โปรแกรมแปลงภาษากลางเป็นภาษาไทย

นาย สมชาย กิติภรณ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2543
ISBN 974-346-507-3
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A UNL TO THAI DECONVERTER PROGRAM

Mr. Somchai Kitiporn

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2000

ISBN 974-346-507-3

หัวข้อวิทยานิพนธ์	โปรแกรมแปลงภาษากลางเป็นภาษาไทย
โดย	นาย สมชาย กิติภรณ์
ภาควิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.วันชัย ริ้วไพบูลย์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สมชาย ประสิทธิ์จูตระกูล)

..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.วันชัย ริ้วไพบูลย์)

..... กรรมการ
(รองศาสตราจารย์ มัณฑนา ปราการสมุทร)

..... กรรมการ
(อาจารย์ ดร.บุญเสริม กิจศิริกุล)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สมชาย กิติภรณ์ : โปรแกรมแปลงภาษากลางเป็นภาษาไทย
(A UNL TO THAI DECONVERTER PROGRAM)
อ.ที่ปรึกษา : รศ.ดร.วันชัย ริ้วไพบูลย์. 166 หน้า. ISBN 974-346-507-3.

วิทยานิพนธ์ฉบับนี้นำเสนอแนวคิดและขั้นตอนวิธีการแปลงจากภาษา UNL (Universal Networking Language) ซึ่งเป็นภาษาทางปัญญาประดิษฐ์ให้เป็นภาษาไทยโดยอ้างอิงภาษา UNL ตามมาตราฐานของ UNU/IAS แนวคิดขั้นตอนการแปลงนั้นได้ดัดแปลงจากระบบการสร้างภาษาไทยจากภาษากลาง (Interlingua) และใช้วิธีการออกแบบฐานข้อมูลพจนานุกรมรวมไปถึงการใช้วิธีการโปรแกรมมิ่งเข้ามาใช้ในการแปลภาษาโดยได้จัดทำโปรแกรมต้นแบบในการแปลงใช้ชื่อว่า "CU-Deconverter" แนวความคิดในการแปลงจากภาษา UNL เป็นภาษาไทยที่นำเสนอแบ่งออกเป็น 6 ขั้นตอนคือ (1) ตรวจสอบความถูกต้องของเอกสารภาษา UNL ที่เป็นต้นฉบับ, ตรวจคำศัพท์ต่างๆ และแสดงผลความผิดพลาด (2) หาความหมายของคำและข้อมูลประกอบทางด้านภาษาไทยจากฐานข้อมูลพจนานุกรม (3) เปรียบเทียบรูปแบบกริยาและดึงค่าของหน้าที่ของคำจากฐานข้อมูล (4) จัดเรียงคำตามค่าที่ได้จากขั้นตอนที่ 3 (5) จัดรูปแบบประโยคและตัดคำซ้ำซ้อนกันและ (6) แสดงผลในรูปแบบประโยคภาษาไทย

ฐานข้อมูลพจนานุกรมที่ได้มีการออกแบบในวิทยานิพนธ์ฉบับนี้ประกอบด้วยตารางต่างๆ ได้แก่ (1) ตารางแสดงลักษณะเฉพาะของคำ (2) ตารางคำสากล (3) ตารางประเภทคำ (4) ตารางแสดงความสัมพันธ์ (5) ตารางแสดงคุณลักษณะคำ (6) ตารางพจนานุกรมยูเอ็นแอล (7) ตารางหน้าที่ของคำ และ (8) ตารางรูปแบบกริยา ในเบื้องต้นได้บรรจุข้อมูลต่างๆ เพื่อใช้ในการแปลงจากภาษา UNL เป็นภาษาไทยแล้วบางส่วน ทั้งนี้ในกรณีที่ไม่พบคำศัพท์, รูปแบบกริยาหรือข้อมูลต่างๆ ก็สามารถเพิ่มคำศัพท์หรือรูปแบบกริยาเข้าไปในฐานข้อมูลเพื่อใช้ในการแปลงภาษาต่อไป โปรแกรมขั้นตอนวิธีการแปลงภาษานี้จะทำงานโดยใช้รูปแบบการติดต่อ ทั้งในแบบของวินโดวส์และแบบเวบโดยผู้ใช้สามารถระบุแฟ้มข้อมูลหรือเวบเพจที่ต้องการแปลงได้

เอกสารของภาษา UNL ที่ใช้ในการทดสอบนี้ได้นำตัวอย่างมาจากหนังสือ UNL ของ UNU/IAS โดยอยู่ในรูปแบบแฟ้มข้อมูล HTML ในแฟ้มข้อมูลจะประกอบด้วยคำสากล, รูปแบบการเชื่อมประโยคและข้อมูลประกอบสำหรับใช้ในการอธิบายคำสากลเพิ่มเติม ผลการทดสอบได้ผลออกมาเป็นภาษาไทยตามความหมายที่ถูกต้องตรงกับรูปประโยคต้นฉบับประมาณ 85%

ภาควิชา .....วิศวกรรมคอมพิวเตอร์..... ลายมือชื่อนิสิต.....
สาขาวิชา .....วิทยาศาสตร์คอมพิวเตอร์... ลายมือชื่ออาจารย์ที่ปรึกษา .....
ปีการศึกษา ...2543..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

407-14884-21 : MAJOR COMPUTER SCIENCE

KEY WORD : MACHINE TRANSLATION, INTERLINGUA,
UNL (UNIVERSAL NETWORKING LANGUAGE), DECONVERTER

SOMCHAI KITIPORN : A UNL TO THAI DECONVERTER PROGRAM
THESIS ADVISOR : ASSOC. PROF. WANCHAI RIVEPIBOON,Ph D.

166 pp. ISBN 974-346-507-3

This thesis presented thinking method and procedure to convert UNL (Universal Networking Language), an intellectual invention language, to Thai language and based on UNL according to UNU/IAS Standard. The thinking method of converting procedure adapted from the system of Thai language invention from Interlingua and uses dictionary database design process including with using programming process to convert language by making a model coverting program named “CU-Deconverter”. The thinking method to convert UNL to Thai language is presented in 6 steps as followings: (1) Checking the correctness of the original UNL document, vocabularies and displaying an error.(2) Searching the meaning of word and additional information of Thai language from dictionary database. (3) Comparing Verb Pattern and the value of function from database. (4) Arranging words in accordance with the value derived from step 3. (5) Forming the sentence and deleting repeated words and (6) Displaying in Thai sentence patterns.

Dictionary Database designed in this thesis is consisted of following tables: (1) Table of Object Classes, (2) Table of Universal Words, (3) Table of Categories, (4) Table of Relation Labels, (5) Table of Attribute Labels, (6) Table of UNL Dictionary, (7) Table of Function and (8) Table of Verb Pattern. At the beginning, it is consisted of data to convert UNL to Thai language partially and if vocabulary, vocabulary pattern and data is not found, the new vocabulary or the new vocabulary pattern can be added to the database in order to use in language convertion. This langauge converting program will be operate with two User Interface both Windows and Webpage styles and the users can specify file or webpage that they want to make it adapted.

UNL’s Document utilized in this test was pickup as sample from UNL textbook of UNU/IAS, in the form of HTML data file, which consists of Universal Word, Sentences combination format and information of additional explanation of Universal Word. Test results in Thai language having accuracy rate of correct meaning with original sentence format at approximately 85 %.

Department …….Computer.Engineering….. Student’s signature ……

Field of study …..Computer.Science…….. Advisor’s signature ……

Academic year …2000……...

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของรองศาสตราจารย์ ดร.วันชัย ริ้วไพบูลย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่างๆซึ่งเป็นประโยชน์อย่างยิ่งต่อการวิจัยมาด้วยดีโดยตลอดตั้งแต่เริ่มต้นทำวิทยานิพนธ์รวมถึงตรวจแก้วิทยานิพนธ์ฉบับนี้อย่างละเอียด ขอขอบคุณศูนย์ประมวลผลภาษาธรรมชาติ (LINK) ศูนย์เทคโนโลยีและวิทยาศาสตร์แห่งประเทศไทย (NECTEC) ที่ได้ให้ความเอื้อเฟื้อให้ข้อมูลต่างๆ เกี่ยวกับการแปลภาษาด้วยคอมพิวเตอร์ ผู้เขียนขอขอบคุณ Dr.Francisco Astudillo-Pacheco UNL Center สำหรับข้อมูลและคำแนะนำในการติดต่อกับทางศูนย์ภาษา (UNL Center) ขอขอบคุณ Dr. Hiroshi Uchida ผู้อำนวยการโครงการศูนย์ UNL Center แห่ง UNU/IAS ที่ได้กรุณาให้ความช่วยเหลือข้อมูลต่างๆ แก่ผู้เขียน

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณบิดาและมารดา ผู้ซึ่งเป็นผู้ให้กำเนิดและคอยให้กำลังใจแก่ผู้วิจัยเสมอจนได้มีโอกาสศึกษาต่อในระดับปริญญามหาบัณฑิตจนกระทั่งสำเร็จการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง. ..... ฌ

สารบัญภาพ..... ฎ

บทที่

1. บทนำ..... 1

1.1 ความเป็นมาและความสำคัญของปัญหา..... 1

1.2 วัตถุประสงค์ของการวิจัย..... 10

1.3 ขอบเขตของการวิจัย..... 10

1.4 ขั้นตอนและวิธีการดำเนินการวิจัย..... 11

1.5 ประโยชน์ที่คาดว่าจะได้รับ..... 12

1.6 ผลงานที่ตีพิมพ์จากงานวิจัย..... 12

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง..... 13

2.1 หลักการแปลภาษาด้วยคอมพิวเตอร์..... 13

2.1.1 วิธีการแปลโดยตรง (Direct Translation Approach) 15

2.1.2 วิธีการแปลโดยการถ่ายทอดความหมาย (Transfer Approach).....16

2.1.3 วิธีการแปลโดยใช้ภาษากลาง (Interlingua Approach).....17

2.2 แนวทฤษฎีทางไวยากรณ์และเทคนิคการวิเคราะห์..... 18

2.3 ภาษากลาง (Interlingua) 20

2.4 ระบบการสร้างภาษาไทยจากภาษากลาง..... 22

2.5 งานวิจัยที่เกี่ยวข้อง..... 30

2.6 การวิจัยระบบเครื่องแปลภาษาภายในประเทศ..... 38

2.7 อนาคตและแนวทางของการวิจัยและพัฒนา..... 40

3. ลักษณะของภาษา UNL และแนวความคิดในการแปลงเป็นภาษาไทย..... 42

3.1 ความหมายของภาษา UNL..... 42

สารบัญ (ต่อ)

หน้า

3.2 ลักษณะทั่วไปของระบบภาษา UNL..... 42

3.3 การแสดงข้อมูลภาษา UNL ด้วยคอมพิวเตอร์..... 44

3.4 จุดมุ่งหมายของภาษา UNL..... 45

3.5 กฎเกณฑ์ของภาษา UNL..... 45

3.6 แนวความคิดในการแปลงเป็นภาษาไทย.....47

4. การออกแบบโปรแกรม CU-Deconverter.....53

4.1 โครงสร้างของโปรแกรม CU-Deconverter..... 53

4.1.1 ลักษณะติดต่อแบบวินโดว์.....54

4.1.2 ลักษณะติดต่อแบบเวบ.....57

4.1.3 โครงสร้างโปรแกรม Deconverter 59

4.1.4 ฐานข้อมูล UNL 60

4.2 ขั้นตอนการทำงานของโปรแกรม.....62

5. การทดสอบโปรแกรม CU-Deconverter..... 71

5.1 วิธีการทดสอบ..... 71

5.2 ผลการทดสอบ..... 71

5.2.1 ผลการทดสอบการแปลงจากภาษา UNL เป็นภาษาไทยทีละขั้นตอน..... 72

5.2.2 ตัวอย่างผลลัพธ์ข้อความผิดพลาดของการทำงานทีละขั้นตอน..... 75

5.2.3 ผลการทดสอบการแปลงด้วยโปรแกรม CU-Deconverter..... 76

6. สรุปผลการวิจัยและข้อเสนอแนะ..... 79

6.1 สรุปผลการวิจัย..... 79

6.2 ข้อดีและข้อเสียของโปรแกรม CU-Deconverter..... 80

6.3 ข้อจำกัด..... 81

6.4 ข้อเสนอแนะและความคิดเห็นเพิ่มเติม..... 81

รายการอ้างอิง..... 84

ภาคผนวก..... 86

ภาคผนวก ก.....87

- ตัวอย่างฐานข้อมูล UNL ตารางข้อมูลลักษณะเฉพาะ, ประเภทคำ, หน้าที่คำ..... 87
- ตัวอย่างฐานข้อมูล UNL ตารางข้อมูลคำเชื่อมโยง, ลักษณะคำ.....92

สารบัญ (ต่อ)

หน้า

ภาคผนวก ข 99
- ตัวอย่างหน้าจอโปรแกรม CU-Deconverter ติดต่อแบบวินโดว์ 99
- ตัวอย่างหน้าจอโปรแกรม CU-Deconverter ติดต่อแบบเวบ.....112

ภาคผนวก ค..... 123
- ตัวอย่าง Source Program.....123
- ตัวอย่างภาษา UNL และผลลัพธ์การแปลงเป็นภาษาไทย..... 141
- Data Dictionary.....155

ประวัติผู้เขียน..... 166

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 1.1 ตารางภาษาที่ทั่วโลกใช้กันมาก..... 3
ตารางที่ 2.1 ตัวอย่างฐานข้อมูลคู่ระหว่างภาษาไทยและภาษาอังกฤษในระดับประโยค.....15
ตารางที่ 2.2 โครงสร้างของพจนานุกรม 2 ภาษา..... 17
ตารางที่ 2.3 ตารางแสดงฐานข้อมูลของการเรียงลำดับคำในภาษาไทย..... 28
ตารางที่ 2.4 ตัวอย่างพจนานุกรมอิเล็คทรอนิกส์..... 31
ตารางที่ 2.5 ตัวอย่างระบบเครื่องแปลภาษาที่พัฒนาขึ้นในมหาวิทยาลัยในช่วงปี ค.ศ. 1970.....31
ตารางที่ 2.6 ตัวอย่างระบบเครื่องแปลที่พัฒนาใช้ในสหรัฐฯ.....32
ตารางที่ 2.7 ตัวอย่างโครงการพัฒนาระบบเครื่องแปลภาษา.....32
ตารางที่ 2.8 ระบบการแปลภาษาด้วยคอมพิวเตอร์ของบริษัทในประเทศญี่ปุ่น.....33
ตารางที่ 2.9 ตารางแสดงการเปรียบเทียบกลวิธีการแปลภาษาด้วยคอมพิวเตอร์.....37
ตารางที่ 3.1 ตัวอย่างของ UNL ในฐานข้อมูลพจนานุกรม..... 48
ตารางที่ 3.2 ตัวอย่างคำ "Book" ในตาราง UNL Dictionary..... 48
ตารางที่ 4.1 เมนูการทำงานของโปรแกรมที่เป็นลักษณะการติดต่อแบบวินโดว์..... 55
ตารางที่ 4.2 ลักษณะโปรแกรมการติดต่อแบบเวบ..... 57
ตารางที่ 4.3 รายละเอียดขั้นตอนการทำงานของแต่ละโมดูล..... 59
ตารางที่ 4.4 รายละเอียดตารางต่างๆ ในฐานข้อมูลพจนานุกรม..... 61
ตารางที่ 4.1 แสดงรหัสหน้าที่และค่าของแต่ละหน้าที่คำ..... 67
ตารางที่ 5.1 ผลการทดสอบแปลงจาก UNL เป็นภาษาไทยจำนวน 70 ประโยค.....76
ตารางที่ 7.1 ตัวอย่างตารางข้อมูลแสดงลักษณะเฉพาะคำ..... 88
ตารางที่ 7.2 ตัวอย่างตารางข้อมูลประเภทคำ..... 89
ตารางที่ 7.3 ตัวอย่างตารางข้อมูลหน้าที่คำ.....90
ตารางที่ 7.4 ตัวอย่างตารางข้อมูลคำเชื่อมโยง..... 95
ตารางที่ 7.5 ตัวอย่างข้อมูลตารางคุณลักษณะคำ..... 97

สารบัญภาพ

หน้า

รูปที่ 1.1 โครงสร้างของระบบการแปลภาษาด้วยคอมพิวเตอร์ทั่วไป..... .5
รูปที่ 1.2 ขั้นตอนของระบบการแปลภาษาด้วยคอมพิวเตอร์โดยทั่วไป..... 5
รูปที่ 1.3 ประวัติและพัฒนาการทางทฤษฎีของการวิจัยและพัฒนาระบบเครื่องแปลภาษา..... 10
รูปที่ 2.1 โครงสร้างการแปลภาษาด้วยการถ่ายทอด……. 13
รูปที่ 2.2 แสดงโครงสร้างของการใช้คอมพิวเตอร์ในการแปลภาษาวิธีต่างๆ..... 15
รูปที่ 2.3 วิธีการแปลโดยตรง……..... 16
รูปที่ 2.4 วิธีการแปลโดยใช้การถ่ายทอดความหมาย…….. 16
รูปที่ 2.5 วิธีการแปลโดยใช้ภาษากลาง……..... 17
รูปที่ 2.6 ตัวอย่างของการกสัมพันธ์ระหว่างคำกริยากับคำนาม..... 19
รูปที่ 2.7 ตัวอย่างของความสัมพันธ์แบบหน่วยหลัก-หน่วยพึ่งพา..... 19
รูปที่ 2.8 โครงสร้างภาษากลางของประโยค “He moves a chair in the room”..... .21
รูปที่ 2.9 ตัวอย่างการแปลจากภาษาอังกฤษเป็นภาษาไทย..... .22
รูปที่ 2.10 โครงสร้างของระบบการสร้างภาษาไทยจากภาษากลาง..... 23
รูปที่ 2.11 หลักการ Traversing แบบ depth-first หรือ top-down & left-right algorithm..... 24
รูปที่ 2.12 ผลลัพท์ที่เป็นส่วนเอาท์พุทของ Generation Planning..... 25
รูปที่ 2.13 ตัวอย่างการกำหนดโครงสร้างทางไวยากรณ์..... 26
รูปที่ 2.14 ผลลัพธ์ของโครงสร้างทางไวยากรณ์ภาษาไทย สำหรับการแปลประโยค..... 26
รูปที่ 2.15 กฎการกำหนดลำดับของโนดทางลอจิก..... . 27
รูปที่ 2.16 ตารางแสดงฐานข้อมูลของการเรียงลำดับคำในภาษาไทย..... 29
รูปที่ 3.1 ลักษณะทั่วไปของระบบภาษา UNL ……..... 42
รูปที่ 3.2 ขั้นตอนการแปลงภาษา UNL เป็นภาษาไทย……..…….. 49
รูปที่ 3.3 ตัวอย่างการแปลงจากภาษา UNL เป็นภาษาไทย……... 51
รูปที่ 3.4 ตัวอย่างตารางในฐานข้อมูล UNL 52
รูปที่ 4.1 โครงสร้างโปรแกรม CU-Deconverter..... 54
รูปที่ 4.2 โครงสร้างโปรแกรมลักษณะติดต่อแบบวินโดว์…..... 56
รูปที่ 4.3 โครงสร้างโปรแกรมลักษณะติดต่อแบบเวบ…..……... 57
รูปที่ 4.4 โครงสร้างโปรแกรมลักษณะติดต่อแบบเวบในส่วนของฐานข้อมูล…..... 58
รูปที่ 4.5 ขั้นตอนการทำงานของโปรแกรม CU-Deconverter……. 59
รูปที่ 4.6 ตารางในฐานข้อมูล UNL …….. 62
รูปที่ 4.7 ตัวอย่างของคำและความหมาย UNL ที่แสดงในบรรทัด…….. 63
รูปที่ 4.8 แสดงขั้นตอนการทำงานของโมดูลตรวจสอบความถูกต้อง…… 64
รูปที่ 4.9 ขั้นตอนการทำงานหาความหมายของคำ…….. 65

รูปที่ 4.10 ตัวอย่างของความสัมพันธ์ระหว่างคำในภาษา UNL..... 66
รูปที่ 4.11 แสดงขั้นตอนการทำงานของโมดูลเปรียบเทียบรูปแบบกริยา..... 66
รูปที่ 4.12 ตัวอย่างผลลัพธ์หลังจากจัดเรียงตามค่าของหน้าที่คำ..... 68
รูปที่ 4.13 ตัวอย่างผลลัพธ์หลังจากขั้นตอนจัดเรียงคำ..... 68
รูปที่ 4.14 แสดงขั้นตอนการทำงานของโมดูลจัดเรียงคำ..... 68
รูปที่ 4.15 ตัวอย่างผลลัพธ์หลังจากขั้นตอนกำหนดคำภาษาไทย..... 69
รูปที่ 4.16 แสดงขั้นตอนการทำงานของโมดูลกำหนดคำภาษาไทย..... 69
รูปที่ 4.17 ตัวอย่างผลลัพธ์สุดท้ายของการแปลงจาก UNL เป็นภาษาไทย..... 70
รูปที่ 4.18 แสดงขั้นตอนการทำงานของโมดูลแสดงผลภาษาไทย..... 70
รูปที่ 7.1 หน้าจอเริ่มเข้าสู่โปรแกรม CU-Deconverter..... 98
รูปที่ 7.2 หน้าจอ UNL Dictionary Explorer..... 98
รูปที่ 7.3 หน้าจอที่แสดงผ่าน Web Browser ในรูปแบบวินโดว์..... 99
รูปที่ 7.4 แสดงเมนูของ View UNL Database..... 99
รูปที่ 7.5 หน้าจอเมนู Tools แสดงขั้นตอนการทำงานของโปรแกรม Phase I..... 100
รูปที่ 7.6 หน้าจอเมนู Tools แสดงขั้นตอนการทำงานของโปรแกรม Phase II..... 100
รูปที่ 7.7 หน้าจอ UNL Input รับแฟ้มข้อมูล UNL..... 101
รูปที่ 7.8 หน้าจอหลังจากข้นตอน UNL Input..... 101
รูปที่ 7.9 หน้าจอแสดงผลในขั้นตอน Pre-Edit..... 102
รูปที่ 7.10 หน้าจอแสดงผล Error ข้นตอน Pre-Edit..... 102
รูปที่ 7.11 หน้าจอแสดงผลขั้นตอน Scan UWs From Dictionary..... 103
รูปที่ 7.12 หน้าจอแสดงผลขั้นตอน Get Verb Pattern..... 103
รูปที่ 7.13 หน้าจอแสดงผลขั้นตอน Word Ordering..... 104
รูปที่ 7.14 หน้าจอแสดงผลขั้นตอน Morphological Generation..... 104
รูปที่ 7.15 หน้าจอแสดงผลลัพธ์ภาษาไทยในขั้นตอน Thai Language..... 105
รูปที่ 7.16 หน้าจอแสดงข้อมูลเกี่ยวกับโปรแกรม CU-Deconverter..... 105
รูปที่ 7.17 แสดงรายชื่อ Text File ที่ใช้ในโปรแกรม..... 106
รูปที่ 7.18 ตัวอย่างข้อมูลที่เก็บไว้ใน Text File..... 106
รูปที่ 7.19 ตัวอย่างหน้าจอ Table Category..... 107
รูปที่ 7.20 ตัวอย่างหน้าจอ Table Sub Category..... 107
รูปที่ 7.21 ตัวอย่างหน้าจอ Table Object Class..... 108
รูปที่ 7.22 ตัวอย่างหน้าจอ Table Function..... 108
รูปที่ 7.23 ตัวอย่างหน้าจอ Table Attribute Type..... 109
รูปที่ 7.24 ตัวอย่างหน้าจอ Table Attribute Labels..... 109
รูปที่ 7.25 ตัวอย่างหน้าจอ Table Relation Labels..... 110

รูปที่ 7.26 ตัวอย่างหน้าจอ Table Universal Words……..... 110
รูปที่ 7.27 ตัวอย่างหน้าจอ About CU-Deconverter Web Page…….…..... 111
รูปที่ 7.28 ตัวอย่างหน้าจอ My CU-Deconverter Web Page .....……..... 111
รูปที่ 7.29 ตัวอย่างหน้าจอ Select A UNL Topic Web Page……..... 112
รูปที่ 7.30 ตัวอย่างหน้าจอ Related-Site Web Page……..……..... 112
รูปที่ 7.31 ตัวอย่างหน้าจอ Site-Index Web Page……..…..... 113
รูปที่ 7.32 ตัวอย่างหน้าจอ UNL Database Web Page……..... 113
รูปที่ 7.33 ตัวอย่างหน้าจอ View Category Table Web Page……..... 114
รูปที่ 7.34 ตัวอย่างหน้าจอ Entry Category Table Web Page ……..... 114
รูปที่ 7.35 ตัวอย่างหน้าจอ List View Category Table Web Page ……..... 115
รูปที่ 7.36 ตัวอย่างหน้าจอ Filter Category Table Web Page ……..... 115
รูปที่ 7.37 ตัวอย่างหน้าจอ Input UNL File Web Page ……..……..... 116
รูปที่ 7.38 ตัวอย่างหน้าจอ Open Input UNL File Web Page ……..... 116
รูปที่ 7.39 ตัวอย่างหน้าจอขั้นตอน Pre-Edit Web Page ……..……..... 117
รูปที่ 7.40 ตัวอย่างหน้าจอขั้นตอน Thai Language Web Page …… 117
รูปที่ 7.41 ตัวอย่างหน้าจอขั้นตอน Word Ordering Web Page …… 118
รูปที่ 7.42 ตัวอย่างหน้าจอขั้นตอน Morphological Generation Web Page …… 118
รูปที่ 7.43 ตัวอย่างหน้าจอ CU-Deconverter Help Web Page ..……..... 119
รูปที่ 7.44 ตัวอย่างหน้าจอ Universal Networking Language Web Page …… 119
รูปที่ 7.45 ตัวอย่างหน้าจอ CU-Deconverter FAQ Web Page …… 120
รูปที่ 7.46 ตัวอย่างหน้าจอ CU-Deconverter Thesis Web Page …… 120
รูปที่ 7.47 ตัวอย่างหน้าจอ Thesis Introduction กิตติกรรมประกาศ Web Page …… 121
รูปที่ 7.48 ตัวอย่างหน้าจอ Thesis Contents สารบัญ Web Page …… 121

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การแปลภาษาด้วยเครื่อง (Machine Translation) จะเป็นการใช้คอมพิวเตอร์เพื่อการแปลภาษาระหว่างภาษาของมนุษย์ เช่น การแปลภาษาระหว่างภาษาไทยกับภาษาอังกฤษเป็นต้น การแปลภาษาจัดเป็นเรื่องสำคัญสำหรับงานสำนักงานในอนาคตมาก เพราะการติดต่อการค้าระหว่างประเทศ การแลกเปลี่ยนสื่อสารข้อมูลเป็นเครือข่ายใหญ่ระหว่างประเทศและข้อสำคัญคือการรวมข่ายงานของภาษาธรรมชาติทั้งในแง่ของการติดต่อด้วยเสียง, การติดต่อฐานข้อมูล (Database Interface) และการแก้ไขข้อความ (Text Editing) เข้าด้วยกันถ้ามีระบบการแปลภาษาอยู่พร้อมแล้ว การแปลภาษาจากภาษาอื่นมาเป็นภาษาไทยหรือจากภาษาไทยไปสู่ภาษาอื่น จะช่วยอำนวยความสะดวกและเร่งการติดต่อแลกเปลี่ยนข้อมูลให้เป็นไปอย่างรวดเร็วเรื่องการแปลภาษานี้ยังมีข้อเสียอยู่บ้าง ตรงที่เราไม่สามารถแปลเอกสารทุกสิ่งทุกอย่างได้ เราจะต้องตีกรอบของการแปลนี้แคบอยู่เฉพาะเรื่อง เช่นการแปลเอกสารทางเทคโนโลยี เป็นต้น ทั้งนี้เพราะการแปลภาษาต้องมีการเก็บคำศัพท์พร้อมวิธีการใช้คำนั้นภายในหน่วยความจำของคอมพิวเตอร์ไว้ ถ้าต้องการให้ความถูกต้องแม่นยำของการแปลสูงมากๆแล้ว พจนานุกรมทางอิเล็กทรอนิกส์เก็บคำนี้ต้องมีขนาดใหญ่มากจริงๆ ดังนั้นจึงต้องมีการแปลเฉพาะสาขาในระยะนี้ แต่ถ้าเทคโนโลยีของคอมพิวเตอร์ขยายไปอีก การแปลมากสาขาก็อาจเป็นไปได้ในอนาคต (การประมวลผลภาษาธรรมชาติ, 2535)

ภาษา UNL (Universal Networking Language) เป็นภาษาทางปัญญาประดิษฐ์ที่สร้างขึ้นเพื่อเป็นภาษากลางสำหรับการติดต่อสื่อสารระหว่างภาษาท้องถิ่นที่แตกต่างกัน ในการแปลภาษาโดยใช้ภาษา UNL เป็นตัวเชื่อมระหว่างภาษาต้นฉบับ (Source Language) กับภาษาเป้าหมาย (Target Language) นั้นเช่น จากภาษาอังกฤษเป็นภาษา UNL แล้วจึงแปลเป็นภาษาไทยนั้น เป็นกระบวนการแปลภาษาแบบถ่ายทอดทางอ้อมแบบหนึ่ง โดยแบ่งเป็น 2 ขั้นตอนใหญ่ๆได้แก่ขั้นตอนการวิเคราะห์ภาษาต้นฉบับไปสู่ภาษา UNL และขั้นตอนการสร้างภาษาเป้าหมายจากภาษา UNL วิธีการแปลภาษาแบบนี้จะช่วยลดความยุ่งยากและลดเวลาในการวิเคราะห์ภาษาต้นฉบับและการสร้างภาษาเป้าหมายเมื่อเปรียบเทียบกับกระบวนการแปลภาษาแบบถ่ายทอดโดยตรง ทั้งนี้เพราะเราไม่จำเป็นต้องคำนึงถึงคู่ภาษาที่จะแปล เนื่องจากการวิเคราะห์ภาษาต้นฉบับใดๆ ก็ตามให้เป็นภาษา UNL แล้วก็จะสามารถแปลเป็นภาษาเป้าหมายอะไรก็ได้ โดยที่ไม่ต้องวิเคราะห์ใหม่และใน

ทำนองเดียวกันทางด้านการสร้างภาษาเป้าหมายใดๆนั้น ก็เป็นการสังเคราะห์จากภาษา UNL ซึ่งเป็นผลของการวิเคราะห์ภาษาต้นฉบับใดๆ เพียงอย่างเดียว เมื่อมีการสังเคราะห์แล้วก็ไม่ต้องทำซ้ำอีกต่อไปนอกจากนั้นการแปลภาษาโดยใช้ภาษา UNL นี้จะใช้พจนานุกรมภาษา UNL เป็นฐานข้อมูลในการวิเคราะห์และสังเคราะห์เพียงชุดเดียวต่อหนึ่งภาษา ซึ่งต่างจากการแปลภาษาแบบถ่ายทอดโดยตรงที่จะต้องใช้พจนานุกรมถึง n ชุดในการแปลจากภาษาต้นฉบับ 1 ภาษาไปสู่ภาษาเป้าหมาย n ภาษา (การสังเคราะห์ภาษาไทยจากภาษากลางในระบบการแปลหลายหลายภาษา, 2536)

ในการสังเคราะห์ประโยคภาษาไทยจากภาษา UNL ของงานวิจัยนี้ ได้ศึกษาถึงโครงสร้างและมาตราฐานของภาษา UNL ซึ่งจะต้องถูกกำหนดขึ้นให้เป็นตัวแทนทางด้านความหมายลึกของทุกภาษา โดยใช้ทฤษฎีทางวากยสัมพันธ์หรือไวยากรณ์เคสของฟิลล์มอร์เป็นพื้นฐานและพิจารณาให้เหมาะสมกับภาษาไทย เพื่อจะได้นำไปสู่การเขียนกระบวนการสร้างภาษาไทยจากภาษา UNL

ประวัติความเป็นมาของการแปลภาษาด้วยเครื่อง (Machine Translation)

นานมาแล้วในเมืองบาบิโลน ผู้คนเริ่มทำการสร้างหอคอยเพื่อที่จะไปให้ถึงสวรรค์ ด้วยความมุ่งมั่น เมื่อพระเจ้าทราบเรื่อง พระองค์ทรงโกรธมากและพระองค์ก็สร้างภาษาต่างๆ ขึ้น เพื่อไม่ให้มนุษย์สามารถติดต่อสื่อสารกันในที่สุดหอคอยก็สร้างไม่สำเร็จที่กล่าวนี้เป็นเรื่องในพระคัมภร์เก่าเกี่ยวกับหอคอยของบาบิโลน “Tower of Babel”

เรื่องดังกล่าวข้างต้น ทำให้เกิดภาษาต่างๆ มากกว่า 3,000 ภาษา ที่ใช้พูดกันในโลกทุกวันนี้ หรืออาจมากกว่านี้ ภาษามากกว่า 13 ภาษาถูกใช้กับคนมากกว่า 100 ล้านคนได้แก่ภาษาจีน, ภาษาอังกฤษ,ภาษาสเปน,ภาษาโปรตุเกส, ภาษาอารบิก, ภาษาอินโดนีเซีย, ภาษาฝรั่งเศส, ภาษารัสเซีย, ภาษาเยอรมันและภาษาญี่ปุ่น และภาษามากกว่า 38 ภาษาที่เป็นภาษาแม่ของคนมากกว่า 10 ล้านคน (UNL Center, 1996)

หากเรียงลำดับภาษาที่มีผู้คนใช้มากที่สุดแล้วลดหลั่นกันไปตามลำดับ จะได้ดังตารางที่ 1.1 ต่อไปนี้

ตารางที่ 1.1 ตารางภาษาที่ทั่วโลกใช้กันมาก (การประมวลผลภาษาธรรมชาติ, 2535)

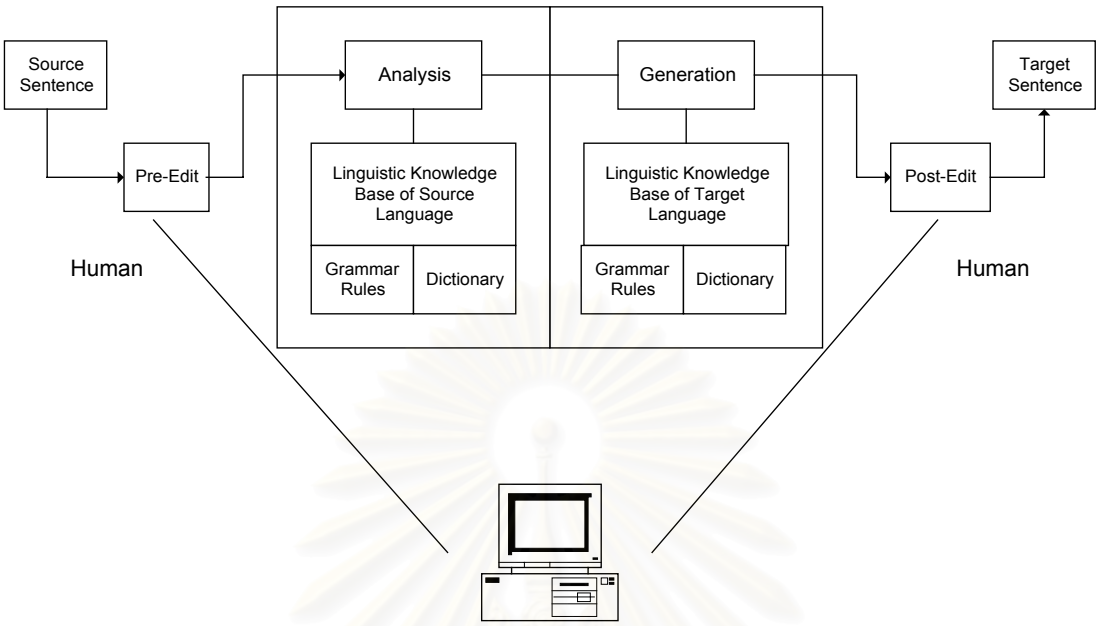
ภาษา	จำนวนคนที่ใช้ (ล้านคน)	ภาษา	จำนวนคนที่ใช้ (ล้านคน)
Mandarin	800	French	103
English	425	Italian	82
Hindi	240	Urdu	80
Russian	235	Cantonese	74
Spanish	230	Javenese	70
German	175	Ukranian	78
Japanese	140	Telugn	78
Bengali	130	Tamil	75
Portugese	120	Wu (China)	75
Arabic	115	Min (China)	61
Malay	105	Thai	60

จากการที่มีภาษามากมายดังกล่าว ได้ทำให้เกิดอุปสรรคในการติดต่อสื่อสารระหว่างชาติต่างๆ ในโลกการถ่ายทอดความรู้เทคโนโลยี การแลกเปลี่ยนวัฒนธรรมและสิ่งต่างๆ จะต้องสิ้นเปลืองค่าใช้จ่ายและเสียเวลาในการแปลภาษาค่อนข้างมาก ถึงแม้ว่าในปัจจุบันจะมีการติดต่อสื่อสารกันโดยใช้ภาษาอังกฤษเป็นภาษากลางแต่จะดีกว่านี้หรือไม่ ถ้ามนุษย์ทุกคนสามารถติดต่อสื่อสารหรือถ่ายทอดความรู้โดยใช้ภาษาที่ตนเองใช้มาแต่กำเนิด และสามารถติดต่อสื่อสารกับบุคคลที่ใช้ภาษาอื่นได้โดยสะดวกที่สุด อุปสรรคต่างๆ ทางด้านภาษาได้ทำให้เกิดงานวิจัยทางด้านการแปลภาษาจากภาษาหนึ่งไปยังอีกภาษาหนึ่ง และมีการพัฒนากันอย่างต่อเนื่อง การใช้คอมพิวเตอร์หรือการแปลภาษาด้วยเครื่องนั้น ได้เริ่มขึ้นมาตั้งแต่ในทศวรรษที่ 1950 โดยการสนับสนุนของกระทรวงกลาโหมของสหรัฐอเมริกา ในการที่จะสร้างโปรแกรมประยุกต์ที่มีความรวดเร็ว และถูกต้องในการแปลเอกสารต่างๆ ทางทหาร ซึ่ง ณ ปัจจุบันก็ถือได้ว่าเป็นทศวรรษที่ 5 ของการใช้คอมพิวเตอร์ช่วยในการแปลภาษาแล้ว ความสนใจการพัฒนาในด้านนี้ก็ได้แพร่หลายไปทั่วโลก และยอมรับกันว่าการใช้คอมพิวเตอร์เข้าช่วยในการแปลภาษานั้น ให้ประโยชน์กับเราได้อย่างมาก ถึงแม้ว่าจะไม่สามารถทำได้สมบูรณ์แบบ 100% ก็ตาม (S. Keretho C. Wongchaisuwat and Y. Poovarawan, 1993)

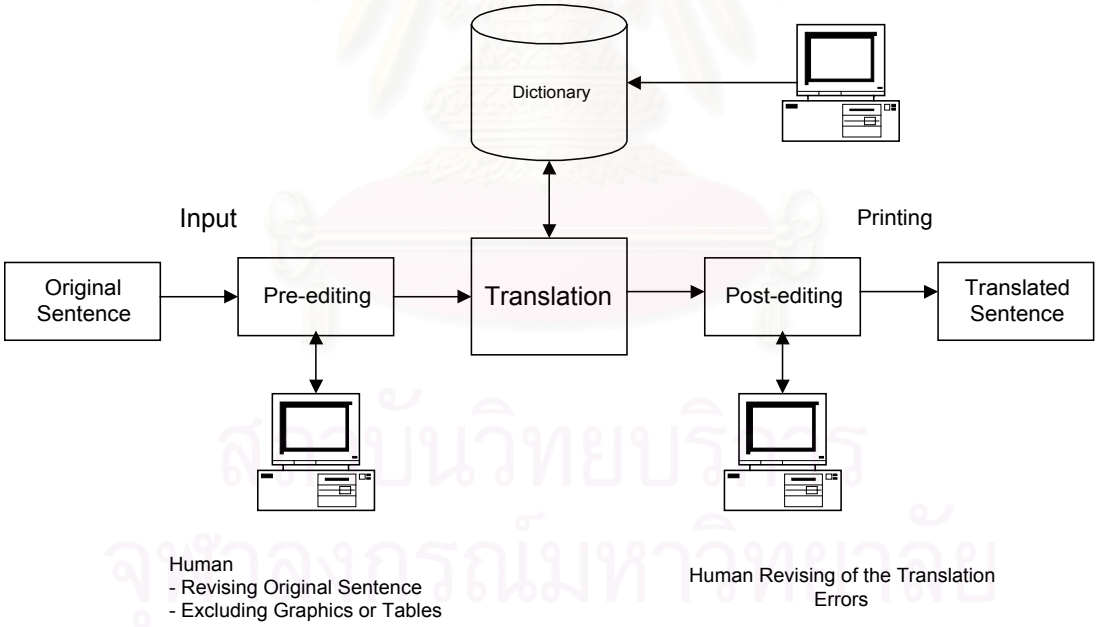
ในการแปลภาษาหนึ่งไปเป็นอีกภาษาหนึ่งด้วยคอมพิวเตอร์นั้นเป็นการเลียนแบบกระบวนการแปลภาษาของมนุษย์ คือ มนุษย์จะต้องอ่านภาษาต้นแบบ และทำความเข้าใจกับภาษาต้นแบบนั้น แล้วจึงแปลมาเป็นภาษาเป้าหมายตามความคิดที่มนุษย์เข้าใจ ซึ่งผู้ที่จะแปลนั้นจะต้องเข้าใจหรือมีความรู้ ทั้งภาษาต้นแบบและภาษาเป้าหมาย ซึ่งมนุษย์จะใช้ความรู้ที่จดจำไว้ในสมองเป็นเครื่องมือในการแปล ส่วนระบบคอมพิวเตอร์จะแปลภาษาได้ต้องอาศัยกระบวนการประมวลผลภาษาธรรมชาติ (Natural language processing) และฐานความรู้ทางภาษาศาสตร์ (Linguistic Knowledge Base) (อรัญญา, 2537)

การแปลภาษาด้วยเครื่องนั้นเป็นการนำทฤษฎีทางด้านการประมวลผลภาษาธรรมชาติ มาเป็นพื้นฐานในการวิเคราะห์ (Analysis) และสังเคราะห์ (Generation) ประโยค ขบวนการแปลนั้นเป็นการเลียนแบบวิธีการแปลโดยธรรมชาติของบุคคลดังแสดงไว้ในรูปที่ 1.1 นั้นคือจะต้องมีการอ่านประโยคของภาษาต้นฉบับ ที่จะต้องการแปลและทำความเข้าใจในความหมายคำและโครงสร้างของประโยค ต่อจากนั้นก็ค้นหาคำศัพท์ของภาษาเป้าหมาย ที่มีความหมายตรงกับคำในภาษาต้นฉบับ แล้วเรียงคำให้ถูกต้องตามหลักไวยากรณ์ของภาษาเป้าหมายนั้น ซึ่งอาจจะต้องมีการตกแต่งประโยคให้ถูกต้องยิ่งขึ้น ดังนั้นปัญหาของการแปลภาษาด้วยเครื่องคอมพิวเตอร์จึงอยู่ที่ว่าจะสามารถทำให้คอมพิวเตอร์เข้าใจประโยคหรือข้อความที่จะแปลได้เหมือนมนุษย์ที่เป็นผู้แปลได้ อย่างไร ปัญหานี้สามารถแก้ไขได้ด้วยการสร้างฐานความรู้ (Knowledge Base) ในการแปลให้ระบบคอมพิวเตอร์ ฐานความรู้ที่ว่านี้ได้แก่ ระบบการวิเคราะห์ประโยค (Text Analysis System) และระบบการสังเคราะห์หรือสร้างประโยค (Text Generation System) ซึ่งบรรจุไว้ด้วยกฎไวยากรณ์ของภาษาตลอดจนพจนานุกรมอิเล็กทรอนิกส์ (Electronic Dictionary) ของคู่ภาษาที่จะแปลดังแสดงไว้ในรูปที่ 1.2 (Machine Translation, 2536)

ในการแปลภาษาด้วยเครื่องนี้ ส่วนสำคัญอย่างยิ่งของระบบคือพจนานุกรมอิเล็กทรอนิกส์ การแปลจะมีคุณภาพหรือไม่ขึ้นอยู่กับพจนานุกรมอิเล็กทรอนิกส์ที่ถูกต้อง สมบูรณ์ และมีประสิทธิภาพซึ่งคอมพิวเตอร์สามารถนำไปใช้งานได้อย่างรวดเร็วและเหมาะสม ดังนั้น การพัฒนาระบบการแปลภาษาด้วยเครื่องจึงต้องมีการพัฒนาพจนานุกรมอิเล็กทรอนิกส์ควบคู่ไปด้วย งานพัฒนาพจนานุกรมอิเล็กทรอนิกส์เป็นงานที่มีขอบเขตกว้างขวาง ต้องใช้เวลาและแรงงานมาก เพราะเป็นงานที่เข้าไปเกี่ยวข้องกับระบบภาษาธรรมชาติทั้งระบบ (โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์, 2532)



รูปที่ 1.1 โครงสร้างของระบบการแปลภาษาด้วยคอมพิวเตอร์ทั่วไป



รูปที่ 1.2 ขั้นตอนของระบบการแปลภาษาด้วยคอมพิวเตอร์โดยทั่วไป

ลักษณะงานต่างๆ ที่มักต้องการการแปลจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่ง สามารถที่จะแบ่งออกเป็นกลุ่มๆ ได้ดังต่อไปนี้ (Makoto Nagao, 1986)

1. งานวรรณกรรมและงานเขียนทั่วไป มักมีจุดประสงค์หลักในการสื่อสารทางด้านอารมณ์กับผู้อ่าน
2. เอกสารและสัญญาต่างๆ ทางกฎหมาย ต้องการภาษาที่มีความชัดเจน ไม่กำกวม
3. เอกสารด้านงานเทคนิคและวิทยาศาสตร์ต้องการสื่อความเข้าใจสำหรับบุคคลกลุ่มหนึ่ง ๆ
4. การค้นหาข้อมูล ชื่อเรื่อง ประโยคต่างๆ เพื่อการสืบค้น ต้องการการแปลที่พอสื่อความเข้าใจกับผู้ใช้นั้นๆ เท่านั้น

นอกจากการแบ่งกลุ่มลักษณะงานต่างๆ ข้างต้นแล้ว การแปลภาษายังอาจแบ่งแยกออกได้ตามวัตถุประสงค์ของการแปลก็ได้คือ

1. การแปลเพื่อการอ่านตรวจสอบระดับแรก หรือ การแปลเพื่อใช้จริงระดับสุดท้าย
2. การแปลเพื่อใช้งานภายในเฉพาะกลุ่ม หรือ การแปลสำหรับกลุ่มคนทั่วไป
3. การแปลเพื่อต้องการรายละเอียด หรือ การแปลแบบสรุปคร่าวๆ
4. การแปลเพื่อเป็นการแนะนำในเบื้องต้น และจะมีรายละเอียดตามมา

การใช้คอมพิวเตอร์ในการแปลภาษา สามารถแบ่งตามความสามารถได้เป็น 3 แบบใหญ่ ๆ คือ

1. ระบบการแปลภาษาด้วยเครื่อง (Machine Translation) ย่อว่า MT เป็นระบบที่ใช้คอมพิวเตอร์หรือเครื่องจักรกลเป็นผู้แปลภาษาให้ทั้งหมดโดยที่ไม่ต้องมีมนุษย์เข้าไปเกี่ยวข้องในขั้นตอนใดๆเลยอย่างไรก็ดีโปรแกรมที่ออกแบบมาสำหรับระบบนี้ต้องมีประสิทธิภาพสูงมีพจนานุกรม และกฎในทางไวยากรณ์ต่างๆ ของภาษานั้นที่ครบถ้วนสมบูรณ์และถูกต้อง
2. ระบบการใช้คอมพิวเตอร์ช่วยในการแปลภาษา (Machine-Assisted Translation) ย่อว่า MAT เป็นระบบที่ต้องให้มนุษย์เข้าไปมีส่วนเกี่ยวข้องระหว่างการประมวลผล เพื่อให้ได้ผลลัพธ์ที่ถูกต้อง สามารถแบ่งได้เป็น 2 รูปแบบย่อย ดังนี้
    2.1 ใช้มนุษย์ช่วยคอมพิวเตอร์ในการแปลภาษา (Human-Assisted Machine Translation) ย่อว่า HAMT ระบบนี้คอมพิวเตอร์จะรับผิดชอบในการแปลภาษาทั้งหมดแต่จะมีการติดต่อกับมนุษย์ในบางขั้นตอน เช่น การเลือกคำที่เหมาะสมในกรณีมีคำที่สามารถใช้ได้หลายคำ หรือการเลือกวางตำแหน่งของคำ หรือวลี ว่าควรจะวางไว้ที่ใดในประโยค เป็นต้น

2.2 ใช้คอมพิวเตอร์ช่วยมนุษย์ในการแปลภาษา (Machine-Assisted Human Translation) ย่อว่า MAHT ในทางกลับกัน ระบบนี้มนุษย์จะเป็นผู้แปลภาษา แต่ จะใช้คอมพิวเตอร์เพื่อช่วยในการค้นหาคำศัพท์จากคลังข้อมูลคำจากที่ต่างๆ แนะนำวิธีการใช้คำไวยากรณ์ต่างๆ หรือ จัดรูปแบบเอกสาร เป็นต้น

3. ระบบคลังข้อมูลคำศัพท์ต่างๆ (Terminology Databanks) เป็นระบบที่ใช้คอมพิวเตอร์ เก็บข้อมูลคำศัพท์ วิธีใช้ หลักไวยากรณ์ต่างๆ เพื่อให้มนุษย์เป็นผู้เรียกใช้

หากพิจารณาการแปลภาษาจากภาษาหนึ่งไปยังอีกภาษาหนึ่งให้ดี จะพบว่าไม่เป็นเรื่องที่ง่ายเลยที่จะแปลภาษา ถึงแม้ว่าบางครั้งจะดูเหมือนว่าแปลในลักษณะหนึ่งต่อหนึ่งได้ เช่น

ภาษาเยอรมัน

Das	Buch	liegt	auf	dem	Tisch
The	Book	is	on	the	table

ภาษาอังกฤษ

หากดูโครงสร้างของภาษาแล้ว จะพบว่าการแปลภาษาจากภาษาหนึ่งไปยังอีกภาษาหนึ่งจะต้องเกี่ยวกับความรู้ทางภาษาศาสตร์ในหลายระดับ ทั้งระดับรูปคำ ระดับวากยสัมพันธ์ และระดับความหมาย เช่น ถ้าหากต้องการแปลข้อความว่า

The coach lost a set ให้เป็นความหมายภาษาไทย เราจะพบปัญหามากมาย เช่น เมื่อนำคำหลักมาดูในรูปวากยสัมพันธ์จะพบว่า

คำว่า coach เป็นได้ทั้ง คำนาม และกริยา

คำว่า lost เป็นได้ทั้งกริยาและวิเศษณ์

คำว่า set เป็นได้ทั้ง คำนาม กริยา และวิเศษณ์

ดังนั้นความเป็นไปได้ของประโยคทางวากยสัมพันธ์เป็นไปได้ถึง 2x2x3 หรือ 12 แบบซึ่งก็จะต้องแยกแยะทางวากยสัมพันธ์ให้ถูกต้องเสียก่อน ถึงแม้ว่าจะดูทางด้านวากยสัมพันธ์แล้วก็ตามก็ยังไม่เพียงพอในการที่จะแปลความหมายได้ เพราะถ้าลองเปิดพจนานุกรมเพื่อดูความหมายจะพบว่า

คำว่า coach มีความหมายได้ 6 ความหมาย

คำว่า lost มีความหมายได้ 10 ความหมาย

คำว่า set มีความหมายได้ 34 ความหมาย

เมื่อรวมกันจะมีความเป็นไปได้ถึง 6x10x34 หรือ 2,040 รูปแบบ และนี่เองคือ ปัญหาของการแปลภาษาด้วยคอมพิวเตอร์

ปัจจุบันคอมพิวเตอร์มีขีดความสามารถในการประมวลผลได้เร็วขึ้นกว่าเดิมมากและสามารถเก็บข้อมูลได้เป็นจำนวนมากเช่นกัน การคิดค้นหลักและวิธีการเกี่ยวกับงานทางด้านปัญญาประดิษฐ์เป็นไปได้อย่างกว้างขวาง จนกล่าวได้ว่าโอกาสของการประยุกต์คอมพิวเตอร์เข้ามาในงานด้านการแปลภาษาดูจะมีทางเป็นไปได้มากขึ้น (ยืน ภู่วรวรรณ, ชัยยงค์ วงศ์ชัยสุวัฒน์, 2535)

สำหรับความเป็นมาของระบบการแปลภาษาด้วยคอมพิวเตอร์ สรุปได้ดังนี้

สมัยที่ 1 ระหว่าง คศ. 1946 - 1952 แนวความคิดในการสร้างระบบขณะนั้น คือ แปลแบบคำต่อคำ อาศัยพจนานุกรมคำศัพท์ (dictionary) ขนาดใหญ่ ต่อมาจึงพัฒนาลักษณะพจนานุกรมคำศัพท์ใหม่ โดยแบ่งเป็น 2 ส่วนคือ

ก. พจนานุกรมรากศัพท์ (stem dictionary)

ข. พจนานุกรมวิภัตปัจจัย (ending dictionary)

และในช่วงปลายได้เกิดการแตกแยกทางความคิดเป็น 2 ทางคือ

1. เห็นว่าควรแปลเป็นภาษาคู่

2. เห็นว่าควรแปลเป็นภาษากลางก่อน

สมัยที่ 2 ระหว่าง คศ. 1952 - 1966 ในช่วงนี้แนวความคิดได้ขยายกว้างขึ้น แต่ยังแตกออกเป็น 2 แนวทาง คือ

1. เห็นว่าต้องใช้อรรถศาสตร์ (Semantic) มาแก้เรื่องความกำกวม

2. เห็นว่าต้องมีการร่วมมือกันระหว่างนักวิจัยสาขาต่างๆ เช่น นักปรัชญา นักมนุษย์วิทยา นักจิตวิทยา นักภาษาศาสตร์ ผู้เชี่ยวชาญการสื่อสาร นักคณิตศาสตร์ เป็นต้น

สมัยที่ 3 ระหว่าง คศ. 1966 ถึงปัจจุบัน การศึกษาได้แผ่ขยายจนมีศาสตร์แขนงใหม่เกิดขึ้นเรียกว่าภาษาศาสตร์คอมพิวเตอร์ แบ่งย่อยเป็น 8 สาขา และ 1 ในสาขานี้คือ การแปล

เหตุผลที่สำคัญของผู้พัฒนางานแปลภาษาด้วยเครื่อง (ยืน ภู่วรวรรณ, ชัยยงศ์ วงศ์ชัยสุวัฒน์, 2535) ได้แก่

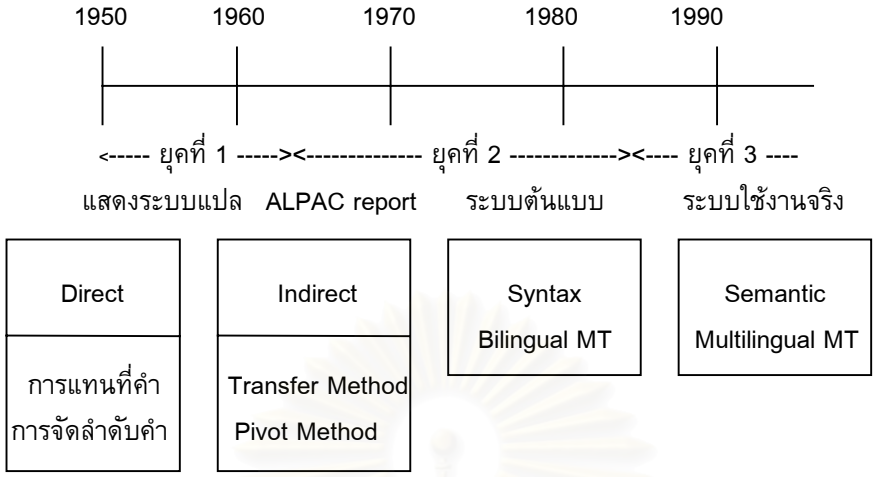
1. ต้องการเอาชนะความหลากหลายของภาษาที่ใช้
2. ราคาและความต้องการของการแปลภาษาสูงขึ้นๆ
3. งานการประมวลผลภาษาธรรมชาติเป็นงานที่ท้าทายความสามารถของมนุษย์ และทำให้มนุษย์สามารถสร้างสิ่งใหม่ได้อีกมาก
4. มีความต้องการสร้างเครื่องจักรให้มีความฉลาดและสามารถโต้ตอบกับมนุษย์ได้ดีขึ้น
5. งานที่เป็นผลพลอยได้จากงานแปลภาษามีอีกมาก

หลังจากที่ได้ผลลัพธ์จากการใช้คอมพิวเตอร์แปลภาษาในรูปแบบต่างๆแล้วการวัดประสิทธิภาพและความถูกต้องของระบบ จะพิจารณาจากผลลัพธ์ดังนี้

1. ประโยคที่ได้ออกมา ถูกต้อง ครบถ้วนสมบูรณ์
2. การแปลประโยคถูกต้อง แต่ยังผิดหลักไวยากรณ์อยู่
3. ประโยคที่ได้ออกมา ยากต่อการเข้าใจ ต้องมีการเปลี่ยนแปลงรูปประโยคบางส่วน
4. บางส่วนของประโยคถูกต้อง แต่บางส่วนยังผิดพลาด
5. การเรียงคำไม่ถูกต้อง และไม่สามารถจัดเรียงให้ถูกต้องได้
6. มีการแปลเฉพาะบางส่วนของประโยคเท่านั้น
7. ไม่มีผลลัพธ์ออกมาให้เลย

การใช้คอมพิวเตอร์ในการแปลภาษาในปัจจุบัน ยังมีความผิดพลาดเกิดขึ้นเป็นเรื่องธรรมดา จึงต้องใช้มนุษย์เข้ามามีส่วนเกี่ยวข้องในการทำให้ผลลัพธ์ที่ได้ออกมาสมบูรณ์แบบ ทั้งนี้ก็เนื่องมาจากการที่แปลภาษานั้นมิได้มีพื้นฐานอยู่ที่ความเข้าใจข้อความที่ได้รับเข้าโดยแท้จริง แต่เป็นการวิเคราะห์จากกฎที่กำหนดให้ ซึ่งภาษาที่ใช้กันอยู่โดยทั่วไปมีความพลิกแพลงอยู่มาก ในบางครั้งอาจไม่ตรงตามกฎที่มีอยู่หรืออาจต้องการภาษาที่สละสลวยมากขึ้น จากการศึกษาผลของการใช้คอมพิวเตอร์ในการแปลภาษานั้น พบว่าจะให้ผลลัพธ์ที่ดีเมื่อเอกสารที่ใช้เป็นเอกสารในด้านเทคนิคและวิทยาศาสตร์ ซึ่งก็เป็นเพราะว่าเอกสารในลักษณะนี้เป็นงานที่เฉพาะเจาะจงและมักใช้ภาษาที่ตรงไปตรงมาเพื่อสื่อความเข้าใจ

สำหรับประวัติและพัฒนาการทางทฤษฎีของการแปลภาษาด้วยเครื่องนั้น แบ่งออกเป็น 3 ยุคดังแสดงไว้ในรูปต่อไปนี้



รูปที่ 1.3 ประวัติและพัฒนาการทางทฤษฎีโดยสังเขปของการวิจัยและพัฒนาระบบเครื่องแปลภาษา (Hirsata Nomuro, 1991)

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาลักษณะและมาตราฐานของภาษา UNL (Universal Network Language)
2. เพื่อออกแบบฐานข้อมูลเพื่อใช้ในการแปลจากภาษา UNL เป็นภาษาไทย
3. เพื่อพัฒนาโปรแกรม “CU-Deconverter”

1.3 ขอบเขตของการวิจัย

1. พัฒนาโปรแกรม “CU-Deconverter” แปลจากเอกสาร UNL ที่อยู่ในรูปแบบ HTML เป็นภาษาไทย
2. ใช้กับบทความหรือข้อความที่มีลักษณะภาษาเป็นทางการ เช่น หนังสือทางวิชาการ, คู่มือการใช้งาน, ข่าว
3. ไม่ใช้กับวรรณกรรมหรือข้อคิดเห็นที่มีอารมณ์ร่วมของผู้แต่ง เช่น นวนิยาย
4. คลอบคลุมเอกสาร UNL ที่ได้รับตัวอย่างจากศูนย์ภาษา UNL หรือที่ได้รับตัวอย่างจากหน่วยปฏิบัติการวิจัยและพัฒนาเทคโนโลยีภาษาและวิทยาการความรู้ ศูนย์เทคโนโลยีอิเล็คทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)

5. ลักษณะของโปรแกรม “CU-Deconverter”

5.1 มีฐานข้อมูล UNL ที่จัดทำขึ้นใหม่เพื่อรองรับการแปลจากภาษา UNL เป็นภาษาไทย

5.2 มีเครื่องมือที่ใช้ในการป้อนหรือแก้ไขฐานข้อมูล UNL ได้แก่

5.2.1 ตารางแสดงลักษณะเฉพาะของคำภาษา UNL

5.2.2 ตารางคำสากลภาษา UNL

5.2.3 ตารางพจนานุกรมยูเอ็นแอล

5.2.4 ตารางประเภทคำภาษา UNL

5.2.5 ตารางคำเชื่อมโยงภาษา UNL

5.2.6 ตารางแสดงคุณลักษณะคำภาษา UNL

5.2.7 ตารางหน้าที่ของคำภาษา UNL

5.3 มีที่ใช้ในการแปลจาก UNL เป็นภาษาไทยโดยมีขั้นตอนการทำงานดังนี้

5.3.1 ตรวจสอบความถูกต้องของเอกสารภาษา UNL

5.3.2 หาความหมายของคำและข้อมูลประกอบทางด้านภาษาไทย

5.3.3 เปรียบเทียบรูปแบบกริยา

5.3.4 จัดเรียงคำ

5.3.5 จัดรูปแบบประโยคและตัดคำซ้ำซ้อน

5.3.6 แสดงผลในรูปแบบประโยคภาษาไทย

5.4 คู่มือประกอบของการใช้เครื่องมือและโครงสร้างของฐานข้อมูล UNL

1.4 ขั้นตอนและวิธีการดำเนินการวิจัย

1. ศึกษามาตราฐานและกฎเกณฑ์ต่างๆ ของระบบภาษา UNL
2. ออกแบบระบบและโมดูลต่างๆ เพื่อจัดทำโปรแกรม “CU-Deconverter”
3. จัดทำฐานข้อมูลพจนานุกรมเพื่อใช้ในการแปลจากภาษา UNL เป็นภาษาไทย
4. พัฒนาโมดูลต่างๆ ของ “CU-Deconverter” เพื่อแปลจากภาษา UNL เป็นภาษาไทย
5. ทดสอบโมดูลและระบบโดยรวมทั้งหมด
6. สรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ต้นแบบของระบบการแปลจากภาษา UNL เป็นภาษาไทย
2. สามารถใช้โมดูลที่พัฒนาขึ้นมาเป็นพื้นฐานในการพัฒนาระบบการประมวลผลภาษาไทยต่อไป

1.6 ผลงานที่ตีพิมพ์จากงานวิจัย

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้ตีพิมพ์และนำเสนอในงานประชุมวิชาการ The Fourth Symposium on Natural Language Processing 2000 เมื่อวันที่ 10-12 พฤษภาคม 2543 ณ โรงแรมเชียงใหม่พลาซ่า จังหวัดเชียงใหม่ ในบทความเรื่อง “The Prototype of UNL to Thai Deconverter Program” โดยผู้นำเสนอคือ Somchai Kitiporn และ Wanchai Rivepiboon

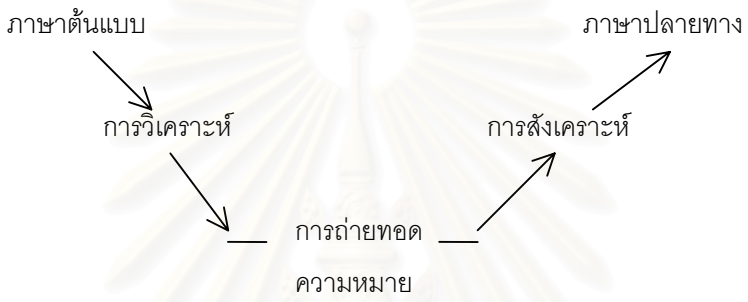
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 หลักการแปลภาษาด้วยคอมพิวเตอร์

หลักการของการแปลภาษาด้วยคอมพิวเตอร์ คือ การใช้คอมพิวเตอร์ช่วยถ่ายทอดความหมายของภาษาต้นแบบไปยังภาษาเป้าหมายที่ต้องการ งานพัฒนาจึงประกอบด้วยการวิเคราะห์ภาษาต้นแบบ การถ่ายทอดความหมาย และการสังเคราะห์ภาษาเป้าหมาย (ยืน ภู่วรวรรณ, ชัยยงค์ วงศ์ชัยสุวัฒน์, 2535)

ระดับการถ่ายทอดจึงเขียนเป็นโมเดลแบบย่อได้เป็น ดังรูปที่ 2.1



รูปที่ 2.1 โครงสร้างการแปลภาษาด้วยการถ่ายทอด

งานแปลภาษาด้วยวิธีนี้ เมื่อแยกเป็นขั้นลำดับแล้วมีขั้นตอนการประมวลผล ดังนี้

ขั้นตอนการวิเคราะห์

เป็นขั้นตอนที่นำเอาภาษาต้นแบบมาวิเคราะห์ในระดับต่างๆ เพื่อแยกความกำกวมของภาษาในระดับต่างๆ ให้ชัดเจน การประมวลผลในขั้นนี้ประกอบด้วย

1. การแยกแยะคำการตัดแบ่งคำการหาขอบเขตของคำตลอดจนดูเครื่องหมายวรรคตอนที่เกี่ยวข้อง
2. การวิเคราะห์ระดับคำเพื่อหาการผันรูปของคำ การเปลี่ยนรูปเนื่องจากวากยสัมพันธ์ของประโยคหารากของคำและลักษณะของคำที่ปรากฏอยู่ในประโยค
3. ทำการแยกแยะความกำกวม ในการจัดแบบชนิดของคำเช่นคำคำหนึ่งเป็นไปได้ทั้งกริยาและคำนาม เราจะมีวิธีการแยกแยะชนิดของคำเหล่านี้ออกมาให้ถูกต้อง

4. การแบ่งแยกอนุประโยคและวลีจัดแบ่งรูปโครงประโยคให้ถูกต้องประโยคบางประโยค มีอนุประโยคประกอบด้วยหลายชั้น จำเป็นต้องแยกออกให้ชัดเจน

5. ประมวลผลในกลุ่มของนามวลี เพื่อแยกแยะส่วนต่างๆ ของนามวลีให้ชัดเจน

6. ทำการประมวลผลในกลุ่มของกริยา ดูรูปโครงสร้างความสัมพันธ์ของกริยาหลักกับคำนามเพื่อกำหนดรูปแบบที่ชัดเจนขึ้น

7.รวมนามวลีและกริยาวลีเข้าด้วยกันเพื่อสร้างโครงสร้างรูปแบบที่เป็นมาตราฐานเดียวกัน

8. แยกแยะความกำกวมในระดับอรรถศาสตร์ (Semantic) ที่อาจเกิดขึ้นได้ในระดับนี้ การวิเคราะห์นี้จำเป็นต้องหาข้อมูลบางอย่างซึ่งต้องเก็บข้อมูลเหล่านั้นไว้ในดิกชันนารี โดยการวิเคราะห์จะเขียนในรูปของกฎเกณฑ์ทางภาษา

ขั้นตอนการถ่ายทอดความหมาย

การถ่ายทอดความหมายเป็นกระบวนการหลังจากการวิเคราะห์การถ่ายทอดความหมายนี้จะใช้ข้อมูลของพจนานุกรมแบบสองภาษา เพื่อถ่ายทอดข้อมูลของภาษาหนึ่งไปยังอีกภาษาหนึ่ง กระบวนการและขั้นตอนของการถ่ายทอดประกอบด้วย

1. จัดรูปแบบโครงสร้างข้อมูลให้เหมาะสมกับการถ่ายทอดกระบวนการนี้มีการจัดทำโครงสร้างความสัมพันธ์ของข้อมูลให้เหมาะสมต่อการแทนความหมาย
2. ถ่ายทอดข้อมูลความหมายระหว่างคู่ภาษา โดยดูจากพจนานุกรมสองภาษา การถ่ายทอดนี้จะต้องมีการเลือกข้อมูลที่เหมาะสมและเป็นไปได้มากที่สุด
3. จัดการเกี่ยวกับความหมายทางด้านปฏิเสธ หรือความหมายที่กำกวม
4. จัดการเกี่ยวกับความหมายของกลุ่มคำนาม
5. ตรวจสอบความถูกต้อง และให้เป็นรูปแบบเดียวกันเพื่อสร้างประโยค
6. ตรวจสอบและหาวิธีการแทนความหมายของคำกริยา และที่เกี่ยวข้องรวมทั้งหารูปแบบของการผันแปรของกริยา
7. ดำเนินการเกี่ยวกับเรื่องคำวิเศษณ์ และคำกริยาวิเศษณ์

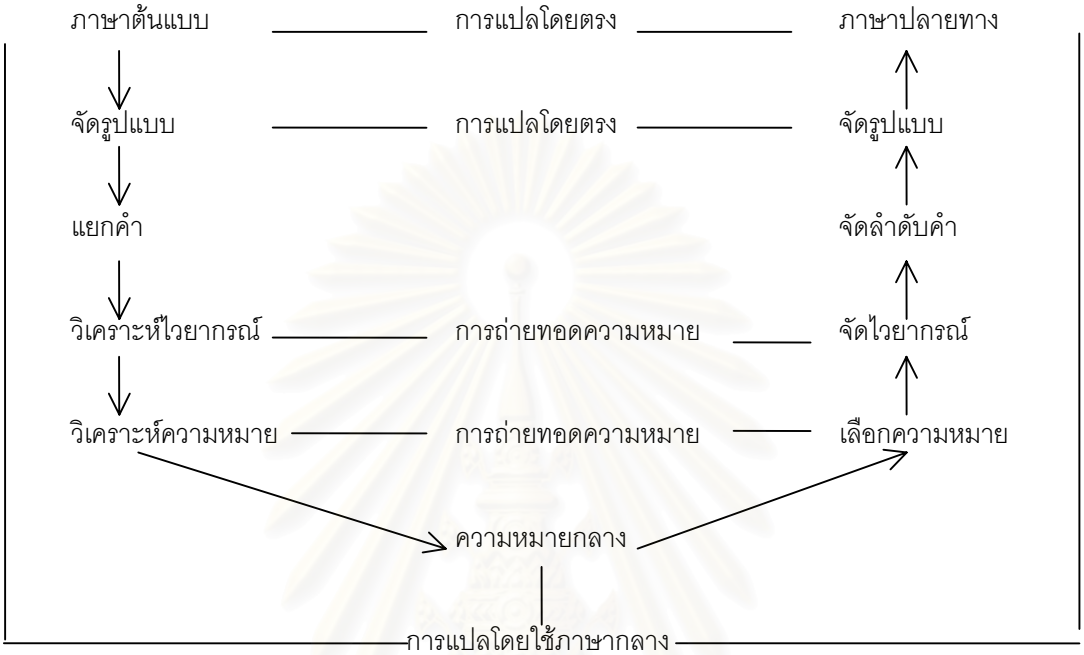
ขั้นตอนการสังเคราะห์ประโยค

การสังเคราะห์ประโยคนับว่าเป็นขั้นตอนที่ง่ายกว่าสองขั้นตอนที่กล่าวมาแล้ว ขั้นตอนย่อยของกระบวนการสังเคราะห์ประกอบด้วย

1. การหารูปแบบไวยากรณ์ที่เหมาะสมของประโยคที่จะสังเคราะห์ดูรูปแบบของคำกริยาที่จะเป็นไปได้จัดเรียงประธานของประโยค
2. จัดการกับคำที่เป็นสำนวน (idiom) หรือวลีเฉพาะบางอย่างที่จำเป็นจะต้องอยู่เป็นกลุ่มคำ
3. จัดเรียงลำดับคำในประโยคให้เหมาะสมกับภาษาปลายทางนั้นๆ ตามการนิยามของภาษา

4. หารูปแบบคำที่สอดคล้องกัน เช่น คำกริยาที่สอดคล้องกับประธาน ดูรูปแบบของคำที่เป็นไปได้ที่เหมาะสมในระดับคำ
5. สร้างคำสรรพนามและคำนามประสมให้ถูกต้อง

แนวความคิดทางด้านการใช้คอมพิวเตอร์ช่วยในการแปลภาษาสามารถแบ่งได้เป็น 3 ประการหลักๆ ดังแสดงเป็นรูปแบบโครงสร้างในรูปที่ 2.2



รูปที่ 2.2 แสดงโครงสร้างของการใช้คอมพิวเตอร์ในการแปลภาษาวิธีต่างๆ

2.1.1 วิธีการแปลโดยตรง (Direct Translation Approach)

เป็นวิธีที่ง่ายที่สุดแต่จะเหมาะสมกับการใช้งานบางประเภทและมีข้อจำกัดอยู่มากวิธีการทำงานก็จะรับข้อความที่ต้องการจะแปล และทำการค้นหาคำที่ตรงกันในฐานข้อมูลเพื่อได้คำแปลออกมา ดังแสดงในรูปที่ 2.3 ระบบงานนี้จะมีการเก็บข้อมูลในระดับคำหรือระดับประโยคไว้ในฐานข้อมูลในลักษณะเป็นคู่ระหว่างภาษาที่สนใจ ดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 ตัวอย่างฐานข้อมูลคู่ระหว่างภาษาไทยและภาษาอังกฤษในระดับประโยค

ประโยคภาษาไทย	ประโยคภาษาอังกฤษ
ผมไปทำงานทุกวัน	I go to work everyday.
คุณต้องการกาแฟบ้างไหม	Do you want some coffee ?
เครื่องเสียงชุดนั้นราคาเท่าไร	How much does that stereo set cost ?
แท็กซี่ มาทางนี้	Taxi! This way, please.

รูปที่ 2.3 วิธีการแปลโดยตรง

2.1.2 **วิธีการแปลโดยการถ่ายทอดความหมาย** (Transfer Approach)

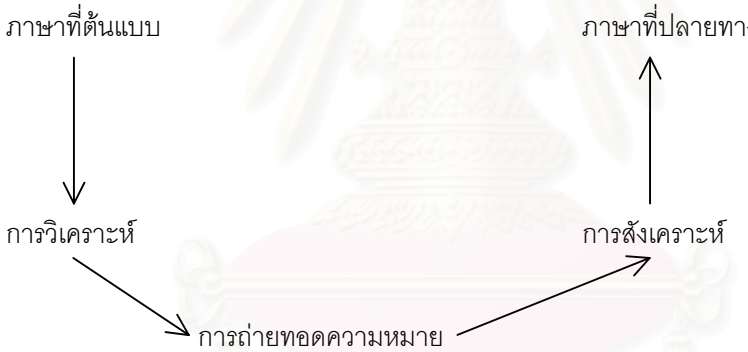
เป็นวิธีที่นิยมใช้กันมากแบ่งออกได้เป็น 3 ขั้นตอนย่อยๆ ดังนี้

2.1.2.1 ขบวนการวิเคราะห์ความกำกวมของประโยคที่เข้ามาทำการแบ่งคำเป็นส่วนค้นหาขอบเขตของคำตรวจสอบรากของคำวิเคราะห์ว่าเป็นคำประเภทใดในกรณีเป็นได้หลายประเภท

2.1.2.2 ขบวนการถ่ายทอดความหมาย โดยอาศัยพนจานุกรมสองภาษาสำหรับการแปลตรวจสอบความถูกต้องของรูปแบบเพื่อเตรียมสร้างประโยคจัดการคำคุณศัพท์และกริยาวิเศษณ์

2.1.2.3 ขบวนการสร้างประโยค จัดการเกี่ยวกับรูปแบบของกริยา จัดเรียงประธานของประโยครวมทั้งคำต่างๆ ให้อยู่ในลำดับที่ถูกต้อง จัดการคำกริยาให้เหมาะสมกับประธาน

วิธีการแปลโดยการถ่ายทอดความหมายนี้ สามารถแสดงเป็นแผนภาพได้ดังในรูปที่ 2.4



รูปที่ 2.4 วิธีการแปลโดยใช้การถ่ายทอดความหมาย

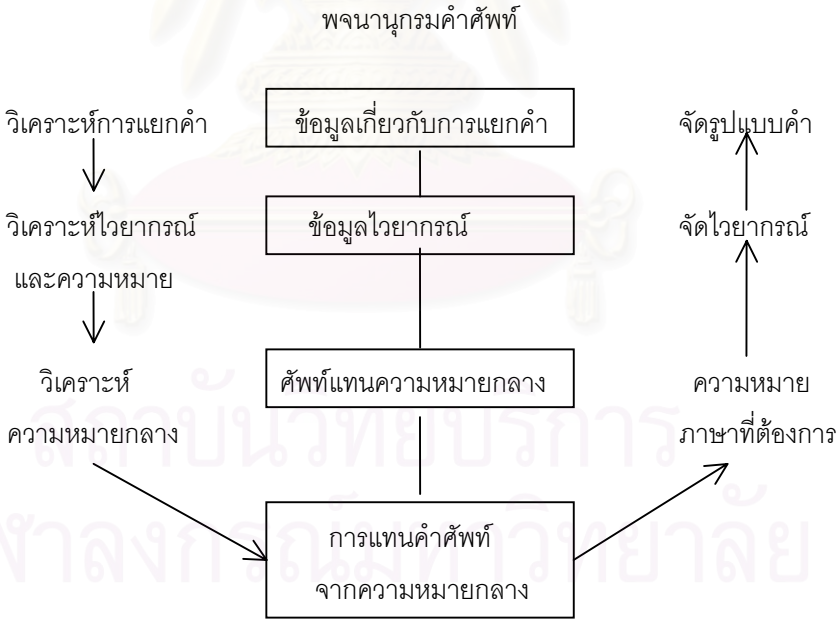
ในการแปลแบบนี้ ก็จะต้องมีฐานข้อมูลเช่นเดียวกัน แต่ข้อมูลที่เก็บไว้จะแตกต่างจากวิธีการแปลโดยตรง หลังจากที่แบ่งคำแล้ว ในขั้นตอนการถ่ายทอดความหมายจะต้องตรวจสอบและวิเคราะห์วิธีการใช้คำนั้นด้วย เช่น คำ “ถูก” เป็นคำที่สามารถมีที่ใช้หลายรูปแบบ ก็ต้องวิเคราะห์ว่ามีประธานและกรรมในประโยคหรือไม่ และคำเหล่านั้นเป็นประเภทใด เมื่อได้ข้อมูลครบถ้วนแล้วจึงทำการแปลคำนั้นให้เป็นภาษาอังกฤษที่เหมาะสมต่อไป ดังในตารางที่ 2.2

ตารางที่ 2.2 โครงสร้างของพจนานุกรม 2 ภาษา

คำศัพท์	ชนิดของคำ	ความหมาย
ถูก	((VERB(SUBJ ANI)(OBJ THING))	TOUCH
ถูก	((ADJ)(SUBJ THING))	CHEAP
ถูก	((ADJ)(SUBJ MAN))	RIGHT
ฉัน	((MAN SUBJ))	I
คุณ	((MAN SUBJ))	YOU
รถ	((VEHICLE))	CAR
คอมพิวเตอร์	((THING))	COMPUTER

2.1.3 **วิธีการแปลโดยใช้ภาษากลาง** (Interlingua Approach)

วิธีการแปลแบบนี้จะมีการกำหนดภาษากลาง (Interlingua) มาก่อนแล้วทำการแปลในลักษณะคล้ายคลึงกับการแปลโดยใช้การวิเคราะห์ประโยคเพื่อให้ออกมาเป็นภาษากลางก่อนแล้วจึงเลือกแปลงเป็นภาษาอื่นๆ ต่อไป ดังแสดงในรูปที่ 2.5 วิธีการนี้เหมาะสมเป็นอย่างยิ่งในระบบที่ต้องการสนับสนุนการแปลหลายภาษาเป็นวิธีการที่ได้รับบความสนใจในหลายๆ ประเทศ เช่น ญี่ปุ่น จีน ประเทศในแถบเอเชียและยุโรป



รูปที่ 2.5 วิธีการแปลโดยใช้ภาษากลาง

2.2 แนวทฤษฎีทางไวยากรณ์และเทคนิคการวิเคราะห์

คอมพิวเตอร์มิได้มีความฉลาดเช่นมนุษย์การที่คอมพิวเตอร์จะสามารถวิเคราะห์ความหมายของภาษาได้ก็จะต้องนำความรู้ เกี่ยวกับการวิเคราะห์ภาษาที่มีอยู่ในสมองของคนเราใส่ให้ ปัจจุบันความรู้ที่ใช้ในการวิเคราะห์มีหลายแนวคิดด้วยกันเช่นแนวคิดเกี่ยวกับไวยากรณ์โครงสร้างวลี (phrase structure grammar) ไวยากรณ์การก (case grammar) ไวยากรณ์พึ่งพา (dependency grammar) ไวยากรณ์หน้าที่ (functional grammar) และไวยากรณ์ศัพท์การก (lexicase) เป็นต้น โครงการวิเคราะห์ภาษาไทยที่คณะผู้เขียนดำเนินการวิจัยอยู่นี้ ประกอบกันเพื่อการวิเคราะห์ความหมายของประโยค โดยใช้การก (case) ที่กำหนดขึ้นในภาษากลางแสดงความสัมพันธ์ของคำ ส่วนตัวระบบอาศัยเทคนิคและซอฟท์แวร์การวิเคราะห์ของระบบ PIVOT ซึ่งได้รับการสนับสนุนจากบริษัท NEC ประเทศญี่ปุ่น (Machine Translation, 2536)

อย่างไรก็ตามการแสดงองค์ประกอบของประโยคเป็นวลีต่าง ๆ ข้างต้น ยังไม่เพียงพอเพราะไวยากรณ์นี้เพียงแต่รวมหน่วยต่างๆขึ้นเป็นวลีและประโยคเท่านั้นไม่ได้แสดงข้อมูลของโครงสร้างทางภาษาทั้งในแง่ของวากยสัมพันธ์และอรรถศาสตร์ ดังนั้น จึงได้นำเอาแนวทฤษฎีของไวยากรณ์การกเข้ามาช่วยในการแสดงความสัมพันธ์ของวลีต่าง ๆ (Machine Translation, 2536)

ตัวอย่างแนวคิดเกี่ยวกับไวยากรณ์

2.2.1 ไวยากรณ์โครงสร้างวลี (Phrase Structure Grammar)

แนวคิดสำคัญของไวยากรณ์โครงสร้างวลีนี้คือ การรวมหน่วยย่อยให้เป็นหน่วยใหญ่กว่า ซึ่งก็คือการรวมคำให้เป็นวลี (phrase) เช่น กริยาวลี (verb phrase) และนามวลี (noun phrase) เป็นต้น และวลีต่างๆจะรวมเป็นหน่วยที่ใหญ่กว่าขึ้นไป ซึ่งก็คือ ประโยค (sentence)

ในการแจงส่วนประโยคได้อาศัยแนวคิดของทฤษฎีนี้มาสร้างกฎโครงสร้างวลี (phrase structure rules) ขึ้นเพื่อรวมหน่วยต่างๆ (lexical items) ให้มีโครงสร้างเป็นวลีและเป็นประโยค ตัวอย่างของกฎโครงสร้างวลีสำหรับภาษาไทยมีรายละเอียดดังนี้

กฎโครงสร้างวลีภาษาไทยประโยคความเดียว

S ======> (NP) VP (NP) (NP) (PP) (ADVP)

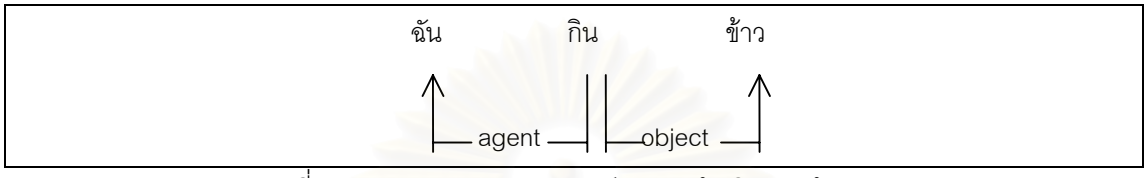
VP ======> (PRE-AUXP) (NEG) V (POST-AUXP)

ข้อความในวงเล็บหมายถึง คำหรือวลีที่สามารถละไว้ได้

2.2.2 ไวยากรณ์การก (Case Grammar)

แนวคิดสำคัญของไวยากรณ์การกก็คือคำนามในประโยคจะมีการกสัมพันธ์ (case relation) อันเป็นความสัมพันธ์ทางความหมายกับคำกริยาหลักเสมอ และไม่ว่าคำนั้นจะปรากฏในตำแหน่งใดในประโยค ก็ย่อมจะมีการกสัมพันธ์เดียวกันกับกริยาทั้งสิ้น แนวคิดนี้เป็นแนวคิดสากลของภาษา (universal language) ซึ่ง

สอดคล้องกับระบบการแปลแบบพหุภาษา การวิเคราะห์ภาษาไทยอาศัยแนวคิดนี้เป็นหลักในการสร้างความสัมพันธ์ทางความหมายของคำกริยาและคำนาม การแสดงการกสัมพันธ์นั้นจะเริ่มจากการหากรอบการก (case frame) ให้กับคำกริยาต่าง ๆ กรอบการกนี้จะแสดงความสัมพันธ์ระหว่างคำกริยาและคำนามที่อยู่ในรูปแบบบังคับเกิด (obligatory case) ส่วนรูปแบบการกให้เลือกเกิด (optional case) นั้นจะปรากฏในกฎการเลือกการกที่ได้จากการวิเคราะห์ความหมายของคำต่าง ๆ ที่สัมพันธ์กัน ระบบการแปลภาษาหลายระบบจึงนำแนวคิดนี้มาประยุกต์ใช้ในการหาความสัมพันธ์ทางความหมายของคำกริยาและคำนาม



รูปที่ 2.6 ตัวอย่างของการกสัมพันธ์ระหว่างคำกริยากับคำนาม

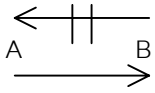
2.2.3 ไวยากรณ์พึ่งพา (Dependency Grammar)

ไวยากรณ์พึ่งพามีแนวคิดสำคัญคือ หน่วยต่าง ๆ ย่อมมีความสัมพันธ์ที่พึ่งพากัน (dependency relation) กล่าวคือ ความสัมพันธ์จะอยู่ในรูปของหน่วยหนึ่งที่ทำหน้าที่เป็นหน่วยหลัก (head) และอีกหน่วยหนึ่งจะเป็นหน่วยขยายหรือหน่วยพึ่งพา (modifier or depender) แนวคิดการแสดงความสัมพันธ์แบบพึ่งพานี้สามารถให้คำอธิบายทางภาษาได้ใกล้เคียงภาษาธรรมชาติมาก เพราะสามารถแสดงการเชื่อมต่อกันระหว่างหน่วยต่าง ๆ เป็นลำดับ (heirarchical link) ซึ่งสอดคล้องกับลักษณะทางวากยสัมพันธ์ (syntactic structure) ของภาษาต่าง ๆ เช่น ข้อตกลง (agreement) ระหว่างคำกริยากับคำนาม การใช้ตัวบ่งชี้ (marker) และการเรียงลำดับของคำ (word order) ผลของการวิเคราะห์โดยใช้ไวยากรณ์พึ่งพานี้ จะได้โครงข่ายทางความหมายของประโยค ดังตัวอย่าง

รูปที่ 2.7 ตัวอย่างของความสัมพันธ์แบบหน่วยหลัก-หน่วยพึ่งพา

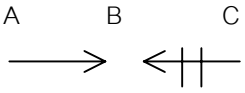
ข้อตกลงเบื้องต้นของการแสดงความสัมพันธ์แบบพึ่งพามีดังต่อไปนี้

2.2.3.1 ความสัมพันธ์แบบพึ่งพาย่อมไม่เป็นแบบที่ทั้งสองโนด (node) ต่างก็เป็นหน่วยหลัก (Head) และหน่วยพึ่งพา (depender) ของกันและกัน เช่น



คำอธิบาย : ในที่นี้ B เป็นหน่วยพึ่งพา A ดังนั้น B จะเป็นหน่วยหลักของ A ไม่ได้

2.2.3.2 หน่วยใด ๆ จะเป็นหน่วยพึ่งพาของหน่วยหลักมากกว่าหนึ่งหน่วยไม่ได้ เช่น



คำอธิบาย : B จะเป็นหน่วยพึ่งพาของทั้ง A และ C ไม่ได้

2.2.3.3.ทุก ๆ โนดย่อมมีความสัมพันธ์กับโนดอื่น ๆ เสมอเช่น ในโครงสร้างความรวม (coordinated structure) ที่มีคำเชื่อม “และ” จะกำหนดให้หน่วยที่อยู่ข้างหน้าเป็นหน่วยหลัก

2.3 ภาษากลาง (Interlingua)

ภาษากลางเป็นภาษาทางปัญญาประดิษฐ์ ที่ถูกกำหนดขึ้นเพื่อแสดงความหมายของประโยคของภาษาต้นฉบับที่จะแปล ซึ่งมีลักษณะเป็นกลางไม่ขึ้นกับภาษาใด (Language Independent) และสามารถใช้แทนทุกภาษา (Multilanguage Representation) การวิจัยทางด้านการกำหนดรูปแบบของภาษากลาง เพื่อแทนความหมายของภาษาต่างๆ นั้นมีอยู่ในระบบแปลภาษาหลายระบบ[1] ซึ่งโดยส่วนใหญ่มีการกำหนดโครงสร้างของภาษากลางเป็น 2 ส่วน ได้แก่ส่วนแสดงหน่วยความหมาย (Unit of meaning) ของคำในภาษาต้นฉบับ เรียกว่า Conceptual Primitive ย่อว่า CP และส่วนแสดงความสัมพันธ์ของหน่วยความหมาย เพื่อแทนความหมายของเนื้อหาในภาษาต้นฉบับ เรียกส่วนนี้ว่า ไวยากรณ์ภาษากลาง (Interlingual Grammar) ในการกำหนดส่วนแสดงความสัมพันธ์ของหน่วยความหมายเพื่อแทนคำในแต่ละภาษา จะต้องพิจารณาความหมายลึก (deep meaning หรือ concept) ของคำนั้น ประกอบกับคุณสมบัติ (property) ของคำนั้นๆ ด้วย โดยการจัด หมวดหมู่ของคำไปตามข้อกำหนดทั้งสองสำหรับภาษากลางที่ใช้ในงานวิจัย ก็มีลักษณะดังกล่าวข้างต้น โดยในส่วนการแสดงความสัมพันธ์ของส่วนแสดงความสัมพันธ์ของหน่วยความหมายนั้นอาศัยทฤษฎีทางการการกสัมพันธ์ (Case Relation) และแสดงโครงสร้างภาษากลาง ในรูปความสัมพันธ์พึ่งพิง (Conceptual Structure หรือ Dependency Tree Representation) ส่วนประกอบภาษากลางที่เป็นอินพุตของระบบการสร้างภาษาไทยของงานวิจัยมีดังนี้ (ดร.รัตติกร วรากูลศิริพันธุ์ และคณะ, 2536)

2.3.1. ส่วนแสดงหน่วยความหมายของคำ (Conceptual Primitive) ย่อว่า CP : เป็นส่วนที่ใช้แทนคำในแต่ละภาษา ที่มีความหมายลึกและลักษณะการใช้ในประโยคเหมือนกัน โดยมีการจัดกลุ่มคำเหล่านั้นไว้ใน concept เดียวกัน เรากำหนดสัญลักษณ์สำหรับแทน “CP” ของกลุ่มคำในแต่ละภาษา ดังกล่าวเช่นไรก็ได้ในงาน

วิจัยจะใช้ภาษาอังกฤษ (อักษรตัวพิมพ์ใหญ่) แทนส่วนแสดงความสัมพันธ์ของหน่วยความหมาย เช่นตัวอย่างในรูปที่ 11 “MOVE” เป็นส่วนแสดงความสัมพันธ์ของหน่วยความหมายของ move, moving, เคลื่อน “ROOM” เป็นส่วนแสดงความสัมพันธ์ของหน่วยความหมายของ room, ห้องเป็นต้น

2.3.2. การก เป็นส่วนแสดงความสัมพันธ์ระดับลึกของคำที่ถูกแทนด้วยส่วนแสดงความสัมพันธ์ของหน่วยความหมาย เช่นตัวอย่างรูปที่ 11 <ACT> เป็น case ที่บ่งว่า “HE” เป็นประธานของ “MOVE” และ <OBJ> เป็น case ที่บ่งว่า “CHAIR” เป็นกรรมของ “MOVE” และ <LOC> เป็น case ที่บ่งว่า “ROOM” เป็นสถานที่เกิดเหตุการณ์ของภาคแสดงคือ “MOVE”

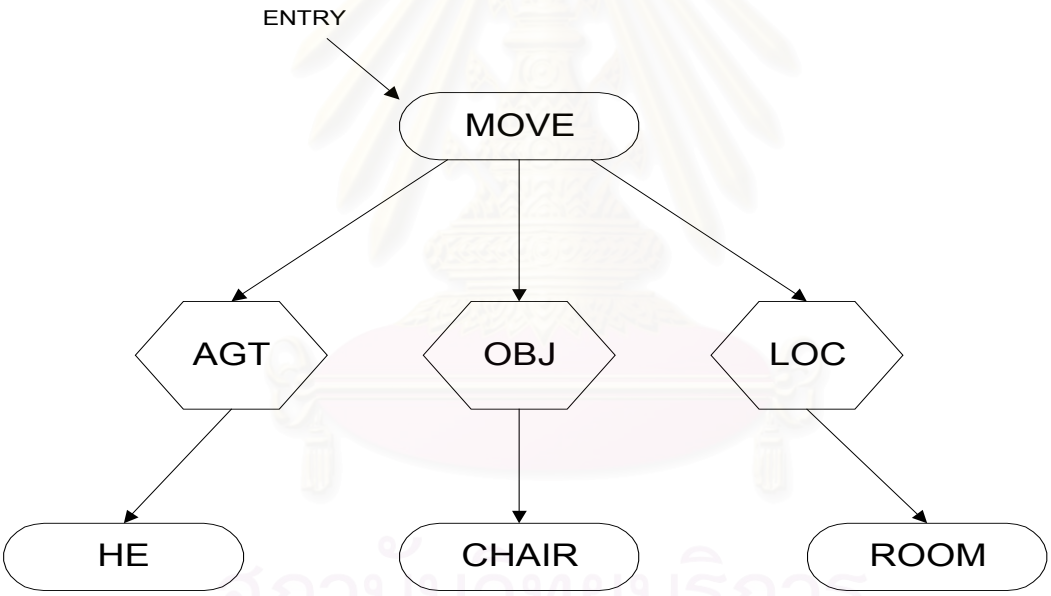
2.3.3. กริยาหลัก (Main Predicate) เป็นส่วนที่เป็นแกนกลางของประโยคโดยจะพิจารณาจาก

2.3.3.1 คำกริยา เช่น เคลื่อน กิน วิ่ง เดิน นั่ง อ่าน ร้อง

2.3.3.2 คำคุณศัพท์ที่ทำหน้าที่คล้ายคำกริยา เช่น สวย หอม ใหญ่ สูง

2.3.3.3 คำกริยาช่วย เช่น เป็น อยู่ คือ verb to be, there be, have

ในโครงสร้างของภาษากลางนี้ กริยาหลักจะเป็นจุด หรือโนดเริ่มต้น (entry node) ของการวิเคราะห์และสร้างประโยค ตัวอย่างรูปที่ 2.8 คือ “MOVE”



รูปที่ 2.8 โครงสร้างภาษากลางของประโยค “He moves a chair in the room”

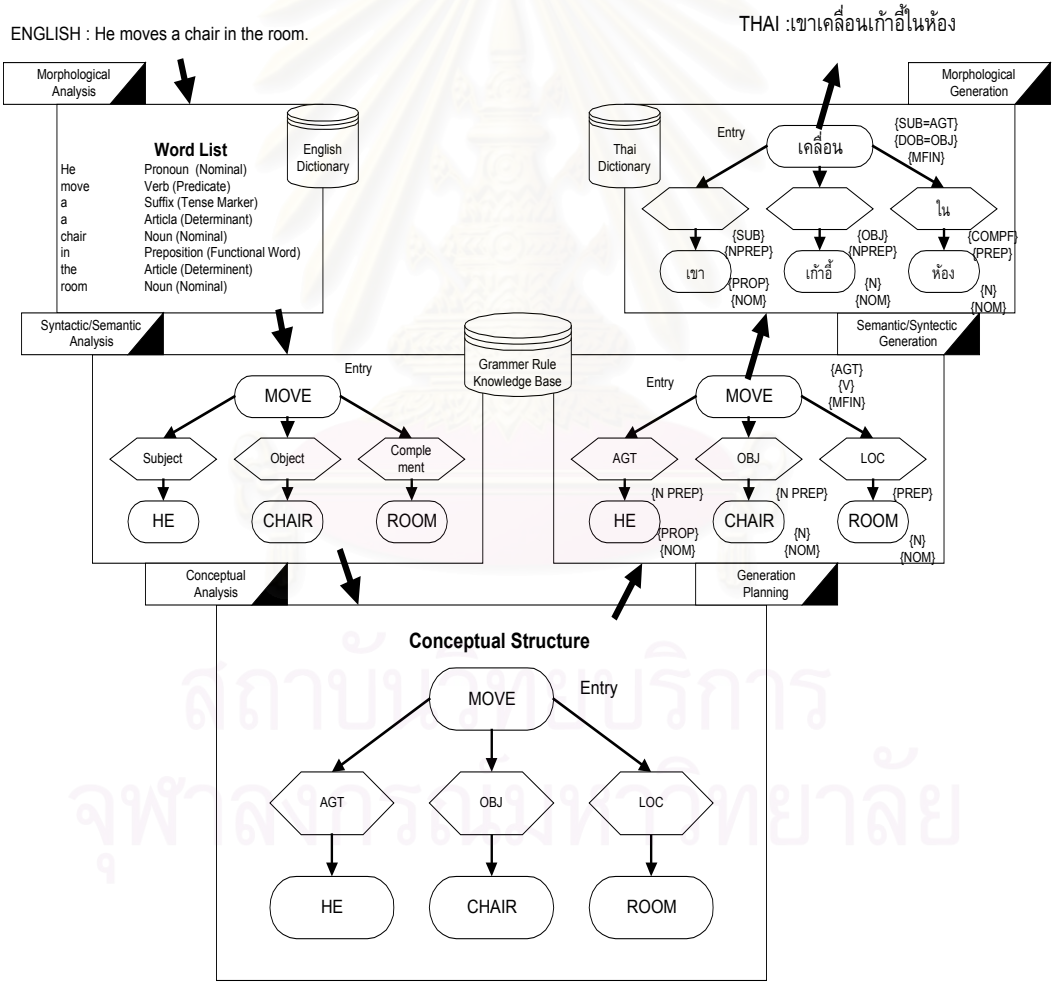
(ดร.รัตติกร วรากูลศิริพันธุ์, 1989)

2.3.4. ส่วนแสดงโครงสร้างในรูปแบบต้นไม้ (Dependency Tree) เป็นส่วนแสดงโครงสร้างของภาษากลางในรูปแบบของต้นไม้ ประกอบด้วย “โนด” และ “อาร์ค” โนดจะแทนส่วนแสดงความสัมพันธ์ของหน่วยความ

หมายและการก ส่วนอาร์คจะแสดงความสัมพันธ์ของโนดในลักษณะของ head-dependent relation ดังตัวอย่างรูปที่ 2.8 “MOVE” เป็น head ของ “HE” ในทางตรงกันข้าม “HE” เป็น dependent ของ “MOVE”

2.4 ระบบการสร้างภาษาไทยจากภาษากลาง (Thai Generation System from Interlingua)

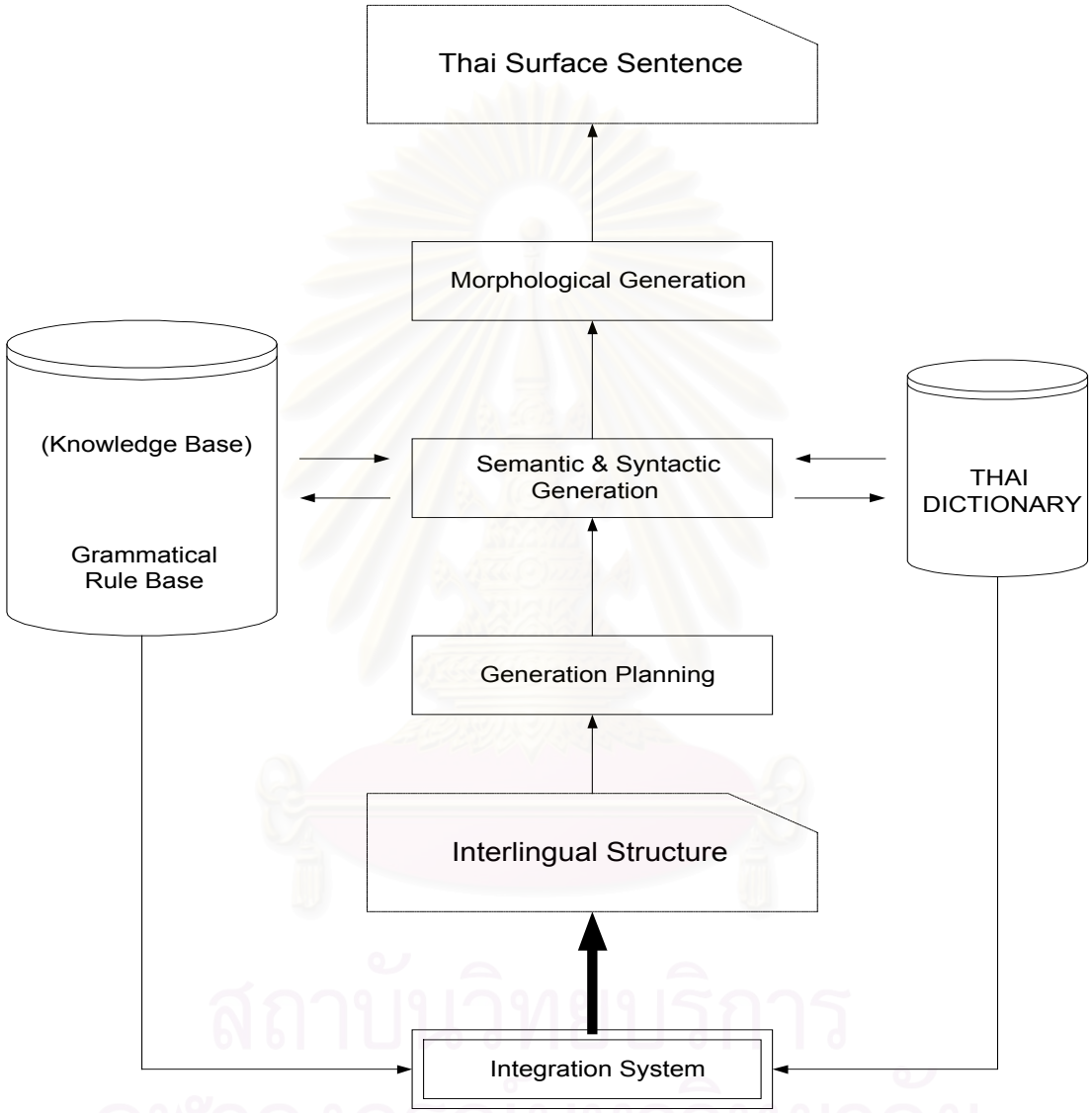
การสร้างประโยคภาษาไทยจากภาษากลางจะอาศัยข้อมูลที่ได้จากฝ่ายวิเคราะห์ภาษาต้นแบบซึ่งเก็บไว้ในภาษากลางเป็นลักษณะสำคัญ (feature) ของส่วนแสดงหน่วยความหมายของคำและการกพร้อมกับข้อมูลของคำเหล่านั้นจากพจนานุกรมภาษาไทย ระบบจะต้องค้นหาคำในภาษาไทยที่ทำหน้าที่และมีความหมายตรงกับส่วนแสดงหน่วยความหมายของคำและการกในภาษากลาง และจะต้องตัดคำกำกวม โดยเลือกเอาเฉพาะคำที่ถูกต้องที่สุด ในขั้นตอนสุดท้ายจะต้องทำการเรียงคำ ให้อยู่ในตำแหน่งที่ถูกต้องพร้อมกับมีการตกแต่งประโยคให้ถูกต้องและสมบูรณ์ ตามหลักไวยากรณ์ภาษาไทย ดังตัวอย่างที่แสดงในรูปที่ 2.9



รูปที่ 2.9 ตัวอย่างการแปลจากภาษาอังกฤษเป็นภาษาไทย แสดงให้เห็นถึงขั้นตอนการทำงานทั้งหมดของระบบการแปลภาษาด้วยคอมพิวเตอร์ของงานวิจัย (ดร.รัตติกร วรากูลศิริพันธุ์, 1989)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โครงสร้างของระบบการสร้างภาษาไทยในงานวิจัย (Text Generation System for Machine Translation) แบ่งออกเป็น 3 ขั้นตอน ได้แก่ การเตรียมข้อมูลสำหรับการสร้างประโยคภาษาไทย (Generation Planning) การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย (Semantic and Syntactic Generation) และการกำหนดคำภาษาไทย (Morphological Generation) ดังแสดงไว้ในรูปที่ 2.10



รูปที่ 2.10 โครงสร้างของระบบการสร้างภาษาไทยจากภาษากลาง

2.4.1. การเตรียมข้อมูลสำหรับการสร้างประโยคภาษาไทย

ในส่วนนี้จะทำการดึงข้อมูลจากพจนานุกรมภาษาไทย โดยใช้ส่วนแสดงหน่วยความหมายของคำจากภาษากลางเป็นคำหลักหรือตัวกำหนดตำแหน่งของข้อมูลจากพจนานุกรม ข้อมูลจะประกอบด้วยคำในภาษาไทยและคุณสมบัติทางด้านไวยากรณ์ของคำนั้นๆ ได้แก่

2.4.1.1. คุณสมบัติทางด้านประเภทของคำ (Type Category) ย่อว่า TCAT ของคำ ได้แก่การเป็นนาม(N) สรรพนาม(PRON) กริยา(V) คุณศัพท์(ADJ) กริยาวิเศษณ์(ADV) บุพบท(PREP) สันธาน(CONJ) เป็นต้น

2.4.1.2. คุณสมบัติทางด้านโครงสร้างที่บ่งความสัมพันธ์ระดับลึกและระดับผิวของคำนั้นๆ กับคำอื่น โดยเฉพาะกริยาหลัก เราเรียกข้อมูลนี้ว่า Type mapping information ย่อว่า TMAP เช่นกริยา “เคลื่อน” ในภาษาไทยถ้าเป็นสกรรมกริยาจะมี TMAP เป็น {SUB=AGT, DOB=OBJ, COMP=TAR} และถ้าเป็นอกรรมกริยาจะมี TMAP เป็น {SUB=OBJ, COMP=TRA}

นอกจากนั้น จะมีข้อมูลอื่นๆ เช่น รูปแบบกริยา ข้อมูลดังกล่าวทั้งหมดจะถูกเก็บอยู่ในกลุ่มอรรถศาสตร์ (Semantic Block) ของพจนานุกรม ซึ่งจะถูกนำมาเปรียบเทียบกับข้อมูลที่ได้รับมาจากการวิเคราะห์ภาษาต้นแบบ เพื่อเตรียมการสร้างภาษาไทยโดยยึด กริยาหลักเป็นหลักในการค้นหาข้อมูลที่ถูกต้องด้วยการ traversing แบบ depth-first algorithm ดังแสดงในรูปที่ 2.11

หน้าที่หลักในส่วนของการเตรียมข้อมูลสำหรับการสร้างประโยคภาษาไทยมีดังต่อไปนี้คือ

1) เปลี่ยนโครงสร้างของ conceptual structure ในกรณีจำเป็น ได้แก่การเพิ่มโนดใหม่ที่เป็นส่วนแสดงหน่วยความหมายของคำและกาลก

2) กำหนดประธานของประโยคให้กับกริยาหลัก

3) กำหนดค่าประเภทของคำให้กับส่วนแสดงหน่วยความหมายของคำแต่ละตัว

4) กำหนดค่า TPLAN ให้กับส่วนแสดงหน่วยความหมายของคำแต่ละตัว ได้แก่

4.1 กำหนดค่า Main Finite ให้กับกริยาหลัก = TPLAN. {MFIN}

4.2 กำหนดค่า Finite ให้กับ กริยาย่อยตัวอื่น = TPLAN. {FIN}

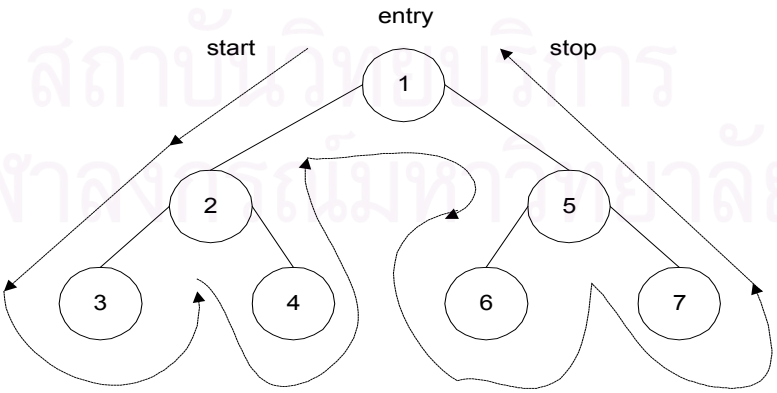
4.3 กำหนดค่า nominal ให้กับ CP ที่เป็นนามและสรรพนาม = TPLAN. {NOM}

5) กำหนดค่า TPLAN ให้กับ CASE แต่ละตัว ได้แก่

TPLAN{PREP} = บุพบท,TPLAN{NPREP} = ไม่มีบุพบท,TPLAN{CONJ} = สันธาน

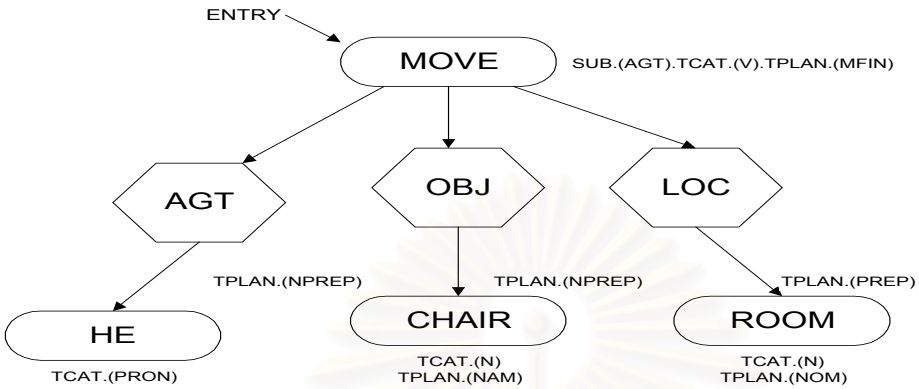
6) กำหนดค่ารูปแบบกริยาให้กับกริยาหลัก

7) ตัดส่วนกำกวมของประโยค



รูปที่ 2.11 หลักการ Traversing แบบ depth-first หรือ top-down & left-right algorithm

ในบทความวิจัย(ดร.รัตติกร วรากูลศิริพันธุ์, 1989) จะใช้ประโยคภาษาอังกฤษ “He moves a chair in the room.” เป็นตัวอย่างที่จะแปลเป็นภาษาไทย ประโยคนี้จะมีรูปแบบภาษากลางดังแสดงในรูปที่ 11 เมื่อภาษากลางนี้ผ่านส่วน Generation Planning จะได้ผลลัพธ์สุดท้ายของประโยคนี้ ดังแสดงไว้ในรูปที่ 2.12



รูปที่ 2.12 ผลลัพท์ที่เป็นส่วนเอาท์พุทของ Generation Planning
สำหรับการแปลประโยค “He moves a chair in the room”

2.4.2 การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย

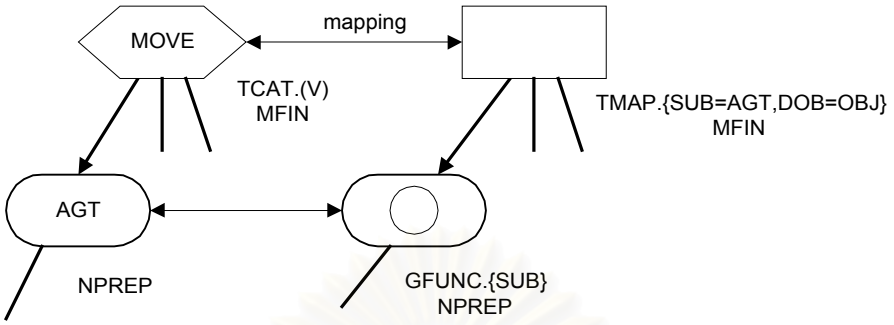
2.4.2.1 การสร้างโครงสร้างประโยคทางไวยากรณ์ (Grammatical Structure Construction)

ในส่วนนี้ทำการสร้างโครงสร้างทางไวยากรณ์ของภาษาไทยให้กับภาษากลาง โดยทำการกำหนดหน้าที่ของส่วนแสดงหน่วยความหมายของคำ แต่ละตัวที่เชื่อมต่อจากกิริยาหลักตามหลักไวยากรณ์ไทย และใส่ค่านั้น ๆ ไว้ที่การกที่แสดงความสัมพันธ์ระหว่างกิริยาหลักกับส่วนแสดงหน่วยความหมายของคำตัวนั้น ๆ ได้แก่

1. GFUNC. (SUB) สำหรับส่วนแสดงหน่วยความหมายของคำที่เป็นประธาน
2. GFUNC. (DOB) สำหรับส่วนแสดงหน่วยความหมายของคำที่เป็นกรรมตรง
3. GFUNC. (IOB) สำหรับส่วนแสดงหน่วยความหมายของคำที่เป็นกรรมรอง
4. GFUNC. (COMP) สำหรับส่วนแสดงหน่วยความหมายของคำที่เป็นส่วนขยายประโยค ที่ขาดไม่ได้
5. GFUNC. (COMPF) สำหรับส่วนแสดงหน่วยความหมายของคำที่เป็นส่วนขยายประโยคทั่วไป

รูปที่ 2.13 แสดงการกำหนดโครงสร้างทางไวยากรณ์ให้กับประโยคตัวอย่าง โดยใช้วิธีการ

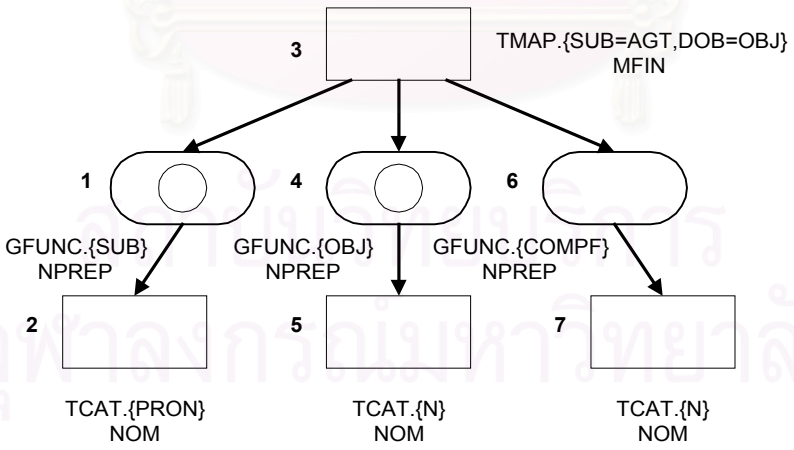
เทียบConceptual CP เข้ากับ Grammatical Node ซึ่งจะได้ผลลัพธ์สำหรับตัวอย่างนี้ในรูปที่ 2.14 หลังจากนั้นระบบจะทำการเรียงลำดับของคำที่ถูกต้องตามหลักไวยากรณ์ไทยในขั้นตอนต่อไป



รูปที่ 2.13 ตัวอย่างการกำหนดโครงสร้างทางไวยากรณ์

2.4.2.2. การเรียงลำดับของคำภาษาไทย (Thai's Words Ordering)

การเรียงลำดับของคำภาษาไทยที่อยู่บนโครงสร้างไวยากรณ์นั้น ก็คือการจัดลำดับของโนดส่วนแสดงหน่วยความหมายของคำและโนดการก เพื่อที่จะนำมาเรียงกันให้เป็นประโยคในแนวเส้นตรง ทั้งนี้โดยอาศัยข้อมูลที่กำหนดตำแหน่งของคำที่อยู่ข้างเคียงกันตามหลักโครงสร้าง (Syntax) ของภาษาไทยมาเป็นตัวตัดสิน ดังนั้นจึงต้องมีการกำหนดกฎการเรียงลำดับคำหรือกลุ่มคำ โดยแบ่งตามประเภทของคำ และหน้าที่ของคำนั้น ๆ และใช้กฎดังกล่าวเป็นฐานข้อมูลในการเปรียบเทียบอ้างอิง ของการกำหนดลำดับของโนดส่วนแสดงหน่วยความหมายของคำและโนดการกที่อยู่ในโครงสร้างต้นไม้ เพราะฉะนั้นการทำงานในส่วนนี้ จะประกอบไปด้วยส่วนสำคัญ 2 ส่วนคือ กฎทางลอจิกเพื่อกำหนดลำดับของโนดที่อยู่ข้างเคียงกันบนโครงสร้างรูปแบบต้นไม้ และส่วนที่เป็นฐานข้อมูลของกฎไวยากรณ์การเรียงลำดับคำและกลุ่มคำในภาษาไทย ซึ่งจะกล่าวโดยละเอียดในลำดับต่อไป



รูปที่ 2.14 ผลลัพธ์ของโครงสร้างทางไวยากรณ์ภาษาไทย สำหรับการแปลประโยค

"He moves a chair in the room."

2.4.2.2.1 กฎเพื่อกำหนดลำดับของโนดข้างเคียง (Neighboring Nodes Ordering Rules)

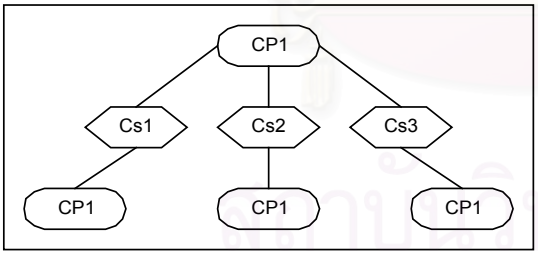
จากโครงสร้างของภาษากลางที่กล่าวมาในข้อ 2.4.2.2 นั้น เราได้กำหนดโนดมี 2 ประเภทคือส่วนแสดงหน่วยความหมายของคำและโนดการก โดยที่โนดแม่ (Parent or Root Node) ของแต่ละกิ่งจะเป็นส่วนแสดงหน่วยความหมายของคำเสมอจากหนึ่งโนด ส่วนแสดงหน่วยความหมายของคำสามารถแตกไปได้หลายกิ่งตามโครงสร้างของประโยค โดยที่โนดถัดไปจะต้องเป็นการกและตามด้วยส่วนแสดงหน่วยความหมายของคำที่สัมพันธ์กัน นั่นคือการกจะเป็นโนดลูก (Daughter or Sister Node) ที่แตกกิ่งมาจากส่วนแสดงหน่วยความหมายของคำที่เป็นโนดแม่เสมอ

นอกจากนั้นโนดแรกสุดของภาษากลางจะเป็นกริยาหลักของประโยคที่ใช้เป็น entry node ของการจัดลำดับของคำดังนั้นเราสามารถกำหนดลำดับของโนดด้วยการเปรียบเทียบความสัมพันธ์ของโนดเป็นคู่ ๆ ที่อยู่ข้างเคียงกัน โดยเริ่มจาก entry node ซึ่งเป็น root node ตัวแรกสุดไปตามกิ่งจากบนลงล่าง และจากซ้ายไปขวา (ตามหลักการ Depth-Fist Algorithm) ตามลำดับจนกระทั่งย้อนกลับมาที่ entry node อีกครั้งหนึ่ง กฎทั้งหมดแบ่งได้เป็น 2 กรณีคือ

1) CP-Case-CP Ordering Rules : เป็นการกำหนดลำดับของโนดที่อยู่ในกิ่งเดียวกัน ระหว่างส่วนแสดงหน่วยความหมายของคำกับการกและ การกกับส่วนแสดงหน่วยความหมายของคำซึ่งจะมีทั้งหมด 4 กฎดังแสดงในรูปที่ 2.15

2) Case-CP-Case Ordering Rules : เป็นการกำหนดลำดับของโนดที่อยู่คนละกิ่ง ทั้งนี้ ด้วยการเปรียบเทียบลำดับของการกในระดับเดียวกันโดยใช้ root ส่วนแสดงหน่วยความหมายของคำของการกทั้งสองเป็นหลัก ซึ่งจะมีทั้งหมด 6 กฎ ดังแสดงในรูปที่ 2.15

เมื่อใส่ฐานข้อมูลของกฎไวยากรณ์การเรียงลำดับคำและกลุ่มคำในภาษาไทย เข้าไปในกฎทางลอจิกดังกล่าวข้างต้น จะสามารถกำหนดลำดับของคำบนโครงสร้างทางไวยากรณ์ได้ถูกต้อง



CP-CASE-CP ordering rules

1) If CP1 => Cs1 & Cs1 => CP2 then CP1 => Cs1 => CP2
2) If CP1 => Cs1 & CP2 => Cs1 then CP1 => CP2 => Cs1
3) If Cs1 => CP1 & Cs1 => CP2 then Cs1 => CP2 => CP1
4) If Cs1 => CP1 & CP2 => Cs1 then CP2 => Cs1 => CP1

CASE-CP-CASE ordering rules

1) If CP1 => Cs1 & CP1 => C82 & Cs1 =>C82 then CP1 => Cs1 => Cs2
2) If CP1 => Cs1 & CP1 => Cs2 & C82 =>Cs1 then CP1 => Cs2 => Cs1
3) If CP1 => Cs1 & C82 => CP1 then Cs2 => CP1 => Cs1
4) If Cs1 => CP1 & CP1 => C82 then Cs1 => CP1 => Cs2
5) If C81 => CP1 & C82 => CP1 & C81 =>C82 then C81 =>C82 => CP1
6) If C81 => CP1 & C82 => CP1 & C82 =>C81 then C82 =>C81 => CP1

รูปที่ 2.15 กฎการกำหนดลำดับของโนดทางลอจิก (CP = Conceptual Primitive, Cs = Case)
เครื่องหมาย A => B หมายถึง A มีลำดับอยู่หน้า B

2.4.2.2.2. ฐานข้อมูลของกฎไวยกรณ์การเรียงลำดับคำในภาษาไทย

(Thai Grammatical Words Ordering Rules Base)

จากที่กล่าวมาในข้อ 2.4.2.2.1 ว่าการเปรียบเทียบเพื่อจัดลำดับของคำหรือโนดในโครงสร้างไวยากรณ์นั้นเรามีโนดอยู่ 2 ประเภทคือส่วนแสดงหน่วยความหมายของคำและการกโดยที่ส่วนแสดงหน่วยความหมายของคำ จะเป็นกลุ่มคำนาม (Noun) คำสรรพนาม (Pronoun) คำกริยา(Verb) คำคุณศัพท์(Adjective) และคำกริยาวิเศษณ์ (Adverb) ส่วนการกนั้นจะถูกแทนที่ด้วยคำบุพบท (Preposition) และคำสันธาน (Conjunction) แต่ในหลักภาษาไทยนั้นยังมีการแบ่งประเภทของคำ (Word Categories) เหล่านี้ปลีกย่อยลงไปอีกตามหน้าที่ของคำเมื่อไปอยู่ในประโยคและการกำหนดประเภทของคำนั้นมีหลายวิธี แล้วแต่ความเหมาะสมและความถูกต้องตามข้อตกลงของนักภาษาศาสตร์ ซึ่งปัจจุบันยังไม่มีการสรุปเป็นแม่แบบที่แน่นอน แต่ในระบบการแปลภาษาโดยเฉพาะการสร้างภาษาไทยจากภาษากลางดังกล่าว เราจะต้องคำนึงถึงการกำหนดประเภทของคำที่สามารถใช้เป็นฐานในการจัดเรียงลำดับของคำ ได้อย่างถูกต้องแม่นยำ ในทุกกรณี และใช้เวลาในการจัดการ (Processing) น้อย ปัจจุบันการศึกษาทางด้านนี้อยู่ในระหว่างการวิจัยที่ยังสรุปเป็นแม่แบบที่แน่นอนไม่ได้ สำหรับงานวิจัยนี้ ได้กำหนดฐานข้อมูลของกฎไวยากรณ์ การเรียงลำดับคำในภาษาไทยในขั้นแรก ดังแสดงในตารางที่ 2.3 จากการทดลองสร้างประโยคที่มีโครงสร้างไม่ยุ่งยากนักได้ผลเป็นที่น่าพอใจ แต่อย่างไรก็ตามกำลังได้มีการทำการปรับปรุงฐานข้อมูลดังกล่าว เพื่อให้ได้กฎการเรียงคำที่สมบูรณ์ยิ่งขึ้น

ตารางที่ 2.3 ตารางแสดงฐานข้อมูลของการเรียงลำดับคำในภาษาไทย

Under CP - Case – CP Ordering	Under Case - CP - Case Ordering
SADV => PRED	SADV => Every thing
SHOD => PRED	SHOD => Every thing
SUB => PRED	DOB => IOB, COMP, COMPF
PRED => DOB, IOB, COMP, COMPF	IOB => COMP, COMPF
PRED => INT	COMP => COMPF
PREP => N, PRON	Every thing => INT
N_1 => V, CLAS, PREP, N_2, PRON	V => CLAS, PREP
CLAS => ORD, DEM, ADI	CLAS => PREP
N_2 => CAD, QUAT, CLAS	
CAD => CLAS	
QUAT => CLAS	

<u>หมายเหตุ</u> เครื่องหมาย A => B หมายถึง A มีลำดับอยู่หน้า B

รายละเอียดประเภทของคำแสดงได้ดังต่อไปนี้

SADV	:	Sentencial Adverb ส่วนขยายความประโยค เช่น โดยทั่วไป, ที่จริง, ตามความจริง , ดังนั้น
SMOD	:	Sentencial Modefier ส่วนขยายประโยค เช่น เมื่อเร็ว ๆ นี้ กาลครั้งหนึ่ง
PRED	:	Predicate CP ได้แก่ กริยาหลัก, คำคุณศัพท์ที่ทำหน้าที่เป็นกริยาหลัก
SUB	:	Subject ประธานของประโยค
DOB	:	Direct Object กรรมตรงของประโยค ได้แก่ N , PRON
IOB	:	Indirect Object กรรมรองของประโยค ได้แก่ N , PRON
COMP	:	Complement เป็นส่วนที่อยู่หลัง IOB
COMP	:	Complement ที่บอกเวลา, บอกเหตุผล
INT	:	Interogative marker เช่น หรือไม่, หรือ, อะไร
N	:	Noun เช่น ไก่, ม้า
PRON	:	Pronoun เช่น เขา, หล่อน, ฉัน
V	:	Verb เช่น เดิน, กิน, นอน
ADJ	:	Adjective เช่น สวย, ใหญ่, เล็ก
PREP	:	Preposition เช่น ที่, บน, ใน
CLAS	:	Classifier เช่น คัน, ตัว, เล่ม, ด้าม
CAD	:	Cardinal Number เช่น 1, 2, สาม, ห้า, ร้อย
ORD	:	Ordinal Number เช่น แรก, คันแรก
DEM	:	Memonsative เช่น นี่, นั่น
QUAT	:	Quantitative เช่น หลาย, บาง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.5 งานวิจัยที่เกี่ยวข้อง

หากจะย้อนไปในอดีตอันไกลโพ้นมนุษย์มีความพยายามในการใช้รหัสแทนความหมายที่ใช้ในการติดต่อกันรหัสที่ว่านี้อาจเป็นสากลที่ทำให้ทุกชาติทุกภาษารู้ได้ทันที จนถึงในปี พ.ศ. 2472 ก็เริ่มมีการส่งสัญญาณโทรเลขด้วยรหัสมอร์ส

ในปี พ.ศ. 2489 วาร์เรน วีเวอร์ (W. Weaver) ได้คิดค้นหลักการทางคณิตศาสตร์เพื่อใช้ส่งรหัสในลักษณะที่ป้องกันการถอดความ

ในปี พ.ศ.2490 บูท(Booth) ได้ทำการเก็บข้อมูลเป็นระบบในรูปแบบพจนานุกรมและหาอัลกอริทึมการเรียกค้นข้อมูลในปี 2491 ไรเคน (Richens) ก็ได้สร้างกฎเกณฑ์ทางภาษาเกี่ยวกับการผันแปรของคำ ในเรื่อง inflection และ derivation

ในปี พ.ศ. 2493 - 2503 มีการพัฒนาทฤษฎีพื้นฐานทางภาษาศาสตร์กันมาก นับว่าเป็นการวิจัยพื้นฐานของการประมวลผลภาษาธรรมชาติที่มีผลต่อการพัฒนาในยุคต่อมา

ระบบการแปลภาษาได้พัฒนากันมามาก ในระยะหลังได้มีบริษัทผู้สนใจหลายบริษัทร่วมพัฒนา เช่น บริษัทซีเมนต์พัฒนาระบบ METAL บริษัทฟูจิตสึพัฒนาระบบ ATLAS บริษัท NEC พัฒนาระบบ PIVOT ฯลฯ จนในปัจจุบันมีการพัฒนางานด้านการแปลภาษาด้วยคอมพิวเตอร์กันมากมายกว่าร้อยระบบ (ยืน ภู่วรวรรณ, ชัยยงค์ วงศ์ชัยสุวัฒน์, 2535)

จะเห็นได้ว่าการวิจัยและขอบเขตของการพัฒนาระบบเครื่องแปลภาษานั้นได้เพิ่มความสลับซับซ้อนมากยิ่งขึ้น โดยเฉพาะหลังจากที่มีรายงานผลของการประเมินผลการวิจัยระบบเครื่องแปลภาษาของ ALPAC (Automatic Language Processing Advisory Commitee) ที่บ่งถึงการสูญเปล่าของงบประมาณที่รัฐบาลสหรัฐให้ทุนอุดหนุนวิจัยอยู่ในขณะนั้น งบประมาณสำหรับการวิจัยจึงได้ถูกตัดไปเป็นจำนวนมากแม้ว่าการวิจัยในสหรัฐฯ ในขณะนั้นจะซบเซาลงไปมากก็ตาม แต่ก็เป็นการดีที่เป็นสิ่งกระตุ้นให้นักวิจัยกลับมาทบทวนและวิจัยในขั้นพื้นฐานกันมากยิ่งขึ้น จนเกิดทฤษฎีและแนวคิดใหม่ๆ ทั้งในด้านภาษาและคอมพิวเตอร์ จนเกิดเป็นสาขาวิชาใหม่ที่เรียกกันว่า Computational Linguistics ตัวอย่างทฤษฎีใหม่ๆ ที่เกิดขึ้น เช่น GB (Government and Binding), FUG (Functional Unification Grammar), LFG (Lexical Functional Grammar), GPSG (Generalized Phrase Structure Grammar), CUG (Categorial Unification Grammar) เป็นต้น นอกจากนี้ก็มีความพยายามที่จะสร้างเป็นระบบ ตลอดจนคลังข้อมูลและคลังศัพท์ขนาดใหญ่มาก

ตัวอย่างพจนานุกรมอิเล็คทรอนิกส์ที่ใช้ในการประมวลผลภาษากับเครื่องแปลภาษาเพื่อใช้กับภาษาในกลุ่มประเทศยุโรป ดังตารางที่ 2.4 แสดงตัวอย่างระบบและโครงการวิจัยพัฒนาที่เด่นและประสบความสำเร็จขั้นหนึ่ง

ตารางที่ 2.4 ตัวอย่างพจนานุกรมอิเล็คทรอนิกส์
(การประมวลผลภาษากับเครื่องแปลภาษา, 1991)

ชื่อ	องค์การที่รับผิดชอบ	จำนวนคำศัพท์	ประเภทภาษา
TERMIUM	รัฐบาลประเทศคานาดา	หลายแสน	อังกฤษ, ฝรั่งเศส
EURODICAUTOM	EC	3 แสน	อังกฤษ, ฝรั่งเศส เยอรมัน-อิตาลี
LEXIS	รัฐบาลประเทศเยอรมัน	3 ล้าน	อังกฤษ, ฝรั่งเศส
TEAM	บริษัท Siemens	2 ล้าน	อังกฤษ, ฝรั่งเศส เยอรมัน, อิตาลี เป็นต้น

ทางด้านการวิจัยในปัจจุบันนี้ก็ได้เจาะลึกลงไปกว่าในระยะเริ่มแรกมีการตีความทางความหมาย มีการใช้ฐานความรู้ขนาดใหญ่เข้าช่วย และยังได้วิจัยต่อออกไป รวมไปถึงการวิเคราะห์และการสังเคราะห์เสียงพูดเพื่อที่จะให้มนุษย์สามารถติดต่อกับเครื่องคอมพิวเตอร์ได้อย่างใกล้ชิดมากยิ่งขึ้น (วิรัช ศรเลิศล้ำวาณิช, ดร.สุรพันธ์ เมฆนาวิน, 1994)

ตัวอย่างระบบเครื่องแปลภาษาที่พัฒนาขึ้นในมหาวิทยาลัยในช่วงปี ค.ศ. 1970 ที่ใช้ในการประมวลผลภาษากับเครื่องแปลภาษาเพื่อใช้กับภาษาในกลุ่มประเทศยุโรป ดังตารางที่ 2.5 แสดงตัวอย่างระบบและโครงการวิจัยพัฒนาที่เด่นและประสบความสำเร็จขั้นหนึ่ง

ตารางที่ 2.5 ตัวอย่างระบบเครื่องแปลภาษาที่พัฒนาขึ้นในมหาวิทยาลัยในช่วงปี ค.ศ. 1970
(การประมวลผลภาษากับเครื่องแปลภาษา, 1991)

ชื่อ	องค์การที่รับผิดชอบ	ประเทศ	หมายเหตุ
TAUM-METEO	ม. Montreal	แคนาดา	พยากรณ์อากาศ (อังกฤษ - ฝรั่งเศส)
ARIANE	ม. Grenoble	ฝรั่งเศส	ฝรั่งเศส, อังกฤษ เป็นต้น
SUSY	Saarbrucken	เยอรมัน	เยอรมัน, อังกฤษ, ฝรั่งเศส

ตัวอย่างระบบเครื่องแปลที่พัฒนาใช้ในสหรัฐฯ ที่ใช้ในการประมวลผลภาษากับเครื่องแปลภาษาเพื่อใช้กับภาษาในกลุ่มประเทศยุโรป ดังตารางที่ 2.6 แสดงตัวอย่างระบบและโครงการวิจัยพัฒนาที่เด่นและประสบความสำเร็จขั้นหนึ่ง

ตารางที่ 2.6 ตัวอย่างระบบเครื่องแปลที่พัฒนาใช้ในสหรัฐฯ
(การประมวลผลภาษากับเครื่องแปลภาษา, 1991)

ชื่อ	ภาษาที่แปล	หมายเหตุ
SYSTRAN	16 คู่ภาษา	Mainfrrame Computer
LOGOS	อังกฤษ, เยอรมัน เป็นต้น	ระบบช่วยแปลภาษา
WEIDNER	อังกฤษ, เยอรมัน, ฝรั่งเศส, อิตาลี เป็นต้น	ระบบแปลแบบโต้ตอบ
ALPS	อังกฤษ, เยอรมัน, ฝรั่งเศส, อิตาลี เป็นต้น	ระบบค้นหาคำศัพท์
SMART		ระบบช่วยแปลภาษา

ตัวอย่างโครงการพัฒนาระบบเครื่องแปลภาษา ที่ใช้ในการประมวลผลภาษากับเครื่องแปลภาษา ดังตารางที่ 2.7 แสดงตัวอย่างระบบและโครงการวิจัยพัฒนาที่เด่นและประสบความสำเร็จขั้นหนึ่ง

ตารางที่ 2.7 ตัวอย่างโครงการพัฒนาระบบเครื่องแปลภาษา
(การประมวลผลภาษากับเครื่องแปลภาษา, 1991)

ชื่อ	ระยะเวลา	ประเทศ	ภาษาที่แปล
Mu	1982 – 1986	ญี่ปุ่น	ญี่ปุ่น, อังกฤษ
Mu2	1986 – 1989	ญี่ปุ่น	ญี่ปุ่น, อังกฤษ
EUROTRA	1982 – 1989	EG	9 ภาษาในยุโรป
TAO-ESOPE	1983 – 1987	ฝรั่งเศส	ฝรั่งเศส, อังกฤษ
ALVEY	1984 – 1987	อังกฤษ	อังกฤษ, ญี่ปุ่น
CRITTER	1986 -	แคนาดา	ฝรั่งเศส, อังกฤษ
CICC	1986 -	ญี่ปุ่น, ไทย, จีน, อินโดนีเซีย, มาเลเซีย	ญี่ปุ่น, ไทย, จีน, อินโดนีเซีย, มาเลเซีย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างระบบการแปลภาษาด้วยคอมพิวเตอร์ที่พัฒนาโดยบริษัทในประเทศญี่ปุ่น ซึ่งทำการแปลจากภาษาอังกฤษเป็นภาษาญี่ปุ่นและจากภาษาญี่ปุ่นเป็นภาษาอังกฤษ ดังตารางที่ 2.8 จากการประชุม MT Summit ณ. เมือง Hakone ประเทศญี่ปุ่น เมื่อเดือน กันยายน ค.ศ. 1987

ตารางที่ 2.8 ระบบการแปลภาษาด้วยคอมพิวเตอร์ของบริษัทในประเทศญี่ปุ่น

บริษัทผู้ผลิต	ชื่อระบบแปล	ระบบคอมพิวเตอร์	ภาษาที่แปล	ความเร็ว (คำ/ช.ม.)
Bravice	MEDIUM-PAK MICRO-PAK	DEC VAX/11 NEC 9801 M3, IBM5550 5200/05 DEC Professional	E to J E to J E to G E to S	3,000 1,500
Fujitsu	ATLAS-I ATLAS-II	Mainframe larger then M34OR & FACOM S-3000 supermini	E to J J to E	60,000 60,000
Hitachi	HICATS-JE	M280 mainframe & HITAC M series computer	J to E E to J	60,000 60,000
NEC	PIVOT	ACOS mainframe & EWS4800	E to J J to E	60,000 60,000
Toshiba	TAURAS	UX700 supermini	E to J J to E	7,000 5,000
OKI Elec	PENSEE	if 1000PC UNITOPIA/MC68010-20	J to E	4,000
Sharp	ยังไม่มีการตั้งชื่อ	OA90DX small business computer	E to J	5,000
Mitsubishi	MELTRAN-J/E	MELCOM PSI (SIMPOS OS)	J to E	5,000
CANNON Inc	LAMB	Symbolics 3620 OS release 6.1	J to E	1,000
Matsushita	PAROLE	Symbolics OS Genera	J to E	
SANYO	SWP-7800	IRMX cpu 80186 word processor	J to E	3,500

(หมายเหตุ) J = Japanese, E = English, G = Geerman, S = Spanish

จุฬาลงกรณ์มหาวิทยาลัย

การวิจัยทางด้านการแปลภาษาด้วยเครื่องมีการดำเนินการกันทั่วโลกมากมายหลายระบบ บางระบบก็ยังดำเนินการต่อเนื่อง บางระบบก็ล้มเลิกเพราะพบอุปสรรค และปัญหาต่างๆ ตัวอย่างของระบบการแปลภาษาด้วยเครื่องที่รู้จักกันทั่วไป ได้แก่

1. ระบบซิสทราน (SYSTRAN System) เริ่มพัฒนาขึ้นที่มหาวิทยาลัย George Town โดย ดำเนินงานตั้งแต่ปี 2493 โดยมีเป้าหมายเพื่อให้กลุ่มประเทศทางยุโรปเป็นผู้ใช้
2. ระบบแอลป์ (ALP System) เป็นระบบค้นหาคำศัพท์ที่พัฒนาขึ้นบนไมโครคอมพิวเตอร์
3. ระบบโลโก้ (LOGOS System) เป็นระบบที่ใช้สำหรับการแปลภาษาระหว่าคู่ภาษาอังกฤษและฝรั่งเศส โดยใช้หลักการที่เรียกว่า direct system
4. ระบบคัลท (CULT System) เป็นระบบที่พัฒนาขึ้นในมหาวิทยาลัยฮ่องกงระบบนี้เน้นการแปลระหว่างภาษาจีนกับอังกฤษ
5. โครงการยูโรตรา (EUROTRA Project) เป็นโครงการร่วมของกลุ่มประเทศประชาชาติยุโรปได้พัฒนาขึ้นเพื่อแปลภาษาหลายภาษาที่เป็นปัญหาที่อยู่ในยุโรป
6. ระบบเก็ทตา (GETA System) เป็นระบบที่พัฒนาขึ้นมาในมหาวิทยาลัย Geonouv ประเทศฝรั่งเศสมีการทดลองกับภาษาหลายภาษารวมทั้งภาษาไทยด้วย ฯลฯ

การแปลภาษาสำหรับนักท่องเที่ยวโดยใช้วิธีการแปลโดยตรง ซึ่งพัฒนาโดยห้องวิจัยปัญญาประดิษฐ์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ เป็นการแปลภาษาโดยใช้การจับคู่ในระดับคำ วลี ประโยค โดยสามารถแปลประโยคที่เป็นประโยชน์สำหรับนักท่องเที่ยวได้ประมาณ 1,000 ประโยค สำหรับสถานการณ์ต่างๆ 20 สถานการณ์ เช่น ในร้านอาหาร ตลาด เป็นต้น

ระบบคอมพิวเตอร์ในการแปลภาษาซึ่งพัฒนาโดยสถาบันเทคโนโลยีพระจอมเกล้าธนบุรี เป็นการสร้างระบบต้นแบบของการวิเคราะห์ประโยคภาษาอังกฤษเพื่อสร้างเป็นภาษาไทย การทำงานจะใช้การแบ่งคำ ลดความกำกวม ตรวจสอบลักษณะประโยค แล้วแปลให้เป็นภาษากลาง ก่อนที่จะสร้างเป็นภาษาไทยออกมา ระบบนี้ได้ถูกทดสอบกับประโยคภาษาอังกฤษ 50 ประโยค จากโครงสร้างภาษอังกฤษที่แตกต่างกัน 7 แบบ มีคำศัพท์ในพจนานุกรม 50 คำ การทดสอบจะทำบนภาษากลางที่มีโครงสร้างต่างกัน 50 แบบ และสร้างประโยคภาษาไทยออกมาได้ 19 รูปแบบด้วยกัน

ระบบคอมพิวเตอร์แปลภาษาอังกฤษเป็นภาษาญี่ปุ่น “ยามาโตะ” (Makoto Nagao, 1986) ในปี ค.ศ. 1982 เอช วาดา (H. Wada) ได้พัฒนาระบบการแปลภาษระบบแรกของญี่ปุ่น ในขณะนั้น คอมพิวเตอร์ไม่สามารถแสดงตัวอักษรแบบคันจิ (Kanji) ได้ ภาษาญี่ปุ่นที่ได้ออกมาจึงแสดงโดยใช้ตัวชุดอักษรที่ใช้ในการออกเสียง คือ คาตากานะ (Katagana)

โครงการยูโรตราเป็นโครงการแปลภาษาเพื่อใช้ได้กับทุกๆ ภาษา เริ่มขึ้นในปี ค.ศ. 1978 โดยกลุ่มนักวิจัยจากมหาวิทยาลัยต่างๆในยุโรปหลักการของระบบยูโรตราก็คือ จะมีโมดูลในการแปลงเป็นภาษากลางและจะถูกออกแบบให้มีขนาดเล็ก จากนั้นในการแปลภาษาแต่ละภาษา จะมีโมดูลการวิเคราะห์แล้วเชื่อมโยงกับโมดูลการแปลงให้เป็นภาษากลาง แล้วจึงมีโมดูลในการสร้างประโยคของแต่ละภาษานั้น

ระบบทาเปสตรี (Tapestry System) เป็นเครื่องมือที่ออกแบบมาสำหรับการพัฒนาโปรแกรมประยุกต์ประเภทการประมวลภาษาธรรมชาติ ประกอบไปด้วยโมดูล (Module) ย่อยๆ ทำหน้าที่ต่างกันออกไป ได้มีการนำระบบทาเปสตรีไปสร้างโปรแกรมประยุกต์สำหรับการแปลภาษาอังกฤษเป็นภาษามาเลย์ และจากภาษามาเลย์เป็นภาษาอังกฤษแล้ว

โครงการเครื่องจักรกลในการแปลภาษาของ NECTEC-CICC เป็นการร่วมมือกันระหว่างญี่ปุ่นและไทยในการพัฒนาระบบคอมพิวเตอร์ช่วยในการแปลภาษา โดยใช้วิธีการแปลเพื่อใช้กับทุกภาษา คือ แปลจากภาษาหนึ่งให้เป็นภาษากลาง จากนั้นก็ทำการวิเคราะห์ประโยค ไวยากรณ์ คำต่างๆ แล้วใช้พจนานุกรมร่วมกับกฎไวยากรณ์ต่างๆ ในการสร้างประโยคของอีกภาษาหนึ่ง ส่วนสำคัญมากที่สุดส่วนหนึ่งของระบบคอมพิวเตอร์ช่วยในการแปลภาษานี้ก็คือ พจนานุกรม เพราะเป็นตัวประสานระหว่างภาษาที่ใช้กับภาษากลางนั่นเอง

การแปลภาษาโดยใช้ภาษากลางเป็นตัวเชื่อมระหว่างภาษาต้นฉบับกับภาษาเป้าหมายนั้นเป็นกระบวนการแปลภาษาถ่ายทอดทางอ้อมแบบหนึ่ง โดยแบ่งเป็น 2 ขั้นตอนใหญ่ๆ ได้แก่

1. ขั้นตอนการวิเคราะห์ภาษาต้นฉบับไปสู่ภาษากลาง
2. ขั้นตอนการสร้างภาษาเป้าหมายจากภาษากลาง

วิธีการแปลภาษาแบบนี้จะช่วยลดความยุ่งยากและลดเวลาในการวิเคราะห์ภาษาต้นฉบับและการสร้างเป้าหมาย เมื่อเปรียบเทียบกับกระบวนการแปลภาษาแบบถ่ายทอดโดยตรง ทั้งนี้เพราะเราไม่จำเป็นต้องคำนึงถึงคู่ภาษาที่จะแปล เนื่องจากการวิเคราะห์ภาษาต้นฉบับใดๆ ก็ตามให้เป็นภาษากลางแล้วก็จะสามารถแปลเป็นภาษาเป้าหมายอะไรก็ได้ โดยที่ไม่ต้องวิเคราะห์ใหม่ในทำนองเดียวกันทางด้านการสร้างภาษาเป้าหมายใดๆ นั้น ก็เป็นการสังเคราะห์จากภาษากลางซึ่งเป็นผลของการวิเคราะห์ภาษาต้นฉบับใดๆ เพียงอย่างเดียว เมื่อมีการสังเคราะห์แล้วก็ไม่ต้องทำซ้ำอีกต่อไปนอกจากนั้นการแปลภาษาโดยใช้ภาษากลางนี้จะใช้พจนานุกรมภาษากลางเป็นฐานข้อมูลในการวิเคราะห์ และสังเคราะห์เพียงชุดเดียวต่อหนึ่งภาษา ซึ่งต่างจากการแปลภาษาแบบถ่ายทอดโดยตรงที่จะต้องใช้พจนานุกรมถึง n ชุด ในการแปลต้นฉบับ 1 ภาษาไปสู่ภาษาเป้าหมาย n ภาษา

ในการสังเคราะห์ประโยคภาษาไทยจากภาษากลางของงานวิจัย ได้ศึกษาถึงโครงสร้างของภาษากลาง ซึ่งจะต้องถูกกำหนดขึ้นให้เป็นตัวแทนทางด้านความหมายลึกของทุกภาษา โดยใช้ทฤษฎีทางวากยสัมพันธ์ หรือไวยากรณ์เคสของฟิลล์มอร์เป็นพื้นฐานและพิจารณาให้เหมาะสมกับภาษาไทย เพื่อจะได้นำไปสู่การเขียนกระบวนการสร้างภาษาไทยจากภาษากลางดังกล่าว กระบวนการสร้างภาษาไทยแบ่งเป็น 3 ขั้นตอนใหญ่ๆ ได้แก่ขั้นตอนการเตรียมข้อมูลที่มากับภาษากลางและข้อมูลทางด้านคำศัพท์จากพจนานุกรม และขั้นตอนเป็นการสังเคราะห์ภาษากลางบนพื้นฐานไวยากรณ์ไทยทางด้านความหมายและโครงสร้างทางไวยากรณ์ เพื่อนำไปสู่การสร้างประโยคภาษาไทยที่ถูกต้องและขั้นตอนสุดท้าย เป็นการเอาท์พุต ผลลัพธ์ประโยคภาษาไทย โดยมีการเรียงลำดับของคำในประโยคที่ถูกต้องตามหลักไวยากรณ์ไทยแล้ว (ดร.รัตติกร วรากูลศิริพันธุ์ และคณะ, 2536)

รูปแทนภาษากลางเป็นโครงข่ายทางอรรถศาสตร์ (semantic network) ของประโยครูปแบบและข้อกำหนดภาษากลางออกแบบโดยอูชิดะ (Uchida) และพัฒนาโดยนักวิจัยจาก 5 ประเทศที่เข้าร่วมโครงการเครื่องแปลภาษาในเอเชียซึ่งได้แก่ประเทศญี่ปุ่น จีน มาเลเซีย อินโดนีเซีย และไทยทุกประเทศที่เข้าร่วมโครงการใช้ข้อกำหนดนี้เพื่อแสดงรูปแทนทางอรรถศาสตร์ (semantic representation) สำหรับเครื่องแปลภาษาของประเทศนั้นๆ ได้มีการเสนอการแทนรูปประโยคภาษาไทยด้วยโครงข่ายทางความหมายตามโครงสร้างของภาษากลาง (วันทนีย์ พันธชาติ, 2536)

งานวิจัยทางด้านการแปลภาษาด้วยเครื่อง นอกจากงานวิจัยเกี่ยวกับระบบวิธีการแปลแล้วยังมีอีกสิ่งหนึ่งที่ควรกล่าวถึง คืองานวิจัยทางด้านการป้อนความรู้แก่ระบบแปลภาษา ซึ่งเกี่ยวข้องกับการวิเคราะห์และสังเคราะห์ภาษาเพื่อสร้างกฎเกณฑ์ขึ้น ขอบเขตของงานวิจัยมีดังนี้คือ

1. งานวิจัยทางด้านคำของภาษา (Lexicon)

การแปลภาษาด้วยเครื่องในยุคแรกๆมักจะมุ่งความสนใจโดยส่วนใหญ่มาทางด้านคำของภาษา งานวิจัยทางด้านนี้เชื่อว่า ความรู้ความเข้าใจภาษาแต่ละภาษาจะเกิดขึ้นได้ถ้าเข้าใจการแสดงออก (language behavior) ของคำศัพท์ในภาษานั้น ดังเช่นการแปลแบบโดยตรงได้พยายามใช้วิธีการนี้ ซึ่งงานวิจัยทางด้านนี้ทำให้มีกฎเฉพาะที่มีผลค่อนข้างสูงในเรื่องของความถูกต้อง แต่ก็ทำให้ฐานข้อมูลใหญ่เกินไป และเป็นการยากที่จะดำเนินการต่อไป ในปัจจุบันการวิจัยทางด้านคำของภาษา ยังมีอุปสรรคใหญ่ที่ขวางหน้าอยู่อย่างหนึ่งก็คือการที่จะพยายามหากฎทั่วไป (ซึ่งกล่าวได้ว่าหายากมาก) เพื่อที่จะมาช่วยเหลือในการวิเคราะห์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. งานวิจัยทางด้านกฎโครงสร้างของประโยค (Syntactic Rules)

เป็นงานวิจัยทางด้านลักษณะความสัมพันธ์ของโครงสร้างประโยค ได้มีการกล่าวกันว่า งานวิจัยของการแปลภาษาด้วยเครื่องในปัจจุบัน โดยส่วนใหญ่แล้วได้มุ่งใช้การวิเคราะห์ทางด้านนี้เป็นหลักในการให้ความรู้กับระบบการแปลภาษาด้วยเครื่อง แต่อย่างไรก็ตามงานวิจัยทางด้านนี้ก็ยังมีอุปสรรคอย่างมากเกี่ยวกับการทำให้กฎต่างๆ ของประโยคสมบูรณ์พอที่จะใช้ได้อย่างถูกต้องกับการแปลภาษา

3. งานวิจัยทางด้านกฎความหมายของคำและประโยค (Semantic Rules)

เป็นงานวิจัยที่เกี่ยวกับการสร้างกฎในเรื่องความหมายของคำและประโยค ซึ่งจะเป็นส่วนที่สามารถนำมาใช้ในการแก้ปัญหาเรื่องความหมายของประโยคที่ไม่ถูกต้องได้เป็นอย่างดี แต่อย่างไรก็ตามการกำหนดเกี่ยวกับความหมายของภาษานั้นไม่ใช่เรื่องที่ง่ายนัก โดยเฉพาะอย่างยิ่งกรณีของประโยคที่มีความหมายค่อนข้างลึกซึ้ง นอกจากนั้นการกำหนดกฎเกณฑ์ของความหมายของภาษานั้นเป็นเรื่องที่ยากและต้องใช้เวลามากจึงจะทำให้สมบูรณ์ได้

4. งานวิจัยทางด้านกฎเกณฑ์ต่างๆของภาษา (Cognitive Rules)

เป็นงานวิจัยที่เกี่ยวกับการใช้ระบบการแปลภาษาด้วยเครื่อง เรียนรู้กฎเกณฑ์ต่างๆ ที่เกี่ยวกับภาษา ซึ่งงานวิจัยทางด้านนี้มีปัญหามากกว่าระบบอื่นๆ ที่กล่าวมาข้างต้น เนื่องจากรูปแบบของงานทางด้านนี้จะเป็นที่ถกเถียงอย่างกว้างขวาง และมีความซับซ้อนมากเพราะว่าระบบการแปลภาษาด้วยเครื่อง จะต้องเรียนรู้กฎอย่างมากมายมหาศาล โครงการการพัฒนาระบบรวมเพื่อการแปลภาษาด้วยคอมพิวเตอร์ (นายบุญเจริญ ศิริเนาวกุล , 2531)

ตารางที่ 2.9 ตารางแสดงการเปรียบเทียบกลวิธีการแปลภาษาด้วยคอมพิวเตอร์

กลวิธีการแปลภาษา การทำงาน	การแปล โดยตรง	การแปล แบบถ่ายทอด	การแปลโดย ใช้ภาษากลาง
1. การวิเคราะห์โครงสร้างของภาษาต้นฉบับ	มี	มี	มี
2. การวิเคราะห์ความหมายของภาษาต้นฉบับ	ไม่มี	มี	มี
3. การแทนภาษาต้นฉบับด้วยภาษากลาง	ไม่มี	ไม่มี	มี
4. การสังเคราะห์โครงสร้างและความหมายของภาษาเป้าหมาย	ไม่มี	มี	มี
5. สร้างภาษาเป้าหมาย	มี	มี	มี
6. ชนิดของพจนานุกรมที่ใช้ในระบบ	ทิศทางเดียว	สองทิศทาง	หลายทิศทาง
7. จำนวนพจนานุกรมใช้ในระบบสำหรับ n ภาษา	2n	n	n
8. จำนวนคู่ภาษาที่แปลได้สำหรับ n พจนานุกรม	n/2	N	n!

2.6 การวิจัยระบบเครื่องแปลภาษาภายในประเทศ

ความพยายามในการจัดการทางภาษาของภาษาไทยด้วยเครื่องคอมพิวเตอร์ เริ่มมีขึ้นตั้งแต่การพัฒนาระบบเวิร์ดโพรเซสเซอร์และในปี ค.ศ. 1981 การวิจัยและพัฒนาระบบเครื่องแปลภาษาในประเทศไทยก็ได้เริ่มขึ้นอย่างจริงจัง โดยมีโครงการที่ได้รับการสนับสนุนจากประเทศฝรั่งเศส ชื่อโครงการ ARIANE แปลจากภาษาอังกฤษเป็นภาษาไทยด้วยวิธีการแปลแบบย้ายข้าม (Transfer Method)

โครงการการแปลภาษาด้วยเครื่องในประเทศไทยนั้น ได้เริ่มขึ้นตั้งแต่ปี พ.ศ.2524 โดยเป็นโครงการวิจัยร่วมของทบวงมหาวิทยาลัย กับมหาวิทยาลัย Grenoble ของฝรั่งเศส และมหาวิทยาลัย Sains ของมาเลเซีย โดยมีวัตถุประสงค์ที่จะทำการแปลจากภาษาอังกฤษ เป็นภาษาไทย มหาวิทยาลัย 7 แห่งในประเทศไทยได้เข้าร่วมทำงานในนามของทบวงมหาวิทยาลัย โดยทำงานวิจัยการแปลภาษาด้วยเครื่องในระบบที่ชื่อว่า ARIANE มี ดร.อุดม วโรดม์สิกขดิตถ์ และดร.นิตยา กาญจนะวรรณ เป็นประธานโครงการโครงการ ARIANE นี้ ขึ้นอยู่กับระบบ GETA ของฝรั่งเศสซึ่งใช้ระบบภาษากลาง และระบบการถ่ายถอดในการแปลขั้นตอนในการแปลมีดังนี้คือ วิเคราะห์คำและประโยค แล้วสร้างรูป tree ขึ้น ซึ่งในโครงสร้างรูป tree นี้ จะมีการบอกความสัมพันธ์ในส่วนต่าง ๆ ของประโยค ทั้งทางด้านตรรก และความหมาย แล้วจึงถ่ายทอดคำและโครงสร้างของภาษาต้นฉบับคือภาษาอังกฤษ เป็นภาษาเป้าหมายคือภาษาไทย หลังจากนั้นจะสร้างโครงสร้างรูปผิวขึ้นจากลักษณะร่วมของโครงสร้างรูปลึก ซึ่งลักษณะร่วมของโครงสร้างรูปลึกในโครงการนี้เรียกว่าภาษากลาง ในปัจจุบันการแปลระบบ ARIANE นี้สามารถแปลประโยคธรรมดาและประโยคซับซ้อนบางชนิดได้บ้างแล้ว โดยประโยคต่าง ๆ นั้นจะต้องขึ้นอยู่กับคำศัพท์ประมาณ 2,000 คำที่ได้รับการวิเคราะห์และบรรจุอยู่ใน Dictionary ของระบบ โครงการ ARIANE ยังคงดำเนินการอยู่จนปัจจุบัน นอกจากโครงการ ARIANE นี้แล้ว มีการทำวิจัยเกี่ยวกับการแปลภาษาด้วยเครื่องอยู่หลายแห่ง เช่น มหาวิทยาลัยมหิดล ได้มีการแปลพระไตรปิฎกออกมาเป็นภาษาไทยโดยการใช้คอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์ และ AIT มีงานวิจัยระดับปริญญาโทและปริญญาเอกอยู่หลายชิ้น ซึ่งทำการวิจัยเกี่ยวข้องกับการแปลภาษาด้วยเครื่องโดยเน้นที่การแปลจากไทย-อังกฤษ และ อังกฤษ-ไทย

ต่อมาในปี ค.ศ. 1986 ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ กระทรวงวิทยาศาสตร์ เทคโนโลยีและสิ่งแวดล้อมได้ทำสัญญาให้ความร่วมมือในการพัฒนาระบบเครื่องแปลภาษากับศูนย์ความร่วมมือด้านคอมพิวเตอร์ระหว่างประเทศ หรือ

Center of International Co-operation of Computerization ย่อว่า CICC กระทรวงการค้าและอุตสาหกรรมระหว่างประเทศของประเทศญี่ปุ่นหรือ Ministry of International Trade and Industry ย่อว่า MITI ปัจจุบันนี้ยังคงดำเนินอยู่ ชื่อโครงการพัฒนาระบบเครื่องแปลภาษาสำหรับภาษาในเอเชีย

โครงการความร่วมมือระหว่างรัฐบาลไทยและรัฐบาลญี่ปุ่นโดยผ่านกระทรวงการค้าและอุตสาหกรรมระหว่างประเทศของรัฐบาลญี่ปุ่นโดยมีหน่วยงานชื่อศูนย์ความร่วมมือด้านคอมพิวเตอร์ระหว่างประเทศ เป็นหน่วยงานประสานงานกับรัฐบาลไทยผ่านกระทรวงวิทยาศาสตร์และเทคโนโลยีและการพลังงาน โดยมีศูนย์เทคโนโลยีอิเล็คทรอนิกส์และคอมพิวเตอร์แห่งชาติ เป็นหน่วยงานประสานงาน ได้ร่วมทำโครงการการพัฒนาระบบคอมพิวเตอร์ช่วยในการแปลภาษา โดยใช้วิธีแปลภาษาต้นแบบให้เป็นภาษากลางแล้วถ่ายทอดภาษากลางเป็นภาษาเป้าหมายอีกทีหนึ่งซึ่งวิธีนี้มีข้อดีคือ เหมาะสำหรับการแปลหลายๆ ภาษาในระบบเดียวกัน เพราะเมื่อวิเคราะห์ภาษาต้นแบบใดๆ ให้อยู่ในรูปของภาษากลางแล้วก็สามารถแปลเป็นภาษาเป้าหมายใดๆ ก็ได้ ทำให้ลดความยุ่งยากและลดเวลาในการวิเคราะห์ประโยคภาษาต้นแบบและการสร้างประโยคภาษาเป้าหมาย

ในประเทศไทย ได้มีการออกแบบและสร้างพจนานุกรมสำหรับการแปลภาษาอังกฤษเป็นภาษาไทยโดยสร้างฐานข้อมูลเก็บคำศัพท์เพื่อใช้ในงานทางด้านการแปลภาษาด้วยคอมพิวเตอร์ ณ ห้องวิจัยปัญญาประดิษฐ์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ต่อมา ได้มีระบบการแปลภาษาสำหรับนักท่องเที่ยวโดยใช้การแปลโดยตรง ซึ่งเป็นการแปลโดยใช้การจับคู่ในระดับคำ วลี และประโยค โดยสามารถแปลประโยคที่เป็นประโยชน์สำหรับนักท่องเที่ยวได้ประมาณ 1,000 ประโยค สำหรับสถานการณ์ต่างๆ 20 สถานการณ์

ปี 2528 สนึก คีรีโต ได้พัฒนาโปรแกรมการแปลภาษาอังกฤษเป็นภาษาไทยโดยใช้โครงสร้างการถ่ายทอดความหมายในระดับโอนเอียงไปทางอรรถศาสตร์ (Semantic Preference) ในงานวิจัยดังกล่าว มีการสร้างพจนานุกรมสองภาษาสำหรับการถ่ายทอดไว้ โครงสร้างพจนานุกรมยึดหลักการแทนความหมายระหว่างคู่ศัพท์ที่เกี่ยวข้องกับสิ่งแวดล้อมของประโยค

ในปี พ.ศ.2531 เป็นต้นมา ศูนย์เทคโนโลยีอิเล็คทรอนิกส์และคอมพิวเตอร์แห่งชาติ กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน ได้มีสัญญาการร่วมมือกับกระทรวงกระทรวงการค้าและอุตสาหกรรมระหว่างประเทศของประเทศญี่ปุ่น โดยผ่านทางศูนย์ความร่วมมือด้านคอมพิวเตอร์ระหว่างประเทศ ทำการวิจัยการแปลภาษาด้วยเครื่องขึ้นโดยมีวัตถุประสงค์แรกคือ การแปลจากภาษาญี่ปุ่น-ไทย และภาษาไทย-ญี่ปุ่น โครง

การนี้เป็นโครงการระยะยาว 6 ปี โดยมหาวิทยาลัย 4 แห่ง ได้ร่วมวิจัยในนามของศูนย์เทคโนโลยีอิเล็คทรอนิกส์และคอมพิวเตอร์แห่งชาติ คือมหาวิทยาลัยเกษตรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี และสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง การวิจัยนี้จะเป็นการแปลภาษาด้วยเครื่อง ระบบภาษากลาง โดยใช้ระบบ VENUS ของประเทศญี่ปุ่น ซึ่งงานของการวิจัยนี้เพ่งเริ่มต้นและยังไม่มีผลคืบหน้าที่เห็นเด่นชัดมากนัก นอกจากความเข้าใจและสนใจเรื่องการแปลภาษาด้วยเครื่อง ได้แพร่หลายในหมู่นักวิจัยไทยในมหาวิทยาลัยต่าง ๆ เพิ่มขึ้น การสัมมนาเกี่ยวกับการแปลภาษาด้วยเครื่องในครั้งแรกซึ่งจัดขึ้นในเดือนกรกฎาคม 2531 นั้น มีผู้เข้าร่วมกว่า 50 คน (ดร.สุรพันธ์ เมฆนาวิน,1994)

ในปี 2534 วันทนีย์ พันธชาติและคณะนักวิจัย ห้องปฏิบัติการวิจัยภาษาและวิทยาการความรู้ ศูนย์เทคโนโลยีอิเล็คทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ กระทรวงวิทยาศาสตร์เทคโนโลยีและสิ่งแวดล้อม ทำการวิจัยระบบการวิเคราะห์ภาษาไทยเพื่อการแปลภาษาด้วยคอมพิวเตอร์ โดยใช้ไวยากรณ์โครงสร้างวลี ไวยากรณ์การกและไวยากรณ์พึ่งพา ประกอบกันเพื่อวิเคราะห์ความหมายของประโยค โดยใช้การกที่กำหนดขึ้นในภาษากลางแสดงความสัมพันธ์ของคำ ส่วนตัวระบบอาศัยเทคนิคและซอฟท์แวร์การวิเคราะห์ของระบบ PIVOT ซึ่งได้รับการสนับสนุนจากบริษัท NEC ประเทศญี่ปุ่น

2.7 อนาคตและแนวทางของการวิจัยและพัฒนา

วิทยาการทางการประมวลผลภาษาธรรมชาติถึงแม้จะได้เริ่มมากว่า 40 ปีแล้วก็ตาม ผลงานที่จะนำมาใช้งานได้จริงๆ นั้นยังไม่ได้ปรากฏให้เห็นชัด เนื่องจากยังขาดทฤษฎีหรือหลักการที่จะสามารถนำมาประยุกต์ใช้กับระบบที่กว้างได้ ส่วนใหญ่ของทฤษฎีนั้นจะจำกัดขอบเขตของปัญหาซึ่งจะใช้ได้กับเฉพาะแต่ละกรณีๆ เท่านั้นเอง แต่กระนั้นก็ตามก็ได้พยายามประยุกต์ให้เกิดระบบต่างๆ ขึ้นมา เช่น ระบบเครื่องแปลภาษาซึ่งมีให้เห็นมากมายทั้งในญี่ปุ่น ยุโรปและสหรัฐฯ ระบบค้นหาข้อมูลด้วยภาษาธรรมชาติ ระบบตรวจไวยากรณ์ ระบบย่อความเป็นต้น

สำหรับประเทศไทยในขณะนี้ยังคงอยู่ในระยะเริ่มต้นเท่านั้นเมื่อเทียบกับต่างประเทศ ซึ่งนับได้ว่ามายุคที่ 3 ของการ Breakthrough ทางเทคโนโลยี ดังนั้นความพยายามส่วนใหญ่จะอยู่ที่การวางรากฐานเทคโนโลยีทั้งในด้านการวิจัยภาษาและการประมวลผลภาษาด้วยเครื่องคอมพิวเตอร์เราจำเป็นที่จะต้องมีแนวคิดใหม่ทางทฤษฎีที่จะจัดการกับภาษาไทยตามแนวทางของภาษาปัจจุบัน

แนวโน้มของความพยายามที่กำลังดำเนินอยู่ในปัจจุบันนี้ของนานาประเทศดังต่อไปนี้อาจเป็นแนวทางสำหรับการวิจัยและพัฒนาสำหรับประเทศไทยได้ดี (วิรัช ศรเลิศล้ำวาณิช,ดร.สุรพันธ์ เมฆนาวิน,1994)

2.7.1. การหาแนวทางใหม่สำหรับการแปลภาษา

2.7.1.1 การแปลโดยใช้ภาษากลาง (Interlingua Based Translation)

2.7.1.2 การแปลโดยใช้ฐานบทความ (Corpus Based Translation)

2.7.1.3 การแปลโดยใช้วิธีการเปรียบเทียบ (Analogy Translation)

2.7.1.4 การแปลโดยใช้ฐานความรู้ (Knowledge Based Translation)

2.7.1.5 การแปลโดยใช้วิธีการย้ายข้าม (Transfer Based Translation)

2.7.2. การหาระบบทางภาษาใหม่

2.7.3. การพัฒนาระบบพจนานุกรมขนาดใหญ่

2.7.4. การรวมทางการประมวลผลภาษาธรรมชาติ

2.7.5. การจัดการระดับปริจเฉท (Discourse)

2.7.6. การติดต่อกับผู้ใช้ (User Interface) ที่เหมาะสม

2.7.7. การต่อระบบแปลภาษาเข้ากับระบบประมวลผลภาษาอื่น

2.7.8. ระบบแปลแบบคำต่อคำ (Speech-to-Speech)

2.7.9. การพัฒนาระบบฐานความรู้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3
ลักษณะของภาษา UNL และแนวความคิดในการแปลงเป็นภาษาไทย

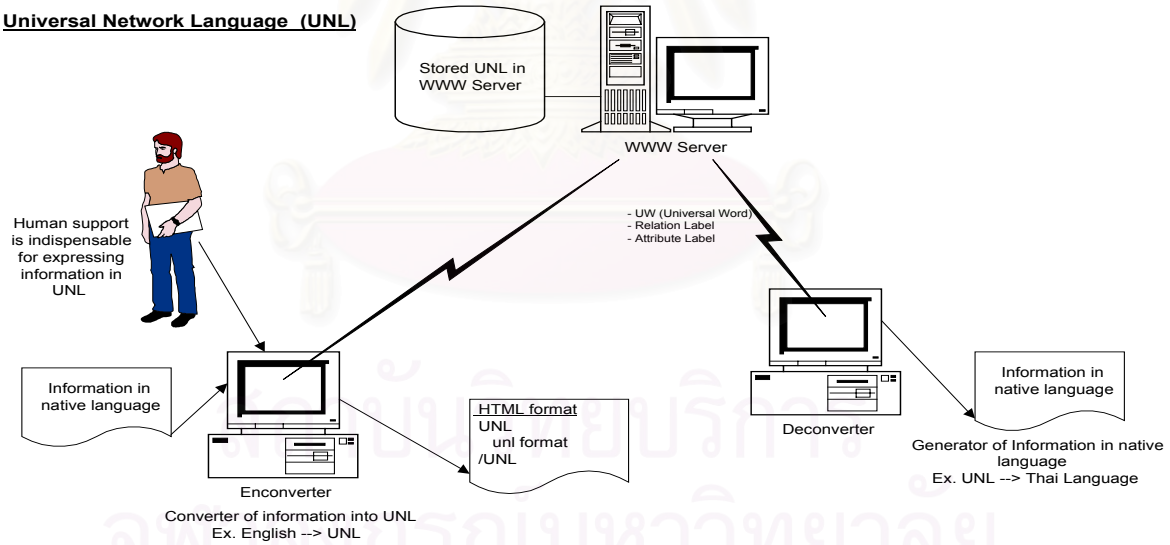
3.1 ความหมายของภาษา UNL (Definition of UNL)

ภาษา UNL เป็นภาษาทางปัญญาประดิษฐ์โดยมีลักษณะเป็นเครือข่ายแบบหลากหลายมีจุดมุ่งหมายเพื่อให้บุคคลสามารถใช้ภาษาท้องถิ่นของตัวเองติดต่อกับบุคคลที่ใช้ภาษาต่างกัน ภาษา UNL เป็นภาษาโดยทั่วไปที่ใช้ร่วมกันสำหรับทุกคนในการสื่อสารหลายภาษา ระบบภาษา UNL มีองค์ประกอบพื้นฐานประกอบด้วยเครือข่ายการสื่อสารและโปรแกรมในการแปลงระหว่างภาษาUNL กับภาษาท้องถิ่น

3.2 ลักษณะทั่วไปของระบบภาษา UNL (UNL General System)

ระบบที่แปลจากภาษาท้องถิ่นไปเป็นภาษา UNL เรียกว่า “Enconverter” และจากภาษา UNL เป็นภาษาท้องถิ่นเรียกว่า “Deconverter” ข้อมูลข่าวสารในแต่ละภาษาจะใช้ถูกแปลงและแลกเปลี่ยนข้อมูลผ่านเครือข่ายการสื่อสารในรูปแบบของภาษา UNL ข้อมูลข่าวสารในรูปแบบภาษา UNL จะถูกแปลงกลับเป็นภาษาท้องถิ่นแต่ละภาษาบนหน้าจอในเครือข่ายการสื่อสารดังแสดงไว้ในรูปที่ 3.1

ในการส่งข้อมูลข่าวสารข้อมูลจะแปลงจากภาษาท้องถิ่นเป็นภาษา UNL อย่างรวดเร็วและสามารถตรวจสอบด้วยการแปลงกลับจากภาษา UNL เป็นภาษาท้องถิ่นด้วยข้อมูลที่มีอยู่ในภาษา UNL



รูปที่ 3.1 ลักษณะทั่วไปของระบบภาษา UNL

3.2.1 ระบบภาษา UNL ในเครือข่ายการสื่อสาร (UNL System in Network)

ผู้แต่งเอกสารภาษา UNL จะทำได้โดยมีส่วนประกอบ ได้แก่ “enconverter” และ “deconverter” ของแต่ละภาษา ผู้ใช้สามารถป้อนข้อมูลเป็นภาษาของตัวเอง ในเวลาเดียวกันจะแปลงเอกสารและจัดทำเป็นภาษา UNL ทันที ผู้แต่งเอกสารภาษา UNL จะสามารถแสดงเป็นภาษาท้องถิ่นโดยแปลงจากภาษา UNL ระบบจะทำความเข้าใจในภาษาต้นฉบับได้เอง เพื่อเป็นการตรวจสอบความถูกต้องของการแปลง (enconversion) ในบางขั้นตอนผู้ใช้สามารถจัดทำเอกสารภาษา UNL ได้ถูกต้องอย่างที่ต้องการ

ข้อมูลข่าวสารที่อธิบายในภาษา UNL จะปรากฏในโฮมเพจ (home page) หรือถูกเก็บไว้ในที่เก็บเอกสารสำคัญ เอกสารภาษา UNL จะถูกแลกเปลี่ยนในเครือข่ายสื่อสารและถูกเก็บไว้ใน WWW เซิฟเวอร์ มันจะถูกส่งไปโดยโปรแกรมโอนย้ายข้อมูล (transfer) เอกสารภาษา UNL จะรับได้ด้วยหน้าจอในเครือข่ายและจะถูกแปลงเป็นภาษาท้องถิ่นและอ่านด้วยเบราเซอร์ที่เหมือนกับเป็นตัวแปลงภาษา “deconverter”

3.2.2 โปรแกรมแปลงจากภาษา UNL เป็นภาษาท้องถิ่น (Deconverter : Generator of Information in Native Language)

“Deconverter” จะทำหน้าที่แปลงเป็นภาษาท้องถิ่นจากภาษา UNL โดยใช้กฎของระบบ ภาษา UNL “Deconverter” เป็นเทคโนโลยีที่ถูกสร้างขึ้น เพื่อใช้กับข้อมูลที่เป็นภาษา UNL โดยมีความคาดหวังว่าจะแปลงเป็นภาษาแต่ละภาษาได้อย่างถูกต้อง 100% และสามารถพัฒนาโดยให้เชื่อมต่อกับระบบภาษา UNL ที่มีอยู่แล้ว

“Deconverter” เป็นระบบอัตโนมัติในการแปลงภาษา UNL เป็นภาษาท้องถิ่น โดยต้องการคุณภาพสูงและผลลัพธ์ที่ถูกต้อง มีความสำคัญอยู่ที่โครงสร้างพื้นฐานของ “deconverter” ทำให้สามารถแบ่งใช้ได้ทั่วไป เทคโนโลยีที่ถูกพัฒนาสำหรับภาษาหนึ่งสามารถนำไปใช้กับภาษาอื่นได้โดยการแบ่งการใช้โครงสร้างพื้นฐาน ถึงแม้ว่าตัวอักษรจะขึ้นอยู่กับแต่ละภาษา โปรแกรมสร้างภาษาและการบริหารพจนานุกรมสามารถใช้ร่วมกันได้ถ้ารูปแบบของกฎและพจนานุกรมได้ตามมาตราฐาน

3.2.3 โปรแกรมแปลงจากภาษาท้องถิ่นเป็นภาษา UNL (Enconverter : Converter of Information into UNL

ถึงแม้ว่าการแปลงภาษา UNL เป็นภาษาหลายๆ ภาษาได้ แต่การแต่งเป็นเอกสารภาษา UNLก็ยังไม่ง่ายพอที่ผู้ใช้จะทำได้เร็วพอที่จะกระจายเอกสารไปเป็นหลายภาษา โดยมี UNL editor เป็นเครื่องมือที่อำนวยความสะดวกในการทำเอกสารภาษา UNL

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาษา UNL เป็นข้อมูลที่ถูกพัฒนาขึ้นสำหรับการประมวลผลด้วยคอมพิวเตอร์ ดังนั้นจึงไม่ใช่เรื่องง่ายพอสำหรับคนที่จะดูแลโดยตรง ผู้ใช้มีความต้องการที่จะป้อนข้อมูลเป็นภาษาของตนเองโดยใช้ UNL editor

ข้อมูลที่ป้อนในภาษาท้องถิ่นถูกแปลง “enconveted” เป็นภาษา UNL ความถูกต้องของผลลัพธ์ในการแปลงถูกตรวจสอบด้วยการแปลงกลับ “deconverted” เป็นภาษาต้นฉบับ ในการตรวจสอบนี้ ความถูกต้องของการแปลงกลับจะต้องเป็น 100%

เมื่อพบว่าผลลัพธ์ไม่ถูกต้องเพียงพอ ผู้ใช้สามารถที่จะแก้ไขเอกสารต้นฉบับหรือจัดทำเอกสารภาษา UNL ได้โดยทันที โดยการใช้เครื่องมือของ editor การกำหนดใจความสำคัญของเรื่องโดยผู้ใช้ จะเป็นสาเหตุที่เป็นอุปสรรคของการติดต่อสื่อสารโดยใช้ภาษา UNL

ด้วยเหตุผลนี้เทคโนโลยีขั้นสูงของการวิเคราะห์ภาษาธรรมชาติได้ถูกนำมาใช้ในการพัฒนาตัวแปลง “enconverter” เทคโนโลยีการวิเคราะห์มีจุดหมายที่จะทำให้เป็นการวิเคราะห์โดยอัตโนมัติ เอกสารที่เป็นภาษาท้องถิ่น บางครั้งก็รวมไปถึงประโยคที่ไม่อาจเข้าใจได้สำหรับมนุษย์, ประโยคที่มีความกำกวม, หรือประโยคที่มีความยาวมาก เป็นประโยคที่คอมพิวเตอร์ไม่สามารถวิเคราะห์ได้ แต่ใช้คนช่วยได้

การแปลงเป็นภาษา UNL จะแก้ปัญหาการกำกวมของประโยคและความหมายของคำ โดยใช้ความช่วยเหลือของคน

3.3 การแสดงข้อมูลภาษา UNL ด้วยคอมพิวเตอร์ (UNL Information is Represented in Computer)

เอกสารในคอมพิวเตอร์จะประกอบไปด้วยข้อมูลหลายประเภท เช่น เอกสารในเวิลด์โปรเซสซิ่ง จะมีข้อมูลเช่น ประเภทการติดต่อ, ขนาดตัวอักษร, รูปภาพเป็นต้น ในการติดต่อโดยผ่านอินเตอร์เน็ท ตัวอักษรที่อยู่ในรูปแบบ HTML มีความสามารถที่จะเชื่อมโยงไปยังเอกสารอื่นๆได้ และผู้อ่านสามารถที่จะอ้างถึงเอกสารได้หลายชนิดโดยใช้การเชื่อมโยง ดังนั้นเอกสารอีเล็คทรอนิคสามารถบรรจุข้อมูลที่เป็นองค์ประกอบในตัวเอกสารเองและกระจายการใช้งานเพิ่มมากขึ้น ข้อมูลภาษา UNL จะเป็นการใช้งานที่เหมือนกันในเครือข่ายการสื่อสาร

ส่วนดีของ HTML เป็นเอกสารที่จัดทำด้วยตัวอักษรแบบง่าย โดยทั่วไปข้อมูลเป็นการแบ่งเป็นตัวอักษรและฝังไว้ในข้อมูล ใน HTML ข้อมูลที่เก็บไว้เป็นการอธิบายด้วยตัวอักษรแบบง่าย ด้วยคุณลักษณะแบบนี้ทำให้ HTML ได้พัฒนาไปในส่วนอื่นในการใช้ข้อดีของ hyper-text ในที่สุดของ HTML รายละเอียดและรูปแบบได้เปิดเผยต่อสาธารณชน รูปแบบของ HTML ยังมีการขยายและพัฒนาต่อไป รูปแบบในการใช้ภาษา UNL คาดหวังไว้ว่าเป็นส่วนขยายส่วนหนึ่งใน HTML

ภาษา UNL เป็นการออกแบบให้เป็นภาษาโดยทั่วไปในเครือข่ายสื่อสาร และข้อมูลที่แสดงในภาษา UNL สามารถจัดการโดยระบบใดๆ ในโลกนี้ เพื่อให้บรรลุถึงจุดหมายนี้ รายละเอียดรูปแบบของภาษา UNL ได้

แสดงไว้เป็นส่วนขยายส่วนหนึ่งในรูปแบบของ HTML ข้อมูลของภาษา UNL ได้ฝังไว้ในเอกสาร HTML ด้วยข้อความที่ระบุเป็นข้อมูลภาษา UNL ที่แนบไว้เป็นส่วนเริ่มต้นและสิ้นสุด ส่วนขยายของรูปแบบตรงกับส่วนที่มีอยู่แล้วของ HTML ดังนั้นการแสดงข้อมูลภาษา UNL ถูกทำเหมือนเอกสารอื่นโดยไม่ทำให้เสียโครงสร้าง hyper-text ของ HTML เป็นการตรงกับรูปแบบของ HTML รายละเอียดของภาษา UNL จะถูกทำขึ้นด้วยตัวอักษรแบบง่าย (plain text) และรูปแบบจะถูกเปิดเผยต่อสาธารณชน (UNL Center, 1996)

3.4 จุดมุ่งหมายของภาษา UNL (Aims of UNL)

ภาษา UNL **ได้ถูกออกแบบเพื่อความมุ่งหมายดังต่อไปนี้**

3.4.1 ภาษา UNL ใช้เพื่อเป็นสื่อกลางในการแทนความหมายของทุกๆ ภาษา

3.4.2 เงื่อนไขของภาษา UNL จะต้องกำหนดไม่ยากเกินไปแต่จะต้องง่ายเพียงพอที่จะเข้า ใจสำหรับที่จะนำไปพัฒนาโปรแกรมในส่วนของ “enconverters” และ “deconverters” ในทุกๆ ภาษา

3.5 กฎเกณฑ์ของภาษา UNL (UNL Rules)

3.5.1 การแสดงรูปประโยคของภาษา UNL

ส่วนประกอบของภาษา UNL ประกอบด้วย

3.5.1.1 คำสากล (Universal Word) ย่อว่า UW ใช้สำหรับแสดงความหมายของคำ

3.5.1.2 คำเชื่อมโยง (Relation label) ใช้สำหรับแสดงความสัมพันธ์ระหว่างคำสากล

3.5.1.3 คำบอกคุณสมบัติ (Attribute label) ใช้สำหรับอธิบายเงื่อนไขหรือข้อมูลเพิ่มเติม ของคำสากล

รูปแบบไวยากรณ์ของภาษา UNL

```
<UNL>              ::= {<binary relation>}...
<binary relation>  ::= Relation label ‘(‘ <UW1> ‘,’ <UW2> ‘)’
<UWn>                    ::= <element UW> | <compound UW>
<element UW>             ::= English head word
                         [ ‘(‘ {{Relation label { ‘<‘ | ‘>‘ }| ‘=‘ } <element UW>
                               {[ ‘,’ {{Relation label { ‘<‘ | ‘>‘ } | ‘=‘ } <element UW> ]}... ‘)’
                                    {[ ‘.’ Attribute label ]}...
                         <compound UW>::={<Binary relation>}...
```

3.5.2 ประมวลศัพท์ของภาษา UNL (Vocabulary of UNL)

3.5.2.1. คำสากล

3.5.2.2. คำเชื่อมโยง คำเชื่อมโยงระหว่างส่วนประกอบของประโยคตัวอย่างเช่น

agt --> Agent : ตัวแทนที่แสดงการกระทำ หรือ

ตัวแทนที่มีทำให้วัตถุเคลื่อนไหวด้วยความตั้งใจ.

3.5.2.3. คำบอกคุณสมบัติ

3.5.2.3.1 ใช้อธิบายลักษณะเฉพาะของคำสากลเช่น @generic บอกความหมายทั่ว ไปของคำสากล

3.5.2.3.2 ใช้อธิบายรูปแบบกาลในไวยากรณ์

3.5.2.3.3 ใช้อธิบายหน้าตาหรือกฎเกณฑ์

3.5.2.3.4 ใช้อธิบายความมุ่งหมายของคำพูด

3.5.2.3.4.1 อธิบายความมุ่งหมายของส่วนประกอบประโยค

3.5.2.3.4.2 อธิบายรูปประโยคโดยรวม

<Aux-verb>@attribute-label e.g. can@ability

3.5.2.3.5 ใช้อธิบายโครงสร้างของประโยค

3.5.3. การแสดงประเภทความหมายหลายๆ ชนิด (Representation of Various Kind of Meaning)

3.5.3.1. แสดงความหมายร่วมกัน (Coordination)

3.5.3.1.1 แสดงความหมายโดยไม่มีขอบเขต

3.5.3.1.2 แสดงความหมายโดยมีขอบเขต

3.5.3.2 แสดงจำนวน : ตัวเลข, จำนวนเลขและปริมาณ

3.5.4 ลักษณะรูปแบบของภาษา UNL (Description Format of UNL)

3.5.4.1. แสดงความหมายด้านตรรกของภาษา UNL

<UNL> ::= {RELATION[':'IDOR] '(' <UW1> ',' <UW2> ')'}...

<UWn> ::= UW | <Scope Node>

<Scope Node> ::= 'S' IDOR

3.5.4.2. รูปแบบที่แสดงของภาษา UNL

[UNL] : แสดงจุดเริ่มต้นของภาษา UNL

[/UNL] : แสดงจุดสิ้นสุดของภาษา UNL

[W] : แสดงจุดเริ่มต้นของชุดคำสากล

[/W] : แสดงจุดสิ้นสุดของชุดคำสากล

[R] : แสดงจุดเริ่มต้นของชุดคำเชื่อมโยง

[/R] : แสดงจุดสิ้นสุดของชุดคำเชื่อมโยง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างประโยค Monkey eats bananas.

UNL

[S]

[W]

eat(icl>action).@present.@entry:00

monkey(icl>animal).@generic:01

banana(icl>food).@generic:02

[/W]

[R]

00agt01

00obj02

[/R]

[/S]

ภาษา UNL ได้จัดเตรียมข้อมูลให้กับทุกๆ ภาษาโดยการใช้ “deconverter” ในการแปลงจากภาษา UNL ให้แสดงเป็นแต่ละภาษาท้องถิ่น (UNL Center, 1996)

3.6 แนวความคิดในการแปลงเป็นภาษาไทย (Concept of Deconverter to THAI Language)

3.6.1 แนวคิดในการแก้ปัญหาการแปลงจากภาษา UNL เป็นภาษาไทย

ปัญหาสำหรับการแปลจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่งนั้นพบว่าคำหนึ่งคำของภาษาหนึ่งนั้นสามารถแปลความหมายไปเป็นอีกภาษาหนึ่งได้หลายความหมาย อีกทั้งยังขึ้นอยู่กับกฎเกณฑ์ของประโยค, วัตถุประสงค์ของผู้เขียน รูปแบบประโยคและปัจจัยอื่นๆ ตัวอย่างของคำว่า “Book” สามารถแปลเป็นภาษาไทยได้ 7 ความหมายไม่นับรวมเมื่อนำไปรวมกับคำอื่น เช่นคำว่า “Book Value” จากปัญหาดังกล่าวพบว่าถ้าใช้วิธีการออกแบบฐานข้อมูลพจนานุกรมในการบรรจุความหมายของคำอย่างชัดเจน ซึ่งจะช่วยให้การใช้คอมพิวเตอร์แปลความหมายได้ถูกต้องชัดเจนยิ่งขึ้นจากการกล่าวถึงลักษณะและรูปแบบของภาษา UNL ดังกล่าวข้างต้นซึ่งจะประกอบด้วยคำคำสากลหนึ่งคำและเชื่อมโยงด้วยคำสากลอีกหนึ่งคำ เมื่อนำมาออกแบบและบรรจุข้อมูลในฐานข้อมูลพจนานุกรมซึ่งจะบรรจุคำหลัก คำสากล(1) และ คำสากล(2) ซึ่งใช้อธิบายความหมายของคำสากล(1) เพิ่มเติม ดังนั้นได้ออกแบบโดยใช้ คำสากล(2) แสดงเป็น “Object Class” ตามตัวอย่างในตาราง 3.1 ซึ่งบรรจุตัวอย่างของคำในฐานข้อมูลพจนานุกรมที่นำเสนอ

ตารางที่ 3.1 ตัวอย่างของ UNL ในฐานข้อมูลพจนานุกรม

คำสากล	รหัสคำเชื่อมโยง	แสดงลักษณะเฉพาะคำ	ภาษาไทย	ประเภทคำ
EAT	Icl	Action	กิน (Kin)	Vi
MONKEY	Icl	Animal	ลิง (Ling)	Noun
BANANA	Icl	Food	กล้วย (Krauy)	Noun

จากตัวอย่างคำว่า “Book” ในภาษาอังกฤษที่เป็นคำสากลตามรูปแบบของภาษา UNL ที่อยู่ในรูปแบบของ HTML File อธิบายได้ดังต่อไปนี้

Book : keeping every possible meaning

Book(icl>publications) : book as publications

Book(=accounts) : written records of a business

Book(obj>room) : to reserve (a room)

จากตัวอย่างข้างต้นได้เสนอแนวคิดวิธีการแปลงจากภาษา UNL เป็นภาษาไทยซึ่งประกอบด้วย :

1. ออกแบบฐานข้อมูลซึ่งใช้บรรจุคำหลักคำสากลในตารางฐานข้อมูล “Universal Words” ซึ่งบรรจุคำหลัก (English head word) หนึ่งคำซึ่งบรรจุทุกความหมายของคำศัพท์ฯ นั้น
2. เพิ่มตาราง “Object Class” สำหรับบรรจุคำเชื่อมโยงคำสากลตัวอย่างเช่น publications, accounts, room.
3. นำ 2 ตารางข้างต้นมารวมกันเป็นตาราง “UNL dictionary” ดังตัวอย่างแสดงไว้ในตาราง 3.2
4. นำคำเชื่อมโยงมาใช้สำหรับออกแบบตาราง "UNL Verb Pattern" เพื่อนำไปใช้สำหรับการหารูปแบบกริยาภาษาไทยและหาค่าของหน้าที่ของคำ ในขั้นตอนการทำหารูปแบบกริยา ซึ่งจะกล่าวในรายละเอียดต่อไป

ตารางที่ 3.2 ตัวอย่างคำ "Book" ในตารางพจนานุกรม UNL

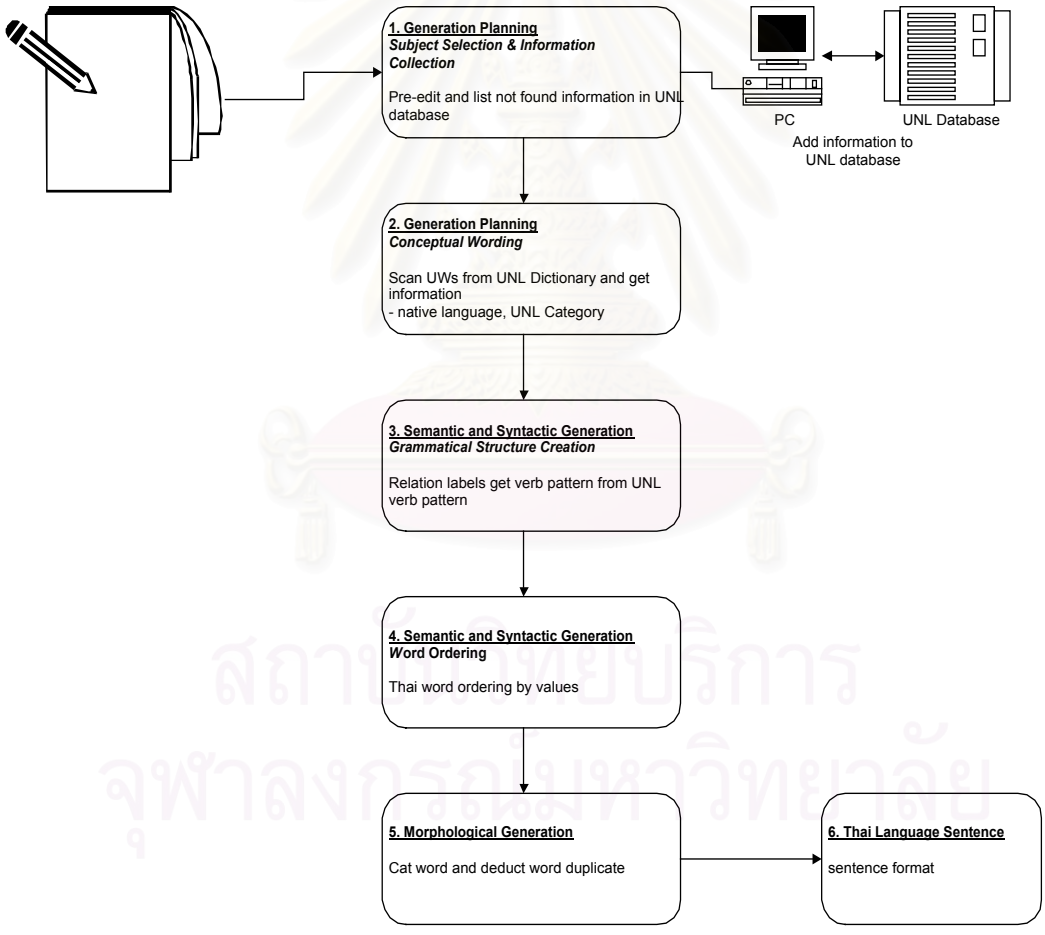
คำสากล	รหัสคำเชื่อมโยง	ลักษณะเฉพาะคำ	ภาษาไทย	ประเภทคำ
Book	Icl	Publications	หนังสือ	Noun
Book	=	Accounts	บัญชี	Noun
Book	Obj	Room	จอง	Vt

3.6.2 ขั้นตอนการแปลงจากภาษา UNL เป็นภาษาไทย (Step of Converter from UNL to THAI)

แนวคิดในการแปลงจากภาษา UNL เป็นภาษาไทยที่วิทยานิพนธ์ฉบับนี้ได้นำเสนอมี 6 ขั้นตอน โดยได้ดัดแปลงมาจากแนวคิดของขั้นตอนระบบการสร้างภาษาไทยจากภาษากลาง (Thai Generation System form Interlingua) โดยมีขั้นตอน 6 ขั้นตอน ซึ่งขั้นตอนต่างๆประกอบด้วย

1. การเตรียมข้อมูลสำหรับสร้างประโยค : จัดเตรียมเนื้อหาและข้อมูล
2. การเตรียมข้อมูลสำหรับสร้างประโยค : หาความหมายของคำ
3. การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย
4. การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย
5. การกำหนดคำภาษาไทย
6. แสดงผลในรูปแบบประโยคภาษาไทย

ขั้นตอนดังกล่าวแสดงได้ด้วยรูปดังนี้



รูปที่ 3.2 ขั้นตอนการแปลงภาษา UNL เป็นภาษาไทย

จากรูปที่ 3.2 อธิบายขั้นตอนต่างๆ 6 ขั้นตอนซึ่งมีรายละเอียดดังต่อไปนี้

3.6.2.1 การเตรียมข้อมูลสำหรับสร้างประโยค : จัดเตรียมเนื้อหาและข้อมูล ในขั้นตอนนี้เรียกว่าโมดูลตรวจสอบความถูกต้อง (Pre-Edit Module) ในขั้นตอนนี้จะทำการอ่านข้อมูลจากเอกสารภาษา UNL ที่อยู่ในรูปแบบ HTML File จากนั้นนำข้อมูลไปค้นหาในฐานข้อมูล UNL เพื่อแสดงข้อความผิดพลาดในกรณีที่ไม่พบในฐานข้อมูล UNL

3.6.2.2 การเตรียมข้อมูลสำหรับสร้างประโยค : หาความหมายของคำ ในขั้นตอนนี้เรียกว่าโมดูลหาความหมายของคำจากพจนานุกรม (Scan Uws From UNL Dictionary Module) ในขั้นตอนนี้จะนำข้อมูลที่ได้จัดเตรียมไว้จากข้อ 1. นำคำศัพท์คำสากลค้นหาในตาราง UNL Dictionary เพื่อดึงความหมายของคำ ประเภทของคำ และจัดเตรียมข้อมูลเพื่อนำไปใช้ในขั้นตอนต่อไป

3.6.2.3 การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย : การสร้างโครงสร้างไวยากรณ์ในขั้นตอนนี้เรียกว่าโมดูลเปรียบเทียบรูปแบบกริยา ในขั้นตอนนี้นำข้อมูลที่ได้จากข้อ 2. และคำต่างๆ ค้นหาและเปรียบเทียบกับตารางรูปแบบกริยา เพื่อหาหน้าที่คำของประโยคและค่าของหน้าที่คำ ของแต่ละคำสากลพร้อมทั้งเตรียมข้อมูลเพื่อนำไปใช้ในขั้นตอนต่อไป

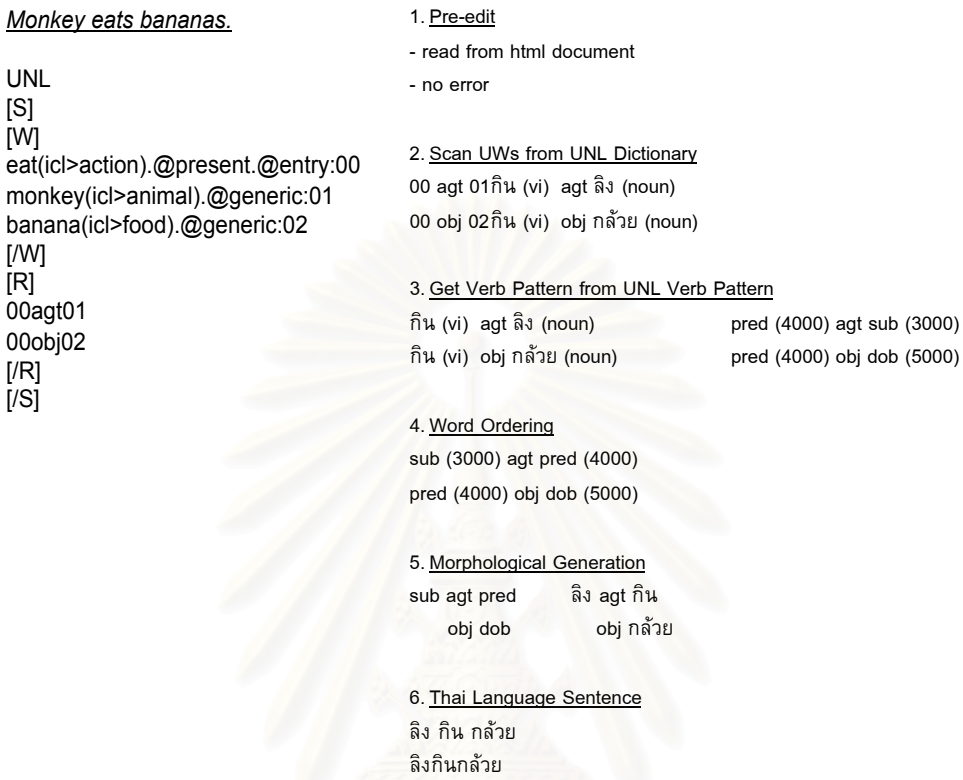
3.6.2.4 การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย : ในขั้นตอนนี้เรียกว่าโมดูลเรียงลำดับคำ (Word Ordering Module) ในขั้นตอนนี้จะนำค่าของหน้าที่คำ มาเปรียบเทียบกันสำหรับในแต่ละประโยคเพื่อจัดเรียงลำดับของแต่ละคำให้ถูกต้องโดยเรียงจากค่าน้อยไปมากตามค่าของหน้าที่คำและจัดเตรียมข้อมูลเพื่อใช้ในขั้นตอนต่อไป

3.6.2.5 การกำหนดคำภาษาไทย : โมดูลจัดรูปแบบประโยคและตัดคำซ้ำซ้อน (Concat word and deduct duplicate word Module) หลังจากจัดเรียงค่าของหน้าที่คำ ถูกต้องตามที่กำหนดแล้ว ในขั้นตอนนี้จะทำหน้าที่ตัดคำที่ซ้ำซ้อนกันในประโยคที่เรียงลำดับติดต่อกัน หลังจากผ่านขั้นตอนนี้จะได้ข้อมูลที่พร้อมจะจัดเป็นรูปแบบประโยคภาษาไทย

3.6.2.6 แสดงผลในรูปแบบประโยคภาษาไทย : โมดูลแสดงผลประโยคภาษาไทย (Thai language sentence Module) ขั้นตอนสุดท้ายนำข้อมูลที่ได้จากข้อ 5. จัดเรียงให้เป็นภาษาไทยที่ถูกต้อง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างขั้นตอนการแปลงจากภาษา UNL เป็นภาษาไทยซึ่งประกอบด้วยขั้นตอน 6 ขั้นตอน ดังตัวอย่างรูปต่อไปนี้



รูปที่ 3.3 ตัวอย่างการแปลงจากภาษา UNL เป็นภาษาไทยประโยค “Monkey eats bananas”

จากรูปที่ 3.3 อธิบายขั้นตอนต่างๆ ได้ดังนี้ ในขั้นตอน 1. ทำการอ่านข้อมูลจากเอกสาร UNL ที่อยู่ในรูปแบบ HTML จากนั้นนำข้อมูลคำสากลไปค้นหาในตารางคำสากลแล้วไม่พบข้อผิดพลาด ในขั้นตอน 2. นำเอาคำสากลไปค้นหาในฐานข้อมูล UNL เพื่อหาความหมายภาษาไทยของคำศัพท์และประเภทของคำศัพท์ ในขั้นตอน 3. นำค่าที่ได้จากขั้นตอนที่ 2 ไปเปรียบเทียบในตารางรูปแบบกริยา ได้หน้าที่ของคำและค่าของหน้าที่คำ จากนั้นนำค่าที่ได้มาจัดเรียงลำดับโดยเรียงจากน้อยไปมาก ซึ่งจะได้ผลลัพธ์ตามข้อ 4. จัดเรียงคำภาษาไทย ขั้นตอน 5. ตัดคำ “กิน” ที่เรียงต่อกันและซ้ำกันให้เหลือ 1 คำ ขั้นตอนสุดท้ายนำความหมายภาษาไทยที่ได้มาจัดเรียงและแสดงผลในรูปแบบภาษาไทย รายละเอียดแต่ละขั้นตอนจะกล่าวถึงในบทถัดไป

UNL Function

Function Code	Name	Native Description	Values
sadv	Sentencial Adverb	ส่วนขยายความประโยค	1000
smod	Sentencial Modefier	ส่วนขยายประโยค	2000
sub	Subject	ประธานของประโยค	3000
pred	Predicate CP	กริยาหลัก	4000
dob	Direct Object	กรรมตรงของประโยค	5000

UNL Universal Words

UN Keyword	Native Description
EAT	กิน, กินอาหาร
MONKEY	ลิง
BANANA	กล้วย

UNL Object Class (A Kind Of)

Object Class	Native Description	Code AKO
action	การกระทำ	2111
animal	สัตว์	13111
food	อาหาร	

UNL Category

Category	Native Description
noun	นาม
vi	อกรรมกริยา
vt	สกรรมกริยา
adv	กริยาวิเศษณ์
adj	คำคุณศัพท์

UNL Dictionary

UW Keyword	RL Code	Object Class	Native Language	Category
EAT	icl	action	กิน	vi
MONKEY	icl	animal	ลิง	noun
BANANA	icl	food	กล้วย	noun

UNL Relation Labels

Code	Name	Description	Native Description	Type	Native Language
agt	Agent	An agent which causes volitional action.	ใช้แสดงความสัมพันธ์ว่าเป็นผู้กระทำเหตุการณ์อย่างตั้งใจ	1	
		Agent is an animated object with intention.	ผู้กระทำที่เป็นสิ่งที่เคลื่อนได้ด้วยเครื่องยนต์	1	
obj	Object	An object of action or change that affects it.	แสดงความสัมพันธ์ที่เป็นผู้ถูกกระทำ หรือผู้ทรงสภาพ	1	
icl	Inclusion	Relation labels between UWs	แสดงความสัมพันธ์ระหว่างคำ	2	

UNL Attribute Labels

Code	Description	Native Description	Type	Native Language
generic	Generic UW	แสดงความทั่วไปของ UW	1	
present	Event happening at present	แสดงเหตุการณ์ว่าเกิดขึ้นในเวลาปัจจุบัน	2	
entry	dominating UW in a UNL for a sentence	UW ใน UNL ที่ครอบครองประโยค	5	

UNL Verb Pattern

Category	AL Code.1	AL Code.2	AL Code.3	AL Code.4	AL Code.5	RL Code	Category	AL Code.6	AL Code.7	AL Code.8	AL Code.9	AL Code.10	Func Code.1	RL Code	Func.Code.2
vi	present	entry				agt	noun	generic					pred	agt	sub
vi	present	entry				obj	noun	generic					pred	obj	dob

รูปที่ 3.4 ตัวอย่างตารางในฐานข้อมูล UNL
สำหรับใช้ในการแปลงประโยค “Monkey eats bananas”

ในรูปที่ 3.4 เป็นตัวอย่างตารางต่างๆในฐานข้อมูล UNL ที่ใช้เป็นข้อมูลสำหรับการแปลจากภาษา UNL เป็นภาษาไทยเพื่อให้ได้ผลลัพธ์ดังแสดงในรูปที่ 3.3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4
การออกแบบโปรแกรม CU-Deconverter

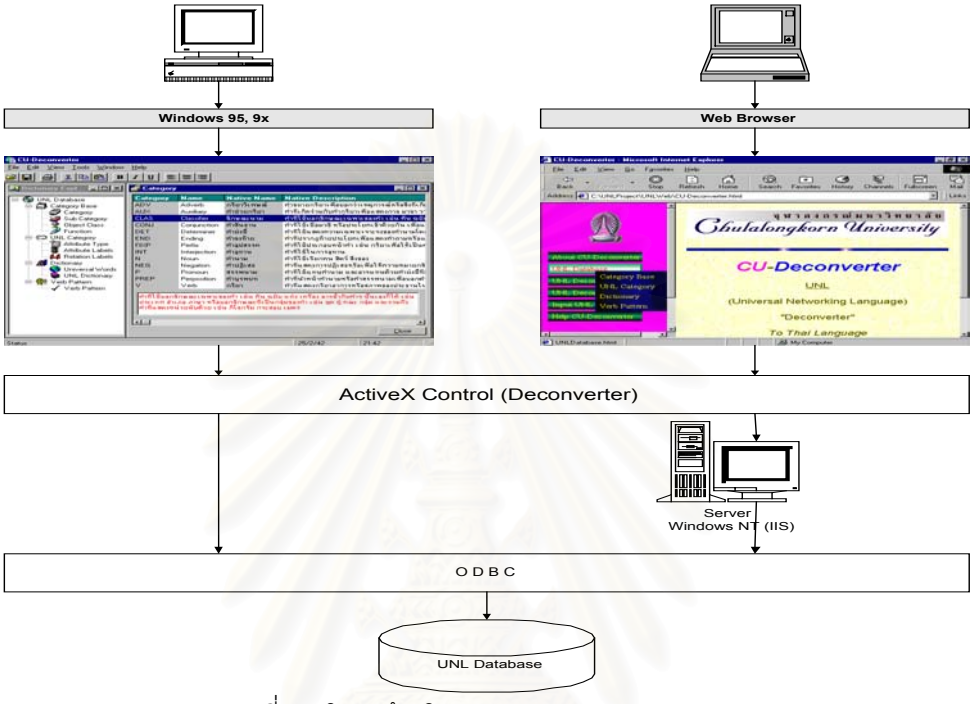
4.1 โครงสร้างของโปรแกรม CU-Deconverter (Structure of Program CU-Deconverter)

จากบทที่แล้วได้นำเสนอแนวความคิดและขั้นตอนในการแปลงจากภาษา UNL เป็นภาษาไทยโดยอ้างอิงจากขั้นตอนการแปลจากภาษากลางเป็นภาษาไทยนั้น ในบทนี้จะกล่าวถึงการออกแบบและโครงสร้างการทำงานของโปรแกรม CU-Deconverter ที่ใช้ในการแปลงจากภาษา UNL เป็นภาษาไทยโดยลักษณะโครงสร้างของโปรแกรมได้ออกแบบให้มีการติดต่อกับผู้ใช้ 2 แบบ คือแบบวินโดว์และแบบเวบ เมื่อจะใช้งานโปรแกรมจะปรากฏหน้าจอให้เลือกแฟ้มข้อมูลที่มีชนิดเป็น HTML หรือเลือกหน้าจอเวบเพจ จากนั้นโปรแกรมจะทำการแปลงทีละขั้นตอนทั้งหมด 6 ขั้นตอนโดยเริ่มตั้งแต่

1. จัดเตรียมเนื้อหาและข้อมูล ทำหน้าที่ตรวจสอบข้อผิดพลาดของแฟ้มข้อมูลหรือหน้าเวบเพจในกรณีที่ค้นหาข้อมูลไม่พบในฐานข้อมูลหรือมีข้อความในแฟ้มข้อมูลที่ไม่ถูกต้องพร้อมทั้งแสดงผลข้อความที่ไม่ถูกต้อง ทั้งนี้ในกรณีที่ค้นหาคำหรือข้อมูลอื่นๆไม่พบในฐานข้อมูลผู้ใช้สามารถที่จะเพิ่มเติมคำหรือข้อมูลเพิ่มเข้าไปในฐานข้อมูลเพื่อให้มีความสมบูรณ์มากขึ้น
2. หาความหมายของคำ ในขั้นตอนนี้จะทำการค้นหาความหมายภาษาไทยของคำและส่วนประกอบของคำจากฐานข้อมูลเพื่อนำข้อมูลมาใช้ประกอบในการแปลงเป็นภาษาไทย
3. เปรียบเทียบรูปแบบกริยา ขั้นตอนนี้เป็นการนำข้อมูลที่ได้จากขั้นตอนที่ 2 ทำการตรวจสอบเปรียบเทียบกับรูปแบบกริยาในฐานข้อมูลรูปแบบกริยา เพื่อทำการหาลักษณะของกริยาที่ถูกต้องตามหลักภาษาไทย
4. จัดเรียงคำ ในขั้นตอนนี้จะทำการเรียงลำดับของคำตามค่าของหน้าที่ของคำ ที่ได้จากขั้นตอนที่ 3 โดยทำการเรียงจากน้อยไปมากพร้อมทั้งนำผลลัพธ์ที่ได้ใช้ในขั้นตอนต่อไป
5. การกำหนดคำภาษาไทย จากผลลัพธ์ในขั้นตอนที่ 4 ขั้นตอนนี้จะทำหน้าที่ในการตัดหน่วยของคำ (Morpheme) ที่ซ้ำซ้อนโดยเปรียบเทียบกับคำก่อนหน้าที่เหมือนกันแล้วตัดให้เหลือ 1 คำพร้อมทั้งสร้างแฟ้มข้อมูลเพื่อใช้ในขั้นตอนสุดท้าย
6. แสดงผลในรูปแบบประโยคภาษาไทย ในขั้นตอนสุดท้ายนี้จะทำหน้าที่ในการจัดเรียงประโยคให้ถูกต้องและแสดงผลเป็นภาษาไทย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โดยลักษณะของโปรแกรมหลัก CU-Deconverter ทั้ง 6 ขั้นตอนนี้ถูกพัฒนาให้เป็นแบบ ActiveX โดยสามารถเรียกใช้โปรแกรมผ่านได้ทั้งวินโดว์และเวบ ในส่วนฐานข้อมูล UNL นั้นจะใช้ ไมโครซอฟท์แอสเซส โดยติดต่อผ่านไมโครซอฟท์เจ็ท หรือ ODBC ตามรูปที่ 4.1. ส่วนประกอบของโปรแกรมแบ่งตามโครงสร้างทั้งในแบบวินโดว์และเวบนั้นแบ่งเป็น 4 ส่วน



รูปที่ 4.1 โครงสร้างโปรแกรม CU-Deconverter

จากรูปที่ 4.1 โครงสร้างโปรแกรมแบ่งออกเป็น 4 ส่วนได้แก่

4.1.1 ลักษณะติดต่อแบบวินโดว์ (Windows Interface)

4.1.2 ลักษณะติดต่อแบบเวบ (Web Interface)

4.1.3 โครงสร้างโปรแกรม Deconverter (Structure of Deconverter Program)

4.1.4 ฐานข้อมูล UNL (UNL Database)

4.1.1 ลักษณะโปรแกรมติดต่อแบบวินโดว์ (Windows Interface)

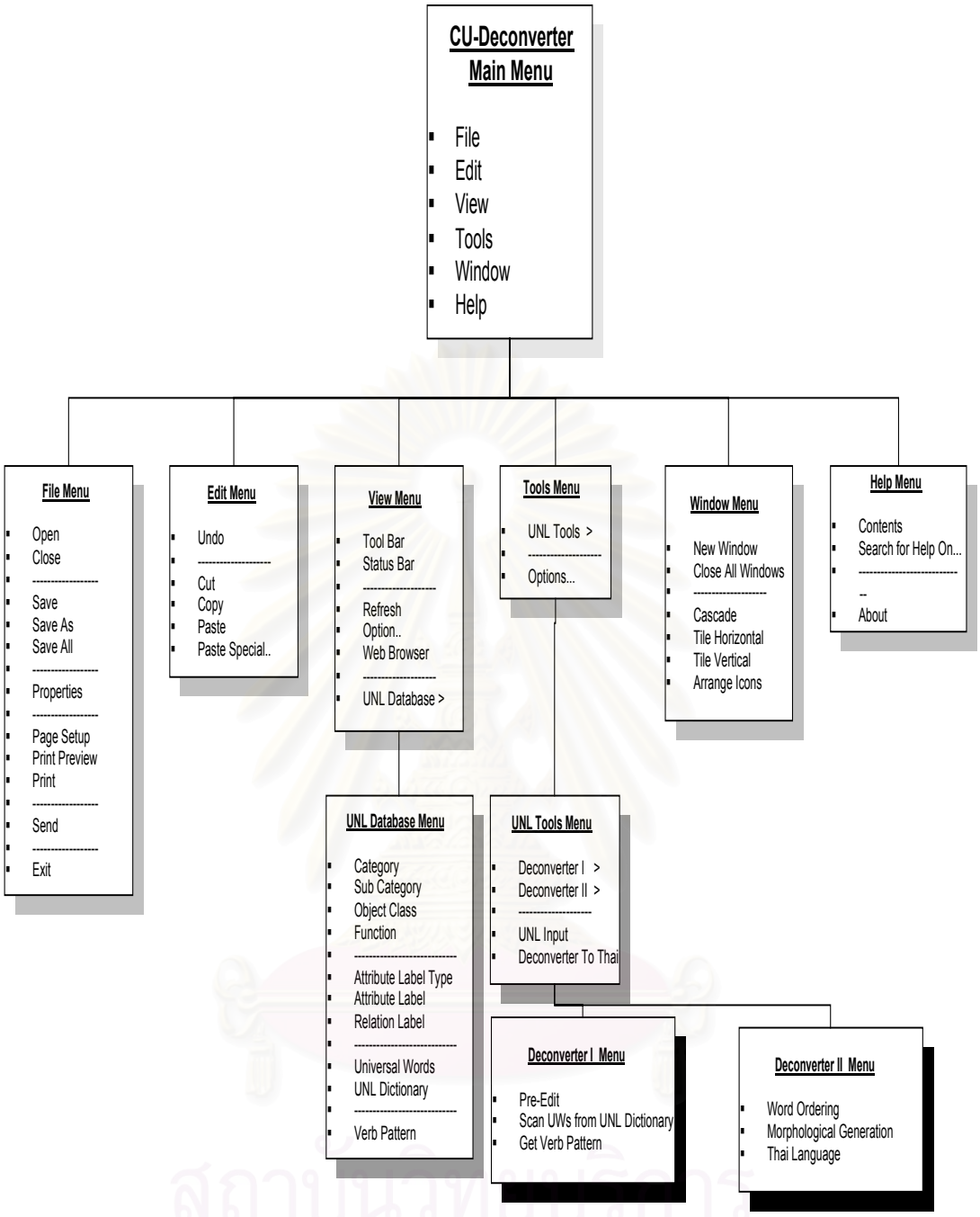
ในส่วนนี้ได้จัดแบ่งการทำงานของโปรแกรมตามลักษณะการติดต่อตามมาตราฐานแบบวินโดว์ โดยทั่วไปตามรูปที่ 4.2 โดยได้ออกแบบแบ่งออกเป็น 6 ส่วนดังตารางที่ 4.1 นี้

ตารางที่ 4.1 เมนูการทำงานของโปรแกรมที่เป็นลักษณะการติดต่อแบบวินโดว์

เมนู	หน้าที่การทำงาน
File Menu	เป็นส่วนที่เกี่ยวข้องกับการติดต่อฐานข้อมูล
Edit Menu	เป็นส่วนที่เกี่ยวข้องกับการแก้ไขข้อมูล
View Menu	เป็นส่วนที่ใช้ในการเรียกดูฐานข้อมูลหรือการเรียกดูผ่าน เวบ
Tools Menu	เป็นส่วนที่ใช้ในการแปลงจากภาษา UNL เป็นภาษาไทย
Window Menu	เป็นส่วนที่ใช้จัดการเกี่ยวกับวินโดว์
Help Menu	เป็นส่วนที่ใช้แสดงข้อความช่วยเหลือและข้อมูลของระบบ

โดยเมนูทั้ง 6 ส่วนนี้จะทำงานในลักษณะของ Poll Down เมนูตามลักษณะของวินโดว์และจะมีปุ่มไอคอนประกอบตามหัวข้อแต่ละเมนู เช่น ปุ่มเปิด (Open) เมื่อทำการคลิกที่ปุ่มเปิดหรือเลือกผ่านเมนูแฟ้มข้อมูล (File Menu) ตัวโปรแกรมจะทำการเปิด UNL Database Explorer ให้ผู้ใช้ทำการเลือกตารางที่ต้องการจะดูข้อมูลหรือแก้ไขข้อมูลโดยในส่วนนี้ทำหน้าที่เหมือนวินโดว์ Explorer หลังจากที่กดเลือกตารางแล้ว ก็จะแสดงผลของข้อมูลในอีกวินโดว์หนึ่ง ในส่วนของเมนูแก้ไข (Edit Menu) จะมีส่วนประกอบที่ใช้ในการเข้าข้อมูลหรือจัดการข้อความเช่น จัดชิดซ้าย จัดข้อความตรงกึ่งกลางเป็นต้น ในส่วนของเมนูเรียกดู (View Menu) นั้นจะมีเมนูสำหรับเลือกดูฐานข้อมูล UNL และมีเมนูสำหรับเรียกโปรแกรม CU-Deconverter ผ่านทางหน้าจอเวบโดยมีลักษณะเหมือนเวบเพจทุกประการ แต่เป็นการเรียกดูผ่านโปรแกรมที่เป็นการติดต่อแบบวินโดว์ในส่วนของเมนูเครื่องมือ (Tools Menu) นั้นประกอบด้วยหน้าจอสำหรับเลือกแฟ้มข้อมูลที่มีชนิดเป็น HTML สำหรับเป็นใช้ในการแปลงจากภาษา UNL เป็นภาษาไทยและมีเมนูในขั้นตอนต่างๆ 6 ขั้นตอนที่ใช้ในการแปลงจากภาษา UNL เป็นภาษาไทย ในส่วนของวินโดว์เมนูนั้นประกอบไปด้วยส่วนที่ใช้จัดการกับวินโดว์ตามมาตราฐานแบบวินโดว์ที่ใช้โดยทั่วไป และในส่วนสุดท้ายที่เป็นเมนูช่วย (Help Menu)นั้นจัดเตรียมไว้สำหรับอธิบายรายละเอียดของโครงการ CU-Deconverter และใช้ในการค้นหาข้อมูลที่เป็นรายละเอียดวิธีการทำงานของการแปลงจากภาษา UNL เป็นภาษาไทยรวมไปถึงประวัติความเป็นมาและปัญหาต่างๆ ของการใช้การแปลภาษาด้วยเครื่อง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.2 โครงสร้างโปรแกรมลักษณะติดต่อแบบวินโดว์

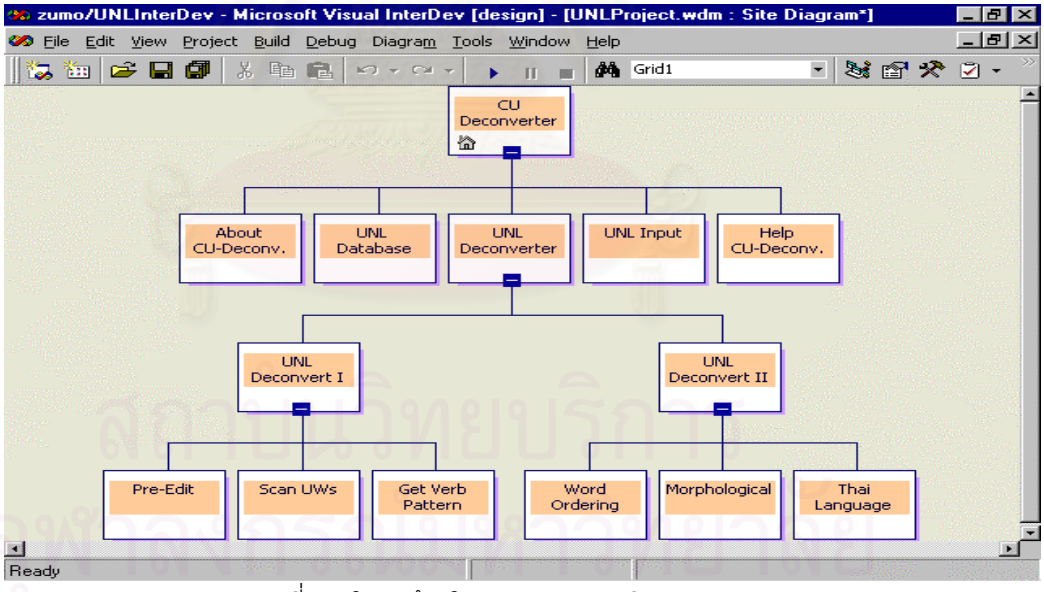
4.1.2 **ลักษณะโปรแกรมติดต่อแบบเวบ** (Web Interface)

ในส่วนของการติดต่อแบบเวบนี้ได้จัดแบ่งการทำงานของโปรแกรมเหมือนลักษณะของเวบโดยทั่วไป โดยได้ออกแบบแบ่งออกเป็น 5 ส่วนดังตารางที่ 4.2 นี้

ตารางที่ 4.2 หน้าจอเวบเพจของโปรแกรมที่เป็นลักษณะการติดต่อแบบเวบ

หน้าเวบเพจ	หน้าที่การทำงาน
About CU-Deconverter	เป็นเพจที่อธิบายเกี่ยวกับรายละเอียดของ Project
UNL Database	เป็นเพจที่ใช้ในการติดต่อกับฐานข้อมูล UNL
UNL Deconverter	เป็นเพจในส่วนของโปรแกรม Deconverter
UNL Input	เป็นเพจในส่วนที่ใช้ติดต่อกับผู้ใช้ในการเปิดแฟ้มหรือเวบเพจที่ต้องการแปลงจาก UNL เป็นภาษาไทย
Help CU-Deconverter	เป็นเพจที่มีคำอธิบายข้อมูลต่างๆ เกี่ยวกับประวัติความเป็นมาของการแปลภาษาด้วยเครื่อง, ข้อมูลเกี่ยวกับ UNL และรายละเอียดต่างๆ

จากหน้าที่การทำงานของแต่ละเวบเพจสามารถจัดโครงสร้างโปรแกรมที่มีลักษณะติดต่อแบบเวบได้ดังรูปที่ 4.3



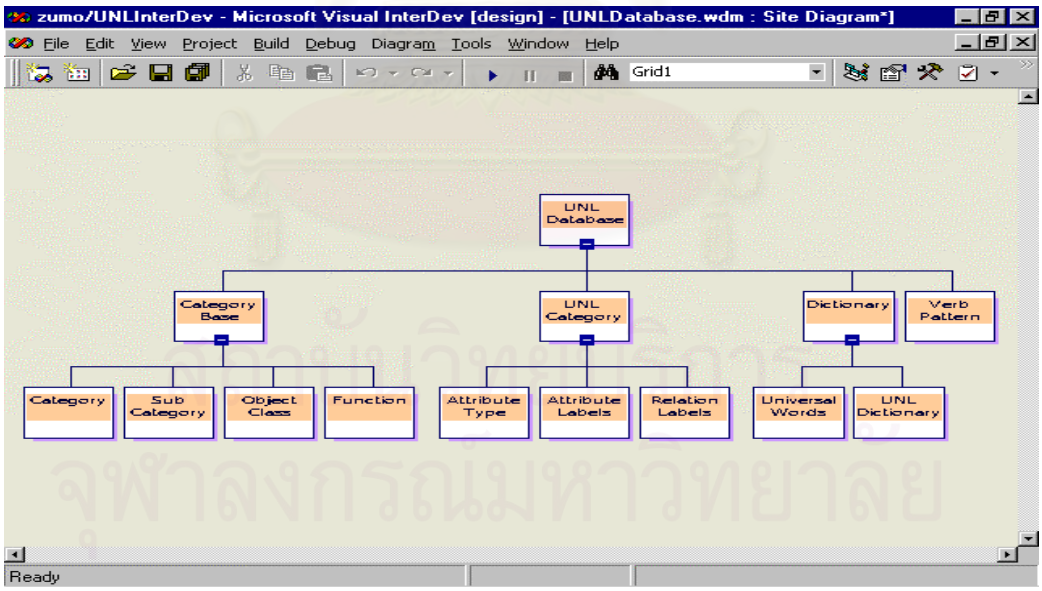
รูปที่ 4.3 โครงสร้างโปรแกรมลักษณะติดต่อแบบเวบ

โดย Home Page ที่เริ่มต้นของเวบนี้จะมีหน้าจอของ CU-Deconverter เป็นหน้าจอเริ่มต้น โดยทางซ้ายมือจะแบ่งเป็นเฟรมของเมนูการทำงานทั้ง 5 ส่วนและทางขวามือจะเป็นเฟรม About CU-Deconverter โดย

ในส่วนของเพจนี้จะอธิบายรายละเอียดของโครงการ ต่อมาในส่วนของเพจ ฐานข้อมูล UNL จะแบ่งออกเป็น 4 ส่วนตามรูปที่ 4.4 คือ

1. ประเภทคำพื้นฐาน (Category Base) ในส่วนนี้เป็นฐานข้อมูลที่เป็นข้อมูลพื้นฐานของโปรแกรม
2. ประเภทคำ UNL (UNL Category) บรรจุข้อมูลพื้นฐานที่เป็นไปตามมาตราฐานและกฎเกณฑ์ของภาษา UNL
3. พจนานุกรม (Dictionary) เป็นฐานข้อมูลพจนานุกรมที่ได้ออกแบบไว้ใช้สำหรับการแปลงเป็นภาษาไทย
4. รูปแบบกริยา (Verb Pattern) บรรจุรูปแบบกริยาที่ใช้ในการเปรียบเทียบเพื่อแปลงเป็นประโยคของภาษาไทย

มาในส่วนของ UNL Deconverter ประกอบไปด้วยขั้นตอนต่างๆ 6 ขั้นตอนเหมือนกับในส่วนแบบวินโดว์ตั้งแต่โมดูลตรวจสอบความถูกต้อง ไปจนถึงโมดูลแสดงผลประโยคภาษาไทยในส่วนของการเลือกแฟ้มข้อมูลภาษา UNL เป็นส่วนที่ใช้สำหรับให้ผู้ใช้เลือกแฟ้มข้อมูลหรือหน้าจอเวบเพจที่ต้องการแปลงเป็นภาษาไทยและส่วนสุดท้ายเป็นเพจของ Help Deconverter ซึ่งจะมีข้อมูลอธิบายเกี่ยวกับขั้นตอนการทำงานของโปรแกรมและรายละเอียดของประวัติความเป็นมาและปัญหาของการใช้เครื่องคอมพิวเตอร์ในการแปลภาษา รวมไปถึงผลงานวิจัยในอดีตที่ผ่านมาเกี่ยวกับ การแปลภาษาด้วยเครื่อง (ดูตัวอย่างหน้าจอในภาคผนวก ข)



รูปที่ 4.4 โครงสร้างโปรแกรมลักษณะติดต่อแบบเวบในส่วนของฐานข้อมูล

4.1.3 **โครงสร้างโปรแกรม** Deconverter (Structure of Deconverter Program)

ในส่วนของโปรแกรมนั้นได้ออกแบบโดยใช้แนวความคิดมาจากขั้นตอนการแปลจากภาษากลางเป็นภาษาไทย ซึ่งจะอธิบายรายละเอียดเปรียบเทียบในหัวข้อการทำงานของโปรแกรม Deconverter แต่ละหัวข้อโดยที่ประกอบด้วย 6 โมดูลตามรูปที่ 4.5 และแนวคิดในการออกแบบโปรแกรมในแต่ละโมดูลจะทำการสร้างแฟ้มข้อมูลเอกสาร (Text File) ในแต่ละโมดูลของตัวเองเพื่อให้โมดูลถัดไปทำงานโดย 6 โมดูลนั้นประกอบด้วย

ตัวอย่างขั้นตอนการทำงานของโปรแกรมแต่ละโมดูลแสดงได้ดังรูปตัวอย่างต่อไปนี้

1. Pre-edit
- read from html document
- no error

2. Scan UWs from UNL Dictionary
00 agt 01กิน (vi) agt ลิง (noun)
00 obj 02กิน (vi) obj กล้วย (noun)

3. Get Verb Pattern from UNL Verb Pattern
กิน (vi) agt ลิง (noun) pred (4000) agt sub (3000)
กิน (vi) obj กล้วย (noun) pred (4000) obj dob (5000)

4. Word Ordering
sub (3000) agt pred (4000)
pred (4000) obj dob (5000)

5. Morphological Generation
sub agt pred ลิง agt กิน
obj dob obj กล้วย

6. Thai Language Sentence
ลิง กิน กล้วย
ลิงกินกล้วย

รูปที่ 4.5 ขั้นตอนการทำงานของโปรแกรม CU-Deconverter

รายละเอียดการทำงานของแต่ละขั้นตอนแบ่งออกเป็นโมดูล โดยแต่ละโมดูลจะประกอบด้วยขั้นตอนย่อยๆ ดังตารางต่อไปนี้

ตารางที่ 4.3 รายละเอียดขั้นตอนการทำงานแต่ละโมดูล

โมดูล	ขั้นตอนการทำงาน
1. จัดเตรียมเนื้อหาและข้อมูล	1. ให้เลือกแฟ้มข้อมูลหรือหน้าเวบเพจสำหรับการแปลงจากภาษา UNL เป็นภาษาไทย 2. อ่านแฟ้มข้อมูล HTML และทำการค้นหาคำศัพท์คำสากลจากฐานข้อมูล UNL

	3. ในกรณีที่ไม่เจอคำศัพท์คำสากลจะแสดงข้อความผิดพลาด 4. จัดเตรียมรูปแบบเป็นแฟ้มข้อมูลเอกสารสำหรับใช้ในการโมดูลถัดไป
2. หาความหมายของคำ	1. ใช้คำศัพท์คำสากลที่ได้ไปค้นหาความหมายของคำศัพท์และประเภทของคำในตารางพจนานุกรม UNL 2. สร้างแฟ้มข้อมูล “Scan Uws.Txt" สำหรับโมดูลถัดไป
3. เปรียบเทียบรูปแบบกริยา	1. ค้นหารูปแบบของประโยคจากตารางข้อมูลรูปแบบกริยา UNL 2. ดึงคำอธิบายของหน้าที่คำและค่าของหน้าที่คำ จากตาราง หน้าที่ของคำ 3. สร้างแฟ้มข้อมูล "Get Verb Pattern.Txt"
4. จัดเรียงคำ	1. เรียงลำดับของคำตามค่าของหน้าที่คำที่ได้จากโมดูลเปรียบเทียบรูปแบบกริยาโดยเรียงจากน้อยไปมาก 2. สร้างแฟ้มข้อมูล "Word Order.Txt"
5. หาความหมายของคำ	1. ในขั้นตอนนี้ทำการตัดคำที่ซ้ำโดยเปรียบเทียบคำศัพท์ที่เรียงต่อเนื่องกันและซ้ำกันจะตัดคำหลังที่ซ้ำทิ้ง 2. แทนค่าของหน้าที่คำที่เรียงลำดับแล้วด้วยความหมายภาษาไทยของคำแต่ละคำสำหรับใช้ในการสร้างเป็นภาษาไทยในโมดูลถัดไป 3. สร้างแฟ้มข้อมูล "Morphological Generation.Txt"
6. แสดงผลในรูปแบบประโยคภาษาไทย	1. แทนค่าของคำเชื่อมโยงด้วยความหมายภาษาไทย 2. จัดคำให้เรียงต่อกันและแสดงผลในรูปแบบประโยคภาษาไทย

4.1.4 **ฐานข้อมูล** UNL (UNL Database)

ในส่วนของฐานข้อมูลที่ใช้ในการแปลงจากภาษา UNL เป็นภาษาไทยนั้นได้ทำการออกแบบฐานข้อมูลของ UNL Database เป็นตารางต่างๆ โดยแบ่งเป็น

1. ส่วนของข้อมูลพื้นฐาน เช่น ประเภทของคำ, หน้าที่และค่าของคำแต่ละประเภท
2. ส่วนของพจนานุกรมสำหรับเก็บข้อมูลที่อธิบายความหมายของคำอย่างชัดเจนและ
3. ในส่วนของรูปแบบประโยคเพื่อใช้ในการเปรียบเทียบรูปแบบประโยคสำหรับการแปลงจากภาษา UNL เป็นภาษาไทย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

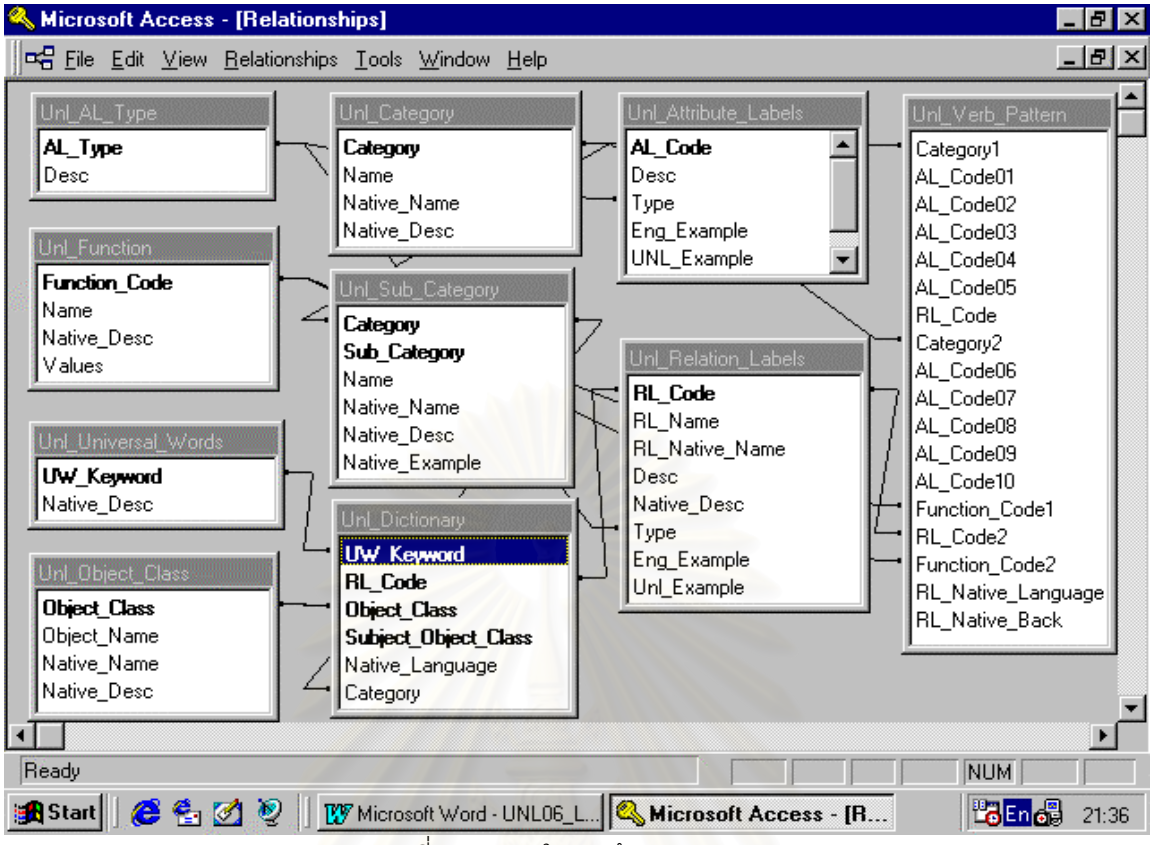
โดยทั้งหมดนี้ในฐานข้อมูลประกอบด้วย 8 ตารางหลักๆ ดังรูปที่ 4.6 และรายละเอียดของแต่ละตาราง โดยแสดงรายละเอียดไว้ที่ตารางที่ 4.4 ดังต่อไปนี้

ตารางที่ 4.4 รายละเอียดตารางต่างๆ ในฐานข้อมูลพจนานุกรม

	ชื่อตาราง	รายละเอียดของตาราง
1.	ตารางลักษณะเฉพาะคำ	ตารางประเภทชนิดของวัตถุต่างๆเพื่อใช้ประกอบสำหรับบอกความหมายของคำเพื่อให้สื่อความหมายได้อย่างชัดเจน เช่น Animal, Man
2.	ตารางคำสากล	ตารางคำศัพท์โดยเป็นคำเดี่ยวๆ อธิบายความหมายของคำโดยทั่วไปในทุกๆความหมาย โดยไม่เฉพาะเจาะจงความหมาย เช่น คำว่า “Monkey” อธิบายได้ 4 ความหมายดังนี้ 1. ลิง รวมทั้งชะนีและเสน 2. เด็กซน, เที่ยวซน 3. เลียน, ล้อ, คนโทน้ำ 4. น้ำหนักที่เลื่อนขึ้นเลื่อนลงบนราง สำหรับตอกเข็ม
3.	ตารางประเภทคำ	ตารางประเภทคำสำหรับบอกประเภทของคำ เช่น Noun, Vi, Vt
4.	ตารางคำเชื่อมโยง	ตารางของคำที่ใช้เป็นตัวเชื่อมระหว่างคำศัพท์ ซึ่งเป็นรูปแบบตามกฎเกณฑ์ของ UNL
5.	ตารางคุณลักษณะคำ	ตารางประเภทบอกลักษณะ ใช้อธิบายลักษณะของคำที่ใช้ในประโยค ซึ่งเป็นรูปแบบตามกฎเกณฑ์ของ UNL
6.	ตารางพจนานุกรม	ตารางพจนานุกรมภาษา UNL เป็นตารางที่ใช้เป็นหลักในการหาความหมายของคำศัพท์ที่อธิบายไว้อย่างชัดเจน โดยจะประกอบด้วยคำ UNL Object Class และคำศัพท์ UNL Universal Words
7.	ตารางหน้าที่คำ	ตารางหน้าที่ของคำโดยประกอบด้วยหน้าที่ของคำ, ความหมายและค่าของหน้าที่คำเพื่อนำไปใช้ในกรณีจัดเรียงคำ
8.	ตารางรูปแบบกริยา	ตารางรูปแบบกริยาเป็นตารางที่ใช้ในการเก็บรูปแบบกริยา โดยอ้างอิงจากประเภทของคำ, คำบอกคุณสมบัติ, คำเชื่อมโยงตามรูปแบบภาษา UNL โดยจะ map ได้หน้าที่คำ และ คำเชื่อมโยง

ตารางข้อมูลต่างๆ จัดทำขึ้นโดยใช้ไมโครซอฟท์แอสเซส ซึ่งประกอบด้วยตารางต่างๆ ดังรูป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.6 ตารางในฐานข้อมูล UNL

โดยในส่วนของตารางต่างๆ ในฐานข้อมูล UNL นั้นจะมีความสัมพันธ์กับขั้นตอนการทำงานของโปรแกรมในแต่ละขั้นตอนซึ่งจะอธิบายรายละเอียดในหัวข้อของขั้นตอนการทำงานของโปรแกรม โดยในส่วนแรกของข้อมูลพื้นฐานนั้นประกอบด้วยตารางประเภทคำ, ตารางหน้าที่คำ, ตารางคำเชื่อมโยงและตารางแสดงคุณลักษณะ ส่วนที่สองของข้อมูลพจนานุกรมประกอบด้วย ตารางแสดงลักษณะเฉพาะของคำ, ตารางคำสากลและตารางพจนานุกรม ส่วนสุดท้ายที่เป็นรูปแบบประโยคนั้นได้แก่ตารางรูปแบบกริยา โดยตารางหลักที่ใช้ในโปรแกรมคือตารางพจนานุกรม ซึ่งเป็นตารางที่บรรจุความหมายของคำอย่างชัดเจนโดยประกอบจากตารางคำสากล, ตารางแสดงลักษณะเฉพาะของคำ และตารางคำเชื่อมโยง

4.2 ขั้นตอนการทำงานของโปรแกรม (Process of Deconverter Program)

แนวคิดของการออกแบบโปรแกรมสำหรับใช้ในการแปลงจากภาษา UNL เป็นภาษาไทยนั้น ได้ออกแบบโดยอ้างอิงจากขั้นตอนการแปลจากภาษากลางเป็นภาษาไทยจากระบบการสร้างภาษาไทยเพื่อการแปลภาษาด้วยคอมพิวเตอร์ (Thai Generation System for Machine Translation) ของผศ.ดร.รัตติกร วรากูลศิริพันธุ์ โดยโครงสร้างของระบบการสร้างภาษาไทยจากภาษากลางแบ่งออกเป็น 3 ขั้นตอนได้แก่ การเตรียมข้อมูล

สำหรับการสร้างประโยคภาษาไทย (Generation Planning) การสร้างโครงสร้างสำหรับประโยคตามไวยากรณ์ภาษาไทย (Semantic and Syntactic Generation) และการกำหนดคำภาษาไทย (Morphological Generation) โดยจะทำการเปรียบเทียบกับขั้นตอนการทำงานของโปรแกรม CU-Deconverter ที่ใช้แปลงจากภาษา UNL เป็นภาษาไทยซึ่งแบ่งขั้นตอนต่างๆ ได้ดังนี้

1. การเตรียมข้อมูลสำหรับการสร้างประโยคภาษาไทยของระบบการสร้างภาษาไทยจากภาษากลาง

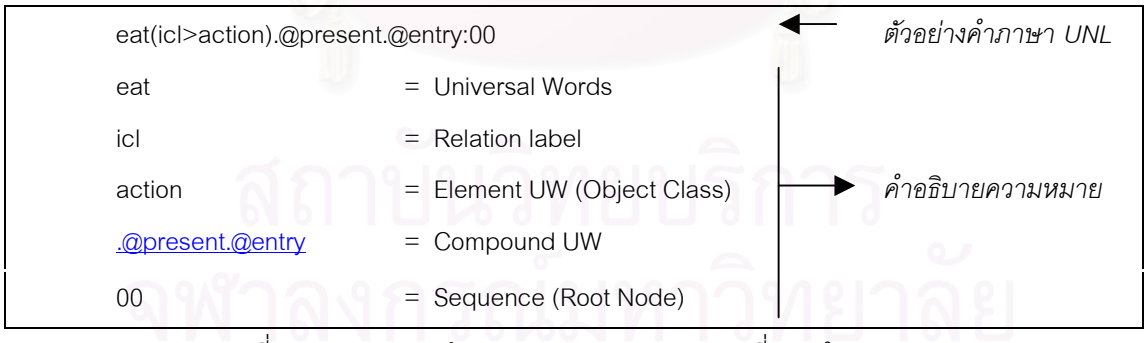
ในส่วนนี้จะทำการดึงข้อมูลจากพจนานุกรมภาษาไทยโดยใช้ส่วนแสดงความหมายของคำหรือการกจากภาษากลางเป็นคำหลักหรือตัวกำหนดตำแหน่งของข้อมูลจากพจนานุกรม ข้อมูลจะประกอบด้วยคำในภาษาไทย และคุณสมบัติทางด้านไวยากรณ์ของคำนั้นๆ ได้แก่

ก) คุณสมบัติทางด้านของคำ ได้แก่การเป็น นาม, สรรพนาม, กริยา, คุณศัพท์, กริยาวิเศษณ์, บุพบท, สันธาน เป็นต้น

ข) คุณสมบัติทางด้านโครงสร้างที่บ่งความสัมพันธ์ระดับลึกและระดับผิวของคำนั้นๆ กับคำอื่นโดยเฉพาะกริยาหลักเราเรียกข้อมูลนี้ว่าการเปรียบเทียบข้อมูล(Mapping Information) ย่อว่า TMAP เช่น กริยา “เคลื่อน” ในภาษาไทยถ้าเป็นสกรรมกริยาจะมี TMAP เป็น {SUB=AGT, DOB=OBJ, COMP=TAR} และถ้าเป็นอกรรมกริยาจะมี TMAP เป็น {SUB=OBJ, COMP=TRA}

นอกจากนั้นจะมีข้อมูลอื่นๆ เช่น รูปแบบกริยาข้อมูลดังกล่าวทั้งหมดจะถูกเก็บอยู่ใน กลุ่มอรรถศาสตร์ของพจนานุกรม ซึ่งจะถูกนำมาเปรียบเทียบกับข้อมูลที่ได้รับมาจากการวิเคราะห์ภาษาต้นแบบ เพื่อเตรียมการสร้างภาษาไทยโดยยึด กริยาหลักเป็นหลักในการค้นหาข้อมูลที่ถูกต้องด้วยการ traversing แบบ depth-first algorithm

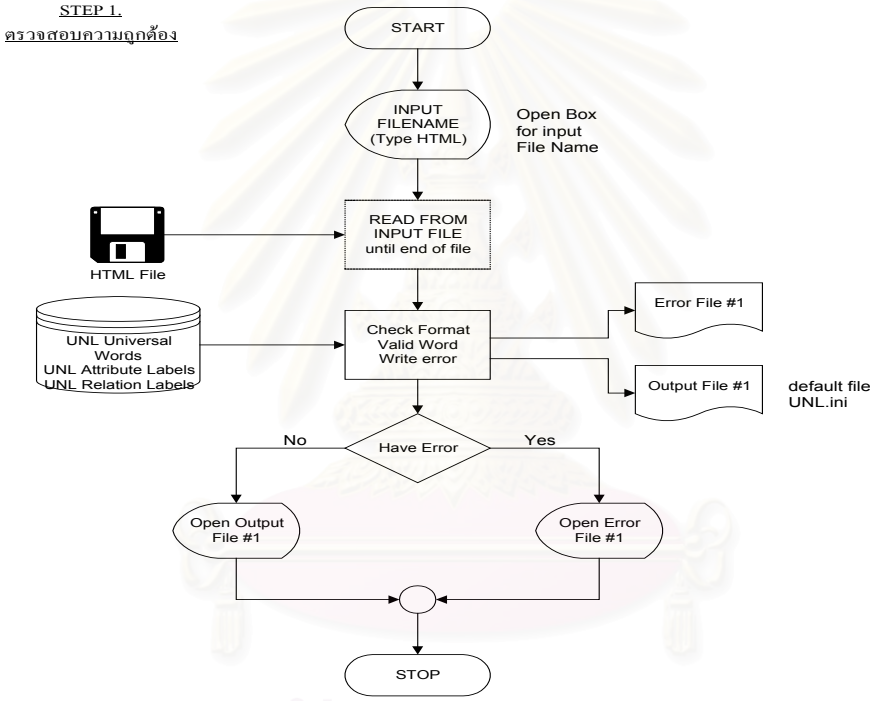
จากขั้นตอนดังกล่าวได้นำมาใช้เป็นแนวคิดและปรับเปลี่ยนเป็น 2 ขั้นตอนได้แก่ 1.1 ตรวจสอบความถูกต้อง, 1.2 หาความหมายของคำ โดยได้ยกตัวอย่างคำศัพท์ของภาษา UNL 1 บรรทัดดังแสดงไว้ในรูปที่ 4.7 เพื่ออธิบายขั้นตอนการทำงานของโปรแกรมดังต่อไปนี้



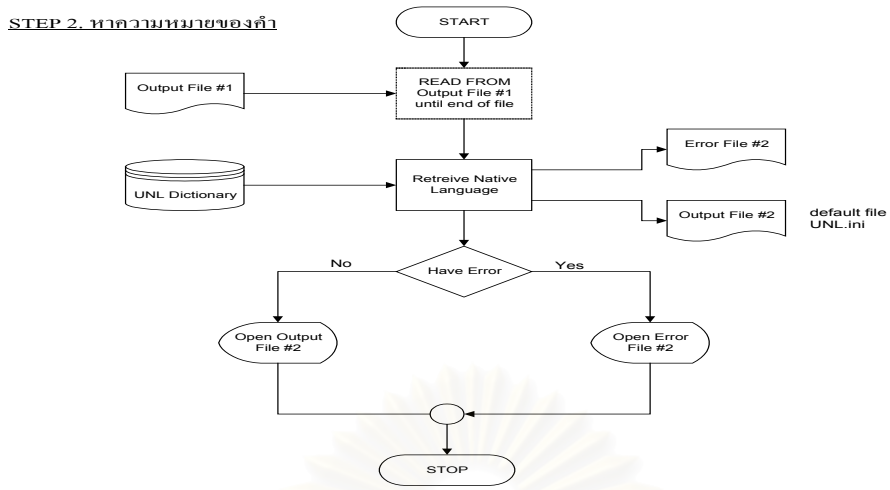
รูปที่ 4.7 ตัวอย่างของคำและความหมายภาษา UNL ที่แสดงในบรรทัด

1.1 ตรวจสอบความถูกต้อง : ในส่วนนี้โปรแกรมจะทำหน้าที่ตรวจสอบข้อมูลที่เป็นการเลือกแฟ้มข้อมูล ซึ่งได้แก่ แฟ้มข้อมูลในรูปแบบ HTML โดยทำการอ่านข้อมูลทีละบรรทัดแบบเรียงลำดับจากนั้นทำการตรวจสอบค่าจนถึงแท็ก (tag) ที่เป็นค่า [W] ซึ่งเป็นการบอกเริ่มต้นคำของภาษา UNL จากนั้นจะอ่านข้อมูลได้บรรทัดดังรูป 4.7 เมื่อตัดคำแรกของบรรทัดนั้นโดยถือเอาเครื่อง หมายวงเล็บเปิด '(' เป็นหลักโดยก่อนหน้านั้นถือเป็นคำสากลเมื่อได้คำสากลแล้วก็จะทำการค้นหาข้อมูลในตารางคำสากล ในกรณีที่ไม่พบข้อมูลในฐานข้อมูลก็จะแสดงเครื่องหมายผิดพลาด (Error) ไว้ จากนั้นจะทำการจัดเตรียมคำศัพท์ของภาษา UNL และนำเอาค่าที่ได้ตั้งแต่คำเชื่อมโยง, คำสากลที่แสดงลักษณะเฉพาะของคำ, คำสากลที่เป็นคำร่วมซึ่งเป็นคู่เชื่อมโยงโดยทุกค่าจะนำไปค้นหาในฐานข้อมูล UNL เพื่อที่จะนำข้อมูลไปใช้ในขั้นตอนที่ 2. ดังแสดงไว้ในรูปที่ 4.8

รูปที่ 4.8 แสดงขั้นตอนการทำงานของโมดูลตรวจสอบความถูกต้อง



1.2. หาความหมายของคำ : หลังจากตรวจสอบค่าความถูกต้องของคำศัพท์ต่างๆ และมีการเพิ่มหรือแก้ไขข้อมูลแล้วในกรณีที่ไม่ถูกต้อง จากนั้นโปรแกรมในส่วนนี้จะทำหน้าที่ดึงเอาคำศัพท์ต่างๆที่ได้จัดเตรียมไว้ในขั้นตอนที่ 1 ไปค้นหาในฐานข้อมูลพจนานุกรมโดยข้อมูลที่ได้จะประกอบด้วยข้อมูลของคำในภาษาไทย และคุณสมบัติทางด้านไวยากรณ์ของคำนั้นๆ ได้แก่ คุณสมบัติทางด้านประเภทของคำ ได้แก่การเป็น นาม, สรรพนาม, กริยา, คุณศัพท์, กริยาวิเศษณ์, บุพบท, สันธาน เป็นต้นดังแสดงไว้ในรูปที่ 4.9 พร้อมทั้งแสดงข้อความผิดพลาดในกรณีที่ไม่พบข้อมูลในฐานข้อมูลพจนานุกรม



รูปที่ 4.9 ขั้นตอนการทำงานของโมดูลหาความหมายของคำ

2. การสร้างโครงสร้างประโยคทางไวยากรณ์ของระบบการสร้างภาษาไทยจากภาษากลาง

ในส่วนนี้ทำการสร้างโครงสร้างทางไวยากรณ์ของภาษาไทยให้กับภาษากลาง โดยทำการกำหนดหน้าที่ของ GFUNC ของส่วนแทนความหมายของคำแต่ละตัวที่เชื่อมต่อจากกิริยาหลักตามหลักไวยากรณ์ไทย และใส่ค่านั้น ๆ ไว้ที่การกที่แสดงความสัมพันธ์ระหว่างกิริยาหลักกับส่วนแสดงความหมายของคำตัวนั้น ๆ ได้แก่

1. GFUNC. (SUB) สำหรับส่วนแทนความหมายของคำที่เป็นประธาน (Subject)
2. GFUNC. (DOB) สำหรับส่วนแทนความหมายของคำที่เป็นกรรมตรง (Direct Object)
3. GFUNC. (IOB) สำหรับส่วนแทนความหมายของคำที่เป็นกรรมรอง (Indirect Object)
4. GFUNC. (COMP) สำหรับส่วนแทนความหมายของคำที่เป็นส่วนขยายประโยคที่ขาดไม่ ได้
5. GFUNC. (COMPF) สำหรับส่วนแทนความหมายของคำที่เป็นส่วนขยายประโยคทั่วไป

จากขั้นตอนดังกล่าวนำมาปรับเป็น 1 ขั้นตอนได้แก่ 2.1 เปรียบเทียบรูปแบบกริยาโดยอ้างถึงตัวอย่างคำศัพท์ของ UNL ตามรูป 4.7 เพื่ออธิบายขั้นตอนการทำงานของโปรแกรมดังต่อไปนี้

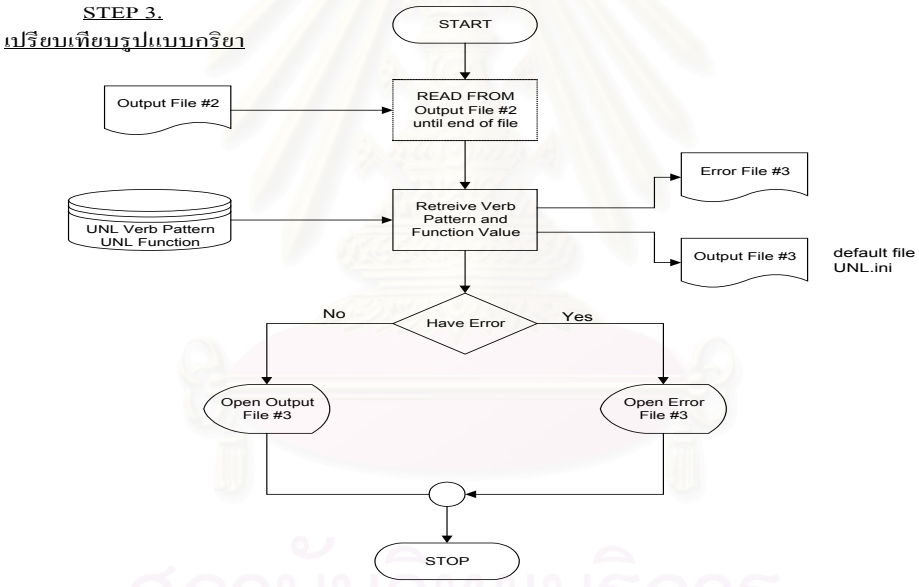
เปรียบเทียบรูปแบบกริยา : เป็นขั้นตอนที่ใช้ในการเปรียบเทียบรูปแบบกริยาจากข้อมูลที่ได้รับจาก 2 ขั้นตอนข้างต้นในส่วนนี้โปรแกรมจะตรวจสอบโดยเริ่มจากบรรทัดที่มีแท็ก [R] เป็นการเริ่มต้นของการบอกความสัมพันธ์ของคำศัพท์ภาษา UNL ดังแสดงไว้ในรูปที่ 4.10 ในระหว่างบรรทัดที่อยู่ในส่วนแรกของประโยค โดยใช้ตัวเลขลำดับที่เป็นตัวบอกคำศัพท์และนำไปใช้ในการหาข้อมูลว่าในคำศัพท์นั้นมีส่วนที่เป็นคู่เชื่อมโยง

ประกอบด้วยอะไรบ้างและใช้เป็นส่วนประกอบสำหรับไปค้นหาเพื่อเปรียบเทียบรูปแบบกริยากับในฐานข้อมูล เพื่อหาผลลัพธ์ของส่วนที่เปรียบเทียบเรียบร้อยแล้วโดยจะได้ออกมาในรูปแบบของรหัสหน้าที่คำ และคำอธิบายความหมายของการเชื่อมประโยคดังแสดงไว้ในรูปที่ 4.11 และหลังจากได้รหัสหน้าที่คำแล้วก็จะนำข้อมูลไปค้นหาในตารางหน้าที่คำเพื่อให้ได้ค่าของหน้าที่คำดังแสดงไว้ในตารางที่4.1 เพื่อจะนำไปใช้ในขั้นตอนต่อไป

[R]	*จากตัวอย่างหลังจากที่ Mapping Verb Pattern เรียบร้อยแล้วจะได้ค่า*	
00agt01	*Function Value ดังต่อไปนี้ :*	*pred (4000) agt sub (3000)*
00obj02		*pred (4000) obj dob (5000)*
[/R]		

รูปที่ 4.10 ตัวอย่างของความสัมพันธ์ระหว่างคำในภาษา UNL

ขั้นตอนนี้จะทำการอ่านแฟ้มข้อมูลเอกสารจากขั้นตอนที่ 2 เพื่อนำค่าต่างๆ มาใช้ในการเปรียบเทียบรูปแบบกริยากับฐานข้อมูล UNL แล้วทำการดึงหน้าที่คำและค่าของหน้าที่คำเพื่อใช้ในขั้นตอนถัดไป ดังแสดงด้วยรูปดังต่อไปนี้



รูปที่ 4.11 แสดงขั้นตอนการทำงานของโมดูลเปรียบเทียบรูปแบบกริยา

ค่าต่างๆ ของหน้าที่ของคำและรายละเอียดเกี่ยวกับหน้าที่คำที่นำไปใช้ในการเรียงคำเพื่อให้ได้รูปประโยคที่ถูกต้องตามหลักภาษาไทย แสดงได้ดังตารางต่อไปนี้

ตารางที่ 4.5 แสดงรหัสหน้าที่คำและค่าของแต่ละหน้าที่คำ

รหัสหน้าที่คำ	ชื่อ	คำอธิบายภาษาไทย	ค่าหน้าที่คำ
ADJ	Adjective	เช่น สวย ใหญ่	22000
CAD	Cardinal Number	เช่น 1, 2, สาม, ห้า, ร้อย	17000
CLAS	Classifier	เช่น คัน, ตัว, เล่ม, ด้าม	19000
COMP	Complement	เป็นส่วนที่อยู่หลัง IOB	7000
COMPF	Complement	ที่บอกเวลา, บอกเหตุผลล	8000
DEM	Memonsative	เช่น นี่, นั่น	21000
DOB	Direct Object	กรรมตรงของประโยค ได้แก่ N, PRON	5000
INT	Interogative marker	เช่น หรือไม่, หรือ, อะไร	9000
IOB	Indirect Object	กรรมรองของประโยค ได้แก่ N, PRON	6000
N	Noun	เช่น ไก่, ม้า	10000
ORD	Ordinal Number	เช่น แรก, คันแรก	20000
PRED	Predicate CP	ได้แก่ กริยาหลัก, คำคุณศัพท์ที่ทำหน้าที่เป็นกริยา	4000
PREP	Proposition	เช่น ที่, บน, ใน	11000
PRON	Pronoun	เช่น เขา, หล่อน, ฉัน	15000
QUANT	Quantitative	เช่น หลาย, บาง	18000
SADV	Sentencial Adverb	ส่วนขยายความประโยค เช่น โดยทั่วไป, ที่จริง,	1000
SMOD	Sentencial Modefier	ส่วนขยายประโยค เช่น เมื่อเร็วๆ นี้, กาลครั้งหนึ่ง	2000
SUB	Subject	ประธานของประโยค	3000
V	Verb	เช่น กิน, เดิน, นั่ง	13000

3. การเรียงลำดับของคำภาษาไทยของระบบการสร้างภาษาไทยจากภาษากลาง

การเรียงลำดับของคำภาษาไทยที่อยู่บนโครงสร้างไวยากรณ์นั้น ก็คือการจัดลำดับของโนด ส่วนแทนความหมายของคำและโนดการก เพื่อที่จะนำมาเรียงกันให้เป็นประโยคในแนวเส้นตรงทั้งนี้โดยอาศัยข้อมูลที่กำหนด ตำแหน่งของคำที่อยู่ข้างเคียงกันตามหลักโครงสร้าง (Syntax) ของภาษาไทยมาเป็นตัวตัดสินดังนั้นจึงต้องมีการกำหนดกฎการเรียงลำดับคำหรือกลุ่มคำ โดยแบ่งตามประเภทของคำ และหน้าที่ของคำนั้น ๆ และใช้กฎดังกล่าวเป็นฐานข้อมูลในการเปรียบเทียบอ้างอิง ของการกำหนดลำดับของโนดส่วนแสดงความหมายของคำ และโนดการกที่อยู่ในโครงสร้างต้นไม้ เพราะฉะนั้นการทำงานในส่วนนี้จะประกอบไปด้วยส่วนสำคัญ 2 ส่วน คือ กฎทางลอจิกเพื่อกำหนดลำดับของโนดที่อยู่ข้างเคียงกันบนโครงสร้างต้นไม้ ในที่นี้คือโครงสร้างภาษากลาง และส่วนที่เป็นฐานข้อมูลของกฎไวยากรณ์การเรียงลำดับคำและกลุ่มคำในภาษาไทย

จากขั้นตอนดังกล่าวนำมาปรับเป็น3 ขั้นตอนได้แก่ 3.1 จัดเรียงคำภาษาไทย, 3.2 กำหนดคำภาษาไทย , 3.3 แสดงผลประโยคภาษาไทย โดยอ้างถึงตัวอย่างคำศัพท์ของภาษา UNL ตามรูป 4.7 เพื่ออธิบายขั้นตอนการทำงานของโปรแกรมดังต่อไปนี้

3.1 จัดเรียงคำภาษาไทย : เป็นขั้นตอนที่ทำหน้าที่ในการนำค่าตามหน้าที่ของคำศัพท์ ใน ตารางที่ 4.5 มาจัดเรียงค่าจากน้อยไปหามากตามค่าของหน้าที่ของคำศัพท์ ซึ่งแต่ละหน้าที่ของคำจะมีค่าหรือค่าหน้าที่คำของแต่ละคำศัพท์ไม่เท่ากันโดยมีขั้นตอนการทำงานดังแสดงไว้ในรูปที่ 4.14 และโปรแกรมในส่วนนี้จะทำการเรียงค่าของคำเป็นคู่ๆ จากตัวอย่างในรูปที่ 4.10 จะทำการสลับค่าดังนี้

00 agt 01	*ผลลัพธ์จากการเรียงค่า*	*sub (3000) agt pred (4000)*
00 obj 02	*"*	*pred (4000) obj dob (5000)*

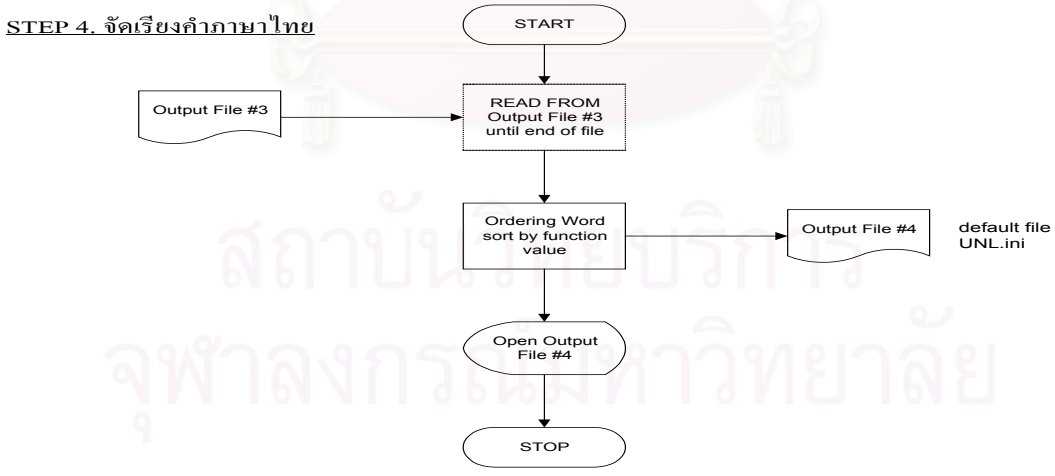
รูปที่ 4.12 ตัวอย่างผลลัพธ์หลังจากจัดเรียงตามค่าของหน้าที่คำ

หลังจากที่ได้ทำงานผ่านโปรแกรมในขั้นตอนนี้แล้วจะมีข้อมูลที่พร้อมจะทำการขั้นตอนการตัดคำซ้ำ ในขั้นตอนต่อไปโดยมีผลลัพธ์ของข้อมูลดังนี้

ลิง (noun) agt กิน (vi)	*sub (3000) agt pred (4000)*
กิน (vi) obj กล้วย (noun)	*pred (4000) obj dob (5000)*

รูปที่ 4.13 ตัวอย่างผลลัพธ์หลังจากขั้นตอนจัดเรียงคำ

ขั้นตอนนี้จะทำการอ่านแฟ้มข้อมูลเอกสารจากขั้นตอนที่ 3 เพื่อนำค่าหน้าที่คำมาใช้ในการจัดเรียงคำให้ถูกต้องตามหลักภาษาไทยและเพื่อใช้ในขั้นตอนถัดไป ดังแสดงด้วยรูปดังต่อไปนี้



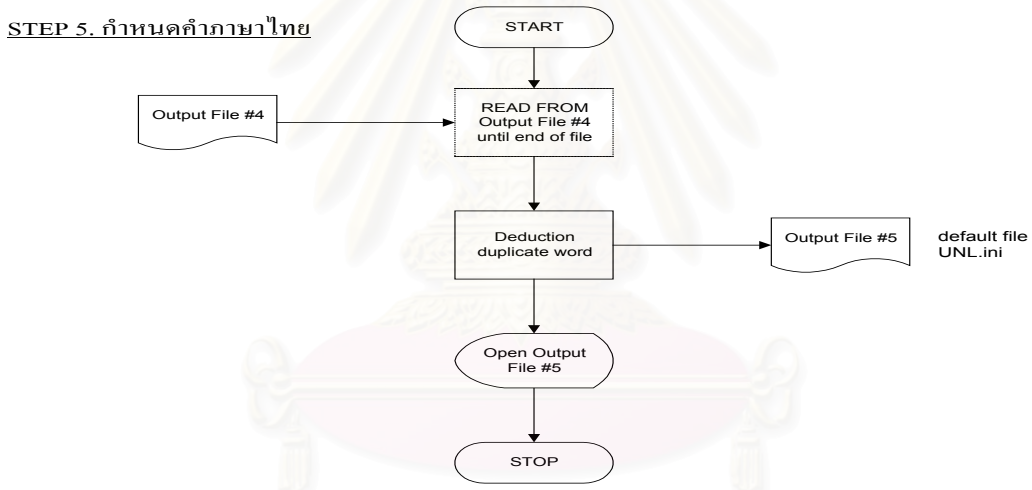
รูปที่ 4.14 แสดงขั้นตอนการทำงานของโมดูลจัดเรียงคำ

3.2 กำหนดคำภาษาไทย : ในขั้นตอนนี้จะทำหน้าที่ในการตรวจสอบคำศัพท์ที่ซ้ำซ้อนกันหลังจากที่ได้มีการจัดเรียงค่าให้ถูกต้องตามค่าหน้าที่คำแล้วโดยพิจารณาจากคำศัพท์ที่อยู่ต่อเนื่องกันจะตัดคำที่ซ้ำซ้อนในคำหลังออกเพื่อให้มีคำศัพท์เพียงคำเดียวดังแสดงขั้นตอนไว้ในรูปที่ 4.16 และจากตัวอย่างผลลัพธ์ในรูปที่ 4.13 หลังจากผ่านโปรแกรมในขั้นตอนนี้แล้วจะมีคำศัพท์ "กิน" (pred, 4000) ที่ซ้ำซ้อนกันโดยโปรแกรมจะตัดคำหลังออกดังแสดงผลลัพธ์ไว้ในรูปที่ 4.15 ดังนี้

ลิง (noun) agt กิน (vi)	*sub (3000) agt pred (4000)*
obj กล้วย (noun)	*obj dob (5000)*

รูปที่ 4.15 ตัวอย่างผลลัพธ์หลังจากขั้นตอนกำหนดคำภาษาไทย

ขั้นตอนนี้จะทำการอ่านแฟ้มข้อมูลเอกสารจากขั้นตอนที่ 4 หลังจากจัดเรียงคำต่างๆเรียบร้อยแล้ว มาใช้ในการดูคำซ้ำซ้อนเพื่อตัดคำซ้ำซ้อนออกไป แล้วจัดเตรียมคำให้ถูกต้องเพื่อใช้ในขั้นตอนถัดไป ดังแสดงด้วยรูปดังต่อไปนี้



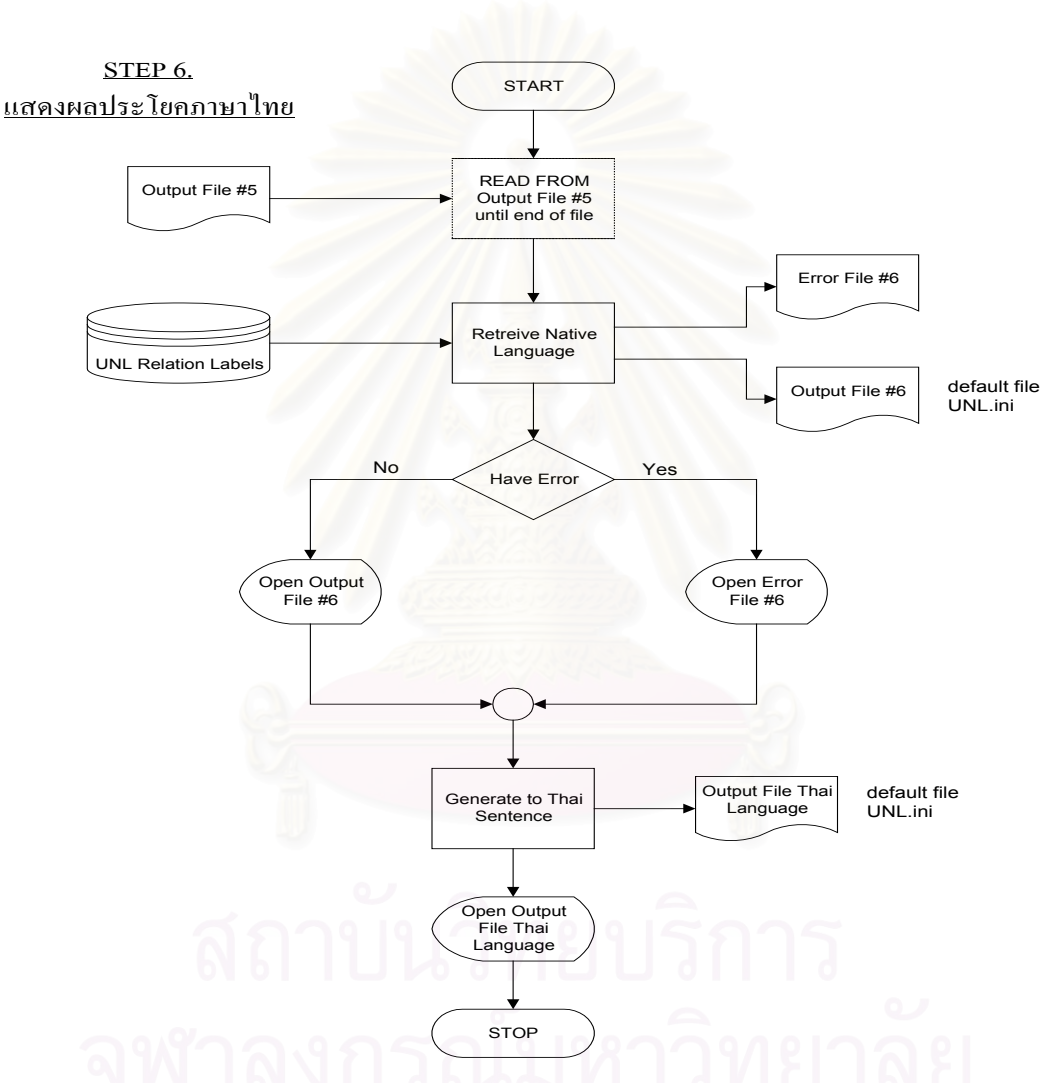
รูปที่ 4.16 แสดงขั้นตอนการทำงานของกำหนดโมดูลคำภาษาไทย

3.3 แสดงผลประโยคภาษาไทย : ขั้นตอนสุดท้ายนี้ออกแบบสำหรับทำการสร้างรูปแบบประโยคภาษาไทยเพื่อแสดงผลลัพธ์การแปลงจากภาษา UNL เป็นภาษาไทย ในขั้นตอนนี้จะนำเอาคำศัพท์ที่ได้จากขั้นตอนที่ 5 มาจัดเรียงต่อเนื่องกันก่อนในบรรทัดเดียวกัน หลังจากนั้นจึงทำการเชื่อมคำให้อยู่ในรูปแบบของประโยคภาษาไทยโดยมีขั้นตอนการทำงานดังแสดงไว้ในรูปที่ 4.18 และผลลัพธ์สุดท้ายของการแปลงจากภาษา UNL เป็นภาษาไทยดังแสดงไว้ในรูปที่ 4.17

ลิง กิน กล้วย ลิงกินกล้วย

รูปที่ 4.17 ตัวอย่างผลลัพธ์สุดท้ายของการแปลงจากภาษา UNL เป็นภาษาไทย

ขั้นตอนนี้จะทำการอ่านแฟ้มข้อมูลเอกสารจากขั้นตอนที่ 5 หลังจากตัดคำซ้ำซ้อนเรียบร้อยแล้ว ในขั้นตอนนี้จะเป็นเตรียมรูปประโยคและแสดงผลลัพธ์ในรูปประโยคภาษาไทย ดังแสดงด้วยรูปดังต่อไปนี้



รูปที่ 4.18 แสดงขั้นตอนการทำงานของโมดูลแสดงผลประโยคภาษาไทย

บทที่ 5
การทดสอบโปรแกรม CU-Deconverter

5.1 วิธีการทดสอบโปรแกรม CU-Deconverter

ผู้วิจัยได้ทำการทดสอบโปรแกรมที่ได้ทำขึ้นโดยใช้แฟ้มข้อมูลภาษา UNL ที่อยู่ในรูปแบบแฟ้มข้อมูลของ HTML โดยใช้ตัวอย่างจากในหนังสือ Universal Networking Language ของสถาบัน The United Nations University Institute of Advanced Studies จำนวน 70 ประโยคโดยในส่วนของฐานข้อมูล UNL ที่ได้ออกแบบไว้บรรจุคำศัพท์คำสากลประมาณ 2,300 คำรวมทั้งข้อมูลพจนานุกรม UNL ซึ่งบรรจุคำศัพท์ที่ใช้ในการแปลงประโยคตัวอย่าง อีกทั้งในส่วนของคำที่เป็นข้อมูลพื้นฐานต่างๆได้ใส่ข้อมูลไว้ใช้สำหรับการแปลงประโยคตัวอย่างในส่วนข้อมูลเบื้องต้นของ ภาษา UNL ได้ใส่ข้อมูลไว้ครบถ้วนตามกฎเกณฑ์ของภาษา UNL ด้วย

ขั้นตอนในการทดสอบจะเป็นการเรียกโปรแกรม CU-Deconverter และทำการเปิดแฟ้มข้อมูลตัวอย่างทีละแฟ้มข้อมูล จากนั้นทำการทดสอบทีละขั้นตอนซึ่งมีทั้งหมด 6 ขั้นตอนโดยในแต่ละขั้นตอนที่มีข้อความแสดงความผิดพลาดก็จะทำการแก้ไขเพิ่มเติมฐานข้อมูล UNL ให้ครบถ้วน จนกระทั่งได้ผลลัพธ์สุดท้ายเป็นประโยคภาษาไทย

สำหรับความถูกต้องในการแปลงจากภาษา UNL เป็นภาษาไทยนั้นได้ใช้วิธีการตรวจสอบเปรียบเทียบกับต้นฉบับที่เป็นประโยคภาษาอังกฤษ โดยในแฟ้มข้อมูลของ UNL ที่อยู่ในรูปแบบ HTML นั้นจะมีประโยคต้นฉบับเป็นภาษาอังกฤษกำกับไว้ เมื่อได้ผลลัพธ์จากการแปลงภาษา UNL เป็นภาษาไทยแล้วก็จะนำผลลัพธ์ภาษาไทยนั้นไปเปรียบเทียบกับต้นฉบับที่เป็นภาษาอังกฤษเพื่อสอบทานความถูกต้อง

5.2 ผลการทดสอบ

จากการทดสอบโปรแกรมในหัวข้อ 5.1 เพื่อหาความถูกต้องในการแปลงจากประโยคของ ภาษา UNL เป็นภาษาไทยนั้น จะได้ผลลัพธ์ซึ่งประกอบด้วยส่วนต่างๆ 3 ส่วนดังนี้คือ

1. แสดงตัวอย่างผลการทดสอบขั้นตอนการแปลงจากภาษา UNL เป็นภาษาไทยทีละขั้นตอน โดยใช้ตัวอย่างประโยคของภาษา UNL อธิบายประกอบ 4 ประโยคตามตัวอย่าง 5.2.1

2. แสดงตัวอย่างผลลัพธ์ข้อความผิดพลาดของการทำงานที่ละขั้นตอนตามตัวอย่างข้อ 5.2.1.1 – 5.2.1.4
3. แสดงผลลัพธ์สุดท้ายของการแปลงด้วยโปรแกรม CU-Deconverter จากภาษา UNL เป็นภาษาไทยจำนวน 70 ประโยคตามที่ได้แสดงไว้ในตารางที่ 5.1

5.2.1 ตัวอย่างผลการทดสอบขั้นตอนการแปลงจากภาษา UNL เป็นภาษาไทยที่ละขั้นตอน

5.2.1.1 ตัวอย่างแสดงการแปลจากภาษา UNL จากประโยค Monkey eats bananas.

UNL Input File HTML Format	1. Pre-Edit
S]	S]
[W]	[W]
eat(icl>action).@present.@entry:00	eat,icl,action,present,entry --> 00
monkey(icl>animal).@generic:01	monkey,icl,animal,generic --> 01
banana(icl>food).@generic:02	banana,icl,food,generic --> 02
[/W]	[/W]
[R]	[R]
00agt01	00agt01
00obj02	00obj02
[/R]	[/R]
[/S]	
Monkey eats bananas.	
monkey <- agt - eat - obj -> banana	
2. Scan UWs From Dictionary	5. Morphological
00 agt 01 กิน (Verb) agt ลิง (Noun)	sub agt pred ลิง (01) agt กิน (00)
00 obj 02 กิน (Verb) obj กล้วย (Noun)	n obj v กล้วย (02) obj
3. Get Verb Pattern	6. Thai Language
กิน (Verb) agt ลิง (Noun) PRED (4000) agt SUB (3000)	ลิง กิน
	กล้วย
กิน (Verb) obj กล้วย (Noun)	
V (13000) obj N (10000)	Thai Language Sentence
	----------
4. Word Ordering	ลิงกินกล้วย
sub (3000) agt pred (4000)	
n (10000) obj v (13000)	

5.2.1.2 ตัวอย่างแสดงการแปลจากภาษา UNL จากประโยค The airplane is about to land.

UNL Input File HTML Format	1. Pre-Edit
UNL	UNL
[S]	[S]
[W]	[W]
land(icl>airplane).@begin-soon.@entry:00	land,icl,airplane,begin-soon,entry --> 00
airplane(icl>finished).@generic:01	airplane,icl,finished,generic --> 01
[/W]	[/W]

[R]	[R]
00agt01	00agt01
[/R]	[/R]
[/S]	
The airplane is about to land.	
airplane <- agt - land.@begin-soon	
2. Scan UWs From Dictionary	5. Morphological
00 agt 01 ลงจอด (Verb) agt เครื่องบิน (Noun)	sub agt pred
	เครื่องบิน (01) agt ลงจอด (00)
3. Get Verb Pattern	6. Thai Language
ลงจอด (Verb) agt เครื่องบิน (Noun)	เครื่องบิน กำลังจะ ลงจอด
PRED (4000) agt SUB (3000)	
	Thai Language Sentence
4. Word Ordering	----------
sub (3000) agt pred (4000)	เครื่องบินกำลังจะลงจอด

5.2.1.3 ตัวอย่างแสดงการแปลจากภาษา UNL จากประโยค

to go to a library and borrow books.

UNL Input File HTML Format	1. Pre-Edit
UNL	UNL
[S]	[S]
[W]	[W]
borrow(icl>action).@present.@entry:00	borrow,icl,action,present,entry --> 00
book(icl>publications).@generic:01	book,icl,publications,generic --> 01
go(icl>action).@present:02	go,icl,action,present --> 02
library(icl>man-made-place).@generic:03	library,icl,man-made-place,generic --> 03
[/W]	[/W]
[R]	[R]
02gol03	02gol03
00seq02	00seq02
00obj01	00obj01
[/R]	[/R]
[/S]	
to go to a library and borrow books	
book <- obj - borrow - seq -> go - gol -> library	
2. Scan UWs From Dictionary	5. Morphological
02 gol 03 ไป (Verb) gol ห้องสมุด (Noun)	pred gol dob ไป (02) gol ห้องสมุด (03)
00 seq 02 ยืม (Verb) seq ไป (Verb)	sadv seq pred seq ยืม (00)
00 obj 01 ยืม (Verb) obj หนังสือ (Noun)	pred obj dob obj หนังสือ (01)
3. Get Verb Pattern	6. Thai Language
ไป (Verb) gol ห้องสมุด (Noun) PRED (4000) gol DOB (5000)	ไป ห้องสมุด
ยืม (Verb) seq ไป (Verb) PRED (4000) seq SADV (1000)	และ ยืม
ยืม (Verb) obj หนังสือ (Noun) PRED (4000) obj DOB (5000)	หนังสือ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. Word Ordering	Thai Language Sentence
pred (4000) gol dob (5000)	----------
sadv (1000) seq pred (4000)	ไปห้องสมุดและยืมหนังสือ
pred (4000) obj dob (5000)	

5.2.1.4 ตัวอย่างแสดงการแปลจากภาษา UNL จากประโยค UNL is a common language that would be used for network communications.

UNL Input File HTML Format	1. Pre-Edit
UNL	UNL
[S]	[S]
[W]	[W]
language(icl>expression).@present.@entry.@pred:00	language,icl,expression,present,entry,pred --> 00
UNL(icl>label).@topic:01	UNL,icl,label,topic --> 01 | *Error Line* |
common(mod>share):02	common,mod,share --> 02
use(icl>action).@present.@pred:03	use,icl,action,present,pred --> 03
language(icl>expression).@present.@entry.@pred:04	language,icl,expression,present,entry,pred --> 04
communication(icl>information-transfer).@pl:05	communication,icl,information-transfer,pl --> 05
network(mod>computer):06	network,mod,computer --> 06
[/W]	[/W]
[R]	[R]
00aoj01	00aoj01
00mod02	00mod02
03obj04	03obj04
03pur05	03pur05
05mod06	05mod06
[/R]	[/R]
[/S]	
UNL is a common language that would be used for network	
communications.	
UNL <- aoj - language - mod -> common	
language <- obj - use - pur -> communication - mod -> network	
2. Scan UWs From Dictionary	5. Morphological
00 aoj 01 ภาษา (Noun) aoj UNL (topic)	sub aoj pred UNL (01) aoj ภาษา (00)
00 mod 02 ภาษา (Noun) mod กลาง (Adverb)	pred mod adj mod กลาง (02)
03 obj 04 ใช้ (Verb) obj ภาษา (Noun)	pred obj dob ใช้ (03) obj
03 pur 05 ใช้ (Verb) pur ติดต่อ (Noun)	pred pur iob pur ติดต่อ (05)
05 mod 06 ติดต่อ (Noun) mod เครือข่าย คอมพิวเตอร์ (Noun)	iob mod comp mod เครือข่ายคอมพิวเตอร์ (06)
	6. Thai Language
3. Get Verb Pattern	UNL เป็น ภาษา
ภาษา (Noun) aoj UNL (topic) PRED (4000) aoj SUB (3000)	กลาง
ภาษา (Noun) mod กลาง (Adverb) PRED (4000) mod ADJ (22000)	ใช้
ใช้ (Verb) obj ภาษา (Noun) PRED (4000) obj DOB (5000)	เพื่อ ติดต่อ
ใช้ (Verb) pur ติดต่อ (Noun) PRED (4000) pur IOB (6000)	ด้วย เครือข่ายคอมพิวเตอร์
ติดต่อ (Noun) mod เครือข่ายคอมพิวเตอร์ (Noun)	
IOB (6000) mod COMP (7000)	Thai Language Sentence

	----------
4. Word Ordering	UNLเป็นภาษากลางใช้เพื่อติดต่อด้วยเครือข่ายคอมพิวเตอร์
sub (3000) aoj pred (4000)	
pred (4000) mod adj (22000)	
pred (4000) obj dob (5000)	
pred (4000) pur iob (6000)	
iob (6000) mod comp (7000)	

จากผลการทดสอบโปรแกรมดังแสดงไว้ข้างต้นอธิบายถึงขั้นตอนการทดสอบทีละขั้นตอนซึ่งเป็นขั้นตอนที่แสดงไว้หลังจากได้ผ่านการแก้ไขเพิ่มเติมฐานข้อมูล UNL ไว้เรียบร้อยแล้ว ส่วนในกรณีที่แต่ละขั้นตอนมีการแสดงข้อความผิดพลาดจะแสดงไว้ในตัวอย่าง 5.2.2 ดังต่อไปนี้

5.2.2 ตัวอย่างผลลัพธ์ข้อความผิดพลาดของการทำงานทีละขั้นตอน

5.2.2.1. ตัวอย่างข้อความผิดพลาดในขั้นตอนตรวจสอบความถูกต้อง

5.2.2.1.1 ตัวอย่างข้อความผิดพลาด “ไม่พบคำสากลในฐานข้อมูล UNL”

1. language(icl>expression).@present.@entry.@pred:00 --> language Not found In Universal Word
2. UNL(icl>label).@topic:01 --> UNL Not found In Universal Word
3. common(mod>share):02 → common Not found In Universal Word
4. use(icl>action).@present.@pred:03 --> use Not found In Universal Word

5.2.2.1.2 ตัวอย่างข้อความผิดพลาด “ไม่พบข้อมูลในพจนานุกรม UNL”

1. language(icl>expression).@present.@entry.@pred:00 --> Not found In UNL Dictionary
2. UNL(icl>label).@topic:01 --> Not found In UNL Dictionary
3. common(mod>share):02 --> Not found In UNL Dictionary
4. use(icl>action).@present.@pred:03 --> Not found In UNL Dictionary

5.2.2.1.3 ตัวอย่างข้อความผิดพลาด “รายละเอียดที่อยู่ในเอกสาร UNL ที่ไม่พบในพจนานุกรม UNL”

1. UW_Keyword = 'language' And RL_Code = 'icl' And Object_Class = 'expression' And Subject_Object_Class = ‘expression’
2. UW_Keyword = 'UNL' And RL_Code = 'icl' And Object_Class = 'label' And Subject_Object_Class = 'label'
3. UW_Keyword = 'common' And RL_Code = 'mod' And Object_Class = 'share' And Subject_Object_Class = 'share'

4. 4. UW_Keyword = 'use' And RL_Code = 'icl' And Object_Class = 'action' And Subject_Object_Class = 'action'

5.2.2.2. ตัวอย่างข้อความผิดพลาดในขั้นตอนเปรียบเทียบรูปแบบกริยา

ตัวอย่างที่ 1. 00 aoj 01 language (present) aoj UNL (topic) Category1 = 'present' And AL_Code01 = 'present' And AL_Code02 = 'entry' And AL_Code03 = 'pred' And RL_Code = 'aoj' And Category2 = 'topic' And AL_Code06 = 'topic' ตัวอย่างที่ 2. 00 mod 02 language (present) mod common () Category1 = 'present' And AL_Code01 = 'present' And AL_Code02 = 'entry' And AL_Code03 = 'pred' And RL_Code = 'mod' ตัวอย่างที่ 3. 03 obj 04 use (present) obj language (present) Category1 = 'present' And AL_Code01 = 'present' And AL_Code02 = 'pred' And RL_Code = 'obj' And Category2 = 'present' And AL_Code06 = 'present' And AL_Code07 = 'entry' And AL_Code08 = 'pred'

5.2.3 ผลการทดสอบการแปลงจากภาษา UNL เป็นภาษาไทยด้วยโปรแกรม CU-Deconverter

ตารางที่ 5.1 ผลการทดสอบแปลงจากภาษา UNL เป็นภาษาไทยจำนวน 70 ประโยค

ประโยคภาษาอังกฤษ	ประโยคภาษา UNL	แปลงเป็นประโยคภาษาไทย
2 days a week	(2 <- num - day) - per -> week	2 วันต่อสัปดาห์
3kg of apples	apple - qua -> kg -> num -> 3	แอปเปิลหนัก3 กิโลกรัม
The airplane is about to land.	airplane <- agt - land.@begin-soon	เครื่องบินกำลังจะลงจอด
to go to a library and borrow books	book <- obj - borrow - seq -> go - gol -> library	ไปห้องสมุดและยืมหนังสือ
Cat and rats are animals.	(cat <- and - rat) <- aoj - animal	หนูและแมวเป็นสัตว์
A cat is bigger than rat.	cat <- aoj - big - bas -> rat	แมวใหญ่กว่าหนู
A child is jumping.	child <- agt - jump.@repeat	เด็กกำลังกระโดด
Children are dying of hunger.	child.@pl <- obj - die.@progress - cau -> hunger	เด็กๆกำลังจะตายเพราะความหิว
The sea is calm.	sea <- soj - calm	ทะเลสงบ
to come from London.	come - src -> London	มาจากกรุงลอนดอน
to come to see me	come - pur ->see - obj -> I	มาเพื่อพบฉัน

ประโยคภาษาอังกฤษ	ประโยคภาษา UNL	แปลงเป็นประโยคภาษาไทย
to compete with a friend	compete - ptn -> friend	แข่งขันกับเพื่อน
Most of these problems a computer could solve easily.	solve - obj -> problem.@topic - mod -> most	คอมพิวเตอร์แก้ปัญหาโดยมากได้ง่าย
to cure by surgery	cure - met -> surgery	รักษาให้หายโดยศัลยแพทย์
to cut with knife	cut - ins -> knife	เฉือนด้วยมีด
to die of cancer	die - cau -> cancer	ตายเพราะโรคมะเร็ง
To eat an apple	eat - obj -> apple	กินผลแอปเปิล
A girl is crying	girl <- agt - cry.@progress	เด็กหญิงกำลังร้องไห้
to go to London.	go - gol -> London	ไปกรุงลอนดอน
Leaves are green.	leave <- aoj - green	ใบไม้เป็นสีเขียว
He did it.	he <- agt - do.@past - obj -> it	เขาเคยทำเหตุการณ์
I know he goes.	I <- agt - know.@pred - obj -> go.@pred - agt -> he	ฉันรู้ว่าเขาไป
His face I despise.	I <- agt - despise - obj -> face.@emphasis	ฉันชังหน้าของเขา
I bought a knife.	I <- agt - buy - obj -> knife.@indef	ฉันซื้อมีด
I bought knives.	I <- agt - buy - obj -> knife.@pl	ฉันซื้อมีดหลายเล่ม
I do not go.	I <- agt - go.@not	ฉันไม่ไป
I have three.	I <- agt - have - obj -> - num -> 3	ฉันมี3
I like apples, bananas and oranges.	I <- agt - like.@pred.@entry - obj -> orange - and -> banana - and -> apple	ฉันชอบส้มและกล้วยและแอปเปิล
I like apples, bananas and oranges.	I <- agt - like.@pred.@entry - obj -> (orange.@sub - and -> banana - and -> apple)	ส้มและกล้วยและแอปเปิลฉันชอบ
John's dog	dog - pos ->John	สุนัขของจอห์น
He has just come home	he <- agt - come.@end-just - gol -> home	เขาเพิ่งมาถึงบ้าน
to keep a cat or dog	keep - obj ->(cat<- or - dog)	เก็บแมวหรือสุนัข
to lunch at noon	lunch - tim -> noon	กินอาหารกลางวันตอนเที่ยงวัน
to make butter from milk	butter <- obj - make - mat -> milk	ทำเนยเหลวด้วยน้ำนม
Snow melts.	snow <- aoj - melt	หิมะละลาย
The monkey fell from a branch.	monkey.@def <- agt - fall - src -> branch	ลิงตกจากกิ่งก้าน
A monkey likes bananas.	monkey.@generic <- agt - like - obj -> banana	ลิงชอบกล้วย
My father is dying.	I <- pos - father <- obj - die.@begin-soon	พ่อของฉันกำลังจะตาย
Monkey eats bananas.	monkey <- agt - eat - obj -> banana	ลิงกินกล้วย
It is not an orange.	It <- aoj - orange.@not	มันไม่ใช่ส้ม
She has played a piano.	she <- agt - play.@complete - obj -> piano	เธอเล่นเปียโนแล้ว
to play in a garden	play - ppl ->garden	เล่นในสวน
The pot has broken.	pot <- obj - break.@state	หม้อแตกแล้ว
A rabbit runs.	rabbit <- agt - run	กระต่ายวิ่ง

ประโยคภาษาอังกฤษ	ประโยคภาษา UNL	แปลงเป็นประโยคภาษาไทย
Representation of sentence structure	representation@title - mod -> structure - mod -> sentence	แสดงโครงสร้างของประโยค
to return home crying	home <- gol - return - coo -> cry	กลับบ้านแล้วกำลังร้องไห้
to return home because of rain	home <- gol - return - con -> rain	กลับบ้านเพราะว่าฝนตก
to run fast	run - man -> fast	วิ่งอย่างเร็ว
She will come at 9.	she <- agt - come.@future - tim -> 9	เธอกำลังจะมาตอน9
She goes to school at 9.	she <- agt - go.@present - gol -> school go.@present - tim -> 9	เธอไปโรงเรียนตอน9
The match has just begun.	match <- obj - begin.@begin-just	การแข่งขันกีฬาเพิ่งจะเริ่มต้น
the third man	man - mod -> third	ผู้ชายที่สาม
three miles	mile - num -> 3	3 ไมล์
a ticket from London to Oxford	ticket - mod -> (London <- fmt - Oxford)	ตั๋วจากกรุงลอนดอนไปออกซฟอร์ด
two pints of beer	beer - qua -> pint - num -> 2	เบียร์หนัก2 ไพนท์
UNL is a common language tthat would be used for network communications.	UNL <- aoj - language - mod -> common language <- obj - use - pur -> communication - mod -> network	UNLเป็นภาษากลางใช้เพื่อติดต่อด้วยเครือข่ายคอมพิวเตอร์
to walk on the road	walk - opl -> road	เดินบนถนน
to walk with a friend	walk - cag -> friend	เดินไปกับเพื่อน
to win in competition	win - lpl ->competition	ชนะในการแข่งขัน
to work for family	work - ben -> family	ทำงานเพื่อครอบครัว
to work from noon	work - tmf -> noon	ทำงานตั้งแต่เที่ยงวัน
to work until noon	work - tmt -> noon	ทำงานจนถึงเที่ยงวัน
Is that you who went?	go.@interrogation -> agt -> you.@focus	คุณหรือเปล่าที่ไป

จากผลการทดสอบที่แสดงไว้ในตารางที่ 5.1 แสดงให้เห็นว่าความถูกต้องของการแปลงจากภาษา UNL เป็นภาษาไทยด้วยโปรแกรม CU-Deconverter นั้นมีความถูกต้องพอสมควร โดยตัวอย่างประโยคที่นำมาใช้ในการทดสอบทั้ง 70 ประโยคนั้นไม่พบปัญหาในการแปลงเป็นภาษาไทยโดยที่รูปประโยคผิดไปจากความหมายต้นฉบับแต่อย่างใด ผู้วิจัยได้พบปัญหาคือในกรณีที่ความหมายของคำศัพท์คำสากล สามารถแปลความหมายในแนวทางเดียวกันได้มากกว่า 1 ความหมาย เช่น คำว่า "Work" ซึ่งแปลความหมายได้ว่า "การทำงาน" แต่สามารถแปลได้หลายความหมายดังนี้ 1. การทำงาน 2. การทำงานด้วยสมอง 3. การทำงานด้วยมือ 4. นวด, ปั้น เป็นต้น ดังนั้นคำศัพท์คำสากลที่นำมาใช้ร่วมในการแปลความหมายสำหรับเก็บไว้ในพจนานุกรม UNL จะต้องระบุเฉพาะเจาะจงให้ชัดเจนมากยิ่งขึ้นเพื่อให้สามารถถ่ายทอดความหมายได้ถูกต้องมากที่สุด และจากผลการทดสอบพบว่าข้อมูลที่อยู่ในรูปแบบภาษา UNL ไม่ได้รองรับสำหรับการแสดงความหมายในรูปแบบประโยคลักษณะสุภาพ และปัญหาที่พบข้อสุดท้ายคือคำชี้เฉพาะ เช่น คำว่า those, a ไม่พบในประโยคภาษา UNL ทำให้เมื่อแปลเป็นภาษาไทยแล้วจะสื่อความหมายได้ไม่ครบถ้วน

บทที่ 6
สรุปผลการวิจัยและข้อเสนอแนะ

สรุปแนวทางที่ใช้ในการวิจัยครั้งนี้จะเป็นการแปลจากภาษา UNL เป็นภาษาไทย ซึ่งผู้วิจัยได้พบว่าปัญหาการแปลจากภาษา UNL เป็นภาษาไทยนั้นได้แก่ ปัญหาในเรื่องการจัดทำฐานข้อมูลพจนานุกรมสำหรับบรรจุคำศัพท์ให้ครบถ้วนทุกคำตามคำสากล, ปัญหาการจัดทำรูปแบบกริยา ให้ครบถ้วนสมบูรณ์, ปัญหาการแปลความหมายไม่ตรงกับคำศัพท์เดิม เช่น "Brown Mixture" แปลเป็น "ยาแก้ไอน้ำดำ" ซึ่งถือเป็นคำเฉพาะ, ปัญหาจากข้อมูลของภาษา UNL ที่ไม่สนับสนุนการแปลที่แสดงข้อความสุภาพ เป็นต้น

ในงานวิจัยชิ้นนี้ได้เน้นเฉพาะในเรื่องของขั้นตอนการแปลงจากภาษา UNL เป็นภาษาไทยโดยใช้แนวความคิดมาจากระบบการสร้างภาษาไทยจากภาษากลางมาดัดแปลงแก้ไขเพิ่มเติม ผู้วิจัยได้ทำการออกแบบฐานข้อมูลพจนานุกรมบรรจุคำศัพท์, คำประกอบต่างๆและใช้โปรแกรมมิ่งแบบโครงสร้าง (Structure Programming) พัฒนาโปรแกรมการแปลภาษา ในส่วนของขั้นตอนการทำงานของโปรแกรมได้จัดทำให้สอดคล้องเข้ากับกฎเกณฑ์ตามมาตราฐานของภาษา UNL โดยจากผลการทดสอบความถูกต้องของการแปลงจากภาษา UNL เป็นภาษาไทยพบความถูกต้องประมาณ 85% จากจำนวน 70 ประโยค

6.1 สรุปผลการวิจัย

ขั้นตอนวิธีการแปลงจากภาษา UNL เป็นภาษาไทยที่ผู้วิจัยได้นำเสนอนี้แบ่งออกเป็น 6 ขั้นตอนเพื่อทำการแก้ไขปัญหาที่พบในการแปลภาษาคือ (1) ปัญหาการแปลความหมายของคำศัพท์ที่คำหนึ่งในภาษาต้นฉบับสามารถแปลความหมายไปยังอีกภาษาได้หลายความหมาย (2) ปัญหาการแสดงรูปแบบกริยาจากภาษาต้นฉบับไปยังภาษาปลายทางให้มีความถูกต้องใกล้เคียงกับภาษาที่แปลไปมากที่สุด โดยผู้วิจัยได้ใช้ความรู้ทางด้านการออกแบบฐานข้อมูล, ความรู้ทางด้านหลักภาษาไทยมาช่วยในการออกแบบการจัดทำโปรแกรมเพื่อให้สามารถแปลงจากภาษา UNL ให้เป็นภาษาไทยได้รูปแบบที่มีความหมายถูกต้องตามหลักภาษาไทยมากที่สุด การจัดทำวิธีการแปลงภาษานั้นผู้วิจัยได้เลือกใช้วิธีการแบบเดียวกับการแปลจากภาษากลางมาเป็นภาษาไทย ทำการเปรียบเทียบรูปแบบกริยา ให้ถูกต้องตามหลักไวยากรณ์การแสดงภาษาไทยเพื่อให้ได้รูปประโยคที่สามารถอ่านได้ความหมายอย่างถูกต้อง และผู้วิจัยได้ทำการทดสอบโปรแกรมแปลภาษาโดยใช้ประโยคของ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาษา UNL จากตัวอย่างในหนังสือ Universal Networking Language มาทำการแปลเป็นภาษาไทยและทำการเปรียบเทียบความถูกต้องกับประโยคต้นฉบับที่เป็นภาษาอังกฤษ ซึ่งมีความถูกต้องประมาณ 85%

จากผลการออกแบบและพัฒนาโปรแกรมต้นแบบที่ใช้ในการแปลงจากภาษา UNL เป็นภาษาไทยและผลจากการทดสอบโปรแกรมตามที่ได้เสนอในบทที่ 5 โดยที่ฐานข้อมูลพจนานุกรมซึ่งบรรจุคำศัพท์ไว้ในไมโครซอฟท์แอสเซส และบรรจุคำศัพท์ที่เป็นคำสากลประมาณ 2,300 คำ จากปัญหาต่างๆ ที่พบในขั้นตอนของการทำทดสอบ ได้แก่ ปัญหาการแสดงรูปแบบกริยาให้ถูกต้องตามหลักภาษาไทย, ปัญหาที่คำศัพท์หนึ่งคำสามารถแปลเป็นภาษาไทยได้หลายความหมายในทำนองเดียวกันหรือมีความหมายใกล้เคียงกัน เช่นคำว่า "Work" และปัญหาการแสดงข้อความสุภาพ

6.2 ข้อดีและข้อเสียของโปรแกรม CU-Deconverter

ข้อดีของโปรแกรม CU- Deconverter ได้แก่

6.2.1. ขั้นตอนวิธีที่นำเสนอมีลักษณะขั้นตอนการโปรแกรมมิ่งแบบโครงสร้าง อาศัยการออกแบบฐานข้อมูลแบบพจนานุกรม ทำให้ระบบสามารถรองรับคำศัพท์ใหม่และรูปแบบกริยาใหม่ๆ

6.2.2. ระบบที่จัดทำนำเสนอนั้นสามารถเพิ่มคำศัพท์หรือรูปแบบกริยา รวมไปถึงข้อมูลพื้นฐานต่างๆ ทั้งในด้านหลักไวยากรณ์ภาษาไทยรวมไปถึงข้อมูลพื้นฐานที่เกี่ยวกับกฎเกณฑ์ของภาษา UNL ที่เพิ่มเติม

6.2.3. เนื่องจากโปรแกรมต้นแบบที่พัฒนาขึ้นได้แบ่งการทำงานออกเป็นโมดูลย่อยๆ ดังนั้นในกรณีที่มีการเปลี่ยนแปลงกฎเกณฑ์หรือวิธีการบางอย่างก็จะกระทบเฉพาะขั้นตอนนั้นๆ

ข้อเสียของโปรแกรม CU-Deconverter ได้แก่

6.2.4. ในกรณีที่คำศัพท์คำสากลหนึ่งคำสามารถแปลเป็นภาษาไทยได้หลายความหมายที่ใกล้เคียงกัน ตอนที่สร้างคำศัพท์ต้องเลือกคำที่ใช้ประกอบแสดงลักษณะเฉพาะของคำให้ชัดเจนและสื่อความหมายให้ถูกต้อง

6.2.5. ฐานข้อมูลพจนานุกรมที่เป็นตารางรูปแบบกริยา จะต้องบรรจุข้อมูลจำนวนมากเพื่อให้ได้รูปแบบกริยาที่ถูกต้องตามหลักภาษาไทย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

6.3 ข้อจำกัด

เนื่องจากวิทยานิพนธ์ฉบับนี้โปรแกรม CU-Deconverter และขั้นตอนการทำงานของโปรแกรมได้อ้างอิงลักษณะตามมาตราฐานของภาษา UNL ดังนั้นหากมีการแก้ไขหรือเปลี่ยนแปลงกฎเกณฑ์ต่างๆดังกล่าวข้างต้นจะมีผลทำให้กระทบกับโครงสร้างและขั้นตอนการทำงานของโปรแกรม อีกทั้งตัวอย่างที่นำมาทดสอบสำหรับวิทยานิพนธ์ฉบับนี้ได้นำมาจากในหนังสือ UNL Center, Institute of Advanced Studies, The United Nations University. UNL Universal Networking Language, 1996 ซึ่งมีตัวอย่างประโยคค่อนข้างจำกัด

6.4 ข้อเสนอแนะและความคิดเห็นเพิ่มเติม

จากการทำวิจัยวิทยานิพนธ์เกี่ยวกับการแปลงจากภาษา UNL เป็นภาษาไทยโดยที่ลักษณะของภาษา UNL นั้นอ้างอิงคำศัพท์หลักมาจากภาษาอังกฤษ ผลจากการทำงานและทดสอบโดยลงลึกในรายละเอียดของคำศัพท์และความหมายของคำ, ชนิดของคำรวมทั้งส่วนประกอบทางภาษาของคำศัพท์ต่างๆ ผู้วิจัยพบว่าโดยหลักการแล้วในแต่ละภาษาทุกชนิดน่าจะมีที่มาจากความต้องการถ่ายทอดสิ่งที่พบเห็นหรือบอกกล่าวการกระทำต่างๆ ซึ่งโดยพื้นฐานของแนวความคิดภาษาแล้วทุกภาษาน่าที่จะมีความหมายใกล้เคียงหรือถ่ายทอดเหมือนกัน แต่เนื่องมาจากขนบธรรมเนียมวัฒนธรรม, ประเพณี ลักษณะภูมิประเทศ, ศาสนาและแนวความเชื่อที่แตกต่างกัน จึงได้ทำให้ภาษาต่างๆ ทั่วโลกเกิดความแตกต่างกันเรื่อยมาตั้งแต่อดีต

ผลจากทดสอบโปรแกรมต้นแบบที่ได้พัฒนาขึ้นพบว่าการแปลจากภาษา UNL เป็นภาษาไทยค่อนข้างมีความถูกต้องในระดับหนึ่ง แต่พบว่าการแสดงลักษณะคำสุภาพหรือข้อความที่บ่งบอกเพศยังไม่มีข้อมูลรองรับทำให้การแปลงเป็นภาษาไทยในกรณีนี้ยังไม่ชัดเจน เช่นคำว่า “Eat” สามารถแปลความหมายในภาษาไทยได้ทั้งคำว่า “กิน” หรือ “รับประทาน” แล้วแต่กรณีหรือคำต่อท้ายที่แสดงทั้งความสุภาพและบ่งบอกเพศ ได้แก่คำว่า “ครับ” หรือ “ค่ะ”

สรุปโดยรวมแล้วผู้วิจัยมีความเห็นว่าการแปลจากภาษาหนึ่งไปเป็นอีกภาษาหนึ่งนั้น ถ้ามีข้อมูลที่ใช้ประกอบคำ เช่น บอกประเภทคำ, ชนิดคำ, ข้อมูลประกอบรูปแบบกริยาหรือการบ่งบอกความสุภาพ ซึ่งข้อมูลต่างๆ เหล่านี้จะทำให้การแปลภาษาสามารถทำได้อย่างง่ายและชัดเจน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ผู้วิจัยคิดว่าการแปลภาษาด้วยเครื่องจะมีประโยชน์มหาศาลสำหรับผู้ใช้งานทั่วโลกเนื่องจากทุกวันนี้การติดต่อสื่อสารที่รวดเร็วและมีความสะดวกสบายมากขึ้นประกอบกับเทคโนโลยีที่ก้าวหน้ารวมไปถึงอินเตอร์เน็ต (Internet) ซึ่งสามารถทำให้ผู้คนบนโลกนี้สามารถติดต่อสื่อสารถึงกันได้อย่างง่ายดายเพียงปลายนิ้วสัมผัส การถ่ายทอดความรู้, เทคโนโลยีและสิ่งต่างๆในปัจจุบันนี้ได้อาศัยภาษาอังกฤษเป็นสื่อกลางในการถ่ายทอด สำหรับบุคคลทั่วไปที่มีการศึกษาหรือบุคคลที่มีความรู้ภาษาอังกฤษเป็นอย่างดีย่อมไม่มีปัญหาในการติดต่อสื่อสาร แต่สำหรับบุคคลทั่วไปหรือประชาชนที่มีความรู้ไม่มากจะประสบปัญหาในการติดต่อสื่อสารหรืออ่านเอกสารในอินเตอร์เน็ต ซึ่งอาจจะบอกกล่าวว่าในเมื่อภาษาอังกฤษมีความสำคัญและจำเป็นอย่างไรเสียก็ต้องเรียนรู้นั้นก็อาจจะกล่าวได้ แต่จะยิ่งก่อให้เกิดประโยชน์ในวงกว้างมากขึ้น รวมไปถึงการถ่ายทอดวัฒนธรรมท้องถิ่นพื้นบ้าน ถ้ามีโปรแกรมหรือตัวกลางที่ใช้ในการถ่ายทอดจากภาษาท้องถิ่นไปเป็นภาษาอังกฤษหรือจากภาษาอังกฤษกลับไปเป็นภาษาท้องถิ่น และด้วยเทคโนโลยีที่ก้าวหน้าขึ้นมากประกอบกับราคาที่ถูกลงจึงมีความคิดว่าต่อไปในอนาคตอันใกล้นี้ ลักษณะของโปรแกรมที่เป็นการถ่ายทอดหรือแปลภาษานี้จะมีประโยชน์เป็นอย่างยิ่งและมีความสำคัญมากขึ้นเรื่อยๆ ต่อมวลมนุษย์ชาติ

อนึ่งถ้าโปรแกรมการแปลภาษาเหล่านี้สามารถแปลหรือถ่ายทอดได้อย่างถูกต้อง และสามารถนำโปรแกรมเหล่านั้นมาดัดแปลงเพิ่มเติมให้สามารถแปลงจากคำพูดเป็นคำพูด จากบทสนทนาระหว่างคนที่ใช้ภาษาต่างกันก็จะยิ่งเป็นประโยชน์ต่อมนุษย์ทุกคนที่ใช้ภาษาต่างกัน เสมือนทำให้โลกใบนี้เป็นโลกที่มีภาษาเดียวกันเหมือนยุคสมัยที่ก่อนเกิดเหตุการณ์หอคอยแห่งบาบิโลน

อนึ่งในอนาคตอันใกล้นี้ผู้วิจัยมีความเห็นว่าการใช้คอมพิวเตอร์แปลภาษาจะยิ่งทวีความสำคัญมากขึ้นเรื่อยๆ โดยมีองค์ประกอบดังต่อไปนี้

1. ราคาของเครื่องต่อประสิทธิภาพ (Price/Performance) จะยิ่งถูกลงมากยิ่งขึ้น ทำให้การจัดเก็บฐานข้อมูลพจนานุกรมสามารถทำได้ด้วยต้นทุนที่ไม่แพง อีกทั้งด้วยประสิทธิภาพของเครื่องที่สูงขึ้นทำให้โปรแกรมที่ใช้ในการแปลภาษาสามารถทำงานได้อย่างรวดเร็วและถูกต้อง

2. อินเตอร์เน็ตและการสื่อสารที่ติดต่อถึงกันได้ทั่วโลกทำให้มีความจำเป็นต้องใช้โปรแกรมแปลภาษา ถึงแม้ว่าปัจจุบันภาษาอังกฤษจะเป็นที่ยอมรับกันโดยทั่วไปและถือเป็น มาตราฐานของการสื่อสารและแลกเปลี่ยนข้อมูลกัน แต่กรณีนี้จะจำกัดกลุ่มคนเฉพาะบุคคลที่มีการศึกษาได้ระดับที่มีความรู้พอควร ไม่สามารถคลอบคลุมบุคคลได้ทั้งหมดทั่วโลก ดังนั้นหากการแปลภาษาสามารถทำได้จริงจะยิ่งขยายจำนวนผู้ใช้ออกไปมากยิ่งขึ้น

3. ในอนาคตอาจจะมีการพัฒนาเครื่องแปลภาษาที่ใช้ได้ในทุกกรณี ไม่เฉพาะที่ใช้ในการแปลเอกสารเท่านั้น ผู้วิจัยมีความเห็นว่าถ้าเป็นไปได้มนุษย์ทุกคนในโลกนี้สามารถติดต่อสื่อสารด้วยภาษาท้องถิ่นของตัวเองกับบุคคลที่ใช้ภาษาต่างกันได้จะยิ่งทำให้อุปสรรคของการติดต่อสื่อสารหมดไป ตัวอย่างในกรณีนี้หมายถึงว่า คนไทยสามารถสนทนาผ่านเครื่องแปลภาษาออกมาเป็นภาษาสเปนหรือภาษาเยอรมัน และคนสเปนหรือคนเยอรมันสามารถโต้ตอบกลับมาได้ด้วยภาษาของตัวเองและคนไทยสามารถเข้าใจได้โดยผ่านเครื่องแปลภาษา ซึ่งถ้าถึงจุดนั้นได้ก็จะเป็นใช้เครื่องแปลภาษาโดยสมบูรณ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

ดวงแก้ว สวามิภักดิ์. การสร้างซอฟท์แวร์วิเคราะห์ไวยากรณ์ไทย ภายใต้ระบบยูนิกซ์. โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2533.

นวลทิพย์ ตันติเศวตรัตน์และคณะ. โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์. ภาควิชาภาษาและสังคม สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี, 2531.

บุญเจริญ ศิริเนาวกุล. รายงานผลงานโครงการการพัฒนาระบบรวมเพื่อการแปลภาษาด้วยคอมพิวเตอร์. สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี, 2531.

ยืน ภู่วรวรรณและชัยยงค์ วงศ์ชัยสุวัฒน์. การประมวลผลภาษาธรรมชาติ. สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี, 2535.

ยืน ภู่วรวรรณ, ชัยยงค์ วงศ์ชัยสุวัฒน์และมนัสวีร์ อยู่ทวี. การพัฒนาระบบการแปลภาษาอังกฤษเป็นไทยด้วยหลักภาษากลาง. มหาวิทยาลัยสงขลานครินทร์, 2534.

รัตติกร วรากูลศิริพันธุ์. ระบบการสร้างภาษาเป้าหมายเพื่อการแปลภาษาด้วยคอมพิวเตอร์. สถาบันเทคโนโลยีพระจอมเกล้าทหารลาดกระบัง, 2532.

รัตติกร วรากูลศิริพันธุ์. ระบบการสร้างภาษาไทยเพื่อการแปลภาษาด้วยคอมพิวเตอร์ (ระยะที่ 2). สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2532.

รัตติกร วรากูลศิริพันธุ์,สมศักดิ์ จันวันและอรัญญา ปรีชาไว. การสังเคราะห์ภาษาไทยจาก ภาษากลางในระบบการแปลหลายหลายภาษา. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , 2532.

วันทนีย์ พันธชาติ. การแสดงรูปแทนภาษากลางสำหรับภาษาไทย. ห้องปฏิบัติการวิจัยภาษาและวิทยาการความรู้ ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2536.

วันทนีย์ พันธชาติ, วรนุช เกิดสินธ์ชัย, ทัศนีย์ เจริญพรและสุดา รังกุพันธ์. ระบบการวิเคราะห์ภาษาไทยเพื่อการแปลภาษาด้วยคอมพิวเตอร์. ห้องปฏิบัติการวิจัยภาษาและวิทยาการความรู้ ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2536.

วิรัช ศรเลิศล้ำวาณิช และสุรพันธ์ เมฆนาวิน. แนวโน้มการวิจัยระบบเครื่องแปลภาษา. ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, 2537.

อรัญญา ปรีชาไว. ระบบการวิเคราะห์ประโยคภาษาอังกฤษไปสู่ภาษากลางเพื่อการแปลด้วยคอมพิวเตอร์ วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าทหารลาดกระบัง, 2537.

ภาษาอังกฤษ

A.A.Hornby. Guide to Patterns and Usage in English, 2nd Edition. Oxford University Press, 1975.

A.B. Tucker. Current Strategies in Machine Translation Research and Development. in Machine Translation Theoretical and Methodological Issues, Cambridge University Press, 1987 : pp. 22-44.

C.J. Fillmore. A Case for Case Universals in Linguistic Theory. New York : Emmon Bach and R.T. Harms, 1968.

J.Solcum. A Survey of Machine Translation: Its History, Current Status and Future Prospects. American Journal of Computational Linguistics, Vol. 11, No.1 January - March 1985 : pp. 1 – 17.

James Allen. Natural Language Understanding. Benjamin Cummings, 1987.

Kereto. S. An English Thai Machine Translator. Proceedings of the Regional Symposium on Computer Science and its Applications At KMITL, 27-29 Jan 1987.

M.Nagao, J.Tsujii, J.Nakamura. The Japanese Government Project for Machine Translation. American Journal of Computational Linguistics, Vol.11, No.2-3, April - September 1985 : pp. 91 - 100.

Makoto Nagao. Machine Translation How Far Can It Go?. Oxford University Press,1989.

P. Thajchayapong et al. A Thai-English Machine Translation. in Proceeding of Regional Symposium on Computer Science and Its Applications At KMITL, 27-29 Jan 1987.

Shunji Ichiyama, Naoyuki Nomura. A Dictionary For Multilingual Translation System. Electronic Information Communication Conference, 1989.

UNL Center. UNL Universal Networking Language. Institute of Advanced Studies, The United Nations University, 1996.

Free Online Translation. SYSTRAN, 1999. http://translationsoftware4u.com

Inter-Tran Language Translate Page. INTER-TRAN, 1999. http://www.flinet.com

Translate via the Internet with Languageware. Language Ware, 1999. http://www.languageware.net/index.html

Universal Networking Language Information. UNL Center, 1999. http://unl.ias.unu.edu

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก
ตัวอย่างตารางข้อมูลฐานข้อมูล UNL

1. ตัวอย่างตารางข้อมูลแสดงลักษณะเฉพาะคำ, ประเภทคำ, หน้าที่คำ

แสดงลักษณะเฉพาะคำ	ชื่อ	ชื่อภาษาไทย	คำอธิบายภาษาไทย
A part of animal or its derivative	A part of animal or its derivative	ส่วนของสิ่งมีชีวิต	หรือสิ่งที่ได้จากสิ่งมีชีวิต เช่น มือ เท้า แขน ไข่ น้ำตา
A part of plant or its derivative	A part of plant or its derivative	ส่วนของพืชหรือสิ่งที่ได้จากพืช	เช่น กิ่งไม้ ราก เปลือก
A part of product or its derivative	A part of product or its derivative	ส่วนของผลิตภัณฑ์	หรือผลที่ได้จากผลิตภัณฑ์ เช่น ขาโต๊ะ ขี้ยางลบ ใบมีด
Abstract	Abstract	นามธรรม	
Abstract matter	Abstract matter	นามธรรมที่คลอบคลุมเหตุการณ์	และกริยาอาการต่างๆ
Abstract relation	Abstract relation	ความสัมพันธ์เชิงนามธรรม	
Abstract thing	Abstract thing	นามธรรมที่มีลักษณะคล้ายสิ่งจับต้องได้	นามธรรมที่มีลักษณะคล้ายสิ่งที่สามารถจับต้องได้
Abstract thing fixed as calture	Abstract thing fixed as calture	สิ่งที่เป็นนามธรรมที่เกี่ยวกับวัฒนธรรม	สิ่งที่เป็นนามธรรมที่เกี่ยวข้องกับวัฒนธรรม เช่น ภาษาไทย รำวงมวยไทย
Abstract thing fixed as socially	Abstract thing fixed as socially	สิ่งที่เป็นนามธรรมที่เกี่ยวข้องกับสังคม	เช่น งานปีใหม่ งานลอยกระทง
Action	Action	การกระทำ	เช่น ปฏิกิริยา เล่น เดินทาง ใช้
Activity	Activity	กริยาและอาการต่างๆ	
Animal	Animal	สัตว์	เช่น หนอน หมา แมว
Appearance	Appearance	ลักษณะที่ปรากฏให้เห็น	เช่น รูปร่าง สี ลาย
Astronomy	Astronomy	สถานที่ที่เกี่ยวข้องกับดาราศาสตร์	เช่น โลก พระอาทิตย์
Attribute	Attribute	คุณสมบัติของวัตถุหรือสิ่งของ	
Concrete	Concrete	รูปธรรม	
Concrete place	Concrete place	สถานที่	
Concrete thing	Concrete thing	สิ่งของ	

Condition	Condition	สภาวะ	เช่น สภาพร่างกาย สถานการณ์
Emotional-action	Emotional action	การกระทำที่เกี่ยวข้องกับความรู้สึก	เช่น รัก หวัง โกรธ เสียใจ
Event	Event	เหตุการณ์	เช่น ดวง โชค ความสำเร็จ
Expression	Expression	แสดงความหมาย, ถ้อยคำ	เช่น ภาษา
Finished	Finished Product	ผลิตภัณฑ์	โต๊ะ ดวงไฟ รถยนต์
Food	Food	อาหาร	
Function	Function	หน้าที่และประสิทธิภาพ	เช่น ความสามารถ ประสิทธิภาพ คุณภาพ
Gas	Gas	ก๊าซ	เช่น อากาศ ไอน้ำ ก๊าซธรรมชาติ
Immaterial man-made thing	Immaterial man-made thing	สิ่งที่ไม่ใช่วัตถุที่มนุษย์สร้างขึ้น	เช่น วรรณคดี เพลง บทความ โปรแกรมคอมพิวเตอร์
Liquid	Liquid	ของเหลว	เช่น น้ำ ปรอท น้ำมัน
Living-thing	Living thing	สิ่งมีชีวิต	
Man-made thing	Man-made thing	สิ่งของที่มนุษย์สร้างขึ้น	
Man-made-place	Man-made place	สถานที่ที่มนุษย์สร้างขึ้น	เช่น ทางด่วน ตึก บ้าน
Material	Material	วัตถุดิบ	เช่น พลาสติก ผ้า
Mental action	Mental action	การกระทำที่เกี่ยวกับความรู้สึกนึกคิด	
Natural phenomenon	Natural phenomenon	ปรากฏการณ์ธรรมชาติ	เช่น พายุฝน สุริยุปราคา
Nature	Nature	สถานที่เป็นธรรมชาติ	
Nature thing	Nature thing	สิ่งของที่เกิดตามธรรมชาติ	
Number/Quantity	Number/Quantity	จำนวน	เช่น ปริมาตร ความสูง
Organization	Organization	กลุ่มหรือองค์กรต่างๆ	เช่น บริษัท กรรมการ กลุ่ม คณะ
Person	Person	บุคคล	เช่น ผู้ปกครอง เขา ผู้ดูแล คู่รัก เพื่อน
Phenomenon	Phenomenon	ปรากฏการณ์	
Physiological-phenomenon	Physiological phenomenon	ปรากฏการณ์ทางสรีระหรือทางร่างกาย	เช่น อาการป่วย การเกิด การตาย การย่อยอาหาร
Place	Place	สถานที่	เช่น ขวา ซ้าย บน แถวๆ
Plant	Plant	พืช	เช่น ดอกไม้ มะนาว กุหลาบ
Product of activity	Product of activity	ผลของกริยาและอาการต่างๆ	
Quality	Quality	ค่าที่แสดงลักษณะ	เช่น ใจดี สูง ง่าย

Quanlity	Quanlity	คุณลักษณะ	เช่น ลักษณะ อุปนิสัย
Rational action	Rational action	การกระทำที่ต้องใช้เหตุผลและความคิด	เช่น คิด ตัดสินใจ ควบคุม
Region	Region	สถานที่	เป็นดินแดนหรือมีขอบเขตที่แน่นอน เช่น จังหวัด อำเภอ หมู่บ้าน
Relation	Relation	ความสัมพันธ์	เช่น ความเกี่ยวข้อง
Result of action	Result of action	ผลของการกระทำ	เช่น คำอธิบาย แผนการ ข้อเสนอ
Result of mental action	Result of mental action	ผลของการกระทำที่เกี่ยวข้องกับความรู้สึก	นึกคิด เช่น ความจำ ความเป็นห่วง ความรัก
Set	Set	กลุ่ม	เช่น เป็นแแถว เรียงราย เป็นกลุ่ม
Social phenomenon	Social phenomenon	ปรากฏการณ์ทางสังคม	เช่น การประท้วง สงคราม การจราจล
Solid	Solid	ของแข็ง	เช่น ทอง เงิน เพชร
Sort/Type	Sort/Type	ประเภทหรือชนิด	เช่น แบบ ชนิด ประเภท เผ่า
Subject	Subject	บุคคลและกลุ่ม	ที่ประกอบด้วยคน แบ่งเป็น
Substance	Substance	สสาร	
Time	Time	เวลา	เช่น เมื่อวานนี้ สัปดาห์นี้ สมัยธนบุรี
Topography	Topography	สถานที่ที่เป็นภูมิประเทศ	เช่น ทุ่ง เทือกเขา ชายทะเล
Value	Value	ค่าของคุณสมบัติหรือสิ่งของ	
Value Appearance	Value Appearance	ค่าที่แสดงลักษณะที่ปรากฏ	เช่น กลุ่ม สี่เหลี่ยม สว่าง แดง
Value Condition	Value Condition	ค่าที่แสดงสภาวะ	เช่น เป็นอันตราย ปลอดภัย
Value Function	Value Function	ค่าที่แสดงหน้าที่และประสิทธิภาพ	เช่น (ประสิทธิภาพ)สูง (คุณภาพ)เยี่ยม
Value Number/Quantity	Value Number/Quantity	ค่าที่แสดงจำนวนหรือปริมาตร	เช่น 10 วัน มาก นาน
Value Relation	Value Relation	ค่าที่แสดงความสัมพันธ์	เช่น ใกล้เคียง คล้าย ติดกัน
Value Sort/Type	Value Sort/Type	ค่าที่แแสดงชนิดหรือประเภท	เช่น รุ่น 2503 แบบดิจิตัล

ตารางที่ 7.1 ตัวอย่างตารางข้อมูลแสดงลักษณะเฉพาะคำ

ประเภทคำ	ชื่อ	ชื่อภาษาไทย	คำอธิบายภาษาไทย
ADV	Adverb	กริยาวิเศษณ์	คำขยายกริยาเพื่อบอกว่าเหตุการณ์หรือสิ่งที่เกิดขึ้น เกิดเมื่อไร ที่ไหน หรืออย่างไร ในภาษาไทยกริยาวิเศษณ์จะไม่ปรากฏหน้าคำกริยาที่ถูกขยายเลย
AUX	Auxiliary Verb	คำช่วยกริยา	คำที่เกิดร่วมกับคำกริยาเพื่อแสดงกาล มาลา วาจก ลักษณาการ และเจตนาของประโยค และสามารถเกิดในตำแหน่งก็ได้
CLAS	Classifier	ลักษณะนาม	คำที่ใช้บอกลักษณะเฉพาะของคำ เช่น คัน ฉบับ แท่ง เครื่อง อาจซ้ำกับคำๆ นั้นเองก็ได้ เช่น ประเทศ อำเภอ ภาษา หรือบอกลักษณะที่เป็นกลุ่มของคำ เช่น ชุด คู่ คณะ กลุ่ม และรวมถึงคำที่แสดงหน่วยนับด้วย เช่น กิโลกรัม กระสอบ เมตร
CONJ	onjunction	คำสันธาน	คำที่ใช้เชื่อมวลี หรือประโยคเข้าด้วยกัน เพื่อแสดงความสัมพันธ์ในแง่ใดแง่หนึ่ง
DET	Determiner	คำบ่งชี้	คำที่ใช้แสดงความเฉพาะเจาะจงของคำนามโดยปรากฏว่าร่วมกับคำนามและคำลักษณะนาม
END	Ending	คำลงท้าย	คำที่ปรากฏท้ายประโยคเพื่อแสดงคำถามหรือแสดงความสุภาพ
FIXP	Perfix	คำอุปสรรค	คำที่ใช้ประกอบหน้าคำ เช่น กริยาเพื่อให้เป็นคำนามหรือคำกริยาวิเศษณ์ คำที่เติมคำอุปสรรคจะเปลี่ยนหน้าที่ของคำ โดยมิได้เปลี่ยนความหมาย เช่น การความ อย่าง โดย ตัวอย่าง การทำงาน ความรัก อย่างวิเศษ โดยตรง
INT	Interjection	คำอุทาน	คำที่ใช้ในการอุทาน
N	Noun	คำนาม	คำที่ใช้เรียกคน สัตว์ สิ่งของ
NEG	Negation	คำปฏิเสธ	คำที่แสดงการปฏิเสธหรือเพื่อให้ความหมายกลับเป็นตรงข้าม เช่น ไม่ ไม่ได้ มิได้ มิ ไม่ใช่
P	Pronoun	สรรพนาม	คำที่ใช้แทนคำนาม และอาจแทนด้วยคำบ่งชี้ที่มีความหมายรวมถึงนามที่ละไว้ด้วย
PREP	Perposition	คำบุรพบท	คำที่นำหน้าคำนามหรือคำสรรพนามเพื่อบอกตำแหน่งสถานที่ และความสัมพันธ์ระหว่างคำกริยาและคำนามหรือสรรพนามที่ตามมา
V	Verb	กริยา	คำที่แสดงกริยาอาการหรือสภาพของประธานในประโยค และถือเป็นคำหลักของประโยค

ตารางที่ 7.2 ตัวอย่างตารางข้อมูลประเภทคำ UNL

รหัสหน้าที่คำ	ชื่อ	ค่า	คำอธิบายภาษาไทย
ADJ	Adjective	22000	เช่น สวย ใหญ่
CAD	Cardinal Number	17000	เช่น 1, 2, สาม, ห้า, ร้อย
CLAS	Classifier	19000	เช่น คัน, ตัว, เล่ม, ด้าม
COMP	Complement	7000	เป็นส่วนที่อยู่หลัง IOB
COMPF	Complement	8000	ที่บอกเวลา, บอกเหตุผล
DEM	Memonsative	21000	เช่น นี่, นั่น
DOB	Direct Object	5000	กรรมตรงของประโยค ได้แก่ N, PRON
INT	Interogative marker	9000	เช่น หรือไม่, หรือ, อะไร
IOB	Indirect Object	6000	กรรมรองของประโยค ได้แก่ N, PRON
N	Noun	10000	เช่น ไก่, ม้า
ORD	Ordinal Number	20000	เช่น แรก, คันแรก
PRED	Predicate CP	4000	ได้แก่ กริยาหลัก, คำคุณศัพท์ที่ทำหน้าที่เป็นกริยาหลัก
PREP	Proposition	11000	เช่น ที่, บน, ใน
PRON	Pronoun	15000	เช่น เขา, หล่อน, ฉัน
QUANT	Quantitative	18000	เช่น หลาย, บาง
SADV	Sentencial Adverb	1000	ส่วนขยายความประโยค เช่น โดยทั่วไป, ที่จริง, ตามความจริง, ดัง
SMOD	Sentencial Modefier	2000	ส่วนขยายประโยค เช่น เมื่อเร็วๆ นี้, กาลครั้งหนึ่ง
SUB	Subject	3000	ประธานของประโยค
V	Verb	13000	เช่น กิน, เดิน, นั่ง

ตารางที่ 7.3 ตัวอย่างตารางข้อมูลหน้าที่คำ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. ตัวอย่างตารางข้อมูลคำเชื่อมโยง, แสดงคุณลักษณะ

รหัส	ชื่อ	ชื่อไทย	คำอธิบาย	คำอธิบายภาษาไทย	ประเภท	ตัวอย่างภาษาอังกฤษ	ตัวอย่างภาษา UNL
=	Synonym	ความหมายเหมือนกัน	Mean synonym	คำที่มีความหมายเหมือนกันหรือใกล้เคียงกัน	2		book=code
agt	Agent	ผู้กระทำ, ตัวแทน	An agent which causes volitional action. Agent is an animated object with intention.	ตัวแทนที่มีกำลังใจในการกระทำเหตุการณ์หรือที่ทำการเคลื่อนไหววัตุโดยมีจุดมุ่งหมาย	1	A rabbit runs.	run-agt->rabbit
and	Conjoint	สัมพันธ์กัน	Conjoint relation of things/events	ความสัมพันธ์ระหว่างสิ่งของกับเหตุการณ์	1	Cats and rats are animals. A cat mews and a dog barks.	(cat<-and-rat)<-obj-animal. Cat<-agt-mew<-and-bark-agt->dog
aoj	Attributive Object	คุณสมบัติของวัตถุ	An object of an attribute.	คุณสมบัติของวัตถุ	1	Leaves are green. / green leaves	green-aoj->leave
bas	Basis	เกณฑ์, หลักเกณฑ์	Criteria for comparison	เกณฑ์สำหรับการเปรียบเทียบ	1	A cat is bigger than a rat.	cat<-obj-big-bas->rat
ben	Beneficiary	ผู้รับประโยชน์	Beneficiary of events	ผู้รับประโยชน์จากเหตุการณ์	1	to work for family	work-ben->family
cag	Co-agent	ผู้ร่วมกระทำ	Concomitant / Co-agent	คู่กัน, มาพร้อมกัน, สิ่งที่มาพร้อมกัน/ผู้ร่วม	1	to walk with a friend	walk-cag->friend

				กระทำ			
cau	Cause	เหตุการณ์	A matter of an event which is caused by it.	เรื่องราวของเหตุการณ์ที่เกิดเหตุการณ์เอง	1	to die of cacer	die-cau->cancer
con	Condition	เงื่อนไข	Condition which brings about events	เงื่อนไขที่นำไปสู่เหตุการณ์	1	to return home because of rain	home<-gol-return-con->rain
coo	Co-occurrence	เกิดขึ้นพร้อมกัน	Simultaneous progress of events	เหตุการณ์ที่ดำเนินต่อไปในเวลาเดียวกัน	1	to return home crying	home<-gol-return-coo->cry
fmt	From-To	จาก-ถึง	Range of things/events	ช่วงของสิ่งของ/เหตุการณ์	1	a ticket from London to Oxford	ticket-mod->(London<-fmt-Oxford)
gol	Goal	จุดหมาย	Last place(physical/logical) of agent/object in an event	จุดหมายที่เป็นสถานที่สุดท้าย (สถานที่จริง/นามธรรม) ของเหตุการณ์/วัตถุในเหตุการณ์	1	to go to London	go-gol->London
icl	Inclusion	ประกอบด้วย	Relation labels between Uws	ความสัมพันธ์ระหว่างคำ Uws	2	dog is animal	dog-icl->animal
ins	Instrument	เครื่องมือ	An instrument of a volitional action	เครื่องมือที่ใช้กระทำเหตุการณ์ที่ตั้งใจ	1	to cut with a knife	cut-ins->knife

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

			which is done with it.				
lpl	Logical Place	สถานที่นามธรรม	A scene of an event which occurs in it.	ฉากหรือสถานที่ของเหตุการณ์ที่เกิดขึ้น	1	to win in competition	win-lpl->competition
man	Manner	แบบ, วิธี	A manner of action or change done with it.	แบบของการกระทำหรือมีการเปลี่ยน	1	to run fast	run-man->fast
mat	Material	วัตถุดิบ	An material of action where it plays.	วัตถุดิบที่นำมาใช้กระทำ	1	to make butter from milk	butter<-obj-make-mat->milk
met	Method	วิธีการ	Relation of means or method to a volitinal action which is done with it.	ความสัมพันธ์ของความหมายหรือวิธีการที่ใช้กระทำอย่างตั้งใจ	1	to cure by surgery	cure-met->surgery
mod	Modifier	ส่วนขยาย	Modification of object	ส่วนขยายของวัตถุ	1	the third man	man-mod->third
num	Number	เลขที่	Numerical quantity of object, unit	เลขที่จำนวนของวัตถุ, หน่วย	1	three miles	mild-num->3
obj	Object	วัตถุ	An object of action or change that affects it.	วัตถุที่ใช้ในการกระทำหรือวัตถุที่มีผลกระทบจากการเปลี่ยนแปลง	1	To eat an apple. Snow melts.	eat-obj->apple. melt-obj->snow
opl	Objective Place	สถานที่ตามวัตถุประสงค์	A place of an action which	สถานที่ใช้ในการกระทำที่มีผลกระทบ	1	to walk on the road	walk-opl->road

			affects it.				
or	Disjunctive	หรือ, ไม่สัมพันธ์กัน	Disjunctive relation of things/events	ความไม่สัมพันธ์กันของสิ่งของ/เหตุการณ์	1	to keep a cat or dog	keep-obj->(cat<-or-dog)
per	Per unit	ต่อหน่วย	Unit to measure object	จำนวนหน่วยนับของวัตถุ	1	2 days a week	<2<-num-day)-per->week
pos	Possessor	มี, การมี, เป็นเจ้าของ	Possessor of the possessed object	ความเป็นเจ้าของวัตถุ	1	John's dog	dog-pos->John
ppl	Physical Place	สถานที่จริง	A place of an event which occurs in it.	สถานที่เกิดเหตุการณ์	1	to play in a garden	play-ppl->garden
ptn	Partnet	ผู้มีส่วนร่วม	Counter-partner of an action which requires symmetrical cooperation.	จำนวนของผู้มีส่วนร่วมของการกระทำที่มีต้องการจุดประสงค์ร่วมกัน	1	to compete with a friend	compete-ptn->friend
pur	Purpose	วัตถุประสงค์	Purpose of a volitional action	วัตถุประสงค์ของการกระทำที่ตั้งใจ	1	to come to see me	come-pur->see-obj->I
qua	Quantity	ปริมาณ, จำนวน	Quantity of object / amount or degree of change	ปริมาณของวัตถุ/จำนวนหรือระดับของการเปลี่ยน	1	3kg of apples. To walk three miles.	apple-qua->kg-num-3. Walk-qua->mile-num-3
seq	Sequential	เรียงตามลำดับ	An event of the	เหตุการณ์ที่เกิดเรียงตาม	1	to go to a library and borrow books	book<-obj-borrow-seq->go-gol-library

	Order		following event	กัน			
soj	Stative Object	สภาพวัตถุ	An object of a state.	สถานะของวัตถุ	1	The sea is calm. / clam sea	calm-soj->sea
src	Source	แหล่งที่มา	Original place (physical/logical) or state of agent/object in an event	สถานที่ต้นกำเนิด (สถานที่จริง/นามธรรม) หรือสถานะของตัวแทน/วัตถุที่อยู่ในเหตุการณ์	1	to come from London	come-src->London
tim	Time	เวลา	Time of an event which occurs at that time.	เวลาที่เกิดเหตุการณ์	1	to lunch at noon	lunch-tim->noon
tmf	Time-From	จากเวลา	Time of an event which begins at that time.	เวลาที่เริ่มต้นเกิดเหตุการณ์	1	to work from noon	work-tmf->noon
tmt	Time-To	ถึงเวลา	Time of an event which ends at that time.	เวลาที่สิ้นสุดเหตุการณ์	1	to work until noon	work-tmt->noon

ตารางที่ 7.4 ตัวอย่างตารางข้อมูลคำเชื่อมโยง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

AL Code	Description	Type	English Example	UNL Example
begin-just	Event has just started.	30	The match has just begun.	match<-obj-begin.@begin-just
begin-soon	Event is about to begin.	30	The airplane is about to land. My father is dying.	airplane<-agt-land.@begin-soon. I<-pos-father<-obj-die.@begin-soon
begin-very-soon	Sub-categorization can be proposed by attaching UW with '-'	30		
complete	Completion of an event	30	She has plyaed a piano.	she<-agt-play.@complete-obj->piano
complete-all	Sub-categorization can be proposed by attaching UW with '-'	30		
complete-part	Sub-categorization can be proposed by attaching UW with '-'	30		
def	UW already	10	The monkey fell from a branch.	monkey@def<-obj-fall-src->branch
emphasis	Emphasized element.	4a	His face I despise.	I<-agt-despise-obj->face.@emphasis
end-just	Event has just finished.	30	He has just come home.	he<-agt-come.@end-just-gol->home
end-soon	Event is about to end.	30	We are approaching the goal.	we<-agt-approach.@end-soon-gol->goal
entry	Root Entry	01	Monkey eats bananas.	eat(icl>action).@present.@entry:00
focus	Focus	4a	Is that you who want?	go.@interrogation->agt->you.@focus
future	Event which will happen in future	20	She will come at 9.	she<-agt-come.@future-tim->9
generic	Generic UW	10	A monkey likes bananas.	monkey@generic<-agt-like-obj->banana

indef	Non-specified class of UW, but not generic	10	I bought a knife.	I<-agt-buy-obj->knife.@indef
not	Complement	10	I do not go., It is not an orange.	I<-agt-go@not. It<-aoj-orange.@not
past	Event happened in the post	20	He dit it.	he<-agt-do.@past-obj->it
pl	Plurality of UW, but not generic.	10	I bought knives.	I<-agt-buy-obj->knife.@pl
present	Event happening at present	20	She goes to school at 9.	she<-agt-go.@present-gol->school.@present-tim->9
progress	Event is progressing.	30	A girl is crying. Children are dying of hunger.	girl<-agt-cry.@progress child.@pl<-obj-die.@progress-cau->hunger
repeat	Repetition of the same event of the same agent/object	30	A child is jumping.	child<-agt-jump.@repeat
state	State of event finished with a remaining result	30	The pot has broken.	pot<-obj-break.@state
theme	Theme	4a		
topic	Topic:Emphasis Topic/ Contrastive Topic/ Given Topic	4a	Most of these problems a computer could solve easily.	solve-obj->problem.@topic-mod->most

ตารางที่ 7.5 ตัวอย่างข้อมูลตารางแสดงคุณลักษณะ

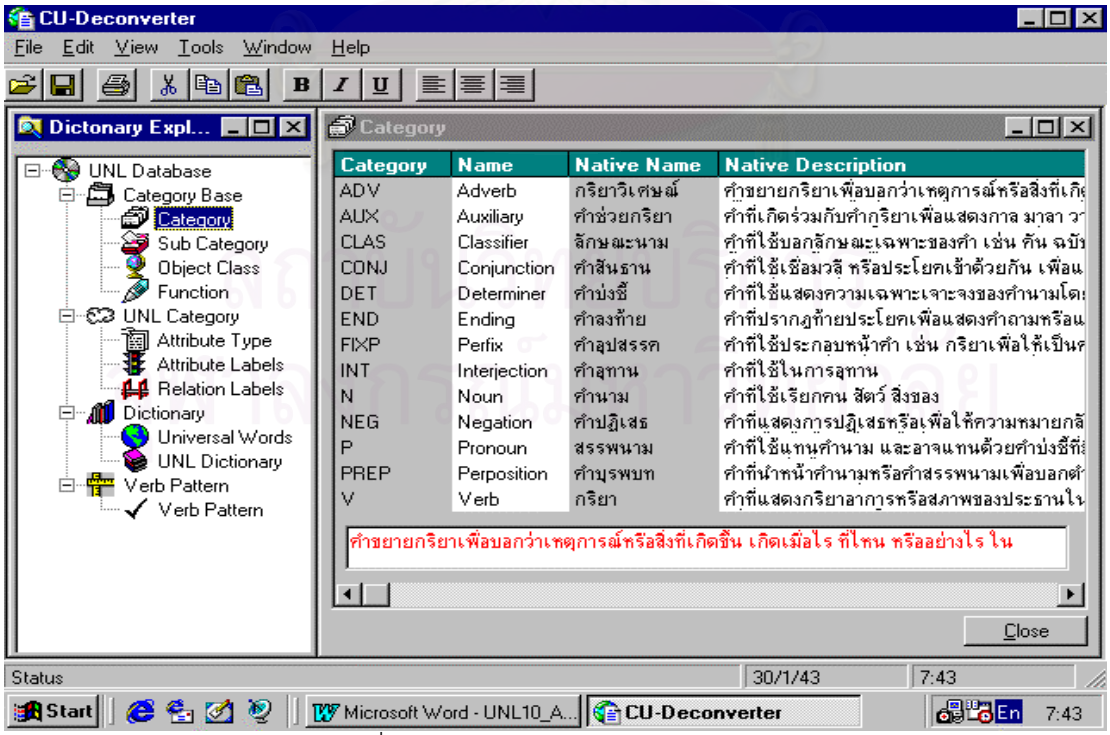
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

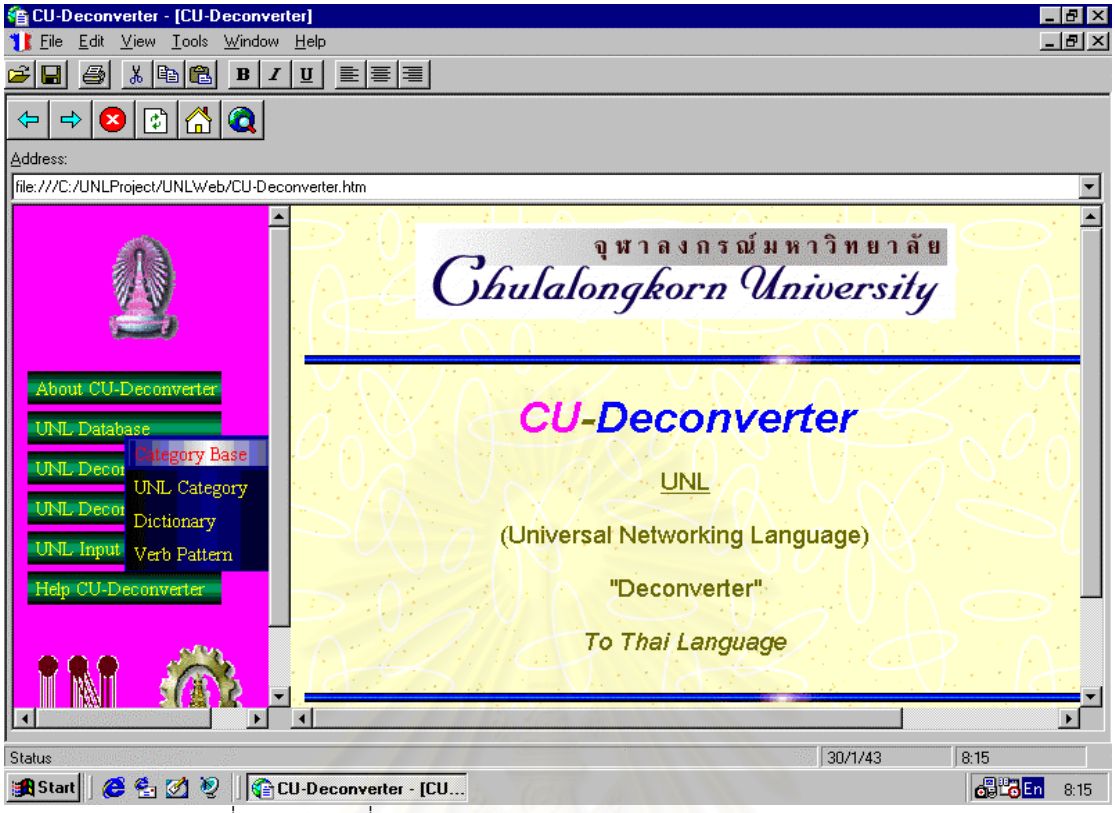
ตัวอย่างหน้าจอโปรแกรม CU-Deconverter

1. ตัวอย่างหน้าจอติดต่อแบบวินโดว์

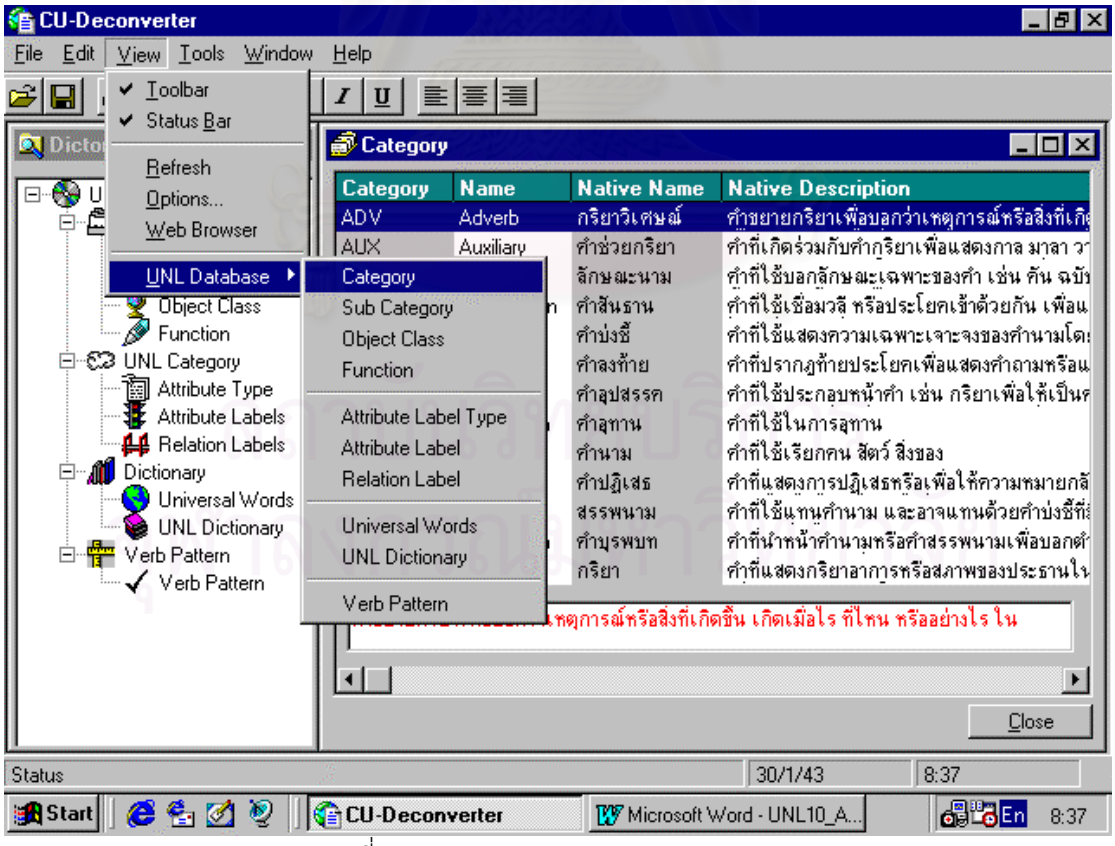
รูปที่ 7.1 หน้าจอเริ่มเข้าสู่โปรแกรม CU-Deconverter



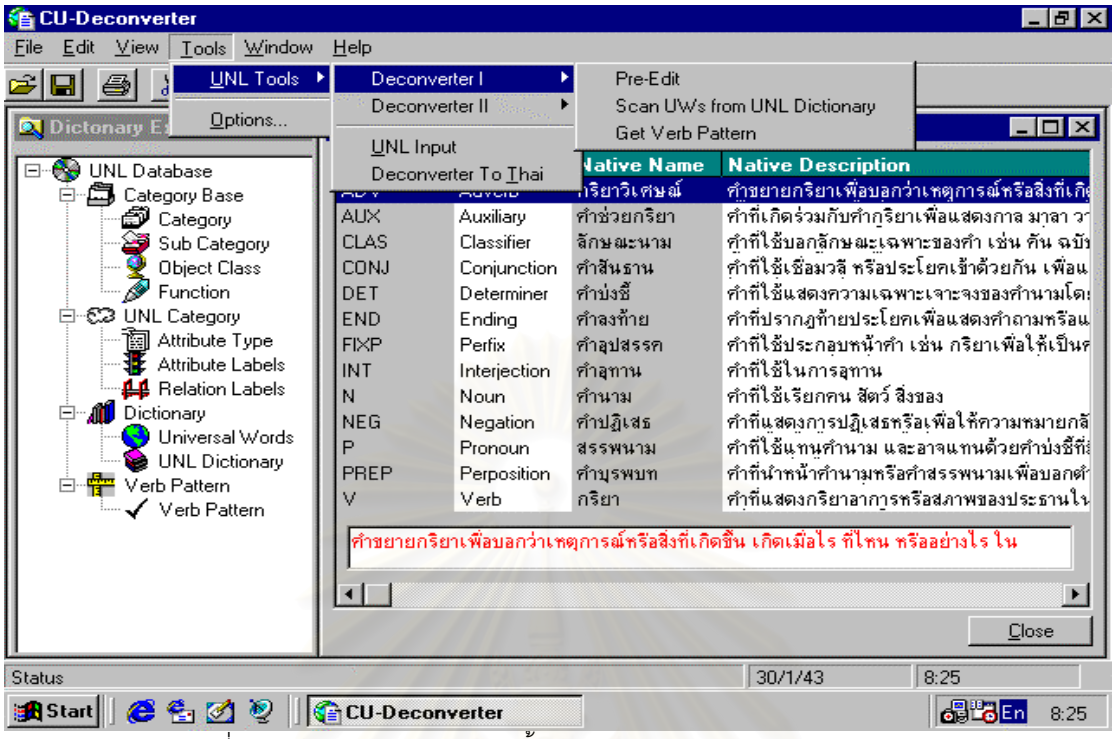
รูปที่ 7.2 หน้าจอ UNL Dictionary Explorer



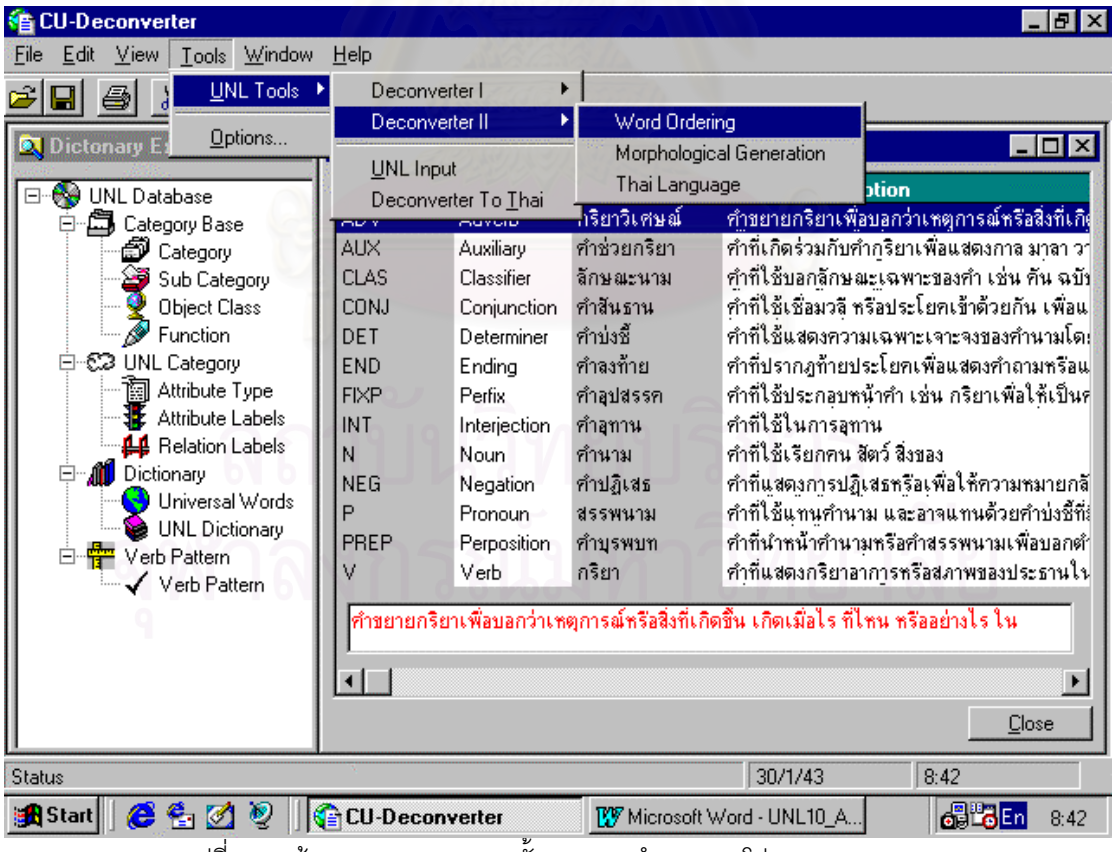
รูปที่ 7.3 หน้าจอที่แสดงผ่าน Web Browser ในแบบ Windows Interface



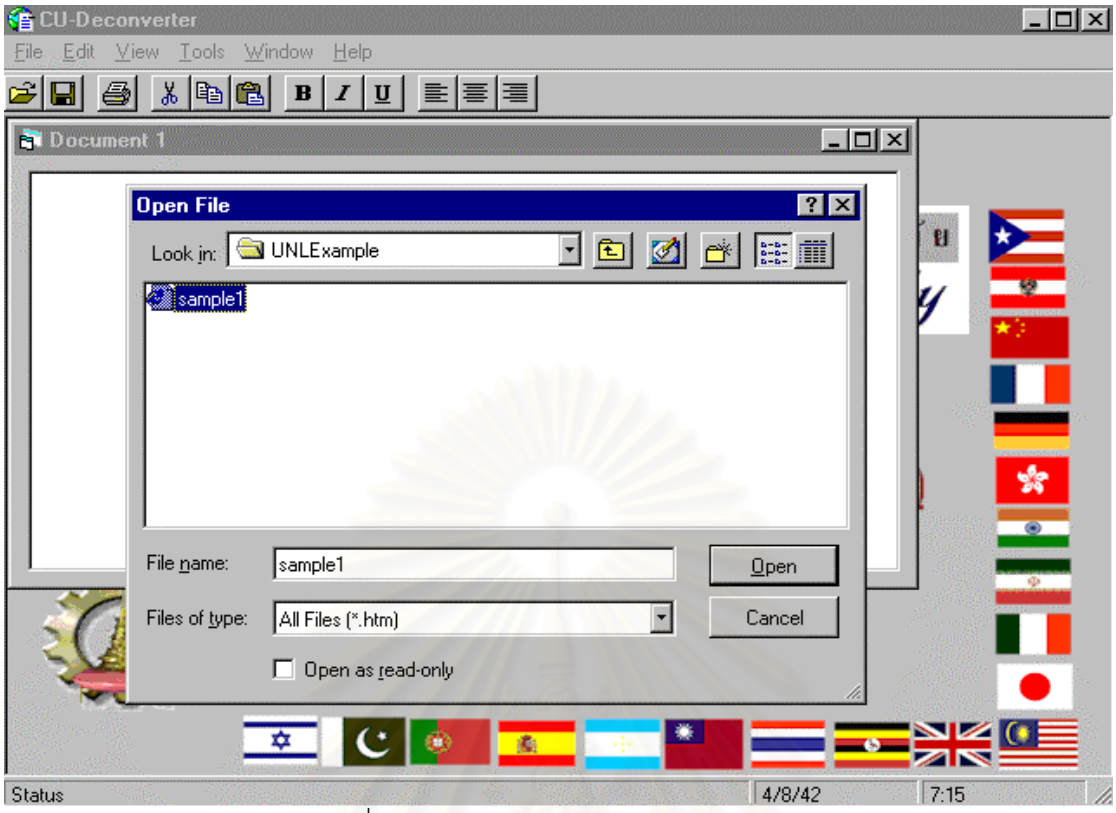
รูปที่ 7.4 แสดงเมนูของ View UNL Database



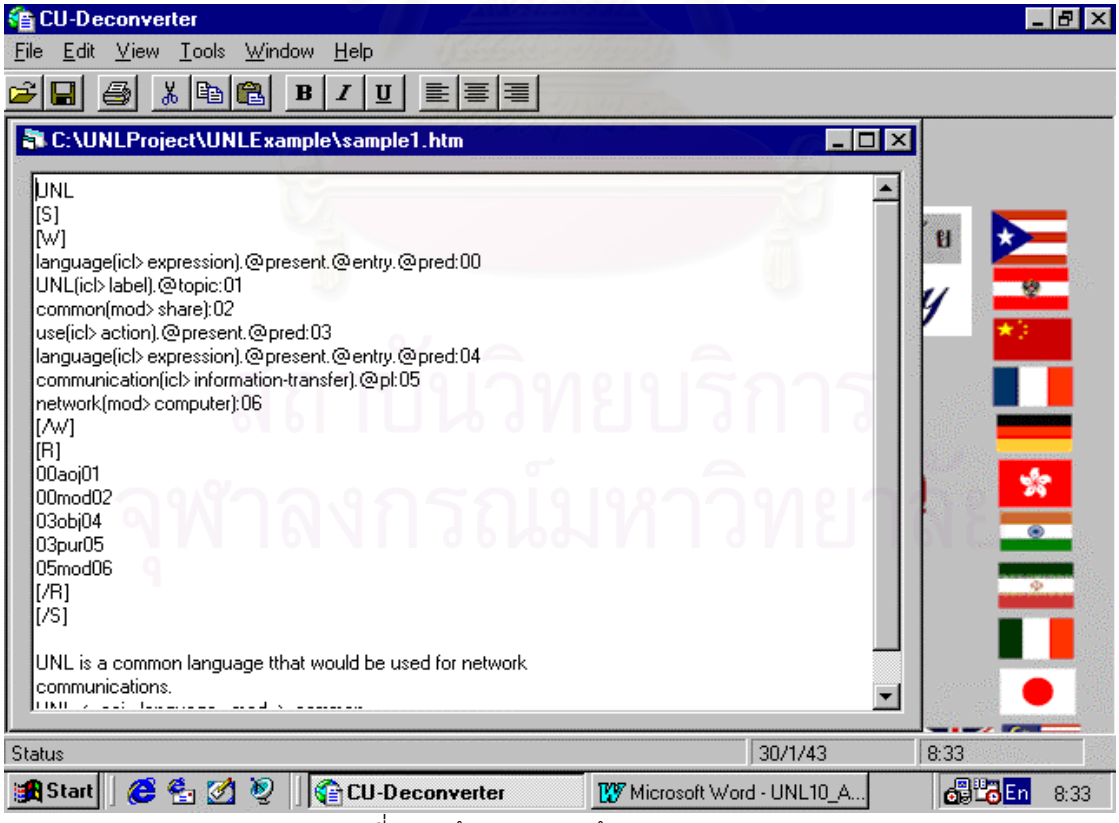
รูปที่ 7.5 หน้าจอเมนู Tools แสดงขั้นตอนการทำงานของโปรแกรม Phase I



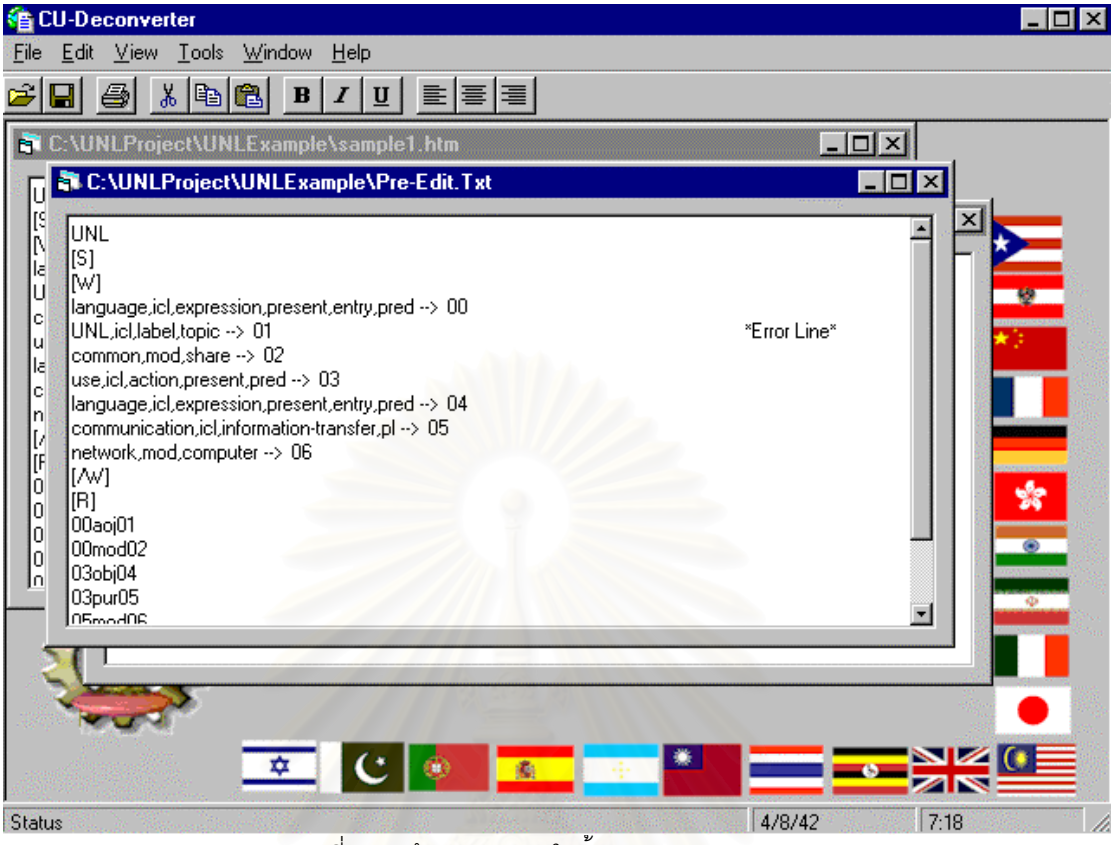
รูปที่ 7.6 หน้าจอเมนู Tools แสดงขั้นตอนการทำงานของโปรแกรม Phase II



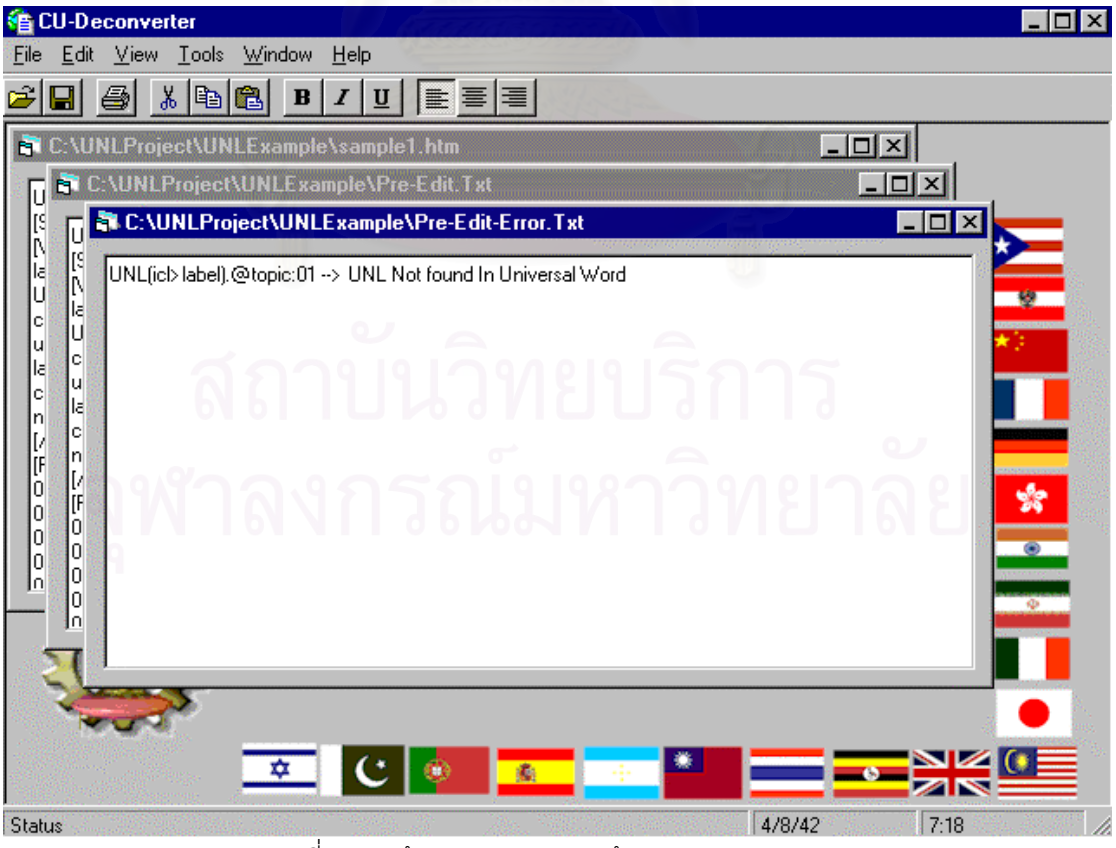
รูปที่ 7.7 หน้าจอ UNL Input รับแฟ้มข้อมูล UNL



รูปที่ 7.8 หน้าจอหลังจากขั้นตอน UNL Input

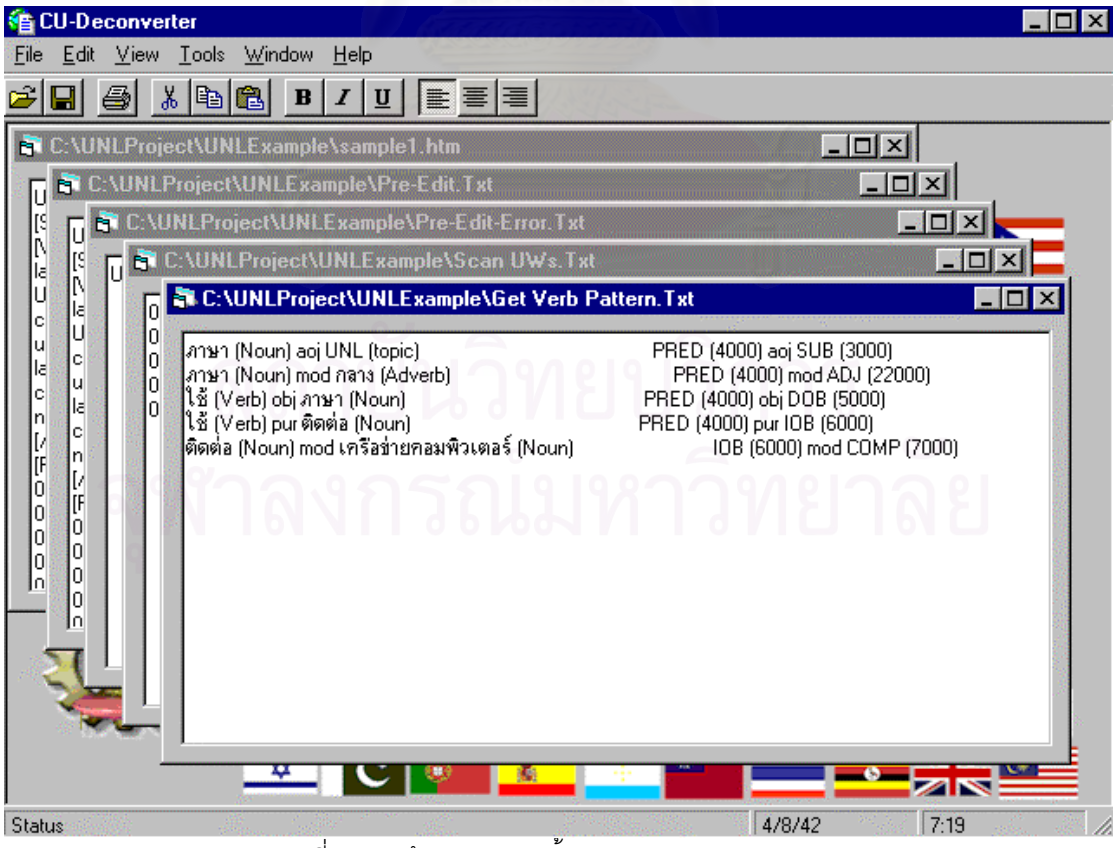


รูปที่ 7.9 หน้าจอแสดงผลในขั้นตอน Pre-Edit

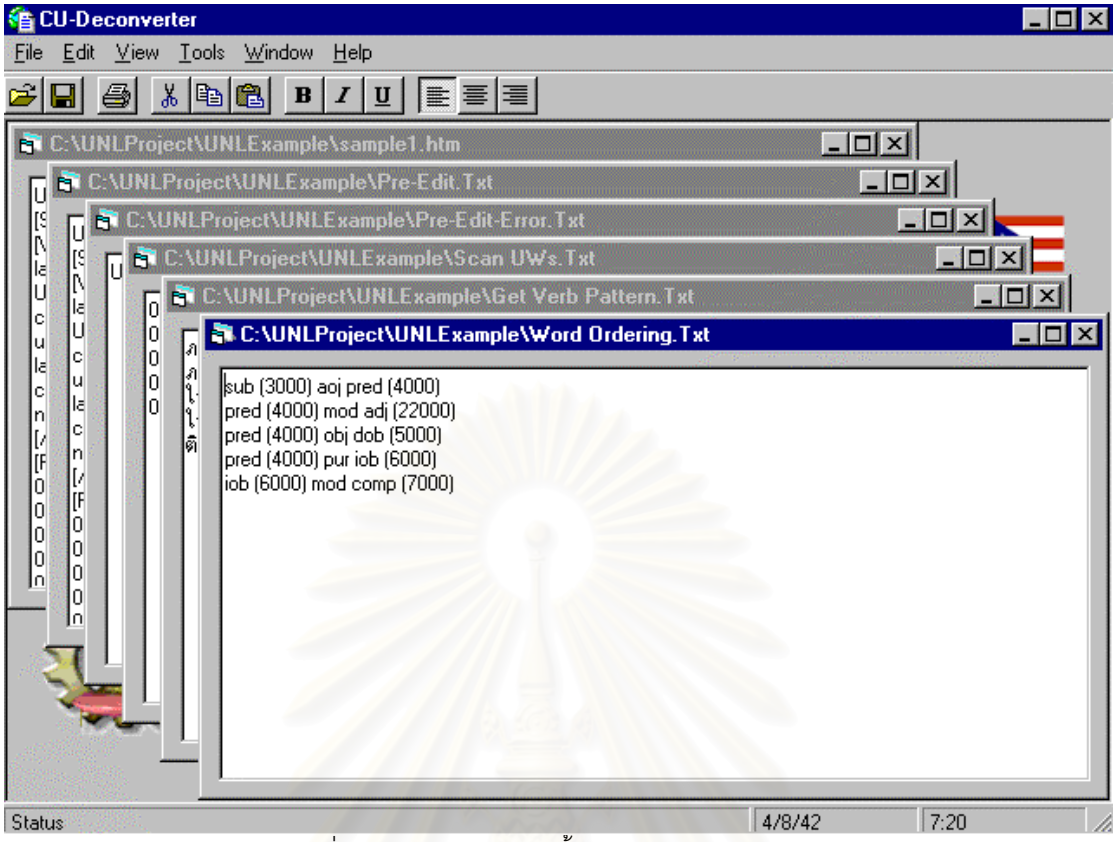


รูปที่ 7.10 หน้าจอแสดงผล Error ขั้นตอน Pre-Edit

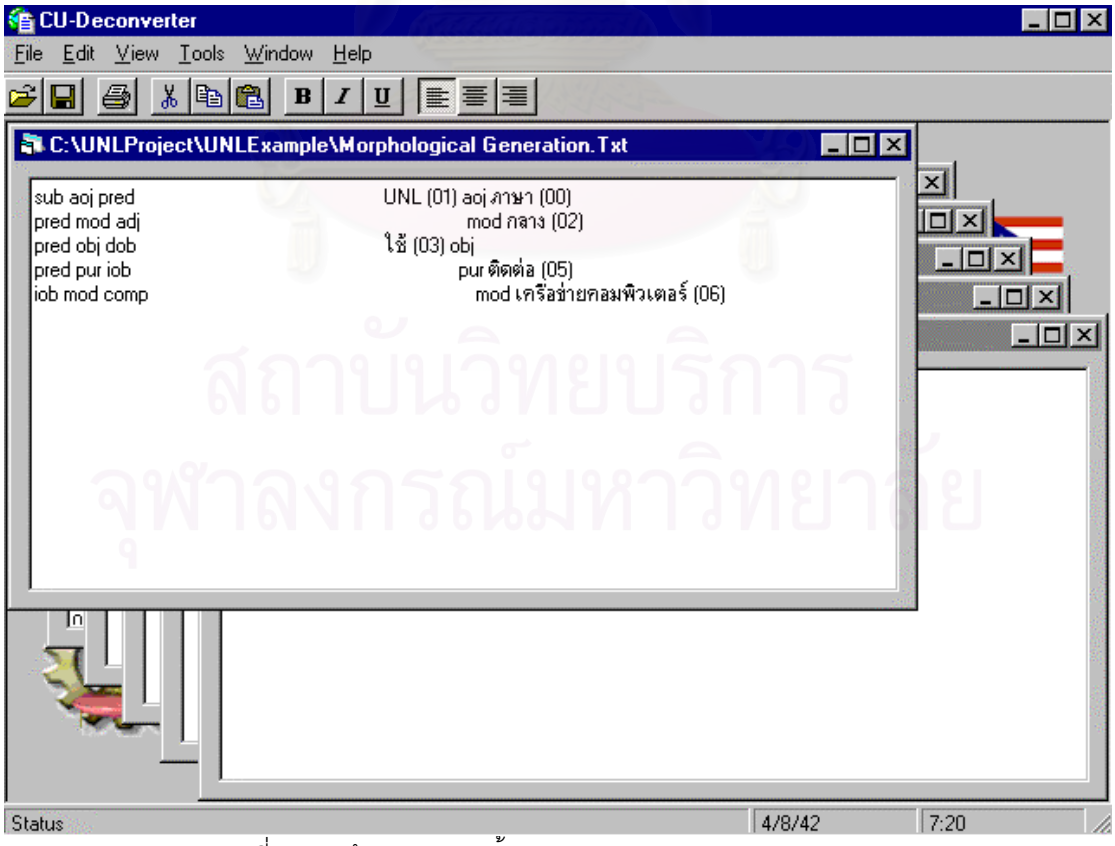
รูปที่ 7.11 หน้าจอแสดงผลขั้นตอน Scan UWs From Dictionary



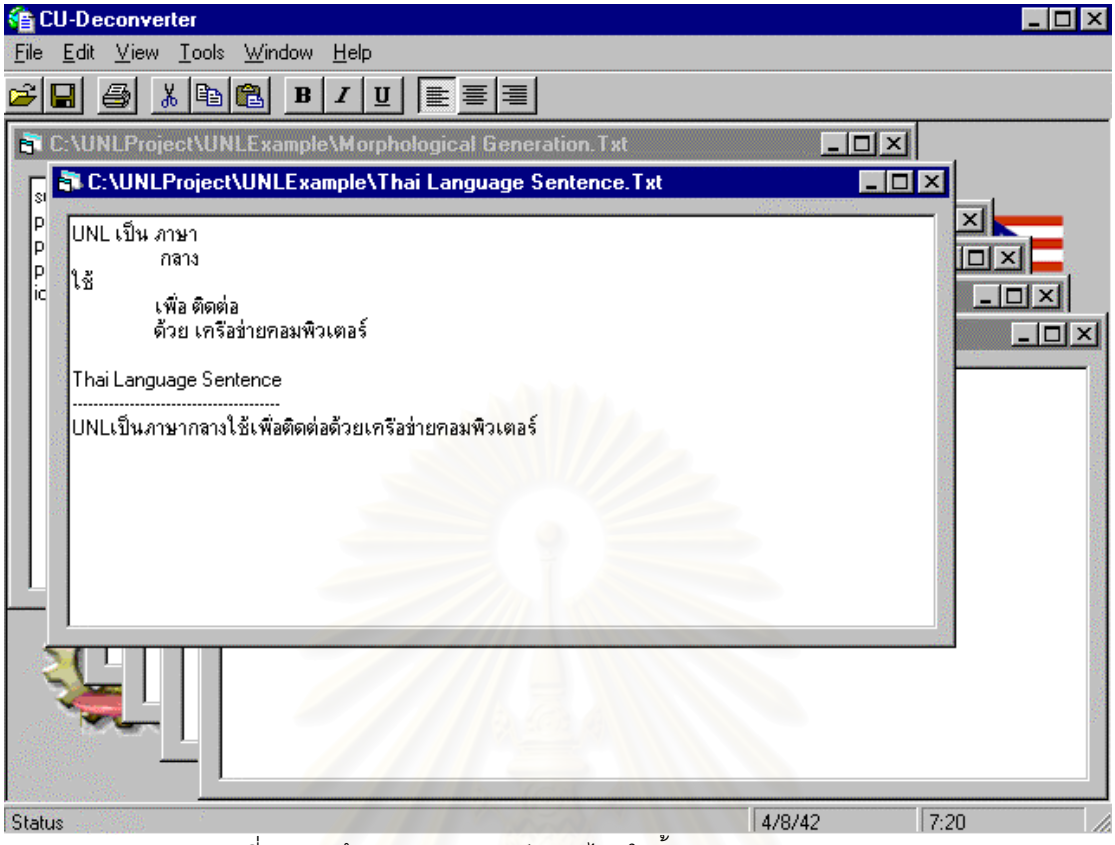
รูปที่ 7.12 หน้าจอแสดงผลขั้นตอน Get Verb Pattern



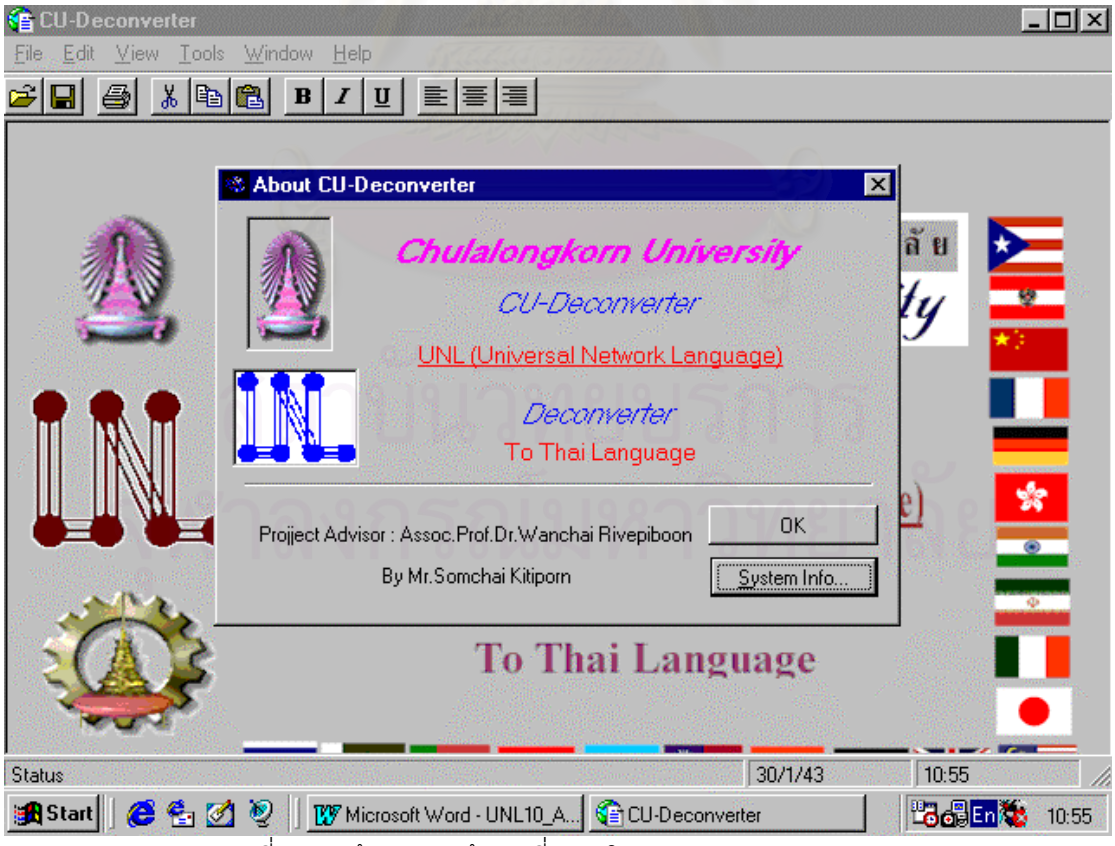
รูปที่ 7.13 หน้าจอแสดงผลขั้นตอน Word Ordering



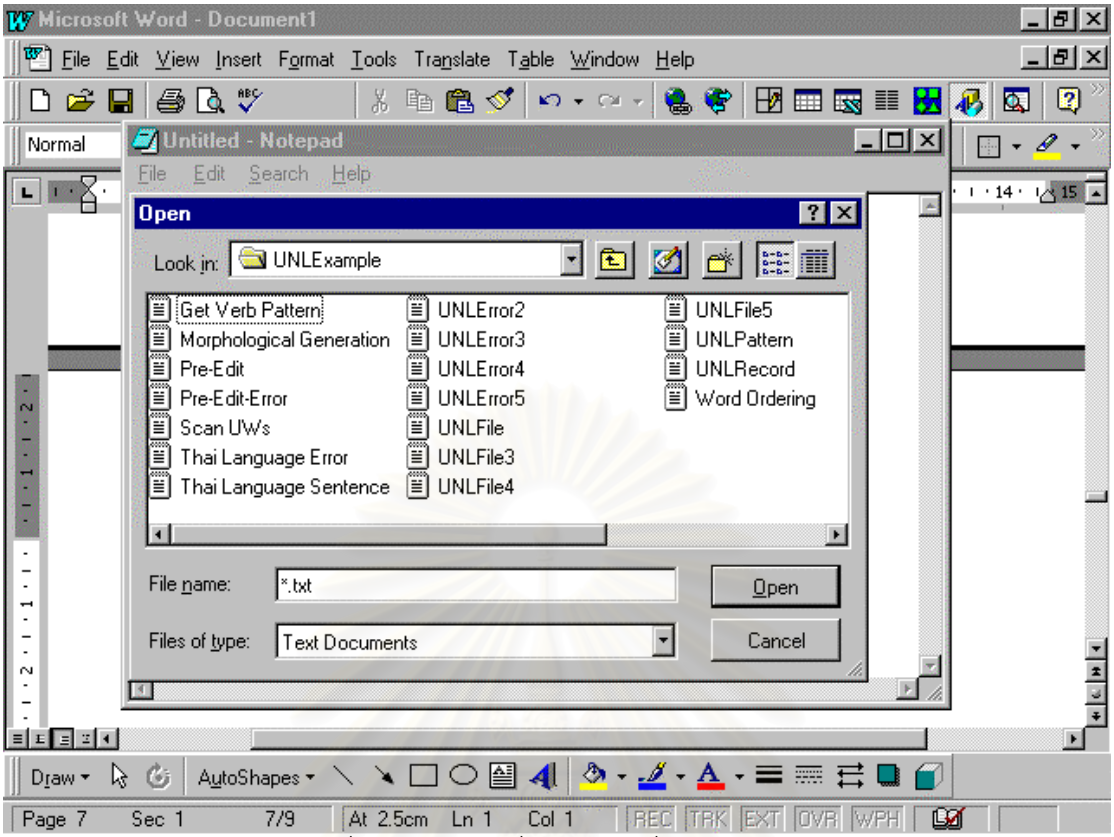
รูปที่ 7.14 หน้าจอแสดงผลขั้นตอน Morphological Generation



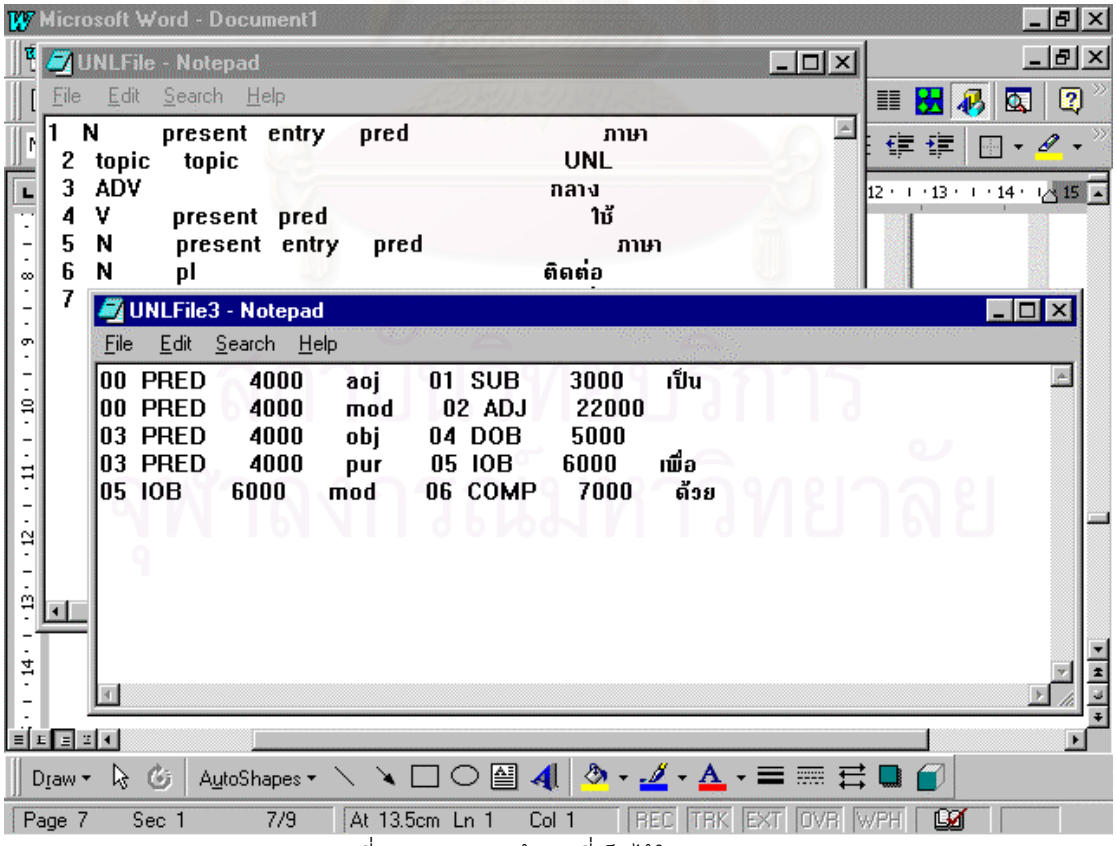
รูปที่ 7.15 หน้าจอแสดงผลลัพธ์ภาษาไทยในขั้นตอน Thai Language



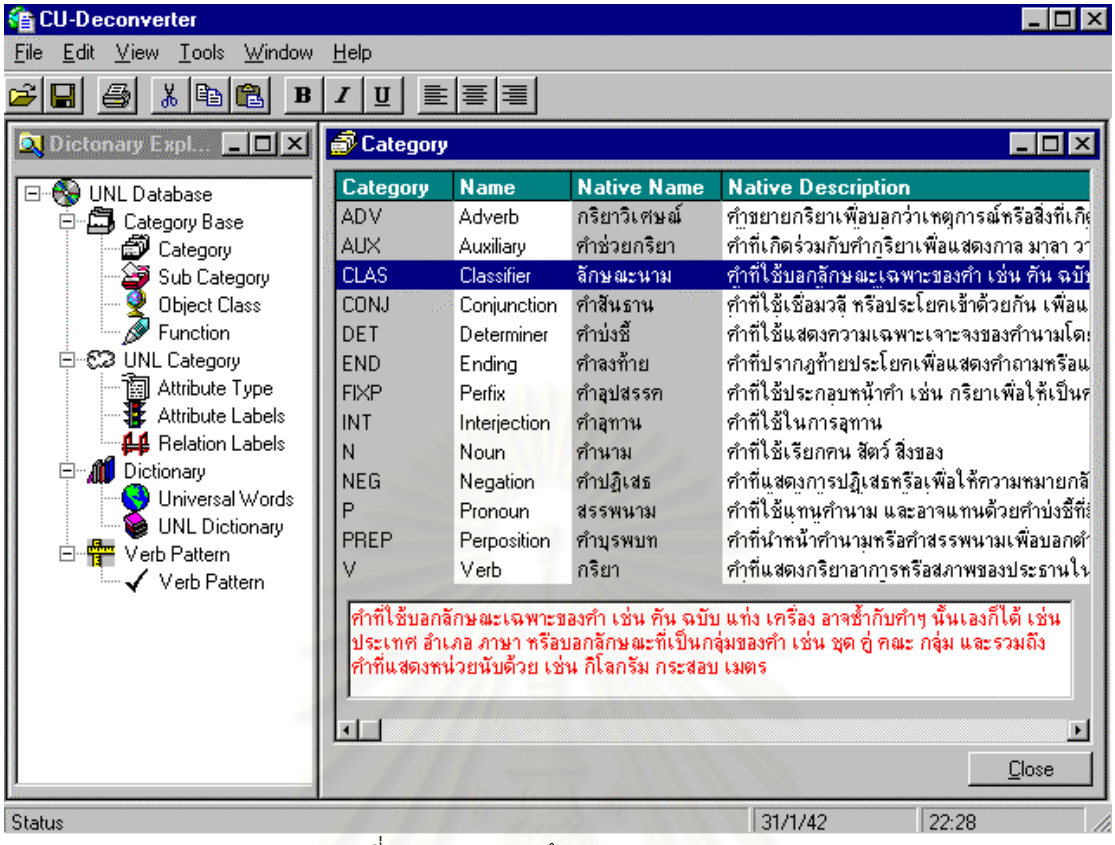
รูปที่ 7.16 หน้าจอแสดงข้อมูลเกี่ยวกับโปรแกรม CU-Deconverter



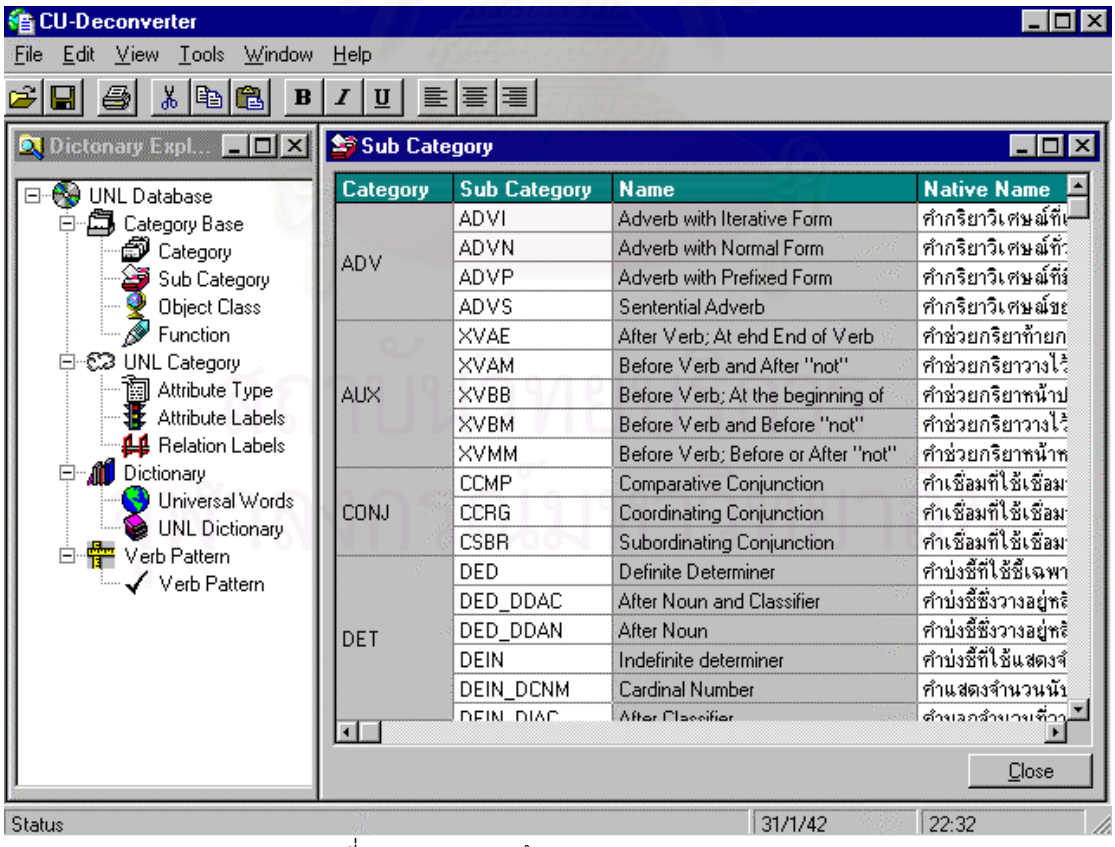
รูปที่ 7.17 แสดงรายชื่อ Text File ที่ใช้ในโปรแกรม



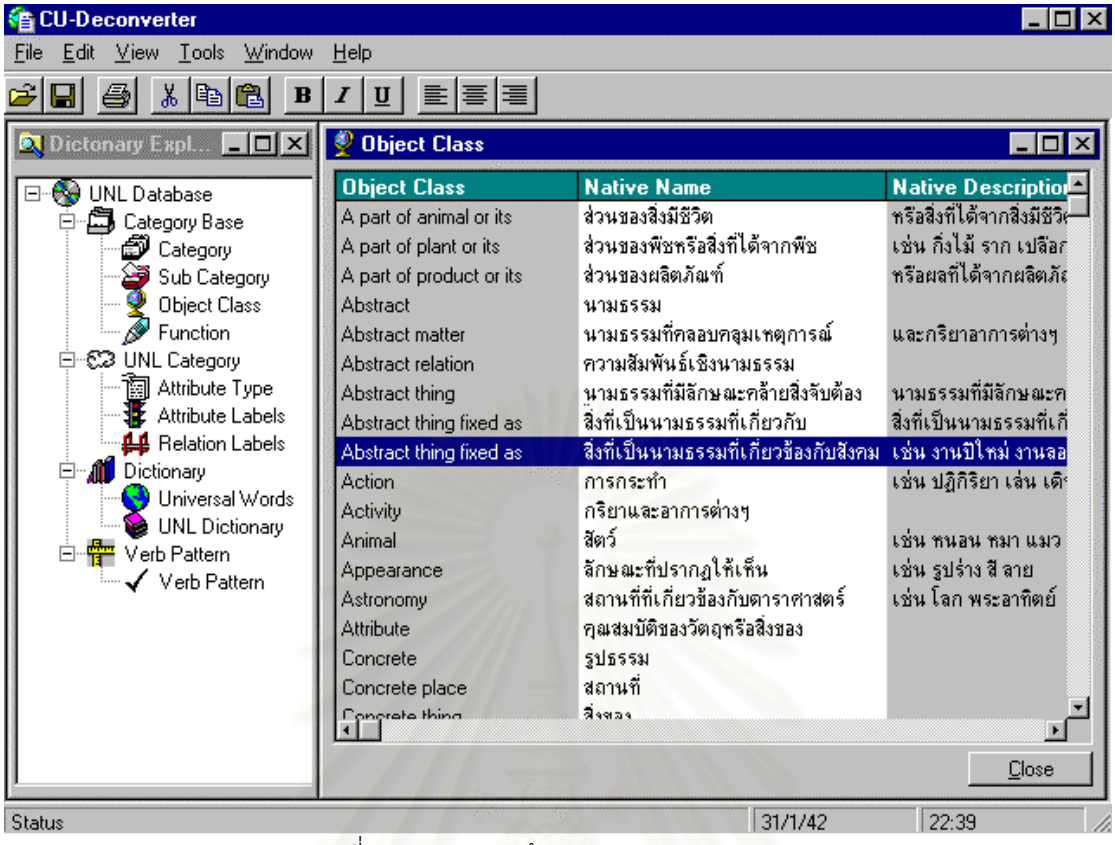
รูปที่ 7.18 ตัวอย่างข้อมูลที่เก็บไว้ใน Text File



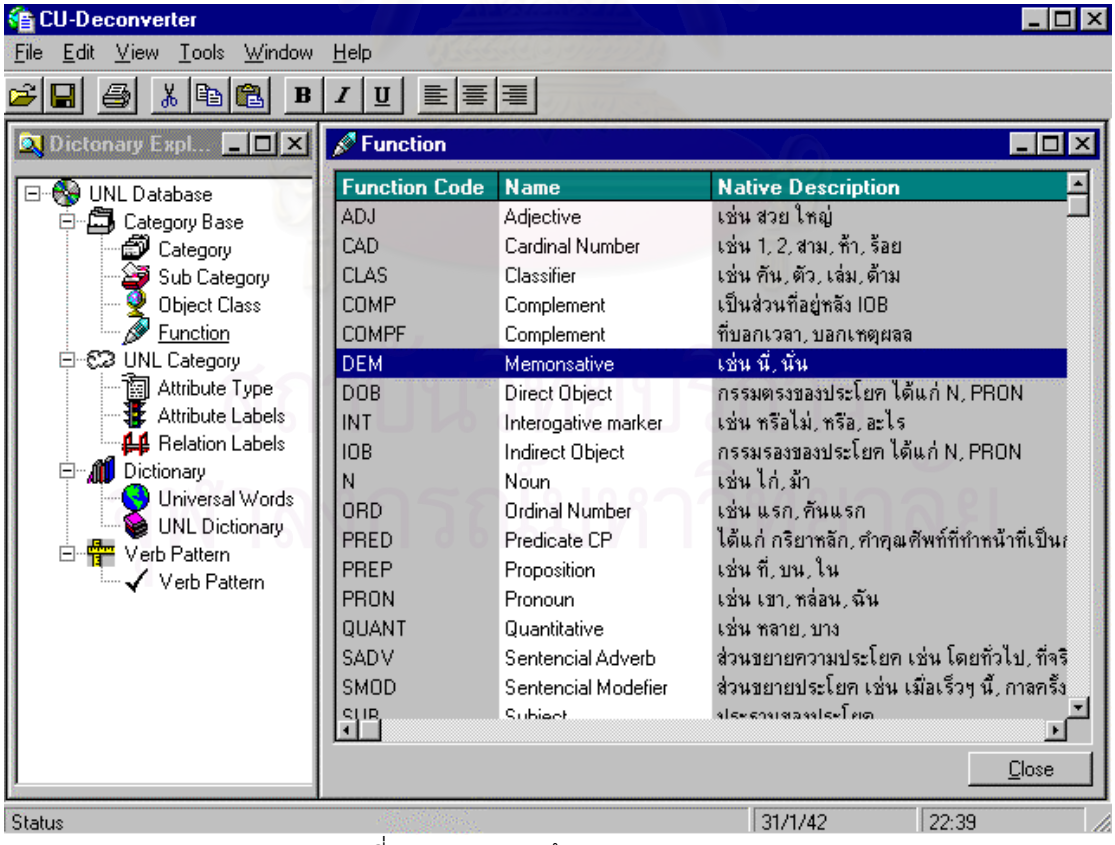
รูปที่ 7.19 ตัวอย่างหน้าจอ Table Category



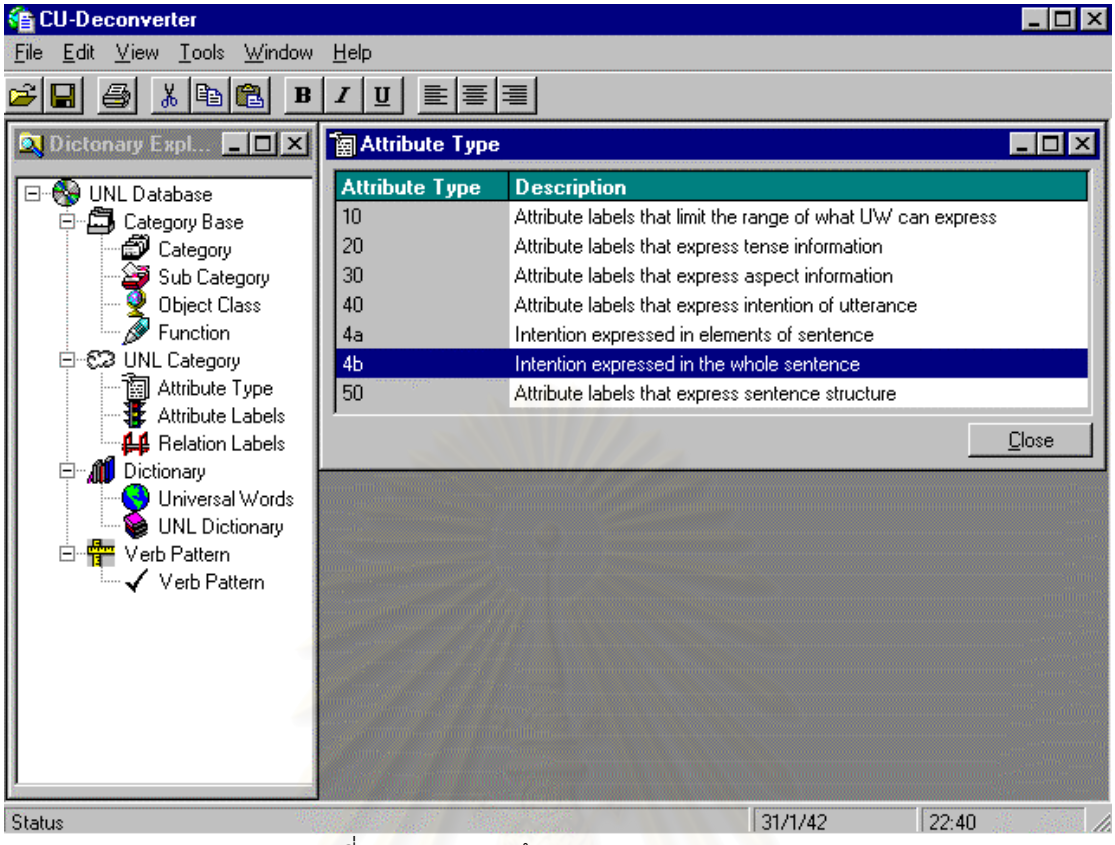
รูปที่ 7.20 ตัวอย่างหน้าจอ Table Sub Category



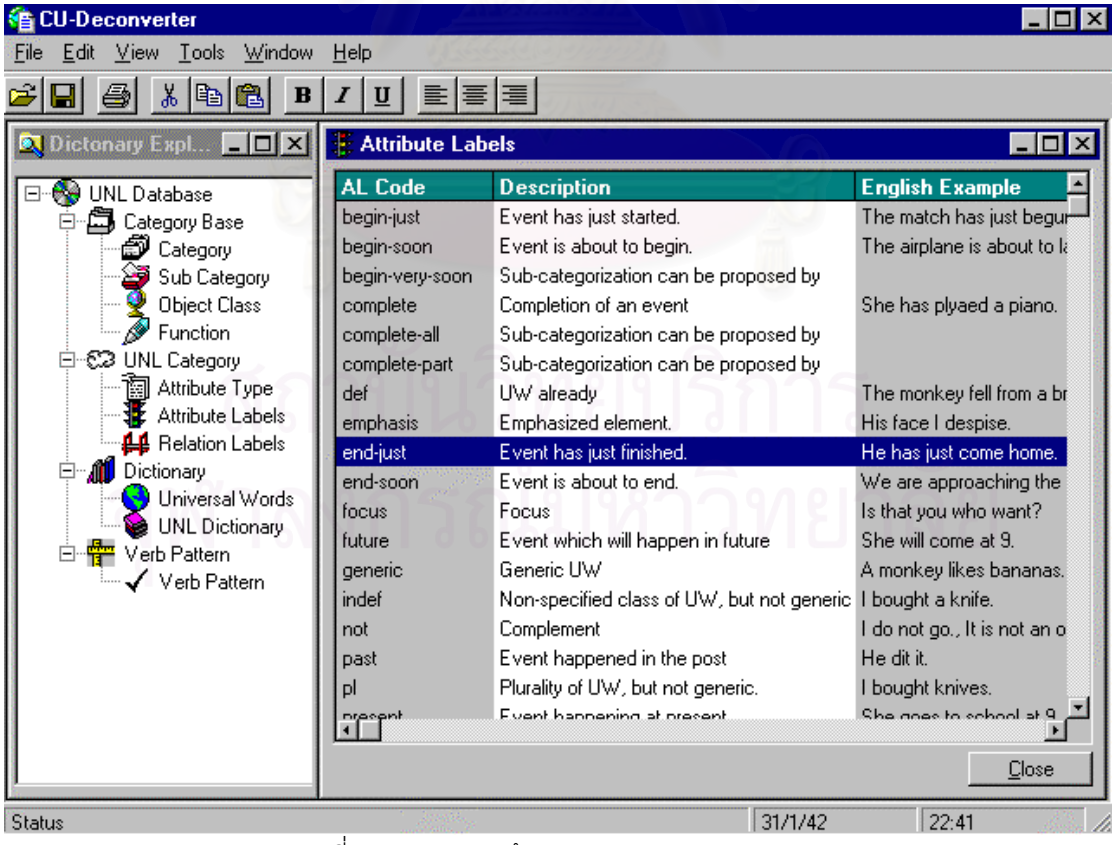
รูปที่ 7.21 ตัวอย่างหน้าจอ Table Object Class



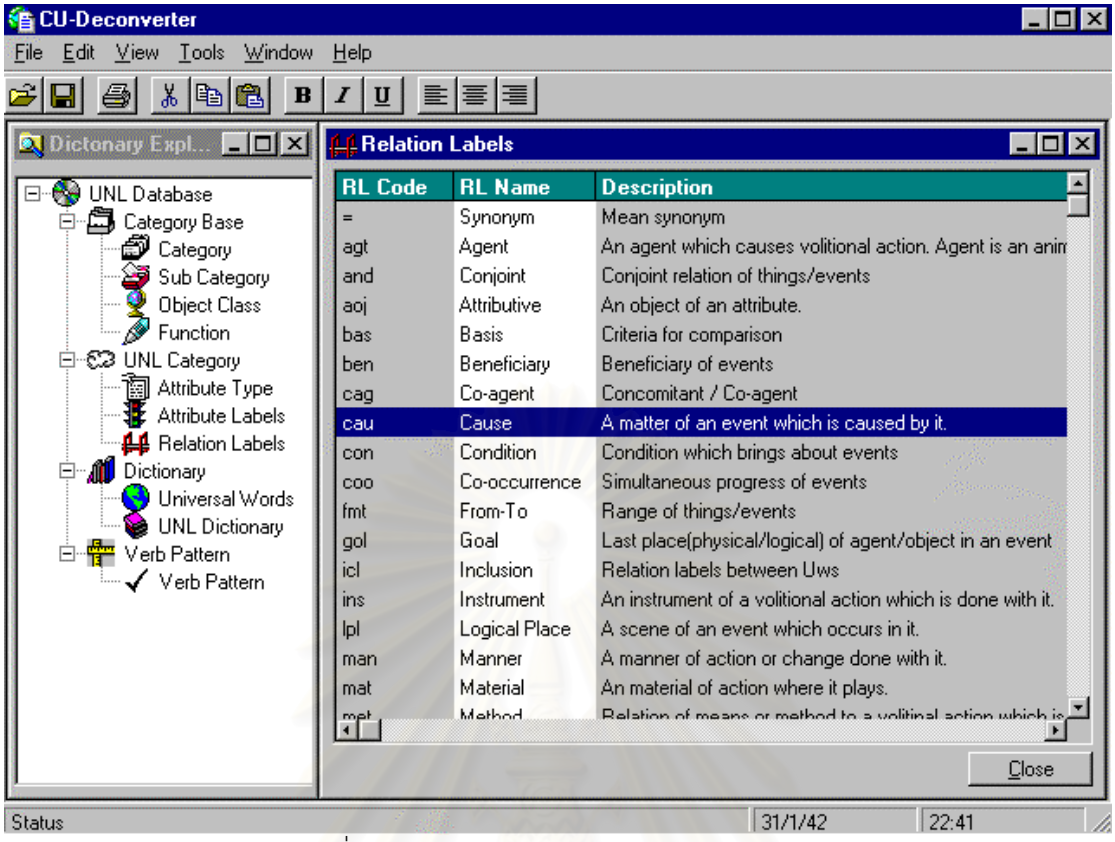
รูปที่ 7.22 ตัวอย่างหน้าจอ Table Function



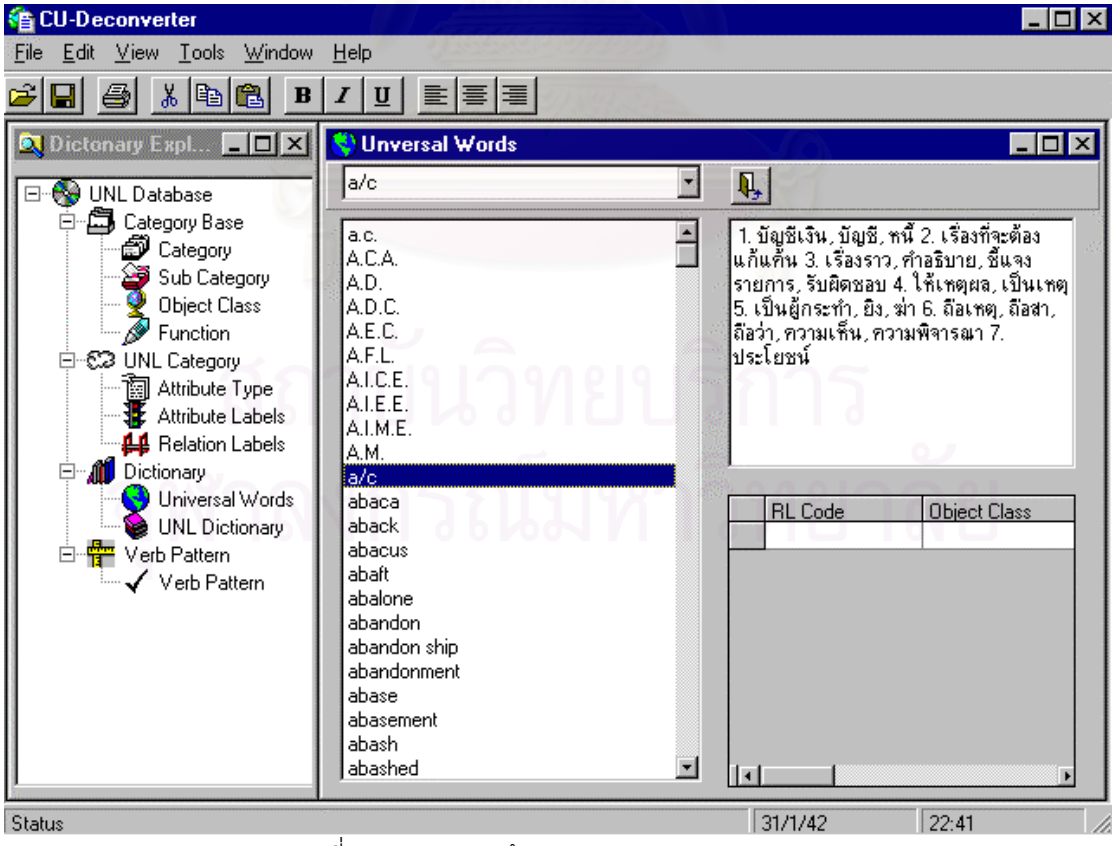
รูปที่ 7.23 ตัวอย่างหน้าจอ Table Attribute Type



รูปที่ 7.24 ตัวอย่างหน้าจอ Table Attribute Labels



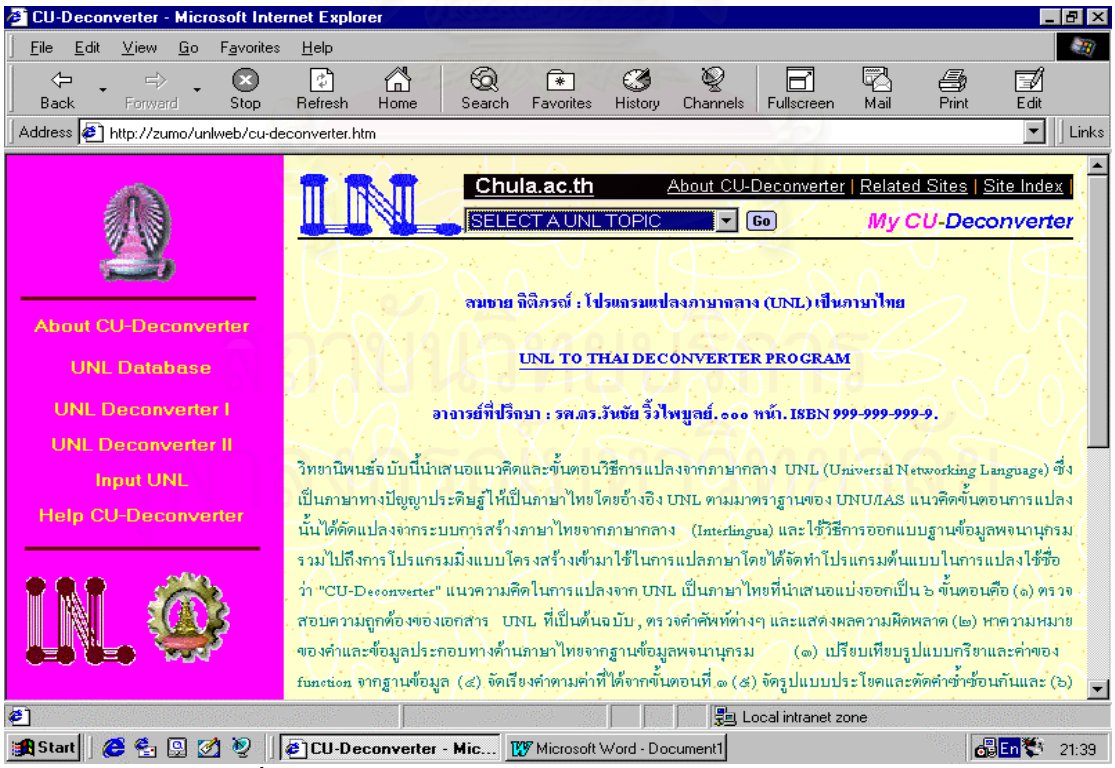
รูปที่ 7.25 ตัวอย่างหน้าจอ Table Relation Labels



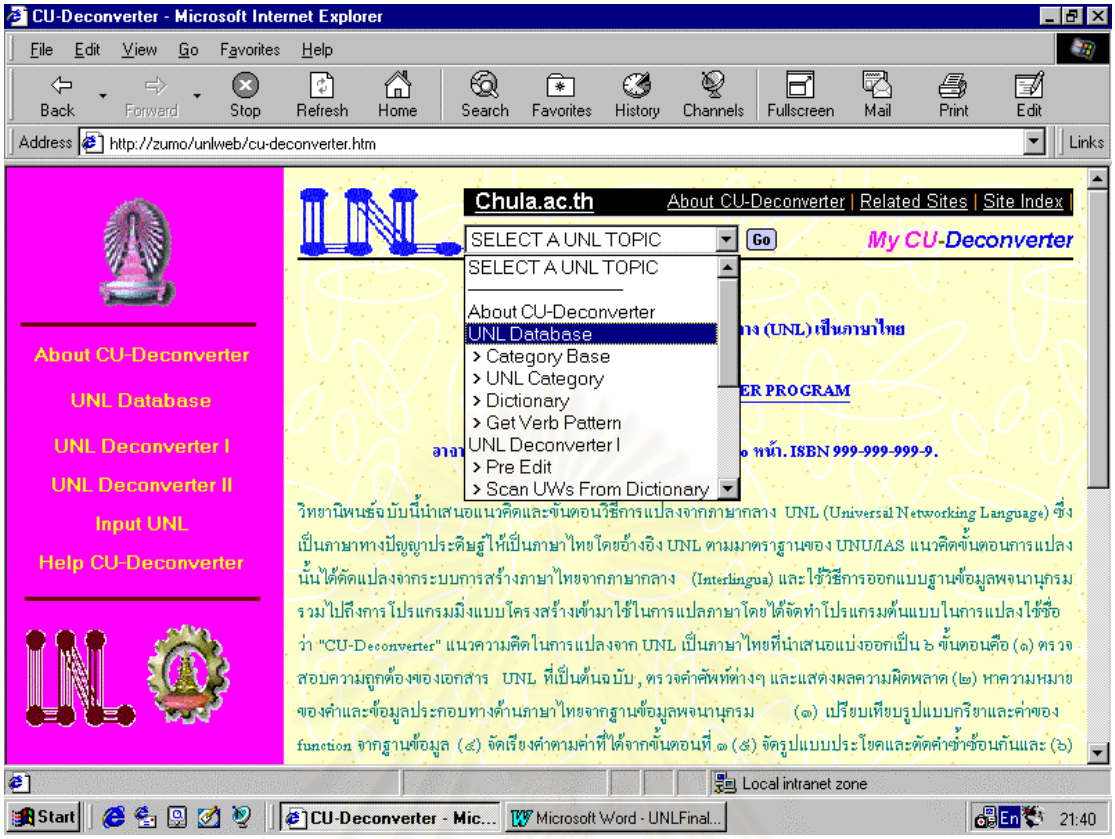
รูปที่ 7.26 ตัวอย่างหน้าจอ Table Universal Words

2. ตัวอย่างหน้าจอติดต่อแบบเวบ

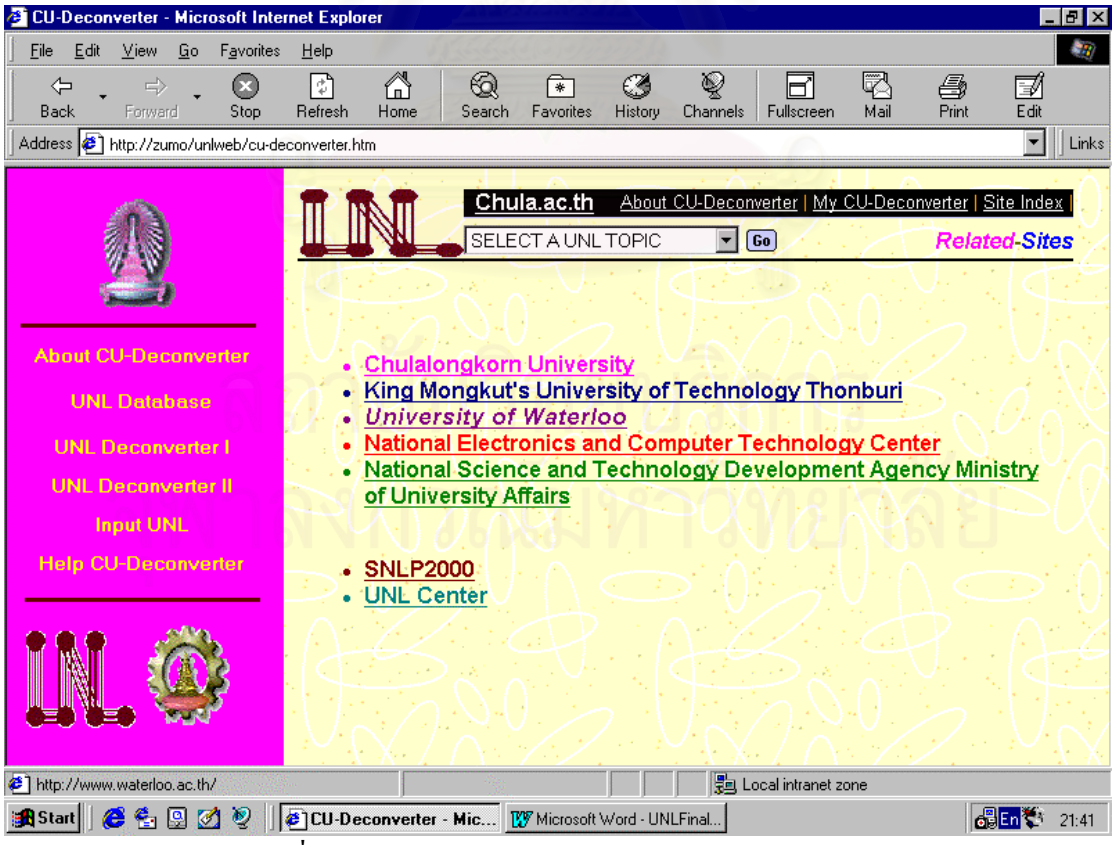
รูปที่ 7.27 ตัวอย่างหน้าจอ About CU-Deconverter Web Page



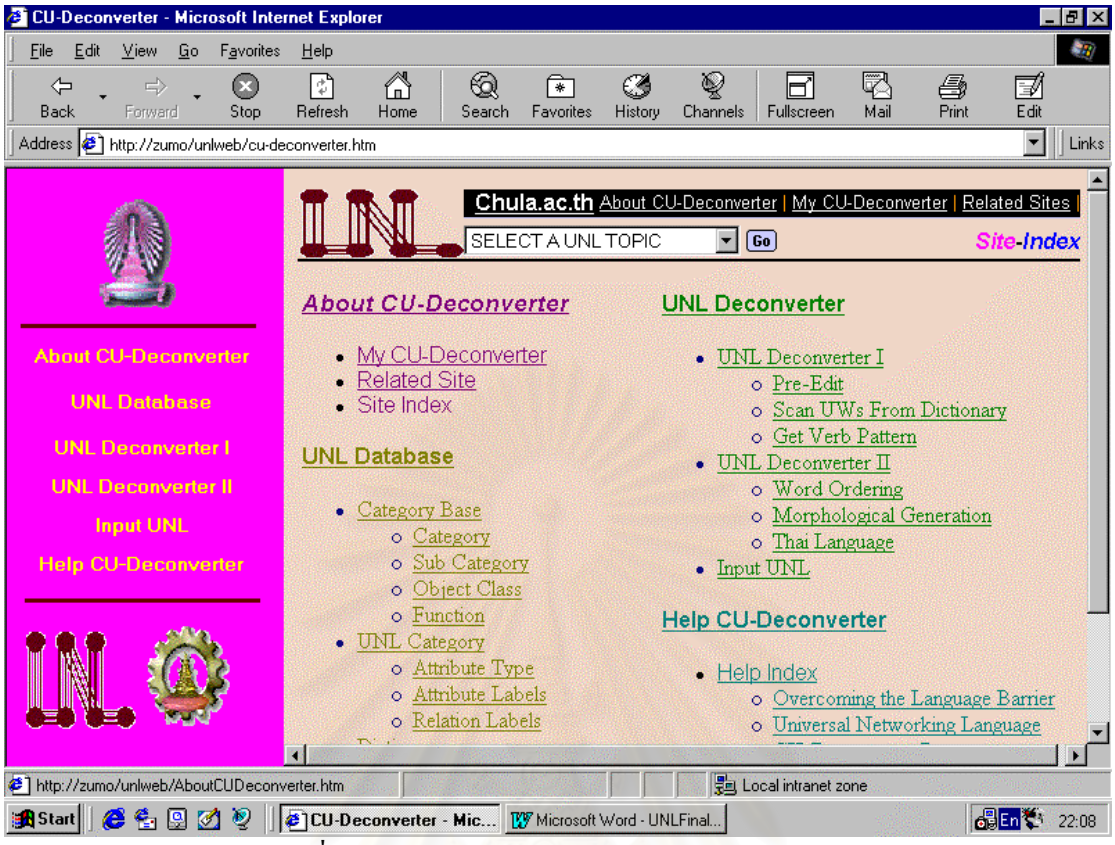
รูปที่ 7.28 ตัวอย่างหน้าจอ My CU-Deconverter Web Page



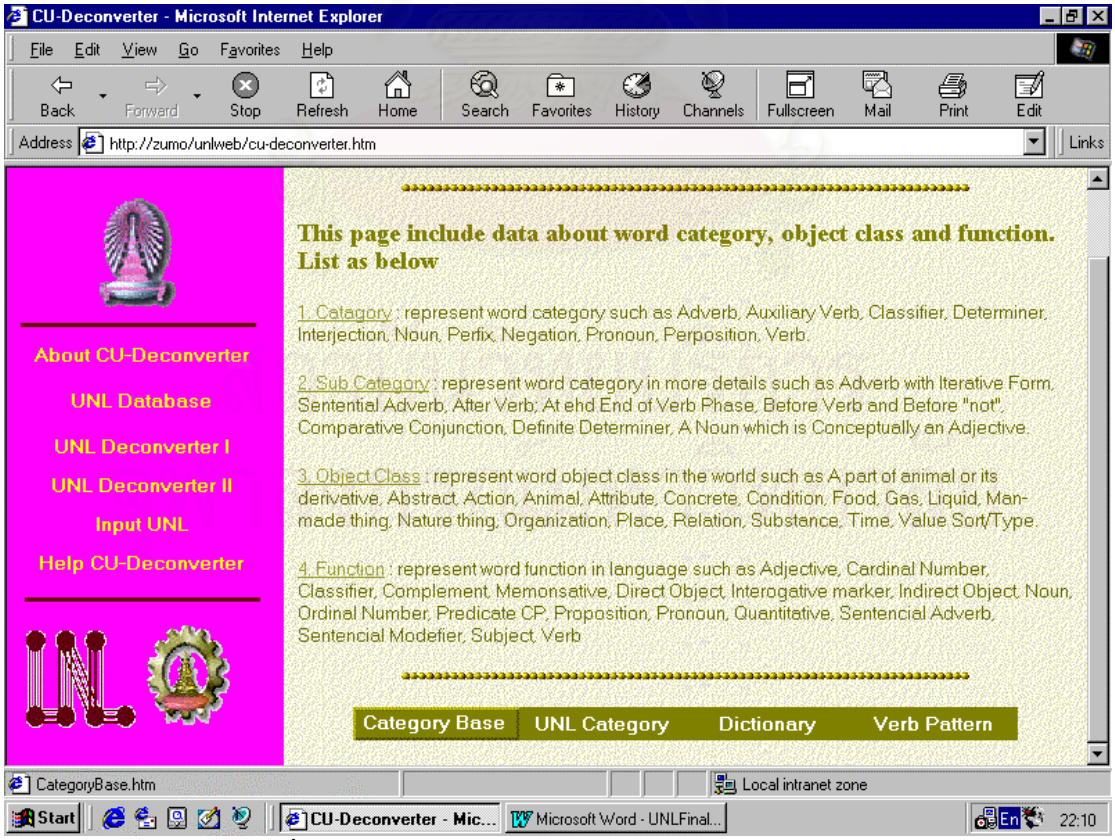
รูปที่ 7.29 ตัวอย่างหน้าจอ Select A UNL Topic Web Page



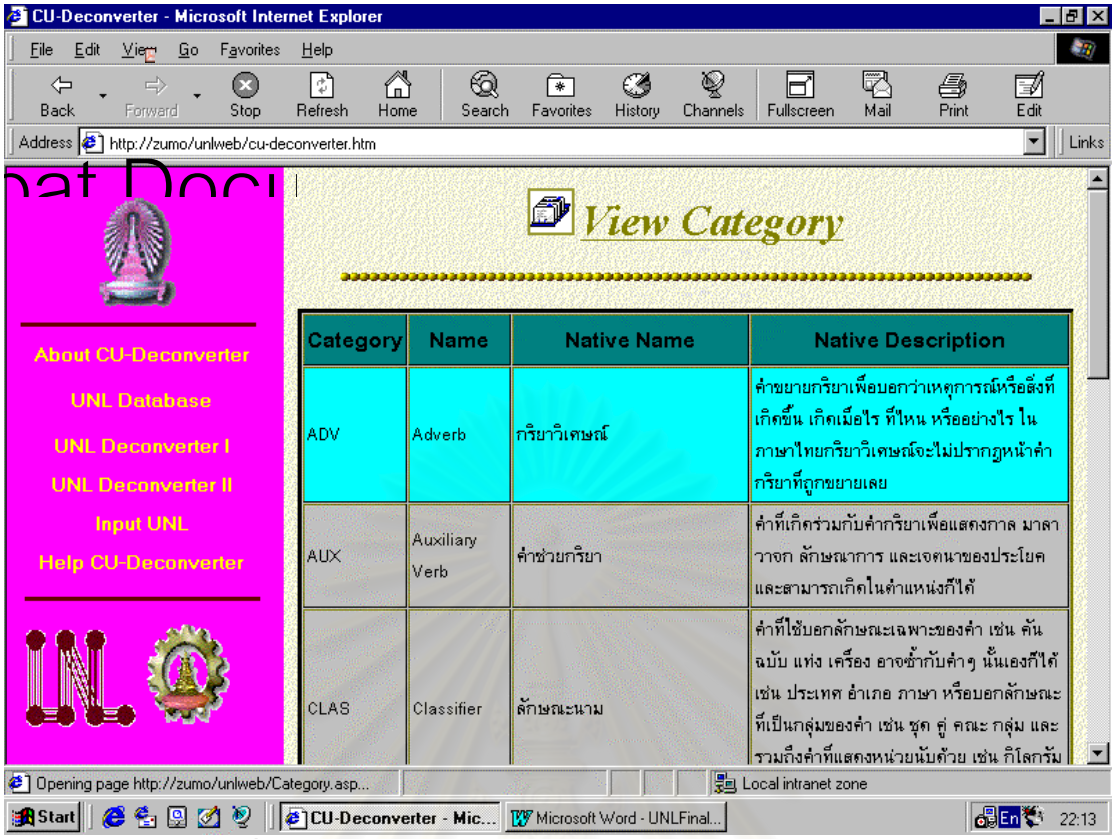
รูปที่ 7.30 ตัวอย่างหน้าจอ Related-Sites Web Page



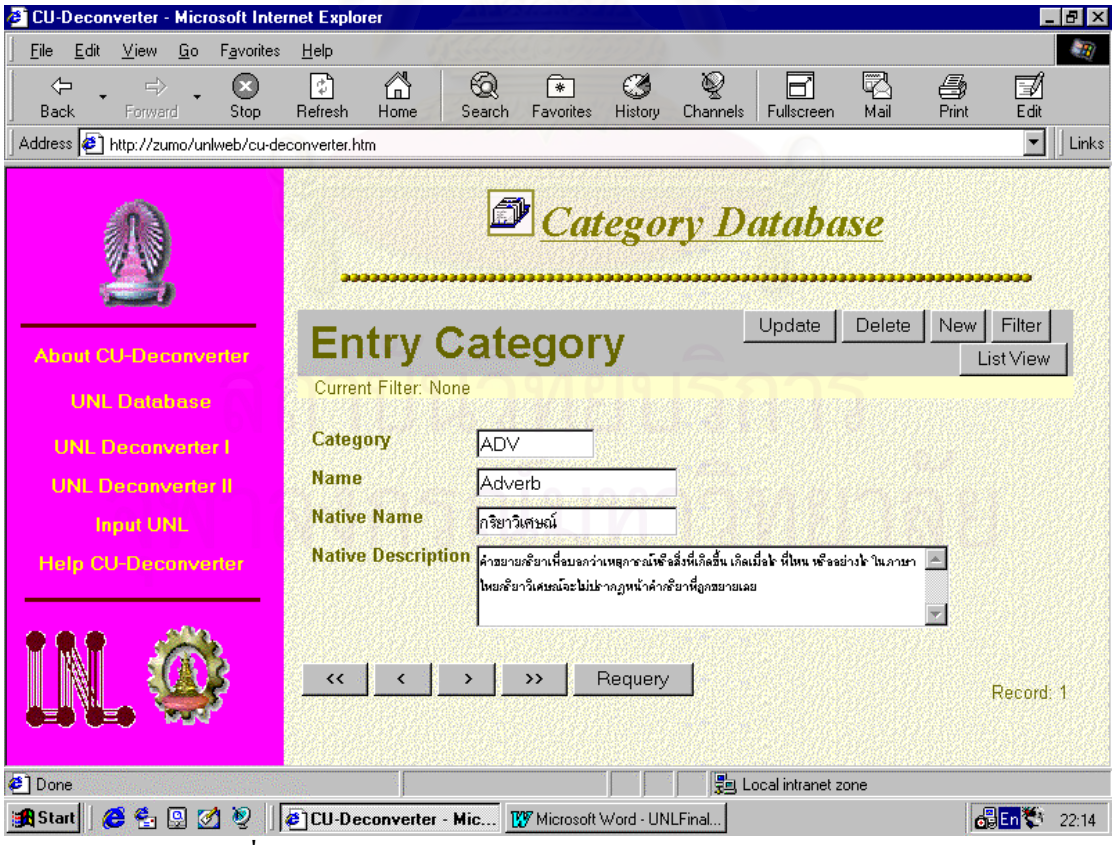
รูปที่ 7.31 ตัวอย่างหน้าจอ Site-Index Web Page



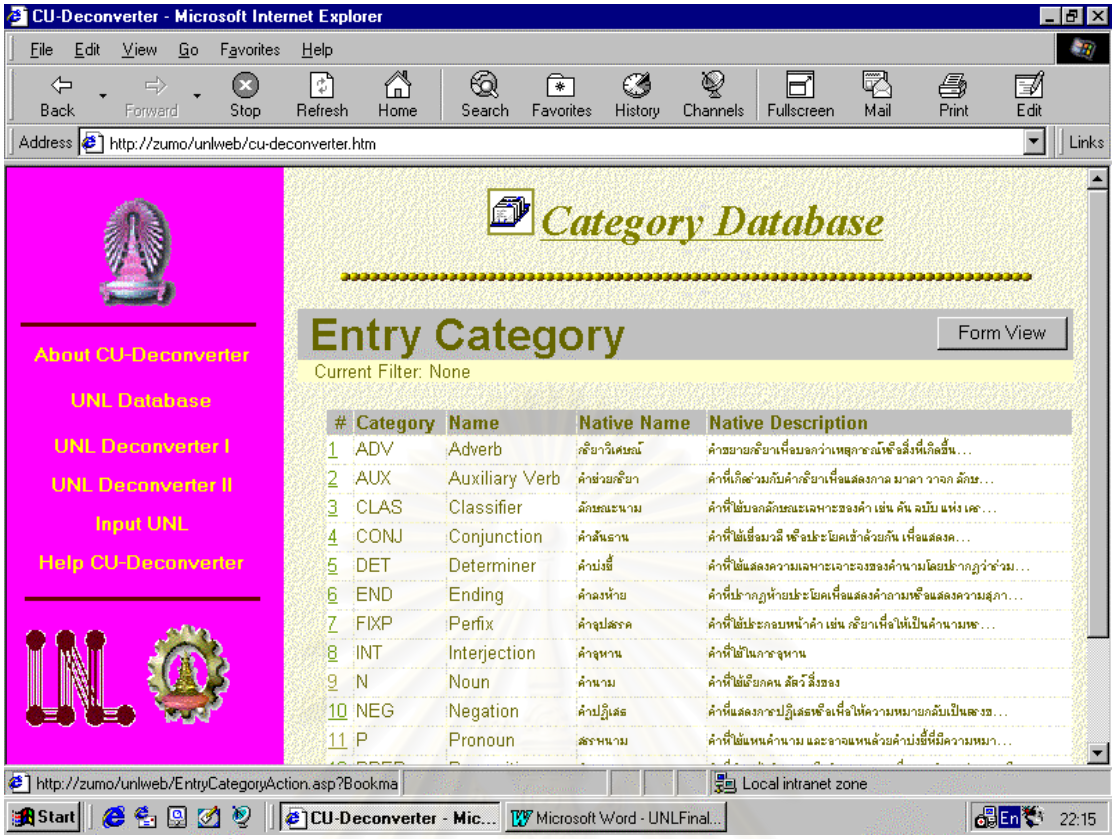
รูปที่ 7.32 ตัวอย่างหน้าจอ UNL Database Web Page



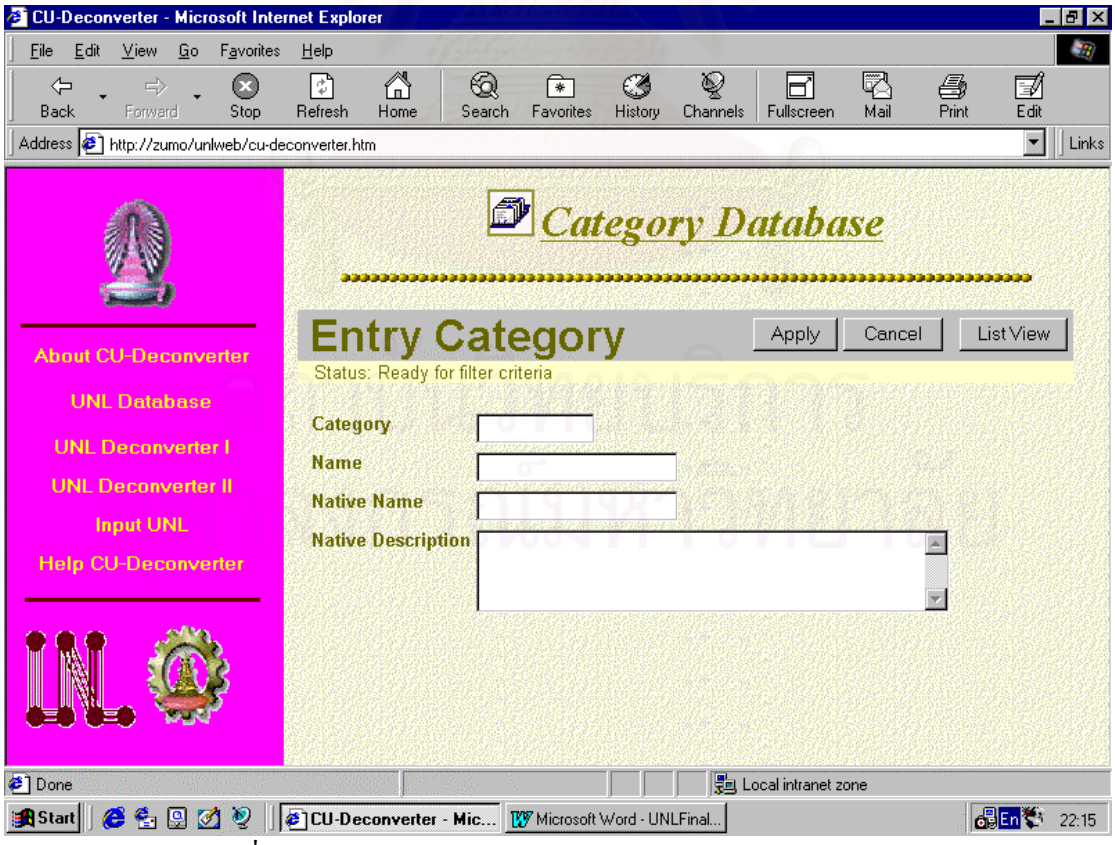
รูปที่ 7.33 ตัวอย่างหน้าจอ View Category Table Web Page



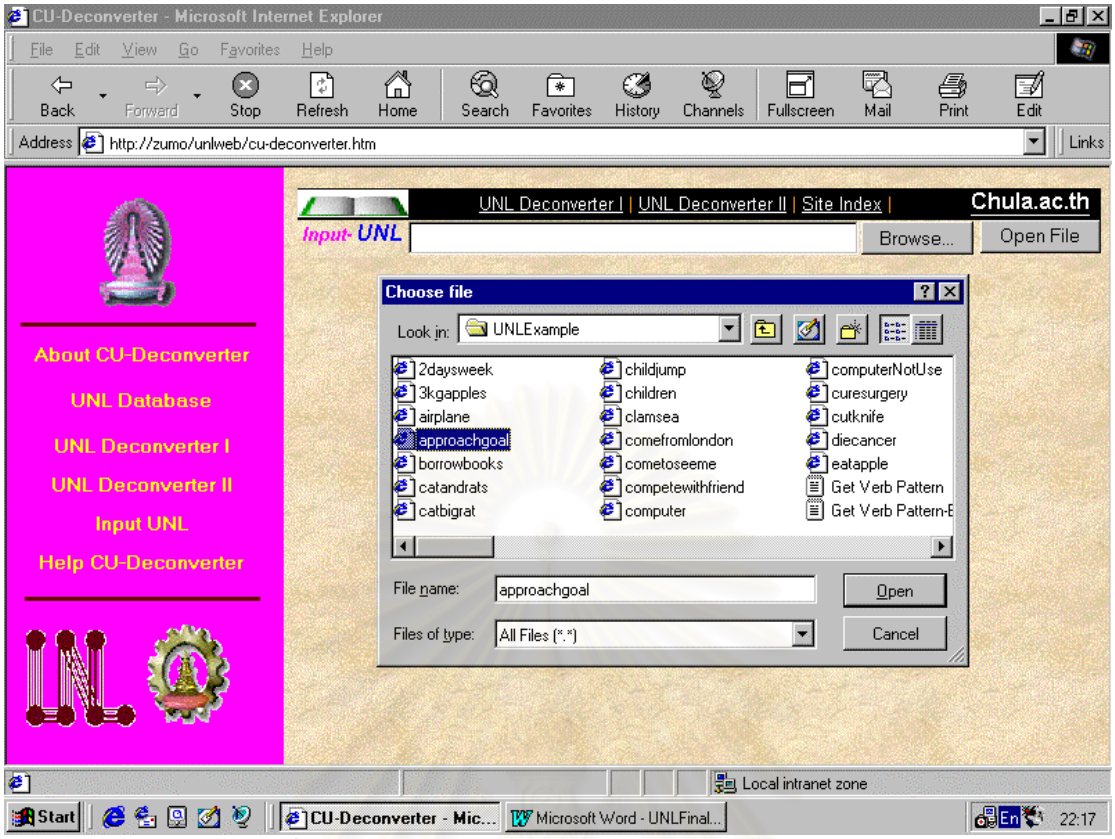
รูปที่ 7.34 ตัวอย่างหน้าจอ Entry Category Table Web Page



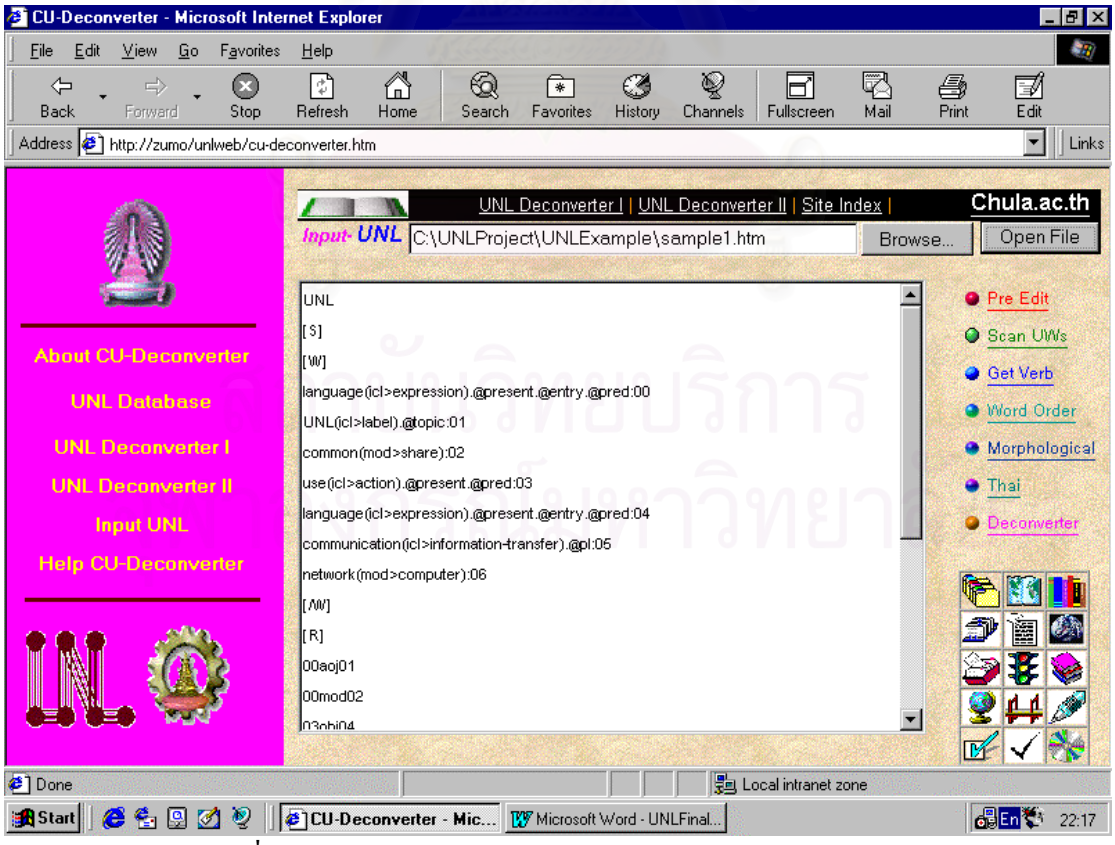
รูปที่ 7.35 ตัวอย่างหน้าจอ List View Category Table Web Page



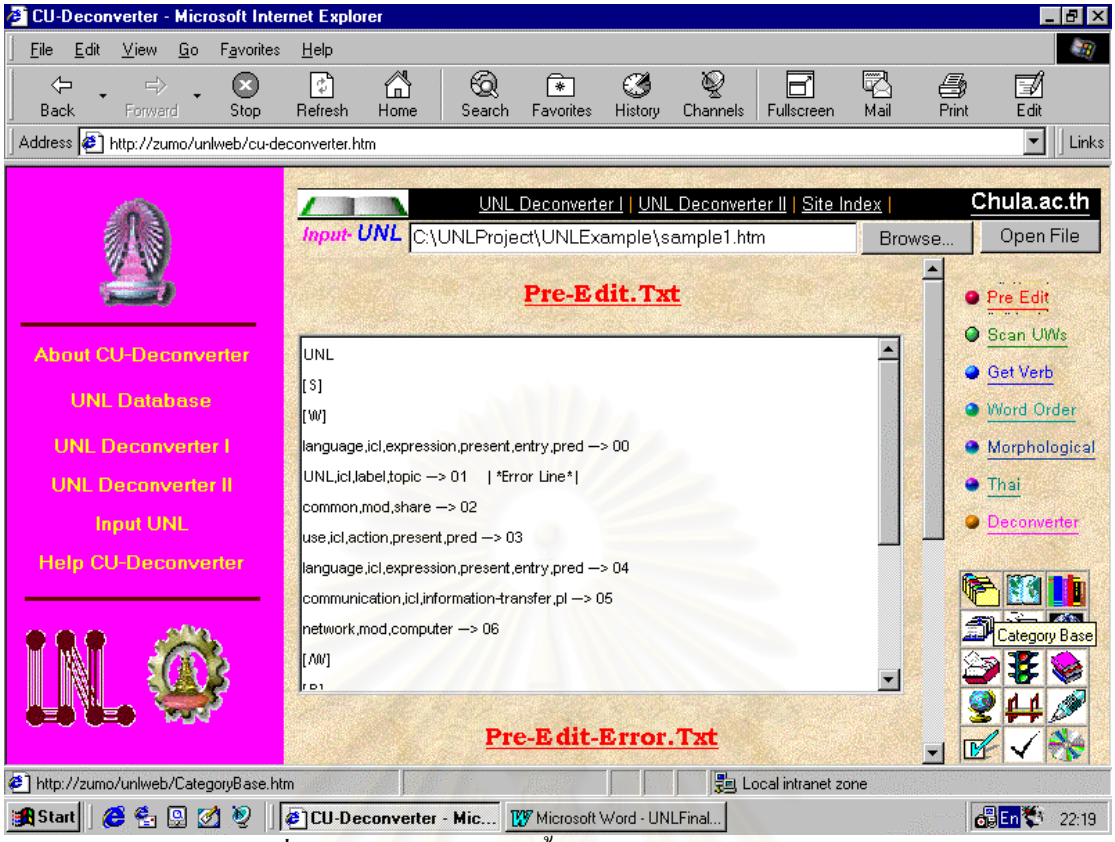
รูปที่ 7.36 ตัวอย่างหน้าจอ Filter Category Table Web Page



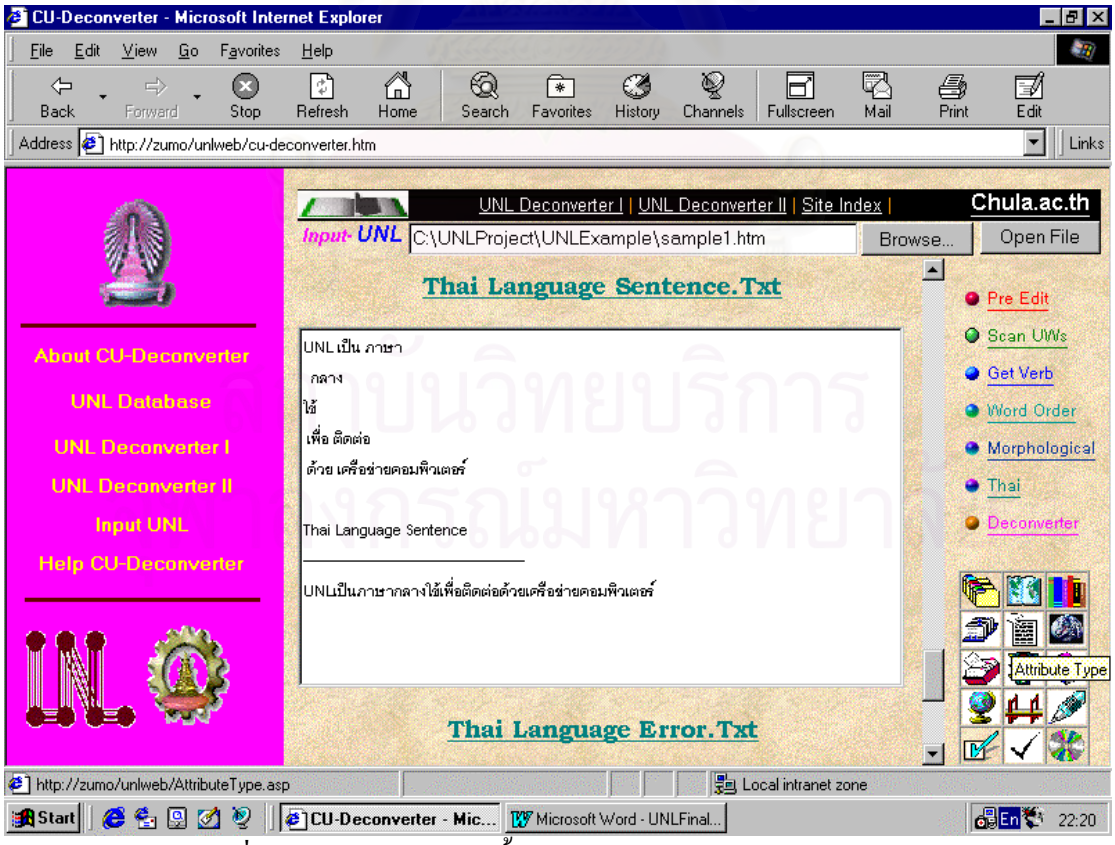
รูปที่ 7.37 ตัวอย่างหน้าจอ Input UNL File Web Page



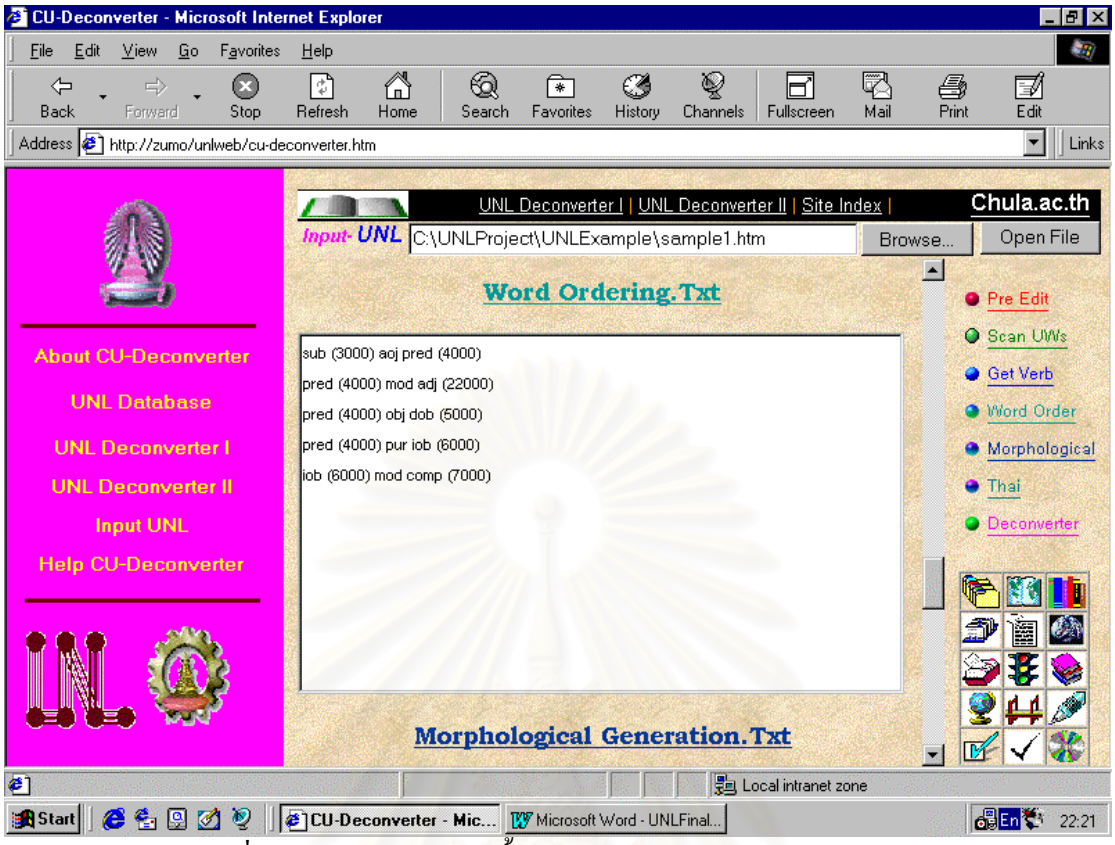
รูปที่ 7.38 ตัวอย่างหน้าจอ Open Input UNL File Web Page



รูปที่ 7.39 ตัวอย่างหน้าจอขั้นตอน Pre-Edit Web Page

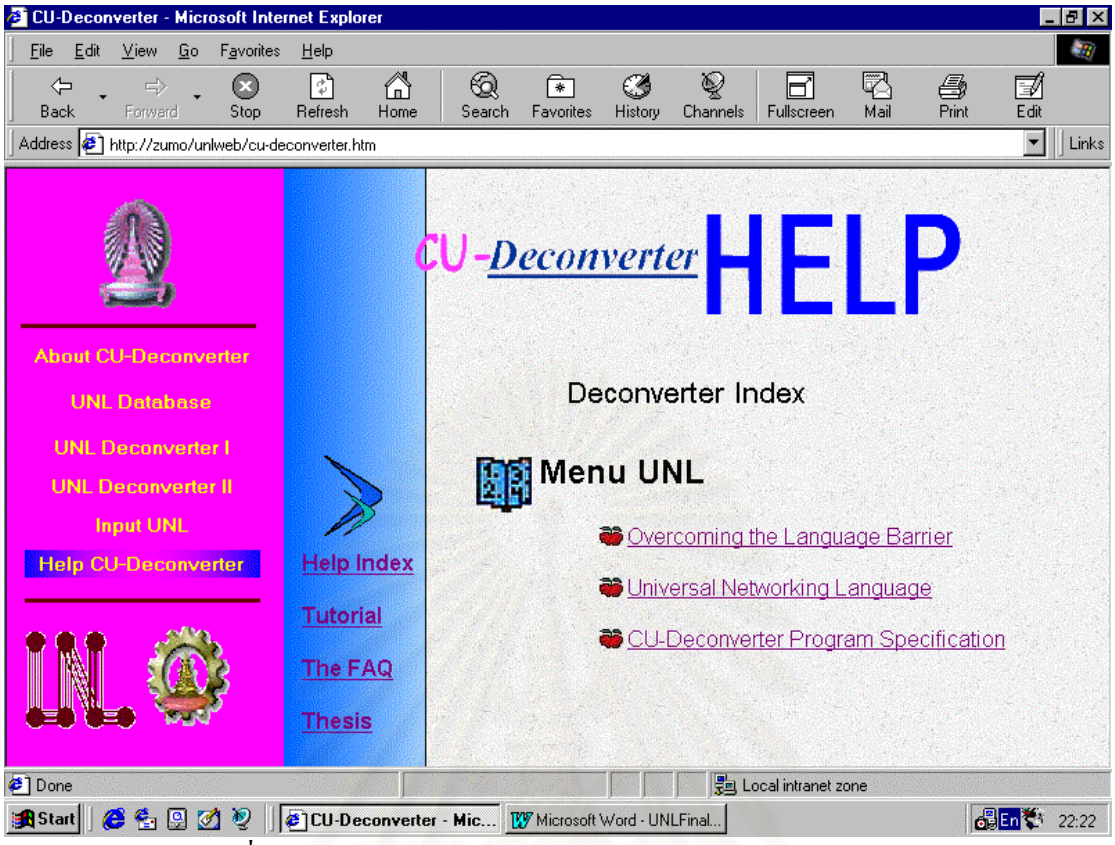


รูปที่ 7.40 ตัวอย่างหน้าจอขั้นตอน Thai Language Web Page

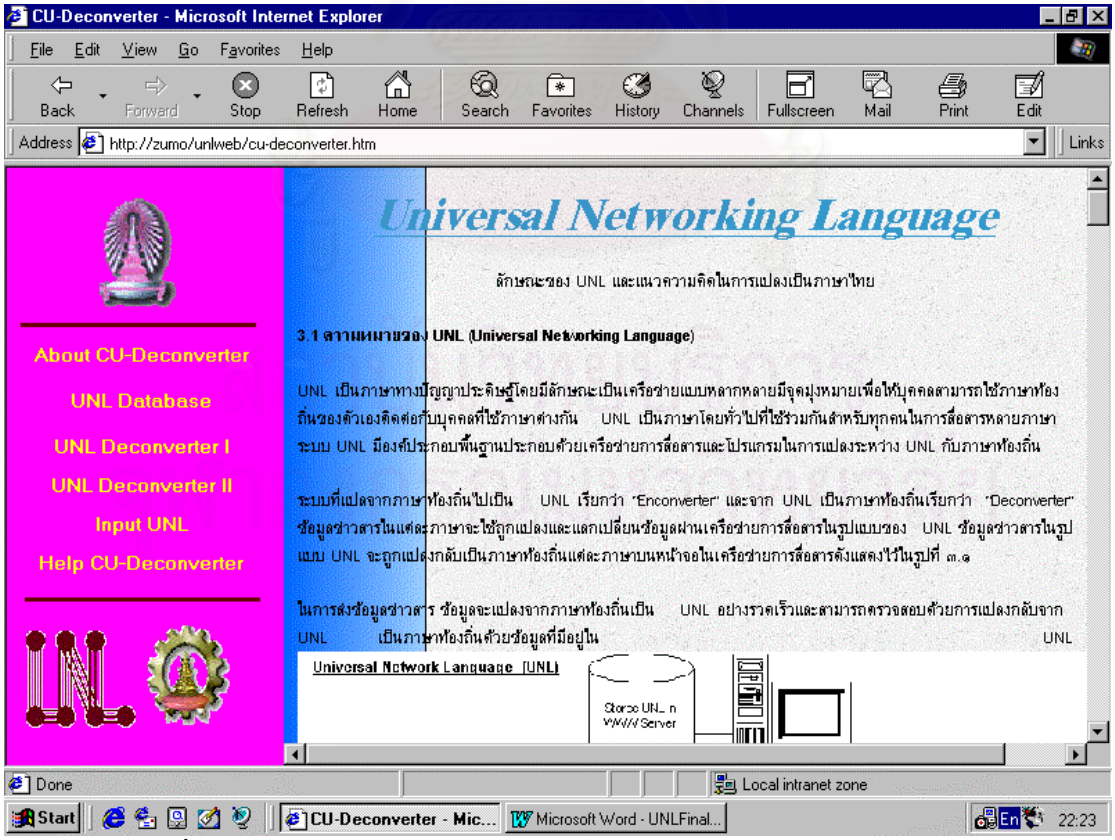


รูปที่ 7.41 ตัวอย่างหน้าจอขั้นตอน Word Ordering Web Page

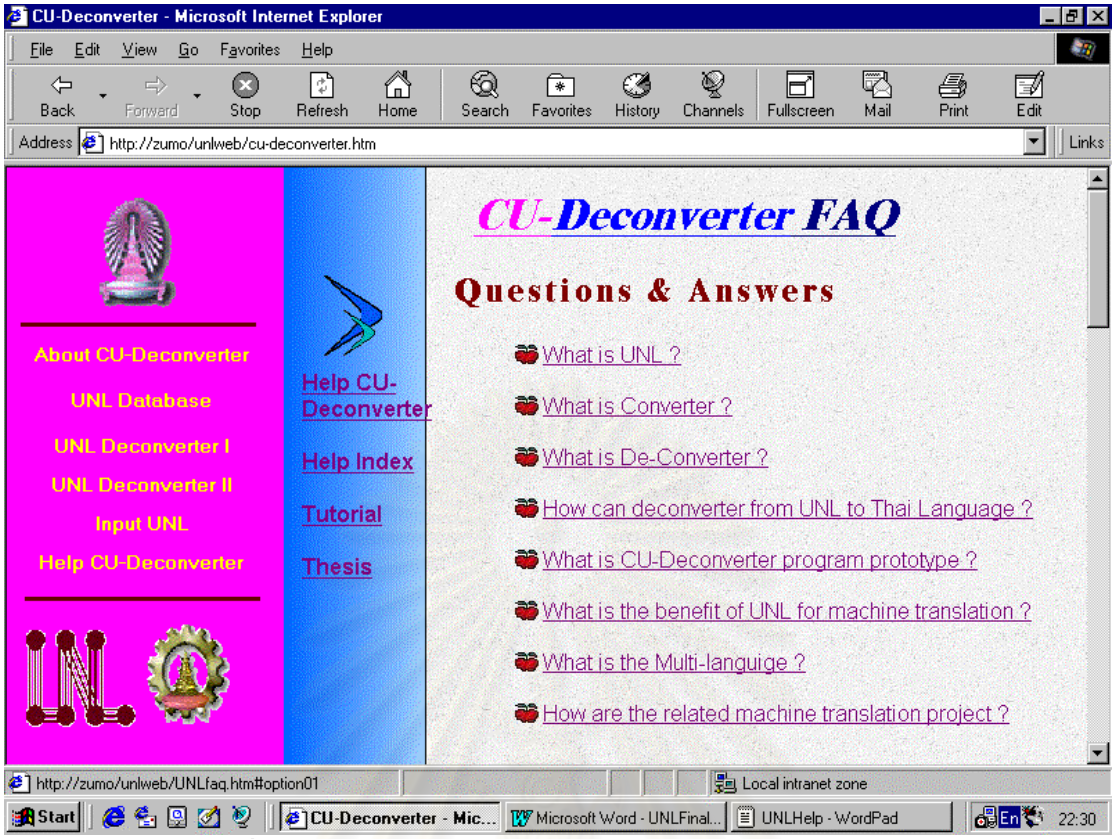
รูปที่ 7.42 ตัวอย่างหน้าจอขั้นตอน Morphological Generation Web Page



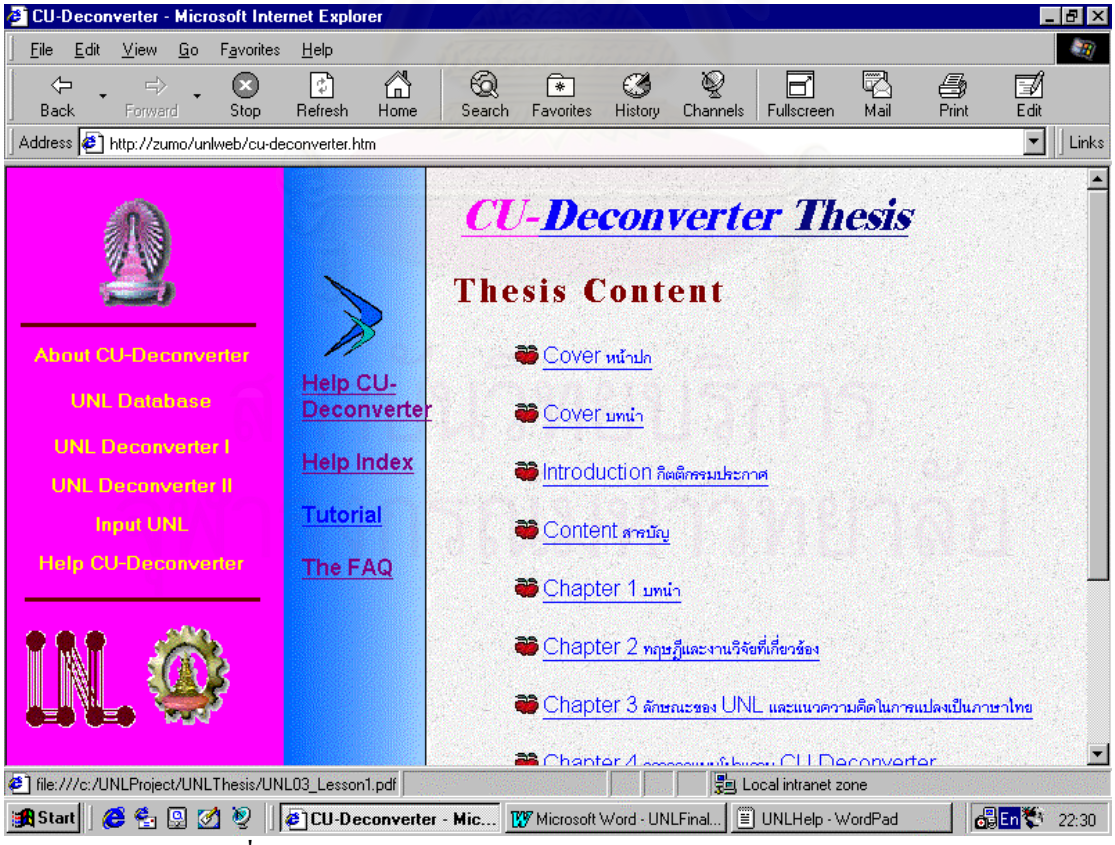
รูปที่ 7.43 ตัวอย่างหน้าจอ CU-Deconverter Help Web Page



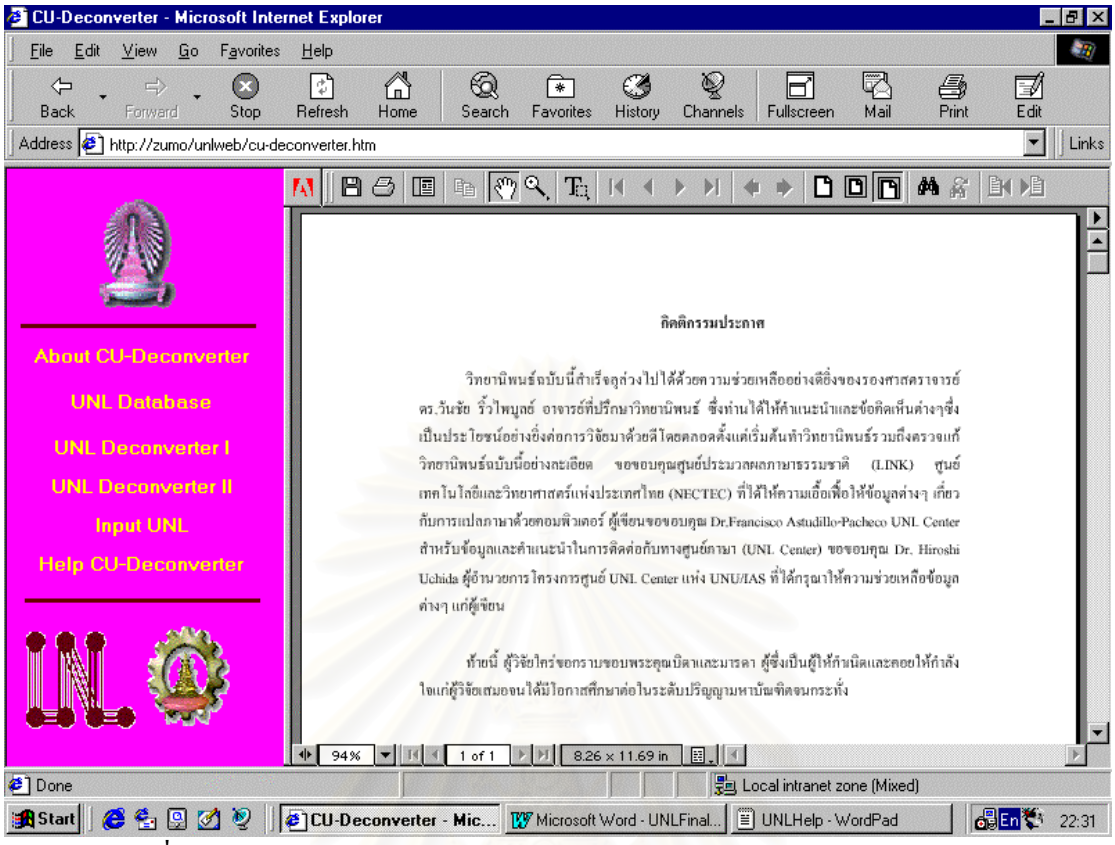
รูปที่ 7.44 ตัวอย่างหน้าจอ Universal Networking Language Web Page



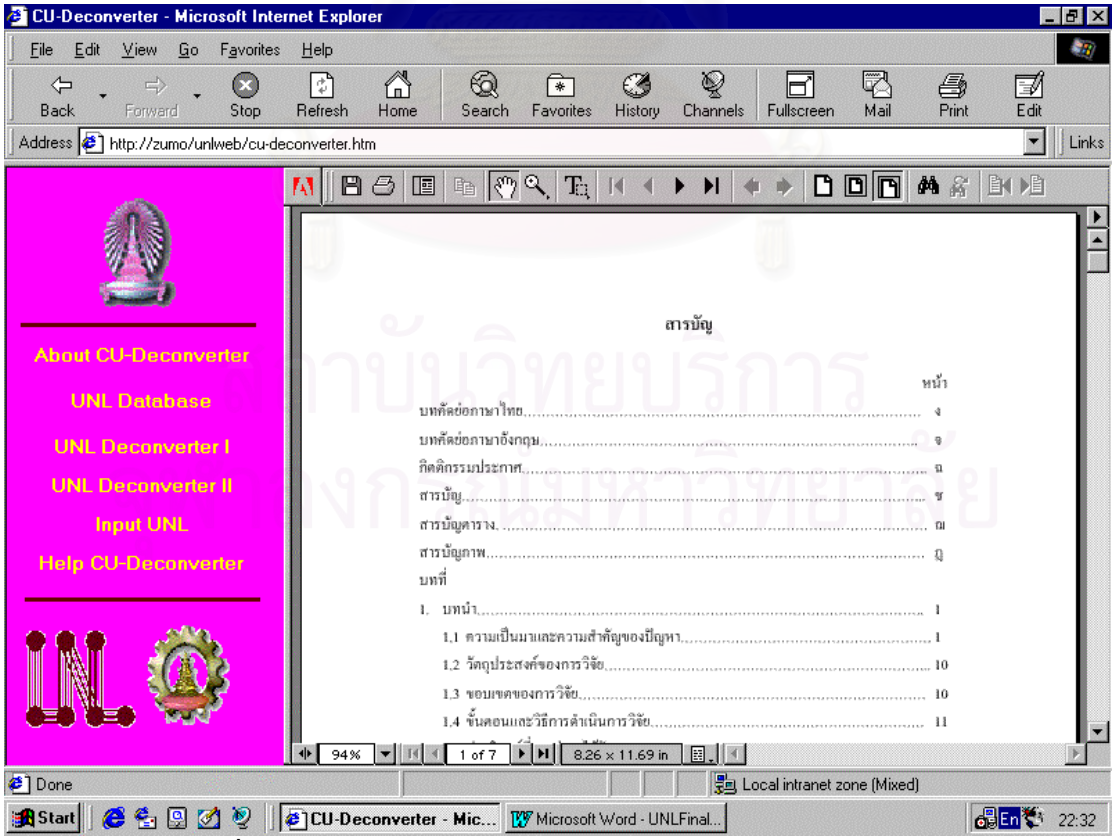
รูปที่ 7.45 ตัวอย่างหน้าจอ CU-Deconverter FAQ Web Page



รูปที่ 7.46 ตัวอย่างหน้าจอ CU-Deconverter Thesis Web Page



รูปที่ 7.47 ตัวอย่างหน้าจอ Thesis Introduction กิตติกรรมประกาศ Web Page



รูปที่ 7.48 ตัวอย่างหน้าจอ Thesis Contents สารบัญ Web Page

ภาคผนวก ค

ตัวอย่าง Source Program

1. Pre-Edit Module

```
Private Sub Pre_Edit_Click()

   OutputFile = "C:\UNLProject\UNLExample\Pre-Edit.Txt"
   OutputError = "C:\UNLProject\UNLExample\Pre-Edit-Error.Txt"
   StrucFile = "C:\UNLProject\UNLExample\UNLRecord.Txt"
   UNLFile = "C:\UNLProject\UNLExample\UNLFile.Txt"
   UNLPattern = "C:\UNLProject\UNLExample\UNLPattern.Txt"

   InpFileNum = FreeFile
   On Error Resume Next
   Open InputFile For Input As InpFileNum
   If Err <> 0 Then
     MsgBox ("File Input Error: " + Str$(Err))
   Else
     sFile = OutputFile
     OutFileNum = FreeFile
     Open OutputFile For Output As OutFileNum
     If Err <> 0 Then
       MsgBox ("File Output Error: " + Str$(Err))
     Else
       sFileErr = OutputError
       OutFileErr = FreeFile
       Open OutputError For Output As OutFileErr
       If Err <> 0 Then
         MsgBox ("File Output Error: " + Str$(Err))
       Else
         sFileStruc = StrucFile
         StrucFileNum = FreeFile
         Open StrucFile For Output As StrucFileNum
         Close StrucFileNum
         Open StrucFile For Random As StrucFileNum Len = Len(UNLRecords)
         If Err <> 0 Then
           MsgBox ("File Structure Error: " + Str$(Err))
         Else
           sFileUNL = UNLFile
           UNLFileNum = FreeFile
           Open UNLFile For Output As UNLFileNum
           Close UNLFileNum
           Open UNLFile For Random As UNLFileNum Len = Len(UNLFileRec)
           If Err <> 0 Then
             MsgBox ("File UNL Open Error: " + Str$(Err))
           Else
             sFileUNLPattern = UNLPattern
             UNLPatternNum = FreeFile
             Open UNLPattern For Output As UNLPatternNum
             Close UNLPatternNum
             Open UNLPattern For Random As UNLPatternNum Len = Len(UNLPatternRec)
             If Err <> 0 Then
               MsgBox ("File UNL Pattern Open Error: " + Str$(Err))
             Else
               Pre_Edit_UNL1
             End If
           End If
         End If
       End If
     End If
   End If
   Close InpFileNum
   Close OutFileNum
   Close OutFileErr
   Close StrucFileNum
   Close UNLFileNum
   Close UNLPatternNum

   LoadNewDoc
   ActiveForm.rtfText.LoadFile sFile
   ActiveForm.Caption = sFile
```

```
    If ErrorCount > 0 Then
      LoadNewDoc
      ActiveForm.rtfText.LoadFile sFileErr
      ActiveForm.Caption = sFileErr
    End If

End Sub
Private Sub Pre_Edit_UNL1()

  Open_Dynaset "DB_READONLY", UNL_DictionaryDs, "UNL_Dictionary", "MASTER"
  Open_Dynaset "DB_READONLY", UNL_Universal_WordsDs, "UNL_Universal_Words", "MASTER"
  Open_Dynaset "DB_READONLY", UNL_Attribute_LabelsDs, "UNL_Attribute_Labels", "MASTER"
  Open_Dynaset "DB_READONLY", UNL_Relation_LabelsDs, "UNL_Relation_Labels", "MASTER"

  ErrorCount = 0
  RecordNum = 0
  Exit_PreEdit = False
  Do Until Exit_PreEdit  ' EOF(InpFileNum)
    Line Input #InpFileNum, TextLine
    If TextLine = "[/S]" Then
      Exit_PreEdit = True
    Else
      RecordNum = RecordNum + 1
      UNLRecords.ALs = " "
      UNLRecords.ObjectClass = " "
      UNLRecords.RecNo = 0
      UNLRecords.RLs = " "
      UNLRecords.UWs = " "
      FindPos = InStr(TextLine, "(")
      LastPos = Len(TextLine) ' InStr(TextLine, ":")
      LineError = 0
      If FindPos > 0 Then ' Find Content
        Pre_Edit_UNL1_Sub
      Else
        TextResult = TextLine
      End If
      If LineError > 0 Then
        TextResult = Trim(TextResult) + "     | *Error Line* |"
      End If
      Print #OutFileNum, TextResult
      UNLRecords.RecNo = RecordNum
      UNLRecords.CrLf = Chr(13) & Chr(10)
      Put #StrucFileNum, RecordNum, UNLRecords
    End If
  Loop

  Close_Dynaset UNL_Universal_WordsDs
  Close_Dynaset UNL_Attribute_LabelsDs
  Close_Dynaset UNL_Relation_LabelsDs

End Sub

Private Sub Pre_Edit_UNL1_Sub()
  TextResult = ""
  StartPos1 = FindPos
  'Universal Words
  Wk_Universal_Word = Mid(TextLine, 1, FindPos - 1)
  UNLRecords.UWs = Wk_Universal_Word
  TextResult = Mid(TextLine, 1, FindPos - 1)
  Wkfind = "UW_Keyword = " + "'" + Wk_Universal_Word + "'"
  Find_Record UNL_Universal_WordsDs, Wkfind, "UW_Keyword", WkRtn, WkFoundRtn
  If Not WkFoundRtn Then
    Wk_Native_Desc = Wk_Universal_Word
    TextError = TextLine + " --> " + Wk_Universal_Word + " Not found In Universal Word"
    Print #OutFileErr, TextError
    ErrorCount = ErrorCount + 1
    LineError = LineError + 1
  Else
    Wk_Native_Desc = UNL_Universal_WordsDs.Fields("Native_Desc")
    Text_Result = Text_Result + " " + Wk_Native_Desc
  End If
  ' Relation Labels
  StartPos2 = InStr(TextLine, ">")
  EndPos2 = InStr(TextLine, ")")
```

```
  Wk_Length = StartPos2 - StartPos1 - 1
  TextResult = TextResult + "," + Mid(TextLine, StartPos1 + 1, Wk_Length)
  UNLRecords.RLs = Trim(Mid(TextLine, StartPos1 + 1, Wk_Length))
  Wk_RLs = Trim(Mid(TextLine, StartPos1 + 1, Wk_Length))
  ' Object Class
  Wk_Length = EndPos2 - StartPos2 - 1
  TextResult = TextResult + "," + Mid(TextLine, StartPos2 + 1, Wk_Length)
  UNLRecords.ObjectClass = Trim(Mid(TextLine, StartPos2 + 1, Wk_Length))
  Wk_ObjectClass = Trim(Mid(TextLine, StartPos2 + 1, Wk_Length))
  ' Find in UNL_Dictionary
  WkfindDict = "UW_Keyword = " + "'" + Wk_Universal_Word + "'" + _
               " And RL_Code = " + "'" + Wk_RLs + "'" + _
               " And Object_Class = " + "'" + Wk_ObjectClass + "'" + _
               " And Subject_Object_Class = " + "'" + Wk_ObjectClass + "'"
  Find_Record UNL_DictionaryDs, WkfindDict, "UW_Keyword", WkRtnDict, WkFoundRtnDict
  If Not WkFoundRtnDict Then
    Dict_Wk_Native_Desc = Wk_Universal_Word
    Dict_Category = " "
    TextError = TextLine + " --> " + " Not found In UNL Dictionary"
    Print #OutFileErr, TextError
    Print #OutFileErr, WkfindDict
    Print #OutFileErr, String(150, "-")
    ErrorCount = ErrorCount + 1
    LineError = LineError + 1
  Else
    Dict_Wk_Native_Desc = UNL_DictionaryDs.Fields("Native_language")
    Dict_Category = UNL_DictionaryDs.Fields("Category")
  End If
  ' Attribute Labels
  FindPos = InStr(TextLine, ".@")
  TextCut = Mid(TextLine, FindPos + 1, LastPos - FindPos + 3)
  If FindPos > 0 Then  ' Found .@
    Wk_ALs = Mid(TextLine, FindPos + 1, LastPos - FindPos - 3)
    UNLRecords.ALs = Wk_ALs
    Pre_Edit_UNL1_Sub2
  Else  ' Not found .@
    Wk_ALs = " "
    FindPos = InStr(TextCut, ":")
    Wk_Seq = Val(Mid(TextCut, FindPos + 1, 2)) + 1
    TextResult = TextResult + " --> " + Mid(TextCut, FindPos + 1, 2)
    If Wk_Seq > 0 Then
      ' For UNL File
      UNLFileRec.RecNo = Val(Wk_Seq)
      UNLFileRec.Native = Dict_Wk_Native_Desc
      UNLFileRec.Category = Dict_Category
      UNLFileRec.AL_Code01 = " "
      UNLFileRec.AL_Code02 = " "
      UNLFileRec.AL_Code03 = " "
      UNLFileRec.AL_Code04 = " "
      UNLFileRec.AL_Code05 = " "
      UNLFileRec.CrLf = Chr(13) & Chr(10)
      Put #UNLFileNum, Val(Wk_Seq), UNLFileRec
      ' For UNL Verb Pattern
      UNLPatternRec.RecNo = Val(Wk_Seq)
      UNLPatternRec.AL_Code01 = Wk_ALs
      UNLPatternRec.CrLf = Chr(13) & Chr(10)
      Put #UNLPatternNum, Val(Wk_Seq), UNLPatternRec
    End If
  End If

End Sub
Private Sub Pre_Edit_UNL1_Sub2()

  StartPos1 = InStr(TextCut, ".@")
  ExitPos = 1
  Wk_Seq = 0
  have_dot = 0
  Do Until ExitPos = 0
    FindPos = InStr(TextCut, ".@")
    If FindPos = 0 Then  ' Last Position
      FindPos = InStr(TextCut, ":")
      Wk_Seq = Val(Mid(TextCut, FindPos + 1, 2)) + 1
      TextResult = TextResult + "," + Mid(TextCut, 2, FindPos - 2)
      TextResult = TextResult + " --> " + Mid(TextCut, FindPos + 1, 2)
      ExitPos = 0
```

```
      If Wk_Seq > 0 Then
        If Len(Wk_ALs) > 0 Then
          i = 0
          ExitPattern = 1
          Wk_Text_ALs = Wk_ALs
          Wk_Long_ALs = Wk_ALs
          Wk_Array_ALs = " "
          Wk_Array_AL01 = " "
          Wk_Array_AL02 = " "
          Wk_Array_AL03 = " "
          Wk_Array_AL04 = " "
          Wk_Array_AL05 = " "
          Wk_Len_ALs = Len(Wk_ALs)
          Do Until ExitPattern = 0
            i = i + 1
            FindPosAL = InStr(Wk_Text_ALs, ".@")
            If FindPosAL > 0 Then
              Wk_Array_ALs = Mid(Wk_Text_ALs, 2, FindPosAL - 2)
              LenAL = FindPosAL
              Wk_Text_ALs = Mid(Wk_Text_ALs, FindPosAL + 1, Wk_Len_ALs - LenAL + 1)
            Else
              Wk_Array_ALs = Mid(Wk_Text_ALs, 2, Len(Wk_Text_ALs) - 1)
              ExitPattern = 0
            End If
            Select Case i
              Case 1
                Wk_Array_AL01 = Wk_Array_ALs
              Case 2
                Wk_Array_AL02 = Wk_Array_ALs
              Case 3
                Wk_Array_AL03 = Wk_Array_ALs
            End Select
          Loop
        End If
        If Dict_Category = " " Then
          Dict_Category = Wk_Array_AL01
        End If
        ' Write UNL File
        UNLFileRec.RecNo = Val(Wk_Seq)
        UNLFileRec.Native = Dict_Wk_Native_Desc
        UNLFileRec.Category = Dict_Category
        UNLFileRec.AL_Code01 = Wk_Array_AL01
        UNLFileRec.AL_Code02 = Wk_Array_AL02
        UNLFileRec.AL_Code03 = Wk_Array_AL03
        UNLFileRec.AL_Code04 = Wk_Array_AL04
        UNLFileRec.AL_Code05 = Wk_Array_AL05
        UNLFileRec.CrLf = Chr(13) & Chr(10)
        Put #UNLFileNum, Val(Wk_Seq), UNLFileRec
        ' Write UNL Pattern
        UNLPatternRec.RecNo = Val(Wk_Seq)
        UNLPatternRec.AL_Code01 = Wk_Array_AL01
        UNLPatternRec.AL_Code02 = Wk_Array_AL02
        UNLPatternRec.AL_Code03 = Wk_Array_AL03
        UNLPatternRec.AL_Code04 = Wk_Array_AL04
        UNLPatternRec.AL_Code05 = Wk_Array_AL05
        UNLPatternRec.CrLf = Chr(13) & Chr(10)
        Put #UNLPatternNum, Val(Wk_Seq), UNLPatternRec
      End If
    Else ' Not Last Position
      LastPos = InStr(TextCut, ":")
      Wk_Seq = Val(Mid(TextCut, LastPos + 1, 2)) + 1
      StartPos2 = FindPos + 1
      TextResult = TextResult + "," + Mid(TextCut, 2, StartPos2 - 3)
      TextCut = Mid(TextCut, StartPos2, LastPos - StartPos2 + 3)
    End If
  Loop
End Sub
```

2. Scan UWs From UNL Dictionary Module

```
Private Sub Scan_UWs_Click()

   OutputFile = "C:\UNLProject\UNLExample\Scan UWs.Txt"
   OutputError = "C:\UNLProject\UNLExample\UNLError2.Txt"

   Open_Dynaset "DB_READONLY", UNL_CategoryDs, "UNL_Category", "MASTER"

   InpFileNum = FreeFile
   On Error Resume Next
   Open InputFile For Input As InpFileNum
   If Err <> 0 Then
     MsgBox ("File Input Error: " + Str$(Err))
   Else
     sFile = OutputFile
     OutFileNum = FreeFile
     Open OutputFile For Output As OutFileNum
     If Err <> 0 Then
       MsgBox ("File Output Error: " + Str$(Err))
     Else
       sFileErr = OutputError
       OutFileErr = FreeFile
       Open OutputError For Output As OutFileErr
       If Err <> 0 Then
         MsgBox ("File Output Error: " + Str$(Err))
       Else
         Scan_UWs_UNL2
       End If
     End If
   End If
   Close InpFileNum
   Close OutFileNum
   Close OutFileErr
   Close_Dynaset UNL_CategoryDs

   LoadNewDoc
   ActiveForm.rtfText.LoadFile sFile
   ActiveForm.Caption = sFile

   If ErrorCount > 0 Then
     LoadNewDoc
     ActiveForm.rtfText.LoadFile sFileErr
     ActiveForm.Caption = sFileErr
   End If
End Sub
Private Sub Scan_UWs_UNL2()

  Dim StartPos1, EndPos1, FindPos, Wk_Length, Textlength As Integer
  Dim StartPos2, EndPos2, EndPos, LastPos, ExitPos As Integer
  Dim TextLine, TextResult, TextCut As String
  Dim Wk_Universal_Word, Wk_Native_Desc As String
  Dim WkRtn As String
  Dim Wkfind As String
  Dim WkFoundRtn As Boolean
  Dim WkselectSQL As String
  Dim UNLFileRec As UNL_Files

  UNLFile = "C:\UNLProject\UNLExample\UNLFile.Txt"

  Open_Dynaset "DB_READONLY", UNL_DictionaryDs, "UNL_Dictionary", "MASTER"

  UNLFileNum = FreeFile
  On Error Resume Next
  Open UNLFile For Random As UNLFileNum Len = Len(UNLFileRec)
  If Err <> 0 Then
    MsgBox ("File UNL Error: " + Str$(Err))
  End If

  ErrorCount = 0
  FindStart = 0
  Do Until EOF(InpFileNum)
    Line Input #InpFileNum, TextLine
    FindPos = InStr(TextLine, "(")
    LastPos = Len(TextLine)
```

```
    ' Find [R]
    If InStr(TextLine, "[R]") Then
      FindStart = 1
    End If
    If FindStart >= 1 And InStr(TextLine, "[/R]") Then
      FindStart = 0
    End If
    If FindStart > 1 Then
      TextResult = Mid(TextLine, 1, 2) + " " + Mid(TextLine, 3, 3) + " " + _
                Mid(TextLine, 6, 2)
      Wk_seq1 = Val(Mid(TextLine, 1, 2)) + 1
      Wk_Rel = Trim(Mid(TextLine, 3, 3))
      Wk_seq2 = Val(Mid(TextLine, 6, 2)) + 1
      Get #UNLFileNum, Wk_seq1, UNLFileRec
      Wk_Category1 = Trim(UNLFileRec.Category)
      Wk_Native1 = Trim(UNLFileRec.Native)
      ' Find in UNL_Category 1
      WkfindCategory = "Category = " + "'" + Wk_Category1 + "'"
      Find_Record UNL_CategoryDs, WkfindCategory, "Category", WkRtnCategory, WkFoundRtnCategory
      If Not WkFoundRtnCategory Then
        Wk_Name_Category1 = Wk_Category1
      Else
        Wk_Name_Category1 = UNL_CategoryDs.Fields("Name")
      End If

      Get #UNLFileNum, Wk_seq2, UNLFileRec
      Wk_Category2 = Trim(UNLFileRec.Category)
      Wk_Native2 = Trim(UNLFileRec.Native)
      ' Find in UNL_Category 2
      WkfindCategory = "Category = " + "'" + Wk_Category2 + "'"
      Find_Record UNL_CategoryDs, WkfindCategory, "Category", WkRtnCategory, WkFoundRtnCategory
      If Not WkFoundRtnCategory Then
        Wk_Name_Category2 = Wk_Category2
      Else
        Wk_Name_Category2 = UNL_CategoryDs.Fields("Name")
      End If

      TextResult = TextResult + String(20, " ") + Wk_Native1 _
                + " (" + Wk_Name_Category1 + ") " + Wk_Rel _
                + " " + Wk_Native2 + " (" + Wk_Name_Category2 + ") "
      Print #OutFileNum, TextResult
    Else
      Scan_UWs_UNL2_Sub
    End If
    If FindStart = 1 Then
      FindStart = 2
    End If
    If LineError > 0 Then
      'LenText = Len(TextResule)
      'LenBlank = 80 - LenText
      'StrBlank = String(LenBlank, " ")
      TextResult = Trim(TextResult) + "     | *Error Line* |"
    End If
  Loop

  Close UNLFileNum
  Close_Dynaset UNL_DictionaryDs

End Sub

Private Sub Scan_UWs_UNL2_Sub()
   LineError = 0
    If FindPos > 0 Then
      TextResult = ""
      StartPos1 = FindPos
      'Universal Words
      Wk_Universal_Word = Mid(TextLine, 1, FindPos - 1)
      TextResult = Mid(TextLine, 1, FindPos - 1)
      Wkfind = "UW_Keyword = " + "'" + Wk_Universal_Word + "'"
      StartPos2 = InStr(TextLine, ">")
      EndPos2 = InStr(TextLine, ")")
      Wk_Length = StartPos2 - StartPos1 - 1
      TextResult = TextResult + "," + Mid(TextLine, StartPos1 + 1, Wk_Length)
      ' Object Class
      Wk_Length = EndPos2 - StartPos2 - 1
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
      TextResult = TextResult + "," + Mid(TextLine, StartPos2 + 1, Wk_Length)
      ' Attribute Labels
      FindPos = InStr(TextLine, ".@")
      TextCut = Mid(TextLine, FindPos + 1, LastPos - FindPos + 3)
      If FindPos > 0 Then
        StartPos1 = InStr(TextCut, ".@")
        ExitPos = 1
        Do Until ExitPos = 0
          FindPos = InStr(TextCut, ".@")
          If FindPos = 0 Then
            FindPos = InStr(TextCut, ":")
            TextResult = TextResult + "," + Mid(TextCut, 2, FindPos - 2)
            TextResult = TextResult + " --> " + Mid(TextCut, FindPos + 1, 2)
            ExitPos = 0
          Else
            LastPos = InStr(TextCut, ":")
            StartPos2 = FindPos + 1
            TextResult = TextResult + "," + Mid(TextCut, 2, StartPos2 - 3)
            TextCut = Mid(TextCut, StartPos2, LastPos - StartPos2 + 3)
          End If
        Loop
      Else
        FindPos = InStr(TextCut, ":")
        TextResult = TextResult + " --> " + Mid(TextCut, FindPos + 1, 2)
      End If
    Else
      TextResult = TextLine
    End If
End Sub
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3. Get Verb Pattern Module

```
Private Sub Get_Verb_Pattern_Click()

   OutputFile = "C:\UNLProject\UNLExample\Get Verb Pattern.Txt"
   OutputError = "C:\UNLProject\UNLExample\Get Verb Pattern-Error.Txt"
   InputFile2 = "C:\UNLProject\UNLExample\Scan UWs.Txt"
   UNLFile = "C:\UNLProject\UNLExample\UNLFile.Txt"
   UNLFile3 = "C:\UNLProject\UNLExample\UNLFile3.Txt"

   Open_Dynaset "DB_READONLY", UNL_Verb_PatternDs, "UNL_Verb_Pattern", "MASTER"
   Open_Dynaset "DB_READONLY", UNL_FunctionDs, "UNL_Function", "MASTER"

   InpFileNum = FreeFile
   On Error Resume Next
   Open InputFile2 For Input As InpFileNum
   If Err <> 0 Then
     MsgBox ("File Input Error: " + Str$(Err))
   Else
     UNLFileNum = FreeFile
     Open UNLFile For Random As UNLFileNum Len = Len(UNLFileRec)
     If Err <> 0 Then
       MsgBox ("File UNL Error: " + Str$(Err))
     Else
       UNLFileNum3 = FreeFile
       Open UNLFile3 For Output As UNLFileNum3
       Close UNLFileNum3
       Open UNLFile3 For Random As UNLFileNum3 Len = Len(UNLFileRec3)
       If Err <> 0 Then
         MsgBox ("File UNL 3 Error: " + Str$(Err))
       Else
         sFile = OutputFile
         OutFileNum = FreeFile
         Open OutputFile For Output As OutFileNum
         If Err <> 0 Then
           MsgBox ("File Output Error: " + Str$(Err))
         Else
           sFileErr = OutputError
           OutFileErr = FreeFile
           Open OutputError For Output As OutFileErr
           If Err <> 0 Then
             MsgBox ("File Output Error: " + Str$(Err))
           Else
             Get_Verb_UNL3
           End If
         End If
       End If
     End If
   End If
   Close InpFileNum
   Close OutFileNum
   Close OutFileErr
   Close UNLFileNum
   Close UNLFileNum3
   Close_Dynaset UNL_FunctionDs
   Close_Dynaset UNL_Verb_PatternDs

   LoadNewDoc
   ActiveForm.rtfText.LoadFile sFile
   ActiveForm.Caption = sFile

   If ErrorCount > 0 Then
     LoadNewDoc
     ActiveForm.rtfText.LoadFile sFileErr
     ActiveForm.Caption = sFileErr
   End If

End Sub
Private Sub Get_Verb_UNL3()

  ErrorCount = 0
  FindStart = 0
  Do Until EOF(InpFileNum)
    Line Input #InpFileNum, TextLine
```

```
    TextResult = TextLine
    Get_Verb_UNL3_Sub
    Print #OutFileNum, TextResult
  Loop

End Sub
Private Sub Get_Verb_UNL3_Sub()

  ' Find Category
  Wk_seq1 = Val(Mid(TextLine, 1, 2)) + 1
  WkRecNo1 = Mid(TextLine, 1, 2)
  Text_Rel = Mid(TextLine, 4, 6)
  FindPos = InStr(Text_Rel, " ")
  Wk_Rel = Trim(Mid(Text_Rel, 1, 6 - FindPos + 1))
  Wk_seq2 = Val(Mid(Text_Rel, FindPos + 1, 2)) + 1
  WkRecNo2 = Mid(Text_Rel, FindPos + 1, 2)
  Get #UNLFileNum, Wk_seq1, UNLFileRec
  Wk_Category1 = Trim(UNLFileRec.Category)
  Wk_AL01 = Trim(UNLFileRec.AL_Code01)
  Wk_AL02 = Trim(UNLFileRec.AL_Code02)
  Wk_AL03 = Trim(UNLFileRec.AL_Code03)
  Wk_AL04 = Trim(UNLFileRec.AL_Code04)
  Wk_AL05 = Trim(UNLFileRec.AL_Code05)

  Get #UNLFileNum, Wk_seq2, UNLFileRec
  Wk_Category2 = Trim(UNLFileRec.Category)
  Wk_AL06 = Trim(UNLFileRec.AL_Code01)
  Wk_AL07 = Trim(UNLFileRec.AL_Code02)
  Wk_AL08 = Trim(UNLFileRec.AL_Code03)
  Wk_AL09 = Trim(UNLFileRec.AL_Code04)
  Wk_AL10 = Trim(UNLFileRec.AL_Code05)

  ' Find in UNL_Verb_Pattern from
  Get_Verb_UNL3_Index
  Find_Record UNL_Verb_PatternDs, WkfindVerb, "Category1", WkRtnVerb, WkFoundRtnVerb
  If Not WkFoundRtnVerb Then
    Wk_Verb_Function1 = Wk_Category1
    Wk_Verb_RLCode = Wk_Rel
    Wk_Verb_Function2 = Wk_Category2
    Wk_Verb_RL_Native = " "
    ErrorCount = ErrorCount + 1
    Print #OutFileErr, TextLine
    Print #OutFileErr, WkfindVerb
    Print #OutFileErr, String(150, "-")
  Else
    Wk_Verb_Function1 = UNL_Verb_PatternDs.Fields("Function_Code1")
    Wk_Verb_RLCode = UNL_Verb_PatternDs.Fields("RL_Code2")
    Wk_Verb_Function2 = UNL_Verb_PatternDs.Fields("Function_Code2")
    Wk_Verb_RL_Native = UNL_Verb_PatternDs.Fields("RL_Native_Language")
    Wk_Verb_RL_Native_Back = UNL_Verb_PatternDs.Fields("RL_Native_Back")
  End If

  ' Find in UNL_Function from Category 1
  WkfindFunction = "Function_Code = " + "'" + Wk_Verb_Function1 + "'"
  Find_Record UNL_FunctionDs, WkfindFunction, "Function_Code", WkRtnFunction, WkFoundRtnFunction
  If Not WkFoundRtnFunction Then
    Wk_Function_Name = Wk_Function1
    Wk_Function_Values1 = "00000"
  Else
    Wk_Function_Name = UNL_FunctionDs.Fields("Name")
    Wk_Function_Values1 = Trim(Str(UNL_FunctionDs.Fields("Values")))
  End If

  ' Find in UNL_Function from Category 2
  WkfindFunction = "Function_Code = " + "'" + Wk_Verb_Function2 + "'"
  Find_Record UNL_FunctionDs, WkfindFunction, "Function_Code", WkRtnFunction, WkFoundRtnFunction
  If Not WkFoundRtnFunction Then
    Wk_Function_Name2 = Wk_Function2
    Wk_Function_Values2 = "00000"
  Else
    Wk_Function_Name2 = UNL_FunctionDs.Fields("Name")
    Wk_Function_Values2 = Trim(Str(UNL_FunctionDs.Fields("Values")))
  End If

  Len_Result = Len(TextResult)
```

```
    TextResult = Trim(Mid(TextResult, 12, Len_Result - 12)) + String(100 - Len_Result, " ") + _
                 "  " + _
                 Wk_Verb_Function1 + " (" + _
                 Wk_Function_Values1 + ")" + _
                 " " + Wk_Verb_RLCode + " " + _
                 Wk_Verb_Function2 + " (" + _
                 Wk_Function_Values2 + ")" + _
                 "  "

    ' Move blank for cut function
    If Wk_Verb_Function1 = "cut" Then
      Wk_Verb_Function1 = " "
      Wk_Function_Values1 = " "
      WkRecNo1 = " "
    End If
    If Wk_Verb_Function2 = "cut" Then
      Wk_Verb_Function2 = " "
      Wk_Function_Values2 = " "
      WkRecNo2 = " "
    End If

    ' Write UNL Pattern
    UNLFileRec3.RecNo1 = WkRecNo1
    UNLFileRec3.Function1 = Wk_Verb_Function1
    UNLFileRec3.Function2 = Wk_Verb_Function2
    UNLFileRec3.RL_Code = Wk_Verb_RLCode
    UNLFileRec3.RecNo2 = WkRecNo2
    UNLFileRec3.Value1 = Wk_Function_Values1
    UNLFileRec3.Value2 = Wk_Function_Values2
    If Not Wk_Verb_RL_Native = " " Then
      UNLFileRec3.RL_Native_Language = Wk_Verb_RL_Native
    Else
      UNLFileRec3.RL_Native_Language = " "
    End If
    If Not Wk_Verb_RL_Native_Back = " " Then
      UNLFileRec3.RL_Native_Back = Wk_Verb_RL_Native_Back
    Else
      UNLFileRec3.RL_Native_Back = " "
    End If
    UNLFileRec3.CrLf = Chr(13) & Chr(10)
    Put #UNLFileNum3, , UNLFileRec3

End Sub

Private Sub Get_Verb_UNL3_Index()
  ' Find in UNL_Verb_Pattern from
    WkfindVerb = "Category1 = " + "'" + Wk_Category1 + "'"
    If Not Wk_AL01 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code01 = " + "'" + Wk_AL01 + "'"
    End If
    If Not Wk_AL02 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code02 = " + "'" + Wk_AL02 + "'"
    End If
    If Not Wk_AL03 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code03 = " + "'" + Wk_AL03 + "'"
    End If
    If Not Wk_AL04 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code04 = " + "'" + Wk_AL04 + "'"
    End If
    If Not Wk_AL05 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code05 = " + "'" + Wk_AL05 + "'"
    End If
    If Not Wk_Rel = "" Then
      WkfindVerb = WkfindVerb + " And RL_Code = " + "'" + Wk_Rel + "'"
    End If
    If Not Wk_Category2 = "" Then
      WkfindVerb = WkfindVerb + " And Category2 = " + "'" + Wk_Category2 + "'"
    End If
    If Not Wk_AL06 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code06 = " + "'" + Wk_AL06 + "'"
    End If
    If Not Wk_AL07 = "" Then
      WkfindVerb = WkfindVerb + " And AL_Code07 = " + "'" + Wk_AL07 + "'"
    End If
    If Not Wk_AL08 = "" Then
```

```
    WkfindVerb = WkfindVerb + " And AL_Code08 = " + """ + Wk_AL08 + """
  End If
  If Not Wk_AL09 = "" Then
    WkfindVerb = WkfindVerb + " And AL_Code09 = " + """ + Wk_AL09 + """
  End If
  If Not Wk_AL10 = "" Then
    WkfindVerb = WkfindVerb + " And AL_Code10 = " + """ + Wk_AL10 + """
  End If
End Sub
```


สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. Word Ordering Module

```
Private Sub Word_Ordering_Click()

   OutputFile = "C:\UNLProject\UNLExample\Word Ordering.Txt"
   OutputError = "C:\UNLProject\UNLExample\UNLError4.Txt"
   UNLFile3 = "C:\UNLProject\UNLExample\UNLFile3.Txt"
   UNLFile4 = "C:\UNLProject\UNLExample\UNLFile4.Txt"

   InpFileNum = FreeFile
   On Error Resume Next
   Open UNLFile3 For Input As InpFileNum Len = Len(UNLFileRec3)
   If Err <> 0 Then
      MsgBox ("File Input Error: " + Str$(Err))
   Else
      sFile = OutputFile
      OutFileNum = FreeFile
      Open OutputFile For Output As OutFileNum
      If Err <> 0 Then
         MsgBox ("File Output Error: " + Str$(Err))
      Else
         sFileErr = OutputError
         OutFileErr = FreeFile
         Open OutputError For Output As OutFileErr
         If Err <> 0 Then
            MsgBox ("File Output Error: " + Str$(Err))
         Else
            UNLFileNum4 = FreeFile
            Open UNLFile4 For Output As UNLFileNum4
            Close UNLFileNum4
            Open UNLFile4 For Random As UNLFileNum4 Len = Len(UNLFileRec4)
            If Err <> 0 Then
               MsgBox ("File Output Error: " + Str$(Err))
            Else
               Word_Ordering_UNL4
            End If
         End If
      End If
   End If
   Close InpFileNum
   Close OutFileNum
   Close OutFileErr
   Close UNLFileNum4

   LoadNewDoc
   ActiveForm.rtfText.LoadFile sFile
   ActiveForm.Caption = sFile

   If ErrorCount > 0 Then
      LoadNewDoc
      ActiveForm.rtfText.LoadFile sFileErr
      ActiveForm.Caption = sFileErr
   End If

End Sub

Private Sub Word_Ordering_UNL4()

  ErrorCount = 0
  FindStart = 0
  Do Until EOF(InpFileNum)
     Line Input #InpFileNum, TextLine
     TextResult = TextLine
     WkRecNo1 = Trim(Mid(TextLine, 1, 4))
     WkRecNo2 = Trim(Mid(TextLine, 35, 4))
     Wk_Function1 = LCase(Trim(Mid(TextLine, 5, 10)))
     Wk_Function2 = LCase(Trim(Mid(TextLine, 39, 10)))
     Wk_RL_Code = Trim(Mid(TextLine, 25, 10))
     wk_value1 = Trim(Mid(TextLine, 15, 10))
     wk_value2 = Trim(Mid(TextLine, 49, 10))
     Wk_RL_Native_Language = Trim(Mid(TextLine, 59, 20))
     Wk_RL_Native_Back = Trim(Mid(TextLine, 79, 20))

     ' Word Ordering by Value
     If Len(Trim(wk_value1)) = 0 Or Len(Trim(wk_value2)) = 0 Then
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
      If Len(Trim(wk_value1)) = 0 Then
        TextResult = Wk_Function2 + " (" + wk_value2 + ") "
      End If
      If Len(Trim(wk_value2)) = 0 Then
        TextResult = Wk_Function1 + " (" + wk_value1 + ") " + _
                Wk_RL_Code + " "
      End If
      UNLFileRec4.RecNo1 = WkRecNo1
      UNLFileRec4.RecNo2 = WkRecNo2
      UNLFileRec4.Function1 = Wk_Function1
      UNLFileRec4.Function2 = Wk_Function2
      UNLFileRec4.Value1 = wk_value1
      UNLFileRec4.Value2 = wk_value2
    Else
      If Val(wk_value1) <= Val(wk_value2) Then
        TextResult = Wk_Function1 + " (" + wk_value1 + ") " + _
                Wk_RL_Code + " " + Wk_Function2 + _
                " (" + wk_value2 + ") "
        UNLFileRec4.RecNo1 = WkRecNo1
        UNLFileRec4.RecNo2 = WkRecNo2
        UNLFileRec4.Function1 = Wk_Function1
        UNLFileRec4.Function2 = Wk_Function2
        UNLFileRec4.Value1 = wk_value1
        UNLFileRec4.Value2 = wk_value2
      Else
        TextResult = Wk_Function2 + " (" + wk_value2 + ") " + _
                Wk_RL_Code + " " + Wk_Function1 + _
                " (" + wk_value1 + ") "
        UNLFileRec4.RecNo1 = WkRecNo2
        UNLFileRec4.RecNo2 = WkRecNo1
        UNLFileRec4.Function1 = Wk_Function2
        UNLFileRec4.Function2 = Wk_Function1
        UNLFileRec4.Value1 = wk_value2
        UNLFileRec4.Value2 = wk_value1
      End If
    End If
    UNLFileRec4.RL_Code = Wk_RL_Code
    UNLFileRec4.RL_Native_Language = Wk_RL_Native_Language
    UNLFileRec4.RL_Native_Back = Wk_RL_Native_Back
    UNLFileRec4.CrLf = Chr(13) & Chr(10)
    Put #UNLFileNum4, , UNLFileRec4

    Print #OutFileNum, TextResult

  Loop

End Sub
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. Morphological Generation Module

```
Private Sub Morphological_Generation_Click()

   OutputFile = "C:\UNLProject\UNLExample\Morphological Generation.Txt"
   OutputError = "C:\UNLProject\UNLExample\UNLError5.Txt"
   UNLFile = "C:\UNLProject\UNLExample\UNLFile.Txt"
   UNLFile4 = "C:\UNLProject\UNLExample\UNLFile4.Txt"
   UNLFile5 = "C:\UNLProject\UNLExample\UNLFile5.Txt"

   InpFileNum = FreeFile
   On Error Resume Next
   Open UNLFile4 For Input As InpFileNum Len = Len(UNLFileRec4)
   If Err <> 0 Then
     MsgBox ("File Input Error: " + Str$(Err))
   Else
     sFile = OutputFile
     OutFileNum = FreeFile
     Open OutputFile For Output As OutFileNum
     If Err <> 0 Then
       MsgBox ("File Output Error: " + Str$(Err))
     Else
       sFileErr = OutputError
       OutFileErr = FreeFile
       Open OutputError For Output As OutFileErr
       If Err <> 0 Then
         MsgBox ("File Output Error: " + Str$(Err))
       Else
         UNLFileNum5 = FreeFile
         Open UNLFile5 For Output As UNLFileNum5
         Close UNLFileNum5
         Open UNLFile5 For Random As UNLFileNum5 Len = Len(UNLFileRec5)
         If Err <> 0 Then
           MsgBox ("File Output Error: " + Str$(Err))
         Else
           sFileUNL = UNLFile
           UNLFileNum = FreeFile
           Open UNLFile For Random As UNLFileNum Len = Len(UNLFileRec)
           If Err <> 0 Then
             MsgBox ("File UNL Open Error: " + Str$(Err))
           Else
             Morphological_UNL5
           End If
         End If
       End If
     End If
   End If
   Close InpFileNum
   Close OutFileNum
   Close OutFileErr
   Close UNLFileNum
   Close UNLFileNum5

   LoadNewDoc
   ActiveForm.rtfText.LoadFile sFile
   ActiveForm.Caption = sFile

   If ErrorCount > 0 Then
     LoadNewDoc
     ActiveForm.rtfText.LoadFile sFileErr
     ActiveForm.Caption = sFileErr
   End If

End Sub

Private Sub Morphological_UNL5()

  ErrorCount = 0
  Prv_Native1 = " "
  Prv_Native2 = " "
  Do Until EOF(InpFileNum)
    Line Input #InpFileNum, TextLine
    TextResult = TextLine
    Wk_RecNo1 = Trim(Mid(TextLine, 1, 4))
    Wk_Recno2 = Trim(Mid(TextLine, 35, 4))
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
    Wk_Function1 = LCase(Trim(Mid(TextLine, 5, 10)))
    Wk_Function2 = LCase(Trim(Mid(TextLine, 39, 10)))
    Wk_RL_Code = Trim(Mid(TextLine, 25, 10))
    wk_value1 = Trim(Mid(TextLine, 15, 10))
    wk_value2 = Trim(Mid(TextLine, 49, 10))
    Wk_RL_Native_Language = Trim(Mid(TextLine, 59, 20))
    Wk_RL_Native_Back = Trim(Mid(TextLine, 79, 20))

    ' Find Native Description 1
    If Len(Trim(Wk_RecNo1)) > 0 Then
      Wk_seq1 = Val(Wk_RecNo1) + 1
      Get #UNLFileNum, Wk_seq1, UNLFileRec
      Wk_Native1 = Trim(UNLFileRec.Native)
    Else
      Wk_Native1 = " "
    End If

    ' Find Native Description 2
    If Len(Trim(Wk_Recno2)) > 0 Then
      Wk_seq2 = Val(Wk_Recno2) + 1
      Get #UNLFileNum, Wk_seq2, UNLFileRec
      Wk_Native2 = Trim(UNLFileRec.Native)
    Else
      Wk_Native2 = " "
    End If

    ' Morphological Generation
    TextResult = Wk_Function1 + " " + Wk_RL_Code + " " + _
              Wk_Function2
    Len_Result = Len(Trim(TextResult))
    TextResult = TextResult + String(60 - Len_Result, " ")
    If Wk_Native1 <> Prv_Native1 And Wk_Native1 <> Prv_Native2 Then
      TextResult = TextResult + Wk_Native1 + " (" + Wk_RecNo1 + ") "
    Else
      TextResult = TextResult + String(15, " ")
    End If
    TextResult = TextResult + Wk_RL_Code + " "
    If Wk_Native2 <> Prv_Native2 And Wk_Native2 <> Prv_Native1 Then
      TextResult = TextResult + Wk_Native2 + " (" + Wk_Recno2 + ") "
    Else
      TextResult = TextResult + String(15, " ")
    End If
    ' Morphological Output File
    UNLFileRec5.RecNo1 = Wk_RecNo1
    UNLFileRec5.RecNo2 = Wk_Recno2
    UNLFileRec5.RLs = Wk_RL_Code
    UNLFileRec5.Native1 = Wk_Native1
    UNLFileRec5.Native2 = Wk_Native2
    UNLFileRec5.RL_Native_Language = Wk_RL_Native_Language
    UNLFileRec5.RL_Native_Back = Wk_RL_Native_Back
    UNLFileRec5.CrLf = Chr(13) & Chr(10)
    Put #UNLFileNum5, , UNLFileRec5
    ' Output File 5
    Print #OutFileNum, TextResult
    Prv_Native1 = Wk_Native1
    Prv_Native2 = Wk_Native2
  Loop

End Sub
```

6. Thai Language Module

```
Private Sub Thai_Language_Click()
   OutputFile = "C:\UNLProject\UNLExample\Thai Language Sentence.Txt"
   OutputError = "C:\UNLProject\UNLExample\Thai Language Error.Txt"
   UNLFile = "C:\UNLProject\UNLExample\UNLFile.Txt"
   UNLFile5 = "C:\UNLProject\UNLExample\UNLFile5.Txt"

   InpFileNum = FreeFile
   On Error Resume Next
   Open UNLFile5 For Input As InpFileNum Len = Len(UNLFileRec5)
   If Err <> 0 Then
     MsgBox ("File Input Error: " + Str$(Err))
   Else
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
      sFile = OutputFile
      OutFileNum = FreeFile
      Open OutputFile For Output As OutFileNum
      If Err <> 0 Then
        MsgBox ("File Output Error: " + Str$(Err))
      Else
        sFileErr = OutputError
        OutFileErr = FreeFile
        Open OutputError For Output As OutFileErr
        If Err <> 0 Then
          MsgBox ("File Output Error: " + Str$(Err))
        Else
          sFileUNL = UNLFile
          UNLFileNum = FreeFile
          Open UNLFile For Random As UNLFileNum Len = Len(UNLFileRec)
          If Err <> 0 Then
            MsgBox ("File UNL Open Error: " + Str$(Err))
          Else
            Thai_Language_UNL6
          End If
        End If
      End If
    End If
    Close InpFileNum
    Close OutFileNum
    Close OutFileErr
    Close UNLFileNum

    LoadNewDoc
    ActiveForm.rtfText.LoadFile sFile
    ActiveForm.Caption = sFile

    If ErrorCount > 0 Then
      LoadNewDoc
      ActiveForm.rtfText.LoadFile sFileErr
      ActiveForm.Caption = sFileErr
    End If

End Sub
Private Sub Thai_Language_UNL6()
  ErrorCount = 0
  Prv_Native1 = " "
  Prv_Native2 = " "
  Last_Thai_Sentence = " "
  Do Until EOF(InpFileNum)
    Line Input #InpFileNum, TextLine
    TextResult = TextLine
    TextResult2 = " "
    Wk_RecNo1 = Trim(Mid(TextLine, 1, 4))
    Wk_Recno2 = Trim(Mid(TextLine, 5, 4))
    Wk_RLs = Trim(Mid(TextLine, 9, 10))
    Wk_Native1 = Trim(Mid(TextLine, 19, 256))
    Wk_Native2 = Trim(Mid(TextLine, 275, 256))
    Wk_RL_Native_Language = Trim(Mid(TextLine, 531, 20))
    Wk_RL_Native_Back = Trim(Mid(TextLine, 551, 20))
    ' Thai Language Sentence
    If Wk_Native1 <> Prv_Native1 And Wk_Native1 <> Prv_Native2 Then
      'Check Digit for 1 Blank
      Len_Native1 = Len(Trim(Wk_Native1))
      Last_Native1 = Trim(Mid(Trim(Wk_Native1), Len_Native1, 1))
      If InStr("0123456789", Last_Native1) > 0 Then
        TextResult = Trim(Wk_Native1) + " "
        TextResult2 = Trim(Wk_Native1) + " "
      Else
        TextResult = Trim(Wk_Native1)
        TextResult2 = Trim(Wk_Native1)
      End If
    Else
      TextResult = String(15, " ")
    End If
    TextResult = Trim(TextResult) + " " + Wk_RL_Native_Language + " "
    TextResult2 = TextResult2 + Wk_RL_Native_Language
    If Wk_Native2 <> Prv_Native2 And Wk_Native2 <> Prv_Native1 Then
      TextResult = TextResult + Wk_Native2
      TextResult2 = TextResult2 + Wk_Native2
```

```
    Else
      TextResult = TextResult + String(15, " ")
    End If
    TextResult2 = TextResult2 + Wk_RL_Native_Back
    Last_Thai_Sentence = Last_Thai_Sentence + Trim(TextResult2)
    ' Output File 6
    Print #OutFileNum, TextResult
    Prv_Native1 = Wk_Native1
    Prv_Native2 = Wk_Native2
  Loop
  TextResult = String(1, " ")
  Print #OutFileNum, TextResult
  TextResult = "Thai Language Sentence"
  Print #OutFileNum, TextResult
  TextResult = String(40, "-")
  Print #OutFileNum, TextResult
  TextResult = Trim(Last_Thai_Sentence)
  Print #OutFileNum, TextResult

End Sub
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Layout Text File for Deconverter Program

```
Type UNL_Types
   RecNo As String * 4
   UWs As String * 40
   RLs As String * 20
   ObjectClass As String * 40
   ALs As String * 30
   CrLf As String * 2
End Type


Type UNL_Files
   RecNo As String * 4
   Category As String * 10
   AL_Code01 As String * 10
   AL_Code02 As String * 10
   AL_Code03 As String * 10
   AL_Code04 As String * 10
   AL_Code05 As String * 10
   UNL_Function As String
   Native As String * 256
   Native_Dictionary As String * 80
   CrLf As String * 2
End Type

Type UNL_Pattern
   RecNo As String * 4
   AL_Code01 As String * 10
   AL_Code02 As String * 10
   AL_Code03 As String * 10
   AL_Code04 As String * 10
   AL_Code05 As String * 10
   CrLf As String * 2
End Type

Type UNL_File3
   RecNo1 As String * 4
   Function1 As String * 10
   Value1 As String * 10
   RL_Code As String * 10
   RecNo2 As String * 4
   Function2 As String * 10
   Value2 As String * 10
   RL_Native_Language As String * 20
   RL_Native_Back As String * 20
   CrLf As String * 2
End Type

Type UNL_File4
   RecNo1 As String * 4
   Function1 As String * 10
   Value1 As String * 10
   RL_Code As String * 10
   RecNo2 As String * 4
   Function2 As String * 10
   Value2 As String * 10
   RL_Native_Language As String * 20
   RL_Native_Back As String * 20
   CrLf As String * 2
End Type


Type UNL_File5
   RecNo1 As String * 4       '1-4
   RecNo2 As String * 4       '5-8
   RLs As String * 10         '9-18
   Native1 As String * 256    '19-274
   Native2 As String * 256    '275-530
   RL_Native_Language As String * 20  '531-550
   RL_Native_Back As String * 20      '551-570
   CrLf As String * 2
End Type
```

ภาคผนวก ค

ตัวอย่างภาษา UNL และผลลัพธ์การแปลงเป็นภาษาไทย

ตัวอย่างประโยคภาษา UNL และผลลัพธ์ภาษาไทย

UNL [S] [W] day(icl>time).@generic.@entry:00 2(icl>number):01 week(icl>time).@generic:02 [/W] [R] 00num01 00per02 [/R] [/S] 2 days a week (2 <- num - day) - per -> week	2 วัน ต่อ สัปดาห์ Thai Language Sentence ---------- 2 วันต่อสัปดาห์
UNL [S] [W] apple(icl>plant).@generic.@entry:00 kg(icl>quantity).@generic:01 3(icl>number).@generic:02 [/W] [R] 00qua01 01num02 [/R] [/S] 3kg of apples apple - qua -> kg -> num -> 3	แอปเปิล หนัก 3 กิโลกรัม Thai Language Sentence ---------- แอปเปิลหนัก3 กิโลกรัม
UNL [S] [W] land(icl>airplane).@begin-soon.@entry:00 airplane(icl>finished).@generic:01 [/W] [R] 00agt01 [/R] [/S] The airplane is about to land. airplane <- agt - land.@begin-soon	เครื่องบิน กำลังจะ ลงจอด Thai Language Sentence ---------- เครื่องบินกำลังจะลงจอด
UNL [S] [W] approach(icl>action).@end-soon.@entry:00 we(icl>organization).@generic:01 goal(icl>event).@generic:02 [/W] [R] 00agt01 00gol02 [/R] [/S] We are approaching the goal. we <- agt - approach.@end-soon - gol -> goal	พวกเรา ใกล้จะถึง จุดหมาย Thai Language Sentence ---------- พวกเราใกล้จะถึงจุดหมาย

UNL [S] [W] borrow(icl>action).@present.@entry:00 book(icl>publications).@generic:01 go(icl>action).@present:02 library(icl>man-made-place).@generic:03 [/W] [R] 02gol03 00seq02 00obj01 [/R] [/S] to go to a library and borrow books book <- obj - borrow - seq -> go - gol -> library	ไป ห้องสมุด และ ยืม หนังสือ Thai Language Sentence --- ไปห้องสมุดและยืมหนังสือ
UNL [S] [W] rat(icl>animal).@generic.@entry:00 cat(icl>animal).@generic:01 animal(icl>living-thing).@generic:02 [/W] [R] 00and01 00aoj02 [/R] [/S] Cat and rats are animals. (cat <- and - rat) <- aoj - animal	หนู และ แมว เป็น สัตว์ Thai Language Sentence --- หนูและแมวเป็นสัตว์
UNL [S] [W] big(icl>quantity).@generic.@entry:00 cat(icl>animal).@generic:01 rat(icl>animal).@generic:02 [/W] [R] 00aoj01 00bas02 [/R] [/S] A cat is bigger than rat. cat <- aoj - big - bas -> rat	แมว ใหญ่ กว่า หนู Thai Language Sentence --- แมวใหญ่กว่าหนู
UNL [S] [W] jump(icl>action).@repeat.@entry:00 child(icl>person).@generic:01 [/W] [R] 00agt01 [/R] [/S] A child is jumping.	เด็ก กำลัง กระโดด Thai Language Sentence --- เด็กกำลังกระโดด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

child <- agt - jump.@repeat	
UNL [S] [W] die(icl>action).@progress.@entry:00 child(icl>person).@pl:01 hunger(icl>physiological phenomenon).@generic:02 [/W] [R] 00obj01 00cau02 [/R] [/S] Children are dying of hunger. child.@pl <- obj - die.@progress - cau -> hunger	เด็ก ๆกำลังจะ ตาย เพราะ ความหิว Thai Language Sentence ---------- เด็กๆกำลังจะตายเพราะความหิว
UNL [S] [W] calm(icl>natural).@present.@entry:00 sea(icl>topography).@generic:01 [/W] [R] 00soj01 [/R] [/S] The sea is calm. sea <- soj - calm	ทะเล สงบ Thai Language Sentence ---------- ทะเลสงบ
UNL [S] [W] come(icl>action).@present.@entry:00 london(icl>region).generic:01 [/W] [R] 00src01 [/R] [/S] to come from London. come - src -> London	มา จาก กรุงลอนดอน Thai Language Sentence ---------- มาจากกรุงลอนดอน
UNL [S] [W] come(icl>action).@present.@entry:00 see(icl>action).@present:01 I(icl>person).@generic:02 [/W] [R] 00pur01 01obj02 [/R] [/S] to come to see me come - pur ->see - obj -> I	มา เพื่อ พบ ฉัน Thai Language Sentence ---------- มาเพื่อพบฉัน
UNL	

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

[S] [W] compete(icl>action).@present.@entry:00 friend(icl>person).@generic:01 [/W] [R] 00ptn01 [/R] [/S] to compete with a friend compete - ptn -> friend	แข่งขัน กับ เพื่อน Thai Language Sentence ---------- แข่งขันกับเพื่อน
UNL [S] [W] solve(icl>action).@present.@entry:00 problem(icl>rational-action).@generic:01 most(icl>function).@generic:02 computer(mod>computer):03 easy(icl>quality):04 [/W] [R] 00agt03 01mod02 01man04 [/R] [/S] Most of these problems a computer could solve easily. solve - obj -> problem.@topic - mod -> most	คอมพิวเตอร์ แก้ ปัญหา โดยมาก ได้ ง่าย Thai Language Sentence ---------- คอมพิวเตอร์แก้ปัญหาโดยมากได้ง่าย
UNL [S] [W] cure(icl>doctor).@present.@entry:00 surgery(icl>person).@generic:01 [/W] [R] 00met01 [/R] [/S] to cure by surgery surgery <- met - cure	รักษาให้หาย โดย ศัลยแพทย์ Thai Language Sentence ---------- รักษาให้หายโดยศัลยแพทย์
UNL [S] [W] cut(icl>knife).@present.@entry:00 knife(icl>finished).@generic:01 [/W] [R] 00ins01 [/R] [/S] to cut with knife knife <- ins - cut	เฉือน ด้วย มีด Thai Language Sentence ---------- เฉือนด้วยมีด
UNL [S] [W]	ตาย เพราะ โรคมะเร็ง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

die(icl>action).@present.@entry:00 cancer(icl>physiological).@generic:01 [/W] [R] 00cau01 [/R] [/S] to die of cancer cancer <- cau - die	Thai Language Sentence --- ตายเพราะโรคมะเร็ง
UNL [S] [W] eat(icl>action).@present.@entry:00 apple(icl>food).@generic:01 [/W] [R] 00obj01 [/R] [/S] To eat an apple apple <- obj - eat	กิน ผลแอปเปิล Thai Language Sentence --- กินผลแอปเปิล
UNL [S] [W] cry(icl>sad).@progress.@entry:00 girl(icl>child).@generic:01 [/W] [R] 00agt01 [/R] [/S] A girl is crying girl <- agt - cry.@progress	เด็กหญิง กำลัง ร้องไห้ Thai Language Sentence --- เด็กหญิงกำลังร้องไห้
UNL [S] [W] go(icl>action).@present.@entry:00 london(icl>region).generic:01 [/W] [R] 00gol01 [/R] [/S] to go to London. go - gol -> London	ไป กรุงลอนดอน Thai Language Sentence --- ไปกรุงลอนดอน
UNL [S] [W] green(icl>color).@present.@entry:00 leave(icl>plant).@generic:01 [/W] [R] 00aoj01 [/R] [/S]	ใบไม้ เป็น สีเขียว Thai Language Sentence --- ใบไม้เป็นสีเขียว

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Leaves are green. leave <- aoj - green	
UNL [S] [W] do(icl>action).@past.@entry:00 he(icl>person).@generic:01 it(icl>event).@present:02 [/W] [R] 00agt01 01obj02 [/R] [/S] He did it. he <- agt - do.@past - obj -> it	เขา เคย ทำ เหตุการณ์ Thai Language Sentence ---------- เขาเคยทำเหตุการณ์
UNL [S] [W] know(icl>action).@pred.@entry:00 i(icl>person).@generic:01 go(icl>action).@pred:02 he(icl>person).@generic:03 [/W] [R] 00agt01 00obj03 02agt03 [/R] [/S] I know he goes. I <- agt - know.@pred - obj -> go.@pred - agt -> he	ฉัน รู้ ว่า เขา ไป Thai Language Sentence ---------- ฉันรู้ว่าเขาไป
UNL [S] [W] despise(icl>emotional-action).@present.@entry:00 i(icl>person).@generic:01 face(icl>body).@emphasis:02 his(icl>person):03 [/W] [R] 00agt01 00obj02 02pos03 [/R] [/S] His face I despise. I <- agt - despise - obj -> face.@emphasis	ฉัน ชัง หน้า ของ เขา Thai Language Sentence ---------- ฉันชังหน้าของเขา
UNL [S] [W] buy(icl>action).@present.@entry:00 I(icl>person).@generic:01 knife(icl>finished).@indef:02 [/W]	ฉัน ซื้อ มีด Thai Language Sentence

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

[R] 00agt01 00obj02 [/R] [/S] I bought a knife. I <- agt - buy - obj -> knife.@indef	---------- ฉันซื้อมีด
UNL [S] [W] buy(icl>action).@present.@entry:00 I(icl>person).@generic:01 knife(icl>finished).@pl:02 [/W] [R] 00agt01 00obj02 [/R] [/S] I bought knives. I <- agt - buy - obj -> knife.@pl	ฉัน ซื้อ มีด Thai Language Sentence ---------- ฉันซื้อมีดหลายเล่ม
UNL [S] [W] go(icl>action).@not.@entry:00 I(icl>person).@generic:01 [/W] [R] 00agt01 [/R] [/S] I do not go. I <- agt - go.@not	ฉัน ไม่ ไป Thai Language Sentence ---------- ฉันไม่ไป
UNL [S] [W] have(icl>action).@present.@entry:00 i(icl>person):01 3(icl>number):02 [/W] [R] 00agt01 00num02 [/R] [/S] I have three. I <- agt - have - obj -> - num -> 3	ฉัน มี 3 Thai Language Sentence ---------- ฉันมี3
UNL [S] [W] like(icl>action).@present.@entry:00 i(icl>person).@generic:01 orange(icl>plant).@generic:02 banana(icl>plant).@generic:03 apple(icl>plant).@generic:04	ฉัน ชอบ ส้ม และ กล้วย และ แอปเปิล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

[/W] [R] 00agt01 00obj02 02and03 03and04 [/R] [/S] I like apples, bananas and oranges. I <- agt - like.@pred.@entry - obj -> orange - and -> banana - and -> apple I <- agt - like.@pred.@entry - obj -> (orange.@sub - and -> banana - and -> apple)	Thai Language Sentence ---------- ฉันชอบส้มและกล้วยและแอปเปิล
UNL [S] [W] like(icl>action).@present.@entry:00 i(icl>person).@generic:01 orange(icl>plant).@sub:02 banana(icl>plant).@generic:03 apple(icl>plant).@generic:04 [/W] [R] 02and03 03and04 00agt01 [/R] [/S] I like apples, bananas and oranges. I <- agt - like.@pred.@entry - obj -> (orange.@sub - and -> banana - and -> apple)	ส้ม และ กล้วย และ แอปเปิล ฉัน ชอบ Thai Language Sentence ---------- ส้มและกล้วยและแอปเปิลฉันชอบ
UNL [S] [W] dog(icl>animal).@generic.@entry:00 john(icl>person):01 [/W] [R] 00pos01 [/R] [/S] John's dog dog - pos ->John	สุนัข ของ จอห์น Thai Language Sentence ---------- สุนัขของจอห์น
UNL [S] [W] come(icl>action).@end-just.@entry:00 he(icl>person).@generic:01 home(icl>man-made-place).@generic:02 [/W] [R] 00agt01 00gol02 [/R] [/S]	เขา เพิ่ง มา ถึง บ้าน Thai Language Sentence ---------- เขาเพิ่งมาถึงบ้าน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

He has just come home he <- agt - come.@end.just - gol -> home	
UNL [S] [W] keep(icl>action).@present.@entry:00 cat(icl>animal).@generic:01 dog(icl>animal).@generic:02 [/W] [R] 00obj01 01orr02 [/R] [/S] to keep a cat or dog keep - obj ->(cat<- or - dog)	เก็บ แมว หรือ สุนัข Thai Language Sentence ---------- เก็บแมวหรือสุนัข
UNL [S] [W] lunch(icl>action).@present.@entry:00 noon(icl>time).@generic:01 [/W] [R] 00tim01 [/R] [/S] to lunch at noon lunch - tim -> noon	กินอาหารกลางวัน ตอน เที่ยงวัน Thai Language Sentence ---------- กินอาหารกลางวันตอนเที่ยงวัน
UNL [S] [W] make(icl>action).@present.@entry:00 butter(icl>finished).@generic:01 milk(icl>part-animal).@generic:02 [/W] [R] 00obj01 00mat02 [/R] [/S] to make butter from milk butter <- obj - make - mat -> milk	ทำ เนยเหลว ด้วย น้ำนม Thai Language Sentence ---------- ทำเนยเหลวด้วยน้ำนม
UNL [S] [W] melt(icl>natural).@present.@entry:00 snow(icl>liquid).@generic:01 [/W] [R] 00aoj01 [/R] [/S] Snow melts. snow <- aoj - melt	หิมะ ละลาย Thai Language Sentence ---------- หิมะละลาย
UNL	

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

[S] [W] fall(icl>action).@present.@entry:00 monkey(icl>animal).@generic:01 branch(icl>plant).@generic:02 [/W] [R] 00agt01 01src02 [/R] [/S] The monkey fell from a branch. monkey.@def <- agt - fall - src -> branch	ลิง ตก จาก กิ่งก้าน Thai Language Sentence ---------- ลิงตกจากกิ่งก้าน
UNL [S] [W] like(icl>action).@present.@entry:00 monkey(icl>animal).@generic:01 banana(icl>food).@generic:022 [/W] [R] 00agt01 01obj02 [/R] [/S] A monkey likes bananas. monkey.@generic <- agt - like - obj -> banana	ลิง ชอบ กล้วย Thai Language Sentence ---------- ลิงชอบกล้วย
UNL [S] [W] die(icl>action).@begin-soon.@entry:00 father(icl>family).@generic:01 i(icl>person).@generic:02 [/W] [R] 01pos02 00obj01 [/R] [/S] My father is dying. I <- pos - father <- obj - die.@begin-soon	พ่อ ของ ฉัน กำลังจะ ตาย Thai Language Sentence ---------- พ่อของฉันกำลังจะตาย
UNL [S] [W] orange(icl>plant).@not.@entry:00 it(icl>thing).@generic:01 [/W] [R] 00aoj01 [/R] [/S] It is not an orange. It <- aoj - orange.@not	มัน ไม่ใช่ ส้ม Thai Language Sentence ---------- มันไม่ใช่ส้ม
UNL [S]	

[W] play(icl>action).@complete.@entry:00 she(icl>person).@generic:01 piano(icl>finished).@generic:02 [/W] [R] 00agt01 00obj02 [/R] [/S] She has played a piano. she <- agt - play.@complete - obj -> piano	เธอ เล่น เปียโน Thai Language Sentence --- เธอเล่นเปียโนแล้ว
UNL [S] [W] play(icl>action).@present.@entry:00 garden(icl>man-made-place).@generic:01 [/W] [R] 00ppl01 [/R] [/S] to play in a garden play - ppl ->garden	เล่น ใน สวน Thai Language Sentence --- เล่นในสวน
UNL [S] [W] break(icl>pot).@state.@entry:00 pot(icl>rice).@generic:01 [/W] [R] 00obj01 [/R] [/S] The pot has broken. pot <- obj - break.@state	หม้อ แตก Thai Language Sentence --- หม้อแตกแล้ว
UNL [S] [W] run(icl>action).@present.@entry:00 rabbit(icl>animal).@generic:01 [/W] [R] 00agt01 [/R] [/S] A rabbit runs. rabbit <- agt - run	กระต่าย วิ่ง Thai Language Sentence --- กระต่ายวิ่ง
UNL [S] [W] representation(icl>action).@title.@entry:00 structure(icl>immaterial).@generic:01 sentence(icl>immaterial).@generic:02 [/W]	แสดง โครงสร้าง ของ ประโยค Thai Language Sentence

[R] 00mod01 01mod02 [/R] [/S] Representation of sentence structure representation@title - mod -> structure - mod -> sentence	---------- แสดงโครงสร้างของประโยค
UNL [S] [W] return(icl>action).@present.@entry:00 home(icl>man-made-place).@generic:01 cry(icl>action).@present:02 [/W] [R] 00gol01 00coo02 [/R] [/S] to return home crying home <- gol - return - coo -> cry	กลับ บ้าน แล้วกำลัง ร้องไห้ Thai Language Sentence ---------- กลับบ้านแล้วกำลังร้องไห้
UNL [S] [W] return(icl>action).@present.@entry:00 home(icl>man-made-place).@generic:01 rain(icl>natural-fall).@generic:02 [/W] [R] 00gol01 00con02 [/R] [/S] to return home because of rain home <- gol - return - con -> rain	กลับ บ้าน เพราะว่า ฝนตก Thai Language Sentence ---------- กลับบ้านเพราะว่าฝนตก
UNL [S] [W] run(icl>action).@present.@entry:00 fast(icl>quality).@generic:01 [/W] [R] 00man01 [/R] [/S] to run fast fast <- man - run	วิ่ง อย่าง เร็ว Thai Language Sentence ---------- วิ่งอย่างเร็ว
UNL [S] [W] language(icl>expression).@present.@entry.@pred:00 UNL(icl>label).@topic:01 common(mod>share):02 use(icl>action).@present.@pred:03 language(icl>expression).@present.@entry.@pred:04	UNL เป็น ภาษา กลาง ใช้ เพื่อ ติดต่อ ด้วย เครือข่ายคอมพิวเตอร์

สำนักวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

communication(icl>information-transfer).@pl:05 network(mod>computer):06 [/W] [R] 00aoj01 00mod02 03obj04 03pur05 05mod06 [/R] [/S] UNL is a common language tthat would be used for network communications. UNL <- aoj - language - mod -> common language <- obj - use - pur -> communication - mod -> network	Thai Language Sentence ---------- UNLเป็นภาษากลางใช้เพื่อติดต่อด้วยเครือข่ายคอมพิวเตอร์
[S] [W] eat(icl>action).@present.@entry:00 monkey(icl>animal).@generic:01 banana(icl>food).@generic:02 [/W] [R] 00agt01 00obj02 [/R] [/S] Monkey eats bananas. monkey <- agt - eat - obj -> banana	ลิง กิน กล้วย Thai Language Sentence ---------- ลิงกินกล้วย
UNL [S] [W] come(icl>action).@future.@entry:00 she(icl>person).@generic:01 9(icl>number):02 [/W] [R] 00agt01 00tim02 [/R] [/S] She will come at 9. she <- agt - come.@future - tim -> 9	เธอ กำลังจะ มา ตอน 9 Thai Language Sentence ---------- เธอกำลังจะมาตอน9
UNL [S] [W] go(icl>action).@present.@entry:00 she(icl>person).@generic:01 school(icl>man-made-place).@generic:02 9(icl>number).@generic:03 [/W] [R] 00agt01 00gol02 00tim03 [/R]	เธอ ไป โรงเรียน ตอน 9 Thai Language Sentence ---------- เธอไปโรงเรียนตอน9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

[/S] She goes to school at 9. she <- agt - go.@present - gol -> school go.@present - tim -> 9	
UNL [S] [W] begin(icl>action).@begin-just.@entry:00 match(icl>social-phenomenon).@generic:01 [/W] [R] 00obj01 [/R] [/S] The match has just begun. match <- obj - begin.@begin-just	การแข่งขันกีฬา เพิ่งจะ เริ่มต้น Thai Language Sentence ---------- การแข่งขันกีฬาเพิ่งจะเริ่มต้น
UNL [S] [W] man(icl>person).@generic.@entry:00 third(icl>number).@generic:01 [/W] [R] 00mod01 [/R] [/S] the third man man - mod -> third	ผู้ชาย ที่สาม Thai Language Sentence ---------- ผู้ชายที่สาม
UNL [S] [W] mile(icl>quantity)@generic.@entry:00 3(icl>number)@generic:01 [/W] [R] 00num01 [/R] [/S] three miles mile - num -> 3	3 ไมล์ Thai Language Sentence ---------- 3 ไมล์
UNL [S] [W] ticket(icl>finished).@generic.@entry:00 oxford(icl>region).@generic:01 london(icl>region).@generic:02 [/W] [R] 00mod02 01fmt02 [/R] [/S] a ticket from London to Oxford ticket - mod -> (London <- fmt - Oxford)	ตั๋ว จาก กรุงลอนดอน ออกซฟอร์ด Thai Language Sentence ---------- ตั๋วจากกรุงลอนดอนไปออกซฟอร์ด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

UNL [S] [W] beer(icl>finished).@generic.@entry:00 pint(icl>quantity):01 2(icl>number):02 [/W] [R] 00qua01 01num02 [/R] [/S] two pints of beer beer - qua -> pint - num -> 2	เบียร์ หนัก 2 ไพนท์ Thai Language Sentence ---------- เบียร์หนัก2 ไพนท์
UNL [S] [W] walk(icl>action).@present.@entry:00 road(icl>man)@generic:01 [/W] [R] 00opl01 [/R] [/S] to walk on the road walk - opl -> road	เดิน บน ถนน Thai Language Sentence ---------- เดินบนถนน
UNL [S] [W] walk(icl>action).@present.@entry:00 friend(icl>person).@generic:01 [/W] [R] 00cag01 [/R] [/S] to walk with a friend walk - cag -> friend	เดิน ไปกับ เพื่อน Thai Language Sentence ---------- เดินไปกับเพื่อน
UNL [S] [W] win(icl>emotional).@present.@entry:00 competition(icl>action).@generic:01 [/W] [R] 00lpl01 [/R] [/S] to win in competition win - lpl ->competition	ชนะ ใน การแข่งขัน Thai Language Sentence ---------- ชนะในการแข่งขัน
UNL [S] [W] work(icl>action-man).@present.@entry:00 family(icl>organization).@generic:01	ทำงาน เพื่อ ครอบครัว Thai Language Sentence

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

[/W] [R] 00ben01 [/R] [/S] to work for family work - ben -> family	---------- ทำงานเพื่อครอบครัว
UNL [S] [W] work(icl>action-man).@present.@entry:00 noon(icl>time).@generic:01 [/W] [R] 00tmf01 [/R] [/S] to work from noon work - tmf -> noon	ทำงาน ตั้งแต่ เที่ยงวัน Thai Language Sentence ---------- ทำงานตั้งแต่เที่ยงวัน
UNL [S] [W] work(icl>action).@present.@entry:00 noon(icl>time).@generic:01 [/W] [R] 00tmt01 [/R] [/S] to work until noon work - tmt -> noon	ทำงาน จนถึง เที่ยงวัน Thai Language Sentence ---------- ทำงานจนถึงเที่ยงวัน
UNL [S] [W] go(icl>action).@interrogation.@entry:00 you(icl>person).@focus:01 [/W] [R] 00agt01 [/R] [/S] Is that you who went? go.@interrogation -> agt -> you.@focus	คุณ หรือเปล่าที่ ไป Thai Language Sentence ---------- คุณหรือเปล่าที่ไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค

Data Dictionary

Table Name : UNL_AL_Type

Description : UNL Attribute Label Type

Columns

Name	Type	Size	Description
AL_Type	Text	2	UNL Attribute Labels Type
Desc	Text	64	Description

Table Indexes

Name	Number of Fields
Primary Key	1
Fields:	AL_Type, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Attribute_Labels

Description : UNL Attribute Labels

Columns

Name	Type	Size	Description
AL_Code	Text	20	Attribute Labels
Desc	Text	64	Description
Type	Text	2	Type
Eng_Example	Memo	-	English Example
UNL_Example	Memo	-	UNL Example
Native_Desc	Text	50	Native Description

Table Indexes

Name	Number of Fields
AL_Code	1
Fields:	AL_Code, Ascending
PrimaryKey	1
Fields:	AL_Code, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Category

Description : UNL Category

Columns

Name	Type	Size	Description
Category	Text	10	Category Type
Name	Text	20	English Name
Native_Name	Text	20	Native Name
Native_Desc	Memo	-	Native Description

Table Indexes

Name	Number of Fields
PrimaryKey	1
Fields:	Category, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Dictionary

Description : UNL Dictionary

Columns

Name	Type	Size	Description
UW_Keyword	Text	20	Universal Words
RL_Code	Text	10	Relation Labels
Object_Class	Text	40	Object Class
Subject_Object_Class	Text	40	Subject Object Class
Native_Language	Text	50	Native Language
Category	Text	10	Category

Table Indexes

Name	Number of Fields
PrimaryKey	4
Fields:	UW_Keyword, Ascending
	RL_Code, Ascending
	Object_Class, Ascending
	Subject_Object_Class, Ascending
Unl_DictionaryCategory	1
Fields:	Category, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Function

Description : UNL Function

Columns

Name	Type	Size	Description
Function_Code	Text	10	Function Code
Name	Text	50	Name
Native_Desc	Text	65	Native Description
Values	Number (Long)	4	Value

Table Indexes

Name	Number of Fields
Function_Code	1
Fields:	Function_Code, Ascending
PrimaryKey	1
Fields:	Function_Code, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Object_Class

Description : UNL Object Class

Columns

Name	Type	Size	Description
Object_Class	Text	40	Object Class
Object_Name	Text	40	Object Name
Native_Name	Text	40	Native Name
Native_Desc	Text	100	Native Description

Table Indexes

Name	Number of Fields
PrimaryKey	1
Fields:	Object_Class, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Relation_Labels

Description : UNL Relation Labels

Columns

Name	Type	Size	Description
RL_Code	Text	5	Relation Labels Code
RL_Name	Text	20	Relation Labels Name
RL_Native_Name	Text	30	Relation Labels Native Name
Desc	Text	100	Description
Native_Desc	Text	100	Native Description
Type	Text	1	Type
Eng_Example	Memo	-	Relation Labels Example
Unl_Example	Memo	-	UNL Example

Table Indexes

Name	Number of Fields
PrimaryKey	1
Fields:	RL_Code, Ascending
RL_Code	1
Fields:	RL_Code, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Sub_Category

Description : UNL Sub Category

Columns

Name	Type	Size	Description
Category	Text	10	Category Type
Sub_Category	Text	10	Sub Category Type
Name	Text	45	English Name
Native_Name	Text	45	Native Name
Native_Desc	Memo	-	Native Description
Native_Example	Memo	-	Native Example

Table Indexes

Name	Number of Fields
Category	1
Fields:	Category, Ascending
PrimaryKey	2
Fields:	Category, Ascending
Sub_Category, Ascending	
Sub_Category	1
Fields:	Sub_Category, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Universal_Words

Description : UNL Universal Words

Columns

Name	Type	Size	Description
UW_Keyword	Text	40	Universal Words
Native_Desc	Memo	-	Native Description

Table Indexes

Name	Number of Fields
PrimaryKey	1
Fields:	UW_Keyword, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Verb_Pattern

Description : UNL Verb Pattern

Columns

Name	Type	Size	Description
Category1	Text	10	Category 1
AL_Code01	Text	10	Attribute Labels Code 1
AL_Code02	Text	10	Attribute Labels Code 2
AL_Code03	Text	10	Attribute Labels Code 3
AL_Code04	Text	10	Attribute Labels Code 4
AL_Code05	Text	10	Attribute Labels Code 5
RL_Code	Text	10	Relation Label Code
Category2	Text	10	Category 2
AL_Code06	Text	10	Attribute Labels Code 6
AL_Code07	Text	10	Attribute Labels Code 7
AL_Code08	Text	10	Attribute Labels Code 8
AL_Code09	Text	10	Attribute Labels Code 9
AL_Code10	Text	10	Attribute Labels Code 10
Function_Code1	Text	10	Function Code 1
RL_Code2	Text	5	Relation Labels Code 2
Function_Code2	Text	10	Function Code 2
RL_Native_Language	Text	20	Relation Labels Native Language
RL_Native_Back	Text	20	Relation Labels Native Language (End of Sentence)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Table Name : UNL_Verb_Pattern

Description : UNL Verb Pattern

Table Indexes

Name	Number of Fields
AL_Code01	1
Fields:	AL_Code01, Ascending
AL_Code02	1
Fields:	AL_Code02, Ascending
AL_Code03	1
Fields:	AL_Code03, Ascending
AL_Code04	1
Fields:	AL_Code04, Ascending
AL_Code05	1
Fields:	AL_Code05, Ascending
AL_Code06	1
Fields:	AL_Code06, Ascending
AL_Code07	1
Fields:	AL_Code07, Ascending
AL_Code08	1
Fields:	AL_Code08, Ascending
AL_Code09	1
Fields:	AL_Code09, Ascending
AL_Code10	1
Fields:	AL_Code10, Ascending
Category1	1
Fields:	Category1, Ascending
Category2	1
Fields:	Category2, Ascending
RL_Code1	1
Fields:	RL_Code, Ascending

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียน

นายสมชาย กิติภรณ์ เกิดวันที่ 13 ธันวาคม 2509 ที่กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาตรีบริหารธุรกิจ สาขาวิชาระบบสารสนเทศ คณะบริหารธุรกิจ สถาบันเทคโนโลยีราชมงคล ในปีการศึกษา 2533 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ ที่จุฬาลงกรณ์มหาวิทยาลัยเมื่อ พ.ศ. 2540

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย