

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 10 日現在

機関番号：32689

研究種目：基盤研究(B)

研究期間：2011～2014

課題番号：23320120

研究課題名(和文)コーパス駆動型研究に基づく学習用ドイツ語語彙

研究課題名(英文)Basic German Vocabulary for Foreign Language Learners: A data-driven Approach

研究代表者

岡村 三郎 (Okamura, Saburo)

早稲田大学・国際学術院・教授

研究者番号：30009724

交付決定額(研究期間全体)：(直接経費) 4,900,000円

研究成果の概要(和文)：インターネットの大規模なテキスト情報を利用し、ドイツ語の基本語彙を方法的に客観的に確定し記述するために、約8億5千万語からなる十分大規模なコーパスを構築した。見出し語の出現頻度、安定性、生産性の3基準をコーパスに適用し基本語彙のランキングリストを作り、基本語彙に属する語の語形の記述、意味の記述、さらに学習用基本語彙のための適切な例文を選び、その訳等をつけている。現在は1000語までその記述が完成していて<http://basic-german.com/beta/>で検索可能である。

研究成果の概要(英文)：Seeking to provide a methodical and objective basis for identifying and delineating what constitutes the basic vocabulary of the German language, we made use of the Internet's vast resources to construct a corpus of some 850 million words. We then compiled a hierarchical list, with the rankings based on three criteria: frequency, stability, and productivity. We further selected usage examples from the Internet and translated them into Japanese. As of now (March 2015), descriptions of the first 1000 words have been completed, searchable at <http://basic-german.com/beta/>.

研究分野：人文学

キーワード：コーパス言語学 コーパス構築 データ駆動型 ドイツ語基本語彙 ドイツ語教育 辞書論 造語論

1. 研究開始当初の背景 従来のドイツ語基本語彙研究には以下のような方法があったが、それぞれ弱点を持っていた。

a) 頻度による方法：語の出現頻度によって基本語彙を選び出そうとする。この方法の弱点は、コーパスが小規模であること、並びに現在のコンピュータが持つ計算機能を十分に利用していなかった。

b) コミュニカティブ・語用論的な方法：語のコミュニケーション重要性、それについての編纂者の個人的な判断によって基本語彙を選ぼうとする。この方法の決定的な弱点は、データの裏打ちがないこと、並びに基本語彙選択の基準を追検査できないことである。

c) 辞書学的方法：既存の辞書、語彙集を利用し、それらのデータの共通部分を求めて基本語彙を選ぼうとする方法である。この方法の弱点として、二次資料による研究であること、さらに統計的数値の処理が不適切である点が上げられる。

2. 研究の目的 (1) 上述のように既存の学習用ドイツ語語彙集は、各語についての編纂者の個人的・経験的な判断、または不十分なデータ量に頼ることが多かった。それに対し本研究は、インターネットの膨大なテキスト情報を使い十分大規模な基準コーパスを構築し、それに基づきドイツ語の基本語彙を確定し、それにより日本人ドイツ語学習者のための学習用ドイツ語語彙集の編纂することを目的とする。本研究はその際に、ドイツ語の実証的な研究に基づいて基本語彙を方法論的に客観的に記述する。

(2) その研究結果は学習者がドイツ語を学ぶ際に役に立つのみならず、教材作成や学校でのカリキュラム作成の際にも指針となる。研究結果はインターネット上にデータベースとして公表し、さらに学習用語彙集として刊行することもめざし、それにより多くの人にとって利用可能なものになる。

3. 研究の方法 (1) 本研究「コーパス駆動型研究に基づく学習用ドイツ語語彙」は、厳密なコーパス言語学の方法(データ駆動型、大規模情報の処理)をとり、a)大量のデータを使用する、及び b)語彙を認定する方法という二つの革新的な特徴を持つ。

a)すなわち膨大な語数を持つ複数のコーパスに基づいて、十分大規模な基準コーパスを構築し、それから基本語彙を算出する。これらのコーパスは広域的なプリントメディアのテキスト(コンセプトとしての書き言葉として)及びインターネット・議論フォーラムの言語(コンセプトとしての話し言葉として)を、少なくとも10年以上の長期にわたり、幅広い専門分野から反映するものである。

(2) b) 基本語彙を選択する基準としては、出現の頻度のみならず、長期間にわたる出現の安定性、分野を越えた出現の安定性、生産性(造語能力)とする。まず頻度はある見

出し語の相対的な出現頻度を意味し、コーパスの中でもっとも出現頻度の高い語との相対的な頻度を計算した頻度クラスを指数とする。

さらにある見出し語の出現の安定性は、コーパスの中でその語がどの程度規則正しく、または不規則に分布しているかによって算出する。安定性は、長期間にわたる語の出現の安定性、及び分野を越えた語の出現の安定性の二つに分けられる。安定性を確認するためにはGries' DP (Gries 2008)を指数とする。生産性とは、ある語彙素が、それ自身出現頻度の高い複合語の一部となりうる能力、すなわちその造語能力である。この生産性の算出には形態素解析プログラムを利用する。

4. 研究成果 (1) 上に述べたように厳密なコーパス言語学の方法(データ駆動型、大規模情報の処理)に従い、ドイツ語のプリントメディア、並びにインターネットフォーラム等のコーパスを元にして、約8億5千万語からなる十分大規模な基準コーパスを構築した

(2) この基準コーパスは、コンセプトとしての書き言葉(Spiegel.deを代表とするプリントメディア)、及びコンセプトとしての話し言葉(インターネットフォーラム)の両方を使って構築されており、プリントメディアのみに偏りがちだったこれまでのコーパスと比べてドイツ語使用の現状をより良く反映したコーパスの構築に成功している。

(3) コンセプトとしての書き言葉、及びコンセプトとしての話し言葉双方の部分コーパスの違いも計算して数値化しており、<http://www.basic-german.com>でその違いを見ることができる。

(4) 方法的な客観性と透明性を保証するために、見出し語の出現頻度、安定性、生産性の3基準を上述の基準コーパスに適用し基本語彙のランキングリストを作り、その結果は部分的にすでにネット上で提供している(<http://www.basic-german.com/>)

(5) 基本語彙のランキングリストに合わせ、言語的に比較的簡単でありしかも典型的な例文を選び出すために、共起および例文の困難度(例文の中にこれまで得られた基本語彙が現れる度合い)を基準とし、基本語彙にふさわしい例文をネットから選出し、例文のリストを完成した。これらのデータは全て、既にオンラインデータベースに入力されている。

(6) それを元に学習用基本語彙のための適切な例文を選び、それに日本語の訳を付け基本語彙辞典編纂の作業を行っている。我々はとりあえず上位数千語までに、その語の基本語彙ランキング、語形の記述、意味のドイツ語、日本語での記述、コーパスから選んだ例文、そしてその訳等を付けている。2015年3月現状では1000語までその記述が完成していて<http://basic-german.com/beta/>で検索

可能な状態にある。それと平行して、多くのドイツ語学習者がウェブで利用することを念頭に、利用しやすい単語カードも開発し、これも利用可能である。

(7) なお我々のコーパス駆動型研究のもたらした既存の基本語彙との差異についての興味深い知見(例えば同じ頻度クラスに属する語が、安定性、生産性の基準を加味することによってランキングリスト上ではいかに差別化されるか)についてはすでに論文(Lange/Okamura/Scharloth (2015))でも報告したし、これからも論文にまとめる予定である。

(8) 研究協力者ドレスデン工科大学シャルロート、ヨアヒム (SCHARLOTH Joachim)のもとで、この研究と密接に関連して以下の博士論文が作成課程にあり、これらはこれまで得られた知見をさらに深化するものである。
Frank Nickel: Die Bedeutung der Englischkompetenz japanischer Deutschlerner für den Wortschatzaufbau DaF in Japan
Nelli Nurgaliev: Anglizismen in der Pressesprache

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 1件)

Noah Bubenhofer, Willi Lange, Saburo Okamura, Joachim Scharloth (2011): Welcher Wortschatz? Korpuslinguistische Untersuchungen zur Wortschatzelektion japanischer Deutschlehrbücher für Anfänger. In: ドイツ語教育 Deutschunterricht in Japan, 16,43-60. (査読有り)

[学会発表](計 3件)

Grundwortschatz Deutsch als Fremdsprache. Ein datengeleiteter Ansatz. Deutscher Germanistentag 2013 (Kiel/Germany)において、Saburo Okamura, Willi Lange, Joachim Scharloth の共同発表、2013年9月23日

Grundwortschatz Deutsch. Ein datengeleiteter Ansatz. XV. Internationale Tagung der Deutschlehrerinnen und Deutschlehrer (Bozen/Italy)において、Saburo Okamura, Willi Lange, Joachim Scharloth の共同発表、2013年8月2日

Methoden der Bestimmung des Kernwortschatzes Deutsch 日本独文学会秋期研究発表会(金沢)において、

Saburo Okamura, Willi Lange, Joachim Scharloth の共同発表、2011年10月16日

[図書](計 6件)

Willi Lange, Saburo Okamura, Joachim Scharloth (2015): Datengeleiteter Grundwortschatz Deutsch. Kriterien für die Lemmaselektion. In: Drumbi, Hans (Hrsg.): Deutsch von innen Deutsch von außen. Tagungsbeiträge der XV. Internationalen Deutschlehrertagung 2013 in Bozen. Teilband "Wortschatz und Phraseologie". Bozen. (in press)

Willi Lange, Saburo Okamura, Joachim Scharloth (2015): Grundwortschatz Deutsch als Fremdsprache: Ein datengeleiteter Ansatz. In: Joerg Kilian/Jan Eckhoff (Hrsg.) Deutscher Wortschatz - beschreiben, lernen, lehren, S. 203-219.

Noah Bubenhofer, Willi Lange, Saburo Okamura, Joachim Scharloth (2015): Wortschatz in Lehrwerken für Deutsch als Fremdsprache: ein frequenzorientierter Ansatz. In: Jana Kiesendahl / Christine Ott (Hrsg.): Linguistik und Schulbuchforschung. Göttingen: V&R unipress. (in press)

Saburo Okamura / Willi Lange / Joachim Scharloth (Hrsg.) (2012): GRUNDWORTSCHATZ DEUTSCH: Lexikografische und fremdsprachendidaktische Perspektiven, 日本独文学会研究叢書 88

Noah Bubenhofer (2012): Lehrwerke und Referenzwortschätze. Der Nutzen frequenzbasierter Grundwortschätze. In: Saburo Okamura / Willi Lange / Joachim Scharloth (Hrsg.) (2012): GRUNDWORTSCHATZ DEUTSCH: Lexikografische und fremdsprachendidaktische Perspektiven, 日本独文学会研究叢書 88, 13-27.

Saburo Okamura / Willi Lange / Joachim Scharloth (2012): METHODEN DER BESTIMMUNG DES KERNWORTSCHATZES DEUTSCH. In: Saburo Okamura / Willi Lange / Joachim Scharloth (Hrsg.): GRUNDWORTSCHATZ DEUTSCH: Lexikografische und

fremdsprachendidaktische
Perspektiven, 日本独文学会研究叢書
88,29-44.

〔その他〕

ホームページ等

基準コーパス Web ページ: (Datengeleiteter
Kernwortschatz Deutsch)

<http://www.basic-german.com>

オンライン・ドイツ語学習用検索ソフト
yomunda!

<http://www.yomunda.com/>

学習用基本語彙(単語カードを含む)

<http://www.basic-german.com/beta/>

6. 研究組織

(1) 研究代表者

岡村 三郎 (OKAMURA Saburo)

早稲田大学・国際学術院・教授

研究者番号: 30009724

(2) 研究分担者

ランゲ、ヴィリィ (LANGE Willi)

早稲田大学商学学術院・教授

研究者番号: 40277752

シャルロット、ヨアヒム (SCHARLOTH
Joachim) (平成24年まで)

獨協大学・外国語学部・准教授

研究者番号: 70585786

(3) 研究協力者

シャルロット、ヨアヒム (SCHARLOTH
Joachim) (平成25年より)

ドレスデン工科大学教授

ブーベンホーファー、ノア (BUBENHOFER
Noah)

IDS, マンハイム 及び ドレスデン工科大学
助手