

WRAPL による機械翻訳のためのテキスト編集と翻訳精度の向上

佐良木 昌

佐良木技術翻訳事務所・機械翻訳研究室

加藤 輝政 小川 清

名古屋市工業研究所

I 英語複文構造のパターン

本稿では、複文構造について、関係節や分詞節などを含む複文に限定してパターン整理を行う。これら複文では、従属節の統語構造が不完全であり、構文解析は困難を伴う。その克服の方途を探る。

複文の分類を以下に示す。この分類は、複文における節同士の結合度という観点を導入し、その結合度に段階的な強弱レベルを設けることによって行う[1]。副詞節を導きその先頭に位置する従属接続詞を含む複文では、その接続詞を分岐点として、主節と従属節とが分離されている（主節内に挿入された従属節は除く）。これを複文のプロトタイプとし、その主節・従属節の結合強度を標準とする。またプロトタイプ複文は、パラタクシス parataxis¹に変換することができる。すなわち、主節・従属節を単文に還元すると共に、従属節相当の後続単文の文頭に、接続副詞を付加することによって文同士の結合関係を表すことができる²。

α 構文明示的な複文

- A. 限定用法の関係節 関係節と主節内の先行詞とは不可分離の統語構造であり、結合度は大変強い。
- B. *It that* 節構文 *It that* 構文は、主節の一部を *that* 節の前に切り出した形であり、結合度は標準より強い。
- C. 相関係節 主節内の先行副詞・形容詞と相関従位接続詞とを介して節同士が相関関係にある *so...that* ~, *such...that* ~などの構文を、相関係節と仮に呼ぶ。プロトタイプ複文と同じくパラタクシスに変換することができるので、結合度は標準である。
- D. 継続用法の関係節 主節内の先行詞や文先行詞について、追加的に説明を加えるにすぎないので、結合度は最も弱い。

β 構文暗示的な複文 (従属接続詞を伴わない)³[2]

- A. 分詞構文 接続詞を伴う従属節は主節との時間的または論理的関係を明示するが、このような結節が際立つのを避けた表現である。主節・従属節が一つの流れをなす表現であることから[3]、結合度は標準よりやや強い。
- B. 形容詞用法で後位用法の分詞を含む文 一般に、実詞の「一時的状態を記述している」[4]ので、結合度は継続用法関係節と同等とする。関係節表現の反復を避けるために用いられる。
- C. 補部を伴う後位用法の形容詞句を含む文 限定用法関係代名詞節ほど意味限定が強くなく、実詞の一時的な特徴的属性を叙述する表現である[4]。このことから、結合度は継続用法関係節と同等とする。
- D. 比較級反復節 *the + 比較級... the + 比較級... 等*の構文をいう。節間の結合を明示する接続詞を欠くものの、文の形式的配列自体が節同士の意味的相関を示す構文である。結合度は標準より弱い。

各パターンと結合度とに応じて、構文形式の意味特性(結合関係)を表現しうる接続副詞や照応関係を明示する語句を補うことで、複文を接続副詞・語句を介した単文結合へと再編する。この接続副詞・語句の活用で、構文形式のもつ情報を保持しながら原文を単文結合に忠実に編集することが可能となる。

本稿ではβ-C パターンの解析と、wrapl による当該複文の編集処理および編集処理によって機械翻訳の精度がどのように向上したか、について報告する。

II 実詞に後続する形容詞節の認識論的意味

英語では、実詞(名詞、代名詞)を修飾する形容詞節が当該実詞に後続する構造が知られている。英語構文別の認識論的構造は、宮下[5][6]が明らかにしているが、本稿では、この解明に踏まえ、未解明であった、補部を伴う後位用法の形容詞句について分析する。

このβ-C パターン構文では、先行実詞において概括的抽象的表現がなされ、形容詞節において具体的再表現が行われる。しかし、限定用法関係代名詞節ほど意味限定が強くなく、実詞の一時的な特徴的属性を叙述する表現である。後位用法の分詞節と同様に、関係節表現の反復を避けた表現である。関係節と共に用いられるときには、関係節の重複や、接続関係や主語従語関係の繰り返しを避けるという表現的意味がある。

¹ 英語においては、疑問代名詞を従属節構文に充当することによって、単文の結合から複文への発展を遂げてきたのであるが、この複文の歴史的形成立ちは、ラテン語における複文成立の歴史的経路を模写し反復したものとして捉えることができる。この意味で、プロトタイプ複文の呼称を用いる。[複文は、言語の歴史においては比較的に後世に発展した。原初の状態では、別々の文が単に平置され、文間の論理的な関係が明確に示されことはなかった。この単なる並置は、パラタクシスとして文法家には知られていない。たとえば、*"I know: he's coming"* この種の英文では、二つの文の間で休止がなく、そのため十分に、文が一つの全体に融合しかつ後置文が前置文に対して従属していることを示している。] LRPALMER [1906] "THE LATIN LANGUAGE" Chapter V "THE DEVELOPMENT OF THE LITERARY LANGUAGE"

² 第2回年次大会で報告した英語構文パターンに、従属接続詞を伴わない「複文」(βB・C)を加えて、パターンを再編成した。

³ 文献[2]では、文のつながりを用いない文型(つながりなしはゼロ記号で表される)をα (doing, done, to do) β (doingly, done-ly, adj-ly, adj, noun, prep. phr) の二パターンに整理している。これらパターンも追加・整理し複文パターンを網羅することが、今後の課題である。

また、文の階層構造を明示的に示す場合にも用いられる。たとえば、限定用法関係代名詞節において、当該実詞の恒常的な態様を表すと共に、当該実詞を構成する各要素については、後位形容詞用法の分詞節や後位用法の形容詞句によって、その態様を表す。例えば、特許英文にその典型をみる(1)。また、属性を規定する際に、 β のA・B・Cの三パターンそれぞれの認識論的特徴を生かした表現も見られる(2)。

(1) A valve assembly for metering exhaust gas to the intake of an internal combustion engine has an electromagnetic solenoid actuator having a magnetic circuit *which* includes a stationary pole piece, an armature *moveable* within said pole piece, and an air gap *separating* the pole piece and the armature.

(2) Metals, such as, platinum (Pt), titanium (Ti), or tantalum (Ta) may be co-deposited with polysilicon *forming* a polysilicide (PtSi, TiSi, TaSi) *which* has a resistivity *that* is approximately one order of magnitude *less* than polysilicon. ¹

III wrapl によるテキスト編集

第2回年次大会で報告したテキスト処理専用言語WRAPLによる、機械翻訳のためのテキスト処理の実際について報告する。複文の編集によって、機械翻訳の精度がどのように向上したかの実験例を報告する(IIで述べた β -Cパターンに限って報告する)。

A ルール生成プロセス

英語においては、形容詞は名詞の前に置かれてその名詞を直接形容する場合が多いが(ゲルマン語的構造)、名詞の後に置かれることもある(ラテン語的構造)[7][8][9][10][11]。後者は、関係節が簡略された結果として形容詞が名詞の後に置かれて後置形容詞になったと捉えることができる。この構文では、実詞と形容詞(句)との統語関係が明示されていないので、構文解析が難しくなる。特に、実詞と補部を伴う形容詞との間に、形容詞にかかる副詞類が存在するときには、構文解析が難しく翻訳精度が低下する一因となっている。換言すれば、後置形容詞のある文章を簡略化されたと予測される元の文に戻す処理を施すことによって機械翻訳の精度を高めることができる。

しかし、形容詞が名詞の後に置かれている場合がすべて複文の簡略表現ではない。このため、形容詞が名詞の後に置かれる場合として以下の場合を検討した。

[後位形容詞の特徴]

(1) -thing, -body, -one で終わる複合型の代名詞を修飾するとき。

Nothing strange happened.

(2) -able, -ible, で終わる形容詞がall, any, every,等を伴う名詞に強意に添える場合。absolutely, entirely, purely, singularly, so, still, too, almost等の副詞をともなって意味を強めながら実詞の後に位置する場合。

every means imaginable, principles purely pacific

(3) フランス語などの影響による場合

the sum total, third person plural

(4) 補部を伴って長くなる場合

a friend worthy of confidence

(5) 叙述用法に近い場合(後位のとき、意味が変わる形容詞もある)

a baby asleep, literature proper

(6) 分詞形容詞に動詞の性質が強い場合

Nick was among the twenty people arrested.

(7) 不完全他動詞の目的格補語として使われる場合

(4)の場合は、複文の簡略型として、元の複文を予測して文を分割処理することで機械翻訳の精度向上が期待される。(7)の場合は、第4文型として使われるわけであるので分割処理できない。上記のほかの場合は、形容詞単独で句構成を取らないことがほとんどである。名詞句として単語登録することで翻訳精度の向上が期待できるので、編集の対処からはずす。

[不完全他動詞の種類]

不完全他動詞として用いられる動詞には、

(1) "make group"

make, get, let, render, set, appoint, declare, elect

(2) "paint group"

bake, burn, beat, color, cut, dye, lick, wash, wipe

(3) "think group" think, consider

(4) 知覚動詞 heard, saw, hear

等がある。

[名詞の同定]

後置形容詞を選別するためには、形容詞の前にある単語が名詞か否かを判別する必要がある。wraplは、現状では、辞書を参照して品詞を特定することはできない。そこで、名詞には、単語の語尾・接尾辞に特徴を持つものがあるので、その特徴を抽出して名詞の一部を特定するルールを記述する。

(1) -ment, -tionなどを除けば名詞である。(ただし、functionなど動詞、名詞が同形については、個別に処理をする必要がある。)

¹ ① Metals, may be co-deposited with polysilicon ② forming a polysilicide ... ③ which has a resistivity ④ that is magnitude less than polysilicon 主節に直属する現在分詞の形容詞節②では、金属とポリシリコンとによって、物質ポリシリサイドが合成されるといふ、この条件の下での物質の性格(恒久的な無条件の性格ではない)を叙述する。続く関係節③④では、合成物質ポリシリサイドについて、具体的性質を説明しているが、限定用法用いているのは、電気抵抗をもつのはポリシリサイドの常態的性格であることを、表現するためである。最初の関係代名詞節(which)③で、その特性一般(has a resistivity)を説明し、第二の関係代名詞節(that)④では、その特性の具体的程度(approximately one order of magnitude less than polysilicon)を説明している。less than polysiliconは後位用法の形容詞句であり、ポリシリコンとの相対的比較における抵抗率の大きさを表現している。絶対的な性格規定ではなく相対的な性格規定である。

² oil producing countries(= countries which produce oil, color-rich patterns (patterns which are rich in color) 複合形容詞が名詞の前にあって修飾しているときには、目的語-動詞、名詞-形容詞の語順となり[12][13]「石油(を)産出(する)国」、「色の豊かな模様」という、日本語と同じ語順になる。これは、古英語期における従属節SOV型の文型に対応する。修飾句が後続するときは、SVO型の文型に対応する。詳細な報告は、別添に行いたい。

(2) -ty は dirty, empty, misty, nasty, pretty, sporty, etc と数字の twenty 等を除けば名詞である。

(3) -or は or, nor, for, major, poor, prior, superior, etc を除けば名詞である

[後位形容詞の同定]

後位形容詞を判別するために、その形態的特徴を整理する。特に、形容詞語尾の特徴抽出が重要である。

(1) 単一語 adjcent, apart, apparent, different, difficult, equal, equivalent, essential, inferior, inherent, pposite, separate, superior, worthy¹等

(2) 名詞 + -ic, -al, -ical, -ar, -ary, -en, -ous, -ish, -y, -ive, like, -ed

(3) 名詞、動詞 + -ful, -less

(4) 動詞 + -able, -ible

(5) 名詞、形容詞 + -ly

形容詞の検索・特定、はこれらの特徴を用いることで可能になる。

[形容詞の後に置かれる前置詞]

後置形容詞はその後に前置詞を伴うことが多い。このために後置形容詞を検索するのに形容詞の後に前置詞があることを調べるのは有効な手段である。形容詞に後続する前置詞には、形容詞の補部を構成する前置詞と、形容詞を修飾する副詞用法の前置詞句とがある。形容詞が補部をとるのは、inherent in, inferior to, superior to, worthy of などがある。

B ルール記述の方法

テキスト編集のルール記述一般論について、まず述べる。自然言語は生動的であり可変的であるので、自然言語の多様性に対応できるルール構成・開かれたプログラム構造が必要だと考える。大規模コーパスによる検証をふまえてルール可変的な構造を構築する必要がある。すなわち、コーパスにより①ルール適用の除外例や、②挿入句を含む文型、変形構文例を発見しながら、その除外のルールや、変形構文などを処理する特殊ルールを追加できる構造とする。一般的ルール構造を示す。

R_{exclusive} ルール適用を除外する構文の指定ルール

R_{specific} 構文変形についての特殊ルール

R_{generic} 一般ルール

wrapl のルール記述では、**R_{exclusive}** と **R_{specific}** とを同一ルールとして記述できる。例えば、部分的に句構成が同じ構文について、ルール適用を除外するルール **R_{exclusive}** を **R_{specific}** に盛り込むことができる (図、C項)。このように、テキスト編集のルール構造は、除外ルール **R_{exclusive}**、特殊ルール **R_{specific}**、一般ルール **R_{generic}** の三層を成す。

C ルール記述の一例

R_{generic}

S+V+O : V ≠ Causative Verb, O = N+Adj+ prep ~ の文型 のとき、

S+V+O → S+V+N;

The above N is/are Adj + prep ~

という構成でルール I を記述する。

1000={10 The|A|An} (11)+(20 ~make*|~making|~made|render*|~,)+{30 a|an|the} {31 *tion|*ment} {40 *ly} ? {41 inherent} {42 in} {43 ~,)+/¥

10, 11, 20, 30, 31, {;¥n}, {The above }, 31, { is }, 40, 41, 42, 43=1001?1001

波線部分が **R_{exclusive}** に相当する。

R_{specific}

S+V+O: S=N + Adj + prep ~, V = be-Verb

の文型 のとき、

S+V+O → N is/are Adj + prep ~;

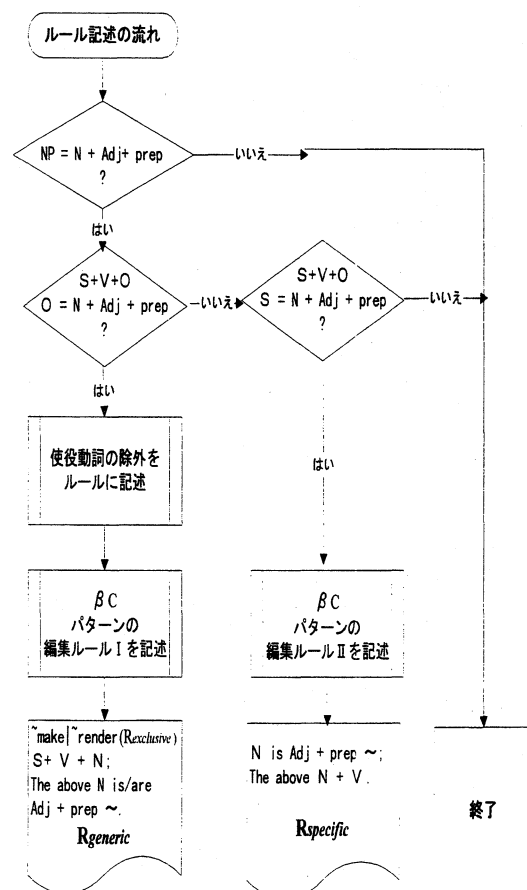
The above N is O.

という構成でルール II を記述する

1001={10 The|A|An} (11)*{12 *tion|*ment} {40 *ly} ? {41 inherent} {42 in} {43 ~,)+{50 is|was|has} {been} ? (60)+/¥

10, 11, 12, { is }, 40, 41, 42, 43, {;¥n}, {The above }, 12, 50, 60=e?e

図



¹ worthy のみがサクソン語源である。他はロマンス語系

IV wrapl によるテキスト編集の実際¹

原文 1

The key cylinder 32 and bezel 30 may be of substantially conventional designs, which will typically include an actuating portion 36 rotatably movable upon insertion of a key 38 into the key slot 34 and rotation thereof in known lock and key manner.

キー・シリンダー32 とベゼル 30 は、挿入にキー38 のキー・スロット 34 とその回転に、周知のロックとキー方法の中で、典型的に回転可能に移動できる動作部分 36 を含む実質上、従来の設計でありえる。

編集文 1

The key cylinder 32 and bezel 30 may be substantially conventional designs;

The above will typically include an actuating portion 36 ;

The above portion 36 is rotatably movable upon insertion of a key 38 into the key slot 34 and rotation thereof in known lock and key manner.

キー・シリンダー32 とベゼル 30 は、実質上、従来の設計でありえる；

上記は、典型的に動作部分 36 を含む；

上記部分 36 は、挿入にキー38 のキー・スロット 34 とその回転に、周知のロックとキー方法の中で、回転可能に移動できる。

(継続用法の関係節も単文に編集されている)

原文 2

they subsist on forage on marginal lands generally unsuitable for sustained farming.

訳文

それらは、持続させられた飼育に不適当な一般にぎりぎりの土地の上で、飼料によって生きて行く。(それら＝家畜)

編集文 2

they subsist on forage on marginal lands;

The above lands are generally unsuitable for sustained farming.

それらは、ぎりぎりの土地で飼料によって生きて行く；

上記土地は、一般に持続させられた飼育に不適当である。

原文 3

However, these private networks, even the most sophisticated, still suffer from the high cost of leased lines in most countries and the difficulties inherent in attempting to create global networks based on a patchwork of services subject to widely varying capabilities and regulation.

しかし、兵卒がネットワークにのせる、最も多くのものは精巧にしたこれらは、大部分の国の専用回線の高いコストとパッチワークに基づいてサービス主題の広く能力と規制を変化させることにグローバル・ネットワークをつくらうとすることに生まれつきの困難で、まだ苦しむ。

編集文 2

However, these private networks, even the most sophisticated, still suffer from the high cost of leased lines in most countries and the difficulties;

The above difficulties are inherent in attempting to create global networks based on a patchwork of services;

The above services are subject to widely varying capabilities and regulation.

しかし、これらの個人のネットワーク(最も洗練されたものでも)は、大部分の国と困難の中で、まだ専用回線の高いコストで苦しむ；

上記困難は、サービスの寄せ集めに基ついてグローバル・ネットワークをつくらうとすることの点で固有である；

上記サービスは、広く能力と規制を変化させることに従う。

(Example 1 is cited from cited from U.S. Patent, examples 2 from Oxford MicroConcord Corpus, and example 3 from NII Agenda for cooperation.)

【参考文献】

- [1] 佐良木・黒野[1996]「構文の節結合強度に応じた書換え規則によるテキスト処理およびテキスト処理専用言語 WRAPL の開発」言語処理学会第2回年次大会発表論文集 pp. 161-164
- [2] 竹内祐治 [1977]「統語形式の二極性—テニヲカの統語形式と非テニヲカの統語形式」英語英米文学第十七 pp.157-178 中央大学英米文学会 「統語形式の基本型—[基本単位]に基本文型のパターンが提案されている。
- [3] 竹内祐治 [1986]「統語論的クライマックス指標—E Hemingway の短編を中心に」pp. 26-32 英語表現研究第3号 日本英語表現学会、[1990]「higher wages」などのネクサス性をめぐって」同第7号
- [4] 安井・秋山・中村 [1976]『辞源語訳』現代の英文法7 研究社
- [5] 宮下眞二 [1982]『英語文法批判』日本翻訳家養成センター
- [6] 宮下眞二 [1985]『英語はどういう言語か』季節社
- [7] 小野・中尾 [1980]『英語史I』英語学体系8 大修館書店
- [8] 中尾・児馬 [1990]『歴史的にさぐる現代の英文法』大修館書店
- [9] 河原・春田 [1992]「形容詞の配列順序の事例研究」『telos』第8号 金沢経済大学人間科学研究所
- [10] Karl Bruner [1960, 1962] "Die englische Sprache: Ihre geschichtliche Entwicklung" 邦訳書名『英語発達史』大修館書店
- [11] 佐良木昌 [1996]「ラテン語関係節の史的発展の論理分析(要旨)」言語・認識・表現 第1回研究会論文集 LACE No.1 pp. 87-95
- [12] Varie Adams [1973] "An Introduction to Modern English Word-Formation" English Language Series Vol. 7 Longman (『現代英語の単語形成』杉浦・岡村共訳 こびあん書房)
- [13] 竹内祐治 [1978]「複合形容詞のSyntax」英語英米文学第十八集 pp.157-181 中央大学英米文学会 [1987]「Adjectivalsと欠落情報」同第二十七集 pp. 323-344

¹ 機械翻には、PC-Transer/ej (株式会社ノバ) を使った。