

Internal Use Only (非公開)

TR-SLT-0076

Acquiring Bilingual Named Entity Translations
from Content-aligned Corpora

Tadashi Kumano Hideki Kashioka
Hideki Tanaka Takahiro Fukusima

2005年3月24日

概要

We propose a new method for acquiring bilingual named entity (NE) translations from non-literal, content-aligned corpora. It first recognizes NEs in each of a bilingual document pair using the NE extraction technique, then finds NE groups whose members share the same referent, and finally corresponds between bilingual NE groups. The exhaustive detection of translation fragments NEs can potentially acquire translation pairs with broad coverage. The correspondences between bilingual NE groups are estimated based on the similarity of the appearance order in each document, and the corresponding performance came up to $F = 71.0\%$ by using small bilingual dictionary together.

(株) 国際電気通信基礎技術研究所
音声言語コミュニケーション研究所
〒619-0288 「けいはんな学研都市」 光台二丁目 2 番地 2 TEL : 0774-95-1301

Advanced Telecommunications Research Institute International
Spoken Language Translation Research Laboratories
2-2-2 Hikaridai "Keihanna Science City" 619-0288, Japan
Telephone: +81-774-95-1301
Fax : +81-774-95-1308

©2005 (株) 国際電気通信基礎技術研究所
©2005 Advanced Telecommunications Research Institute International