

論文 / 著書情報  
Article / Book Information

題目(和文)	
Title(English)	CONTEXTUALLY AWARE WRITING ASSISTANCE SYSTEM FOR JAPANESE
著者(和文)	HODOSCEKBor
Author(English)	Bor HODOSCEK
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9334号, 授与年月日:2013年9月25日, 学位の種別:課程博士, 審査員:室田 真男,中山 実,野原 佳代子,山元 啓史,赤間 啓之
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9334号, Conferred date:2013/9/25, 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

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

専攻：人間行動システム 専攻  
Department of  
学籍番号：  
Student ID Number  
学生氏名：  
Student's Name Hodoscek Bor

申請学位(専攻分野)：博士 (工学)  
Academic Degree Requested Doctor of  
指導教員(主)：室田真男  
Academic Advisor(main)  
指導教員(副)：  
Academic Advisor(sub)

### 要旨(和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

本論文「Contextually Aware Writing Assistance System for Japanese (作文のコンテキストに対応する日本語作文支援システム)」は、日本語学習者のための文脈特性を用いた作文支援システムの開発と評価を目的としたものである。

第1部では文脈特性(レジスター、トピック、リーダビリティ)を意識した作文支援システムの開発に必要な基本的な知識、すなわち、書きことばとしての日本語の特性、科学技術文書の日本語、作文支援システムの現状と問題点、第2言語としての日本語作文教育、コロケーション(共起)の特性を述べている。その上で、本論文における問題提起と研究仮説2点の説明、および方法論を解説し、本学位論文のアウトラインを示している。第2部の中心的課題は、作文支援システムを実現するための開発可能性を検討することである。コーパスによる数理的手法を用いて検討し、3つのモデルを提案している。1)レジスター：ある特定の場面、たとえば科学文書、手紙、新聞記事など、その文章が実際に使われる場面に応じて変化する言語の形態のこと。本研究においては、学術論文の文体や形式を扱い、その場面に応じた形式の適切さを取り扱う概念である。2)トピック：文法的な文の主語としての話題ではなく、文章全体に散らばる内容を反映する語群をいう。3)リーダビリティ：相対的な読みやすさの概念である。相対的とは文章の書き手と読み手との間の共有経験の違いであり、文章の難易度だけを指摘するものではない。この3つのモデルが、どのように互いに関わりあいながら、書くべき内容は作り上げられているのか、ウィルバー・シュラムのコミュニケーションモデルに即して議論している。シュラムのモデルを踏まえ、書きことばにおける伝達モデル、すなわち書き手と読み手の間で行われる情報伝達の仕組みについて述べている。また、各モデルの作文支援システムへの応用可能性について評価を行っている。

第3部では、第1部で概説した問題への有効な解決手法として開発した2つのシステム(ナツメ、ナツメグ)のアーキテクチャ、評価、発展性について詳しく検討した。2つのシステムの仕様および連携は以下の通りである。ナツメはオンラインシステムで、共起パターンの検索および共起の出現パターンがどのジャンルに分布しているのかを明示し、そのパターンに即した例文を提示する。これによりジャンル毎の比較を可能にしている。ナツメの運用実験では、学習者が自分の作文について問題ありと感じているパターンについては、システムが明示した共起パターンに気づくようになったが、学習者自身が問題ありと感ぜない限り、システムによって支援できないという欠点が観察された。それを補う目的でナツメグは設計された。ナツメグも、ナツメ同様のオンラインシステムではあるが、その欠点を補うために学習者が自ら入力した作文を自動解析し、その入力誤用パターンと評価されれば、学習者が問題ありと感ぜないパターンであったとしても注意を促すものである。誤用パターンの判定には、ナツメに実装してある共起データベースを利用している。学習者の作文に見られる共起パターンとデータベースのパターンを比較し、レジスターの逸脱度を計算している。この手法の評価は、ナタネと呼ばれる日本語学習者作文コーパスを利用して行った。さらに、第2部で解説した3つのモデルを2つのシステムへ統合できるかどうか、それにより運用改善が期待できるかどうか検討した。

第4部では結論として、本研究における関係領域への貢献、今後の研究への発展可能性について示唆した。本研究結果の概要は大きく以下の2点に要約できる。1)提案した2つのモデル、トピックおよびレジスターはコーパスにおけるテキストの違いを弁別するのに有効であった。残る1つのモデル、リーダビリティは文単位では、十分に予測はできなかったが、文章単位では、提案した予測関数(文構造の複雑さと語彙の難易レベルを含む関数)は強力ではないが、予測を肯定するものであった。2)開発したシステムは従来のシステムには見られない、レジスターの概念を実装したものである。それは、学習者自身がシステムを利用して、いかに自律的にレジスターの概念を発見するようになるかを意図したものである。特に科学技術文書、学術論文などへの応用として設計されている。最後に、本論文の最も大きな貢献はレジスターを作文支援システムへ応用したことにある。すなわち、レジスターは母語では「当然」のように感じ、意識すらしないものではあるが、第2言語ではそれが活用できない。この問題点を明確にし、学術論文のレジスターにおいての第2言語学習者の持つ負担を軽減させるためのシステムを開発し、評価を行ったことである。

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 2 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 2 copies of 800 Words (English).



(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

専攻 : Department of	人間行動システム	専攻	申請学位 (専攻分野) : Academic Degree Requested	博士 Doctor of	( 工学 )
学籍番号 : Student ID Number			指導教員 (主) : Academic Advisor(main)		室田真男
学生氏名 : Student's Name	Hodoscek Bor		指導教員 (副) : Academic Advisor(sub)		

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words )

Part I of this thesis provided some of the background knowledge necessary for understanding the attempts undertaken within the thesis to implementing various writing assistance systems for second language (L2) Japanese writers, including: 1) the four scripts of the Japanese writing system, 2) the current state of computer-aided writing assistance with a focus on Japanese writing, 3) the current situation of L2 Japanese learners both within Japan and abroad, and 4) the importance of collocations for L2 learning. Part I also framed the two research questions that shaped and informed the various research projects conducted for this thesis.

Part II of the thesis turned to introduce the key concept of writing context and, by taking inspiration from Schramm's model of communication, to trace out how the linguistic concepts of register, topic, and readability interact together in forming the writing context. The primary objective for Part II was to demonstrate the feasibility of utilizing quantitative methods for drawing on corpora to realize systems of writing assistance. In that connection, three distinct, but highly integrated models of register, topic, and readability were proposed and individually evaluated in terms of their respective potential contributions to the overall performance of systems of writing assistance.

As attractive solutions to some of the problems outlined in Part I, Part III of the thesis examined in detail the construction, evaluation, and evolution of the two writing assistance systems, Natsume and Nutmeg. Part III also sought to demonstrate the potential performance improvements that can be realized through the incorporation of the various models developed in Part II.

Finally, Part IV of the thesis provides a summary of the various research contributions proposed throughout the thesis and their relationships to the two research questions posed in Part I, and reflects on some interesting suggestions for future research.

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 2 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 2 copies of 800 Words (English).