

用言の機能に基づいた修飾語句の分類に関する考察

1M-7

鷲崎 誠司

NTT 情報通信網研究所

1 はじめに

リアルタイム情報を利用者に提供するための情報提供文の自動生成を目指している [1] [2] [3]. リアルタイムに変化する数値情報を文章に変換することを考えた場合、数値の変化に見合う修飾語句(形容詞、形容動詞、副詞)を状況に応じて適切に選択することができれば、少ない文字数で効果的な数値情報の説明文章の生成が可能となる。

従来の修飾語句(程度表現)に関する研究としては、程度表現を5段階の離散的表現と捉えてモデル化したもの [4]、多言語の修飾語句を比較し修飾語句が表す程度の範囲をメンバーシップ関数を用いてモデル化したもの [5]、そして大規模な言語コーパスから形容詞の類似度を自動的に計算し、クラスタリング手法を用いて形容詞を分類したもの [6] などがある。

本稿の目的は、被験者に対して数値情報と程度を表す修飾語句・被修飾語句に相当する用言(動詞、形容詞、形容動詞、助動詞)をペアで与え、与えられた数値に応じて適切な表現を選択させる認知実験を行なうことで、数値情報が与えられた場合の修飾語句が表す程度の範囲を明らかにし、程度表現としての修飾語句の捉えられ方を考察することである。

以下では、実験の概要と実験結果を統計的手法を用いて分析して得られた結果を報告する。

2 実験と結果分析

2.1 実験概要

数値データが与えられた時に、どのような語句選択を行なうかを確認するために、次のような実験を計画/実施した。

概要

被験者に対して数値情報を表形式で与え、表中の数値に関する説明として修飾語句を用いた選択枝を含む文章を用意しておき、それらの中からの最適なものを選択する選択判断実験。

被験者に与える数値情報の分野は、経済情報、電話料金情報、スポーツ情報の3種類を与え、分野間の差異を考慮に入れた。また、被験者に与える文章の総数が多い分、類似した内容の文章が多くなり、回答する際に学習する可能性があるため、パソコンを用いて問題をランダムに与えた。更に被験者が同一問題を再読することで選択の判断が揺れることも考えられるため、一問当たりの回答時間を制限した。

また、パソコンで問題を与えることの対照実験として、問題を別の人により発声させ、被験者にはそれを聞かせることのみで問題を与え回答してもらった実験を行ない、パソコン画面による視覚からの情報入力時との差異を調査した。

実験後には、被験者に対して回答選択の理由をアンケート形式で記入してもらい、回答選択の動機などの分析に用いた。

提示した情報の例

今回の実験では、合計10個の異なる表を用いた。表1に被験者に提示した表の例として、過去10年間の公衆電話数の推移表の一部を示す。

表1: 表の例

年度	街頭公衆電話	点頭公衆電話	100円公衆電話
1981	175,998	481,221	256,705
1982	161,530	439,676	327,049
...
1990	39,837	149,315	638,094

注: 表中の単位は、契約数を表す

提示した文章の例

1990年の街頭公衆電話数は1982年と比較すると、

- [1. やや、2. わずかに、3. かなり、
4. 多少、5. 非常に、6. ...、10. 大幅に]

減少している

被験者に与える文章中の選択枝は、以下のように決定した。まず、事前実験により収集した文章に対して形態素解析を行ない、各分野で多用される用言と修飾語句を抽出する。次に、係受け解析を行なった結果を用いて、修飾語句の係先である用言を確定し実験で用いる修飾語句と用言の候補を決定する。この時点では、修飾語句の中はかなり口語的な表現も含まれているため、各々の用言により一般的に使用される修飾語句を予め絞っておき、その結果を用いて全ての文章に対して同じ修飾語句を利用した選択枝を作成した。

その結果として、実験では用言(増加する、減少するなど)20種類、修飾語句(やや、わずかになど)10種類を用いることとし、一人当たりの総質問数は500問となった。

2.2 分析概要と分析結果

分析概要

以下のステップに従って、用言と修飾語句の関わり合いを分析した。

1. 修飾語句の選択時における数値情報の参照部分を明確にするために、アンケート分析と回帰分析を行なう
2. 各用言はどのような修飾語句の使用ができるのか、更に修飾語句の表す程度の範囲に関して分析するために、判別分析を用いる
3. 用言毎の修飾語句の使用の差異の分析のために、クロス集計を用いて修飾語句の程度の強弱を用言毎に判定する
4. クラスタ分析を用いて修飾語句の程度の範囲に関する傾向に基づいた用言の分類を行なう

A study of the classification of modifiers based on the function of declinable words

Seiji SUZAKI

NTT Network Information Systems Laboratories

分析結果

まず、被験者は与えられる数値情報のどの部分を参照して、修飾語句を選択しているかを分析した。この点を明らかにするために、実験後に行なった各被験者に対して回答に関するアンケートを利用した。このアンケートの結果から、与えられる表を平均や各数値の差異などを判断しながら行なうと予想し、各数値の平均からの差を標準偏差を用いて表した数値と修飾語句の選択結果との間で回帰分析を行なった。

その結果、上記の数値は、各用言において修飾語句の選択結果とかなりの部分で一致することが確認できた ($F=2.53$ 、 $p<0.03$)。このことから、被験者は数値情報が与えられた場合、頭の中で簡単に統計処理を行なっていると想像できる。

次に、各用言に対するデータを判別分析にかけ、その結果から各用言に対する修飾語句の選択結果を大局的に見た結果としては、修飾語句は、修飾の程度によって順序づけることが可能であることがわかった。また、各用言に対して、同程度を表すような修飾語句はいくつか存在する(高い程度を表す「とても」と「かなり」など)。それらは、修飾語句の機能から大きく2つに分類可能である。一つは、特定用言に対して共起確率の低い語句(用言「少ない」に対する修飾語句「少し」)で、もう一つは使用する用言に限定されず一般的に使用されるもの(かなり、多少)である。

このような特定用言に対してその使用頻度に強い相関があるような修飾語句を除き、一般的に使用される修飾語句の使用の程度を基にしてクロス集計を行なった。その結果、多くの用言に対する各修飾語句はそれが表す程度に関して半順序関係を保持していることがわかった。ただ、意味的に近い修飾語句(「かなり」と「とても」など)は、この順序が交差する用言もあり、語句の使用に個人の傾向が入った可能性もある。

最後に、各用言に対する修飾語句の選択に関して、クラスター分析を行なった。修飾語句が表す程度の範囲に従って、用言に関する3つのクラスターが得られた。表2は、各々の用言の特徴から、各クラスターをgroup1、group2、group3と名付け修飾語句の使用パターンをまとめたものである。

表2：修飾語句の分類結果

程度		group1	group2	group3
程度低	level1	まったく ほとんど	まったく ほとんど	まったく ほとんど
	level2	わずかに 少し	わずかに やや わりと	わずかに 少し
⇕	level3	やや わりと	かなり とても	やや わりと
	level4	かなり とても	非常に	かなり とても
程度高	level5	大幅に 非常に	*	大幅に 非常に

注：group2における程度のレベルは4つ

表2において、level1-level5は修飾の程度の強弱を表す。group1とgroup3との間の差異がないように見えるが、実際の分析結果では、修飾語句の程度の範囲に差異が見られたため、異なるグループとした。

また、各グループに含まれる用言の特徴としては、用言が

持つ機能に関して分類したところ、以下のような特徴を持っていることがわかった。

- group1 . 異なる2点の比較に用いた用言
例：増加する、減少する
- group2 . group1と同じだが、中立の意味に用いた用言
例：同じ、同様だ、等しい
- group3 . 連続点の比較に用いた用言
例：変動する、変わる

3 考察

分析の結果として、数値情報を説明するための用言が与えられた時の修飾語句の選択は、用言が持つ機能を基に3種類に分類したのに対して、修飾語句はその選択パターンとして4~5つのグループに分けられるものやそうでないものなどいくつかの選択基準が明らかになった。

この結果は、過去の研究において形容詞(修飾語句)を5段階の程度に分類した結果とよく一致する。ただ、group2だけは4段階の程度になり、多少異なっている。これは、実験の精度の問題もあるだろうが、全ての修飾語句を一様に同じ数に分類するだけでは、うまくいかない部分がある可能性があることを示唆していると考えられる。

また、今回の分析によって修飾語句の選択に関しては、異なる程度を表すものに関しては、使用する用言と修飾する程度がわかれば使用するべき修飾語句の一群がわかった。今後の課題として、修飾語句の程度がわかった時に、同じ程度を表す修飾語句の中から、最適な語句を選択する必要があると考えられる。これを解決するためには、分野に依存した語用論や、語句の使用履歴情報などを参考にすることが考えられるだろう。

4 まとめ

本稿では、修飾語句に関する認知実験を行なった結果を報告した。その結果、修飾語句が表す程度表現は、用言の機能毎に4段階から5段階と離散的に理解されていた。

今後は、修飾語句を用いた程度表現の生成に関して検討を行なう予定である。

参考文献

- [1] 大内他：情報提供文生成のための個別情報抽出、電子情報通信学会春季全国大会 6-112(1991)
- [2] 鷺崎他：わかりやすさを考慮した情報提供文生成手法の検討、電子情報通信学会春季全国大会 6-113(1991)
- [3] 鷺崎他：情報提供文における接続表現の特徴分析、情報処理学会第44回全国大会 3-95(1992)
- [4] 亀井他：程度表現のモデル化、電子情報通信学会言語理解とコミュニケーション研究会、NLC88-6(1988)
- [5] 勝守他：自然言語の意味測定によるファジイ命名関係、情報処理学会第43回大会 3-107(1991)
- [6] Hatzivassiloglou et al. : Toward the Automatic Identification of Adjectival Scale: Clustering Adjectives According to Meaning, in *Proceedings of the 31th Annual Meetings of ACL*, pp172-182(1993)