

vector. But the user's intended meaning is obviously better supported by a word like マグネット[magnetto] ("magnet"). Even worse, the lower the frequency of 磁石[jishaku] ("magnet"), the more weight will be given to 接着[setchaku] ("adhesion")'s context vector and 磁石[jishaku] ("magnet")'s context vector will be neglected. This inevitably leads to a decrease in translation accuracy.

Another reason why the context vector of one query term does in general not reliably express one meaning is that the query word can be ambiguous. An ambiguous word's context vector, which contains correlation information related to different senses, leads to correlation information which can be difficult to compare across languages.

The user might for example input the ambiguous word フード[fūdo] ("food" or "hood"). The resulting context vector will be noisy, since it contains the context information of both meanings, "food" and "hood", which will lead to lower translation accuracy. This problem is neither addressed by Non-Patent Document 1, nor Non-Patent Document 2.

The problem of a single term's unreliable context vector is addressed by the following invention.

DISCLOSURE OF INVENTION

An exemplary object Embodiments of the present invention is seek to provide a term translation acquisition method and a term translation acquisition apparatus that may solve the aforementioned problems.

An exemplary aspect of the present invention is a term translation acquisition apparatus which includes: a creation unit which creates a statistical model based on a set of input terms' context vectors, wherein the set of terms, including at least two terms, are in the same source language and describe the same concept; and a ranking unit which

uses the created statistical model to score terms in a target language that are considered as translation candidates for the concept.

Another exemplary aspect of the present invention is a term translation acquisition method which includes: creating a statistical model based on a set of input
5 terms' context vectors, wherein the set of terms, including at least two terms, are in the same source language and describe the same concept; and using the created statistical model to score terms in a target language that are considered as translation candidates for the concept.

Yet another exemplary aspect of the present invention is a computer-readable
10 recording medium storing a program that causes a computer to execute: a creation function of creating a statistical model based on a set of input terms' context vectors, wherein the set of terms, including at least two terms, are in the same source language and describe the same concept; and a ranking function of using the created statistical model to score terms in a target language that are considered as translation candidates for
15 the concept.

According to embodiments of the present invention, the problem of sparse context vectors related to low-frequent terms, as well as the problem of noisy context vectors related to ambiguity of input terms can be mitigated. As a consequence,
translation accuracy ~~is~~ may be improved.

20

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a block diagram showing the functional structure of the term translation system related to Non-Patent Document 1.

Figure 2 is a block diagram showing the functional structure of the term
25 translation system related to Non-Patent Document 2.