

## 文の論理意味表示の視覚化とその応用

吉本 啓<sup>1)</sup>\*, アラステア・J・バトラー<sup>1)</sup>

1) 東北大学高度教養教育・学生支援機構

### 要旨

筆者たちはこれまで、自動文意味解析について研究し、それにもとづいて日本語や英語のテキストに対して文統辞・意味解析情報をタグ付けした新しいタイプのコーパス構築法を確立して、実際に開発を行ってきた。その成果を利用しやすくするためにしている文の論理意味表示の新たな視覚提示法の開発とそれを利用した外国語教育等への応用法の研究について経過を報告する。

### 1 はじめに

従来、形式統辞理論を用いて複雑な構文をコンピュータ解析しようとする試みが行われてきた(Clark et al. 2010 を参照のこと)が、十分な成果は挙がっていない。これに対して筆者たちの方法では、表層的な統辞情報をバトラーの提案する形式意味理論であるスコープ制御理論(SCT; Butler 2010, Butler 2011)を実装したシステムに入力することにより、論理意味情報(述語論理式)の形で深い文法情報を出力することが出来る(Butler and Yoshimoto 2012, 吉本他 2013)。名詞句の指示情報を表すインデックスも不要である。理論の検証のために大量の日本語データの意味解析が必要であるが、言語処理に広く利用されている「京都大学テキストコーパス」(Kurohashi and Nagao 2003)を加工して利用しようとしても文節の壁があまりに高く、他に大量の日本語データの入手も困難なことから、ついに自らツリーバンク(統辞・意味解析情報付きコーパス)の構築を開始した。

筆者たちが開発してきたツリーバンクの特色は以下

の通りである。「京都大学テキストコーパス」は文節を基本的単位として、それらの間の修飾情報も含めて統辞情報を付加したコーパスである。しかし、文を文節に分割してそれらと関係づけるやり方では文の意味構造は捉えられない。一例を挙げると、

- (1) a. 日本へ行きたいとトムは言っている。  
b. その道を行くと彼に会った。

の2つの文において、下線部分はともに述語の終止形に助詞「と」が付加されたものである。しかし、意味構造の観点から見ると、最初の文の下線部分は引用された発話内容であり、述語「言う」の項として埋め込まれるのに対し、2番目の文における「と」は条件表現のオペレーターとして2つの文をつなぐ役割を果たしている。上記の統辞情報だけでは、このような意味論上の重大な違いを捉えることは出来ない。

同様の例として、2種類の関係節—寺村の(1975)言う「内の関係」と「外の関係」—を挙げることが出来る。

- (2) a. 昨日撮った 写真  
b. 子供が泳いでいる 写真

上の2つの名詞句はともに関係節による修飾が行われているが、(2a)では主名詞が関係節の中で述語「撮る」の目的語の格役割を果たす「内の関係」となっているのに対し、(2b)はそのような役割が認められず、関係節が主名詞の内容を表す「外の関係」の例である。論理表示においては、ここでも後者は関係節の内容が

\*) 連絡先: 〒980-8576 宮城県仙台市青葉区川内41 東北大学高度教養教育・学生支援機構 kei@compling.jp

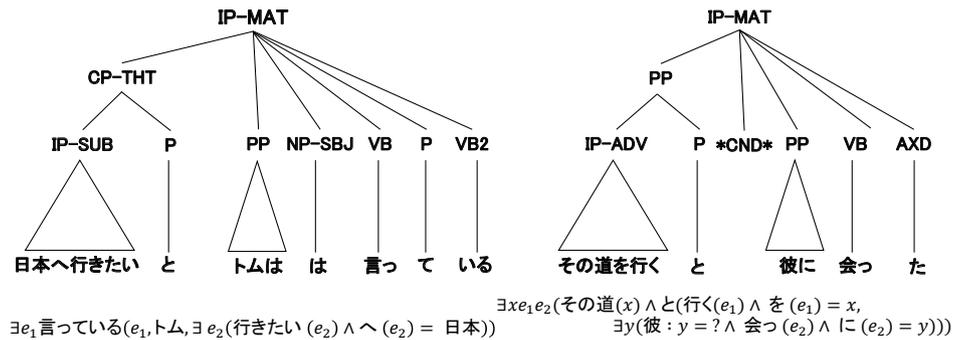


図1：文(1a, b)の統辞・意味解析

主名詞の意味に埋め込まれるのに対し、前者では連言結合子  $\wedge$  により両者が結合されるという対照をなしている。

要するに、日本語に関して現在利用可能なコーパスでは、言語形式の生起や言語形式間の共起関係 (co-occurrence) については知ることが出来るが、統辞・意味情報の基礎をなす語句間のリンク情報である依存関係 (dependency) についての情報は得られないし、言語研究者がそれらを使っても結局は人手による処理が不可欠であり、大量のデータ処理には利用できない。

これに対し筆者たちの方法では、日本語・英語の文に対し表層統辞解析情報をタグ付けし、それをSCTシステムで処理することによって、文の種類を問わず論理意味表示を出力できる。これにより得られる語句間の依存関係は、様々な形で利用することが出来る。しかし、意味情報自体は述語論理式で書かれており、その理解には論理学の知識が必要である。せっかくの情報をより広く活用できるようにするため、意味関係の視覚化を思い立った。

## 2 研究の目標

図1に上記(1a, b)の2つの文のコーパスにおけるアノテーションを示す。a文の統辞解析木では、後置詞(P)である「と」は節「日本へ行きたい」を従えて引用節(CP-THT)を構成し、全体が動詞「言う」の項の1つとなっており、論理式にもこのことが反映されている。これに対してb文では「と」の導く従属節が条件節であることがノード \*CND\* の付加によって示され、これにもとづいて論理式では全体が条件文となり、従属節の意味はその前件に相当する。

新しい形のコーパス開発の結果は、機械翻訳等の自然言語処理技術として利用することが出来る。また、法律情報処理や外国語教育等、関連分野への影響も大きい。例えば、外国語テキストの自動意味解析は、様々な形で学習者のサポートや学習支援システムに利用できる。このように一定の精度の文自動意味解析を可能にしたことの意義は理論的にも応用面でも大きい。残念ながらその成果はまだ十分には発揮されていない。そのもっとも大きな理由は、特に日本の文系言語研究者や学生の中に論理学を十分に理解する人が稀で、せっかくの論理意味情報を活用できないことにある。そこで、本研究では、論理意味情報の中核をなす、語句間の依存関係情報を処理し、視覚表示して利用しやすくすることを課題とする。

第一に試みるのは、学校教育で行われていて馴染み深い「係り受け」の表示である。(3a, b)は前ページの(1a, b)文の係り受け関係を矢印によって表示したものである。助詞「と」の導く節「日本へ行きたいと」が動詞「言っ」に引用格の成分として依存(後者を修飾)していることが示されている。

- (3) a. 日本へ行きたいと トムは 言っ ている  
└──────────┬──────────┘ ↑ 引用格
- b. その道を行くと 彼に会った  
└──────────┬──────────┘ ↑ 条件

これに対して(3b)では、節「その道を行くと」が節「彼に会った」に係っており、しかもそれが条件節であることが明示されている。両者の違いは、このように矢印による修飾・被修飾関係の表示で表すことができる

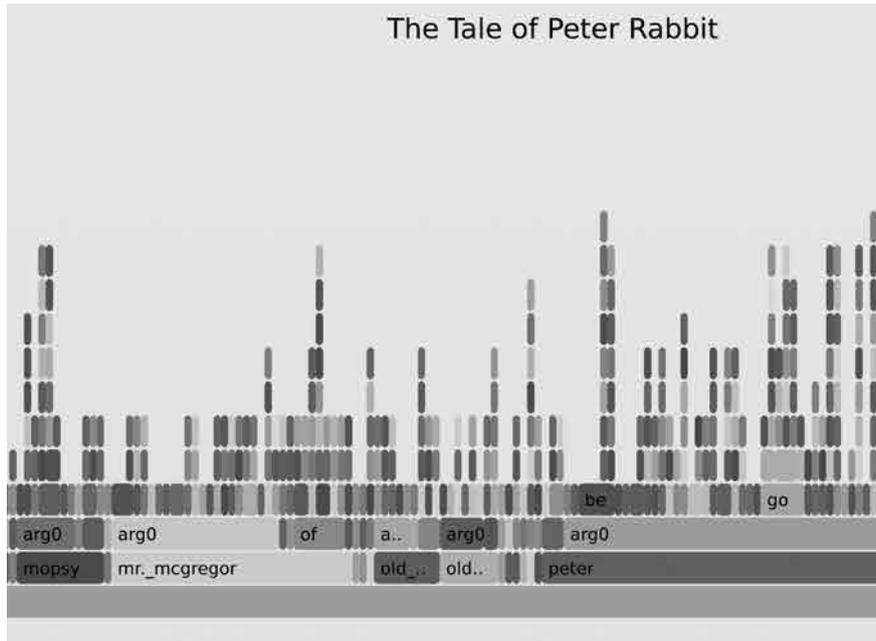


図 2：フレーム・グラフの例

が、この他にも、コンピュータ・ディスプレイ上で当該の句にマウスのポインタを置くと「引用格; 修飾先: 言っ」や「条件節; 修飾先: 彼に会った」と表示するなどの工夫が考えられる。なお、文節間の係り受けのみが表示されるのではない。(3a, b)では、係り先は文節ではなく、単語や節である。また、文節文法では扱うことのできない、関係節中の動詞と主名詞との関係も扱うことができる。さらに、「内の関係」と「外の関係」とを区別して表示できる。

第二の課題は、コンピュータ・システムのスタック・トレースのデータを表示して管理するための手段である Flame Graph を利用した意味解析情報の視覚化である (Butler and Yoshimoto 2014)。図 2 に、約 1000 の単語からなる *The Tale of Peter Rabbit* (Beatrix Potter 著) の内容をフレーム・グラフとして表示した結果の一部を示す。図の一番下の層には、Peter, Mr. McGregor 等の固有名詞が並んでいる。そのすぐ上には arg0, arg1 等が表示されているが、これは、述語の第 1 格 (主格), 第 2 格 (直接目的格), 等ということであり、最下層の名詞が文の中で主語や目的語等、どのような格役割を果たしているかを示している。下から 3 番目には、go, catch 等、名詞が直接依存する動詞が並んでいる。さらに上層には、下部の名詞を支配する動詞や役割が順番に並んでいる。このように、個体

や述語など一定のグループの語ごとに依存関係の階層を作って文章の意味を再構成し、直感的に理解しやすくすることが出来る。なお、印刷では示せないが、本来はセルごとに赤, オレンジ, 黄色等に色分けして表示される (色分けはランダムに行なわれ, 意味は無い)。

フレーム・グラフのより具体的な利用法を、次の文を例に取って考えてみよう。

(4) 花子は果汁を凍らせて, デザートを作った。

これはコントロール構文の一種であり、文の主語「花子は」が従属節の「凍らせる」および主節主語「作った」によって共有されている。これらの「花子-凍らせる」や「花子-作った」におけるような依存関係もスコープ制御理論を利用した意味解析情報のアノテーションによって初めて把握することが出来る。図 3 に

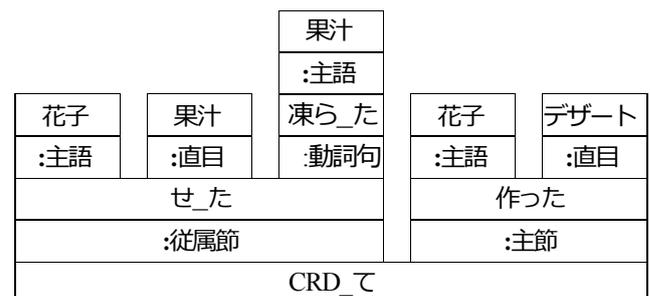


図 3：例文 (4) の意味のフレーム・グラフ

例文(4)の意味のフレーム・グラフによる視覚化の一例を挙げる。ここでは、述部の意味を下にして、その上に格情報を介して名詞句等の依存関係を視覚化してある。図2の :argo や :arg1 等は、ここでは:主語、:直目(直接目的語)など、より分かりやすい表現に言い換えてある。このような表示により、動詞句や助動詞等の用言の結合価(格依存関係)が一目瞭然である。

以上、論理意味情報にもとづく係り受け表示およびフレーム・グラフにもとづく視覚表示という2つのテーマに関して、その理論的・技術的困難を克服し、研究者や学生の利用を可能にすることが本研究の課題である。

### 3 本研究の特色・意義

語句間の係り受け関係視覚表示は、関係節を含む非有界依存(unbounded dependency)や照応等の深い意味情報を分かりやすい形で提供するという特色がある。しかもそれは、深い意味情報をタグ付けしない表層的な統辞解析情報をSCTシステムに入力することにより得られるので、どのような言語テキストからも容易に出力することが出来る。したがって、外国語教育における読解のサポートや法律の条文のような難解・複雑な文章の理解の促進等の応用に適している。

フレーム・グラフを利用した視覚表示は、文だけでなくテキスト全体を対象として、名詞句の意味(個体)等から出発したボトム・アップ的な、直感的に理解しやすい意味表示を行うという利点がある。これを使えば、どのような登場人物がどのような行為をなしたのか一目瞭然である。また、この手法は言語データからの文法情報の抽出に利用することも出来る。例えば、図3に示したやり方で、データ中の全動詞の各々について格フレーム情報を出力することが簡単に行える。従来の形態素情報と文節にもとづくコーパスでは、日本語の格助詞・係助詞が表示する文法格が曖昧であることと、コントロール構文や関係節等の非有界依存構文が存在するため、自動的な抽出は不可能である。

さらに、近年他にも文自動意味解析の研究が行われ、隆盛の兆しがあるが(例えば、Bos 2008)、同一の意味を持つ論理式でも変項を初めとして多様な表記が許されるため、式自体を検索したり必要に応じて処理す

ることは難しい。また、それらの間の比較もきわめて困難である。フレーム・グラフによる表示は依存関係を上下階層として単純化して示すので、異なる表記法の論理式を標準化することになり、検索、処理や相互の比較を容易にする。これにより、論理意味表示をともなうコーパスから高度の言語情報を得ることが可能になり、また文意味解析の様々な理論・方法に関し優劣を論じるための共通の基盤を提供できる。

### 4 研究の方法

これまでに確立した論理意味表示の自動解析(Butler and Yoshimoto 2012)にもとづいて日本語ツリーバンクを構築し、係り受けおよびフレーム・グラフという2種類の視覚表示法を開発し、ツリーバンクの意味情報の視覚提示に利用する。また、これらにもとづいて日本語テキスト読解実験、異なる自動意味解析法の比較と評価、および日英語の構文の差異の類型論的検討を行う。

提案する研究では、筆者たちの文自動意味解析法を利用して、本研究の目的に合致した日本語テキストを選び、各文に対して統辞・意味解析情報をタグ付けする。さらに、得られた意味解析情報より係り受け関係およびフレーム・グラフという2種類の視覚化情報を抽出するための方法を確立する。

テキスト中の文は、まず統辞解析プログラムにかけられる。プログラムとしては、主として、無料で入手可能な形態素・統辞解析器を利用する。その解析結果に対して人手による修正を行って、タグ付けのための文統辞解析情報が得られる。また、統辞解析結果は構造変換プログラムへと入力され、スコープ制御理論(SCT)に従った中間表示へと変換される。SCT意味解析システムはこの中間表示を入力とし、出力として意味表示(述語論理式)が得られる。図4を参照のこ

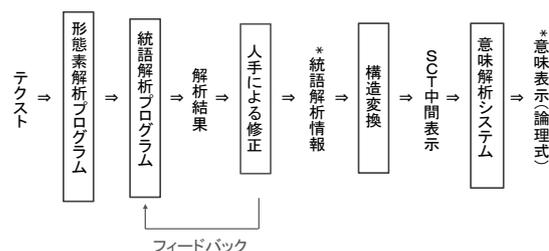


図4：ツリーバンク開発の過程

と、述語論理式を含む依存関係の情報にもとづき、係り受け関係およびフレーム・グラフの視覚情報を構築する。

開発に当っては、Annotation Manual for the Penn Historical Corpora and the PCEEC (Santorini 2010)の方針に従う。これは世界におけるツリーバンク研究開発のさきがけとなったPenn Treebankの解析規約を修正したものであり、フラットな統辞構造の採用と、名詞句・動詞句・節等に対して主語・目的語・関係節等の様々な機能を付加することを特色としている。動詞句、名詞句、形容詞句など句の種類を問わず、X-バー理論に従って同一のフラットな構造として解析される。句のヘッド (N, P, ADJ等) がつねにそれと同一カテゴリーの句 (NP, PP, ADJP等) を投射する。このように木構造を簡単にすることによって、その検索や変換がきわめて簡単に行われる。またスコープの出現順位に従い、最初のものほど広いスコープを持つというデフォルトのスコープ包含関係を設定し、これに反するもののみを記述することによって、柔軟なスコープ包含関係の指定を可能にする。また、句や節に対するより正確な機能情報の付加により、統辞構造の曖昧性を克服することが出来、述語-項関係にとどまらない、より複雑な構文により表現された意味情報の抽出が容易になる。

筆者たちの自動意味解析においては、語句間のすべての依存関係が与えられる。そのため、これらすべての係り受け関係を表示するときわめて煩瑣なものになり、一般人の利用が著しく困難になる。そこで、利用者にとって必要なものを中心として視覚化し、他を省く必要がある。日本語教育の観点からは動詞、名詞のような実質語、および助詞・助動詞といった機能語がヘッドとなって構成される句や節を主として取り扱うのが適切である。さらに、ゼロ代名詞を含む照応情報はこれまで十分にタグ付けできていないが、文章の理解に不可欠なので、そのタグ付けおよび視覚化の課題についても解決する。

フレーム・グラフの図2の例は物語中にあらわれる個体を最下層に置き、各々の個体ごとにそれと依存関係にある格役割、述語、オペレーター等をスタックとして示したものである。最下層には個体以外にも、何

でも選ぶことが出来、文章の意味の柔軟な抽出や視覚化を可能にする。例えば文法に関しては、助動詞や従属節の種類ごとにその内部に表れる構成素を表示し、南 (1974) の文階層説を検証することが出来る。この点で様々な表示法を試してみれば比較し、理論言語学的に、また応用面でそれぞれがどのように貢献できるかを検討する。

フレーム・グラフの抽象的意味表示はまた、同一の意味の形態・統辞論的な手段が言語間で異なる場合 (例えば、英語の that 節に対する、日本語のコトによる名詞節)、両者の共通性を捉えるのに利用することが出来る。日英語間のそのような事例について、共通の表記法を検討する。格フレームやオントロジー関連のデータベースも利用する。

## 5 研究計画

意味表示の目的に供する、約2万文からなるツリーバンクを構築する。そのノウハウは、これまでの研究を通じて (Butler and Yoshimoto 2012, 吉本他 2013) 十分に蓄積されてきている。まず入手したテキストを形態素解析器および Probabilistic Context-Free Grammar にもとづく統辞解析器にかけ、文の文節への分割およびそれらの間の依存関係に関する情報を得る。これをさらに、木構造の検索・変換用ツール *tsurgeon* のスクリプトを用いて句構造へと変換する。その結果のほとんどすべてについて人手による修正が必要であり、これが全体の中で主要な作業を占める。得られた統辞解析結果は、構造変換を経て、SCTシステムに入力されて論理意味表示 (新デイヴィッドソン式表示) が得られる。その結果は、統辞解析情報の修正へとフィードバックされる。

得られた論理意味解析情報について、どのような文法情報を選んで係り受け情報を表示すべきかについて考察し、適切と思われる文法情報表示の候補を数件選ぶ。これらによる日本語テキストの係り受け関係の視覚表示を与えたテキスト読解実験を日本語学習者に対して行い、理解度のテストやアンケートにより、最適のものを選ぶ。

フレーム・グラフについては、最下層の意味グループとして採用する、個体、述語、オペレーター等の意

味的次元ごとに日本語に特有の構文や機能語の差異を除去していく。これにもとづいて、異なる意味解析法や異なる言語間の比較を行う。

## 6 結論

統辞・意味情報付きコーパスの成果を利用しやすくするために、論理意味表示を視覚化して提示する研究の経過について報告した。今後、テキスト読解実験を通じて最適のシステム構築を行う予定である。

## 引用文献

- Bos, Johan. Let's not argue about semantics. *Proceedings of the 6th Language Resources and Evaluation Conference (LREC 2008)*. Marrakech, Morocco. 2008.
- Butler, Alastair. *The Semantics of Grammatical Dependencies, Current Research in the Semantics/Pragmatics Interface*, Volume 23, Emerald. 2010.
- Butler, Alastair. Semantically restricted argument dependencies. *Journal of Logic, Language and Information* 20 (1): 69-114. 2011.
- Butler, Alastair and Kei Yoshimoto. Banking meaning representations from treebanks. *Linguistic Issues in Language Technology - LiLT* 7 (6): 1-22. 2012.
- Butler, Alastair and Kei Yoshimoto. Semantic visualisation with Flame Graphs. *Proceedings of the Eleventh International Workshop of Logic and Engineering of Natural Language Semantics*, pp. 205-215. JSAI International Symposia on AI, the Japanese Society for Artificial Intelligence, Ochanomizu University and Keio University. 2014.
- Clark, Alexander, Chris Fox and Shalom Lappin. *The Handbook of Linguistics and Natural Language Processing*. Wiley-Blackwel. 2010.
- Kurohashi, Sadao and Makoto Nagao. Building a Japanese parsed corpus - while improving the parsing system, A. Abeille, ed., *Treebanks: Building and Using Parsed Corpora*, chap. 14. Kluwer Academic Publishers. 2003.
- Santorini, Beatrice. Annotation Manual for the Penn Historical Corpora and the PCEEC (Release 2). Tech. rep., Dep. of Computer and Information

- Science, University of Pennsylvania. 2010.
- 寺村秀夫「連体修飾のシンタクスと意味その1」『日本語・日本文化』4, 外国語大学留学生別科, 71-119, 1975.
- 南不二男『現代日本語の構造』大修館書店, 1974.
- 吉本啓, 周振, 小菅智也, 大友瑠璃子, Alastair Butler.「日本語ツリーバンクのアノテーション方針」『言語処理学会第19回年次大会発表論文集』, pp. 924-927, 名古屋大学, 2013.