

質問コーパスを利用した疑問詞の推定とその効果 Estimation of Interrogatives Using a Question Corpus and its Utilization

千野 愛実花¹⁾ 寺岡 丈博²⁾
Amika Chino Takehiro Teraoka

1 背景と目的

近年発展してきた大規模言語モデルの応用により、対話システムへの関心が高まっている。対話システムでは、タスク指向型と非タスク指向型があり、前者は実用化されているが、後者に関しては課題が多く残っている。これまでに、ペルソナと呼ばれるプロフィール情報をシステムに与え、整合性のある応答を目指すことを目的とした研究 [1] や、対話の破綻を回避することでユーザとシステムが対話を継続させることを目的とした研究 [2] がされてきている。しかし、ユーザとシステムとのコミュニケーションという点においては、十分ではなく、課題が残っている。対話や会話に関するコミュニケーション研究においては、初対面での話題導入には質問が大半を占めている [3] ことや、対話中にいくつか質問と返答を繰り返すことで話題を確立することなどが明らかにされているが、現在の対話システム研究では、それらの知見が十分に活かされていない。そのため、人間同士の対話で見られる質問のやりとりや相槌、発話や応答の間、さらにはジェスチャーなどの活用が今後は必要になると考えられている。

本研究では、システムとユーザの対話においても質問のやりとりは重要であるという考えのもと、システム側からの質問生成とその連鎖について必要な質問コーパスを構築する。その応用として、質問の生成とその連鎖を行い (図 1)、推定した疑問詞を用いて質問を再生成することで精度の向上を図る。

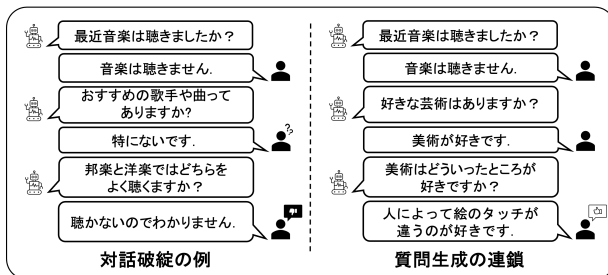


図 1: 質問生成の連鎖

2 関連研究

これまで、対話システムはユーザとより自然な対話を目指すために、様々なアプローチから研究されてきた。ここでは、システム発話の一貫性、質問生成、質問応答、さらに初対面における対話分析の研究について取り上げる。

発話の一貫性に関する研究では、Gao らの研究が挙げられる。Gao らの研究では、ペルソナを統合してナラティブの一貫性を向上させることは重要という考えの

- 1) 拓殖大学大学院工学研究科 Graduate School of Engineering, Takushoku University
- 2) 拓殖大学工学部情報工学科 Department of Computer Science, Faculty of Engineering, Takushoku University

と、ペルソナと知識の相互関係を含んだ体系的な知識フレームを作成した。これにより、人工知能によるペルソナ推論を可能にし、従来の対話システムより、一貫性のある魅力的な応答の生成が可能となった [4]。この研究は、対話システムの発話に一貫性を保たせるためのものであり、本研究とは目的が異なる。

質問生成の研究では、片山らの研究と、Horiuchi らの研究が挙げられる。片山らの研究では、相手の発話に対して正しく深掘りした質問生成を試みている。この研究では、学習データに Twitter コーパスのみを利用している [5]。しかし、一問一答形式のため対話の流れがわかりにくいという課題がある。一方、Horiuchi らは、大規模コーパスで事前学習し、“What did you do?” というような特定の質問を行うための発話生成モデルの構築を試みた [6]。この研究の課題として、日本語データへの拡張と人間が特定の質問を行う際の実際の誘導方法の調査を挙げている。また、Chino らは、頻出する質問文からテンプレートを作成し、質問を生成しているが、精度に大きな課題が残っている [7]。

質問応答の研究では、Hwang らの研究や Zhu らの研究が挙げられる。Hwang らの研究では、高品質な会話型質問応答のデータを自動で生成するための新たなフレームワークを提案している。この研究では、文章から質問に適したフレーズを抽出し、過去の発話を考慮して対応する質問を生成した [8]。さらに、質問生成後に抽出された回答を修正し、回答がペアとなる質問と正確に一致させた。回答修正アプローチが、データの質の大幅な向上に繋がった。Zhu らの研究では、質問応答タスクにおける質問の連鎖を行っている。他段階の質問を生成するフレームワークを提案している [9]。

初対面における対話分析の研究では、お互いの情報を共有するために質問と応答を繰り返し、共通の興味・関心を引き出すことで対話の継続が可能となる [10]。

現在の対話システムでは、ユーザ側からの質問に正しく応答できるかについて焦点が当てられており、システム側からの質問に関しては十分とはいえない。本研究では、人間同士の対話と同様に、システムとユーザの対話においても質問のやりとりは重要であると考え、システム側からの質問生成とその連鎖に必要な質問コーパスを構築する。

3 質問コーパスの構築と拡張

3.1 使用したデータ

質問コーパスの構築には、大阪大学マルチモーダル対話コーパス Hazumi [11]、日本語日常対話コーパス [12]、JPersonaChat [13] 及び RealPersonaChat [14] の 4 つのコーパスを利用する。Hazumi は、人とエージェントとの雑談対話が収録されているコーパスである。実験参加者がいくつかの話題に対し対話する様子が収録されている。収録時間は、1 対話あたり 15 分程度である。JPersonaChat は、PersonaChat に基づいて収録されている

Question1	i_1	Reply	Related	Question	i_j	i_2	Label	corpus
おすすめの曲は何ですか？	What	えーと色々あるのですが、SHAMROCKとかは、一度は聞いたことあるかな？	['音楽', '聞いて']	いつも音楽は何で聞いてますか？	['何']	Why	5	RPC
こんな時、占いたという結果になりますか？	What	外国生まれなのでね。彫りが深い顔立ちには気にしない方がいいです。	['星座']	星座などはどうでしょうか？	['どう']	What	1	JPC
通話は何でされてますか？	What	電車です	['全部', '通勤時間', 'かかり']	全部で通勤時間はどれくらいかかりますか？	['どれくらい']	How	0	RPC
曲は何ですか？	What	私には「大きな栗の木の下で」を演奏します。	['何', '音', '担当', 'する']	何の音を担当するのですか？	['何']	What	1	JDD
どんな話ですか？	What	原作が伊藤計劃さんので、SFで、スチームパンク！	['作品']	いつ頃の作品ですか？	['いつ']	When	2	RPC
普段の時間って何で使ってます？	What	最近はマンガ読んでますね	['ジャンル']	どんなジャンルを？	['どんな']	What	1	RPC
何回目の結婚記念日ですか？	What	覚えていないので、数えてみないと分かりません。	['結婚', 'し', '時', 'お母さん', '歳']	結婚した時、お母さんは何歳でしたか？	['何']	When	2	JDD
何という芝居ですか？	What	『悪魔をめぐらせ』というタイトルのブラック・コメディです。	['芝居']	どのような芝居ですか？	['どのような']	What	1	JDD
例えば何が面白いですか？	What	たくさんありますよ、観るの楽しいです、！！	['愛の言葉']	愛の言葉はどうですか？	['どう']	How	0	RPC
飛騨高山には何があるのですか？	What	飛騨高山は古い城下町なので、町並みがきれいですよ。私も町並みを	['どれ', '城下町']	どれくらい古い城下町なのですか？	['どれくらい']	How	0	JDD
何が面白いですか？	What	肉ではサンマとサバとサケとイワシがあります。肉では鶏のもも肉と	['ワイン', '一緒']	ワインと一緒にどうですか？	['どう']	How	0	JDD
何がありますか？	What	ビーフシチューとかカレー、それにバスタが何種類もあります。	['よう', '種類', 'あり']	どのような種類がありますか？	['どのような']	What	1	JDD
これは何ですか？	What	スマートフォンに付けるレンズです。これをつけることで撮影が	['これ', '何', '撮る']	これで何を撮るのですか？	['何']	What	1	JDD

図 2: 質問コーパスの抜粋

コーパスであり、実験参加者には、対話の基盤となるペルソナ情報を与えている。与えられたペルソナに沿って 2 者間で対話した内容を収集したコーパスである。平均発話数は 15 発話程度である。一方、RealPersonaChat は、事前に話者のペルソナと性格特性を収集し、それらの情報を含む雑談対話が収録されている。初対面同士の対話であり、1 対話あたり 30 発話程度収録されている。話題は決められていない。日本語日常対話コーパスは、定められた 5 つのトピックに対し、人間同士の雑談対話が収録されている。1 対話は 4 発話以上からなり、平均 8 発話収録されている。

3.2 構築の手順

3.1 節で記載した 4 つのコーパスを利用し、質問コーパスを構築する。

はじめに、既存の 4 つのコーパス質問文を抽出する。質問コーパスには、質問文とそれまでの対話履歴を発話ターンの長さごとに取得する。コーパスに含まれる全ての発話から文末にクエスチョンマークのあるものを質問文として取得する。ここでは、オープンクエスチョンのみコーパスに含めるため、抽出した質問文をオープンクエスチョンとクローズドクエスチョンに分類する。オープンクエスチョンとは、回答範囲を制限しない質問のことで、クローズドクエスチョンとは、「はい」、「いいえ」のように選択肢から回答する質問である。本研究では、話題を継続させるのではなく、発展させることを目的としているため、分類した質問文からオープンクエスチョンのみを抽出し、コーパスを構築する。これらの分類には、Meta 社が公開している多言語翻訳モデル [15] を用いて翻訳を行う。翻訳結果から 5W1H が含まれるものをオープンクエスチョンとし、コーパスにまとめる。その際、質問と対話履歴だけでなく、疑問詞と質問に含まれる名詞と動詞についてもまとめる。このとき、形態素解析器には MeCab[16]、辞書には NEologd[17] を用いた。なお、「なんでですか？」や「いつですか？」といった動詞・名詞のない発話については、先行研究 [5] の課題として挙げられていたように、疑問詞のみを生成してしまふことで、単調な発話が生成される可能性があるためここでは除外する。本研究で構築した質問コーパスの規模を表 1 に示す。

表 1: 質問コーパスの規模

	2 turns	4 turns	6 turns	連鎖
Hazumi	114	108	102	19
JapaneseDailyDialogue	1,455	752	282	270
JPersonaChat	1,179	953	639	109
RealPersonaChat	9,664	8,160	7,143	917
Total	12,412	9,973	8,166	1,315

また、本研究では、質問のやりとりに着目しているため、1 つ目の質問だけでなく、対話として質問が連鎖するデータも質問コーパスとしてまとめる。連鎖するデータは、1 つ目の質問のデータと同様、オープンクエスチョンのみを収集する。また、コーパスに含める情報も 1 つ目の生成と同様である。ただし、質問が What あるいは How の疑問詞となった場合、返答を考慮した疑問詞の同定 [18] を行い、疑問詞が What あるいは How かどうかを明確にする。表 2 に連鎖となる質問コーパスの規模を示す。

表 2: 質問の連鎖

質問 \ 連鎖	What	How	Why	When	Where	Who	Total
What	567	165	30	25	36	17	840
How	136	89	7	14	17	1	264
Why	12	3	2	3	2	2	24
When	33	10	1	4	8	0	56
Where	57	31	5	5	22	0	120
Who	7	2	0	0	0	2	11
Total	812	300	45	51	85	22	1,315

表 2 より、全ての疑問詞に対し、連鎖となる疑問詞は What が最も多くなることがわかる。

また、疑問詞の同定を行った際に、What から How に変化する質問文より、How から What に変化する質問文の方が多かった。How から What に変化した例を以下に示す。

例: “How” から “What” に変化

質問文 : ゲーム会社はどんな感じですか？
 返答 (How) : やりがいがあって楽しいです。
 返答 (What) : 何時間でも働けます。

4 質問コーパスの応用

4.1 生成と連鎖

はじめに、対話履歴から対話の話題となる語 (話題語) と話題から関連する語 (関連語) を取得する。話題語の取得には、TF-IDF 法を用いる。ここでは、「名詞」のみ TF-IDF 値を求める。発話が増えるごとに逐次的に TF-IDF 値を計算し、直前の TF-IDF 値から値が大きくなった単語と新しく取得した単語を話題語とする。

上記の方法で取得した話題語から関連語を取得する。関連語の取得には、TF-IDF 値と連想概念辞書 [19] を用いる。この連想概念辞書には、小学校の国語の教科書で用いられる名詞を刺激語とした連想実験により、得られた連想語と刺激語に対する連想語の連想されやすさを表した連想距離がまとめられている。この連想語は、刺激語に対して上位語となる「上位概念」や下位語となる「下位概念」などの 8 種類の連想課題に関して実際に人

表 3: 質問の連鎖の結果. 括弧内の数字は正解率を示す. アスタリスク (*) は, t 検定 ($p < 0.05$) によりベースラインと提案手法との間に有意な差があることを示す.

	Input	Avg. scores	Positive	Neutral	Negative	Avg. scores	Positive	Neutral	Negative	Recovery
ベースライン	2 turns	3.32	0.50 (25/50)	0.16	0.34	3.56	0.54 (21/39)	0.21	0.26	0.36 (4/11)
	4 turns	2.92	0.42 (21/50)	0.12	0.46	3.00	0.42 (16/38)	0.16	0.42	0.42 (5/12)
提案手法	2 turns	3.48	0.56 (28/50)	0.12	0.32	3.41	0.54 (21/39)	0.13	0.33	0.64 (7/11)
	4 turns	3.80*	0.64 (32/50)	0.12	0.24	4.00*	0.71 (27/38)	0.08	0.21	0.42 (5/12)

間が連想した単語である. 質問を行う直前の発話に否定の言葉がある場合, 話題語の上位語となる語を取得し, 対話中の関連語とする. 上位語の取得方法は, 話題語が連想概念辞書の上位概念の刺激語としてある場合, 連想語を取得し, 下位概念の連想語としてある場合は, 刺激語を取得する. どちらにも含まれる場合には, 刺激語から得られた上位語を使用する. 一方, 否定がない場合には, 対話コーパスから, 話題語が含まれる次の発話に含まれる語に対する TF-IDF 値を用いる. このとき最も値が高い語を本研究の関連語とする.

3 章で作成した質問コーパスから, 2,442 セット [20] に対して, 日本語学習済み T5[21] を用いてファインチューニングを行う. なお, 1 セットは, 対話履歴, 1 つの対話に対する質問に含まれる語及び質問としている. この際, 出力に取得した関連語が必ず含まれるように出力を補正する. ファインチューニングしたモデルに, 対話履歴と関連語を入力することで, 関連語を含む質問が出力される.

ユーザからの返答を受け連鎖となる質問の生成を行う. 連鎖の生成についても日本語学習済み T5 を用いて, ファインチューニングを行う. ここでの, 学習データは, 3 章で作成した質問コーパスから, 質問の直前の発話, 質問に含まれる名詞, 質問の 2,442 セットを用いる. 対話履歴と話題語を用いて, 連鎖となる質問を生成する. ここでは, ユーザの返答から話題が離れてしまうのを防ぐため話題語を用いる.

4.2 質問生成と連鎖の結果

コーパス内の対話から 2 発話を抽出し 50 対話を実験の入力とする. ここでは, 実験で使用する対話に対して, 関連語, 生成した質問 (質問 1) 及び連鎖となる質問 (質問 2) について評価を行う.

それぞれの対話に対する関連語を 5 段階で評価する. 5 段階評価の項目は, 「5: 適切である», 「4: やや適切である», 「3: どちらともいえない», 「2: あまり適切でない», 「1: 適切でない」とする.

質問 1 と質問 2 については, ベースラインと比較し, 評価を行う. ベースラインは, 日本語学習済み T5 を対話と質問のペアでファインチューニングしたモデルである. 50 対話において, ベースラインと提案手法でそれぞれ生成した質問 (質問 1) と, 連鎖となる質問 (質問 2) に対して, 5 段階評価のアンケートを行う. 質問 1 と質問 2 は, 「5: 文脈に沿っている», 「4: やや文脈に沿っている», 「3: どちらともいえない», 「2: あまり文脈に沿っていない», 「1: 文脈に沿っていない」の 5 段階とする. 質問 1 に対する返答に関しては, アンケートで 1 つの生成文に対し, 5 人に回答してもらい, 5 人が回答した中で最も使われた単語を含む発話で, トークン数の一番多い発話を本研究の実験に採用する.

評価を行った全ての項目に対し, 5 人中 3 人の回答が

4 以上のものを “Positive”, 3 のものを “Neutral”, 2 以下のものを “Negative” とする.

質問の生成に対する 5 段階評価の結果を表 4 に示す. ベースラインの評価では, Positive と判断されたものが, 約 6 割であったが, 提案手法は 7 割を超えており, 良い評価を得ることができた. このことから, 関連語を用いて質問を生成することに一定の効果があることを確認した.

表 4: 質問生成の結果. プラス記号 (+) は, 符号検定 ($p < 0.05$) によって, ベースラインと提案手法に統計的有意差があることを示す.

	Input	Avg. scores	Positive	Neutral	Negative
ベースライン	2 turns	3.68	0.62 (31/50)	0.10	0.28
	4 turns	3.84	0.66 (33/50)	0.10	0.24
提案手法	2 turns	4.20+	0.78 (39/50)	0.04	0.18
	4 turns	4.12	0.76 (38/50)	0.04	0.20

連鎖となる質問の結果を表 3 に示す. 表 3 の左側は 50 対話全体の評価結果で, 右側は連鎖となる質問 2 を行う前の発話で対話破綻が起きているものを除いた結果である. “Recovery” は, 質問 2 を生成することで, 対話破綻から復帰した対話の割合である. 全体の結果は 5 割程度であるが, ベースラインを上回ることができている. 破綻していない対話に対する結果は, 全体の結果とほとんど変わらない.

一方で, Recovery については, ベースラインが 4 割弱であり, 提案手法は 6 割を超えている. これらの結果から, 話題語を複数用いる提案手法は対話の破綻から回復できることを示す.

4.3 連鎖のエラー分析

連鎖の評価が良くなかった対話について分析を行った. 分析結果を表 6 と表 7 に示す. これらの表から, 同じ疑問詞が続いている場合に, 連鎖の生成がうまくいっていないと考えられる. 表 2 から What に関しては, What から What が約 4 割と最も多いためここでは考慮しない. そのため, What 以外の疑問詞について分析すると, 2 ターンでは約 6 割 (10/17) が 1 つ目の生成と同じ疑問詞で連鎖の質問を生成することで, うまく質問を連鎖させることができていない. これは, 同じ疑問詞を繰り返すことにより, 1 つ目の質問と意図が同じである質問が生成されていることが要因である.

4.4 再生成

4.3 節におけるエラー分析から質問の連鎖がうまくいっていない場合に, 同じ疑問詞が連続して生成されることが散見された. 連鎖の精度を向上させるために, 疑問詞が連続している場合, 再生成を行う. 表 2 からどの疑問詞でも 2 つ目の疑問詞は What となることが最も多い. そのため, 同じ疑問詞が生成された場合, What にあたる疑問詞を入力に使用する. ここでは, 3.1 節で構

表 5: 生成例. A と B は発話者を示し, (A) はシステムでの出力結果で, (B) はユーザの返答である.

ID	提案手法	再生成
1	対話履歴 A	今日はここまで何で来られましたか?
	対話履歴 B	ここまではあの電車を乗り継いで来ました.
	質問 1 (A)	どうして電車に乗ったのですか?
	返答 B	電車が一番速いからです.
	質問 2 (A)	なぜ電車に乗ったのですか? 何の電車に乗ったのですか?
2	対話履歴 A	最近芸術作品に触れる機会がありましたか?
	対話履歴 B	最近はありませんけれども, 昔はよく映画とか美術館に行っていました.
	質問 1 (A)	好きな作品などはありますか?
	返答 B	モネの作品が好きです. 一気に世界観に引き込まれる感じが素敵です.
	質問 2 (A)	好きな作品などはありますか? モネの世界観って何ですか?
3	対話履歴 A	海外旅行に興味はありますか?
	対話履歴 B	海外はイタリアに行きました. あと韓国とこの間台湾に行ってきました.
	質問 1 (A)	台湾はどうでしたか?
	返答 B	日本から行くのがお手軽でいい旅行が出来ました.
	質問 2 (A)	手軽な旅行といえば日本はどうでしょうか? 手軽に旅行できる日本とはどんなところですか?

表 6: 連鎖の評価が悪かったもの (2 turns). 再生成の対象を太字で示す.

	What	How	Why	When	Where	Who	Closed	Total
What	5	2	1				1	9
How		3					1	4
Why			1					1
When								
Where					1			1
Who						1		1
Closed		2					4	6
Total	5	7	2		1	1	6	22

表 7: 連鎖の評価が悪かったもの (4 turns). 再生成の対象を太字で示す.

	What	How	Why	When	Where	Who	Closed	Total
What	4	1	1				2	8
How		1					2	3
Why			1					1
When								
Where								
Who								
Closed	3						3	6
Total	7	2	2				7	18

築したコーパスから, What にあたる疑問詞として最頻出の語である“何”を使用する. 再生成のモデルは, 4.1 節と同様, 日本語学習済み T5 を用いてファインチューニングを行う. ここでは, これまでと同様, 対話履歴と疑問詞を入力とする. それに加え, 出力される疑問詞を制御する必要があるため, 入力には疑問詞も用いる.

4.5 結果と考察

再生成を行った後の結果を表 8 に示す.

表 8: 再生成の結果

	連鎖	再生成
2 turns	0.56 (28/50)	0.68 (34/50)
4 turns	0.64 (32/50)	0.64 (32/50)

2 ターンでは, 50 対話中 10 対話が再生成に該当した.

そのうち, 6 対話に対する評価が Negative から Positive に変化した. 残りの 4 対話に関しては, Neutral あるいは Negative から変化はなかった. 全体としては, 再生成を行う前の 56%(28/50) から 68%(34/50) と 12% 精度が高くなった.

一方, 4 ターンでは, 50 対話中 5 対話が再生成に該当した. そのうち, Negative から Positive に変化した発話と, Positive から Negative に変化した発話がそれぞれ 1 つずつだったため, 全体としての結果に変化は見られなかった.

再生成した 2 ターンでの例を表 5 に示す. ID1 と ID2 は, Negative から Positive に変化した例であり, ID3 は Negative から変化しなかった例である. Positive に変化した例では, 質問の意図が全く同じであるものから, 疑問詞を変化させることで, 対話の内容がより深くなる質問となっている. そのため, 連鎖として成り立っている. 一方で, ID3 は, 疑問詞が変化したことにより, 質問の意図は変化しているが, “台湾”に関する質問から“日本”に関する質問に切り替わったことに変化がなかったため, 依然として適切でない質問であった.

5 まとめ

本研究では, 既存の対話コーパスから質問とそれまでの対話履歴を発話ターンの長さごとに取得し, 質問コーパスを構築した. それに基づいて質問の生成とその連鎖, さらに疑問詞を利用したエラー分析後の再生成まで行った. 再生成の結果から 2 ターンでは, 連鎖の精度を 1 割程度向上することができ, コーパスの効果を確認した.

今後の課題として, コーパスに含めている品詞を名詞のみに限定しているが, 共起される疑問詞に限られるため, 動詞も含めてコーパスに収録する. また, 本研究では, 補正する疑問詞を What (“何”) に限定していたが, それ以外にも What に当たる疑問詞は複数あるため, 細分化し, 生成に利用していきたい.

さらに, 現状では, 4 ターンまでの実験となっているため, ターン数を増やして質問コーパスによる効果を検証したい.

謝辞

本研究はJSPS 科研費 JP22K00646 の助成を受けたものです。また、国立情報学研究所の IDR データセット提供サービスにより大阪大学から提供を受けた「大阪大学 マルチモーダル対話コーパス Hazumi」を利用しました。

参考文献

- [1] Shuai Liu, Hyundong Cho, Marjorie Freedman, Xuezhe Ma, and Jonathan May. RECAP: Retrieval-enhanced context-aware prefix encoder for personalized dialogue response generation. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8404–8419, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [2] 坪倉和哉, 岡田真依, 入部百合絵, 北岡教英. 対話破綻修復コーパスの収集と分析 - ユーザの個人特性とシステムとの関係性を考慮した修復文生成に向けて -. 言語処理学会 第30回年次大会 発表論文集, pp. 1436–1440, 2024.
- [3] 方敏. 初対面以降の会話における話題選択及び変化に関する一考察. ことば, Vol. 41, pp. 89–105, 2020.
- [4] Silin Gao, Beatriz Borges, Soyoung Oh, Deniz Bayazit, Saya Kanno, Hiromi Wakaki, Yuki Mitsufuji, and Antoine Bosselut. Peacock: Persona commonsense knowledge for consistent and engaging narratives, 2023.
- [5] 片山太一, 大塚淳史, 光田航, 齋藤邦子, 富田準二. 相手の発話を深掘りするための質問生成技術. 人工知能学会全国大会論文集, Vol. JSAI2018, pp. 4G103–4G103, 2018.
- [6] Sota Horiuchi and Ryuichiro Higashinaka. Learning to ask specific questions naturally in chat-oriented dialogue systems. In *The 12th International Workshop on Spoken Dialog System Technology*, 2021.
- [7] Amika Chino and Takehiro Teraoka. Relevance-Aware Question Generation in Non-task-Oriented Dialogue Systems. In Gino Chen, Jessie Y. C. and Fragomeni, editor, *Virtual, Augmented and Mixed Reality (HCI International 2023)*, pp. 344–358, Cham, 2023. Springer Nature Switzerland.
- [8] Seonjeong Hwang and Gary Geunbae Lee. Conversational QA Dataset Generation with Answer Revision. In *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 1636–1644, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.
- [9] Wang Zhu, Jesse Thomason, and Robin Jia. Chain-of-questions training with latent answers for robust multistep question answering. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 8845–8860, Singapore, December 2023. Association for Computational Linguistics.
- [10] 鵬群趙. Topic initiation process in japanese conversation in japanese-chinese contact situations and native situations : An analysis of first-encounter and subsequent conversations. 神戸大学大学院国際文化学研究科, pp. 116–125, 2023.
- [11] 駒谷和範, 岡田将吾. 大阪大学 マルチモーダル対話コーパス hazumi. 国立情報学研究所情報学研究データリポジトリ. (データセット), 2022.
- [12] 赤間怜奈, 磯部順子, 鈴木潤, 乾健太郎. 日本語日常対話コーパスの構築. 言語処理学会 第29回年次大会 発表論文集, pp. 108–113, 2023.
- [13] Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. Empirical analysis of training strategies of transformer-based japanese chat systems. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pp. 685–691, 2023.
- [14] Sanae Yamashita, Koji Inoue, Ao Guo, Shota Mochizuki, Tatsuya Kawahara, and Ryuichiro Higashinaka. RealPersonaChat: A realistic persona chat corpus with interlocutors' own personalities. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation*, pp. 852–861, 2023.
- [15] Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. Beyond english-centric multilingual machine translation, 2020.
- [16] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In Dekang Lin and Dekai Wu, editors, *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230–237, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [17] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会 第23回年次大会 発表論文集, pp. 875–878, 2017.
- [18] 千野愛実花, 寺岡丈博. ユーザの回答に依存した疑問詞の同定とその効果. 人工知能学会全国大会論文集, Vol. JSAI2023, pp. 2O5OS2a04–2O5OS2a04, 2023.
- [19] 岡本潤, 石崎俊. 概念間距離の定式化と既存電子化辞書との比較. 自然言語処理, Vol. 8, No. 4, pp. 37–54, 2001.
- [20] 千野愛実花, 寺岡丈博. 対話継続を目指した質問生成の連鎖. 人工知能学会全国大会論文集, Vol. JSAI2024, pp. 3Xin2102–3Xin2102, 2024.
- [21] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-text Transformer. *J. Mach. Learn. Res.*, Vol. 21, No. 1, jan 2020.