

機関番号：32621

研究種目：基盤研究（B）

研究期間：2007～2010

課題番号：19300284

研究課題名（和文）協調学習の情報抽出とグループを超えた再利用の研究

研究課題名（英文）Study on Information Extraction and Reuse on CSCL

研究代表者

田村 恭久 (TAMURA YASUHISA)

上智大学・理工学部・准教授

研究者番号：30255715

研究成果の概要（和文）：協調学習(CSCL)発言の自動参照支援を目的とし、自然言語処理および人手による参照情報付加を行うシステムを開発した。まず、単文中の SVO 情報（主語-述語-目的語）を付加する機能を実現した。既存の自然言語処理結果を補完するため、発言者に SVO 情報を提示し、それを修正するインタフェースを開発した。次に、複数文の意味関係を顕在化させるため、修辞構造理論に基づいた文間関係候補を提示し、発言者が修正する機能を開発した。

研究成果の概要（英文）：Authors developed an NLP (Natural Language Processing) based CSCL utterance analysis and semantic metadata attachment system. Conventional NLP software does not identify SVO metadata on single sentence, and relationship metadata among sentences. The proposing system supposes the metadata with use of conventional NLP and helps to identify them with use of analogical inference and user's data input.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	5,300,000	1,590,000	6,890,000
2008年度	4,100,000	1,230,000	5,330,000
2009年度	3,800,000	1,140,000	4,940,000
2010年度	2,000,000	600,000	2,600,000
年度			
総計	15,200,000	4,560,000	19,760,000

研究分野：教育工学

科研費の分科・細目：科学教育・教育工学、教育工学

キーワード：協調学習、CSCL、Web サービス、自然言語処理、再利用

1. 研究開始当初の背景

従来の CSCL 研究は、(1)発言入力環境の提供（電子掲示板、Wiki、SNS 等）、(2)発言役割の人手入力による発言シーケンスの視覚化、(3)発言順序や役割を強制的に指定する環境の提供、に大別され、発言内容をコンピュータが分析し、協調学習支援を行うものがない。このため、表面的な支援に留まり、発言の内容や意味に踏み込んで支援できなかった。

2. 研究の目的

前述の背景を踏まえ、CSCL の発言を自然言語処理を用いて分析し、また発言者の人手入力により既存記述の不足分を補い、発言の内容や意味をコンピュータが分析できる程度の説明情報を発言に付加する。これにより、関連発言の自動抽出や、より精度の高い発言間関連の可視化を目指す。

3. 研究の方法

研究開始当初、既存の自然言語処理を用いた

場合の精度や意味把握の限界が不明であったので、Chasen などの NLP オープンソースソフトを導入し、機能を検証した。この結果、単語（形態素）の品詞特定は正しく行えるが、文中の単語間の関係（係り受け関係）が正しく行えない場合が多く、また複数文の意味的な関係を把握することはコンピュータでは不可能であることが判明した。このため、以下の研究開発を行った。

3.1 オントロジー構造

本稿で提案するオントロジーは、まず行っている議論の中心となる単語（Main Word）を1つ決定する。次に、この中心単語に以下の関連のある単語をピックアップする。ここで中心単語に関連する概念を以下のように定める。なお、このオントロジーは後述する入力インターフェースから入力され、自動で XML として保存される。この XML で保存される際に使用されるタグを括弧内に示す。

- (1) 上位概念 (Upper Concept) : is-a の関係の上位。中心単語を含む、より一般的、より総称的、より抽象的なもの
- (2) 下位概念 (Lower Concept) : is-a の関係の下位。中心単語の指すもののうち、より特定の、より個別の、より具体的なもの
- (3) 同義語 (Synonym) : 同一概念。中心単語と同じ意味のもの
- (4) 母体 (Integration) : has-a の関係の上位。中心単語を含んでいる（持っている）もの
- (5) 部品 (Parts) : has-a の関係の下位。中心単語が含んでいる（持っている）もの
- (6) 対象 (Target) : 中心単語が使用される対象または目的
- (7) その他 (Other Relation) : 上記以外で中心単語となんらかの関係があるもの（類似概念など）

上記のオントロジーの対象となる単語は名詞であり、かつ代名詞などを除去する必要がある。これらのオントロジーに使用される単語は、議論の内容を表現するため議論中に出現した単語のみを用いて構築される。ただし、同義語は中心単語と同じ意味の単語であるため、また対象は学習者内で暗黙の了解として議論される可能性があるため、議論中に出現していなくても用いることができるものとする。また、各関連単語はそれぞれ最大5つずつ設定可能となっている。中心単語は必ず1つ決定しなければならないが、各関連単語については当てはまる単語が存在しなければ空白でかまわないものとする。

3.2 類似議論取得までの流れ

図1に、ある学習者グループが別の学習者グループが行った類似する議論のログを取得するまでの流れを示す。

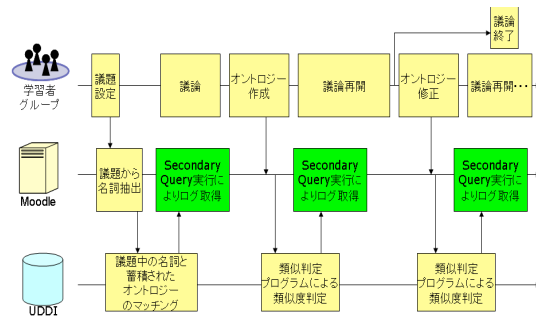


図1 処理の流れ

本稿のシステムではオープンソースの形態素解析プログラム MeCab を用いて、発言者が入力したテキストから形態素を抽出する。これにより、テキストに含まれる各々の形態素とその品詞を得ることができる。オントロジーは名詞が対象となるが、代名詞は示す対象が抽象的でオントロジーとしては不適切である。このため、MeCab の解析結果からまず名詞を抽出し、次に作成した代名詞リストを用いて代名詞を除外して、オントロジー候補を抽出する。

まず、学習者グループが議題を設定したとき、この議題に対し上記の処理を行い、オントロジー候補を抽出する。これを Primary Query で UDDI に通信し、すでに UDDI に登録されている議論のオントロジーとマッチングを行い、マッチングする議論があれば Secondary Query を実行し、類似する議論を取得する。

この後、学習者グループの議論がある程度進んだ時点で、すでに議論に登場した単語を用いてオントロジーを構築する。このオントロジー構築には学習者グループ内の1名が代表して入力を行うが、入力する内容は学習者グループ内のほかのメンバーと議論を行い、合意のもとに入力を行う。なお、議論進行と同時に発言内容に対してプログラムによる形態素解析が行われ名詞が抽出されており、中心単語はすでに抽出された名詞のリストから選択することにより設定される。ここで構築されるオントロジーは XML として保存されるが、学習者はこの入力インターフェースを用いることにより XML を意識することなくオントロジーを構築することが可能である。

オントロジーの構築が終了すると、再び Primary Query が実行され UDDI に構築したオントロジーが通信される。このとき、類似判定プログラムにより構築したオントロジーとすでに登録されているオントロジーの比較を行い、類似していると判定されれば Secondary Query を実行し類似議論を取得される。学習者グループはこの取得された議論を参照しながら議論を再開する。もし、議論を再開した後に構築したオントロジーを修

正する必要がある場合は、オントロジーを修正し、その修正したオントロジーを用いて再度類似議論を UDDI に対し検索を行う。

3.3 議論の類似判定

本稿で実装した機能は、学習者がオントロジーの作成したオントロジーと、既に UDDI に登録されている他のグループの議論を表現したオントロジーとの比較を行う。この比較では、各関連単語同士のマッチングを行い、これを表 1 に示す規則によってパターン 1 からパターン 4 に分類する。

表 1 オントロジーマッチング規則

	オントロジー-A	オントロジー-B
1	中心単語 or 同義語	中心単語 or 同義語
2	中心単語 or 同義語	上位概念 or 下位概念 or 母体 or 部品
3	上位概念 or 下位概念 or 母体 or 部品	中心単語 or 同義語
4	対象	対象
	上位概念	上位概念
	下位概念	下位概念
	母体	母体
	部品	部品
	その他	その他

表 1 の規則に従いパターンを分類し、このパターンの組み合わせにより暫定類似度を決定する手法を考案した。暫定類似度を決定する規則を表 2 に示す。本稿では、(1) 規則 1 を最優先、規則 2 を次席とし、他の規則より類似度判断より重点的に用いる、(2) 優先マッチング規則が下位であっても、複数のマッチングがあった場合にはその個数を優先する、というルールを用いた。ただし、規則 1 と規則 2 が同時に当てはまる事例がなかったため、類似度 10 から 6 を規則 1 適用時、類似度 5 から 1 を規則 2 適用時とした。

表 2 暫定類似度決定規則

	オントロジー-A	オントロジー-B
1	中心単語 or 同義語	中心単語 or 同義語
2	中心単語 or 同義語	上位概念 or 下位概念 or 母体 or 部品
3	上位概念 or 下位概念 or 母体 or 部品	中心単語 or 同義語
4	対象	対象
	上位概念	上位概念
	下位概念	下位概念
	母体	母体
	部品	部品
	その他	その他

表 2 の規則により得られた暫定類似度に加え各関連単語同士のマッチングした回数を考慮し最終的な類似度を算出する。類似度を算出する式を以下に示す。

- ・ (マッチング回数) ≤ 9 の場合
 $(\text{類似度}) = (\text{暫定類似度}) + (\text{マッチング回数}) \times 0.9$
- ・ (マッチング回数) > 9 の場合
 $(\text{類似度}) = (\text{暫定類似度}) + 0.9$

上式により、2つのオントロジーの類似度を 0~10.9 の数値で算出し、この数値が高ければより類似しているものとみなす。

文中の係り受け関係については、形態素解

析の結果を表示し、誤りがある部分についてはインタフェースを通じて発言者が修正する機能を開発し、コンピュータの解析結果を補完することに成功した。また、複数文の意味的な把握については接続詞などを用いて推定し、この候補を発言者が修正する機能を開発した。

また、上記により分析・補完されたデータは、<議論>-<発言>-<文>-<フレーズ>-<単語>という階層から成るツリー構造として表現できる。この構造を XML で記述するための XML 言語 COLCOX を開発した。

3.4 発言役割を基にした議論の可視化

協調学習における自グループおよび他グループの議論構造を分析し可視化する電子掲示板システムを提案した。協調学習に用いられる一般的な電子掲示板は「発言者名」、「タイトル」、「本文」などの項目を入力することで発言し、また他のグループメンバーの発言を参照することで議論が行われる。これに加え、以下の機能を追加した。

- (1) 多文間関係を分析することにより、より詳細な議論構造を明確にする
- (2) (1)の分析結果をツリー構造で可視化する
- (3) 発言者が他の議論を参照しながら発言する

図 2 にシステムの概要を示す。発言者は本システムを使うことで、一般的な電子掲示板のような発言が可能だけでなく、議論全体をツリー構造で視覚的に理解し、自らの発言と他者発言の参照を並列に繰り返すことで、より効率的な発言や参照が可能となる。

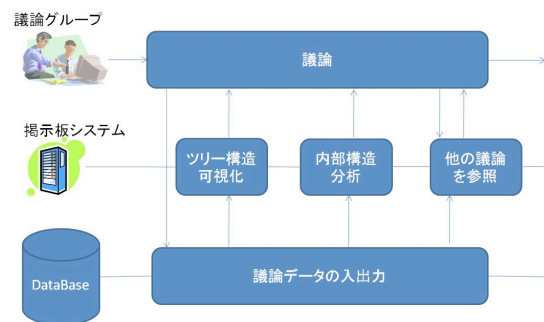


図 2 ツリー表示システム概要

本稿のシステムでは、発言者が発言を入力する際に、併せて発言役割を選択させる。この情報をもとに議論構造を可視化し、リアルタイムにグループメンバーに表示する機能を提案する。選択可能な発言役割は、「肯定」「否定」「意見」「疑問」「理由」「主張」の 6 種類とした。これら 6 種類の発言役割をもとに、以下のように議論の構造を分析する機能を実装した。

(1) 補足発言：発言 A に対し、発言 B が「疑問」を持ち、発言 B に対し発言 C に対して「肯定」の発言を述べた場合、発言 C は発言 B に対する「補足」発言であると判定する。

(2) 対立発言：発言 A に対し発言 B が「肯定」の発言をし、発言 A に対し発言 C が「否定」の発言をした場合、発言 B と発言 C は「対立」の発言であると判定する。

(3) 反例発言：発言 A に対し発言 B と発言 C が「対立」の関係にあるとき、発言 B と発言 C の各々に加えられた「理由」の発言である発言 D と発言 E は「反例」の発言であると判定する。

(4) 二重否定：発言 A に対し発言 B が「否定」の発言をし、発言 B に対し発言 C が「否定」の発言をした場合、二重否定の関係となっており、発言 C は発言 A に「肯定」であると判定する。

議論全体の構造を把握しやすくするため、議論の構造をツリー状に表示することで可視化する。図 3 にツリー表示を含めた協調学習議論のインタフェースを示す。画面は上下左右の 4 部分に分かれる。自グループの発言内容は上側に表示され、下側は参照する議論が表示される。左側は発言の参照と入力（自グループのみ）、右側には可視化した発言のツリー構造が表示される。発言を入力する際は、引用元の発言を選択し、発言役割をプルダウンメニューで選択し、内容をテキストボックス内に入力する。ツリー表示部分の個々の発言をマウスクリックすると、発言内容がポップアップ形式で表示される。

構造全体を表示するだけではツリーが大きくなりすぎ、全体の把握できない危険性がある。そこで、議論が活発に行われる元となる発言を分析し、その発言のまとまりごとにツリー構造を表示する機能を実装した。議論が活発に行われるもととなる発言とは、議論全体の中でもっとも深い階層まで続く発言の流れの中から議論が横に広がった発言とした。また、前述した 4 種類の発言関係をツリー中に図示し、発言関係を迅速に把握する支援をおこなった。

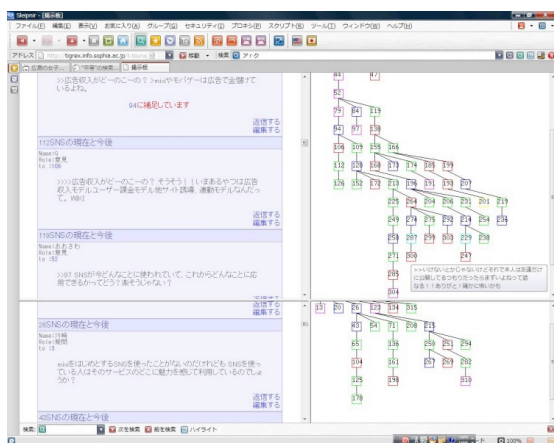


図 3 インタフェース画面

また、これらのツリー状に描写された番号の上にマウスを置くと、その番号に対応した発言がポップアップ表示される。これによって議論構造だけでなく、発言内容を容易に参照し把握することが可能である。

他のグループの議論を参考とする場合、発言者はどの発言を利用して発言したのか引用を明確化する必要がある。そこで、引用する議論の番号を入力することで、参照先の内容を反映する機能を実装した。これによって、議論の参加者は他の参加者と、発言の根拠となる参照元を確認することができ、議論内容の共有を容易に行うことが可能である。

4. 研究成果

本研究により、以下の事柄が明らかになった。

(1) 現行の自然言語処理の限界：自然言語処理の研究が進展し、この成果が Chasen、Mecab、Juman などのオープンソースプログラムとして公開されている。これらは上述のように形態素の品詞特定はほぼ完璧に行うが、文中の形態素の係り受け関係の解析については不完全である。特に、意味を踏まえた SVO(主語、述語、目的語)解析については、解析結果の大半に問題がある。また、SVO のネストから成る文(例：僕がウナギが好きなのを君は知らない)のネスト構造の解析も不完全である。しかし、このネスト構造が正確に解析できないと、文の意味的な把握が困難である。さらに、CSCL で入力される文は、主語や目的語が省略されている場合が多く、これらを補完してから解析をはじめめる必要がある。また、複数文の間の関連性については、現行の自然言語処理の限界を超えており、理論的な仮説を立てるに留まっている。

(2) 人手による解析結果の補完：上記の限界を超えるため、自然言語処理による形態素や係り受けの推定を、発言入力者により修正・追加するインタフェースを開発した。現在このインタフェースの試用版が完成しているが、まだ使いやすい表現形式となっていないため、今後これを改善する必要がある。ただ、マウスなどによる直感的なインタフェースにより、修正や入力が容易である。このインタフェースの評価は今後進める必要がある。

(3) メタデータ COLCOX の開発：協調学習の発言を、メタデータを含めて構造的に表現するため、<議論>-<発言>-<文>-<フレーズ>-<単語>という階層を規定し、これらの各々の階層に必要なメタデータを規定した XML 言語 COLCOX を開発した。この COLCOX に発言データを

格納することにより、形態素レベルでのみならず、節や文単位で、類似・反対・根拠など文間・発言間の関係性をコンピュータにより解析することが可能となる。現在、(2)のインタフェースを用いた分析・補完結果を COLCOX に格納する機能を開発中である。

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)

[学会発表] (計 11 件)

- (1) 富田、沖林、田村、アクセス数や議論内容から E ラーニングの学習効果は予測できるか、日本認知科学会全国大会、2010 年 9 月 18 日、神戸大学
- (2) 沖林、富田、田村、2 講義 Moodle を用いた同期的・非同期的ディスカッションによる実践、日本教育工学会全国大会、2010 年 9 月 20 日、金城学院大学
- (3) 竹中、新井、池田、米山、田村、自然言語処理と説明情報付与による CSCL の発言入力・参照支援、教育システム情報学会全国大会、2010 年 8 月 27 日、北海道大学
- (4) 田村、鷺見、議論ログ共有システムにおける類似議論判定機能、教育システム情報学会研究会、2010 年 3 月 13 日、畿央大学
- (5) 田村、劔、2 文間関係に基づく協調学習議論の構造可視化、教育システム情報学会研究会、2010 年 3 月 13 日、畿央大学
- (6) 田村、CSCL 視点の集合知形成支援、電子情報通信学会 集合知シンポジウム、2010 年 1 月 25 日、広島市まちづくり市民交流プラザ
- (7) 鷺見、田村、Web サービスを用いた CSCL 間データ共有機能の改善、JSiSE 全国大会 2009 年 8 月 21 日、名古屋大学
- (8) 前島、劔、鷺見、岸本、田村、沖林、富田、萱、伊藤、西端、姫田、岡本、CSCL データの構造化と Web サービスによる LMS 間共有、教育システム情報学会全国大会、2008 年 9 月 4 日、熊本大学
- (9) 鷺見、田村、Web サービスによる CSCL データの LMS 間共有、教育システム情報学会全国大会、2008 年 9 月 4 日、熊本大学
- (10) 劔、田村、2 文間関係からの CSCL ログの構造化、教育システム情報学会全国大会、2008 年 9 月 4 日、熊本大学
- (11) 前島、田村、格解析を用いた CSCL データの省略解析、教育システム情報学会全国大会、2008 年 9 月 4 日、熊本大学

6. 研究組織

(1) 研究代表者

田村恭久 (TAMURA YASUHISA)

上智大学・理工学部・准教授

研究者番号：30255715

(2) 研究分担者

なし

(3) 連携研究者

岡本敏雄 (OKAMOTO TOSHIO)

電気通信大学大学院・情報システム学研究科・教授)

姫田麻利子 (HIMETA MARIKO)

大東文化大学・外国語学部・准教授

西端律子 (NISHIBATA RITSUKO)

畿央大学・教育学部・准教授

伊藤健二 (ITO KENJI)

慶應義塾大学・DMC 統合研究機構・准教授

萱 忠義 (KAYA TADAYOSHI)

学習院女子大学・国際文化交流学部・専任講師

富田英司 (TOMIDA EIJI)

愛媛大学・教育学部・専任講師

沖林洋平 (OKIBAYASHI YOHEI)

山口大学・教育学部・専任講師

