

Acquisition of Word Translations Using Local Focus-based Learning in Ainu-Japanese Parallel Corpora

Hiroshi Echizen-ya¹ Kenji Araki² Yoshio Momouchi³ and Koji Tochinai⁴

¹ Dept. of Electronics and Information, Hokkai-Gakuen University, S26-Jo,
W11-Chome, Chuo-ku Sapporo, 064-0926 Japan
echi@eli.hokkai-s-u.ac.jp,

TEL: +81-11-841-1161 (ext. 7863), FAX: +81-11-551-2951

² Division of Electronics and Information, Hokkaido University, N13-Jo, W8-Chome,
Kita-ku Sapporo, 060-8628 Japan

araki@media.eng.hokudai.ac.jp,

TEL: +81-11-706-6534, FAX: +81-11-706-6534

³ Dept. of Electronics and Information, Hokkai-Gakuen University, S26-Jo,
W11-Chome, Chuo-ku Sapporo, 064-0926 Japan

momouchi@eli.hokkai-s-u.ac.jp,

TEL: +81-11-841-1161 (ext. 7864), FAX: +81-11-551-2951

⁴ Division of Business Administration, Hokkai-Gakuen University, 4-Chome,
Asahi-machi, Toyohira-ku Sapporo, 060-8790 Japan

tochinai@econ.hokkai-s-u.ac.jp,

TEL: +81-11-841-1161 (ext. 2753), FAX: +81-11-824-7729

Abstract. This paper describes a new learning method for acquisition of word translations from small parallel corpora. Our proposed method, **Local Focus-based Learning (LFL)**, efficiently acquires word translations and collocation templates by focusing on parts of sentences, not on entire sentences. Collocation templates have collocation information to acquire word translations from each sentence pair. This method is useful even when frequency of appearances of word translations is very low in sentence pairs. The LFL system described in this paper extracts Ainu-Japanese word translations from small Ainu-Japanese parallel corpora. The Ainu language is spoken by the Ainu ethnic group residing in northern Japan and Sakhalin. An evaluation experiment indicated that the recall was 57.4% and the precision was 72.0% to 546 kinds of nouns and verbs in 287 Ainu-Japanese sentence pairs even though the average frequency of appearances of the 546 kinds of nouns and verbs was 1.98.

1 Introduction

In recent years, many studies have addressed methods for building bilingual dictionaries from bilingual corpora [1–4]. Using bilingual corpora, such methods can obtain natural equivalents. However, these methods require large parallel corpora to acquire many word translations that are corresponding words of source language words and target language words because they cannot acquire many word