想できるが、本論文での分析で得られた結果は、その抽象度の度合いを一定の段階まで明示できたものとする。今後は、各コミュニケーション形態に特有の表現パターンについて、やり取りの中での生起位置や文脈等の情報と照らし合わせることで、コミュニケーション形態と文脈に適応した言語表現パターンのリストアップをおこなっていくことで、英語教育にも応用可能な表現パターンの抽出を目指したい。

■ 謝辞

本論文は、JSPS 科研費 (若手研究 (B), 研究課題番号: 17K17943) の助成を受けたものです。

参考文献

Fillmore, Charles J (1979) "On Fluency," *Individual Differences in Language Ability and Language Behavior*, pp. 85–101.

Iwasaki, Shoichi (2015) "A multiple-grammar model of speakers' linguistic knowledge," *Cognitive Linguistics*, Vol. 26, No. 2, pp. 161–210.

Kecskes, Istvan (2010) "Situation-bound utterances as pragmatic acts," *Journal of Pragmatics*, Vol. 42, No. 11, pp. 2889–2897.

Kuroda, Kow (2009) "Pattern Lattice as a Model for Linguistic Knowledge and Performance.," in *PACLIC*, pp. 278–287.

Kuroda, Kow and Yoichiro Hasebe (2009) "Modeling (human) knowledge and processing of natural language using pattern lattice," in *15th Annual Meeting of Japanese Society of Natural Language Processing*, pp. 670–673.

McLuhan, Marshall (1962) *The Gutenberg galaxy: The making of typographic man*: University of Toronto Press.

土屋智行 (2017) 「N-gram にもとづくメディアごとの定型表現出現傾向の分析」, 『言語処理学会第 13 回年次大会論文集』, 711–713 頁.

資料: 代理業者による対面 (f2f), 電話 (phone), およびメール (mail) でのやりとり特有の頻出表現パターン

下のリストでは、各コミュニケーション形態で特に出現の度合いが高いパターンを抽出した。たとえば、対面によるやり取りでは z 値が高いものの、他のコミュニケーション形態 (e.g., 電話とメール) ではその値が 1 以下のものを抽出している。括弧内の値は z 値を示す。

f2f・agent 特有の頻出表現パターン ($z \geq 6.16$)

want to go ... (14.09), it ... be a (10.92), let's see what ... (10.92), you want ... go (10.92), you ... to go (10.92), you ... get a (10.92), you can get ... (9.33), then we can ... (7.75), can do is ... (7.75), to ... in japan (7.75), want to ... to (7.75), the ... thing about (7.75), a ... of people (7.75), let's have a ... (7.75), to ... you know (7.75), want

... go to (7.75), lot of people ... (7.75), the hot springs ... (7.75), hot springs and ... (7.75), you ... like ... (6.41), so ... have a (6.16), is there any ... (6.16), are ... interested in (6.16), the good thing ... (6.16), there's a ... of (6.16), a ... of a (6.16), kind of ... of (6.16), if you book ... (6.16), if you ... with (6.16), of ... you know (6.16), then you know ... (6.16), be you know ... (6.16), you know you ... (6.16), to the ... and (6.16), one of your ... (6.16), you can go ... (6.16), you can ... that (6.16), i ... if you're (6.16), if you're looking ... (6.16), oh yeah yeah ... (6.16), going to ... a (6.16), you ... on the (6.16), a ... for you (6.16), a lot ... people (6.16), which is ... popular (6.16), be a great ... (6.16), be a nice ... (6.16), can get a ... (6.16), it sounds like ... (6.16), days in osaka ... (6.16), a week before ... (6.16), book with us ... (6.16)

phone • agent 特有の頻出表現パターン ($z \geq 5.09$)

send you a ... (15.68), some of the ... (10.39), i'll send ... a (10.39), you a list ... (8.62), you ... to spend (8.62), so i will ... (8.62), a bit more ... (8.62), would be ... in (8.62), or did ... want (8.62), the ... okay so (8.62), i'll ... you a (8.62), if you're willing ... (8.62), if ... willing to (8.62), want to spend ... (6.86), you ... list of (6.86), a ... of different (6.86), um you know ... (6.86), you can find ... (6.86), list of different ... (6.86), your email address ... (6.86), would be interested ... (6.86), would ... interested in (6.86), you're ... to spend (6.86), yeah that's a ... (5.09), is a very ... (5.09), we could do ... (5.09), is there ... in (5.09), in hokkaido and ... (5.09), you want ... be (5.09), you know ... of (5.09), have a ... day (5.09), you know ... and (5.09), that could be ... (5.09), i ... you know (5.09), of your trip ... (5.09), you ... you can (5.09), willing to ... a (5.09), if ... go to (5.09), to get ... the (5.09), you'll be flying ... (5.09), you want ... spend (5.09), want ... stay in (5.09), so i ... send (5.09), i will ... you (5.09), in front of ... (5.09), in ... of you (5.09), send you ... list (5.09), you ... information about (5.09), how long ... you (5.09), send ... a list (5.09), a list ... different (5.09), you ... fly direct (5.09), up to hokkaido ... (5.09), could i ... your (5.09), take your email ... (5.09), take your ... address (5.09), take ... email address (5.09), would ... like to (5.09), that kind of ... (5.09), okay so ... will (5.09), or are you ... (5.09), are you know ... (5.09), there ... in particular (5.09), the top of ... (5.09), your ... would be

(5.09), send ... a few (5.09), you a few ... (5.09), uh ... you know (5.09), would you like ... (5.09), you like to ... (5.09), first of all ... (5.09), to osaka ... then (5.09), on the mainland ... (5.09), then back to ... (5.09), if i ... take (5.09), i ... take your (5.09)

mail • agent 特有の頻出表現パターン ($z \geq 5.83$)

my name is ... (11.82), to ... from you (8.82), japan rail pass ... (8.82), do ... best to (8.82), you ... your mother (8.82), thank you for ... (8.82), in the attachment ... (8.82), once you have ... (8.82), we can ... a (5.83), if you ... interested (5.83), i look ... to (5.83), i ... forward to (5.83), you ... your family (5.83), i look forward ... (5.83), look forward to ... (5.83), look forward ... hearing (5.83), look ... to hearing (5.83), forward to hearing ... (5.83), and your family ... (5.83), i've ... a list (5.83), i would strongly ... (5.83), japan travel consultant ... (5.83), japan travel ... dear (5.83), do our best ... (5.83), do our ... to (5.83), our best to ... (5.83), you and ... mother (5.83), and your mother ... (5.83), to this email ... (5.83), to choose from ... (5.83), all the best ... (5.83), you will have ... (5.83), list of ... options (5.83), provide you with ... (5.83), for getting in ... (5.83), you won't ... to (5.83), japan ... consultant dear (5.83), travel consultant dear ... (5.83), included in the ... (5.83), the naha ... resort (5.83), i ... strongly recommend (5.83), would strongly recommend ... (5.83), free to ... me (5.83), the first thing ... (5.83)