名詞間の意味的共起情報を用いた複合名詞の解析

小林　義行[†]　徳永　健伸[†]　田中　穂積[†]

複合名詞は名詞を結合することによって数限りなく生成できるので，全てを辞書に登録することは不可能である．したがって，辞書に登録されている名詞の組み合わせとして複合名詞を解析する手法が必要である．そのためには，複合名詞をそれを構成している名詞に分割し，名詞間の係り受け構造を同定しなくてはならない．これらの処理は統語的な手係りが少ないために難しく，何らかの意味的な情報が必要である．しかし，大規模な意味的情報を人手で構築し保守することはコストが大きいため，計算機によって自動的に知識を獲得することが望ましい。本論文では，コーパスから自動的に抽出した名詞間の意味的共起情報を用いて複合名詞の構造を解析する方法を提案する．この方法では，共起情報を統計的に処理して名詞間の意味的関係の強さを評価し，係り受け関係の曖昧性解消に利用する．まず，4 文字漢字語 16 万語から意味クラスの共起データを抽出した．抽出した共起データから統計的に名詞間の意味的関係の強さを計算する．そのための尺度として相互情報量を基にした評価尺度を提案する．この尺度と複合名詞の構造に関するヒューリスティクス，機械可読辞書から得られる言語知識を用いて複合名詞を解析する．評価のために新聞や用語集から抽出した漢字複合名詞を解析し，平均語長 5.5 文字の漢字複合名詞を約 78%の精度で解析できた．

キーワード:　統計的自然言語処理，複合名詞解析, 共起

Analysis of Japanese Compound Noun Using Collocational Information

KOBAYASI Yosiyuki [†] , TOKUNAGA Takenobu[†]
and TANAKA Hozumi[†]

Analyzing compound nouns is one of the crucial issues for natural language processing systems. Registering all compound nouns in a dictionary is an impractical approach, since we can create a new compound noun by combining nouns. Therefore, a mechanism to analyze the structure of a compound noun from the individual nouns is necessary. However, the analysis are difficult only when using syntactic knowledge. Therefore, we have to use semantic knowledge. It is hard to construct and maintain a large semantic knowledge_base, so we need a method to acquire semantic knowledge and use such the knowledge for the analysis. In this paper, we propose a method to analyze structures of Japanese compound nouns by using word collocational information and a thesaurus. The collocational information is acquired from a corpus of four *kanzi* character words. For each possible structure of a compound noun, the preference is calculated based on this collocational information. An experiment is conducted with 160,000 word collocations to analyze compound nouns of with an average length of 5.5 characters. The accuracy of this method is about 78%.

KeyWords:　*Statistical NLP, Compound noun analysis, Collocation*

† 東京工業大学 情報理工学研究科 計算工学専攻, Department of Computer Science, Tokyo Institute of Technology