

5S-8

言語行為に基づく メッセージのセグメント分割

堀井統之 加藤恒昭 大山芳史
NTT情報通信処理研究所

1. はじめに

メッセージ、特に電報文を対象として、その効率的な蓄積方法について検討を行なっている。

できるだけ少ない蓄積メッセージから多様なメッセージを生成するためには、蓄積されている複数のメッセージの各部分を組み合わせて新たなメッセージを合成できると都合がよい。これは、たとえば、

(例1) 新郎、新婦の誕生バンザイ。人生は七転八起。二人三脚で長い旅路のゴールを目指そう。

(例2) ご結婚おめでとうございます。北の町の春は、お二人の愛から始まります。末長くお幸せに。

という2つの電報文が蓄積されているときに、これらから、

ご結婚おめでとうございます。人生は七転八起。二人三脚で長い旅路のゴールを目指そう。末長くお幸せに。

を合成できるということである。

このような処理を考えた場合、メッセージをそのまま蓄積するのではなく、新たなメッセージを合成する際の一部分となり得るような単位、すなわち「人生は七転八起。二人三脚で長い旅路のゴールを目指そう。」などのような、意味的につながりを持つまとまりに分割して蓄積する必要がある。

我々はそのような単位のことをセグメントと呼んでいるが、本稿では、各文の言語行為に着目して、メッセージを自動的にセグメントごとに分割する手法について述べる。

2. 言語行為に基づくセグメント分割手法

2.1. 統語的な情報を用いたセグメントの認識

セグメントを認識する方法としては、指示語、接続詞、同語の反復といった統語的な情報を用いて、文間の意味的なつながりを認識する方法がある。

(例3) 幾つ春を数えても、幾つ冬を数えても、二人でいたい。そんな夫婦になってください。

(例4) 右のポケットに幸福を。左のポケットに愛情を。お腹のポケットには赤ちゃんを。

この方法は、例3、例4のように、文間の論理的関係を表すような指標が存在する場合には有効である。しかし、メッセージは書き言葉でありながら口語的な文章であるため、統語的な情報が省略されたり、存在しなかったりすることが多い。

また、蓄積するメッセージをセグメントごとに分割する場合には、その分割点の決定も行なわなくてはならない。

2.2. 本手法の概要

例1の電報文の一部を利用して新たな電報文を合成する場合、第1文は1文だけで利用することができるが、第2文と第3文はまとめて用いなくてはならない。したがって、

第1文/第2文・第3文

(「/」は分割、「・」は未分割を表す)

という形で蓄積しておくのであるが、ここで、各文についてその文の言語行為に着目すると、

祝福/陳述・依頼

である。ここで、接続する2つの言語行為「祝福-陳述」間は分割され、「陳述-依頼」間はつながっているが、他の電報文でも「祝福-陳述」間は分割され、「陳述-依頼」間は未分割であることが多い。すなわち、接続する2つの言語行為の組み合わせにより、その間の分割または未分割が決定できると考えられる。

今回提案するセグメント分割の手法は、このような規則性を利用したものである。

2.3. 文パターンとその決定方法

言語行為の分類については、いくつか提案がなされているが、ここでは山梨[1]によるものを、セグメント分割という観点で整理した、文パターンという概念を用いる。文パターンの分類を表1に示す。

表1 文パターンの分類

文パターン	該当する言語行為
陳述型	陳述、報告
要望型	依頼、祈願、命令、忠告、警告
約束型	約束
祝福型	祝福
感謝型	感謝
陳謝型	陳謝
質問型	質問

各文の文パターンの決定に際しては、言語行為を明示的に特徴づける遂行動詞を含めた主用言が重要な指標となる。しかし、日本語は話者の意図や態度が主用言以外の文末表現で決まることも多く[2]、主用言だけから文パターンを決定するのは困難である。そこで、文パターンの決定は主として文末の用言文節によって行なった。次ページの表2に各文パターンを決定する指標となる文末の形態的特徴の例を示す。

表2 文パターンを決定する文末の形態的特徴

文パターン	文末の形態的特徴例
陳述型	自動詞連用形+ます、用言連用形+ています [てくれます]
要望型	用言連用形+てください、お [ご] +用言連用形 [用言性名詞] +願います、用言連用形+てね、用言終止形+な
約束型	他動詞連用形+ます、用言連用形+てもいい (わ) よ
祝福型	おめでとう (ございます)、お祝い+します [いたします]
感謝型	ありがとう、ご苦労さまです
陳謝型	ごめんなさい、すいません
質問型	文+か [かしら/かね/かな]

表3 セグメント分割ルール

接続する2文間 分割される組み合わせ	
要望型	祝福型
質問型	祝福型
祝福型	祝福型
祝福型	祝福型
祝福型	陳述型
祝福型	陳述型
祝福型	約束型
陳述型	約束型
約束型	約束型

また電報文に現れる話し言葉においては、体言止め、倒置、固有名詞の付加(「～してくださいね、○○さん。」などの場合)なども多くみられるので、それらも考慮した。

2. 4. セグメント分割ルール

セグメントという概念は、「意味的につながりを持つまとまり」という、ある程度主観的なものである。したがって、どのようにセグメント分割すべきかの判断には、個人により多少の差があると考えられる。

そこで、電報例文集から収集した電報文217電文(いずれも結婚の祝電)のセグメント分割結果を決定するために、3人の被験者にその217電文についてセグメント分割を行なってもらった。3人の分割結果が一致したのは、217電文中161電文(約74%)であり、これらに関してはその分割結果を用いることにした。また残りの56電文については、さらに他の3人の被験者にも分割を行なってもらい、合計6人の被験者の分割結果の中でいちばん多かった結果を用いた。

このようにして得られた分割結果と各文の文パターンから、接続する2文間が分割される割合の高い文パターンの組み合わせを、セグメント分割ルールとした(表3)。

3. 評価及び考察

3. 1. 評価

本手法の有効性を評価するために、形態素情報から文パターンを決定する文パターン抽出部と、各文の文パターンが正しく与えられた場合に表3のルールを用いて自動的にセグメント分割を行なうセグメント分割部をコーディングし、実験を行なった。形態素情報は、日英翻訳システムALT-J/E[3]の形態素解析部を用いた。実験に用いた電報文は、2.4.でセグメント分割ルールを作成するために用いたものと同じの217電文である。

この実験の結果を以下に示す。

- 抽出した文パターンの正解率
84.5% (535文中452文)
- セグメント分割の成功率
87.2% (321箇所中280箇所)

「セグメント分割の成功率」は、2文以上からなる電報文を対象に、接続する2文間について正しい分割または未分割が行なわれた割合である。

文パターンの抽出に関しては、形態素情報のみを利用して得られた結果としては、高い正解率である。意味カ

テゴリなどの利用により、さらに高い正解率が期待できる。

また、セグメント分割については、セグメント分割ルールを抽出したのとは別の電報文144電文を用いた机上検討も行なった。その結果、

- セグメント分割の成功率
77.1% (402箇所中310箇所)

が得られ、本手法の有効性が示された。

3. 2. 考察

本手法により分割または未分割に失敗した例から、本手法の改善点について検討する。

(1) 文パターンの細分類

(例5) 結婚おめでとう。夫婦ドライブ、今日から出発。上手に運転してね。末長くお幸せに。

例5は、

第1文/第2文・第3文/第4文
祝福 陳述 要望 要望

が人間によるセグメント分割結果であるが、本手法では第3文と第4文が分割されない。これは、「要望型-要望型」の場合に、分割・未分割の割合がほぼ等しいので、今回は未分割としていることによる。しかし、このような失敗の場合について2文の文パターンを細分類してみると、「依頼-祈願」である場合が多い。したがって、文パターンの細分類、さらにはそれに応じたセグメント分割ルールの決定を行なわなくてはならない。

(2) 統語的な情報の併用

(例6) 独身学校無事卒業、夫婦学校入学、おめでとうございます。なお卒業はありません。

例6は、接続詞「なお」により、第1文と第2文は未分割であるが、本手法では「祝福型-陳述型」のため、分割されてしまう。統語的な情報の併用も必要であろう。

4. おわりに

各文の言語行為に着目したセグメント分割手法を提案し、その有効性を示した。

今後は、3.2.で示した改善点を踏まえて、セグメント分割手法の確立を目指す。

[参考文献]

[1] 山梨正明：「発話行為」、大修館書店、1986
 [2] 久米、吉本：「意図を表す言語形式の語用論的制約について」、情報処理学会第38回全国大会、1989
 [3] 宮崎正弘：「日本文音出力のための言語処理に関する研究」、東京工業大学学位論文、1986