



**THE RULE-BASED MACHINE TRANSLATION SYSTEM  
FROM PALI TO THAI**

**NATTHAWAN PHONSON**

ฉบับนี้หนาวาร  
จาก  
บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF SCIENCE  
(COMPUTER SCIENCE)  
FACULTY OF GRADUATE STUDIES  
MAHIDOL UNIVERSITY**

**2001**

**ISBN 974-04-0624-6**

**COPYRIGHT OF MAHIDOL UNIVERSITY**

TH  
N282 r  
2001  
C.2

4137844 SCCS/M : MAJOR : COMPUTER SCIENCE ; M.Sc.(COMPUTER SCIENCE)

KEY WORDS : RULE-BASED / MACHINE TRANSLATION / PALI LANGUAGE / THAI LANGUAGE

NATTHAWAN PHONSON : THE RULE-BASED MACHINE TRANSLATION SYSTEM FROM PALI TO THAI. THESIS ADVISORS: SUPACHAI TANGWONGSAN, Ph.D., SUKANYA PHONGSUPHAP, Ph.D., 211 p. ISBN 974-04-0624-6

The objective of this research is to explore the viability of a Rule-Based Machine Translation system from Pali to Thai. A Rule-Based Machine Translation system from Pali to Thai (RB-MT) RB-MT is a process of translating a source language into a target language by the computer, based on grammatical rules established for this particular purpose. The knowledge source consists of a set of rules and a dictionary of Pali language to Thai while the machine translation (MT) is an application of theories on natural language processing which deals with an analysis of the language structure comprising of words built up into sentences. This research continuous from a previous study and expands the previous version, which dealt only with simple sentences. The newly developed system is expected to translate both simple and complex Pali (subordinate clause and a main clause) sentences into Thai.

In the process of translating the computer will use 3 sets of data: an input text (sentence); the translation program; and, the permanent resident knowledge source. The most significant part of the knowledge source is the dictionary. There will be ongoing matching between a new input Pali word and the stored Pali word in the dictionary. The knowledge source is also composed of a set of rules used in the processing in each step of the translation. For the development of RB-MT, the system begins with a search for the word data in the Pali-Thai dictionary by using rules of declension of words. Next, it performs type checks, then it analyses the Pali sentence structure. Subsequently, it performs transformation of the Pali sentence structure into Thai structure through established rule. Finally, it searches for the meaning of words from the Pali-Thai dictionary, including an insertion of correct conjunctions and prepositions. The outcome will then be displayed on the screen.

The software tool used in the development is Delphi II, and Paradox VII as a database management system. To find the results in the experiment, both simple and complex Pali sentences were put to test with the newly developed system. The experimental results consist of 52 examples which can be divided into 2 groups ; 20 examples of simple Pali sentences and 32 examples of complex Pali sentences. It was found that the results yielded quite satisfactorily results when compared to those of BUDSIR, and much better to those from the machine translation system in the previous version. However, future research an ongoing development is recommended to further improve this system.

4137844 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์ ; วท.ม. (วิทยาการคอมพิวเตอร์)

ณรรจวรรณ พูลสน :ระบบการแปลภาษาบาลีเป็นภาษาไทยด้วยคอมพิวเตอร์โดยใช้กฎไวยากรณ์ (THE RULE-BASED MACHINE TRANSLATION SYSTEM FROM PALI TO THAI).  
คณะกรรมการควบคุมวิทยานิพนธ์ : ศุภชัย ตั้งวงศ์สานต์, Ph.D., สุกัญญา พงษ์สุภาพ, Ph.D. 211 หน้า.  
ISBN 974-04-0624-6

งานวิจัยนี้ได้นำเสนอระบบการแปลภาษาบาลีเป็นภาษาไทยด้วยคอมพิวเตอร์โดยใช้กฎไวยากรณ์ (THE RULE-BASED MACHINE TRANSLATION SYSTEM FROM PALI TO THAI : RB-MT) RB-MT เป็นกระบวนการในการแปลภาษาต้นฉบับไปเป็นภาษาเป้าหมายด้วยคอมพิวเตอร์ซึ่งอาศัยกฎไวยากรณ์ที่กำหนดขึ้น โดยมีแหล่งความรู้ คือ ชุดของกฎไวยากรณ์และพจนานุกรมภาษาบาลีเป็นภาษาไทย ในขณะที่การแปลภาษาด้วยเครื่องคอมพิวเตอร์ (Machine Translation : MT) เป็นการประยุกต์ใช้ทฤษฎีด้านการประมวลภาษาธรรมชาติ (Natural Language Processing : NLP) มาใช้วิเคราะห์โครงสร้างของภาษาที่ประกอบขึ้นเป็นประโยค งานวิจัยนี้เป็นการขยายงานต่อจากงานวิจัยเดิมที่สามารถแปลภาษาบาลีได้เฉพาะประโยคความเดียวในขณะที่ระบบใหม่สามารถแปลภาษาบาลีได้ทั้งประโยคความเดียวและประโยคความซ้อน ซึ่งประโยคความซ้อนของภาษาบาลีประกอบด้วยส่วนย่อยของประโยคและส่วนหลักของประโยค

ในกระบวนการทำงานของระบบการแปลภาษานี้คอมพิวเตอร์จะใช้ข้อมูล 3 ชุดด้วยกัน คือ ประโยคที่รับเข้าไปในระบบ , โปรแกรมการแปล และแหล่งความรู้ที่ถูกสร้างขึ้น ส่วนของแหล่งความรู้ที่สำคัญที่สุดคือพจนานุกรม โดยจะมีการจับคู่ระหว่างคำบาลีที่ใส่เข้ามากับคำบาลีที่อยู่ในพจนานุกรม โดยแหล่งความรู้ยังประกอบด้วยชุดของกฎเกณฑ์ที่ใช้ในการทำงานในแต่ละขั้นตอนของกระบวนการแปล ในการพัฒนา RB-MT มีกระบวนการทำงานเริ่มจากการค้นหาข้อมูลของคำศัพท์จากพจนานุกรมบาลี-ไทย (BUDSIR and Pali-Thai Dictionary) และใช้กฎเกณฑ์การแจกวิภัติของคำศัพท์ ต่อมาจึงมีการตรวจสอบประเภทของประโยคที่เข้ามาโดยใช้กฎเกณฑ์การหาประเภทของประโยค ในลำดับถัดไปคือการวิเคราะห์โครงสร้างของประโยคภาษาบาลีด้วยกฎเกณฑ์การหาหน้าที่ของคำและ Context-free grammars หลังจากนั้นจึงทำการแปลงจากโครงสร้างประโยคภาษาบาลีให้อยู่ในรูปของโครงสร้างประโยคภาษาไทยด้วยกฎเกณฑ์ที่กำหนดขึ้น ขั้นตอนสุดท้ายคือการหาความหมายของคำศัพท์จากพจนานุกรมบาลี-ไทยและใส่คำเชื่อมหรือคำบุพบทที่ถูกต้องซึ่งผลลัพธ์ที่ได้จะถูกแสดงออกมาทางหน้าจอ

การพัฒนาโปรแกรมระบบนี้ได้ใช้โปรแกรม Delphi II สำหรับการพัฒนา และใช้ Paradox VII เป็นฐานข้อมูล ผลการทดลองได้นำประโยคความเดียวและประโยคความซ้อนเข้าทดสอบกับระบบจำนวน 52 ตัวอย่าง สามารถแบ่งได้เป็น 2 กลุ่มคือ ประโยคความเดียวจำนวน 20 ตัวอย่างและประโยคความซ้อนจำนวน 32 ตัวอย่าง ผลการทดลองเป็นที่น่าพอใจเมื่อเปรียบเทียบกับพระไตรปิฎกฉบับคอมพิวเตอร์ (BUDSIR) และเปรียบเทียบกับงานวิจัยเดิม