

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 19 日現在

機関番号：32401

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24720185

研究課題名(和文)日独「句例」対訳データベースの構築

研究課題名(英文) Construction of the database of parallel translations of Japanese-German "Phrases" database

研究代表者

阿部 一哉 (Abe, Kazuya)

跡見学園女子大学・文学部・助教

研究者番号：50570420

交付決定額(研究期間全体)：(直接経費) 2,500,000円、(間接経費) 750,000円

研究成果の概要(和文)：「『レストランを予約する』 - 'einen Tisch im Restaurant reservieren'」のように、句レベルで日本語とドイツ語を対応付けた「句例」は、我々日本人が日本語の語感に頼りつつドイツ語の文をつくらうとした場合、とりわけ、ドイツ語教育の分野における作文指導などにおいて、非常に有用であると考えられる。このような見込みに基づいて、本研究ではコーパス言語学などの知見を援用しつつ、効率的にドイツ語句例を収集し、日本語の句例と対応付ける方法論の検討を行った。なお、その成果として、パイロット版「日独句例パラレルコーパス」を構築し、オンライン公開している。

研究成果の概要(英文)：The phrase pairs like "'to reserve a table at the restaurant' - 'einen Tisch im Restaurant reservieren'" between Japanese and German correspondences is supposed to be very useful when a Japanese wants to write in German, especially at the teaching of German composition. Under such a prospect, referring to the knowledges like corpus linguistics, the methodology for conveniences in collection of German phrases and translating them in to Japanese were developed and examined. As one of the results, the pilot version of 'the parallel corpus of the Japanese-German phrases' were developed and published on-line.

研究分野：人文学

科研費の分科・細目：言語学・外国語教育学

キーワード：ドイツ語学 ドイツ語教育学 翻訳 独作文 コーパス言語学 応用言語学 Webサービス

## 1. 研究開始当初の背景

### (1)

現在多種多様な(電子)コーパスが、利用可能になっている。ICT (Information Communication Technology) の発達により、個人の研究者でも、大量の言語データを比較的容易に処理することが可能になった。また、これに合わせて様々な言語において、電子テキストがコーパスとして公開され、有償・無償で利用可能になっている。

### (2)

ドイツ語におけるコーパスも、多種多様なものが利用可能になっている。中でも、現在ドイツ語研究所の DeReKo(Deutsche ReferenzKorpora)や、ベルリン・ブランデンブルク人文科学アカデミーの DWDS(Digitales Wörterbuch der deutschen Sprache) などが、オンラインで無償利用できるよう、公開されている。

### (3)

日本のドイツ語研究におけるコーパス利用の段階ははまだ黎明期にある。コーパスは、辞書研究などの応用分野においても、積極的に活用されている。しかし、日本におけるドイツ語研究の分野では、まだ辞書研究といった観点からの、コーパス利用の検討はまだ本格的には着手されていないのが現状であった。

### (4)

日本人がドイツ語を書く際に有用なツールを開発することは、我々ドイツ語研究者にとっての喫緊の課題である。パーソナル・コンピュータを使用してドイツ語を書く、というニーズは増大している。しかしその一方で、日本語母語話者がドイツ語を書くために、本当に有用なオンライン辞書というものは今現在もって、まだ存在していない。したがって、有用なオンライン和独・独和辞書ツールの整備は、我々ドイツ語研究者にとっての喫緊の課題であった。

## 2. 研究の目的

### (1)

本研究の第一の目的は、「日独『句例』対訳データベース」パイロット版をオンライン公開することであった。研究開発当初の背景を踏まえ、本研究では、オンライン和独・独和辞書ツールとして、「日独『句例』対訳データベース」を想定した。これは、「『レストランを予約する』-'einen Tisch im Restaurant reservieren」のように、句レベルで日本語とドイツ語を対応付けた「句例」を大量に所収するオンラインデータベースである。句例データベースを構築するためには、小規模のサンプルデータを用いて、妥当なデータ構造

を検討する必要がある。そのためにパイロット版のシステムの構築を目指した。

### (2)

本研究の第二の目的は、効率的な、句例収集手法を確立することであった。データベースに収録する言語データの基本単位として、句を想定したのは次の理由による。日独対訳ペアを作成する際、語では多義性により一対一の対応を取ることが難しい一方、文は個別性が高すぎて、使用者の多様なニーズに合致するデータを提供することが難しいと考えられる。これに対して句レベルであれば、一対一で対応した、一般性の高い対訳ペアを作成することが期待できる。ただし、こういった句を実際の言語使用データから収集し、日独対訳ペアを作成するための、効率的な手法が欠如していた。したがって、句例収集の効率的な手法の確立は喫緊の課題であったと言える。

## 3. 研究の方法

### (1)

まず研究全体の計画段階として、「日独『句例』対訳データベース」構築の構想を打ち立てた。ここでは(イ)可能な限り大量かつ網羅的な実例収集、(ロ)使用頻度の導入、(ハ)句例分析、(ニ)ICT技術を用いたパラレルコーパス、(ホ)ドイツ語から分析する根拠、という大きく5つの構想の柱が得られた。

(イ)は、使用者の多様なニーズに応えるために掲げられた構想である。(ロ)は句例といっても、なんでもかんでも収集するのではなく、ある程度一般性が認められるものに限定するため、使用頻度の高いものを収集するという考えに基づいている。(ハ)は、やはり句例といっても闇雲に収集を行うのではなく、例えば「他動詞+目的語」といった形で、句例分析を行いつつ学問的な整理も行いながら研究を進めるという構想である。(ニ)は、(イ)で挙げたように可能な限り大量かつ網羅的なデータを扱うために、ICT技術を積極的に援用するという構想である。(ホ)は、日本語からドイツ語を書くというニーズに応えるためには、日本語から句例を収集するのが順当であるが、作業上の効率を視野に入れた場合、まずはドイツ語における一般的な句例を収集すべきだという構想である。研究はこれらの構想にしたがって進められた。

### (2)

次に、句例のサンプルデータに基づき、「日独『句例』対訳データベース」パイロット版のシステム構築を行った。サンプルデータは研究協力者の在間進氏が収集整理を行ってきた、日独対訳事例集を使用させていただいた。この研究の目的は、句例データベースを

オンライン公開する際の、データ構造を設計することであった。その際、開発段階におけるデータ構造の変更などに柔軟に対応できるよう、スキーマレスなデータ管理が特色の MongoDB などバックエンドのデータベースシステムとして採用した。

### (3)

また、句例収集手法の確立を目指して、オンラインコーパスや、Web データといった様々なリソースに応じた、手法の検討を行った。具体的に検討を行ったのは、次の3つである。

(イ) マンハイム・ドイツ語研究所の DeReKo から動詞に基づき事例を収集し、得られた事例のコンコーダンスライン分析に基づき、句例を効率的に抽出する手法 (ロ) ベルリン・ブランデンブルク人文科学アカデミーの DWDS を利用し、名詞+動詞コロケーションを取り出し、得られたコロケーションに着目し句例を抽出する手法 (ハ) FU-ベルリンで開発している Web コーパス DECOW2012 から、名詞+動詞コロケーションを取り出し、得られたコロケーションに着目し句例を抽出する手法

## 4. 研究成果

### (1)

成果の一点目は、句例データベース構築の構想である。冒頭で述べたように、日本のドイツ語辞書研究におけるコーパス活用手法についての知見はまだ不足していると言わざるを得ない。このような状況のなかデータ収集から活用法までの一連のプロセスについての構想を打ち立てた。

### (2)

成果の二点目は、開発段階におけるオンライン辞書システム構築手法の確立である。今回、「日独『句例』対訳データベース」のオンラインシステムのバックエンドには、JavaScript によるデータ問い合わせが可能な MongoDB を採用した。これにより、開発段階における様々な変更にも、柔軟に対応することが可能になった。

### (3)

成果の3点目は、ドイツ語句例抽出におけるデータ収集及び処理に関する手法の確立である。ここでは、コーパスに基づく質的研究の主要なツールであるコンコーダンスラインを使用した句例収集手法や、量的研究における主要概念であるコロケーション(言語使用において頻繁に用いられる、語と語の結合体)に基づく句例収集手法が整備された。

### (4)

この研究によって得られた直接的成果から派生した、次に挙げるあらたな研究の可能性も、本研究の成果として認めることができる

だろう。(イ) ディー・ツァイトなどのオンラインニュースデータを効率的に収集し、講読授業の基礎データを作成するなどの形で、コーパスとして活用する手法の検討。(ロ) Twitter などの発信型サービスにおける表現集の作成。

### (5)

なお、「日独『句例』対訳データベース」のパイロット版を、下で挙げる URL において公開し、今後もデータを拡充しつつ、本研究成果の社会への還元を推進していく所存である。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- ① 阿部一哉、在間進、句例パラレルコーパスの構築と諸問題、跡見学園女子大学文学部紀要、査読なし、48号、2013、A95-A104
- ② 阿部一哉、書くためのパラレルコーパス構築と頻度、日本独文学会研究叢書、査読なし、098号、2014、65-79
- ③ 阿部一哉、動詞+名詞コロケーション分析に基づく句例抽出手法、東京外国語大学大学院ドイツ語学文学研究会編『DER KEIM』、37号、2014、19-42
- ④ 阿部一哉、新聞記事によるドイツ語コーパス構築、ドイツ語情報処理、査読有、24号、2014 出版予定、1-8

[学会発表] (計3件)

- ① 阿部一哉、句例パラレルコーパス構築とその諸問題、2012年10月13日、日本独文学会2012年度秋季研究発表会
- ② Kazuya Abe, Twitter-Corpus and collection of German phrases, 2014/3/4, A Workshop at the 36th Annual Conference of the German Linguistic Society 2014
- ③ 阿部一哉、読書報告：シュテファン・ミュラー著、コーパスの質的研究、2013年4月27日、ドイツ語言語理論研究会

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

○取得状況 (計0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

[その他]  
ホームページ等  
<http://atomilang.atomi.ac.jp/parac/>

## 6. 研究組織

### (1) 研究代表者

阿部一哉 (ABE, Kazuya)  
跡見学園女子大学・文学部人文学科・助教  
研究者番号：50570420

### (2) 研究協力者

在間進 (ZAIMA, Susumu)  
東京外国語大学・名誉教授  
研究者番号：30117709

シュテファン・ミュラー (Müller, Stefan)  
FU-Berlin・Institute for German and  
Dutch Philology・教授  
研究者番号：