

<修士論文要旨>

JAITS

論文題目: タグ付きパラレルコーパス作成利用支援ツール「Alpha」の設計と開発、およびそれを用いた分析手法の提示

(英文題目: Alpha; Tool for Assisting in the Creation and Utilization of a Tagged Parallel Corpus: Its Design and Development including a Method for Conducting Analysis via the Tool)

提出者: 北村富弘

授与機関: 関西学院大学大学院

取得学位の名称: 修士(日本語教育学)

学位取得の方法: 課程

取得年月日: 2013年3月16日

要旨

本研究では、タグ付きパラレルコーパスの作成、利用を簡便にするためのコンピュータアプリケーション Alpha を開発した。

単言語コーパスの場合、タグ付きコーパスを作成し、利用することは各種コンピュータアプリケーションの使用により容易になっている。しかしながら、パラレルコーパスの場合、その作成と利用において、従来公開されているコンピュータアプリケーションを用いても特にアラインメント(対訳対応付け)の工程ならびにタグ付与に困難が伴う。

そこで本研究では、この点を解消することを目的に、タグ付きパラレルコーパス作成利用支援ツール Alpha を開発した。なお、Alpha の開発における要件として多言語対応としたが、特に日本語、中国語への対応を第一義とした。設計における主眼は、タグ付きパラレルコーパスの作成と利用を簡便にし、研究者の注意力を言語学的判断に置けるようにすることである。

Alpha 開発へのアプローチとして、本研究ではまず、タグ付きパラレルコーパスの有用性を論じ、それに関する現状での課題を明らかにした。次いで、それにもとづき Alpha の構想と設計を行った。

パラレルコーパスは対照言語学、翻訳研究、言語教育、翻訳教育等の研究および教育実践に用いられている。コーパス利用の有用性については、研究の精密化、新しい研究領域・手法の開発、内省では得られなかった現象への気付き、観察した現象の再現性等があげられる。またタグ付きコーパスの特徴である付帯情報の意義として、付帯情報そのものが知的資源

KITAMURA Tomihiro, "Alpha; Tool for Assisting in the Creation and Utilization of a Tagged Parallel Corpus: Its Design and Development including a Method for Conducting Analysis via the Tool," *Interpreting and Translation Studies*, No.13, 2013. pages 153-156. © by the Japan Association for Interpreting and Translation Studies

であること、それが分析の手掛かりとなること、また、研究の客観性向上に寄与すること等があげられる。

これらの有用性を生かすためのパラレルコーパス作成利用支援ツールを開発するという観点から、日本語、中国語への対応を要件として先行研究を検討すると、既存のコンピュータアプリケーションを参考にしつつ、新たに設計する必要があることが明らかになった。

Alpha の設計思想としては、言語研究に紙媒体を用いた場合の手軽さの部分を継承しつつ、紙媒体ではなしえない効率の高いデータ整理、分析を実現することを理想とした。すなわち、コンピュータ上における以下のような利用方法を想定した。ST、TT において対訳対応するセンテンスごとに、画面の左右に ST、TT それぞれが表示されること。ユーザがマウスを使って、対訳研究上有用な対応箇所(語または語の組合せ)にマーカー色を付けることができること。またマーカーの箇所には必要な注釈を書き込むことができること。これらのマーカーおよび注釈がコーパスにおけるタグおよびその付帯情報となること。マーカー付作業を終えた後、全文検索・抽出、注釈(すなわち付帯情報)の検索・抽出ができること。また、これら検索・抽出は複数の言語資料にわたって行うことができること。

本研究で開発した Alpha 初期バージョンの主要な構成は、アラインメント支援、付帯情報付与支援、および、文字列および付帯情報を検索条件とした KWIC コンコーダンス表示である。また、付帯機能として、インターネット上に存在する一連のウェブページから効率的にコーパスを作成するための HTML インポート機能を持つ。

なお、本研究の一部として、Alpha を言語研究に応用する例を示した。

まずその1つとして、日本語・中国語で構成されるパラレルコーパスを作成した。ここでは、Alpha の HTML インポート機能を用いて効率よくコーパスを作成する方法を示した。具体的には、東京大学東洋文化研究所田中明彦研究室作成による「日本政治・国際関係データベース」のうち「日中関係資料集」を言語資料として、日本語版と中国語版の対応のある外交文書 132 文書(66 ペア)を収集し、日中両言語間のセンテンスごとに意味的対応のついたパラレルコーパスを作成した。これを日中外交文書コーパスという。

次に、パラレルコーパスへのタグ付与の例、および付帯情報を利用した分析例を示した。具体的には、上記で作成した日中外交文書コーパスの一部である「戦略的互惠関係」の包括的推進に関する日中共同声明」の日本語版及び中国語版を用いて、日中両言語間に見られる名詞句の明示表現と潜在表現の対照に関する分析を行った。その詳細は以下のとおりである。

分析の目的: 日中両言語による外交文書に見られる対照的特徴を分析する試みとして、名詞句の明示表現と潜在表現の対比を行う。

用語の定義: 明示表現 (explicitation) とは、1 つの事象について特定化・詳述化がより強い表現であり、潜在表現 (underspecification) とは、1 つの事象について特定化・詳述化がより弱い表現である。たとえば、「日本、アメリカ合衆国、韓国の外務大臣による会合」という事象が、英

語版および日本語版において以下のような表現であったとする。

- ① 日本語版：日米韓外相会合
- ② 英語版：a ministreal trilateral meeting (三か国閣僚会合)

この場合、日本語版がより特定化・詳述化の強い表現をとっている。したがって、①を明示表現、②を潜在表現とする。

分析動機：「日米韓三か国共同声明」の日本語版及び英語版を資料として、このような明示表現・潜在表現の対比を名詞句について調べると、日本語版が明示表現、英語版が潜在表現であるケースが多く、その内訳で最も多くみられるのは英語版の名詞句に定冠詞があるケースであった。この場合、定冠詞による前方照応が働くので、名詞句の一部が省略され潜在的な表現となるという説明ができるであろう。

しかしながら、明示表現・潜在表現の対比は、いずれも定冠詞のない言語である日本語・中国語の対照にも見られる。そこで、日中両言語における明示表現・潜在表現の対比を生む要素について明らかにすることを考えた。その分析の第一歩として、日中共同声明のテキストを資料として名詞句の対比を試みた。

言語資料と分析手順：本分析に用いた日本語版資料は「戦略的互惠関係」の包括的推進に関する日中共同声明(外務省 2008 年 5 月 7 日。文字数 2848 字。)であり、中国語版資料は「中日关于全面推进战略互惠关系的联合声明」(中华人民共和国外交部 2005 年 5 月 7 日。文字数 1905 字。)である。

分析手順として、この日中両言語版が互いに等価であると仮定し、意味上対応する箇所の名詞句の表現に着目し、明示表現・潜在表現の対比が見られる箇所にタグ付与・付帯情報付与をおこなった。その後、明示表現・潜在表現の対比が見られる箇所を抽出した。

結果と考察：今回の分析資料においては、日本語版と中国語版に現れる明示表現の例数はほぼ拮抗していることがわかった。ただし、中国語版が明示表現、日本語版が潜在表現であった箇所を詳細に見ると、その多くにおいて、中国語版の1センテンスが比較的短く、日本語版の1センテンスが比較的長いことが観察された。例を挙げると以下のような対照である。

日本語版：

3. 双方は、歴史を直視し、未来に向かい、日中「戦略的互惠関係」の新たな局面を絶えず切り開くことを決意し、将来にわたり、絶えず相互理解を深め、相互信頼を築き、互惠協力を拡大しつつ、日中関係を世界の潮流に沿って方向付け、アジア太平洋及び世界の良き未来を共に創り上げていくことを宣言した。

中国語版：

三、双方决心正视历史、面向未来、不断开创中日战略互惠关系新局面。双方将不断增进相互理解和相互信任、扩大互利合作、使中日关系的发展方向与世界发展潮流相一致、共同开创亚太地区和世界的美好未来。(双方は、歴史を直視し、未来に向かい、日中戦略的互惠関係の新たな局面を絶えず切り開くことを決意した。双方は、絶えず相互理解と相互信頼を増進し、互惠協力を拡大しつつ、日中関係の発展方向を世界の発展の潮流と一致させ、アジア太

平洋地区及び世界の良き未来を共に創り上げていく。)

この例では、日本語版の1センテンスに対して中国語版の2センテンスが対応している。ここで、中国語版の第2センテンス目の主語「双方(双方)」が日本語版のセンテンスでは明示されておらず、その箇所の日本語版 - 中国語版の対比は「 φ : 双方(双方)」すなわち「潜在表現: 明示表現」になっている。それぞれ言語ごとに見ると、日本語版ではセンテンスの主題として「双方」を立て、この主題が文末の「宣言した」まで影響を及ぼしているが、一方、中国語版では、第1センテンス、第2センテンスにそれぞれ主語として「双方(双方)」を立てている。すなわち、日本語において主題の影響が当該センテンスの最後まで継続しているが、中国語版においてはその間に主語の立て直しが行われているのである。

これらのことから、本資料の明示表現・潜在表現には、それぞれ個別言語のテキスト言語学的性質が影響していることが示唆された。なお、本分析のデータ数が非常に限られているため、以上の結論を一般化することはできない。

今後、様々なテーマのもとでタグ付きパラレルコーパスが構築され、翻訳研究、対照言語学等に活用されることが望まれる。これらの研究に資するべく、順次、Alpha の機能追加・拡張を行っていきたい。

.....
【著者紹介】北村富弘 (KITAMURA, Tomihiro)。2013年3月関西学院言語コミュニケーション文化研究科博士前期課程修了。現在、法務省大阪刑務所国際対策室法務技官(中国語、英語担当)。中国語実務翻訳に関する研究を志向している。
.....