

話し言葉への間の挿入と評価

ガルシア ユウジ[†] 森 博彦[†] 岡 誠[‡]東京都市大学大学院工学研究科[†] 東京都市大学知識工学部[‡]

1 研究背景

世界中では多くの音声合成エージェントが発売されており、自然言語処理を扱う音声合成エージェントが身近になってきた。自然言語処理の研究では文脈の理解などの研究が注目され、精度の高い返答を行えるようになってきた。一方で、音声合成エージェントの話し方に着目した研究は少ない。音声合成エージェントは主に「書き言葉」といった、話す内容が正確に伝わるよう句読点のはっきりした話し方である事が多い。しかし、自然な発話では文法的に支離滅裂だが、意味が伝わる「話し言葉」が扱われている。特に、音声合成エージェントが間を置く位置に関しては書き言葉であれば句読点を置く位置になるが、書き言葉の場合は議論がされていない。本研究では音声合成エージェントがより話し言葉に近づくため発話中に置く間の位置に着目して研究を行う。

2 関連研究

村田ら[1]は統計的アプローチで書き言葉の日本語テキストへ読点を自動挿入した。読点を挿入するためのアルゴリズムとして、新聞のテキストデータを係り受け解析し文節単位に分け、各文節に対してどのような素性があるかラベルを付与し、それら素性ラベルで求められる確率を用いて最大エントロピー法を適用した確率モデルを作成した。テストデータの各文節に対して読点を挿入する確率を求め、動的に読点を挿入する実験の結果、読点の再現率70.66%、適合率84.65%となった。本研究では、村田らの読点挿入手法を参考に、話し言葉での素性ラベルを定義し、最大エントロピー法を用いた確率の算出により、間の挿入を行う。

3 目的

話し言葉を扱う音声合成システムが置く間の最適なタイミングの推定し、音声合成による評価を行う。

4 検証

4.1 間の位置の検証

間の適切な挿入箇所の検討にあたって実際に人が

扱う話し言葉ではどのように間が扱われているか、音声データを聴き検証した。音声データでは慶應義塾大学・中央大学・東京都市大学の学生の中で初対面の2人が15分間対談する音声进行分析した。分析の結果、話し言葉には構文的な間と非構文的な間の2種類の間の用法があることが確認できた。構文的な間は、文節間に入る間であり、村田ら[1]の研究で定義された間と同じような間が扱われている。一方で、非構文的な間では、文節ではなく全ての形態素間に間が挿入される可能性がある。これら非構文的な間は2種類に分類でき、固有名詞や次に話す内容を思い出す間と、自らの言い間違えに気付いた瞬間の間を確認できた。

本研究では、これら構文的な間と非構文的な間を再現することで、自然な話し言葉を話す発話エージェントの作成を目指す。また、発話エージェントがいい間違えを訂正する場面はないと考え、訂正の間は本研究では扱わないとする。

4.2 音声合成による検証

4.2.1 間の挿入アルゴリズムの提案

表1 間の素性ラベル表

構文的	文節の最終形態素の節の分類
非構文的	形態素の品詞
	間と間の形態素数
	固有名詞の直前かどうか

村田らが提案した読点の用法の分類と間の位置の検証を参考に、間の素性ラベル表を作成し、それぞれの素性を定義した(表1)。構文的な間で「文節の最終形態素の節の分類」に関しては、村田ら[1]の研究を参考に文節の最終形態素に対して文節情報を素性ラベルとして付与することにした。素性ラベルは現代日本語の節の分類を行なっている佐藤ら[2]の定義した節境界認定規則を利用した。非構文的な間で「形態素の品詞」に関しては、各形態素の持つ品詞を素性ラベルとして付与する。「間と間の文字数」に関しては、村田ら[1]の研究同様に、一文が長くなりすぎないように、定期的に文字数が区切られることが話し言葉でも考えられるため、各形態素に対して直前の間からの形態素数を4つに分類して(0~1, 2~3, 4~6, 7~)、素性ラベルを与える。「固有名詞の直前かどうか」に関しては、事前検証で見つけた「内容を思い出す間」であり、固有名詞を思

Insert interval to spoken language

[†] Tokyo City University Graduate School of Integrative Science and Engineering[‡] Tokyo City University Faculty of Knowledge Engineering

い出す直前に間を取ることが確認されているため、固有名詞である形態素の直前の形態素には、固有名詞の直前であるという素性ラベルを与える。これら素性ラベルの出現を元に求められる確率モデルに最大エントロピー法を用いて、間の挿入確率を推定する確率モデルを作成した。

4.2.2 音声合成による比較検証

作成した確率モデルを比較し、話し言葉への間の挿入に対して効果の有無を検証する。本研究のモデルと比較対象として、ランダムに間が挿入されるもの、文章の本来の位置にある間が挿入されるもの、3つを比較する。評価基準として、Amazon が提供する音声合成 Alexa に間のない文章を音声合成させ、自然に出現する間を基準の3として聞かせた。評価項目はそれぞれ5段階、自然に聞こえたか、話す速度は自然だったか、不自然な間はあったかを評価してもらう。

表 2 挿入した間の評価結果の平均

	自然さ	速度	不自然な間
基準	3.00	3.00	None
ランダム	1.83	3.00	4.67
本研究	3.33	4.17	3.83
本来	4.67	5.00	3.33

表 2 に示した評価の結果、自然さに関しては、本研究のアルゴリズムはランダムに間を挿入するよりも自然だが、本来の間の状態ほど良くはないということがわかった。これはランダムでは非構文的な間に挿入されることが多く、基準は間がなさすぎて、本研究が良い結果となった。一方で、本来の間に及ばなかった理由として、文の終了を示す間がない箇所があり違和感に繋がった。話す速度に関しては、適切な間の位置になる程、聴きやすい速度になることがわかった。これはランダムでは不自然な間が多く、基準は間が少なすぎるので速く聴こえたと考える。不自然な間に関しては、全体的に不自然な間があったという結果になった。本来の間であっても全ての人にとって自然な間は存在しないことがわかるが、「あまりなかった」に可能な限り近くことはできる。

定性的なアンケートでは、具体的な間の改善方法が記述されていることが多く、ルールベースの検討などの必要性を考える。また、内容を思い出す間として、素性ラベルを付与したが、音声合成がエピソードや固有名詞を思い出しているという状態を被験者に想起させられることはなく、その間自体に意味を持たせることはできなかった。しかし、全体を通して話し言葉には特有の間の表現があり、多少のルールを無意識のうちに感じて話し言葉を聞いているということもアンケートからわかった。

4.3 言い方に特化した確率モデルの検証

表 3 言い方重視の間の素性ラベル表

構文的	形態素の種類
非構文的	形態素の品詞
全体	間と間の形態素数

音声合成による比較検証の結果より、話し言葉での間の位置に重要な要素は、思い出しているなどの用法よりも、ある言い方の時は間が入り、ある言い方の時は間が入らないなど、無意識の内に存在するルールのような基準が各々設けられているということがわかった。表 1 で提案した素性ラベル表で、構文的な間が入る位置での素性は節を分類してしまっているが、これは用法に基づいてしまっているため、これを形態素の種類まで分解し、素性のラベルとする。こうすることで、確率モデル計算の際に、用法ではなく、言い方に対して構文的な間を挿入することができる。また、間と間の形態素数の素性ラベルに関しては構文的、非構文的に関わらず使える素性であったため、全ての形態素に対してラベルを付与することにした。固有名詞の直前のラベルに関しては、用法であり、言い方ではないため削除した。更に詳細なルールを定義した。言い方のルールとして、接続助詞の「て」と「それで」の省略形「で」の間には間が存在することがわかったため、必ず間が挿入されるようにした。また間の直前に母音を伸ばす傾向があるため、音声合成で間が入る直前の 1 文字の母音を伸ばすように音声合成を調整した。

新しく作成した言い方に特化した音声合成を 2 人の被験者に聴かせて定性的な調査をした所、全員に共通して間が多いという意見が出た。そこで確率モデルで推定した間の中から 10%取り除いた、すると間は自然な位置になっているという評価を得た。

5 今後の方針

間の量に関して、人間と音声合成では違和感を感じる量に違いがあるため、より適切な間の量の検討したい。また、本来の話し言葉で間投詞が多かったが、間の挿入を優先し、本研究では扱わなかったため、話し言葉での間投詞の効用を検討したい

考察文献

- [1] 村田匡輝, 大野誠寛, 松原茂樹: 「読点の用法的分類に基づく日本語テキストへの自動読点挿入」, 電子情報通信学会論文誌, vol.I95-D, No.9, (2012)
- [2] 佐藤理史, 丸山岳彦, 夏日和子: 「現代日本語書き言葉均衡コーパスに対する節境界付与」, 言語処理学会, 第 22 回年次大会発表論文集, (2016)