

ブラウザベースの動詞語義及び意味役割付与作業システム

上野 真幸 (岡山大学大学院自然科学研究科)

竹内 孔一 (岡山大学大学院自然科学研究科)

Construction of a Browser-based Annotation Tool for Verb Meanings and Semantic Role Labels

Masayuki Ueno(Graduate School of Natural Science and Technology,
Okayama University)

Koichi Takeuchi(Graduate School of Natural Science and Technology,
Okayama University)

1. はじめに

文中の動詞の語義を同定し、さらに、動詞と係り関係にある単語との意味的關係を人手で付与する述語項構造付与システムの構築を行った。

動詞語義概念とは動詞がとりえる1つの異なる意味について共通属性としてまとめたもので、例えば「走る」なら「電車が走る」「閃光が走る」などがそれぞれ異なる語義概念であり、1つの語義概念「電車が走る」の場合なら、【移動】であり、「移動する」「歩く」などを同じ動詞の類語として扱う単位である。また意味役割とは述語と係り元との關係を表したものであり、人により判断の揺れるものが存在するため信頼性の高いアノテーションを行うことは困難である。本稿で構築した動詞語義及び意味役割付与作業システムはBCCWJのCOREの文章に対し、動詞語義及び意味役割を人の手で付与する際の補助を目的としたシステムである。動詞語義と意味役割については無料で公開されている動詞項構造ソーラス[1]のものを用いる。また複数人による付与を可能とするため、Webブラウザベースで構築を行う。そのためにCakePHP[2]を用いて構築を行った。

本システムの構造とインターフェースについて解説し、実際に動詞語義付与及び意味役割付与を行い、テキスト入力による付与と比べてどの程度作業効率が上がったか、そして各作業員間の付与の一致率を報告する。

2. 関連研究

関連研究としてFrameNet[3]やSlate[4]が挙げられる。FrameNetはフレーム意味論に基づいて構築された英語の語彙辞書である。FrameNetはある語を想起させる意味概念をframeとして仮定し、意味概念を持つ単語を結びつけることで構築されている。日本語においても日本語FrameNet[5]の開発が進んでいる。また佐藤[6]によってframeNetのデータをWebブラウザ上で検索表示できるソフトウェアであるFrameSQLが開発された。単語列ベースで詳細な付与が可能である一方で、システム自身は公開されていない。Slateは徳永ら[7]によって開発された汎用アノテーションツールであり、多様なアノテーションに対応できる。クライアントサーバ型のアプリケーションであり、複数の作業員でアノテーションを行うことができる。汎用性の高いツールであるため本研究で扱う動詞語義及び意味役割の付与もできるが、本研究で提案する動詞語義及び意味役割は種類が多いため、Slateで実行するには煩雑になり適していない。また、本システムの特徴である、複数人作業員の付与結果の記録と決定を行うことができない。

3. 動詞語義及び意味役割付与作業システム

本研究で扱う動詞語義及び意味役割付与作業システムとは、文章から動詞を選択し、選択した動詞の語義決定を行い、意味役割の付与を行うものである。システムの構成、利用している言語資源、及びシステムの特徴を述べる。また実際の作業の観点から必要と感じた機能についても説明する。

3.1. 人手による動詞語義及び意味役割付与の問題点

動詞語義及び意味役割の人手による付与を行う際の問題点を整理する。本システムで扱う意味役割は 95 種類、動詞語義は動詞ごとに複数存在する。動詞語義及び意味役割は種類が多く人手で記述を行う場合の処理は非常に煩雑なものとなる。既存のツールで本システムで扱う動詞語義及び意味役割の付与を人手で行った場合、膨大な時間を要し、間違いも発生しやすい。また動詞語義及び意味役割は人によって判断の揺れるものが存在し、信頼性の高いデータを作成するのは困難である。例えば「探索で見つかる」という文の動詞語義及び意味役割付与を考えると、「見つかる」は【発見】という語義であるが、一方「探索」は、「探索」という[原因]で【発見】したのか、「探索」という[手段]で【発見】したのか人によって判断の異なるところである。

3.2. 動詞語義及び意味役割付与作業システムの設計

前節で述べたように、動詞語義及び意味役割付与は煩雑で人によって判断が異なることがあるため、一人で作業を行うのは好ましくない。よって複数人で作業を行うために Web 上での閲覧・作業を実現する。動詞語義などは大量のデータ検索を行うのため、利用するデータベースは検索が高速である MySQL[8] を利用する。フレームワークは高速開発に適している CakePHP を利用する。本システムでは複数人による付与をひとつに集約するために、動詞語義及び意味役割が複数付与された場合、付与事例を考慮して、最終的に 1 人の作業員（作業リーダー）が判断して決定を行う。これにより、1 人で付与した結果よりも信頼性の高いデータが得られると考える。

3.3. データベース構築

本システムは BCCWJ コーパスの文章データ、LUW 単位でのデータ及び動詞項構造シソーラスを用いる。さらに LUW から動詞のみを抽出し、その頻度（出現回数）を格納したテーブル、付与された動詞語義及び意味役割のテーブルを作成した。

図 1 の左側は本システムで構築したデータベース図である。luw と sentence は BCCWJ コーパスの CORE の luw と文章データをそれぞれテーブルとしたものである。semantic と vth は動詞項構造シソーラスの意味役割と動詞語義をそれぞれ格納したものである。luw_verb は luw テーブル内に出現した動詞とその回数 (freq) を格納している。luw_verbid_sentenceid は luw_verb テーブルの動詞がどの文で使われているかと作業状態か否か (sence) を示したテーブルである。luw_semantic 及び luw_vth は付与された意味役割及び動詞語義を格納するテーブルである。図 1 の右側は「詰め将棋の本を買ってきました」という文の「買う」に対し、語義【購入】と係り元である「詰め将棋の本を」に意味役割 [対象] を付与した例である。

3.4. 動詞語義および意味役割付与の手順

本システムの付与作業の手順を説明する。意味役割は動詞と項との意味的關係であり、動詞は複数の語義を持つため、語義により意味關係が異なり、語義を先に決定する必要がある。また各動詞語義の例文が複数必要であるが、作業員が付与対象となる文を見なければ動詞語義を決定するのは困難であるため、作業員が BCCWJ の文を見て作業対象となる文を決める必要がある。よって動詞語義付与を行う前に作業文の選択を行う必要がある。そして作業対象に選ばれた文に対して、動

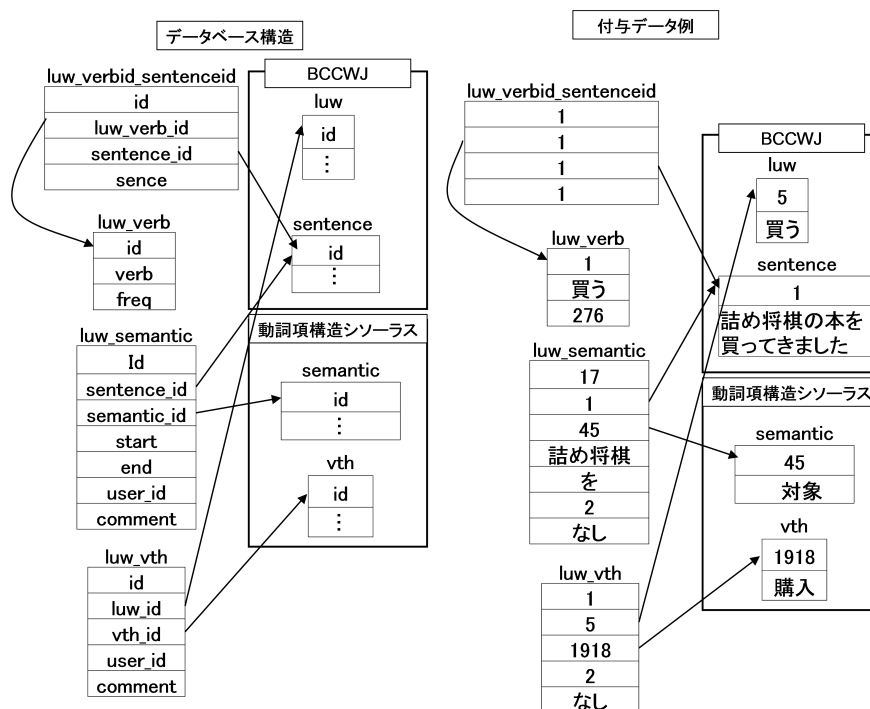


図1 データベース構造と付与データ例

詞語義付与及び意味役割付与を行った後、付与された動詞語義及び意味役割から作業リーダーが決定を行う。以上の観点から、付与作業の手順は以下の5つに集約できる。

- A) 作業文の選択.
- B) 動詞語義付与.
- C) 動詞語義の決定.
- D) 意味役割付与.
- E) 意味役割の決定.

この順に各作業者が作業を行う。それぞれに作業用 Web ページを作成した。基本作業画面は前回 [9] で述べたのでここでは割愛する。

3.5. 作業効率向上のための機能

実際に本システムを使って5ヶ月間作業者を雇い、付与作業を行った所以下のような意見が寄せられた。

1. 前回自分がどこの作業をしていたのか覚えていない.
2. 次の作業文にすすむときにページの移動が多い.
3. 当てはまる動詞語義、意味役割がない.

作業の負担を減らすためにこれらの意見を参考に改善を行った。具体的には以下のようにシステムを改善した。

1. 前回作業（語義付与、意味役割付与）を行った文へのリンクを作成.
2. 同じ作業を次の作業文で行う「次の文へ」のリンクを作成.

3. 動詞語義，意味役割選択の際に適合語義，意味役割なしの選択肢とコメント欄を追加。
4. 動詞語義追加の機能を作成。

これらの改善により作業負担が減るとともに，動詞項構造シソーラスの拡張にもつながると考えられる。作業者が適合する語義なしと判断した場合，語義を追加し適合語義を作成することにより，動詞項構造シソーラスを拡張できる。この作業では一般作業者が語義を判断し，それを確認した作業リーダーが新たな語義を追加するといった流れになると考えられる。

4. 作業結果

2012年9月から2013年1月25日の間で，本システムを用いて4人の作業者を雇い付与を行った。作業者は3人の学部学生(文系2名理系1名)と1人の大学院生であり，作業期間や意味役割分類の経験等はバラバラである。分類経験が無い作業者に対しては分類についての説明を行った。各作業者の動詞語義と意味役割付与数を表1に示す。

表1 動詞語義及び意味役割付与数

	作業者 A	作業者 B	作業者 C	作業者 D
語義付与数	2324	6770	2220	6376
意味役割付与数	696	12346	3558	13766
意味役割付与文数	295	5672	1578	6755

各作業者間で付与数に差があるのは作業期間に差があるためである。2012年9月から作業を開始した作業者Bと作業者Cにおいては，約6000文に対して語義と意味役割を付与している。各作業者間の一致率を表2に示す。

表2 各作業者間の一致率

	語義一致率	意味役割一致率
作業者 A, B	92% (984/1064)	56% (177/312)
作業者 A, C	91% (932/1023)	68% (209/306)
作業者 A, D	82% (714/869)	65% (457/696)
作業者 B, C	81% (900/1101)	63% (920/1458)
作業者 B, D	84% (4647/5523)	65% (7781/11830)
作業者 C, D	82% (951/1151)	54% (1704/3129)

この一致率は両作業者が付与を行った文のみを対象としている。この結果では動詞語義は80%以上の一致率を記録している。動詞語義は一動詞につき約3~4つの語義から選択する形式であり，作業者は他作業者の付与結果を見ることができないため，高い一致率であるといえる。意味役割は95種類の意味役割から選択するため，全体的に語義に比べ低い結果となっている。現在もこの作業は継続中であり，意味役割の決定作業が進んでいないため，作業リーダーとの一致率は不明である。これらの一致率は本システムで算出することができる。そのため作業が進んでいるリアルタイムでの一致率を常に見ることができる。一致率閲覧画面の例を図2に示す。

図2は作業者A, Dの一致率表示画面である。両作業者が作業した文の数や一致した数を表示している。語義平均選択肢数は両作業者が付与した語義の選択肢数の平均値であり，語義付与の曖昧性を示している。位置が一致した意味役割数は選択された意味役割の種類は違うが，選んだ係り

語義	2324
語義が無いと思ったもの	131(0.056368330464716)
二人が語義を付与した動詞数	869
二人が一致した語義数	714/869(0.82163406214039)
語義選択肢平均数	2.9747899159664(2124/714)
付与された意味役割数	696
二人が意味役割を付与した動詞数	295
位置が一致した意味役割数	573/696(0.82327586206897)
一致した意味役割数	457/696(0.6566091954023)

図2 付与数と一致率の表示例

先が同じものの数を示している。この画面は計算量が多いため他の動作に比べると表示が遅い。

5. 考察

付与作業の効率について考察を行う。今回の付与作業での付与速度は、9月時点での作業員2名の付与データを100件ほど人手で見たとこ、1語義付与約15～30秒程度で行われている。本システムを用いずに語義をテキストで入力する場合慣れた作業員でも1～2分程度要するためかなりの効率が上がったといえる。効率が上がった主な要因として以下の点が挙げられる。

- (1) 語義や意味役割を選択するだけで文字入力する必要が無い。
- (2) 次の作業に移る際の手間が少ない。

(1) はシステムの補助をなしにこの付与作業を行う場合、例えば「買う」に対する語義であれば、【状態変化あり-位置変化-位置変化（物理）（人物間）-他者からの所有物の移動-購入-彼女が高級品を専門店から買う】等を入力しなければならないため非常に手間がかかる。これらの入力を大きく減らしたことが作業の効率が上がった大きな要因であると考えられる。(2) は作業の量が非常に多いため、各作業間の画面クリックが例え一つでも多い場合、作業の際ストレスがかかるとの意見が作業員から寄せられた。この点を改善し、なるべく作業間の動作を減らした結果、作業の効率上昇につながったと考えられる。

6. まとめ

本研究では文章の意味を同定するために、動詞語義および意味役割の付与を補助するシステム構築を行った。データベースとフレームワークを用いて、Web上で作業、閲覧を実現した。この手法により複数人による付与作業が可能となり、インターネットブラウザさえあればシステムのダウンロードを必要とせず、複数人で1つの文に対する付与を実現することで、データの信頼性向上を図った。また5ヶ月間4名の作業員に作業をを行ってもらい、語義では80%、意味役割では60%程度の一致を得た。

参考文献

- [1] 動詞項構造シソーラス. <http://cl.cs.okayama-u.ac.jp/rsc/data/index.html>.
- [2] CakePHP. <http://cakephp.jp/>.
- [3] FrameNet. <http://framenet.icsi.berkeley.edu/>.
- [4] Slate. <http://www.cl.cs.titech.ac.jp/slate/>.
- [5] Japanese FrameNet. <http://jfn.st.hc.keio.ac.jp/ja/index.html>.
- [6] 佐藤弘明. FrameSQL で利用する日本語フレームネット. 特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ (研究成果報告会) 予稿集, pp. 143–146, 2010.
- [7] Dain Kaplan, 飯田龍, 徳永健伸. 汎用アノテーションツール Slate. 言語処理学会 第 17 回年次大会 発表論文集, pp. 619–622, 2011.
- [8] MySQL. <http://www-jp.mysql.com/>.
- [9] 上野真幸, 竹内孔一. 動詞語義及び意味役割付与作業システムの構築. 第 2 回日本語コーパスワークショップ, NL-207-15, pp. 69–76, 2012.