九州大学学術情報リポジトリ
Kyushu University Institutional Repository

定型表現のメディア依存性 ： 「旅行課題遂行会話データベース」とその構築に向けた理論的基盤

土屋, 智行
九州大学大学院言語文化研究院 ： 助教

https://doi.org/10.15017/1804170

出版情報：言語文化論究. 38, pp.1-15, 2017-03-03. 九州大学大学院言語文化研究院
バージョン：
権利関係：

定型表現のメディア依存性
――「旅行課題遂行会話データベース」とその構築に向けた理論的基盤――

土　屋　智　行

1．はじめに

本研究（科研費プロジェクト「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」）では、対面・電話・メールなどのコミュニケーションメディア（以降「メディア」）と、そのメディアを利用した円滑かつ効率的なコミュニケーションを実現するために、主体が何を記憶し、その記憶をどのように利用しているかを探っている。本稿では、本研究で構築しているデータベース構築に至る理論的基盤を説明した後、データベースの内容に若干の観察と分析をおこなう。

1.1　メディアに応じた表現の「使い分け」

はじめに、分かりやすい例として、英語圏の旅行代理店で旅程の相談を電話でおこなう場面を想定してみる。本研究で収集した会話（「旅行課題遂行会話データベース」、Jon-Finn）では、以下のような会話がおこなわれている。下の会話およびメールのやりとりでは、Jon が旅行者、Finn が旅行代理業者である。

Finn: Hello, you've reached japantravel.com. This is Finn. How can I help you today?
Jon: Hi, Finn. My name is Jon and I am looking to travel to Japan.
Finn: Oh, okay, brilliant! Well, you've come to the right place, Jon.
Jon: Yeah.
Finn: So, do you have any plans for your trip so far or you're completely starting afresh?
Jon: Well, I'm travelling with my uncle, so there's just the two of us.
Finn: Okay.
Jon: And we don't want to go to Tokyo. We want to see more sort of historical side of Japan.
Finn: Right, right.
Jon: And the one thing we have in mind is to go to Hiroshima to - to specifically see the Atomic Bomb Dome.
…

一方で、同様のトピックをメールでおこなう場合には、以下のような文面となる。
まず、旅行者（Jon）は、以下の様なメールの文面を送っている。

> Hello JapanTravel,
> My name is Jon and I am looking to travel to Japan.
> We are three people looking to go, me and my two sisters. They are huge winter sports fans and really want to go to Hokkaido during wintertime. We are planning to travel on the 1st of November until the 9th….

次に、代理業者（Finn）は、以下の様なメールの文面を送っている。

> Hi Jon!
> Such a pleasure to hear from you. Thank you for choosing Japan Travel as your gateway to Japan and all the wonders it has to offer! My name is Finn, and I'll do my very best to make sure that you and your sisters have a fantastic trip filled with unforgettable experiences….

上記の２つの例は、いずれも同一人物によっておこなわれたコミュニケーションであるが、「旅行代理店での旅程の相談」という同一のトピックであっても、使用される言語表現にはそれぞれに異なる部分が存在する。

本研究の大きな問いは次のとおりである ── コミュニケーションの主体は、使用するメディアに応じて言語表現を使い分けているが、この言語表現が選択されるまでに、主体の中でどのような過程が生じているのだろうか。本稿では、この問いに答えるべく、理論的基盤および観察・分析手法を提示し、その手法を実施することで得られた知見と展望を述べ、今後の詳細な分析の土台とする。

２．定型表現について

本節では、まず分析の対象である定型表現について、その定義と理論的・教育的示唆について述べる。次に、定型表現の分布や用法がメディアによって異なる現象について触れ、その現象を説明しうる手がかりとして Gibson（1979）のアフォーダンス理論および McLuhan（1962）の感覚比率の概念を紹介する。最後に (a)「個人内・メディア間」の言語使用の観察・分析の重要性、(b) メディアおよび文脈への適応過程を記録したデータベース構築の必要性の２点を主張し、次節に向けた議論の基盤とする。

定型表現の定義

「定型的 *formulaic*」という語は、Pawley & Syder（1983）や Fillmore（1979）でも使用されているが、これが「定型表現 *formulaic expression*」または「定型的言語 *formulaic language*」という言語現象として注目され始めたのは Wray（2002）による体系的な議論の後である。 Wray による定型表現の定義は、(1)のようなものである。

(1) 連続的または非連続的に、複数の語あるいは他の要素が既成的に組み上げられた（ようにみえる）表現である。つまり、文法的な産出や分析がされるよりも、全体として記憶に貯蔵され、また使用では全体として取り出される[1]。（日本語訳、訳は本稿著者による）

この定義は、Wray も自ら述べているように、イディオムやことわざ、決まり文句等の下位カテゴリー全体を括ったものであり、具体的な言語形式や意味・機能への言及をしない曖昧なものである。しかし、この定義が示唆しているのは、定型表現が言語的な記憶とその利用の実態に最も直接的に

かかわり、それゆえ言語学にとって決して周辺的な現象ではないばかりか、むしろ認知科学等の関連領域とのつながりを深める重要なトピックであるということである。本稿では、この「言語的な記憶」という認知科学的観点の妥当性から、この定義の一部を、本稿における定型表現の定義にも加えることとする。

本稿における、定型表現の定義は(2)のとおりとする。

(2) コミュニケーションがおこなわれる環境の中で、発信内容を効果的・効率的に伝達するために、主体が記憶し、利用する複合的な表現

(2)の「環境」は、言語的な情報がやりとりされるコミュニケーション・メディアのことを指す。メディアは、コミュニケーションにおける言語・非言語情報を媒介する道具である（cf. 石田 2003: 83-87)。たとえば、対面での音声コミュニケーションであれば、音声を媒介する空間そのものがメディアであり、したがって、コミュニケーションがおこなわれる環境は物理的な空間となる。対して、メールでのコミュニケーションであれば、文字情報を媒介するパソコンがメディアであるので、この場合の環境は、パソコンの画面の表示および操作部としてのキーボードとマウス（またはトラックパッド）となる。

また、「複合的」が意味するところは、表現が複数の語または形態素の連続的または非連続的な連結であり、音韻的または意味的に抽象化されない状態で記憶されているという点である。たとえば "kick the bucket" は、定型表現の下位カテゴリーであるイディオムに分類されるが、この表現は、それぞれの単語 kick, the, bucket という語および配列そのもので記憶されているものであり、'CVCC CCV CVCCVC' という母音と子音の組み合わせでも、"kick the NP" という構文へと抽象化されて記憶されているものでも、またその構文的知識から具体化されているものでもない。あくまで kick, the, bucket という各単語の連結および［DIE］という意味として記憶されている表現である[2]。

以下では、上の定義にしたがって、定型表現の理論的・教育的示唆およびメディアと定型表現の関係について述べる。

定型表現の理論的示唆 ── 慣習・母語話者らしさ・流暢性 ──

近年、コーパスをはじめとした大規模な言語データの構築とそれを利用した言語研究が進められており、定型表現は、これらの言語データの研究手法の発展とともに注目されてきた経緯がある。大規模な言語データの存在は、他にも N-gram モデルやコロケーション、構文文法等の研究の発展にも貢献したが、これらはいずれも言語の慣習的側面にかかわる領域である。

定型表現をはじめとした慣習的な表現の理論的な重要性は、これら言語データからの分析をとおして認識されている。たとえば Erman & Warren (2000) は、複数のコーパスを用いた調査によって、書き言葉、話し言葉それぞれで約50％が既成表現 (prefab) であることを示している。この分析結果が正しければ、コミュニケーションの実現のかなり多くの部分を主体の定型的な記憶が担っており、定型表現研究は、「主体は表現をどのようなかたちで、どれだけ記憶しているのか」という問いに最も直接的なアプローチを試みる領域といえる。

Pawley & Syder (1983) は、言語表現の「母語話者らしい流暢性と選択性」の実現に定型的な表現がかかわっているとしている。Pawley & Syder は、結婚のプロポーズをする際の表現を挙げ、その自然さの違いについて述べている。

(3) I want to marry you.

(4) a. I wish to be wedded to you.

b. I desire you to become married to me.

c. Your marrying me is desired by me.
d. My becoming your spouse is what I want.
e. I want marriage with you.
f. What is desired by me is to wed you.
g. I, who am speaking, want to marry you, whom I am addressing. （Pawley & Syder 1983: 196）

(3)-(4)は、結婚のプロポーズをおこなう際の表現として挙げられたものであるが、この中で(3)が最も母語話者らしい自然な表現である一方、(4a)-(4g)は不自然さやぎこちなさが生まれるとしている。ある言語表現に対して違和感を覚えるか否かには、より自然な表現がすでに社会的もしくは個人的な慣習として定着していることが関係している。

また Fillmore (1979) は、母語話者らしさに関連する概念として流暢性を挙げている。Fillmore は、言語の流暢性は大きく４つに分類され、またそれぞれに定型的な表現が関わっているだろうと述べている。 Fillmore の流暢性の４分類は、以下の項目である (cf. *ibid*: 93)。

・発話内または発話間の休止が少ない流暢性
・意味的に「濃い」発話をおこなう流暢性
・状況や文脈に「適切な」発話をおこなう流暢性
・創造的な発話をおこなう流暢性

この４つの流暢性は、発話をおこなうにあたっての言い間違い、吃音、冗長性、誤解やナンセンスさなど、円滑なコミュニケーションにとって障害となる要素を排除し、発話および情報の価値を高めることを目指して実現されるものである。

以上のように、定型表現は主体による言語の記憶と、それによる円滑なコミュニケーションの実現に直接関わる領域であり、かつ 母語話者らしさと流暢性の実現に中心的な役割を果たすものである。

定型表現研究の教育的示唆

また、定型表現の研究は、英語教育等の言語習得・言語教育の領域にも貢献しうる。

上でも述べたように、定型表現は、主体の「母語話者らしい」言語使用を実現する役割を担っているが、様々なメディアや状況において、英語母語話者が一般的に使用している定型表現とその用法の目録を作成すれば、それは各メディアにおいて母語話者らしい会話に近づくことを望む英語学習者にとって、学習のための貴重な資源となりうる。本研究は、この言語資源の構築も視野に入れ、データの収集と分析をおこなうものでもある。

以上で述べたように、定型表現研究は、コーパスの開発とともにその重要性を増しており、「言語の記憶」および「母語話者らしさの実現」という側面に強く関わるトピックとして、理論的な発展が求められる。加えて、非母語話者にとって、「母語話者らしい」表現の目録の構築は期待されるものであり、本研究も、英語教育への応用を視野に入れてデータの分析と考察をおこなうものである。

2.1 定型表現のメディア依存性

これまで異なるメディアに応じた定型表現の使い分けの現象があることに触れたが、以下では、この現象の分析にあたって解決すべき課題を、具体的なデータにもとづいて検討していく。まず、定型表現のメディア依存性とその分析の課題について、British National Corpus（以降 BNC）を例にあげて説明する。

BNC は、イギリス英語の話し言葉と書き言葉のデータを有する約１億語の英語均衡コーパスであり、英語データの分析では中心的に用いられている。このコーパスから、５つの語の連結 (5-grams)

表１：話し言葉と書き言葉の5-grams の表現の頻度

Spoken 5-grams	frequency	Written 5-grams	frequency
be gon na[3] have to	322	there be no doubt that	465
will not be able to	212	do not want to be	428
do not think there be	193	there be no need to	406
be not go to be	185	do not have to be	391
be go to have to	183	do not seem to be	355
be go to be a	182	be more likely to be	341
be gon na be a	175	research grant award ref no	327
be not gon na be	173	will not be able to	315
do not know what to	166	be not go to be	310
do not want to be	146	do not know what to	293
there be go to be	145	turn out to be a	274
do not want to go	139	the hundred share index close	272
be gon na have a	134	be go to have to	271
do not know what be	133	have not be able to	270
do not think that be	131	be go to be a	268
that be gon na be	119	there be no reason why	267
that be a good idea	117	would not be able to	243
do not seem to be	117	do not know how to	233
do not have to be	117	be not to say that	230

を網羅的に収集し、その単純な出現頻度を取得すると、話し言葉と書き言葉では出現する表現の分布に違いがあることが分かる。

上の表は、BNC を話し言葉と書き言葉データに分け、それぞれの 5-gram（格変化の違いを除去したもの）を出現頻度順に上位20まで並べたものである。表１上の “frequency” は粗頻度を表し、話し言葉または書き言葉のみで使用されている表現は下線で強調している。表１を見ると、両方のデータに共通して頻度が高い表現と、話し言葉もしくは書き言葉で特に頻度の高い表現がそれぞれ約半数を占めていることを確認できる。これは、話し言葉・書き言葉という形態に依らず、頻繁に使用される定型表現がある一方、それぞれの形態に特化した定型表現が一定の割合で存在していることを示唆している。しかしながら、BNC の収集元となるデータは、性別、年齢、社会的属性などが異なる複数の著者または話し手による、多様なトピックにもとづいた言語活動から構成されているため、この結果をそのままメディアごとの違いと捉えることはできない。言語表現の使い分けを直接観察するためには、主体が異なるメディアを用いた際におこなわれるコミュニケーションを観察・分析する必要がある。つまり、１人の人間に、あるトピックについてのコミュニケーションを、異なるメディアを用いておこなってもらい、そのコミュニケーションを収録したデータを観察し、それぞれのメディアでどのような定型表現が選ばれているかを分析する、という手続きをとることが重要となる。

メディアの違いによる定型表現の枠組みの違い

現時点で、メディアごとの言語表現の使い分けを的確に示せるコーパスデータは構築されていないとはいえ、主体がメディアに応じて言語表現を使い分けているということは、私達の言語活動にもとづいた直感とも一致する。したがって、ここでは、先ほど述べた Fillmore（1979）の流暢性の概念と、メディアにおける身体性に関係する概念として Gibson（1979）のアフォーダンスと McLuhan

(1962) の感覚比率を挙げながら、定型表現の記憶とその使用によるコミュニケーションの円滑化について、「主体によるメディアへの適応」という観点から論じる。

まず、2節冒頭でも述べたように、Fillmore (1979) は発話の流暢性と定型表現には密接な関連があることを指摘したが、本稿では、この流暢性の実現は、円滑なコミュニケーションにとっての障害を乗り越え、発話および情報の価値を高めることを目指していると考える。メディアが変われば、当然何がコミュニケーションの障害となりうるかが異なってくるため、発信内容の価値を高めるための方略も異なってくることが考えられる。

たとえば、通常の対面会話である場合、円滑なコミュニケーションの障害となるのは言い間違いや吃音、冗長性、誤解やナンセンスさなどである。したがって、主体は、発話や情報の価値を高めるために、流暢性の高い発話の実現、すなわち発話間および発話内休止を調整したり、内容を的確かつ簡潔に言い表すような表現を選択する等の方略をとることとなる。

対して、電話やメールなどの異なるメディアを用いたコミュニケーションでは、また異なる方略をとることとなる。たとえば、電話会話においては、そのコミュニケーションの形態上、相手と直接対面しておらず、相手の視線等の非言語情報を得られない状況が、対面会話にはない障害となる。また、ほとんどの場合、話者は互いに空間的に異なる場所にいるため、空間情報を互いに共有していないことも障害となる。同様に、メールでの会話においては、非言語情報が含まれないことに加え、音声情報もなく、リアルタイムなコミュニケーションがおこなわれない状況がある。主体は、それぞれの環境において、最初の段階で名前を名乗ったり、受け手への丁寧かつ好意的な表現をより明示的に使用するなど、発話や情報の価値を高めるような方略をとる。

以上に述べたことを踏まえると、コミュニケーションで用いるメディアはそれぞれに特有の障害が存在しており、主体はその障害を乗り越えるようなコミュニケーション方略をとっていると推測できる。したがって、主体が用いる定型表現も、各メディアによってそのバリエーションや用途が異なってくると予測できる。

では、各メディアでのコミュニケーションにおいてはどのような障害が存在し、主体はその障害を乗り越えるために何をしているのか、またどのようなやりとりを重要なものと考え、記憶しているのだろうか。本稿では、主体によるコミュニケーション方略の観察・検討指標の理論的基盤として、Gibson (1979) のアフォーダンス理論と McLuhan (1962) の感覚比率の概念を簡単に説明する。

アフォーダンス

Gibson (1979) は、視覚に関する様々な現象を挙げながら、生物の身体と環境との関係性から知覚をとらえるべきであると主張した。視覚を持つ生物は、環境から網膜に差し込む光学的配列とその変化を、自らの身体運動感覚と統合することで、自身の移動や外部の変化を認識している。したがって、環境における様々な物理的刺激が、主体の知覚を支える情報として、環境から主体に提供されているとする。この環境と主体の関係性をアフォーダンスと呼ぶ。

感覚比率

McLuhan (1962) は、人類の文明が文字を発明したことによって、文化形態が聴覚型から視覚型へと移行していったことや、グーテンベルグが印刷技術を発明したことによって、文字文化が広く普及したことについて述べている。その中で彼は、文字を掲載するメディアによって、人間の五感が変容していると指摘し、その変容のしかたを説明するひとつの枠組みとして、「感覚比率 *ratio of senses*」という概念を用いている。

> ある文化圏の内部から、もしくは外部から一つの技術が導入され、その結果としてわれわれのもつ五感のうち特定の感覚だけがとくに強調され、優位を与えられる場合、五感がそれぞれに務める役割比率に変化が生じ

るのだが、そのときわれわれの感受性はもとのままではありえないのだ。眼も耳も他の感覚も以前とは違った働き方をしはじめる。(*ibid*, 森訳 : 41)

文字を掲載するメディアのみならず、電話やテレビ、ラジオなどの通信技術や映画などの娯楽に関する技術にも、人間はこの感覚比率を調整することによって順応していくものであると、McLuhan は述べている。コミュニケーションをおこなうメディアが変わるとき、主体の身体、特に感覚器官とその統合状態も変化するという点は、先に紹介したアフォーダンス理論と軌を一にするものであろう。

メディアにおける身体性

メディアは、人間がコミュニケーション形態を拡張するために使用する道具である。その一方で、コミュニケーション形態の拡張は、主体がそのメディアに適応するよう、自らの身体感覚や身体動作を調整することで実現される。これは、主体が自らの身体を軸として環境からの刺激を認識し、適応していくというアフォーダンスと、メディアに応じて自らの五感を再構築していくという感覚比率の基本的理念と一致する。メディアは道具であると同時に、コミュニケーションをおこなう環境であるともいえるだろう。

本研究では、特定のメディアを用いたコミュニケーションについて、主体は大きく２つの適応をおこなっていると考える。ひとつは、メディアへの感覚的・身体的な適応、もうひとつは、メディアおよび文脈ごとに適したコミュニケーションを実現する言語的記憶の形成である。対面の音声会話を例にとれば、前者は空間的な制約（聞き手との位置関係、空間の広さ、環境音の多さ・大きさ、聞き手の身体の状況、等）に合わせて注意の対象を選択し、また聞き手に届くように声量や口の動きを調整することにあたる。対して後者は、様々な状況と文脈における対面の音声会話の経験を通して、伝達内容を円滑かつ効率的に伝えられるような言語表現および表現手段を総体的に記憶していくことにあたる。本研究では、特に後者で述べた言語的記憶の形成と、それによる特定のメディアおよび文脈に適したコミュニケーションの実現という側面から、対人的なコミュニケーションを見ていく。

３．データベースの構築

3.1　概要

「旅行課題遂行会話データベース」は、先述の科研費プロジェクトの活動として、本稿著者が平成27年度10月から構築を開始している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、電子メールのやりとりの３種類を収録している。以下に、データベースの構成および収録の手法について説明する。

3.2　構築の目的

「旅行課題遂行会話データベース」は、以下の２点の実現を期待して構築をおこなっている。ここではそれぞれの項目について簡単に説明する。

(A)「１トピック・多メディア」コーパス構築とそれによる理論的貢献

(B) 円滑なコミュニケーションにかかわる定型表現リスト作成とそれによる教育的貢献

(A) について、本研究ではデータ収録にあたって、会話する個人（協力者）とトピックを固定しながらも、コミュニケーションに用い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。

(B) について、本研究は、データベースの構築によって、特定のメディアおよび文脈で使用される定型表現の特定と抽出を目指している。これによって、対面会話や電話会話、電子メールで使用される定型表現のリストが作成できれば、英語学習者にとって貴重な言語資料の提供が期待できる。

3.3 構成

データベースの構成の特徴として、以下の項目が挙げられる。

- *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）の合計4名
- *時　　間*：合計約10時間（独話約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（合計約5時間）
- *内　　容*：「印象に残った旅」に関する独話および日本国内の旅程決定に関する課題遂行会話

協力者

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。1名の英語非母語話者（アイスランド出身）は、発音、内容ともにネイティブレベルに近く、他の協力者とのコミュニケーションも問題なくおこなえている。参加者にはそれぞれ Ted, Jon, Mitch, Finn という仮名が与えられ、Ted と Jon には「旅行者役」が、Mitch と Finn には「代理業者役」が割り当てられた。協力者の名前、出身国、担当役の情報は、表2のとおりである。

表2：協力者の出身国および担当役

協力者名	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

内容

収録内容は、大きく「協力者による自然発話」と「課題遂行会話」の2つに分かれている。自然会話では、「印象に残った旅」についての自発会話を10分程度してもらい、その音声と映像を収録した。後日おこなわれた課題遂行会話では、協力者に「日本国内の旅程の相談」を課題として与え、そこでおこなわれる一対一コミュニケーションの音声と映像を収録した。

時間

独話のデータには、各協力者1回ずつ、合計4回の収録をおこなった。データには、各回約8～10分程度の音声および映像が収録されている。課題遂行会話のデータは、旅行者役1名と代理業者役1名による一対一の会話を収録した。収録は、対面会話6回、電話会話6回、メール会話6往復(12通)、合計24回おこなった。なお、メールのデータは、送信者である旅行者役と、返信者である代理業者役のメール作成中の様子をそれぞれ収録したので、実際のやりとりの回数は、6回となる。

課題遂行会話の収録時間の内訳は、表３のとおりである。

表３：課題遂行会話の収録時間

使用メディア	収録回数	１回あたりの収録時間	合計時間
対面会話	６回	20～25分	約２時間
電話会話	６回	15～25分	約２時間
メール会話	12回 （送信者６回、返信者６回）	20～25分	約５時間

3.4　手法

次に、課題遂行会話の実施の手法について説明する。

期間

収録は、４月の中旬から約３週間の間に、ほぼ全ておこなった。音声と映像は、自然発話４回、課題遂行会話24回、合計28回収録した。

収録ペアの組み合わせ方

各協力者には、対面、電話、メールの３種類のメディアを用いて、自分とは異なる役（旅行者役であれば代理業者役）との会話をおこなってもらった。会話は、メディア１種あたり３回、全体で９回おこなってもらった。対面・電話・メールの会話を１セッションとすると、各協力者は収録を３セッションおこなったこととなる。なお、本研究では、それぞれの旅行者役、代理業者役が各２名しかいなかったため、Ted は Mitch と、Jon は Finn と２セッションおこなってもらった。

収録の手続き

収録は、音声会話と電話会話では音声および映像を、メールのやりとりではメールを作成する協力者の映像およびメール作成中のウィンドウの映像を収録した。電話会話では、協力者はそれぞれ別室で電話機の無線式子機の内線電話機能を利用して会話してもらった。音声の収録には無線のヘッドマウント型マイクを、映像の収録には小型のアクションビデオカメラを、メール作成中のウィンドウ映像の収録には画面録画のアプリケーションをそれぞれ使用した。なお、ヘッドマウント型マイクは、電話での会話中でも使用している。

参加者への説明内容

旅行者役が希望する旅行内容については、「旅行期間」「同行者情報」「観光目的」「目的地」「リクエスト」の５項目だけを収録実施者側で指定し、他の情報については、各協力者で自由に決めてもらった（表４）。これに対し、代理業者役には、各都市に関する基本的な情報および各都市間の移動時間を資料として手渡し、宿泊施設や移動にかかる費用などの詳細な情報については、インターネットで自由に調べられるようにした。

表４：収録の際に指定した旅行内容と例

指定項目	例
１．旅行期間	４日～10日間（６月１日着・６月６日発）
２．同行者情報	一人旅、両親との旅行、息子・娘との旅行、等
３．観光目的	史跡巡り、日本料理、温泉、ショッピング、等
４．目的地	北海道、東京、京都、大阪、沖縄の５都市
５．リクエスト	刺身をメニューに含めない、布団ではなくベッド、等

なお、今後の英語教育への利用も視野に入れ、協力者には、明朗かつ丁寧な人物としてやりとりをおこなうように依頼した。

4．分析状況

現在は、音声と映像の書き起こしが終了し、アノテーションの作業をおこなっている段階である。データの内容の詳細な分析は別稿にて述べることとするが、本稿では基礎的な分析として、課題遂行会話のうち、メールの書き言葉データと対面・電話の２つを統合した会話データについて、代理業者役と旅行者役の定型表現の使用状況を比較する。具体的には、各データの N-gram 集計の結果およびその傾向をみていく。

4.1　データの N-gram 分析

まず、メールデータと会話データのデータ量について、代理業者役と旅行者役の点からみる。メールデータは合計で5,489 トークン、会話データは合計で37,084 トークン、全体で42,573 トークンとなった（表５）。全体の比率では、メールデータが全体の12.89％、会話データが全体の87.11％を占めることとなる。

表５: メールデータと会話データのデータ量

	Agent	Traveler	Total
Mail	3,536 (64.42％)	1,953 (35.58％)	5,489
Spoken	26,937 (72.64％)	10,147 (27.36％)	37,084
Total	30,473 (71.58％)	12,100 (28.42％)	42,573

いずれのデータも、代理業者役の人物によるデータがコミュニケーション全体の約７割を占めており、代理業者役のアウトプット量の方がより多いことが確認できる。また、メールデータと会話データで代理業者役と旅行者役のアウトプット量を比較すると、メールでは代理業者役のアウトプットの割合が比較的少ない。メールデータと会話データにおける代理業者役・旅行者役のアウトプット量を、χ 二乗検定を用いて分析したところ、両者の間には統計的な有意差が確認できた（p 値 <.001）。したがって、少なくとも本データでは、会話でのコミュニケーションの方が、代理業者役のアウトプット量は多くなる傾向にあることが分かる。

4.1.1　会話データの N-gram 分析

次に、会話データにおいて、代理業者役・旅行者役がそれぞれどれだけ定型的な表現を使用しているのかを、N-gram 集計の結果から見ていく。

まず、会話データのトークン数は、表５でも示したとおり、代理業者役が26,937、旅行者役は10,147であり、全体で37,084であった。つまり、代理業者役と旅行者役のアウトプットが約７対３の割合となっている。このデータから５grams の集計をおこない、４回以上生起したものを抽出したところ表６のとおりとなった。

会話データから４回以上生起したものは、代理業者役から19例、旅行者役から１例確認できた。左の “ngram” 列は実際に出現した表現、中央の “count” は生起数、右の “frequency” は、全トークン数のうちにその表現が占めている割合をパーセンテージで表したものである。それぞれの表現を見ると、多くの表現が会話の冒頭に生起するもので、かつ旅行者の情報を聞き出す発話であることが確認できる。たとえば、これまでの日本への旅行歴に関する質問（“... is it your first time in Japan”）、

表６：会話データにおける代理業者役・旅行者役の５gram 集計

Agent			Traveler		
ngram	**count**	**frequency**	**ngram**	**count**	**frequency**
if you're going to be	7	0.03	looking to travel to Japan	4	0.04
this time of the year	6	0.02			
you want to stay in	5	0.02			
it your first time in	5	0.02			
How can I help you	5	0.02			
your first time in Japan	5	0.02			
did you want to stay	5	0.02			
is it your first time	4	0.01			
can do is I can	4	0.01			
don't have to worry about	4	0.01			
I can – I can	4	0.01			
you want to go to	4	0.01			
send that free to you	4	0.01			
a – there is a	4	0.01			
at this time of the	4	0.01			
or did you want to	4	0.01			
what we can do is	4	0.01			
send you a list of	4	0.01			
popular at this time of	4	0.01			

旅行日程を旅行者役から聞いた後の応答（"…popular at this time of the year"）、旅行のリクエストに関する質問（"…or did you want to stay in…"）である。対して、旅行者役が発話している表現としては、日本への旅行を希望している旨を伝達する内容（"looking to travel to Japan"）があるが、４回以上生起したものはそれだけであった。

4.1.1　書き言葉データの N-gram 分析

次に、メールデータにおいて、代理業者役・旅行者役がそれぞれどれだけ定型的な表現を使用しているのかを、N-gram 分析によって見ていく。会話データのトークン数は、表５でも示したとおり、代理業者役が3,536、旅行者役は1,953であり、全体で5,489であった。つまり、代理業者役と旅行者役のアウトプットが約6.5対3.5の割合となっている。このデータから５grams の集計をおこない、２回以上生起したものを抽出したところ表７のとおりとなった。

代理業者役は、全体のやりとりのうち、冒頭の挨拶表現（"Thanks for getting in touch"）、結辞にあたる表現（"I look forward to hearing from you", "feel free to contact me"）、旅程内容に向けた意気込み（"…for you and your…", "We'll do our best to…"）が確認できた。それに対し、旅行者役は、冒頭の代理業者への宛名が多く、次いで旅行に関する希望に言及する表現（"I am wanting to travel…", "…We want to stay…", "…would be great if the…"）、業者の取り計らいへの感謝（"Thank you very much for…"）が確認できた。

表 7 : メールデータにおける代理業者役・旅行者役の5gram 集計

Agent			Traveler		
ngram	**count**	**frequency**	**ngram**	**count**	**frequency**
for you and your family	3	0.08	I am wanting to travel	3	0.15
I look forward to hearing	3	0.08	To JapanTravel.com Dear JapanTravel Employee	3	0.15
for you and your girlfriend	2	0.06	Thank you very much for	2	0.10
you and your family In	2	0.06	experience We want to stay	2	0.10
for you In terms of	2	0.06	Dear JapanTravel Employee Hi my	2	0.10
me know and I can	2	0.06	would be great if the	2	0.10
let me know and I	2	0.06	JapanTravel Employee Hi my name	2	0.10
you like what you see	2	0.06	Employee Hi my name is	2	0.10
if you like what you	2	0.06	until the 9th I am	2	0.10
and if you like what	2	0.06	rest of our time in	2	0.10
We'll do our best to	2	0.06	the rest of our time	2	0.10
I've included a list of	2	0.06	JapanTravel.com Dear JapanTravel Employee Hi	2	0.10
is a great place to	2	0.06			
where you will be travelling	2	0.06			
you will be travelling from	2	0.06			
feel free to contact me	2	0.06			
Hiroshima I would strongly recommend	2	0.06			
from you Thank you for	2	0.06			
hear from you Thank you	2	0.06			
to hear from you Thank	2	0.06			
The Tokyo Metropolitan Train Museum	2	0.06			
Thanks for getting in touch	2	0.06			
for getting in touch My	2	0.06			
look forward to hearing from	2	0.06			
If you are interested in	2	0.06			
forward to hearing from you	2	0.06			

4.1.2　会話・メールデータの比較および収録段階で見られた傾向

会話データとメールデータを比較すると、いずれも代理業者役の人物のほうが、まとまった表現を使用していることが確認できる。ただし、単純に代理業者役のアウトプット量に比例している可能性は否めないので、その点については精査する必要がある。また、定型表現を使用する場面については、代理業者役・旅行者役ともに一定の偏りがあることが分かる。たとえば、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。

また、データの収録段階でも、代理業者役のほうが定型表現を多用していることが確認できているが、個人によって積極的・優先的に使用している表現が異なることが確認できる。たとえば、Finn は音声会話の中で *It's not for everyone* という発話を２度おこなっているが、この表現は、下の例にあるように、旅行者役から「刺身をメニューに含めないで欲しい」等、旅行中に避けたい日本の文化的項目が提示された返答として使用されている。この表現は、同じ代理業者役の Mitch の会話からは得られなかった。

Ted: That's the thing. Ah, I've never been there before, and the only place I know I want to really go to is Tokyo. Ah, the – the extent of the research that I've done is little bit into the cuisine and I know for sure that I don't want any raw fish. No sashimi.
Finn: Okay.
Ted: That's, ah – that's I guess the one thing that I really don't want because I've tried it before and it's – that's, yeah.
Finn: Yeah, it's not for everyone, yeah.
Ted: No, I suppose not. No.

4.1.3　考察および今後の課題

「旅行課題遂行会話データベース」の構築と基礎的な分析をおこなったところ、4.1.2節で示したように、(a)旅行者役よりも代理業者役のほうが、そのアウトプット量の多さの影響は否めないものの、定型表現を多く使用している、(b)全ての場面で等しく定型表現が使われているのではなく、一部の場面で限定的に使われていることが分かった。つまり、コミュニケーションの参与者は、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選択・使用していることを示していると考えられる。

本稿では、定型的な表現を使用することでコミュニケーションの円滑性や効率性が実現されると述べたが、分析の結果、特定のメディアや場面等の条件において、この主張は正しいといえるだろう。しかし一方で、バリエーションに富んだ表現を使用したほうがより円滑かつ効率的にコミュニケーションが実現されるような条件があることも否めない。この点については、今後分析をさらに進めることで明らかにしていきたい。

注

1 ...a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (*Ibid:* 9)

2 この説明では「音韻的・意味的に抽象化されない状態で記憶される」としているが、これは「主体による音韻的・意味的な心的操作の対象ではない」ことを含意しない。定型表現は、既成的な表現として記憶されている一方、文脈への適応や創造的な営みの一環として、形式的・意味的な変化を容認する。“kick the bucket” の例でも、“kick*ed* the bucket” のように文法的な変化をともなうことができ、さらに “kicked the *bit* bucket” と創造的に拡張することで、パソコンおよび周辺機器の破損・故障を意味することができる。これは、“kick the bucket” の形式および意味が抽象化されずに記憶されながらも、形式的・意味的な心的操作をともなって創造的に利用されうることを示している。

3 BNC の形態素解析では、*gonna* は *going to* の縮約形として 2 形態素に分割される。本稿では、BNC における他形態素の解析状況との整合性を維持するため、*gonna* を 1 形態素とはせずに、解析結果をそのまま転用する。

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ある。

参 考 文 献

Erman, Britt and Beatrice Warren. (2000). “The idiom principle and the open choice principle”, Text, 20 (1), 29-62.

Fillmore, Charles J. (1979). “On Fluency”, In C. J. Fillmore, D. Kempler, and W. S-Y. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior*, 85-101. New York/San Fransisco/London: Academic Press.

Gibson, James J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin. (古崎 敬・古崎 愛子・辻 敬一郎・村瀬 旻［訳］1986.『生態学的視覚論 ── ヒトの知覚世界を探る ──』, 東京 : サイエンス社)

石田 英敬. (2003).『記号の知／メディアの知 ── 日常生活批判のためのレッスン ──』, 東京 : 東京大学出版会.

McLuhan, Marshall. (1962). *The Gutenberg Galaxy: The Making of Typographic Man*, University of Toronto Press. (森 常治［訳］1986.『グーテンベルクの銀河系 : 活字人間の形成』, 東京 : みすず書房)

Pawley, Andrew and Frances H. Syder. (1983). “Two puzzles for linguistic theory: Nativelike selection and nativelike fluency”, In J. C. Richards and R. W. Schmidt (Eds.) *Language and Communication*, 191-225. New York: Longman Inc.

Wray, Alison. (2002). *Formulaic Language and the Lexicon*, Cambridge: Cambridge University Press.

使用コーパス

British National Corpus (BNC) from Sketch Engine: https://the.sketchengine.co.uk

Media-dependent Characteristics of Formulaic Expressions

Tomoyuki TSUCHIYA

This paper aims to present a theoretical basis of formulaic expression research and propose an observation and analysis method through the construction and basic analysis of *the Database of Travel Planning Task Dialogues*. The author argues that the subject is able to select appropriate expressions at a wide variety of communication media types because of his/her physical and sensory adaptation to the media and his/her knowledge of the formulaic expressions obtained through communication.

With the cooperation of four English speakers, the author constructed *the Database of Travel Planning Task Dialogues*. In this database, the cooperators were divided into two groups –travel agency and traveler– and carried out itinerary consultation tasks using different media (face-to-face conversation, telephone conversation, and e-mail exchange). Accordingly, all 18 tasks (approximately. 10 hours) were recorded and transcribed.

As a result of their basic analysis (5-grams aggregation), it became clear that agents use formulaic expressions more frequent than travelers and the scenes where formulaic expressions are used are relatively limited. It is suggested that the subject carries out smooth and efficient communication by using formulaic expressions at certain types of scenes, while using varied expressions at the other types. Further cross-sectional analysis of media and situation would be required.