

Natural Languages Translation Using an Intermediate Language

Keerthi Kamal Adusumilli, *Member, IAENG*

Abstract— English is the official language of only seven countries comprising little more than five percent of the world's population. Yet the majority of online content is in English. While many business people speak English as a second language, most prefer to communicate in their native tongue. Using machine translation one could easily translate the content in one language to the other largely in real time. This paper describes the role of an Intermediate language in machine translation of natural languages, how modular the work can be accomplished in translating a one language to other.

Index Terms— intermediate language, knowledge representation, natural language processing, modularity, translation.

I. INTRODUCTION

Natural languages translation in a traditional way requires a one-one translation of both languages which is a tedious process, requiring translators who are proficient in both the languages.

With ever increasing needs a language needs to be translated into many other languages and vice-versa. Therefore, a translator package need to be developed for pair of languages for both to & fro of translation, but building a package takes at least 2-3 yrs of time with lots of effort & money.

Developing Machine translation packages in this traditional way is cumbersome. In order to reduce the effort the concept of intermediate language has been put forward, being reusable not only reduces the effort and time but also the need for finding a translator proficient in a given pair of language.

II. TRANSLATION USING TRADITIONAL METHOD

In this method if there are N languages to translate say L1, L2, L3, L4.....LN. A language need to be translated into N-1 other languages, as we are given N languages, so totally we have a job of translating $N*(N-1)$ languages, obtained by counting permutations which implies that we have to make $N*(N-1)$ translation packages. With the increase in N the effort required greatly increases.

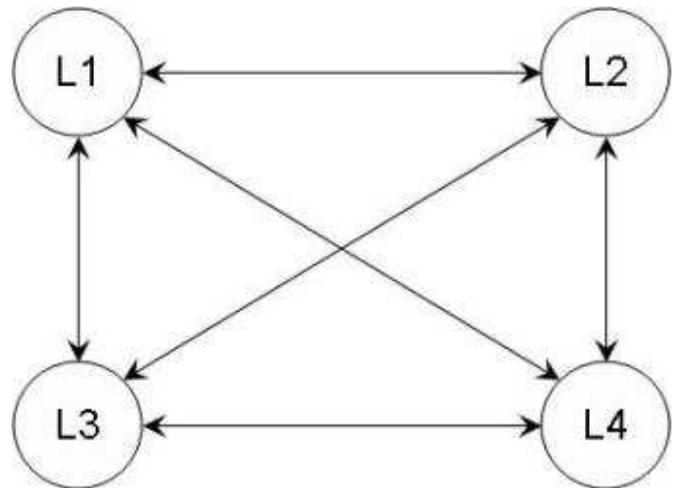


Figure 1: A bi-directional arrow represents a translation of given language into target language and vice-versa. This graph addresses that 12 translation packages are required for translating 4 languages into one another which is $N*(N-1)$, where N is the number of given languages.

III. TRANSLATION USING A SINGLE INTERMEDIATE LANGUAGE

In this method every language is translated into an intermediate language(X) and vice-versa where

Manuscript received February 17, 2006.

Keerthi Kamal Adusumilli is a 3rd year Integrated Post Graduate Student at ABV-Indian Institute of Information Technology & Management, Gwalior, India 474010(phone: +91-9893317446; fax: +91-751-2460313; e-mail:akkkamal@gmail.com).

X is a widely used natural language. Like this by counting the number of translations required we have N translations of languages L1, L2, L3, L4.....LN into Intermediate language(X) and X into given languages as N, totally we have a job of 2*N translations.

Suppose if we want to make a translation package for the language pairs like L1 & L5 by using this approach we can make use of the translation packages of L1-X and L5-X.

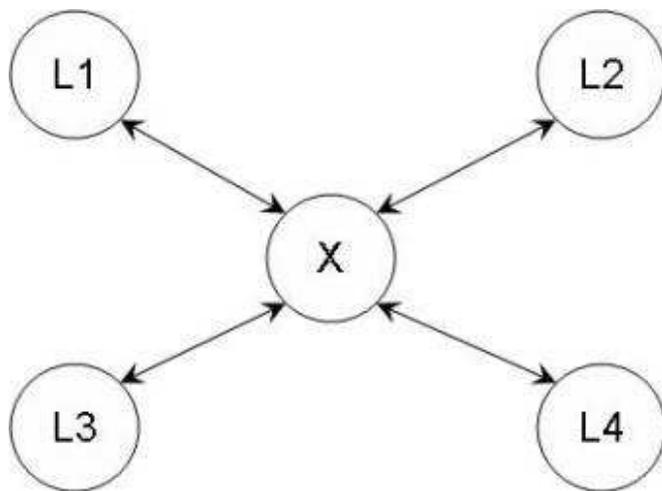


Figure 2: A bi-directional arrow represents a translation of given language into intermediate language and vice-versa. This graph addresses that 8 translation packages are required to be used for translating 4 languages into one another which is $2*N$, where N is the number of given languages.

Suppose there is a need for translating content in language L1 to L5, but a direct translation package does not exist currently but packages like L1-L3 and L3-L5 translation packages exist so making use of these packages the required translation can be done.

IV. TRANSLATION USING TWO INTERMEDIATE LANGUAGES

Instead of constraining to one intermediate language we can have two intermediate languages. Suppose X1 & X2 are the two most widely used languages then some languages can have X1 as their intermediate language and some other can have X2

as their intermediate language depending on the ease and compatibility. There might be translation packages already available for one of the intermediate language and the language of the other group. Some least used languages can have the advantage by using the double translation instead of no solution.

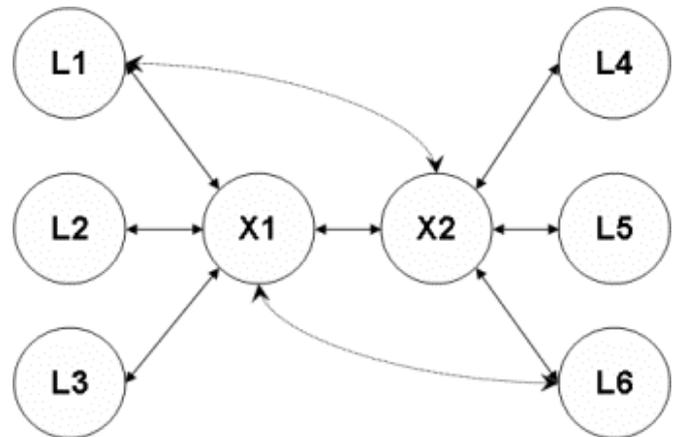


Figure 3: This Figure depicts that L1, L2& L3 are having X1 as their intermediate language and L4, L5& L6 are having X2 as their intermediate language and solid lines represent the translation between two intermediate languages and also translation between the languages with their corresponding intermediate language, while Dotted lines represent the direct translation between one intermediate language with the language of the other group. Suppose L1 want to be translated into Languages of the other group so the paths like L1-X2-L4, L1-X2-L5 and L1-X2-L6 or L1-X1-L6 are the solutions, suppose L2 & L5 are the least used languages then we have to use the path of L2-X1-X2-L5, unavoidably double translation instead of developing a pair specific package for least used languages.

V. ADVANTAGES

Using an intermediate language approach we can cope up with the ever increasing needs of translating one language to many other languages. The work of developing a package can be broken down into two prime modules, developing a package for the pair of given language and the intermediate language and another package for the pair of intermediate language and the target language, it also eliminates

the need for finding a translator proficient in a given pair of languages.

If a package has been developed for translation of one language into the intermediate language, by reusing the package as a module assisting the translation of given language into a target language we can cover many other languages with ease when compared to the development of one-one translation packages.

Some least used languages also had a solution using the intermediate language instead of no solution.

VI. CONCLUSION

Machine translation may not perfectly translate a given language as there are breaches in language syntaxes but do allow readers to understand the gist of a document.

In spite of having several advantages like less effort, modularity in using this intermediate language there is are little drawback; loss of information as a result of two translations taking place in the path of given language to target language and the two translation need to occur sequentially which is time consuming.

Still advantages outweigh the drawbacks.

ACKNOWLEDGMENT

Keerthi kamal Adusumilli thanks his professor G.K.Sharma and his mentor Prashant Singh.

REFERENCES

- [1] Napier, M. "The Soldiers are in the Coffee - An Introduction to Machine Translation", Cultivate Interactive, issue 2, 16 October 2000
URL: <<http://www.cultivate-int.org/issue2/mt/>>.
- [2] Donald Knuth. The Art of Computer Programming, Volume 3: Sorting and Searching, Third Edition. Addison-Wesley, 1997. ISBN 0-201-89685-0. Section 5.1: Combinatorial Properties of Permutations, pp.11-72

- [3] Kenneth Goodman, Sergei Nirenburg, the KBMT Project, First Edition: "A Case Study in Knowledge-Based Machine". Morgan Kaufmann 1 July 1991 ISBN: 1558601295.
- [4] Michael Lerner " Make sales (or just friends) with a Web site that speaks the visitor's language", Building worldwide Web sites 1 September 1999, URL: <<http://www-128.ibm.com/developerworks/library/web-localization.html>>
- [5] Dan Maxwell ,Multilingual Machine Translation, Esperantic Studies, Number 3 Summer 1992
URL: <<http://esperantic.org/esf/es3.htm>>
- [6] Miklos Bona, Bona Bona,"Combinatorics of Permutations",CRC Press,1 Jun 2004,ISBN1584884347