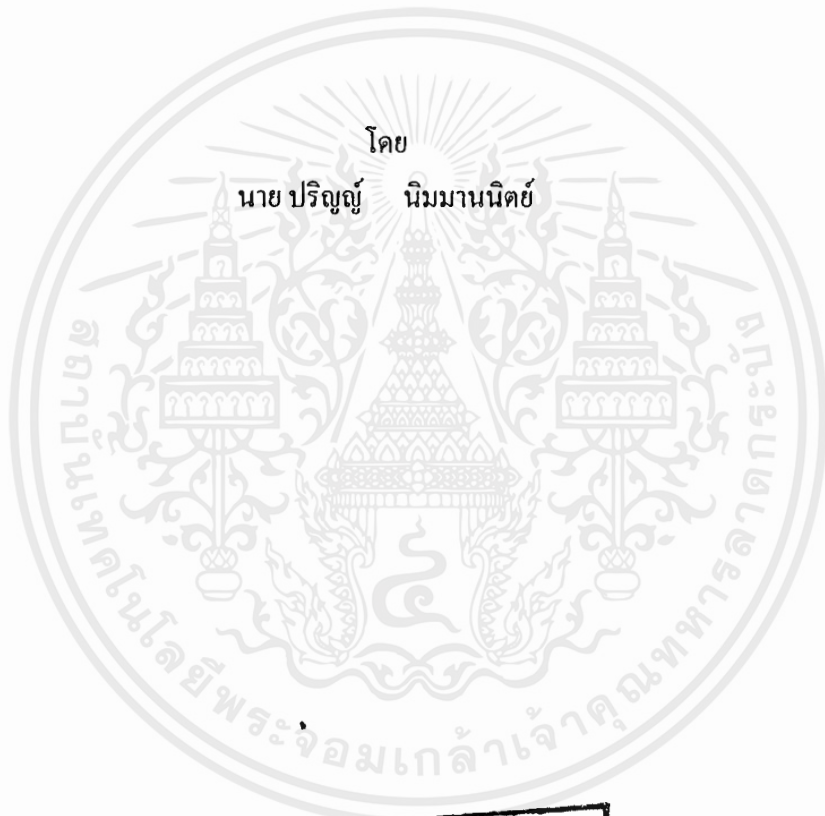




การแปลภาษาโดยวิธีการเทียบเคียงฐานข้อมูลชนิดเทียบเคียงตัวอย่างจากภาษาอังกฤษเป็นภาษาไทย
(Language Translation Using Corpus-Base Example-Base from English to Thai Language)



วัน เดือน ปี..... 15 ค.ค. 2541
เลขทะเบียน..... 038992
เลขเรียกหนังสือ..... T.40233 ฝ.1157ก.

ปริญญาานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิศวกรรมคอมพิวเตอร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2540

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์อื่นใด
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

038992

การแปลภาษาโดยวิธีการเทียบเคียงฐานข้อมูลชนิดเทียบเคียงตัวอย่างจากภาษาอังกฤษเป็นภาษาไทย
(Language Translation Using Corpus-Base Example-Base from English to Thai Language)



ปริชญ์นิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิศวกรรมคอมพิวเตอร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2540

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2540

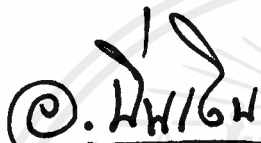
ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การแปลภาษาโดยวิธีการเทียบเคียงฐานข้อมูลชนิดเทียบเคียงตัวอย่างจากภาษาอังกฤษ
เป็นภาษาไทย

ผู้จัดทำ

1. นาย ปริญญ์ นิมมานนิตย์


.....อาจารย์ที่ปรึกษา
(.....me ได้ดน..... ปนไน.....)



การแปลภาษาโดยวิธีการเทียบเคียงฐานข้อมูลชนิดเทียบเคียงตัวอย่างจากภาษาอังกฤษเป็นภาษาไทย

นายปริญญา นิมนานนิตย์

อ. เอื้อน ปิ่นเงิน
ปีการศึกษา 2540

บทคัดย่อ

ในโครงการนี้จะทำการศึกษาเกี่ยวกับการแปลภาษาซึ่งในการแปลภาษานั้นมีด้วยกันหลายวิธี แต่ในโครงการนี้จะทำการศึกษาเน้นไปทางด้าน Corpus-Based statistic-Oriented Transfer and Generation Model for Machine Translation, Corpus-based จะเป็นวิธีที่ทำเพื่อหากฎในการแปลและเลือกหาวิธีที่จะใช้ในการแปลระหว่างคู่ภาษาที่ต้องการ Transfer Score (คือการกำหนดเพื่อใช้ในการวัดหาวิธีที่น่าจะเหมาะสมของวิธีการแมปिंगที่แตกต่างกันระหว่างประโยคต้นแบบ (Source Sentence) กับประโยคผลลัพธ์) และ Generation Score (คือการกำหนดเพื่อที่จะจัดหาวิธีในการหาความน่าจะเป็นสำหรับหาแบบแผนที่ดีที่สุด) อัลกอริทึม (Algorithm) นี้มีจุดประสงค์เพื่อที่จะหา Transfer Units ที่เหมาะสมใน Syntax Tree และกฎในการแปล โดยการประยุกต์อย่างเช่น Algorithm, กฎการสร้าง (Generation Rules) ต้องการสามารถหาได้โดยตรงจากประโยคผลลัพธ์ (Target Language) ดังนั้นในการสร้าง Target Sentences สามารถที่จะปรับไปใช้ตามไวยากรณ์ (Grammar) และรูปแบบของประโยคผลลัพธ์แทนที่จะสร้างโดยการยึดติดกับการวิเคราะห์ไวยากรณ์ของภาษาต้นแบบ (Source Language)

Language Translation Using Corpus-Based Example-Based from English to Thai Language

Mr. Prin Nimmannitya

Mr. Aurn PinNgurn

1997

Abstract

In this project, we are studying about how to translate a language. In this reason, there are many ways to translate a language but in this project, we are studying only Corpus-based statistical-oriented transfer and generation model for machin translation. Corpus-based is a method that we used to find translation rules and choosing the way to use with language pair that would like transfer score (a transfer score is used to measure an appropriate method of mapping transfer that different between source sentence and target sentence) and generation score (a generation score is used to find an appropriate probability for good plan). The objective is to find transfer unit that good for syntax tree and rule-based by applying, for instance, algorithm generation rules can find directory from target language. So when we generate target sentence, we can tune to use with grammar and model of target sentence instead of generated by stuck with analyze grammar of source language.

กิตติกรรมประกาศ

IV

ขอขอบคุณ อาจารย์เอื้อน ปิ่นเงิน ที่ให้โอกาสผมได้ทำโครงการวิจัยนี้ โดยได้รับความคิดริเริ่มจากวิชาของอาจารย์

ขอขอบคุณ ดร. สุรพันธ์ เมฆนาวิน ที่ช่วยสอน และให้คำแนะนำ รวมทั้งหนังสือและเอกสารที่ใช้ทำการศึกษา

ขอขอบคุณ อาจารย์ อภิเนตร อุนากุล ที่ช่วยแนะนำเกี่ยวกับโครงการวิจัยนี้ ทั้งขอบเขตของโครงการและผู้เชี่ยวชาญ

ขอขอบคุณ นายศิริพงศ์ วิเศษมงคลชัย (คุณต้น) และนายสิริโรจน์ รัตนวัฒน์กุล (คุณโรจน์) ที่ช่วยให้กำลังใจและช่วยพิมพ์ ขอบคุณมาก ๆ นะ

ขอขอบคุณ นายเริงชัย ดวงแก้ว (คุณอ๊ว) ผู้มากวนใจและส่งข้าว ส่งน้ำ

ขอขอบคุณ น.ส. ขวัญ โปษะกฤษณะ ผู้เอื้อเฟื้อ (บริจาค) หนังสือกาฬิเวอร์

ขอขอบคุณ เพื่อน ๆ อีกหลายคน ที่มากอยเป็นกำลังใจ share resource ต่าง ๆ ภายในห้องพัก

ขอขอบคุณ น้ำแมว น้ำเบญ น้ำศิริชัย น้ำวิชัย ที่คอยให้กำลังใจและปรึกษา

และขอขอบคุณ น.ส.ปิติพีร์ รวมเมฆ (คุณเอม) สำหรับกำลังใจที่ดีเวลาเมื่อรู้สึกท้อแท้ และหมดหวัง (ขอขอบคุณจากใจ)

สุดท้ายนี้ ขอกราบขอบพระคุณ คุณพ่อ (รศ.นพ. เจษฎา นิมมานิตย์) คุณแม่ (รศ.พรทิพย์ นิมมานิตย์) ผู้ให้กำเนิด น้องหวาน และน้องปรีดี คอยให้ความรัก และกำลังใจ

ผู้จัดทำ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญภาพ.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและที่มา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตของงานวิจัย.....	2
1.4 วิธีการดำเนิน.....	2
บทที่ 2 หลักการและทฤษฎี.....	4
2.1 หลักการทั่วไปของรูปแบบการแปลง (Transfer Model).....	4
2.1.1 การแปลโดยอ้างอิงกฎ (Rule Base).....	4
2.1.2 การแปลโดยเทียบเคียงฐานข้อมูล (Example Base).....	5
2.1.3 การแปลโดยใช้ข้อมูลทางสถิติ (Purely Statistical).....	5
2.1.4 การแปลแบบรวมกัน (A Cooperative).....	6
2.2 Translation Model and Translation Score.....	7
2.2.1 Transfer Score.....	8
2.2.2 Generation Score.....	10
2.3 Learning Transfer Unit and Operations.....	11
2.4 Learning Target Generation Rules.....	14
2.5 Experiment.....	15
2.6 Conclusions.....	16
บทที่ 3 การเขียนและการทำงานของโปรแกรม.....	21
3.1 ขั้นตอนการทำงานของซอฟต์แวร์.....	21
3.2 ตัวอย่างการทำงานของซอฟต์แวร์.....	23
3.3 ข้อดีและข้อเสียของการแปลด้วยวิธีนี้.....	25
3.4 ปัญหาและแนวทางแก้ไข.....	25
บทที่ 4 บทวิจารณ์และสรุป.....	28

ภาคผนวก ก เนื้อเรื่องและลักษณะของคำในเรื่อง กาลีเวอร์.....	29
บรรณานุกรม.....	45



สารบัญตาราง

ตารางที่		หน้า
2-1	ตารางการแยกคำสั่งในการแปล.....	17
2-2	ตารางตัวอย่างคำสั่งในแต่ละคำสั่งของการแปลภาษาอังกฤษเป็นภาษาจีน.....	17



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

รูปที่		หน้า
2 - 1	Concept Model สำหรับ Transfer และ Generation Process.....	18
2 - 2	Normal Form สำหรับประโยค “He put the book on the desk”.....	18
2 - 3	Syntax tree with attached word index list.....	19
2 - 4	An example of English Transfer process.....	20
3 - 1	ขั้นตอนการทำงานของซอฟต์แวร์ที่ใช้วิธีเทียบเคียงฐานข้อมูลชนิดเทียบเคียงตัวอย่าง.....	22



กันบ้างแล้ว ในโครงการวิจัยนี้ได้สังเกตเห็นความสำคัญของการแปลภาษา จึงได้ทำการศึกษาการแปลในรูปแบบวิธีรวบรวมฐานข้อมูลชนิดเทียบเคียงตัวอย่างข้อมูล (Example Base) เนื่องจากวิธีนี้เป็นวิธีที่มีโอกาสที่จะแปลภาษาได้สมบูรณ์มากที่สุด และคนส่วนใหญ่ไม่ค่อยมีการศึกษาวิธีนี้ เนื่องจากสมัยก่อนทรัพยากร (Resource) ต่าง ๆ ค่อนข้างแพง และวิธีนี้จะใช้ทรัพยากรเปลืองมากที่มีการรวบรวมข้อมูลที่ผ่านมามีการรวบรวมข้อมูลประมาณ 2 - 3 GB ซึ่งเป็นจำนวนที่เกินความสามารถที่จะทำได้ในสมัยก่อน แต่เนื่องจากในปัจจุบันทรัพยากรที่ใช้ยังมีราคาถูกลง และความเร็วของคอมพิวเตอร์ได้เพิ่มขึ้นอย่างมาก วิธีนี้จึงเป็นวิธีที่น่าสนใจและน่าจะเหมาะสมที่สุดสำหรับการแปลภาษา และในโครงการวิจัยนี้ได้มีการสร้างตัวอย่างซอฟต์แวร์ในการแปลภาษาไว้ เพื่อเป็นตัวอย่างเครื่องแปลง โดยจะใช้วิธีแปลแบบรวบรวมฐานข้อมูลชนิดเทียบเคียงตัวอย่าง ซึ่งจะมีการกล่าวต่อไป

1.2 วัตถุประสงค์ของงานวิจัย

- 1.2.1 ทำการศึกษาวิธีการแปลภาษา การแปลภาษาโดยเน้นการแปล โดยใช้วิธีรวบรวมฐานข้อมูลชนิดเทียบเคียงตัวอย่างเป็นหลักสำคัญในการแปล
- 1.2.2 เพื่อเป็นแนวทางในการศึกษาการแปลภาษาในระดับอื่นต่อไป
- 1.2.3 เพื่อเป็นแนวทางให้ผู้สนใจการแปลภาษาโดยวิธีรวบรวมฐานข้อมูล
- 1.2.4 เพื่อให้ผู้สนใจเห็นภาพรวมของเครื่องแปลงภาษาถึงลักษณะการทำงาน

1.3 ขอบเขตของงานวิจัย

งานวิจัยนี้จะทำการศึกษากระบวนการแปลภาษาโดยวิธีรวบรวมฐานข้อมูล โดยเน้นชนิดเทียบเคียงตัวอย่างในขั้นต้นจะทำการศึกษารูปแบบการแปลภาษาแบบต่าง ๆ แล้วทำการศึกษาเฉพาะเจาะจงลงไปทำการแปลแบบรวบรวมฐานข้อมูล เมื่อทำการศึกษาเสร็จแล้วจะทำการกำหนดขอบเขตของซอฟต์แวร์ที่ทำการเขียน และทำการกำหนดจะใช้รูปแบบภาษาของภาษาที่อยู่ในเรื่องกาลิเวอร์ (Gulliver) เป็นขอบเขตของการแปล โดยการแปลนั้นจะสามารถใช้ศัพท์ต่าง ๆ ในเรื่องกาลิเวอร์ กับรูปแบบประโยคที่อยู่ในเรื่องกาลิเวอร์เป็นสำคัญ นอกจากนี้ในซอฟต์แวร์ที่ทำการเขียนจะมีรูปแบบประโยคที่เป็นตัวอย่างที่ทำให้เห็นจุดเด่นของการแปลภาษาวิธีรวบรวมฐานข้อมูลชนิดเทียบเคียงตัวอย่างเป็นอย่างดี ในที่นี้ซอฟต์แวร์ที่ทำการเขียนจะใช้ภาษาปาสคาล (Pascal) สำหรับทำการเขียน

1.4 วิธีการดำเนิน

งานวิจัยโครงการนี้จะเริ่มจากการศึกษาทฤษฎีต่าง ๆ โดยทั่ว ๆ ไปของการแปลภาษา โดยจะมีการศึกษาในทางทฤษฎีตัวอย่างจากซอฟต์แวร์ที่มีผู้อื่นเคยทำวิจัยไว้แล้วในที่ต่าง ๆ หลังจากนั้นจะทำการศึกษาจากผู้เชี่ยวชาญในสาขานี้ ขั้นตอนต่อไปเป็นการกำหนดขอบเขตของซอฟต์แวร์ที่จะทำการเขียน เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น เมื่อผู้ใดเห็นประโยชน์หรือต้องการนำเอกสารนี้เป็นเอกสารใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นตัวอย่างสำหรับงานวิจัยนี้ ซึ่งจะมีการเลือกขอบเขตของการแปล ในที่นี้ได้อำนาจการเลือกหนังสือเรื่อง กาลิเวอร์ ทำการแปลทำการออกแบบโครงสร้างคร่าว ๆ ของซอฟต์แวร์ จากนั้นเริ่มทำการเขียนโดยจะทำการเลือกวิธีที่จะทำการแปล โดยสมมติว่าสิ่งที่เราจะทำการแปลนั้นมีวากยสัมพันธ์ (Syntax) และ ไวยากรณ์ (Grammar) ว่าถูกต้องอยู่แล้ว ดังนั้นในส่วนของการเขียนซอฟต์แวร์จะไม่มีเขียนส่วนของการตรวจสอบวากยสัมพันธ์ หลังจากเริ่มต้นทำการเขียนโปรแกรมด้วยภาษาปาสคาล เมื่อทำการเขียนเสร็จแล้ว จะทำการตรวจสอบซอฟต์แวร์ว่าถูกต้องหรือไม่ เมื่อทำทุกอย่างเสร็จสมบูรณ์จะทำการสรุปการทำงานต่าง ๆ แล้วทำการเขียนรายงานรวมถึงผลที่ได้รับและแนวทางในการศึกษาต่อไป



บทที่ 2

หลักการและทฤษฎี

2.1 หลักการทั่วไปของรูปแบบการแปลง (Transfer Model)

ใน transfer-base machine translation system, source sentence เป็นการวิเคราะห์ใน intermediate representation สำหรับ source language intermediate representation คือการที่จะแปลในตัว ของ target sentence สิ่งที่จะเจอจะดีกว่านั้นคือ งานหลัก ๆ สำหรับการ transfer และการ generate ยังรวมถึง

1. ลดการวิเคราะห์ผลลัพธ์ในรูปแบบ intermediate ที่เหมาะสมสำหรับ transfer
2. เลือกศัพท์ของ target ที่เหมาะสมสำหรับศัพท์ของ source
3. สร้างรูปแบบการ mapping ที่เหมาะสมสำหรับ รูปแบบ sentence ไปเป็น target structure
4. สร้างรูปแบบที่สมดุลของ target จาก target representation (คำว่า transfer ในบางครั้ง อาจจะหมายถึงทั้ง transfer และ generation)

2.1.1 การแปลโดยอ้างอิงกฎ (Rule Base)

ลักษณะสำคัญพื้นฐานของการแปลง คือ การหาค่าการ mapping จาก tree ไป tree ทั้งใน ระดับการแปลประโยคและการแปลตามความหมายคำ โดยการใช้ set ของ source-target transfer pattern action point เพื่อตอบสนองการเปลี่ยนแปลงในประโยคย่อย และในการเรียงลำดับของคู่ภาษา ปัญหาที่สำคัญอย่างเช่นความใกล้เคียงที่ครอบคลุมของกฎในการแปล (หรือรูปแบบ) และความเหมือนกัน ระหว่างกฎในบางรูปแบบที่สามารถเปลี่ยนแปลงได้อย่างกว้างขวางมันเป็นการยากและมีคุณค่าที่จะได้ การแปลที่สมบูรณ์และถูกต้องของกฎการแปลด้วยตนเอง โดยเฉพาะมันเป็นเรื่องยากที่จะกำหนดหน้าที่ โดยตรงของแต่ละ atomic unit สำหรับการ transfer เนื่องจากมีกฎที่มากมายของการแปล ด้วยเหตุนี้การ ที่จะทำการแก้ไขปัญหาที่ยากและปรับปรุงปัญหาต่าง ๆ ที่เกิดขึ้นกับการแปลจะเพิ่มความใหญ่ของระบบ เพิ่มขึ้น

อีกเรื่องหนึ่งใน transfer-base system ส่วนมากที่ทำตามแบบทางเดียว analysis-transfer-generation process, generation sentence โดยส่วนใหญ่แล้วเป็นการยากมาก ๆ ที่จะวิเคราะห์ grammar ตั้งแต่ generation rules ที่ได้ส่วนมาก source language generation grammar อาจจะรักษารูปแบบส่วนใหญ่ของ source language จนกระทั่งที่ถูกสร้างขึ้นไม่เหมือนประโยคที่ใช้พูดกันทั่วไปจากประสบการณ์ในพฤติกรรมในการแปลแสดงให้เห็นว่า คุณภาพของการแปลนั้นยังห่างไกลมาก ๆ จากที่ผู้ใช้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมคาดหวังไว้สำหรับผู้ทั่วไป ถึงอย่างไรผู้ใช้งานจะทำประโยคนั้นให้ถูกและสามารถเข้าใจในเหตุผลสำคัญได้อย่างหนึ่ง สำหรับการแปลที่ทำให้การแปลนั้นไม่คืออาจจะขึ้นอยู่กับความจริงที่ว่า generation grammar ไม่ได้ถูกออกแบบจากมุมมองของ target language เพื่อที่จะลดแรงงานมนุษย์ในการหากฎในการแปลและเมื่อสร้าง generation rules ที่มี source-bounded น้อย ๆ systematic approach สำหรับการปรับปรุงกฎในการแปลจาก source side และวิธีสำหรับ generation จาก target side ตามความต้องการ

2.1.2 การแปลโดยเทียบเคียงฐานข้อมูล (Example Base)

Example-Bass MT เป็นที่ยอมรับอย่างกว้างขวางเหมือนเป็นวิธีที่มีขีดความสามารถสำหรับการแปลอยู่ในช่วงเวลาเมื่อไม่กี่ปีที่ผ่านมา ขั้นตอนการแปลนั้นได้ทำโดยการจับคู่ (Matching) ของประโยคที่ต้องการจะแปลกับฐานข้อมูล (Corpus) ขนาดใหญ่

ตัวอย่างของการแปลโดยวิธีวัดความเหมือนในการแปลนั้นจะทำการหาตัวอย่างการแปลที่เหมือนกันมากที่สุดกับข้อมูลที่เข้ามา ปัญหาโดยทั่วไปของวิธีนี้คือการวัดด้วยความเหมือนนั้นอาจจะไม่สามารถที่จะหาตัวอย่างที่เหมือนกันอย่างมากกับประโยคที่เราต้องการแปล ยิ่งไปกว่านั้นข้อมูลตัวอย่างขนาดใหญ่อาจจะมีข้อมูลมากเกินไป แต่ขังใจก็ตามก็อาจจะสามารถลดลงถ้าใช้รูปแบบ (model) ในการออกแบบได้เหมาะสม นอกจากนั้นตัวมันเองจะลดลงหลายอย่าง เช่น ข้อมูลที่มากเกินไปจะลดและที่ใช้ในการหาข้อมูลถ้าเกิดมันมีการสร้างฐานตัวอย่างข้อมูลขนาดใหญ่

2.1.3 การแปลโดยใช้ข้อมูลทางสถิติ (Purely Statistical)

อีกทางหนึ่ง นอกจากสองวิธีข้างบน คือ การนำ translation model มาใช้ในแค่ความต้องการ วิธีทางสถิติ แม้ว่าในทางทฤษฎีจะเป็นที่น่าสนใจ แต่ก็ยังมีปัญหาในทางปฏิบัติอยู่มาก ในตอนแรกตั้งแต่มีการใช้รูปแบบการแปลภาษาแบบง่าย ๆ ในระดับของ string มันจะไม่สามารถทำได้เมื่อพบกับระยะทางยาว ๆ ที่อื่นกับขนาดของหน้าต่าง (window) ยิ่งไปกว่านั้นอาจจะเป็นเพราะว่า source model การใช้งานที่น้อยของข้อมูลทางด้านกายกยสัมพันธ์ และข้อมูลการแปลความหมายของคำโดยส่วนใหญ่ที่เหมาะสมกับระบบเครื่องแปล (Machine Translation : MT) มันจะทำการให้ source model มี order-free และสร้างตัวเลขที่เป็นไปได้ในการแปล ซึ่งบางทีอาจจะหาขอบเขตได้จากข้อมูลของวากยสัมพันธ์กับข้อมูลของการแปลซึ่งสามารถทำได้ ผลที่ตามมาคือช่องว่างของ parameter จะเพิ่มขึ้นเป็น exponential กับขนาดของหน้าต่าง ขอบเขตนั้นจะมุ่งไปที่ language model เมื่อเทียบเคียงกันแล้วฐานข้อมูลที่มีขนาดใหญ่ จำเป็นและจะต้องมีช่องว่างของตัวแปลขนาดใหญ่ด้วย นอกจากนี้มันอาจจะไม่สามารถนำไปใช้กับ model บาง model อย่างเช่น model สำหรับระบบที่มากความรู้ทางด้านกายกยสัมพันธ์และการแปลความหมายค่าอยู่แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.4 การแปลแบบรวมกัน (A Cooperative)

จากมุมมองของผู้ออกแบบวิธี statistical จะเหมาะสมสำหรับเป็นอุปกรณ์เริ่มต้นสำหรับความต้องการที่จะมี fine-grained probabilistic rule ที่มีขนาดใหญ่และไม่เปลี่ยนแปลง โดยส่วนใหญ่แล้วความรู้ด้านการเรียนภาษาจะได้มานั้นขึ้นอยู่กับคำแนะนำที่มาจากปฏิบัติที่ไม่เหมือนกัน เพราะว่าความรู้ทางด้าน Linguistical ส่วนมากได้มาจากการสังเกตหลาย ๆ ตัวอย่าง มีกลไกที่มากมายของกระบวนการสุ่มตัวอย่างสามารถวัดที่จะเข้ากับแบบจำลอง (Statistical Model) เพราะฉะนั้นการที่จะอธิบายแบบจำลองได้จะต้องใช้แบบจำลอง Stochastic Model เพื่อที่จะลดความไม่แม่นยำของข้อมูลเดิมในโครงสร้างเดิมที่ส่งผ่านการแปลแบบใช้กฎ และจดจำการทำงานลงใน rule induction อย่างไรก็ตาม Statistic ไม่จำเป็นต้องใช้ Statistical Model อย่างเดียว การที่จะได้มาซึ่งพื้นที่พารามิเตอร์ขนาดใหญ่ นั้นเป็นเพราะว่าทั้ง syntactic และ semantic ซึ่งเป็นข้อมูลที่ได้ทั่วไปในเครื่องแปล

พื้นที่พารามิเตอร์จะสามารถลดความแตกต่างได้ ถ้ากระบวนการโอนย้ายสามารถที่จะออกแบบโดยใช้ syntactic และ semantic ซึ่งจะเป็นลักษณะสำคัญที่ต้องนำมาพิจารณาพร้อมกัน ยกตัวอย่างเช่น แม้ว่าในคำอธิบาย (context) ของระบบเครื่องแปลจากภาษาอังกฤษไปเป็นภาษาจีน ซึ่งจะเกี่ยวข้องกับ ความแตกต่างของทั้งสองภาษาอย่างมาก มีพื้นที่การแปลได้เพียง 4 - 5 คำสั่งในการแปลทั่วไป ความหลากหลายของวลีที่จำเป็นอย่างยิ่งที่จะนำมาแปลด้วยประโยคที่มีความยาวไม่มากนัก ซึ่งแสดงให้เห็นว่าโครงสร้างของข้อมูลจะทำให้เนื้อหาจำนวนของข้อมูล ซึ่งสามารถนำผลนั้นมาใช้สำหรับการแปลตามหลักของเรานั้นเกี่ยวกับโครงสร้าง Statistical Model มาใช้เราจะสามารถทราบข้อมูล semantic และ syntactic ดังนั้นเราก็จะได้เปรียบอย่างมากจาก Statistical Model

ในส่วนนี้ Corpus Base จะเกี่ยวข้องกับข้อมูล Syntactic และ Semantic ด้วย โดยกลไกของ statistical สำหรับจุดมุ่งหมายของการแปลโดยทั่วไปในแต่ละหน้าที่ ในทำนองเดียวกันถ้าเราให้ความสนใจกับ

1. กลไกสำหรับการสร้างสกอร์ (Scoring Mechanism) เพื่อที่จะประเมินค่าระหว่างการแปลและการสร้างที่เป็นไปได้
2. อัลกอริทึมสำหรับหา Transfer Unit และ Transfer Rule ในส่วนที่ช่วยของในแต่ละหน้าที่ของอัลกอริทึมนี้ จะหา substantial help ในการลดการวิเคราะห์ผลลัพธ์ไปเป็นรูป normalized form ที่เหมาะสมสำหรับการแปล มันยังคงถูกยอมรับโดย Transfer Rule ที่ยังต้องการ Systematically และเพื่อให้ได้ Generation Rule เพื่อใช้สำหรับหาค่าในฐานข้อมูลของไวยากรณ์ของผลลัพธ์ (Target Grammar) จะทำการ annotated bilingual corpus ด้วยโครงสร้างข้อมูล จะนำไปใช้เพื่อสร้างโมเดล

(Model) Preliminary Result แสดงให้เห็นว่าส่วนประกอบที่อยู่ใน Transfer Rule จะเป็นการค้า เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น ไม่สามารถนำออกจำหน่ายหรือทำซ้ำโดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประกอบด้วย Transfer Operation ขนาดเล็กเท่านั้นที่สามารถนำไปใช้ได้ ดังนั้น Transfer Operation จะแสดงให้เห็นถึงความสามารถในการแปลที่มีอยู่อย่างมากนั้น สามารถที่จะสร้างการเปรียบเทียบของแต่ละการแปลที่แตกต่างกัน โดยเทียบกับ Decomposition ของ Source Tree ไปเป็นรูป Local Transfer Unit

2.2 Translation Model and Translation Score

จากรูปที่ 2 - 1 แสดง model สำหรับขั้นตอนการแปลและขั้นตอนการสร้างในรูปแบบความคิดนี้กฎการแปลความหมายคำต่าง ๆ แสดงสำหรับส่วนวิเคราะห์จะสมมติว่าได้ถูกวิเคราะห์ใน annotated syntax tree (AST) โดยที่ syntax tree จะอธิบายร่วมกับวากยสัมพันธ์ และการแปลความหมายทำในแต่ละจุด สำหรับความง่ายในการแปลและสร้างก็คือ มันควรที่จะเริ่มจากการแนะนำ normalized version ที่เราเรียกว่า Normal Form (NF) โดยที่ AST ที่ประกอบด้วย atomic transfer units เพียงอย่างเดียว ซึ่งหมายความว่า transfer operation นั้นสามารถทำให้เสร็จโดยง่าย ใช้ local operations ในแต่ละ unit นอกจาก conceptual model แล้ว ขั้นตอน source normalization ก็จะทำเพื่อลด source AST ไปเป็น normal form source NF ไปเป็น target NF จะถูกทำหลังจากนั้น และ target generation grammar

จากที่มันอาจจะมีตัวเลือกหลายตัวเลือกในขั้นตอนการแปลกลวิธี scoring จะถูกใช้ เพื่อจะหาข้อแตกต่าง แต่ละทางเลือกขึ้นอยู่กับว่า transfer-base transition process สามารถที่จะ model เหมือนกับว่าเป็นขั้นตอนที่ดีที่สุด ในกรณีที่มีการแปลความหมายคำ และการ transformation ที่ดีที่สุด คือ การค้นหาเพื่อค่า translation score ที่มากที่สุด ตัวอย่างเช่น ในการแปลภาษาอังกฤษ เช่น ภาษาจีนในระบบเครื่องแปล

ถ้า CW^{nC} , EW^{nE} , ELM, TM, CLM เป็นภาษาจีน (ie target) ความยาวของประโยคจะเป็น n_C ในภาษาอังกฤษ (ie source) ความยาวประโยคจะเป็น n_E ภาษา English (source) language model transfer model และ Chinese (target) language model จะแสดงโดย I_C , I_E ตามลำดับ เช่น intermediate representation สำหรับประโยค target และ source ดังนั้น translation score สามารถ define เหมือนกับตัวอย่างดังนี้

$$\begin{aligned} & P(CW_1^{nC} | EW_1^{nE}, ELM, TM, CLM) \\ &= \sum_{I_C} \sum_{I_E} P(CW_1^{nC}, I_C, I_E | EW_1^{nE}, ELM, TM, CLM) \\ &= \sum_{I_C} \sum_{I_E} P(G_1^{nC} | I_C, I_E, EW_1^{nE}, ELM, TM, CLM) \end{aligned}$$

$$\begin{aligned}
& \times P(I_C | I_E, Ew_1^{n_E}, ELM, TM, CLM) \\
& \times P(I_E | Ew_1^{n_E}, ELM, TM, CLM) \\
& \approx \sum \sum P(Cw_1^{n_C} | I_C, CLM) \\
& \times P(I_C | I_E, TM) \\
& \times P(I_E | Ew_1^{n_E}, ELM) \\
& = \sum_{I_C} \sum_{I_E} P_G(.) \times P_T(.) \times P_A(.)
\end{aligned}$$

อันที่จริงแล้ววิธีการวางแผนในการค้นหา คือ การทำเพื่อหาการวิเคราะห์ที่เป็นไปได้มากที่สุดของ

$$I_{E_{\max}}^{\Delta} = \arg \max P(I_E | Ew_1^{n_E}, ELM)$$

จะขึ้นอยู่กับค่าที่มากที่สุดจะเป็น Analysis score PA(.) ในส่วนของกรวิเคราะห์ จากนั้นค้นหาค่าที่เป็นไปได้มากที่สุดของ intermediate representation IC_{\max} สำหรับประโยคภาษาจีนจะมาจาก

$$I_{C_{\max}}^{\Delta} = \arg \max_{I_C} P(I_C | I_{E_{\max}}, TM)$$

ค่าที่มากที่สุดจะเป็น generation score PT(.) จากนั้นหาประโยคภาษาจีน

$$CW_{\max}^{\Delta} = \arg \max_{CW_1^{n_C}} P(CW_1^{n_C} | I_{C_{\max}}, CLM)$$

Analysis score หรือ Score function คือจะในกรณีที่เกิดถ้ากำหนดหรือ Z ความหมายในช่วง analysis phase transfer score และ generation score จะถูกกำหนดในช่วงของ fronster และ generation process

2.2.1 Transfer Score

งานหลักของช่วง transfer คือ ทำหน้าที่เป็นตัวเชื่อมช่องว่างของโครงสร้างของ source จัดโครงสร้างของ target โดยที่อาจจะมีการเชื่อมได้หลายวิธี ดังนั้นมันจึงจำเป็นต้องมีวิธีหา score เพื่อที่จะเลือกวิธีการสร้าง target ที่เหมาะสมที่สุด ตัวอย่างหนึ่งในการแปลจากภาษาอังกฤษเป็นภาษาจีน เช่น

ต้องการเปลี่ยน noun ให้เข้ากับ verb ในภาษาจีน เช่น The use of NP... มีการแปลได้หลายวิธีและทุกวิธีก็สามารถยอมรับได้ อย่างไรก็ตามมันควรจะตรงในภาษาจีน ถ้ารูปปรกติของมันเปลี่ยนเป็นเท่ากันกับ Using NP... หรือ To use NP...

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การวิเคราะห์อย่างพิถีพิถันขึ้นอยู่กับกระบวนการแยกของตัวอย่างการแปลจริงๆ แจ้งให้ translation process ที่ต้องการควรมี transfer score สำหรับวัดว่าค่าเพื่อจะหาว่าวิธีไหนต้องการจากวิธีที่แตกต่างกัน และมีโครงสร้างที่ปรับเปลี่ยนระหว่างประโยค source กับ target

อย่างไรก็ตาม ปัจจัยทางด้านเนื้อที่สำหรับ transfer score น่าจะใหม่มาาก ๆ ถ้า underlying transfer mechanism ต้องขึ้นอยู่กับ การ mapping แบบ tree-to-tree จาก source AST ไปเป็น Target AST ในกรณีนี้ อาจมีอันหนึ่งมีโครงสร้างฐานข้อมูล (Data-Base) ขนาดใหญ่ สำหรับ source target ที่มีการเปลี่ยนไปเปลี่ยนมา และ scores ของมันในทางตรงกันข้ามค่าของ transfer score และ generation score จะพุ่งถึงในส่วนถัดไป อาจจะเร็วไปมาก ๆ ถ้าแต่ละ AST สามารถลดมาเป็น normalized version ที่สามารถแยกออกเป็น set ของ atomic transfer units หรือที่เรียกว่า local transfer units (LTU'S) และ transfer กับ generation process สามารถทำให้สำเร็จได้โดยง่ายโดยใช้ local operation อย่างเช่น local permutation หรือ deletion บท transfer unit นั้น ๆ จะไม่ขึ้นอยู่กับแต่ละ transfer units และ transfer สามารถหาค่าได้ง่ายโดยใช้การแยกออกเป็นส่วน

สำหรับตัวอย่างถ้า I และ I' เป็น target และ source normal forms ที่ประกอบด้วย set ของ target LTU'S x_i และ source LTU'S x'_i ที่แต่ละ LTU x_i ที่แต่ละ LTU ประกอบด้วย syntactic category (X_i) สำหรับ unit หัวของมัน (h) ความสัมพันธ์กันจะต้องกับ unit พวกนี้ จากนั้น transfer score สามารถกำหนดได้โดยง่าย โดยการจำแนกกับข้างล่าง

$$\begin{aligned} P(I_c | I_E, TM) \\ &= P(X'_1, X'_2, \dots, X'_n | X_1, X_2, \dots, X_n) \\ &\approx \prod_{i=1}^n P(X'_i(h', a', s', m') | X_i(h, a, s, m)) \end{aligned}$$

เพื่อให้ได้ term สุดท้ายตรงสมมติว่า X_i ขึ้นอยู่กับเฉพาะ LTU, X สูตรนี้จะใช้เพื่อหา Target LTU'S ที่ทำจะเป็นไปได้ดีที่สุดสำหรับ set ของ source LTU'S และที่สามารถทำให้สำเร็จโดยหา mapping ที่ดีที่สุดระหว่างแต่ละคู่ LTU โดยวิธีการแยก โดยปรกติแล้วตัวเลขของ LTU'S จะมีปริมาณที่แน่นอนและมีขนาดที่สมควรจากหลายค่าร่วมกัน โดยโครงสร้างวากยสัมพันธ์ของช่องทางที่ต้องการและ training corpora เก็บด้วยเหตุของเล็กน้อย

หมายเหตุ syntactic และ semantic เป็นหน้าที่ของ LTU และทั้ง 4 tuple สามารถเปลี่ยนระหว่าง transfer mode ยังคงสามารถใส่หรือลบตามที่ต้องการสูตรนี้จะทำการหาค่าถึงความเร็วที่เพียงพอสำหรับ transfer process และ enable drastic stylistic และการเปลี่ยนแปลงโครงสร้างระหว่างภาษา 2 ภาษาที่ถูกทำงานอยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2 - 2 จะแยกให้เห็นตัวอย่างของ normal form สำหรับสูตรที่พูดถึงข้างบนในความเป็นจริงโครงสร้าง head-argument ที่เหมาะสมที่กำหนดคุณสมบัติที่ดีของ local transfer สามารถใช้ได้เหมือน LTU วิธีนี้จะสรุปคอนทักซ์เพื่อหาทุกทางที่เป็นไปได้สำหรับแต่ละ nit โดยอัตโนมัติจาก corpus ของภาษาที่ควมกัน

2.2.2 Generation Score

เพื่อที่จะร่วม target sentence จาก target normal form ต้องการทำ structure generation สำหรับเรียงลำดับ target lexicon ที่เลือกไว้แล้วจะจัดหา morphological generation สำหรับสร้าง target-specific morphemes, particles และอย่างอื่น ยกตัวอย่างเช่น คำภาษาจีนพิเศษคำว่า “ba”, “bei” (passive voice marker) และจัดระดับสำหรับ noun ต้องถูกสร้างระหว่างช่วง morphological generation แปลประโยคที่แตกต่าง คือ ที่ต้องการต่างกันในแต่ละสถานะการที่แตกต่างกัน generation score ต้องการวัดค่า ความโน้มเอียงสำหรับแต่ละรูปแบบ sentence ที่เป็นไปได้ต่างกัน

ให้ target sentence $Cw,^{wc}$ แสดงโดย set ของ sentence segment t_i , ระดับการเปลี่ยนแปลงจะกำหนดโดย function เปลี่ยนแปลง $u(x;)$ และ (morphological) generation จะแสดงโดย generation frame $f(X)$ สำหรับ LTU X_i จากนั้น generation score สามารถกำหนดอย่างข้างล่างขึ้นอยู่กับการรวมของ local transfer

$$\begin{aligned} & P(Cw_1^{wc} | I_C, CLM) \\ & \approx P(t_1, t_2, \dots, t_n | X_1', X_2', \dots, X_n') \\ & \approx \prod_i P(f(X_i'), u(X_i') | X_i'(h', a', s', m')) \\ & = \prod_i P(u(X_i') | X_i'(h', a', s', m')) \\ & \times P(f(X_i') | u(X_i'), X_i'(h', a', s', m')) \end{aligned}$$

Target sentence ที่ดีที่สุดจะทำให้ค่า generation score สูงสุด อาจจะต้องเลือกรูปหลายประโยคที่เป็นไปได้แตกต่างกันโดยปรกติแล้ว เราจะสมมติว่า sentence segment ลำดับที่ I จะขึ้นอยู่กับเฉพาะกับ LTU ลำดับที่ I ต้องจำไว้ว่า product term แรกในสูตร สุดท้ายจะขึ้นกับโครงสร้าง generation (หรือโครงสร้าง transfer) process และ second term ขึ้นอยู่กับช่วง morphological generation permutation Function จะเป็นตัวกำหนดความสัมพันธ์ทางบวกอีกทางหนึ่ง บอกตามความแตกต่างว่า target-specific morphemes อันใหม่เป็นตัวที่จะใส่ในแต่ละตำแหน่งของมันในระหว่างแต่ละ elements ของ 4-tuple ตัวอย่างเช่น เมื่อ normalized verb phase

$$VP(h, a1, a2, s, m) \\ = VP(put, the - book, on - the - desk, NIL, NIL)$$

ใช้ในการสร้างภาษาจีนที่เหมือนกัน

$$VP(BA, book, put, on - desk)$$

Permutation function จะเป็น $U(h, a1, a2, s, m) = (a1, h, a2, s, m)$ และ generation frame จะเป็น $f(a1, h, a2, s, m) = (BA, a1, \phi, h, \phi, a2, \phi, s, \phi, m, \phi)$ สำหรับส่วนในภาษาจีน “ba” (ϕ เป็นเครื่องหมาย null)

2.3 Learning Transfer Unit and Operations

ตัวแปรหนึ่งที่สำคัญในความสำเร็จของ transfer และ generation mode) จะขึ้นอยู่กับความคงอยู่ของ localized transfer units และ consistent set of transfer operations (transfer rule) ที่ควบคุม transfer process อย่างเช่น unit จะแสดงค่าการคำนวณของ transfer หรือ generation score และงานง่าย ๆ โดยการจำแนก syntax tree หรือ normal form ในทางปฏิบัติแล้วอาจจะไม่หา ideal underlying transfer mechanism และ conduct transfer โดยใช้เฉพาะ local operation อย่างไรก็ตาม transfer unit จะปรากฏสำหรับหลาย ๆ โครงสร้าง (กรณีที่เกี่ยวข้องก็คือการใช้ประโยชน์ทั้งหมดกัน transfer unit) เพื่อจะกำหนด normal form ที่มีคุณสมบัติของ local transfer เราสามารถนำไปใช้กับจาก algorithm “index permutation” ไป bilingual tree bank เพื่อหา candidate transfer unit และ corresponding tree bank เพื่อหา candidate transfer unit และ corresponding transfer rules เพื่อความสอดคล้องกับ model ในรูปที่หนึ่งที่จะทำการเรียนคือขบวนการ transfer (I) ระหว่างขบวนการ Source AST และ source normal form ระหว่างคู่ normal form ของและระหว่าง target normal form และ target AST อย่างไรก็ตามสำหรับความง่ายเพื่อแสดงให้เห็น เราจะแสดงว่าจะหา candidate ของ LT’s

Corresponding transfer rules จาก sources syntax tree และ corresponding target translation ของตัวเองอย่างไร การเรียนรู้เกี่ยวกับ algorithm สำหรับคู่ tree อื่น ๆ คือ ปัจจัยที่จะบ่งชี้และเป็นความจริงได้โดยง่ายขึ้น

ในตอนนี้จะอ้างถึงรูปที่ 2 - 3 สำหรับ source AST และ Target translation pair (T1, T2, ...) เป็น equivalent target words สำหรับ source words ที่อยู่ติดกันโดยไม่จัดลำดับใหม่ ดังนั้น mode บรรพบุรุษที่เหมือนกันของ source word นี้ จะต้องมีลำดับของคำที่ติดกันเป็นเครื่องใช้ใน index list ของตัวมันเอง index list นั้น จะมีการเรียงลำดับข้อมูลของ transfer process mode ของ source syntax tree สามารถระบุตาม index list ของมันเป็น 3 แบบ คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. index list คือ ลำดับที่ติดต่อกันและเรียงลำดับโดยใช้จำนวนเต็ม (ตัวอย่างเช่น D(3, 4), E(1, 2))
2. ตัวชี้ค่าจะเป็นติดกันแต่ไม่เรียงกัน (เช่น B(3, 4, 1, 2), C(5, 7, 6, 8), A(3, 4, 1, 2, 5, 7, 6, 8))
3. ตัวชี้ค่าจะยังคงไม่มีจำนวนเต็มติดต่อกัน นอกจากหลังจากได้รับการเรียงลำดับแล้ว เช่น F(5, 7), G(6, 8) ชนิดของ mode ที่ต่างกันต้องการ transfer operations ที่แตกต่างกัน เพื่อที่จะสร้างตัวติดกันของมัน และเรียงลำดับเหมือนกับ ลำเนา target ของพวกมัน

ถ้า mode อยู่ใน type-1 แล้ว terminal strings rooted ทั้งหมด ที่ node นี้ ควรจะต้อง translated เป็นลำดับของ target word ที่ติดกันใน target sentence โดยไม่เปลี่ยน linear order ของมัน ระหว่าง transfer ตัวอย่างเช่น Node D เช่น type - 1 node และ Children ของมัน คือ (s3, s4) ถูก translated ใน (T3, T4) โดยที่ไม่เปลี่ยนความสัมพันธ์แบบ linear order ของมันใน target sentence

Type - 2 node เช่น candidate ที่เป็นไปได้ของ local transfer unit terminal string rooted ที่ node ถูก translate ไปเป็นลำดับของ target ว่า index list ได้ sorted หมายความว่า children ของมันจะเรียงลำดับระหว่าง transfer ยิ่งไปกว่านั้น operation ต้องการจะเปลี่ยน linear order ของ source words ไปเป็น order ของ target equivalent ของมันเพื่อจะสามารถ ได้มาโดยง่าย โดย ตรวจสอบความสัมพันธ์ของ ตำแหน่งของ index list ของ children ของมัน ตัวอย่าง terminal string (S3, S4, S1, S1) rooted ที่ node B สามารถที่จะ transfer ไปเป็น equivalent target word (T1, T2, T3, T4) ของพวกมัน โดย transfer Operation ข้างล่างนี้

$$B(3,4,1,2) \rightarrow D(3,4) E(1,2)$$

$$\Downarrow \text{Local Transfer}$$

$$B(1,2,3,4) \rightarrow E(1,2) D(3,4)$$

สิ่งที่ต้องการในการ transformation มีลักษณะเกี่ยวกับการเปลี่ยนแปลงของ children ที่สร้าง ตัวชี้ค่าของ children ปรากฏในลำดับขึ้น เราจะเรียก transfer Operation อย่างนี้ว่า XCHG (EXCHANGE or permutation) operation สุดท้าย ถ้า node เป็นของ type - 3 ถ้าชี้ค่าที่ไม่ติดกันหมายความว่า source word rooted ที่ node นั้น ต้องถูก translated ไปเป็นตำแหน่งที่ไม่ต่อกัน target word ของพวกมันต้องผสมกับ target word อื่นยกตัวอย่าง เช่น children ของ type - 3 node F(5, 7) ถูก translated ไปเป็นตำแหน่งที่ 5th และ 7th เพื่อเป็น locally transferable type - 3 node จำเป็นต้องรวมกัน (MERGE) กับ node อื่นๆ ดังนั้น children ของพวกมัน สามารถเรียงลำดับโดยวิธีของ local permutation ควรจะจำไว้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ว่าไม่ว่าเมื่อใดก็ตาม ถ้า node นั้น ๆ ถูก merged เพื่อสร้าง unit ที่สามารถเรียงลำดับโดยวิธี local permutation ดังนั้น merged node ต้องเป็นของ type -2 ดังนั้น node ใน type - 3 ทั้งหมดที่มี common type - 2 บรรพบุรุษ node สามารถ merged เพื่อสร้าง transfer unit เพื่อจะเก็บรักษาตัวเลขของ children ของ merged node ให้เล็กเท่าที่จะเป็นไปได้มากที่สุด merged type - 3 nodes จะยังคงสร้าง candidate ของ local transfer unit ตัวอย่างเช่น node ใน type - 3 F(5, 7) and (6, 8) สามารถ merged เพื่อสร้าง node ใหม่อย่างเช่น $c(5, 7, 6, 8) \rightarrow S_5, S_7, S_6, S_8$ XCHG operation สามารถที่จะจัดหาดังหลังจากที่ generation อันดับที่ถูกต้อง

algorithm นี้สามารถที่จะขยายเมื่อ source word และ target words ไม่เป็น one to one correspondent (1) ถ้า source word corresponds มากกว่าหนึ่ง target words source word สามารถมีความสัมพันธ์กับ index list ประกอบด้วย ตัวทั้งหมดที่ติดกันของ target words (2) ถ้ามีมากกว่าหนึ่ง source words ที่ถูก translated ในหนึ่ง target word ดังนั้น source words สามารถที่จะมีความสัมพันธ์กับ word index ของ single target word algorithm นี้จะสร้างเหมือนในกรณีที่เป็น one to one สำหรับทั้ง 2 case ข้างบน (3) ถ้า source word mapped เป็น Zero target word source word สามารถมีความสัมพันธ์กับ null word index mother node ของ node ที่ต้องการ DELETION (DEL) operation เพื่อ delete source word (4) ในทางตรงกันข้าม ถ้า target token ไม่มีความสัมพันธ์กับ source word ดังนั้นตัวสุดท้าย type -3 node ด้วย target word index หายไปจะปรากฏใน source syntax tree ในกรณีนี้ INSERTION (INS) operation จะต้องการที่จะใส่ pseudo node ไปเป็น type -3 node ที่โค้วกับ source terminal words มากที่สุด (5) ในกรณีพิเศษบางกรณี source word อาจจะมีการแทนที่ กับ equivalent source word ตัวอื่น ก่อนที่จะทำให้มันอยู่ในแนวเดียวกันกับ target word สถานะการอย่างที่จะเกิดขึ้น ตัวอย่างเช่น เมื่อ สรรพนามภาษาอังกฤษถูก translated ไปเป็นคำภาษาอื่น มีความสัมพันธ์กับ referent ของตัวมันเอง การเปลี่ยนที่สำคัญอย่างอื่นเรา จะเรียกว่า REPL (REPLACEMENT) operation

รูปที่ 2 - 4 แสดง transfer operation ที่เกิดขึ้นในการแปลส่วน ver phase ในภาษาอังกฤษ “.....select SAVE from the menu” ใน สำเนาในภาษาจีนของตัวเอง สัญลักษณ์ทั้งหมดของ C-XXX จะแสดงกับ Chinese equivalent ของสัญลักษณ์ ภาษาอังกฤษ XXX index o จะแสดงอย่างชัดเจนใน source syntax tree เพื่อแสดงตำแหน่งของสิ่งที่ไม่ได้กำหนด จะแสดงให้เห็นในภายหลัง transfer operation ส่วนใหญ่ และ transfer rules สามารถอธิบายในรูปของ NOP, INS, DEL, XCHG, REPL, และ MERGE operation (หรือวัตถุประสงค์มัน) ในทาง mechanical และ consistent ยิ่งไปกว่านั้น source tree node และ target tree node กับ index list ที่เหมือนกันจะสร้าง translation pair ตัวอย่างเช่น ถ้า target tree node B' ปรากฏและ B'-- $\rightarrow T1T2T3T4$ ดังนั้น sub tree rooted ที่ B'(1,2,3,4) จะสำเนาของ B(3,4,1,2) ตั้งแต่ B--- $\rightarrow s3s4s1s2$ และในทำนองเดียวกัน อย่างเช่น translation pairs เช่น B และ B' ดังเช่นจะจัดหา candidate ที่ดีสำหรับเราเพื่อที่จะกำหนด normal form มันเป็นไปได้ที่ transfer operation อย่าง INS และ MERGE จะประยุกต์เพื่อจะติดกับ node ในทางที่สร้าง source tree และ target tree เหมือนกันมากที่สุด เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อย่างเช่น การวางแผนอาจจะดูกรณีที่เป็นไปได้ที่จะเป็น nonlogical operation เป็นไปได้ที่การวัดเหมือนกัน จะเป็นตัวอย่างของ matched translation pair

โดยการซึ่งใน tree ด้วยวิธี bottom - up จะสามารถหา minimal number ของ transfer operation ต้องการที่จะ transfer จาก tree ไปสู่ target equivalent ของมันเอง และหา candidate สำหรับ local transfer units unit นี้ จะจัดหา normal form ไม่ถูกกำหนดขอบเขตจริง ๆ โดย syntax LRU candidate ที่ได้มากับ algorithm ที่ผ่านมาสามารถลดลงเพื่อให้ได้โครงสร้างที่ลึกขึ้น โดยการลดขั้นตอน จะเริ่มการทำการทดสอบความสามารถร่วมกับ candidate นี้ และสร้างการปรับปรุงที่เหมาะสม

2.4 Learning Target Generation Rules

Algorithms ที่ผ่านมาสามารถที่จะปรับให้กับ tree ไปเป็น normal form mapping และ normal form ไปเป็น normal form mapping เพราะพวกมันเป็นเพียงชนิดที่แตกต่างกันของ tree ในปัจจัยที่ใช้ เมื่อนำมาใช้ algorithm ระหว่าง target normal form และ target AST ในทิศทางที่กลับกัน เราสามารถได้ set ของ generation rules ที่เกิดขึ้นกับ target grammar generation rules ต้องการในแตกต่างจากสำเนาของมันสำหรับ conventional transfer based system เมื่อมองลึกเข้าไป transfer operations ระหว่าง target normal former target syntax tree จะหาได้จาก annotated syntax trees ของ target language ในที่นี้ Corresponding generation grammar เป็น analysis grammar โดยเฉพาะของ target language ไม่ใช่ grammar ที่ปรับจาก source grammar ผ่าน ขั้นตอนทางเดียวของ analysis transfer - generation generation process ใช้ generation rules เป็นกระบวนการย้อนกลับอย่างง่ายของการวิเคราะห์ target sentence กับ target grammar ในที่นี้ generation sentence จะถูกบังคับให้ทำตาม grammar และรูปแบบของ target language

Algorithm นี้จะหา important mechanism ของการสร้าง translation ที่มีคุณภาพมาก ๆ มัน เป็นไปไม่ได้สำหรับ generation grammar ที่จะปรับจาก source language side ในกรณีนี้ รูปแบบของ generation sentence จะกำหนดมีความสัมพันธ์กับ source grammar ที่อาจจะไม่ปรกติ ที่จะใช้กับภาษาพูดปกติของ target language เพื่อที่จะสร้าง translation ที่มีความสามารถสูง ๆ เราสามารถที่จะนำมาใช้ เป็น hi directional strateg ในกรณีที่มีความต้องการ transfer rules เป็น นำมาจาก source AST และ source normal form และ generation rules จะหาได้จาก target AST (ได้มาโดย (วิธีวิภาค) target sentence และ target normal form ในเวลานี้ transfer score และ generation score สามารถที่จะปรับเพื่อให้ได้ตัวเลือกที่ดีที่สุด

2.5 Experiment

เพื่อที่จะแสดงความสามารถอย่างสูงของ model ข้างบนการทดลองข้างต้นจะทำกับ bilingual corpus เล็ก ๆ corpus เล็ก ๆ นี้จะเก็บ 111 คู่ประโยคภาษาอังกฤษกับภาษาจีน และ corresponding parse tree ของมันโดยที่ถูกสร้างโดย Behavior Trans English - Chinese MT system ความยาวโดยปรกติของ source (ภาษาอังกฤษ) language คือ 13 คำ

Source tree จะประกอบด้วย node 6126 ที่ local operation สามารถทำได้ อย่างไรก็ตามจะมีเพียง 564 transfer operations จะแสดงในตารางที่ 2 - 1 จะต้องได้มาโดย target translation ของ sentence ทั้งหมดจะอยู่ในลักษณะประมาณ 5.1 operation แต่ประโยคจากที่เราได้สังเกตและหาเหตุผลจากการสมมติครั้งก่อนที่ว่า structure information สามารถลดการหาของ transfer operation ที่เป็นไปได้มากที่สุดของความขัดแย้งกับ purely statistical model ที่รูปแบบการเรียงตัวเพิ่มขึ้นเป็น exponentially กับความยาวของคำ

ระหว่าง transfer operations เฉพาะ MERGE operations ต้องการ nonlocal movement โชคดีที่ nonlocal operation ประกอบเพียงแค่ประมาณ 8-9% ของ transfer operation ดังนั้น merged node จะเจอน้อยกว่า 2 ระดับ ขึ้น tree ผ่านไปในกรณีมากที่สุดเดียวในกรณีส่วนใหญ่ transfer operation สามารถที่จะปฏิบัติ locally และสมมติในการลด transfer score และ generation score ขึ้นกับ local transfer unit อย่างมีเหตุผลทีเดียว

ตารางที่ 2.2 จะแสดง typical transfer operation เล็ก ๆ น้อย ที่ถูกโดย algorithm แต่ละหน้าที่ transfer operation จะแสดงในหลัก (column) "OP" และความถี่ที่เกิดขึ้น จะแสดงในหลัก "Freg" หลัก transfer unit จะแสดงถึง node บนแต่ละ transfer operation ว่าถูกใช้งานลบหรือใส่ที่ประกอบจะอ้างถึงในวงเล็บแบบแหลม (" $\langle \rangle$ ") สิ่งที่เป็นพิเศษของ transfer operation "XI" (XCH รวมกับ INS) ที่จะเกิดขึ้นในเวลาเดียวกันสำหรับการสร้างแสดงดังนี้

DEL operation หัวตัวแรก ตัวอย่างเช่น แสดงว่า article หรือ determiner (DET) ภาษาอังกฤษอยู่ใน noun phase (NZ) บาง prepositions (P) อยู่ใน preposition phase (P1) ของกริยา "be" และกริยาช่วยตัว (AUX) จะอยู่ก่อน verb phase (V2) complementizer หรือ relative pronoun (CMPR) ใน post nominal modifier (NMOF) จะรวมประโยคที่ไม่สมบูรณ์ (SI - ZA) จะถูกบปใน English - Chinese translation INS operation ส่วนใหญ่จะขึ้นอยู่กับการใส่ของ Chinese "de" particle (CDE), classifier ($\langle \text{CL} \rangle$), locative morphemes ($\langle \text{LOC} \rangle$) English equivalent ของ "please" ($\langle \text{please} \rangle$), และอื่น ๆ อีก XCHG operations ขึ้นกันอย่างหยาบกับ Structure transfer process พบมันควบคุมการเปลี่ยนกลุ่มของ post - nominal modifiers (NMF*) หรือกลุ่มของ verb modifiers (ADV) ของประโยคภาษาอังกฤษที่อยู่หน้า noun phases หรือ verb phases (V1) พวกมันส่วนใหญ่จะควบคุมการเปลี่ยน passive voice ในเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่บนเว็บไซต์เป็นการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

English declarative sentences (SDEC) ใน active voice ที่แสดงใน Chinese sentences และที่ได้ก็คือ subject noun phrases (N3) จะเปลี่ยนเป็นตัวส่งของ verb phases (V4)

เมื่อดูลึกเข้าไปอีกที่ transfer operations ในการสร้าง transfer rules แสดงว่ามากกว่า 75% ของ transfer operation จะครอบคลุมเหนือ 24 transfer operation ที่ได้พูดอย่างคร่าว ๆ พอวันที่ในตารางที่ 2 มันคงให้เห็นว่าการสร้าง transfer operation จะสอดคล้องกับ general linguistic observation ได้ดีทีเดียว proposed model ที่จัดหาสิ่งที่มาสนับสนุนการสร้าง model transfer process และ transfer rules ที่ได้มาในทาง systematic และ consistent สำหรับการทดสอบในปัจจุบัน transfer operation ส่วนใหญ่สามารถอธิบายได้ในรูปของ syntactic features และ syntax tree จะใช้เพื่อแสดงเป็น normal form โดยตรงสำหรับ domain ที่ใหญ่กว่านี้จะคาดเดาได้ว่าจะมี semantics - orient normal form มากขึ้นเหมือนที่อธิบายมาก่อนหน้านี้แล้ว จะต้องการในลำดับที่ต้องเกี่ยวข้องกับ transfer operation ที่ประหลาดขึ้นและมี syntactic ที่เปลี่ยนแปลงมากขึ้น แต่กระนั้นก็ตาม procedure ที่ถูกสร้างขึ้นมาสามารถที่จะใช้งานโดยไม่ต้องเปลี่ยนแปลงในข้อของความสัมพันธ์กับ semantic level ในตอนนี้ยังอยู่ในการค้นและทดสอบ

2.6 บทสรุป

จะเห็นว่าได้ transfer score และ generation score มีหน้าที่โดยเฉพาะของมันเพื่อที่จะวัดสำหรับ transfer และ generation processes corpus - base จะเป็นการพูดถึงการขึ้นกับกรรมวิธีที่จะสามารถให้พวกเราสามารถรู้ถึงข้อแตกต่างในการเทียบสิ่งที่เปลี่ยนไปในแต่ละการ transfer $<0=$ generation algorithm นี้จะมีจุดประสงค์เพื่อให้ได้ set ของ local transfer unit และ transfer rules ของคู่ภาษาที่ทำการแปลใน algorithm นี้จะยอมให้ค่า transfer process ได้สำเร็จ และ generation processes เพียงอันหนึ่งโดยได้ local transfer operation นอกจากนี้โดยการประยุกต์ algorithm เพื่อศึกษา generation rule จาก target language โดยตรง generation phase สามารถสร้าง target sentences ที่ยังก็ตาม grammar และรูปแบบของ target language ระบบสามารถที่จะปรับเพื่อสร้าง target sentence ที่ขึ้นกับ target grammar แทนที่การเปลี่ยนแปลง version ของ source grammar

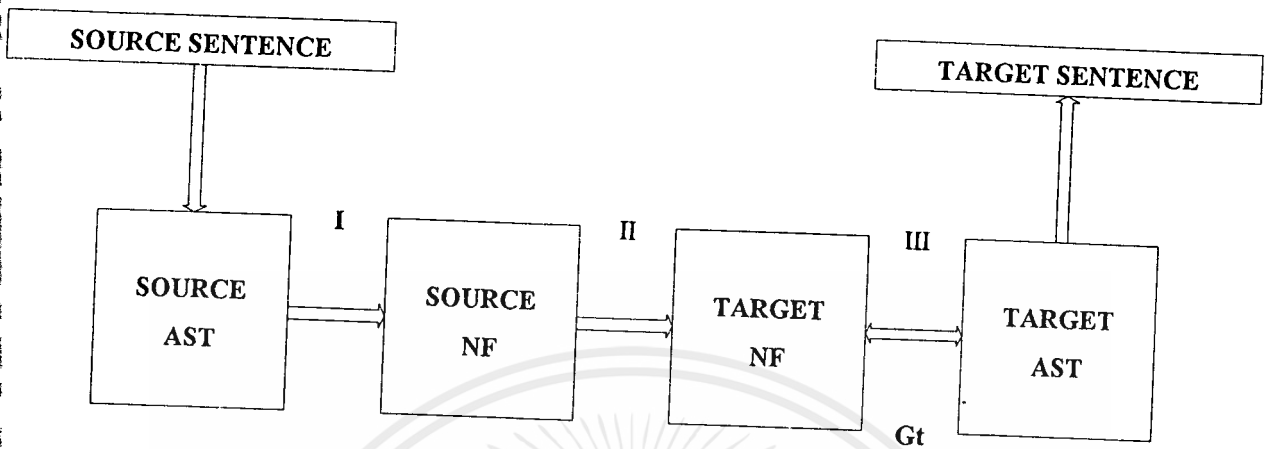
Operation	Frequency	Percentage
DEL	230	40.8%
INS	158	28.0%
XCHG	122	21.6%
MERGE	50	8.9%
REPL	4	0.7%

ตารางที่ 2.1 ตารางการแยกคำสั่งในการแปล

Freq	OP	Transfer Unit	Configuration after Transfer
114	DEL	N2 -> <DET> N1	=>N1
42	DEL	P1 -> <P*> N3	=>N3
16	DEL	V3 -> <AUX> V2	=>V2
15	DEL	N2 -> <DET> NLM*N1	=>NLM* N1
10	DEL	NMOF -> <CMPR> SI-2A	=>SI-2A
20	XI(+)	N3-A -> N2 NMF*	=>NMF* <DE> N2
16	INS	SDEC -> N3 V4	=>N3 V4 <DE>
15	INS	S2 -> SAC* S	=>SAC* <please> S
10	INS	N2 -> DET N1	=>DET <CL> N1
8	INS	P1 -> P* N3	=>P* N3 <LOC>
75	XCHG	N3-A -> N2 NMF*	=>NMF* N2
19	XCHG	V2A -> V1 ADV*	=>ADV* V1
5	XCHG	SDEC -> N3 V4	=>V4 N3
(+)XI: XCHG and INS			

ตารางที่ 2.2 ตารางตัวอย่างคำสั่งในแต่ละคำสั่งของการแปลภาษาอังกฤษเป็นภาษาจีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



I : Source Normalization

II : Transfer

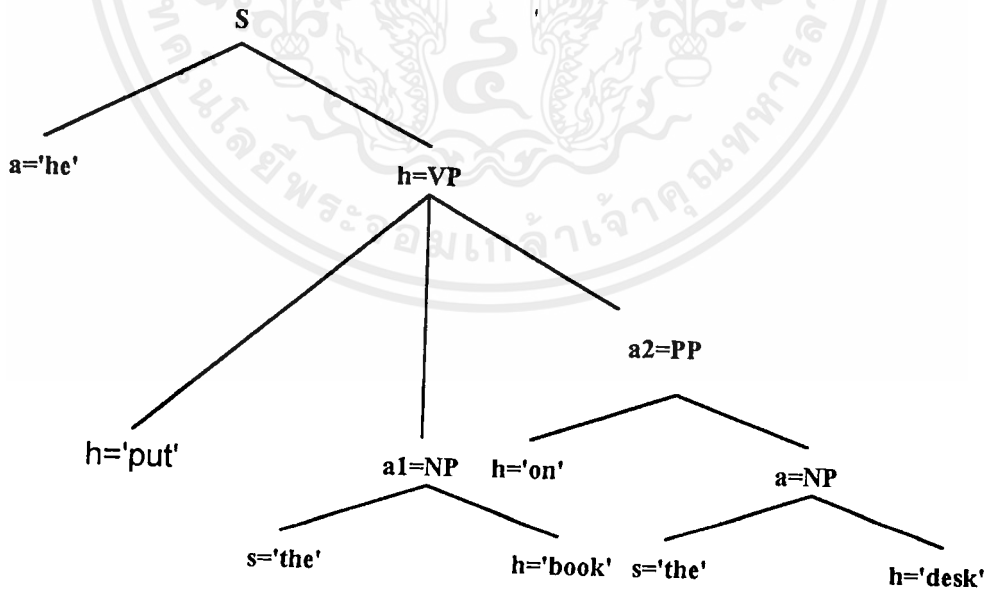
III : Target Generation

Gt : Target Grammar

AST : Annotation Syntax Tree

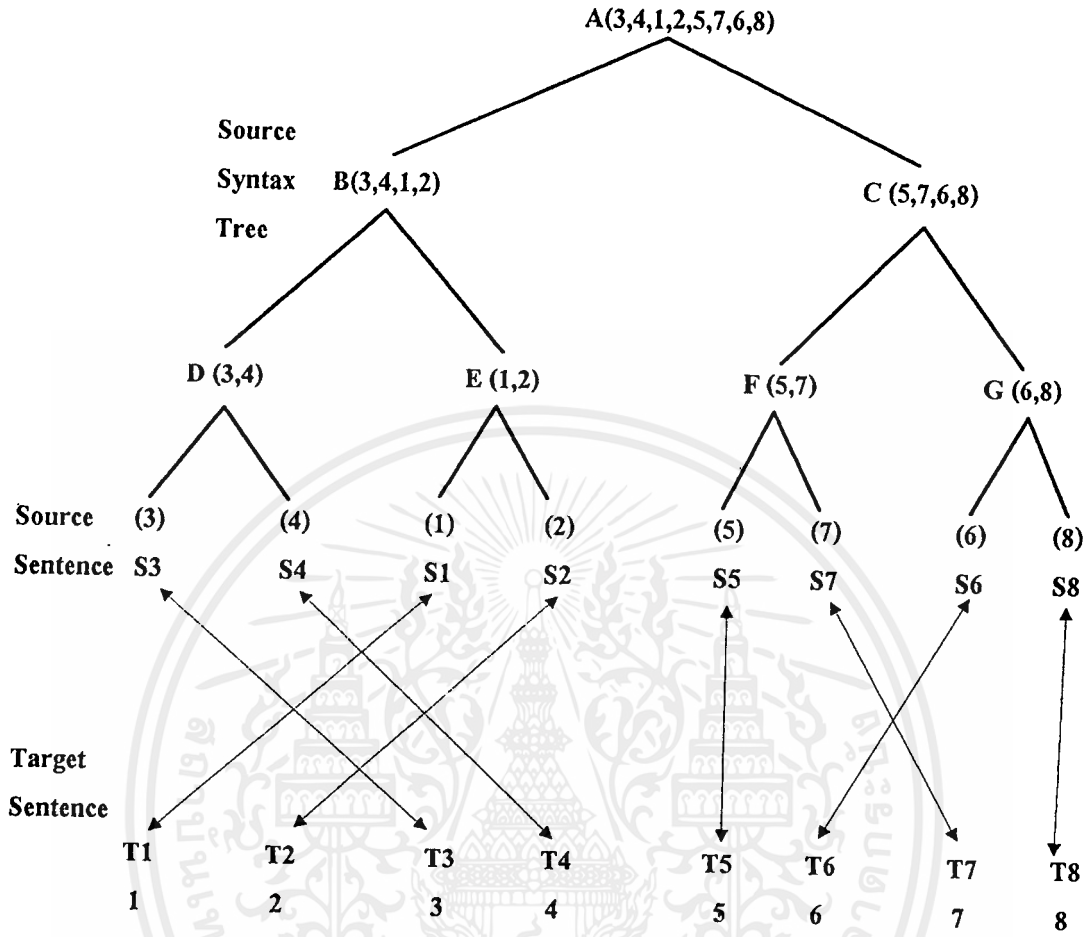
NF : Normal Form

รูปที่ 2.1

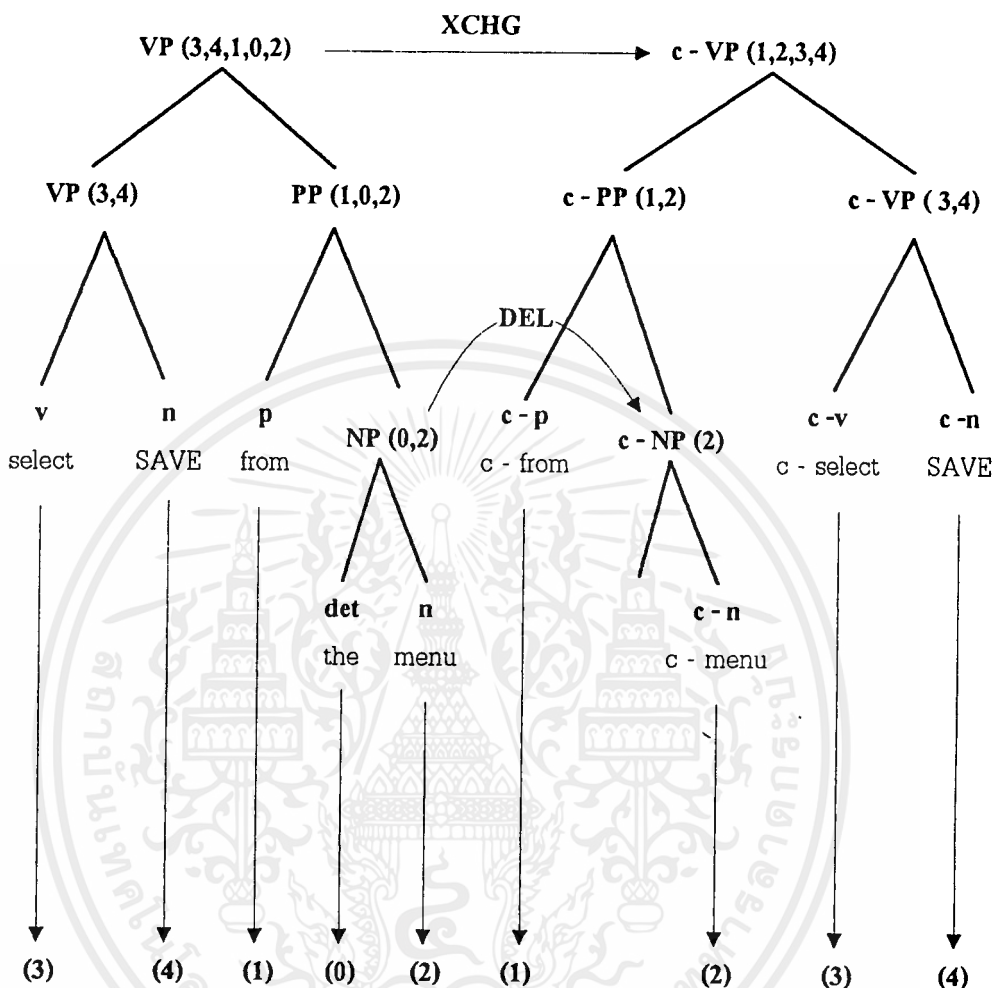


รูปที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 กฎเกณฑ์ควบคุมการสร้างภาษาเครื่องกับรายชื่อดรรชนีคำ
(Syntax tree with attached word index list)



รูปที่ 2.4 ตัวอย่างการแปลภาษาอังกฤษเป็นภาษาจีน

(An example of English Chinese Transfer process)

บทที่ 3

การเขียนและการทำงานของโปรแกรม

ซอฟต์แวร์ที่ทำการเขียนนี้จะอ้างอิงการทำงานบางส่วนของโปรแกรม โดยวิธีรวบรวมฐานข้อมูลชนิดเทียบเคียงตัวอย่าง เพื่อให้เหมาะสมกับภาษาที่ใช้ในการเขียนซอฟต์แวร์นี้จะไม่ใช้รูปแบบที่แท้จริงของการแปล โดยวิธีรวบรวมฐานข้อมูลชนิดเทียบเคียงตัวอย่าง จะเป็นเพียงตัวอย่างการแปลภาษาเท่านั้น

3.1 ขั้นตอนการทำงานของซอฟต์แวร์

ในการทำงานของซอฟต์แวร์นี้ จะมีขั้นตอนตามลำดับดังต่อไปนี้

1. รับประโยคที่จะทำการแปล
2. แยกประโยคออกเป็นคำและตัดคำที่ไม่ใช่ออก ตัวอย่างเช่น a, an, the
3. ทำการพิจารณาถึงชนิดของคำที่แยกออกมาในแต่ละคำ
4. ตรวจสอบว่า ประโยคนั้นมีคำใดที่ทำหน้าที่เป็นตัวเชื่อมหรือบ่งบอกว่ามีมากกว่า 1 ประโยค ในประโยคนั้น ถ้าไม่มีให้ไปทำต่อในขั้นตอนที่ 6 ถ้ามีให้ไปทำต่อในขั้นตอนที่ 5
5. ทำการแยกประโยคออกจากกัน นำประโยคที่แยกออกจากประโยคหลักไปใส่ไว้ใน temporary
6. ทำการหาประโยคตัวอย่างในฐานข้อมูล เพื่อมาเป็นตัวอย่างในการแปล
7. ทำการแปลตามตัวอย่าง
8. ตรวจสอบว่ายังมีประโยคหลงเหลือใน temporary หรือไม่ ถ้ามีไปขั้นตอนที่ 9 ถ้าไม่มีให้ไปทำในขั้นตอนที่ 10
9. นำประโยคจาก temporary มาใส่ในประโยคหลัก แล้วกลับไปยังขั้นตอนที่ 5
10. ออกจากโปรแกรม

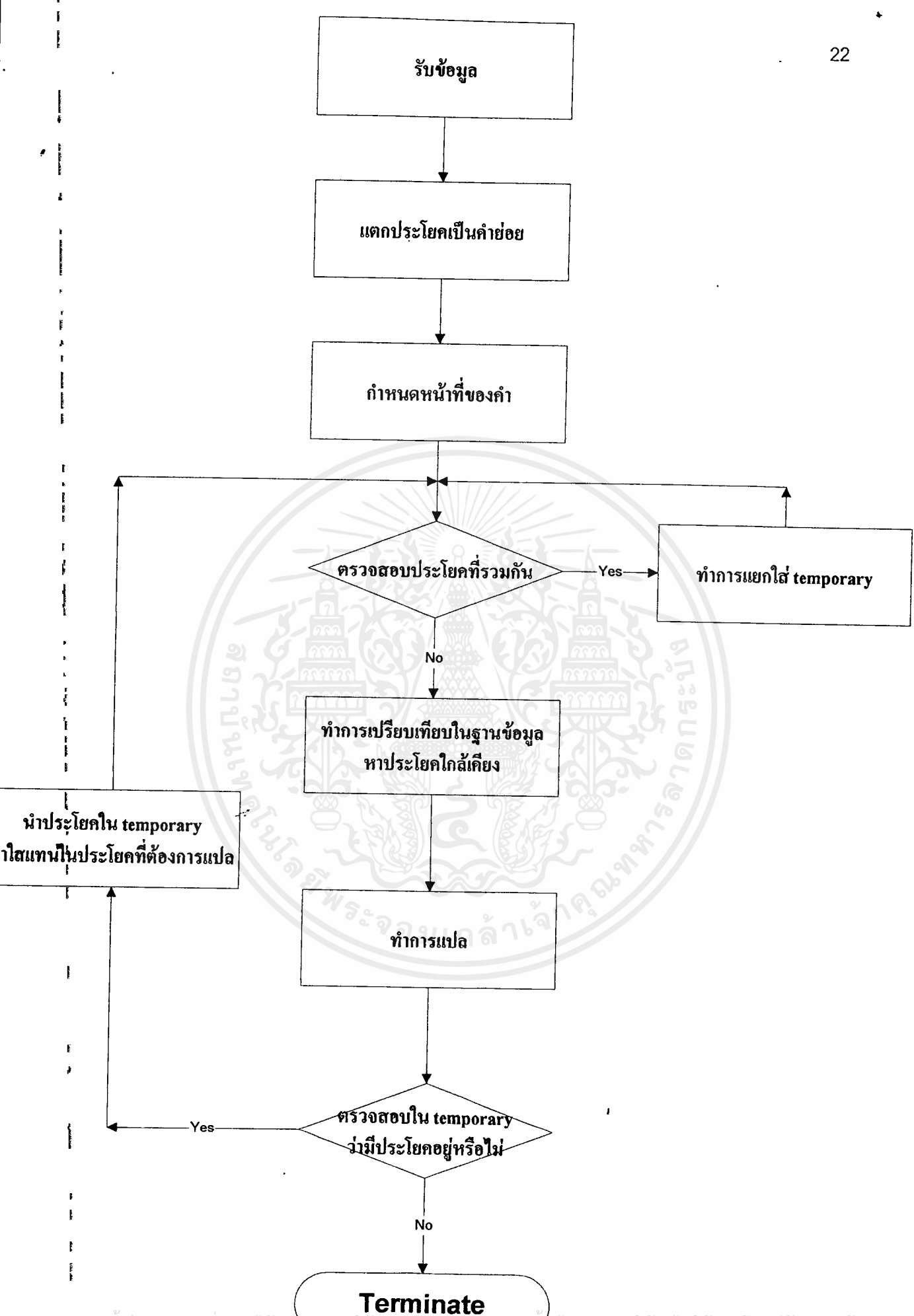
สำหรับขั้นตอนการทำงานของซอฟต์แวร์ ดูได้ดังรูปที่ 3.1

ในการเขียนซอฟต์แวร์นี้จะมีส่วนสำคัญที่ต้องทำการระบุว่า การทำงานของคำแต่ละคำเป็นลักษณะใด ไม่ว่าจะเป็นคำถาม, กริยา, กริยาช่วย, คำคุณศัพท์ ยกตัวอย่างเช่น

* boat เป็น คำนาม ประเภทยานพาหนะ

* cat เป็น คำนาม ประเภทสัตว์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้ภายในเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น ถ้าทั้งท่านมีให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
รูปที่ 3.1 ขั้นตอนการทำงานของซอฟต์แวร์ที่ใช้วิธีเทียบเคียงฐานข้อมูลชนิดเทียบเคียงตัวอย่าง

* love เป็น กริยา ประเภทความรู้สึก

โดยส่วนนี้จะช่วยในการแปลประโยคคำถามและมีหลายหน้าที่ หลายความหมาย ยกตัวอย่างเช่น “เขากินข้าว”

เราจะเห็นได้ว่าซอฟต์แวร์ไม่ว่าจะรู้ว่าเขานั้นหมายถึงอะไร อาจจะเป็น

1. เขา ที่มีความหมายว่าภูเขาเล็ก ๆ
2. เขา ที่เป็นบุรุษสรรพนาม
3. เขา ที่หมายถึง เขาว่า เขาควาย

เมื่อซอฟต์แวร์วิเคราะห์คำที่ตามมา จะได้ว่ากริยาที่ตามหลักคำนามนั้นเป็นกริยาประเภทการกระทำ เพราะฉะนั้นคำนามที่นำหน้ากริยานี้ จะต้องเป็นคำนามที่เป็นสิ่งมีชีวิตหรือเป็นบุรุษสรรพนาม ดังนั้น เขาในที่นี้จะหมายถึงเขา ที่เป็นบุรุษสรรพนาม ซอฟต์แวร์จะทำการแปล “เขากินข้าว” ได้เป็น “They eat rice.”

3.2 ตัวอย่างการทำงานของซอฟต์แวร์

ต่อไปนี้จะเป็นการยกตัวอย่างประโยคที่ทำการแปลในซอฟต์แวร์

Giant waves washed over the boat.

เมื่อซอฟต์แวร์ทำการรับข้อความนี้เข้ามา จะทำการแยกคำได้เป็น

Giant	waves	Washed	over	The	boat
-------	-------	--------	------	-----	------

ซอฟต์แวร์จะทำการตัดคำที่ไม่ใช้ในภาษาไทยออก ซึ่งในที่นี้คือ the จะได้ว่า

Giant	waves	Washed	over	Boat
-------	-------	--------	------	------

จากนั้นทำการกำหนดหน้าที่ของแต่ละคำ จะได้ว่า

Giant	waves	Washed	over	Boat
adj.	Noun natural	verb	adv.	Noun transport

จากนั้น ทำการหาลักษณะการแปลของประโยค จะได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

adj., noun natural, verb, adv., noun transport



noun natural, adj., verb, adv., noun transport

เมื่อหาได้แล้วทำการเปลี่ยนตำแหน่งคำตามที่ได้ แต่ไม่เปลี่ยนตำแหน่งของชนิดของคำและ
ใส่ค่าคำที่ต้องการจะแปล ถ้าเป็น 1 ให้ทำการแปล ถ้าเป็น 0 ไม่ต้องทำการแปล จะได้

Giant	waves	Washed	over	Boat
adj.	Noun natural	verb	adv.	Noun transport
1	1	1	1	1

ทำการแปลตามคำศัพท์ จะได้

Giant	waves	Washed	over	Boat
adj.	Noun natural	verb	adv.	Noun transport
1	1	1	1	1
คลื่น	ยักษ์	สาด	เหนือ	เรือ

ทำการแสดงผลออกทางหน้าจอ

อันที่จริงแล้วในการใช้วิธีนี้จะทำให้ภาษานั้นสะดวกมากกว่าวิธีอื่น ๆ จะเห็นได้จาก
ประโยคบางประโยคที่ทำการแปล ยกตัวอย่างเช่น

gulliver turned his head to look around

ถ้าทำการแปลโดยทำการแปลแบบตรงตัวจะได้

กาลิเวอร์หมุนหัวของเขาเพื่อมอรอบ ๆ

ถ้าทำการแปลโดยทำการแปลแบบเที่ยงเคียง จะได้

กาลิเวอร์หันหน้าของเขาเพื่อมองดูไปรอบ ๆ

ทั้งนี้เนื่องจากเราทำการแปลโดยเทียบเคียงกับฐานข้อมูลจำนวนมาก ในเมื่อฐานข้อมูลนั้น บรรจรูปแบบของประโยคหลายแบบ จนเกือบทุกรูปแบบของประโยคของภาษาที่ทำการแปล จะทำให้เราสามารถทำการแปลได้ตามที่เราต้องการ

3.3 ข้อดีและข้อเสียของการแปลด้วยวิธีนี้

ข้อดีของการแปลด้วยวิธีนี้ คือ

1. ทำให้ภาษาที่ใช้กับประโยคที่เราแปลออกมา ใช้คำได้สละสลวยกว่าวิธีอื่น ๆ
2. การแปลด้วยวิธีนี้ จะสามารถทำการแปลคำกำกวม ประโยคสำนวน คำลักษณะแปลก ๆ ได้
3. วิธีนี้เป็นวิธีที่มีโอกาสแปลได้อย่างสมบูรณ์ที่สุด ถ้าเราสามารถเก็บข้อมูลได้ครบทุกรูปแบบประโยค
4. ถ้าประโยคที่ทำการแปล ขอเพียงมีลักษณะ โครงสร้างใกล้เคียงกับประโยคที่มีอยู่ในฐานข้อมูล โปรแกรมจะสามารถสร้างประโยคในภาษาผลลัพธ์ได้ออกมาเป็นอย่างดี

ข้อเสียของการแปลด้วยวิธีนี้ คือ

1. เนื่องจากทำงานกับฐานข้อมูลขนาดใหญ่ จึงทำให้ใช้ทรัพยากรเปลืองมาก
2. ถ้าฮาร์ดแวร์ไม่ดีนั้น จะทำให้การแปลใช้เวลานาน (ดังนั้นถ้าจะใช้วิธีนี้ควรจะใช้ฮาร์ดแวร์ที่มีความเร็วทางด้านการสืบค้นข้อมูลพอสมควร)
3. จะต้องทำการเขียน code ขึ้นมาเป็นจำนวนมาก เพื่อที่จะทำการแปลในแต่ละประโยค
4. ถ้าตัวอย่างของประโยคไม่มี ประโยคนั้นจะไม่สามารถทำการแปลได้เลย

3.4 ปัญหาและแนวทางแก้ไข

ปัญหาที่เกิดขึ้นกับการแปลนี้ อาจเกิดจาก

1. คำที่มีหลายความหมาย
2. คำที่เป็นภาษาแสลง
3. คำที่เป็นประโยคคำถาม และสามารถเป็นคำเชื่อมในเวลาเดียวกัน
4. ประโยคสำนวน

ตัวอย่างของปัญหาพร้อมทั้งวิธีแก้ไข

1. คำที่มีหลายความหมาย

ในลักษณะคำที่มีหลายความหมาย ยกตัวอย่างง่าย ๆ เช่น They ในประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The football players walk along the road.

They try to climb to the top of the mountain

กับประโยค

The ant army try to catch a food.

They use their needle to kill their enemy.

จะเห็นว่าทั้งสองประโยคใช้ They เหมือนกัน แต่ในประโยคแรก They จะหมายถึง พวกเขา แต่ในประโยคที่สอง They จะหมายถึง พวกมัน

การที่จะหาว่าในแต่ละประโยค จะใช้ความหมายอะไรนั้น เราจะต้องอ้างอิงกับประโยคก่อนหน้า

ดังนั้นเราจึงจำเป็นต้องเก็บค่าทั้งประธานและกรรมของประโยคก่อนหน้าไว้ใช้

ในกรณีนี้เราอาจจะใช้ตัวแปรบางตัวมาเป็นตัวช่วยในการเก็บค่า

โดยจะต้องทำการแยกค่าของประธาน และกรรม เป็นสองค่า เมื่อได้เช่นนี้เราจะสามารถระบุได้ว่า ประโยคถัดไปที่เรากำลังทำการแปลอยู่นั้นจะใช้ตัวแปลตัวไหน ส่วนคำศัพท์ที่มีหลายความหมาย แต่อยู่ในตำแหน่งที่ไม่เหมือนกันจะสามารถทำการแปลได้อยู่แล้ว ถ้าเรามีรูปประโยคนั้นอยู่ในฐานข้อมูลอยู่แล้ว

2. คำที่เป็นภาษาแสดง

ในคำศัพท์ที่เป็นภาษาแสดงที่ใช้ ตัวอย่างเช่น คำศัพท์หรือรูปแบบของประโยคที่ใช้กันเฉพาะในภาษาพูด เช่น What is up, Not much จะเห็นได้ว่า คำแสดงเหล่านี้เป็นการผสมคำที่มีอยู่แล้วมาเป็นภาษาใหม่ ในกรณีนี้เราสามารถที่จะกำหนดเฉพาะแบบลงไปได้

อีกกรณีหนึ่ง เช่น Atta boy จะเป็นคำแสดงที่ไม่ได้เกิดจากคำศัพท์ที่มีอยู่แล้ว เพราะฉะนั้นในกรณีนี้ จะต้องทำการตรวจสอบลักษณะคำศัพท์กับผู้ใช้ แล้วใส่ลงไปฐานข้อมูล

3. คำที่เป็นประโยคคำถามและสามารถเป็นคำเชื่อมในเวลาเดียวกัน

ในคำลักษณะนี้ ตัวอย่างเช่น When, Where, Why เราจะเห็นได้ว่า คำศัพท์เหล่านี้สามารถเป็นได้ทั้งประโยคคำถามและตัวเชื่อมประโยคในเวลาเดียวกัน เช่น

When did he go to school.

กับประโยค

when he wake up, he was take a bath.

ในประโยคแรกนั้น คำว่า **When** จะทำหน้าที่เป็น Question Tag ซึ่งมีความหมายว่า “เมื่อไหร่”

ในประโยคที่สอง คำว่า **when** จะทำหน้าที่เป็น conjunction ซึ่งมีความหมายว่า “เมื่อ” จะเห็นได้ว่าคำสองคำนี้เป็นคำเดียวกัน แต่หลายหน้าที่และหลายความหมาย การที่เราจะกำหนดว่าแต่ละคำแปลได้อย่างไรนั้น เราจำเป็นจะต้องมีข้อมูลส่วนหนึ่งของประโยคที่มีทั้งเป็น Question Tag และเป็น conjunction เพื่อทำการเทียบเคียง

4. ประโยคสำนวน

ประโยคสำนวน ตัวอย่างเช่น

Like a piece of cake.

อาจจะแปลได้ว่า ชอบเด็กเป็นจีน หรืออาจจะแปลได้ว่า ง่ายเหมือนปอกกล้วยเข้าปาก รูปแบบประโยคเหล่านี้เราจะต้องทำการตรวจสอบหาให้ได้ทำการแปลก่อนวิธีอื่น โดยถ้าเจอลักษณะที่เหมือนกับประโยคสำนวนเหล่านี้จะต้องทำการกำหนดให้เป็นคำศัพท์แค่ชุดเดียว ยกตัวอย่างเช่น

This examination is easy like a piece of cake.

จะทำการกำหนดให้ **like a piece of cake** เป็นแค่ **adverb** ตัวเดียว แล้วทำการแปลแยกกับคำอื่น

บทที่ 4

บทวิจารณ์และสรุป

จากที่ได้ทำการศึกษากการแปลภาษา มีอุปสรรคอยู่บ้างเล็กน้อย เนื่องจากมีผู้ทำการศึกษาน้อยมากในประเทศไทย จึงเป็นการยากที่จะหาแหล่งข้อมูลที่สมบูรณ์ได้ โดยเฉพาะการแปลโดยวิธีรวบรวมฐานข้อมูล ยังไม่มีผู้ใดได้ทำการศึกษาสำเร็จในประเทศไทย จึงเป็นการยากอย่างยิ่งที่จะหาตัวอย่างและข้อมูล

แต่ก็ได้ทำการศึกษาโดยสังเกตเพื่อเป็นแนวทางสำหรับผู้ที่จะทำการศึกษาต่อไป ในส่วนของตัวอย่างและซอฟต์แวร์ที่ได้ทำออกมาถือว่าสำเร็จสมบูรณ์ตามที่ไดวางเป้าหมายไว้

ในงานนี้ ถ้าผู้ใดจะทำการศึกษาต่อ ควรจะมีข้อมูลและพื้นฐานในวิชาปัญญาประดิษฐ์ (Artificial Intelligence : AI) การประมวลผลภาษาทั่วไป (Natural Language Processing) และเครื่องแปลภาษา (Machine Translation)

งานวิจัยนี้ได้หวังว่าจะได้นำไปใช้ประโยชน์เป็นพื้นฐานและตัวอย่างในการพัฒนาต่อไปในอนาคต

ในส่วนการเขียนโปรแกรมนั้น ในระหว่างทำการเขียนได้เกิดปัญหาในการหาหน้าที่ของคำและความหมายที่ตรงตามลักษณะประโยคที่ได้มา เพราะจริง ๆ แล้วความหมายของคำที่เราพูดกันหรือใช้กันโดยทั่วไปนั้น จะไม่ตรงกับประโยคที่เราอ่าน ดังนั้นถ้าต้องการให้ทำการแปลออกมาได้อย่างสวยงามจะต้องใช้การเทียบเคียงเกือบทุกประโยคและตัวอย่างจะต้องมีมาก ๆ ในการเขียนซอฟต์แวร์ครั้งนี้ไม่ได้ทำการใส่การวัดค่าของ Transfer Score ไว้ เพราะสิ่งที่วัด คือ ตัวผู้เขียนเอง นอกจากนี้จะต้องทำการเขียนโปรแกรมให้เทียบเคียงกับรูปแบบของวิธีที่ทำการศึกษา ซึ่งวิธีนี้จะเป็นการยากเมื่อใช้ภาษาระดับสูงเขียนเนื่องจากการเทียบระหว่างประโยคกับประโยค ดังนั้นผู้เขียนจึงได้ดัดแปลงลักษณะอัลกอริทึมไป โดยจะใช้การไล่ไปในลักษณะของชนิดของคำแทน ซึ่งวิธีนี้จะทำได้ง่ายกว่า และสามารถที่จะใช้แทนวิธีการเทียบเคียงฐานข้อมูลโดยตรงได้

ภาคผนวก ก.

เนื้อเรื่องและลักษณะของคำในเรื่อง Gulliver

เนื้อเรื่อง Gulliver

one night gulliver was at sea in his little sailboat.

Suddenly a terrible storm struck.

Giant waves washed over the boat.

Lightning flashed across the sky.

Thunder roared Boom!crash!

all at once a huge wave swept over the boat.

It tossed gulliver high into the air and threw him across the water.

The next thing he knew, he was rolling head over heels on a sandy beach.

By this time gulliver was so tired he fell asleep in the wet sand.

When gulliver opened his eyes the next morning.

He did not know where he was.

He tried to sit up, but he could not move.

I must be dreaming

thought gulliver.

He wiggled his arm.

They were tied to the ground by tiny ropes.

He wiggle his legs.

They were tied to the ground, too.

Then gulliver turned his head to look around.

An army of tiny soldiers was marching toward him.

Some sat on tiny horses and carried tiny swords.

Others were pulling little cannons over the sand.

A few of the soldiers crept right up to gulliver.

They crawled into his pockets and pulled out his watch.

Then they opened his watch.

Help!

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Cried a little old man whose beard had caught in the wheels.

The soldiers had to chop off his long white beard.

Now gulliver wiggled as hard as he could.

The tiny ropes broke and he sat up.

All at once a hundred tiny cannonballs hit his chest

ping!

Ping!

But they were too small to hurt gulliver.

They bounced off his chest and rolled back to the soldiers.

A fat little man on a fine brown horse began to shout

load the cannons and fire again!

That must be the leader

thought gulliver.

He lifted him ringt off the ground, horse and all.

The angry little man pulled out his sword.

Put me down this minute!

He said.

I am the emperor of littiput.

Gulliver smiled at the tiny emperor.

Don't beafraid of me, your highness,

said gulliver,

I do not want to hurt you.

Then he told the emperor about the strom that had blown him to lilliput.

All I want now

he said

is something to eat.

Gulliver put the little man back on the ground.

Bring food for the giant!

The emperor shouted.

He will not hurt us.

He is a friend.

The little soldiers hurried away.

They knocked on every door in lilliput.

they cried.

“bring all the food you can spare!”

soon the lilliputians came with food for gulliver.

They brought wagons full of cheese and baskets of bread.

They had carts filled with fruit and barrels of milk.

Gulliver scooped up all the food as the people passed by.

He ate everything.

At first the children were afraid of gulliver.

But they soon found out that he was a friend.

Gulliver played with the children every day.

They made seesaws across his shoes and played leapfrog over his fingers.

They played hide-and-peek in his pockets.

They climbed up to his knees and slid down his legs.

On morning gulliver went to the beach to sit by himself and think.

He had been in lilliput for many weeks.

And the lilliputians were very kind to him.

Still, gulliver was lonely.

He thought of his own home far across the sea.

He thought about all his old friends.

How I wish I could see them again

he said.

Suddenly there was a shout.

Ahoy!

Cried the look out on the castle tower.

Enemy boats headed this way!

The emperor’s soldiers got busy once.

Some of them jumped on their horses and pulled out their swords.

Others rolled out the cannons and loaded them.

Gulliver heard the shouting.

Let me help!

He cried.

And he waded out to meet the enemy boats.

The little boats came closer and closer.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ They began to fire their cannons. ไม่นานญาติให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Ping!

Ping!

The tiny cannonballs bounced off gulliver's chest.

Gulliver took a deep, deep breath.

Then, with all his might...whoosh!

He blew the enemy boats ringt out of singt.

When gulliver came back the lilliputians cheered.

You have saved our country

said the emperor.

Now we must do something for you.

What do you want more than anything else in the world?"

you have all been very kind to me

said gulliver.

But still I miss my own country.

More than anything else I want to go home

perhaps we can help you

said the emperor.

For two whole days the lilliputians worked.

They choopped down trees and dragged them to the beach.

They mended ropes.

They split logs.

They hammered and measured and sawed.

They sewed dozns of sheets together to make a giant sail.

Then they put the sil on a long pole and raised it high into the air.

When they were all done

they showed gulliver what they had built for him.

It was a raft!

Now he had a way to get home again.

Oh!, thank you

said gulliver.

The lilliputains filled their rowboats with barrels of ffod and water.

They loaded everything onto the raft and gulliver climbed aboard.

The tiny emperor stood on the shore.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการ We will miss gulliver ท่านั้น ไม่นุญตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

he thought.

As gulliver sailed out to sea

he waved good-by the lilliputians.

He was sorry to leave his new friends.

But he was happy to be on his way home at last.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การหาความหมายของคำแต่ละคำในเรื่อง Gulliver

one night

number ntime

gulliver was at sea in his little saiboat.

person be prep nplace nperson adj ntransport

suddenly

adv

a terrible storm struck.

adj nnet v

giant waves washed over the boat.

adj nnet v prep ntransport

lightning flashed across the sky.

nnet v adv nplace

thunder roared.

Nnet v

Boom!

!

Crash!

!

all at once

adv prep adv

a huge wave swept over the boat.

adj nnet v prep ntransport

it tossed gulliver high into the air

naot v person adv prep ntransport

and threw him across the water.

Con v nperson prep nnet

the next thing he knew

adj net nperson v

he was rolling head over heels on a sandy beach.

Nperson be v nbody prep nbody prep nt nplace

by this time

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

prep adj ntime

gulliver was so tired

person be adv adj

he fell asleep in the wet sand.

Nperson adj adv prep adj nt

the gulliver opened his eyes the next morning

con person v Nperson nbody adj ntime

he did not know where he was

Nperson aux adv v con Nperson be

he tried to sit up

Nperson v prep v adv

but the could not move.

con Nperson aux adv v

I must be dreaming

Nperson aux be n

thought gulliver.

v person

he wiggled his arm.

Nperson v Nperson nbody

they were tied to the ground too.

Nperson v Nperson nbody

they were tied to the ground too.

Nperson be v adv n adv

then gulliver turned his head to look around.

adv Nperson v Nperson nbody prep v adv

an army of tiny soldiers was marching toward him.

Nperson prep adj person be v prep pron

some sat on tiny horses

pron v prep adj n

and carried tiny swords.

con v adj nt

others were pulling little cannons over the sand.

Porn be y adj nt prep nt

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

a few of the soldiers crept right up to gulliver.

Adj prep person v adv adv prep person

they crawled into his pockets

nperson v prep nperson nt

and pulled out his watch.

Con v adv nperson nt

then they opened his watch.

Adv nperson v nperson nt

help!

!

cried a little old man whose beard had caught in the wheels.

V adj adj nperson pron nbody aux v prep n

the soldiers had to chop off his long white beard.

Nperson aux adv v adv nperson adj adj nbody

now gulliver wiggled as hard as he could.

adv nperson v adj adj adj nperson v

the tiny ropes broke and he sat up

Adj n v con nperson v adv

all at once

adv prep adv

a hundred tiny cannonballs hit his chest

number adj n b nperson nbody

ping!

!

ping!

!

but they were too small to hurt gulliver.

Con nperson be adv adj prep v nperson

they bounced off his chest

nperson v adv nperson nbody

and rolled back to the soldiers

con v adv prep nperson

a fat little man on a fine brown horse began to shout

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

adj adj nperson prep adj adj na v prep v

load the cannons and fire again

v n con v adv

that must be the leader

con aux be n

thought gulliver.

V nperson

he lifted him right off the ground

nperson v pron adv adv n

horse

na

and all

Con adv

the angry little man pulled out his sword.

Adj adj nperson v adv nperson n

put me down this minute

v pron adv adj n

he said.

Nperson v

I am the emperor of lilliput

nperson be nperson prep nplace

gulliver smiled at the tiny emperor.

Nperson v prep adj nperson

don't be afraid of me

aux be v prep nperson

your highness

nperson n

said gulliver

V nperson

I do not want to hurt you

nperson aux adv v prep v pron

them he told the emperor about the storm that had blown him to lilliput.

Con nperson v nperson con n con aux v pron prep nplace

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

all I want now

adv nperson v adv

he said

nperson v

is something to eat

be nt prep v gulliver

gulliver put the little man back on the ground.

Nperson v adj nperson adv prep n

bring food for the giant

V n prep n

the emperor shouted.

Nperson v

he will not hurt us.

Nperson aux adv v pron

he is a friend.

Nperson be n

the little soldiers hurried away.

Adj nperson v adv

they knocked on every door in lilliput.

Nperson v pre adj n prep nplace

a giant has landed on the beach

n aux v prep nplace

they cried.

Nperson v

bring all the food you can spare

v adj n nperson aux v

soon the lilliputians came with food for gulliver.

adv nperson adv n prep nperson

They brought wagons full of cheese

Nperson v n adj n

and baskets of bread.

Con n prep n

they had carts filled with fruit

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

nperson v n v prep n

and barrels of milk.

Con n prep n

Gulliver scooped up all the food as the people passed by.

Nperson v adv adj n prep n v adv

He ate everything.

Nperson v nt

At first the children were afraid of gulliver.

Prep number nperson be a aux nperson

But they soon found out

con nperson adv v adv

that he was a friend.

Nperson be n

Gulliver played with the children every day.

Nperson v prep nperson adj ntime

They made seesaws across his shoes

nperson v n adv nperson n

and played leapfrog over his fingers.

Con v n adv nperson n

They played hide-and-seek in his pockets.

Nperson v v con v prep nperson n

They climbed up to his knees

nperson v adv prep nperson nbody

and slid down his legs.

Con v adv nperson nbody

On morning gulliver went to the beach to sit by himself

prep ntime nperson v prep nt prep v prep nperson

and think.

Con v

He had been in lilliput for many weeks.

Nperson aux be prep nplace pre adj ntime

And the lilliputians were very kind to him.

Con nperson be adv adj prep pron

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Still, gulliver was lonely.

Adv nperson be adv

He thought of his own home far across the sea.

Nperson v prep nperson adj nplace adj prep n

He thought about all his old friends.

Nperson v prep adv nperson adj nperson

How I wish I could see them again

qt nperson v nperson aux v nperson adv

he said.

Nperson v

Suddenly

adv

there was a shout.

Nperson be v

Ahoy!

!

Cried the look out on the castle tower.

V nperson prep nplace nplace

Enemy boats headed this way

nperson ntransport v adv adv

The emperor's soldiers got busy once.

Nperson nperson v adv adv

Some of them jumped on their horses

adj prep nperson v prep nperson na

and pulled out their swords.

Con v adv nperson nt

Others rolled out the cannons

adj v adv nt

and loaded them.

Con v nperson

Gulliver heard the shouting.

Nperson v n

. Let me help

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

v pron v

He cried.

Nperson v

And he waded out to meet the enemy boats.

Con nperson v adv prep v nperson ntransport

The little boats came closer

adj ntransport a adv

and closer

con adv

They began to fire their cannons.

Nperson v pre v nperson nt

Ping!

!

Ping!

!

The tiny cannonballs bounced off gulliver's chest.

Adj nt v adv nperson nbody

Gulliver took a deep

nperson v adj

deep breath.

Adj v

Then

con

with all his might

con adj nperson aux

whoosh!

!

He blew the enemy boats ringt out of singt.

Nperson v nperson nperson adv adv prep adv

When gulliver came back the lilliputians cheered.

Con nperson v adv nperson v

You have saved our country

nperson aux v person nplace

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

said the emperor.

V nperson

Now we must do something for you.

Adv nperson aux v nt prep nperson

What do you want more than anything else in the world?"

qt aux nperson v adv prep nt adv prep nplace

you have all been very kind to me

nperson aux adv be adv adj prep nperson

said gulliver.

v nperson

But still I miss my own country.

Con adv nperson v nperson adj nplace

More than anything else I want to go home

adv prep nt adv nperson v prep v nplace

perhaps we can help you

adv nperson aux v nperson

said the emperor.

V nperson

For two whole days the lilliputians worked.

Prep number adj ntime nperson v

They chooped down trees

nperson v adv nt

and dragged them to the beach.

Con nt prep nplace

They mended ropes.

Nperson v nt

They split logs.

Nperson v nt

They hammered

nperson v

and measured

Con v

and sawed.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Con v

They sewed dozns of sheets together to make a giant sail.

Nperson v adj prep nt adv prep v adv nt

Then they put the sil on a long pole

prep nperson v nt prep adj nt

and raised it high into the air.

Con v naot adv prep nnet

When they were all done

con nperson be adv v

they showed gulliver what they had built for him.

Nperson v nperson con nperson aux v prep nperson

It was a raft

naot be nt

Now he had a way to get home again.

Prep nperson v adv prep v nplace adv

Oh!

!

thank you

v nperson

said gulliver.

v nperson

The lilliputains filled their rowboats with barrels of ffood

nperson v nperson ntransport con nt prep nt

and water.

Con nt

They loaded everything onto the raft and gulliver climbed aboard.

Nperson v nt prep nt con nperson v adv

The tiny emperor stood on the shore.

Adj nperson v prep nplace

We will miss gulliver

nperson aux v nperson

he thought.

Nperson v

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

As gulliver sailed out to sea

adv nperson v adv prep nplace

he waved good-by the lilliputians.

Nperson v adv prep nperson

He was sorry to leave his new friends.

Nperson be adv prep v nperson adj nperson

But the was happy to be on his way home at last.

Con nperson ve adv prep be prep nperson adv nplace prep adv



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



- [1] Bennett, W.S. and J. Slocum, 1985. "The LRC Machine Translation System," Computational Linguistics, vol. 11, no. 2-3 pp. 111-119. 1985.
- [2] Brown, P.F., et al. 1990. "A Statistical Approach to Machine Translation," Computational Linguistics, vol. 16 no.2, pp. 79-85. 1990.
- [3] Chang, J. -S., Y.-Fl. Luo and K.-Y. Su, 1992. "GPSM: A Generalized Probabilistic Semantic Model for Ambiguity Resolution," Proceedings of ACL-92, pp. 177-184, 30th Annual Meeting of the Association for Computational Linguistics, University of Delaware, Newark, DE, USA, 28 June-2 July, 1992.
- [4] Chen, S.-C., J-S. Chang, J.-N. Wang, and K.-Y. Su, 1991. "ArchTran: A Corpus-Based Statistics-Oriented English-Chinese Machine Translation System," Proceeding of Machine Translation Summit III, pp. 33-40, Washington, D.C., USA, July 1-4, 1991.
- [5] Jing-Shin Chang and Keh-Yih Su., 1993. "Proceedings of the Fifth National Conference on Theoretical and Messadological Issue in MT TMI 93," A Corpus-Based Statistics-Oriented Transfer and Generation Model for Machine Translation, pp. 3-13, Department of Electrical Engineering, National Tsing-Hua University, Hsinchu, Taiwan, R.O.C., 1993.
- [6] Walt Disney's, 1975. "Gulliver Mickey" Book Club Edition, pp. 38, 1975.