

機関番号：32634

研究種目：基盤研究(C)

研究期間：2007～2010

課題番号：19520434

研究課題名（和文） フレーム意味論に基づく多言語語彙検索ソフトウェアの開発

研究課題名（英文） Development of software tools for multilingual lexicons based on frame semantics

研究代表者

佐藤 弘明 (SATO HIROAKI)

専修大学・商学部・教授

研究者番号：60187226

研究成果の概要（和文）：本研究で開発した「多言語語彙検索ソフトウェア」の1つは **FrameSQL** である。このソフトウェアは、(i-iv)の4つの言語の語彙データを同じ検索画面からまるで1つの統合されたデータベースのように検索を行うことができる。(i) 米国・バークレー・フレームネット・プロジェクトの英語データ、(ii) 慶応義塾大学の日本語フレームネット・プロジェクトの日本語データ、(iii) スペイン・バルセロナ自治大学のスペイン語データ、(iv) ドイツ・ザールラント大学のドイツ語データ。

研究成果の概要（英文）：One of the “software tools for multilingual lexicons based on frame semantics” that I developed for this research is **FrameSQL**, a web-based application available at http://sato.fm.senshu-u.ac.jp/frameSQL/fn2_15/notes/index2.html. FrameSQL can search (i) the English lexical database built by the Berkeley FrameNet project, (ii) the Japanese lexical database built by the Japanese FrameNet project at Keio University in Japan, (iii) the Spanish lexical database built at Autonomous University of Barcelona in Spain, and (iv) the German lexical database built by the Saarbrücken Lexical Semantics Acquisition Project. FrameSQL puts together these four lexical databases, and the user can access them seamlessly, as if they were a unified database.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,000,000	300,000	1,300,000
2008年度	600,000	180,000	780,000
2009年度	500,000	150,000	650,000
2010年度	500,000	150,000	650,000
年度			
総計	2,600,000	780,000	3,380,000

研究分野：コーパス言語学、自然言語処理

科研費の分科・細目：言語学・英語学

キーワード：フレーム意味論、FrameNet、語彙、コーパス、項構造、日本語、英語、スペイン語

1. 研究開始当初の背景

言語学者チャールズ・フィルモアは、話者が言語を用いる際に想起する背景知識を捉えるフレーム意味論と呼ばれる言語理論を提唱した。その理論に基づきフィルモアの研究組織 Berkeley FrameNet Project は、大規模

コーパスから採取した英文に意味情報と文法情報を付与した FrameNet と呼ばれるオンライン語彙データベースの構築を続けている。フレーム意味論が扱う背景知識には、言語が使用される場面(Scene)が含まれているが、現在、公開されているほとんどのコーパ

スは文字データのみを扱っており、言語が使用される場面を効率的に検索することができなかった。また、フレーム意味論の枠組みで、スペイン語、日本語、ドイツ語の語彙データベースの構築が進んでいるおり、それらのデータベースを統合する検索ソフトウェアの必要性があった。

2. 研究の目的

フレーム意味論に基づく英語語彙研究のために自己開発してきたソフトウェアと映画コーパス検索システムをスペイン語、ドイツ語、日本語などの多言語に対応させる。それぞれの言語間での対照研究を可能とする検索機能を開発して、世界の研究者に対してインターネット上で無料公開をして、フレーム意味論関連分野の研究発展に寄与する。

3. 研究の方法

本研究は、語彙検索のためのソフトウェア開発を主たる目的としているため、語彙データ入力のための予算は計上していない。そのため、語彙データ入力を行っている他の研究組織との協力関係が重要となる。英語、日本語、スペイン語に関しては、それぞれの言語の語彙データを入力している組織がそれぞれ過去3年間に大きな予算を獲得してきた。それぞれの組織が新しい語彙データの入力作業を行った結果、そのデータを本研究で改良した検索ソフトウェアで利用することが可能となり、新しいデータベースとしてインターネット上に無料で公開できた。また、これらの組織の研究者と意見交換を行うために、研究所を訪れ、直接、研究者と会ってソフトウェアの改良を行った。

4. 研究成果

(1) 開発した2種類のソフトウェア

本研究で開発した「多言語語彙検索ソフトウェア」は2種類ある。1つは、FrameSQLと呼ばれるソフトウェアで、英語、日本語、スペイン語、ドイツ語の語彙データをWebブラウザから検索を行うことが可能である。もう1つは、DVD映画から複数言語の映画字幕・音声データを検索するための多言語対応映画DVD検索ソフトウェアである。

(2) 成果の公開方法の違い

2つのソフトウェアの成果公開方法には、1点大きな違いがある。FrameSQLに関しては、「5. 主な発表論文等」に記載した研究論文と研究発表に加えて、FrameSQLのソフトウェア自体をインターネットで全世界の研究者に公開している。これは、データの提供を受けている研究機関からデータのインターネット上での公開の許諾を受けているため、データを含めた完全公開が可能となった。

一方、多言語対応映画DVD検索ソフトウェアに関しては、複数言語の字幕・音声データを収録した映画DVD資料を購入し、そのデータを検索し、映画で場面を再生するソフトウェアの開発には成功し、その成果を雑誌論文の項目に記載した④「DVD映画を利用した日英マルチメディア・コーパスの構築」と②「iPod touchを利用した多言語マルチメディア・パラレル・コーパスの構築」で発表した。しかし、ソフトウェア自体をインターネット上には公開していない。その理由は、映画字幕・音声データを一般公開するためには、著作権者である映画会社に高額な著作物使用料金を支払わなければならないからである。

(3) 多言語対応 FrameSQL

このソフトウェアは、(i-iv)の4つの言語の語彙データを同じ検索画面から1つの統合されたデータベースのように検索を行うことができる。(i) International Computer Science Institute(米国・バークレー市)のバークレー・フレームネット・プロジェクト(<http://framenet.icsi.berkeley.edu/>)が開発を続けているバークレー・フレームネット・データベース(英語)、(ii) 慶応義塾大学の日本語フレームネット・プロジェクト(<http://jfn.st.hc.keio.ac.jp/>)が開発を続けている日本語フレームネット・データベース、(iii) スペイン・バルセロナ自治大学のスペイン語フレームネット・プロジェクトが開発を続けているスペイン語フレームネット・データベース、(iv) ドイツ・ザールラント大学のSALSA IIプロジェクト(<http://www.coli.uni-saarland.de/projects/salsa/page.php?id=index>)が開発したドイツ語語彙データベース。



図1: FrameSQL 検索画面

(i-iv)のプロジェクトは、それぞれのサイトで語彙データを配布しているが、FrameSQL

を使用すれば、これらのデータをユーザーがダウンロードしなくても、FrameSQLのサイト (http://sato.fm.senshu-u.ac.jp/frameSQL/fn2_15/notes/index.html) で検索が行える。

FrameSQLは、無料で全世界の研究者に公開されており、多くの国からインターネットを通じて利用されてきた。また最新のフレーム意味論について論じた論文(v)の337ページでも、私の開発したFrameSQLに関する解説があり、現在でも米国・パークレー・フレームネット・プロジェクトの研究者がFrameSQLを評価していることが分かる。

(v) Charles J. Fillmore and Collin F. Baker (2010) “A Frames Approach to Semantic Analysis,” *the Oxford Handbook of Linguistic Analysis* by Bernd Heine and Heiko Narrog, 2010, pp.313-339, Oxford University Press.

(4) 多言語対応映画DVD検索ソフトウェア雑誌論文の項目に記載した④「DVD映画を利用した日英マルチメディア・コーパスの構築」では、本研究課題で開発したソフトウェアを使用して、DVD映画日本語字幕を効率的にコーパスとして利用する方法を示した。このソフトウェアを利用して作成した日英パラレル・コーパスからの検索例を示しながら、日英の文字情報だけではなく、音声・映像情報が活用できるDVD映画日英マルチメディア・コーパスについて論じた。

④「iPod touchを利用した多言語マルチメディア・パラレル・コーパスの構築」では、本研究課題で開発したソフトウェアを利用して、多言語DVDから多言語マルチメディア・パラレル・コーパスを構築する方法を述べた。構築するコーパスは、アップル社の携帯音楽プレーヤーiPod touchで利用可能である。



図2: iPod touchで検索結果

このコーパスには、多言語DVDから取り出した英語、スペイン語、イタリア語、フランス語、ドイツ語の字幕・音声データが納められており、どの言語の文字検索もiPod touch上で実行できる。図2では、英語の*good evening*をDVD映画『ハリー・ポッターと賢者の石』(2001年)から検索し、それに対応するスペイン語*buenas noches*、イタリア語*buona sera*、フランス語*bonsoir*、ドイツ語*Guten Abend*を表示している。検索結果は、文字だけでなく音声でも確認できるため、複数の言語の文字と音声を対照研究することが可能である。さらに、コーパスで使用されているすべての言語の3文字以上の長さの単語を電子辞書で自動的に調べて、その単語の意味を確認できる機能を備えており、iPod touchで語句の意味の確認と音声の確認を行いながら、効率的な対照研究が行える。

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5件)

① Hiroaki Sato, “How FrameSQL Shows the Japanese FrameNet Data,” *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, 2010年, USB記憶媒体収録のためページ番号なし, 査読有り.

④ 佐藤弘明, 「iPod touchを利用した多言語マルチメディア・パラレル・コーパスの構築」, 『ことばの普遍と変容』第5号, pp.67-74, 専修大学社会知性開発研究センター, 2010年, 査読無し.

③ Hiroaki Sato, “Browsing Japanese FrameNet with FrameSQL,” *Proceedings of Pacific Association for Computational Linguistics 2009 (PACLING 2009)*, 2009年, CD-ROM収録のためページ番号なし, 査読有り.

④ 佐藤弘明, 「DVD映画を利用した日英マルチメディア・コーパスの構築」, 『ことばの普遍と変容』第4号, pp.111-122, 専修大学社会知性開発研究センター, 2009年, 査読無し.

⑤ Hiroaki Sato, “New Functions of FrameSQL for Multilingual FrameNet,” *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, 2008年, CD-ROM収録のためページ番号なし, 査読有り.

[学会発表] (計4件)

① Ohara, Kyoko Hirose & Hiroaki Sato, “Investigating Japanese FrameNet Data

withFrameSQL,” 2010年9月5日, チェコ・プラハ

② Hiroaki Sato, “Browsing Japanese FrameNet with FrameSQL,” Pacific Association for Computational Linguistics (PACLING 2009), 2009年9月3日, 札幌・北海道大学.

③ Hiroaki Sato, “Search Functions of FrameSQL for FrameNet,” International Conference on Middle and Modern English Corpus Linguistics, 2009年7月9日, オーストリア・インスブルック.

④ Hiroaki Sato, “New Functions of FrameSQL for Multilingual FrameNet,” The Sixth International Conference on Language Resources and Evaluation, 2008年5月28日, モロッコ・マラケシュ.

[その他]

ホームページ等

http://sato.fm.senshu-u.ac.jp/frameSQL/fn2_15/notes/index2.html

6. 研究組織

(1) 研究代表者

佐藤 弘明 (SATO HIROAKI)

専修大学・商学部・教授

研究者番号 : 60187226