

TR-1-0186

A T R 対話データベースの内容
Contents of the ATR Dialogue Database

江原暉将 井ノ上直己 幸山秀雄
長谷川敏郎 庄山富美* 森元暉
Terumasa EHARA Naomi INOUE Hideo KOHYAMA
Toshirou HASEGAWA Fumi SHOHYAMA
Tsuyoshi MORIMOTO

1990.10.30

内容梗概

A T R では電話およびキーボードを用いて行われる対話を収集し、その結果を A T R 対話データベース (A D D) として蓄積している。本文では、この A T R 対話データベースに収録されている言語情報および、収集した対話の内容について述べる。A D D に収集されている対話は、模擬対話実験による対話およびラジオ放送で用いられた、実際の対話であり、データベース化に先だって、各種事前分析が施されている。これによって、付加される言語情報が豊富になり、抽出できる言語現象も多種に亘ることが出来る。特に、日本語と英語の対応データが付与されていることで、彼我の表現の相違などを抽出することが出来る。対話そのものの内容については、大まかな内容分類 (タスク) 毎にいくつかのトピックを定め、対話内容に偏りが出ないように注意している。また、各対話には、キーワードが付与されており、対話内容の概要がキーワードを見ることによって分かるようにしている。

ATR Interpreting Telephony Research Laboratories
A T R 自動翻訳電話研究所

* (株) 東洋情報システム