

動詞語義及び意味役割付与作業システムの構築

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 an 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て、さらに、動詞と係り関係にある単語との意味的關係を付与する述語項構造付与システムの構築を行っている。動詞語義とは文章内の動詞の意味を同定したものであり、同じ表層の動詞でも複数の意味を持つ場合が多く、既存の辞書では対応できない語義も存在する。また意味役割とは文章内の各項が文章中でどのような意味を持つかを同定したものであり、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および意味役割付与作業システムは BCCWJ の CORE の文章に対し、動詞語義およ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シソーラス [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 上でのアプリケーションとして CakePHP [2] を用いて構築を行った。

本稿では本システムの構造や使い方を解説し、実際に 800 文程度に対して動詞語義付与を行い、作業者の行動から必要な機能の割り出しを行った。

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。

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 89 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。

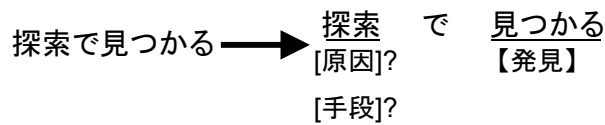


図 1 人によって判断が揺れる意味役割の例

図 1 は「探索で見つかる」という文の動詞語義及び意味役割付与例である。「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。Web 上でデータ入力、編集などができるデータベース管理ツールとしてはファイルメーカーなどがあるが、利用形態が制限されており、言語処理ツールを利用するなどといった拡張もできない。そこで、フリーでダウンロードできるデータベース管理ツールとフレームワークを用いて開発を行う。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。また Web 上でシステムを利用する場合、インターネットブラウザから作業を行うことができるためシステムのダウンロードを行う必要がない。ここで複数人で複数文に対して付与を行う場合、次のような問題が発生する。

- (1) 対象となる文の数が膨大なため、付与される文が統一性をもたない。
- (2) 同じ文に対して複数の付与が行われる。

まず (1) を解決するために作業対象を指定する必要がある。つまりデータベース上に存在する文すべてに作業対象か否かの状態を持たせるということである。次に (2) であるが、本システムでは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、(2) の問題を解決し、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 のなかから動詞のみを抽出し、その頻度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。

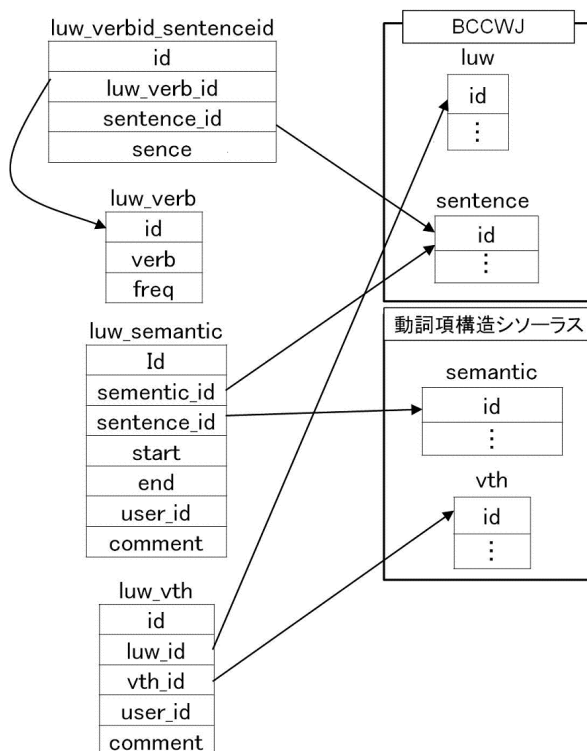


図 2 データベース構造

図 2 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。

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。文に動詞語義及び意味役割を付与する場合、はじめに動詞語義を決定する必要がある。動詞語義は動詞項構造シソーラスのデータを利用するため、動詞項構造シソーラスに語義の存在しない動詞は付与できない。また各動詞語義の例文が複数必要であるが、人が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、人が文を見て付与対象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。このため付与作業の手順としては以下の 5 つがあげられる。

- 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。それぞれの作業において、最適な操作環境とデータを表示する必要があるため、それぞれにおいて、作業用 Web ページを作成した
以下にそれぞれについて説明する。

A) 作業文の選択

最初の作業として、付与作業を行いたい動詞を選択する。動詞を選択すると選択された動詞を含んだ文の一覧を表示する。その中から作業対象としたい文を選択する。選択された文は作業対象となり、付与作業が可能となる。

test

Sentence.id	Sentence.article Number	Sentence.text	Sentence.exists	LuwVerbidSentenceid.sence
1	1	詰め将棋の本を買って来ました。	0	<input checked="" type="checkbox"/>
52	15	グランツーリスモ4を新品で買おうと思っているのですが、どこで買ったら一番安いでしょうか？	1	<input type="checkbox"/>
52	15	グランツーリスモ4を新品で買おうと思っているのですが、どこで買ったら一番安いでしょうか？	1	<input type="checkbox"/>
56	15	家の近くには、TOKUJIROUがあるから、自分はそこで買おうとっ！	1	<input type="checkbox"/>

図 3 作業文の選択例

図 3 は「買う」という動詞の作業文選択画面である。「買う」という動詞を含んだ文の一覧を表示する。各文にはチェックボックスを設置しており、作業対象とする文を作業者が選択できるようにしている。チェックボックスにチェックを入れることで語義付与対象となる。またチェックを外すことで付与対象から外すこともできる。

B) 動詞語義付与

作業対象が存在する場合、語義の付与を行うことができる。付与作業を行いたい文を選択すると、短単位または長単位に分割された文を表示する。分割された文章の動詞を選択すると、選択した動詞の語義一覧を表示する。表示した語義から選択を行うことで語義の付与を行う。

図 4 は「詰め将棋の本を買って来ました。」という文の「買う」に対する動詞語義の付与画面であ

例文
詰め将棋の本を買ってきました。
選択動詞:買う

test

Actions

<input checked="" type="radio"/> 0	1918	買う	状態変化あり	位置変化	位置変化(物理)(人間間)	他者からの所有物の移動	購入
<input type="radio"/> 1	1919	買う	状態変化あり	位置変化	位置変化(物理)(人間間)	他者からの所有物の移動	購入
<input type="radio"/> 2	1920	買う	状態変化あり	位置変化	位置変化(物理)(人間間)	他者からの所有物の移動	購入
<input type="radio"/> 3	5251	買う	状態変化あり	主体の変化	心理的变化	感情変化の引き起こし	嫌悪
<input type="radio"/> 4	5959	買う	状態変化あり	主体の変化(判断・認識の変化)	判断(認識)	心理的立場	
<input type="radio"/> 5	9726	買う	状態変化なし(活動)	対人・対物的動作	働きかけに対する応対	応じる	

図 4 動詞語義付与例

る。各語義にラジオボタンを設置しており、各語義のラジオボタンを選択することで対象となる動詞の語義を選択できる。語義一覧表示は動詞項構造ソーラスの分類のみを表示しているが、各語義の id を選択することで例文などの語義詳細を閲覧できる。

C) 動詞語義の決定

複数人の作業により動詞語義が付与された後、作業リーダーは動詞語義の決定を行う。よってこの作業では、複数人による作業の付与結果を表示し、どの付与結果を決定するかを選択するページを構築した。付与された動詞語義の一覧から意味役割の決定を行う。

選択文:詰め将棋の本を買ってきました。

動詞:買う

決定中の語義:1920:状態変化あり-位置変化-位置変化(物理)(人間間)-他者からの所有物の移動-購入

decision

	id	vth_id	分類	user	comment	Actions
<input type="radio"/>	11	1920	状態変化あり-位置変化-位置変化(物理)(人間間)-他者からの所有物の移動-購入	ueno		Delete
<input type="radio"/>	10	1918	状態変化あり-位置変化-位置変化(物理)(人間間)-他者からの所有物の移動-購入	root		Delete
<input type="radio"/>	1921	5959	状態変化あり-主体の変化(判断・認識の変化)-判断(認識)-心理的立場-	ueno		Delete

図 5 動詞語義決定の例

図5では「詰め将棋の本を買ってきました」という文の「買う」という動詞に対して、3つの語義が付与されている。各語義の左側にラジオボタンを設置している。決定したい語義のラジオボタンにチェックを入れることで動词语義を決定することができる。

D) 意味役割付与

作業対象である文の動词语義が決定されている場合、意味役割の付与を行うことができる。付与対象となる項と意味役割の種類を選択することで意味役割の付与を行う。また付与した意味役割にコメントをつけることができる。

The interface displays a grid of semantic role options, each with a radio button. The selected role is '対象' (Target). Below the grid is a 'comment' input field. The main part of the interface shows a sentence '詰め将棋の本を買ってきました。' with a table of grammatical analysis. A red box highlights the '買う' (buy) entry, which is currently selected with ID 1918. A large black arrow points down to a second screenshot where the '買う' entry is now selected with ID 1918 and the label '意味役割付与' (Meaning Role Assignment) is visible.

図6 意味役割付与例

図6は「詰め将棋の本を買ってきました。」という文の「買う」に対する意味役割の付与画面である。各項にチェックボックス、各意味役割にラジオボタンをそれぞれ設置している。まず対象となる項のチェックボックスにチェックを入れる。そして、意味役割一覧から付与したい意味役割のラ

ジオボタンにチェックを入れることで意味役割を付与できる。図6では「詰め将棋の本」に対して[対象]を付与している。

E) 意味役割の決定

複数人の作業により意味役割が付与された後、作業リーダーは意味役割の決定を行う。よってこの作業では、複数人での作業の付与結果を表示し、どの付与結果を決定するかを選択するページを構築した。付与された意味役割の一覧から意味役割の決定を行う。

id	sentence_id	luw_id	text	semantic	comment	author	Actions
3	詰め将棋の本を買ってきました。	買う	詰め将棋の本	対象		ueno	決定中 <input type="button" value="決定解除"/>
18	詰め将棋の本を買ってきました。	買う	詰め将棋の本	結果		root	<input type="button" value="決定"/> <input type="button" value="Delete"/>
19	詰め将棋の本を買ってきました。	買う	詰め将棋の	対象		root	<input type="button" value="決定"/> <input type="button" value="Delete"/>

図7 意味役割決定の例

図7では「詰め将棋の本を買ってきました」という文の「買う」という動詞に対して、「詰め将棋の本」に[対象]、「詰め将棋の本」に[結果]、「詰め将棋の」に[対象]という3つの意味役割が付与されている。決定を選択することにより、意味役割を決定状態にすることができる。また決定解除で決定状態を解除できる。決定状態にある意味役割は消去などの操作を行うことができない。図7では「詰め将棋の本」に対する意味役割として[対象]が決定状態にあることを示している。

4. 付与作業のユーザーインターフェースの改善

実際に本システムを使って3人の作業で動詞語義付与を800文程度行った所以下のような意見が寄せられた。

1. 自分の作業が終わっているかどうかを簡単に確認したい。
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. 作業例文を選ぶ際に付与が終わっている文には「OK」を表示。
2. 同じ作業を次の文で行う「次の文へ」のリンクを作成。
3. 動詞語義選択の際に適合語義なしの選択肢とコメント欄を追加。

これらの改善により作業負担が減るとともに、適合語義なしの選択肢を追加したため、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。しかし今後の課題として動詞項構造シソーラスの拡張という観点から考慮した場合、適合語義なしという分類では不十分であると考えられる。改善法としては「動詞語義の追加」という動作を実装することである。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。

5. 考察

作成したシステムの動作速度についての考察を行う。本システムはサーバクライアント方式で作成しており、レンタルサーバで独立した CPU を優先して利用しているため、膨大な量のデータを利用するにもかかわらず、各作業は操作から 1, 2 秒程度で動作する。しかし、付与データが増えると動作速度は低下すると考えられる。動作速度を下げる要因として、データの取得にかかる時間が長くなることがあげられるためである。さらに本システムは一つの動作に複数のテーブルを用いている動作が多い。取得するデータ数が同じであってもテーブル数が増えると動作が遅くなる。これはフレームワークの仕様であるが、データを取得する場合にテーブル選択、データ検索、データ取得という手順で行っているためである。またフレームワークの機能として、結合させたテーブルのデータを自動で取得するというものがあるが、本システムのように扱うデータが多い場合、必要なデータ以外のデータも多く取得してしまうため、動作が非常に遅くなる。このため本システムではこの自動取得の機能をなるべく用いず、必要データのみを取得している。意味役割付与動作では、フレームワークの機能で自動取得した場合に比べ、4分の1程度の時間でデータを取得できた。

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。また実際にシステム利用者の意見を取り入れることでユーザーインターフェースの改善を行った。

参考文献

- [1] 動詞項構造ソーラス. <http://cl.it.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. 619–622, 2011.
- [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. 143–146, 2010.
- [8] MySQL. <http://www-jp.mysql.com/>.