タスク指向型対話システムにおける多様な自然発話を収集する手法に関する研究

A Study on Methods for Collecting Diverse Natural Utterances in Task-Oriented Dialogue Systems

東京工科大学

工学研究科

嶋　和明

目次

用 語

GiFT	Gift for the Tin man の略名．本研究で提案するコーパス収集手法.
HMI	Human machine interface
HUD	Head up display
IVI	In-vehicle infotainment
IoT	Internet of things
POI	Point of interest
POS	Part of speech
PTT	Push to talk
SNS	Social networking service
TTS	Text to speech
WOZ	Wizard of Oz 手法の略称.

記 号

$\bar{x}$　そのコーパスにおける一発話あたりの形態素数.

x　そのコーパスに含まれる延べ形態素数.

n_{utt}　そのコーパスに含まれる発話数.

TTR　異なり形態素数比 (type / token ratio).

V　xに含まれる異なり形態素数.

TUR　一発話あたりの異なり形態素数比 (type / utterance ratio).

C_m　収集されたコーパスに含まれる形態素が，ログデータにどの程度含まれるかを示すカバレッジ.

N_{cl}　収集されたコーパスの形態素で，ログデータにも同じ形態素が出現しているかをカウントした一致数.

N_l　ログデータに含まれる延べ形態素数.

C_k　収集されたコーパスに含まれる異なり形態素が，ログデータにどの程度含まれているかを示すカバレッジ.

K_{cl}　収集されたコーパスの異なり形態素数で，ログデータにも同じ異なり形態素が出現しているかをカウントした一致数.

K_l　ログデータに含まれる異なり形態素数.

I_{wt}　一発話収集するのに要する作業時間に基づく時間的負荷指数.

I'_{wt}　I_{wt}に作業者数の要素を追加した時間的負荷指数.

C_{utt}　被験者一人あたりの平均発話数.

T_{wt}　発話取集のための作業時間.

C_p　コーパス収集作業に要する作業者数.

n_i　発話数（i はグループ A または B を意味する）.

$\bar{x}_i$　そのグループの一発話あたりの形態素数.

s_i　$\bar{x}_i$ に対する標準偏差.

$x_i(j)$　各発話の形態素数（グループ A では j = 1, 2, …, n_A，グループ B では j = 1, 2, …, n_B）.

CV_i　$\bar{x}_i$, s_i より得られる変動係数.

図 の 目 次

表 の 目 次

第1章　　序論

本章では，車載機器を例にシステムの音声操作に関する技術紹介をするとともに，既往研究および近年の技術動向を紹介し，本研究が掲げる課題を挙げ，それをもとに本研究の目的と目標について述べる．また本論文の構成について述べる．

1.1. はじめに

音声認識操作の Human machine interface (HMI) としての利便性は，複雑な機器の操作を音声で行えることである．機器操作ボタンの配置や操作順序などを覚える必要はなく，実行したい機能名称を発話することにより操作が可能になる．

車載機器においては，音声操作の利便性はさらに高まる．カーナビゲーション（カーナビ），カーオーディオと呼ばれていたものは，昨今では In-vehicle infotainment (IVI) とも呼ばれるようになり，テレマティクスやインターネットに接続することにより，経路案内や音楽再生に留まらず，電話操作，メール送受信，天気予報など，様々な機能を提供するようになった [1]．しかし，多機能化に伴い，操作の複雑化が課題となっている．それに加え，車室内は空間的制約により，ボタンなどのハード的な操作部用スペースは限られる．こうした課題の解決に各カーメーカ，車載機器メーカは取組み，別体のリモートコントローラ，画面ごとにキー配置を変えられるタッチパネル，Head up display (HUD), ジェスチャー操作などを採用し HMI の向上に努めてきた [2], [3].

音声認識による車載機器の操作はその代表的な一例である．たとえば，経路案内機能として恐らく最も多用される発話は「自宅へ帰る」と予想するが，このための一般的な操作は Push to talk (PTT) キーを押し，「自宅へ帰る」と発話するのみである．また，昨今の IVI や車両では PTT キーさえなく，ウェイクワードを発話することにより，音声操作のみで機能を実行させることが可能であり，これによりドライバはハンドルから手を放す必要がない製品もある [4].

車載機器では，組み込み型音声認識による音声操作が実用化されてきた．組み込み型では音声認識に必要なグラマや辞書を事前に用意し，車載機器内で管理している [5]．すなわち，車載機器が受け付けられる音声コマンドには限りがあり，結果的にユーザは車載機器が受け付けられる音声コマンドを覚える必要があった．しかし，この問題も昨今ではクラウド音声認識技術を活用し，高性能なサーバ上で処理することにより，多様な語彙を高精度で認識する技術を用いて HMI を向上させる手法が提案され，一つの解決策となっている．こうしたクラウド音声認識技術を活用した動きが広まっている [6]-[8].

多様な語彙を音声認識可能な技術（音声をテキスト化する技術）と，多様な言い回しでもユーザの意図を推定し適切な機器操作へと結び付ける言語理解技術が組み合わされば，より高いユーザビリティを提供することが可能である．しかし，音声から多様な言い回しをテキスト化したとしても，言語理解器が多様な言い回しを受け付けられるとは限らない．たとえば，「自宅へ帰る」という音声と「お家へ帰る」という音声を，音声認識技術によりテキスト化したとしても，言語理解器が「自宅」と「お家」が同義であることを理解していなければ「自宅を目的地として設定する機能」に結び付けることはできない．また，「自宅」と同義でも「拙宅」といった単語をユーザが発話するとは考えにくいため，現実的に使用される発話を言語理解器の学習データとする必要がある．

それに加え，ユーザが高性能な音声認識技術を使用し，多様な表現を音声認識可能と捉えたことによる，音声操作に対する「慣れ」に伴い，発話の表現が定型音声コマンドのような「機械が理解しやすい言い方」から，より自然で自由な発話へと多様化することが予想される．さらに，音声操作への慣れの視点のみならず，ユーザの機器に対する「認識」が「単純な機械」から「高度な言語処理技術を搭載した機器」へと変化すれば，発話もより自然で自由な発話へと多様化することも考慮する必要がある．機器の技術発展が継続していくことを考えれば，将来的には人間同士の会話のような話し方で機器操作したいというニーズが生まれると予想する．

こうした，自然な発想の発話による機器操作の利点は，ユーザのニーズを満たすのみに留まらず，機器操作の負荷低減にも寄与する．たとえば，前述した

ような車載機器においては，手を使った機器操作は安全な運転の妨げとなるため，音声操作を行うことにより負荷が低減することは望ましい．その上，自然に思いついた発話を機器が受け付けることができれば，ユーザは「機器が理解しやすい言い方」を考える必要がなくなり，認知的な負荷も低減することにより，安全な運転に寄与することが可能である．

機器操作による負荷には，「わき見」や「ヴィジュアルワークロード」と呼ばれる視覚情報に関する負荷と，音声操作やハンズフリー通話など視線はわき見をしていないが「意識のわき見」や「メンタルワークロード」と呼ばれる負荷があり，ドライバの運転阻害となりうる負荷として考えられている [9]-[11]．わき見運転は，物理的動作のためドライバの視線検出などからメインタスクである運転作業に対し，どの程度視線を車載機器に向けていたかといった調査をすることが可能である [12]．また，2004 年には「画像表示装置の取り扱いについて 改訂第 3.0 版」という指針が日本自動車工業会から出されており，ドライバの視線を表示器に向ける時間などが公開されている [13]．一方，意識のわき見やメンタルワークロードと呼ばれる負荷については，目に見える負荷ではなく容易には定量化できないため，過去の研究では，R-R interval variance (RRV) 法と呼ばれる，心拍間隔の変動より内面的な負荷を計測する手法がとられた研究がある [14]-[16]．また，森田らの研究では，マルチモーダル操作により，視線で機器を特定し，音声で操作を行い，人間の処理負荷を分散させることにより，全体的なワークロードの視点から負荷低減させる提案を行っている [17]．

これらを踏まえると，実際に操作するのは機器であっても，人間が機器に合わせ理解されやすいようにした定型音声コマンドベースの発話のみでなく，より自然で自由な発想による多様な発話でも機器操作が可能となるよう，機器の開発・評価などに用いられる発話データも多様な発話を含んでいる必要がある．本研究では，こうした発話例を収集する手法について提案し検証する．

1.2. 既往研究

音声による機器操作は，組み込み型音声認識のように限られたスペックによる音声認識から，クラウド技術を活用したサーバの潤沢なスペックによる音声認識技術まで，幅を広げている．それは車載機器のみでなく，たとえば，ロボットもスマートフォンなどの通信端末を介してクラウドに接続することにより，クラウド音声認識技術を活用することが可能である [18].

同様の手法を取れば，機器自体はクラウドとの通信機能を持っていなくとも，スマートフォンなどの別の通信端末と連携することにより，様々な機器が Internet of things (IoT) 機器としてクラウドと接続した機器となりうる．また，一般家電など移動体ではない機器においては，スマートフォンなどの通信端末を用いずとも，モデム・ルータなどから直接インターネットに接続可能なシステムが提案されている．これによる対話，空調制御，照明制御といったことが実現され，さらには，より効率的な制御が可能となるよう，デバイス上，ローカルネットワーク上，クラウド上にある推論器（入力に対し適切な応答を推論するもの）に制御を分散させる手法も提案されている [19].

このような技術革新により，様々な機器が音声認識操作に対応するようになった．しかし，それぞれの機器に向けて発せられる発話内容は必然と異なる．たとえば，IVI と家電照明機器に対し発する発話が異なるように，機器により発せられる発話の内容も異なるため，それぞれの機器に応じた対話システムが必要であり，機器に応じた学習データや言語理解器も必要となる．

対話システムには大きく分けて二種類あり，タスク指向型対話システムと非タスク指向型対話システムがある．なお，対話システムにはジェスチャーや表情といった非言語的なものもあるが，本研究では言語的なものを指す．

タスク指向型対話システムとは，話者発話の意図・意味を理解し，何かのタスクを達成（機能を実行）することである．たとえば，IVI 機器に対し「自宅に帰る」と言えば，経路案内機能により目的地を自宅に設定することがタスクであり，家電照明機器に対し「明かりをつけて」と言えば照明の電源を入れることがタスクとなる．本研究で着目しているのもタスク指向型である．これに対し，非タスク指向型システムとは，何かの操作をするのではなく，雑談など

対話そのものを行うシステムである [20]．非タスク指向型対話システムの代表的なものとして ELIZA [21] や，ALICE [22] などがある．

音声操作による対話システムに使用される言語理解器に関する研究も様々行われている．時代を遡れば，1981 年に実用化された ANSER システム [23] は，初期段階の対話システムであり，当時としては画期的なシステムであった．しかし，決まった単語を決まったタイミングで発話しなければ，システムが対応できないといった制約があった．近年では，統計的な手法に基づく機械学習により，未知な入力に対しても確率的に意図を推定するといった技術が検討され，統計には主に発話に含まれる単語（形態素）が用いられる [24]．たとえば，Homma らの研究では，カーナビの目的地検索用の発話を学習データとして機械学習させた言語理解器を作成した [25], [26]．この研究の特色は，音声認識の認識誤りデータを学習データに混ぜる optimal doping の手法を用いて意図推定精度を向上させている点である．

昨今の言語処理における機械学習では，Devin らの Bidirectional encoder representations from transformers (BERT) と呼ばれる，言語表現モデルが広く知られている [27]．BERT は，コーパスのラベルのないコンテキストを左右の方向から学習する．このモデルは質疑応答や言語推論など，様々な状況に対応するモデルを生成可能である．しかし，この技術を扱うには高度な処理能力が必要であり，文献 [27] によると，学習処理が完了するのに要した時間は，テンソルプロセッシングユニットを 16 個使ったシステムで四日間必要であった．さらに Lan らは，a lite BERT (ALBERT) と呼ばれるモデルを開発し，BERT から調整するパラメータを大幅に削減（ALBERT-large と BERT-large の比較では約 18 分の 1 に削減）した上で精度を向上させた [28].

- **疑似環境を用いて収集したデータを用いた研究**

言語理解器の学習に用いられる発話例の収集に関する研究も従来から行われてきた．特定の機器操作のための発話収集手法として，典型的な手法の一つが，システムを模倣した実験環境により，被験者には機械制御された対話システムを操作しているように思わせ，実際には別室の人間が対話システムを制御して

いる装置（いわゆる Wizard of Oz (WOZ) システム）との対話により発話を収集する WOZ 手法である [29]．WOZ 手法による先行研究として，Okamoto らは，京都のツアーガイドを行う Web エージェントを構築した [30]．この研究では，最初は WOZ システムを制御しているオペレータがユーザに対しシステムとしての応答を担うが，システムがオペレータとユーザの対話事例を集めながら次第に対話モデルを獲得し，人間のオペレータの代わりにツアーガイドを行える学習対話型エージェントを紹介している．アメリカ DARPA のプロジェクトでは，フライト情報サービス (ATIS) をタスクとして，WOZ 手法によるコーパス収集が行われた [31], [32]．このコーパスは ATIS コーパスとして知られ，多数の言語理解の研究に用いられた [33]-[35]．WOZ 手法は従来からある典型的な手法ではあるが，現在でも活用されており，人間が裏で制御を代行することにより，今はまだ存在しない（あるいは準備が困難な）システムを疑似的に動作させることが可能なため，対話システム以外にも活用されている．近年では，自動車シミュレータによる自動運転制御システムを WOZ システムで代用し，被験者（ドライバ）の状態を観測するといった研究にも活用されている [36], [37].

- Internet, SNS, **など既存のデータを用いた研究**

昨今では Internet を活用したデータ収集も行われている．Web クローリングなどにより，Internet 上にある既存のコンテンツを巡り，テキストデータを収集する手法が活発に用いられている．たとえば，Ahmaed らはマルチリンガルコーパスを収集するため，同様の意味を持つ英語とフランス語，また，英語とスペイン語の文章を収集した [38]．彼らは，この研究のために政府，保険，銀行をドメインとするバイリンガルのサイトを活用した．Mishra らの研究では，Wikipedia をクローリングすることにより，ソフトウェアエンジニアリングに関連する専門用語のテキストデータを収集した [39]．この研究で収集されたテキストデータに基づき，ソフトウェアエンジニアリングに関する専門用語（「クッキー」や「ウィルス」など）がソフトウエア開発業界と，他の業界では異なる意味合いになることを認識するモデルを機械学習により作成した．また，Internet の中でも特に Social Networking Service (SNS) を活用し，テキストを収集，解析する研究も盛んに行われている．たとえば，小池らの研究では動画共有サ

ービスにおいて人気の出るタイトルジェネレータを開発した．この研究では，収集したタイトルの文長を調査したところ 98%が 11 から 60 の文長であったため，SNS 投稿から同文長の投稿のみを抽出し，ジェネレータの学習データとして活用している [40]．Liu らは，SNS から収集したテキストデータを活用し，中国語の文章の類似性を算出するモデルを作成した [41]．Le らは，SNS 上にある多言語により記入されたレストランのレビューより感情分析を行った [42]．その結果，ユーザのレーティングは言語や場所によって影響を受けることを示した．松本らは，下関市周辺の観光に関する SNS への投稿を収集した．地元住民の投稿ではなく，観光客の投稿にフォーカスするようにプロフィール情報などからテキストマイニングを行い，投稿内容から新たな観光資源の発見を試みている [43]．泉らは，災害に関連するキーワードを用いた SNS 投稿を抽出し，その中から有益な情報と，不必要な情報を分類し，BERT に学習させることで有益な投稿を判別するモデルを構築した [44]．

- **Internet, SNS, 以外で既存のデータを用いた研究**

この他にも類似した研究として，Internet や SNS に限らず既存のコンテンツを活用した研究という意味では，安藤らは LaboroTVSpeech というコーパスを構築している．LaboroTVSpeech は，テレビ番組の音声・字幕データを基にしたコーパスであり，構築手順を自動化することにより約 2000 時間ものデータを収集し，大規模なコーパスを構築する研究を行っている [45]．また，製品・サービスのレビューやユーザ利用ログデータ（本研究では「ログデータ」とは Internet, SNS, などで一般に公開されたものではなく，ベンダーが管理する製品，サービスの利用ログデータを指す）も膨大な既存データと言うことができ，ログデータを解析しユーザ嗜好に合わせたレコメンドシステムなどのサービス向上に用いられている [46]．

1.3. 本研究が掲げる既存の発話収集手法の分類とその特徴

本研究は発話収集に関するものであり，本研究ではその手法を三種類に分類

した．一種類目は，WOZ 手法に代表されるような「疑似的な環境を用いた発話収集手法」である．二種類目は，ホームオーディオやスマートフォンのアシスタント機能などに発せられた発話ログデータを用いた「対象となるドメインの発話ログデータを活用する手法」である．なお，本研究における「発話ログデータ」とは，発話を音声認識した結果のテキスト文を意味する．三種類目は，Internet, SNS, などの多種多様で膨大なデータを用いた「関連性を問わず膨大なデータより抽出する手法」である．以下にそれぞれの手法と比較した場合の長所と短所を述べる．

- **「疑似的な環境を用いた発話収集手法」の長所短所**

この手法は，WOZ 手法に代表される手法であり，収集環境は目的とするシステムを疑似的に構築し，目的とするドメインと共通のドメインにより収集する手法である．

・長所

この手法の長所は，収集した発話はそのままで目的とするドメインの発話例であるという点である．WOZ 手法にしても，目的のドメインを疑似的なシステムにより模倣することで発話収集することを目的とする手法であるため，必然と必要な発話例を収集可能であることが他の手法にはない利点である．

・短所

この手法の短所となりうるのが，他の手法に比べ収集作業に手間がかかる点と，収集量を増やしづらい点である．まず，ドメインとなる操作環境を模倣する実験環境を用意する必要がある．WOZ 手法であれば，モニター，カメラ，マイク，スピーカといった機器を備えた WOZ システムを構築する必要がある．次に，収集量を増やすには被験者数を増やす必要があるが，被験者数を増やせば，その分，収集作業時間が多くかかる関係がある．作業時間短縮のため，並行作業を行う方法もあるが，その場合，WOZ システムなどの実験環境や，システムを裏で制御するオペレータを並行作業分追加で用意する必要もある．

- **「対象となるドメインの発話ログデータを活用する手法」の長所短所**

この手法は，発話ログデータを用いたものであり，その発話ログデータは，目的とするシステムと同じ，あるいは類似するシステムを用い，目的とするドメインと同じ，あるいは類似するドメインにより収集する手法である．

・長所

この手法の長所は，一般ユーザの発話ログデータであるため，言い淀みによる発話の失敗などをスクリーニングする必要はあるが，基本的には発話ログデータは，そのまま発話例となる点である．そのドメインのサービスが継続する限り，発話ログデータは増加するため，言語理解器の追加学習データとすることも可能である．また，そのサービスを提供するベンダーにとって無料で扱える点も利点である．

・短所

この手法の短所となる点は，対象ではないドメイン，すなわち，ジャンルが異なるようなドメインの発話となると，目的とする発話はログデータに表れにくくなることが挙げられる．「次へ」などの汎用的な発話であれば他のドメインにおいても流用可能であるが，たとえば，「航空機の音声予約ドメイン」と「楽曲再生の音声操作ドメイン」では，使われる単語は必然と異なるため，単純には発話ログデータの流用はできない．この点は，目的のドメインのための発話を収集する「疑似的な環境を用いた発話収集手法」の方が有利である．

また，発話ログデータはあくまでログデータであるため，過去の発話である点に留意する必要がある．話者が発話対象の機器を，「単純な機械」と認識すれば「機械に理解しやすい発話」を行い，「高度な言語処理技術を搭載した機器」と認識すれば「より自然で自由な発話」をする．したがって，現状の発話ログデータが，将来に渡るユーザの発話を反映しているとは限らない．そのため，「機械に理解しやすい発話」から「より自然で自由な発話」へ変化することが予想される新たな開発を行う場合は，「疑似的な環境を用いた発話収集手法」により，新たに検討しているシステム，サービスに沿った収集が必要となる．

- **「関連性を問わず膨大なデータより抽出する手法」の長所短所**

この手法は，目的とするシステム，ドメインとの関連性を問わない膨大なデータ資源より，求めるデータを抽出する手法である．

・長所

この手法の長所は，大量のデータを収集可能という点である．特には，Internet, SNS, といった膨大なデータより抽出する手法であり「疑似的な環境を用いた発話収集手法」と比較してデータ量の視点で有利である．著作権やプライバシーといった考慮は必要であるが，基本的には無料で扱える点も利点である．

・短所

この手法の短所となりうるのが，膨大なデータ量であるために，何が重要なデータか検討し抽出しなければならない点である．確固たる見識がないまま収集したのでは，そのデータの良し悪しを検証しようもなく，使い物にならない可能性もありうる [47]．したがって，特定のドメインにおいて，書き言語，話し言語の違い [48], [49] などを考慮した上，「発話例」となるテキストのみを抽出するといった条件を綿密に検討する必要性を考慮すると，収集したものがそのまま発話例となる「疑似的な環境を用いた発話収集手法」や「対象となるドメインの発話ログデータを活用する手法」より困難である．

本手法の別視点の用途として，大量のデータから意味ある情報を引き出すために，マイニングツールを活用する用途がある．昨今のテキストマイニングでは，テキストデータを単語分割し数量化することにより，統計的な単語間の関連性を出力するものがある．このようにして得られた解析結果は，発話例ではないが，目的とするドメインと関連性が強い単語，文章を調査するのに役立つ．これまでに紹介した Internet, SNS, を活用した研究においても，「発話例」の収集ではなく，テキストデータの調査，分析に関するものである [38], [41], [44].

以上の特徴を踏まえ，各手法の比較を行ったものが表 1 である．

表 1．各発話収集手法の長所短所比較

手法	開発対象との差異	収集作業時間	発話データとしての質	収集量
疑似的な環境を用いた発話収集手法	システム：疑似 ドメイン：共通	△	◎	○
対象となるドメインの発話ログデータを活用する手法	システム：類似 ドメイン：類似	○*1	○	◎/×
関連性を問わず膨大なデータより抽出する手法	不問	○*1	△	◎

◎：必要十分な成果を得られる．

○：必要を満たす程度の成果を得られる．

△：不可能ではないが課題がある．

×：不可能である．

*1：当該手法でも作業時間を要するが，「疑似的な環境を用いた発話収集手法」より収集量が膨大であり，収集量に対する作業時間としては有利とし「○」とした．

表 1 と各手法の長所短所から説明を補足する．

- 「疑似的な環境を用いた発話収集手法」は，目的とするドメインの発話を収集するための手法であるため，必要な発話を収集する手法としては最良と言える．しかし，収集作業には手間も要する．
- 「対象となるドメインの発話ログデータを活用する手法」にて収集量を「◎/×」，としたのは，ログデータはベンダーの情報資産であり，ユーザのプライバシー保護の観点からも第三者が利用可能なわけではない．したがって，「◎」は，ログデータを所有するベンダーの場合であり，ログデータを持たない第三者にとっては「×」となる．

- 「関連性を問わず膨大なデータより抽出する手法」は，膨大なデータ量を扱えるが，発話例を抽出するというよりは，膨大なデータ量から特徴を見出す分析，調査に向いている．

以上から，発話収集手法としては「疑似的な環境を用いた発話収集手法」，「対象となるドメインの発話ログデータを活用する手法」が適当な手法と言える．表 1 より，有利な手法は「対象となるドメインの発話ログデータを活用する手法」である．ただし，対象ではないドメインや，新規ドメインの開発においては「疑似的な環境を用いた発話収集手法」を検討することになる．また，発話ログデータは過去の発話であるため，将来に渡る発話を網羅しているとは限らない点を留意する必要がある．

1.4. 本研究における目的・目標

1.1 節で述べたように，ユーザの発話は，高性能な音声認識技術を使用するに連れ，定型音声コマンドのような「機械が理解しやすい言い方」から，より自然で自由な発話へと多様化することが予想される．また，ユーザがシステムを高度な言語処理技術を搭載した賢い機器と認識すれば，発話もより自然で自由な発話へと多様化する．将来的には人間同士の会話のような話し方で機器操作したいというニーズが生まれると予想される．そのため，**多様化した発話でも操作可能なシステムの開発・評価のため，より自然で自由な発想による多様化した発話例を整備すること**が本研究の目的である．

この目的の課題は，単純に WOZ 手法などの被験者に発話対象を機械と意識させるような「疑似的な環境を用いた発話収集手法」では，より自然で自由な発想による多様化した発話にはなりえない点である．また，「対象となるドメインの発話ログデータを活用する手法」により収集した発話は，ホームオーディオやスマートフォンといった機械に向けた過去の発話である．そのため，発話ログデータから発話例を収集するというよりは，「発話ログデータより自然で自由な発想による多様化した発話を収集する」とする方が本研究の目的に即

している．なお，発話ログデータを人為的な加工により多様化させ新たな発話とするのは「より自然な発話」という視点で適切とは言えない．

そこで本研究では，収集環境やプロセスに工夫の余地がある「疑似的な環境を用いた発話収集手法」を用いることとし，以下の特徴を持つ手法を提案することを目標と定めた．

<< 目標 1 >>

より自然で自由な発想による多様化した発話を収集可能

<< 目標 2 >>

手法の実施負荷低減

目標 1 を達成するため，本研究では，WOZ システムといった機器を用いないことで，被験者に発話対象は機械という認識を低下させ，代わりに被験者の発話の聞き手に人間を介在させることで，より自然で自由な発想による多様化した発話を想起させる．

目標 2 の達成においては，まず，目標 1 の WOZ システムといった特殊な機器を用意する必要がない点が実施の容易さに繋がる．さらに，限られた作業時間において被験者により多くの発話を想起させることにより，発話あたりの収集作業時間の視点から負荷低減を行う．

1.5. 本論文の構成

本論文は以下のように構成されている．

第 2 章では，コーパスを用いた研究について調査研究を行う．またその中で本研究における「コーパス」と，「発話の多様性」の定義を行う．

第 3 章では，新しい「疑似的な環境を用いた発話収集手法」として「インタビュー形式による多様性の高い機器操作発話収集手法」を提案する．ドメインはカーナビの目的地検索 [point of interest (POI) 検索] である．提案手法により収集したコーパスと，商用カーナビの製品ログデータを比較し，本手法の有用性を示す．

第 4 章では，インタビュー形式によるコーパス収集手法の中で用いた，probing の有用性と時間的負荷低減効果について検証する．Probing は，被験者より，1 シチュエーションから複数の回答を収集可能にする．これにより，従来手法の第一発話のみ収集する手法に対し，どの程度被験者数を減らせるか検証する．また，商用カーナビの製品ログデータに含まれる形態素のカバレッジを検証し，有用性視点の検証を行う．最終的に，probing により発話数を増加した場合と，従来の第一発話のみを収集した場合の時間的負荷を比較し，作業負荷をどの程度低減可能かを示す.

第 5 章では，新たな「疑似的な環境を用いた発話収集手法」として，「発話対象を人間と想定した Web アンケートによるコーパス収集手法」を提案する．本提案では，「介護を受ける」ドメインのもと，Web アンケート上においてシチュエーションの提示を行い，そのシチュエーションにおける発話をアンケートの回答文として収集する．まず，小規模での収集実験を行い，その知見を基に被験者数を増加した収集実験を行う．また，本実験の収集作業に要した時間と第 3 章のコーパス収集作業に要した時間を比較し，どの程度時間的負荷を低減できたか示すとともに，形態素解析による検証により，本手法の有用性を示す.

第 6 章では，結論として，本研究のまとめと成果，また今後の課題について述べる.

第2章　コーパスに関する研究と発話の多様性

本章では，コーパス関連研究について調査した内容を示すとともに，コーパスを用いた研究における本研究の位置づけと，本研究におけるコーパスと発話の多様性の定義について述べる．また，関連する用語についても説明する．

2.1. コーパスを用いた研究と本研究におけるコーパスの定義

コーパスとは，記載された言語や話された言語を記録した言語資料である．昨今では，コンピュータの発展・普及により，こうした情報を電子化することにより蓄積可能となった．こうしたコーパスは電子コーパスと呼ばれ，一般化しつつあるため，単にコーパスと呼ぶこともある．収集した情報をそのままに，付加情報を加えていないものを「生コーパス」と呼び，品詞や構文構造など何らかの情報を付加したものを「タグ付けコーパス」と呼ぶ．

さらにコーパスの大別として，「テキストコーパス」と「音声コーパス」がある．テキストコーパスとは，テキストを収集したものであり，新聞，小説などに記載された言語や，演説，会話などを書き起こした文を集めたコーパスである．音声コーパスとは，音声を収集したものであり，新聞，小説などに記載されたものを読み上げたものや，演説，会話などの録音音声を集めたコーパスである．

テキストコーパスと音声コーパスの違いとして，テキストコーパスには，文，単語の意味や構文構造といった言語情報がよく表される．一方，音声コーパスには，音声の抑揚や声の高さ，話速といったものが含まれ，感情などの非言語情報・パラ言語情報が含まれる．例として，静かな環境の中で発話した場合と，騒音環境下で発話した場合とではロンバード効果により音声特性に変動が生じる [50]．こうした音としての情報は，テキストコーパスでは発話内容は同じため同一文となり特性の差が見えにくいが，音声コーパスであれば非言語情報・パラ言語情報として記録される．

また，書き言語を収集したものは「文字言語コーパス」と呼ばれ，話し言語を収集したものは「音声言語コーパス」と呼ばれる．

これらのコーパスは「生 – タグ付き」,「音声 – テキスト」,「文字言語 – 音声言語」のそれぞれが直交するような関係にある．この区分のイメージを示したものが図 1 である．

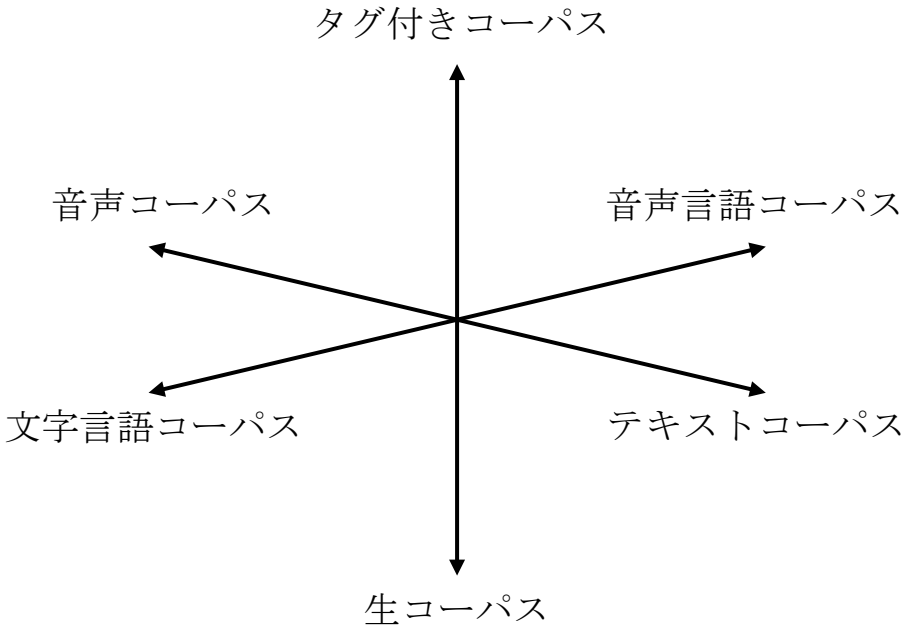


図 1．コーパスの区分

テキストコーパスは，米国英語の調査を目的に 1964 年に公開された Brown corpus (Brown University Standard Corpus of Present-Day American English) が初期のコーパスとして有名であり，米国出版物を対象に約 100 万語より構成されている．これに対し，英国英語を対象としたものが 1978 年の LOB corpus (Lancaster/Oslo-Bergen Corpus of British English) である．Brown corpus と比較可能にするため，収集年や規模などが共通となるように構築されている [51]．音声コーパスにおいては，音声資源に関する国際協調を推進するため，Coordinating Committee for Speech Databases and Speech I/O Systems Assessment (COCOSDA) [52] が 1991 年に設立された．これをもとにして，米国では Linguistic Data Consortium (LDC) [53] が 1992 年に設立され，TIMIT, RM, ATIS, SWITCHBOARD,

WSJ, などの音声コーパスを販売配布している．同様の機関としてヨーロッパでは 1995 年に European Language Resources Association (ELRA) [54] が設立し，BABEL, EUROM1/MULTEXT, GLOBALPHONE, などの音声コーパスを販売提供している．こうした団体は LDC, ELRA, のみではなく，アジアを含め各地域にて類似した組織が立ち上がっている [55], [56].

日本におけるコーパスの歴史としては，1952 年に新聞に使われた用語を調査することが始まりと言われている．その後，総合雑誌など様々な媒体で用語利用の調査は行われたものの，コーパスの構築には至らなかったとされている．この一つの要因は，JIS 漢字コードの制定は 1978 年であり，これ以前の調査では，コンピュータ上では日本語を適切に扱えないという課題であった．また，日本語は分かち書きされない言語であるため，用語利用調査を目的として文例は収集したものの，当時の段階では単純には単語数を調べることができなかったとされている [47].

その後，他の言語資料，コンピュータによる形態素解析技術の発展に伴い，現在では，国立国語研究所より「現代日本語書き言葉均衡コーパス (BCCWJ: Balanced Corpus of Contemporary Written Japanese)」，「日本語話し言葉コーパス (CSJ: Corpus of Spontaneous Japanese)」など各種大規模コーパスを入手可能になった [57]．また，BCCWJ, CSJ は現代日本語の書き言語，話し言語の全体像を表した基準的立ち位置にあるコーパスであり，2011 年 (BCCWJ)，2004 年（CSJ の第一刷）に公開され，研究に用いられている [58]-[61].

この他にも，形態素解析ツールなどの発展にともない，研究者自身が研究目的に沿った独自のコーパスを用意し解析可能になった．昨今では，1.3 節で示したように，新たにコーパスを収集するのではなく，Internet, SNS, などの既存のデータを活用する研究も行われている．特にログデータでは，発話ログデータに限らずユーザのあらゆるログデータ（アクティビティ）から，協調フィルタリングなどを活用し，ユーザの嗜好に合わせたサービス提供資源としての役割も持つようになってきている [46], [62].

- **コーパスを言語資料として用いた研究**

コーパスを言語資料として用いた研究では，小河らは，小学生の単語学習に

よる発達過程の検討資料として，4 年ごとに改定される小学校の国語の教科書をコーパスとし，学年別に教科書に出現する品詞の延べ単語数，単語の種類数，出現頻度などを解析した [63]．李らは，現代書き言葉の原型が確立した時期と見なされる 1910 年から 2014 年に出版された小説において，各年で発刊された 5 から 6 作品より各冊で約 5000 文字を目安に抽出しコーパスとして構築した．このコーパスにおいて助詞に注目し，時間経過とともに使用率の高くなった助詞，低くなった助詞を調査し語学，文体学の資料とする考察を行っている [64]．村井らは，日本語は，音声としての発話（話し言語）と文章（書き言語）では違いがあるということに着目し，その中でも自称詞（一人称で自分を指す言語）と対称詞（二人称で相手を指す言語）が会話文の中でどのように出現するかをコーパスより統計的に検証を行った [65]．李は，日本語教育において教師の経験に基づく勘も重要な要素としつつ，語彙・単語の出現頻度といった，計量学的なコーパス調査に基づく教育方針の検討も有効であると提起している [66]．

このように，言語資料であるコーパスは言語学，文学，教育といった分野で多く活用されている [67]-[69]．

- **コーパスを開発資源として用いた研究**

昨今では，コーパスを言語資料（文学・言語学などのための資料）としてではなく，対話システムなどの開発資源に応用した研究も盛んに行われている．たとえば，音声合成システム，あるいは，テキスト読み上げは Text to speech (TTS) とも呼ばれ，機械に発話させる装置として利用されている．TTS は対話システムのシステム側の発話機能として利用されている他，ニュースや書籍などテキストコンテンツを読み上げることで，ユーザにはテキストから音声情報へ変換して提供するといった用途もある [70]-[73]．

TTS の従来の手法では，規則合成方式と呼ばれる音声・言語のパラメータより規則形式で実装した手法が用いられていた．その後，音声データを音節型，音韻連鎖型といった特定の母音や子音の並びで音を区切った波形素片をあらかじめコーパスから抽出し，その中から音声を合成するコーパスベースの手法（データドリブン方式）が扱われていた [74]．その後の研究で，Sagisaka らは波形素片の長さを限定しない TTS 開発手法を紹介している [75], [76]．さらに，

動的に波形素片の並びを変動させる手法による音質向上が行われ，この研究は後のコーパスを基に開発される TTS の糸口となり v-Talk と呼ばれる TTS が開発された [77], [78]．その後，XIMERA といった後継のコーパスベース TTS 開発へとつながっている [79].

音声認識技術の開発においてもコーパスは活用されており，ディクテーション向上のため N-gram による言語モデル構築にはテキストコーパスが用いられている [80]．一方，音響モデルも踏まえて検討すると，記載されたものを読み上げた音声（朗読音声）と，自然に発した音声では，文法的特徴のみでなく，音声の特徴も異なる．自然に発した音声を認識させるためには，学習に用いるコーパスも「音声言語コーパス」を用いるのが望ましい．文字言語コーパスであれば，記載された文章を読めば音声もテキストも揃うが，音声言語コーパスは，発話された音声を書き起こす分，手間を要するため大規模なコーパスを用意するのは困難である [81]．そのため，WSJ, JNAS [82], といった，新聞などを読み上げた音声コーパス（文字言語音声コーパス）を用いた開発から始まり，CSJ のような話し言語のコーパス（音声言語音声コーパス）へと移り，さらに雑音対策が行われていった [83].

コーパスを言語処理・言語理解器の学習・評価データとする研究も行われている．倉田らは，被験者に状況テンプレートと呼ばれる，被験者の置かれているシチュエーションを示した教示文，イラスト，図を用意し車のオーディオ，エアコン，カーナビを操作対象としたコーパスを，言語理解器の学習データと評価用データとして収集している [84]．この実験では，各ドメインのタスク達成率は 92%以上となった．この他にも，1.2 節で紹介している言語処理・言語理解器に関する研究もコーパスを用いたものである．たとえば，Homma らの研究は独自で収集したコーパスを用いた言語理解器に関する研究である [25], [26]. Okamoto らの研究は，WOZ 手法によるコーパスを収集しつつ学習も並行する研究である [30]．目的に応じたコーパスを収集するのではなく，既存のコーパスとして ATIS を活用した研究が文献 [33]-[35] である．

特許の明細書なども，機械翻訳のための言語処理を検討するコーパスとして

研究に用いられている．こうした資料は，日本語と英語等で意味が一致するパラレルコーパスとして貴重な資源である．特許明細書には，その分野の専門用語や，まだ対訳辞書化されていない未知語が用いられるため，これらを適切に認識することにより，機械翻訳精度を向上させることが可能である [85]．下畑らの研究では，こうした用語の対訳辞書作成のため，特許関連の書類をコーパスとして用語の自動抽出を行った [86].

- **本研究におけるコーパスの定義**

本研究はタスク指向型，つまり特定の機器操作を意図した発話例の収集に関するものである．したがって，コーパスを開発資源とする研究に分類される．

用いるコーパスは，「疑似的な環境を用いた発話収集手法」を採択しているため，被験者の発話（話し言語）を収集したものであり「音声言語コーパス」となる．さらに，多様な発話を言語理解させるために学習させるデータはテキストベースであるため，発話を書き起こしたもの（以後，発話文と呼ぶ）を集めたテキストコーパスである．これに，教師データとなるよう発話文の意図などのタグ付けを行う．したがって，本研究で用いるコーパスは「タグ付き音声言語テキストコーパス」と言える．これが本研究におけるコーパスであり，以後，これを単にコーパスと呼ぶ.

IVI のカーナビ機能において POI 検索することを想定した発話の場合，本研究におけるコーパスのイメージは表 2 のようなものである．

表 2．本研究における POI 検索ドメインを想定したコーパス例

Utterance	Intent	Query
近くのコンビニ	周辺検索	コンビニ
ここら辺のレストラン	周辺検索	レストラン
海に行きたい	単純検索	海
きれいな公園	単純検索	きれいな公園

各項目について説明する．Utterance とは話者が発した発話をテキストに書き起こした発話文である．Intent とは，話者がその発話により IVI に期待する操作

内容であり，数個から数十個存在する．一方，Query とは話者が POI 検索として具体的に探したい目的地に関するキーワードとなる部分を抽出したものであり，語彙を限定せず受け付ける．

このように本研究におけるコーパスは，発話文に対し Intent タグを付与し，Query 抽出を行ったデータベースのような構造になっている．実際のコーパスには，さらに話者の素性を表す管理コードや，その他属性もあるが，本研究とは関連が薄いため割愛する．

2.2. 形態素と形態素解析

日本語の場合，文章により単語の活用が変化するため，活用によって変化する部分と変化しない部分に分かれる．このようにして分けたものを「形態素」と呼ぶ [81]．ただし，研究者によっては，必ずしもこれが形態素の定義として一致しておらず，単に「単語」が形態素を意味する場合もある．日本語では漢字，ひらがな，カタカナなど多様な表現方法があるため，表記上の多様性が高く一単語がどこまでか，そもそも「単語」の定義は何かを決めるのは困難であるとされている [63].

「形態素解析」とは，文章を形態素単位に区切り，品詞や読みなど，形態素に付帯される情報を付与する技術である [47]．従来は手作業にて解析を行い，情報を付与していたが，昨今では，コンピュータにより自動で品詞や読みを割り振ることが可能となり，言語処理を行う上で重要な役割を担っている．特に日本語においては，分かち書きを行わない言語である．すなわち，英語などでは単語の区切りは空白で示されるため，形態素解析も各単語（各形態素）に対し行われる．一方，日本語では一目では単語の区切り（形態素の区切り）が分からないため，形態素解析ツールは形態素としての解析を行う前に，連続する文字列を区切った後，品詞付与といった処理が必要である [87]．同様に分かち書きをしない言語としては中国語やタイ語などがある．

こうした手間のかかる形態素解析であるが，日本語用に自動で形態素解析を行う無償ツールがあり，JUMAN, ChaSen, MeCab などがある [88]-[90]．前述した

ように，形態素の厳密な定義には不明瞭な部分があり，文章，使用したツールによっては形態素解析結果に差が生じることが予想される．そのため，本研究では統一して MeCab を用いて形態素解析を行う．

一般的な言語処理においてはこのようなツールで解析した結果を用いる．本研究もこれに該当する．高精度な形態素解析を必要とする場合，たとえば，形態素解析ツールに学習させるためのコーパス作成を目的とするといった場合は，高信頼性のあるコーパスが必要であるため，まず現状の形態素解析ツールで解析結果を得た後，さらに人手による確認，修正を加える場合もある [91], [92].

このように形態素解析結果を用いて，後段のシステムにて解析した発話の意図推定，トピック推定，専門書の索引選定や，形態素解析に要する新たな辞書の作成などに用いられている [26], [93]-[95]

2.3. 本研究における発話の多様性の定義

本研究では「より自然で自由な発想による多様化した発話」を表すため，「発話の多様性」や「多様性の高い発話」といった表現を用いる．「発話の多様性が高い」とは，「より自然で自由な発想による多様化した発話がされている」という意味を示す．反対に，「発話の多様性が低い」とは，定型音声コマンドのような，決まった単語で端的な発話の傾向を示す．

発話の多様性を検証するため，本研究では形態素の量の視点と，形態素のバリエーションの視点より検証する．

● 形態素の量視点による検証指標

人間は発話対象が人間か機械かによって発話の特性が異なることが分かっている．具体的には一回の発話に含まれる形態素数が異なり，機械に対する発話に比べ，人間に対する発話の方が一発話あたりの形態素数が多いことが Takeda ら [96]，河口ら [97] の研究で示されている．

次に，一発話あたりの形態素数$\bar{x}$は以下の式で表される．

$$\bar{x} = \frac{x}{n_{utt}} \tag{2.1}$$

ここで，xは延べ形態素数を示し，n_{utt}は発話数を意味する．一発話あたりの形態素数が多い方が，より人間との会話に近い自然で自由な発想による発話がなされているものとする．

- **形態素のバリエーション視点による検証指標**

人間に対する発話の方が一発話あたりの形態素数が多くなるため，この値が大きければ，そのコーパスは人間との会話に近い発話が含まれていると言える．しかし，各被験者（あるいは各発話）が似通った単語を用いた発話をしていたのであれば，一発話あたりの形態素数が増加したとしても多様な発話を収集したとは言い切れない．

そこで，本研究ではコーパスに含まれる異なり形態素に着目する．異なり形態素数が多いということは，そのコーパスには多様な形態素（多様な単語表現）が含まれていると言えるため，異なり形態素数自体が発話の多様性を示す一つの指標と言える．この他，異なり形態素数を用いた多様性検証には，異なり形態素数比 [type / token ratio (TTR)] が用いられている [98]．TTR は以下の式で表される．

$$\mathrm{TTR} = \frac{V}{x} \tag{2.2}$$

ここで，Vは延べ形態素数xに含まれる異なり形態素数を示す．

鈴木らの研究でも，所信表明演説など政治テキスト分析の中で多様性の指標として TTR を用いている [98]．しかし，本研究においては TTR を使用するには問題がある．演説やスピーチなどでは，基本的に一つの大きな議題に対し，一人の話者が話し続ける．これに対し，本研究のコーパスは，タスク指向型の発話であるため，特定のドメインにおける機器操作発話を，複数の被験者より複数収集するというものである．したがって，解析は一回の演説・スピーチではなく，収集した複数の発話に対し行われるため，分母とするのは延べ形態素数

ではなく発話数とするのが適切である．そこで本研究では，TTR から変更を加え「一発話あたりの異なり形態素数比 [type / utterance ratio (TUR)]」を以下の式にて定義する．

$$\mathrm{TUR} = \frac{V}{n_{utt}} \tag{2.3}$$

TUR が大きければ，そのコーパス，あるいは収集手法は，異なり形態素が出現しやすい傾向を示し，反対に，TUR が小さければ各発話は，似た単語表現が用いられていることを示している．TUR を比較することにより，「疑似的な環境を用いた発話収集手法」の限られた発話数の中で，どれだけ効率的に多様な形態素（異なり形態素）を収集しているかを検証することが可能となる．また，用いられた収集手法が異なり形態素を出現させやすい手法であるかを確認する指標ともなる．

ただし，複数のコーパス，収集手法に対し TUR を比較する場合は，被験者数といった発話数の基となる指標を揃えた上で比較する必要がある．

以上より，本研究では発話の多様性を確認するにあたり，上記に挙げた複数の指標により，提案手法が従来手法に対し，より多様な発話となっているかを検証する．すなわち，一発話あたりの形態素数より，人間同士の会話にみられるような，より自然で自由な発想による発話かどうかを確認し，異なり形態素数からそのコーパスに含まれる形態素の多様性を確認し，TUR によりその収集手法が多様な形態素を効率的に収集する手法かを確認する．

第3章　インタビュー形式による多様性の高い機器操作発話収集手法

本章では，多様性の高い発話の収集手法として，インタビュー形式による多様性の高い機器操作発話収集手法の提案について述べる．本手法の特徴は二点あり，一つは WOZ システムなどの特別な実験環境を用いず，オペレータがシチュエーションを提示し被験者より回答発話を得ることで「疑似的な環境を用いた発話収集手法」の課題である作業の手間を低減させている点である．もう一つは，一人の被験者より 1 シチュエーションから複数の発話を収集する probing という工夫を行い，発話の収集量を増加させている点である [99][*1].

3.1. はじめに

昨今の IVI, カーナビを始めとする車載機器では，スマートフォンなどの通信端末を介し，クラウドサービスの提供が行われている [100]．逆説的には，スマートフォンなどの端末に比べ，車載機器のクラウド進出は一歩遅れた格好である．こうしたクラウドサービスの中には，クラウド音声認識技術を用いた POI 検索機能もある [101], [102]．POI 検索の仕方には，「自車位置周辺の検索」，「目的地周辺の検索」など様々な検索方法があるが，ユーザからすれば，これらの音声操作をサーバスペックによる高精度な音声認識技術を用いて行うという経験は，スマートフォンに比べまだ浅いことになる．こうした状況は，本研究が想定する「機械が理解しやすい言い方」から多様性の高い発話へと変化していく状況と合致する．

そのため本章では，POI 検索ドメインをターゲットとし，「インタビュー形式による多様性の高い機器操作発話収集手法」を提案・実施することとした．

---

*1 本章は研究業績 [c]の内容に基づいています．

3.2. 背景と課題

クラウド音声認識によって音声認識性能が向上し，様々な発話を音声認識（音声のテキスト化）することが可能になった．また，スマートフォンやホームオーディオなどでは，すでにクラウド音声認識を活用したアシスタント機能が一般消費者向けに提供されている [103], [104]．IVI, カーナビなどの車載機器においてもクラウド音声認識技術を活用したサービスが提供されている [100]-[102]．こうした高性能な音声認識技術に触れたユーザは，次第に機器に向けて多様な言い回しをすることが予想される．また，カーナビから IVI のように，より高度な技術を用いた製品になれば，ユーザの認識も高度な言語処理技術が使われた機器と捉え，発話の多様性が高まることが予想される．すなわち，従来の音声操作として定型音声コマンドベースの発話をしていたユーザが，人に話しかけるような多様性の高い発話で機器を操作したいといったニーズや状況が発生すると考えられ，こうしたニーズ・状況に対応するには，開発に用いるコーパスにも多様性の高い発話を含んでいる必要がある．

先行研究におけるコーパス収集では，倉田らの研究で用いた手法が従来手法の例となる．彼らの試みでは，被験者に状況テンプレートと呼ばれる被験者の置かれているシチュエーションを示した教示文，イラスト，図を用意し「あなたは車を運転しています．この車の機能は音声認識で操作できます．今から提示される状況になったとき，自分の要求を満たすために，あなたなら何と命令しますか．ただし，相手はただの機械です．コマンドの認識にしか対応していません．また，一文で入力してもらわないと対応できません.」といった指示を行い，車のオーディオ，エアコン，カーナビを対象としたコーパス収集を行っている [84]．また，同文献によれば，「オーディオオン」，「次の曲」と言った実際のコマンドを例示したとある．この手法は，本研究視点においては WOZ 手法に近く，発話対象は機械であると意識させた疑似的な環境を用いた発話収集手法として有効な手法である．その一方，WOZ 手法と同様に，被験者は発話対象が機械であると認識しているため，定型音声コマンドベースの発話が収集されたと予想される．

カーナビ（IVI の経路案内機能とも言えるが，本章ではまとめて「カーナビ」と呼ぶ）といった機器に対する発話でありつつも，多様性の高い発話収集が可能な手法を検討する必要がある．

3.3. 提案手法の概略

カーナビの音声操作による POI 検索をドメインとし，「インタビュー形式による多様性の高い機器操作発話収集手法」を提案する．以後，本手法を「インタビュー形式による収集手法」と表す．本手法の最大の特徴は，WOZ システムのような特殊な機器は用いず，代わりに被験者にシチュエーションを提示するオペレータを置くことであり，インタビュアとして「○○の時，どのように言いますか？」と被験者にインタビューし，回答発話を収集し書き起こすことでコーパスとすることである．

前述したように，人間が機械に向かって発話する場合は，一発話あたりの形態素数が減少する [96], [97]．そのため，機械に対する発話を人間であるオペレータが聞き手になることにより，機械向けの発話でありつつも，発話の多様性が高まることを想定した手法である．また，被験者はカーナビに対する発話を問われているため，収集される発話は，機械向けの発話と，人間同士の会話のような発話との間にあたる発話を想定している．本提案手法によるコーパスの立ち位置の概念を図 2 に示す．

これに加え，発話収集量を増加させるため，probing という手法をインタビューの中で用いる．被験者数はそのままに，発話収集量が増加すれば，発話あたりの作業時間としては減少させることが可能になる．Probing の詳細については次節にて述べる．

オペレータが被験者にインタビューを行い，発話例を収集するため，類似したコーパスとして「会話コーパス」あるいは「対話コーパス」と言われるものがある [105], [106]．これらとの違いについて述べる．「会話コーパス」，「対話コーパス」は，まさに人間同士の「会話」となるコーパスである．実際，提案手法は，オペレータと被験者がインタビューという状況下において会話を行う．

しかし，提案手法のインタビュー形式による収集手法のコーパスは，オペレータがシチュエーションを被験者に提示し，回答発話のみを抽出し書き起こす．この書き起こした発話文をコーパスとする．すなわち，提案手法はインタビューにより抽出する「各シチュエーションに対する回答発話のコーパス」である．

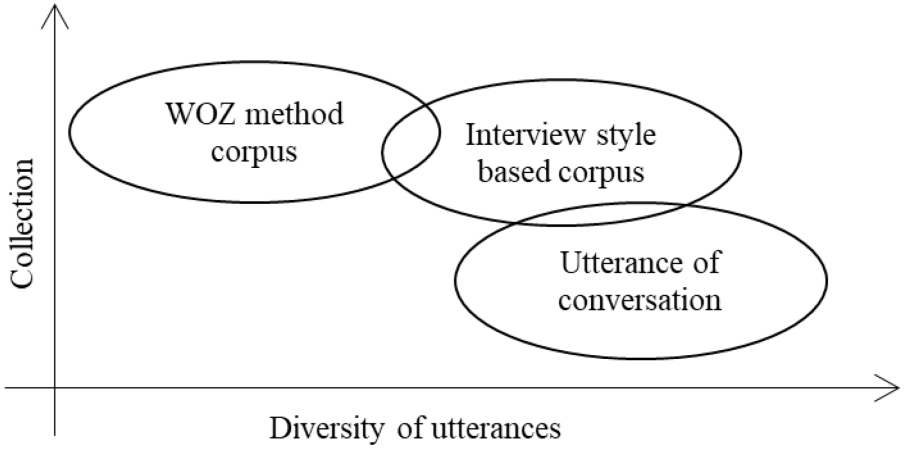


図 2．インタビュー手法によるコーパスのイメージ [Copyright (C) 2018 IEICE]

3.4. 提案手法の詳細と収集結果

準備作業として，インタビューにて被験者に提示するシチュエーションの検討を行う必要がある．そのため，ユーザが音声操作により POI 検索を行い，検索結果から希望の POI を目的地に設定，その後，カーナビがルート案内を開始する．この一連の作業のユースケースを検討し，約 50 設問からなるシチュエーションを用意した．各シチュエーションでは，一発話のみ収集するのではなく，probing により，被験者が自然に思いつく発話であれば複数発話回答可能になるようにインタビューの仕方を工夫した．

これらの設問の意図が，実際のインタビューにおいて適切に被験者に伝わるか確認するため，最初にテストとして数名の被験者にインタビューを実施し，被験者が設問に対し誤解するようであれば，表現見直しといった設問の修正を行った．なお，このテストの被験者の発話は以後の検証には用いていない．

実際のインタビュー実施時には，オペレータは，用意されたシチュエーションをもとに被験者にフェイス・トゥ・フェイスによるインタビューを行い，各シチュエーションにおける回答発話の書き起こしを行う．

また，インタビューの仕方によっては，被験者に対し発話を誘導する恐れがある．たとえば，「自宅に帰るためには何と言いますか？」という質問をした場合は，「自宅」という単語に誘導され，被験者は「自宅に帰る」というありきたりな発話へと誘導されてしまう可能性がある．ここで，「お家に帰りたい」，「帰宅」，「もう帰る」など，多様な言い回しを収集するためにはインタビューの仕方にも注意を要する．このため，オペレータにはインタビュアとしての経験を積んだ者が対応するようにした．

本実験の被験者には，日常的に自動車を運転する首都圏在住の 18 歳以上 69 歳以下の民間人 100 名（男女比 1:1）を選定した．被験者には，インタビューに対応するにあたり，一定額の報酬を用意した．

3.4.1. 目的に沿った発話収集のための教示

インタビューにおいて，話題が想定ドメインから外れないようにコントロールすることが重要である．本章のインタビュー形式による収集手法では POI 検索をドメインとしているため，多様性の高い発話を収集することが目的とは言え，「明日の天気は？」という発話は目的範囲外である．したがって，こうした目的外の無効な発話を避け，所定の目的範囲内に被験者の発話を収めなければならない．また，人間が発話をする際，発話対象により発話の内容が変動する [96], [97]．発話対象がカーナビ（機械）であると強く意識した状態では，定型音声コマンドのような発話が収集されることが予想され，発話の多様性が低くなる恐れがある．

以上の問題を回避し，多様性の高い発話を収集しつつもカーナビの POI 検索のための発話という目的範囲内に収められるように，収集作業の最初にオペレータから被験者に以下のように説明を行った．

- **インタビューの最初に行った被験者への説明内容**

「カーナビの音声認識機能は日々進歩してきており，今までの予め決められた言葉ではなく，いろいろな言い方にも対応できるようになりつつあります．そこで本日は，いろいろな状況を示して，○○さんならそれをどのような言葉で表現するか，そういった事例を集めるのが目的です．もちろんそのお答えに『合っている』,『間違っている』はありません．○○さんご自身の自然な発想でお答えください．」

1 つ目のアンダーライン部では，発話対象がカーナビであることを明確に説明することにより，目的範囲外の発話を回避するための説明である．

2 つ目のアンダーライン部では，従来のカーナビの音声認識では，認識される発話に制限があるという認識を持った被験者に対し，そうした制約がないことを説明し自由な発想による発話が可能であることを説明している．

3 つ目のアンダーライン部は，2 つ目のアンダーライン部をさらに強調するための説明である．

3.4.2. コーパスの収集と収集結果

インタビュー時には，最初に被験者から最近の旅行のドライビングスケジュールの事前ヒアリングを行い，被験者に当時のことを思い出させるとともに，その旅行の目的地などを聞き出した．次に，事前ヒアリングした情報をインタビューのシチュエーションに組込み，その被験者が事前ヒアリングで示した目的地を，カーナビを用いて検索するためにどのような発話をするか聞き出し，回答発話を書き起こした．このように，被験者のドライビングスケジュール情報に基づき，目的地を設定したい状況であるというシチュエーションを定義した上で，被験者主導により目的地を設定する発話を聞き出している．

一連のインタビューの例は以下である．ここで，オペレータは「オ」，被験者は「被」とする．

- **ドライビングスケジュールの事前ヒアリング**

オ：最近は車でどこか旅行などされたんですか？

被：山梨の【目的地 1】に行きましたよ．

オ：【目的地 1】いい所ですよね．御家族で行かれたんですか？

被：ええ，親戚が住んでるので，毎年行くんですよ．

オ：【目的地 1】に行くまでにどこかに寄ったりはされたんですか？

被：そういえば，休憩に【目的地 2】にも行きましたね．

- **発話収集のためのインタビュー**

オ：あなたは先ほど設定した【目的地 1】に向かっています．そこで，休憩を取ることにしました．休憩場所は最初におうかがいした【目的地 2】です．そんな時，カーナビになんと言いますか？

被：うーん．そうですね．カーナビに対して言うんですよね？

オ：そうですね．

被：じゃあ，【目的地 2】探して．

オ：実際にカーナビに言うときは「【目的地 2】探して」にしますか？「じゃあ，【目的地 2】探して」までにしますか？

被：**「【目的地 2】探して」**だけですね．

オ：<u>他にどのような言い方がありますか？</u>

被：近くで【目的地 2】

オ：**「近くで【目的地 2】」**だけですね？

被：そうですね．

オ：<u>他にありますか？</u>

被：ちょっと思いつかないですね．

オ：わかりました．ありがとうございました．

オ：では次の質問ですが，今仰られた「【目的地 2】探して」といった発話で，このような絵（カーナビ画面のイメージ図）が出てきました．その場合なんと仰いますか？

⋮

⋮

以上のように，オペレータと被験者が会話をしていく中で被験者のシチュエーションに応じた発話は収集された．太字で表した回答発話がコーパスとして収集された部分である．

- **Probing の実施**

「発話収集のためのインタビュー」上にあるアンダーラインにて示したオペレータの会話は，probing と呼ばれる質問 (probing question) であり，被験者に同じシチュエーションにおける別の言い回しを引き出している．1 シチュエーションから複数の発話を収集し，限られた被験者数からより多くの発話を収集するための工夫である．収集発話量を増やすことにより，「疑似的な環境を用いた発話収集手法」の課題であった収集作業の時間的負荷を低減している．なお，被験者の probing への回答は任意であり，強制的に聞き出すことはしない．強制的に聞き出す場合，被験者は無理に発話を考えることを迫られ，実際には用いないような発話がコーパスに含まれるのを防ぐためである．

以上のようにして収集したコーパスのことを「自由発話コーパス」と呼ぶ．自由発話コーパスの各発話に対し，発話の意図を示す Intent をタグ付けした．また，発話に POI 検索ための検索ワード（施設名，ジャンル，カテゴリといった多種多様なワード，および「夜景の見えるきれいなレストラン」といった複数単語からなる検索ワードも含む）が入っているのであれば，検索ワードとなる部分を Query として人手によりタグ付けを行った．タグ付けは，日本語を母国語とし，POI 検索ドメイン，言語処理にも精通した一人の研究者により人手で行われた．

Intent と Query について例を挙げる．たとえば，発話文が「ここら辺でおいしいレストラン探して」の場合，以下のようになる．

Intent: 周辺検索

Query: おいしいレストラン

すなわち，Intent は発話文の意図でありカーナビの検索機能の種類を示す．Query とは発話文の中でも検索したい施設名やジャンルといったものが入り，

Intent の引数のような情報として後段の POI 検索アプリケーションに問い合わせるためのものである．また，発話文によっては Query が存在しない場合もある．たとえば「自宅へ帰る」という発話文では以下となる．

Intent: 帰宅

Query:（なし）

この場合「自宅」は検索ワードではなく，あらかじめ決められた固定の POI を意味している．そのため，検索ワードはなく，Intent の「帰宅」というタグのみが付与される．

以上の処理を行い最終的に，8452 発話をコーパスとした．提案手法で収集した発話文のタグ付け例を表 3 に示す．

表 3．タグ付与されたコーパス例 [Copyright (C) 2018 IEICE]

Intent	Utterance	Query
周辺検索	ここら辺でコンビニ探して	コンビニ
周辺検索	ガソリンスタンドを自車位置周辺で検索	ガソリンスタンド
周辺検索	最寄りのファミレス	ファミレス
帰宅	自宅に帰る	ー
帰宅	家まで道案内して	ー
帰宅	おうちに帰りたい	ー
単純検索	お腹がすいた	お腹がすいた
単純検索	ラーメンが食べたい	ラーメン
単純検索	コンビニ探して	コンビニ
目的地周辺検索	目的地近くでガスリンスタンド	ガソリンスタンド
目的地周辺検索	到着したらお土産買いたい	お土産買いたい

8452 発話の発話量としての妥当性を考察する．倉田らは，車のオーディオ，エアコン，カーナビ操作に関する発話例を収集している [84]．これによると，ナビ操作ドメインの 125 の機能数に対し，28130 発話を収集し，タスク達成率は 92.2%と述べている．平均発話数は 225 発話である．これに対し，提案手法では

用意したシチュエーション数は約 50 であり，被験者平均では 40.6 である（被験者の回答次第でシチュエーション数は変動）．平均発話数は 208 発話となり，同程度の発話量であるため妥当な発話量を収集できている．

なお，インタビューが終了した後，それぞれの被験者に，インタビューで自身が回答した発話文を再度発話させ，音声を収録した．ただし，本研究では当該音声データは分析対象とせず，発話文のテキストのみを分析対象とするため割愛する．

3.5. 自由発話コーパスの検証準備

収集された自由発話コーパスが，多様性の高い発話であることを検証した．この解析のため，自由発話コーパスと，実際の製品においてユーザが発話したログデータとの比較を行った．自由発話コーパスの比較対象とするデータには，クラリオン株式会社（現在のフォルシアクラリオン・エレクトロニクス株式会社）の車載機器向けクラウドサービス Smart Access [100] にある Intelligent VOICE [101] の市場ログデータを用いた．

Intelligent VOICE は，サードパーティ製のクラウド型音声認識技術を用いた車載機器音声操作機能であり，POI 検索，電話発信，音楽再生などを提供している．大語彙連続音声認識により，たとえば，「夜景の見える綺麗なレストランを探して」といった，語彙や文型の制限を受けない自然な発話から POI 検索を行うことが可能である．また，音声認識結果の確認や，検索結果の絞り込みのための質問などの対話制御を行っている [101]．本研究では Intelligent VOICE に発話された POI 検索のための発話のみを扱った．また，比較検証の対象として，POI 検索に関連した「単純検索」と「周辺検索」という 2 つの Intent に該当する発話を解析対象とした．「周辺検索」は，自車位置周辺で POI 検索する意図に対応した Intent である．「単純検索」は，特に検索条件の制約を設けずに POI 検索する意図に対応した Intent である．この 2 つを解析対象とした理由は，POI 検索が Intelligent VOICE の主機能であること，またその中でも，「単純検索」と「周辺検索」の機能は，一般的に利用頻度が高いためである．Intelligent

VOICE の概略システムを図 3 に示す.

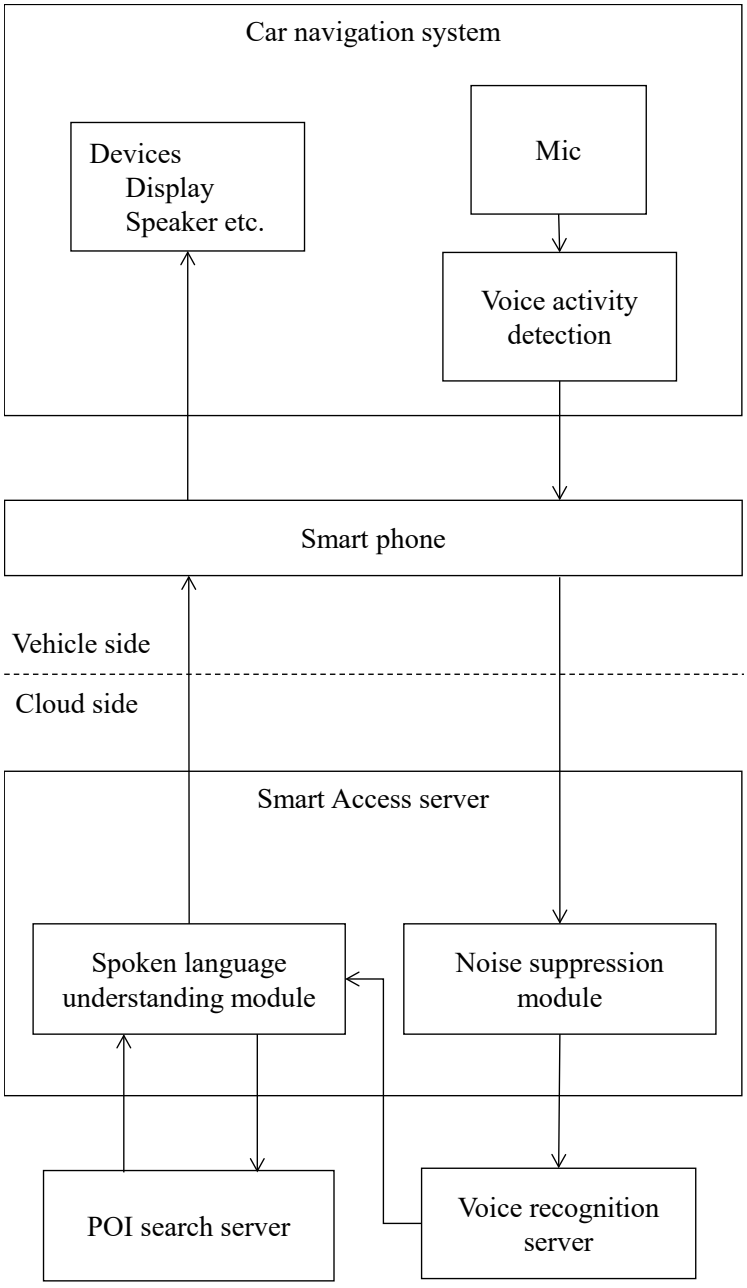


図 3. Intelligent VOICE システム [Copyright (C) 2018 IEICE]

3.5.1. 検証用の自由発話コーパス

タグ付けを行った提案手法による自由発話コーパスの 8452 発話には，POI 検索作業を行う中で，システムからの確認依頼にユーザが回答する Yes や No を

Intent とする発話や，何かの処理を止める Exit を Intent とする発話も含まれている．この中から「単純検索」と「周辺検索」の Intent に絞った 1597 発話を検証対象とした．内訳は「単純検索」が 1287 発話であり，「周辺検索」が 310 発話である．自由発話コーパスより「単純検索」，「周辺検索」 Intent を抽出した 1597 発話のコーパスを「検証用データ 1」と呼ぶ.

3.5.2. 検証用ログデータ

検証用データ 1 に合わせ，Intelligent VOICE の発話ログデータから無作為に「単純検索」を意図する 1287 発話，「周辺検索」を意図する 310 発話，合計 1597 発話を抽出し，比較対象とするデータとした．このデータの書き起こし文，正解 Intent，正解 Query は，日本語を母国語とし，POI 検索ドメイン，言語処理にも精通した一人の研究者が人手により修正・付与した．これを「検証用データ 2」と呼ぶ.

3.5.3. テストデータ

検証用データ 2 とは別に，Intelligent VOICE の発話ログデータから各 Intent にて同数の合計 1597 発話を用意した．これは，後述する言語理解器よる比較実験にて，本章において用意した検証用データ 1，検証用データ 2 を学習用コーパスとして利用した場合，どちらが言語理解器にとって有利であるか検証するための入力用テストデータとして使用する．このデータに対しても各検証用データと同様に，人手により，書き起こし文，正解 Intent，正解 Query を修正・付与した．以降，これを「テストデータ」と呼ぶ.

3.6. 形態素解析と言語理解器による比較検証

提案手法で得たコーパスが，多様性の高い発話であることを検証するため，

検証用データ 1 と検証用データ 2 に対し形態素解析による比較を行った.

さらに，提案手法で得たコーパスの有用性を示すため，検証用データ 1 と，検証用データ 2 を，それぞれ言語理解器に学習させ，テストデータに対する言語理解精度の比較を行った.

3.6.1. 形態素解析による発話の多様性検証

前述したように Takeda ら，河口らの研究 [96], [97] から，発話対象が人間か機械かでは，発話対象が人間である方が一発話あたりの形態素数は増加する. そのため，検証用データ 1 と検証用データ 2 を，それぞれ MeCab [90] を用いて形態素解析を行い，各品詞 [part of speech (POS)] と一発話あたりの形態素数を示したのが表 4 である. この表から，延べ形態素数は検証用データ 1 の方が 639

表 4. 形態素解析結果 [Copyright (C) 2018 IEICE]

POS	Verification data set	
	Set 1 (interview)	Set 2 (logdata)
Noun（名詞）	3649	3854
Particle（助詞）	934	582
Verb（動詞）	428	178
Adjective（形容詞）	169	67
Auxiliary verb（助動詞）	158	77
Prefix（接頭詞）	49	31
Adnominal（連体詞）	44	15
Filler（フィラー）	1	5
Adverb（副詞）	16	5
Conjunction（接続詞）	0	0
Interjection（感動詞）	2	2
Symbol（記号）	5	0
Total（合計：延べ形態素数）	5455	4816
No. of morphemes per utterance（一発話あたりの形態素数）	3.42	3.02

個多いことを確認した．また，一発話あたりの形態素数では，検証用データ 1 の方が 0.40 個多く，比率にして 11.7%多いことを確認した．

図 4 は形態素解析結果をヒストグラムにしたものである．図表より，頻出する POS の上位 3 位は，頻度順に，名詞，助詞，動詞であった．検証用データ 2 に対する検証用データ 1 の POS の比率は，名詞では約 95%とほぼ同等であり，助詞では 160%，動詞では 240%であった．POS 全体としても名詞以外は検証用データ 1 の方が形態素数は多いことが示された．

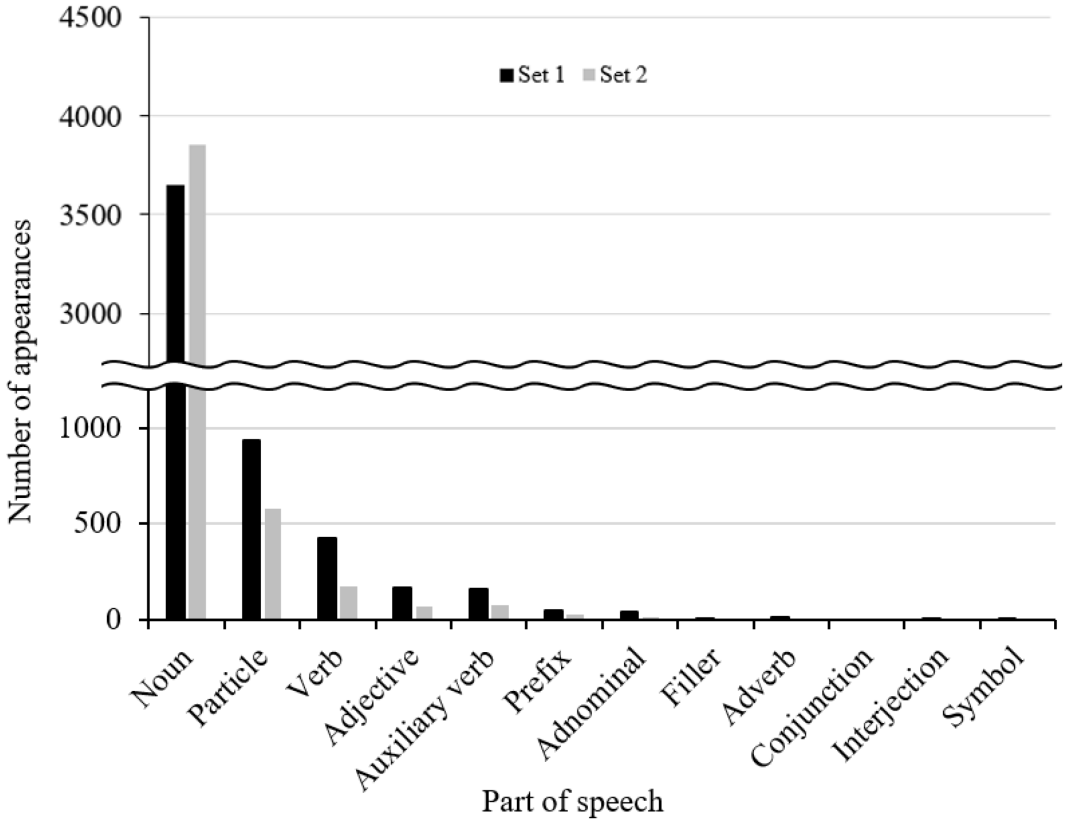


図 4．形態素解析結果のヒストグラム [Copyright (C) 2018 IEICE]

~~具体的に個別の発話を見ると，検証用データ 1 の方が~~検証用データ 2 より，より多種類の POS を含んだ発話が多く見られた．その例を以下に挙げる．

- **検証用データ 1（提案手法）**
  「雪が積もっているスキー場」（単純検索）

「この近くのショッピングモールを探す」（周辺検索）

- **検証用データ 2（発話ログデータ）**

  「『地名』　レストラン」（単純検索）

  「近くのコンビニ」（周辺検索）

以上のように，検証用データ 2 の方がより簡略化された発話になっている．本検証では，単純検索，周辺検索の二種類の Intent の発話を用いているため，基本的に名詞とは POI 名称に関するものとなる．検証用データ 2 は 1 に比べ，延べ形態素数が低いにも関わらず，名詞のみが多い理由は，「『地名』　レストラン」といった名詞を並べた発話が多くなされたためである．一方，検証用データ 1 では助詞，動詞も多く，さらに，一発話あたりの形態素数も多いため名詞を並べた発話ではなく，様々な形態素を用いた多様な発話が含まれている．

次に，異なり形態素に着目し，各数値をまとめたものが表 5 である．各数値について説明する．全 POS を対象とした異なり形態素数は，検証用データ 1 では 817 個，検証用データ 2 では 1079 個であり，検証用データ 2 の方が多かった．しかし，先に述べたように，検証用データ 2 は POI 名称に関する名詞を並べた発話が多い．多種多様な POI 名称があるが，そのバリエーションの多さは発話の多様性とは無関係である．そのため，名詞を除いた異なり形態素数を比較すると，検証用データ 1 は 159 個，検証用データ 2 では 86 個となり，検証用データ 1 の方が異なり形態素数は約 85%多く，検証用データ 2 より多様な形態素を含んでいることが確認された．さらに，一発話あたりの異なり形態素数比 TUR を算出した結果，検証用データ 1 は 0.100，検証用データ 2 は 0.054 であり，検証用データ 1 の方が約 85%多いことを確認した（なお，本実験では発話数を揃

表 5．名詞を除いた異なり形態素数と TUR 比較

	Verification data set	
	Set 1 (interview)	Set 2 (logdata)
No. of different morphemes (except for nouns)	159	86
No. of utterances	1597	1597
TUR	0.100	0.054

えているため，異なり形態素数の比と TUR は同一になる）．

以上より，提案したコーパス収集法は，実際の製品に向かって話されるユーザ発話より，一発話あたりの形態素数，異なり形態素数，TUR が大きく，多様性の高い発話を収集できているとともに，手法自体も効率的に多様な形態素を収集可能な手法であることが示された．

3.6.2. 言語理解器による比較検証

「提案手法の検証用データ 1」と，「ログデータの検証用データ 2」をそれぞれ言語理解器に学習させ，「評価用ログデータのテストデータ」に対する言語理解精度の比較を行う．

言語理解器には，Intent 推定，Query 抽出という二種類の作業を行うものを使用した．Intent 推定の処理では，発話文を形態素に分割し，単語表記の 1-gram, 2-gram, を素性とするベクトルに変換する．そして，機械学習を利用した多クラス分類を用い，このベクトルが各 Intent に属する確率を計算し，確率が最も高かった Intent を採用する．多クラス分類のアルゴリズムとしては，最大エントロピー法を利用した．最大エントロピー法に基づく学習および推定の実装には，Classias を使用した [107].

Query 抽出の処理では，発話文の形態素列に対して，機械学習を利用した系列ラベリングを行う．それぞれの学習文の形態素に対して，Query 先頭 (B), Query 内 (I), Query 外 (O) を示す IOB2 タグを付与し，CRF によってモデルを学習する．Query 抽出時には，発話文の各形態素に対して IOB2 タグを推定し，B ラベルから始まり I ラベルが終わるまでの範囲を Query として採用する．ただし，B ラベルの後続の形態素が B ラベルまたは O ラベルだった場合，また，B ラベルが文末の形態素に付与された場合には，B ラベルが付与された形態素のみを Query として抽出する．

CRF の学習と推定の実装には，CRF++ を使用した [108]．CRF の素性テンプレートは，注目形態素と前後二形態素から構成され，表記または POS の 1-gram, 2- gram, 3-gram, 直前および注目形態素の出力ラベルからなる．

上記の言語理解器に，それぞれの検証用データを学習させ，テストデータの発話文を評価し，Intent 推定正解率，Query 抽出正解率を算出した．学習データを検証用データ 1 にした場合，学習データを検証用データ 2 にした場合，における正解率を比較したものが表 6 である．

表 6．テストデータに対する正解率 [Copyright (C) 2018 IEICE]

Verification data	Percentage of correct answer	
	Query	Intent
Set 1 (interview)	97.5%	99.8%
Set 2 (logdata)	97.6%	98.7%

Intent の推定精度を見るとほぼ同等の値だが，数値としては検証用データ 1 の方が 1.1%高かった．McNemar Test により有意差検定をしたところ，検証用データ 1 の性能が検証用データ 2 より有意に上回っていることが示された ($p < 0.01$)．一方，Query の抽出精度は，有意差はなかったものの ($p = 0.76$)，検証用データ 2 の方が 0.1%上回っていることを確認した．

検証用データ 1 の Query の正解率が検証用データ 2 より高くならなかった要因について考察する．本実験のコーパス収集では，カーナビの POI 検索を目的に被験者の知る POI 名称を用いて，様々な言い回しを回答させた．そのため，本実験の被験者の知る POI 名称と，全国のユーザの発話が含まれるログデータに含まれる POI 名称では，後者の方がバリエーションの幅は広い．このバリエーションの差が，検証用データ 1 を学習データとした時に，Query 抽出精度が高くならなかった要因と考えられた．そこで，この POI 名称のバリエーションの差の影響を低減するため，これまでの実験には用いていないログデータから，1597 発話の 50%の数となる，単純検索 644 発話，周辺検索 155 発話，合計 799 発話を追加で用意し，検証用データ 1 と 2 にそれぞれに追加し，言語理解器に学習させた．追加した学習データによる，テストデータに対する正解率を示したものが表 7 である．

表 7. 追加学習によるテストデータの正解率 [Copyright (C) 2018 IEICE]

Verification data	Percentage of correct answer	
	Query	Intent
Set 1 (interview)	99.2%	100%
Set 2 (logdata)	98.4%	99.2%

この表より，両データとも精度は向上し，僅差ではあるものの検証用データ 1 の方が数値上の精度は上回った．具体的には，検証用データ 1 の Intent 推定精度は 100%となり，Query の抽出精度は 1.7%向上し，同実験の検証用データ 2 の Query 抽出精度より 0.8%上回る結果を得られ，Intent 推定，Query 抽出ともに，検証用データ 1 の方が検証用データ 2 より有意に上回った ($p < 0.01$)．この結果から，本研究で収集したコーパスをログデータと合わせて使用することで，ログデータのみの学習，すなわち，実際の製品に話されている発話文のみを学習するよりも推定精度が向上することを確認した．

注：CRF を用いたラベル付けでは，各形態素のラベルが O, I の順序で出力される可能性があり，その場合には I ラベルの形態素を Query には採用しない制約を設けたが，この実験では発生しなかった．また，一発話から複数の Query が抽出される場合（複数の形態素に B ラベルが付与される場合）が，きわめて少数ながら発生した（表 6, 表 7 の各実験条件にて 1〜2 発話）．その場合，発話文の最初に現れた Query を抽出結果として採用した．

- **提案手法の方がログデータより推定精度が優位になる結果の考察**

検証用データ 2 とテストデータは，ともにログデータであるため，一般的な考えとして検証用データ 2 の方が優位な結果が得られる予想がある．しかし，実際には提案手法の検証用データ 1 の方が高い言語理解精度となった．その要因を考察する．

単純な要因としては，ログデータの検証用データ 2 よりも，提案手法の検証用データ 1 の方が延べ形態素量は多いことが表 4 より分かる．すなわち，形態

素の量視点では，提案手法の方が学習量は多い．その上，異なり形態素数は単純計算では，提案手法は 817 個，ログデータは 1079 個であったが，POI にあたる名詞を除いた場合では，提案手法は 159 個，ログデータは 86 個であった．POI は施設名などであるため，Intent 推定には影響を及ぼさないと考えると，名詞以外で（POI 以外で）異なり形態素を多く持つ提案手法の方が，未知の形態素を減らせるため言語理解器にとって優位であったと言える [109].

次に，各検証用データとテストデータに含まれる，形態素のカバレッジ視点の考察を行う．検証用データ 2 と比べ，検証用データ 1 にて正解に転じたテストデータの発話文の内訳を見ると，正解 Intent が周辺検索の発話文における推定精度の向上が顕著であった．具体的には，検証用データ 2 と比べた検証用データ 1 の正解発話数の増加を見ると，Intent 推定では，正解に転じた発話文 8 個全てが周辺検索であり，Query 抽出では，正解発話数の増加分 13 個のうち 10 個が周辺検索であった．そのため，顕著な差が現れた正解 Intent が周辺検索であった発話文に着目し，分析を行った．

最初に，検証用データ 1 のみで Intent, Query ともに正解となったテストデータの発話文を形態素に分解し，各形態素が学習データに出現していたか否かを調べた．その結果，検証用データ 1 に出現していた形態素の割合は 65 個中 62 個 (95%)，検証用データ 2 に出現していた形態素の割合は 65 個中 57 個 (88%)であった．すなわち，**検証用データ 1 では，検証用データ 2 より 7%高い割合でテストデータの形態素を網羅していることが確認された**.

次に，検証用データ 1 と検証用データ 2 の双方において Intent, Query, ともに正解となったテストデータの発話文の形態素についても，同じ解析を実施した．その結果，検証用データ 1 に出現していた形態素の割合は 1127 個中 1040 個 (92%)，検証用データ 2 に出現していた形態素の割合は 1127 個中 1045 個 (93%)であった．すなわち，**テストデータの形態素の網羅率には 1%の差しかなく，双方の学習データは同程度にテストデータの形態素を網羅している**ことを確認した．

以上の分析結果より，検証用データ 1 のみで正解したテストデータの発話文の形態素が，検証用データ 2（ログ＋ログ）には存在せず，検証用データ 1（コ

ーパス＋ログ）にのみ存在するという差があることが確認できた．この差は，検証用データ 1 の方が発話の多様性が高い点，また，probing により別の言い回しを収集したことにより，異なり形態素を収集しやすかった点が要因と考えられる．

以上より，テストデータの発話文の形態素が学習データ内に多く現れるほど，機械学習を用いた Intent 推定や Query 抽出では，推定の手がかりが多くなるため，正解率が向上する．このことから，検証用データ 1 の正解率が向上したと言える．

3.7. まとめ

本章の研究では，POI 検索をドメインとし，多様性の高い発話を収集することを目的にインタビュー形式によるコーパス収集手法を提案した．今回の検証により，提案手法は，実製品の発話ログデータより，多様性の高い発話を収集可能であるとともに，TUR から提案手法自体が異なり形態素を効率的に収集可能な手法であることが示された．

また，言語理解器の学習データとして，提案手法で得たコーパスとログデータを混合させることにより，実際の製品のユーザ発話に対して高い言語理解精度が得られることが確認された．

以上より，提案手法は多様性の高い発話を収集可能な手法であり，さらに，言語理解器開発のためのコーパスとしても有用であることが示された．

- **今後の課題**

今後の課題として，インタビューによる時間的負荷低減が挙げられる．インタビューの被験者の平均作業時間は，音声録音作業を含めると約半日であった．より短時間で，効率的に多様性の高い発話を収集可能な手法が必要と考える．

第4章　インタビュー形式による収集手法における probing の有用性と時間的負荷低減効果の検証

本章では，第3章のインタビュー形式による収集手法の中で行われた，1シチュエーションから複数の回答発話を収集する probing にフォーカスした検証を行い，本手法の有用性と時間的負荷低減効果について述べる [110], [111].

4.1. はじめに

本章では，第 3 章のインタビュー形式による収集手法において用いられた probing について詳細を述べる．Probing とは，用語の意味としては，より詳細な調査を行うといった意味であるが，本研究にて行った probing とは，簡潔に言えば同じシチュエーションにおいて追加の発話を促すことである．インタビューにおいて発話収集する際，追加の発話を促すことにより，被験者は複数回回答することが可能になる．これにより，1 シチュエーションあたりの発話収集量は増えるため，作業時間をかけずに発話収集量を増やせる手法と言える．しかし，追加発話を促された被験者の発話が有用な発話であったかは第 3 章では検証されていない．本章にてその効果と妥当性について述べる．また，発話収集量が増加したことによる，収集作業負荷の度合いを，被験者数の低減可能数と，時間的負荷の視点から述べる．

4.2. 背景と課題

第 3 章のインタビュー形式による収集手法や，従来手法である WOZ 手法と

いった「疑似的な環境を用いた発話収集手法」では，被験者から直接発話を収集する．そのため，WOZ 手法では実際の製品を模倣した WOZ システムを被験者に使用させることで，実際の製品でも発せられると予想される発話を収集でき，さらに，インタビュー形式による収集手法では，人間のオペレータによるインタビューの会話上で被験者の回答発話を収集することにより，多様性の高い発話を収集できた．

一方，インタビュー形式による収集手法では，インタビューのため被験者にどこかの会議室に来させたり，インタビューを行うオペレータを用意したりする必要がある．WOZ 手法においても，実際の製品開発とは別にコーパス収集のために専用の WOZ システムを準備する必要がある．さらに，収集作業を効率化するために，複数の被験者から並行して収集する場合には，両手法とも並行作業分のオペレータや，WOZ システムが必要であり，1.3 節で述べたように「疑似的な環境を用いた発話収集手法」は，「対象となるドメインの発話ログデータを活用する手法」や「関連性を問わず膨大なデータより抽出する手法」と比較すると手間のかかる収集手法と言える．

第 3 章では，インタビュー形式による発話文と音声の収集作業時間は，被験者一人あたり半日であった．被験者数は 100 名であったため，仮にオペレータが 2 名で 1 日に 4 名の被験者にインタビューを行い，連日インタビューを実施した場合の日数は最短でも 25 日間要することになる．しかし，現実的には被験者の都合（仕事の都合により対応可能なのは休日のみなど）を考えれば，25 日間以上の作業日数が必要になることは想像に難くない．こうした負荷を踏まえるとコーパス収集作業の負荷低減が求められる．

負荷低減の手段として，インタビュー形式による収集手法では，インタビュー中にオペレータから probing を行い，被験者より 1 シチュエーションから複数の発話を収集した．本章のキーアイディアとなる probing は，インタビュー手法において用いられる一つの情報収集の手段である．Probing の目的はインタビューの被験者から，さらに情報を収集することである．Probing は沈黙や，ジェスチャーなどの非言語的信号や，口頭による追加の質問により実施可能である [112]．第 3 章の実験では，probing question と呼ぶ「他に思いつく言い方はあり

ますか？」と質問をすることにより probing を実施している．これにより，発話収集量を増加させることが可能となる．

WOZ システムは実際の製品を模倣した仮のシステムであるため，「他にどのような言い方がありますか？」と，再度被験者に追加質問するのは実際の製品を模倣しているとは言い難い．一方，インタビュー形式による収集手法では，インタビューの会話中に probing が行われるが，会話中に「他にどのような言い方がありますか？」といった別の表現を質問することに不自然さはないため，相性が良い点が利点である．

本章の検証では，probing により発話収集量を増加させる場合，従来の 1 シチュエーションより一発話のみ収集する場合に対する被験者数の低減可能数を算出する．さらに，発話収集量の増加に伴う，発話あたりの時間的負荷の低減効果を定量的に検証する．

また，第 3 章の検証にてコーパス全体の有用性は検証されたが，probing により促された被験者の第二発話，第三発話は，コーパスとして有用であったのかは検証されていない．第一発話はシチュエーションに応じて被験者が最初に想起した発話であるため，実際に発話しうる第一候補の発話として有用な情報である．一方，第二発話，第三発話は，別の言い方をすれば，あるシチュエーションにおいて被験者が発する第二候補，第三候補の発話である．すなわち，極端な表現をすれば，第一発話が優秀な発話であり，コーパス全体の有用性を支えていたが，第二発話，第三発話は，実際には被験者が機器に用いない不要な発話である可能性を含んでいる．本章では，第 3 章のコーパス収集実験にて probing により収集された第二発話，第三発話といった後続の発話が有用な発話であったかを検証する．

4.3. Probing 実施方法

Probing の実施方法は非常にシンプルであり，第 3 章のインタビュー形式によるコーパス収集の実験内にて活用されている．

インタビュー形式による収集手法では，オペレータより被験者にシチュエーションを提示し，そのシチュエーションにて機器を音声操作するために発する発話を被験者より聞き出す．これが第一発話の収集となる．次に，「他にどのような言い方がありますか？」と，同じシチュエーションにおいて他に想起可能な発話を被験者に問いかける．この問いかけが probing であり，「他にどのような言い方がありますか？」という質問が probing question である．被験者が他の発話を思いつけば，それが第二発話として収集される．同様の作業を繰り返し，第三発話，第四発話を収集する．ただし，probing question は被験者にとって強制的な質問ではなく，発話が思いつかないのであれば，そのシチュエーションの回答は終了となり，オペレータはインタビューを進め，次のシチュエーションの発話収集へと移る．インタビューにおいてどのように probing が実施されるかについては，3.4.2 項のインタビュー例にてアンダーラインで示されている．以上の作業をフローチャートにより示したものが図 5 である．図上，太線の部分が probing によるフローを示している．

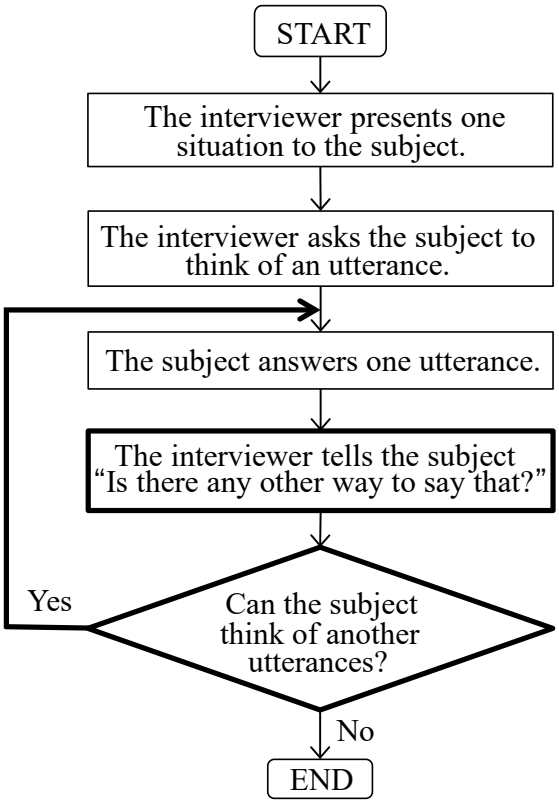


図 5．インタビューにて probing により複数発話収集する手順

4.4. Probing の有用性検証

Probing の有用性を示すため，下記の視点から検証を行う．

1. 形態素解析による必要被験者数の低減効果の検証
2. ログデータによる有用性の検証
3. 時間的負荷の低減効果の検証

それぞれの検証視点について説明する．

- **1. 形態素解析による必要被験者数の低減効果の検証**

多様性のある発話を言語理解器が高精度に意図推定するには，学習に用いるコーパスの発話の多様性が重要である．Probing は限られた被験者数から多くの発話を収集可能な手法であり，言語理解器にとって有利なコーパスとは，コーパスに多くの発話があることが一つの要件ではある．文献 [113] においても大量の発話が言語理解器の性能向上に寄与していることが示されている．しかし，これは一つの要件にすぎない．発話数が多いのみではなく，コーパスの発話の多様性を保つには，そのコーパスには多様な形態素を含んでいる必要がある．そのため，従来手法に則ったコーパスである「第一発話のみのコーパス」と，提案手法の「probing を用いたコーパス」の異なり形態素数を比較し，提案手法のコーパスが，従来手法のコーパスと形態素の視点で同品質となる被験者数を示すことにより，提案手法が負荷を低減できているかを，必要被験者数の削減効果として示す．

- **2. ログデータによる有用性の検証**

1 の検証により異なり形態素数の視点から必要被験者数の低減効果を示すことができたとして，提案手法のコーパスが有用なコーパスであるためには，ユーザが機器に対し実際に発話する形態素を含んでいる必要がある．すなわち，提案手法のコーパスに含まれる形態素に，実際のユーザ発話の形態素が含まれているかを示すことにより，probing により収集されたコーパスの有用性を示す

ことが可能となる．そのため，発話ログデータに現れる形態素に対する，従来手法の「第一発話のみのコーパスの形態素」と，提案手法の「probing を用いたコーパスの形態素」のカバレッジを比較検証する．

- **3. 時間的負荷の低減効果の検証**

提案手法の時間的負荷について検証する．提案手法は，コーパス収集の際，従来手法のように第一発話のみ収集する手法に比べ，多くの発話を集めることを可能にする．したがって，1 の検証にて，同じ発話数を集めるのであれば，提案手法の方が少ない被験者数よりコーパス収集が可能という検証結果が得られることになる．その一方，probing をする分，被験者あたりの質問数は増加しており，インタビューの作業時間も増加するはずである．もし，提案手法により，発話収集作業に大きな時間を要していれば，全体の作業負荷は，従来手法より増加している可能性がある．こうしたデメリットがないか検証するため，従来手法に則った第一発話のみ収集するインタビューのタイムテーブルと，提案手法の probing を実施したインタビューのタイムテーブルを比較検証する．タイムテーブルにより示されるインタビュー作業時間と発話収集量より，提案手法の時間的負荷の低減効果を示す．

4.4.1. 検証に用いるデータセット

Probing の有用性を検証するためのコーパスとして，第 3 章にてインタビュー形式による収集手法により収集されたコーパス [99] をもとに，付与したタグを再考し最適化したものを使用する．このコーパスの概略は以下である．

- 被験者：

  100 人の日常的に自動車を運転する日本人 100 人．男女比 1:1．年齢 18～69 歳．

- ドメイン：

  カーナビを使用した音声操作による POI 検索．

- 使用する発話：

  単純検索を意図する 608 発話．周辺検索を意図する 242 発話の合計 850 発話．

上記のコーパスには，被験者が目的地を意図するワードが含まれている．たとえば，「近くのコンビニ」であれば「コンビニ」の部分がユーザの求める目的地を意味する部分 (Query) である．この他にも，地域名称やジャンルといった情報がこの Query 部分に入る．これらの Query は，その発話の自然さや発話の多様性とは関係ない．たとえば「近くのラーメン」と「近くのみそラーメン」という発話では，前者では「ラーメン」が Query で単語数は 1 個である．後者は「みそラーメン」の部分が Query となり，単語数は「みそ」と「ラーメン」の 2 個である．同様に「近くの道の駅」という発話であれば「道の駅」が Query だが，単語数は「道」，「の」，「駅」の 3 個となる．しかし，いずれの発話も実態は「近くの○○」であり，同じ表現である．このように，Query の部分の違いが発話の多様性の違いになってしまう不都合を回避するため，コーパスの Query の部分は共通の凡庸な名称に置き換え，検証への影響を回避した．なお，コーパスをもとに検証用に用意したデータのことを，「データセット」と呼び分ける．

4.4.2. 形態素解析による必要被験者数減少効果の検証

4.4.1 項にて準備したデータセットより形態素解析による発話の多様性の検証を行うため，二種類の検証用データセットを準備した．

- **データセット 1：第一発話のみのデータ**

  被験者の第一発話のみを集めたデータセットであり，従来手法による収集手法に相当する．したがって，被験者の各シチュエーションに対する発話数も一回となる．

- **データセット 2：Probing を用いた全発話**

  提案手法の probing への回答発話を含めたデータセットであり，第一

発話のみではなく第二発話，第三発話といった後続の発話も含んだ全ての発話を含んでいる．したがって，被験者の各シチュエーションに対する発話数は一回か，それ以上である．

両方のデータセットに対し形態素解析を行い，ある特定の被験者数における異なり形態素数の比較を行った．なお，形態素解析には MeCab [90] を使用した．比較のための数値入手順は以下である．

1. 特定の被験者数を決める．
2. 全ての被験者をランダムにグループ分けする．各グループの人数は 1 にて決定した特定の被験者数である．
3. 被験者の発話を形態素に分割する．
4. 各グループのメトリクスを計算する．
5. 計算された各グループのメトリクスから平均とバラつき（分散）を算出する．
6. 1 にて決定した被験者数よりこの手順を 1000 回繰り返し，平均したこれまでの平均値と分散から最終的な平均値と分散を得る．

この手順を疑似コードにて表したものが Algorithm 1 である．また，この手順により得られた被験者数と異なり形態素数との相関を図により示したものが図 6 である．図において各ポイントは平均値，エラーバーは標準偏差を示す．標準偏差は，被験者数が 50 人以下の場合にのみ計算した．

Algorithm 1. メトリックス計算の疑似コード

```
1: input: n                          ▷ n: the number of subjects to be used.
2: begin
3:     S ← {s_1, s_2, ⋯, s_N}
         ▷ N: the number of all subjects, s_i: the identifier of each subject.
4:     A ← {},  V ← {}
              ▷ Initialize arrays to store averages (A) and variances (V).
5:     for m ← 1; m ≤ M; m ← m + 1 do          ▷ Do M iterations.
6:         T ← {}          ▷ Initialize an array (T) to store sample values.
7:         S′ ← shuffle(S)      ▷ Shuffle the order of subjects randomly.
8:         for k ← 1; k + n − 1 ≤ N; k ← k + n do
9:             U ← {s′_k, s′_{k+1}, ⋯, s′_{k+n−1}}
                                  ▷ Choose one group containing n subjects.
10:            t ← calculate(U)
                        ▷ Calculate a sample value on chosen n subjects.
11:            T ← T ∪ t                    ▷ Save the sample value.
12:        end for
13:        A ← A ∪ average(T)
                              ▷ Save an average on the sample values.
14:        V ← V ∪ variance(T)
                    ▷ Save an unbiased variance on the sample values.
15:    end for
16:    a ← average(A), v ← average(V)
                          ▷ Calculate the overall average and variance.
17:    d ← √v                          ▷ Calculate the standard deviation.
18:    return a, d
19: end
```

この手順は，本章に示されているすべてのメトリクス，すなわち，データセット内の延べ形態素数，データセット内の異なり形態素数，および形態素カバレッジを計算するために用いられる．

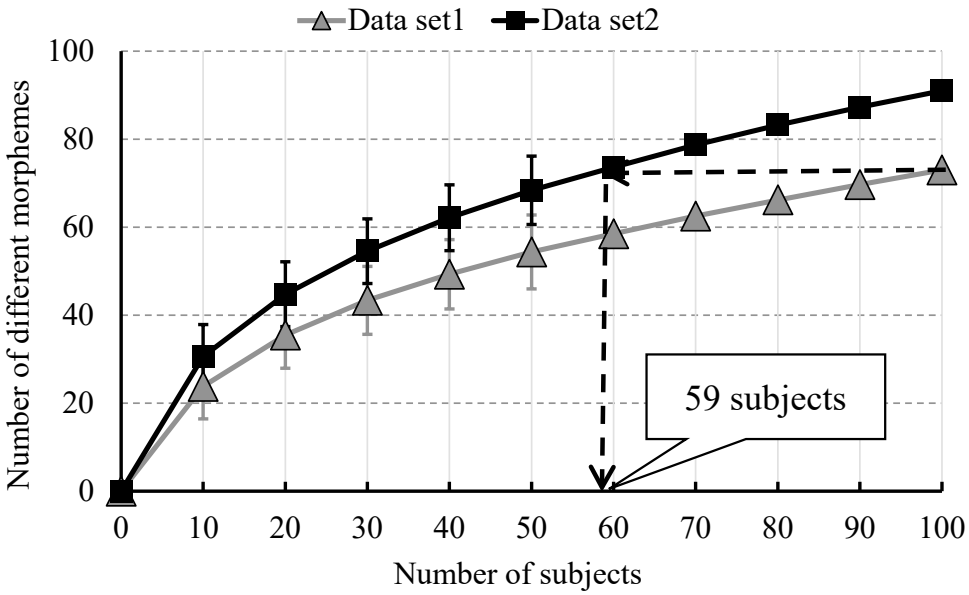


図 6．被験者数と特定の被験者数から得た異なり形態素数との相関性

従来手法であるデータセット 1 では，グループ内の被験者数が 100 人の時，異なり形態素数が 73 個であった．提案手法であるデータセット 2 では，被験者数が 59 人の時，異なり形態素数が 73 個に達することが確認された．すなわち，データセット 2 では 59 人分のデータが，データセット 1 の 100 人分のデータと同等の形態素のバリエーションを得られることを示している．これにより，提案手法は，同等の異なり形態素数を保ちながら被験者数を 41%低減させることが可能と示された．

この検証にて得られたデータセット 2 の 59 人時のデータをデータセット 2-1 とする．表 8 は，データセット 1 と 2-1 の延べ形態素数と異なり形態素数を示したものである．

表 8．データセット 1 と 2-1 の形態素解析結果

	Data set 1 (100 subjects)	Data set 2-1 (59 subjects)
Total No. of morphemes	559	565.5
No. of different morphemes	73	73.1

4.4.3. 発話ログデータによる有用性の検証

4.4.2 項の検証により，データセット 1 とデータセット 2-1 は延べ形態素数，異なり形態素数の視点において同等のデータセットであることが示された．しかし，この検証にて示されたのは，延べ形態素数と異なり形態素数といった各データセットに含まれる「形態素の量」の視点によるものである．それらが実際のユーザ発話に出現する形態素でなければ，最悪な場合，実際の発話には用いられない形態素で構成された「ゴミデータ」となる恐れがある．そのようなデータセットより学習した言語理解器では，適切な意図推定は叶わない．したがって，提案手法より構築したデータセットの形態素が，実際のユーザ発話の中に存在するかを確認することが重要である．

この検証を行うため，クラリオン株式会社（現在のフォルシアクラリオン・エレクトロニクス株式会社）が商用として提供している，クラウド音声認識を活用したクラウドサービスの一つである Intelligent VOICE [100] の POI 検索機能の発話ログデータを使用した．

日本語を母国語とし，POI 検索ドメインに精通した一人の研究者により，正解の書き起こし文，Intent 付与，Query 抽出を行ったログデータから，検証用として周辺検索を意図する 310 発話と，単純検索を意図する 1287 発話の合計 1597 発話を抽出した．この発話ログデータ上にある Query も，データセット 1 と 2（および 2-1）と同様に Query 部分を共通の凡庸な名称に置き換えた．Query を置き換えた後，抽出したログデータ上に出現する形態素が，データセット 1 と 2-1 にも出現しているかカバレッジを比較した．

カバレッジの比較は 2 つの視点から実施した．「延べ形態素」と「異なり形態素」の視点からである．延べ形態素のカバレッジC_mは以下の式にて計算される．

$$C_m = \frac{N_{cl}}{N_l} \tag{4.1}$$

ここで，N_lは発話ログデータの延べ形態素数である．N_{cl}は発話ログデータに含まれる形態素が，データセット 1 と 2-1 において同じ形態素が出現した場合にカウントした一致数である．異なり形態素のカバレッジC_kは以下の式により得

られる．

$$C_k = \frac{K_{cl}}{K_l} \tag{4.2}$$

ここで，K_lは発話ログデータの異なり形態素数である．K_{cl}は発話ログデータに含まれる異なり形態素が，データセット 1 と 2-1 のそれぞれにおいて同じ形態素が出現した場合にカウントした一致数である．以上よりカバレッジを算出したものが表 9 である．

表 9．発話ログデータに含まれる形態素のカバレッジ

	Data set 1 100 subjects	Data set 2-1 59 subjects
Total morpheme (C_m)	94.8%	94.4%
Different morpheme (C_k)	73.3%	70.6%

延べ形態素の視点で，データセット 1 のカバレッジC_mは 94.8%であり，データセット 2-1 では 94.4%であった．その差は僅か 0.4%である．異なり形態素のカバレッジC_kではどちらもおおむね 70%であり，データセット 1 は 73.3%，データセット 2-1 では 70.6%で，その差は 2.7%あった．

これらの結果から，提案手法により得られた 59 人のデータセットのカバレッジは，おおむね従来手法による 100 人のデータセットと同程度であることが示された．

4.4.4. 時間的負荷減少効果の検証

従来手法と提案手法において，発話収集による時間的負荷の比較検証を行う．インタビュー形式による収集手法のプロセスは，シチュエーション検討，被験者のリクルーティング，インタビューのための準備作業，インタビュー作業，各種連絡などである．これらの作業の中で，probing の有無により差が生じるの

は，実際にインタビューを実施する作業である．そのため，インタビュー作業の総合時間にフォーカスした．具体的には，一回のインタビューにおけるタイムテーブルに注目した．一人の被験者に対するインタビューは次のような作業により構成されている．

1. 概略説明：
   オペレータにより被験者にインタビューの概略について説明する．インタビューの目的，シチュエーション例，本日の予定など．

2. シチュエーションの紹介：
   オペレータより，音声操作によりカーナビの特定の機能を操作するシチュエーションについて説明する．その後，被験者は，その特定の機能を実行させるためにカーナビに発する発話を考える．

3. シチュエーションに対する回答：
   被験者がカーナビに発する発話をオペレータに回答する．

4. Probing：
   被験者が回答を一回した後，オペレータが「他にどのような言い方がありますか？」と probing を行い，同じシチュエーションにおける他の言い方がないか被験者に確認する．

5. Probing への回答：
   被験者が他に発話を思いつけばオペレータに回答する．

6. 休憩：休憩時間を取る．

上記の 2〜5，すなわち，シチュエーションの紹介から probing への回答までは各シチュエーションにおいて繰り返し行われる．各項目に要する平均時間は，シチュエーションの説明に 0.5 分，被験者が回答するのに 0.2 分，probing への回答に 0.1 分程度である．

この発話収集作業では，オペレータは平均して 40.6 個のシチュエーションを被験者に提示した．なお，シチュエーション数が整数ではない理由は，オペレ

ータは被験者の回答内容によってシチュエーション数を変動させる場合があるためである．

一回のインタビューに被験者が回答した発話数は平均して 84.5 発話であった．そして，インタビューで行われた probing も 84.5 回となる．これらを踏まえた一回のインタビューのタイムテーブルを表 10 に示した．

表 10．Probing を用いた一回のインタビューのタイムテーブル

Item	Time [min]	Detail
Outlining	15.0	
Introducing situation	20.3	40.6 situations × 0.5 min
Answering	16.9	84.5 utterances × 0.2 min
Probing	8.5	84.5 utterances × 0.1 min
Break	10.0	
Total	70.7	

提案手法はインタビュー作業において一貫して行われた．したがって，この表は提案手法のタイムテーブルを示す．これより，probing を用いなかった場合，すなわち，従来手法のタイムテーブルを推定した．具体的には，probing と，被験者の probing に対する回答に要した時間を除外した．これにより，被験者は 1 シチュエーションに対し一発話のみ回答したことになり，被験者の回答数とシチュエーション数は同じ値（40.6 回）になる．これらの条件を踏まえ，従来手法に相当する probing を用いない場合のインタビュー作業のタイムテーブルを推定したものが表 11 である．

表 11．Probing を用いない一回のインタビューのタイムテーブル

Item	Time [min]	Detail
Outlining	15.0	
Introducing situation	20.3	40.6 situations × 0.5 min
Answering	8.1	40.6 utterances × 0.2 min
Break	10.0	
Total	53.4	

これらの表から，提案手法と従来手法の時間的負荷を定量的に比較する．具体的には，作業時間を発話数で割った「時間的負荷指数」を比較する．時間的負荷指数は小さいほど良く，一発話を収集するために要したインタビュー作業時間が短時間であったことを意味する．すなわち，時間的負荷指数が小さいほど低負荷な発話収集手法と言える．時間的負荷指数I_{wt}は以下の式より算出する．

$$I_{wt} = \frac{T_{wt}}{C_{utt}} \tag{4.3}$$

ここで，被験者一人あたりのインタビューの作業時間をT_{wt}とし，一人あたりの平均発話数をC_{utt}とする．提案手法において，一人の被験者より収集した平均発話数は 84.5 発話であり，一人の被験者のインタビューに要した時間は 70.7 分であった．従来手法では，一人の被験者より収集した平均発話数は 40.6 発話であり，一人の被験者のインタビューに要した時間は 53.4 分であった．

これらの値から，時間的負荷指数を算出すると，従来手法では 1.32 であり，提案手法では 0.836 であった．時間的負荷指数は提案手法の方が従来手法より 37%少ない．以上から，probing **を用いた提案手法のインタビュー形式による収集手法は**，1 **シチュエーションから一発話のみ集める従来の収集作業に比べ 37%負荷が少ない**ことが確認された．

こうした負荷低減は発話収集作業における費用面においても効果がある．従来手法も，提案手法も，発話収集作業（インタビュー）としては，ほぼ同じプロセスである．すなわち，シチュエーション検討，被験者のリクルーティング，インタビューのための準備作業，各種連絡といった作業である．唯一の違いは，提案手法ではインタビュー中に各シチュエーションで probing を行うことである．双方の作業プロセスがほぼ同等のため，費用面においても価格はそのままに収集発話数を増加させることが可能となる．

最後に，提案手法を設けたことによる被験者の心理的負荷について考察する．提案手法はより多くの発話を収集するため，同じ被験者に対し 1 つのシチュエーションから複数の回答発話を得る手法である．被験者からすれば従来手法に比べ probing を受ける分，作業量としては増加している．これはオペレータに

とっても同様である．しかし，これに伴い増加する負荷は軽微なものと考える．なぜならば，人間同士の日常の会話において，「他にどのような言い方がありますか？」と相手に尋ね，その発話の別の言い方や，他の表現方法を訊くことは会話として自然であり，日常会話として慣れているためである．その上，被験者にとっての probing に回答することによる心理的負荷は，新たなシチュエーションを紹介され，その状況を理解し，イメージした上で新たな発話を考える負荷に比べれば少ない．対照的に，probing に回答する場合は，すでにオペレータに紹介されたシチュエーションの理解は終えているため，新たにシチュエーションをイメージするような負荷は発生しない．また，probing への回答はそもそも強制ではない．もし，被験者が何も思いつかないのであれば，「他には思いつきませんね.」などとオペレータに伝えればよく，それにより probing をスキップさせることが可能である．以上から，本提案手法は，心理的な負荷の視点においても増加させずに実施可能な手法である．

4.5. まとめ

本章では，収集負荷を低減する発話収集手法について検証を行った．この提案手法は probing により，被験者より 1 シチュエーションから複数の発話を収集可能にする．本提案手法自体は第 3 章の実験の中で実施されており，収集されたコーパス全体の有用性としてはすでに確認されていた．しかし，probing により収集された第二発話，第三発話といった後続の発話が，被験者より複数発話収集したために発話の多様性の劣化といった悪影響がある可能性を否定できなかった．

そのため本章では，最初に，提案手法で得たデータセットの形態素の多様性について検証した．その結果は，提案手法は従来手法に対し 41%被験者数を削減した状態でも，従来手法と同程度の形態素の多様性を保持していることが示された．

次に，データセットに出現する形態素が，実際のユーザ発話の中でも使用されているかを検証した．具体的には，実際の製品の発話ログデータを用いて，

ログデータに含まれる形態素が，従来手法のデータセット（100 人分）と，提案手法のデータセット（59 人分）に，どの程度含まれているかを検証した．その結果は，被験者数を 59 人に削減した提案手法のデータセットでも，従来手法のデータセットとほぼ同等のカバレッジを保っていることが示された．

従来手法に対し probing が加わったことによる時間的負荷への影響についても検証した．インタビュー作業時間と，収集された発話数から導き出される時間的負荷指数から，probing を用いた提案手法による発話収集作業の方が，従来手法より 37%時間的負荷が少ないことを確認した．

以上より，probing を用いた本提案手法は，被験者数，時間的負荷を抑える効果を持つコーパス収集手法であることが示された．

- **今後の課題**

今後の課題として，probing を用いて収集したデータセットと，従来手法の第一発話のみのデータセットを，それぞれ言語理解器に学習させ，推定精度に差があるかを検証することが挙げられる．表 9 を見る限り，両コーパスはログデータに対しほぼ同等の形態素のカバレッジを持っている．しかし，厳密に言えば，probing を用いた提案手法のコーパスの方が，延べ形態素数のカバレッジでは 0.4%低く，異なり形態素のカバレッジでは 2.7%低い．これは形態素のカバレッジという視点であれば僅差である．しかし，言語理解器の推定精度という，実際の製品のユーザビリティに直結する指標から，必要被験者数の低減効果を改めて導き出すことには，より厳密な検証が行えるため意義があると考える．

第5章　発話対象を人間と想定した Web アンケートによるコーパス収集手法

本章では，より容易に実施可能であり，多様性の高い発話を収集可能な新たなコーパス収集手法として，発話対象を人間と想定させた被験者より Web アンケートを活用したコーパス収集手法について述べる [114], [115].

5.1. はじめに

これまで，インタビュー形式による収集手法や，probing による収集作業の負荷低減の提案，検証を行ってきた．しかし，これらの検証は第 3 章の実験をベースとしており，probing も第 3 章の実験の中で実施されたものであった．したがって，probing による作業負荷低減は示されたが，第 3 章の実験で要した作業時間からの改善には至っていない．本章では，さらに作業負荷を低減しつつ，多様性のある発話収集手法として，発話対象を人間と想定させた被験者より Web アンケートを活用したコーパス収集手法について述べる．

5.2. 背景と課題

第 3 章においてインタビュー形式による収集手法を提案した．この提案手法では機械に向けて発する発話であっても，人間であるオペレータが被験者とのインタビューの中にて，各シチュエーションに応じた発話を被験者から引き出す．これにより，多様性の高い発話を収集可能であること示すとともに，言語理解器による推定精度による検証でも，提案手法のコーパスと発話ログデータを混合することにより推定精度が向上することを第 3 章にて示した．

第 4 章では，第 3 章で実施したインタビュー形式による収集手法の中で実施

した probing について検証した．Probing は，被験者から 1 シチュエーションより複数の発話を収集可能にする．追加で収集された発話が有用であるかを検証し，結果として probing を実施しなかった場合の被験者 100 人のデータセットに対し，probing を実施した場合は被験者数 59 人で同程度のデータセットとなることを示した．さらに，時間的負荷指数による比較検証においても，probing を用いた方が一発話に対する時間的負荷が少ないことを示した．

このように，本研究の目的である多様性の高い発話を整備するための新たな収集手法と，「疑似的な環境を用いた発話収集手法」の課題であった作業負荷を低減する手法を提案，実施，検証してきた．従来手法である WOZ 手法を含め，「疑似的な環境を用いた発話収集手法」では，オペレータが被験者と対になって発話を収集する手法である．そのため，収集効率を上げるため並行作業にて発話収集を行うにはオペレータも並行作業分必要である．さらに，WOZ 手法であればシステムを模倣した WOZ システムといった収集環境を複数台用意する必要があり，インタビュー形式による収集手法においては WOZ システムなどの特別な装置は不要ではあるものの，会議室などインタビューをするための会場を用意する必要がある．そのため，probing による改善は確認できたものの，依然少なくない労力が必要な収集手法である．実際，第 3 章で述べたインタビュー形式による収集手法では，被験者あたりの収集作業は，インタビューと音声録音を含め約半日を要している．より作業負荷を抑えたコーパス収集手法が求められる．

5.3. 提案手法が達成すべき目標

背景を踏まえ，本章の提案手法で達成する目標を以下のように定めた．

1. 被験者より直接発話を収集すること．
2. インタビュー形式の収集手法よりも実施が容易であること．
3. 多様性の高い発話を収集すること．

各目標の理由について述べる．

- **目標 1：被験者より直接発話を収集すること．**

1.4 節で述べたように，本研究の目的である「より自然で自由な発想による多様化した発話例を整備すること」を達成するために「疑似的な環境を用いた発話収集手法」を取る必要がある．この手法は，疑似的な環境を用いて，被験者より直接的に発話収集する手法である．すなわち，既存のデータなどから発話文を生成するのではなく，被験者から直接収集することでより自然で自由な発想による多様化した発話を収集する．

発話収集手法としては，「対象となるドメインの発話ログデータを活用する手法」もあるが，発話ログデータから発話例を収集するというよりは，「発話ログデータより自然で自由な発想による多様化した発話を収集する」とする方が本研究の目的に即している．そのため，収集環境やプロセスに工夫の余地がある「疑似的な環境を用いた発話収集手法」の方が適しており，本研究の方針でもある．

- **目標 2：インタビュー形式の収集手法よりも実施が容易であること．**

1.3 節で述べたように，「疑似的な環境を用いた発話収集手法」の課題として，作業負荷が他の手法に比べ大きいことが挙げられる．Probing による改善を行った第 3 章のインタビュー形式の収集手法においても，被験者一人あたりの発話文のみの収集作業時間（インタビュー時間）は，第 4 章の検証より，およそ 70.7 分と推定され，音声録音作業を含む被験者一人あたりの総作業時間は約半日であった．この場合，100 人の被験者に対し，一日に 4 人の被験者に実験を実施した場合，最短でも 25 日かかる試算である．これを改善できれば，「疑似的な環境を用いた発話収集手法」であっても，作業負荷が改善する．

- **目標 3：多様性の高い発話を収集すること．**

目標 1 と重複する部分があるが，1.4 節で述べたように，本研究の目的は「より自然で自由な発想による多様化した発話例を整備すること」である．「疑似的な環境を用いた発話収集手法」であっても，発話の多様性が低下してはなら

ない．インタビュー形式による収集手法では，「インタビューにて人間のオペレータが被験者の発話を聞き取る」という工夫を行うことにより発話の多様性を高め，事実，発話ログデータより多様性の高い発話を収集した．これに代わる工夫が必要となる．

5.4. 提案手法の概要

2.1 節で述べたように，本研究におけるコーパスとは「タグ付き音声言語テキストコーパス」であり，3.6.2 項にて実施した検証用として言語理解器に学習させるコーパスもテキストコーパスである．本章にて提案する新たなコーパス収集手法ではこの点に注目した．テキストデータのみの収集であれば（音声を収集しないのであれば）インタビュー形式による収集手法のように，被験者にインタビュー会場に来させるのではなく，Internet やパソコンを活用したリモート作業により，発話文を収集する手法を用いることが可能である．いわゆるリモート会議のように，オンライン上でインタビューを行う手法もあるが，それではインタビューの作業時間は実質的には変わらない．そこで，Web 上にアンケートサイトを設け，そこにシチュエーションを示し，そのシチュエーションに応じた発話を被験者自身に回答として記載させる方法を取り入れる．この手法であれば，オペレータは収集作業に立ち会う必要もなく，並行作業する場合も Web アンケートシステムが受け付けられる限り対応可能である．

Web アンケートを用いたコーパス収集は過去の研究でも行われている．藤本らは，Web アンケートを活用したコーパス収集を行い，そのコーパスを学習データとして機械学習させたカーナビエージェント用の言語理解器を開発している [116]．ただし，多様性の高い発話を収集目的とした場合，Web アンケートによる収集にも懸念がある．前述したように，人間の発話は相手が機械であると一発話あたりの形態素数が減少する [96], [97]．被験者はパソコンなどを使用して Web アンケートに回答するため，インタビュー形式による収集手法では人間のオペレータが対応していた発話の聞き手が，Web アンケート・パソコンといった機械に置き換わることになる．そのため，発話の簡略化が行われ発話の多

様性が損なわれる恐れがある．

そこで，本提案では被験者に対し Web アンケート上で発話対象は人間であると教示することで，収集される発話の多様性を高める工夫を行う．

5.5. 提案手法の詳細

前述した 3 つの目標をいかに満たすかという視点で本提案の詳細について述べる．

5.5.1. 被験者より直接発話を収集・実施を容易にする方策

本提案手法は Web アンケートを活用した手法であり，これにより，目標 1 と 2 を満たす．

まず，目標 1「被験者より直接発話を収集すること」をいかに満たすかについて述べる．Web アンケートでは，インタビュー形式による収集手法と同じように，シチュエーションを提示し「この時どのような発話をしますか？」といった質問を行う．つまり，シチュエーションがそのまま設問となる．被験者は，シチュエーションに対する自身の発話を発話文として回答するため，被験者の発話を直接収集している．

次に，目標 2「インタビュー形式の収集手法よりも実施が容易であること」をいかに満たすかについて説明する．Web アンケートを活用することにより，WOZ 手法やインタビュー形式による収集手法とは異なり，WOZ システムの台数や，オペレータの人数を気にせず，並行作業により複数の被験者から同時に発話文を収集することが可能となる．さらに，そもそも WOZ システムといった実験環境や，インタビューを行うための会場，インタビューを行うオペレータも，Web アンケートにて代用するのであれば用意不要であるため実施が容易である．また，インタビュー形式による収集手法の利点であった probing との相性も Web アンケートにおいても同等と考える．シチュエーションに対する他

の言い回しを質問することは，インタビューにおいてもアンケートにおいても特段の差はないと考えるためである．

5.5.2. 多様性の高い発話収集に関する方策

目標 3「多様性の高い発話を収集すること」をいかに満たすかについて述べる．まず，課題として，前述したように，本提案手法は Web アンケートを活用した収集手法であるため，被験者はパソコンなどの機械を用いて回答する．すなわち，被験者にとっての直接的な回答先は，インタビューの時のような人間のオペレータではなく，パソコンなどの機械である．発話対象が機械であると，一発話あたりの形態素数が減少する [96], [97].

そのため，被験者に対し発話対象は人間であることを Web アンケート上で教示する工夫を行う．これにより，発話の回答先（聞き手）はパソコンなどの機械であるが，発話対象（ドメインに対応した発話対象）は人間という構図になり，多様性の高い発話を収集可能と考えた．具体的には，Web アンケート上にイラストにより発話対象が人間であることを被験者にイメージさせる．ただし，「発話対象は人間である」といった直接的で過度な教示は行わず，被験者の自然な認識により，相手が人間であると認識されるようにする．これは，過度に発話対象を人間と意識したことにより，発話の自然さが損なわれることを懸念したためである．

WOZ 手法では，**被験者に発話対象を機械と想定させ**，機械向けの発話を収集する．これに対し，本提案手法は，**被験者に発話対象を人間と想定させ**，実際には機械向けの（発話の多様性の高い）発話を収集する．このように，提案手法は，WOZ 手法とは対照的な特徴を持っていることから，本章の提案手法を「Web アンケートを活用した GiFT (gift for the Tin man) 手法」あるいは単に GiFT 手法と呼ぶこととする．

5.5.3. 提案手法を実施する上での懸念

5.5.1, 5.5.2 項の検討内容をもとに提案手法は実施可能と考えるが懸念もある．Web アンケートの設問に対し，被験者が誤解する，あるいは設問自体に誤解を招くような不備があったとしても，実験者は回答を確認してみないことには期待しない発話が収集されていることを判断できない点である．

インタビュー形式による収集手法であれば，常にオペレータがいるため，被験者が疑問を感じた場合はオペレータに確認を取りながら作業を進めることが可能である．また，シチュエーション設定に不備があった場合においても，その場にて判断することができ，確認や修正を行うことが可能である．したがって，インタビュー形式による収集手法は，オペレータが一つ一つの発話を確認しながら収集する手法とも言える．

Web アンケートでは収集作業を開始すれば，被験者の都合さえ合えば一度に大量の発話を収集可能という利点もあるが，もし，Web アンケートの設計に不備があり，目的の発話を収集できていなかった場合は，それらの回答はゴミデータになる可能性があるのみならず，被験者を改めて募る必要も発生する．

こうした懸念を払拭するため，本格的な収集実験を行う前に，小規模な収集実験を行うことにした．

5.6. 小規模実験による提案手法の実施

5.5.3 項の懸念を踏まえ，小規模な収集実験を実施した．被験者に想定させたドメインは「介護を受ける」であり，本実験においても被験者主導による発話となるよう，Web アンケート上に被験者が要求する介護内容を前提条件として定義し，各介護を要求する被験者の発話文を収集した．

POI 検索ドメインから，「介護を受ける」ドメインへ変更した理由について述べる．本実験ではコンタクトしやすさから，主な被験者は東京工科大学の学生であり，一部はその関係者から募った．この場合，特に学生の場合は運転免許を所持していない，所持していても車を運転しない（カーナビを操作しない）

可能性もありうる．文献 [117] によれば，若者が自家用車に対する失費を，より安価な移動手段に置き換える潜在的ニーズがあるとされている．すなわち，若者の車離れという背景を懸念した．このため POI 検索をドメインとするのは不適当と考えたためである．

「介護を受ける」というドメインであれば，POI 検索と同様にタスク指向型対話であり，学生であっても病気・ケガなどで介護を受ける，または介護をする機会の可能性を持ち，仮にそれらの経験がなかったとしても，使用経験の薄い機械（カーナビ）を操作するシチュエーションをイメージするよりは，介護を受けるシチュエーションの方がイメージしやすい．

さらに，「介護を受ける」ドメインにて高精度に音声操作可能な介護ロボットといった機器の一般消費者向けの普及は，本研究にて調査した限りまだない．こうした状況は，介護ロボットといった機器を多様性の高い発話で操作するという考えが広まっていない段階であり，本研究が想定する「機械が理解しやすい言い方」から，多様性の高い発話へと変化していく状況とも合致する．

Web アンケートに設けるシチュエーション数の決定においては，被験者には無報酬で対応させることを想定したため，回答作業時間を考慮した．長時間に渡りアンケートを回答し続けるのは被験者の負担が大きい点と，アンケートの後半になるに連れ，疲労から発話の自然さが損なわれる点を懸念したためである．そのため，事前に数名の被験者に Web アンケートを受けさせ，10 分程度で回答可能なように，シチュエーション数や，設問の仕方を調整した．以上より，表 12 の 9 シチュエーションを用意した．なお，この事前確認にて収集した回答は以後の検証には用いていない．

各シチュエーションにおいて，最低一発話は回答しなければならないようにアンケートを設計した．また，各シチュエーションには probing question 用の回答枠を設け，被験者は，1 シチュエーションに対し複数回答可能となるようにアンケートを設計した．ただし，probing question への回答は任意であり無回答でもアンケートを提出可能にした．

表 12．Web アンケート上に用意した介護のシチュエーション

No.	Situation	Sample utterance
1	寝ている状態から起こしてもらう場合	体を起してください
2	座っている状態から立ち上がらせてもらう場合	立ち上がらせてください
3	ベッドから車いすへ移動させてもらう場合	車いすへ移動したい
4	車いすを押してもらう場合	車いすを押して
5	車いすで移動中，気になるものがあり車いすを止めてもらう場合	ちょっと止めて
6	車いすからベッドへ移してもらう場合	ベッドへ移動したい
7	ご飯を食べさせてもらう場合	ご飯が食べたい
8	味噌汁を飲ませてもらう場合	味噌汁が飲みたい
9	食事をさげてもらう場合	食事を下げて

5.7. 小規模実験にて収集したコーパスの詳細

東京工科大学の学生とその関係者 30 人をボランティアの被験者とし Web アンケートによる発話収集を実施した．収集された発話は 461 発話であり，この内，17 発話は被験者の誤解，実験者が発話文を見ても意図を理解できない，などの理由により無効とした発話である．これらの判定は，日本語を母国語とする研究者（実験者）一人が行った．したがって，検証に使用できる有効な発話数は 444 発話である．この実験結果をまとめたものが表 13 である．

表 13．発話対象を人間と教示した被験者 30 人からの発話収集結果

	Results
Total No. of utterances	471
Invalid No. of utterances	17
Valid No. of utterances	444
Availability rate	94.3%

収集された発話のいくつかには「○○に行きたい.」といった，特定の場所は示さず，ワイルドカードとして「○○」といった発話文を回答する場合があった．こうした場合，後の検証に影響が出ないように「そこ」といった凡庸な単語に置き換えた．こうした調整を行った結果，収集された発話の有効率は約90%を超えており，約 95%に迫っている．このことから，Web アンケート，および，Web アンケート上で実施した probing を用いた本手法は発話文を収集する手法として有効な手法であることが示された.

5.8. インタビュー形式のコーパスとの比較検証

収集された発話文の多様性を検証するため，第 3 章にて収集したコーパス（検証用データ 1 の 1597 発話）と，提案手法により収集されたコーパスにて，形態素解析による比較を行う.

第 3 章のコーパスを比較対象としたのは，インタビュー形式による収集手法は，人間であるオペレータが会話の中で被験者の発話を収集することにより，より多様性の高い発話を収集することを目的としており，発話の多様性が高まるように工夫し収集する視点において本章の実験と目的が同じためである.

インタビュー形式によるコーパスと，GiFT 手法によるコーパスを MeCab [90] を用いて形態素解析した結果が表 14 である．また，両コーパスは発話数に違いがあるため，GiFT 手法のコーパスの発話数を 1597 発話分（1597 / 444 倍）にした値も併記した.

さらに，表 14 よりインタビュー形式によるコーパスと，1597 / 444 倍した GiFT 手法によるコーパスの結果をヒストグラムにて示したものが図 7 である.

表 14. インタビュー形式のコーパスと GiFT 手法により被験者 30 名から収集したコーパスの形態素解析結果

	Interview style-based	GiFT (1597/444 times)	GiFT (original)
Noun（名詞）	3649	1637	455
Particle（助詞）	934	3201	890
Verb（動詞）	428	2741	762
Adjective（形容詞）	169	306	85
Auxiliary verb（助動詞）	158	1489	414
Prefix（接頭詞）	49	22	6
Adnominal（連体詞）	44	14	4
Filler（フィラー）	1	0	0
Adverb（副詞）	16	144	40
Conjunction（接続詞）	0	18	5
Interjection（感動詞）	2	223	62
Symbol（記号）	5	611	170
Total（合計：延べ形態素数）	5455	10406	2893

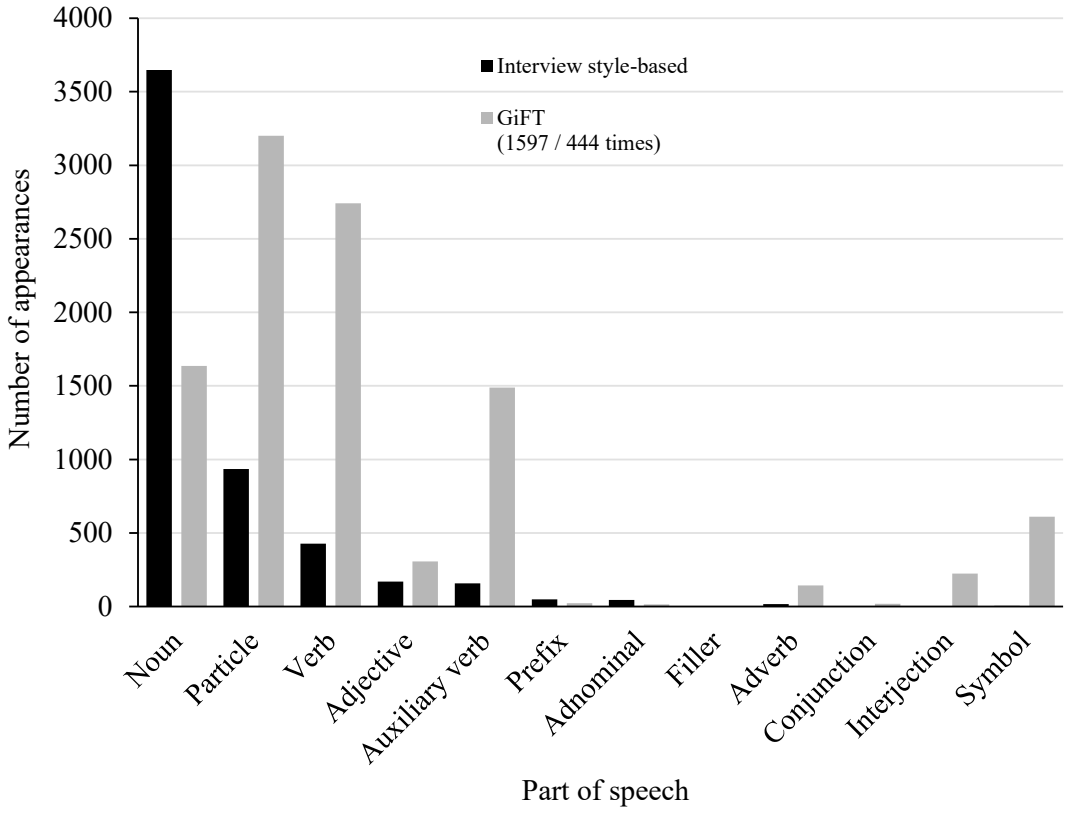


図 7. インタビュー形式によるコーパスと GiFT 手法による被験者 30 名のコーパスにて発話数を同数とした場合の POS 比較ヒストグラム

これらの図表からわかるように，発話数を同数とした場合の GiFT 手法によるコーパスはインタビュー形式によるコーパスより，約 91%延べ形態素数を多く持っていることを確認した．また，インタビュー形式によるコーパスは顕著に大量の名詞を含んでおり，他の POS では GiFT 手法のコーパスより少ない傾向であった．その理由は，インタビュー形式による収集手法により収集されたコーパスはカーナビの POI 検索がドメインであるためである．POI は施設名や地域名といった名詞で構成される．そのため発話内に多くの名詞が出現している [99].

次に，一発話あたりの形態素数を比較したものが表 15 である．表 14 で，発話数を揃えた比較を行っているため，延べ形態素数の比と同じ値になるが，改めて算出した.

表 15．インタビュー形式によるコーパスと GiFT 手法による被験者 30 名より収集したコーパスの形態素数比較

	Interview style-based	GiFT
Total No. of morphemes	5455	2893
No. of utterances	1597	444
No. of morphemes per utterance	3.42	6.52

この表が示すように，インタビュー形式によるコーパスの一発話あたりの形態素数は 3.42 個であり，GiFT 手法によるコーパスの一発話あたりの形態素数は 6.52 個であった．したがって，GiFT 手法により一発話あたりの形態素数は約 91%増加した．

5.9. 小規模実験結果を踏まえ被験者数を増やした検証実験

5.8 節での検証により，第 3 章のインタビュー形式によるコーパスと，提案手法の GiFT 手法によるコーパスでは，提案手法の方が多様な POS を収集できており，一発話あたりの形態素数も多く，発話の多様性が高いと言える．

しかし，同様に述べたように，インタビュー形式によるコーパスは POI 検索をドメインとしているため，施設名などの名詞が多く発話に含まれている特徴がある．そのため，ドメイン違いにより POS の特性に違いがある可能性がある．

ドメイン違いによる実験結果への影響がある場合，GiFT 手法の特徴である，「被験者に発話対象を人間であると想定させること」が発話の多様性を高めたのか，ドメイン違いによる差なのか，その要因を断定できない．

そのため，小規模実験の結果を踏まえ，GiFT 手法の特徴にフォーカスした検証を行うため，さらなる実験を行った．具体的には，9 シチュエーションや probing といった Web アンケートの設計はそのままに，被験者を 30 人から 200 人に増加し，これを下記のように 2 グループに分けた．

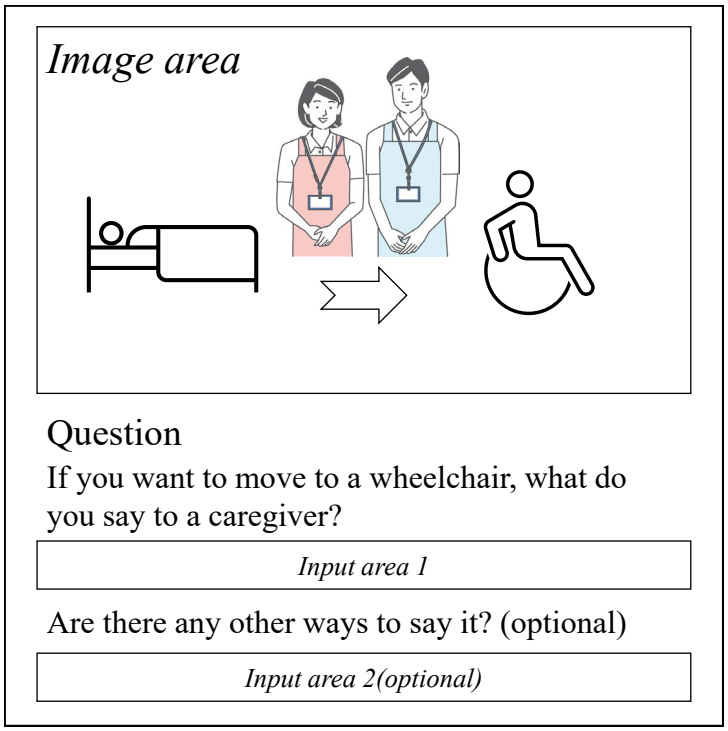
- **グループ A（発話対象は介護士：GiFT 手法）**
  発話対象は「人間の介護士」であると教示したグループ
- **グループ B（発話対象は介護ロボット）**
  発話対象は「介護ロボット」であると教示したグループ

グループ A は，発話対象を人間の介護士と被験者に教示したグループであり，提案手法の GiFT 手法によるグループである．また，5.7 節で収集したコーパスと同条件のまま被験者数を 30 人から 100 人にしたグループとも言える．このグループにおいても，発話対象が人間の介護士であると教示するにあたっては，単純な「介護士にどのように言いますか？」といった質問表現や，介護イメージとして介護士が介護しているイラストを添える程度にし，被験者の自然な認識で発話対象が人間だと認識させるようにした．また一方では，被験者が過度に発話対象を人間だと意識しないように「相手は人間なので自然な会話で回答してください.」といった教示は避けた．

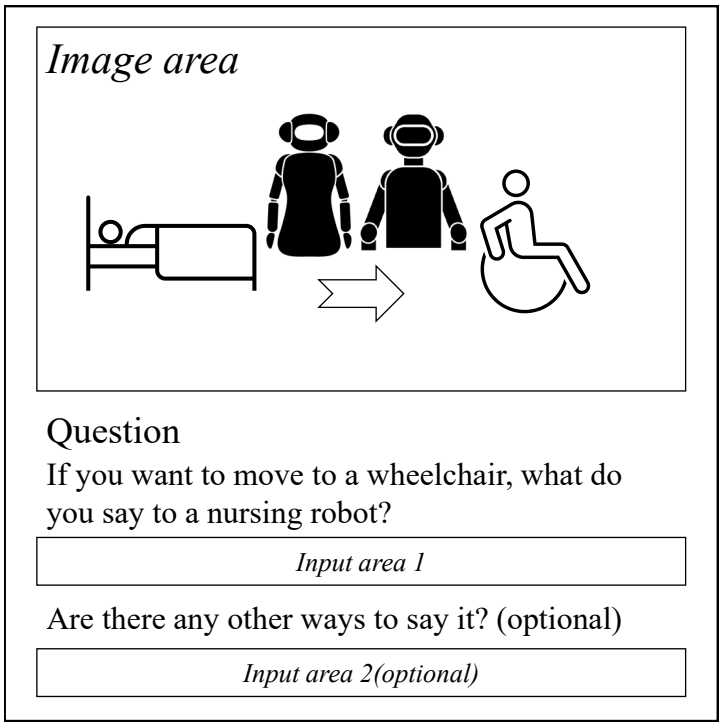
グループ B は，発話対象を介護ロボットと被験者に教示したグループである．グループ B においても発話対象が介護ロボットだと教示するにあたっては，単純な「介護ロボットにどのように言いますか？」といった質問表現や，介護ロボットが介護するイラストにより，被験者が自然に発話対象は介護ロボット（機械）であると認識するようにした．

図 8 はグループ A，B の被験者が受けたアンケートのイメージ図である．(a) は発話対象を人間の介護士と想定した場合のアンケートイメージである．(b) は発話対象が介護ロボットと想定した場合のアンケートイメージである．

この実験の被験者は，ボランティアで協力した日本人の 200 名である．このうち 181 名は東京工科大学の 18-23 歳程度の学生であり，男女比は約 7:3 である．残りの 19 名は学外者である．個人情報保護の視点から，学外者からは個人情報に相当する情報は収集しなかった．以上の被験者を無作為に 100 名ずつグループ A と B に分けた．



(a)



(b)

図 8．Web アンケートのイメージ図

5.9.1. 新たに収集されたコーパスの詳細

200 人の被験者を対象とした実験により，グループ A から 1380 発話を得られた．このうち 31 発話は被験者によるシチュエーションの誤解などにより検証に利用できない無効発話であった．結果として有効な発話は 1349 発話であった．同様にグループ B からは 1546 発話を収集でき，無効発話は 25 発話であった．有効な発話は 1521 発話である．これらの判定も，日本語を母国語とする日本人の一人の研究者（実験者）が行った．以上をまとめたものが表 16 である．また，本章の**付録**に各グループの発話例の一部を示す.

表 16. グループ A と B の発話収集結果

	Group A (person)	Group B (robot)
Total No. of utterances	1380	1546
No. of invalid of utterances	31	25
No. of valid of utterances	1349	1521
Availability rate	97.75%	98.38%

この結果から，グループ A と B ともに発話の有効率が 95%を超えていることが示された．これにより，本実験においても提案手法が発話収集手法として有効であり，目標 1 の「被験者より直接発話を収集すること」が満たされた.

また，この実験にて収集された発話のいくつかにおいても「○○に行きたい」の「○○」のように単語を特定せずワイルドカードとする回答が見られた．こうした発話は後述の検証のため「そこ」などの凡庸な単語に置き換えた.

5.9.2. インタビュー形式との時間的負荷比較と形態素解析による検証

ここでは，提案手法が目標 2 の「インタビュー形式の収集手法よりも実施が容易であること」と，目標 3 の「多様性の高い発話を収集すること」を満たしているかを検証する.

目標 2 は，従来手法であるインタビュー形式による収集手法の時間的負荷と，提案手法の時間的負荷を比較することにより，提案手法の方が負荷は低いことを示す．目標 3 はグループ A と B の形態素解析による比較検証にて提案手法の有用性を示す．

5.9.2.1. 時間的負荷指数による収集負荷の比較

第 4 章にて用いた一発話あたりの時間的負荷を算出する時間的負荷指数を本章でも用いる．第 4 章にて，インタビュー形式による収集手法では，被験者一人あたりに要した時間は表 10 にて推定されており，70.7 分であった．一人あたりの作業時間を平均発話数で割った時間的負荷指数は，0.836 であった．ただし，これは被験者のみに着目した時間的負荷指数である．インタビュー形式による収集手法ではオペレータが常時被験者と共同で作業している．一方，Web アンケートでは被験者が一人で作業するという作業者数の違いがある．そこで，式(4.3)で定めた時間的負荷指数I_{wt}を改め，I'_{wt}として以下のように定めた．

$$I'_{wt} = \frac{T_{wt} C_p}{C_{utt}} \tag{5.1}$$

ここで，C_pは作業者数を示し，インタビュー形式による収集手法では 2 名となり，Web アンケート形式では 1 名となる．式(5.1)をもとに，改めてインタビュー形式による収集手法の時間的負荷指数I'_{wt}を算出すると，作業者数は 2 名となるため，時間的負荷指数I'_{wt}は倍の 1.67 となる．

一方，グループ A と B では被験者一人あたりの Web アンケートの回答時間は約 10 分である．グループ A では 1349 発話を 100 人から収集したため，被験者一人あたりの平均発話数は 13.5 発話となる．グループ B では 1521 発話を 100 人から収集したため，被験者一人あたりの平均発話数は 15.2 発話となる．時間的負荷指数I'_{wt}を算出するとグループ A では 0.740 であり，グループ B では 0.658 であった．以上の結果を表にしたものが表 17 である．

表 17. インタビュー形式の収集手法とグループ A, B の時間的負荷指数比較

	Interview-based method (For an operator and a subject)	Group A (person)	Group B (robot)
I'_{wt}	1.67	0.740	0.658

この表より，インタビュー形式による収集手法の時間的負荷指数I'_{wt}に対し，グループ A は約 56%負荷少なく，グループ B は約 61%負荷が少ないことを確認した．したがって，Web アンケートを用いたグループ A, B は時間的負荷指数の比較により，インタビュー形式による収集手法より低負荷であり，目標 2 の「インタビュー形式の収集手法よりも実施が容易であること」を満たすことが示された．

- **顕著な時間的負荷低減となった考察**

インタビュー形式による収集手法の時間的負荷に対し Web アンケートを用いた手法では，グループ A, B ともに約半減と顕著な時間的負荷指数の低下が得られた要因を考察する．一つの理由は，インタビュー形式による収集手法では，オペレータは被験者と会話し，これから行われるインタビューの説明を行う必要がある．対照的に，提案手法では，被験者が Web アンケートを通して一方的に説明されるのみである．すなわち，Web アンケートでは被験者は一人でアンケート回答作業を行い，なおかつ他者との会話は発生しない点が挙げられる．二つ目に，比較する手法間に作業者数の違いがあった点が挙げられる．インタビューの時間的負荷は，被験者とオペレータの分を合わせて考慮する必要があるが，Web アンケートでは被験者が一人で作業するため大きな差となる．そのため，時間的負荷は半減以下という顕著な時間的負荷指数の差が発生したと考える．

さらに，インタビュー形式による収集手法では，並行して作業可能な数はオペレータの人数に制約され，100 人の被験者を同時にインタビューするにはオペレータも 100 人必要である．これに対し，Web アンケートでは被験者は一斉に回答が可能であり，理論上では被験者の都合さえつけば，Web アンケート開始から 10 分程度で収集作業は完了することが可能である．そのため，理論上は，

提案手法はより時間的負荷を低減可能な収集手法である．

- **グループ A と B 間の時間的負荷の差に関する考察**

これまでの検証にて，提案手法のグループ A はインタビュー形式による収集手法より，時間的負荷を約半減する効果があることが確認できた．これにより，目標 2「インタビュー形式の収集手法よりも実施が容易であること」は達成できたが，グループ A と B 間の時間的負荷指数I'_{wt}を比較すると，介護ロボットを発話対象としたグループ B の方がわずかに負荷は少なかった．この点について考察する．その理由は，グループ B は介護ロボット（機械）に向けた発話であったため，発話の単純化が発生したことにより「単純な発話＝アンケートでの入力量が少ない」という回答発話文の入力負荷の少なさがある．この差が，グループ B の時間的負荷指数の方が少なかった要因と考える．

インタビューのように，口頭にて発話を伝えるには大差ない負荷と考えるが，それをアンケート上でタイプ入力する場合では，形態素数（単語数）の多いグループ A の方が被験者にとって負荷が大きい．これは収集された発話数にも表れており，表 16 においてグループ B の方が収集された発話数が多いが，シチュエーションに対し第一発話は必須回答として Web アンケートを設計しているため，この発話数の差は任意回答としている probing に対する回答数の差である．Probing への回答数は，グループ A は 449 発話（全発話の 33%），グループ B は 621 発話（全発話の 40%）である．発話数の差は 172 発話であり，パーセンテージとしては約 28%の差である．発話対象を介護ロボットとするグループ B の方が，発話文の入力負荷が少ない分回答しやすく，回答数の増加に繋がったと考える．

5.9.2.2. グループ A とグループ B の形態素解析による比較

グループ A と B のコーパスに対し形態素解析を行った．形態素解析には MeCab を用いた [90]．表 18 に形態素解析結果を示す．これをヒストグラムにしたものが図 9 である．

表 18. グループ A と B の形態素解析結果

Part of speech	Group A (person)	Group B (robot)
Particle（助詞）	2774	2090
Verb（動詞）	2384	1953
Auxiliary verb（助動詞）	1395	400
Noun（名詞）	1285	1290
Symbol（シンボル）	596	176
Adjective（形容詞）	279	84
Interjection（感動詞）	165	29
Adverb（副詞）	129	86
Prefix（接頭辞）	23	20
Conjunction（接続詞）	7	1
Adnominal（連体詞）	4	2
Filler（フィラー）	0	1
Total（合計：延べ形態素数）	9041	6132

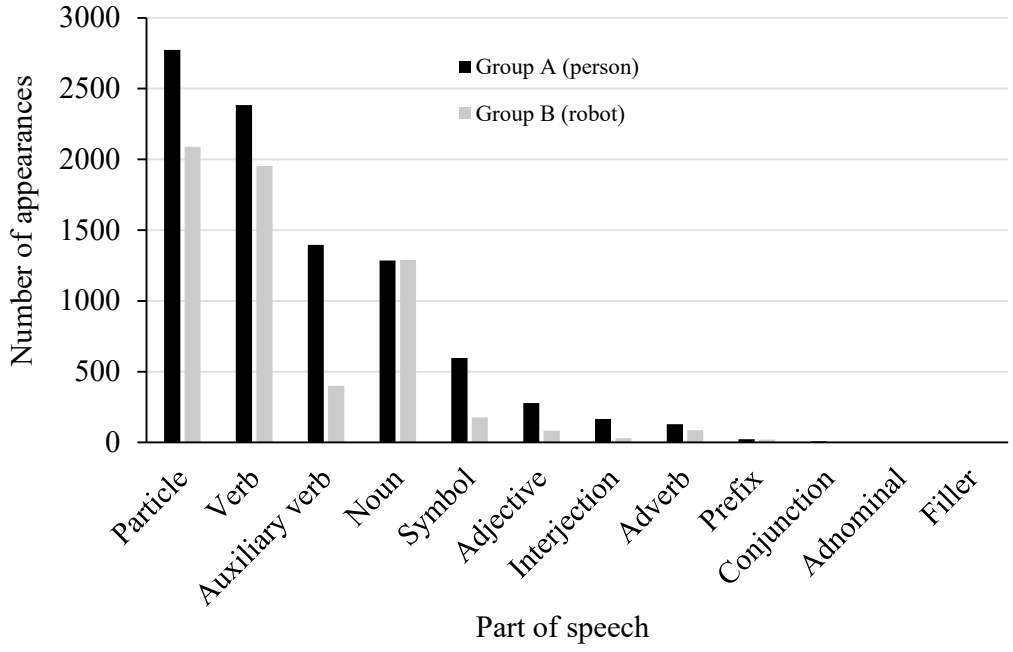


図 9. グループ A と B の POS を比較した形態素解析結果のヒストグラム

図表から，各 POS において，提案手法のグループ A の方が多くの POS を含んでいることが確認された．これは多様な POS を用いた，より人間同士の会話に近い表現がなされていると推測される．また，助動詞に関しては顕著にグループ A の方が多いことが分かる（約 3.5 倍）．グループ A, B の助動詞において支配的な上位 3 つを上げる．グループ A では 18 種の助動詞があり，上位三種は「です」が 427 個，「たい」が 349 個，「ます」が 186 個であった．これら三種で助動詞の 69.0%の占めている．一方，グループ B は 10 種の助動詞があり，「たい」が 253 個，「です」が 66 個，「た」が 24 個であり，これらが助動詞の 85.8%を占めている．両グループで共通する「たい」，「です」に注目すると，「たい」は話し手の希望を意味するものであり，「です」は断定の丁寧な言い方である．これらの数が，グループ A では「です」＞「たい」であるのに対し，グループ B では「です」＜「たい」となっている．このことから，グループ A では発話対象が人間の介護士であるため，タスク達成のための希望を伝えつつも丁寧な表現を用いた「〇〇たいです」といった表現により，助動詞が特に多くなったものと予想される．実際に各グループにて「〇〇たいです」と表現している発話文を確認したところ，グループ A で 138 発話，グループ B で 41 発話であった．

続けて，グループ A と B の各統計結果と，式(2.3)より算出した一発話あたりの異なり形態素数比 TUR をまとめたものが表 19 である．

表 19．グループ A と B の各統計結果と TUR 比較

	Group A (person)	Group B (robot)
Total No. of morphemes	9041	6132
n_i	1349	1521
$\bar{x}_i$ (s_i)	6.70 (±3.46)	4.03 (±1.70)
CV_i	51.6%	42.2%
No. of different morphemes	346	333
TUR	0.256	0.219

ここで，$i = A, B,$ とし，グループ A の発話数n_A，グループ B の発話数n_Bとするとき，各発話の形態素数を$x_i(j)$とする（グループ A では$j = 1, 2, ..., n_A$，グループ B では$j = 1, 2, ..., n_B$）．このとき，x_i（そのグループの延べ形態素数）の一発話あたりの形態素数を$\bar{x}_i$とする．また，$\bar{x}_i$に対する標準偏差をs_iとする．このときに求められる変動係数CV_iは以下の式である．

$$CV_i = \frac{s_i}{\bar{x}_i} \times 100\%, i = A, B, \tag{5.2}$$

これらのデータを考察する．グループ A は B に比べ少ない発話数であった$(n_A > n_B)$．しかし，延べ形態素数において，グループ A は B よりも 47%多い．これは，グループ A の方がコーパスに含まれるデータ量が多く，一発話あたりの形態素数も多いことを示している．

一発話あたりの形態素数$\bar{x}_i$は，延べ形態素数を発話数n_iで割ることで得られる．一発話あたりの形態素数$\bar{x}_i$および標準偏差s_iを算出すると，$\bar{x}_A$は 6.70 (±3.46)であり，$\bar{x}_B$は 4.03 (±1.70)であり，やはりグループ A の方が多い．より確かな検証のため，統計学的にこれらの平均値に差があるのかを検証した．

標本数（発話数）が 30 以上であれば中心極限定理より，標本平均の分布は正規分布に近似すると言われている [118]．また標本の分散も母分散に近似すると言われている．本実験では標本数はグループ A, B ともに十分大きい（ともに 1000 以上）ため，正規分布に近似していると言える．母分散が分かり，正規分布しているのであれば標準化することにより検定統計量zより，z検定が可能である．

帰無仮説を二標本の平均に差はないとし，対立仮説を二標本の平均に差があるとする．そのときの統計量zは以下の式で求められる．

$$z = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{6.70 - 4.03}{\sqrt{\frac{3.46^2}{1349} + \frac{1.70^2}{1521}}} = 25.7 \tag{5.3}$$

標準正規分布表より片側で $p < 0.01$ となり，平均値に差はないとする帰無仮

説は棄却され，対立仮説である平均値に差があるということを採択する．以上から，グループ A はグループ B より 66%多く一発話あたりの形態素を含んでいる ($\bar{x}_A > \bar{x}_B$).

次に，異なり形態素数と，一発話あたりの異なり形態素数比 TUR を確認する．異なり形態素数が多ければ，被験者は多様な形態素を用いて発話していることを意味している．反対に，異なり形態素数が少ない値であれば，各被験者の発話に出現する形態素が同様であることを意味する．すなわち，各被験者は似たような発話をしていることを意味する．言語理解器にとって発話の意図推定には異なり形態素数が多いことがアドバンテージとなる．なぜなら，多くの単語を学習することにより，未知の単語を減らせるためである [109].

表 19 より，グループ A の発話数はグループ B の発話数の 88.6%であったのに対し，異なり形態素数はグループ A の方が 4%多いことが確認された．これにより，グループ A の被験者はグループ B よりも多様な形態素を用いて発話していることが示された．また，TUR に着目すると，グループ A の方が約 17%大きいことから，グループ A で用いた提案手法の方が，異なり形態素を効率的に収集可能な手法であることが示された.

次に，標準偏差s_iに着目してみると，グループ A の方が値は大きい．これは一発話あたりの形態素数の変動幅が大きいことを意味している．すなわち，一発話が短文なものから長文のものまで様々なバリエーションの発話を含んでいるということを示している．しかし，グループ A とグループ B 間の一発話あたりの形態素数の差が大きい．そのため，一発話あたりの形態素数の変動係数CV_iを比較する．変動係数CV_iの比較により，CV_AはCV_Bより 9.4%大きいことを確認した ($CV_A > CV_B$)．これにより，グループ A に含まれる形態素数の変動範囲はグループ B の変動範囲よりも大きく，グループ A の方が多様な長さの発話文を収集できていることが示された.

これまでの内容をまとめると以下のようになる.

- グループ A の一発話あたりの形態素数が，グループ B よりも約 66%多い ($\bar{x}_A > \bar{x}_B$)．これは，提案手法であるグループ A の方が，グループ B

よりも，より人間との会話に近い自然で自由な発想による発話を収集できていることを示している．

- グループ A の方がグループ B よりも異なり形態素数が約 4%多い．これは，提案手法のグループ A はグループ B よりも多様な形態素を用いた発話を含んでいることを示している．

- グループ A の方がグループ B より TUR が約 17%大きい．これは，提案手法の方が異なり形態素を効率的に収集可能な手法であることを示している．

- グループ A の変動係数はグループ B よりも約 9.4%大きい ($CV_A > CV_B$)．これは提案手法であるグループ A の方が，グループ B よりも発話文の文長に多様性があることを示している．

以上から，提案手法によるグループ A のコーパスは，グループ B のコーパスに比べ，一発話あたりの形態素数，異なり形態素数ともに多く，多様性の高い発話を収集できていることが示された．さらに，発話文の文長においても，グループ B より短い発話から長い発話まで多様性があることも示された．また，TUR はグループ A の方が 17%大きかったことにより，発話対象を人間と想定する GiFT 手法は，異なり形態素を効率的に収集可能な手法であると言え，これはグループ A の方が発話数は少ないにも関わらず異なり形態素数が多い点にも表れており，本手法自体が多様な形態素を収集可能な手法であることが示された．

5.10. まとめ

本章の研究では，発話対象を人間と想定した発話収集を Web アンケートを活用して実施し，有用性の検証を行った．この手法を「Web アンケートを活用した GiFT 手法」と呼ぶこととした．この手法の特徴は，Web アンケートを実施したのでは，被験者にとっての発話対象がパソコンなどの機械にあたるため，

発話の多様性が高まらないことを考慮し，被験者に Web アンケートを回答するにあたり，発話対象は人間だと意識させる点である．

提案手法を実施するにあたり「介護を受ける」ドメインで 9 シチュエーションを設定した．

Web アンケートの特性上，シチュエーションに誤解を招くような不備があったとしても，インタビューのようにオペレータはいないため，都度被験者を軌道修正することは難しい．そのため，最初に小規模実験として 30 人の被験者に Web アンケートを回答させ，そこで得たコーパスとインタビュー形式のコーパスとを比較した．その結果，POS の分布に大きな差があることが確認された．この実験から，ドメインの違いが評価結果に影響していることが考えられ，本手法の特徴である「発話対象を人間と教示すること」の効果にフォーカスするさらなる実験を実施することにした．

小規模実験の結果を踏まえた続く実験では，Web アンケートに設けたシチュエーションはそのままに，被験者を合計 200 人用意した．被験者は，「発話対象は人間の介護士だと教示した提案手法（GiFT 手法）によるグループ A」と，「発話対象は介護ロボットだと教示したグループ B」にそれぞれ 100 名とし，Web アンケートにより回答発話文を収集しコーパスとした．

時間的負荷指数を用いた負荷の検証では，インタビュー形式による収集手法より，グループ A は約 56%少なく，グループ B は約 61%少ないことを確認した．時間的負荷がおよそ半減したのは，インタビューでは被験者とオペレータの二人の作業者がいるのに対し，Web アンケートでは被験者が一人で回答する違いが大きい．また，グループ B の方が提案手法のグループ A より低負荷である理由は，機械向けの発話であったため，発話が単純化され，Web アンケート回答時の入力負荷が少なかったためと予想される．以上から，提案手法は，インタビュー手法と比べ時間的負荷は約半減させられることが示された．

発話の多様性の検証では，グループ A, B の両コーパスを形態素解析による比較をした結果，提案手法によるグループ A のコーパスは，延べ形態素数，一発話あたりの形態素数，異なり形態素数が，グループ B のコーパスより多く含ま

れていることが確認され，多様性の高い発話が収集されていることが示された．また，一発話あたりの異なり形態素数比 TUR を比較した結果，提案手法が多様な形態素を収集しやすい手法であることが確認され，提案手法の有用性が示された．

- **今後の課題**

今後の課題として，シチュエーションの検討が挙げられる．具体的には，介護の中でも「入浴」や「お手洗い」といった状況に対し，求められる介護の内容を精査し，アンケートの設問を潤沢にしていくことが挙げられる．

また，本章の実験では，最終的にボランティアの被験者 200 人に Web アンケートを実施した．そのため，無償で対応可能な拘束時間として約 10 分と定め，その中で回答可能な設問数（シチュエーション数）が 9 問という考えのもと，Web アンケートを実施した．シチュエーション数を増やすことにより，発話文の収集量は増加可能と考えるが，被験者にとっては負担が大きくなるため，無効発話が多く収集される恐れもある．そのため，被験者のモチベーションを維持したまま Web アンケートを完了させる工夫も合わせて必要になる．

設問数の増加に伴うエラーチェックも検討課題となる．本章の実験では，大学にて教員より学生に任意でのアンケート案内を行った．また，作業時間も 10 分程度であり，収集されるデータの品質は担保される見込みがあったため特に実施しなかったが，設問数の増加に伴い被験者が回答することを億劫に感じるなどにより，無効発話（不真面目な回答など）が増加する予想がある．さらに，教員から学生への案内ではなく，一般に幅広い被験者より収集するため，ポイントサイトなどを活用し収集する場合は，依頼主の顔が見えない第三者からのアンケート依頼となり，さらに無効発話が増加する恐れがある．また，被験者がさらに増えた状態では，人手により無効発話の選別を行うのは作業者の大きな負担になりえる．この対策の一つとして，設問の途中で誰でも正解するような簡単な設問を設ける案が考えられる．たとえば，「今年は西暦何年か？」といった設問である．こうした簡単な設問に正解できない被験者の回答データは，そもそも設問の確認さえしていないと判断し，その被験者のデータは検証から

除外するという方法である．こうしたフィルタリングを行うことにより，データの品質を上げ，作業者の負担を抑えるだけでなく，無効となった被験者に対する報酬の発生を抑えることも可能である．ただし，この対策は副次的な案であり，被験者全体の無効発話を抑える工夫の方が主とすべき検討課題と考える．

「介護を受ける」というドメインにおいては，最適な被験者の選別と，回答方法の改善が課題として挙げられる．本章の実験では，コンタクトしやすい学生を主な被験者とした．対象を広げ世代均衡とする方法もあるが，高齢者を対象とする方が適切とする考えもある．その一方，Web アンケートに回答するためには，パソコンなどを扱える一定のスキルが必要だが，パソコンの扱いに慣れていない場合は，回答作業に強い負担を感じたり，操作に不慣れなために不自然な回答をしたりする懸念がある．したがって，パソコンなどの扱いに不慣れな被験者からでも収集可能な工夫を合わせて検討する必要がある．

GiFT 手法と名付けた提案手法は「発話対象を人間と想定する」という特徴があるが，対象となる人間の属性による影響を検討することが挙げられる．たとえば，本実験では介護士を発話対象としたが，家族に介護してもらう場合は異なる発話になると考える．また，同じ介護士でも，初対面の介護士と，面識のある介護士では発話内容が異なると予想されるためである．

付録：収集した発話例

グループ A（発話対象が介護士の場合）

Situation No.	Utterance
1	体を起こしていただけませんか。
1	体を起こしてほしい
1	起こしてもらってもいい？
2	介護士さん　立ち上がらせてもらいたい
2	立ち上がりたい
2	腰を上げたい
3	車いすに座りたいので手伝ってほしいです
3	すみません、車椅子に移動したいのですが、 手を貸してもらえませんか？
3	すいません、あの、車いすに移していただけたいんですけど
4	手が疲れたので押してもらってもいいですか？
4	すまんが、車椅子押してくれんかのう
4	車いすを押してくれませんか
5	ちょっと止めてください
5	ちょっと止まってもらっていいですか
5	すこし止まってください
6	ベッドで休みたいので手をかしてください
6	ベッドで寝たい
6	横になりたい
7	ご飯を頂きたいです
7	ご飯をください
7	ご飯を食べさせてください
8	味噌汁飲みたいので持ってきてください
8	味噌汁が飲みたいから飲ませてください
8	味噌汁が飲みたい
9	食べ終わったので片づけてもらえますか？
9	食べ終わったので、下げてください。
9	食事を下げてもいいです

グループ B（発話対象が介護ロボットの場合）

Situation No.	Utterance
1	上体を起こして
1	体を起こしてください
1	起きたい
2	立ちたい
2	立たせて下さい
2	立たせて
3	車いすに乗りたい
3	車いすに座りたい
3	車椅子に移せ
4	車椅子押して
4	車いすを押してください
4	押してほしい
5	止めてください
5	車いすを止めて
5	止めて
6	ベッドまで運んで
6	ベッドに移りたい。
6	ベッドに戻してください
7	ご飯を食べさせてください。
7	ご飯を食べさせて
7	ごはん食べたい
8	味噌汁を口に入れて.
8	味噌汁が飲みたい
8	味噌汁
9	ご馳走様
9	下げて大丈夫です
9	ごちそうさま

無効発話例

グループ A（発話対象が介護士の場合）

Situation No.	Utterance	Reason
1	おはようございます。	発話文からタスクを想定できず.
5	目的の事を言って止めてもらう。	発話文ではない.
6	今日もありがとうございました、お疲れ様です。	発話文からタスクを想定できず.

グループ B（発話対象が介護ロボットの場合）

Situation No.	Utterance	Reason
1	システムナンバー００２！	発話文からタスクを想定できず.
4	こんにちは	発話文からタスクを想定できず.
5	ご使用ありがとうございました	自分を介護ロボットと勘違いしていると思われる.

第6章　　結論

最後に，本研究の総括として，第 1 章の問題提起から，第 5 章の Web アンケートを用いた GiFT 手法によるコーパス収集に至るまでの本研究の内容をまとめ，その成果と今後の課題について述べる．

6.1. 本研究のまとめと成果

本研究では，多様性の高い発話を収集する収集手法について研究したものである．昨今のクラウド音声認識技術に触れたユーザは，システムが許容する発話の自由さが広がったと考える，あるいは，機器に対する認識が「単純な機械」から「高度な言語処理技術を搭載した機器」と変化することにより，単純な定型音声コマンドベースの発話から，より人間との会話に近い発話で機器を操作したいニーズを考慮し，多様性の高い発話を整備することを目的としている．

第 1 章では，IVI などの車載機器を中心に音声認識操作に関する技術動向，既往研究について述べた．これらから，コーパスの重要性について説明するとともに，本研究の目的が多様性の高い発話の整備であるとした．この目的を達成するための方針として，収集環境やプロセスに工夫の余地がある「疑似的な環境を用いた発話収集手法」を採択し，多様性の高い発話の収集手法を提案することを目標とした．また，こうした発話の収集手法では多くの手間を要していたことから，より容易な方法で収集可能な新たな収集手法を提案することも目標の一つとした．

第 2 章では，コーパスには様々な種類があり，音声認識操作に限らず言語学，教育各分野でも研究が行われている．本研究では，特定のシチュエーションにおいて機器に対しどのような音声発話を行うかを調査し収集した，タスク指向型対話システムのための「タグ付き音声言語テキストコーパス」が研究対象であることを示した．また，発話の多様性は，一発話あたりの形態素数，異なり形態素数，一発話あたりの異なり形態素数比 TUR を用いて検証することを定義

した．

第 3 章では，インタビュー形式による多様性の高い機器操作発話収集手法を提案し，その有用性を検証した．具体的には，商用利用されている，クラウド音声認識技術を使った POI 検索機能の発話ログデータとの比較を行った．ログデータとの比較により，提案手法の発話は，一発話あたりの形態素数，異なり形態素数が多く発話の多様性が高いことを確認した．一発話あたりの異なり形態素数比 TUR の検証においても，提案手法は異なり形態素を効率的に収集可能な手法であることが示された．

また，言語理解器を用いた意図推定精度の検証では「発話ログデータと提案手法のコーパスを混ぜた混合学習データ」が最も高い意図推定精度が得られることを確認し，提案手法のコーパスは実用性においても有用であることが示された．

第 4 章では，コーパス収集作業の負荷低減手法として，probing を検証した．Probing は，被験者より 1 つのシチュエーションから複数の発話を得ることを可能にする．これにより限られた被験者数から多くの発話を収集することが可能である．この試みは第 3 章にて実施していたが，probing それ自体による効果は検証されていなかった．

検証においては，まず，従来手法の第一発話のみのコーパスの異なり形態素数と一致する提案手法の被験者数を算出した．その結果，従来手法の 100 人分のコーパスに対し，提案手法では 59 人分の発話で同程度の異なり形態素数を持つことが確認された．また，商用システムの発話ログデータに含まれる形態素をどの程度網羅できているかを比較検証した結果，従来手法の 100 人分の第一発話のみのコーパスと，提案手法の 59 人分のコーパスは，ほぼ同程度のカバレッジであることが示され，従来手法の 100 人分の第一発話のみの発話収集に対し，提案手法を用いることにより 41%被験者数を削減可能な効果があることが確認された．さらに，提案手法では，probing を実施した分，作業量は増加していることを考慮し，時間的負荷指数を用いた時間的負荷の検証を行った．その結果，提案手法は発話収集において第一発話のみ収集する場合より 37%時間的負荷が少ないことが確認され，本手法は時間的負荷の視点でも低負荷であることが示された．

第 5 章では，発話の多様性を保ちつつ，さらなるコーパス収集作業負荷を低減させる手法として，発話対象を人間と想定した Web アンケートによるコーパス収集手法を提案し，その有用性を検証した．この手法の特徴は，被験者に対し発話対象を人間と想定させることにある．本提案手法を「Web アンケートを活用した GiFT 手法」と名付けた．

「介護を受ける」というドメインのもと，小規模実験として被験者 30 人で提案手法を実施した．収集したコーパスと，第 3 章にて収集したインタビュー形式によるコーパスを形態素解析による比較を行ったところ，ドメインの違いにより POS の特徴が変動していることが考えられた．

小規模実験での知見を活かし，より詳細な検証を行うため，被験者を 200 人に増やした．また，提案手法の特徴である「発話対象が人間であると教示すること」による，発話の多様性への影響が明確になるように，被験者を 2 グループに分けた．グループ A は発話対象を人間の介護士と教示した提案手法のグループであり，グループ B は発話対象を介護ロボットと教示したグループである．

まず，この提案手法の実施の容易さを示すため，インタビュー形式による収集手法との時間的負荷指数による時間的負荷を比較した結果，Web アンケートを用いた各グループの時間的負荷は，グループ A にて 56%，グループ B にて 61%負荷を低減していることが示された．この顕著な低減は，Web アンケートではインタビューのようなオペレータが不要である点が大きい．

グループ A とグループ B のコーパスを形態素解析し，延べ形態素数，異なり形態素数，一発話あたりの形態素数，一発話あたりの異なり形態素数 TUR と変動係数の視点から，提案手法のグループ A の方が多様性の高い発話を収集できていることが示された．また，一発話あたりの異なり形態素数比 TUR を比較した結果でも提案手法は効率的に異なり形態素を収集可能な手法であることが示され，これは提案手法のグループ A の方が発話数は少ないにも関わらず，異なり形態素数が多い点にも表れていた．

以上から，本研究では多様性の高い発話収集手法として，インタビュー形式による収集手法を提案し，その時間的負荷低減手法として probing を用いた．さらには，より容易な収集手法として，GiFT 手法と名付けた発話対象を人間と

想定した Web アンケートに基づいた収集手法を提案した．これらにより，多様性の高い発話を収集しつつ，収集作業を時間的負荷指数にしてインタビュー形式による収集手法より 56%負荷低減させることに成功した．

6.2. 今後の課題

- **発話収集を拡大させるための課題**

多言語のコーパス収集方法の検討が挙げられる．本研究では，Web アンケートを活用した GiFT 手法により作業負荷を低減することが示されたが，この実験は日本語のみの実験であった．多言語のコーパスを収集するにはこうした Web アンケートを各言語分用意する必要があり，作業負荷は言語数分掛け算式に増すことが予想される．たとえば，自動翻訳ツールを活用したアンケート自動生成ツールの開発などが解決策となるが，アンケート文をその言語の自然な表現にするなどの考慮が必要である．

- **発話の多様性の定義に関する課題**

発話の多様性の検証方法の詳細化が挙げられる．本研究では，一発話あたりの形態素数，異なり形態素数，一発話あたりの異なり形態素数比 TUR といった指標から発話の多様性を検証した．しかし，これらの指標では，著しく長い発話に対し，多様性の高い発話として良いという判定になる．たとえば，「今日は気分がいいから山梨にドライブに行って，山中湖を見てきたいんだけど.」という POI 検索のための発話は，各指標からすると優位と判定する．この発話は第 3 章で示した一発話あたりの形態素数よりも，明らかに多くの形態素数を含んでいる．一方，発話自体は不自然というわけではない．こうした発話をどのように判断するかは検討課題である．一つのアイディアとしては，発話文に複数のタグを付け，各タグを意味する発話文を抽出し，検討をする方法が挙げられる．前述の発話文では，「今日は気分がいい」という利用者情報と，「山梨にドライブ」という一つ目のタスク，「山中湖に行く」という二つ目のタスク，

と三種類のタグ情報があると言える．このように，各タグを意味する発話文ごとに検証するアプローチが考えられる．

また，本研究では，異なり形態素数，一発話あたりの異なり形態素数比 TUR にて多様性の検証を行ったが，異なり形態素は一定量の発話文を収集すれば，そのタスクに必要な形態素はおおむね網羅可能である．しかし，将来的に真に「ユーザの思いついたままの発話」を考慮した場合，形態素に基づく検証とは全く異なるアプローチも検討する必要があると考える．たとえば，「あそこ行きたいんだよね.」という発話における「あそこ」は，話者によって異なるが，形態素としては変わらない．この例はユーザーカスタマイズ機能の一例に過ぎないが，形態素の量や異なり形態素による検証とは別軸の検証方法も求められると考える．

多様性の高い発話が必要な条件についても検討が必要と考える．本研究では，認知負荷を低減するため，思いついたままの発話による機器を操作可能とすることを想定し，多様性の高い発話に着目した．一方，人間と遜色ない超高度な言語処理技術が確立し，あらゆる機器にその技術が活用されたとしても，発話対象が産業用ロボットのような作業機器であれば，結局は端的な発話に収束する予想もある．対照的に，コミュニケーションを重視した，いわゆるヒューマノイドロボットであれば，同じタスクを達成する発話であっても，より多様性の高い発話が用いられる予想もある．このように，発話対象，ドメイン，発話の目的等に応じ，求められる発話の多様性も変化すると考える．

- **提案手法の使い分けに関する課題**

第 3 章のインタビュー形式による収集手法と，第 5 章の Web アンケートを活用した GiFT 手法の使い分けについて検討が必要と考える．5.5.3 項で述べたように，Web アンケートの提案手法にも懸念があり，被験者がアンケートに疑問を持った場合，オペレータに確認を取るといった行為は行えない．そのため，被験者は誤解したまま回答する懸念がある．こうしたリスクを低減するため，インタビューにおいても，Web アンケートにおいても，最初に数名の被験者にテストすることで，設問の調整を行った．しかし，こうした工夫を行っても設

問を誤解する被験者は一定数存在する．たとえば，Web アンケートには「発話を収集する実験である」といった説明があるにも関わらず，回答には「おじぎする」といった発話ではなく行動を回答した場合があった．また，5.10 節の今後の課題で述べたように，シチュエーション数が増えれば被験者の負担が増え無効発話（不真面目な回答）が増加することも予想される．対照的に，インタビューであればオペレータが誤解を訂正したり，監視者として無効発話の発生を抑える効果を期待できる．

被験者の誤解については，ドメインや設問が複雑な程発生しやすい．そのため，ドメインや設問が複雑であれば，オペレータのサポートを受けられるインタビューの方が適当と考える．こうした手法の選択手段を検討することが今後の課題として挙げられる．一つのアイディアとしては，収集対象のドメインに近しいコーパスがあるのであれば，そのパープレキシティを調べるという方法がある．これにより，続く単語の予想が難しい（様々な単語が入りうる）ならば，そのドメインや設問は複雑と考える一つの指標となる．

- **Probing の仕方に関する課題**

Probing の仕方に関する検討が挙げられる．本研究における probing は「他にどのような言い方がありますか？」と質問し，追加の発話を収集する手法であった．今後の検討課題として，状況・目的に応じた probing 方法の検討が挙げられる．Probing にはジェスチャーなどの非言語的な方法により，追加の情報を引き出す方法もある．インタビューであれば，オペレータが被験者の回答の意味を理解できていないような仕草をすることにより，被験者は他の表現を考えるといった方法も挙げられる．こうした状況は，話者がロボットに話しかけ，ロボットが適切に発話を理解できない場合に，話者が他の表現を用いる状況に近い．そのため，ロボットの言語処理がまだ完全ではない状況では，このような「言い直した発話」を学習データに含めることも言語処理向上に役立つと予想する．

また，Web アンケートにおいても probing の仕方に考慮が必要である．本研究では「他にどのような言い方がありますか？」と質問し，その結果は有効発話率が 95%を超えていた点からも適当な実施方法であった．ただし，「他にど

のような言い方がありますか？」と同様な意味合いの別表現として「他に思いつことはありますか？」,「他にありますか？」が思いつくが，これらの表現は曖昧性を含んでおり，たとえば「他にありますか？」では，被験者は「あと，〇〇もしてほしい」といった追加の要求を回答する可能性もありうる．Web アンケートでは，オペレータのサポートはなく，アンケート上に表示されたテキストやイラストをもとに，被験者のみの解釈により回答を行うため，明示的な表現をしつつ，発話の多様性を阻害しない表現を用いる必要がある．

参考文献

[1] 大須賀和美，大須賀博，“最新版 自動車用語辞典，” 精文館，ISBN 978-4-88102-048-7, Feb. 2016.

[2] 由雄淳一，“6-6 カーナビゲーション，” 画像電子学会誌 6.装置動向，vol. 46, no. 1, pp. 103-105, Jun. 2017.

[3] 横山利夫，波多野邦道，小沢浩一郎，樋山智，小高賢二，“自動車業界における自動運転実用化に向けた取り組み，” 学術の動向，vol. 25, no. 5, pp. 17-21, May. 2020.

[4] BMW インテリジェント・パーソナル・アシスタント，https://www.bmw.co.jp/ja/topics/offers-and-services/bmw-digital-services-and-connectivity/intelligent-personal-assistant.html [accessed on Jan. 6, 2023].

[5] 平沢純一，村上久幸，“組み込み機器向け音声インタフェース技術の開発プロセス，” 情報処理，vol. 51, no. 11, pp. 1464-1471, Nov. 2010.

[6] F. Weng, P. Angkititrakul, E.E. Shriberg, L. Heck, S. Peters, and J.H.L. Hansen, “Conversational in-vehicle dialog systems: The past, present, and future,” IEEE Signal Processing Magazine, vol. 33, no.6, pp. 49-60, Nov. 2016.

[7] 大淵康成，“カーナビゲーション向け音声処理技術の最前線，” 信学技報，vol. 114, no. 274, pp. 3-8, Oct. 2014.

[8] 本間健，額賀信尾，大淵康成，“クラウド型音声認識サービスの車載機利用のための音声処理技術開発，” 情処学研報，vol. 2013-SLP-98, no. 6, pp. 1-4, Oct. 2013.

[9] 古賀光，阿部正明，“発話自由度のある高度な音声操作システムによる運転行動への影響，” 自動車技術会論文集，vol. 50, no. 2, pp. 524-529, Mar. 2019.

[10] 大桑政幸，江部和俊，稲垣大，土居俊一，“ドライバの視聴覚認知に伴う負担度評価，” 計測自動制御学会論文集，vol. 36, no. 12, pp. 1079-1085, Dec. 2000.

[11] 朝尾隆文，和田隆広，“自動車運転におけるワークロードと安全，” 計測と制御，vol. 45, no. 8, pp. 671-676, Oct. 2006.

[12] 山口晶広，山本新，“ドライバの脇見検知のための視線の追跡と検出，” 電気学会論文誌 C（電子・情報・システム部門誌），vol. 121, no. 3, pp. 693-694, Mar. 2001.

[13] 日本自動車工業会，“画像表示装置の取り扱いについて 改訂版第 3.0 版，” https://www.jama.or.jp/operation/safety/display_system/pdf/jama_guidelines_v30_jp.pdf [accessed on Jan. 6, 2023].

[14] M. Hirose, T. Ishii, “A Method for Objective Assessment of Mental Work,” Bulletin of JSME, vol. 29, no. 253, pp. 2330-2335, Jul. 1986.

[15] 橋本洋志，竹田大祐，大山恭弘，石井千春，新妻実保子，橋本秀紀，“まつわりつきロボット動作の心理的評価，” 電気学会論文誌 C（電子・情報・システム部門誌），vol. 126, no. 1, pp. 83-90, Jan. 2006.

[16] 伊藤敏彦，甲斐充彦，岩本善行，水谷誠，由浅裕規，小西達裕，伊東幸宏，“目的地設定タスクにおける対話状況の違いによる言語・音響的特徴の比較，”情報処理学会論文誌，vol. 43, no. 7, pp. 2118-2129, Jul. 2002.

[17] 森田祐介，古屋友和，田村総，平尾章成，“視線と音声を用いたマルチモーダルインタフェースによるワークロードの低減，” 自動車技術会論文集，vol. 51, no. 5, pp. 944-949, Sep. 2020.

[18] S. Yamamoto, J. Woo, W. H. Chin, K. Matsumura, and N. Kubota, “Interactive Information Support by Robot Partners Based on Informationally Structured Space,” J. Robot. Mechatron., vol. 32, no. 1, pp. 236-243, Feb. 2020.

[19] 鯵坂志門，中村壮亮，“マルチレイヤ AI を用いた分散型スマートハウスの実装. 知能と情報，” vol. 33, no. 1, pp. 572-581, Feb. 2021.

[20] 小林 峻也，萩原 将文，“ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム，” 人工知能学会論文誌，vol. 31, no. 1, pp. 1-10, Jan. 2016.

[21] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” Communications of the ACM, vol. 9, no. 1, pp. 36-45, Jan. 1966.

[22] Richard S Wallace, “The anatomy of ALICE,” Parsing the Turing Test, ISBN 978-1-4020-6710-5, pp. 181-210, Jan. 2009.

[23] 速水悟，管村昇，“音声対話システムの研究と実用化の動向(<小特集>音声によるコンピュータとの対話を目指して)，” 日本音響学会誌，vol. 50, no. 7, pp. 574-580, Jul. 1994.

[24] 狩野芳伸，“コンピューターに話が通じるか:対話システムの現在，” 情報管理，vol. 59, no. 10, pp. 658-665, Jan. 2017.

[25] T. Homma, K. Shima, and T. Matsumoto, “Robust utterance classification using multiple classifiers in the presence of speech recognition errors,” Proc. 2016 IEEE Workshop on Spoken Language Technology (SLT), pp. 369-375, San Diego, U.S.A. , Dec. 2016.

[26] T. Homma, Y. Obuchi, K. Shima, R. Ikeshita, H. Kokubo, and T. Matsumoto, “In-vehicle voice interface with improved utterance classification accuracy using off-the-shelf cloud speech recognizer,” IEICE Trans. Inf. & Syst., vol. E101-D, no. 12, pp. 3123-3137, Dec. 2018.

[27] J. Devlin, M-W Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), pp. 4171-4186, 2019. https://arxiv.org/pdf/1810.04805.pdf [accessed on Jan. 6, 2023].

[28] Z. Lan, M.Chen, S. Goodman, K. Gimpel, P. Sharma, “Albert: A lite bert for self-supervised learning of language representations,” https://arxiv.org/abs/1909.11942, Feb. 2020. [accessed on Jan. 6, 2023].

[29] N. M. Fraser, G. N. Gilbert, “Simulating speech systems,” Computer Speech & Language, vol. 5, no. 1, pp. 81-99, Jan. 1991.

[30] M. Okamoto, N. Nakayama, “Wizard of Oz Method for Constructing Conversational Web Agents,” Transactions of the Japanese Society for Artificial Intelligence: AI, vol. 17, no. 3, pp. 293-300, 2002.

[31] C. T. Hemphill, J.J. Godfrey, and G. R. Doddington, “The ATIS spoken language systems pilot corpus,” Proc. 3rd DARPA Speech and Natural Language Workshop, pp. 96-101, Hidden Valley, U.S.A. , Jun. 1990.

[32] L. Hirschman, “Multi-site data collection for a spoken language corpus,” Proc. 5th DARPA Speech and Natural Language Workshop, pp. 7-14, New York, U.S.A. , Feb. 1992.

[33] G. Tur, D. Hakkani-Tur, L. Heck, “What is left to be understood in ATIS?,” Proc. IEEE Workshop on Spoken Language Technology (SLT), pp. 19-24, Berkeley, U.S.A. , Dec. 2010.

[34] C. Chelba, M. Mahajan, and A. Acero, “Speech utterance classification,” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 280-283, Hong Kong, China, Apr. 2003.

[35] G. Mesnil, X. He, L. Deng, and Y. Bengio, “Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding,” Proc. Annual Conference of the International Speech Communication Association (Interspeech), pp. 3771-3775, Lyon, France, Aug. 2013.

[36] G. Pai, S. Widrow, J. Radadiya, C. D. Fitzpatrick, M. Knodler, A. K. Pradhan, “A Wizard-of-Oz experimental approach to study the human factors of automated vehicles: Platform and methods evaluation,” Traffic injury prevention, vol. 21, no. sup1, pp. 140-144, Aug. 2020.

[37] A. Sahaï, J. Barré, M. Bueno, “Urgent and non-urgent takeovers during conditional automated driving on public roads: The impact of different training programmes,” Transportation Research Part F: Traffic Psychology and Behaviour, vol. 81, pp. 130-143, Aug. 2021.

[38] M. Ahmed, C. Dixit, R.E. Mercer, A. Khan, M. R. Samee, F. Urra, “Multilingual corpus creation for multilingual semantic similarity task,” In Proceedings of The 12th Language Resources and Evaluation Conference, pp. 4190-4196, May. 2020.

[39] S. Mishra, A. Sharma, “Crawling Wikipedia Pages to Train Word Embeddings Model for Software Engineering Domain,” In 14th Innovations in Software

Engineering Conference (formerly known as India Software Engineering Conference), pp. 1-5, Feb. 2021.

[40] 小池隆斗，萩原将文，“スタイル変換を用いた YouTube 動画タイトルの生成支援システム，” 日本感性工学会論文誌，vol. 21, no. 1, pp. 49-55, Feb. 2022.

[41] L. Liu, Q. Wang, Y. Li, “Improved Chinese Sentence Semantic Similarity Calculation Method Based on Multi-Feature Fusion,” J. Adv. Comput. Intell. Intell. Inform., vol. 25, no. 4, pp. 442-449, Jul. 2021.

[42] J. Le, H. Yamana, “Cross-lingual Investigation of User Evaluations for Global Restaurants,” Information and Media Technologies, vol. 10, no. 2, pp. 317-322, Jun, 2015.

[43] 松本義之，井上仙子，“下関地域における Twitter を利用した観光情報分析，” バイオメディカル・ファジィ・システム学会誌，vol. 22, no. 2, pp. 59-66, Jul. 2020.

[44] 泉翔太，堀太成，山根達郎，全邦釘，藤森祥文，森脇亮，“Deep Learning を用いたマイクロブログ投稿文の災害情報分類，” AI・データサイエンス論文集，vol. 1, no. J1, pp. 398-405, Nov. 2020.

[45] 安藤慎太郎，藤原弘将，“テレビ録画とその字幕を利用した大規模日本語音声コーパスの構築，” 研究報告音声言語情報処理 (SLP)，vol. 2020-SLP-246, no. 8, pp. 1-7, Dec. 2020.

[46] 滝本清仁，廣瀬英雄，“マトリックス分解を用いた推薦アルゴリズム：Netflix への適用，” 日本計算機統計学会第 22 回大会論文集，1B-2, pp. 17-22, May. 2008.

[47] 小木曽智信，“日本語コーパスの今(<特集>デジタル時代の日本語)，” 情報の科学と技術，vol. 64, no. 11, pp. 463-468, Nov. 2014.

[48] 滝島雅子，“話し言葉と書き言葉における敬語名詞の語彙比較 CEJC と BCCWJ のデータを用いて，” 計量国語学，vol. 32, no. 6, pp. 315-330, Sep. 2020.

[49] 中尾亮太，黒橋禎夫，“日本語話し言葉書き言葉変換による大学講義の日英翻訳の精度向上，” 自然言語処理，vol. 28, no. 4, pp. 1034-1052, Dec. 2021.

[50] E. Lombard, “Le signe de l’elevation de la voix,” Ann Maladies de L'Oreille et du Larynx , vol. 37, no. 2, pp. 101-119, 1911.

[51] 竹沢寿幸，末松博，“音声・テキストコーパスとその構築技術,標準化動向 (<特集> 「コーパスに基づく音声・自然言語処理」)，” 人工知能学会誌，vol. 10, no. 2, pp. 168-180, May. 1995.

[52] International Committee for Co-ordination and Standardisation of Speech Databases, http://www.cocosda.org/ [accessed on Jan. 6, 2023].

[53] Linguistic Data Consortium, https://www.ldc.upenn.edu/ [accessed on Jan. 6, 2023].

[54] European Language Resources Association, http://catalogue.elra.info/en-us/ [accessed on Jan. 6, 2023].

[55] 板橋秀一，“音声コーパス事始め――日本を取り巻く状況――，” 日本音響学会誌, vol. 72, no. 10, pp. 669-670, Oct. 2016.

[56] 板橋秀一，“多言語音声処理に向けた音声資源の現状と課題，” 日本音響学会誌， vol. 64, no. 12, pp. 721-726, Dec. 2008.

[57] 国立国語研究所，https://www.ninjal.ac.jp/ [accessed on Jan. 6, 2023].

[58] 加藤祥，菊地礼，浅原正幸，“『現代日本語書き言葉均衡コーパス』に基づく指標比喩データベース，” 自然言語処理，vol. 27, no. 4, pp. 853-887, Dec. 2020.

[59] 乙武北斗，内田ゆず，高丸圭一，木村泰知，“BERT による周辺文脈を考慮したオノマトペの語義分類手法の提案，” 知能と情報，vol. 32, no. 1, pp. 518-522, Feb. 2020.

[60] 井上昂治，山本賢太，中村静，高梨克也，河原達也，“アンドロイド ERICA の傾聴対話システム-人間による傾聴との比較評価-，” 人工知能学会論文誌，vol. 36, no. 5, H-L51_1-12, Sep. 2021.

[61] 西村良太，森雷太，太田健吾，北岡教英，“音声対話システムのための自由発話に対応した照応解析による入力発話への話題補完手法，” 人工知能学会論文誌，vol. 37. no. 3, IDS-F_1-13. May. 2022.

[62] 小林恭輔，高久雅生，“楽曲探索を支援するための類似楽曲提示手法，” 情報知識学会誌，vol. 32, no. 2, pp. 287-293. Sep. 2022.

[63] 小河妙子，藤田知加子，“小学校国語教科書に掲載されている単語の種類数と出現頻度，” 読書科学，vol. 60, no. 4, pp. 241-253, Nov. 2018.

[64] 李広微，金明哲，“モデリングから見る小説における助詞の経時変化，” 情報知識学会誌，vol. 31, no. 3, pp. 371-383, Sep. 2021.

[65] 村井源，松本斉子，“会話文での自称詞と対称詞の出現傾向と役割-話し言葉と書き言葉での相違から，” 情報知識学会誌，vol. 32, no. 1, pp. 3-14, Nov. 2021.

[66] 李在鎬，“日本語教育学の課題に対して計量分析は何ができるか，” 計量国語学，vol. 32, no. 7, pp. 372-386, Dec. 2020.

[67] 宮内拓也，浅原正幸，“日本語における名詞句の情報構造と語順の相関についての 統計的検討，” 自然言語処理，vol. 27, no. 2, pp. 361-381, Sep. 2020.

[68] 許家瑶，“日本語でどのように物語を導入するか―前置き連鎖が利用されていない事例に着目して―，” ことば，vol. 42, pp. 216-232, Dec. 2021.

[69] 西垣知佳子，星野由子，安部朋世，神谷昇，小山義徳，石井雄隆，“小学生のためのデータ駆動型英語学習支援サイトの開発と公開，” 小学校英語教育学会誌，vol. 20, no. 1, pp. 367-382, Mar. 2020.

[70] Amazon Polly, Amazon.com, Inc. https://aws.amazon.com/jp/polly/ [accessed on Jan. 6, 2023].

[71] Ruby Talk, 日立ソリューションズ・テクノロジー, https://www.hitachi-solutions-tech.co.jp/iot/solution/voice/Ruby_Talk/index.html [accessed on Jan. 6, 2023].

[72] 原功，“サービスロボットのためのコミュニケーションフレームワーク，” 計測と制御．vol. 57, no. 10, pp. 706-710, Oct. 2018.

[73] 吉田直人，矢野美由紀，米澤朋子，“アナウンサエージェントの対話的ニュース読み上げ手法による寄り添い感・信頼感の向上と理解導入効果，” ヒューマンインタフェース学会論文誌，vol. 23, no. 2, pp. 165-176, May. 2021.

[74] 小林隆夫，“小特集にあたって(<小特集>音声合成に関する研究の動向)，” 日本音響学会誌，vol. 67, no. 1, pp. 15-16, Dec. 2010.

[75] 匂坂芳典，“種々の音韻連接単位を用いた日本語音声合成，” 信学技報，vol. SP87-136, pp. 47-52. Mar. 1987.

[76] Y. Sagisaka, “Speech synthesis by rule using an optimal selection of non-uniform synthesis units,” ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing. IEEE Computer Society, pp. 679-682. Apr. 1988.

[77] Y. Sagisaka, N. Kaiki, N. Iwahashi, K. Mimura, “ATR ν-Talk speech synthesis system,” Proc. ICSLP, pp. 483-486, Oct. 1992.

[78] 党建武，“第 3 編音声合成：（総説）音声合成技術の現状と展望，” 伊福部達，西阪仰，小坂哲夫，仲山豊秋，古井貞熙，伊藤彰則，辻川剛範，笹嶋宗彦，光吉俊二，藤田覚，能勢隆，金田隆志，加藤集平，広瀬啓吉，駒谷和範，北岡教英，荒木健治，小林優佳，坂井誠，名田徹，片桐恭弘，中川拓哉，額賀信尾，成合功一郎，大畑佳織，藤堂栄子，上田裕市，向川啓太，坂口毅雄，谷田部智之，“進化するヒトと機械の音声コミュニケーション，” 株式会社エヌ・ティー・エス，ISBN 978-4-86469-065-2, pp. 125-140, Sep. 2015.

[79] 河井恒，戸田智基，山岸順一，平井俊男，倪晋富，西澤信行，津崎実，徳田恵一，“大規模コーパスを用いた音声合成システム XIMERA，” 電子情報通信学会論文誌 D, vol. J89-D. no. 12, pp. 2688-2698, Dec. 2006.

[80] 中川聖一，赤松裕隆，西崎博光，“音声認識用言語モデルのためのタスク適応化と定型表現の利用，” 自然言語処理，vol. 6, no. 2, pp. 97-115. Jan. 1999.

[81] 古井貞熙，“人と対話するコンピュータを創っています-音声認識の最前線-，” 角川学芸出版，ISBN 978-4-04-621640-3. Feb. 2009.

[82] Speech Resources Consortium, http://research.nii.ac.jp/src/JNAS.html [accessed on Jan. 6, 2023].

[83] H, Hans-Günter; P, David, “The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions,” ISCA ITRW ASR2000, Paris, France, pp. 18-20, Sep. 2000.

[84] 倉田岳人，市川治，西村雅史，“ユーザの発話傾向分析に基づく車載機器操作のための音声入力手法の検討，” 電子情報通信学会 D，vol. J93-D, no. 10, pp. 2107-2117, Oct. 2010.

[85] 下畑さより，山崎貴宏，坂本仁，北村美穂子，村田稔樹，“特許翻訳における専門用語辞書構築，” 言語処理学会第 11 回年次大会論文集，A2-4, Mar. 2005.

[86] 下畑さより，井佐原均，“日英特許コーパスからの専門用語対訳辞書の自動獲得，” 自然言語処理，vol. 14, no. 4, pp. 23-41, Sep. 2007.

[87] 相良かおる，小野正子，小作浩美，鈴木隆弘，高崎光浩，嶋田元，“分かち書き用辞書 ComeJisyo の評価，” 医療情報学，vol. 32, no. 6, pp. 301-307, 2012.

[88] 京都大学大学院情報学研究科知能情報学専攻知能メディア講座言語メディア分野 黒橋・河原研究室，“日本語形態素解析システム JUMAN，” https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN [accessed on Jan. 6, 2023].

[89] 奈良先端科学技術大学院大学情報科学研究科，“ChaSen,” https://chasen-legacy.osdn.jp/ [accessed on Jan. 6, 2023].

[90] T. Kudo, “MeCab: Yet another part-of-speech and morphological analyzer,” http://taku910.github.io/mecab/ [accessed on Jan. 6, 2023].

[91] 近藤明日子，“『明六雑誌コーパス』『国民之友コーパス』の構築——形態論情報を付与した近代雑誌コーパスの設計——，” 日本語の研究，vol. 12, no. 4, pp. 167-174, Oct. 2016.

[92] 高丸圭一，内田ゆず，乙武北斗，木村泰知，“地方議会会議録コーパスにおけるオノマトペ出現傾向と語義の分析，” 人工知能学会論文誌，vol. 30, no. 1, pp. 306-318, Jan. 2015.

[93] 松河秀哉，大山牧子，根岸千悠，新居佳子，岩﨑千晶，堀田博史，串本剛，川面きよ，杉本 和弘，“トピックモデルによるテキスト分析を支援するソフトウエアの開発，”日本教育工学会論文誌，vol. 42, no. Suppl, pp. 37-40, Dec. 2018.

[94] 竹﨑あかね，細羽見喬，法隆大輔，木浦卓治，“農林水産分野における自動索引付けに有効な言語資源の開発と評価，”農業情報研究，vol. 19, no. 1, pp. 10-15, Apr. 2010.

[95] 小木曽智信，小町守，松本裕治，“歴史的日本語資料を対象とした形態素解析，” 自然言語処理，vol. 20, no. 5, pp. 727-748, Dec. 2013.

[96] K. Takeda, H. Fujimura, K. Itou, N. Kawaguchi, S.Matsubara, and F. Itakura, “Construction and evaluation of a large in-car speech corpus,” IEICE Trans. Inf. & Syst., vol. E88-D, no. 3, pp. 553-561, Mar. 2005.

[97] 河口信夫，松原茂樹，山口由紀子，武田一哉，板倉文忠，“CIAIR 実車走行車内音声データベース，”情処学研報，vol. 2003-SLP-49, no. 24, pp. 139-144, Dec. 2003.

[98] 鈴木崇史，影浦峡，“名詞の分布特徴量を用いた政治テキスト分析，”行動計量学, vol.38, no.1, pp. 83-92, Jul. 2011.

[99] 嶋和明，本間健，池下林太郎，小窪浩明，大淵康成，佘錦華，“インタビュー形式によるカーナビ向け自由発話コーパス収集法，”電子情報通信学会論文誌 D, vol. J101-D, no. 2, pp. 446-455. Feb. 2018.

[100] Faurecia Clarion Electronics Co., Ltd., Smart Access, http://www.smart-acs.com/ [accessed on Jan. 6, 2023].

[101] Faurecia Clarion Electronics Co., Ltd., Intelligent VOICE, https://www.clarion.com/jp/ja/products-personal/service/IntelligentVoice/index.html [accessed on Jan. 6, 2023].

[102] Panasonic Corporation. Drive P@ss, https://panasonic.jp/car/guide/voice_recognition/ [accessed on Jan. 6, 2023].

[103] Google LLC., Google アシスタント，https://assistant.google.com/intl/ja_jp/ [accessed on Jan. 6, 2023].

[104] Apple Inc., Siri, https://www.apple.com/jp/siri/ [accessed on Jan. 6, 2023].

[105] 小磯花絵，“『日本語日常会話コーパス』モニター公開版の構築，” 計量国語学，vol. 32, no. 2, pp. 133-142, Jun. 2019.

[106] 小磯花絵，“『日本語日常会話コーパス』の設計と構築，” SIG-SLUD, vol. 5, no. 2, pp. 1-2, Oct. 2017.

[107] N. Okazaki, “Classias: a collection of machine-learning algorithms for classification,” http://www.chokkan.org/software/classias/ [accessed on Jan. 6, 2023].

[108] T. Kudo, “CRF++: Yet another CRF toolkit,” https://taku910.github.io/crfpp/ [accessed on Jan. 6, 2023].

[109] 内元清貴，関根聡，井佐原均，“最大エントロピーモデルに基づく形態素解析 未知語の問題の解決策，” 言語処理学会，自然言語処理，vol. 8, no. 1, pp. 127-141, Jan. 2001.

[110] K. Shima, T. Homma, M. Motohashi, R. Ikeshita, H. Kokubo, Y. Obuchi, and J. She, “Efficient Corpus Creation with Reduced Number of Subjects for Natural Language Understanding,” Proc. of the 8th Int. Symp. on Computational Intelligence and Industrial Applications and the 12th China Japan Int. Workshop on Information Technology and Control Applications (ISCIIA & ITCA 2018), No.3M1-3-5, Nov. 2018.

[111] K. Shima, T. Homma, M. Motohashi, R. Ikeshita, H. Kokubo, Y. Obuchi, and J. She, “Efficient Corpus Creation Method for NLU Using Interview with Probing Questions,” J. Adv. Comput. Intell. Intell. Inform., vol. 23, no. 5, pp. 947-955, Sep. 2019.

[112] K. J. Roulston, “Probes and probing,” in The SAGE Encyclopedia of Qualitative Research Methods, L.M. Given, Ed., SAGE Publications, ISBN 978-1-4129-4163-1, pp. 681-683, Aug. 2008.

[113] T. Homma, A. S. Arantes, M. T. G. Diaz, M. Togami, “Maximizing SLU performance with minimal training data using hybrid RNN plus rule-based

approach," Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, pp. 366-370, Melbourne, Australia, Jul. 2018.

[114] K. Shima, J. She, Y. Obuchi, and A. M. Iliyasu, "Web-Questionnaire-Based Method for Creating Corpus Containing a Large Number of Morphemes." Proc. the International Conference of IWACIII2021, no. T2-3-3, Beijing, China, Nov. 2021.

[115] K. Shima, J. She, Y. Obuchi, and A. M. Iliyasu, "Web-Questionnaire-based Corpus Creation Under Assumption of Human as Speech Targets," J. Adv. Comput. Intell. Intell. Inform. vol. 26, no. 4, pp. 513-520, Jul. 2022.

[116] 藤本拓，原隆浩，西尾章治郎，"自然な発話により操作可能なカーナビゲーションシステムの開発，"電子情報通信学会論文誌 D， vol. J96-D, no. 11, pp. 2815-2824, Nov. 2013.

[117] 小木津武樹，"自動運転による無人移動サービスに関する取り組み，" The Japanese Journal of Rehabilitation Medicine, vol. 57, no. 2, pp. 149-153, Feb. 2020.

[118] 白砂堤津邪， "例題で学ぶ初歩からの統計学，" 株式会社日本評論社, ISBN 978-4-535-55498-6, pp. 133.166, Jan. 2009.

本研究に関連して発表した研究業績

雑誌論文

No.	年	著者名	掲載誌名	論文名	掲載詳細情報
[a]	2022	**K. Shima**, J. She, Y. Obuchi, and A. M. Iliyasu	Journal of Advanced Computational Intelligence and Intelligent Informatics	Web-Questionnaire-based Corpus Creation Under Assumption of Human as Speech Target	Vol. 26, No. 4. pp. 513-520, Jul. 2022.
[b]	2019	**K. Shima**, T. Homma, M. Motohashi, R. Ikeshita, H. Kokubo, Y. Obuchi, and J. She	Journal of Advanced Computational Intelligence and Intelligent Informatics	Efficient corpus creation method for NLU using interview with probing questions	Vol. 23, No. 5, pp. 947-955, Sep. 2019.
[c]	2018	**嶋和明**，本間健，池下林太郎，小窪浩明，大淵康成，佘錦華	電子情報通信学会論文誌 D	インタビュー形式によるカーナビ向け自由発話コーパス収集法	Vol. J101-D, No. 2, pp. 446-455, Feb. 2018.

学会発表

No.	年	著者名	国際会議名	論文名	掲載詳細情報
[d]	2021	**K. Shima**, J. She, Y. Obuchi, and A. M. Iliyasu	International Workshop on Advanced Computational Intelligence and Intelligent Informatics	Web-Questionnaire-Based Method for Creating Corpus Containing a Large Number of Morphemes.	No. T2-3-3, Beijing, China, Nov. 2021.
[e]	2018	**K. Shima**, T. Homma, M. Motohashi, R. Ikeshita, H. Kokubo, Y. Obuchi, and J. She	International Symposium on Computational Intelligence and Industrial Applications and International Workshop on Information	Efficient Corpus Creation with Reduced Number of Subjects for Natural Language Understanding	No. 3M1-3-5, Tengzhou, China, Nov. 2018.

表彰

No.	年	受賞名	発行元
[f]	2021	Session Best Presentation Award	International Workshop on Advanced Computational Intelligence and Intelligent Informatics
[g]	2018	Session Best Presentation Award	International Symposium on Computational Intelligence and Industrial Applications and International Workshop on Information

謝辞

本論文は，タスク指向型の音声操作に用いるコーパスの収集手法に関する研究を纏めたものになります．本論文を作成するにあたり，多くの方のご指導を賜れたことを心から光栄に思います．

東京工科大学教授，福島 E. 文彦博士，大淵 康成博士，大野 澄雄博士，亀田 弘之博士，佘 錦華博士には，ご多忙な中，本論文の審査と熱意溢れるご指導を頂き心より御礼申し上げます．

特に，大淵 康成教授は前職における尊敬する上司でもあり，音声認識や言語理解といった分野に私が関心を持つきっかけを頂きました．本論文を含め，これまでの研究においても随所で適格なご指導を頂き，行き詰った中でも突破口となる方向性を指し示して下さいました．

佘 錦華教授は，私が東京工科大学在学時の恩師でもあり，当時より本論文作成に至るまで正に手取り足取りご指導を頂きました．十数年にもわたるご指導を頂けたことは感謝という言葉に収まるものではありません．

Prince Sattam bin Abdulaziz University の Professor Abdullah M. Iliyasu，（株）日立製作所の本間 健博士，小窪 浩明博士，池下 林太郎氏（共著当時（株）日立製作所所属），フォルシアクラリオン・エレクトロニクス（株）の本橋 将敬氏からは，これまで論文の共著を頂く中で，手厚いご指導を賜れましたことを心より光栄に思います．

良き師に出会えることは生涯を通した宝とも言うべきことであり，本論文を作成するにあたりそれを実感できたことは，本論文を作成する中で得られた一つの大きな成果です．私の周りには常に多くの良き恩師，先輩，同僚に囲まれ，至らぬ点の多い私に多くの啓発，触発，気づきを与えて頂きました．これらなくして本論文もありえず，心より御礼申し上げます．

最後に，本論文作成にあたり家族の精神的支えがあったことは言うまでもありません．支えてくださった父母兄弟に心より感謝いたします．