表層表現中の情報に基づく文章構造の自動抽出

黒橋　禎夫[†]　長尾　眞[†]

テキストや談話を理解するためには，まずその文章構造を理解する必要がある．文章構造に関する従来の多くの研究では，解析に用いられる知識の問題に重点がおかれていた．しかし，量的 / 質的に十分な計算機用の知識が作成されることはしばらくの間期待できない．本論文では，知識に基づく文理解という処理を行なわずに，表層表現中の種々の情報を用いることにより科学技術文の文章構造を自動的に推定する方法を示す．表層表現中の情報としては，種々の手がかり表現，同一 / 同義の語 / 句の出現，2 文間の類似性，の 3 つのものに着目した．実験の結果これらの情報を組み合わせて利用することにより科学技術文の文章構造のかなりの部分が自動的に推定可能であることがわかった．

キーワード: 文脈処理，文章構造，結束関係，表層表現

Automatic Detection of Discourse Structure by Checking Surface Information in Sentences

Sadao Kurohashi [†] and Makoto Nagao[†]

To understand a text or dialogue, one must track the discourse structure. While work on discourse structure has mainly focused on knowledge employed in the analysis, detailed knowledge with broad coverage availability to computers is unlikely to be constructed for the present. In this paper, we propose an automatic method for detecting discourse structure by a variety of keys existing in the surface information of sentences. We have considered three types of clue information: clue expressions, occurrence of identical/synonymous words/phrases, and similarity between two sentences. Experimental results have shown that, in the case of scientific and technical texts, considerable part of the discourse structure can be estimated by incorporating the three types of clue information, without performing sentence understanding processes by giving knowledge to computers.

KeyWords: *Discourse, Coherence relation, Surface information*

† 京都大学工学部 電気工学第二教室, Department of Electrical Engineering, Kyoto University