

定性的な意味構造の分析サイクル実現を目指したシステムの提案

Proposal of a Cyclic Analysis System of Semantic Structure

レボウィッツ 紀子*¹

Noriko LEBOWITZ

松村 敦*¹

Atsushi MATSUMURA

宇陀 則彦*¹

Norihiko UDA

*¹筑波大学大学院図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

In order to extract the semantic structure, it is necessary to apply not only a quantitative analysis but also a qualitative analysis. In this paper, we propose a system that analyses text from both aspects. The system performs four cyclic processes, and makes a semantic structure model. The cycle consists of qualitative analysis, rulemaking, quantitative analysis and creating a model. The model can be revised according to the result of the cyclic analysis, if it is inappropriate.

1. はじめに

意味は言語によって表現されるが、表現されたテキストの意味構造をその表層構造のみから得ることはできない。なぜならテキストには文脈があり、それは多くの場合に言語では表現されないものだからである。このため、テキストが内包する意味構造を機械的に抽出することは非常に困難な作業となる。一方、人は文脈を加味した上で、テキストの意味を理解することができる。つまり、テキストの意味構造を得るためには、文脈を読み取ることでできる人の関与を欠くことはできない。

そこで本研究では、人による定性的分析と機械による定量的分析を組み合わせた意味構造分析サイクルを提案する。本研究で実現を目指す定性的な意味構造分析サイクルとは、テキストが内包する意味構造を手により定性的に抽出・分析し、それをもとに作成したルールにそってシステムを用いた定量的な分析を行い、その結果をもとに意味構造をモデル化し、さらにモデル化された意味構造を人が評価し修正を行うという一連の繰り返しを指す。この意味構造分析サイクルをシステムとして実装することで、定性的特徴を反映した客観性のある意味構造の抽出を目指す。

2. 意味構造分析の方法

ここで、分析の対象となるテキストの意味構造と文脈の関係について整理しておく。文脈とは、テキスト全体に整合性を持たせる内的な要素であると考えられる。Halliday は、テキストを単に文の連続ではなく意味的に整合性のある体系として保証するものとしてテキストの構成要素間の結束性をあげている [Halliday 94]。この点から、結束性と文脈には相互関係があることが予想される。つまり、結束性を用いてテキストを分析することが、文脈を考慮した分析にもつながると考えられる。

結束性を用いたテキストの意味構造の分析方法には、大きくわけて以下の2つがある。

- テキストの構成要素間の局所的な関係の分析から始めて、テキスト全体の構造を明らかにしていくマイクロ分析
- テキスト全体に対する視点から、文や節などの構成要素が果たす機能やその関係を捉えるマクロ分析

結束性を用いたテキストのマイクロ分析には、構成要素間の接続関係・係り受け関係・修辞関係を分析する手法が確立されている。しかしテキストとは、あるアイデアについて語られる体系としてみるべきであり、その体系のもつ意味構造は、構成要素間の関係を局所的に分析した結果から得られるものとずれがあると考えられる。なぜなら、体系としてのテキストにおいては構成要素間の局所的な関係に加えて、構成要素とテキスト全体との関係まで考慮されるべきであり、マイクロ分析のみではその関係を十分に分析することができないからである。したがって、マクロ分析の視点からテキストが有する意味構造の抽出を実現する方法が必要となってくる。

そこで本研究では、テキストの各構成要素がテキスト全体に対してどのように機能し、それらの機能はテキスト全体とどのような結束性を持っているのかを表現する意味構造モデルを生成し、それに基づいてテキストをマクロ分析する手法のシステム化を提案する。

3. 意味構造の分析サイクル

3.1 分析サイクルの概要

本章で述べる意味構造分析の目的は、テキストをマクロ分析するための指標となるモデルを生成することである。ここでいうマクロ分析とは、文や節などの構成要素がテキスト全体に対して果たす機能やその結束性を捉える分析方法のことである。

表1にHallidayが提唱する5種類の結束性を示す。また表2にテキストの構成要素の果たす機能を示す。ここでは旅行案内を例とした。ここでいう構成要素の果たす機能とは、表層的にはテキストを構成しているパーツである。たとえば、観光スポットについて記述されたテキストを見ると、スポット自体に関する情報・スポットの所在地やアクセス方法に関する情報・スポットについて知るための付加的な情報から構成されている。しかし、これらは単なる表層的なパーツとしてテキスト上にあるだけでなく、各パーツが意味的に結束することによって、テキスト全体としての意味構造をつくりあげているため、これらのうちいずれかが機能しなくなると、体系としての意味構造が崩れることになる。本研究では、Hallidayの提唱する結束性をもとに、これら機能間の関係を分析するサイクルを形成する。

連絡先: レボウィッツ紀子, 筑波大学大学院図書館情報メディア研究科, 茨城県つくば市春日 1-2, nlebo@slis.tsukuba.ac.jp

表 1: Halliday の提唱する結束性

| | |
|-------|----------------------|
| 照応 | テキスト中の意味的な参照関係 |
| 代用 | 代名詞などによるつながりの関係 |
| 省略 | テキスト中では省略されている意味的な関係 |
| 接続 | 明示された接続表現による意味的な関係 |
| 語彙結束性 | 同一語や上位語などの語彙的な関係 |

表 2: テキストの構成要素が果たす機能例

| |
|---------------------------|
| 観光スポットに関する記述 |
| —主要部：スポットに関する主要な内容 |
| —地理的情報：スポットの場所やアクセスに関する情報 |
| —付記：スポットに関する付加的な情報 |
| —地理的：地理的な背景情報 |
| —歴史的：歴史的な背景情報 |
| —言語的：言語的な付加情報 |
| —慣習的：慣習的な付加情報 |

3.2 各フェーズにおける作業

本研究で提案する意味構造分析サイクルの概念図を図 1 に示し、分析サイクルの各フェーズにおいて行われる具体的な作業と、それに対応する提案システムの機能について述べる。

定性的分析

定性的分析とは、人によるテキストの意味構造分析のことである。このフェーズにおける具体的な作業は、テキストからの命題の抽出 命題の果たす機能分析 命題間の結束性の分析である。ここでいう命題とはある文が表現している意味とし、実際に抽出するものは意味を成す文という単位とする。また、定性的分析において行われる結束性(照応, 代用, 省略)の処理および接続と語彙結束性を機械的に処理するためのルール作成へとつながる規則性を見出す必要がある。

この定性的分析に対応するシステムの機能については、分析対象とするテキストの読み込み、命題抽出のための単文の切り出しの仕組みに加え、命題の機能タイプおよび命題間に関する情報のタグ付けし、それらをデータベースに蓄積する仕組みが必要である。

ルール作成

ルール作成とは、前段階の定性的分析で得たデータをもとに定量的分析のためのルールを作成するフェーズである。作成されるルールの具体的な内容には、命題の機能タイプ、命題間の結束関係の定義がある。さらに、定性的分析によって抽出した命題、機能、結束関係についてのスキーマ作成を行うことと、命題間の接続と語彙的結束性を機械的に抽出するためのルールが必要である。

定量的分析

前段階で作成したルールに則り、多くのテキストを定めた視点で処理するフェーズが定量的分析である。このフェーズの最大の役割は、命題間の接続と語彙的結束性の抽出と、意味構造モデルの初期値を得るための機械処理によるデータ作成である。

この定量的分析に対応するシステムの機能については、テキストの読み込み、命題抽出のための単文の切り出し、情報タグ

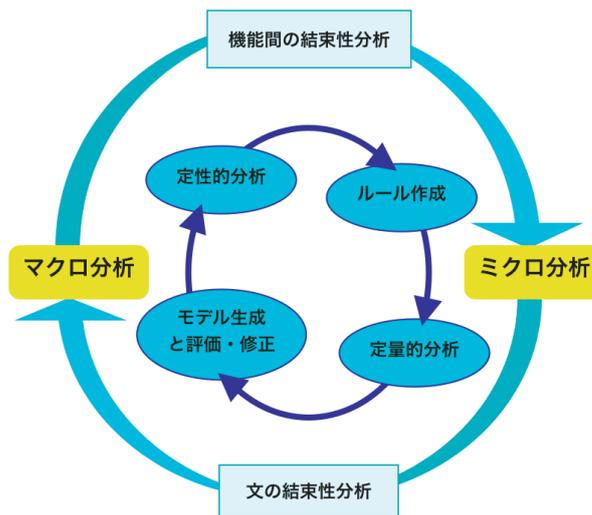


図 1: 意味構造分析サイクル

の付与に加え、データとして蓄積した命題と命題間関係をグラフ構造で可視化し、分析結果を命題マップとして提示する仕組みが必要である。その上で、複数の命題マップをマージし、意味構造モデルの初期値を生成するフェーズへとつなげる。

モデル生成・評価・修正

定性的分析結果および定量的分析結果のデータをもとに、意味構造モデルを生成し評価するフェーズである。生成したモデルは人によって査定され、修正が必要な点について検討され、修正が施された上で、マクロ分析のための意味構造モデルとして次のサイクルにおける定性的分析に用いられる。

定量的分析によって得られたデータの量は定性的分析によって得られたもの比べ圧倒的に多くなるので、命題間の結束性の抽出も接続と語彙結束性によるものが多くなるのが予想される。したがって、意味構造モデルの初期値に反映される結束性は偏りの強いものと思われる。モデル修正のための機能としては、モデルに反映されている結束性の偏りを明示的に示す仕組みが必要である。

4. おわりに

本稿では、テキストの意味構造の分析サイクルとそのサイクルを基に意味構造モデルの生成を実現するシステムの提案を行った。ここで生成する意味構造モデルとは、テキストを構成する各文の意味を抽出しつなぎ合わせたものではなく、テキストが内包する体系的な意味構造を形式化するものである。全体論的見地からすると、このようにテキストの意味構造をマクロな視点から分析することは非常に意味のあることだといえる。今後は、提案システムを実装し、システムを用いて実際にテキストの意味構造を分析していくことで、具体的な機能の検討などを行っていきたい。

参考文献

[Halliday 94] Halliday, M. A. K.: An Introduction to Functional Grammar, Edward Arnold (1994), (邦訳: 機能文法概説, 山口 登・笈 壽雄 訳, くろしお出版 (2001)).