

連語辞書の構築による機械翻訳の訳質改善

佐良木 昌

日本大学

古賀 勝夫

クロスランゲージ

あらまし 英語における修辞技法は修辞学として体系化されていることから、自然言語処理の背景的知識として活用できる。対照・列挙・前辞反復などという修辞技法により創案された対句は、等位接続型 A and/or/but B などの固定的な形式を採っており、特許英文に多用されている。固定形式により統一性が強まって反義語や同義語の対句という修辞的連語となると共に、文法的には一つの品詞として機能する。また特許英文では、前置詞単体がきわめて多義であることから、その多義性を一意に絞るために複合前置詞が、愛好されている。等位接続型などの一般形式や複合前置詞の構成パターンを用いて、特許英文から修辞的連語や複合前置詞を大量に抽出することができ、それらを辞書として編集し機械翻訳システムに実装することで、特許英文翻訳の訳質を改善可能である。

キーワード 修辞技法、対照・同義語畳用・列挙・前辞反復、複合前置詞、機械翻訳

Compiling the dictionary of rhetorical collocations and complex prepositions in English patent documents

Masashi SARAKI

Nihon University

Katuo KOGA

Cross Language Inc.

Abstract We have retrieved a lot of rhetorical collocations and complex prepositions, using the formulas of rhetorical expressions in English patent documents, and have compiled the dictionary of the rhetorical phrases. Our attention in this paper is focused on the method of vocabulary-building for rhetorical phrases and complex prepositions in the patent documents. Rhetoric includes many rules which can control all prose composition and many techniques for expressing all matters relating to beauty or forcefulness of style. Writers prefer in patent rhetoric to invent rhetorical phrases such as *Antithesis*, *Catalog*, *Synonymy*, or *Anadiplosis*. Alliteration and end rhyme are the repetition of sounds in positions close enough to be noticed. Another aspect of such rhyming means that the words within the rhetorical collocations have the same affixes which can be used for the purpose of retrieval of the rhetorical collocations. By implementing the dictionary on a machine translation system, the system will be able to increase the quality of the translation thereof.

Keyword patent rhetoric, rhetorical collocations, complex prepositions, antithesis, synonymy, catalog, anadiplosis, machine translation

1. はじめに

1.1. 法律と特許英文

特許英語は技術英語の一種ではあるが、例えば、米国特許制度では、特許明細書を法律文書としており、クレームの解釈は裁判所が行うこともあり、米国特許明細書は法律文書として取り扱われている。あるいは、「特許英語は技術英語と法律英語の複合乃至混血」[1]とも考えられている。

欧米の特許文書においては、近代論理学の適用を背景とした上位概念・下位概念に基づく語彙の使用といった論理学的修辞を観ることができる。たとえば、文献公知のように、ポッツのクレーム理論[2]における周辺限定主義／中心限定主義をめぐるクレーム論議は、論理学論争の様相を呈していたし[3, 4]、

ポッツが指摘するようにクレーム定義方法を巡る論議は、J. S. ミル[5]と W. ヒューエル[6]の論理学論争を源とする、属と種差とによる定義法を論理学的背景としている。

1.2. 特許英文における論理と修辞

このような論理学的厳密性は、たとえば米国特許書式に反映されている。クレームを本質論とすれば、明細書、とりわけ *embodiment* の記述は具体的現実論であり発明実施の現実性がなければならないと規定されている。論理学的厳密性は、特許英語表現にも反映している。例えば、クレーム記述を表す言葉は *recite* であり、明細書の記述を指す言葉は *description* であると言われている [7]。 *recite*, *description* は、ともにラテン語起源であるが、前者はもともと

法律文書において事実を述べるという原義であり、後者はフラン語から借用され言葉で説明するという原義である。原義から見て、この二語の使い分けには一般—個別の論理、換言すれば属(genus)—種(species)の論理系統樹に基づいていると言える。発明のエッセンスを短縮して記述したものを、abstract (of disclosure)と呼び、明細書とは別個に添付するが、明細書においては、クレームに基づく発明の特徴を summary (of the invention) [MPEP § 608.01(d)]と呼ぶことも、同じ論理、すなわち属(genus—一般)と種(species—個別)である。明細書記述においても同じく一般と個別との論理に則った記述が採られる傾向を観察することができる。

特許実務家によると、米国特許明細書の作成においても、包括的概念と個別概念との論理関係を踏まえて、用語を慎重に選択しながら起草すべきというプラクティスが重視されている。以下、引用する。

実施例を記載するばあい、・・・各要素、比率ないし条件に関する包括的概念、個別概念を支持する十分な数だけ入れておくべきである。……………

包括的な記述の裏づけとなる具体的実施例は少なくとも一つは入れておくべきである。とりわけ包括的用語を十分に裏づけるためには種々の具体例ないし条件を用いた実施例を多数記載するのが望ましい。包括的用語で表現した要素の具体例を、さらに中位概念的表現を用いていくつかのグループに分類したばあいは、各グループについて少なくとも1種の具体例を用いた実施例を記載しておくのが望ましい。[8]。

個別事例を列記するには、表現能力として、同義語・類義語・関連語についての豊富な語彙知識が求められる。換言すれば、個別事例同士の種差を見極めて、それを語彙的に充分表現できる能力が、求められるのである。

上記のように、属と種との概念階層を利用して修辭的的技巧を凝らす必要がでてくる。論理と修辭とを統合して特許英文の表現を工夫する傾向にある。こういった論理＝修辭の観点から、特許英文の語彙運用の諸特徴を視るならば、特徴的な語彙連関[9]が見やすくなると思われる。前項において、属と種との論理が特許文書の基底的論理であると述べたが、この論理は、レトリックの知識があれば、修辭学的には、部分と全体とを置き換える換喩 metonymy ないし提喩 synecdoche であると分かる。類概念と種概念との置換を、提喩と定義するレトリック論[10]を採用するならば、シネドキは Patent Rhetoric の基本と言えるだろう。ただし、提喩を捉えるには、用語のタクソノミ taxonomy が必要であるが、この点は本稿の範囲外にあり別稿の課題とする。

2. 修辭的連語の抽出基準

レトリックは、自然言語処理の背景的知識として活用できる。対照・列挙・前辞反復などの修辭技法により創案された対句は、たとえば、等位接続型”A and/or/but B”という固定

的な形式を採っている。この固定形式により統一性が強まって修辭的連語となると共に、文法的には一つの品詞として機能する。同義語を反復する同義語量用 synonymy や反意語を並べてる矛盾語法 oxymoron という表現がある。これらは冗長に過ぎる表現ではない。同義語や反意語の重ね方には規則性があり、語義の算術的総和を超え出る意味を表現している。この種のレトリックは、英語が同義語豊かな言語であることから頻用される。レトリックの知識があれば、同義語・反意語などにより構成された連語を同定する手掛かりが得られる。

2.1. 修辭的連語の定義

連語 collocation の定義については、様々な論議がある。ここでは BBI 連語辞書[11]を手がかりに、自然言語処理の観点から連語について考えたい。

2.1.1 連語の定義

BBI 辞書では idioms については、句全体の意味に、個々の語の意味が反映していないとの見解から、これを含まないとしている。一方 simile については、collocations と idioms との中間にあって過渡的なもの(transitional)で、句全体としての意味に、個々の語の意味が部分的に反映しているとの見解から、これは collocations として採用している。しかし BBI では、simile を、"miscellaneous"として扱っている。本稿では、この BBI 見解を修辭学の観点から見直し、修辭的連語 rhetorical collocation の概念を導入する[12]。

2.1.2 修辭学の導入

単なる結合を越えて統一性が強固であることが collocations の成立条件である。free combinations—collocations—similes—idioms、これらにおける結合度に関して、その結合強度の測定基準として、「頻度」「共起制限」「意味の特殊化」[13]が提案されている。ここでは、この提案項目に加えて、固定的な形式と修辭技法的な要素との二つを、修辭的連語の基準として設定する。

英語には、対句 "A and/or/but B", "A, B, and/or C" や、比喩表現 "(as) ADJ as NP" などがあり、等位接続詞により同一品詞の語が結合される、あるいは従位の接続詞により形容詞と名詞(句)とが結合される、という固定的な形式を前提としている。この固定形式によって、対句・比喩として統一性が強まり、文法的には一つの品詞として機能する。

ここでは、語彙的かつ形態的な規定を設ける。たとえば、等位接続型の修辭的連語について、以下の規定を定める。

- α) 反対語の並列
- β) 同義語の並列
- γ) 類義語・関連語の並列
- δ) 同じ語の反復

2.2. 特許英文における修辭的連語の事例

ここで、個々の修辭表現について、解説すると共に具体例を挙げる。

2.2.1 「対照 Antithesis」

英語では、対極をなす二つの語(反義語・対照語や反義の前置詞・副詞など)を and で結合することで、両者を統合して対句の形で表す。

1) 例えば、go up and down のように反意の副詞を結合することで、統合的な表現を工夫している。対立する二語が明示されているので、単一の抽象語 (fluctuate) で表すよりも動的な表現となる(日本語では漢字二文字「増減」で一括した表現をするが[14]、これに対応するものといえる)。

2) 装置や処理過程に備わる反復動作は、その動作方向や動作位置のベクトルが正・逆の両方であることが多く、特許英文では、両方向の動作を一括して表わすことが顕著である。装置の技術的構造や動作・処理の過程の説明は、反復動作や、互いに反対方向に進む経過を表現することがしばしばである。こうした技術的表現には、二極対立を表す反対語同士や、その上位概念の語を用いた修辭技法が求められる。USP から抽出した対照の対句表現(加除・増減の過程や開閉動作など)を、下に示す。

[動詞対] add and delete, attach and detach, increase and decrease, mount and demount, plug and unplug; turn on and off

[形容詞対] enable and disable, open and closed, positive and negative; directly and indirectly

[名詞対] acceleration and deceleration, activation and deactivation, addition and deletion, connection and disconnection, compression and decompression, encryption and decryption, inflation and deflation, thermal expansion and contraction, transmission and reception

2.2.2 「同義語疊用 synonymy」

同じ事柄を別の面から述べるために同義語同士や類語同士を重ねる表現が、英語修辭法にはある。特許英文では、一連の処理過程や手順を述べる時、関連語を重ねる。USP から抽出した同義語疊用の組、および拡張した同義語疊用と見なしうる関連語の組の事例を、同じく USP から挙げる。

[動詞の組] decompile and decompress, plug and play, have been disclosed and described (herein/ in U.S. Pat. No.)

[形容詞の組] open and uncovered

[名詞の組] activation and initialization, disconnection and reconnection, extraction and purification

2.2.3 「列挙 Catalog」

同義語・類義語・関連語を挙げて同一対象の構成や性質を網羅的に説明する、これが列挙である。同義語疊用の拡張とも言え、古典ギリシャ語 cataloge が語源で、これに

は一覧の意がある。三個一対の形式 triad が最も均衡がとれており、同語源の語で統一される傾向が顕著である[15]。特許英文では、連鎖連動・継続共起・同時進行などの表現に使われている。USP から抽出した列挙の事例を、下に示す。

[動詞リアード] be accumulated, categorized, and disseminated; collecting, organizing, and distributing the signals; be formulated, executed, and coordinated; be re-compiled or re-assembled, relinked, and reloaded

[名詞リアード] addition, subtraction and multiplication; application portability, interoperability, and interchangeability; stability, maneuverability and followability

2.2.4 「前辞反復 Anadiplosis」

同じ語を繰り返す形で対句を構成する。

[名詞対] devices and device drivers, both operating data and debug data, key and key cylinder

上で述べた修辭技法の表現では、接辞が揃う傾向が顕著である。接辞が同じであれば、同時に、頭韻あるいは脚韻が揃う。韻律を兼ね備えると、すなわち韻を踏むと、その文体的効果が高まり、句のまとまりが強められる。接辞の同一性は、修辭的連語の抽出に利用できる[16]

3. 複合前置詞の定義と来歴

ここでは、複合前置詞 complex preposition の定義と来歴、そして諸特徴を簡単に述べる。

3.1. 複合前置詞の定義と諸特徴

3.1.1 定義

本稿では、複合前置詞を、Quirk, Greenbaumら(以下では、単に Quirk, et al. と記す)の定義に従い、「複数の語から成る前置詞」[17]と定義する。複合前置詞は、従来は群前置詞 group preposition とともに句前置詞 phrasal preposition と呼ばれていた。また、学校英文法でも群前置詞と呼ばれている。

3.1.2 構成

Quirk, et al. は、複合前置詞を、2語構成と3語構成と分けて整理しているが、本稿では、3語からなる複合前置詞に P1 言及する。3語構成を、一般式 P1+NP+P2 (P1, P2:前置詞)にて表す。名詞句 NP とした理由は、名詞に冠詞や修飾語が付くものも含めるため、複合前置詞の範囲拡張を考慮している。ここで、複合前置詞の構成パターン[18][19]の一部を示す。

in+NP+of (in advance of)

in+NP+with (in love with)

at+NP+of (at the expense of)

for+NP+of (for the sake of)

on+NP+of (on account of)

3.1.3 用法

語の結合である句が、一つの前置詞であるかのように機能する(文法化 grammaticalization の結果)。複合前置詞内の名詞については、意味が漂白化 bleaching し、それに伴い、無冠詞に推移する傾向が認められる。

Quirk1, et al.の指摘では、法律英語では目立って用いられており、その法律文・官庁文に顕著である。

3.1.4 語彙的特徴

P1+NP+P2 の複合前置詞の中核を成す名詞は、多くは、抽象名詞でありフランス語からの借用語である [20]。本稿では、動詞派生の抽象名詞を核とした複合前置詞を中心に採りあげる。

3.1.5 意味的特徴

英語の GP の意味分類を、以下に示す。

場所／原因・理由／結果／関係
目的／利益／その他

単一前置詞の多くは多義であるが、複合前置詞は一意である。たとえば、with と in cooperation with とは、前者が多義(日本語の「と／と共に」に当たる)であるのに対し、後者は一意(日本語の「と共同して」に当たる)である。

複合前置詞を構成する前置詞は、もともと多義の部類にはいるが、複合前置詞が法律文でよく用いられるということは、その多義を避けるという実務的目的によるものであろうと推定可能である。

3.2. 複合前置詞の事例

本項では in+NP+with のパターンについてのみ事例を挙げる。

in accordance with, in collaboration with
in collusion with, in combination with
in comparison with, in compliance with
in concert with, in cooperation with
in connection with, in deliberation with
in partnership with, in league with
in line with, in step with

4. 連語辞書による訳質改善

以下に修辭的連語・複合前置詞を収録した辞書を用いた機械翻訳の具体例を従来の処理と比較して示す。同じ英日機械翻訳システムを使って、修辭的連語および複合前置詞を収録した辞書の非実装時／実装時の訳文を比較した。

クロスランゲージ社の特許英文専用翻訳システム PAT-Transer を使い、ユーザ辞書に修辭的連語・複合前置詞などを登録して USP 和訳実験を行った。実験の手順は以下

のとおりである。システム基本辞書・専門用語辞書と共に、連語ユーザ辞書を使うが、最優先は連語ユーザ辞書で、これを最上段に置く。次は、基本辞書を置く。ここまでの順位は不動とする。第三および第四には、専門用語辞書を置くという序列にした。

[試験文1] 実施例

In one preferred embodiment of the present invention, five icons, each representing one of the subject matter categories, are displayed and may be selected and activated by the viewer manipulating the remote control.

A) 連語辞書なしの機械翻訳による実施例の和訳は、「1つの本発明の好適な実施例において、5つのアイコン(内容カテゴリーの各々を表している一つ)は、表示されて、選んでもよくて、通って遠隔制御を操作している視聴者をアクティブにした。」

[評価] A)の和訳は、

1) ”be selected and activated” の and の構文解析に失敗、助動詞が activated を支配せず、と解析。

2) by the viewer manipulating the remote control が、受動態の意味上の主語であることを解析できず、the viewer を activated の目的語とし、能動態・過去形として解析してしまった。

3)したがって、by を副詞として解析することになり「通って」と訳出した。

B) これに対して連語ユーザ辞書を使用した場合、同じ MT システムでの訳文出力は、

「1つの本発明の好適な実施例において、5つのアイコン(内容カテゴリーの各々の表している一つ)は、表示されて、遠隔制御を操作している視聴者によって選択・起動されてもよい。」

[評価] B)の和訳は、

1) selected and activated が受動態で、may に支配されていると解析した。連語配下の諸語を単体として and の構文解析は回避されている。

2) by が受動態の動作主を目的語にとる前置詞として正確に解析がなされている。

なお、助動詞 may の訳語はシステム基本辞書により「できる」に変更できる。

[試験文2] クレーム

12. A pressure controller device according to claim 1 wherein a grip is provided on the outside of the housing in close adjacency to the handle.

C) 連語ユーザ辞書なしの MT によるクレーム文和訳は、「グリップがハンドルに緊密な隣接関係のハウジングの外に設けられている請求項1に記載の圧力コントローラ装置。」

D) これに対して連語辞書を使用した同じ MT の訳文は、

「グリップがハンドルに近接するハウジングの外に設けられている請求項1に記載の圧力コントローラ装置。」

[評価]

C)の和訳は、in close adjacency to を構成する語の逐語訳「緊密な隣接関係の」である。

D)の和訳は、in close adjacency to を前置詞として構文解析している。訳文は「近接して」という訳語が生成されて、C)のように逐語訳にならない。

5. 結論的所感

・修辭的連語および複合前置詞を登録した辞書を参照することで、等位接続の句形態を一つの形態素として認識し、分解することなく構文解析を行うことができる。その結果、統語解析における誤りの発生を抑えることができる。

・さらに、修辭的連語の場合、辞書に格納されていない表現形態であっても、当該表現形態の接頭辞・接尾辞が同一であるという条件判断を解析プログラムに加えることにより、当該の表現形態を一つの句形態として認識することが可能になる。

・上記の二つの方法により、分解的に解析すべきではない句形態について、その統語解析を回避することができる。解析を回避することにより、文全体の統語解析の精度を向上させることができる。

参考文献

- [1] 藤芳寛治(2004).「特許英語通信文と英文明細書作成へのガイド(12)－米国型英文クレームの作成－」『パテント Vol.57, No.7, 62-84
- [2] Potts, H. E. (1944). The Definition of Invention in Patent Law. The Modern Law Review Vol. 7, No. 3 (Jul., 1944), Blackwell.113-123
- [3] 田辺 徹(1986)『エンジニア・翻訳者のための英文特許入門』アイピーシー 318-330, 332-346
- [4] Wegner, H. C. (1997). The Doctrine of Equivalents after Warner-Jenkinson. D.C. Bar PTC Section meeting at George Washington University, March 14, 1997. Asia Pacific Legal Institute, International Comparative and Intellectual Property Studies, Retrieved August 31, 2010 from <http://apli.org/ftp/warner.pdf>
- [5] Mill, J. S. (1874). *A System of Logic* (4th ed.). Harper & Brothers. 105-120, 214-220
- [6] Whewell, W. (1860). *On The Philosophy of Discovery*. Reprinted in 1971. Lenox Hill Pub. 238-291
- [7] 藤芳寛治(2004). 前掲同書
- [8] 朝比奈宗太(1992).「国際的調和時代における外国出願の留意点と対応－米国を中心として－」通商産業調査会 41
- [9] Collins, K. E. (2008, December). The Reach of Literal Claim Scope into After-Arising Technology: On Thing Construction and the Meaning of Meaning. Connecticut Law Review. Vol. 41 No. 2. 493-559
- [10] 佐藤信夫(1983). 『レトリック感覚』講談社 138-175
- [11] Benson, M., Benson, E. & Robert, I., Eds. (2010). The BBI Dictionary of English Word Combinations: Your Guide to Collocations and Grammar John Benjamins Pub Co; 3 Revised
- [12] 佐良木昌(1999).「修辭学的手法による英語コロケーションの抽出と修辭的連語辞書の構築」電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション 99(88), 9-14, 1999-05-26
- [13] 中本恭平(1997).「コロケーション再考 頻度、共起制限、意味の特殊化」英語表現研究第14号 58-66
- [14] 林四郎(1987).『漢字・語彙・文章の研究へ』明治書院 112-151, 252-269
- [15] 佐良木昌(2000).「修辭的連語とその語源的特徴」日本英語表現学会第29回大会ハンドアウト
- [16] 佐良木昌(1999). 前掲同書
- [17] Quirk, R., Greenbaum, S. & Leech, J. (1985). *A Comprehensive Grammar of The English Language*. Longman. 669-673
- [18] 秋元実治(2000).『文法化とイディオム化』ひつじ書房 162-191
- [19] Hoffmann, S. (2005). *Grammaticalization And English Complex Prepositions: A Corpus-based Study*. Routledge.
- [20] 秋元実治(2005).「複合前置詞とイディオム化」講演レジュメ 名古屋大学国際開発研究科 2008年8月1日