

ChatGPT を活用した日本語コーパス文生成の基礎検討

石川 真大^{1,a)} 小田 恭央^{2,b)} 森勢 将雅^{1,c)}

概要: テキスト音声合成の研究では大量の音声データが必要であり、同じ話者が安定した声質で発話することが重要であることから、コーパス文の朗読が実施されている。ここで、例えば読みやすさ等を加味し、利用する音声合成タスクに適したコーパス文を柔軟に構築することができれば、合成音声の品質向上に寄与すると考えられる。本研究では、ChatGPT を利用することで条件に沿った文を作成し、朗読用のコーパスとして利用するための基礎検討を実施した。本稿では、主に文字数、含めるべきモーラ、読みやすさを条件としてコーパス文を作成できないか検討し、ChatGPT を利用する有効性について議論する。

1. はじめに

1.1 研究の背景と目的

統計的パラメトリック音声合成 [1] によるテキスト音声合成 (Text-to-Speech; TTS) は、数百文程度のクリーンな発話音声から、人と等価ではないものの聞き取ることに十分な明瞭度を持つ音声合成を実現していた。現在は Deep neural network (DNN) に基づく方法が主流であり、WaveNet [2] が提案された後に品質が大幅に向上した。2018 年に提案された Tacotron 2 [3] では、人間と等価な音質のテキスト音声合成を達成したものの、学習データ量が数十時間要求するようになってきている。このように、最新のテキスト音声合成では音声データの質に加え量が重要であり、音声認識のように多数の話者ではなく、1 名の話者が安定した声質で十分な量の発話をするのが望ましいとされている [4]。

日本語 TTS で用いられる音声データは、自由会話ではなく特定の文をまとめたコーパスを朗読することが一般的である。その際、統計的パラメトリック音声合成が主流の時代では、503 文から構成される ATR 音素バランス文 [5] が利用されていた。類似したコーパスでは、音素バランスが調整された 100 文からなる声優統計コーパス [6] が提案されている。利用するためのライセンスに着目し、パブリックドメインで利用できる音素バランスを勘案したコーパスでは、筆者らが構築した ITA コーパス [7] が存在する。日本語の TTS では、音素バランス等の条件を与えたコーパ

スが利用されてきたといえる。

2010 年代の後半からは DNN による TTS が主流になりつつあり、それに伴い学習に必要なデータ量も増大している。既存のコーパスでは十分な量を確保することは困難となったため、新たなコーパス作りに対する需要が生まれることとなった。JSUT [8] では、BASIC5000 と呼ばれる 5,000 文からなる大規模なコーパスを含んでおり、常用漢字と読みを含むという音素バランスとは異なる条件を満足している。ROHAN コーパス [9] は BASIC5000 と同様に常用漢字を含む性質を持ちつつ、「キェ」など日本語では出現する確率の低いモーラを含む 4,600 文から構築される日本語コーパスである。大規模なコーパスは日本語のみならず英語でも存在しており、代表的なものとして LJSpeech [10] や LibriTTS [11] が提案されている。LJSpeech はパブリックドメインであるため誰でも自由に利用できるという利点があり、これは ROHAN コーパスや ITA コーパスも同様である。利用ライセンスの設定は、幅広いユーザに利用してもらうためにも重要である。

1.2 関連研究と本研究の位置づけ

本研究では、大規模なコーパスの構築には時間がかかることに対応するため、GPT (Generative Pre-trained Transformer) を用いた日本語コーパスの作成について基礎的な検討を行う。関連する研究として、自動コーパス生成 [12], [13] や Web コーパス [14] の提案がすでになされている。本研究は、テキスト音声合成に向けたコーパスであり話者が朗読することを前提としている点でコンセプトが異なる。具体的には、文の長さや量、含まれるモーラの種類等をどのように調整したコーパスにするかが課題設定である。このようなコーパスの質に関する条件を満たす文を ChatGPT

¹ 明治大学
² SSS 合同会社
^{a)} ev200530@meiji.ac.jp
^{b)} Oda@zunko.jp
^{c)} mmorise@meiji.ac.jp

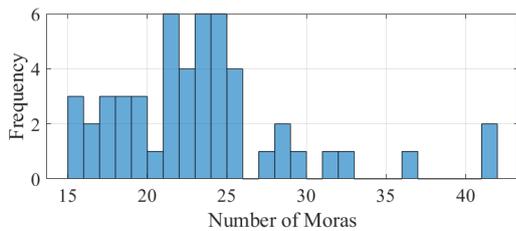


図 1 生成した 50 文の長さに関するヒストグラム。

により生成することにより、コーパス文生成におけるユーザのコストを抑制することが本研究の狙いである。本発表では、まずは小規模なテストとして 100 文を対象とし、コーパス文生成において必須の条件となるモーラ、文の長さ、文の数を指定する入力条件の検討及び、読み上げやすいコーパス文の生成について報告する。

2. GPT-3.5 を用いたコーパス生成

2.1 コーパス作成における条件

生成するコーパス文の条件は日本語のモーラをある程度バランス良く含んだ 15~25 文字の 100 文と設定し、発話者の読み上げやすさに配慮したコーパス文の作成を目指した。使用した OpenAI の GPT-3.5 default は、一度の質問に対する生成の最大文字数に 2,048 文字という制限があり、一度に 100 文全てを生成する事は出来ない。そこで、条件を分割することで文字数制限に影響することなく生成することを目指した。

字数や特定のモーラに関する条件の指定は、比較的簡単に評価が可能である。一方、読みやすさについては発話者個人による影響も存在する可能性があり [15]、条件を満足しているかの判断は容易ではない。それでも発話しやすさは朗読させることを考えると重要な条件であることから、具体的な評価はしないものの検討項目には入れることとした。

2.2 生成に向けた入力条件の検討

はじめに、文を構成するモーラ数に関する指定方法をテストすることとした。文の長さの単位としてはモーラが適切と考えられるが、モーラの指定では条件を満たさないものが多かったことを鑑み、本稿では文字数を対象としている。「20 文字以下」という条件を指定し 50 文生成した結果のヒストグラムを図 1 に示す。この条件で生成されたモーラ数は文字数と厳密には一致しないが、漢字を含むため文字数はモーラ数よりも少なくなることが想定される。しかしながら、生成された 50 文のうちモーラ数が 20 以下である文は 15 文であり、過半数の文では文字数を指定しても条件が満たされないことが確認された。

次の検討として、発話者の読みやすさに配慮するため、入力条件に用いるキーワードについて 2 つテストを実施した。まず「簡潔な」というワードを用いた結果、図 2 に示

M 50音と拗音のパターンを全て使って、簡潔な15~25文字の文章を10作成

- 
1. あしたの天気は雨。
 2. いい子にしましょう。
 3. うそをついてはいけません。
 4. えきについたら電話。
 5. お酒は飲み過ぎに注意。
 6. かぜをひかないように。
 7. きょうは忙しいです。
 8. くつは履き替えましょう。
 9. けが人は病院へ。
 10. ここに立ち入り禁止。

図 2 字数と「簡潔な」をキーワードに文を生成した結果。字数の条件を満たしていないことが確認できる。

M 50音と拗音のパターンを全て使って、適切な文体の文章を10作成

- 
1. 私は学校に行きます。(普通体)
 2. 彼女は本を読んでいます。(現在進行形)
 3. あなたはその本を読んだことがありますか？(過去完了形)
 4. 先生は優秀な教育者です。(尊敬語)
 5. 社長からご挨拶をいただきました。(謙譲語)
 6. ここは自由に入退室ができます。(丁寧語)
 7. 今日の天気はどうでしたか？(疑問文)
 8. これは私が作った料理です。(強調)
 9. 昨日はたくさん勉強したので、今日はゆっくり休みます。(論理的接続詞)
 10. 美味しい料理と素敵な音楽で、楽しい夜を過ごしました。(並列構造)

図 3 「適切な文体」をキーワードに生成した結果。ですます調の文が生成されているが、字数については上限・下限の範囲を超えたものも含まれる。

すとおり 15~25 文字を指定しているにも関わらず、指定された字数よりも短い文を生成していることが確認できる。指定する条件によっては文字数の条件と衝突する可能性が考えられるため、異なる条件として「適切な文体の」で検証を進めた。「適切な文体」の場合、文体の種類について記載されており、モーラ数については上限・下限の範囲を超えた文が生成されていることが確認できる。なお、筆者によるアンオフィシャルな発話テストではこの条件を指定することが読みやすさに寄与するとは言い難く、この条件は読みやすさの制御には適切とは言い難いと考えられる。

最後の検討として、ROHAN コーパス構築時に特定モーラを含めるという条件を設定していたことに倣い、拗音と濁音を全て網羅できるかの検証を実施した。具体的には、「50 音と拗音のパターンを全て使って」や「濁音、半濁音と拗音の組み合わせを全て使い」などの条件で生成を試みた。これは、1 文で特定の 1 モーラを加えるような条件では、多数の文を構築するためには多数の条件を設定しなければならないためである。今回は「濁音、半濁音と拗音の組み合わせを全て使い」という条件に加え「以下の制約に従って生成」と続けて必要条件を箇条書きで「15~25 文字、35 文、あ~そを使う」のように追加することとした。ただ

し、これでも全てのモーラの出現を満足することはできなかったことから、1回で全文を生成するのではなく、濁音、半濁音の条件を除いた「拗音を全て使って」に絞って生成することとした。この条件でも指定した全てのモーラが含まれる文が生成されなかったため、最終的には手動で特定モーラを含む単語に入れ替える等の作業が必要であった。

2.3 生成されたコーパス文の例

ここまでの検討に基づき、入力条件「拗音を全て使って20文字以下の文を50作文」により生成されたコーパス文を図4に示す。一度に生成する文は50文、1文を構成する文字数は20以下としている。得られた文から使用しているモーラのリストを構築し、不足しているモーラについて図5に示す条件を与え再度生成した。字数の条件は、50文生成された時点での文字数の情報をまとめ、そこから幅を持たせるように設定している。また、必要なモーラを全て指定し、モーラの数に十分に得られるような文の量について条件を与えた。ここまでの自動生成によって、ROHANコーパスのように日本語ではほぼ出現しない一部のモーラを除き、幅広いモーラをカバーしたコーパス文を構築することが可能であった。本稿では、コーパス文の読みやすさ等の評価は実施しておらず、次節で定性的な評価により有効性を議論する。

3. 有効性に関する議論

3.1 コーパス文生成の可能性

一度に大量の文を生成することは難しいものの、50文程度であれば同時に生成できることが確認できた。その際に、文字数指定等は条件を満たさない文を生成する可能性があるものの、入力方法の工夫によりある程度条件を揃えた文を作れる可能性が示された。読みやすさについては評価していないものの、アンオフィシャルな発話のテストでは既存のコーパスと比較して特段読みにくさは生じなかった。プロであればある程度難読であっても滑らかに発話できる可能性があるため、発話訓練を受けていないアマチュアによる評価が必要になるだろう。

生成された文の著作権の扱いについては現時点で決められないものの、生成AIによる生成結果に著作権を認めない場合は、自動的にパブリックドメインのコーパス文となる。ROHANコーパスのように全文を手動で作文する手間と比較すると、ChatGPTにより生成した結果を手動で修正の方がコストは少なく抑えられると考えられる。現時点で妥当な入力条件を得ているとは言い難いため、条件を厳密に満たす適切な入力条件について検討することは今後検討する必要がある。現状では手作業での修正が必要があるが、全てを作文することと比較すると修正元となる文があることは作文のコストをある程度抑制できていると考えられる。

M 拗音を全て使って、20文字以下の文を50作文して。

- 
1. きょうはびよんきちとみょうこちゃんとお遊んだ。
 2. ぎゅっと握りしめた手から、力が抜けた。
 3. しゃべりが上手い人に憧れる。
 4. じゃまされたくないから、ぼくはひとりりで遊ぶ。
 5. ちょっときついけど、がんばって歩く。
 6. びょうきのときは、薬を飲んで早く治したい。
 7. あの人のひどい言葉が、ぼくのところを痛めた。
 8. びりびりとした辛い味が、好きじゃない。
 9. ぐずぐずしていたら、時間ももたない。
 10. みかんの皮をむくのが、苦手だ。
 11. しょっちゅうねんどが手について、大変だ。
 12. つらいことがあっても、たえずに前に進もう。
 13. くしゃみをしたときは、マスクをするようにしよう。
 14. ぐずぐず言わずに、ささとやりなさい。
 15. かんたんなことだと思って、うかつにやると危ない。
 16. むずかしい問題を、ひとつずつ解決していこう。
 17. はしやぎすぎて、足をくじいてしまった。
 18. あの子のむずかしい名前を、なかなか覚えられない。
 19. いつも通りに歩いていたら、いきなりぶつかった。
 20. あの人のうそを見抜けなかった自分がかかりそう。
 21. にぎやかな場所にいると、ぼくは疲れてしまう。
 22. ぼくはあの人と仲良くなりたいたいと思っている。
 23. しゃべりすぎて、のどがかわいてしまった。
 24. びっくりしたときは、胸がどきりと動く。
 25. あのお店のカレーは、とってもおいしいと評判だ。
 26. ふつうの猫とちがって、ポテイルはしっぽが短い。
 27. あの人のうるさい話に、とうとううんざりしてしまっただ。
 28. しょうがないから、こつこつと仕事を頑張るしかない。
 29. いつもそばにいてくれる友だちに、感

M 続きを生成

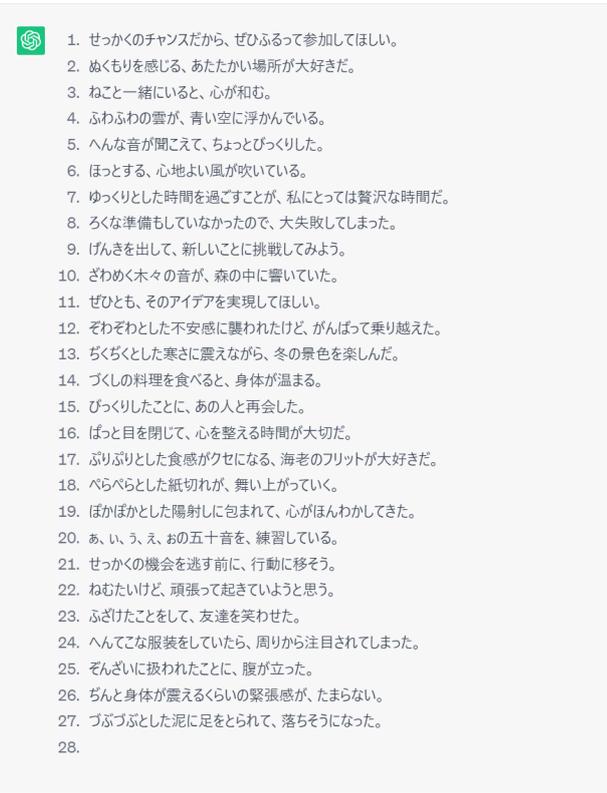
- 
- 謝りたい気持ちがある。
30. ぐちゃぐちゃになっていた机を、片付けなければ。
 31. うまいかないことがあっても、あきらめずに頑張ろう。
 32. かたくなになって、自分の意見を曲げないでいると、周りを傷つけてしまうこともある。
 33. あの人のぶりっ子な態度に、いらだててしまった。
 34. ぐっすり眠って、疲れをとりたい。
 35. かわいい子犬のピョンピョンとした動きが、癒しになる。
 36. じっと見つめられると、照れくさくなってしまっただ。
 37. しゃっくりが止まらず、困っている。
 38. ひどい天気の日には、家でゆっくり過ごしたい。
 39. がまんができるように、じっと耳を傾ける。
 40. ときには、思い切って行動することも必要だ。
 41. あの人のむずかしい話を理解するのに、苦労している。
 42. ぐずれ落ちそうになった心を、支えてくれた人に感謝している。
 43. ぐるぐると回り続ける回転寿司のお皿に、目が回ってしまった。
 44. しんどくても、目標に向かって歩き続ける。
 45. はっきり言葉を伝えることが、人間関係を良好に保つ秘訣だ。
 46. あの人のことを考えると、ときどき泣いてしまっただ。
 47. にぎやかな町並みを散策して、楽しい時間を過ごしたい。
 48. ひとりでいると、寂しさが募ってしまうこともある。
 49. すっきりとした空気を吸って、心と体をリフレッシュさせたい。
 50. わがままを言わずに、周りの人に感謝することが、人間関係を築く第一歩だ。

図4 最終的に使用した入力条件と生成結果

3.2 生成が難しい条件

今回はまだ基礎検討レベルであるが、現時点で生成が難しいケースもいくつか確認されたので紹介する。初めに、「ア、イ、ウ、エ、オ」を捨て仮名として発音できる文の生

M 以下の文字全てをそれぞれ3回以上用い、文字数15~25の文章を30文作成。
「せ、ぬ、ね、ふ、へ、ほ、ゆ、ろ、げ、ざ、ぜ、ぞ、ぢ、づ、び、ぼ、ぶ、べ、ぼ、あ、い、う、え、お」

- 
1. せっかくのチャンスだから、ぜひふるって参加してほしい。
2. ぬくもりを感じる、あたたかい場所が大好きだ。
3. ねこと一緒にいると、心が和む。
4. ふわふわの雲が、青い空に浮かんでいる。
5. へんな音が聞こえて、ちょっとびっくりした。
6. ほっとする、心地よい風が吹いている。
7. ゆっくりとした時間を過ごすことが、私にとっては贅沢な時間だ。
8. ろくな準備もしていなかったため、大失敗してしまった。
9. げんきを出して、新しいことに挑戦してみよう。
10. ざわめく木々の音が、森の中に響いていた。
11. ぜひとも、そのアイデアを実現してほしい。
12. そわそわとした不安感に襲われたけど、がんばって乗り越えた。
13. ぢくぢとした寒さに震えながら、冬の景色を楽しんだ。
14. づくしの料理を食べると、身体が温まる。
15. びっくりしたことに、あの人と再会した。
16. ぱっと目を閉じて、心を整える時間が大切だ。
17. ぷりぷりとした食感がクセになる、海老のフリットが大好きだ。
18. べらべらとした紙切れが、舞い上がっていく。
19. ぼかぼかとした陽射しに包まれて、心がほんわかしてきた。
20. あ、い、う、え、おの五十音を、練習している。
21. せっかくの機会を逃す前に、行動に移そう。
22. ねむたいけど、頑張って起きてみようと思う。
23. ふざけたことをして、友達を笑わせた。
24. へんてこな服装をしていたら、周りから注目されてしまった。
25. ぞんざいに扱われたことに、腹が立った。
26. ちんと身体が震えるくらいの緊張感が、たまらない。
27. つぶつぶとした泥に足をとられて、落ちそうになった。
28.

M 続きを生成

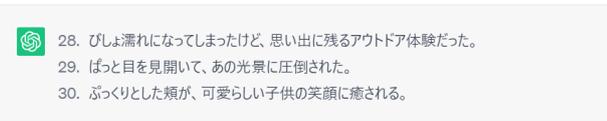
- 
28. びしょ濡れになってしまったけど、思い出に残るアウトドア体験だった。
29. ぱっと目を見開いて、あの光景に圧倒された。
30. ぷっくりとした頬が、可愛らしい子供の笑顔に癒される。

図 5 一度目の生成に対する不足しているモーラの補完

成が挙げられる。ただし、テキスト音声合成に利用することを前提とすると、現状の日本語テキスト音声合成は音素系列を入力としていることから、コーパス文も音素系列で表現できる文で構築することが望ましい。したがって、音素系列に変換することが容易であるモーラ単位で入力条件で指定して生成することが可能であれば、この問題は考慮しなくても良いと考えている。

もう1つは、文字数の制限や使用する文字を入力条件として与えたとしても、条件を満たさない文が生成されることである。今回の試行では、文を構成するモーラ数の条件を字数で指定したものの、条件を満たしていない生成結果が多々存在し、コーパス文に特定のモーラを含めるための条件についても同様の傾向が認められた。この問題は、指定する条件を工夫することで解決することも可能であるが、生成結果を手動で調整することでも解決できると思われる。

3.3 残された課題

今回のコーパス文生成については、読みやすさや音素バランスについての評価を実施していない。テキスト音声合成への利用を考えると、訓練を受けたプロの発話者による発声による読みやすさの評価、音素バランスの評価、テキスト音声合成システムとして利用した結果の合成音声の品質の評価等を実施し、総合的に妥当性を判断する必要がある。音素バランスについては、ROHAN コーパスでは考慮しているものの JSUT の BASIC5000 では考慮しておらず、近年の End-to-End を目指した方式においてどこまで考慮すべきかは定かではない。音素バランスを意識して読みやすさが損なわれるデメリットと、音素バランスを重視せず特定の組み合わせが出現しないデメリットのどちらがより大きな問題であるかは今後検討する必要がある。

本報告では、OpenAI の GPT-3.5 を使用した際の入力条件の検討及び問題点について言及しているが、これらの課題は GPT-4 を使用することで解決する可能性もある。GPT-4 の性能は一般に GPT-3.5 よりも優れているとされているが、コーパス文生成のようなタスクにおいて有効であるかについては現時点では検討が十分とは言い難い。今後は、GPT-4 を前提に入力条件を検討し、手動での修正を減らしたコーパス文生成に取り組むことも必要である。

4. おわりに

本稿では、ChatGPT を用いてコーパス文を作成するための基礎検討を実施した結果を報告した。ある程度の条件を満たした文の生成自体は可能であるものの、生成結果が条件を満たさないこと等の問題があるため、手動で修正することや入力条件を調整することが必要になると考えられる。生成された文の読みやすさについては、本手法で構築したコーパスを実際に発話させることで検証することが必要になる。

今回のテストは GPT-3.5 により実施しているため、これを GPT-4 にすることで状況が変化することは考えられる。したがって、今後は GPT のバージョンにより生成されるコーパス文にどのような変化が生じ具体的な発話のしやすさに影響するかなど、コーパス文そのものの評価を行うことを予定している。そのためには、コーパス文に対する適切な評価指標を策定することが重要である。加えて、テキスト音声合成への利用が目標であることから、合成音声の品質によりコーパスの量・質の妥当性についても検討する。

謝辞 本研究の一部は、JSPS 科研費 JP21H04900, JP21K19794 の支援を受けました。

参考文献

- [1] H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis," *Speech Communication*, vol. 51, no. 11, pp. 1039–1064, 2009.

- [2] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “WaveNet: A generative model for raw audio,” arXiv preprint arXiv:1609.03499, 2016.
- [3] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. J. Skerry-Ryan, R. A. Saurous, Y. Agiomyrgiannakis, and Y. Wu, “Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions,” in Proc. ICASSP 2018, pp. 4779–4783, 2018.
- [4] 山本龍一, 高道慎之介, “Python で学ぶ音声合成,” 株式会社インプレス, 2021.
- [5] 小林哲則, 板橋秀一, 速水悟, 竹澤寿幸, “日本音響学会研究用連続音声データベース,” 日本音響学会誌, vol. 48, no. 12, pp. 888–893, 1992.
- [6] “声優統計コーパス,” <https://voice-statistics.github.io/>
- [7] 小口純矢, 金井郁也, 小田恭央, 齊藤剛史, 森勢将雅, “ITA コーパス: パブリックドメインの音素バランス文からなる日本語テキストコーパスの構築と基礎評価” 情報処理学会研究報告, vol.2021-SLP-137 no.31, pp. 1–4 (2021).
- [8] S. Takamichi, R. Sonobe, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji, and H. Saruwatari, “JSUT and JVS: Free Japanese voice corpora for accelerating speech synthesis research,” Acoust. Sci. Tech, vol. 41, no. 5, pp. 34–45 (2021).
- [9] 森勢将雅, “ROHAN: テキスト音声合成に向けたモーラバランス型日本語コーパス,” 日本音響学会誌, vol. 79, no. 1, pp. 9–17 (2023).
- [10] K. Ito and L. Johnson, “The LJ Speech Dataset,” <https://keithito.com/LJ-Speech-Dataset/>.
- [11] H. Zen, V. Dang, R. Clark, Y. Zhang, R. J. Weiss, Y. Jia, Z. Chen and Y. Wu, “LibriTTS: A corpus derived from LibriSpeech for text-to-speech,” in Proc. INTER-SPEECH 2019, pp. 1526–1530, 2019.
- [12] 藤原奈々美, 山内真樹, 内山将夫, 隅田英一郎, “自動コーパス生成による少量対訳コーパスからの統計的機械翻訳,” 言語処理学会第 22 回年次大会発表論文集, pp. 219–222 (2016).
- [13] 今出昌宏, 藤原奈々美, 山内真樹, “ニューラルネット機械翻訳における自動コーパス生成適用,” 人工知能学会第 31 回大会, pp. 1–4 (2017).
- [14] 関口洋一, 山本和英, “Web コーパスの提案,” 情報処理学会研究報告, 2003-NL-157, pp. 123–130 (2003).
- [15] 山本泰我, 小口純矢, 森勢将雅, “プロ声優を対象とした大規模コーパス朗読におけるリテイク数の比較,” 音講論(秋), pp. 1401-1402 (2022).