

Title	深い意味処理のための統合日本語コーパスの開発と言語研究
Sub Title	Linguistic study and development of an integrated Japanese corpus for deep semantic processing
Author	小原, 京子(Ohara, Kyoko)
Publisher	慶應義塾大学
Publication year	2018
Jtitle	学事振興資金研究成果実績報告書 (2017.)
JaLC DOI	
Abstract	<p>本研究の究極的な目的は、研究代表者が構築した認知言語学に基づく詳細な意味情報を付与した言語資源「日本語フレームネット」(JFN)を、言語研究のみならず自然言語処理にとってもこれまで以上に有益な言語資源とする方策を検討することである。そのためには、JFNの規模を大きくさせることと、JFNの持つ豊富な意味情報に加え統語情報をも含める、という2つの方向性が考えられる。本研究では特に、国立国語研究所で構築中のNINJAL Parsed Corpus of Modern Japanese (NPCMJ)とリンクさせることを検討した。NPCMJは文の統語情報のみで意味情報は含まないが規模は大きいのに対しJFNは詳細な意味情報を含むが規模が小さく、両者を統合することで深い意味処理を行う自然言語処理アプリケーションのための言語資源とすることができると考えたからである。</p> <p>当初計画したJFNとNPCMJのリンク方法は2つあった。一つ目は、JFNで意味情報付与の際の入力データとしてNPCMJの統語情報付きデータを用いることである。二つ目は、JFNとNPCMJを統合させ、認知科学的にも妥当な詳細な意味情報と文の統語情報の両方を含んだコーパスとすることである。前者に関しては、ツールを開発した。後者については実現までの目処がたった。</p> <p>本研究の経過・成果は、国際認知言語学会、マルチリンガルフレームネット会議、言語処理学会第24回年次大会にて発表した。</p> <p>本研究によりNPCMJとJFNの統合は実現可能であることが明らかとなった。今後はJFNを深い意味処理を行う自然言語処理アプリケーションにとってさらに有益な資源とするため、より規模の大きく自然言語処理分野で定評のある言語資源とのリンクを検討したい。</p> <p>The principal investigator (PI) of this research project has been developing Japanese FrameNet (JFN), a language resource that contains rich semantic information which is based on the theory of Frame Semantics. The ultimate goal of this research project is to enable JFN to be useful for natural language processing (NLP) applications, especially for those that involve deep semantic processing. In order to achieve this goal, there are two directions. One is to make the size of JFN larger. The other is to add syntactic information to JFN, in addition to the existing semantic information. This research project pursued the latter approach and focused on ways to link the JFN data with the NINJAL (National Institute for Japanese Language and Linguistics) Parsed Corpus of Modern Japanese (NPCMJ). NPCMJ does not contain semantic information at all but includes detailed syntactic information of sentences and its size is larger than that of JFN. Therefore, it was expected that the combined language resource would be useful for various NLP applications. There are two ways to link the JFN data with that of NPCMJ. One is to use the NPCMJ data as input to the JFN annotation (tagging) process. The other is to integrate JFN and NPCMJ to build a combined corpus. As for the former, a tool was implemented to allow the NPCMJ data as input to the JFN annotation process. As for the latter, how to combine the two resources was investigated. The accomplishments of the project were presented at ICLC14 (Tartu, Estonia ; July 2017), the Multilingual FrameNet meeting (Vancouver, Canada ; Aug. 2017), and NLP2018 (Okayama, Mar. 2018).</p> <p>Through this research project, we were able to confirm that NPCMJ and JFN can be combined. Future plans include attempting to link JFN with a language resource which is more suitable for semantic-processing NLP applications and which is larger than NPCMJ.</p>
Notes	
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=2017000001-20170071

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

研究代表者	所属	理工学部	職名	教授	補助額	500（特B）千円
	氏名	小原 京子	氏名（英語）	Kyoko Ohara		
研究課題（日本語）						
深い意味処理のための統合日本語コーパスの開発と言語研究						
研究課題（英訳）						
Linguistic study and development of an integrated Japanese corpus for deep semantic processing						
1. 研究成果実績の概要						
<p>本研究の究極的な目的は、研究代表者が構築した認知言語学に基づく詳細な意味情報を付与した言語資源「日本語フレームネット」(JFN)を、言語研究のみならず自然言語処理にとってもこれまで以上に有益な言語資源とする方策を検討することである。そのためには、JFNの規模を大きくさせることと、JFNの持つ豊富な意味情報に加え統語情報をも含める、という2つの方向性が考えられる。本研究では特に、国立国語研究所で構築中のNINJAL Parsed Corpus of Modern Japanese (NPCMJ)とリンクさせることを検討した。NPCMJは文の統語情報のみで意味情報は含まないが規模は大きいのに対しJFNは詳細な意味情報を含むが規模が小さく、両者を統合することで深い意味処理を行う自然言語処理アプリケーションのための言語資源とすることができると考えたからである。</p> <p>当初計画したJFNとNPCMJのリンク方法は2つあった。一つ目は、JFNで意味情報付与の際の入力データとしてNPCMJの統語情報付きデータを用いることである。二つ目は、JFNとNPCMJを統合させ、認知科学的にも妥当な詳細な意味情報と文の統語情報の両方を含んだコーパスとすることである。前者に関しては、ツールを開発した。後者については実現までの目処がたった。</p> <p>本研究の経過・成果は、国際認知言語学会、マルチリンガルフレームネット会議、言語処理学会第24回年次大会にて発表した。本研究によりNPCMJとJFNの統合は実現可能であることが明らかとなった。今後はJFNを深い意味処理を行う自然言語処理アプリケーションにとってさらに有益な資源とするため、より規模の大きく自然言語処理分野で定評のある言語資源とのリンクを検討したい。</p>						
2. 研究成果実績の概要（英訳）						
<p>The principal investigator (PI) of this research project has been developing Japanese FrameNet (JFN), a language resource that contains rich semantic information which is based on the theory of Frame Semantics. The ultimate goal of this research project is to enable JFN to be useful for natural language processing (NLP) applications, especially for those that involve deep semantic processing. In order to achieve this goal, there are two directions. One is to make the size of JFN larger. The other is to add syntactic information to JFN, in addition to the existing semantic information. This research project pursued the latter approach and focused on ways to link the JFN data with the NINJAL (National Institute for Japanese Language and Linguistics) Parsed Corpus of Modern Japanese (NPCMJ). NPCMJ does not contain semantic information at all but includes detailed syntactic information of sentences and its size is larger than that of JFN. Therefore, it was expected that the combined language resource would be useful for various NLP applications.</p> <p>There are two ways to link the JFN data with that of NPCMJ. One is to use the NPCMJ data as input to the JFN annotation (tagging) process. The other is to integrate JFN and NPCMJ to build a combined corpus. As for the former, a tool was implemented to allow the NPCMJ data as input to the JFN annotation process. As for the latter, how to combine the two resources was investigated. The accomplishments of the project were presented at ICLC14 (Tartu, Estonia; July 2017), the Multilingual FrameNet meeting (Vancouver, Canada; Aug. 2017), and NLP2018 (Okayama, Mar. 2018).</p> <p>Through this research project, we were able to confirm that NPCMJ and JFN can be combined. Future plans include attempting to link JFN with a language resource which is more suitable for semantic-processing NLP applications and which is larger than NPCMJ.</p>						
3. 本研究課題に関する発表						
発表者氏名 (著者・講演者)	発表課題名 (著書名・演題)	発表学術誌名 (著書発行所・講演学会)	学術誌発行年月 (著書発行年月・講演年月)			
河原大輔, 小原京子, 関根聡, 乾健太郎.	クラウドソーシングによる日本語FrameNetと自動構築した格フレームとの対応付け	言語処理学会第24回年次大会予稿集. pp.706-709.	2018年3月			
小原京子, 大久保佳子	日英対訳絵本の語彙から見た日本語フレームネットの評価	言語処理学会第24回年次大会予稿集. pp. 1112-1114.	2018年3月			
小原京子	語彙意味論	『人工知能学大事典』(共立出版・人工知能学会編), pp. 635-637.	2017年7月			
Ohara, Kyoko Hirose	A Frame-based constructicon: A Case study in Japanese	The 14th International Cognitive Linguistics Conference (ICLC-14)	2017年7月			
Ohara, Kyoko Hirose	FrameNet vs. Constructicon: A Case study in Japanese FrameNet	Multilingual-FrameNet Meeting	2017年8月			