日本語文章のための話し言葉・書き言葉変換コーパス

庵 愛　　高島 瑛彦　　増村 亮
日本電信電話株式会社，NTT メディアインテリジェンス研究所
{mana.ihori.kx, akihiko.takashima.dg, ryou.masumura.ba}@hco.ntt.co.jp

1 はじめに

音声認識を用いたアプリケーションの増加に伴い，音声言語を精緻に捉える需要性が高まっている．しかし，音声認識の出力結果は，自然会話で生成される音声言語をそのまま書き起こすため，言い淀みや冗長表現を含む話し言葉テキストとなっており，可読性が低いことが課題である．そこで，音声認識結果の可読性を向上させるため，話し言葉を書き言葉に変換する技術に我々は着目する．この技術は，音声認識の後段の処理である翻訳や要約など，書き言葉テキストの入力が望ましいタスクにおいても有用だと考えられる．

話し言葉から書き言葉へ変換するタスクは，同一言語内の翻訳問題として考えられ，ニューラル機械翻訳の成功を受けてニューラル系列変換モデルに基づく研究が盛んに行われている [4, 6, 8]．しかし，このようなニューラル系列変換モデルでは，入力と出力の関係性を end-to-end でモデル化するために，入力と出力の対データが大量に必要となる．そのため，話し言葉から書き言葉へのニューラル系列変換モデルを作成するためには，話し言葉と書き言葉の文対を大量に用意する必要がある．

話し言葉を書き言葉へ変換するためには，様々な要素を考慮する必要がある．例えば，フィラーや言い淀み，冗長表現の削除や，句読点の付与が必要となる．これらの要素を網羅的に考慮するほど，可読性の高い書き言葉テキストに変換できると考えられる．しかし，従来の研究ではこれらを独立に捉え，それぞれの要素に適したコーパスを使用していた [1,7]．そのため，従来のコーパスではこれらの要素を同時に考慮した変換が実現できない．

また，日本語には話し言葉，書き言葉それぞれに特有の表現が存在する．例えば，話し言葉テキストでは助詞が省略されることがしばしばあるが，書き言葉テキストでは省略することができない．また，話し言葉テキストでは文体の統一が厳密に行われていないが，書き言葉テキストでは読者の混乱を防ぐために文体を統一する必要がある．このように，日本語の文章を対象とする場合，従来考えられていた要素に加えて，日本語特有の要素についても考慮する必要がある．

以上より，本稿では，話し言葉テキストの可読性を向上させることを目的とし，従来の要素と日本語特有の要素の両方を考慮した，日本語文章のための話し言葉から書き言葉へ変換するコーパスを作成する．話し言葉テキストは，複数の要素を網羅的に考慮するほど可読性の高い書き言葉テキストに変換できると考えられるため，本コーパスでは複数の要素を同時に変換する．日本語の話し言葉を扱ったコーパスとして，日本語話し言葉コーパス [3] が存在するが，このコーパスにはフィラーと言い淀み，一定時間の休止区間のみがアノテーションされており，日本語特有の要素については全く考慮されていない．そのため，本コーパスは，日本語における話し言葉・書き言葉変換に関する複数の要素を同時に考慮する初めてのコーパスである．

本稿の貢献を以下に示す．(1) 日本語の文章における話し言葉・き言葉変換コーパスを作成するためのルールを構築した．(2) クラウドソーシングを用いて 4 つのドメインにおける話し言葉，書き言葉の文対を作成した．(3) 作成したコーパスで学習したニューラル系列変換モデルを用いて話し言葉・書き言葉変換におけるベースラインを確認した．

2 話し言葉・書き言葉の変換ルール

本節では，日本語の話し言葉・書き言葉変換を行うためのルールについて詳細に説明する．このルールは，クラウドソーシングで採用した日本人作業者が話し言葉テキストを書き言葉テキストへ変換するときに提示されるものである．我々は，日本語特有のルールを 3 個，一般的なルールを 4 個の計 7 個のルールを作成した．これらのルールをすべて適用した話し言葉・書き言葉変換の例を表 1 に示す．以下，それぞれのルールについて述べる．

2.1 日本語特有のルール

(1) 文体の統一 日本語の文体には，文末に “だ”，“である” などを用いる常体と文末に “です”，“ます” などの丁寧語を用いる敬体が存在する．書き言葉では一般的に常体が用いられるが，我々の用意した話し言葉テキストは，発話を書き起こしたものであるために文末を “だ”，“である” に統一すると不自然な文章となる．さらに，本コーパスでは厳密に書き言葉へ変換することを目的としているわけでなく，可読性の高い文章へ変換することを目的としている．そのため，本コーパスでの文体は，話し言葉でも書き言葉でも用いられる敬体に統一する．

また，日本語には，話し言葉，書き言葉それぞれに特有の表現が存在する．接続詞を例にとると，話し言葉で使われる “でも”，“だから” などの表現は，書き言葉では “しかし”，“したがって” という表現に変換される．しかし，このような変換を厳密に行うことは個人の知識や経験に左右されるため，文章執筆経験の乏しい作業者にとって敷居の高い作業になりかねない．そのため，これらの変換について簡単な例を示し，上記と合わせて以下のルールに従って変換してもらうこととした．**「話し言葉」（口語）を丁寧な言葉遣い（ですます調など）に修正してください．（例）〜みたい → 〜のよう，こっち → こちら，とか → など，だよね → ですよね，〜だったっけ →〜でしたでしょうか，だったかも → だったかもしれません**

(2) 助詞の復元 日本語の話し言葉では，しばしば助詞が省略される．しかし，助詞は名詞と動詞，形容詞の意味関係を示す役割を果たしているため，正しく文章の意味を伝えることを目的とした書き言葉では助詞を省略することができない．そのため，助詞が省略されている場合に以下のルールに従って復元してもらうこととした．**「助詞」（〜が，〜は，〜に，〜を）が不足している場合には，加えてください．**

(3) かな漢字表記の統一 我々の用意した話し言葉テキストは，発話を人手で書き起こしたものであるため，しばしば表記の揺れが存在する．例えば，数字が漢数字に変換されていたり，英語表記がひらがな表記に変換されている．そのため，以下のルールに従ってかな漢字表記の修正を行ってもらうこととした．**読みづらい英語，数字表記，ひらがな表記を修正してください．（例）一二三の四五六の七八九〇 →123-456-7890，えぬてぃーてぃー →NTT**

表 1: 話し言葉・書き言葉変換の例

話し言葉テキスト	書き言葉テキスト
はいはい，それはそうですね 私なんかは運動をたくさんしているので，ご飯もそれほど食べていないので，だいえっとする必要ってないですね いわゆるメタボとは無縁ちゃ無縁ですが，糖尿病にはきをつけてます	それはそうですね． 私は，運動をたくさんしていますし，ご飯もそれほど食べません．よって，ダイエットする必要はないですね． メタボとは無縁ですが，糖尿病には気を付けてます．

2.2 一般的なルール

(4) 句読点の付与 我々の用意した話し言葉テキストには一定時間の休止区間に従ってアノテーションがされているため，それらを読点に変換している．しかし，それらは可読性向上を観点に付与されているわけではないため，抜けや誤りが存在する．そのため，以下のルールに従って読点の誤りを修正，必要な箇所には新たに句読点を付与してもらうこととした．**接続語（そして，しかし，また，つまりなど）の後や，漢字やひらがなが続く場合は，読みやすくなるように読点「,」を加えてください．また，句点「.」と読点「,」の付け方に誤りがありましたら，修正してください．**

(5) 言い淀み表現の除去 フィラーや言い淀みのある文章は可読性が低いため，以下のルールに従って除去してもらうこととした．**「言い淀み」表現を除去してください．**

(6) 冗長表現の削除や文章の簡略化 話し言葉では，思いつくまま話された発話をそのまま書き起こすために，冗長な文章や文法的に誤りのある文章が存在する．そのため，それらの文を以下のルールに従って修正してもらうこととした．**同じ表現が繰り返される場合，無駄な表現を削除，あるいは文章を区切るなどして，読みやすい文章に修正してください．（例）このデザート，甘いといえば甘いよね→このデザート，甘いですよね．安いと思って，お手軽だと思って買った→安いと思いました．また，お手軽だと思ったので買いました．**

(7) 音声認識誤りの修正 我々の用意した話し言葉テキストは，発話を人手で書き起こしたものであるため，しばしば認識誤りが存在する．そのため，以下のルールに従って認識誤りを修正してもらうこととした．**文脈から誤字と認識される言葉は修正してください．**

3　本コーパスの作成方法

本コーパスを作成するにあたり，まず 4 つのドメインの話し言葉テキストを用意した．次に，クラウドソーシングサービスを用いて日本語作業者を採用し，話し言葉テキストを書き言葉テキストへ書き換えてもらった．最後に，それらのデータから不要なデータを削除することで本コーパスを作成した．これらの詳細について以下に述べる．

3.1　話し言葉テキスト

本稿で用いた話し言葉テキストは，4 つのドメインの発話を人手で書き起こしたものである．これらのテキストは，20 文字以上が含まれるものを採用している．各ドメインの詳細について以下に示す．

- コールセンタ対話 : コールセンタで行われるオペレータと顧客の対話をシミュレーションした音声であり，オペレータと顧客，両方の音声を書き起こしたものである．ここでは，3,965 文を用意した．
- 自由会話 (1) : 趣味や旅行などについて 4 人の被験者に自由に会話してもらった音声であり，それぞれの音声を書き起こしたものである．ここでは，3,962 文を用意した．
- 自由会話 (2) : ライフイベントや趣味などについて 2 人の被験者に自由に会話してもらった音声であり，それぞれの音声を書き起こしたものである．ここでは，4,501 文を用意した．
- 留守番電話 : 留守番電話の音声を書き起こしたものである．ここでは，12,567 文を用意した．

3.2　クラウドソーシングでの書き言葉変換

話し言葉テキストを書き言葉テキストへ変換する日本人作業者を雇うためにクラウドソーシングサービスを利用した．このサービスではアンケート形式の Web ページを用いた．Web ページ上には，話し言葉テキスト，3 章に示した変換ルール，表 1 に示した変換例，話し言葉テキストがあらかじめ入力されているテキストボックスを表示した．作業者には，3 章で構築した変換ルールに従ってテキストボックス上の原文を編集する形で書き言葉へ変換してもらった．このときに確実に編集してもらうようにするため，1 文を編集するたびに確認ページに遷移し，自身の回答を確認すると次の文の編集ができるような設計にした．このサービスを用いて，15-88 歳の男女 9,002 人の日本人作業者を採用した．確実に話し言葉テキストを収集するために，1 つの原文に対して 3 人以上の作業者を割り当てた．また，作業者 1 人につき 10 文を編集してもらった．

表 2: コーパスの詳細

ドメイン	訓練	検証	テスト
コールセンタ対話	8,169	584	1,475
自由会話 (1)	5,328	381	996
自由会話 (2)	8,123	581	1,150
留守番電話	15,129	1,081	2,794

本コーパスを作成するにあたり，話し言葉テキストから句読点は除去した．また，上記の方法で収集した文のうち，全く編集されていない文やフィラーが含まれている文，丁寧語に修正されていない文は人手で除外した．収集した文は，訓練データ，検証データ，テストデータの 3 種類に分割した．各ドメインにおけるそれぞれの文数を表 2 に示す．

4　ニューラル話し言葉・書き言葉変換のベースライン評価

実験設定　本コーパスを用いた話し言葉・書き言葉のニューラル系列変換モデルを作成するために，attention-based encoder-decoder network [2] と pointer-generator network [5] の 2 種類のネットワークを構築した．pointer-generator network は，原文からの単語のコピーを可能としたコピー機構を有するため，入出力で複数の単語を共有する同一言語内の翻訳問題で高い性能を示すことが期待されている．各ネットワークは，すべてのドメインのデータを合わせたデータで学習した．これらのネットワークの構造を以下に示す．エンコーダでは，512 ユニットを持つ 2 層の双方向 LSTM-RNN を採用した．デコーダでは，512 ユニットを持つ単方向 LSTM-RNN を採用した．注意機構には，additive attention mechanism を採用した．出力層のユニットサイズは，全ての学習データで 10 回以上出現する文字数である 1,763 に設定した．学習には，勾配ノルムクリッピングを 1.0 に設定したミニバッチ確率勾配降下法を採用した．各 LSTM-RNN では，ドロップアウトを 0.2 に設定した．全ての学習可能なパラメータはランダムに初期化した．ミニバッチ学習をするために，各文 200 文字で truncate した．ミニバッチサイズは 64 に設定した．また，デコード時にはビームサーチアルゴリズムを採用し，ビームサイズは 4 に設定した．

表 3: 話し言葉・書き言葉変換の評価値

ドメイン		BLEU	ROUGE-L	METEOR
コールセンタ対話	a)	0.591	0.693	0.737
	b)	0.705	0.775	0.867
	c)	**0.723**	**0.784**	**0.886**
自由会話 (1)	a)	0.595	0.674	0.765
	b)	0.568	0.689	0.810
	c)	**0.766**	**0.763**	**0.839**
自由会話 (2)	a)	0.629	0.695	0.763
	b)	0.656	0.718	0.828
	c)	**0.672**	**0.726**	**0.843**
留守番電話	a)	0.629	0.731	0.756
	b)	0.748	0.799	0.879
	c)	**0.752**	**0.801**	**0.884**

a) 話し言葉テキスト
b) attention based encoder-decoder network
c) pointer-generator network

評価方法　上記のネットワークを用いて生成された書き言葉テキストに対して，BLEU，ROUGE-L，METEOR を用いて評価値を算出した．BLEU の計算には，N=3 を採用した．ここで，本コーパスは 1 つの話し言葉テキストに対して複数の書き言葉テキストの正解が存在するため，最も評価値が高い書き言葉テキストを正解とした．また，書き言葉テキストに対しても各値を算出した．

結果　各ドメインにおける話し言葉テキスト，各ネットワークの出力結果に対する各値の結果を表 3 に示す．表 3 より，pointer-generator network を用いた書き言葉テキストの性能が最も良いことがわかる．実際に，pointer-generator network から出力された書き言葉テキストの例を表 4 に示す．表 4 より，3 章で設定した話し言葉・書き言葉変換のルールについて網羅的に変換できていることがわかる．以上より，本コーパスを用いることで，話し言葉・書き言葉変換に関する複数の要素を同時に考慮するニューラル系列変換モデルが構築可能なことが示唆された．

5　おわりに

本稿では，4 つのドメインにおける新しい日本語文章における話し言葉・書き言葉変換コーパスを提案した．既存のコーパスとの大きな違いは，変換に必要な従来の要素と日本語特有の要素をすべて同時に考慮している点である．本コーパスを作成するために，話し言葉・書き言葉変換のルールを作成し，クラウドソーシングサービスを利用して日本人作業者を採用した．また，本コーパスの質を調査するため，ニューラル系列変換モデルを用いて書き言葉を生成し，複数の要素を同時に考慮した書き言葉が生成できていることを確認した．このコーパスは，話し言葉から書き言葉へ変換するだけでなく，書き言葉から話し言葉への変換にも活用できると考えられるため，今後調査を行う．

表 4: 書き言葉テキストの生成例

コールセンタ対話	
入力	えーあのもうあんまり使わなくなったということなんですねあの先日からちょっとお習い事を始めまして
出力	あんまり使わなくなったということなんですね．先日からお習い事を始めました．
自由会話 (1)	
入力	プロ野球じ野球ほとんどでもルールわかんないんですよおふほとんどわかんないですよ高校野球は見ますけど
出力	プロ野球は，ほとんどルールがわかんないんですよ．ほとんどわからないんですよ．高校野球は見ますけど．
留守番電話	
入力	あもしもしえーと午前中の打ち合わせ終わりましたので連絡ですえーとー打ち合わせ終わりまして秋葉原の駅まで来ましたえーこれからえー昼食を取りつつえ次の打ち合わせ高田馬場に向かいますえーその連絡でしたえーまた後ほど連絡いたしますよろしくお願いいたします
出力	もしもし，午前中の打ち合わせ終わりましたので，連絡です．打ち合わせが終わりまして，秋葉原の駅まで来ました．これから昼食を取り，次の打ち合わせ，高田馬場に向かいます．また後ほど連絡いたします．よろしくお願いいたします．

参考文献

[1] John J Godfrey, Edward C Holliman, and Jane McDaniel. Switchboard: Telephone speech corpus for research and development. In *Proc. ICASSP*, pp. 517–520, 1992.

[2] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In *Proc. EMNLP*, pp. 1412–1421, 2015.

[3] Kikuo Maekawa, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara. Spontaneous speech corpus of japanese. In *Proc. LREC*, pp. 947–9520, 2000.

[4] Ernest Pusateri, Bharat Ram Ambati, Elizabeth Brooks, Ondrej Platek, Donald McAllaster, and Venki Nagesha. A mostly data-driven approach to inverse text normalization. In *Proc. INTERSPEECH*, pp. 2784–2788, 2017.

[5] Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. In *Proc. ACL*, pp. 1073–1083, 2017.

[6] Ottokar Tilk and Tanel Alumäe. Bidirectional recurrent neural network with attention mechanism for punctuation restoration. In *Proc. INTERSPEECH*, pp. 3047–3051, 2016.

[7] Nicola Ueffing, Maximilian Bisani, and Paul Vozila. Improved models for automatic punctuation prediction for spoken and written text. In *Proc. INTERSPEECH*, pp. 3097–3101, 2013.

[8] Shaolei Wang, Wanxiang Che, and Ting Liu. A neural attention model for disfluency detection. In *Proc. COLING*, pp. 278–287, 2016.