

知識に基づく構造的言語処理の確立と知識インフラの構築

黒橋 禎夫^{†1}

概要: テキストには、言明内の述語項構造および言明間の談話構造がある。しかし、現状の言語解析で高精度にとらえられるのは一文内の明示された述語項構造のみであり、日本語に類する省略された述語項関係や、談話構造の解析は困難である。一方で、近年、ウェブテキストからの基本的な知識の獲得が進展し、大規模な格フレーム、同義表現知識、事態間知識、事実型知識などが利用可能となりつつある。人間の文章の理解を考えれば、人間はその時点ですでに膨大な知識をもっており、その知識に基づいて文章の構造を理解し、そこから(わずかな)新たな知識を得る。このような枠組みを計算機の言語処理においても実現すること、すなわち、すでに獲得した知識を柔軟に利用することにより、文章の解析を高精度化し、これによってテキストの各論述から直接的に言明とその間の因果関係等の知識を抽出することが本研究の中心課題である。これに加えて、その基盤として、言明間あるいは言明と実世界の関係を規定するモダリティ、前提・含意、時制などの意味的關係を正確に捉える意味計算モデルを構築し、その応用として、テキストから直接抽出される知識をテキスト横断的に相互に関連付け、検索・比較・俯瞰を可能とする巨大な知識インフラを構築する。

^{†1} 京都大学