

# 日本語研究におけるタグ付与とタグ付与ソフト TNR\_JapaneseCorpusTagger Tagging and its software 'TNR\_JapaneseCorpusTagger' in Japanese Studies

于 康・田 中 良<sup>1</sup>  
Kang Yu and Ryo Tanaka

One of the central issues on Japanese studies is how efficiently we can analyze the data. In this study, we have developed the tagging software 'TNR\_JapaneseCorpusTagger'. This software makes the tagging analysis accurate and speedy. In addition, this software has proved that it is a very powerful device to find out the generalizations of word formations or constructions based on the statistic results.

キーワード：タグ付与、タグの種類、自動タグ付与機能、タグ付与ソフト

**Key Words** : Tagging, Varieties of tagging, Automatic tagging function, Tagging software

## 1. 問題提起

日本語について、記述法や帰納法といった手法で研究を行うとすれば、用例の収集と用例の分析が非常に重要になってくる。用例の収集については、これまで数多くのコーパスが開発されてきた。とりわけ、国立国語研究所の研究成果が顕著である。例えば、『現代日本語書き言葉均衡コーパス(BCCWJ:Balanced Corpus of Contemporary Written Japanese)』、『日本語話し言葉コーパス

(Corpus of Spontaneous Japanese : CSJ)』、『ことばに関する新聞記事見出しデータベース』等は、何れも国立国語研究所が中心になって、独自か他大学や研究機構の協力の下で開発した日本語の用例を収集するためのコーパスである<sup>2</sup>。そのほか、『青空文庫』<sup>3</sup>や『国会会議録検索システム』<sup>4</sup>等のように、検索機能付きのデータベースも広く日本語の用例収集に利用されている。要するに、今日、少し課題は残っているものの、大凡日本語の用例収集には困ることがなく、必要な用例が収集でき

- 1 著者の于康は陣内正敬先生のご要請で数年間陣内先生と共に総合政策研究科で博士課程前期課程や後期課程の院生の指導に携わっていました。言語コミュニケーション文化研究科では共に日本語教育日本語教育学領域に所属し、公私とも陣内先生に大変お世話になりました。著者の田中良は、陣内先生のご指導を求めるため、大学院の入試を受けて、博士課程後期課程に入学したのです。入学後、陣内先生のご体調が不良のため、ご回復なさるまで、しばらくの間、于康が田中良の指導を担当することになりました。しかし、残念ながら、田中良の希望が叶わず、于ゼミに残ったままです。追悼号の執筆依頼を受けた時点では、文法研究の内容にする予定でしたが、せめて田中良の成長ぶりだけでもいいので、陣内先生にご報告し、喜んでいただきたいと思い、急速テーマを変え、共同で研究してきたタグ付与ソフトに変更したのです。これをもちいて陣内正敬先生にご報告し、謹んでご冥福をお祈り申し上げます次第です。
- 2 内容の詳細は、大学共同利用機関法人人間文化研究機構国立国語研究所のホームページを参照されたい。 <http://www.ninjal.ac.jp/database/>
- 3 内容の詳細は、青空文庫のホームページを参照されたい。 <http://www.aozora.gr.jp/>
- 4 内容の詳細は、国会会議録検索システムのホームページを参照されたい。 <http://kokkai.ndl.go.jp/>

る状況になっていると言えよう。

しかし、用例の収集ができたとしても、用例の分析や研究ができたことにはならない。用例の分析は、当然ながら日本語の研究にとって、研究の成果を見出すために、必要不可欠な研究プロセスである。これまでの用例分析では、目視による分析、Excelを活用した分析がほとんどであった。用例を目視やExcelを活用して分析する手法では、分析対象の成分にいろいろなシルシを付けて分析するといった方法もよく聞かれる。このシルシを付けるという方法は、ある意味では、ここで言うタグに共通性を有するものである。

上述のように、収集してきた用例を如何にして効率よく分析できるかが課題である。目視やExcelより他に有効的な方法がないのか、そのような方法があるとすれば、どのような方法が求められているのかといったことを考えなければならぬであろう。そこで、本研究は、以上の点を踏まえ、タグ付与のソフトを開発し、それが目視やExcelより用例の分析に寄与できることを明らかにしたい。

## 2. タグ付与の目的

用例にタグを付与する際に、2つの方法が考えられる。1つは、用例の全ての成分にタグを付与すること、もう1つは、必要な成分だけにタグを付与することである。すべての文の構成成分にタグを付与するのは、今後の作業の手間暇を省くことができるという点からすれば、非常に有効なように見えるし、プログラミング上でも実現できないことではない。しかし、タグと言っても、統語論のレベルのものもあれば、意味論のレベルのものもあり、更に、語用論のレベルや談話文法のレベルなど、様々なものがある。それらのすべての研究分野を網羅的にタグ設定の射程に入れることは至極難しいので、すべての文の構成成分に、す

べての研究分野を含むタグを付与することは、実現できそうもない。従って、特定の研究分野(複数であっても可能)を絞り、できるだけその、またはそれらの研究分野に使われている各種の分類をタグとして活用し、研究対象になる必要な成分だけにタグを付与するほうが現実的であろう。

タグ付与は、一回性のものではない。すなわち、一回付与すれば終わりということではない。二次タグ付与や三次タグ付与も視野に入れなければならない。二次タグ付与と三次タグ付与とは、一回のタグ付与で問題の解決に至らない場合、付与したタグに更に他のアプローチや立場からタグを付与するということである。例えば、一回目のタグは統語論レベルのものを付けて分析し、納得できる分析結果が得られない場合、または、更に別の視点からその問題を観察することが必要と思われる場合は、意味論か語用論レベルのタグを二次タグまたは三次タグとして付与する。

さて、観察したい対象にタグを付与することは、そのタグを付与しなければ達成できない目的があるからと考えられる。その目的を達成するために、タグの選択について指針の策定が必要になってくる。これまでの目視などで行ってきた研究のプロセスからすると、その指針の策定に参考になり得るものは、およそ次の3点かと思われる。

- ①なぜタグ付与が必要か
- ②タグの統計結果や分析の中からどのような規則性を見出そうとしているか
- ③研究の目的を達成するために、どのレベルや種類のタグが必要か

文法や構文の研究を対象とするならば、成分の生起条件や成分と成分との共起条件がしばしば観察の対象になる。その意味で、タグを付与する場合は、生起や共起に関わる成分にタグを付与することがまず求められるであろう。すなわち、生起や共起の条件を明らかにするためなら、それぞれの成分のスロットにどのような成分なら生起する

か、どのような成分とどのような成分ならよく共起するかまたはあまり共起しないかといった条件を見出すことが求められるのである。その生起や共起の条件は、時には生起や共起の規則性とも言われている。当然ではあるが、生起や共起の条件を求める際に、生起や共起ができる条件の他に、生起や共起ができない条件も考えなければならない。その生起や共起ができない条件は、生起や共起ができる条件を踏まえた上で、テストによって明らかになることが多い。

要するに、以上のように、日本語を研究の対象とする場合は、タグ付与に伴うタグの種類の判別の難易度やタグ付与の現実性からしても、まず優先して考えるべきものは、統語論レベルのタグと意味論レベルのタグであると言えるであろう。

### 3. タグの種類

統語論レベルのタグにしても、意味論レベルのタグにしても、タグの選択や設計を行う際に、直面せざるを得ない重要な問題がもう1つある。それは、タグ判断の現実性である。タグ判断の現実性とは、使用者がある対象についてそれにどのようなタグを付与すればよいかを判断することができるかどうかで、躊躇なく判断できる場合はタグ判断の現実性があるのに対し、判断に非常に迷う場合は、タグ判断の現実性が薄れるということである。

確かに、非対格自動詞や非能格自動詞のように、基本的な術語についてもなかなか判然とした定義が示されていない中では、タグの選択の判断に迷わないことのほうがむしろ少ないであろう。更に、同一の動詞にしても、説によって、分類が異なるし、意味役割にしても、説によって、名称だけではなく、分類の数やそれぞれの分類に含まれるものも随分異なる。従って、タグの選択や設

計を行う際は、個性の強い説よりは、より通説や定説に近いものを選択しなければならないことになるであろう。それにしても、個性が強い説であってもそれなりに理が通ったものもあるので、すべてを無視するわけにはいかない。従って、タグの選定は、通説や定説に加えて、個性は強いがよく使われる説を並列するのが理想的であろう。

以上のことを踏まえ、様々の文法説の中で、日本語記述文法研究会編の『現代日本語文法1 第1部総論 第2部形態論』(2010)、『現代日本語文法2 第3部格と構文 第4部ヴォイス』(2009)、『現代日本語文法3 第5部アスペクト 第6部テンス 第7部肯否』(2007)、『現代日本語文法4 第8部モダリティ』(2003)、『現代日本語文法5 第9部とりたて 第10部主題』(2009)、『現代日本語文法6 第11部複文』(2008)、『現代日本語文法7 第12部談話 第13部待遇表現』(2009)は、いろいろな批判があるものの、現段階では、記述文法という視点から行ってきたものの中で、以上の研究に示されている分類が最も精緻であり体系化されたものであると言っても過言ではなかろう。従って、それらの著書にまとめられ、提案されている分類は、タグとして十分機能できると思われるので、以上の説を中心に、タグの選択を行えば、タグの体系が保たれるだけではなく、内容の確認も容易にできるようになるであろう。

従って、すべてのタグを取り入れることが不可能であるという現実を踏まえつつ、次のような各レベルの基本を反映するタグがあれば、一応基礎的な分析や研究ができるかと思われる。動詞のタグには奥田説と工藤説も取り入れ、それ以外のタグには他の通説や定説になったものもすこし取り入れている<sup>5</sup>。

主語、主題、補語、状況語、名詞的修飾、副詞的修飾、連体修飾、VP、AP

独立語文、NP述語、人称詞述語、指示詞述語、

5 【】と【】内の内容はタグではない。

疑問詞述語、数詞述語、助数詞述語、名詞述語 NP、人称詞、指示詞、疑問詞、数詞、助数詞、連体詞、オノマトペ、る、た、ている、ていた

接頭辞、接尾辞、テンス・アスペクト、敬語・ポライトネス

格助詞、複合格助詞、連体助詞、並列助詞、とりたて助詞、接続助詞、終助詞、接続、助動詞、感動詞

一項動詞、二項動詞、三項動詞、ゼロ項動詞、自動詞、他動詞、非能格自動詞、非対格自動詞、有対動詞、無対動詞、補助動詞、相互動詞

【日本語記述文法研究会の分類】動き動詞、継続動詞、継続動詞\_\_主体動作、継続動詞\_\_主体変化、継続動詞\_\_均質主体動作、継続動詞\_\_特定時点\_\_主体動作、継続動詞\_\_特定時点\_\_結果維持、継続動詞\_\_特定時点\_\_結果残存、継続動詞\_\_特定時点\_\_非可逆的变化、継続動詞\_\_主体変化、継続動詞\_\_主体変化\_\_結果残存、継続動詞\_\_主体変化\_\_非可逆的变化、瞬間動詞、瞬間動詞\_\_主体動作、瞬間動詞\_\_主体動作\_\_変化なし、瞬間動詞\_\_主体動作\_\_結果維持、瞬間動詞\_\_主体動作\_\_結果残存、瞬間動詞\_\_主体動作\_\_非可逆的变化、瞬間動詞\_\_主体変化、瞬間動詞\_\_主体変化\_\_結果維持、瞬間動詞\_\_主体変化\_\_結果残存、瞬間動詞\_\_主体変化\_\_非可逆的变化、状態動詞、形状動詞\_\_スル、状態動詞\_\_シテイル、状態動詞\_\_スル・シテイル<sup>6</sup>

【奥田靖雄氏の分類】もようがえ動詞、とりつけ動詞、とりはずし動詞、うつしかえ動詞、ふれあい動詞、結果的なむすびつき動詞

【工藤真由美氏の分類】客体状態位置変化動詞、所有関係変化動詞、主体変化・主体動作動詞、意志\_\_位置姿勢社会的変化動詞、無意志\_\_状態位置変化動詞、主体動作・客体動き動詞、主体動作・客体接触動詞、認識\_\_言語\_\_表現\_\_活動動詞、意志的動作動詞、長期的動作動詞、非意志\_\_現象動詞、思考動詞、感情動詞、知覚動詞、感覚動詞、存在動詞、空間的配置動詞、関係動詞、特性動詞<sup>7</sup>

難易構文、比較構文、程度構文、認識動詞構文、変化構文、所在構文、付帯状況構文、数量構文、受身文、使役文、使役受身文、可能構文、自発構文、相互構文、再帰構文、否定文、授受文、使役授受文、義務文、許可文、禁止文、推量・判断文、意志・願望文、命令・依頼・勧誘文、疑問文、分類できない表現

有情物、無情物、有界、非有界、定指示、不定指示

主体、対象、相手、起点、着点、経過域、場所、起因・根拠

手段、時、領域、目的、役割、割合、限界、状態、内容

【主体】意志動作の主体、受身的動作の主体、自然現象の主体、変化の主体、性質の主体、所有

【奥田靖雄氏の分類】もようがえ動詞、とりつけ動詞、とりはずし動詞、うつしかえ動詞、ふれあい動詞、結果的なむすびつき動詞

【工藤真由美氏の分類】客体状態位置変化動詞、所有関係変化動詞、主体変化・主体動作動詞、意志\_\_位置姿勢社会的変化動詞、無意志\_\_状態位置変化動詞、主体動作・客体動き動詞、主体動作・客体接触動詞、認識\_\_言語\_\_表現\_\_活動動詞、意志的動作動詞、長期的動作動詞、非意志\_\_現象動詞、思考動詞、感情動詞、知覚動詞、感覚動詞、存在動詞、空間的配置動詞、関係動詞、特性動詞<sup>7</sup>

難易構文、比較構文、程度構文、認識動詞構文、変化構文、所在構文、付帯状況構文、数量構文、受身文、使役文、使役受身文、可能構文、自発構文、相互構文、再帰構文、否定文、授受文、使役授受文、義務文、許可文、禁止文、推量・判断文、意志・願望文、命令・依頼・勧誘文、疑問文、分類できない表現

有情物、無情物、有界、非有界、定指示、不定指示

主体、対象、相手、起点、着点、経過域、場所、起因・根拠

手段、時、領域、目的、役割、割合、限界、状態、内容

【主体】意志動作の主体、受身的動作の主体、自然現象の主体、変化の主体、性質の主体、所有

【奥田靖雄氏の分類】もようがえ動詞、とりつけ動詞、とりはずし動詞、うつしかえ動詞、ふれあい動詞、結果的なむすびつき動詞

【工藤真由美氏の分類】客体状態位置変化動詞、所有関係変化動詞、主体変化・主体動作動詞、意志\_\_位置姿勢社会的変化動詞、無意志\_\_状態位置変化動詞、主体動作・客体動き動詞、主体動作・客体接触動詞、認識\_\_言語\_\_表現\_\_活動動詞、意志的動作動詞、長期的動作動詞、非意志\_\_現象動詞、思考動詞、感情動詞、知覚動詞、感覚動詞、存在動詞、空間的配置動詞、関係動詞、特性動詞<sup>7</sup>

難易構文、比較構文、程度構文、認識動詞構文、変化構文、所在構文、付帯状況構文、数量構文、受身文、使役文、使役受身文、可能構文、自発構文、相互構文、再帰構文、否定文、授受文、使役授受文、義務文、許可文、禁止文、推量・判断文、意志・願望文、命令・依頼・勧誘文、疑問文、分類できない表現

有情物、無情物、有界、非有界、定指示、不定指示

主体、対象、相手、起点、着点、経過域、場所、起因・根拠

手段、時、領域、目的、役割、割合、限界、状態、内容

6 このタグは、略したものがあ。原文との対応関係は次の通りである。

「継続動詞\_\_主体動作=主体動作動詞」「継続動詞\_\_主体変化=主体変化動詞」「継続動詞\_\_均質主体動作=均質型」「継続動詞\_\_特定時点\_\_主体動作=特定時点成立型」「継続動詞\_\_特定時点\_\_結果維持=結果が維持されるタイプ」「継続動詞\_\_特定時点\_\_結果残存=結果が残存するタイプ」「継続動詞\_\_特定時点\_\_非可逆的变化=変化が非可逆的なタイプ」「継続動詞\_\_主体変化=主体変化動詞」「継続動詞\_\_主体変化\_\_結果残存=結果が残存するタイプ」「継続動詞\_\_主体変化\_\_非可逆的变化=変化が非可逆的なタイプ」「瞬間動詞\_\_主体動作=主体動作動詞」「瞬間動詞\_\_主体動作\_\_変化なし=変化がないタイプ」「瞬間動詞\_\_主体動作\_\_結果維持=結果が維持されるタイプ」「瞬間動詞\_\_主体動作\_\_結果残存=結果が残存するタイプ」「瞬間動詞\_\_主体動作\_\_非可逆的变化=変化が非可逆的なタイプ」「瞬間動詞\_\_主体変化=主体変化動詞」「瞬間動詞\_\_主体変化\_\_結果維持=結果が維持されるタイプ」「瞬間動詞\_\_主体変化\_\_結果残存=結果が残存するタイプ」「瞬間動詞\_\_主体変化\_\_非可逆的变化=変化が非可逆的なタイプ」「形状動詞\_\_スル=スル形状動詞」「状態動詞\_\_シテイル=シテイル形状動詞」「状態動詞\_\_スル・シテイル=スル・シテイル形両用状態動詞」

7 このタグは、略したものがあ。原文との対応関係は次の通りである。

「客体状態位置変化動詞=客体の状態変化・位置変化をひきおこす動詞」「所有関係変化動詞=所有関係の変化をひきおこす動詞」「主体変化・主体動作動詞=主体変化・主体動作動詞」「意志\_\_位置姿勢社会的変化動詞=人の意志的な(位置・姿勢)変化動詞」「無意志\_\_状態位置変化動詞=ものの無意志的な(状態・位置)変化動詞」「主体動作・客体動き動詞=主体動作・客体動き動詞」「主体動作・客体接触動詞=主体動作・客体接触動詞」「認識\_\_言語\_\_表現\_\_活動動詞=人の認識活動・言語活動・表現活動動詞」「意志的動作動詞=人の意志的動作動詞」「長期的動作動詞=人の長期的動作動詞」「非意志\_\_現象動詞=ものの非意志的な動き(現象)動詞」「特性動詞=特性動詞思考動詞」

の主体、動きの主体、動作の主体、状態の主体、同定関係の主体、事態の主体、働きかける主体、関係の主体、移動の主体、存在の主体、能力・知覚の主体、能力の主体、言語活動・知覚の主体、心的状態の主体、心的活動・態度の主体、心的活動の主体

【対象】形状変化の対象、位置変化の対象、状況変化の対象、算出の対象、働きかけの対象、言語活動の対象、働きかけられる対象、心的状態の対象、能力の対象、所有の対象、変化の対象、動作の対象、心的活動の対象、働きかけによって変化する対象、働きかけによって出現する対象、結びつけられる対象、移動の目標としての対象、方向性をもつ対象、能力・知覚の対象、移動する対象

【相手】動作の相手、授与の相手、受身的動作の相手、基準としての相手、共同動作の相手、相互動作の相手、基準としての相手、必須の相手

【起点】移動の起点、方向の起点、範囲の始点、変化前の状態

【着点】移動の着点、変化の結果、移動の方向、範囲の終点

【経過域】空間的な経過域、時間的な経過域

【場所】存在の場所、出現の場所、動きの場所、対象が移動し接触する場所

【起因・根拠】感情・感覚の起因、継続的状态の起因、出来事の原因、判断の根拠、変化の原因、行動の理由

【手段】内容物、付着物、道具、方法、材料、構成要素

【時】時点

【領域】認識の成り立つ領域、評価の成り立つ領域

【目的】移動の目的、動作の目的

【役割】名目

【割合】割合の基準

【限界】範囲の上限

【様態】動きの様態

【内容】内容

#### 4. 用例分析に求められるタグ付与の機能

以上に示したタグリストのように、基本的なタグだけに限ってもかなりの数になる。それらのタグを手動で付与するのは容易なことではないとすぐに想像できるし、タグと前後の文の成分やタグとタグを区切る記号を入力しなければならないということまで考慮すると、作業の難度は格段に上がる。

タグ付与の手間暇を少しでも軽減し、手動でタグを付与する際に生じるミスも減らすために、タグ付与ソフトの開発が強く求められるのも容易に理解できるであろう。しかし、研究の現場では、強いニーズがあるにもかかわらず、現在公開されているコーパス関連のソフトなどを調べてみればわかるように、タグ付与ソフトはなかなか見つからないのが現状である。

タグ付与ソフトを開発するとすれば、どのような機能が求められるかを明らかにしなければならない。聞き取りの調査の結果、少なくとも次のような機能が求められているようである。

- ①選択型のタグ付与機能
- ②内容確認機能
- ③自動的に区切る記号付与機能
- ④自動タグ付与機能

「選択型のタグ付与機能」とは、画面に「3.」で示されたタグのリストをボタン式で並べ、自由選択できるという機能のことである。「内容確認機能」とは、どの成分がどの分類に入るかを確認するためのものである。「自動的に区切る記号付与機能」とは、タグとそのタグの前文と後文の区切りや、2つ以上のタグを付与する場合、タグとタグの区切りを示す記号を付与するという機能のことである。「自動タグ付与機能」とは、分析の対象を選択

するだけで、自動的にタグが付与されるという機能のことである。

以上のニーズに応えるために、試行錯誤を繰り返しながら、本研究の内容となるタグ付与ソフト TNR\_JapaneseCorpusTagger を開発した。タグ付与ソフト TNR\_JapaneseCorpusTagger は、「2.」の方針に基づいて「3.」で示したタグの種類をすべて取り入れ、「4.」で明らかになった実際に求められている機能をすべて使うことができるように、開発したあと、幾たびの修正と改善を重ねてきたものである。

## 5. タグ付与ソフト TNR\_JapaneseCorpusTagger

タグ付与ソフト TNR\_JapaneseCorpusTagger は、前例を見ないソフトであるため、タグの種類やタグ付与の機能などにおいても、様々な課題やさらに改善すべきところが残っている。今後、使用者からのご提案やご意見などを取り入れながら、さらなる改良や改善を図っていきたいと考えている。

このソフトに掲載しているタグリストは、主に統語論や意味論を中心としたものであるので、統

語論や意味論の立場でのタグ付与はまずできるであろう。語用論や認知言語学、談話文法など、他の分野のタグについては、使用者が自分でタグリストを作成し、タグ一覧に搭載させる機能も設けているので、それを活用すれば、理論上では、どの分野のタグも自由に付与することができるようになっている。

TNR\_JapaneseCorpusTagger のメイン画面が図1のようになる。

画面の一番上の左側のエリアは、言語を選択するためのものである。このタグ付与ソフトは、日本語だけではなく、中国語のタグ付与もできるようになっている。現段階では取り入れていないが、必要に応じて、多言語を取り入れることも可能である。言語選択のエリアの右側は、[ディレクトリ]であり、タグ付与予定の文章を収納するフォルダのことである。複数のフォルダがある場合、下向きの▼をクリックすれば、必要なフォルダを選択することができる。[ファイル]において、下向きの▼を選択すれば、[ディレクトリ]のフォルダに保存されている文章を選択することができる。[読み込み]ボタンをクリックすれば、選択された文章が下の表示エリアに表示される。[編集



図1

保存]ボタンは、タグ付与を行った文章を保存するためのものである。

二列目のボタンにおいて、[元に戻す]は、付与したタグを削除し、文章を元の状態に戻すためのものであり、[やり直し]は現状に戻したものをタグ付与した状態に復元するためのものである。[タグ]をクリックすれば、タグ一覧が表示される。[検索]は、作業用の検索エリアに表示された文章内で検索するためのものであり、[全検索]は[ディレ

クトリ]のフォルダに含まれているすべての文章を対象に検索するためのものである。[オノマトベ]は、小野正弘編(2007)『擬音語・擬態語4500日本語オノマトベ辞典』(小学館)に収録されている4500の擬音語と擬態語を含んだもので、そのボタンをクリックすれば、文章におけるオノマトベと思われる箇所がすべて赤字で表示される。[全タグ確認]は複数のファイルにおけるタグ付与の正否について一括で確認するものである。

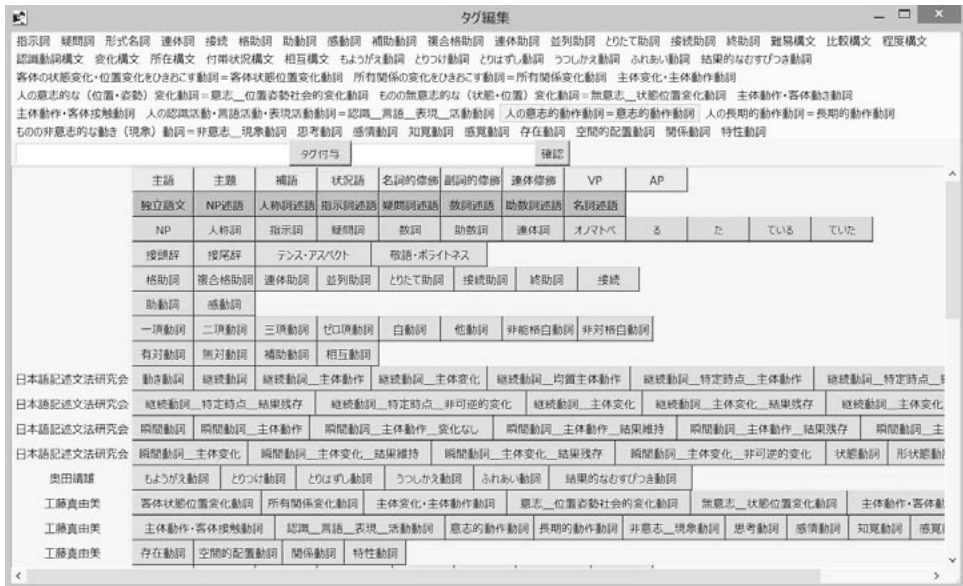


図2

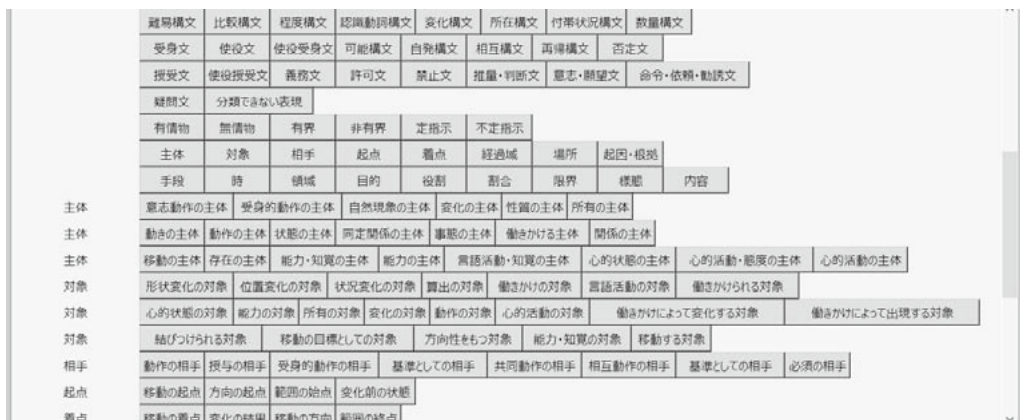


図3

経過域	空間的な経過域		時間的な経過域		
場所	存在の場所	出現の場所	動きの場所	対象が移動し、接触する場所	
起因・根拠	感情・感覚の起因		継続的状態の起因	出来事の原因	判断の根拠
			変化の原因	行動の理由	
手段	内容物		付着物	道具	方法
時	時点		材料		
領域	認識の成り立つ領域		評価の成り立つ領域		
目的	移動の目的		動作の目的		
役割	名目				
割合	割合の基準				
限界	範囲の上限				
様態	動きの様態				
内容	内容				

図4

[タグ]のボタンの話に戻る。その[タグ]のボタンをクリックすれば、図2～図4の画面が現れてくる。一括で表示できないため、分割して表示している。

上のリストと下のタグ付与のボタンは、それぞれの役割が異なっている。図5のように、上のリストの部分は、タグの内容を確認するためのコーナーである。例えば、「指示詞」をクリックすれば、その「指示詞」に含まれる指示詞たるものが示され

る。指示詞かどうか判断がつかない場合、この機能を活用すれば、タグの選択に繋がる。

上部のタグリストに含まれているものであるならば、タグ付与の対象を選択すれば、図6のように自動的に選択肢が現れ、さらに、その選択肢をクリックすれば、タグが付与されるようになっている。

TNR\_JapaneseCorpusTaggerにおいては、次のタグなら、自動的にタグの選択ができるように

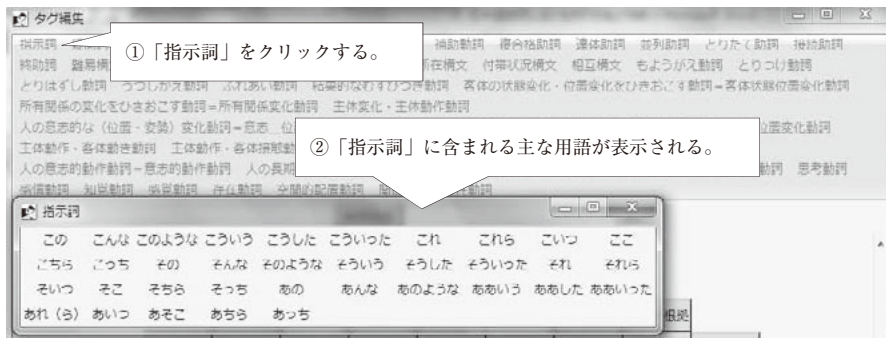


図5

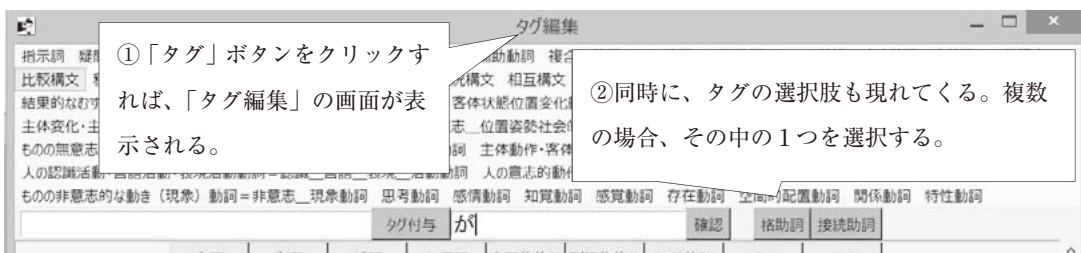


図6



なっている。

「指示詞」「疑問詞」「形式名詞」「連体詞」「接続」「格助詞」「助動詞」「感動詞」「補助動詞」「複合格助詞」「連体助詞」「並列助詞」「とりたて助詞」「接続助詞」「終助詞」「難易構文」「比較構文」「程度構文」「認識動詞構文」「変化構文」「所在構文」「付帯状況構文」「相互構文」「もようがえ動詞」「とりつけ動詞」「とりはずし動詞」「うつつかえ動詞」「ふれあい動詞」「結果的なむすびつき動詞」「客体状態位置変化動詞」「所有関係変化動詞」「主体変化・主体動作動詞」「意志\_\_位置姿勢社会的変化動詞」「無意志\_\_状態位置変化動詞」「主体動作・客体動き動詞」「主体動作・客体接触動詞」「認識\_\_言語\_\_表現\_\_活動動詞」「意志的動作動詞」「長期的動作動詞」「非意志\_\_現象動詞」「思考動詞」「感情動詞」「知覚動詞」「感覚動詞」「存在動詞」「空間的配置動詞」「関係動詞」「特性動詞」

タグ付与ソフトTNR\_JapaneseCorpusTaggerに設けている自動タグ選択機能は、タグ選択の迷いやミスを防ぎ、より正確に研究対象の成分にタグを付与することに必要不可欠であろう。今後、通説や定説または広く使われている個性の強い説に基づき、さらに補充していく予定である。

## 6. まとめ

以上のように、タグ付与ソフトは、人間の脳の代わりに、より正確でスピーディにタグの分析や統計結果から語構成や構文などの規則性を見出すための非常に有力なツールとなることが明らかになった。

TNR\_JapaneseCorpusTaggerはあくまでもタグ付与の1つのツールであり、タグ付与をより円滑に行うための土台作りである。従って、そこに添付されているタグリストは、1つの提案としては成り立つが、使用者にとってなにが最も必要なタグなのかは使用者が自分で創意工夫をして、そ

の土台の上で築いていかなければならない。

今後は、「自動タグ付与機能」をさらに強化し、デフォルトのリストを充実させ、より軽便なタグ付与ソフトになるよう、改善を図っていきたい。

## 参考文献

- 石綿敏雄 1999 『現代言語理論と格』 ひつじ書房  
 于康 2012 『現代日語語言学叢書 語法学』 高等教育出版社  
 于康 2013 『方法工具与日語教学研究叢書I 語料庫の制作与日語研究』 浙江工商大学出版社  
 于康・田中良・高山弘子 2014 『方法工具与日語教学研究叢書II 加注標簽軟件の使用与日語、日語偏誤和翻譯教学研究』 浙江工商大学出版社 (刊行予定)  
 奥田靖雄 1983 『格のかたちをとる名詞と動詞とのくみあわせ』 『日本語文法・連語論(資料編)』 むぎ書房  
 木村睦子 1997 『国立国語研究所報告 113 日本語における表層格と深層格の対応関係』 三省堂  
 工藤真由美 1995 『アスペクト・テンス体系とテキスト—現代日本語の時間の表現』 ひつじ書房  
 仁田義雄編 1993 『日本語の格をめぐって』 くろしお出版  
 仁田義雄 2009 『日本語の文法カテゴリをめぐって(仁田義雄日本語文法著作選 第1巻)』 ひつじ書房  
 仁田義雄 2009 『日本語のモダリティとその周辺(仁田義雄日本語文法著作選 第2巻)』 ひつじ書房  
 仁田義雄 2010 『語彙論的統語論の観点から(仁田義雄日本語文法著作選 第3巻)』 ひつじ書房  
 仁田義雄 2010 『日本語文法の記述的研究を求めて(仁田義雄日本語文法著作選 第4巻)』 ひつじ書房  
 日本語記述文法研究会 2010 『現代日本語文法1 第1部総論 第2部形態論』 くろしお出版  
 日本語記述文法研究 2009 『現代日本語文法2 第3部格と構文 第4部ヴォイス』 くろしお出版  
 日本語記述文法研究 2007 『現代日本語文法3 第5部アスペクト 第6部テンス 第7部肯否』 くろしお出版  
 日本語記述文法研究 2003 『現代日本語文法4 第8部モダリティ』 くろしお出版  
 日本語記述文法研究 2009 『現代日本語文法5 第9部とりたて 第10部主題』 くろしお出版  
 日本語記述文法研究 2008 『現代日本語文法6 第11部複文』 くろしお出版  
 日本語記述文法研究 2009 『現代日本語文法7 第12部談話 第13部待遇表現』 くろしお出版  
 早津恵美子 1989 『有対他動詞と無対他動詞の違いについて』 日本言語学会『言語研究』 95  
 益岡隆志 1987 『命題の文法』 くろしお出版  
 村木新次郎 1991 『日本語動詞の諸相』 ひつじ書房  
 宮島達夫 1972 『動詞の意味・用法の記述的研究』 秀英出版