A Hybrid Approach of Noun Phrase Translation in Cross-Language Information Retrieval

Thanh C. Nguyen, Hieu V. Nguyen, and Tuoi T. Phan

Computer Science Faculty, HCMC University of Technology,
268 Ly Thuong Kiet, HCMC, Vietnam
thanh@cse.hcmut.edu.vn, nguyenvanhieu_danang@yahoo.com,
tuoi@hcmut.edu.vn

Abstract. At present, many researches of noun phrase translation are proposed in Natural Language Processing field, but most of them are in dictionary-based with word-by-word translation and similarity selection. The paper proposes a hybrid approach for noun phrase translation, by combining the set theory and grammar's pattern, and its algorithms, to apply to Vietnamese–English translation. The finding also has good experimental results when applying on Vietnamese noun phrases.

Keywords: Set theory, grammar's pattern, hybrid approach.

1 Introduction

The noun phrase translation from L_s to L_t language often hits difficulties because of its depending on complexity of language's grammar, then it impacts on result quality. There have been published papers addressing approaches of noun phrase translation. The grammar rule-based approach (a), is applied in machine translation in long-term. J. Gao [1] published researches with better result for pair of Chinese-English. With set theory-based approach (b), F.L. Ostenero [2] proposed their solution in 2004, but in which they did not mention their disambiguation enough. The difficulties, which impact on effect of translation process, are many source noun phrases cannot be translated because of 'out-of-dictionary' issue, complexity of language's grammar in grammar rule-based, and number of calculation in set theory-based.

2 Proposed Hybrid Approach of Noun Phrase Translation

The paper proposes a hybrid approach, which applies solutions of (a) and (b), to translate a noun phrase and resolve ambiguity of its result (in below flow chart).
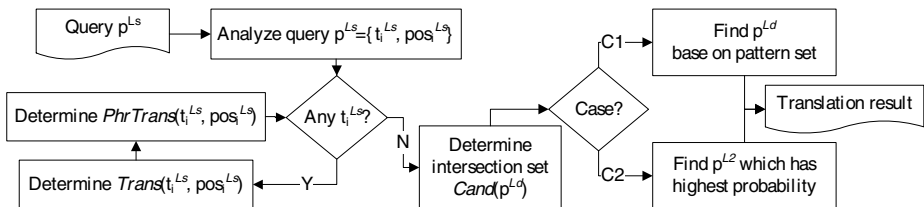
In initial iteration, result of our experiment, which includes bilingual dictionary from [7] and data from [3][4][5][6], achieves 72% of precision in Vietnamese-English noun phrase translation. Detailed result is shown as follows:

Group 1 (2 words per phrase): 171 phrases, 86% of precision
Group 2 (3 words per phrase): 126 phrases, 64% of precision
Group 3 (4 words per phrase): 53 phrases, 43% of precision

3   Conclusion and Future Work

In the paper, we propose a hybrid approach of noun phrase translation, which combine both methodologies of set theory-based and grammar rule-based, and its algorithms. Especially, from source noun phrase, these algorithms focus on two cases of searching its existing correspondent translated noun phrases destination noun phrases or analyzing bi-gram probabilities to select best results. In initial iteration, we apply the approach for Vietnamese-English translation and retrieve many propitious results. Our next step is to focus on reducing the calculation time of intersectional set also system resource to improve the speed and processing-time of these algorithms, which, we believe, will make significant contributions to NLP community.

References

1. Gao, J., Nie, J.-Y., Xun, E., Zhang, J., Zhou, M.: Changning Huang: Improving Query Translation for Cross - Language Information Retrieval using Statistical Model. In: ACM (2001)
2. López - Ostenero, F., Gonzalo, J., Verdejo, F.: Noun phrases as building blocks for cross - language Search Assistance. Information Processing and Management: an International Journal archive 41(3), 549–568 (2005)
3. Vietnamese Ministry of Trade, http://www.mot.gov.vn
4. Text REtrieval Conference (TREC), http://trec.nist.gov
5. PC World, http://www.pcworld.com
6. GATE, A General Architecture for Text Engineering, http://gate.ac.uk
7. HO NGOC DUC, Open-source dictionary, http://www.informatik.uni-leipzig.de/duc/Dict/