

令和 5 年 5 月 9 日現在

機関番号：62618

研究種目：基盤研究(A) (一般)

研究期間：2017～2021

課題番号：17H00917

研究課題名(和文) 日本語歴史コーパスに対する統語・意味情報アノテーション

研究課題名(英文) Syntactic and Semantic Information Annotation on the Corpus of Historical Japanese

研究代表者

浅原 正幸 (ASAHARA, Masayuki)

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・研究系・教授

研究者番号：80379528

交付決定額(研究期間全体)：(直接経費) 33,080,000円

研究成果の概要(和文)：本研究では『日本語歴史コーパス』の平安～明治初期の資料に対して分類語彙表番号アノテーションを進め、64万語規模のデータを構築した。『現代日本語書き言葉均衡コーパス』新聞・書籍・雑誌に対する分類語彙表番号アノテーションデータ 34万語とともに合わせて、おおよそ 100万語規模の語義ラベルつきデータを構築した。
さらに日本語の統語構造データの基準として Universal Dependencies の言語資源の整備をつづけた。Universal Dependencies の団体との基準の協議を進めるとともに、日本語分かち書きの問題について新しい基準を提案した。

研究成果の学術的意義や社会的意義

本研究の意義は、既存の形態論情報タグつきデータに対して『分類語彙表』体系に基づいて単語の意味ラベルを付与したことにある。海外の取り組みをみても、古典-現代を通して 98万語規模の単語の意味ラベルがついたデータは類を見ない。
さらに統語的な研究として、日本語 Universal Dependencies の基準の検討を進めた。特に日本語の分かち書きの問題について取り組み、新しい分かち書きの基準について提案を行い、国際会議にて発表した。

研究成果の概要(英文)：In this study, we annotated the materials from the Heian to the early Meiji period in the "Japanese Historical Corpus" with the Classification of Japanese Vocabulary (Bunrui Goi Hyo) codes, and constructed a dataset of 640,000 words. Combined with the annotated dataset of 340,000 words for the "Balanced Corpus of Contemporary Written Japanese" (BCCWJ) consisting of newspapers, books, and magazines, we constructed a dataset of approximately 1 million words with sense labels.

Furthermore, we continued to develop Universal Dependencies language resources as a standard for Japanese syntactic structure data. We proceeded with discussions on standards with the Universal Dependencies organization and proposed new standards for Japanese word segmentation issues.

研究分野：コーパス言語学

キーワード：分類語彙表 Universal Dependencies

1. 研究開始当初の背景

データ駆動による日本語学研究に資する基礎データとして国立国語研究所にて『日本語歴史コーパス』の整備が継続的に進められている。2017年当時、日本語学分野において『日本語歴史コーパス』は既に非常に重要なデータベースとして位置付けられていた。同コーパスは、上代から近代の多くの資料を含み、形態論情報を付与し、形態論情報に基づく検索系「中納言」を介して公開されている。これらのテキストを分析することで、日本語の文法や語彙の変化を研究できる。日本語の歴史的な変化を計量的に調べるために、既に形態論情報などを手がかりとして利用できていたが、統語・意味情報については Oxford NINJAL Corpus of Old Japanese (ONCOJ) を除くと限定的であった。古典言語のテキストには、現代の言語学的知見をもとにアノテーションすることが求められる、現代にはない文法や語彙が多数存在するため、アノテーションするには専門家の知見が必要となる。古典言語のテキストに対する統語・意味情報のアノテーションにおいては、テキストの内容に対して様々な解釈が可能であるため、統一された方法論が必要となる。しかしながら、内省が効かないテキストに対して、統語・意味情報をアノテーションする方法論が確立しておらず、統語・意味情報について計量的な研究ができない状況であった。

2. 研究の目的

本研究の目的は、歴史的な文献を見る目を持つ研究者と、現代語で培った統語・意味情報アノテーションの研究者とが、知恵を出し合いながら「日本語歴史コーパス」に対するアノテーションの方法論を確立することにある。機械学習器が出力する尤度や、アノテーションの修正が簡便な GUI などを駆使しながら、内省が効かないテキストに対して集合知としてのアノテーションをいかに付与していくかを検討する。

研究の課題として Universal Dependencies に基づく係り受け情報アノテーションと分類語彙表に基づく語義情報アノテーションの2つを設定した。

Universal Dependencies (UD) とは、自然言語処理における言語現象の体系化を目的とした国際的な取り組みである。UD は、世界中の言語の構文解析を比較可能にするため、依存構造の表現方法を統一的に定義したものである。依存構造とは、単語間の係り受け関係を表した構造のことで、自然言語処理において文の構造を理解するために重要な情報である。UD では、依存構造をグラフとして表現し、各単語がどのような構文的な役割を持つかを明確にする。UD は、言語学的な知見に基づいて構築されており、世界中の言語に対して適用可能な体系となっている。日本語においても、UD に基づいた依存構造のアノテーションが行われ、様々な自然言語処理のタスクに応用されている。

分類語彙表は、日本語の語彙を体系的に分類するためのシソーラスである。この分類語彙表を利用して、単語や表現の語義をアノテーションすることは、自然言語処理において重要なタスクの一つである。分類語彙表を利用することで、多義語や同音異義語などの曖昧性を解消することができる。また、語彙の分類に基づいた語義のアノテーションは、単語の意味を明確に定義し、文脈からの解釈に依存しない一定の基準を提供することができる。

3. 研究の方法

Universal Dependencies (UD) は、言語学者や自然言語処理研究者による国際的な共同プロジェクトであり、各言語における統一的なデータ形式での依存構造のアノテーションを目指している。このような統一的なデータ形式は、異なる言語間の比較や統合が容易になり、自然言語処理の研究やアプリケーションにおいて大きなメリットがある。UD に基づくアノテーションを行うには、UD に準拠したアノテーションツールやガイドラインを使用する必要がある。また、各言語に対する UD のガイドラインは、国際的なコミュニティによって策定されるため、海外との交渉や協力が必要となる。例えば、UD の日本語版のガイドラインは、日本語を母語とする言語学者や自然言語処理研究者が中心となって策定されているが、それでも海外の専門家との議論やレビューが行われた。また、UD のガイドラインは定期的に更新されるため、海外との交流や共同作業が必要となる。海外との交渉が必要な理由としては、UD のような国際的な共同プロジェクトでは、言語によっては複数の国で話されることや、異なる国で異なる言語が話されることがある。そのため、多様な文化や言語背景を持つ人々が協力して、共通のアノテーションツールやガイドラインを策定する必要があるためである。海外との交渉が十分に行われ、共通理解が得られた上で、UD に基づくアノテーションが行われることで、より精度の高い自然言語処理が可能になる。

UD のアノテーションは内省が効かない古典語の作業は達成できず、近代語の作業が中心となった。現代語の Universal Dependencies の作業を進めながら、「明六雑誌」のサンプルに対して、文節係り受けを付与して、変換プログラムに基づき UD の体系に変換した。

分類語彙表アノテーションは2種類の言語資源を整理することからはじめた。一つは UniDic-分類語彙表対応表である。UniDic-分類語彙表対応表は、国語研内で整備されていた UniDic と呼ばれる日本語の形態素解析辞書と、『分類語彙表』との対応表である。分類語彙表は、日本語の単語を約3万のカテゴリーに分類する辞書であり、UniDic は日本語の形態素解析に広く用いられている辞書である。UniDic-分類語彙表対応表は、UniDic で扱われる語彙と分類語彙表で扱われるカテゴリーの対応を表しており、それぞれの単語がどのカテゴリーに分類されるかを確認することができる。これにより、分類語彙表に基づいた語義のアノテーションの効率化を図ることができる。もう一つは『古典対照分類語彙表』である。『古典対照分類語彙表』は、日本語の古典文献や歴史文書などに出現する語彙の分類語彙表である。現代語の分類語彙表と同様に、単語をある一定の分類カテゴリーに分類し、同じカテゴリーに分類された単語を集めたりリストを構成する。「竹取物語」「土佐日記」「方丈記」「徒然草」のアノテーションを進めながら、UniDic-古典対照分類語彙表の対応表も構築した。これらに基づき、分類語彙表番号を通時的にアノテーションする技術確立し、作業を進めた。

4. 研究成果

UD 関連については、日本語の Universal Dependencies 基準について継続的にメンテナンスを行った。近代語のデータについては UD Japanese Modern として公開するとともに JADH-2017 で発表を行った。関連して、UD Japanese GSD, UD Japanese PUD, UD Japanese BCCWJ, UD Japanese CEJC など多様な現代語の Universal Dependencies リソースの構築・保守を進めた。さらに UD のアノテーション基準の構築を進め、日本語の UD リソースにおける、単語分かち書きの問題についての論文を発表した。

分類語彙表関連は、まず『現代日本語書き言葉均衡コーパス』34万語に分類語彙表番号をアノテーションした。次に、UniDic-分類語彙表対応表・UniDic-古典対照分類語彙表対応表などを援用しながら、『日本語歴史コーパス』64万語に分類語彙表番号をアノテーションした。類(体・用・相・他)に基づく統計は下表の通りである。なお、未定義は付属語などである。

	体	用	相	他	未定義	総計
0900 竹取	2,318	2,252	706	72	7,409	12,757
0934 土佐	1,710	1,272	453	45	4,728	8,208
1100 今昔	40,687	29,498	8,518	1,189	95,706	175,598
1212 方丈	1,433	792	342	100	2,735	5,402
1220 宇治	24,214	21,336	6,290	716	68,149	120,705
1252 十訓	19,808	13,039	3,974	460	52,896	90,177
1336 徒然	8,876	6,138	2,688	213	22,919	40,834
1642 虎明	1,255	811	369	88	2,925	5,448
1895 太陽	13,256	6,131	3,116	853	23,038	46,394
1904 小読	10,846	5,312	2,620	794	25,762	45,334
1910 小読	28,915	12,833	6,388	1,135	47,623	96,894
総計	153,318	99,414	35,464	5,665	353,890	647,751

同内容は、以下の国際会議で発表を行った：

Masayuki Asahara, Nao Ikegami, Tai Suzuki, Taro Ichimura, Asuko Kondo, Sachi Kato, and Makoto Yamazaki. 2022. CHJ-WLSP: Annotation of ‘Word List by Semantic Principles’ Labels for the Corpus of Historical Japanese. In Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, pages 31-37, Marseille, France. European Language Resources Association.

5. 主な発表論文等

〔雑誌論文〕 計22件（うち査読付論文 18件 / うち国際共著 0件 / うちオープンアクセス 18件）

1. 著者名 加藤 祥, 浅原 正幸, 森山 奈々美, 荻原 亜彩美, 山崎 誠	4. 巻 28
2. 論文標題 『分類語彙表』に対する反対語情報付与	5. 発行年 2021年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 60~81
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.28.60	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 浅原 正幸, 加藤 祥	4. 巻 26,2
2. 論文標題 読み時間と統語・意味分類	5. 発行年 2019年
3. 雑誌名 認知科学	6. 最初と最後の頁 219-230
掲載論文のDOI (デジタルオブジェクト識別子) 10.11225/jcss.26.219	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 浅原 正幸	4. 巻 26,2
2. 論文標題 日本語の読み時間と節境界情報ー主辞後置言語における wrap-up effect の検証	5. 発行年 2019年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 301-328
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.26.301	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 鈴木 類, 古宮 嘉那子, 浅原 正幸, 佐々木 稔, 新納 浩幸	4. 巻 26,2
2. 論文標題 概念辞書の類義語と分散表現を利用した教師なし all-words WSD	5. 発行年 2019年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 361-380
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.26.361	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 浅原 正幸	4. 巻 26,3
2. 論文標題 単語埋め込みに基づくサプライザル	5. 発行年 2019年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 635-652
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.26.635	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 浅原 正幸	4. 巻 27,1
2. 論文標題 Bayesian Linear Mixed Model による単語親密度推定と位相情報付与	5. 発行年 2020年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 133-150
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.27.133	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 加藤 祥, 浅原 正幸, 山崎 誠	4. 巻 15,2
2. 論文標題 分類語彙表番号を付与した『現代日本語書き言葉均衡コーパス』の書籍・新聞・雑誌データ	5. 発行年 2019年
3. 雑誌名 日本語の研究	6. 最初と最後の頁 134-144
掲載論文のDOI (デジタルオブジェクト識別子) 10.20666/nihongonokenkyu.15.2_134	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Masayuki Asahara	4. 巻 なし
2. 論文標題 Word Familiarity Rate Estimation by Bayesian Linear Mixed Model	5. 発行年 2019年
3. 雑誌名 Proceedings of Aggregating and analysing crowdsourced annotations for NLP (AnnoNLP)	6. 最初と最後の頁 6-14
掲載論文のDOI (デジタルオブジェクト識別子) 10.18653/v1/D19-5902	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 松林 優一郎, 浅原 正幸	4. 巻 35.1
2. 論文標題 自然言語処理-言語資源・意味解析-	5. 発行年 2020年
3. 雑誌名 人工知能	6. 最初と最後の頁 89-105
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/jjsai.35.1_89	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 服部 隆	4. 巻 37
2. 論文標題 尾崎紅葉『多情多恨 前編』における文章展開 節 (clause) を用いた文体分析の試み (七)	5. 発行年 2020年
3. 雑誌名 上智大学国文学科紀要	6. 最初と最後の頁 266-249
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 市村 太郎	4. 巻 188
2. 論文標題 副詞「ほんとうに」の展開と「じつに」「まことに」 近代語から現代語へ	5. 発行年 2019年
3. 雑誌名 国文学研究	6. 最初と最後の頁 112-98
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 浅原正幸, 金山博, 宮尾祐介, 田中貴秋, 大村舞, 村脇有吾, 松本裕治	4. 巻 26-1
2. 論文標題 Universal Dependencies 日本語コーパス	5. 発行年 2019年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 3-36
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.26.3	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 浅原正幸	4. 巻 25-5
2. 論文標題 名詞句の情報の状態と読み時間について	5. 発行年 2018年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 527-554
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.25.527	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 宮内拓也, 浅原正幸, 中川奈津子, 加藤祥	4. 巻 -16
2. 論文標題 『現代日本語書き言葉均衡コーパス』への情報構造アノテーションとその分析	5. 発行年 2018年
3. 雑誌名 国立国語研究所論集	6. 最初と最後の頁 19-33
掲載論文のDOI (デジタルオブジェクト識別子) 10.15084/00001606	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 浅原正幸, 田中弥生	4. 巻 15
2. 論文標題 修辞ユニット分析における脱文脈化指数の妥当性の検証	5. 発行年 2018年
3. 雑誌名 国立国語研究所論集	6. 最初と最後の頁 1-15
掲載論文のDOI (デジタルオブジェクト識別子) 10.15084/00001593	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 浅原正幸, 松本裕治	4. 巻 25-4
2. 論文標題 『現代日本語書き言葉均衡コーパス』に対する文節係り受け・並列構造アノテーション	5. 発行年 2018年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 331-356
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.25.331	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 浅原正幸	4. 巻 14-3
2. 論文標題 数理的研究 (2016年・2017年における日本語学界の展望)	5. 発行年 2018年
3. 雑誌名 日本語の研究	6. 最初と最後の頁 99-104
掲載論文のDOI (デジタルオブジェクト識別子) 10.20666/nihongonokenkyu.14.3_99	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Masayuki Asahara	4. 巻 24
2. 論文標題 NWJC2Vec: Word embedding dataset from 'NINJAL Web Japanese Corpus'	5. 発行年 2018年
3. 雑誌名 Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication	6. 最初と最後の頁 7-25
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 浅原正幸, 河原一哉, 大場寧子, 前川喜久雄	4. 巻 59
2. 論文標題 『国語研日本語ウェブコーパス』とその検索系『梵天』	5. 発行年 2018年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 299-305
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 浅原正幸	4. 巻 10
2. 論文標題 『国語研日本語ウェブコーパス』とその検索系『梵天』	5. 発行年 2018年
3. 雑誌名 JSL漢字学習研究会誌	6. 最初と最後の頁 26-30
掲載論文のDOI (デジタルオブジェクト識別子) 10.20808/jslk.10.0_26	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 新納浩幸, 浅原正幸, 古宮嘉那子, 佐々木稔	4. 巻 24
2. 論文標題 nwjc2vec: 国語研日本語ウェブコーパスから構築した単語の分散表現データ	5. 発行年 2017年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 705-720
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.24.705	査読の有無 有
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 Omura Mai, Wakasa Aya, Asahara Masayuki	4. 巻 30
2. 論文標題 Universal Dependencies for Japanese Based on Long-Unit Words by NINJAL	5. 発行年 2023年
3. 雑誌名 Journal of Natural Language Processing	6. 最初と最後の頁 4~29
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.30.4	査読の有無 有
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

〔学会発表〕 計64件 (うち招待講演 2件 / うち国際学会 22件)

1. 発表者名 浅原正幸, 加藤祥
2. 発表標題 『日本語歴史コーパス』に対する文脈化単語埋め込みに基づく意味空間
3. 学会等名 人文科学とコンピュータシンポジウム「じんもんこん2020」
4. 発表年 2020年

1. 発表者名 加藤祥, 浅原正幸
2. 発表標題 多義語語義調査を目指した IPAL 形容詞例文への印象評定情報付与
3. 学会等名 言語処理学会第27回年次大会 (NLP2021)
4. 発表年 2021年

1. 発表者名 Teruo Hirabayashi, Kanako Komiya, Masayuki Asahara, Hiroyuki Shinnou
2. 発表標題 Automatic Creation of Correspondence Table of Meaning Tags from Two Dictionaries in One Language Using Bilingual Embedding
3. 学会等名 Proceedings of the 13th Workshop on Building and Using Comparable Corpora (国際学会)
4. 発表年 2020年

1. 発表者名 Kanako Komiya, Daiki Yaginuma, Masayuki Asahara and Hiroyuki Shinnou
2. 発表標題 Generation and Evaluation of Concept Embeddings Via Fine-Tuning Using Automatically Tagged Corpus
3. 学会等名 Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation (PACLIC 34) (国際学会)
4. 発表年 2020年

1. 発表者名 Teruo Hirabayashi, Kanako Komiya, Masayuki Asahara and Hiroyuki Shinnou
2. 発表標題 Composing Word Vectors for Japanese Compound Words Using Bilingual Word Embeddings
3. 学会等名 Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation (PACLIC 34) (国際学会)
4. 発表年 2020年

1. 発表者名 Takuya Miyauchi and Masayuki Asahara
2. 発表標題 Statistical Approaches to a Correlation between Information Structure and Word Orders of Noun Phrases in Japanese
3. 学会等名 16th International Conference of the Pacific Association for Computational Linguistics (国際学会)
4. 発表年 2019年

1. 発表者名 呉 佩珣, 近藤 森音, 森山 奈々美, 荻原 亜彩美, 加藤 祥, 浅原 正幸
2. 発表標題 『分類語彙表』と『岩波国語辞典第五版タグ付きコーパス2004』の対応表
3. 学会等名 言語資源活用ワークショップ2019
4. 発表年 2019年

1. 発表者名 西内 沙恵, 加藤 祥, 浅原 正幸
2. 発表標題 語義間類似度の双方向評定に基づくプロトタイプの意味の解明ークラウドソーシングを用いた量的調査による多義的形容詞分析ー
3. 学会等名 日本認知言語学会第20回全国大会
4. 発表年 2019年

1. 発表者名 加藤 祥, 西内 沙恵, 浅原 正幸
2. 発表標題 多義語用例の類似度による語義の分類; 「遠い」と「近い」を例に
3. 学会等名 日本認知言語学会第20回全国大会
4. 発表年 2019年

1. 発表者名 浅原 正幸
2. 発表標題 クラウドソーシング結果の可視化手法と統計処理
3. 学会等名 日本言語学会第158回大会
4. 発表年 2019年

1. 発表者名 加藤 祥, 浅原 正幸, 山崎 誠
2. 発表標題 『現代日本語書き言葉均衡コーパス』新聞・書籍・雑誌データの助動詞に対する用法情報付与
3. 学会等名 日本語学会2019年度春季大会
4. 発表年 2019年

1. 発表者名 加藤 祥
2. 発表標題 クラウドソーシングによる語義調査
3. 学会等名 日本言語学会第158回大会
4. 発表年 2019年

1. 発表者名 西内 沙恵
2. 発表標題 クラウドソーシングによる述定・装定の用法分析
3. 学会等名 日本言語学会第158回大会
4. 発表年 2019年

1. 発表者名 西内 沙恵, 加藤 祥, 浅原 正幸
2. 発表標題 ヒトによる多義的形容詞に対する類似性の評価データベース構築-「長い」と「短い」の事例から-
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 浅原 正幸, 加藤 祥
2. 発表標題 BERTed-BCCWJ: 多層文脈化単語埋め込み情報を付与した『現代日本語書き言葉均衡コーパス』データ
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 松田 寛, 若狭 絢, 山下 華代, 大村 舞, 浅原 正幸
2. 発表標題 UD Japanese GSD の再整備と固有表現情報付与
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 浅原 正幸, 西内 沙恵, 加藤 祥
2. 発表標題 NWJC-BERT: 多義語に対するヒトと文脈化単語埋め込みの類似性判断の対照分析
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 河村 宗一郎, 久本 空海, 真鍋 陽俊, 高岡 一馬, 内田 佳孝, 岡 照晃, 浅原 正幸
2. 発表標題 chiVe 2.0: SudachiとNWJCを用いた実用的な日本語単語ベクトルの実現へ向けて
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 山崎 誠
2. 発表標題 意味分類の客観性を探る：『分類語彙表増補改訂版』と『新明解類語辞典』との比較
3. 学会等名 2019年語彙研究会大会
4. 発表年 2019年

1. 発表者名 浅原正幸
2. 発表標題 名詞句の情報の状態と読み時間について
3. 学会等名 言語処理学会第25回年次大会（招待講演）
4. 発表年 2019年

1. 発表者名 荻原亜彩美，森山奈々美，浅原正幸，加藤祥，山崎誠
2. 発表標題 『分類語彙表』に対する反対語情報付与
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 森山奈々美，荻原亜彩美，近藤森音，浅原正幸，相澤彰子
2. 発表標題 BCCWJ-EyeTrack-2: 書籍と教科書の読み時間データ
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 加藤祥, 田邊絢, 浅原正幸, 古宮嘉那子, 新納浩幸
2. 発表標題 多義語の語義分布と語義間の派生関係調査の試み--相の類を中心に
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 浅原正幸
2. 発表標題 読み時間と述語項構造・共参照情報について
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 浅原正幸
2. 発表標題 テキストの読みやすさについて
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 浅原正幸
2. 発表標題 クラウドソーシングによる単語親密度推定
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 浅原正幸
2. 発表標題 単語埋め込みに基づくサプライザルのモデル化
3. 学会等名 日本言語学会第157回大会
4. 発表年 2018年

1. 発表者名 浅原正幸, 南部智史, 佐野真一郎
2. 発表標題 日本語の二重目的語構文の基本語順について
3. 学会等名 言語資源活用ワークショップ2018
4. 発表年 2018年

1. 発表者名 Sachi Kato, Masayuki Asahara, Makoto Yamazaki
2. 発表標題 Annotation of `Word List by Semantic Principles' Labels for the Balanced Corpus of Contemporary Written Japanese
3. 学会等名 Proceedings of PACLIC 32 (国際学会)
4. 発表年 2018年

1. 発表者名 Masayuki Asahara
2. 発表標題 Between Reading Time and Clause Boundaries in Japanese - Wrap-up Effect in a Head-Final Language
3. 学会等名 Proceedings of PACLIC 32 (国際学会)
4. 発表年 2018年

1. 発表者名 Mai Omura, Masayuki Asahara
2. 発表標題 UD-Japanese BCCWJ: Universal Dependencies Annotation for the Balanced Corpus of Contemporary Written Japanese
3. 学会等名 Proceedings of UDW 18 (国際学会)
4. 発表年 2018年

1. 発表者名 Hiroshi Kanayama, Na-Rae Han, Masayuki Asahara, Jena D. Hwang, Yusuke Miyao, Jinho D. Choi, Yuji Matsumoto
2. 発表標題 Coordinate Structures in Universal Dependencies for Head-final Languages
3. 学会等名 Proceedings of UDW 18 (国際学会)
4. 発表年 2018年

1. 発表者名 Masayuki Asahara, Satoshi Nambu, Shin-ichiro Sano
2. 発表標題 Predicting Japanese Word Order in Double Object Constructions
3. 学会等名 Proceedings of CogACLL 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Satomi Matsumoto, Masayuki Asahara, Setsuko Arita
2. 発表標題 Japanese clause classification annotation on the 'Balanced Corpus of Contemporary Written Japanese'
3. 学会等名 Proceedings of Asian Language Resources 13 (国際学会)
4. 発表年 2018年

1 . 発表者名 Masayuki Asahara, Hiroshi Kanayama, Takaaki Tanaka, Yusuke Miyao, Sumire Uematsu, Shinsuke Mori, Yuji Matsumoto, Mai Omura, Yugo Murawaki
2 . 発表標題 Universal Dependencies Version 2 for Japanese
3 . 学会等名 Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC-2018) (国際学会)
4 . 発表年 2018年

1 . 発表者名 Rui Suzuki, Kanako Komiya, Masayuki Asahara, Minoru Sasaki, Hiroyuki Shinnou
2 . 発表標題 All-words Word Sense Disambiguation Using Concept Embeddings
3 . 学会等名 Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC-2018) (国際学会)
4 . 発表年 2018年

1 . 発表者名 Masayuki Asahara, Sachi Kato
2 . 発表標題 Between Reading Time and Syntactic/Semantic Categories
3 . 学会等名 Proceedings of the The 8th International Joint Conference on Natural Language Processing (国際学会)
4 . 発表年 2017年

1 . 発表者名 Masayuki Asahara
2 . 発表標題 Between Reading Time and Information Structure
3 . 学会等名 Proceedings of The 31st Pacific Asia Conference on Language, Information and Computation PACLIC 31 (2017) (国際学会)
4 . 発表年 2017年

1 . 発表者名 Mai Omura, Yuta Takahashi, Masayuki Asahara
2 . 発表標題 Universal Dependency for Japanese Modern Languages
3 . 学会等名 JADH 2017, The Japanese Association for Digital Humanities Conference 2017 (国際学会)
4 . 発表年 2017年

1 . 発表者名 Masayuki Asahara, Nao Ikegami, Yutaka Hara, Sachi Kato, Tai Suzuki
2 . 発表標題 Annotation of ' Word List by Semantic Principles ' Labels for `Corpus of Historical Japanese ' Heian Period Series -- Trial Annotation on Tosa Nikki and Taketori Monogatari --
3 . 学会等名 JADH 2017, The Japanese Association for Digital Humanities Conference 2017 (国際学会)
4 . 発表年 2017年

1 . 発表者名 Takuya Miyauchi, Masayuki Asahara, Natsuko Nakagawa, Sachi Kato
2 . 発表標題 Annotation of Information Structure on ``The Balanced Corpus of Contemporary Written Japanese''
3 . 学会等名 Proceedings of PACLING 2017, the 15th International Conference of the Pacific Association for Computational Linguistics (国際学会)
4 . 発表年 2017年

1 . 発表者名 Masayuki Asahara
2 . 発表標題 Word Familiarity Rate Estimation for `Word List by Semantic Principles' -- a Case Study of Adjective --
3 . 学会等名 Mental Architecture for Processing and Learning of Language (MAPLL) and Tokyo Conference on Psycholinguistics (TCP) 2017 (国際学会)
4 . 発表年 2017年

1. 発表者名 浅原正幸
2. 発表標題 『国語研日本語ウェブコーパス』とその検索系『梵天』
3. 学会等名 韓国日本学会（招待講演）（国際学会）
4. 発表年 2018年

1. 発表者名 浅原正幸, 近藤明日子, 加藤祥
2. 発表標題 国語研で開発している 語義タグ付きデータと辞書
3. 学会等名 「all-words WSD システムの構築及び分類語彙表と岩波国語辞典の対応表作成への利用」研究発表会
4. 発表年 2018年

1. 発表者名 菊地礼, 加藤祥, 浅原正幸
2. 発表標題 「感じる」を指標とするメタファー用例の収集とその分析
3. 学会等名 日本語用論学会 メタファー研究会 2018年3月18・19日 2-Day シンポジウム「身体性」
4. 発表年 2018年

1. 発表者名 田中弥生, 浅原正幸
2. 発表標題 児童による作文の修辞ユニット分析における中核要素認定
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

1. 発表者名 大村舞, 浅原正幸
2. 発表標題 UD Japanese BCCWJ: 現代日本語書き言葉均衡コーパスのUniversal Dependencies
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

1. 発表者名 平林照雄, 鈴木類, 古宮嘉那子, 浅原正幸, 佐々木稔, 新納浩幸
2. 発表標題 『岩波国語辞典』の語義タグを用いたall-wordsの語義曖昧性解消
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

1. 発表者名 加藤祥, 浅原正幸
2. 発表標題 読み手が共通の認識を得るための情報とその表現 小説のタイトルと帯から読み手が取得する情報
3. 学会等名 第41回社会言語科学会研究大会
4. 発表年 2018年

1. 発表者名 浅原正幸
2. 発表標題 読み時間と節境界について
3. 学会等名 シンポジウム「日本語学習者はどのように文章を理解しているのか 目の動きから見えてくるもの」
4. 発表年 2018年

1. 発表者名 浅原正幸
2. 発表標題 Universal Dependencies プロジェクトと日本語チームの活動
3. 学会等名 「言語における系統・変異・多様性とその数理」シンポジウム
4. 発表年 2018年

1. 発表者名 加藤祥, 浅原正幸
2. 発表標題 意味分野の結合類型を用いてコーパスから隠喩用例収集を試みる
3. 学会等名 日本認知言語学会第18回全国大会
4. 発表年 2017年

1. 発表者名 浅原正幸, 加藤祥
2. 発表標題 読み時間と統語・意味分類
3. 学会等名 日本認知科学会第34回大会
4. 発表年 2017年

1. 発表者名 大村舞, 浅原正幸
2. 発表標題 現代日本語書き言葉均衡コーパスのUniversal Dependencies
3. 学会等名 言語資源活用ワークショップ2017
4. 発表年 2017年

1. 発表者名 加藤祥, 浅原正幸
2. 発表標題 分類語彙表番号を用いた比喻表現収集の試み
3. 学会等名 言語資源活用ワークショップ2017
4. 発表年 2017年

1. 発表者名 浅原正幸, 田中弥生
2. 発表標題 修辭ユニット分析における脱文脈化指数の妥当性の検証
3. 学会等名 言語資源活用ワークショップ2017
4. 発表年 2017年

1. 発表者名 浅原正幸
2. 発表標題 読み時間と節境界について
3. 学会等名 日本言語学会第154回大会
4. 発表年 2017年

1. 発表者名 Mai Omura, Aya Wakasa, and Masayuki Asahara
2. 発表標題 Word delimitation issues in UD Japanese
3. 学会等名 Proceedings of the Fifth Workshop on Universal Dependencies (UDW, SyntaxFest 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Masayuki Asahara, Nao Ikegami, Tai Suzuki, Taro Ichimura, Asuko Kondo, Sachi Kato, Makoto Yamazaki
2. 発表標題 CHJ-WLSP: Annotation of `Word List by Semantic Principles' Labels for the Corpus of Historical Japanese
3. 学会等名 Proceedings of 2nd Workshop on Language Technologies for Historical and Ancient Languages (国際学会)
4. 発表年 2022年

1. 発表者名 Masayuki Asahara, Nao Ikegami, Tai Suzuki, Taro Ichimura, Asuko Kondo, Sachi Kato, Makoto Yamazaki
2. 発表標題 The Annotation of Antonym Information in the `Word List by Semantic Principles'
3. 学会等名 Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation (国際学会)
4. 発表年 2021年

1. 発表者名 浅原正幸・池上尚・鈴木泰・市村太郎・近藤明日子・加藤祥・山崎誠
2. 発表標題 分類語彙表番号を付与した『日本語歴史コーパス』データ
3. 学会等名 日本語学会2022年度春季大会
4. 発表年 2022年

1. 発表者名 松田 寛, 大村 舞, 浅原 正幸
2. 発表標題 UD Japanese に基づく国語研長単位解析系の構築
3. 学会等名 言語処理学会第28回年次大会発表論文集
4. 発表年 2022年

1. 発表者名 大村 舞, 若狭 絢, 浅原 正幸
2. 発表標題 国語研長単位に基づく UD Japanese
3. 学会等名 言語処理学会第28回年次大会発表論文集
4. 発表年 2022年

1. 発表者名 浅原 正幸, 若狭 絢
2. 発表標題 『日本語日常会話コーパス』に対する係り受け情報アノテーション
3. 学会等名 言語処理学会第28回年次大会発表論文集
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

wisp2unidic https://github.com/masayu-a/WLSP2UniDic WLSP-antonym https://github.com/masayu-a/WLSP-antonym WLSP2UniDic_historical https://github.com/masayu-a/WLSP2UniDic_historical BCCWJ-WLSP https://github.com/masayu-a/BCCWJ-WLSP

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	服部 隆 (Hattori Takashi) (10289598)	上智大学・文学部・教授 (32621)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	古宮 嘉那子 (Kanako Komiya) (10592339)	東京農工大学・工学(系)研究科(研究院)・准教授 (12605)	
研究分担者	市村 太郎 (Taro Ichimura) (10701352)	京都府立大学・文学部・准教授 (24302)	
研究分担者	山崎 誠 (Yamazaki Makoto) (30182489)	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・研究系・教授 (62618)	
研究分担者	宮内 佐夜香 (Miyouchi Sayaka) (30508502)	中京大学・文学部・准教授 (33908)	
研究分担者	加藤 祥 (Kato Sachi) (40623004)	目白大学・外国語学部・専任講師 (32414)	
研究分担者	池上 尚 (Ikegami Nao) (50739125)	埼玉大学・教育学部・准教授 (12401)	
研究分担者	近藤 明日子 (Kondo Asuko) (30425722)	東京大学・人文社会系研究科・助教 (12601)	
研究分担者	岡 照晃 (Oka Teruaki) (50782942)	東京都立大学・システムデザイン学部・特任助教 (22604)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	鈴木 泰 (Suzuki Tai) (70091832)	東京大学・人文社会系研究科・名誉教授 (12601)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関