構文付きコーパスの作成と類似用例検索システムへの応用

兵藤　安昭[†]　河田　実成[†]
応　江黔[†]　池田　尚志[†]

本論文では，まず，表層的情報のみを用いて安定的・高精度に構文解析を行う骨格構造解析の方法について述べる．次に，これを用いて行った 8 万用例文に対する構文付きコーパスの作成について触れ，骨格構造解析の有効性について述べる．さらに，この構文付きコーパスを対象として構築した類似用例検索システムについて述べる．本システムは，(1) 構文的制約 (係り受け構造) を指定して検索できるので，単語レベルの検索では検索されてしまうような多くの不適切な用例を絞り込むことができる (2) 分類語彙表を利用した意味コード化により，類似用例の検索も可能である (3) インデックスに構文情報を含めることにより，高速な構造検索を実現している (4) ユーザはウインドウ上で，三角表を用いた構文構造のブラウジング，検索パターンの指定などができ，使いやすいインタフェースを実現している，などの特徴がある．

キーワード:　構文付きコーパス，骨格構造解析, 類似用例検索，シソーラス

Building a Large Corpus with Skeltal Syntactic Structure and its Application to Similar Sentence Retrieval System

Yasuaki Hyodo [†] , Mithunari Kawada [†] , Jiangqian Ying [†]
and Takashi Ikeda [†]

In this paper we discuss firstly a method for skeltal syntactic analysis which uses only surface information. By this method we can successfully abstract the skeltal syntactic structure of a long Japanese sentence. Secondly we discuss its use on building a large corpus with skeltal syntactic structure consisting of about 80 thousand example sentences from two Japanese-English dictionaries. Thirdly we discuss our development of a similar sentence retrieval system over this corpus. The advantage of this system is that (1) Syntactic structure as input search pattern enables the system to dismiss unnecessary example sentences. Thus it raises the precision of the sentence retrieval. (2) Similar sentence retrieval realized on the basis of word encoding according to the thesaurus hierarchy. (3) Index table including syntactic information enables the system to make high speed retrieval. (4) The system has an excellent graphical user interface using the triangle representation of syntactic structure.

KeyWords: *tagged corpus, skeltal syntactic structure, similar sentence retrieval, thesaurus*

† 岐阜大学工学部 電子情報工学科, Department of Electronics and Computer Engineering, Faculty of Engineering, Gifu University