

6G-7

文のおもしろさを決定する 言語的な要因の分析

堀井統之 今村賢治 大山芳史
NTT情報通信網研究所

1. はじめに

電報文のような通信文をデータベースに蓄積し、その中からユーザの要望に合った文を検索しようと考えた場合、データベース中の各通信文にどのような属性を付与しておけばよいかが問題になる。

我々は通信文の検索方式を検討し、通信文にどのような属性を付与すべきかを文献[1]で示した。丁寧度、ユーモア度、送り手・受け手の上下関係といった約20の属性を設定したが、新たな通信文をデータベースに追加するには当然その通信文にすべての属性を付与しなくてはならない。しかし、付与すべき属性が多いため付与者の作業工数がかかる。また人間の主観によって決まる属性が多いため、一定の付与基準が設けにくく、付与者によって異なった属性付与が行なわれてしまう。したがって蓄積したい通信文を計算機で解析し、各属性を自動的に取得できることが好ましい。

本稿では、通信文に付与する属性のうちユーモア度、すなわちその文のおもしろさを計算機によって取得する方法について検討した結果を報告する。

2. 文のおもしろさを決定する言語的な要因

文のおもしろさは一般に主観的に決まるものであり、「何がおもしろいのか」を明確に示すことは難しい。特に通信文の場合、送り手と受け手の共通の話題、知識なども大きく影響するため、不特定多数を対象とした文よりもさらに困難である。

しかし、おもしろい文を生成するという立場で考えてみると、我々はいくつかの言語的な技法を用いることができればある程度文のおもしろさを定量的に表すことができると考えられる。ここでいう言語的な技法とは、言語の表層に現れる情報(形態素情報、構文情報など)を利用する手法のことであり、比喻のような意味の領域に入り込んだ手法は含まない。

以下では、実際に我々が構築した電報文データベース(文献[2]のお祝いの例文が蓄積されている)中の各電報文のうち、付与されているユーモア度が高いもの(1~5の5段階で4、5が付与されているもの)を言語的な側面から分析して得られた結果を示す。(例はすべて1電報文中の該当部分のみである。)

反復現象

1文内または文間で同語、同音、同意味属性が反復して使われている。

- (1) 亭主入門、奥様入門、さらにおめでとう。
(同語、文内)

- (2) 我々はきびしい仕事で徹夜。君は嬉しい仕事で徹夜。
(同語、文間)

- (3) 努力は足し算、家計は引き算、愛情は掛け算、家事は割り算、子どもは安産が良い。(同音、文内)

- (4) どこで彼女をキャッチホン。どうしてハートをプッシュホン。二人で築く、明るいマイホームテレホン。
来年会う時は親子電話。(同意味属性、文間)

(2)は"仕事で徹夜"という形態素列の反復というように考えることもできるし、同一構文の反復と考えることもできる。また(3)、(4)もそれぞれ最後の部分を除けば、"計算"という同意味属性の反復、"ホン"という同音の反復と考えることもできる。

音韻の利用

音の類似(洒落)、七五調が利用されている。反復現象のうち同音反復も音韻の利用である。

- (5) 三三九度は、さんざんくどいた後の約束や、・・・
(音の類似)

- (6) 寒くとも、春は近くに、二人かな。(七五調)

方言、口語表現

通常書き言葉では現れない方言や、口語表現の中でも特にくださったものが用いられている。

- (7) じいちゃん、ばあちゃんに、早よ孫ば抱いて貰うごつ。頑張りなせ。(方言)

- (8) 大人じゃん。(口語表現)

- (9) なんちゃって。(口語表現)

カタカナ書き、カタカナの多用

通常ひらがなや漢字で書くべき部分がカタカナ書きされている。またはカタカナの語が多用されている。

- (10) ついにヤッタネ。(カタカナ書き)

- (11) レモンのようにフレッシュで、ビールのようにコクがあり、ジュースのように甘い家庭を、二人で仲良く、つくってください。(カタカナの多用)

特別な接続詞の利用

"なお"、"ただし"、"ところで"といった接続詞が用いられている。

- (12) 卒業証書。貴方は本日をもって、独身課程を優秀な成績で修了された事を、ここに証します。なお副賞の旅行は、自前で行ってください。本当におめでとう。

(13)ご結婚おめでとう。今夜は頑張れ。ただし仕事のため、スタミナは残せ。

これらの接続詞はその存在自体が文のおもしろさに影響を与えているわけではない。しかし、電報文中で上記の接続詞が使われた場合、その後ろに「オチ」とか「冗談」が現れることが多く、結果的にはおもしろい文になっている。

3. 計算機による実現とその問題点

ここでは、2章で示した各言語現象を計算機によって取得する方法とそれに生じる問題点について述べる。

各言語現象を計算機で取得する場合に必要な情報及びテーブルを表1に示す。字種判定とは使われている文字の種類(漢字、ひらがな、カタカナなど)を判定する機能である。方言テーブル・口語表現テーブルはおもしろさに影響を与えるような方言、口語表現を登録したテーブルであり、また特別接続詞テーブルには2章で示したような“なお”、“ただし”といった接続詞が登録されている。

表1を見てわかる通り、今回の処理では形態素解析部の性能が重要となる。我々の用いている形態素解析システム[3]では意味属性もかなり細かく付与されるので、同意味属性反復もある程度認識できると考えられる。ただし、この形態素解析システムは新聞記事等の解析を目的にしたものであるため、多くの場合方言、口語表現の解析では失敗する。話し言葉用の形態素解析を用いないとすると、形態素解析の前段で方言テーブル、口語表現テーブルを利用した書き言葉への書き換えなどの処理が必要になろう。

表1 各言語現象の解析に必要な情報及びテーブル

言語現象	解析に必要な情報及びテーブル
同語反復	形態素解析結果の字面情報
同音反復	形態素解析結果の読み情報
同意味属性反復	形態素解析結果の意味属性情報
音の類似	形態素解析結果の読み情報
七五調	形態素解析結果の読み情報
方言	方言テーブル
口語表現	口語表現テーブル
カタカナ書き	字種判定 及び 形態素解析結果
カタカナの多用	字種判定
特別な接続詞	形態素解析結果の字面情報 及び 特別接続詞テーブル

2章で示した言語現象の出現によってその文がおもしろいと判断できるかどうかを検証するために、電報文データベース中でユーモア度が低いもの(1~3)の中にそれぞれの言語現象が現れるかどうかを机上で検討した。方言、特別な接続詞は全く現れなかった。口語表現はいくつか現れたが、それは終助詞の“ね”とか“よ”が主であり、くだけた口語表現はほとんどなかった。

しかし、反復現象に関しては、以下のようなケースがいくつかあった。

(14)結婚おめでとうございます。ご結婚にあたり、愛情、理解、信頼、三つの言葉を贈ります。

(15)一才歳をとって、あなたはまたひとつ先輩になりました。

(14)の“結婚”の同語反復、(15)の“才”と“歳”の同意味属性反復(“年齢”という意味属性)のいずれも文のおもしろさとは無縁である。

このような過剰認識は、音韻の利用(音の類似、七五調)を認識する場合にも起こることが予測できる。偶然にある一部分の音が類似したり、七五調になってしまう場合が容易に起こり得る。

本稿で示した反復現象、音韻の利用といった文のおもしろさを決定する要因は、「読んでおもしろいと思う言語的要因は何か」という解析の立場ではなく、「おもしろい文を生成するためにはどのような言語的技法を用いるか」という生成の立場で検討した結果であるため、上記のような問題点が生じたと考えられる。しかし、実際にはユーモアを表そうとして反復、音の類似、七五調を生成する場合にも、何の制約もなくこれらを生成するのではなく、言語的制約^{*1}、意味的制約、またはさまざまな知識による制約が課されている。今回は言語的な現象だけに着目していたため、これらの制約に関しては全く考慮していなかったが、ユーモアを表している部分のみを抽出するためには、このような制約を明らかにし、それらを記述したルールを増やしていくことが不可欠となる。

4. おわりに

電報文を対象として、文のおもしろさを決定する要因について言語的な側面から分析し、いくつかの言語現象をおもしろさの指標として抽出した。まだ粗い分析にすぎないが、文のおもしろさを定量的に表すための第1段階と言える。今後さらに検討を進めていきたい。

[参考文献]

- [1] 今村、大山：「メッセージ検索方式の検討」、1991年信学会春季全国大会(D-109)、1991
- [2] NTT監修：「メッセージひんと800」、電気通信共済会、1991
- [3] 宮崎、大山：「日本語音声出力のための言語処理方式」、情処論文誌(Vol. 27, No. 11)、1986

*1 ここでいう言語的制約とは、たとえば同語反復の場合に、自立語を含んだ語の反復だけに限るとか、(1)(2)の例のように「語と句読点の連続が反復する」といったようなヒューリスティクスを用いるといったような、今回の解析レベルよりもう少し細かいレベルでの制約を指す。