社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

どう言ったかを何を言ったかで表す
フォーカスを含んだ発話及びその含意を反映したテキストを含む英語コーパス

鱸　　尚晃† 　中村　　哲†

† 奈良先端科学技術大学院大学先端科学技術研究科　〒630–0101 奈良県生駒市高山町 8916–5
E-mail: †{suzuki.naoaki.sg4,s-nakamura}@is.naist.jp

あらまし　音声対話において，意図は「何を言ったか（言語情報）」と，「どう言ったか（パラ言語情報）」で伝えられる．本研究では，パラ言語情報の中でも強調の一種であるフォーカスを扱い，フォーカスを含む音声を同一言語内でテキストへと言い換えることでパラ言語情報を言語情報へ変換した英語コーパスを作成，分析，公開した．コーパスとその分析結果は，パラ言語情報の言語情報化の可能性を実証するもので，今後パラ言語翻訳などへの応用が期待される．
キーワード　英語, パラ言語情報, フォーカス, 音声とテキスト, コーパス, 音声翻訳

Representing how it is said with what is said
Creation and analysis of an English corpus of focused speech and text reflecting paralinguistically expressed implications

Naoaki SUZUKI† and Satoshi NAKAMURA†

† Faculty of Science and Technology, Nara Institute of Sciencne and Technology　8916–5 Takayama-chou, Ikoma-city, Nara, 630–0101 Japan
E-mail: †{suzuki.naoaki.sg4,s-nakamura}@is.naist.jp

Abstract　In speech communication, people convey intentions through what is said (linguistic information) and how it is said (paralinguistic information). In this study, we address focus, a kind of emphasis among paralinguistic information, and create, analyse and publish an English corpus, which contains speech that differed in the placement of focus and text reflecting the corresponding implications. The corpus and its analysis demonstrate the possibility of converting paralinguistic information into linguistic information, which is expected to be applied to paralinguistic translation in the future.
Key words　English, paralinguistic information, focus in speech and text, corpus, speech translation

1. 研 究 背 景

音声コミュニケーションにおいて，人は自分の意図を伝えるために，「何を言うか（言語情報）」と「どのように言うか（パラ言語情報）」という２種類の情報を利用している [1]．パラ言語情報は，継続時間，強度，ピッチなどの超分節的な特徴によって表現される．同じ言語情報であっても，これらの韻律的特徴の変化により，異なる意味合いを伝えることができる．本研究では，パラ言語情報の１つである強調に注目し，その中でも特にフォーカス（focus）について扱う．代替意味論 [2] と呼ばれる意味論の枠組みでは，フォーカスは言語表現の解釈に関係する代替の存在を示唆するものとされる [3].

(1)　a.　**John** bought the apple.
　　b.　John **bought** the apple.

（1a）では，太字で示されているように *John* にフォーカスが置かれている．話し手は，*It was John who bought the apple* と示唆し，*Peter* や *Mary* といった，文脈的に考えられる他の存在を示すのと同時に，それらを *bought the apple* の動作主として否定している．一方，（1b）ではフォーカスは *bought* にあり，話し手は例えば *What John did was not sell the apple, but buy the apple* を含意する．（1a）と（1b）は言語情報的には同一であるが，フォーカスの位置の違いによって，異なる含意が表される．英語の音声では，フォーカスは核音節，すなわち文において最も顕著なアクセントによって表現される [4]．コミュニケーションにお

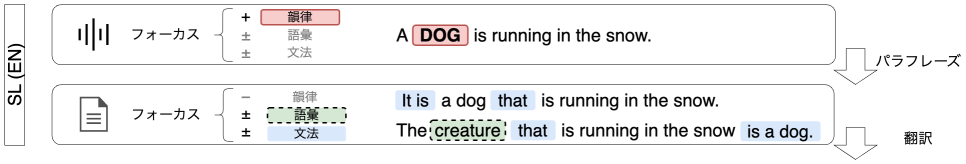

図 1: パラフレーズを用いたフォーカス翻訳の例．SL にてフォーカス音声を語彙と文法の変化を用いてパラフレーズし，そのテキストを TL へと翻訳している．

いて誤解を避けるためには，聞き手は核音節を中心に表されるフォーカスを認知し，それに付随する解釈を理解する必要がある．このタスクは，英語母語話者にとっては容易であるが，非英語母語話者には難しい [5]．したがって，異言語間でのコミュニケーションでは，このタスクを自動的に行うシステムが必要である．

音声翻訳システム（Speech Translation; ST）は，ソース言語（Source Language; SL）の音声をターゲット言語（Target Language; TL）のテキストあるいは音声に自動的に翻訳するシステムである．近年，ST は言語情報の翻訳において大きな進歩を遂げているが，これまで開発された ST の多くは，フォーカスを含むパラ言語的な情報を考慮する機能を持たない．従来の ST では，自動音声認識（Automatic Speech Recognition; ASR）による音声の書き起こしを元に翻訳が行われる．ASR は音声の内容のみを書き起こすようにデザインされているため，パラ言語情報は失われてしまう．結果として，先の例（1a），（1b）は異なる含意を持つにもかかわらず，現在の ST では訳出は全く同一のものとなる．

最近では，パラ言語情報の重要性が広く認識されるようになり，強さ，長さ，基本周波数など，SL の韻律を TL の音声に対応させ，声質 [6]，感情 [7], [8]，強調 [7], [9]〜[15] などパラ言語情報の翻訳に取り組む研究がなされている．しかし，韻律–韻律のマッピング手法は，SL の韻律に対応する韻律が TL で存在するような場合のみに制限される．例えば，英語の音声ではフォーカスには主に韻律が用いられる [16] が，言語一般では語彙や文法も用いられ，それぞれの手段への依存度は言語によって異なる [16]〜[18]．したがって，例えばフォーカスが主に語彙で表されるような言語が TL であった場合，韻律–韻律の手法ではフォーカスを正しくマッピングできない．

そこで，本研究では，韻律–韻律ではなく，韻律–言語情報（語彙と文法）マッピング，すなわち韻律をテキストへパラフレーズすることによるフォーカス変換の可能性を検討する．この手法を ST へ組み込み，パラフレーズされたテキストを翻訳モジュールの入力とすれば，パラ言語情報を保持したまま翻訳が可能になる（図 1）．韻律–言語情報へのフォーカス変換を実現するためには，モデルを学習するためのデータが必要である．しかし，著者の知る限り，異なる要素をフォーカスとする音声と，その含意を反映するテキストがペアになったコーパスは存在しない．最も関連性の高い研究では，強調の度合いが異なるテキスト項目，例えば（*It is a little bit hot* / *It is extremely hot*）と，それに対応する音声，例えば（*It is* ***hot*** / *It is* ***HOT***）を，テキストと同じ強調度合いになるように録音した英語コーパス [19] がある．しかし，強調の度合いに着目しているため，強調する要素は形容詞に固定されている．このデータは，後に音声からテキストへの強調変換のモデル構築に利用された [20]．また，[21] では，フォーカスの配置が異なる発話，例えば（***Sarah*** *closed the door*, *Sarah* ***closed*** *the door*）のような音声が収集されたが，研究の目的は音声合成において韻律の制御を実現することであったため，データには発話に対応する含意は含まれていない．さらに，*the, is* などの機能語にもフォーカスを当てることも可能であるが [22]，そのようなケースはまだ扱われていない．これらを踏まえ，本コーパスでは，機能語を含む文中全ての語をフォーカスの対象とした音声と，それぞれを含意が伝わるようにテキストへ言い換えた英語コーパスを構築[1]する．コーパス作成に加えて，収集データを基に，以下の 2 つの観点（A, B）からの分析も行う．**A. 発話においてフォーカスが置かれにくい語の特徴** 特別な場合を除き，英語においてフォーカスは機能語よりも内容語に置かれる傾向がある [4] が，今回の音声収集では，前述の通り機能語/内容語の区別関係なく全ての語をフォーカスの対象としている．過度に不自然な発話の収集を防ぐため，収集時において，録音者にはその語を強調することが不自然ならば録音をスキップするよう指示を与えた（2．2．2 節参照）．録音がスキップされた語，すなわちフォーカスを置くことが不自然だと判断された語についての分析は，音声処理における複数のタスク——テキストからフォーカス位置を推測し [23], [24]，音声合成にてフォーカスを制御するタスク [21] や，ASR においてフォーカス位置を特定 [20]——において，今後性能向上のための追加情報として利用できる可能性がある．**B. 韻律–言語情報へのパラフレーズ方法** これまで，一般的な言い換え，すなわち元のテキストがほぼ同じ意味を持つ別のテキストにされる際どのようなパターンがあるか [25]〜[27]，フォーカスを表すのにどのような言語表現が用いられるか [28], [29]，についての研究は行われてきた．しかし，フォーカスがパラ言語領域から言語領域へとどのように変換されるのかについては明らかではない．パラ言語領域から言語領域へのフォーカス情報の変換方法について，定量的・定性的な分析を行う．尚，本稿におけるコーパス収集及び分析 A の主要部分については，[30] に基づく．

1.1 英語におけるフォーカス

本節では，英語における韻律・語彙・文法それぞれを用いたフォーカスの表し方について要約する．

(注1)：https://dsc-nlp.naist.jp/data/speech/paralinguistic_paraphrase/ にて利用可能

1.1.1 韻　　律

前節で述べたように，英語の音声は，フォーカスを伝えるために，核音節を使用する．英語母語話者は，次のような手順で特定の情報を強調し，聞き手の注意を引きつける [4]．まず，意図に応じて，発話をイントネーション句（Intonation Phrase; IP）と呼ばれる小さな塊に分割する．そして，各 IP の中で最も重要な単語を選択し，その単語のアクセントのある音節を核音節とする．文脈に応じて，フォーカスは narrow focus と broad focus に分かれ，前者は聞き手の注意を IP 全体へ向けるもの，後者は IP の一部のみへ注意を向けさせるものとされる [4].

1.1.2 語　　彙

語彙によってフォーカスを表す際には，focus particle と呼ばれる一群の単語を用いることができ [31]，*only*，*even*，*alone* などが含まれる [29]．例えば *only* では，文脈的に考えられるその他の選択肢の中で，それが唯一、文の断言（assertion）を成立させる要素であることを示す [3].

(2) a: John touched the painting.
　　b: John only touched the painting.

（2a）と比較して，（2b）では，*broke* や *stole* など，*John* が *the painting* に対して行える動作の中で *touched* のみが成立することを示し，*touched* がフォーカスされる．同様に，*let alone* [32]，再帰代名詞 *himself/herself* [33]，*particularly, mainly* やその他多くの語彙 [31] がフォーカスを伝える語彙として用いられる．

1.1.3 文　　法

文構造を変化でもフォーカスを表せる．強調構文（*It was Simon who kicked the door.*）, 疑似強調構文（*What Mary bought was an apple*）, 倒置（*And then appears a bear*）, 受動態（*I was bit by a dog*） [22] などへの構造変化が可能である．文構造の再構成では，フォーカスのある語を文の後ろへ移動させることが多い．これは，英語の情報構造は主に Given-Before-New と End Weight という原則に支配されているためである [34].

2. コーパス作成

2.1 テキストデザイン

コーパスのベースとなるテキストについては，Flickr8k [35] を用いた．Flickr8k は，人間や動物の動きに関する 8000 の画像と，画像につき 5 つのキャプションがついた計 40,000 キャプションを含むコーパスである．ベーステキストとして適切なものを選ぶため，下記の条件に当てはまるキャプションを削除した: ピリオド以外の句読点を含む; 文法誤り修正モデル GECToR [36] で修正された; 名詞句; 他のキャプションと同一; 同一の画像に対して複数のキャプションが残っていた場合は，重複を削除した．また，単純化のため，本研究では 1 文につき IP が 1 つ，すなわちフォーカスを置ける箇所が 1 単語のみと設定した．1 文の長さの設定については，50 万語からなる London-Lund コーパス [37] を用いて，IP あたりの単語数を計算し，結果の第 3 四分位に相当する 6 単語を最大単語数として選択した．フィルタリングの結果 1375 キャプションが残り，そのうち最初の 196 キャプションを本コーパスのベースのテキストとして採用した．例えば，*A biker enjoys a coffee* が含まれていた．コーパス全体のデザインを図 2に示す.

2.2 音 声 収 集

データ収集には，クラウドソーシングプラットフォームの Amazon Mechanical Turk (MTurk)[2]を利用した．収集者は HIT(Human Intelligence Task) と呼ばれるタスクを公開し，匿名の参加者（Worker）が HIT の完了とともに報酬を受け取る仕組みである.

2.2.1 録音アプリ

MTurk には音声を収集するインターフェイスが実装されていないため，web ベースの録音アプリを作成し，参加者の音声をインタラクティブに録音し，結果はバックエンドのサーバで保存できるようにした．基本的な機能として録音・停止・再生が可能で，追加の機能として Google speech-to-text API[3]を有効にし，入力音声が英単語として認識できない場合に'Speak Clearly' と即座にフィードバックを行えるようにした．さらに，発話ごとに最後の 2 秒間を環境音の録音時間とし，参加者にはこの間は無音でいるよう指示した.

2.2.2 録音プロセス

キャプションあたり 3 人の録音者に割り当て，フォーカスの位置が異なる発話の録音を試みた（図 2）．また，同一の話者からフォーカスの場所を指定しない通常の読み方での発話も収集した．録音 HIT は英国在住の Worker に依頼し，次のような指示を与えた: 表示された文を確認する; 下線が引かれた単語を強調して読み上げて録音する; その単語を強調するのが不自然であれば録音をスキップする; 下線が引かれていなければ，通常の読み方で録音する.

MTurk のようなクラウドソーシングでは，収集データの質を担保することが必須である [35], [38]．そのため，本収集プロセスでは，資格テストとして機能するタスク（3 キャプションの録音）を事前に用意し，人手による音声確認の後，指示通りに録音した Worker のみを本タスクへと誘導した．収集では Worker の同意のもと，年齢や性別，アクセント，学歴についての話者情報も収集した．3423 音声（通常: 588, フォーカス有: 2835）が 9 人の Worker によって録音された．Worker には，2 つのキャプションを処理する場合は$1.2，3 つの場合は$2.0 を支払った．平均時給は$30.9 となり，MTurk の Worker の間で公正とされる $15.0 [39] を上回った．表 1に，9 名の話者情報と収集音声全体に占める発話数の割合を示す.

表 1: 話者情報とコーパスにおける発話割合

性別	年齢（代）	アクセントの由来地域	発話数の割合（%）
男性	30	London	30.79
男性	30	Rochdale	26.41
女性	30	Essex	21.21
男性	60	South West	9.99
男性	20	South East Central Scotland	5.52
女性	40	North East Scotland	4.62
女性	40	Yorkshire	0.56
女性	20	Glasgow	0.56
女性	50	Liverpool	0.35

(注2)：https://www.mturk.com/
(注3)：https://cloud.google.com/speech-to-text

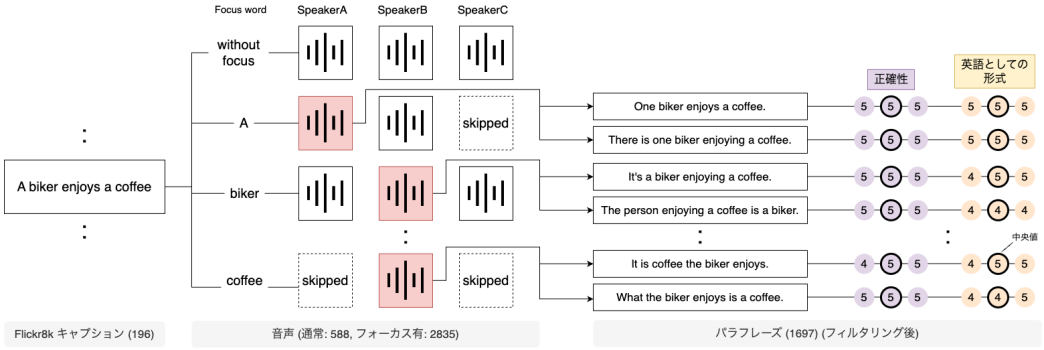

図 2: *A biker enjoys a coffee* を例としたコーパスデザイン

2.3 パラフレーズ収集

パラフレーズなど自由形式回答の収集では，参加者が母語話者であることを保証する必要がある [35]．英語母語話者選別のため，最初に英語圏（US, CA, UK, AU, NZ）在住の Worker に対して母語を回答するタスクを用意した．600 人が参加し，そのうち英語と回答した 550 人を英語母語話者とみなし，半数ずつ 2 つのグループ（G1, G2）に分けた．図3に，パラフレーズ収集時に Worker に与えた指示を示す． まずは，音声収集で録音されたフォーカス有りの音声の中からランダムに選択された 1 つの発話（図 2赤色）を聞いてもらい，その音声で表される含意を明確に伝える文に言い換えるよう指示した．さらに，文法や語彙の変更は促す一方で，大文字や感嘆符の使用，文脈を想像した文章の追加などは行わないようにという指示も追加した．G1 の参加者を対象として事前の資格テストとなる HIT を用意した．この HIT ではパラフレーズのタスクと，与えられた指示を要約する問題を提示し，音声収集時と同様，人手によって結果を確認した後，指示を正しく理解していると考えられる Worker を本タスクに誘導した．発話あたり 2 人の Worker がパラフレーズを行い，2130 パラフレーズを 16 人の Worker から収集した．パラフレーズあたりの報酬は$0.15 で，平均$16.5 であった．

2.4 パラフレーズ評価とフィルタリング

Worker によって作成されたパラフレーズの質は，今後，本コーパスを使った研究に影響する．パラフレーズの質を評価するため，主観評価を行なった．Worker にはフォーカスされた発話と，対応するパラフレーズのペアについて，次の 2 つの観点それぞれについて 5 段階（1–5）で評価するよう指示した: （X）パラフレーズは発話中の意味をどれだけ正確に伝えられているか; （Y）パラフレーズが英語としてどれくらい整っているか．パラフレーズ収集時と同様に，資格テストを今度は G2 の Worker に実施した．資格テストをクリアした Worker に対して，1 パラフレーズあたり 3 名の Worker を割り当てた（16 名参加）．報酬はパラフレーズあたり$0.1 とし，平均時給は$25.0 であった．評価結果の収集後，観点（X）と観点（Y）それぞれについて中央値を計算し（図 2），（X）での値が 3 以上かつ（Y）が 4 以上であったパラフレーズのみを残した．条件を満たさなかったパラフレーズは削除され，再度収集と評価，削除を行なった．さらなるフィルタリングとして，下記基準に該当するパラフレーズを削除した: 観点（X）のスコアの分散が比較的大きい（$\sigma^2 > 1.6$）（例：スコア：2，5，5）; 時制が現在から過去に変化（*The men are climbing. / The men were climbing on something.*）．フィルタリングの結果，1697 パラフレーズを得た．表 2にフォーカス音声とパラフレーズのペアの例を示す．

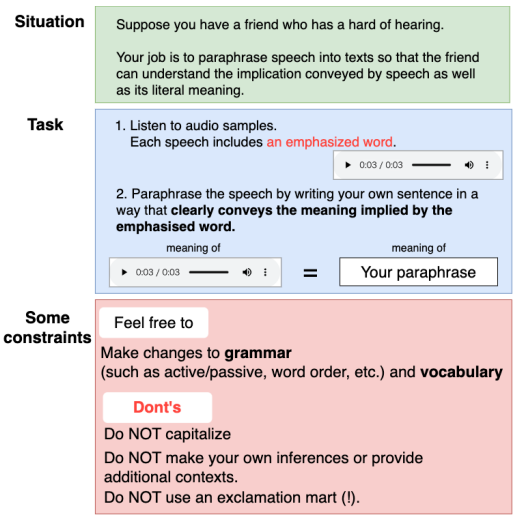

図 3: パラフレーズ収集における指示.

表 2: コーパス内のフォーカス発話とパラフレーズのペア例.

Focused speech	Paraphrase
A dog splashes in the water	One dog splashes in the water
A dog splashes in the water	A dog, not two, splashes in the water
A **dog** splashes in the water	The animal splashing in the water is a dog
A **dog** splashes in the water	A dog, not a child, splashes in the water
A dog **splashes** in the water	What a dog is doing in the water is splashing
A dog splashes **in** the water	A dog splashes in the water, not at the edge of it
A dog splashes **in** the water	A dog immersed in water splashes
A dog splashes in **the** water	A dog splashes in some specific water
A dog splashes in the **water**	Water is what a dog splashes in
A dog splashes in the **water**	A dog splashes in the water, not in the mud

3. コーパス分析

3.1 録音時におけるスキップされた focus 語の特性

全ての語にフォーカスを置いた場合，3264 の発話が可能であったが，429 （13.1%）の録音がスキップされた．例えば，*a boy runs through the grass* では，*a* で 1 人，*the* で 2 人，*grass* で 1 人がスキップした．全体を通して，この例のように冠詞，または文末の語がスキップされる傾向が観察されたため，フォーカス語の品詞及び文中での位置とスキップ率を調べた．結果を表

表 3: フォーカス語の品詞及び位置と録音スキップ率の関係

品詞	位置	スキップ率		
		mean	median	std
前置詞	0.40	0.00	0.00	0.00
	0.50	0.00	0.00	0.00
	0.60	0.00	0.00	0.04
	0.75	0.00	0.00	0.00
	0.80	0.04	0.00	0.12
補助動詞	0.20	0.00	0.00	0.00
	0.40	0.00	0.00	0.00
	0.50	0.00	0.00	0.00
	0.67	0.00	0.00	0.00
冠詞	0.00	0.31	0.33	0.12
	0.60	0.35	0.33	0.17
	0.75	0.38	0.33	0.21
	0.80	0.46	0.33	0.19
数詞	0.00	0.01	0.00	0.05

品詞	位置	スキップ率		
		mean	median	std
名詞	0.00	0.00	0.00	0.00
	0.20	0.00	0.00	0.00
	0.25	0.00	0.00	0.00
	0.33	0.03	0.00	0.10
	0.40	0.00	0.00	0.00
	0.50	0.00	0.00	0.00
	0.60	0.00	0.00	0.00
	0.67	0.00	0.00	0.00
	0.75	0.00	0.00	0.00
	0.80	0.00	0.00	0.00
	1.00	0.27	0.33	0.31
代名詞	0.80	0.00	0.00	0.00
	1.00	0.33	0.33	0.33
副詞	1.00	0.29	0.33	0.26

品詞	位置	スキップ率		
		mean	median	std
形容詞	0.00	0.00	0.00	0.00
	0.20	0.00	0.00	0.00
	0.25	0.00	0.00	0.00
	0.75	0.00	0.00	0.00
	0.80	0.00	0.00	0.00
動詞	0.20	0.00	0.00	0.00
	0.25	0.00	0.00	0.00
	0.40	0.00	0.00	0.03
	0.50	0.00	0.00	0.00
	0.60	0.00	0.00	0.00
	0.67	0.00	0.00	0.00
	0.75	0.00	0.00	0.00
	0.80	0.00	0.00	0.00
	1.00	0.29	0.33	0.34

3に示す．品詞タグ付けは Spacy[4]を用い，文中での位置は文頭を 0.00，文末を 1.00 として正規化した値とした．例えば，5 単語の文においてフォーカス語が 3 番目であれば，この語の値は 0.50 となる．また，該当の品詞及び位置における出現回数が 2 回以下の場合は表から除いた．結果から，フォーカス語が冠詞である場合，及び文末にあった場合（名詞・代名詞・副詞・動詞）に，そうでない場合に比べてスキップ，すなわちフォーカスを置くことが不自然になると判断されやすかったことが分かる．

3.2 パラフレーズ方法分析

フォーカスが音声から言語情報領域へどのように変換されたのかを調べるため，発話–パラフレーズペアについて人手で分析を行い，以下のように，語彙的，文法的観点から変換パターンを分類した（コーパス内の全ての変換を網羅するものではない）．

- 語彙的変換
  - **置換**: フォーカス語を類義語で置き換え, e.g. （***dog*** / *canine*），（***on*** / *on top of*）．
  - **修飾**: フォーカス語やそのフレーズを副詞や節で修飾, e.g. （***play*** / *play actively*），（***is*** / *is indeed*）．
  - **否定**: フォーカス語の代替語を明示し，否定 （*A* ***man*** / *A man, not a woman*）．
- 文法的変換
  - **左方向シフト**: フォーカス語を文頭方向へ移動; 強調構文，疑似強調構文, 倒置などを利用． e.g. （***People*** *sit .. / It's people who sit ..*），（*.. play* ***baseball*** / *baseball is what .. play*）．
  - **右方向シフト**: フォーカス語を文末方向へ移動． 疑似強調構文，倒置，その他を使用． e.g. （*Children* ***play*** *.. / what children do .. is play*），（ *.. is* ***rock*** *climbing* / *.. is climbing a rock*），（*an* ***old*** *man sits* / *a man sits and he is an old man*）
  - **アスペクト変化**: 現在形から現在進行形，またその逆へのアスペクト変化, e.g. （***is*** *walking* / *walks*）．

さらに，フォーカスが当たった品詞ごとのパラフレーズ方法の違いの分析を行なった．[26] を参考に，まずフォーカスを受けた回数が 50 回以上であった品詞について，ランダムに 50 パラフレーズ選んだ．次に，パラフレーズごとに，上記分類の手法が起こった回数を，フォーカスのある語あるいはその語を含む句のみを対象に数えた．例えば，次のパラフレーズについて考える．（*a dog trots through* ***the*** *grass* / *the only grassy area is what a dog trots through*）．この場合，フォーカスのある *the* だけでなく，それを含む名詞句（下線部）を分析対象とする; 語彙的な変化では，*only* が挿入され（修飾），*grass* が *grassy area* に置き換えられた（置換）; 文法的な変化では，reversed-pseudo-cleft と呼ばれる強調構文の一種が用いられた（左側シフト）．表 4は，フォーカス語の品詞ごとに変換方法の平均発生頻度を表したものである．アスペクト変化については，フォーカス語が動詞または補助動詞の場合のみに数えた．

表 4: フォーカス語の品詞別パラフレーズ方法の平均出現頻度

		名詞	動詞	形容詞	数詞	補助動詞	前置詞	冠詞
語彙的	置換	0.28	0.12	0.10	0.18	0.08	0.42	0.72
	修飾	0.00	0.02	0.12	0.06	0.60	0.18	0.50
	否定	0.10	0.06	0.06	0.12	0.10	0.12	0.00
文法的	左側シフト	0.10	0.20	0.04	0.08	0.00	0.10	0.00
	右側シフト	0.46	0.44	0.70	0.52	0.08	0.26	0.02
	アスペクト変化	-	0.28	-	-	0.20	-	-

4. 考察と結論

本研究では，ST への適用を念頭に，音声のフォーカスの位置が異なる発話と，フォーカスによってパラ言語的に表された含意も含めて言い換えたテキストのペアが含まれる英語コーパスの作成を目指した．コーパスの分析では，冠詞及び文末の語にはフォーカスを置くことが不自然とされやすいこと; パラ言語領域から言語領域へのフォーカス情報の変換には，様々な語彙的・文法的手法が用いられ，これらの手法への依存度はフォーカス語の品詞によって異なることを明らかにした．本研究の限界の１つとして文脈，すなわち「なぜその単語をフォーカスするのか？」について考慮がないことが挙げられる．通常のパラフレーズタスクでは，パラフレーズとその評価は文脈によって変化することが報告されている [40]．文脈の重要性は，今回のパラフレーズ収集と評価にも当てはまると考えられ，1.1.1 に挙げた broad focus と narrow focus の区別も含め，今後コーパスを拡張する上で文脈を考慮したデータ収集について検討してい

（注4）：https://spacy.io/

る．文脈についての限界があるものの，本研究はパラ言語翻訳のさらなる進展に向けて新たな方向性を示した．このコーパスとその分析から得られた知見は，パラ言語情報を保持する ST モデルの構築につながるものである．

5. 謝　　辞

本研究の一部は，JSPS 科研費（JP21H05054）の助成を受けた．

文　　献

[1] V. Mitra, S. Booker, E. Marchi, D.S. Farrar, U.D. Peitz, B. Cheng, E. Teves, A. Mehta, and D. Naik, "Leveraging acoustic cues and paralinguistic embeddings to detect expression from voice," arXiv preprint arXiv:1907.00112, pp.••–••, 2019.

[2] M. Rooth, "A theory of focus interpretation," Natural Language Semantics, vol.1, no.1, pp.75–116, 1992.

[3] M. Krifka, "Basic notions of information structure," Acta Linguistica Hungarica, vol.55, no.3-4, pp.243–276, 2008.

[4] J.C. Wells, English intonation: An introduction, Cambridge University Press, 2006.

[5] T.M. Derwing, R.I. Thomson, J.A. Foote, and M.J. Munro, "A longitudinal study of listening perception in adult learners of English: Implications for teachers," Canadian Modern Language Review, vol.68, no.3, pp.247–266, 2012.

[6] Y. Jia, R.J. Weiss, F. Biadsy, W. Macherey, M. Johnson, Z. Chen, and Y. Wu, "Direct speech-to-speech translation with a sequence-to-sequence model," arXiv preprint arXiv:1904.06037, pp.••–••, 2019.

[7] P. Aguero, J. Adell, and A. Bonafonte, "Prosody generation for speech-to-speech translation," Proc. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), vol.1IEEE, pp.557–560 2006.

[8] M. Akagi, X. Han, R. Elbarougy, Y. Hamada, and J. Li, "Emotional speech recognition and synthesis in multiple languages toward affective speech-to-speech translation system," Proc. 10th International Conference on Intelligent Information Hiding and Multimedia Signal ProcessingIEEE, pp.574–577 2014.

[9] T. Kano, S. Sakti, S. Takamichi, G. Neubig, T. Toda, and S. Nakamura, "A method for translation of paralinguistic information," Proc. International Workshop on Spoken Language Translation (IWSLT), pp.158–163, 2012.

[10] T. Kano, S. Takamichi, S. Sakti, G. Neubig, T. Toda, and S. Nakamura, "Generalizing continuous-space translation of paralinguistic information.," Proc. INTERSPEECH, vol.445, pp.25–29, 2013.

[11] G.K. Anumanchipalli, L.C. Oliveira, and A.W. Black, "Intent transfer in speech-to-speech machine translation," Proc. IEEE Spoken Language Technology Workshop (SLT), pp.153–158, 2012.

[12] A. Tsiartas, P.G. Georgiou, and S.S. Narayanan, "Toward transfer of acoustic cues of emphasis across languages.," Proc. INTERSPEECH, pp.3483–3486, 2013.

[13] Q.T. Do, S. Sakti, G. Neubig, and S. Nakamura, "Transferring emphasis in speech translation using hard-attentional neural network models.," Proc. INTERSPEECH, pp.2533–2537, 2016.

[14] Q.T. Do, T. Toda, G. Neubig, S. Sakti, and S. Nakamura, "Preserving word-level emphasis in speech-to-speech translation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.25, no.3, pp.544–556, 2016.

[15] Q.T. Do, S. Sakti, and S. Nakamura, "Sequence-to-sequence models for emphasis speech translation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.26, no.10, pp.1873–1883, 2018.

[16] A. Cruttenden, Intonation, Cambridge University Press, 1997.

[17] L.J. Downing, "The prosody and syntax of focus in chitumbuka," ZAS Papers in Linguistics, vol.43, pp.55–79, 2006.

[18] K. Hartmann, "Focus and tone," Acta Linguistica Hungarica, vol.55, no.3-4, pp.415–426, 2008.

[19] Q.T. Do, S. Sakti, and S. Nakamura, "Toward multi-features emphasis speech translation: Assessment of human emphasis production and perception with speech and text clues," Proc. 2018 IEEE Spoken Language Technology Workshop (SLT), pp.700–706, 2018.

[20] H. Tokuyama, S. Sakti, K. Sudoh, and S. Nakamura, "Transcribing paralinguistic acoustic cues to target language text in transformer-based speech-to-text translation," Proc. Interspeech 2021, pp.2262–2266, 2021.

[21] S. Latif, I. Kim, I. Calapodescu, and L. Besacier, "Controlling prosody in end-to-end TTS: A case study on contrastive focus generation," Proc. 25th Conference on Computational Natural Language Learning (CoNLL), pp.544–551, 2021.

[22] S. Greenbaum, A student's grammar of the English language, Pearson Education India, 1990.

[23] A. Nenkova, J. Brenier, A. Kothari, S. Calhoun, L. Whitton, D. Beaver, and D. Jurafsky, "To memorize or to predict: Prominence labeling in conversational speech," Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational LinguisticsCiteseer, pp.9–16 2007.

[24] A. Talman, A. Suni, H. Çelikkanat, S. Kakouros, J. Tiedemann, and M. Vainio, "Predicting prosodic prominence from text with pre-trained contextualized word representations," NODALIDA, pp.••–••, 2019.

[25] C. Boonthum, "iSTART: Paraphrase recognition," Proc. ACL Student Research Workshop, pp.31–36, 2004.

[26] W.B. Dolan, C. Quirk, and C. Brockett, "Unsupervised construction of large paraphrase corpora: Exploiting massively parallel news sources," COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics, pp.350–356, 2004.

[27] M. Vila, M.A. Martí, H. Rodríguez, et al., "Is this a paraphrase? What kind? Paraphrase boundaries and typology," Open Journal of Modern Linguistics, vol.4, no.01, p.205, 2014.

[28] J. Taglicht, Message and emphasis: On focus and scope in English, no.15, Addison-Wesley Longman Limited, 1984.

[29] M.E. Rooth, "Association with focus," PhD thesis, University of Massachusetts Amherst, 1985.

[30] N. Suzuki and S. Nakamura, "Representing 'how you say' with 'what you say': English corpus of focused speech and text reflecting corresponding implications," Proc. Interspeech 2022, pp.4980–4984, 2022.

[31] E. König, The meaning of focus particles: A comparative perspective, Routledge, 2002.

[32] C.J. Fillmore, P. Kay, and M.C. O'connor, "Regularity and idiomaticity in grammatical constructions: The case of let alone," Language, pp.501–538, 1988.

[33] E. König and V. Gast, "Focused assertion of identity: A typology of intensifiers," Linguistic Typology, vol.10, pp.223–76, 2006.

[34] B. Aarts, Oxford modern English grammar, Oxford University Press, 2011.

[35] C. Rashtchian, P. Young, M. Hodosh, and J. Hockenmaier, "Collecting image annotations using amazon's mechanical turk," Proc. NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, pp.139–147, 2010.

[36] K. Omelianchuk, V. Atrasevych, A. Chernodub, and O. Skurzhanskyi, "GECToR–grammatical error correction: Tag, not rewrite," 15th Workshop on Innovative Use of NLP for Building Educational Applications, pp.163–170, 2020.

[37] J. Svartvik, The London–Lund corpus of spoken English: Description and research, vol.82, Lund University Press, 1990.

[38] R. Kennedy, S. Clifford, T. Burleigh, P.D. Waggoner, R. Jewell, and N.J. Winter, "The shape of and solutions to the MTurk quality crisis," Political Science Research and Methods, vol.8, no.4, pp.614–629, 2020.

[39] M.E. Whiting, G. Hugh, and M.S. Bernstein, "Fair work: Crowd work minimum wage with one line of code," Proc. AAAI Conference on Human Computation and Crowdsourcing, vol.7, pp.197–206, 2019.

[40] R. Barzilay and K. McKeown, "Extracting paraphrases from a parallel corpus," Proc. 39th annual meeting of the Association for Computational Linguistics, pp.50–57, 2001.