片方向の共起性による述語型定型表現の自動抽出

新納　浩幸[†]　井佐原　均[††]

本論文では「目を盗む」や「かたずを飲む」などの述語型定型表現をコーパスから自動抽出することを目的に，従来の相互情報量の条件を緩める方向で，名詞動詞間の共起性を測る新たな基準を提案する．概略，名詞，動詞のどちらかを固定して，その単語と共起する集合内の各単語に，どの程度特異な頻度になっているかの数値を与える．この数値は集合内のその単語の頻度の割合と，集合内の単語の種類数から計算される．この数値の上位のものを取り出すことで定型表現の抽出を行う．本手法の特徴は，名詞を固定した場合に抽出できる表現と，動詞を固定した場合に抽出できる表現はほとんど共通のものがなく，しかもどちらの場合も相互情報量による抽出程度の正解率を得られることである．このため，目的の抽出数の半数づつを名詞固定と動詞固定の各々の場合から取り出せば，相互情報量を用いて抽出する場合よりも高い正解率が得られる．

キーワード:　定型表現，慣用表現，自動抽出，相互情報量，共起性

Automatic Acquisition of Predicative Frozen Patterns on One Directional Association

Hiroyuki Shinnou [†] and Hitoshi Isahara[†]

This paper presents an alternative method to measure word association strength on predicative patterns in order to automatically extract predicative frozen patterns and idioms from a corpus. For this aim, mutual information is traditionally used. We improve the method on mutual information from a view of linguistics. The proposed method are realized by following steps. First, a verb (or noun) is fixed. Next, the set of nouns (or verbs) which associates the verb (or noun) is built up. Last, nouns (or verbs) which have peculiar frequency are chosen from this set. The peculiarity is confirmed from two characteristics, which are ratio of the word frequency for total frequency of the set, and the number of kind of word in the set. Predicative frozen patterns are constructed from chosen words and the fixed word. The advantage of this method is that patterns extracted by fixing a verb and patterns extracted by fixing a noun have few common patterns, and each extraction has equivalent ratio of correctness to extraction by mutual information. Therefore, extracting same number patterns, this method can get more correct patterns than mutual information.

KeyWords: *frozen pattern, idiom, automatic extraction, mutual information, collocation*

† 茨城大学工学部システム工学科, Department of Systems Engineering, Faculty of Engineering, Ibaraki University
†† (投稿時) 電子技術総合研究所知能情報部自然言語研究室, Natural Language Section, Machine Understanding Division, Electrotechnical Laboratory
(現在) 郵政省通信総合研究所関西先端研究センター知的機能研究室, Intelligent Processing Section, Kansai Advanced Research Center, Communications Research Laboratory, Ministry of Posts and Telecommunications