

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

土屋智行

九州大学 言語文化研究院

tsuchiya@flc.kyushu-u.ac.jp

1 はじめに

電話やメール, チャットをはじめとするコミュニケーション・メディア (以降「メディア」) が多様化している現代において, それぞれのメディアでの円滑かつ効率的なコミュニケーション方法がどのようなものであるのかを分析・考察する意義は大いにあるといえる. [1] は, 定型表現とそれによって実現される流暢性に, 効率的なコミュニケーションの基盤があると述べている. 定型表現とは, コミュニケーションの参加者が複数の要素をまとまりとして記憶している表現である [2].

しかし, 多様化しているメディア形態において, 定型表現および定型表現の記憶がコミュニケーションにおいてどのように使用され, どのような機能を持つのかを再検討する必要がある. 本発表では, 複数のメディアを取り上げ, それぞれのメディアにおける定型表現の使用実態を分析する手法と, 実際のデータの分析をおこなう.

2 「旅行課題遂行会話データベース」

「旅行課題遂行会話データベース」は, 科研費プロジェクト (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」) の活動として, 本発表者が平成 27 年度 10 月からコーパス化を目指して構築したデータベースであり, 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. データベースには, 全体で 10 時間程度の会話の音声と映像が収録されている. 会話の内容は, 個人の旅行に関する自然発話 (合計約 40 分) と, 協力者を「旅行者役」と「代理業者役」の 2 つに分けたうえでの課題遂行会話 (合計約 9 時間) である. 課題遂行会話は, 直接対面式の会話, 電話による会話, 電子メールを用いた会話の 3 種類を収録している.

2.1 特徴と構築方法

【特徴】 データベースは, 協力者および会話のトピックを固定しながらも, 使用するメディアを変えるという「1トピック・多メディア」の手法をとっている. したがって, 収録内容としては, 特定の個人が, 特定のトピックのみについてのコミュニケーションを何度もおこなうものであるが, 異なるメディアを用いる結果, メディアごとに選択されるコミュニケーション方略の違いや共通点を分析することが可能となる. さらに, 同じメディアによるコミュニケーションであっても, 時間をおいて数回おこなっていることから, 個人が言語表現を選択・記憶・利用する過程の一端を明らかにすることも期待できる.

【構築方法】 収録では, 4 名の英語話者から協力を得た. 4 名はいずれも男性であり, うち 3 名は英語母語話者である. 2 名は, 後述する課題遂行の中で「旅行者役」, 残り 2 名は「代理業者役」を担った. 協力者についての情報は, 表 1 のとおりである.

まず協力者には, 事前調査として, 「印象に残った旅」についての語りを 10 分程度してもらい, その音声と映像を収録した. 後日, 協力者に「日本国内の旅行プラン」を課題として与え, 対一のコミュニケーションを実施してもらった. 3 つのメディア形態 (対面・電話・メール) でのコミュニケーションを 1 セッションとし, 各協力者には, 自分とは異なる役と 3 セッションおこなってもらった. したがって各協力者は, 3 つの形態による収録を 3 回ずつ, 合計 9 回おこなうこととなった (表 2 参照). 収録した音声と映像は, 全体で 18 回分, 収録時間は, 全体で約 10 時間となった. 音声と映像に加え, メールでのやりとりではパソコンの画面の映像を記録した.

3 分析

3.1 メディアと役割ごとの語数の分布

まず、それぞれのコミュニケーション形態で収集した会話の規模を表3に示した。メディア間を比較すると、対面会話、電話会話、メールのやりとりの順で語数が多く、対面会話で最も会話がおこなわれていることが確認できる。また、役割間を比較すると、いずれのメディアでも代理業者役の発話が旅行者の発話を大幅に上回っている。特に電話会話では、代理業者役が旅行者役よりも3倍近い語数を発話していることが確認できる。全体的には、全体のやりとりのうち3割を旅行者、7割を代理業者が占めていた。

3.2 データベースのN-gram集計とその観察

かならずしも、定型性が高い表現ほど出現頻度が高いとはかぎらないが、本発表では今後の詳細な分析に向けた足がかりとして、より長い連結でより頻繁に出現する表現を定型性が高い表現、すなわち、コミュニケーションの参加者が「まとまり」として記憶・利用している表現と考えて分析する。したがって、本発表では、定型性の高い表現の抽出方法として、N-gramを利用する。

まず、書き起こしデータをメディアと役割(旅行者・代理業者)に従って6つのグループ(対面・旅行者、対面・代理業者、電話・旅行者、電話・代理業者、メール・旅行者、メール・代理業者)に分け、それぞれのグループで5-gramsからbigramまでの表現を抽出した。さらに、その中から2回以上生起したものを集計し、各グループの全体の語数に対しての割合を算出した(表4~7)。

まず、表4~表7までの各数値をメディア間で比較したが、n-gramのタイプおよび役割ごとに数値にばらつきが見られたため、各メディアの定型性の高さは判断できなかった。しかし、役割間で比較すると、旅行者よりも代理業者のほうが一貫して数値が高く、特に

表1: 協力者の出身国および担当役

協力者	出身国	担当役
Ted	ニュージーランド	旅行者
Jon	アイスランド	旅行者
Mitch	英国	代理業者
Finn	英国	代理業者

表2: 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 (送・返信各6回)	20~25分	約5時間

表3: 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%)

電話での会話においては両者の差が大きく、逆にメールでのやりとりにおいては両者の差が小さかった。

3.2.1 1つの会話のデータからみる定型表現の生起位置

次に、予備的な調査として1つの会話データを取り上げ、その会話において、定型表現が出現する位置を観察した。具体的には、旅行者・電話グループの会話データから抽出された生起頻度2以上の5grams~bigramsの表現について、当該グループの会話1回分のどの位置に出現しているかを調査した。

調査の結果を図1に示す。図の横軸の左端は会話開始時点であり、横軸の右端は会話終了時点である。縦軸は、5-gramsからbigramsの生起位置であり、もともと縦軸が高いのを5-gramsの表現とし、順番に4-grams, 3-grams, bigramsとなっている。

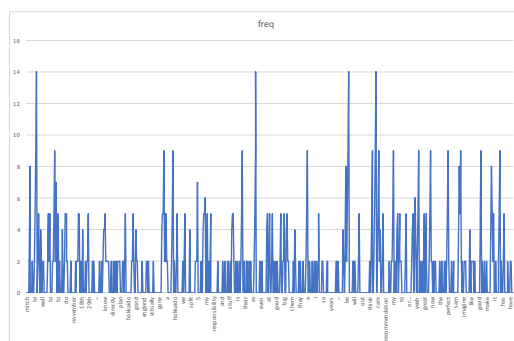


図1: データ (PhoneTravelr06) における定型表現の生起位置の分布

図1を見ると、bigramsの表現は、会話中に一貫して出現し続けているのに対し、3-grams以上の表現はからずしも均等に分散して出現していない。また、5-gramsの表現は会話の冒頭と会話の中盤に出現している。

表 4: 各グループの語数に占める 5-grams の分布

	旅行者	代理業者
対面	0.84%	1.70%
電話	0.54%	2.26%
メール	1.33%	1.53%

表 5: 各グループの語数に占める 4-grams の分布

	旅行者	代理業者
対面	3.13%	5.88%
電話	2.46%	6.56%
メール	3.23%	4.04%

表 6: 各グループの語数に占める 3-grams の分布

	旅行者	代理業者
対面	11.99%	19.16%
電話	10.34%	20.55%
メール	10.50%	10.97%

表 7: 各グループの語数に占める bigrams の分布

	旅行者	代理業者
対面	42.87%	54.62%
電話	39.95%	56.72%
メール	33.95%	34.67%

この生起状況を確認するために、旅行者による電話会話に出現する 5-grams の表現を下に列挙する。

旅行者・電話会話, 5-grams

looking to travel to Japan (3)^a, one cultural place do some (2), my – my dad’ s friends (2), – it might take a (2), we have the rail pass (2), a – this is a (2), is a – this is (2), Yeah Hmm Yeah Sure Uh (2), be much of an issue (2), to travel to Japan Yeah (2), – if – if we (2)

^a括弧 () 内の数字はトークン頻度を示す

それぞれの表現を見ると、日本への旅行を希望する表現に加え、旅程に関する希望を述べる際の表現が確認できる。

4 考察

まず、コミュニケーションにおいて、2つ程度の語の結合は頻出しており、特定のコミュニケーション上の語用論的機能を有するというよりも、むしろ流暢性や円滑性などを高める機能があるかと考えられる。

なお、5-grams をはじめとした大きいまとまりの表現の傾向は、上の例だけでは判断しかねるが、旅行者

が対面会話をおこなった場合の 5-grams では、異なる表現が確認できる。旅行者による対面会話の 5-gram 表現を下に列挙する。

旅行者・対面会話, 5-grams

the realm of possibility Yeah (2), And so we want to (2), seeing shrines I think is (2), sons they really want to (2), my sons they really want (2), how do you – how (2), to see that as well (2), but I’ m not going to (2), take a look at that (2), we’ ll take a look at (2), at 2 o’ clock in the (2), can find a nice place (2), 2 o’ clock in the morning (2), I didn’ t expect this level (2), didn’ t expect this level of (2), expect this level of freedom (2), that I want to see (2), I think I would be (2), I have to go to (2), we can fit it in (2), I think I have to (2), we can find a nice (2), Yeah thank you very much (2), me in the right direction (2), point me in the right (2), like to go to Osaka (2)

それぞれの表現を見ると、電話会話で出現しているものに加え、代理業者の提案に対しての肯定的な反応や感謝を示す表現が多くを占めている。したがって、同じ音声会話であっても、電話よりも対面による会話の方が、相手に肯定的反応を返す場面で定型表現を多用している可能性が示唆される。

紙面の都合上詳細な内容は割愛するが、代理業者の場合、冒頭の挨拶、旅行者についての情報の確認、メールでは結辞を述べる場面で定型表現を使用し、旅行者役は冒頭の挨拶と旅行内容の希望を述べる際に定型表現を使用している。これは、逆にそれ以外の場面では、使用する表現は多様化していると判断できる。つまり、定型表現は全ての場面で等しく使用されているのではなく、参加者がコミュニケーションを円滑化できると感じる一部の場面で使われており、コミュニケーションの参加者は、メディアや会話における役割、コミュニケーションの場面等に応じて定型表現を限定的に選択・使用していることを示していると考えられる。

【謝辞】 本研究は JSPS 科研費（研究活動スタート支援、課題番号: 15H06462）の助成を受けたものです。

参考文献

- [1] Charles J Fillmore. On fluency. *Individual Differences in Language Ability and Language Behavior*, pp. 85–101, 1979.
- [2] A. Wray. *Formulaic Language and the Lexicon*. Cambridge University Press, 2002.