

平成 30 年 5 月 27 日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K02734

研究課題名(和文) 語彙習得と言語使用：ドイツ語基本語彙の認知的習得モデルの実証的な基盤研究

研究課題名(英文) Language Usage and Vocabulary Acquisition: Empirical Foundations for Cognitive Models of the Acquisition of the Basic German Vocabulary

研究代表者

岡村 三郎 (Okamura, Saburo)

早稲田大学・国際学院・教授

研究者番号：30009724

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：本研究は、ドイツ語基本語彙をそれがMental Lexiconの構造に反映されているように学習者のために示そうとする。このために巨大なコーパスをデータ駆動型の方法により分析、得られたデータを、共起する語、類似した響き、同じテーマ、感情表現の類似のネットワークにまとめた。これらネットワークはMental Lexiconの中心構造なので、我々は相互に関連する構造を持つハイパーテキストとして基本語彙学習ポータル(<http://japanisch.basic-german.com/>)を構築し、CCライセンス(CC BY-SA 4.0)を得た最初のドイツ語基本語彙として既にネット上で公開している。

研究成果の概要(英文)：The purpose of our research is to represent the basic German vocabulary for learners in a way, that corresponds with the structure of the mental lexicon. In order to do so, we modeled findings of data-driven analyses of large corpora as networks of co-occurring, similar sounding, topically coherent, and similarly emotionally charged lexemes. As networks are considered the dominant structure of the mental lexicon, we modeled the vocabulary as a hypertext with complex reference structure. The basic German vocabulary has already been published online (<http://japanisch.basic-german.com/>) as the first basic German vocabulary published under a creative commons license (CC BY-SA 4.0).

研究分野：人文学

キーワード：Mental lexicon「脳内辞書」、ドイツ語基本語彙、認知的習得モデル、データ駆動型、基本語彙の中に存在する種々のネットワーク、ドイツ語基本語彙学習ポータル

1. 研究開始当初の背景

語彙知識の伝達を目的とする伝統的な外国語教授用教科書は学習者のニーズに対応できない。これまでのように学習語彙のリストに頼り、またそれを覆い隠して覚えようとする方法は学習者から単に単調であると受け取られるのみならず、それぞれの語が我々の脳の中に記憶される仕組みも反映していない。すなわち語は一語一語個別にではなく、種々の分類クラスの要素として示されており、そして分類クラスはまた互いにネットワークを築いている (例えば概念のネットワーク、語親族、語の結合、連想のネットワーク、情動のネットワークなど)。

2. 研究の目的

上記の不足を取り除くのが本研究の目的である。

第一に、データ駆動型の手法を使い、先行研究で構築した巨大なコーパスおよび既存の他のドイツ語辞書(とりわけ GermaNet)を利用して、ドイツ語基本語彙の構造の中心的部分をなす以下のネットワークを計算して値を得る。

- 語のテーマ別のネットワーク
- 語親族 (複合語、派生語の分析)
- 共起する語
- 響きの類似
- 感情表現

次にこれらの計算結果を利用し、日本人ドイツ語学習者のための語彙学習ポータル(共通参照枠レベル A1-B1)を相互に関連する構造を持つハイパーテキストの形式で作りに上げる。

3. 研究の方法

語彙の構造はデータ駆動型の方法を使い大きなコーパスから抽出し、目に見える形にした。このデータ駆動型の方法では前もって規定されたアルゴリズムをコーパスに適用し、その際に現れる全てのパターンが計算される。それにより言語使用に関しての新しい包括的な見方が可能になる。これらの分析には先行研究で構築した 10 億語を超える語数を持つコーパスを使った。

語のネットワークが持つ次元を実証的に把握するために以下の方法を利用した：

(1) 響きのネットワーク：全ての可能な語のペアの国際音声記号 (IPA) 表記間での Levenshtein-distance (Damerou 1964) を計算することにより響きが似通った語を同定する。

(2) 意味的なネットワーク

a) テーマのネットワーク：topic models (Steyvers/ Griffiths 2007) を使ってテーマの近似性を算出する。

b) 語親族：Morphisto (ツール) を使って合成語、派生語の分析を行う。

c) それ以外の意味的な関連は GermaNet から取り出す。

(3) 共起する語のネットワーク：文肢に特有な n-gram コロケーションの類似性によって決める。

(4) 感情表現のネットワーク：ライプチヒ大学が開発した SentiWS を利用し、同義語間における感情表現を対比する方法をとる。

このようにして得られたネットワークはハイパーテキストとして互いに結びつけられる。

4. 研究成果

この研究の成果として構築され既に公開されている語彙学習ポータル (<http://japanisch.basics-german.com/>) によって、日本人ドイツ語学習者にとって以下の重要な特徴を持つドイツ語基本語彙が初めて利用できることになった。

(1) この基本語彙は実証的な基盤に立脚している。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)にのっとり、大量のデータを使用し、さらにこれまでに例の無い語彙認定方法を適用するという二つの革新的な特徴を持っている。すなわち膨大な語数を持つドイツ語のプリントメディア、および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 億語を超える十分大規模な基準コーパスを構築し、それをもとに基本語彙を算出した。語彙選定の基準としては、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(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)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得た。

(2) それを元に学習用基本語彙のために、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、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た。我々はとりあえず上位数千語まで、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ることを目標とした。2018 年 5 月現在では 2000 語までその記述が完成している。

(3) それに続き、ドイツ語基本語彙の認知的習得モデルの確実な経験的基盤を得るために、Mental lexicon の中でドイツ語基本語彙の構造の中心的部分をなす以下のネットワークをデータ駆動型の方法で計算して値を得た。

- 語のテーマ別のネットワーク
- 語親族 (複合語、派生語の分析)
- 共起する語
- 響きの類似
- 感情表現

これらの基本語彙の中に存在するネットワークはハイパーテキストによって表現することにより、Mental lexicon が持つ多様な構造に関する認識を適切に示すことができる

ようになった。

(4) 最終的な目標である基本語彙学習ポータル構築に関しては、既にできている基本語彙 2000 語までの「基礎語彙コンパクト」、「フラッシュカード」、「基礎語彙辞書」に加え、これらの基本語彙を「テーマのネット」、「ワードファミリーのネット」、「響きのネット」のグラフの中に位置づけ、さらに「語の結合」を加え、それに「感情表現」をも加えている。これにより日本人ドイツ語学習者が語彙を習得しようとするときに自分の中に語彙のネットワークを組み立てながら効率的に学習するための重要な支えが構築されたことになる。

(5) これまで実験的に使ってみた学習者の意見も取り入れ、基本語彙学習ポータルをより使いやすいものへと改良し、加えてスマートフォンでの使用も視野に入れさらに改良を加えている。

(6) 構築された基本語彙学習ポータル (<http://japanisch.basic-german.com/>) は、Creative Commons license (CC ライセンス) を得た最初のドイツ語基本語彙 (CC BY-SA 4.0) として既にネット上で公開している。

(7) 研究協力者ドレスデン工科大学シャルロット、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及びバチエラー論文が作成された。これらはこれまでに得られた知見をさらに深めるものである。

(博士論文) Frank Nickel (2018): Mit Englisch zum Deutschen. Zur Rolle von Kognaten für den Wortschatzaufbau bei Japanischen Deutschlernern.

(バチエラー論文) Anna Bonazzi (2017): Grundwortschatz Deutsch_ Ein Korpusansatz. Analysen zur Textabdeckung unterschiedlicher Grundwortschätze

Elisabeth Muntshick (2018): Zur Bewertung von Textschwierigkeit am Beispiel von Kinder- und Jugendliteratur - Eine korpuslinguistische Analyse im Kontext DaF.

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)

原口 厚 (2017) 〔資料〕「ドイツ語読解の戦略と戦術 (4) 語彙力の拡充」『文化論集』査読無、第 50 号 早稲田商学同攻会. S. 35-178.

〔学会発表〕(計 2 件)

Grundwortschatz als Hypertext. FaDaF Technische Universität Berlin/ Germany(国際学会)において Scharloth, Joachim / Saburo Okamura / Willi Lange の共同発表、2017 年 4 月 1 日

Wortschatzlernen in Netzen - Interaktive Einführung in den datengeleiteten Grundwortschatz Deutsch für japanische Deutschlerner. 2018 年日本独文学会春季研究発表会における Saburo Okamura, Atsushi Haraguchi, Joachim Scharloth の共同発表、2018 年 5 月 26 日

〔図書〕(計 5 件)

Scharloth, Joachim / Saburo Okamura / Willi Lange / Atsushi Haraguchi (2018): Grundwortschatz Deutsch: Empirische Basis, datengeleitete Modellierung und Präsentation. In: Anna Gryszko, Kristina Pelikan, Thorsten Roelcke (Hrsg.): DaF Für Berlin - Perspektiven für Deutsch als Fremd- und Zweitsprache in Schule, Beruf und Wissenschaft. 44. Jahrestagung des Fachverbandes Deutsch als Fremd- und Zweitsprache an der Technischen Universität Berlin 2017. Göttingen: Universitätsverlag. (in press)

Scharloth, Joachim / Saburo Okamura / Willi Lange (2016): Gibt es einen Kernwortschatz? Datengeleitete Perspektiven auf die Erstellung von Grundwortschätzen für Deutsch als Fremdsprache. In: Simona Brunetti et al. (Hrsg.): Versprachlichung von Welt. Il mondo in parole. Festschrift zum 60. Geburtstag von Maria Lieber. Tübingen: Stauffenburg. S. 273-284.

Lange, Willi / Saburo Okamura / Joachim Scharloth (2016): Datengeleiteter Grundwortschatz Deutsch. In: Peter Colliander / Hans Drumbl / Doris Höhmann / Svitlana Ivanenko / Dagmar Knorr / Sandro Moraldo (Hrsg.): Linguistische Grundlagen für den Sprachunterricht. Bozen-Bolzano University Press. S. 221-230.

Lange, Willi / Saburo Okamura / Joachim Scharloth (2015): Grundwortschatz Deutsch als Fremdsprache: Ein datengeleiteter Ansatz. In: Jörg Kilian/Jan Eckhoff (Hrsg.): Deutscher Wortschatz - beschreiben, lernen, lehren. Beiträge zur Wortschatzarbeit in Wissenschaft, Sprachunterricht, Gesellschaft. Frankfurt am Main u.a.: Peter Lang. S. 203-219.

Bubenhofer, Noah / 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. S. 85-110.

〔その他〕

ホームページ

Datengeleiteter Kernwortschatz Deutsch
<http://www.basic-german.com>

Datengeleiteter Kernwortschatz Deutsch
(基本語彙ポータル)
<http://japanisch.basic-german.com>

Yomunda!

<http://www.yomunda.com>

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

原口 厚 (HARAGUCHI Atsushi)

早稲田大学商学学術院・教授

研究者番号：90247239

(3) 研究協力者

シャルロート、ヨアヒム (SCHARLOTH Joachim)

早稲田大学国際学術院・教授

(4) 研究協力者

ランゲ、ヴィリー (LANGE Willi)

前早稲田大学商学学術院・教授