

機関番号：12608

研究種目：基盤研究（B）

研究期間：2008～2010

課題番号：20300074

研究課題名（和文）知識共有のための価値指向型オントロジーの多分野多言語化

研究課題名（英文）Multi-disciplinary and multi-language development of the value oriented ontology

研究代表者

往住 彰文（TOKOSUMI AKIFUMI）

東京工業大学・大学院社会理工学研究科・教授

研究者番号：50125332

研究成果の概要（和文）：

人類が蓄積してきた膨大な知識資源を効果的に利用する方策のひとつとして、人間の知的活動や感性的活動にできるだけ近似した知識表現形を機械可読な形で提供したい。文学、芸術、政治、宗教といった、人間の能力の最高の活動場面で流通している言語テキストを対象としてオントロジーの構築を試み、多分野、多言語における検討をおこなった。

研究成果の概要（英文）：

Properly designed knowledge and affective representation in machine readable form will be a powerful tool to utilize the large-scale text resources we are currently confronting. We analyzed various types of value-laden texts from literature, art, politics and religion, and discussed aspects of ontology design principles.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	4,400,000	1,320,000	5,720,000
2009年度	3,800,000	1,140,000	4,940,000
2010年度	3,600,000	1,080,000	4,680,000
年度			
年度			
総計	11,800,000	3,540,000	15,340,000

研究分野：総合領域

科研費の分科・細目：情報学 ・ 感性情報学・ソフトコンピューティング

キーワード：知識表現，オントロジー，テキスト分析，人文工学

1. 研究開始当初の背景

情報爆発の時代を迎え、価値ある情報の取捨選択は大きな課題となりつつある。しかしながら、旧来の検索エンジンなどでピックアップ、あるいは要約提示される結果は、情報が置かれていたコンテキストから切り離し、断片化した知識に過ぎない。人類的な規模で莫大に蓄積されてきた情報を、知識の断片としてではなく、本来それが置かれていたコンテキストに沿って理解するためには、その情報が生成された分野・言語・コミュニティ特有の価値構造の理解が不可欠である。従来の

情報処理にあわせて、分野・言語・コミュニティに特有の価値情報を解析・提示することで、人類的コンテキストに基づいた深遠な意味処理が実現する。また、ある価値体系における知識を、他の価値体系における語彙で再表現することによって、知識に新たな意味を付加する豊穡な多様化が可能となる。

これまで、テキストから価値情報を抽出する研究は、情報工学の評判分析などの分野で少なからずおこなわれているが、価値極性が明確な語彙（e.g. 良い、悪い）と統語構造を利用したテキスト解析システムが中心であり、

人間が持つ多次元的な価値構造に対応するには不十分である。

これに対して、認知的文化人類学の知見では感情表現語彙は認知特性に分解可能で、多言語間の価値表現語彙を認知特性によって対応させることのできる可能性が示唆されている。

本研究では、認知科学をはじめとして、文学・哲学などの文系の諸学の知見を活用することで、従来の情報学のみでは実現できなかった、より高次で深遠な意味の処理を実現することを旨とする。

2. 研究の目的

人類的な規模で莫大に蓄積されてきた情報を、知識の断片としてではなく、その情報が生成された分野・言語・コミュニティ特有の価値構造と共に提示することが目的である。従来の情報処理にあわせて、コミュニティ特有の価値情報を解析・提示することで、人類的コンテキストに基づいた深遠な意味処理が実現する。価値観が多様化した現代社会で、多コミュニティの価値構造を容易に可視化することで、知識の豊穡化が可能となる。本研究では、人文科学及び社会科学における知識情報集積の将来的な共同研究拠点の形成に向け、複数の領域におけるケーススタディ的な知識豊穡化の実現と、システム的な基盤構築およびその検証を行う。

(1) 多分野のテキスト解析をおこない、政治・イデオロギー、芸術・文学、思想の各分野における価値オントロジーを構築する。具体的には、価値を表現する言語表現（単語・フレーズ）について共起・依存関係とともに構造化し、表示する。

(2) 多言語のテキスト解析をおこない、日本語、英語、フランス語、ラテン語、ギリシア語の各言語における価値オントロジーを構築する。具体的には、各言語における価値に関わる言語表現について共起関係とともに構造化し、表示する。

(3) 普遍的な価値表現、分野に依存する価値表現、言語に依存する価値表現とを抽出し、価値の伝播モデルを構築し、価値の形成メカニズムを解明する。

3. 研究の方法

人類の知の資産は、言語によるテキストという形でもっともよく継承されているといえる。画像、映像、音、さらには身体といったメディアも知の活動に広がりや深みを与えているのは間違いないが、それらの意味を捉え、固定化するためには言語テキスト以上のメディアは存在しない。したがって、本研究の目的をもっともよく達成するためには、言語テキストの分析が王道である。

テキストの分析に関わる以下の3つの方法を同時並行的にすすめ、対象テキストの根底

にある概念を抽出し、オントロジーの構築をおこなった。互いの方法論の依存性を確認しながら、方法自体の調整もおこなった。

(1) テキスト解析とオントロジー構築を支援するためのソフトウェアツールの開発をおこなった。入力テキストに対し、計量分析、ネットワーク分析、グラフ表示をおこなう統合的テキスト解析ツールと、オントロジーのデータベースをGUIで入力するオントロジー・エディタである。

(2) 多分野の具体的対象として、政治・イデオロギー、芸術・文学、思想、宗教の各分野を選び、計量分析とネットワーク分析を基本とするテキストの分析をおこなった。具体的には、テキストに対して、形態素解析と統語解析をおこなった上で、単語の頻度分布、テキスト上での遷移、単語間の依存関係、単語クラスターの抽出の分析をおこなった。これらは、自動的な意味分析への近似として、現在可能な手法を尽くしたものである。

(3) 多言語の具体的対象として、日本語、中国語、韓国語、英語、フランス語、ギリシア語を選び、各言語における概念の特徴を明らかにするために言語間の比較をおこなった。研究期間内におこなった比較は、比較的少数の概念を対象としたもので、精妙で微妙なニュアンスまでをも明らかにするために、既存辞書の分析、大規模コーパスの利用、心理実験を組み合わせている方法論を用いた。

4. 研究成果

(1) テキスト解析とオントロジー構築のための人工学ツールは、研究期間の初期の段階で完成し、研究のほぼ全期間を通じてこれらのソフトウェアツールを活用した(図1)。

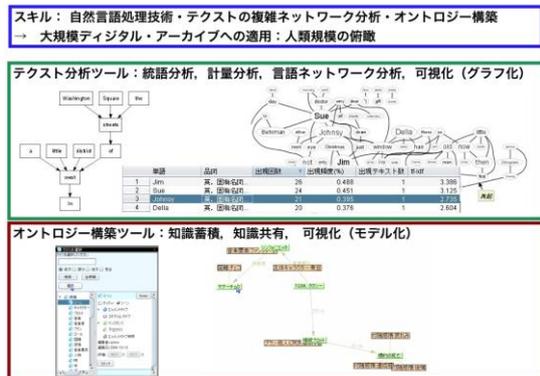


図1 人工学ツール

テキスト解析ツールは、テキストの入力から、分析前の整形、形態素解析、統語解析、単語単位の計量分析、ネットワーク分析を経て、分析結果の可視化までを支援する。形態素解析(MeCab, TreeTagger)、統語解析(CaboCha, MaltParser)、グラフ表示(GraphViz)は他の研究者によって開発および公開され、高い評価を得ているものを利用し、計量分析とネットワーク分析の部分を開発して統合したものである(川島ら, 2010)。本ツール(Text Seer)はBSDライセンスにもとづくオープンソフトウェアとして公開し、バイナリとソースコード

ドを配布している。オントロジー・エディタは、上位下位、全体部分、同義などの関係を定義した上で概念の内部構造と概念間の関係をGUI上の操作で入力する概念データベースである。

こうしたソフトウェアツールの効果は予想をはるかに上回るもので、以後の研究を極めて効率よく推進できることとなった。以下にその主なものをまとめる。

(2) 多分野としてとりあげた分野のひとつが、文学である。テキスト解析とオントロジーの方法論で、ある作家の文学的思考空間を、小説テキストの分析によって抽出することを試みた。工藤ら(2009, 2010, 2011)は、村上春樹の長編小説を対象として、現時点において可能な限り精緻な計量分析をおこなった。その結果、12編の長編小説の文体上の特徴をクラスタ分析で特定し、これまで作風の変化として語られていた現象の定量的な確認することができた。また複数の物語が章の交代とともに進行する並行形式をもつ作品(e.g. 1Q84)の内部構造を語彙の分布と因子分析によって緻密に特定し、この特異な小説形式の効果を描きだすことができた。こうした一連の研究の結果として明らかになってきたのは、たとえば村上春樹がその知的所産、すなわち小説作品を産み出す源泉となっている思考空間が、われわれが村上春樹オントロジーとして捉えようと試みているものであるが、それは静的固定的に存在しているのではなく、動的な変動をもつものであること、その変動は、村上春樹の生涯といった時間軸とともに、作品の内部構造という時間軸でも捉え得るということである。研究期間終了の時点で、村上春樹の動的オントロジーを独立して構築するという段階には至らなかったが、文学領域のオントロジー設計における極めて重要な示唆を得ることができたと考えている。

(3) 同じく文学領域での対象として、村井ら(2011)は、星新一のほぼ全短編を対象とし、物語プロットの定量的分類とテーマ語(特徴語)の自動抽出をおこなった。その結果、物語プロット(いわゆる「落ち」の構造)とテーマ(主たる内容)の間の強い関連性を見いだすことに成功した。この研究例も、文学的分析の相当に深い部分までもが、客観的な基準による分析で捉えられることを示している。また、星新一オントロジーは、語彙の形をとる概念だけではなく、物語プロットの形をとる概念も不可欠のものとして含むことを示唆している。

(4) 政治・イデオロギーの分野での対象として、村井ら(2008)は、国会議員615人のWebサイトからテキストを収集し、政治感性語データベースを作成して、政治的ストの全体傾向と、政党ごとの語彙的特徴を明らかにした。政治テキストは、量(大, 小)、様相(良い, 悪い)、時間(新しい, はやい)といった特徴づけを持つこと、政党間の差は、伝統や文化に関わるニュアンスに現れることなど、政治オントロジーの興味深い特徴を含むものである。

(5) 多言語オントロジーの分析例として、往住

ら(2010)は、「かわいい」という概念の日中比較を試みている。オントロジー構築のために認知的感情理論に基づいて審美感情の理論を整備し、認知的要素とオントロジー項目の対応づけを提案した。広く流通している辞書の分析、大規模コーパスにおける「かわいい」と共起する語の利用、心理実験を複合して用い、「かわいい」の概念内容と語彙使用者の心的状態をオントロジーという形で明示化した。小さい、丸い、女性的という認知的要素の適用条件が緩く、極めて広い対象物に対して「かわいい」を使う傾向がある日本語と、適用条件が厳しく、対象が限定されている中国語の違いを明確に説明することができた。さらに、このオントロジーを文学作品の分析に使用した例も報告している。

(6) このように、本研究によるアプローチによって、データに基づく文学研究、政治イデオロギーの研究が可能であることを示すことができた。これらの他にも、医療に関連するデータ、宗教テキストの分析をおこない、成果を発表した。3年間という短期間のうちに、相当数のテキスト分析をおこない成果をあげることができたのは、テキスト分析とオントロジー構築の人文工学ツールによるところが大きい。

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20件)

1. 村井源・松本斉子・佐藤知恵・往住彰文, 物語構造の計量分析に向けて: 星新一のショートショート物語構造の特徴, 情報知識学会誌, 21, [1], 2011, pp. 6-17, 査読有.
2. 工藤彰・村井源・往住彰文, 計量分析による村上春樹長篇の関係性と歴史の変遷, 情報知識学会誌, 21, [1], 2011, pp. 18-36, 査読有.
3. 村井源, 漸近的対応語彙推定法に基づく翻訳文の解釈的特徴の抽出: 日本語翻訳聖書の計量的比較, 情報知識学会誌, 20, [3], 2010, pp. 293-310, 査読有.
4. 川島隆徳, 村井源, 往住彰文, ゲーム批評から見たゲームの「面白さ」: レビューテキストの計量解析による叙述対象の自動抽出, デジタルゲーム学研究, 4, [1], pp. 69-80, 2010, 査読有.
5. Kawase, A. and Tokosumi, A. Regional Classification of Traditional Japanese Folk Songs: Classification Using Cluster Analysis. Kansei Engineering International Journal, 10, [1], pp. 19-27, 2010, 査読有.
6. Leith Morton, 文学的構築物としての恋愛, FLC言語文化論集POLYPHONIA, 2, pp. 95-110, 2010, 査読有.
7. Leith Morton, Modernity and Japanese

Culture, 言語文化論叢, 15, pp. 39-63, 2010, 査読有.

8. 村井源, 川島隆徳, 徂住彰文, 医療の質・安全研究における関心領域の分析, 医療の質・安全学会誌, 4, pp. 25-33, 2009, 査読有.

9. 村井源, マルコ福音書が多層集中構造, 日本カトリック神学会誌, 20, pp. 65-95, 2009, 査読有.

10. Leith Morton, 与謝野晶子の『みだれ髪』と鉄幹の『紫』のモダニズム, 言語文化論叢, 14, pp. 45-62, 2009, 査読有.

11. 高岸輝, 室町絵巻の環境と表現: 土佐行広から土佐光信・土佐光茂へ, 日本文学, 58, pp. 41-48, 2009, 査読有.

12. 高岸輝, 土佐光信のコミュニケーション: 絵師と画料をめぐる, 文学, 10, pp. 174-180, 2009, 査読無.

13. 村井源, 松本齊子, 山本竜大, 徂住彰文, Webの計量言語学的分析からみた政治的感性の特徴, 感性工学会研究論文集, 7, [3], pp. 561-569, 2008, 査読有.

14. 村井源, 山本竜大, 徂住彰文: Webサイトデータを活用した計量的人間関係解析のための指針: 日本の国会議員Webサイトからみた政治家の中心性とグループ, 理論と方法, 23, [1], pp. 110-128, 2008, 査読有.

15. 高岸輝, 十五世紀絵画のバースペクティブ-土佐光信のリアリズム, 文学, 9, [3], pp. 99-108, 2008, 査読有.

16. Leith Morton, A Tradução do Tanka [Translating Tanka], Dialogia [San Paulo], 7, [1], pp. 31-40, 2008, 査読無.

17. Leith Morton, 『みだれ髪』「白百合」の研究: 世紀末のテキスト交錯の一例として, 言語文化論叢, 13, pp. 39-56, 2008, 査読有.

18. Leith Morton, 鉄幹・晶子・登美子の歌垣, 言語文化論叢, 12, pp. 1-20, 2008, 査読有.

19. Leith Morton, 19世紀末における恋愛文学の編成: 人情本から小説へ, 日本研究, 37, pp. 259-291, 2008, 査読有.

20. Leith Morton, 戦争詩論, 比較文学研究, 91, pp. 43-63, 2008, 査読有.

[学会発表] (計43件)

1. 藤文娜・川島隆徳・村井源・徂住彰文, エンターテインメントコンテンツ作品の相互関係性. 第6回日本感性工学会春季大会, 2011年3月3日, 福岡.

2. 村井源, デジタルアーカイブへのIntra-textualityの導入に向けて, 人文科学とコンピュータ・シンポジウム, 2010年12月12日, 東京.

3. 工藤彰・村井源・徂住彰文, 村上春樹の計量的変遷と共時フィクションの語彙形成. 人文科学とコンピュータ・シンポジウム, 2010年12月11日, 東京.

4. 河瀬彰宏・徂住彰文, ネットワーク表現を

用いた民謡の音楽的概念の比較, 人文科学とコンピュータ・シンポジウム, 2010年12月11日, 東京.

5. Murai, H., The Parallel Concentric Structures within Exodus, Society of Biblical Literature Annual Meeting, 2010年11月21日, Atlanta, USA.

6. 徂住彰文・川島隆徳・工藤彰・村井源, デジタル人文工学の必要性. 第12回日本感性工学会大会, 2010年9月12日, 東京.

7. 藤文娜・徂住彰文, 「かわいい」オントロジーの日中比較. 第12回日本感性工学会大会, 2010年9月12日, 東京.

8. 川島隆徳・村井源・徂住彰文, デジタル人文工学のためのテキスト解析ソフトウェア. 第12回日本感性工学会大会, 2010年9月12日, 東京.

9. 工藤彰・村井源・徂住彰文, デジタル人文工学における文学解釈の可能性. 第12回日本感性工学会大会, 2010年9月12日, 東京.

10. 河瀬彰宏・徂住彰文, 音楽認知研究に関する近年の動向: 計算機を用いた音楽構造抽出の研究. 第12回日本感性工学会大会, 2010年9月12日, 東京.

11. 成田彩夏・徂住彰文, SNSコメントの大量分析による摂食障害の特徴づけ, 第12回日本感性工学会大会, 2010年9月11日, 東京.

12. Kudo, A., Murai, H. and Tokosumi, A., Historical Shifts in Writing Style within the Works of Haruki Murakami Identified using Semantic Analysis, The Seventh International Conference on Cognitive Science, 2010年8月18日, Beijing, China.

13. Kawase, A., Murai, H. and Tokosumi, A., Extracting Conceptual Features and Observing Relations for Composers from Music Criticism, The Seventh International Conference on Cognitive Science, 2010年8月18日, Beijing, China.

14. 村井源・徂住彰文, 文芸批評の計量解析による批評行為の背景的特徴の抽出. 情報知識学会第18回年次大会, 2010年5月15日, 東京.

15. 佐藤知恵・村井源・徂住彰文, 星新一ショートショート文学の物語パターン抽出. 情報知識学会第18回年次大会, 2010年5月15日, 東京.

16. 河瀬彰宏・村井源・徂住彰文, 音楽評論文にみる作曲家の感性的特徴. 情報知識学会第18回年次大会, 2010年5月15日, 東京.

17. 工藤彰・村井源・徂住彰文, 計量分析による村上春樹の語彙構成と歴史的変遷. 情報知識学会第18回年次大会, 2010年5月15日, 東京.

18. 徂住彰文・藤文娜, 「かわいい」オントロジー: デジタル人文工学からのアプローチ, 日本感性工学会かわいい人工物研究部会第1回研究会, 2010年5月15日, 東京.

19. Kawase, A. and Tokosumi, A., Estimating the Minimum Structures of Musical Schemas from Traditional Japanese and Chinese Folk Songs, The International Conference on Kansei Engineering and Emotion Research, 2010年3月2日, Paris, France.
20. 村井源, 聖書における多層集中構造の妥当性の検討, 日本カトリック神学会第21回学術大会, 2009年9月14日, 東京.
21. Tokosumi, A. and Murai, H., Extracting High-Order Aesthetic and Affective Components from Composer's Writings, Human Computer Interaction International, 2009年7月28日, San Diego, USA.
22. 村井源, 往住彰文, テキスト批評の計量化に向けて: 書評の計量分析, 情報知識学会第17回年次大会, 2009年5月16日, 東京.
23. 河瀬彰宏, 村井源, 往住彰文, 音楽評論論文にみる概念構造の変遷: ネットワーク中心性を用いた音楽概念の抽出, 情報知識学会第17回年次大会, 2009年5月16日, 東京.
24. 工藤彰, 村井源, 往住彰文, 村上春樹の初期三部作における構造解析, 情報知識学会第17回年次大会, 2009年5月16日, 東京.
25. 佐藤知恵, 村井源, 往住彰文, 文学作品群の特徴的語彙と概念カテゴリーの抽出: 星新一ショートショート of 計量分析, 情報知識学会第17回年次大会, 2009年5月16日, 東京.
26. 齊藤香里, 村井源, 往住彰文, 心の状態と言語的特徴: ブログにおける商品紹介文の分析, 情報知識学会第17回年次大会, 2009年5月16日, 東京.
27. 山内理嗣, 往住彰文, 発話の内容分析からみた法情報検索の認知メカニズム, 第56回日本図書館情報学会研究大会, 2008年11月15日, 奈良.
28. 村井源, 松本斉子, 往住彰文, コーパスに基づく感性と論理のネットワーク表現, 日本認知科学会第25回大会発表論文集 2008年9月6日, 京田辺.
29. 大楠哲平, 村井源, 往住彰文, テキスト解析によるアイニシュタインの論文の分類, 第10回日本感性工学会大会, 2008年9月9日, 東京.
30. 佐藤さやか, 松本斉子, 村井源, 往住彰文, テキストの内容分析による文芸批評の認知構造の推定, 第10回日本感性工学会大会, 2008年9月9日, 東京.

〔図書〕(計5件)

1. 高岸輝, 竹林舎, 中世における絵巻の収集享受と権力, 高橋亨編: 王朝文学と物語絵. 平安文学と隣接諸学10, 2010, pp. 75-90.
2. 高岸輝, 青簡舎, 「清水寺縁起絵巻」の空間と国土, 佐野みどり・新川哲雄・藤原重雄編: 中世絵画のマトリックス, 2010,

pp. 349-360.

3. Leith Morton, Hawaii University Press, The Alien Within: Representations of the Exotic in Twentieth-Century Japanese Literature, 2009, 272.

4. 高岸輝, 吉川弘文館, 室町絵巻の魔力: 再生と創造の中世, 2008, 199.

5. 高岸輝, 笠間書院, お伽草子百花繚乱, 2008, pp. 587-600.

〔その他〕

ホームページ等

<http://www.valdes.titech.ac.jp/~akt/research/>

6. 研究組織

(1) 研究代表者

往住 彰文 (TOKOSUMI AKIFUMI)

東京工業大学・大学院社会理工学研究科・教授

研究者番号: 50125332

(2) 研究分担者

村井 源 (MURAI HAJIME)

東京工業大学・大学院社会理工学研究科・助教

研究者番号: 70452018

井口 時男 (IGUCHI TOKIO)

東京工業大学・大学院社会理工学研究科・教授

研究者番号: 80232164

モートン リース (MORTON LEITH)

東京工業大学・外国語研究教育センター・教授

研究者番号: 40361787

高岸 輝 (TAKAGISHI AKIRA)

東京工業大学・大学院社会理工学研究科・准教授

研究者番号: 80416263