

言語行為列をシナリオとした文生成

2K-5

堀井統之 大山芳史
NTT情報通信研究所

1. はじめに

本稿では、文章の構造を各文の意図からなる言語行為の列として捉え、その言語行為列をシナリオとして文章を生成する手法について述べる。本手法が電報文の生成に有効であることを示し、さらに生成された電報文候補にスコア付けをするための制約について考察する。

2. 概要

2.1. 電報文の生成

我々は、ユーザの要望に合った電報文を作成支援するシステムについて検討しており、その第1段階として、ユーザが電報文に盛り込みたい1文を入力するとその文の前後に適当な文を自動的に付加して電報文を完成させるという試みを行なっている(図1)。

図1のしくみを実現するためには、入力文を解析して何らかの情報を取得し、その情報をもとに前後に付加できる文を決定しなくてはならない。

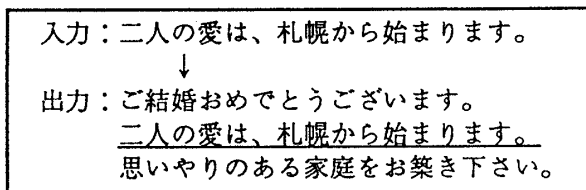


図1 1文からの電報文作成支援例

2.2. 言語行為列シナリオ

文献[1]の「結婚の電報例文」を対象に電報文の構造について分析すると、図2のようなある決まりきったパターンが多く見られる。図2におけるどの電報文も、構成する3文の言語行為に着目すると、

祝福-陳述-要望^{*1}

というパターンになっている。図2に示す例文以外でも、上記のパターンで構成されている電報文は多く、また上記のパターンでない場合でも、

祝福-要望

祝福-要望-要望

といった決まったパターンが用いられている場合が多い。

このことは、電報文の生成を考える上で、電報文を構成する各文の制約として利用することができる。

*1 ここでいう“要望”は、文献[2]に従い、“依頼”、“祈願”、“命令”などを含んでいる。

ご結婚おめでとうございます。(祝福) 北の町の春は、お二人の愛から始まります。(陳述) 末長くお幸せに。(要望)
--

ご結婚おめでとうございます。(祝福) 夫婦は自転車の輪と同じ。(陳述) 両輪そろってこれからの人生を完走しましょう。(要望)
--

結婚おめでとう。(祝福) 手に手をとって、ハワイ焼けとはうらやましい。(陳述) 帰ってきたらみんなで押し掛けるから、たくさん話題を仕入れて来いよ。(要望)

図2 電報文の典型的なパターン

すなわち、3文からなる電報文を生成するならば、それぞれの文の言語行為が、

祝福-陳述-要望

祝福-要望-要望

のようになることを1つの制約とすることができる。

我々はこのような各文の言語行為の並びを「言語行為列」と呼んでいる。本稿で提案する文生成の手法は、この言語行為列を生成する文章のシナリオとして用いるものである。

2.3. 本手法の利用例

2.1. で述べた図1のしくみを実現するためには、言語行為列をシナリオとした文生成の手法が有効である。入力文の前後に付加する文はデータベース等から取得してくるものとする、以下の手順で電報文を生成することができる。

1. 任意に入力された1文を解析してその文の言語行為を得る。

2. 入力文の言語行為を含んでいるシナリオを選択する。

3. 選択されたシナリオに沿って入力文の前後に付加する文をデータベースから取得することにより電報文全体を完成させる。

1.の言語行為の取得方法については、文献[2]にあるように各文の文末の形態的特徴を利用することができる。

3. 実験及び考察

3.1. 実験

2章において本稿で提案する手法の概要を説明したが、その有効性を確認するため、実験を行なった。

実験対象とする電報文は、文献[1]の「結婚の電報例文」のうち、3文から構成される電報文68電文とした。各文の言語行為は全て付与されていることを前提とし、実験に用いるシナリオとしては、一番頻りに現れる“祝福-陳述-要望”を選んだ。

以下に実験の概要を示す。

[実験]

各電報文の j 文目の集合を A_j ($j=1,2,3$) とし、文 a の言語行為を $SPact(a)$ で表わすとすると、 $a_2 \in A_2 \cap SPact(a_2) = \text{"陳述"}$ である文 a_2 を 1 つ取り出し、以下の 2 つの方法で電報文の生成を行なう。

- (1) $a_1 \in A_1$ 、 $a_3 \in A_3$ をそれぞれランダムに選び、" $a_1 a_2 a_3$ " を生成する。
- (2) $a_1 \in A_1 \cap SPact(a_1) = \text{"祝福"}$ である文 a_1 、 $a_3 \in A_3 \cap SPact(a_3) = \text{"要望"}$ である文 a_3 をそれぞれランダムに選び、" $a_1 a_2 a_3$ " を生成する。

文 a_2 として取り出した 1 文に対し、(1)、(2) それぞれの方法で 30 電文ずつを生成させ、自然な電報文の生成率を比較する。

上記の方法(1)は、使われる文の位置に制約はある(1文目にはオリジナルの電報文で1文目に使われている文がくる)が、3文を単に並べただけである。それに対し方法(2)は、言語行為列(祝福-陳述-要望)をシナリオとして電報文を生成している。

表1に実験結果を示す。表中の文1、文2は、文 a_2 として取り出した文である。表1の通り、言語行為列をシナリオとした方法(2)では、いずれもかなり高い精度で自然な電報文を生成しており、本手法の有効性が示されている。特に文2のようにやや凝った文の場合には、その効果が顕著に現われている。

表1 実験結果(自然な電報文の生成率)

	文1	文2
方法(1) [ランダム生成]	73% (22文/30文)	43% (13文/30文)
方法(2) [言語行為列シナリオ]	87% (26文/30文)	77% (23文/30文)

文1: 北の町の春は、お二人の愛から始まります。
文2: 手に手をとって、ハワイ焼けとはうらやましい。

3. 2. 考察

本稿で提案した手法は3. 1. で述べたとおり電報文の生成に関して有効であることは示されたが、実際に電報文をユーザに提示する場合には、

- ・不自然な生成文を除去する
 - ・生成文候補にスコア付けをし優先順位をつける
- といった処理が必要になる。ここでは、3. 1. の実験結果から、どのような観点でこれらの処理を行なっていけばよいかについて考察する。

今回は、"祝福-陳述-要望"というシナリオについてのみ実験を行なったわけだが、この言語行為列の場合、「祝福」の部分は独立性が高いため、結局「陳述-要望」の部分が電報文全体の自然・不自然を決定している。この部分について検討してみると、以下の2点が重要であると考えられる。

1. 2文間の結束性
2. ユーザモデルの限定度

1.については、従来の「接続詞」や「指示語」を指標とする結束性の解析では不十分である。

(例) 北の町の春は、お二人の愛から始まります。
そんな夫婦になってください。[X]

上記の例では、「そんな」という指示語があっても結束性はない。「そんな」が指す対象が何かを解析できなくては、自然な文か、不自然な文(非文)かは判定できない。かなり深い解析が必要となろう。別のアプローチとしては、電報文において「新婚旅行」「子供」といった特定の話題がよく用いられることから、話題の継続性を解析する試みも考えられる。

2.の「ユーザモデルの限定度」とは、各文がどのくらいユーザモデルを限定するかという度合である。たとえば、

「末長くお幸せに。」

は、送り手・受け手のユーザモデルをあまり限定することなく、ほとんどの場合無難に使うことができる。それに対し、

「帰ってきたらみんなで押し掛けるから、たくさん話題を仕入れて来いよ。」

は、おそらく送り手・受け手は友人の関係にあるというようにユーザモデルが限定され、接続できる文の制約となる。もちろん、1.に関して触れた話題の継続性という観点から、この文の前文は「新婚旅行」の話題でなくてはならないという制約も生じる。

1.及び2.を総合して、電報文をユーザに提示する場合の方法を以下に示す。

- ・「ユーザモデルの限定度が高い文」を用いて、ある程度凝った生成文を提示する。この際には、話題の継続性、ユーザモデルの限定度等からスコア付けを行ない、生成文候補に優先順位をつける。

- ・「ユーザモデルの限定度が低い文」を用いて、無難な文を提示する。

なお、話題の解析手法、ユーザモデルの限定度の数値化等は、今後の課題である。

4. おわりに

文章を構成する各文の言語行為に着目し、文章を言語行為の列として捉え、その言語行為列をシナリオとして文章を生成する手法を提案した。また、本手法が電報文の生成に有効であることを示し、生成された電報文候補のスコア付けの方法についても考察を行なった。

今回は定型文の組み合わせということで議論を進めたが、実際に文生成を行なう場合にも、言語行為列のシナリオを利用することができると考えられる。

電報文作成支援システムの構築を目指し、今後も検討を続けていく予定である。

[参考文献]

- [1] NTT監修: 「メッセージひと800」、電気通信共済会、1991
- [2] 堀井、加藤、大山: 「言語行為に基づくメッセージのセグメント分割」、情報処理学会第41回全国大会(5S-8)、1990