

# 基本語彙英語に基づく意味理解システムの試み Towards a Meaning Comprehension System Utilizing Basic English

良峯 徳和

Norikazu Yoshimine

多摩大学グローバルスタディーズ学部  
School of Global Studies, Tama University

This paper describes an attempt to construct a meaning comprehension system which utilizes Basic English. Basic English, developed by Ogden about 80 years ago, could be a model for a language system that allows computers and human users to exchange information. The core vocabulary of Basic English is deliberately chosen based on an ontological analysis where almost every event and relationship can be described. The system aims to simulate some aspects of the compact, yet powerful, functions of Basic English on semantic network systems such as SNePS.

## 1. はじめに

人間が誰でも日常的に使い、母国語が異なる人との間でも十分に意思疎通がはかれるような種類の言語で、しかもコンピュータがその言葉の意味をある程度処理できるような言語ははたして可能だろうか。インターネットをはじめとした情報のデジタル化と情報通信技術の急速な進歩に伴い、世界中で蓄積される情報量が膨大化し、従来の検索・抽出技術では有効性を維持できなくなる可能性が生じてきている。さらに人間の世界でも、社会全体の急速なグローバル化に伴い、世界中の誰とであれ、ある程度の訓練によって、相互に意思疎通し、情報交換できる実際的な国際語の流通が熱望されている。こうした要請に答える形で、近年急速に活性化しつつあるプロジェクトがある。それは人間のための制限言語をコンピュータに適用する試みである。

本研究では、Basic English などの制限英語をもとに、人間のみならずコンピュータにとっても、意味に沿った自然言語の処理を可能にする言語理解システムの構築を目指した基礎研究、とりわけ人間が行っているような、コンテキストに沿った深いレベルの意味理解を可能にする基本語彙レキシコンの構築、および処理システムの設計に関する試みを行っている。

## 2. 制限言語に基づく既存の自然言語理解システム

こうしたシステム構築を目指した試みは、すでにいくつかのプロジェクト(UNO, Attempto Controlled English(ACE), UNL, CDL.nl など)で実施されている。いずれの試みにおいても、制限された特定の語彙内で簡単な構文の英語文を入力すると、その意味に応じた適切な返答を、自然言語の文章として出力するという自然言語による意味理解応答システムの要求をある程度満たすようになっている。これにより自然言語をインターフェイスに用いた人間とコンピュータとの情報交換が可能になり、コンピュータやネットワーク上に蓄積された膨大なデータベースへのアクセシビリティやユーザビリティが、大きく向上する成果が期待できる。さらに、そうしたシステムに各国語に対応した語彙辞書や文法規則を組み込むことで、将来は各国語の制限言語同士の間でも、コンピュータを媒介とした通訳、翻訳、コミュニケーションといった応用可能性が開かれるだろう。

ただし、構文レベル、語彙レベルの両方のレベルについて言えることであるが、機械での処理のしやすさのみを考慮して、制

限を厳しくしすぎると、自然言語としての表現力、表現範囲が大幅に損なわれ、実際の使用には耐えられない乏しい言語体系になってしまう。これに対し、ACE の場合は、利用可能な語彙数が約 10 万語、UNL は約 20 万語とかなり大規模な語彙数を持っている。対応可能な構文については、単文や重文に対してはほとんど問題なく処理できる一方で、複文については現状で対処できるパターンが少数に限定されるなどの制約がある。

対応可能な構文パターンを増やすことに比べ、利用可能な語彙数を増やす作業は、ユーザによる比較的単純な手続きによって可能とされている。例えば、ACE の名詞語彙の場合、singular, plural の区分と、nouns, mass nouns, propernames, measurement nouns の区分がされていれば、使用可能な語彙として登録される(e.g., noun\_sg gallery; noun\_mass concussion)。その際、各語彙が持つ他の語彙や表現との包含関係や同義、反意関係などの情報は含まれない。単に、構文的に処理できる語彙の範囲を増やす手続きがとられているにすぎない。こうした語彙間の関係をシステムの処理に組み込むためには、語彙の使用や置き換えに関するさまざまな規則を設定しなくてはならないが、ACE の場合、1階の述語論理に準拠した規則が適用できると同時に、こうした規則の入力についても、制限言語の範囲で自然言語構文を用いた入力ができるようなインターフェイスを備えている。ただし、現在のところ、規則入力のインターフェイスにおける構文的制約が厳しく、期待通りの機能を発揮できるところまでには到達していない。

一方、Iwanska [00]が試みた UNO の場合、自然言語理解における推論の役割を重視し、とりわけスカラー表現と呼ばれる言語表現によって含意もしくは随意する前提内容を、システムに取り込む試みを行っている。また、UNL や UNL の拡張版である CDL.nl の場合には、自然言語表現をカバーする実用レベルの基本語彙(記述に必要な関係語彙セットも含む)オントロジーをあらかじめ備えており、たとえば「著者」と「人間」、「著者」と書かれた「書物」との関係を理解し、その関係に適応した処理を行えるようなシステムを組み込んでいとされる。

しかしながら、システムが扱うことができる語彙すべてについて、当の語やそれを含むフレーズ間の関係をオントロジーや知識ベースとして構築し、システムに組み込むことは、実際問題としてきわめて困難といわなくてはならないだろう。こうした語彙やフレーズ間の関係を記述するためには、構文知識や語彙知識のみならず、ある程度の常識的な世界知識までも包含しなければならない。もし UNL が約 20 万語の語彙を扱おうとしても、それらの表現すべてについて、上記のような概念的関係を付加するとすると、その組み合わせは天文学的な数字となってしまふ。

連絡先: 良峯徳和, 多摩大学グローバルスタディーズ学部,  
〒252-0805 神奈川県藤沢市円行 802, Tel: 0466-82-4141,  
Fax: 0466-82-5070, email: yoshimine@tama.ac.jp

こうした困難を克服する方法として、大規模コーパスなどを利用した語彙やフレーズ関係の機械処理による抽出が有効な手段とも思えるが、コーパスの対象となるテキストの種類や質によって、その結果に不必要な情報が含まれたり、必要な情報が含まれない可能性が残る。システムが汎用的に利用するレキシコンとしては、語彙やフレーズを使用する文脈や状況に関わらない概念的な含意や論理的、必然的關係のみを、データとして含むべきであり、偶発的事実や間違った記述に基づくデータをその中に取り込むべきではない。得られたデータがどのようなテキスト記述に基づいていたかを逐一取り出し、評価する必要性が再び困難な壁となるだろう。

### 3. 制限言語としての Basic English の特徴

こうした困難を打開するひとつの方策として、本研究が取り上げるのは、自然言語理解応答システムのモデルとなる制限言語の基本語彙を可能な限り、必要最少限にとどめることである。そのようなモデルの最有力候補として、Basic English の制限言語体系を挙げてみたい。

Basic English は、The Meaning of Meaning(『意味の意味』)の共著者の一人である言語学者 Ogden が 1930 年に発表した制限英語体系である。それは、構文的には一般の英語と変わらないが、使用する語彙数がわずか 850 語に限定されている。語彙数が 850 語に制限されているために、多少は回りくどい表現が用いられたり、微妙なニュアンスを表現することが困難とはいえ、日常生活に必要な範囲については、通常英語とほとんど変わらない内容を語るができる。特別な構文規則や語彙を導入せず、VOA の Special English などと同様に、Basic English によって表現された文章は通常英語とほとんど区別がつかないため、通常英語の部分集合として扱うことが可能である。

その最大の特徴は、第 1 に、徹底した語彙数の削減である。Basic English では、ある表現がそれと同じ意味をもった別の表現で置き換え可能な場合には、そのうち、基本概念としての性格が強いより汎用性をもった語彙のみを残し、それ以外の語彙を削除していくという徹底した分析的手法を用いて、語彙数の制限を行った。こうした意味の原子要素としての基本語彙の考え方は、同時代に数学の原理をわずかな原子要素に還元して再構築しようとした Russell と Whitehead の試みや、Wilkins や Leibniz, Bentham など Ogden が強く影響を受けた先駆者の哲学思想に裏付けされたものである。

Basic English の第 2 の特徴は、いわゆる動詞の排除である。Basic English では、基本動作語として認められている 16 個の動詞以外は語彙から排除され、20 個ほどの方位詞と呼ばれる方向を示す前置詞や副詞、および対象を表す名詞などと組み合わせ、4000 語程度の動詞の働きを可能にしている。

第 3 の特徴は、メタファーとして使いやすい語彙を基本語彙として選定している点である。語彙数が厳しく制限された場合、表現の原意あるいは文字通りの意味によって指し示すことのできる対象範囲はどうしても狭くなる。そのため、Basic English ではその制約を言葉のメタファー的使用によって克服し、より豊かな表現の可能性をもたせようとしている。言葉のメタファー的使用を容認することは、語彙や語彙の組み合わせに対し、状況に合わせて異なる意味を持つことを認めることであり、自然言語処理システムのモデルとしては、好ましい要素とはならない。したがって、何らかの制約を加える必要がある。

4 つ目の特徴は、語彙の品詞転換の最大利用である。動詞 cry を名詞の cry として用いたり、名詞の fire に分詞形の接尾語-ed をつけて、形容詞 fired として扱えるようにするといった工夫である。これにより、基本語彙はわずか 850 語でも、品詞を

めて実質的な語彙数を大幅に拡張することができる。しかしながら、一般に自然言語処理システムでは品詞が異なれば、異なる意味や機能をもつ別の言葉として扱わなくてはならないため、人間のユーザには当然と思われるこうした恩恵は得られない。

以上のように、自然言語理解応答システムのモデルとしては、必ずしも望ましくない側面も有する Basic English の制限言語体系であるが、にもかかわらず Basic English をモデルとすることを正当化する理由としては、以下の点があげられよう。

- 動詞や用言の数が制限されているため、構文のパターンが単純化され、統語的な分析に伴う曖昧さの削減、誤解釈の可能性を低減させる。
- 動詞や用言の数が制限されているため、文の意味を語またはフレーズ間の論理的関係として取り出しやすい。
- Basic English の語彙は基本的にそれ以上単純な別の語彙によって置き換えることの出来ない基本語彙すなわち無定義語であるため、語の意味は内包的定義に依存しない。したがって通常言語の辞書等に見られるような語彙の定義間の循環的關係や不整合は存在しない。
- 言語体系としての強力な表現力を有している一方で、全体の語彙数が制限されているため、語の意味を他の語彙との関係すなわち外延的關係を用いて規定することが、可能となる。語彙間の関係を意味ネットワークなどで関連付けることで、語彙相互の意味連関を記述したオントロジカルな機能も重ね持ったレキシコンの構築が可能になる。
- Basic English の語彙間の相互關係を規定したレキシコンが作成されれば、言語体系に対する新たな語彙の追加は、それを既存の語彙に関係づける(既存の語彙で定義づける)だけでよい。

### 4. 今後の課題: Basic English をモデルとした自然言語理解システムの試み

以上のような特徴を持つ Basic English の語彙体系を、意味理解システムの基本語彙レキシコンのモデルとして採用することで、これまでの制限言語を核とした自然言語理解システムでは試みられてこなかった語彙の意味關係を相互ネットワークに基づいて規定するレキシコンの構築を行っている。インプリメンテーション環境としては、1階の述語論理を用いて意味關係を規則化することに優れた SNePS を用いている。Basic English をモデルとすることで、制限言語を用いた既存の自然言語処理システムに比べ、大幅に少ない語彙数で日常的な使用に耐えうる意味処理を行えるシステムの構築が可能になると期待される。

#### 参考文献

- [ACE] <http://www.ifi.unizh.ch/attempto/>  
 [相沢 95] 相沢佳子: ベーシックイングリッシュ再考, リーベル出版(1995).  
 [Basic English] <http://ogden.basic-english.org/>  
 [CDL.nl] <http://www.instsec.org/tr/>  
 [石塚 06] 石塚満: 自然言語テキストの共通的概念記述, 人工知能学会誌, Vol.21, No.6, pp.691-698 (2006.11)  
 [Iwanska 00] Iwanska, L. M.: Natural Language Is a Powerful Knowledge Representation System: The UNO Model, in L. M. Iwanska, and S. C. Shapiro(eds): Natural Language Processing and Knowledge Representation, The MIT Press 2000), chap.1, pp.7-64.  
 [Ogden 23] オグデン.C & リチャーズ.I(石橋幸太郎 訳): 新版意味の意味, 新泉社(2001).  
 [UNDL] <http://www.undl.org/>