Automatic Detection of Translation Errors: The State of the Art

Graham Russell and **Ngoc Tran Nguyen**
IIT–ILT, National Research Council Canada
RALI-DIRO, Université de Montréal[*]
{russell,nguyentt}@iro.umontreal.ca

George Foster
IIT–ILT, National Research Council Canada[†]
george.foster@nrc-cnrc.gc.ca

1 Background

The demonstration presents TransCheck, a translation quality-assurance tool developed jointly by the RALI group at the University of Montreal and the Interactive Language Technologies section of the Canadian National Research Council's Institute for Information Technology.

The system differs from other similar tools in the range of error-types targeted, and the underlying mechanisms employed. The demonstration illustrates the operation of the system and gives the rationale for its design and capabilities. The version demonstrated accepts input in English and French.

2 System Overview

A modular architecture promotes flexibility (ease of adaptation to new domains, client requirements and language pairs) and extensibility (incorporation of new error-detector components as they become available).

In a transparent preprocessing stage, source and target texts are read and aligned. The resulting stream of alignment regions is passed to a set of independent error-detection modules, each of which records errors in a global table for subsequent report generation. Certain of the error-detection components make use of external data in the form of lexical and other language resources.

3 Translation Errors

The difficulty of general-case translation error detection is discussed. Several classes of feasible errors are identified, and the technological capabilities required for their successful detection described.

Detection of incorrect terminology usage, for example, requires the ability to recognize correspondences between source and target language expressions, and to generalize over different realizations of a given term; inflection, coordination and anaphora combine to render inadequate solutions based solely on simple static lists of term pairs. 'Negative terminology', covering false friends, deceptive cognates, Anglicisms, etc., is rather more challenging, and can benefit from a more precise notion of translational correspondence. Proper names pose a range of problems, including referential disambiguation and varying conventions regarding transliteration, while a broad class of paralinguistic phenomena (numbers, dates, product codes, etc.) raise yet others in the area of monolingual analysis and translational equivalence. Omissions and insertions constitute a final error class; these present particular difficulties of recognition and interpretation, and are best addressed heuristically.

The current TransCheck system targets the error types mentioned above. Each is exemplified and discussed, together with the elements of language technology which permit their detection: dictionaries, shallow parsing, alignment, translation models, etc.

Experience gained in preliminary user trials is briefly reported and a variety of usage scenarios considered. Finally, some comparisons are made with other translation tools, including other proposals for translation error detection.

---

[*]C.P. 6128, succ. Centre-ville, Montréal QC, Canada H3C 3J7
[†]University of Quebec en Outaouais, Lucien Brault Pavilion, 101 St-Jean-Bosco Street, Gatineau QC, Canada K1A 0R6