T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	コーパスに基づく辞書記述の精緻化の研究
Title(English)	
著者(和文)	柏野和佳子
Author(English)	Wakako Kashino
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第10284号, 授与年月日:2016年7月31日, 学位の種別:課程博士, 審査員:高村　大也,新田　克己,寺野　隆雄,三宅　美博,奥村　学
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第10284号, Conferred date:2016/7/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

Powered by T2R2 (Tokyo Institute Research Repository)

（論文博士）

論　文　要　旨 （和文2000字程度）

報告番号	乙 第　　号	氏　名	柏野　和佳子

（ 要　旨 ）

本論文では，コーパスを活用した辞書記述方法を提案する。一般に，辞書の構築は，見出し語選定，語義の区分と記述，見出し語／語義レベル別の情報の記述，という3つの工程に大別される。本論文では，これらの各工程におけるコーパス活用の有効性を具体的な事例を示しながら考察した上で，従来よりも精緻な辞書記述の方法を示す。

まず，第2章で，先行研究と最近の研究を通し，従来の国語辞典と計算機用辞書の問題点を述べる。

次に，第3章で，指摘した問題に対する解決方法を検討する。まず，3.1節において，近年構築され利用が可能になった『現代日本語書き言葉均衡コーパス(Balanced Corpus of Contemporary Written Japanese, 略称BCCWJ)』を中心にコーパスの概要を述べる。次に，3.2節において，コーパス分析に基づくことでどのような辞書記述の改善が期待できるかを先行研究や最近の研究の事例により検討する。これらを踏まえ，3.3節において，本論文で提案する辞書記述方法の基本設計を行い，提案法の新規性と期待される効果について述べる。

続いて，第4章では，辞書記述の骨格となる見出し語選定と語義の区分に関わる検討を行う。即ち，まず，4.1節において**見出し語選定**に関する問題点を指摘し，現代語と古語の間に位置する「古風な語」を例に，コーパスの頻度情報を使った選定方法について述べる。次に，4.2節において**語義の区分**について論じる。従来の「国語辞典」の多義語の記述の分析を述べ，多義語の記述に編者の主観的判断が反映されている現状を示す。その上で，客観的に区分する方法として，コーパス分析に基づく語義の区分方法を示す。最後に，4.3節において**多義構造の記述**について論じる。多義構造(語義間の意味的関係)の明示的記述方法を示し，記述の客観性が向上し得ることを示す。

第5章では，構造化された辞書の記述内容に関わる検討を行う。まず5.1節では，**コロケーション(共起情報)の記述**について論じる。従来，計算機による言語理解において多義性解消は主要な問題の1つと認識されてきた。しかし従来は，主として語の多義性が注目されており，複数の語にまたがる多義についてはほとんど検討されていなかった。例えば「骨が折れる」には骨折するという意味と苦労するという意味とがあるが，後者の意味は「骨」「折れる」のいずれの語の多義としても捉えにくく，語と語とが結び付いたコロケーションの多義性という観点から捉えるのが適当である。5.2節では，まずそのことを指摘し，さらに，コロケーションの多義性の要因が述語(動詞／形容詞)による多義性か，名詞やコロケーション自体の多義性によるものかを見極める必要があることを述べ，そのための辞書記述を提案する。次に，5.2節では，**使用域の記述**について論じる。より自然な言語生成を行うには，内容や文脈に相応しい適切な語を選択することが必要である。しかし，従来の辞書には必ずしも十分な位相情報が記述されていなかった。この問題に対し，まず，5.3節のはじめにおいて，使用域のうち，文体情報と言われるものを辞書記述に盛り込むことを可能にするためには，コーパスの側に文体情報を付与しておく

ことが有用であることを指摘し，その具体的な情報付与の方法を提案する。その次に，「古風な語」を例に，コーパスから得られる出典情報を活用して辞書記述を行う方法を提案する。即ち古風な語の場合，(1)古典の引用での使用，(2)明治期から戦前までの使用，(3)時代・歴史小説での使用，(4)現代文脈での使用，という4分類に基づく辞書記述が，従来の辞書記述に比べて語の文体的特徴をより豊富に記述できることを示す。最後に，5.3節において**頻度情報による時間的使用推移の記述**について論じる。まず，従来の辞書に頻度情報の記述が不足していることによる問題点と頻度情報の有用性を指摘する。次に，外来語を例に，新聞記事データベースより得られる頻度情報により時間的使用推移を明らかにし，それを辞書に記述することで，それぞれの語が流行語であるのか，廃れつつある語であるのか，安定して使われている語であるのかといった区別ができるようになることを示す。

第6章で本論文の結論を述べ，さらに，残された課題を述べる。

本論文は，従来の国語辞書および計算機用辞書の記述において，様々な制約によって十分な記述が行われていなかった側面，および編者の主観的判断に強く依存していた側面のうち，辞書学上あるいは応用上の観点から特に重要と考えられる点に順次光を当て，それぞれ，目的に応じた適切なコーパスを使用することによって，それらの辞書記述がどのように精緻化できるかを明らかにし，その期待される効果を具体的に示したものである。

コーパスの意義は，単に用例を抽出できるだけではなく，統計的手法により，言語に関する定量的な指標を抽出できることにもある。コーパスを活用する辞書記述により，従来の辞書の課題であった，母語話者利用を前提とする情報の省略や生成に必要な情報の不足などの問題が軽減された。また，辞書記述方法の手続き化により，プロの辞書編集者に頼らず，作業ベースでの辞書構築が可能になることにも実際上の意義がある。

備考：論文要旨は、和文2000字と英文300語を1部ずつ提出するか、もしくは英文800語を1部提出してください。
Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).
注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

（論文博士）

論　文　要　旨（英文）　（300語程度）

（Summary）

報告番号	乙 第　　号	氏　名	柏野　和佳子

（要　旨）

This thesis proposes novel methods of building Japanese lexicons based on corpora. The lexical building process can be decomposed into three steps: (1) selection of lexical entries, (2) analysis and structural grouping of meanings for each of those selected lexical entries, and (3) description of meaning, use cases, and other auxiliary information such as register and temporal transitions of usages. Up to now, each of these steps must mostly depend on huge amount of labor work of lexicographers, as well as their subjective decisions or intuitions. In recent years, several corpora have been developed in the world and also in Japan, but little study has been reported, especially for Japanese language, that specifically clarifies how lexical building processes and the resulting lexical descriptions can be improved by utilizing information obtained from such corpora. Against this background, this thesis aims to develop a novel approach to take advantages of the corpus-based method toward Japanese lexicon building, highlighting several important aspects both from a viewpoint of lexicology and from that of engineering. After a brief introduction in chapter 1, chapter 2 first identifies the problems in existing Japanese dictionaries, both ordinary ones that have been made for humans, and the ones built for computers. Chapter 3 then introduces some recently-developed Japanese corpora, and proposes the basic framework of the lexical descriptions. Chapter 4 focuses on the steps of the lexical entry selection and the structural analysis of the meanings, and shows how semantic relations between subentries are objectively identified. Chapter 5 specifically illustrates how the lexical descriptions can be enriched within the identified meaning structures. Finally, chapter 6 concludes this thesis, by summarizing its contributions and the remaining future works.

備考：論文要旨は、和文2000字と英文300語を1部ずつ提出するか、もしくは英文800語を1部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ（T2R2）にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).