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.