INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Generate and Repair Machine Translation

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE
UNIVERSITY OF REGINA

By

Kanlaya Naruedomkul
Regina, Saskatchewan
April, 2000

© Copyright 2000: Kanlaya Naruedomkul
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

CERTIFICATION OF THESIS WORK

We, the undersigned, certify that Kanlaya Naruedomkul, candidate for the Degree of Doctor of Philosophy, has presented a thesis on Generate and Repair Machine Translation, that the thesis is acceptable in form and content, and that the student demonstrated a satisfactory knowledge of the field covered by the thesis in an oral examination held on June 23, 2000.

External Examiner:

Dr. G. LaPalme, University of Montreal

Internal Examiners:

Dr. N. Cerbone, Supervisor

R. S. Maguire

(teleconference call)
UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

PERMISSION TO USE POSTGRADUATE THESIS

TITLE OF THESIS: Generate and Repair Machine Translation

NAME OF AUTHOR: Kanlaya Naruedomkul

FACULTY: Faculty of Graduate Studies and Research

DEGREE: Doctor of Philosophy

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Regina, I agree that the Libraries of this University shall make it freely available for inspection. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the Faculty in which my thesis work was done. It is understood that with the exception of UMI Dissertations Publishing (UMI) that any copying, publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Regina in any scholarly use which may be made of my material in my thesis.

SIGNATURE: _____________________________

DATE: ________________

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
To my parents, sisters, brothers, and my two lovely nephews
Abstract

We propose Generate and Repair Machine Translation (GRMT), a constraint-based approach to machine translation (MT) that focuses on accurate translation output. The architecture of GRMT was designed to take advantage of, and have advantages over, the three classic strategies (Direct MT, Interlingual MT and Transfer MT), the nonlinguistic information strategies (Example-Based MT and Statistics-Based MT), and the hybrid strategies (Knowledge-Based MT and Shake-and-Bake MT) with respect to several translation aspects: simplicity, accuracy and multilingualism.

GRMT performs the translation by generating a Translation Candidate (TC), verifying the syntax and semantics of the TC, and repairing the TC when required. GRMT comprises three modules: Analysis Lite Machine Translation (ALMT), Translation Candidate Evaluation (TCE), and Repair and Iterate (RI).

In generating the TC, GRMT refines the scope of translation choices of each input word by taking into account the differences between languages in a unique way. In selecting an appropriate word for each input word, GRMT considers the semantic relationship between words. This semantic relationship is based on the Word Association (WordAsso) number. WordAsso number is assigned to word class. Words are classified according to the meaning of words and their usage. Word classification is designed and used not only in the word selection process but also in the classifier selection process and in semantic representation.

GRMT is highly modular and extendible in the following respects: each component is separated, not only the translation process components (ALMT, TCE, RI), but also in the knowledge-bases, each component can be extended easily to a larger domain. The adding of new languages is also possible since the source language (SL) and the target language (TL) are treated separately. The SL and the TL are connected via the SL-TL dictionary which contains simple information and is manageable.
An English-Thai translation system has been implemented to illustrate the performance of GRMT. The system has been developed and run under SWI-Prolog 3.2.8. The English and Thai grammars have been developed based on the Head-Driven Phrase Structure Grammar (HPSG) and implemented on the Attribute Logic Engine (ALE).

This English-Thai MT system was evaluated and it performs in the way we intended. ALMT generated acceptable translations (grammatically correct, correct word usage and convey the original meaning) for 47 out of the 90 sentences in the test corpus without repair. TCE and RI improved 15 sentences using our current HPSG based grammars and lexicons. Twenty-one sentences which contain logical connections are first separated into linguistic units before the repair can be performed due to a current inadequacy in HPSG's semantic representation. However, each linguistic unit was then repaired successfully. Seven sentences faced with the problems of adding linking words and classifiers in Thai also require further research in order to develop ways to repair these sentences.
Acknowledgments

I would like to acknowledge the following people who, at different times, have helped me with this thesis. My deepest gratitude and appreciation goes to my advisor, Prof. Nick Cercone, who has given me a great deal of freedom to pursue ideas and research topics and who is always there when I need help and guidance. His unfailing support, his patience, his constructive criticism and his vast experience and invaluable advice have made the writing of this thesis possible and enjoyable.

I would like to thank Dr. Guy Lapalme, the external examiner, Dr. Brent Galloway, Dr. Howard Hamilton and Dr. Brien Maguire, the members of my thesis committee, for the remarkable feedback received during my viva. I also would like to thank Dr. Klaus H. Burmeister for chairing the defense.

I am thankful to Dr. Booncharoen Sirinaovakul, whose thoughtful, inventive suggestions are woven into the fabric of this work, and to whom I am deeply grateful for countless conversations and generous sharing of ideas and for explaining complex HPSG ideas with clarity and enthusiasm.

I would like to thank Dr. Douglas Arnold, Dr. Gerald Gazdar, Dr. John Luis Beaven, Dr. Carl Pollard, and Dr. Fred Popowich, for providing me advice and the materials which are very helpful to my research, and Dr. Bob Carpenter and Dr. Gerald Penn for their help with my questions about ALE.

I would like to thank the Department of Computer Science, University of Waterloo for providing me with the office and computer resources, to the academic, administrative, computing and research staff in the Department of Computer Science and Graduate Studies at University of Regina for all the support they have given me, to Dr. Paitoon Tontiwachwuthikul and Dr. Christine Chan for making my first few days in Regina a very pleasant experience.
My gratitude also goes to Assoc. Prof. Nardtida Tumrasvin and Assoc. Prof. Chinda Achariyakul, Department of Mathematics, Mahidol University, Prof. Amaret Bhumiratana, Doctor Kattaporn and Ms. Patrarawadee for supporting me to pursue my doctorate study in Computer Science in Canada. I am also thankful to Dr. Virach Sornlertlamvanich, Nisanard Tannin and Kaewchai Chancharoen for providing me with the information needed for the study.

It has been a privilege indeed to have worked with these people. Also, my stay in Canada, the beautiful country, would not have been as pleasant without these fine people—Ursula Thoene, Chris Blair, Doreen Thiessen, Marnie Jeworski, Sel Murray, Dr. Brian Sveinson and Dr. Larry Saxton, I am grateful for their friendship and support.

I thank also my friends over the years, Weerapong Kritpiphat, Narate Lertpalangsunti, Amornvadee Veawab, Adisorn Aroonwilas, Teeradet Supap, Aijun An, Jessie Hou, Jianchao Han, Jian Jianna Zhang and Yingwei Wang.

I also would like to thank Margaret and Gordon, the Cooper family, for their kindness which will be forever remembered.

I also acknowledge the generous financial support I received while writing this dissertation, including The Royal Thai Government, Natural Sciences and Engineering Research Council (NSERC) and The Institute for Robotics and Intelligent Systems (IRIS).

I am thankful to the improvements in Computer Science and related fields which are very helpful to my research and thesis writing.

Lastly, I would like to express my thanks and appreciation to my parents Kiat and Somjit, my sisters Patcharee and Daranee, my brothers Taworn and Pornchai, and my nephews Sirawit and Potchara for their love and understanding.
# Contents

Abstract                                                                                           i  
Acknowledgements                                                                                   iii  
Table of Contents                                                                                  v  
List of Tables                                                                                      ix  
List of Figures                                                                                     xi  
List of Abbreviations and Acronyms                                                                 xiv  

## Chapter 1 Introduction
1.1 Why Machine Translation                                                                   1  
1.2 Objectives                                      3  
1.3 Outline of the thesis                       4  

## Chapter 2 Machine Translation
2.1 What is Machine Translation?                                      6  
2.2 A Brief History of Machine Translation                                 8  
2.3 Machine Translation Methodologies/Technologies                         12  
  2.3.1 Methodologies used to develop MT systems                               12  
  2.3.2 Techniques used to increase the performance of MT                   14  
  2.3.2.1 Post-editing, Pre-editing and Interaction                         14  
  2.3.2.2 Lexicon constructions and parsing techniques                      17  
  2.3.3 Discussion of some existing MT methodologies                           18  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CONTENTS

2.3.3.1 Three Classic Strategies 18
2.3.3.2 Nonlinguistic Information Strategies 23
2.3.3.3 Hybrid Strategies 25
2.4 Our approach: Generate and Repair Machine Translation 28

Chapter 3 Generate and Repair Machine Translation 30
3.1 Why Generate and Repair, Not Analyze and Generate? 32
3.2 The Three Phases 35
3.3 Knowledge-bases Required 36
3.4 Translation Examples by GRMT 36
3.5 Key Features of GRMT 39

Chapter 4 Differences Between Languages: The English and Thai Languages 41
4.1 Differences between Thai and English 41
4.2 Word Classification 46
  4.2.1 Word Classification Criteria 52
  4.2.2 The semantic relationship development in GRMT 56

Chapter 5 Analysis Lite Machine Translation 58
5.1 Word Treatment 59
  5.1.1 SL Constraints Application 59
  5.1.2 Dictionary Look-up 60
  5.1.3 Dictionaries 62
5.2 Word Selection 63
5.3 Word Ordering 71
  5.3.1 Word Addition 71
  5.3.2 Word Ordering 73
5.4 Example of ALMT 75
CONTENTS

Chapter 6 Translation Candidate Evaluation 80
  6.1 The Analyzer 81
    6.1.1 Parsing 81
      6.1.1.1 About Head-Driven Phrase Structure Grammar 81
      6.1.1.2 Words and Phrases as Feature Structures 82
      6.1.1.3 Lexicon 85
      6.1.1.4 Some Universal Principle and Grammar Rules 87
      6.1.1.5 Attribute Logic Engine 92
      6.1.1.6 English and Thai Grammars in GRMT 94
    6.1.2 Semantic Extraction 104
      6.1.2.1 Semantic Representation 104
      6.1.2.2 Semantic Extraction Examples 106
  6.2 Semantic Comparison 109

Chapter 7 Repair and Iteration 116

Chapter 8 Machine Translation Evaluation 125
  8.1 What needs to be evaluated? (Types of Evaluation) 126
  8.2 MT Evaluation Strategies 128
    8.2.1 Error Analysis 129
    8.2.2 Test Suite 129
    8.2.3 The ARPA MT Evaluation 131
  8.3 GRMT Evaluation 132

Chapter 9 Discussion 142
  9.1 Contributions 142
  9.2 Future Research 145

Bibliography 147
## CONTENTS

### Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>WordAsso hierarchy</td>
<td>161</td>
</tr>
<tr>
<td>B</td>
<td>SL Dictionary</td>
<td>185</td>
</tr>
<tr>
<td>C</td>
<td>SL-TL Dictionary</td>
<td>190</td>
</tr>
<tr>
<td>D</td>
<td>Types, their features and their value types</td>
<td>195</td>
</tr>
<tr>
<td>E</td>
<td>Lexical Rules</td>
<td>198</td>
</tr>
<tr>
<td>F</td>
<td>Grammars</td>
<td>201</td>
</tr>
<tr>
<td>G</td>
<td>Intelligibility and Fidelity Evaluation</td>
<td>205</td>
</tr>
<tr>
<td>H</td>
<td>Comprehensibility Evaluation</td>
<td>225</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Translation examples by the commercial MT system 22
3.1 The TC and selected words generated for Example 1 34
3.2 The selected words and the TC generated for Example 2 37
3.3 Three corresponding Thai words to the word spring 38
3.4 The selected words and the TC generated for Example 3 38
4.1 Descriptions and the English corresponding words of KHON, KHAW and THII 42
4.2 The additional words that express Tense, Passive Voice and Plurality in Thai 43
4.3 Some information on the words NAM, SONG, PAH and PATRA 49
4.4 Semantic relationships assigned for noun-noun combination in C1CC MMT 49
4.5 The subgroup of vehicle 54
4.6 Some classified English and Thai words 55
4.7 Examples of words and their contexts 56
4.8 Classes and their sample words 57
5.1 Some SL Constraints 59
5.2 Example result of Dictionary Look-up of Example 3 61
5.3 The possible corresponding Thai words for SPIRIT, WILDLING, FLESH and WEAK 69
5.4 The possible corresponding Thai words for LEFT, DUCKLING and MARSH 70
5.5 A fraction of classifier classes 72
5.6 The selected words in Thai for the words woman, cat and hen 73
5.7 ALMT steps applied to the sentence of Example 11 76
5.8 ALMT steps applied to the sentence of Example 12 78
6.1 Some types, their features and their value types 84
6.2 Features and their value types of the types noun and verb 86
6.3 The class of cover 115
7.1 A part of Word Treatment output of Example 1 119
7.2 A part of Word Treatment output of Example 2 123
8.1 The intelligibility and fidelity evaluation result of evaluating ALMT on the 90 sentences test corpus 136
8.2 The comprehensible evaluation result 136
9.1 Advantages and disadvantages of the existing MT approaches 143

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## List of Figures

2.1 Direct MT 18  
2.2 Some translations of the word *open* 19  
2.3 Interlingual MT 20  
2.4 Transfer MT 21  
2.5 Example based MT 24  
3.1 GRMT Architecture 30  
3.2 The translation between languages of two language families 40  
4.1 Examples of *nî* (kin) and its subjects 43  
4.2 Examples of additional words to indicate different tenses and aspects 44  
4.3 Examples of passive voice in Thai 44  
4.4 Examples of classifiers 45  
4.5 Examples of Thai attributive constructions 45  
4.6 Examples of Thai constructions without auxiliary verb 46  
4.7 A part of Thai Concept Classification (CICC MMT) 47  
4.8 A part of Indonesian Concept Classification (CICC MMT) 48  
4.9 A co-occurrence of the word *iymt* (phra*) [CICC 1995i] 49  
4.10 A partition type hierarchy for *thing* [Cercone et al. 1992] 51  
4.11 A part of word classification in GRMT 53  
5.1 ALMT Architecture 58  
5.2 SL constraints applications examples 60  
5.3 Example of SL Dictionary 62  
5.4 Example of SL-TL Dictionary 62
5.5 The relationship between the proximity words -Examples 4 and 5 63
5.6 Examples of WordAsso relations 64
5.7 Example of selection steps 67
5.8 Examples of WordAsso relations 69
5.9 Examples of classifier relations 72
5.10 Word orders of 74
5.11 Word orders of 75
5.12 A part of "The ugly duckling" and its translation 79
6.1 TCE Architecture 80
6.2 The lexical entry hide 83
6.3 The phrase Leslie drinks milk 85
6.4 The 3rd-singular verb lexical rule 86
6.5 Example output, hides, of the 3rd-singular verb lexical rule 87
6.6 HFP applied to Leslie drinks milk 88
6.7 Semantic principle applied to Leslie drinks milk 90
6.8 Head_complement rule applied to likes John 91
6.9 Specifier_head rule applied to the book 92
6.10 Example of result of parsing in ALE representation 94
6.11 The lexical entry for hide 98
6.12 The representation of hide in ALE with some macros 99
6.13 The parse syntax of Example 1 and that of its translation 100
6.14 The representation of under in ALE with some macros 101
6.15 The representation of "khiirëe- ugly" in ALE 101
6.16 The representation of the classifier tua in ALE 102
6.17 The parse syntax of Example 2 and that of its translation 103
6.18 The partial description of the lexical entry the 104
6.19 The partial description of the lexical entry Joan 105
6.20 The semantic information, in ALE's internal structure, of the SL in Example 3 107
6.21 The semantic information of the SL in Example 3 107
List of Abbreviations and Acronyms

AI Artificial Intelligence
ALE Attribute Logic Engine
ALMT Analysis Lite Machine Translation
ALPAC Automatic Language Processing Advisory Committee
ALPS Automated Language Processing Systems
ATR-ITL Advanced Telecommunications Research, Interpreting Telecommunication Research Laboratories
AVM Attribute-Value Matrices
BYU Brigham Young University
CAT Computer-Aided Translation
CEC The Commission of the European Communities
CG Categorial Grammar
CICC Center of the International Cooperation for Computerization
CKY Cocke-Kasami-Younger
CLRU The Cambridge Language Research Unit
CT Correct Translation
DBMT Dialogue-based MT
EBMT Example-Based Machine Translation
GPSG Generalized Phrase Structure Grammar
GRMT Generate and Repair Machine Translation
HAMT Human-Aided Machine Translation
HFP Head Feature Principle
HPSG Head-Driven Phrase Structure Grammar
ISI University of Southern California's Information Sciences Institute
ITS Interactive Translation System
KBMT Knowledge-Based Machine Translation
LFG Lexical Functional Grammar
MAHT Machine-Aided Human Translation
MAT Machine-Aided Translation
MCE Multinational Customised English
MIT The Massachusetts Institute of Technology
MMT Multilingual Machine Translation
MT Machine Translation
NLP Natural Language Processing
PAHO Pan American Health Organization
RI Repair and Iterate
SAM Script Applying Mechanism
SL Source Language
TAUM Traduction Automatique de l'Université de Montréal
TC Translation Candidate
TCE Translation Candidate Evaluation
TL Target Language
TSNLP Test Suites for Natural Language Processing
WordAsso Word Association (Number)
Chapter 1

Introduction

Translation, from one language into another, presents subtle, important, ancient, and difficult problems.
— Warren Weaver (1955)

A primary objective of this research is to uncover and deploy new solutions that apply when developing an efficient machine translation (MT) system. We propose an alternative constraint-based approach to machine translation that focuses on accurate translation output. We call this approach Generate and Repair Machine Translation (GRMT) and we have implemented an advanced GRMT prototype for demonstration and evaluation purposes to translate English into Thai.

Successful results from this research should prove beneficial to machine translation studies currently underway around the world. GRMT’s application should minimize the language barrier and simplify communication between peoples who speak different languages.

1.1 Why Machine Translation?

Country to country exchanges in trade, technologies, politics, telecommunications, culture, etc., continue to grow rapidly. Language plays a significant role in the communication processes between nations. To understand what is communicated, it is necessary to understand the language used in the communications process, its nuances and subtleties, hence, accurate machine translation assumes importance in order to facilitate this communications process from one language into another. When exchanges become more
globalized, the requirement for accurate machine translation increases. In multilingual countries like Canada (French-English), Switzerland (German, Italian, French, Swiss), India (32 official languages), the need for translation exists apart from increasingly globalized communications.

In countries where multiple official languages are used, translation is necessary to communicate with others who speak different languages. Translation enables people to express themselves in the way in which they wish to express themselves and to obtain the type of information they desire. Choice of language for communicative and other aspects is important in order for people to retain and enhance distinctive cultures and distinctive ways of thinking; language loss should matter to everyone.

The advance in computer technologies and the explosive growth of the World Wide Web has brought people closer together. Multilingual text has reached nearly to everyone with a computer. Information is readily accessible in growing numbers of languages. Machine translation has been integrated into man-machine communication systems which include electronic mail, information retrieval and the internet. The demand for machine translation systems has, therefore, not surprisingly, increased.

It would appear that successful machine translation efforts become a social and political necessity for modern societies (intelligent communities) that do not wish to impose a common language on their constituents. Past efforts at standardization, e.g., Esperanto, have not been successful.

For over a half century, MT research has drawn attention from people in different fields: cognitive psychology, linguistics, philosophy, cultural studies, computer science, computational linguistics. Several paradigms have developed. However, they still need improvement. No current operational MT systems can produce good quality output without either placing restrictions on input texts or involving human assistance before, during or after the translation processes. MT Research continues unabated, focused primarily on producing translations of greater accuracy.

It was for these reasons that we decided to discover a new approach to develop a machine translation system, an approach that can generate a reliable and accurate translation.
1.2 Objectives

Our goal is to find a new approach to MT which generates accurate translations, simplifies the design and the implementation of MT systems, and encourages others to adopt our approach as a usable tool.

In this thesis we describe the design of the approach we propose, GRMT. GRMT performs the translation process as follows: generate Translation Candidate (TC), verify the syntax and semantics of the TC and repair the TC when required. The tasks involved in generating the TC for the source language (SL) include:

- developing a set of constraints which are specific to resolve the syntactic differences between the source and the target languages.
- developing a SL dictionary to be used together with a set of constraints to refine the scope of the translation choices.
- developing the SL-target language (TL) dictionary to be used in relating SL and TL.
- classifying words into appropriate categories.
- developing a semantic relationship between words based on word classification to be used in selecting an appropriate translation for each input word.
- developing a set of constraints which are specific enough to resolve the syntactic differences between the target and the source languages to be used in completing the structure of the TC.
- developing the ordering rule. This rule is based on the syntactic structure differences between the SL and TL.

Tasks involved in verifying the TC requires development of the SL and TL grammars and lexicons. The grammars are used to analyze the SL and the TC to determine the accuracy of the TC. The grammars are developed based on a well-known grammar formalism, Head-Driven Phrase Structure Grammar (HPSG), since it is an integrated theory of natural language syntax and semantics. Implementation on the Attribute Logic Engine version 3.2 Beta (ALE) provides a mechanism which was designed to facilitate the use of HPSG grammar. The repair process is called upon when necessary to improve the quality of the TC.

Our English-Thai translation system is developed and implemented based on GRMT, in order to investigate the feasibility of GRMT.
GRMT is evaluated using our English-Thai translation system to examine whether the system performs in the way intended, whether the translation is acceptable and whether the system is extendable.

1.3 Outline of the thesis

In chapter 2, an introduction to MT including a brief history of MT and a discussion of the methodologies used in developing MT system is provided. Following this introduction we present the idea behind Generate and Repair Machine Translation.

Chapter 3 begins with a description of the architectural design of the GRMT system, the motivation of the design, similarities and dissimilarities between GRMT and other approaches, the knowledge-bases required in GRMT, advantages of GRMT, GRMT and multilingual MT.

We examine, in chapter 4, the difference between source language and target language, which are English and Thai respectively in this study, and in particular, on the aspects which will concern us in this work. The rest of this chapter addresses the word classification criteria. Of course, the notion of word classes is as old as the study of language itself. However, word classification has not been, in general, exploited fully in MT applications when the criteria used in classifying words includes both the meaning of the words and their usage. In addition, words in different languages are classified under the “same umbrella” even though there are differences at some levels between different languages. Finally the word association number, which plays a crucial role in GRMT, is introduced; the development of semantic relationship between words based on word classification and its usage are discussed.

Analysis Lite Machine Translation, the first phase of GRMT which generates the translation candidate for GRMT, is presented in chapter 5. Each component of ALMT is described in detail. The chapter includes some TC examples generated by ALMT.

Chapter 6 depicts the architecture of Translation Candidate Evaluation (TCE), the second phase of GRMT. After a brief description of HPSG and ALE we proceed to describe the English and Thai grammars we developed based on HPSG. The idea of how TCE extracts the semantic information from the source language input and the TC and how TCE determines the similarity between them are discussed.

Chapter 7 examines how Repair and Iteration, the last phase of GRMT, improves the quality of the translation candidate.
In Chapter 8, we describe the evaluations carried out on intelligibility, fidelity and comprehensibility of the translations for the English-Thai MT system.

Chapter 9 summarizes our contribution and concludes with some remarks about some efficiency considerations and possible improvements to GRMT.

Appendices are provided to illustrate SL Dictionary, SL-TL Dictionary, WordAsso hierarchy, Types, their features and their value types, Lexical Rules, Grammars, Intelligibility and Fidelity Evaluation and Comprehensibility Evaluation.
Chapter 2

Machine Translation

When I look at an article in Russian, I say
"This is really written in English,
but it has been coded in some strange symbols.
I will now proceed to decode it."
— Warren Weaver (1949)

2.1 What is Machine Translation?

Machine translation can be categorized under the more general heading of automatic
natural language processing. MT is described in several ways as follows:

- MT is the application of computers to the translation of texts from one natural
  language into another [Hutchins 1986].
- MT is the transfer of meaning from one natural (human) language to another with
  the aid of a computer [Goshawke et al. 1987].
- Systems which perform a syntactic analysis of a source text and then generate a
  target language rendering thereof which seeks to preserve and reconstitute its
  semantic and stylistic elements are described as machine translation systems [Newton
  1992].
- MT is the name for computerized systems responsible for the production of
  translations from one natural language into another, with or without human
  assistance [Hutchins and Somers 1992].
- MT is the translation of texts by means of computer [Rosetta 1994].
- MT is to automate all, or part of the process of translating from one human language to another [Arnold et al. 1994].

All descriptions, however, promote machine translation as an attempt to translate from one language into another (or others in the case of multilingual systems) with the aid of a computer.

The term MT usually refers to fully automatic machine translation. In the case where a computer "collaborates" with a human translator to some degree during the translation process, the system is then referred to as Machine-Aided Translation (MAT) or Computer-Aided Translation (CAT). We can further classify MAT into two categories as follows:

- Machine-Aided Human Translation (MAHT): The human performs the translation process with the assistance of a machine, e.g., consulting an on-line/automatic dictionary or terminological database, using multilingual word processors.
- Human-Aided Machine Translation (HAMT): The machine performs the translation process and consults a human user at many stages to complete the process, e.g., asking a user to make a choice in case of lexical ambiguity.

The major task of any MT system is to map the source language text onto an appropriate target language text that conveys the meaning of the original text. Therefore, the MT task includes language understanding and language generation tasks. In order to perform these tasks, MT efforts require linguistic and non-linguistic knowledge.

Required linguistic knowledge includes:

- **Morphological knowledge**- To provide the properties and meaning of an individual word.
- **Syntactic knowledge**- To recognize and form the structural relationship between words.
- **Semantic and world knowledge**- To understand and form the relations between the entities and events designated by words.
- **Pragmatic and stylistic knowledge**- To understand and form relations between the structure of texts and their communicative functions and also the idiomatic usage of the languages.
2.2 A Brief History of Machine Translation

The idea of using a machine in translation began in 1933, before the invention of the computer, and it remains a primary motivating problem for computational linguistics. Details of the MT history can be found in many books e.g., [Locke and Booth. 1955], [Hutchins 1986], [Slocum 1988], [Nirenburg et al. 1992], [Hutchins and Somers 1992], [Arnolds 1994]. Some pivotal earlier events are summarized as follows:

1933 George Artsruni (a French engineer, Armenian by birth) was issued a certificate patent for a translation machine which was called a Mechanical Brain by its inventor. [Henisz-Dostert et al. 1979].

P. P. Trojanskij, in the USSR, envisaged the three stages (see section 2.3.1) of mechanical translation [Henisz-Dostert et al. 1979]. He was issued a patent for a Machine for selection and printing of words while translating from one language into another or into several others simultaneously [Hutchins 1986].

1946 Warren Weaver (Vice-President for the Natural and Medical Sciences, Rockefeller Foundation) and Andrew D. Booth (a British Crystallographer) discussed the possibility of using computers for translation. This is the first discussion on MT [Henisz-Dostert et al. 1979].

Weaver brought the idea of MT to general notice and suggested some techniques to be used in developing MT systems. [Henisz-Dostert et al. 1979].

1948 Andrew D. Booth and R. H. Ritchens implemented a word-by-word French-to-Spanish dictionary translation program.

---

1 The name is spelled Petr Petrovich Smirnov-Troyanskii according to Hutchins 1986 and P. P. (Smirnov-) Troyansky according to Bar-Hillel, 1964.
1949 The appearance of *Translation*, Warren Weaver's memorandum [Weaver 1955 reprinted].

This memorandum drew the interest of researchers all over the United States. It is considered the starting point for research in MT.

1949 - 1952 The studies on MT started at a number of research institutions, universities and industries in the United States, e.g., University of Washington in Seattle, University of California at Los Angeles, The Massachusetts Institute of Technology (MIT) and The RAND Corporation in Santa Monica.

1952 Yehoshua Bar-Hillel organized the first MT conference at the Research Laboratory of Electronics- MIT, outlining future research in MT.

1954 Leon Dostert at Georgetown University collaborated with IBM on the Russian-English MT project and demonstrated their MT system to public. It is regarded as “the first generation of MT”.

The journal of *Mechanical Translation* was founded. It was edited and published by William N. Locke and Victor Yngve at MIT.

1955 - 1960 MT projects initiated elsewhere in the world, e.g., Cambridge Language Research Unit in the United Kingdom, University of Milan in Italy, Charles University in Czechoslovakia, Institute of Precision Mechanics and Computer Technology in the Soviet Unions and Kyoto University in Japan [Hutchins 1986].

1959 Yehoshua Bar-Hillel published his critique of contemporary MT research. He concluded [Bar-Hillel 1964]:

- The fully automatic, high quality translation was not really attainable in the near future so that a less ambitious aim is definitely indicated.
- The empirical approach\(^2\) seemed somewhat wasteful in practice and not sufficiently justified in theory.

\(^2\)The empirical approach emphasized the need to base procedures on actual linguistic data. It was distrustful of existing grammars and dictionaries. The approach stressed statistical and distributional analyzes of texts [Hutchins 1986].
1964 The government sponsors of MT in the United States formed the Automatic Language Processing Advisory Committee (ALPAC) to provide directed technical assistance as well as to contribute independent observations in mechanical translation. [ALPAC 1966].

1966 The ALPAC report concluded that MT was slower, less accurate and twice as expensive as human translation. The report also stated that “There is no immediate or predictable prospect of useful MT”. This report brought a virtual end to MT research in the USA and damaged the public perception of MT for many years afterward [Hutchin and Somers 1992].

MT research initiated in Canada.

1969 In Middletown, New York, Charles Byrne and Bernard Scott found Logos to develop MT systems [Demos and Frauenfelder 2000].

1970 A renaissance of MT research projects in the United States after ALPAC report [Hutchins, 1986].

1972 The beginnings of AI-oriented research on MT [Wilks 1972].

1976 The Commission of the European Communities (CEC) purchased the “Systran” system for the development of translating systems for languages of the community [Hutchins 1988].

1978 The first multilingual MT “Eurotra project” had been developed. This system’s goal was to translate back and forth between nine languages; Danish, Dutch, English, French, German, Greek, Italian, Portuguese, and Spanish.

The first public MT system “METEO”, developed by the Université de Montreal, for translating weather reports from English to French had been used. It is one of the most successful MT systems.

1980 - 1982 The first commercial system, “Automated Language Processing Systems (ALPS)”, was developed.

1982 Speech translation had begun at British Telecom Research Laboratories in Great Britain and Advanced Telecommunications Research Laboratories in Japan.

A number of commercial MT systems had begun development, e.g., Globalink, PC-Translator.
1986 Japan launched the Advanced Telecommunications Research, Interpreting Telecommunication Research Laboratories (ATR-ITL), to study multilingual speech translation.

1988 Researchers at IBM’s Thomas J. Watson Research Center revived statistical MT methods [see section 2.3.3.2].

1991 The first translator dedicated workstations appear, including STAR’s Transit, IBM’s TranslationManager, Canadian Translation Services’ PTT, and Eurolang’s Optimizer [Demos and Frauenfelder 2000], [Hutchins, 1997].

1992 ATR-ITL founds the Consortium for Speech Translation Advanced Research (C-STAR) which demonstrated the phone translation between English, German, and Japanese to the public.

1993 The German-funded Verbmobil project gets under way. Researchers focus on portable systems for face-to-face English-language business negotiations in German and Japanese.

BBN Technologies demonstrates the first off-the-shelf MT workstation for real-time, large-vocabulary (20,000 words), speaker-independent, continuous-speech-recognition software [Demos and Frauenfelder 2000].

1994 Free Systran machine translation was available in selected CompuServe chat forums.

1997 AltaVista’s Babel Fish offers real-time Systran translation on the Web.

1999 Logos released e.Sense Enterprise Translation, the first Web-enabled multiple translator operating from a single server.

Kevin Knight, of the University of Southern California’s Information Sciences Institute (ISI), led a multi-university team that developed Egypt, a software toolkit for building statistical MT systems. Egypt examines bilingual texts for statistical relationships, analyzes those patterns, and applies what it has "learned" to its translation functions.

2000 At MIT’s Lincoln Laboratory, Young-Suk Lee and Clifford Weinstein demonstrate an advanced Korean-English speech-to-speech translation-system prototype.
2.3 Machine Translation Methodologies/Technologies

2.3.1 Methodologies used to develop MT systems

The translation process with the aid of a machine was proposed by P. P. Trojanskij in 1933 [Henisz-Dostert et al. 1979]. Trojanskij's translation process begins by analyzing the input words in the source language to obtain their base forms and syntactic functions. Next, the process transforms sequences of those base forms and syntactic functions into equivalent sequences in the target language. Only transformation step is performed by machine. Then the process converts the output of the transformation step into TL.

In 1946, Warren Weaver suggested that war-time cryptography techniques could be used to detect the basic elements which might be contained in all languages and A. D. Booth suggested that any digital computing machine having the necessary storage capacity could make a bilingual dictionary. Weaver also suggested some other techniques used in a translation process, e.g., statistical analysis, Shannon's communication theory, and exploration of the underlying logic and universal features of language.

A Brute-force approach was used to develop a Russian-English system [Garvin 1972]. The Brute-force approach is based on the assumption that, given a sufficiently large memory, MT can be accomplished without a complex algorithm, i.e., that MT can be accomplished with a dictionary containing not only words but also phrases, or a dictionary and a table of grammar rules. The negative opinion of ALPAC regarding the achievements of machine translation is based largely on a study of these systems, and is, of course, justified to the extent to which it concerns them [Garvin 1972].

The approach employed in nearly all MT systems until the late 1960's was the Direct translation approach. The Direct translation approach is a word-to-word replacement approach (see section 2.3.3.1). The morphology and syntax of the SL text need not be analyzed any more than is strictly necessary for the resolution of ambiguities and the specification of the TL word order. Erwin Reifler, at University of Washington in Seattle, recognized the problem of multiple meanings of word as an obstacle to successful direct translation. He began to attack the central problem of semantic ambiguity in 1950 [Bar-Hillel 1964]. Ten years later, the requirements of MT gave impetus to significant theoretical developments in linguistics and what would later become known as the disciplines of computational linguistics and artificial intelligence [Nirenburg et al. 1992].
recognized by 1960, that the achievement of acceptable MT requires a very detailed and extensive knowledge of the languages concerned, particularly with respect to the fundamental semantic problems of translation. Therefore, research on the Indirect Translation approach was begun. The treatment of linguistic structure became more sophisticated. The source language text analysis and target language text generation were treated separately and conversion was achieved via an "interlingual" representation or a "transfer" component operating on abstract representations of the SL and TL. These approaches are now known as the interlingual and transfer approaches respectively (see section 2.3.3.1).

Statistical methods were first suggested by Weaver to resolve problems of multiple meanings of words. A detailed study of statistical techniques was conducted at RAND Corporation by Abraham Kaplan in 1950. This study encouraged the belief that problems of ambiguity in MT could be overcome, however, the approach was quickly abandoned due to machine limitations and the critique of Bar-Hillel [Bar-Hillel 1964]. A decade later, the statistical approach had proven successful in speech recognition and parsing which drew the interest of the MT researchers at IBM's T. J. Watson Research Center [Brown et al. 1988]. In 1988, a statistical approach to MT was implemented at IBM. The basic idea of the statistical approach is to carry out a translation based on complex cooccurrence and distribution probability calculations over very large aligned bilingual text corpora (see section 2.3.3.2).

An artificial intelligence approach to MT was developed by Yorick Wilks at Stanford University in 1972 [Wilks 1972]. Wilks' system is based on an interlingual approach with a semantic grammar and inference rules (see section 2.3.3.3). A conceptually similar approach was carried on by J. Carbonell, R. E. Cullingford and A. V. Gershman in 1981 under the name of Knowledge-Based Machine Translation (KBMT) [Carbonell, Cullingford and Gershman 1981]. Other experiments were conducted at Yale University during 1978-1982, and by S. Nirenburg, V. Raskin and A. Tucker in 1985 [Nirenburg, Raskin and Tucker, 1986]. KBMT is different from the interlingual approach in terms of the expected depth of SL analysis, the building of semantic representations of the meanings of the SL, and the reliance of KBMT systems on the explicit representation of world knowledge [Hutchins 1986; Nirenburg et al. 1992; Whitelock and Kilby 1995].

A framework for MT by the analogy principle or by example-guided inference was first proposed by M. Nagao in 1984 [Nagao 1984]. His idea is that translation is achieved by imitating the translation of a similar example in a database. The task becomes one of
matching new input to the appropriate stored translation. In 1990 Sumita, Iida and Kohyama proposed a similar approach called Example-Based Machine Translation (EBMT) [Sumita, Iida and Kohyama 1990]. The basic idea of EBMT is to search for the closest match of the input text in a bilingual text archive, where text in the SL are stored and aligned with their translations into a TL. The translation of the closest match is accepted as the translation (see section 2.3.3.2).

Lexically-driven machine translation develops the translation upwards from the lexicon without translation rules or interlingual operations. Each lexical entry contains information such as orthography, syntax and semantics. Shake-and-Bake MT developed by Pete Whitelock and Mike Reape [Beaven 1992] is one sample system of a Lexically-driven MT (see section 2.3.3.3).

2.3.2 Techniques used to increase the performance of MT

Two techniques can be explored to enhance the performance of MT systems. The first technique is to accommodate the translation process by forming a collaboration between human and machine in post-editing, pre-editing and interaction. The other technique is to study the nature of the linguistic knowledge required and the approach used in acquiring it e.g., building the lexicon, parsing the SL, generation of TL text.

2.3.2.1 Post-editing, Pre-editing and Interaction

The concepts of post-editing and pre-editing were introduced by Erwin Reifler in 1950 [Locke and Booth, 1955]. Both post-editing and pre-editing are performed by a human who need not be bilingual. A post-editor reduces the translation output to a reasonable literary form while a pre-editor prepares the original text for the machine by removing ambiguities. Interaction involves human collaboration during translation processes.

Post-editing The task of the post-editor is to resolve residual problems (e.g., to select the appropriate choice from the alternative translations) appearing in the raw MT output and to tidy the style of the translation up. The post-editing technique was used at RAND Corporation in the late 1950's with a MT system that translates from French to English [Hutchins 1986]. This idea has continued to be used in a number of MT systems, e.g., AMPAR (English to Russian) developed at the USSR Centre [Marchuk 1989]), SPANAM (Spanish to English) developed by the Pan American Health Organization (PAHO) [Pan...
American Health Organization 1999], and Systran (e.g., English to Chinese, French to
German, English to Italian) developed by Systran [Systran 1999].

The difficulty of post-editing and the time required for it correlates well with the quality
of the raw MT output: the better the output, the less the post-edit effort. The post-editing
task time and cost increases as translation quality gets poorer.

Within the past few years there have been dramatic improvements in hardware (e.g.,
storage capabilities, processor speed, memory), software (e.g., programming languages,
database management) and also developments in linguistic theory. These advances lead MT
researchers to believe that good quality MT output is achievable; therefore, current research
is searching for theories and methods to accomplish perfect translation and thus reduce the
need for post-editing.

Pre-editing The SL text is adjusted before it is entered into the system by someone who
knows the input restrictions. The text is written in a “controlled language” or “sublanguage”
[Rosetta 1994]. A controlled language is a form of language usage restricted by grammar and
vocabulary rules [Arnold 1994]. A sublanguage is the language used by a particular community
of speakers concerned with a particular subject matter [Sager 1986]. By restriction to a
controlled language or sublanguage, the problems of lexical homography and syntactic
ambiguity can be reduced and, at the same time, higher levels of comprehensiveness and
completeness in dictionaries are enabled. As an added benefit, the original text becomes
clearer as well. The pre-edited text can sometimes be translated without the need for post-
editing. The pre-editing technique is particularly cost-effective when there is a need to
translate one document into many languages. The first really successful use of this technique
was made by Xerox Corp. Xerox developed “Multinational Customized English3” (MCE) for
writing unambiguous English. The Systran system with MCE was installed to translate from
English into French, Spanish and Italian [Hutchins 1986]. In 1976, Météo was developed by
the researchers at Traduction Automatique de l’Université de Montréal (TAUM) for
translating weather reports from English to French, and is now in daily use. In Météo, the
sublanguage grammar was written for analyzing English texts and for generating the
corresponding French texts. However, the practice of “sublanguage” is now rarely adopted

---

3 MCE was developed based on a general and technical vocabulary of 3000 words and rules [Hutchins 1986].
since few texts are limited to single subject fields and also for economic reasons, since output inevitably requires revision as well.

**Interaction** Interactive MT systems involve human intervention for resolving doubts and uncertainties about the structure of the SL text or about the correct choice of the translation word in the TL texts during the translation processes. The final (raw) output does not require post-editing because ambiguities and other problems are solved. The philosophy of interactive systems was that unaided computer analysis can never be satisfactory and that difficult problems of linguistic analysis can be circumvented by calling upon human expertise at appropriate moments [Hutchins 1989a].

In the mid-1960s, the Cambridge Language Research Unit (CLRU) investigated the possibilities of an interactive English-French MT system [Hutchins 1986]. The Management of Information through Natural Discourse\(^4\) (MIND) was developed at The RAND Corporation in early 1970 [Bisley and Kay 1972]. MIND relies on a monolingual user to resolve ambiguities in the translation process. The Interactive Translation System\(^5\) (ITS) was developed at the Translation Sciences Institute of Brigham Young University (BYU) in 1970 [Hutchins 1986]. An Interactive Japanese Parser for MT, proposed in 1990, was developed at IBM Research, Tokyo Research Laboratory. This system allows a user to intervene during parsing to help the system to produce a correct parse [Maruyama and Watanabe 1990]. The Dialogue-based MT (DBMT) conducted at Grenoble allows a monolingual user (with limited ability in TL) to interact with the translation system [Boitet 1990]. InterTran, a free web translation service provided by Translation Experts Ltd\(^6\), allows a user to select an appropriate TL word for each SL if there are choices.

The efficiency of an interactive MT system depends on the number of questions asked during the translation process. More questions means less efficiency since a user needs to go through the entire SL text to accurately answer the questions. In addition, the time used during the interaction corresponding between a user and a system should be taken into account.

---

\(^4\) The chief designers of MIND were Martin Kay and Ronald Kaplan. Martin Kay was a member of the CLRU.

\(^5\) The leadership of ITS development was Eldon G. Lytle at BYU. He and other members of the ITS project left BYU and continued the development of their interactive approach to MT, the first commercial MT system Automated Language Processing System (ALPS), in 1980.

\(^6\) InterTran is available at “http://www.tranexp.com:2000/InterTran” as of December, 1999.
2.3.2.2 Lexicon constructions and parsing techniques

Lexicons and parsers are vital components of any MT system. An accurate lexicon and good syntactic and semantic analyses are required to ensure a good translation. Therefore, the lexicon construction method and the parsing technique must be considered not only in the translation process, but also to improve the efficiency of a MT system. Much current research is directed at improving existing MT approaches by developing a better linguistic analysis for the analysis-based MT approach (e.g., Direct MT, Interlingual MT, Transfer MT, Knowledge-Based MT and Shake-and-Bake MT) or by developing better tools to construct and modify an accurate lexicon. Some interesting parsing techniques and lexicon construction tools are summarized below.

PARSEC, the connectionist parsing system, is one of the parsers used in JANUS [Woszczyna et al. 1995]. JANUS is a speech-to-speech translation system developed at Carnegie Mellon University. PARSEC is used as a fall-back module if the symbolic high precision parser fails to analyze the input. The important aspect of the PARSEC system is that it learns to parse sentences from a corpus of training examples.

Steven Abney proposed the idea of parsing by chunks\(^7\) in 1994 [Abney 1994]. In Abney’s chunking parser, the syntactic analyzer comprises the chunker and the attacher. The chunker converts a stream of words into a stream of chunks, and the attacher converts the stream of chunks into a stream of sentences. Abney also compared his parser and chart parsing [Covington, 1994]. In 1996, Michael John Collins at University of Pennsylvania proposed a statistical parser which employs a simple statistical model based on dependencies between words [Collins 1996].

I. Dan Melamed, at the University of Pennsylvania, proposed a new method used in constructing a lexicon [Melamed 1998]. Melamed’s method was designed to generate a word-to-word translation lexicon, and each lexical entry is tagged with its relative frequency of co-occurrence. Davide Turcato, at Simon Fraser University, proposed a method to build a bilingual lexicon [Turcato 1998]. This lexicon contains not only word equivalence but also the syntactic and semantic descriptions.

---

\(^7\) Chunks correspond to prosodic patterns. It appears, for instance, that the strongest stresses in the sentence fall one to a chunk, and pauses are most likely to fall between chunks. Chunks also represent a grammatical watershed of sorts. The typical chunk consists of a single content word surrounded by a constellation of function words, matching a fixed template [Abney 1994].
2.3.3 Discussion of some existing MT methodologies

The methodologies used for developing MT systems can be classified into three different categories: the three classic strategies (Direct MT, Interlingual MT and Transfer MT), the nonlinguistic information strategies (Example-Based MT and Statistics-Based MT), and the hybrid strategies (Knowledge-Based MT and Shake-and-Bake MT).

2.3.3.1 Three Classic Strategies

Three widely articulated strategies for developing MT systems include Direct MT, Interlingual MT and Transfer MT. *Direct MT* (Figure 2.1) is a "word-to-word" replacement approach that replaces the SL word with the TL word. Therefore, the accuracy of the translation depends primarily on the bilingual dictionary. Accuracy is rather limited because the Direct MT system takes into account only morphological information. The following examples, using the word *open* with English as the SL and Thai as the TL, illustrate this limitation (Figure 2.2).

![Diagram of Direct MT](Image)

The meaning of each *open* is different because of the circumstances of usage. Therefore, the translations of the word *open* into the Thai language which are shown underlined in Figure 2.2 depend on which Thai word will cover each circumstance. In phrase (1), the word *open* is translated as *īm* (pɔ̄’d)*8* in Thai because it means *not shut*. However, the word *open* in phrase (2) is translated as *īm* (rɔ̄m) since *open fire* means *to start shooting*. Nevertheless, *open*
the translation of the word ล้ำ (p’a’d), phrase (5), into English is turn on, not open. The word ล้ำ (lyym), in phrase (6), cannot be translated as open either because it has a complete different meaning from the word ล้ำ (lyym) in phrase (4). The translation of the word ล้ำ (lyym), phrase (6), must be forget. The meaning of a word or phrase depends, in part, on the situation in which it is used. The relationship between the basic meaning of a word and its circumstances of usage is very complex. This complexity is one reason why the mapping from one language to another is far from direct.

The direct approach was adopted by most MT systems in the first generation of MT, e.g., the Georgetown system of Garvin 1967 [Goshawke et al. 1987]. The translation accuracy of this approach is limited because the linguistic analysis of the SL is inadequate. In addition, the design of Georgetown system was orientated towards a specific language pair, and emphasis on a bilingual dictionary as knowledge source. The addition of a new language pair means that entire new modules have to be written. The number of rules tend to be very large, and rule interaction difficult to monitor, so that the whole system is difficult to extend and maintain. However, the Direct MT is still appealing in certain circumstances. Traces of the direct approach are found in an indirect approaches such as Météo and Systran [Hutchins and Somers 1992].
To generate more accurate translations, a second approach, *Interlingual MT* (IL MT) was developed. The treatment of linguistic information became more sophisticated in this approach. IL MT analyzes and represents the SL in an intermediate language-independent conceptual representation called *interlingua* (Figure 2.3). IL MT then generates the TLs (in any language) from the interlingua. The accuracy of the IL approach is based on the ability to generate an accurate interlingua. Since interlingua is a language-independent representation of the meaning of the text, the analysis and generation modules are independent from each other. The advantage is that the addition of a new language involves only the creation of two new modules: the analyzer and the generator of the new language. The all-ways translation between \( n \) languages requires \( n \) analysis and \( n \) generation modules whereas the direct approach requires \( n(n-1) \) translation modules.

The IL approach is the most attractive to multilingual machine translation (MMT) systems. However, to define neutral concepts for different languages is rather a chimera. Interlingua must capture all necessary information of the analyzed SL so that the generator can generate an accurate translation in all languages. A number of interlingual systems were developed e.g., ATLAS of Fujitsu [Uchida 1989], PIVOT of NEC [Muraki 1989], Rosetta of Phillips [Landsbergen 1987a], KANT [Mitamura, Nyberg and Carbonell 1991], UNITRAN [Dorr 1991] and CICC Interlingual [CICC 1995e]. The CICC interlingual system was developed and implemented in the project "The research and development cooperation..."
project on a machine translation system for Japan and its neighboring countries [CICC 1995g]. The CICC MMT project was aimed at developing MMT, an all-ways translation between five languages: Japanese, Chinese, Malaysian, Indonesian and Thai. Some limitations and problems were encountered in developing the interlingual system for this project. For example, interlingua theory assumes that sentences in different languages which carry the same meaning must be represented by the same interlingua. However, in the CICC MMT project, the interlingual representations of these sentences are different [CICC 1995d] for several reasons [Cercone 1975; CICC 1995c; CICC 1995g; CICC 1995h; CICC 1995i]:

- The scopes and concept classification of each language are different (e.g., the concept of culture, the concept of the supernatural world, and the concept of unit).
- A single concept in one language can be mapped into many concepts in another language.
- More than one concept can be mapped into only one single concept in another language.
- Some concepts do not exist in some languages.

Different languages and different cultures result in different circumstances of language usage. Thus, it is very difficult to define a neutral concept. Interlingual then, is still an ideal despite being considered one of the best approaches.

Because of these difficulties, the compromise between the direct approach and the interlingual approach led to the idea of the third approach, \textit{Transfer MT} (Figure 2.4). The

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{transfer_mt_diagram}
\caption{Transfer MT}
\end{figure}

\begin{itemize}
\item China, Indonesia, Malaysia and Thailand.
\end{itemize}
transfer approach involves "transferring" the linguistic information of SL to TL. The transfer approach analyzes and represents the SL with a machine-readable form and maps this form into a machine-readable form of TL, then generating the TL from that representation. The accuracy of transfer approach depends primarily on the transfer module.

Unlike the intermediate representation in the interlingual approach, the intermediate representation in the transfer approach strongly depends on language pairs. Therefore, the all-ways translation of $n$ languages requires $n$ analysis, $n$ generation and $n(n-1)$ transfer modules. The transfer approach is not attractive to MMT systems. However, it is still preferred to the interlingual approach among MT researchers for two reasons. There are difficulties in defining neutral concepts as mentioned earlier. Furthermore the intermediate representation of the transfer approach depends on language pairs so the analysis and generation processes are less complicated when compared to the interlingual approach (the relative complexity of the analysis and generation modules in a transfer system is much reduced because the intermediate representations involved are still language dependent). The

Table 2.1: Translation examples by the commercial MT system

<table>
<thead>
<tr>
<th>English</th>
<th>Translation to French</th>
<th>Translation back to English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hatchery officials are having to teach the fish to like worms.</td>
<td>1. Les fonctionnaires de l'incubateur doivent apprendre le poisson pour aimer des vers.</td>
<td>1. Civil servants of the incubator must learn fish for aimerdes toward.</td>
</tr>
<tr>
<td>2. It does not matter if you are born in a duck yard.</td>
<td>2. Il n'importe pas si vous naîssez dans un jardin du canard.</td>
<td>2. He/it doesn't import if you are born in a garden of the duck.</td>
</tr>
<tr>
<td>3. Only a life lived for others is a life worth while.</td>
<td>3. Seulement une vie pour les autres vaut pendant que.</td>
<td>3. Only a life for other is worth while.</td>
</tr>
<tr>
<td>4. I never think of the future.</td>
<td>4. Je ne pense jamais du futur.</td>
<td>4. I never think the future.</td>
</tr>
<tr>
<td>5. You can take a fish to school, but you cannot make them think.</td>
<td>5. Vous pouvez prendre un poisson pour scolariser, mais vous ne pouvez pas les faire penser.</td>
<td>5. You can take a fish to school, but you don't can pasles to make think.</td>
</tr>
<tr>
<td>6. BOISE, Idaho (AP) - Trout in Idaho are not just swimming in schools- they are going to school.</td>
<td>6. BOISE, Idaho (AP) - Truite dans Idaho ne nage pas dans les écoles juste - they are aller scolariser.</td>
<td>6. BOISE, Idaho (AP) - Trout in Idaho doesn't swim rightly in schools-they are to be going to school.</td>
</tr>
</tbody>
</table>

10 This is how the translation actually appeared. This type of error appears through aout these examples.
interlingual approach necessarily requires complete resolution of all ambiguities and anomalies of SL texts so that the translation into any other language is possible whereas in the transfer approach, only those ambiguities inherent in the language pair are tackled. Some examples of MMT systems which were developed based on the transfer approach include: Eurotra which has been developed by European Community11 since 1978; Susy which was developed at the Université des Saarlandes in Saarbrücken during 1972-1986; and METAL which has been developed at the Linguistic Research Center at the University of Texas (Austin, Texas) since 1961. Details of these MT systems can be found in Hutchins and Somers, 1992.

The transfer of structure information in the transfer approach may cause inaccuracies because of the different structures between languages and thus the transfer approach may lose some information during processing. Table 2.1 illustrates this problem. The second column in French represents the translation of the first column from English12. Each translation fails to maintain close fidelity to the meaning of the original sentence. The translations are incorrect in both words selected and in grammar. The third column shows the translation back into English of the French translation in the second column13. The third and the first columns should convey the same meaning but they do not because accumulative errors occurred during the translation processes.

2.3.3.2 Nonlinguistic Information Strategies

There is a possibility that different languages require different amounts of information to be expressed. For example, the Spanish word llegó means either he arrived or she arrived, but in order to translate it into English, we have to choose “he” or “she” on the basis of context and background knowledge. Japanese and Thai have “polite” markers. To translate from English to these languages, the “polite” markers need to be added according to the knowledge of culture. This observation led to nonlinguistic information strategies. Nonlinguistic information strategies are currently popular, e.g., Example-based MT [Jones 1996], Statistical strategies [Brown et al. 1992].

---

11 Eurotra project involves research groups in all member states in Europe e.g., GETA in Grenoble, IAI in Saarbrücken, and ISSCO in Geneva [Hutchins and Somers 1992].
12,13 These translations were provided by the commercial MT system.
Example-Based machine translation (Figure 2.5) attempts to achieve translations by using examples of translation in the translation process. EBMT performs a translation by mapping the input text to an example in the bilingual corpus: a collection of translation pairs. If the entire input phrase can be matched, then it will be translated to the translation pair of that matched example. If the entire input phrase cannot be matched then EBMT takes two approaches to translate such a sentence: the closest mapping and recombining the matched segments.

The first approach, closest mapping, looks for the closest example for the input text. If it is found, the system will use its translation pair as a template and replace the SL words with the corresponding words in TL. In case the closest example is not found, there is a possibility that its segments can be found in a number of examples. Therefore, the second approach, recombining the matched segments, is applied. All matched segments will be translated and recombined to be the translation of the input text. The measurement of similarity between an input text (segments) and a set of examples can be considered according to both surface and content. The similarity measurement can be performed by using a thesaurus. One advantage of EBMT is that translations come with scores which are measurements of similarities to known examples. These scores could be useful to a post-editor. Further details of the EBMT can be found in [Nagao 1984], [Sato and Nagao 1990], [Sumita, Iida and Kohyama 1990] and [Jones 1996].

The feasibility of EBMT depends crucially on the availability (suitability) of the translation examples. EBMT requires a large collection of translation examples (pairs) to
produce an accurate output. The coverage of EBMT system depends on the size and diversity of the collection of the translation pairs.

Statistics-Based MT tries to do away with formulating linguistic knowledge, but applies statistical techniques instead. Statistics-based MT views every sentence in one language as a possible translation of any sentence in other languages [Brown et al. 1990]. This approach attempts to search for a sentence S in the source language that maximizes Pr(S)Pr(T|S), whereas Pr(T|S) is the translation model probability and Pr(S) is the language model probability.

The translation model probability, Pr(T|S), is assigned to every pair of sentences (S,T). It is the probability that a translator will produce T in the target language given S in the source language. The translation model probability is calculated from the probability of word alignment. The word alignment corpus, contained the aligned translated text, is used to predict which elements of one language of the corpus are most likely to be equivalent to elements in the other. The language model probability, Pr(S), is the probability of words occurring in a particular order. Further details of the computation and searching method can be found in [Brown et al. 1988], [Brown et al. 1990], [Arnold 1994], and [Jones 1996].

One sample system of the statistics-based approach has been carried out by a research group at IBM. The researchers had access to three million sentence pairs from the Canadian (English-French) Hansard which contains the Proceedings of the Canadian Parliament.

This approach, while interesting, requires a very large and reliable aligned bilingual corpus to produce an accurate translation. As with EBMT, the statistics-based approach relies heavily on the availability of a large and diverse quality corpus; currently such corpora for most languages are underdeveloped.

2.3.3.3 Hybrid Strategies

Knowledge-Based Machine Translation is based on the interlingual approach using AI techniques. Reasoning-induction and abduction are used in interpreting the SL text by reference to world knowledge, not only a knowledge of language. The meaning representation should be sufficient for translation to a number of languages, rather than sufficient for total understanding [Nirenburg et al. 1992].

Basic Components of KBMT systems include: a parser, a reasoning system, and a generator. The parser (Semantic-based) recognizes semantic features of lexical items in SL text rather
than grammatical categories. Wilks’ Preference Semantics (MT) system\(^\text{14}\) represents the
meaning of the SL text (interlingual representation) with “binary decomposition trees” while
KBMT, developed at Yale University, adopted an IL representation based on the “conceptual
dependency theory” developed by Roger Schank\(^\text{15}\). The reasoning system (inference
mechanism) applies contextual world knowledge which augments the representation, to solve
ambiguities and uncertainties in the SL text. The generator maps the representation of the
SL into the TL.

J. Carbonell, R. E. Cullingford and A. V. Gershman combined their system with a Script
Applying Mechanism (SAM) to understand and represent the meaning of the SL. SAM is the
script-based story understander developed by R. Schank and R. E. Cullingford in 1978.
Carbonell, Cullingford and Gershman remarked that their system did not perform strict
translation, but rather retold or summarized the SL text in the TL [quoting Carbonell,
Cullingford and Gershman 1981; Carbonell and Tomita 1987]. The domain of the translation
is English newspaper articles.

KBMT-89 is a larger-scale KBMT developed by researchers at the Center for Machine
Translation at Carnegie Mellon University and the researchers from IBM’s Tokyo Research
Laboratory in 1987 [Goodman and Nirenburg 1991]. Its interlingual representation is a
frame notation and can be viewed as a kind of a semantic network. The world knowledge is
organized as a multiply interconnected, hierarchical network of frames. The domain of the
translation is limited to personal computer manuals.

Other KBMT systems have been implemented, e.g., ATLAS-II [Uchida 1989], PIVOT

One advantage of the KBMT approach is solving some ambiguity problems. However,
the disadvantage of this approach is the expense of either abductive or inductive reasoning,
even on small domains. Building the knowledge representation is also time consuming and
expensive.

\(\text{13}\) Details discussion on Wilks’ system can be found in [Wilks 1972] and [Whitelock and Kilby 1995].
\(\text{14}\) Schank’s conceptual dependency theory, developed in 1972, is rich in semantic representation. It is
designed to provide a representation of meaning of natural language in terms of which paraphrase,
infrence, and MT could be carried out. There are some similarities between Wilks’ formulas and Schank’s conceptual
dependency representations (Wilks and Schank were both at Stanford in the early 1970). Wilks’ formulas bear
much similarity to the (binary) decomposition trees developed by Lakoff. Further discussion on Wilks’ formulas
and Schank’s conceptual dependency theory can be found in [Cercone 1975] and [Hutchins 1986].
We doubt that a pure interlingua (semantic representation) is appropriate for MT systems since the pure IL does not contain syntactic information. The abstractness of “content” representations results in information loss about “surface” structures of text; it seems that a pure interlingua is more suitable for the paraphrase task than for the translation task.

**Shake-and-Bake MT** developed by Pete Whitelock and Mike Reape is Lexically-driven machine translation. In Shake-and-Bake MT, the SL is analyzed by using a SL monolingual grammar, then each SL word is replaced with its equivalent in the TL using the bilingual dictionary, and finally the translation is generated from a set of TL words by shift-reduce parsing algorithms according to the TL grammar [Whitelock 1992].

One advantage of Shake-and-Bake MT is that it relates the SL and the TL via the bilingual lexicon alone, without a transfer module or an IL representation. The semantic information of the SL is copied across by the coindexation specified in the bilingual lexicon (the variables in the semantic indices are copied from the SL word as prescribed by the bilingual lexicon).

Shake-and-Bake MT is not efficient since it generates the translation by trying all possible permutations of the TL words. The system parses an arbitrary permutation to determine its grammatical well-formedness. If that permutation is well-formed, the system indicates success. If not, another permutation is tried. The system might have to verify all possibilities to find the appropriate translation or find none at all. A number of techniques were proposed to improve the efficiency of the generation process in Shake-and-Bake MT e.g., Brew uses a heuristic method to constrain the solution space [Brew 1992]. Beaven employs a Cocke-Kasami-Younger (CKY) baking\(^{16}\) to avoid recalculating the same combinations of words more than once during the generation process [Beaven 1992].

Popowich’s generation process is also based on a chart parsing algorithm with a variety of different search strategies [Popowich 1996]. Poznanski, Beaven and Whitelock proposed the idea of considering the information from the previous stages, analysis and dictionary look-up, in the generation process [Poznanski, Beaven and Whitelock 1995]. However, none of these algorithms is guaranteed to avoid protracted search times in order to achieve an accurate translation result.

\(^{16}\) Beaven calls his algorithm the “CKY baking” because of the great similarity between his algorithm and the CKY chart parsing algorithm.
The approaches to MT discussed thus far do not explicitly recognize the notion of incremental repair of a mistranslating and subsequent iteration through an evaluation phase to further repair and iteration.

The notion of "repair and iterate" is a well accepted methodology in computer science. It seems to us that this notion should prove useful in a MT system as we iterate to a better and better translation. we note that this concept is not unlike Verbmobil’s notion of “negotiations” [Kay, Gawron and Norvig 1994] for their speech to speech system.

We discuss Generate and Repair Machine Translation in the next section. GRMT is a novel MT paradigm that takes advantage of the best features of the more traditional MT approaches (see Table 9.1) and combines them in a Repair and Iterate architecture.

2.4 Our approach : Generate and Repair Machine Translation

There are advantages and disadvantages in each approach. The efficiency inherent in the direct MT approach is limited because linguistic elements of the languages are considered only at the morphology level. The transfer of structure information in the transfer MT approach may cause inaccuracy because of the different structures between languages and thus may lose some information during processing. In addition, the transfer approach does not appear appropriate for multilingual systems. Of the three approaches, although the interlingual MT approach appears to be the most attractive, the interlingual representation is still an ideal. Example-Based MT and Statistics-Based MT approaches requires a very large and reliable bilingual corpus to produce an accurate translation; these corpus are not available for most languages. To build the knowledge for KBMT is time consuming and expensive. The Shake-and-Bake MT approach lacks efficiency due to its generation process.

Our goal is to overcome shortcomings of earlier approaches and to increase the accuracy of the overall translation. Generate and Repair Machine Translation is a proposed solution.

The GRMT approach is designed to increase the accuracy and efficiency of MT. The architecture of GRMT was designed to take advantage of, and have advantages over, the Direct, Transfer, Interlingual, Non-linguistic and the existing hybrid approaches to MT with respect to several translation aspects: simplicity, accuracy and multilingualism. The GRMT integrates the best features of each approach. The GRMT process is relatively simple and straightforward, not unlike the Direct method. However, the GRMT is more concerned with preserving linguistic information to produce an accurate translation result, like the Transfer
approach. GRMT also treats the SL and TL separately for easy management in multilingual MT systems, like the Interlingual approach. GRMT uses simple statistical information, in part, in the word selection process like the statistic-based approach. GRMT builds the semantic relationship between words based on world knowledge, like the KBMT approach. GRMT associates the SL and TL via the bilingual lexicon, like the Shake-and-bake MT approach.

Unlike the transfer or interlingual approaches, GRMT generates a translation candidate (TC) directly from the SL input to avoid information loss during the transfer process of the transfer approach and to avoid difficulties in defining neutral concepts for different languages in the interlingual approach. GRMT does not require any bilingual corpus in the translation process, unlike EBMT or the statistical based MT approach. Unlike the Shake-and-bake MT approach, GRMT verifies the generated TC and revises it if required to achieve an accurate translation without trying all possibilities.

GRMT takes into account the differences between languages in a unique way, hence a further advantage. If languages can be grouped according to various characteristics which they share, then the translation between groups can be performed more simply by GRMT. For example: Group 1 consists of English, French and Spanish, Group 2 consists of Chinese, Japanese and Thai. To perform the translation between these two groups, the transfer approach requires six SL analyzers, six TL generations and eighteen sets of transfer rules while GRMT requires six analyzers and two sets of constraints.

The English-Thai translation system has been developed based on the GRMT approach. The initial experiments of ALMT [Naruedomkul and Cercone 1997] indicate that the translation candidates can be generated quickly with relatively high accuracy. We have greatly expanded the dictionaries used in the earlier experiment, five fold [Naruedomkul, Cercone and Sirinoavakul 1999]. All necessary knowledge bases required were updated. ALMT has been improved to increase the performance of GRMT. The English-Thai translation system we developed has been used to translate a number of phrases. These sample phrases were collected from different sources: newspapers, text books, articles, bedtime stories and fortune cookies. This dissertation presents the design and some experiments of GRMT.

As far as we are aware two commercial products were claimed to translate English to Thai, Tapestry (http://knowledge.krdl.org.sg/MT/) and Language Force (www.wforce.com), however,
Chapter 3

Generate and Repair Machine Translation

'When I use a word,' Humpty Dumpty said in rather a scornful tone,
'it means just what I choose it to mean - neither more or less.'
— Through the Looking Glass
by Lewis Carroll

Generate and Repair Machine Translation is an efficient approach to machine translation distinct from the direct, transfer, interlingual, statistical, example-based, knowledge-based approaches. The architecture of GRMT (Figure 3.1) is designed to serve two purposes: first, to generate an accurate translation and second, to be amenable to multilingual translation. An accurate translation corresponds to a translation which retains the linguistic meaning of the source language. To achieve an accurate translation, GRMT performs the translation in

Figure 3.1 GRMT Architecture

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
three phases: The first phase, *Analysis Lite Machine Translation* (ALMT) generates a translation candidate for the SL. The accuracy of the generated TC is evaluated by analyzing both the TC and the SL in the second phase, *Translation Candidate Evaluation* (TCE). TCE compares the semantic information of the TC with that of the SL. If there is any dissimilarity then the TC will be “repaired” in the third phase, *Repair and Iterate* (RI). The repaired TC will again be analyzed and again be compared until there is no appreciable dissimilarity between the meaning of the SL and the TC. By performing the TCE and RI processes, GRMT ensures an accurate translation.

To make GRMT attractive to a multilingual MT system, we pay attention to simplicity in accommodating new languages and the associated systems knowledge required. System maintenance also adds further complications. Therefore, GRMT treats SLs and TLs separately and relates them by using information provided in a SL-TL lexicon. The SL-TL lexicon contains simple but necessary information for the translation process, and this SL-TL lexicon is simple to construct (see section 5.1.3). GRMT organizes knowledge required for the translation process into separate modules e.g., SL constraints, TL constraints, grammar and analysis lexicon for each language so that they can be developed easily and maintained simply. SL and TL constraints represent the syntactic differences between SL and TL which GRMT considers, in part, in generating a TC. The translation between languages in different families of languages that share some characteristics should be able to take good advantage of GRMT architecture (see section 3.5).

To illustrate the performance of GRMT, translation from English to Thai has been implemented. The English-Thai MT system translates isolated sentences (sentence by sentence). This English-Thai translation system has been developed and run under SWI-Prolog 3.2.8. The English and Thai grammars (see section 6.1.1.6) have been developed based on Head-Driven Phrase Structure Grammar formalism [Pollard and Sag 1987; Pollard and Sag 1994] and implemented on the Attribute Logic Engine version 3.2 Beta [Carpenter and Penn 1999].

GRMT was tested to generate the translations for a number of sentences/phrases by using a dictionary which contains 314 English words and their translations into 731 Thai words. The sample sentences/phrases were collected from different sources: newspapers, textbooks, articles, bedtime stories and fortune cookies.
3.1 Why Generate and Repair, Not Analyze and Generate?

GRMT (Figure 3.1) generates a TC directly from the SL input, analyzes and repairs (if necessary) the TC to verify its accuracy. The generate and repair methodology draws inspiration from the following sources:

- "...and since word-for-word translations are surprisingly good, it seems reasonable to accept a word-for-word translation as a first approximation and then see what can be done to improve it." quoting Victor H. Yngve [Yngve 1955].
- "If someone asks me how I translate, ... I can describe the physical process: I make a very rapid first draft, put it aside for a while, then go over it at a painfully slow pace, pencil- and eraser-in hand ..." quoting William Weaver [Weaver 1989].

Based on Victor H Yngve's investigation on the direct MT approach and William Weaver's (human) translation procedure, the GRMT approach begins by generating an initial translation which we called the translation candidate and then revises the TC, if necessary, to ensure its accuracy.

The TC, the initial translation, must be generated simply, quickly and efficiently. Therefore, we generate the TC directly from the SL. Our TC generation process is different from the old fashioned Direct MT in several important aspects. The differences between languages have been used as constraints to limit the choices of the corresponding words in the TL. We select the most appropriate choice of TL word for each SL word by considering the relationship between the close proximity words (see section 5.2). All selected words are arranged into grammatical order according to the similarities and the dissimilarities between the structure of the SL and that of the TL. Once the TC generation is completed, the revision process begins by analyzing the TC to determine whether it is grammatically correct and conveys the linguistic meaning of the SL. To verify the meaning of the TC, we compare its parse with the parse of SL in terms of semantics only since there are differences in the syntactic level between languages. Any dissimilarity appearing in the TC parse will be repaired according to the semantic information of the SL parse.

---

1 Victor H. Yngve investigated a word-to-word translation and this quote is a part of his conclusion.
2 William Weaver is the author of several books about opera and a translator of modern Italian literature [Weaver 1989].
Example 1, one of the most popular examples during the first generation of MT, was translated into Russian and then back from Russian into English by the direct MT approach. The three different translations are shown below. None of these translations conveys the linguistic meaning of the original sentence.

Example 1

SL: The spirit is willing but the flesh is weak.\(^3\)

Translated back from Russian into English:

- The whisky is all right but the meat has gone bad.
- The vodka is good but the meat is rotten.
- The liquor is holding out all right, but the meat has spoiled.

Table 3.1 illustrates the selected words (shown in the third column) and the TC generated by our generation process for the Example 1. If we translate back from this TC into English we will achieve the following translation: “The soul is ready but the body is weak.” This result is because the corresponding Thai word, ณิจฉัย (winjaan), selected for the word spirit means “thought of as remaining alive without appearing in physical form.” (the descriptions of each word and its corresponding translations are provided in section 5.2).

The TC is verified and repaired in the revision process. The appropriate sense of spirit in the Example 1 is “mental or moral strength to resist.” which corresponds to Thai ใจแข็ง. Therefore, the word ณิจฉัย (winjaan) is replaced with the word ใจแข็ง in the repair process. The repaired TC is shown in Table 3.1. If we translate back from the repaired TC into English, we will achieve “The mind is ready but the body is weak”.

Our TC generation process is distinct from the direct MT approach, not only the process of translation but also its modularity and extendibility (see section 3.5). Our TC generation process is a modular system in that each component can be modified and changed independently of each other and is intended to be easily extendible to any other language.

A good translation must be generated from a precise source. To generate a good translation from the SL parse, the SL parse should contain consistent information (in terms

\(^3\) In theory of 3D semantics [Galloway 1989], both “spirit” and “flesh” have allosemes in the domain of FOOD and in the domain of ANATOMY [Galloway 1993]. The reason FOOD is wrong for “flesh” is that the alloseme “meat” is archaic and rarely used; the alloseme in ANATOMY is thus the most likely one and this makes a domain match in ANATOMY for the alloseme of “spirit” also most likely. Thus “mind” and “body” are good alloseme choices.

33

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
of both syntax and semantics) required in the generation process. Unfortunately, there is no parser which can provide such parses available in all languages. Therefore, there is a possibility that a generation process generates the translation from the input, the SL parse, and the SL parse may contain errors. The errors accumulate during the generation process, depending on how the translation is generated. For example; in the transfer MT paradigm, to generate the translation from a SL parse, the structure and lexicon of the SL is transferred to those of the TL. This transfer process results in the information-loss problem we discussed in section 2.3.3.1.

Table 3.1: The TC and selected words generated for Example 1

<table>
<thead>
<tr>
<th>Input</th>
<th>Possible corresponding words in Thai</th>
<th>Selected Word</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The spirit</td>
<td>นั้น (nān)</td>
<td>นั้น (nān)</td>
<td>determiner</td>
</tr>
<tr>
<td></td>
<td>วิญญาณ (winjaan), จิตใจ (cidcaj),</td>
<td>วิญญาณ (winjaan)</td>
<td>thought of as remaining alive without appearing in physical form.</td>
</tr>
<tr>
<td></td>
<td>ร่างกาย (banjaakād),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>เหล่า (lāw),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>น้ำเปื้อนอย่างดีดี (nampajjanya'glāb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The spirit</td>
<td>นั้น (nān)</td>
<td>นั้น (nān)</td>
<td>determiner</td>
</tr>
<tr>
<td></td>
<td>วิญญาณ (winjaan), จิตใจ (cidcaj),</td>
<td>วิญญาณ (winjaan)</td>
<td>thought of as remaining alive without appearing in physical form.</td>
</tr>
<tr>
<td></td>
<td>ร่างกาย (banjaakād),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>เหล่า (lāw),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>น้ำเปื้อนอย่างดีดี (nampajjanya'glāb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>willing</td>
<td>พร้อม (phrātām), พร้อมใจ (temcaj)</td>
<td>พร้อม (phrātām)</td>
<td>regarding favorably the possibility of doing something, ready</td>
</tr>
<tr>
<td>but</td>
<td>แต่ (te'e), นอกจาก (nākkācāaag)</td>
<td>แต่ (te'e)</td>
<td>in spite of this</td>
</tr>
<tr>
<td>the flesh</td>
<td>นั้น (nān)</td>
<td>นั้น (nān)</td>
<td>determiner</td>
</tr>
<tr>
<td>is</td>
<td>นั้น (nān), ร่างกาย (rāŋkaaj)</td>
<td></td>
<td>the physical human body as opposed to the mind or soul</td>
</tr>
<tr>
<td>weak</td>
<td>นั้น (nān), ร่างกาย (rāŋkaaj)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


34

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
In shake-and-bake MT, the generation process replaces the SL words in the parse with their corresponding TL words. A number of algorithms were proposed to combine these TL words according to the TL grammar (see section 2.3.3.3). However, none of these algorithms is guaranteed to avoid protracted search times during the generation process.

In interlingual MT, the analysis produces a pure interlingua which does not contain any syntactic information. Generating the TL from this parse could result in a paraphrase, not a translation (see section 2.3.3.3).

In our approach, we generate the translation candidate directly from the SL to avoid these problems: an inaccurate input of the generation process (the SL parse), the information lost during the transfer process, protracted search times during the generation process, the doubt about paraphrase/translation. The TC is revised to ensure that the translation is grammatically correct and preserves the meaning of the SL.

3.2 The Three Phases

*Analysis Lite Machine Translation* generates a TC by considering the differences between language pairs in terms of syntax and semantics without performing any sophisticated analysis. ALMT first simplifies the structure of SL and then considers TL words which correspond to all possible meanings of each SL word. The most appropriate TL word is selected by applying a semantic relationship between words. To make the TC grammatically correct in the TL and to retain the meaning of the SL, some words in the TL may be added and then all words are rearranged according to the grammar of the TL. ALMT comprises three modules: word treatment, word selection, and word ordering.

*Translation Candidate Evaluation* analyzes the generated TC, which is an output of the previous step, ALMT, to determine whether the TC retains the meaning of the SL. TCE performs the evaluation by analyzing both the TC and the SL in terms of syntax and semantics in parallel. Then only the semantic results are compared (there are syntactic level differences between languages). If the semantic results are the same, that TC will be deemed an acceptable translation. If not, any part of the TC which causes its semantics to differ from that of SL will be repaired in the next phase, RI.

*Repair and Iterate* will replace inappropriate words or rearrange words in the TC depending on the analysis results of the previous step, TCE. The analysis results will identify
the incorrect (inappropriate) parts and RI will use the analysis result of the SL as a pattern for repairing the TC.

3.3 Knowledge-bases Required

The essential knowledge-bases required for GRMT include: Constraints, Dictionaries, and Grammar and Lexicon. The SL and TL knowledge-bases are developed separately for easy modification, extendibility and management. The details of each knowledge-base will be discussed in later sections where appropriate.

Constraints, the different characteristics between languages, are required in ALMT. There are two sets of constraints: SL constraints and TL constraints. The SL constraints are the characteristics of the SL which are different from those of the TL. They are used to simplify the structure of the SL and to narrow the scope of possible TL words which correspond to each SL word (see section 5.1.1). The TL constraints are the characteristics of the TL which are different from those of the SL. They are required, not only to retain the meaning of SL but also to make them grammatically correct (see section 5.3.1).

The SL dictionary is required in ALMT. It is used in the constraint application and inflectional analysis steps. Each entry contains its syntactic categories. The SL and the TL are put into correspondence via the SL-TL dictionary. The SL-TL dictionary contains the SL entry and its all possible corresponding words in the TL. Each corresponding TL word is related to its Word Association (WordAsso) number (see section 4.2). All TL words are ordered according to the frequency of usage.

Grammars and lexicons of both SL and TL are required in the analysis process of TCE. Grammars and lexicons of SL and TL are developed principally based on the linguistic theories applied to GRMT, Head-Driven Phrase Structure Grammar [Pollard and Sag 1987; Pollard and Sag 1994]. Each lexical entry contains morphological, syntactic and semantic information (see section 6.1.1.3).

3.4 Translation Examples by GRMT

In this section we present the results of applying GRMT to the translation of some English example sentences into Thai. The input symbols of Example 2 are shown in the first column labeled input (Table 3.2). The second column illustrates the output after applying
the constraints. In this example, no constraint is applied. Each word in this column is used as a keyword to search for the corresponding words in Thai. All possible translations in Thai of each input word are shown in the third column. The most appropriate words are selected by ALMT (see section 5.2) and the choices are shown in the fourth column. The classifier \( \text{tua} \) (tua, in shaded box) is added to complete the TC (details are described in section 5.3.1).

This generated TC and the SL were analyzed in the TCE, the second phase of GRMT. The parses show that the TC requires no repair (see section 6.1.1.6). Thus, the generated TC is a complete translation of Example 2. Table 3.3 presents four different senses of the word \( \text{spring} \) in the classification system (see section 4.2) and their corresponding words in Thai: ฤุุไวีนั้น (ry’duubajmáajphli), สัมบ (sapring), กระไอด (kradòod) and ๕ัพ (námphû). Each word is classified into different classes according to its meaning as illustrated in the fourth column of Table 3.3 (see section 4.2). Consider the case that the word ฤุุไวีนั้น (ry’duubajmáajphli) is chosen for the word \( \text{spring} \) in Example 3 by ALMT as shown in Table 3.4. The TCE indicated the inappropriate selection by comparing

### Table 3.2: The selected words and the TC generated for Example 2

<table>
<thead>
<tr>
<th>Input Constraint Application</th>
<th>Dic. Lookup &amp; Inflec. Analysis</th>
<th>Selected Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td>The</td>
<td>ฤุุไวี (nán)</td>
</tr>
<tr>
<td>ugly</td>
<td>ugly</td>
<td>ฤุุไวี (khùiùìe), ฤุุไวี (náakliad)</td>
</tr>
<tr>
<td>duckling</td>
<td>duckling</td>
<td>ฤุุไวี (túuaptédp)</td>
</tr>
<tr>
<td>hides</td>
<td>hides</td>
<td>ฤุุไวี (dãíjìak), ฤุุไวี (dãíjìdáá),</td>
</tr>
<tr>
<td>his</td>
<td>his</td>
<td>ฤุุไวี (khùa’ùúkhàw)</td>
</tr>
<tr>
<td>head</td>
<td>head</td>
<td>ฤุุ (hu’a)</td>
</tr>
<tr>
<td>under</td>
<td>under</td>
<td>ฤุุ (taaj), ฤุุ (phaajtaaj),</td>
</tr>
<tr>
<td>his</td>
<td>his</td>
<td>ฤุุ (khùa’ùúkhàw)</td>
</tr>
<tr>
<td>wing</td>
<td>wing</td>
<td>ฤุุ (piig), ฤุุ (bin), ฤุุ (kàaøbin)</td>
</tr>
</tbody>
</table>

---

5 In Thai, **classifiers** are used to express not only a quantity but also to modify a noun (see section 4.1)
6 In our experiment, ALMT selected the correct corresponding Thai word for the word \( \text{spring} \) in this expression. This inappropriate selection is to illustrate the repair process (see section 6.1.2.2).
Table 3.3: Three corresponding Thai words to the word spring

<table>
<thead>
<tr>
<th>English</th>
<th>Thai</th>
<th>Description</th>
<th>Word Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>spring</td>
<td>คุ้บในไม้ต้น (ry'duubajmáajphli)</td>
<td>The season between winter and summer</td>
<td>Season</td>
</tr>
<tr>
<td></td>
<td>สริง (sapring)</td>
<td>An elastic device</td>
<td>Elastic device</td>
</tr>
<tr>
<td></td>
<td>กระโดด (kradood)</td>
<td>To move suddenly upward or forward</td>
<td>Move forward as if by jumping</td>
</tr>
<tr>
<td></td>
<td>น้ำพุ (námphú)</td>
<td>A place where water comes up naturally from the ground</td>
<td>Geology</td>
</tr>
</tbody>
</table>

the SL and the TC parses. The SL parse determined the appropriate meaning of the word spring in this expression is "An elastic device" (see section 6.1.2.2). But in the TC parse, the word คุ้บในไม้ต้น (ry'duubajmáajphli) means "The season between winter and summer". According to the analysis of analyzing the SL, the RI module replaced the incorrect selection คุ้บในไม้ต้น (ry'duubajmáajphli) with the word สริง (sapring) to generate the new TC for Example 3: จอห์น พูด เกี่ยวกับ ชู้ในไม้ต้น ที่ พัก ฉัน. This new TC was analyzed and compared to

Table 3.4: The selected words and the TC generated for Example 3

Example 3
SL: John talks about the broken spring.
TC: จอห์น พูด เกี่ยวกับ ชู้ในไม้ต้น ที่ พัก ฉัน
(จอห์น- John) (พูด- talk) (เกี่ยวกับ- about) (ชู้ในไม้ต้น- spring, the season) (ที่- modifying) (พัก- broken) (ฉัน- the)

<table>
<thead>
<tr>
<th>Input</th>
<th>Constraint Application</th>
<th>Dict. Lookup &amp; Inflec. Analysis</th>
<th>Selected Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>John</td>
<td>จอห์น (caan)</td>
<td>จอห์น (caan)</td>
</tr>
<tr>
<td>talks</td>
<td>talks</td>
<td>ชู้ (phuud), บรรยาย (banjaaj), ชู้ (khuj)</td>
<td>ชู้ (phuud)</td>
</tr>
<tr>
<td>about</td>
<td>about</td>
<td>เกี่ยวกับ (kiawkab), ประมาณ (pramaan), ในตัว (pajthua), ถูกจะ (ky'abca)</td>
<td>เกี่ยวกับ (kiawkab)</td>
</tr>
<tr>
<td>the</td>
<td>the</td>
<td>ฉัน (nán)</td>
<td>ฉัน (nán)</td>
</tr>
<tr>
<td>broken</td>
<td>broken</td>
<td>พัก (hâg)</td>
<td>พัก (hâg)</td>
</tr>
<tr>
<td>spring</td>
<td>spring</td>
<td>คุ้บในไม้ต้น (ry'duubajmáajphli), สริง (sapring), กระโดด (kradood), น้ำพุ (námphú)</td>
<td>คุ้บในไม้ต้น (ry'duubajmáajphli)</td>
</tr>
</tbody>
</table>

7 The classifier ฉัน (?tan) is added during the repair process (see chapter 7).
the SL parse again and the TCE found no semantic difference between them. Therefore, this new TC is an appropriate translation for Example 3.

3.5 Key Features of GRMT

The key features of GRMT are: simplicity, modularity, extendibility and multilinguality, by which we mean the following:

*Simplicity* Each step performed by GRMT is straightforward and simple to carry out.

*Modularity* GRMT's translation process is separated into three modules: ALMT, TCE and RI. Each module is comprised of sub-modules for easy modification and maintenance. ALMT is comprised of word treatment, word selection and word ordering. Word treatment deals with the syntax of the SL while word ordering is concerned with the syntax of the TL. The SL and TL are treated separately and related via the SL-TL dictionary. The SL and the TC analysis modules in the TCE are separated. The repair process in the RI module does not depend on languages, it can be applied to any languages pair. The knowledge-bases: constraints, dictionaries, grammars and lexicons required in GRMT are developed separately.

*Extendibility* GRMT is intended to be easily extendible to any other language. Since each component is separated not only in the translation process components but also in the knowledge-bases, each component can be extended easily to a larger domain. The constraints of the SL and those of the TL are developed separately. The SL-dictionary contains only the word form and its syntactic categories information. The SL-TL dictionary contains the SL word form and its corresponding words in the TL with its WordAsso numbers. The grammars we developed are based on the HPSG formalism which provides for adequate coverage of the language with few grammar rules. HPSG theory is a lexically oriented theory and lexical entries are complex. The lexicons are very rich in information and well organized. The lexical rules reduce the number of lexical entries (see section 6.1.1.3). Therefore, the knowledge-bases of new languages can easily be integrated into GRMT which is suitable for multilingual machine translation.

*Multilinguality* GRMT is designed to be amenable to a multilingual MT system. The multilinguality feature depends on the modularity and extendibility features of the system. GRMT is highly modular and extendible in two major respects. The first respect is the treatment of the SL and TL are independent of each other which means the same output of the SL constraints application module can be an input to more than one TL Word selection
module, provided that those TLs share common characteristics. The second respect is the knowledge-bases of the SL and TL required are developed separately, hence it is easy to add the new languages to the system. For example, the SL-constraints (e.g., Plurality, Continuous tense, Adjective and so on; see section 5.1.1) required in translating from English into Thai can be applied to the translation from English into Chinese and Japanese since Chinese, Japanese and Thai share those characteristics. Where French and Spanish share the same syntactic features with English, features which differ from those of Thai, Chinese and Japanese, then GRMT requires six analyzers and two sets of constraints to perform the translation between the two language families (Figure 3.2) while the transfer MT approach requires six SL analyzers, six TL generations and eighteen sets of transfer rules.

![Figure 3.2: The translation between languages of two language families](image)

We now go onto considerable detail of how GRMT works, first considering the English and Thai languages used to illustrate GRMT.

---

8 While Chinese is a member of the Sino-Tibetan language family and Thai and Japanese are not: Japanese is not proven to be related to any language family, but the most discussed possibility is that it could be a distant relative of the Altaic language family. Thai is a member of the unrelated Tai family [Crystal 1987]. However, Chinese, Japanese, and Thai are typologically similar in a number of ways (share certain features of phonology and syntax), which is why they have been grouped here. English, French, and Spanish, in contrast are all members of the same language family. Indo-European (This information provided by Dr. Brent D. Galloway).
Chapter 4

Differences Between Languages: The English and Thai Languages

An advantage employed by GRMT is its treatment of the differences between languages. Some significant characteristics of the source language and the target language, English and Thai in this experiment, are described in this chapter and exploited by GRMT. We conclude this chapter with the word classification and their usage in GRMT.

4.1 Differences between Thai and English

Thai is the standard spoken and literary language of Thailand. Thai is basically monosyllabic in word form. In addition to consonants and vowels\(^1\), each syllable has one of five phonemically differentiated tones (middle, low, falling, high and rising). The Thai language is written as a string of words with no explicit boundary marker between words. Different segmentations result in different meanings or no meaning. In this dissertation, a space is used to specify the word boundary in order to make interpretation simple and clear to the reader.

In Thai, morphology is restricted to derivation\(^2\) since the language makes no use of inflection\(^3\). The word form does not change according to its parts of speech. In other words, the same word form can be a noun, verb, adjective and so on (some examples are shown in

---

\(^1\) Thai has 44 consonants, 21 vowel forms and 32 vowel sounds.

\(^2\) \textit{Inflection} provides the various forms of any single word while \textit{Derivation} creates new words from the old ones [Covington 1994].
Table 4.1: Descriptions and the English corresponding words of ฅน (khon), ช่า (khaw) and ที่ (thii)

<table>
<thead>
<tr>
<th>Thai</th>
<th>Part of Speech</th>
<th>Semantic Description</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>ฅน- khon</td>
<td>noun</td>
<td>human being- a man, woman or child, not an animal</td>
<td>human, person</td>
</tr>
<tr>
<td></td>
<td>verb</td>
<td>to move (esp. something mainly liquid) around and mix something into it with a spoon or similar object.</td>
<td>person, stir</td>
</tr>
<tr>
<td></td>
<td>classifier⁵</td>
<td>classifier of human being (see page 43)</td>
<td>-</td>
</tr>
<tr>
<td>ช่า- khaw</td>
<td>pronoun</td>
<td>The male/female person already mentioned.</td>
<td>he/she</td>
</tr>
<tr>
<td></td>
<td>noun</td>
<td>a very high hill.</td>
<td>mountain</td>
</tr>
<tr>
<td>ที่- thii</td>
<td>noun</td>
<td>a particular area or position in space in relation to others.</td>
<td>place, space</td>
</tr>
<tr>
<td>relative</td>
<td></td>
<td>a particular part of the Earth’s surface, such as a stretch of land, a town, a building.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a position that can be used by someone for a particular purpose.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>something measurable in length, width or depth; area or volume that can be used or filled by a physical object; room.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>corresponds to the following English pronouns:</td>
<td>who, which, where</td>
</tr>
<tr>
<td></td>
<td>pronoun</td>
<td>who- person, which- thing, where- place.</td>
<td>who, where place</td>
</tr>
<tr>
<td></td>
<td>conjunction</td>
<td>used for introducing various kinds of clauses.</td>
<td>that</td>
</tr>
<tr>
<td></td>
<td>classifier</td>
<td>classifier of place, space.</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>preposition</td>
<td>shows a point in space.</td>
<td>at</td>
</tr>
</tbody>
</table>

Table 4.1). There is no inflection due to “verb agreement”, “tense”, “passive voice” and “number”. The non-inflection feature leads to the number of constraints which are used in the first phase, ALMT, of GRMT (see section 5.1.1).

The Thai verb does not have an inflection resulting from verb agreement. In Figure 4.1, it does not matter what the subject is, the word *eat* is translated as นุ- กิน. Tense, passive voice and number can be indicated by additional words as shown in Table 4.2. The words คัมลัน (kamlan), แล้ว (le’ew), ได้ (daj), ฆ่า (kh-th) and ฯ (ca?) are used to indicate different time aspects.

---

⁴ Semantic descriptions are provided from Longman Dictionary of Contemporary English [Longman 1992] and Thai-English Student’s Dictionary [Haas 1964].

⁵ Classifiers indicate the unit of a countable noun (see page 43).

⁶ Gerunds (e.g., parking, fishing,) and present participles as adjectives (e.g., boring, terrifying) are exceptions.
I eat rice every day.  
He eats rice every day.  
They eat rice every day.  
John eats rice every day.

Figure 4.1: Examples of kin (kin) and its subjects

- กิน (kamlan) is used before the verb to mark an on-going or progressive action.
- เล่าว (le’ew) shows past tense and denotes a completed action. เล่าว (le’ew) appears immediately after the verb and its object or right after the verb if there is no object.
- ได้ (daj) also shows past tense. It immediately precedes the verb. It is often used in conjunction with เล่าว (le’ew).
- เธว (khaoj) also precedes the verb and is used to designate an action in the indefinite past.
- จะ (ca?) is used to mark an action to be completed in the future. It always appears directly before the verb.

Some examples are illustrated in Figure 4.2. Tenses can also be conveyed by context, or by means of adding time indicators e.g., today, tomorrow, last year, yesterday. A phrase can refer to an action at any time, without a time reference.

Table 4.2: Additional words to express Tense, Passive Voice and Plurality in Thai

<table>
<thead>
<tr>
<th>Thai</th>
<th>English</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>กิน (kamlan) + V</td>
<td>V to be + V-ing</td>
<td>aspect</td>
</tr>
<tr>
<td>V + (obj) + เล่าว (le’ew)</td>
<td>V2</td>
<td>past tense</td>
</tr>
<tr>
<td>ได้ (daj) + V + เล่าว (le’ew)</td>
<td>V2</td>
<td>past tense</td>
</tr>
<tr>
<td>จะ (ca?) + V</td>
<td>will + V</td>
<td>future tense</td>
</tr>
<tr>
<td>ทำ (thúug) + V</td>
<td>V to be + V3</td>
<td>passive voice</td>
</tr>
<tr>
<td>N + เหลา (la’aj)</td>
<td>N-s, N-es</td>
<td>plurality</td>
</tr>
</tbody>
</table>
In Thai, passive voice is rarely used. It is used mainly when discussing an unpleasant situation; Figure 4.3 illustrates some examples. The word ณ (thuug) denotes the passive voice in Thai. In the third example, the first of the two Thai sentences, which is passive, is not generally used in Thai even though it is grammatically correct. The second sentence carries the same meaning but it is in active voice.

I am swimming.

He bought a book.

The duckling hid in the farmyard.

The wedding will take place in December.

He was killed.

He was arrested.

The book was taken by him.

Figure 4.2: Examples of additional words to indicate different tenses and aspects.

He bought a book.

The duckling hid in the farmyard.

The wedding will take place in December.

He was killed.

He was arrested.

The book was taken by him.

Figure 4.3: Examples of passive voice in Thai.

In Thai, to express a quantity of countable nouns, either by quantifiers e.g., แฝด (la’aj), บาง (baan) or numbers, classifiers are required. A classifier indicates the unit of a noun, each countable noun is compatible with a specific classifier (see Figure 4.4).

7 แฝด - la’aj corresponds to the word *many* in English.
8 บาง - baan corresponds to *some* in English.
The structure of a Thai phrase is similar to that of English in that the typical sentence contains subject, verb and object, in that order. However, some structures are different e.g., in Thai the head (noun) must precede its attributes, possessive pronouns and determiners. The negative of “can”, and “could” is formed in the reverse order of the order in English. Some examples are shown in Figure 4.5. These differences must be considered in the ordering process (see section 5.3).

Figure 4.4: Examples of classifiers

In addition, auxiliary verbs in English are needed in many cases, e.g., in front of an adjective or negative “not”, but they are not used in Thai for the same expression as shown in Figure 4.6. The English pronoun “there” followed by auxiliary “to be” shows that something or someone exists or happens (see Figure 4.6). “There” plus auxiliary “to be” corresponds to the word มี (mii) in Thai. These differences are considered in the constraint application process.

Figure 4.5: Examples of Thai attributive constructions

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Of course there are additional significant differences between Thai and English. The differences illustrated in this section are to provide the reader with some insight of differences GRMT takes advantage of in generating accurate translations.

### 4.2 Word Classification

In chapter 3, we sketched the overall architecture of GRMT in which the relationship between words was employed in several modules e.g., TL word selection and classifier selection in ALMT (the classifier selection is applied to some languages which require classifiers e.g., Thai, Chinese, Japanese), analysis in TCE, comparison and repair in RI. The semantic relationship between words was built based on the words' classes. In this section, we discuss our word classification technique and criteria by which we created the semantic relationship between words and word classes.

\[
\begin{array}{ll}
\text{I am glad} & \text{ฉัน ดีใจ} \\
\text{He does not eat} & \text{เขา ไม่กิน} \\
\text{His neck was long.} & \text{คอ มีหัวยาว} \\
\text{There is a book.} & \text{มี หนังสือหนึ่ง เล่ม} \\
\end{array}
\]

Figure 4.6: Examples of Thai constructions without auxiliary verb

In designing our word classification, we have considered the *A Kind Of* (AKO) relation which was designed to be used in the CICC MMT project (see section 2.3.3.1), the topic hierarchy for physical objects [Cercone et al. 1992] and Hypernym in WordNet 1.6 for English [Cognitive Science Laboratory 1998] with significant introspection\(^9\).

\(^9\) Different point of view will result in different epistemologies (word classification). What important is that independent of epistemologies, the same classification be use across languages to whatever extent possible. Our word classification is the first attempt based on the CICC AKO, topic hierarchy for physical objects, and Hypernym in WordNet 1.6, as GRMT evolves so will the word classification.
In the CICC MMT project, AKO, co-occurrence information and concept identifier (concept-id) were developed. AKO is basically a relationship between a word and its concept class (or semantic domain). Each participating country developed the concept classification for its language based on its own criteria; therefore, the resulting AKO systems are different. For example, the Thai concept classification was developed by grouping words with a specific definitions into the same class [CICC 1995f]. A part of Thai concept classification is shown in Figure 4.7. Each class was assigned a number. The digit which comes after a hyphen "-" is the subclass number, the digit which comes before a hyphen is the superclass number. For example, Animal with AKO number 1-1-2 is a second subclass of the superclass 1-1 which is an AKO number of a class Living Thing. The Indonesian concept classification (Figure 4.8) was developed based on syntactic categories, e.g., nouns are classified into two main classes: concrete thing and abstract thing; verbs are classified into class event with three subclasses state, process and action. Adjective and adverb are classified into classes properties and manner respectively [CICC 1995b]. The numbering system of the Indonesian AKO is the same as that of Thai AKO. However, AKO specifications for Thai, Indonesian and Chinese have the same problems regarding consistency and coverage [CICC 1995a; CICC 1995g; CICC 1995h].

![Figure 4.7: A part of Thai Concept Classification (CICC MMT)](image)

Co-occurrence information indicates the syntactic relationship between words e.g., subject and verb relation, verb and second complement relation [CICC 1995j]. Thai co-
occurrence information was designed from the co-occurrence of two to five adjacent words. The Indonesian version only considered co-occurrence from two adjacent words [CICC 1995c].

Concepts (variant meanings or allosemes) of words in each language were assigned a concept-id. For example, the concept-id 3be7d8 is assigned to the Thai word ขยะแคร์ (khry'ankraari) which corresponds to the English word filter in the sense of “a porous material through which a fluid is passed to separate out matter in suspension.” [CICC 1995g]. Each country designed its own concept-id system. Therefore, a concept-id linking table was needed to link the concepts of each language to IL.

In the Thai analysis process, AKO and syntactic information were used to determine the semantic relationship between two nouns [CICC 1995f]. For example, the syntactic categories of นางพ่อ (naajphe'Aed) and ผู้ชาย (su'nthaan), Example 1, are title noun and proper noun (Table 4.3); therefore, the semantic relationship between นางพ่อ (naajphe'Aed) and ผู้ชาย (su'nthaan) is “name” according to the information available in Table 4.4. The semantic relationship between ชาย (pl'a'ag) and หญิง (påagkaa), Example 2, is “part-of” since their syntactic categories are common nouns with the AKO numbers 1-2-6 and 1-2-5-1-2 respectively.

Co-occurrence information was used to solve lexical ambiguity problems [CICC 1995j]. In the Thai generation process, co-occurrence was used in selecting the surface form of verbs and adjectives. For example: the word beautiful in “beautiful lady” corresponds to the
wordสุจริ (su'aj) which means “having beauty”. However, the word beautiful in “beautiful song” corresponds to the word เปริง (phra') in the sense of “very good” according to the co-occurrence information available in Figure 4.9 [CICC 1995i].

Example 1: นนท์- Dr. (naajphe'ed- Dr.) (su'nthaan- Sunthorn)

Example 2: ปากกา- pen (pla'ag- cap) (paagkaa- pen)

<table>
<thead>
<tr>
<th>Thai</th>
<th>Corresponding Word in English</th>
<th>Syntactic Categories</th>
<th>AKO</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>นนท์- Dr.</td>
<td>title noun</td>
<td>-</td>
<td>Abbreviation of doctor.</td>
<td></td>
</tr>
<tr>
<td>Sunthorn</td>
<td>proper noun</td>
<td>-</td>
<td>person’s given name.</td>
<td></td>
</tr>
<tr>
<td>cap</td>
<td>common noun</td>
<td>1-2-6</td>
<td>a movable cover</td>
<td></td>
</tr>
<tr>
<td>pen</td>
<td>common noun</td>
<td>1-2-5-1-2</td>
<td>an instrument for writing or drawing with ink.</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: Semantic relationships assigned for noun-noun combination in CICC MMT

<table>
<thead>
<tr>
<th>Semantic relationship</th>
<th>Syntactic category and AKO number</th>
<th>Syntactic category and AKO number</th>
</tr>
</thead>
<tbody>
<tr>
<td>modifier</td>
<td>collective noun 2-3-4</td>
<td>common noun</td>
</tr>
<tr>
<td>part of</td>
<td>common noun AKO=1-3-1-1-2, 1-3-1-1-4, 1-3-2-3, 1-2-6</td>
<td>common noun AKO=1-1-1,1-3-1-1-3, 1-2-5-1-2</td>
</tr>
<tr>
<td>name</td>
<td>title noun</td>
<td>proper noun</td>
</tr>
</tbody>
</table>

Figure 4.9: A co-occurrence of the word เปริง (phra') [CICC 1995i]
AKO was used in the analysis process of Indonesian and Chinese also but no public document was produced to show how AKO was used. In the Indonesian generation process, co-occurrence was used to determine whether the occurrence of two adjacent words in a sentence was correct [CICC 1995d].

Developing an AKO system required great effort; however, AKO application was not fully utilized in the CICC MMT system. Our explorations revealed that suitably constructed AKO information is useful in selecting the surface form of TL words in a translation process and it helps to determine the occurrence of adjacent words in the expression. In addition, if AKO information for each language was designed using the same criteria, the AKO information might also be used in linking concept-id.

The topic hierarchy for physical objects was designed to be used in the English Conversation System (ECO) family formalism of semantic network [Cercone et al. 1992]. ECO was developed at University of Alberta during 1974-1978 by Len Schubert, Nick Cercone, Randy Goebel and Al Covington. ECO represents an extended semantic network formalism which is used for representing troublesome natural language constructions, e.g., logical connectives, quantifiers, descriptions, modalities. ECO classifies propositions according to their topic e.g., color proposition, location proposition, size proposition and so on. ECO structures propositions associated with each concept according to a topic hierarchy. A topic is defined as a predicate over proposition-concept pairs. For example, for the proposition *A zebra has black and white stripes*, a topic predicate *coloring* is instantiated in relation to the concept *zebra*. Topic predicates which are linked by subtopic and supertopic form a topic hierarchy. Topic hierarchies provide a basis for organizing the propositions attached to a node of a particular kind. Topic hierarchies are useful in organizing the retrieval of information relevant to the implications of concepts. To accelerate a deductive inference for *Types* in the ECO family, the partitioning hierarchies of type predicates are used as a logical representation of the relationships among the predicates that appear in the hierarchy. Figure 4.10 illustrates a partitioning type hierarchy for *thing*. The numbers following the predicates are assigned when the hierarchy is created or changed. This number is used to determine the
subsumption or disjointness relationships. For example: *wolf* and *human* are disjoint because there is no overlap between the numbering range associated with wolf [38,38] and human [17,26]. *Wolf* is subsumed by *creature* because wolf [38, 38] is ranged within creature [16,40]. The same predicate can appear in several hierarchies, all connected at that predicate. Overlap hierarchies are also possible.}

*WordNet 1.6* is a program that allows users to explore an on-line dictionary on the basis of semantic, rather than alphabetic, similarities [Cognitive Science Laboratory 1998]. WordNet does not contain syntagmatic relations linking words from different syntactic categories because the syntactic categories noun, verb, adverb and adjective are treated separately due to the semantic differences between the relations that link words and concepts from those syntactic categories [Fellbaum 1998]. For example, the relation that links nouns differs from the relation that links verbs. The semantic relation that is most important in organizing nouns is a relation between lexicalized concepts. This relation is called *Hypernym*. Hypernym is the generic term used to designate a whole class of specific instances. Y is a hypernym of X if X is a (kind of) Y. For example: *bird* is a hypernym of *robin*. This semantic relation organizes nouns into a lexical hierarchy. A lexical hierarchy can be reconstructed by following the trail of hypernymically related synonym sets. The semantic relation between sets is *is-a* or *is-a-kind-of*. This hierarchy goes from many specific terms at the lower levels to a few generic terms at the top. The nouns in WordNet form a lexical
inheritance system where by all of the properties of the superordinate are assumed to be properties of the subordinate as well. For example: according to the WordNet; robin is-a thrush, a thrush is-an oscine, an oscine is-a passerine, a passerine is-a bird. Therefore, robin inherited perching from passerine, singing from oscine, and flying and migration from thrush (This example is taken from [Miller, 1998]). In WordNet, verbs are grouped together as sets of synonyms, like nouns. However, the semantic relation between set of verbs is expressed by is one way to. For example: see is-one-way-to perceive, drink is-one-way-to consume or ingest.

We further developed our notion of word classification by classifying words into categories based on shared characteristics. From ECO we borrow the notion that subclasses are subsumed by their superclasses, cf., the topic hierarchy for physical objects. Words in our classification form a lexical inheritance system like nouns in WordNet. Since we were concerned with consistency as well and we wanted to apply this classification in MT, we needed to classify words in any language under the “same umbrella” even though there were differences at some levels between different languages. Our solution is to classify words into categories according to the same criteria, regardless of language as discussed in the next section.

4.2.1 Word Classification Criteria

In GRMT, word classification information is applied to be used in word selection, classifier selection, TCE analysis, comparison and repair processes.

To select the most appropriate word for each input word in ALMT, we consider the relationship between the close proximity words. We must know which word can occur with which word in the same expression. To obtain that information, semantic information of each word should be considered and also the semantic relationship between proximity words. Thus, the criteria used in classifying words considers the definition or meaning of words and their usage. Words which share significant features are classified into the same group. With these criteria, words in GRMT dictionaries are classified into thirteen categories: Thing, Abstraction, Phenomenon, Event, State, Be, Reference term for thing, Group/Grouping, Be able to, Get/Acquire, Have, Accumulate/Collect, and Attribute value. New categories can be added in future if it is necessary to do so. A classification fragment is shown in Figure 4.11 (see also appendix A).
Since we classified words not only according to their meaning but also their usage, words in the same class will automatically be in the same syntactic category, e.g., words in the class *thing* are nouns, words in the class *act* (a subclass of *abstraction*) are verbs, words in the class *attribute value* are adjectives and adverbs. The semantic relation between classes and their superclass is *is-a*. For example: *turkey* is-a *fowl*, *fowl* is-a *bird*, *bird* is-a *vertebrate*, *vertebrate* is-an *animal*, *animal* is-a *living thing*, *living thing* is-a *concrete thing*, *concrete thing* is-a *thing*; *say* is-a *way to utter*, *utter* is-a *way to communicate*, *communicate* is-a *way to interact*, *way to interact* is-a *way to act*, *act* is-an *abstraction*. The semantic relation forms a lexical inheritance hierarchy which generally means all properties of the superclass are assumed to be

![Figure 4.11: A part of word classification in GRMT.](image)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
properties of the subclasses as well. This property is useful in the comparison step of TCE, the second phase of GRMT.

A word can be classified into more than one class depending on context and usage. For example: turkey is also classified into the class of poultry, poultry is-a meat, meat is-a foodstuff, foodstuff is-a food, food is-a substance, substance is-an object, object is-a non-living thing, non-living thing is-a concrete thing, concrete thing is-a thing.

Words which are used to link phrases with clauses or connect clauses together or provide the basis for a descriptive phrase e.g., prepositions, conjunctions, definite determiner are classified as well. A separate classification is built for those words, see appendix D for examples.

The scope of any class can be narrowed, if necessary, by subdividing a class into subclasses. For example, consider the class “vehicle”. If we classify “vehicle” into two subgroups by mode of transportation as shown in Table 4.5: by land, and by sea or by air. Then the classifiers จำนวน (khabuan) and จำนวน (khabuan) are compatible with the first subgroup, by land, while the classifier จำนวน (khabuan) is compatible with the second subgroup, by sea or by air. The group “by land” can be classified further into “Train” and “Road/Rails vehicle (but not train)”. The classifiers จำนวน (khabuan) and จำนวน (khabuan) are compatible with the group “train” and “Road/Rails vehicle” respectively. This classification is used to select an appropriate classifier for a noun if it is required, in the word addition step of ALMT and the analysis process of TCE.

Table 4.5. The subgroup of vehicle.

<table>
<thead>
<tr>
<th>Class</th>
<th>Subclass</th>
<th>Examples</th>
<th>Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land</td>
<td>Train</td>
<td>รถไฟ (รอดผา-train)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td>Road/Rail</td>
<td>รถจักรยานยนต์ (รอดข้าว-automobile)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td>Rail vehicle</td>
<td>รถจักรยานยนต์ (รอดบ้าน-trailer)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>รถจักรยานยนต์ (รอดเล็ก-rickshaw)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>รถจักรยานยนต์ (รอดม้า-horse carriage)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>รถจักรยานยนต์ (รอดคักก้าน-attachment)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>รถจักรยานยนต์ (รอดถ้า-streetcar, tram)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td>Air or Sea</td>
<td></td>
<td>เรือใบ (ryabay-sailboat)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>เรือสะพานยาว (ryaha-anjaaw-“long-tailed boat”)</td>
<td>จำนวน (khabuan)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>เครื่องบิน (khry-“airplane”)</td>
<td>จำนวน (khabuan)</td>
</tr>
</tbody>
</table>
For efficient MMT, words in different languages must be classified into the same classification system, some examples are shown in Table 4.6. For example, we can use the classification information of each word as a part of the semantic information of any phrase which contains that word. The last issue that we must be concerned with in classifying words is "consistency".

Each class is assigned a number which is called *Word Association* (WordAsso) number. WordAsso number represents a relation between a word and its class. Words which belong to the same class are assigned the same WordAsso number. We number classes using the convention that was used in CICC AKO for Thai. The digit which comes after a hyphen "-" is the subclass number, the digit which comes before a hyphen is the superclass number. For example, in Table 4.6, *Imaginary creature* with WordAsso number 1-2-1-2-2 is a second subclass of the superclass 1-2-1-2 which is a WordAsso number of a class *Imaginary being*. The class *Imaginary being* is the second subclass of a class *Supernatural thing* with WordAsso number 1-2-1. This numbering system is used for easy modification.

<table>
<thead>
<tr>
<th>Class</th>
<th>Sample Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Thing</td>
<td></td>
</tr>
<tr>
<td>1-2 Abstract thing</td>
<td></td>
</tr>
<tr>
<td>1-2-1 Supernatural thing</td>
<td></td>
</tr>
<tr>
<td>1-2-1-1 Supernatural being</td>
<td>Jesus, God, พระเจ้า (phaphûdthacaaw-The Lord Buddha)</td>
</tr>
<tr>
<td>1-2-1-1-1 Spirit</td>
<td>ghost, ผี (phi'ii- ghost) ผิวิจี (winjaan- spirit, soul)</td>
</tr>
<tr>
<td>1-2-1-1-2 Evil spirit</td>
<td>satan, devil, demon, ผี (phi'ipo'ab- ghoul)</td>
</tr>
<tr>
<td>1-2-1-2 Imaginary being</td>
<td>Santa Claus, tooth-fairy, mermaid,</td>
</tr>
<tr>
<td>1-2-1-2-1 Imaginary person</td>
<td>น้ำมัน (naanny-ag- mermaid)</td>
</tr>
<tr>
<td>1-2-1-2-2 Imaginary creature</td>
<td>Pegasus, unicorn, กระทุ่ม (khrûd - Garuda¹⁰), นaga (nâag-Naga¹¹)</td>
</tr>
</tbody>
</table>

¹⁰ Garuda is a supernatural half-man and half-bird vehicle or bearer of Vishnu. It is a state symbol of Thailand.
¹¹ Naga is a legendary serpent (Longman Dictionary).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
4.2.2 The semantic relationship development in GRMT

Since the word classification is based on semantic information of words and their usage, a link between WordAsso numbers indicates a relationship between words. This relationship indicates which word can occur with which word in the same expression. Based on this observation, we have built a relationship between words in the form of WordAsso numbers according to the contexts in which the words occur.

The possible contexts in which a given word can occur are part of a word’s meaning. Therefore, we make use of the contextual knowledge to disambiguate polysemous words. For example, we interpret the meaning of *quack* in the first and the second sentences in Table 4.7 (the first column) differently. In the first sentence, the relationship between *duck* (in the sense of “a common swimming bird with short legs and a wide beak, either wild or kept for meat, eggs and soft feathers.”) and *quack* indicates the meaning of *quack* as “to make the sound that ducks make” shown in the second column. However, the relationship between *quack* and *physician* in the second sentence determine the meaning of *quack* as “a person dishonestly claiming to have medical knowledge or skills”. Based on this observation, the WordAsso relation between the classes *animal* and *animal made sound*, *human’s occupation* and *human’s occupation* are established as shown in the third column of Table 4.7. The class *animal* and *animal made sound* are related since *duck* belongs to the class *aquatic bird* with the WordAsso number 1-1-1-2-1-2-1-2-1 and *quack* belongs to the class *Animal made sound*.

<table>
<thead>
<tr>
<th>Sample sentences in English/Thai</th>
<th>Description of a given word</th>
<th>WordAsso relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>“quack, quack”, said the mother duck.</td>
<td>to make the sound that ducks make.</td>
<td>1-1-1-2-1-2-1-2, 4-1-1-2</td>
</tr>
<tr>
<td>A quack is an untrained physician.</td>
<td>a person dishonestly claiming to have medical knowledge or skills.</td>
<td>13-5-1, 13-5-1</td>
</tr>
<tr>
<td>How high is the mountain?</td>
<td>having a top that is some distance, especially a large distance, above the ground.</td>
<td>-</td>
</tr>
<tr>
<td>She likes that tall man.</td>
<td>having a greater than average height.</td>
<td>13-8-9-2, 1-1-1-1-1-1</td>
</tr>
</tbody>
</table>

Table 4.7: Examples of words and their contexts
Table 4.8: Classes and their sample words

<table>
<thead>
<tr>
<th>WordAsso Number</th>
<th>Class</th>
<th>Sample words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1-1-2-1-2-1</td>
<td>Aquatic Bird</td>
<td>duck, เต่าน้ำ (ped- duck), duckling, นกเกี้ยน (làngphèng- duckling), swan, นกยูง (hô- swan), stork, นกกระเรียน (nòkkrása- stork)</td>
</tr>
<tr>
<td>4-1-1-2</td>
<td>Animal made sound</td>
<td>quack, แวดน้อย (khweé-g- quack), peep, นิ้ว (pip- peep)</td>
</tr>
<tr>
<td>13-5-1</td>
<td>Human’s occupation</td>
<td>nurse, นวดประจำ (phajaabaan- nurse), quack, หมอเต่อ (má- thý-an- quack), physician, หมอ (má- doctor, physician)</td>
</tr>
<tr>
<td>13-8-9-2</td>
<td>Attribute value of</td>
<td>tall, ตALL (sùn- tall), short, ตี (tia- short)</td>
</tr>
<tr>
<td>1-1-1-1-1</td>
<td>Male12</td>
<td>boy, น่าสดใส (dèngphùuchaaaj- boy), man, ผู้ชาย (phûuchaaaj- man)</td>
</tr>
</tbody>
</table>

with the WordAsso number 4-1-1-2 (Table 4.8). Both quack and physician belong to the same class Human’s occupation with the WordAsso number 13-5-1.

In Table 4.7, the attribute values (of height) high and tall in the third and the fourth sentences are selective about the nouns they modify. High is used for measurements of distance above ground while tall is used for height measurements, e.g., for people [Longman 1992]. According to their usage and implied reference set, the relationship between the classes Attribute value of height used for people and Male (the third column of Table 4.7) is developed. tall belongs to the class Attribute value of height used generally for people with the WordAsso number 13-8-9-2 and man belongs to the class Male with the WordAsso number 1-1-1-1-1.

In addition to consider the possible contexts in which a given word can occur, the associations between words that are interchangeable in a context and the associations between words that frequency occur together are considered in developing the WordAsso relations for GRMT (see section 5.2).

The construction of WordAsso relation and its use will be presented in chapter 5, some examples will be provided.

---

12 The classes Male (1-1-1-1-1) and Female (1-1-1-1-1-2) are subclasses of Individual (1-1-1-1). This classification is for associating words in Male and Female with the particles ผู้- khráb ค่ะ- kha in Thai. Particles are sentence-ending words. ผู้- khráb and ค่ะ- kha are used to show politeness, courtesy and respect to another person. ผู้- khráb is used by men and ค่ะ- kha is used by women.
Chapter 5

Analysis Lite Machine Translation

...and since word-for-word translations are surprisingly good, it seems reasonable to accept a word-for-word translation as a first approximation and then see what can be done to improve it.

— Victor H. Yngve (1955)

Analysis Lite Machine Translation, the first phase of GRMT, generates an appropriate translation candidate for an input text. ALMT was designed around two simple notions: the more accurately ALMT generates a TC, the less work is required in the latter phases; and generating the TC must be done quickly. Therefore, ALMT generates a TC by considering the differences between language pairs in terms of syntax and semantics without performing any sophisticated analysis. ALMT performs its task by judiciously selecting a few efficient heuristics, constraints, and semantic principles to apply when appropriate. ALMT first applies

![Figure 5.1: ALMT Architecture](image)

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the SL constraints to simplify the structure of the SL, maps each SL word into all possible meanings in the TL. The most appropriate TL word is selected by applying a semantic relationship between words and then rearranging the selected words according to the grammar of the TL. ALMT comprises three modules as shown in Figure 5.1, word treatment, word selection, and word ordering.

**5.1 Word Treatment**

There are two steps performed by *word treatment* module: SL-Constraints Application and Dictionary look-up (Figure 5.1). The SL constraints are used to simplify the structure of the SL. After simplification, the word treatment maps each SL word to all possible corresponding words in TL by consulting the SL-TL dictionary. An inflectional analysis might be needed in the dictionary look-up stage. This analysis is quite simple but useful as it saves dictionary space and reduces search time. The knowledge-bases required in the word treatment module are: the SL constraints, the inflectional rules, the SL dictionary and the SL-TL Dictionary.

**5.1.1 SL Constraints Application**

The SL constraints, the morphological and syntactic characteristics of the SL which differ from those of the TL, are incorporated in this step to narrow the scope of possible TL words which correspond to each SL word. Table 5.1 illustrates some SL constraints we considered according to the differences between English and Thai as discussed in section 4.1.

<table>
<thead>
<tr>
<th>SL-Constraints</th>
<th>Descriptions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plurality</td>
<td>noun_(e)s -* noun + plural</td>
<td>The books ... -* The book + plural</td>
</tr>
<tr>
<td>Continuous tense</td>
<td>V to be + V-ing -* V + ing</td>
<td>I am swimming. -* I + ing + swim</td>
</tr>
<tr>
<td>Passive voice</td>
<td>be + V3 -* passive + V</td>
<td>He was arrested. -* He passive + arrest</td>
</tr>
<tr>
<td>Adjective</td>
<td>V to be + adj -* adj</td>
<td>I am glad. -* I glad</td>
</tr>
<tr>
<td>Negative</td>
<td>do + not -* not</td>
<td>He does not eat -* He not eat</td>
</tr>
</tbody>
</table>

1 This constraint does not apply in the case of the irregular plural forms, e.g., children, feet, men, women, and so on. The irregular plural forms are entries in the dictionaries.
Since Thai word form does not have an inflection resulting from plurality or tense, the inflections of plural noun and present participle forms in English are removed in this step. Once the inflections are removed, the features plural and ing are added to preserve the meaning of the original sentence. The structure of a passive voice is examined as well; the verb inflection and auxiliary to be are removed. The feature passive is added to indicate the passive voice. The features plural, ing and passive will be replaced with the appropriate corresponding TL words in the ordering step of ALMT.

The auxiliaries to be which precedes an adjective and to do which precedes a negative “not” are discarded since they are not required in the same expression in Thai. Figure 5.2 illustrates the outputs of SL constraints applications of Example 1 and Example 2. The auxiliary is is removed from Example 1 according to the adjective constraint. The inflection -s of the word symbol in Example 2 is removed and the feature plural is added, the passive voice form are used is replaced with passive use, the auxiliary do is discarded and the continuous tense form is replaced with ing talk.

Example 1:
SL: The spirit is willing but the flesh is weak.
SL Constrains- output: The spirit _ willing but the flesh _ weak.

Example 2:
SL: Algebraic symbols are used when you do not know what you are talking about.
SL Cons- output: Algebraic plural symbol passive use when you not know what you ing talk about.

Figure 5.2: SL constraints applications examples

5.1.2 Dictionary Look-up

After the SL constraints are applied, each SL word will be used as a keyword to search for its corresponding words in the TL. If the keyword used is found in the SL-TL dictionary, all possible corresponding TL words will be attached to that SL word. If the keyword is not found, inflectional analysis is performed and the search is restarted. The inflections of each entry are not entries in the dictionaries, therefore the size of the SL dictionary and the search times are reduced.

In our case in which English is regarded as SL, the following inflectional forms are considered: past simple tense (-d, -ed), gerund (-ing), subject-verb agreement (-s, -es) and
comparatives (-er, -est); plurality and present participle are treated in the SL Constraints Application. An output of this step is a list of all possible corresponding words of SL words in the TL with their WordAsso numbers. The second column (Table 5.2) shows the dictionary look-up result for Example 3 based on the information available in SL-TL dictionary. The descriptions of the words are provided in the third column.

Table 5.2: Example result of Dictionary Look-up of Example 3

Example 3:
SL: The spirit is willing but the flesh is weak.
SL Constraints-output: The spirit willing but the flesh weak.
Dictionary Look-up-output:

<table>
<thead>
<tr>
<th>SL Constraints output</th>
<th>Dictionary Look-up</th>
<th>Descriptions</th>
</tr>
</thead>
</table>
| The spirit            | ความหมาย [1-2-1-1, 1-2-4] | • thought of as remaining alive without appearing in physical form.  
                           |                    | • a life-giving force  
                           |                    | • mental or moral strength to resist |
| but                   | ต่อ [13-16-3] | ready |
| the flesh             | เนื้อ [1-4-1-3-1] | the soft parts of an animal’s body; especially muscular tissue |
| weak                  | ยืดหยุ่น [13-21-1, 13-10-3] | • not strong; especially not strong enough to work or last properly  
                           |                    | • not strong or firm in character; lacking power |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
5.1.3 Dictionaries

The SL dictionary provides the information required in the constraint application and the inflectional analysis steps. Each lexical entry contains its syntactic categories. The inflections of each entry are not included. Figure 5.3 illustrates a fragment of SL dictionary which is in a Prolog representation: the categories of the word *all* are adjective, adverb, pronoun, noun and determiner. *Hid* is an adjective and a past simple tense form of *hide*, therefore one of its categories is *vpast*. *Vpast* is a category of the verbs which are in a past simple tense form (see also appendix B).

```
entry_e('Egyptian',[n]).
entry_e('all',[adj,adv,pron,n,det]).
entry_e('dream',[v,n]).
entry_e('duckling',[n]).
entry_e('hid',[adj,vpast]).
entry_e('should',[aux]).
entry_e('that',[pron,conj,adj]).
```

Figure 5.3: Example of SL Dictionary

The SL-TL dictionary is required in the dictionary look-up step. It is used to map the SL word to all possible corresponding words in the TL. The SL-TL dictionary contains the SL entry and all of its possible corresponding words in the TL. The TL words are ordered according to the frequency of their usage. Each TL word is related to its WordAsso numbers. Figure 5.4 illustrates a part of SL-TL dictionary which is in a Prolog representation: the word

```
entry('Egyptian',[wordasso('n')]).
entry('all',[wordasso('adj,adv,pron,n,det')]).
entry('dream',[wordasso('v,n')]).
entry('duckling',[n]).
entry('hid',[adj,vpast]).
entry('should',[aux]).
```

Figure 5.4: Example of SL-TL Dictionary
duck corresponds to four Thai words; ติ่ง (pêd) with WordAsso 1-1-1-2-1-2-1 (Aquatic Bird) and 1-1-2-1-2-1-2-4-1 (Poultry), นิ่ว (kôm) with WordAsso 2-1-5-2 (A way to move so as of a body part), หนักสั่ง (câbkôdlôm) with the WordAsso 2-1-5-4-1 (a way to move thing with force) and ผ้าอินคานา (phâalininkhaaw) with the WordAsso 1-1-2-1-1-1 (Fabric), see also appendix C.

5.2 Word Selection

One difficult task in the MT process is to select the most appropriate word in the TL which preserves the meaning of the corresponding word in the SL if there are more than one possibility. To select the most appropriate word in TL for each SL input word, we need to know the meaning of the input word. The meaning of the input word can be determined by the context in which the input word occurs as we discussed in section 4.2.2. In Example 4 (Figure 5.5), the semantic relationship between the word ติ่ง (pêd) and ติ่ง (ตู่) indicates that the word ติ่ง (pêd) means "not shut". Therefore, its corresponding word in English is open; this expression means "He opens the closet." The word ติ่ง (pêd) in Example 5, Figure 5.5, cannot be translated as "open" since the semantic relationship between ติ่ง (pêd) and (ติ่ง) is stronger than the relationship between (ติ่ง-pêd) and (ติ่ง-tuu). Thus, the meaning of the sentence in Example 5 corresponds to "He turns on the light in the closet". This example and others like it lead to the notion that physical proximity implies a stronger connection. Therefore, the word selection process in ALMT is carried out by considering the relationship between the close proximity words.

Example 4: ติ่ง (pêd-open) (ตู่-closet)

Example 5: ติ่ง (pêd-turn on) (ตู่-light) (ตู่-closet)

Figure 5.5. The relationship between the proximity words -Examples 4 and 5.

Figure 5.6 illustrates some Wordasso relations we developed for ALMT (see section 4.2.2), the word with WordAsso number in the first argument can occur with the word which has a WordAsso number shown in the second argument. For example: words with the
WordAsso number 1-1-1-2-1-1 can occur with any word which has WordAsso number 13-8-2 or 13-8-5-2.

```
wordasso_rel('1-1-1','[2-1-5-1']).
wordasso_rel('1-1-1-2','[2-2]').
wordasso_rel('1-1-1-2-1-1','[13-8-2,13-8-5-2]').
wordasso_rel('1-1-1-1-1-2','[13-7-1]').
wordasso_rel('2-5-5','[r12]').
```

Figure 5.6. Examples of WordAsso relations

As well as applying the semantic relationship between words in the form of the WordAsso relation, the selection step continues. According to a linguistic theory, a collocation\(^2\) consists of basically two or three lexical words [Newmark 1981] and our investigation of a number of sample sentences; we consider the relationship between words in the range of four consecutive words in our selection step. Words in close proximity are considered to have stronger connections. Our selection steps are designed as follows:

```
Program word_selection;
X and newX are arrays of xi;
x_i is a list of the TL words which corresponds to the ith word of the SL input sentence;
n is the number of words in the SL input sentence;
i, j, m and n are integers;
x_i is null if i < 1;
x_i is null if i > n;

% Use x_j to determine the most appropriate word for x_i,j and x_i-1,j;
Procedure select (j, newX);
begin
  for i = 1 to n do
```

\(^2\) collocation is a group of words in a sentence in which a word with particular meaning is supplanted by another word because of other words in the environment. For example: the word pay in to pay attention means to use: the word commit in commit a crime means to do.
begin
    if $x_i$ contains only one word
    then begin
        use the word in $x_i$ to select the appropriate TL word(s) in $x_{i,j}$;
        if wordasso_re ($x_i$, $x_{i,j}$) = true
        then begin
            $x_{i,j} = \text{new } x_{i,j}$;
            end;
        else begin
            newX = $x_i$;
            use the word in $x_i$ to select the appropriate TL word(s) in $x_{i+1,j}$;
            if wordasso_re ($x_i$, $x_{i+1,j}$) = true
                then begin
                    $x_{i+1,j} = \text{new } x_{i+1,j}$;
                end;
            else newX = $x_i$;
            end;
        end;
    end;
    $i = i + 1$;
end;

% Procedure word_check ($j$, $X$);
begin
    do while ($j \leq 3$)
    begin
        select ($j$, newX);
        if $X \neq \text{new } X$
            then begin
                $X = \text{new } X$;
                $j = 1$;
                word_check ($j$, $X$);
            end;
        else begin
            if $j < 3$
                then begin
                    $j = j + 1$;
                    word_check ($j$, $X$);
                end;
            else begin
                select_final ($j$, $X$);
                $j = 1$;
                word_check ($j$, $X$);
            end;
        end;
    end;
% Select the first meaning appearing on the list of meanings of each word
Procedure select_final (j, newX);
    begin
        for i = 1 to n do
            begin
                if \( x_i \) contains only one word
                    then \( i = i + 1 \)
                else select the first word in \( x_i+j \) as the most appropriate word;
            end;
    end;

% main
begin
    X = input sentence;
    newX = X;
    if each \( x_i \) contains only one word
        then return X as an output
        stop.
    else begin
        j = 1;
        word_check (j, X);
    end;
end.
Figure 5.7: Example of selection steps

Figure 5.7 illustrates an example of the selection step process. In this illustration an eight (input) words sentence is examined with words A, B, C, ...H. Some words correspond to more than one TL word e.g., A corresponds to three TL words, A_1, A_2 and A_3. Some words correspond to one TL word e.g., B, C and F correspond to B_1, C_1 and F_1 respectively. Based on our selection step and the WordAsso relation template, the appropriate meaning of each word is selected in ten steps as follows:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
1. B1 is used to determine the meanings of A and C which are adjacent to B on the left and right. A has three different meanings, A1, A2 and A3 while C has one meaning. Therefore, we select only the meaning of the first word by considering the WordAsso relation between B1 and A1, B1 and A2, and B1 and A3 respectively. Suppose A2 and A3 have meaningful relations with B1, then keep both of them. Next, we use C1 to determine the meanings of B and D in a similar manner. In the case that there is no relationship between C1 and the corresponding words of D (D1, D2 and D3), the three meanings (D1, D2 and D3) remains the same. F1 is used to determine the meaning of E and G, in the case that F1 has a relationship with E4; E4 is selected.

2. E4 is used to determine its adjacent words (D and F). There is no relationship between E4 and D, therefore the string remains the same.

3. B1 is used to determine the meaning of D (one word apart); D1 and D2 are selected. C1 is used to determine the meanings of A and E (which corresponds to E4 alone at this step), there is no relationship between C and A. G2 is selected according to its relationship with E4.

4. G2 is used to determine its adjacent words (F and H). G2 has a relationship with H3 and H4. then H3 and H4 are selected.

5. B1, C1, E4, F1 and G2 are used to determine the words which are one word apart from each of them. There is no relationship between any of them, the string remains unchanged.

6. E4 is used to determine B and H (two words apart form E4), B already has one meaning and there is no relationship between E4 and H, the string remains the same.

7. A2 is selected for A since A2 is the first meaning appearing at this step according to the SL-TL dictionary.

8. A2 is used to determine the meanings of D, D2 is selected.

9. H3 is selected for H since H3 is the first meaning (most common of those remaining) appearing at this step according to the SL-TL dictionary.

10. A2, B1, C1, D2, E4, F1, G2 and H3 are selected.

The first column of Table 5.3 illustrates the constraint application result of Example 6. The dictionary look-up result is shown in the second column. Each word spirit, willing, but and flesh has more than one meaning (the descriptions of each meaning are shown in Table 5.2). To select the appropriate words for spirit, willing, but and flesh, ALMT first considers the relationship between the and its adjacent words since the has a unique meaning. However, there is no relationship between any of them. ALMT then considers the relationship between weak and its adjacent word, flesh. The word राक्षस (राक्षस) is selected for flesh (the fourth column) according to the WordAsso relation, 13-21-1 and 1-1-2-1-4-2, see Figure 5.8. The first meaning left on the list of the words spirit, willing and but; राक्षस
(winjaan), พระ (phra'am) และ เท่า (te'e) are selected since there is no relationship between spirit, willing, but and their proximity words according to our WordAsso relation template. The selected words for example 6 are shown in the fourth column of Table 5.3.

To associate the input words so, ugly and left in Example 7 (Table 5.4) with the most appropriate words in Thai, ALMT considers the relationship between those words with words

Table 5.3 The possible corresponding Thai words for spirit, willing, flesh and weak

Example 6: The spirit is willing but the flesh is weak.

<table>
<thead>
<tr>
<th>SL Constraints output</th>
<th>Dictionary look-up output</th>
<th>WordAsso number</th>
<th>Word Selection Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>The spirit</td>
<td>นั่น-นัน</td>
<td>a20</td>
<td>นั่น-นัน</td>
</tr>
<tr>
<td></td>
<td>วิชชา- winjaan</td>
<td>1-2-1-1-1</td>
<td>วิชชา- winjaan</td>
</tr>
<tr>
<td></td>
<td>จตุใจ- cidcaj</td>
<td>2-5-6-1</td>
<td>จตุใจ- cidcaj</td>
</tr>
<tr>
<td></td>
<td>ชาว- khwa nun</td>
<td>2-5-6-2</td>
<td>ชาว- khwa nun</td>
</tr>
<tr>
<td></td>
<td>ความยุ่น- khwaamnum'ηmān</td>
<td>2-5-1-1</td>
<td>ความยุ่น- khwaamnum'ηmān</td>
</tr>
<tr>
<td></td>
<td>บรรยายภาษา- banjaakaad</td>
<td>5-6</td>
<td>บรรยายภาษา- banjaakaad</td>
</tr>
<tr>
<td></td>
<td>เทิ้ง- law</td>
<td>1-1-2-1-2-1-3-1</td>
<td>เทิ้ง- law</td>
</tr>
<tr>
<td></td>
<td>น่าไปอย่างดีที่มัน-</td>
<td>2-24-2</td>
<td>น่าไปอย่างดีที่มัน-</td>
</tr>
<tr>
<td></td>
<td>nampaijanphony'glāb</td>
<td></td>
<td>nampaijanphony'glāb</td>
</tr>
<tr>
<td>willing</td>
<td>พระ- phra'am</td>
<td>13-16-3</td>
<td>พระ- phra'am</td>
</tr>
<tr>
<td></td>
<td>เติมใจ- temcaj</td>
<td>13-30</td>
<td>เติมใจ- temcaj</td>
</tr>
<tr>
<td>but</td>
<td>แต่- te'e</td>
<td>r10</td>
<td>แต่- te'e</td>
</tr>
<tr>
<td></td>
<td>หลากหลาย- no'egcāag</td>
<td>r11</td>
<td>หลากหลาย- no'egcāag</td>
</tr>
<tr>
<td>the</td>
<td>นั่น-นัน</td>
<td>a20</td>
<td>นั่น-นัน</td>
</tr>
<tr>
<td>flesh</td>
<td>เนื่อง- ny'a</td>
<td>1-4-1-3-1</td>
<td>เนื่อง- ny'a</td>
</tr>
<tr>
<td></td>
<td>ข่างกาย- rānkaaj</td>
<td>1-1-2-1-4-2</td>
<td>ข่างกาย- rānkaaj</td>
</tr>
<tr>
<td>weak</td>
<td>ต่อมา- ?a'ah?e'</td>
<td>13-21-1,</td>
<td>ต่อมา- ?a'ah?e'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13-10-3</td>
<td></td>
</tr>
</tbody>
</table>

wordasso_rel('1-1-2-1-2-1-2-1', ["2-1-1-6","2-3-1","2-1-25"]).
wordasso_rel('1-1-2-1-5-1', ["2-13"]).
wordasso_rel('13-21-1', ["1-1-2-1-4-2"]).
wordasso_rel('1-1-2-1-1-2', ["2-2-8"]).
wordasso_rel('1-2-1-2-1', ["2-8-3","2-9-1"]).

Figure 5.8: Examples of WordAsso relations
in their proximity. There is a relationship between the words duckling (ลูกนก (ลูงพ็ด), 1-1-1-2-1-2-1) and left as indicated in the WordAsso relation template (Figure 5.8). The WordAsso number 1-1-1-2-1-2-1 can occur with any word with WordAsso number 2-1-1-6 or

Table 5.4. The possible corresponding Thai words for left, duckling and marsh.

Example 7: So the ugly duckling left the marsh.

<table>
<thead>
<tr>
<th>SL Constraints output</th>
<th>Word Treatment</th>
<th>Description</th>
<th>WordAsso number</th>
<th>Selected word</th>
</tr>
</thead>
<tbody>
<tr>
<td>So</td>
<td>ตั้งนัง- dopnan</td>
<td>with the result that</td>
<td>รฉ0</td>
<td>ตั้งนัง- dopnan</td>
</tr>
<tr>
<td>the ugly</td>
<td>นัน- nam</td>
<td>determiner</td>
<td>อฉ4</td>
<td>นัน- nam</td>
</tr>
<tr>
<td>duckling</td>
<td>ซูคี พี- ลูงพ็ด</td>
<td>A young duck.</td>
<td>1-1-1-2-1-2-1</td>
<td>ซูคี พี- ลูงพ็ด</td>
</tr>
<tr>
<td>left</td>
<td>ชวย- saaj</td>
<td>The left side or direction</td>
<td>2-35-18</td>
<td>ชวย- saaj</td>
</tr>
<tr>
<td></td>
<td>จเว่- cag</td>
<td>To stop being in or with.</td>
<td>2-1-1-6</td>
<td>จเว่- cag</td>
</tr>
<tr>
<td></td>
<td>หัว- thin</td>
<td>To go without taking.</td>
<td>2-19</td>
<td>หัว- thin</td>
</tr>
<tr>
<td></td>
<td>ออกกยน- agcag</td>
<td>To go away (from)</td>
<td>2-13</td>
<td>ออกกยน- agcag</td>
</tr>
<tr>
<td>the</td>
<td>นัน- nam</td>
<td>determiner</td>
<td>อฉ4</td>
<td>นัน- nam</td>
</tr>
<tr>
<td>marsh</td>
<td>หนะวะ- naaen</td>
<td>low land that is soft and wet.</td>
<td>1-1-2-1-5-1</td>
<td>หนะวะ- naaen</td>
</tr>
</tbody>
</table>

2-13. Therefore, the words จเว่ (cag, 2-1-1-6) and ออกกยน (agcag, 2-13) are selected.

Next, to select the most appropriate word from these two possibilities; จเว่ (cag) and ออกกยน (agcag), the relationship between the words marsh and left is considered. There is no relationship between ugly and left, or the and left according to the WordAsso relationship template. Therefore, marsh which is one word apart from left and has only one meaning, หนะวะ (naaen), with the WordAsso 1-1-2-1-5-1 is considered. According to the WordAsso

---

3, 4, 5 The meanings of left which is a past tense and a participle of leave [Longman 1992].
WordAsso number 2-13. Then, the word แอนเจล (ā'agcāag) is selected for left.

The word นิ้วมือ (khiilee) is selected for ugly according to the relationship between ugly and duckling as indicated in the WordAsso relationship template (Figure 5.8). There is no relationship between the word สิ่ง (danān), therefore the first meaning on the list, ตัวนี้ (daññ), is selected. The selected words for example 7 are shown in the fifth column of Table 5.4.

5.3 Word Ordering

The syntactic level differences between SL and TL are reconsidered at this point to complete the TC. Some words which are necessary in the TL, not only to retain the meaning of the SL but also to make the TL grammatically correct, are added to the string before the ordering can be performed. For example, the words ฉัน (le'ew) and le6 are required to indicate the past tense in Thai and Chinese respectively (see section 4.1). After all necessary words are added, ALMT rearranges the selected words without performing any analysis in a TL grammatical order according to the ordering rule. Therefore, two steps are performed by word ordering module: word addition and word ordering (Figure 5.1).

5.3.1 Word Addition

The morphological and syntactic characteristics of the SL which are removed in the SL constraints application step will be replaced with the appropriate corresponding words in the TL in the word addition step. In the case that English is the SL and Thai is the TL, the features plural, ing, and passive which are used to indicate the plurality, continuous tense and passive voice features of the SL are replaced with the words แนว (lav aj), ทราบ (kamlan) and อน (thùug) respectively (see section 5.1.1).

In case of pluralities, classifiers are required (see section 4.1) to express the quantity. Each noun relates to a specific classifier; therefore, the classifier relation was designed in the form of WordAsso numbers to be used to select the appropriate classifier for each noun.

---

*Phonetic transcription of Chinese.*
In Thai, a noun and its classifier were stipulated by the Thai Royal Institute. To date, there are approximately 3000 different classifiers [The Thai Royal Institute 1995]. From our studies, we classified the classifiers into classes as shown in Table 5.5. With this classification together with the specification of the Thai Royal Institute, we then developed the relation between nouns and their classifiers, some examples are illustrated in Figure 5.9. The noun with the WordAsso number in the first argument is compatible with a classifier with a WordAsso number shown in the second argument.

Table 5.5 A small fraction of classifier classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Sample Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4-2 Classifier</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1 Classifier of object</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-1 Classifier of living thing</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-1-1 Classifier of human</td>
<td>คุณ</td>
</tr>
<tr>
<td>2-4-2-1-1-2 Classifier of animal</td>
<td>ตัว</td>
</tr>
<tr>
<td>2-4-2-1-1-3 Classifier of plant</td>
<td>ต้น</td>
</tr>
<tr>
<td>2-4-2-1-2 Classifier of non-living thing</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-2-1 Classifier of tool</td>
<td>ผู้</td>
</tr>
<tr>
<td>2-4-2-1-2-2 Classifier of room</td>
<td>ห้อง</td>
</tr>
<tr>
<td>2-4-2-1-2-3 Classifier of external body part</td>
<td>ข้าว ผัก มือ</td>
</tr>
<tr>
<td>2-4-2-1-2-5 Classifier of root vegie</td>
<td>หัว</td>
</tr>
<tr>
<td>2-4-2-2 Classifier of collection</td>
<td></td>
</tr>
<tr>
<td>2-4-2-2-1 Classifier of human</td>
<td>กลุ่ม</td>
</tr>
<tr>
<td>2-4-2-2-2 Classifier of animal</td>
<td>ผู้</td>
</tr>
<tr>
<td>2-4-2-3 Classifier of time period</td>
<td>วัน วัน</td>
</tr>
<tr>
<td>2-4-2-5 Classifier of frequency</td>
<td>ครั้ง</td>
</tr>
</tbody>
</table>

Figure 5.9. Examples of classifier relations.

Example 8 illustrates how to select the appropriate classifiers for the words woman, cat and hen. The selected words in Thai corresponding to the words woman, cat and hen in Example 8 are shown in the second column of Table 5.6. The indefinite determiners a and
an in this expression, corresponding to the word ญ้ (ny'n) in Thai, indicate the need for classifiers for the words นิ ญจิ (phuuji'n), แมว (meew) and ไก่ (kajj) respectively. The word นิ ญจิ (phuuji'n) belongs to the class Female (1-1-1-1-2), a subclass of Human (1-1-1-1). A noun which belongs to the class 1-1-1-1 is compatible with a classifier with the WordAsso number 2-4-2-1-1-1 based on the classifier relation illustrated in Figure 5.9. Therefore, the classifier หน (khon) with 2-4-2-1-1-1 in Table 5.5 is selected for the word นิ ญจิ (phuuji'n-woman).

Table 5.6. The selected words in Thai for the words woman, cat and hen

<table>
<thead>
<tr>
<th>English</th>
<th>Selected Word in Thai</th>
<th>WordAsso</th>
<th>Class</th>
<th>Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>woman</td>
<td>นิ ญจิ (phuuji'n)</td>
<td>1-1-1-1-2</td>
<td>Female</td>
<td>หน (khon)</td>
</tr>
<tr>
<td>cottage</td>
<td>กระท่ำ (krath'd'm)</td>
<td>1-1-2-1-2</td>
<td>Housing</td>
<td>ห (la'n)</td>
</tr>
<tr>
<td>cat</td>
<td>แมว (meew)</td>
<td>1-1-2-1-1</td>
<td>Mammal</td>
<td>ต (tua)</td>
</tr>
<tr>
<td>hen</td>
<td>ไก่ (kaj)</td>
<td>1-1-1-2-2</td>
<td>Fowl</td>
<td>ต (tua)</td>
</tr>
</tbody>
</table>

The words แมว (meew-cat) and ไก่ (kaj-hen) belong to the classes mammal (1-1-1-2-1-1) and fowl (1-1-1-2-1-2-2) respectively. Both classes are subclasses of animal (1-1-1-2). Since a noun with the WordAsso number 1-1-1-2 relates to a classifier with 2-4-2-1-1-2 according to the classifier relation shown in Figure 5.9, the classifier ต (tua) with 2-4-2-1-1-2 is selected for the words แมว (meew-cat) and ไก่ (kaj-hen).

The definite determiner the corresponding to the word น (nán) in Thai indicates the need for a classifier for the word กระท่ำ (krath'd'm). The word กระท่ำ (krath'd'm) belongs to the class Housing (1-1-2-1-2-2) which is compatible with a classifier with the WordAsso number 2-4-2-1-2-10 based on the classifier relation illustrated in Figure 5.9. Therefore, the classifier ห (la'n) with the WordAsso 2-4-2-1-2-10 is selected for the word กระท่ำ (krath'd'm).

5.3.2 Word Ordering

In some MT systems, e.g., Shake-and-Bake MT, the ordering process has been performed by generating valid combinations of words and analyzing (parsing) to see whether the combination is grammatically correct. If the combination is correct then it will be selected as an output; if not, another combination is tried until the grammatically correct sentence is
The disadvantage is that the generator has to consider a number of possibilities which is factorial in the number of signs in the TL sentence. In GRMT, the selected words are ordered according to rules without performing any analysis. These ordering rules account for syntactic level differences between languages. The structure of the SL which is similar to that of the TL remains the same, only the sentence fragments that are different will be rearranged into the grammatical order of the TL.

The structure of the English sentence which contains subject, verb and object in that order is similar to that of the Thai sentence. Therefore, when English is translated into Thai, its structure remains the same as illustrated in Example 9 (Figure 5.10). However, in the case that the English phrase contains attributes, possessive pronouns or determiners, the word order must be rearranged to make translation of the phrase grammatically correct since the structure of such a phrase is different between English and Thai (see section 4.1). The first row of Figure 5.11 illustrates the corresponding words in Thai of Example 10 selected by ALMT. The classifier ตัว (tua) is added in the word addition step. The translations of the noun phrase *The ugly duckling, his head and his wing* are formed in reverse order of the order in English according to the ordering rules (the third row of Figure 5.11).

Some other structures which are different between English and Thai are examined as well. For example, the negative of *can* and *could* (cannot, could not); the adverb *very* which modifies an adjective (*very good, very warm*) and their translations are formed in the reverse order of the order in English.

Example 9: John eats rice every day.

<table>
<thead>
<tr>
<th>Selected words</th>
</tr>
</thead>
<tbody>
<tr>
<td>John (caan) eats (kin) rice (khaw) every (thug) day (wan)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word ordering output</th>
</tr>
</thead>
<tbody>
<tr>
<td>John (caan) eats (kin) rice (khaw) every (thug) day (wan)</td>
</tr>
</tbody>
</table>

Figure 5.10: Word orders of จอมัน กิน ข้าว ทุก วัน
Example 10: The ugly duckling hid his head under his wing.

duckling ugly clas hid head under wing his
(nán) (khiiree) (luugped) (dajsoa'an) (khō'ankan) (hu'a) (tāaj) (spā'akhan) (piig)

Selected words

The ugly duckling hid his head under his wing
(nán) (khiiree) (luugped) (dajsoa'an) (khō'ankan) (hu'a) (tāaj) (spā'akhan) (piig)

Word addition output

duckling ugly clas The hid head his under wing his
(luugped) (khiiree) (tua) (nān) (dajsoa'an) (hu'a) (tāaj) (piig) (spā'akan)

Word ordering output

duckling ugly clas The hid head his under wing his
(luugped) (khiiree) (tua) (nān) (dajsoa'an) (hu'a) (tāaj) (piig) (spā'akan)

Figure 5.11: Word orders of ตัว ของ ได้ นิ้ว ไก่ หัว ใต้ บิด ของ เช้า.

5.4 Example of ALMT

In this section we present the results of applying ALMT to some example sentences. The SL of Example 11 is shown in the first column of Table 5.7. The second column illustrates the output after applying the constraints e.g., the inflection -s triggers the “plural constraint”, are used triggers the “passive voice” constraint, do not triggers the “negative” constraint and are talking triggers “present continuous” constraint as discussed in section 5.1.1. Each word in the second column is used as a keyword to search for the corresponding words in Thai. All possible meanings of each SL word are shown in the third column. Some of the words have more than one meaning e.g., symbol, when, know, etc. The appropriate meaning of use, know and about can be selected by considering the semantic relationship between words and the choice for each of these is shown in the fourth column. However, the appropriate words for symbol and when cannot be selected in the same manner because all possible meanings of each word have the same WordAsso number. Therefore, the first meaning appearing on the list of meaning of each word is selected. All selected words are shown in the fourth column. Before performing the ordering step, the features plural, passive and ing are replaced with the appropriate words; หลาย (la'aj), เป็น (thuug) and กำลัง (kamlan) respectively (the fifth column). The classifier ตัว (tua) is added to express the
Table 5.7: ALMT steps applied to the sentence of Example 11

Example 11: Algebraic symbols are used when you do not know what you are talking about.

<table>
<thead>
<tr>
<th>English Constraints output</th>
<th>Dictionary look-up output</th>
<th>Selected words in Thai</th>
<th>Word Addition</th>
<th>Word ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebraic plural symbols</td>
<td>ทางพีชคณิต (thaanphichakanid)</td>
<td>ทางพีชคณิต (thaanphichakanid)</td>
<td>ลัญยุติ (sanjalag)</td>
<td>ลัญยุติ (sanjalag)</td>
</tr>
<tr>
<td>used you do not know what</td>
<td>ใช้ (chaj),</td>
<td>ใช้ (chaj)</td>
<td>ใช้ (chaj)</td>
<td>ใช้ (chaj)</td>
</tr>
<tr>
<td>when know</td>
<td>เมื่อ (my'a),</td>
<td>เมื่อ (my'a)</td>
<td>เมื่อ (my'a)</td>
<td>เมื่อ (my'a)</td>
</tr>
<tr>
<td>what you are talking</td>
<td>อะไร (araj)</td>
<td>อะไร (araj)</td>
<td>อะไร (araj)</td>
<td>อะไร (araj)</td>
</tr>
<tr>
<td>are talking</td>
<td>คุณ (khanahiti)</td>
<td>คุณ (khanahiti)</td>
<td>คุณ (khanahiti)</td>
<td>คุณ (khanahiti)</td>
</tr>
<tr>
<td>about</td>
<td>ประมาณ (pramaan),</td>
<td>ประมาณ (pramaan)</td>
<td>ประมาณ (pramaan)</td>
<td>ประมาณ (pramaan)</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
quantity of symbol. The word ว่า (wâa) is added to combine clauses. The word ว่า (wâa) is a translation of the word that which is omitted in this sentence, however, it cannot be omitted in Thai otherwise the translation in Thai will be grammatically incorrect. The selected words are put into the correct order according to the Thai grammar as shown in the last column (The shaded area). In Example 10, the generated TC formed the Correct translation (CT) for the input sentence without the necessity of performing any correction.

In generating the TC for Example 12, once all words are selected (the fourth column, Table 5.8), the word ได้ (dâj) is added to clarify the past tense (lived). The classifiers ณ (khon), หลัง (la'η) and ตัว (tua) are also added as we discussed in section 5.3.1. The selected words in each noun phrase; an old woman, the cottage, a fat black cat and a plump brown hen are rearranged into the Thai grammatical order. The generated TC is exactly the same as the CT.

Figure 5.12 illustrates the translation of a fraction of The ugly duckling (version written by Ronne Randall, 1987) by ALMT. In the fourth sentence, ALMT selected the first translation word in the list, ที่ไหน (thiina'j), for the word where since there is no significant relationship between where and its proximity words. This selection is inappropriate for this expression and it will be repaired in the later phases, TCE and RI. The word ได้ (dâj) was added to clarify the past tense (lived). It retains the meaning of the original sentence and is grammatically correct, however, ได้ (dâj) is usually omitted in such expression in Thai. In the eighth and fifteenth sentences, the translations of the phrases You are certainly not one and One evening (the shaded areas) are misordered and will be rearranged in the TCE and RI phrases. The word กับ (kab) is an inappropriate translation for the word to in the fifteenth sentence since it means “for the attention or possession of/ intended to be given to”. This selection will be replaced with the more appropriate word ถึง (thy'η) which is a translation of the word to in the sense of “reaching as far as” in the TCE and RI phrases.

Example 12: An old woman lived in the cottage, with a fat black cat and a plump brown hen.

TC: ผู้หญิง แก่ คนหญิง ได้ อยู่ ใน กระท่อม หลัง นั้น กับ แนว สีดำ ยืน ตัว หนึ่ง และ ได้ สืบคามา อยู่ ตัว หนึ่ง
CT: ผู้หญิง แก่ คนหญิง ได้ อยู่ ใน กระท่อม หลัง นั้น กับ แนว สีดำ ยืน ตัว หนึ่ง และ ได้ สืบคามา อยู่ ตัว หนึ่ง

Table 5.8: ALMT steps applied to the sentence of Example 12

<table>
<thead>
<tr>
<th>English</th>
<th>Constraint Application output</th>
<th>Dictionary look-up output</th>
<th>Selected Word in Thai</th>
<th>Word Addition</th>
<th>Word ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td>An</td>
<td>An</td>
<td>หนึ่ง (ny’η)</td>
<td>หนึ่ง (ny’η)</td>
<td>หนึ่ง (ny’η)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>old woman</td>
<td>old woman</td>
<td>ผู้ (ke’η)</td>
<td>ผู้ (ke’η)</td>
<td>ผู้ (ke’η)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>lived</td>
<td>lived</td>
<td>หญิง (phuuji’η)</td>
<td>หญิง (phuuji’η)</td>
<td>หญิง (phuuji’η)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>in</td>
<td>in</td>
<td>ใน (naj)</td>
<td>ใน (naj)</td>
<td>ใน (naj)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>the</td>
<td>the</td>
<td>หนึ่ง- น้ำ</td>
<td>หนึ่ง- น้ำ</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>cottage</td>
<td>cottage</td>
<td>กระท่อม (krathə’m)</td>
<td>กระท่อม (krathə’m)</td>
<td>กระท่อม (krathə’m)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>with</td>
<td>with</td>
<td>หญิง (sy’η)</td>
<td>หญิง (sy’η)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>หนึ่ง (ny’η)</td>
<td>หนึ่ง (ny’η)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>fat</td>
<td>fat</td>
<td>อ้วน (?uan)</td>
<td>อ้วน (?uan)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>black cat</td>
<td>black cat</td>
<td>สีดำ (sir’dam)</td>
<td>สีดำ (sir’dam)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>and</td>
<td>and</td>
<td>และ (le’η)</td>
<td>และ (le’η)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>หนึ่ง (ny’η)</td>
<td>หนึ่ง (ny’η)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>plump brown</td>
<td>plump brown</td>
<td>อ้วน (?uab)</td>
<td>อ้วน (?uab)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
<tr>
<td>hen</td>
<td>hen</td>
<td>ไก่ (kaj)</td>
<td>ไก่ (kaj)</td>
<td>คำ (khon)</td>
<td>คำ (khon)</td>
</tr>
</tbody>
</table>
The ugly duckling. Written by Ronne Randall.

1. The ugly duckling hid in a corner of the farmyard and cried. 2. The next morning, the ugly duckling decided to run away. 3. "I will find a place where no one will make fun of me," he said to himself. 4. He left the farmyard and ran to the great marsh, where the wild ducks lived. 5. "Perhaps they will be kind to me," he said sadly. 6. But the wild ducks were not kind. 7. "You are very ugly," they said. 8. "You are certainly not one of us." 9. And the wild ducks laughed at the duckling. 10. So the ugly duckling left the marsh. 11. The wind was blowing, and it was cold. 12. Winter was coming. 13. The ugly duckling was very tired. 14. He had to find a place to stay where he could be warm and safe. 15. One evening, he came to a little cottage. 16. The door was open and he went inside. 17. An old woman lived in the cottage, with a fat black cat and a plump brown hen....

Figure 5.12: A part of “The ugly duckling” and its translation

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Chapter 6

Translation Candidate Evaluation

The second phase of GRMT, Translation Candidate Evaluation, determines whether the generated translation candidate retains the meaning conveyed by the source language. TCE analyzes both the SL and the TC in parallel, then compares the results of the parses in terms of semantics alone since there are syntactic level differences between languages (Figure 6.1). If the semantic information in each of the SL and the TC parses are acceptably similar, the TC will be deemed an appropriate translation. If their semantic information is different, the TC and their semantic differences will be reconciled in the third phase, Repair and Iterate. TCE comprises two modules as illustrated in Figure 6.1, the analyzer and semantic comparison.

Figure 6.1: TCE Architecture
6.1 The Analyzer

The analysis module analyzes the TC to examine its syntax and semantics: whether the TC is grammatically correct according to the TL grammar and whether the TC retains the meaning of the source language. Therefore, two steps are performed by the analyzer: Parsing and Semantic Extraction.

6.1.1 Parsing

 Parsing is applied to both the SL and the TC. We have developed English and Thai grammars within the Head-Driven Phrase Structure Grammar formalism [Pollard and Sag; 1987; Pollard and Sag, 1994]. Our grammar is a modification to the grammars which were provided by Gerald Penn [Penn 1993] and Colin Matheson [Matheson 1996]. The grammars we developed have been implemented using the Attribute Logic Engine version 3.2 Beta. ALE is an integrated phrase structure parsing and definite clause logic programming system in which the terms are typed feature structures [Carpenter and Penn 1999].

In the next section, a very brief overview of HPSG and some notation are introduced. Further details on HPSG can be found in [Pollard and Sag 1987], [Pollard and Sag 1994], [Sag and Wasow 1999], HPSG homepage at http://hpsg.stanford.edu/ [Center for the Study of Language and Information 1999] and the typed feature formalism is described in [Carpenter 1992].

6.1.1.1 About Head-Driven Phrase Structure Grammar

Head-Driven Phrase Structure Grammar, the grammatical theory developed by Carl Pollard and Ivan Sag during the mid 1980s [Pollard and Sag 1987; Pollard and Sag 1994], was developed by synthesizing several contemporary linguistic theories, including Categorial Grammar (CG), Generalized Phrase Structure Grammar (GPSG) [Gazdar et al. 1985], and Lexical Functional Grammar (LFG) [Bresnan and Kaplan 1982], from which it borrows some interesting ideas. The HPSG architecture has been largely pursued by an increasing number of linguists, since the formally well-defined framework allows for an explicit formalization of a linguistic theory.

HPSG, an integrated theory of natural language syntax and semantics, is a feature-based grammatical framework which is characterized by a modular specification of linguistic
generalizations through extensive use of principles and lexicalization of grammatical information.

HPSG is formulated in terms of order-independent constraints. In other words, the grammar is formulated as a declarative system of constraints. These constraints provide partial grammatical information that can be flexible, consulted in a variety of language processing models based on the notion of incremental, on-line integration of heterogeneous types of information. Key advantages of HPSG which are appropriate for use in MT are as follows:

- **Small number of rules and rich-information lexicon**
  Based on the assumption of a universal syntax, a small number of highly schematic syntactic rules are assumed to apply universally; the task of explaining the variations between languages must be carried out in the lexicon [Carpenter 1991]. The detailed lexical entries of HPSG are concisely expressed within a multiple inheritance hierarchy and lexical rules. Such hierarchical lexicons allow cross-cutting generalizations about words to be expressed in a highly efficient and compact organization.

- **Unification-based constraints**
  Recent work in the unification framework [Shieber 1986] has shown that a formalism based on unification is particularly well-suited for declarative modes of problem solving. In addition, a declarative mode allows the construction of grammars which can be neutral as to the processes applying to them, whether parser or generator [Estival 1994].

- **Language-independent principles**
  The modular design of HPSG offers a large degree of flexibility for applying the framework to new languages and changing individual components of a grammar.

### 6.1.1.2 Words and Phrases as Feature Structures

Utterances in HPSG are modeled in terms of feature structures of type *sign*, with its two immediate subtypes *word* and *phrase* [Ginzburg and Sag 1998]. The lexical entries are descriptions of feature structures of type *word* while phrase structure rules are partial descriptions of feature structures of type *phrase*.

A feature structure is a description of an object; it specifies some or all of the information that is asserted to be true of the object. The features CATEGORY (CAT) and CONTENT (CONT) (Figure 6.2) specify the syntactic and semantic information of an object
respectively. Figure 6.2 represents a partial description of the lexical entry *hide* in an attribute-value matrix (AVM) diagram¹ (the detailed explanation of this representation can be found in section 6.3.1). Each feature takes a value of a particular and appropriate type. The grammar must include a specification of what types are included, which features are appropriate for which type, and what type of value is appropriate for each feature. Some types, their features and their value types are shown in Table 6.1 (see also appendix D).

![Figure 6.2: The lexical entry hide](image)

Figure 6.3 presents a (simplified) partial description of the phrase *Leslie drinks milk* (this example is taken from Ginzburg and Sag, 1998) in terms of typed feature structures. This representation indicates that *Leslie drinks milk* is the type of head-subject phrase² (hd-subj-ph). The hd-subj-ph is a subtype of the type headed-phrase (hd-ph). The hd-ph is the type of

---

¹ The AVM diagram is the standard method of representing grammatical information in modern computation grammar theories.

² There are two types of phrases in English: headed phrase (e.g., head-subject phrase, head-complement-phrase, head-modifier-phrase) and nonheaded phrase (e.g., imperative-phrase, coordinate-phrase). Further details can be found in [Ginzburg and Sag 1998] and [Sag and Wasow 1999].
the phrase in which one daughter is assigned special status as the HEAD DAUGHTER (HD-DTR). Other subtypes of hd-ph are head-specifier-phrase (hd-spr-ph), head-complement-phrase (hd-comp-ph), head-adjunct-phrase (hd-adjunct-ph). Instances of the hd-ph are governed by the feature declarations shown in the last row of Table 6.1. The indices \( 0, 0, 0 \) specified in the feature PHONOLOGY (PHON) called tags indicate the structure sharing between feature values: two or more different features within the feature structure may have their values specified by one and the same feature structure. The effect of structure sharing is to force two (or more) distinct nodes in a tree admitted by a rule to have identical values for a given feature. In Figure 6.3, the first phonology (phonology here means phonological shape) of the hd-subj-ph is the same as the phonology of its NON-HD-DTRS which, in this case, is Leslie, the second and the third phonologies are the same as those of its HD-DTR which are drinks and milk respectively. The syntactic category of Leslie drinks milk is a sentence (synsem: s). The syntactic category of the NON-HD-DTRS (Leslie) of the given phrase is a noun phrase (NON-HD-DTRS:SYNSEM: np). The syntactic category of its HD-DTR (drinks milk) is a verb phrase (HD-DTR:SYNSEM: vp). The type of this verb phrase drinks milk is a hd-comp-ph with the HD-DTR, drinks, and the NON-HD-DTRS, milk.

Table 6.1: Some types, their features and their value types

<table>
<thead>
<tr>
<th>Type</th>
<th>Features/Type of value</th>
<th>Immediate Supertype</th>
</tr>
</thead>
<tbody>
<tr>
<td>sign</td>
<td>[PHON list(speech-sound)</td>
<td>SYNSEM synsem]</td>
</tr>
<tr>
<td>phrase</td>
<td>...</td>
<td>sign</td>
</tr>
<tr>
<td>word</td>
<td>...</td>
<td>sign</td>
</tr>
<tr>
<td>synsem</td>
<td>[LOC loc]</td>
<td>mod_synsem</td>
</tr>
<tr>
<td>loc</td>
<td>[CAT cat</td>
<td>CONT cont</td>
</tr>
<tr>
<td>cat</td>
<td>[HEAD head</td>
<td>SPR list_synsem</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>hd-ph</td>
<td>[HD-DTR sign</td>
<td>NON-HD-DTRS list(sign)]</td>
</tr>
</tbody>
</table>
6.1.1.3 Lexicon

The lexical types (e.g., noun, verb, adj), a type inheritance and the lexical rules allow complex lexical information (as shown in Figure 6.2) to be derived via the logic of the lexicon, rather than simply stipulated.

The lexical types were introduced to define feature appropriateness, to avoid having to specify values for features that are irrelevant to particular classes. Table 6.2 illustrates the type constraints that state general properties (in terms of particular feature-value specifications) of particular lexemic types, noun and verb. Noun and verb are subtypes of substantive (subst). The feature CASE is appropriate only for nouns (in English) and its value is either nominative (nom) or accusative (acc). The features VERB-FORM (VFORM), INVERTED (INV) and AUXILIARY (AUX) are specifiable only for verbs. The forms of verbs can be classified as base, finite, gerund, infinite, passive participle, present participle and past participle form [Pollard and Sag 1987]. The feature INV and AUX are used to distinguish auxiliary (helping) verbs from all others. The values of INV and AUX are the type of boolean. In interrogatives, a finite auxiliary verb precedes the subject, therefore, the INV and AUX values of a finite auxiliary verb are plus while the INV and AUX values of other kinds of verb

---

3 According to Pollard and Sag, noun, verb, adjective, preposition, relativizer are subtypes of substantive, whereas marker (e.g., complementizers) and determiner are subtypes of functional [Pollard and Sag 1994].
are minus.

Each type in the lexical hierarchy has constraints associated with it - some inviolable, and others that are default in nature. The inheritance of constraints in this type hierarchy is default. The *default inheritance* allows contradictory information associated with a subtype to take precedence over (or override) constraints that would otherwise be inherited from a supertype. In other words, lexical items have many properties in common but may differ from one another in terms of particular constraints that override the general constraints governing their supertypes. By organizing the lexicon in terms of a type hierarchy and the use of *default inheritance* of constraints, the stipulations associated with particular lexical entries can be minimized and the shared properties of different word classes can be expressed.

### Table 6.2: Features and their value types of the types noun and verb

<table>
<thead>
<tr>
<th>Type</th>
<th>Features/Type of value</th>
<th>Immediate Supertype</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun</td>
<td>[CASE case]</td>
<td>subst</td>
</tr>
<tr>
<td>verb</td>
<td>[VFORM vform]</td>
<td>subst</td>
</tr>
<tr>
<td></td>
<td>[INV boolean]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[AUX boolean]</td>
<td></td>
</tr>
</tbody>
</table>

The *lexical rule* is a mechanism for further reducing redundancy and stipulation in the lexicon by using information in one lexical entry as the basis for generating another lexical entry. The lexical rules are used for deriving predictably related lexical entries, e.g., inflected forms of verbs and nouns. A lexical rule applies to a lexical entry (of type *word*) and produces as output a new lexical entry whose (morphological) form, syntactic category and semantics are systematically related to the input [Sag and Wasow, 1999]. Figure 6.4 illustrates the (simplified) rule that applies to verb bases in English, giving their 3rd-singular

![Figure 6.4: The 3rd-singular verb lexical rule](image)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
verb form. This 3rd-singular verb lexical rule is taken from hpsg.pl written by Gerald Penn [Penn 1993]. The rule suffixes an -s or -ies (in case the verb ended with y), thus it generates _hides_ for _hide_ and _flies_ for _fly_. The rule says that for every verb whose form is uninflected (VFORM: bse), not an auxiliary verb (AUX: minus) and it takes a nominative as its subject (SUBJ: NP (nom)), there is a corresponding lexical entry for a 3rd-singular verb whose form is dictated by the 3rd-singular verb lexical rule: finite (VFORM: fin) and it takes a subject which is nominative and singular (SUBJ: NP (nom, sg)). Figure 6.5 illustrates _hides_, the output of the 3rd-singular verb lexical rule applied to the lexical entry _hide_ shown in Figure 6.2. See appendix E for additional lexical rules.

![Figure 6.5: Example output, _hides_, of the 3rd-singular verb lexical rule](image)

6.1.1.4 Some Universal Principles and Grammar rules

In this section, we explain some of the universal principles proposed in HPSG and consider some phrasal constructions with HPSG grammar rules.

The **Head Feature Principle** (HFP) can be formulated as a constraint on phrases of the type hd-ph. This principle restricts the sharing of the HEAD feature between a mother sign and its head daughter as shown in (1). HEAD always takes as its value a feature structure appropriated to each category as discussed in the last section (see the values of the HEAD...
The effect of the sharing restriction is to guarantee that headed phrases really are a projection of their head daughters. In Figure 6.6, the **HD-DTR** *drinks*, of the **hd-comp-ph**, *drinks milk*, passes its **HEAD** feature's value to the **HEAD** feature of its mother (marked by the dotted ellipse). The **HD-DTR** *drinks milk*, of the **hd-subj-ph** *Leslie drinks milk*, passes its **HEAD** feature's value to the **HEAD** feature of its mother. **HFP** also indicates that the category of verb phrases are verb because they have verbal heads and that of noun phrases are nominal because they have noun heads. This principle is applied to all grammar rules.

\[
(1) \quad \text{hd-ph} \Rightarrow \text{SYNSEM | LOC | CAT | HEAD 1} \\
\text{HD-DTR} \quad \text{SYNSEM | LOC | CAT | HEAD 1 1}
\]

*Figure 6.6: HFP applied to Leslie drinks milk*

The **Semantics Principle** (second version) [Pollard and Sag 1994] guarantees that the semantics of a mother sign constrained by the feature **CONT** are identified with that of the adjunct daughter (**ADJUNCT-DTR**) if the phrase is the type of head-adjunct-phrase (**hd-adjunct-ph**). However, if the phrase is not the type of **hd-adjunct-ph** then the semantics of a mother sign are identified with that of the **HD-DTR** as in (2). The features **CONT** is concerned
principally with linguistic information that bears directly on semantic interpretation. The CONT value of nominals (e.g., lexical nouns and their phrasal projections) is the feature structure of INDEX and restriction (RESTR).

(2) \[
\text{hd-ph} \implies \begin{cases} 
\text{SYNSEM} | \text{LOC} | \text{CONT} & \mathbb{1} \\
\text{HD-DTR} & \left[ \text{SYNSEM} | \text{LOC} | \text{CONT} & \mathbb{1} \right]
\end{cases}
\]

with the exception of the hd-adjunct-ph

\[
\text{hd-ph} \implies \begin{cases} 
\text{SYNSEM} | \text{LOC} | \text{CONT} & \mathbb{1} \\
\text{ADJUNCT-DTR} & \left[ \text{SYNSEM} | \text{LOC} | \text{CONT} & \mathbb{1} \right]
\end{cases}
\]

The type INDEX is classified into three subtypes: referential (ref), there and it [Pollard and Sag 1994]. The types there and it are used only for the expletive pronouns there and it while the type ref are used for nouns. INDEX introduces the three agreement features: PERSON (PER), NUMBER (NUM) and GENDER (GEN). The INDEX values of Leslie and milk are shown in Figure 6.7. According to the AVM diagram representation illustrated in Figure 6.7, the linguistic meaning of Leslie is that the object is referred to as Leslie, its quantity is one and its gender is masculine. The PERSON feature is unspecified for Leslie. Milk is a 3rd person, singular and neutral. The INDEX feature plays an important role in HPSG account of binding theory. For example: the indices for John and himself, in the sentence John shaved himself, should be shared [Matheson 1996].

The value of the feature RESTR is a set of parametrized states-of-affairs (psoas). States-of-affairs (soa, the term used in situation semantics, it is what philosophers call propositions) are possible ways the world might be construed [Pollard and Sag 1987]. Psoa is like a soa except that certain of its argument roles (its parameter) have not yet been anchored to determinate objects. The idea behind the feature RESTR is that it contains the basic semantic information on a referent [Matheson 1996], e.g., the RESTR value of the lexical entry milk contains the basic semantic information that the object concerned is milk. In addition, the values of the INSTANCE and INDEX features are unified (the restriction imposed by the common noun is associated with the marker in the discourse as discussed in Footnote 3) [Matheson 1996].

---

4The INDEX feature is used in the HPSG analysis of discourse structure. For example, in the sentences: "the man likes books, he often reads them", the two markers representing the man and books in the first sentence should be identified with the markers for he and them, in the second sentence, respectively [Matheson 1996].

89
Figure 6.7: Semantic principle applied to Leslie drinks milk

The CONT value of verb is the feature structure of predication specified: what kind of relation is involved and who (or what) is participating in the relation. The CONT value of the lexical entry drinks is illustrated in Figure 6.7, it corresponds to the condition that 5 drinks 4.

The effect of the semantic principle is illustrated in Figure 6.7 (marked by the shaded ellipse). The CONT value of the hd-comp-ph drinks milk is passed from its HD-DTR drinks (indicated by the tag 4) and the CONT value of drinks milk is passed to that of hd-subj-ph (indicated by the tag 2) since drinks milk is the HD-DTR of Leslie drinks milk. The semantic representation in Figure 6.7 means Leslie drinks milk depicts a proposition which satisfies the conditions that 5 is a masculine referred to as Leslie, 4 is milk, and Leslie drinks milk. This
principle is applied to all grammar rules. For more principles e.g., Spec Principle, Marking Principle, Nonlocal Feature Principle and Relative Uniqueness Principle, consult [Pollard and Sag 1987], [Pollard and Sag 1994], and [Matheson 1996].

The **Head complement rule** allows a lexical head to combine with the zero or more complements that it selects to form a phrase of type hd-comp-ph. This kind of phrase is subject to the constraint as in (3). The COMPLEMENTS (COMPS) value of the head must match a sequence of SYNSEM values of NON-HD-DTRS. The COMPS value is a list of feature structures. The list specifies a sequence of categories corresponding to the complements that the word combines with. For example: a verb *likes* (Figure 6.8), a HD-DTR, combines with its (only one) complement *John* in a hd-comp-ph. This means that the SYNSEM value of the NON-HD-DTRS of the phrase *likes John* is identified with $\mathbb{2}$.

$$
\text{(3)} \quad \text{hd-comp-ph} \Rightarrow \begin{bmatrix}
\text{HD-DTR} & \begin{bmatrix}
\text{word} & \begin{bmatrix}
\text{SYNSEM} & \text{LOC} & \text{CAT} & \text{COMPS} & \begin{bmatrix}1, 2, \ldots, n\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\text{NON-HD-DTRS} & \begin{bmatrix}
\text{SYNSEM} & \mathbb{1}, \mathbb{1}, \ldots, \mathbb{1}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{array}
$$

$$
\begin{array}{l}
\text{hd-comp-ph} \\
\text{PHON} & \begin{bmatrix}
\text{likes, John}
\end{bmatrix}
\end{array}
\begin{array}{l}
\text{SYNSEM} & \text{LOC} & \text{CAT} \\
\text{COMPS} & \begin{bmatrix}1\end{bmatrix}
\end{array}
\begin{array}{l}
\text{HD-DTR} \\
\text{PHON} & \begin{bmatrix}
\text{like}
\end{bmatrix}
\end{array}
\begin{array}{l}
\text{SUBJ} & \begin{bmatrix}1, \text{NP}\end{bmatrix}
\end{array}
\begin{array}{l}
\text{COMPS} & \begin{bmatrix}2\end{bmatrix}
\end{array}
\begin{array}{l}
\text{NON-HD-DTRS} \\
\text{PHON} & \begin{bmatrix}
\text{John}
\end{bmatrix}
\end{array}
\begin{array}{l}
\text{COMPS} & \begin{bmatrix}2\end{bmatrix}
\end{array}
\end{array}
$$

Figure 6.8: Head complement rule applied to *likes John*

The **Specifier head rule** (spr_hd) allows a lexical head to combine with its specifier that it selects to form a phrase of type spr-hd-ph as in (4). The spr_hd rule requires a two-way restriction: specifiers restrict their heads and heads restrict their specifiers [Matheson 1996]. For example, a specifier *this* selects singular nouns while common nouns select determiners as their specifiers. Therefore, the features SPECIFIER, SPR and SPEC, are introduced. SPR restricts the co-occurrence for heads and the specifiers they select. SPEC restricts the heads,
the kinds of things they specify. A noun book (Figure 6.9), a HD-DTR of a spr-hd-ph the book, combines with its determiner a, SPEC-DTR, which instead selects singular nouns to be specified.

(4)

```
(4)
spr_hd_ph ⇒ [spec-dtr \[ synse | loc | cat | head | spec 2 \] ]
HD-DTR \[ phrase \{ synse | loc | cat | spr 1 \] ]
```

```
hd-spr-ph
PHON \{ a, book \}
SYNSEM|LOC|CAT \{ subj \}
SPR \{ \}
HD-DTR \[ word \{ book \}
PHON \{ \}
SUBJ \{ \}
SPR \{ 2 det \}
CONT | INDEX | NUM | sg
SPEC-DTR \[ PHON \{ a \}
DET \{ \}
SPEC \{ 1 np [sg] \}
```

Figure 6.9: Specifier_head rule applied to the book

6.1.1.5 Attribute Logic Engine

The Attribute Logic Engine 3.2 Beta, was written in SICStus Prolog and developed by Bob Carpenter and Gerald Penn. ALE is a grammar parsing system which was developed to facilitate the use of HPSG and any typed feature structure grammar. Every structure used in ALE comes with a type. These types are arranged in an inheritance hierarchy whereby type constraints on more general types are inherited by their more specific subtypes [Carpenter and Penn 1999]. ALE differs from other approaches to feature structures in at least three notions. First, there is a strong type discipline enforced on feature structures. This means each type must specify which features it can be defined for and which types of values such features can take. The second notion is the type inference whereby types for structures which are underspecified can be automatically inferred. The third notion is that types in ALE are semantic which means that types are used at run-time. Even though ALE employs

---

5 ALE 3.2 Beta also contains a generator.
semantic typing, a type system is employed to detect type errors in grammars.

ALE also offers macros, to be used in description. Macros allow the user to define a
description once and then use a shorthand for it in other descriptions. Macros can also
contain other macros. The advantages of macros is not only to make the grammars more
readable (especially to organize lexical information) but also to reduce the number of simple
typing errors that lead to inconsistencies. Some sample macros and their usage are illustrated
in section 6.1.1.6.

The format ALE provides for stating lexical rules is similar to that found in HPSG.
However, the lexical rule system of ALE is productive in that it allows lexical rules to apply
sequentially to their own output or the output of other lexical rules. Thus, it is possible to
derive the nominal runner from the lexical entry run by the nominal lexical rule, and then
derive the plural nominal runners from runner by the plural noun lexical rule.

The ALE system employs a bottom-up active chart parser [Carpenter and Penn 1999].
The rules are evaluated from left to right, with the mother description coming last. The chart
is filled in using a combination of depth-first and breadth-first control. The parser proceeds
breadth-first in the sense that it incrementally moves through the string from right to left,
one word at a time, recording all of the inactive edges that can be created beginning from the
current left-hand position in the string.

ALE provides a number of routines for displaying and debugging grammar specifications
as well. In executing ALE, the input string must be entered as a Prolog list of atoms. In
particular, it must have an opening and closing bracket, with words separated by commas. No
variable should occur in the query, nor anything other than an atom. The first part of the
parsing output repeats the input string, separated by numbers (nodes in the chart) which
indicates positions in the string for using in inspecting the chart directly (Figure 6.10). The
second part of the output is a category which is derived for the input string. If there are
multiple parses, these can be iterated through by providing positive answers (yes) to the query
"Another?". The parsing stops when no more parses are found. Further details on ALE can
be found at ALE homepage “http://www.sfs.nphil.uni-tuebingen.de/~gpenn/ale.html #Obtain”,
available as of May, 1999.
6.1.1.6 English and Thai Grammars in GRMT

In HPSG, the grammatical categories are broken down into features and the feature values are broken down into features as well. In this process, more and more syntactic information is moved out of the grammar rules and into the lexicon. This philosophy results in rules with very general patterns that cut across grammatical categories. The following seven grammar rules we developed cover a wide range of cases in both English and Thai (see also appendix F).

The subject_head rule (subj_hd), the head_complement rule (hd_comp), and the coordination rule cover the phrasal structures not only in English but also in Thai (the similarity between the typical phrasal structures of English and those of Thai was discussed in section 4.1). The head_subject_complement rule (hd_subj_comp) is used to license an English inverted sentence. The hd_subj_comp can also be used to analyze the phrasal structures which have verb-subj-object order in any language. The specifier_head rule (spr_hd) combines a lexical head with its selected specifier in English. The adjunct_head and head_adjunct rules allow a phrase to consist of a phrasal head and its modifier. The headSpecifier (hd_spr) rule combines a lexical head with its specifier in Thai.

\[
(5) \quad \text{subj-hd-ph} \Rightarrow \begin{bmatrix}
\text{SUBJ-DTR} & [\text{SYNSEM} | \text{LOC} | \text{CAT} | \text{HEAD} \ [1 \ ]]
\text{HD-DTR} & [\text{SYNSEM} | \text{LOC} | \text{CAT} | \text{SUBJ} \ [1 \ ]]
\end{bmatrix}
\]

Figure 6.10: Example of result of parsing in ALE representation
The subj_hd rule allows a lexical head to combine with its selected head according to the constraint as in (5). The HEAD value of the SUBJ-DTR must match the SUBJECT value of the HD-DTR.

\[
\text{hd-comp-ph} \Rightarrow \begin{cases} 
\text{HD-DTR} & \left[ \begin{array}{c} \text{word} \\ \text{SYNSEM} | \text{LOC} | \text{CAT} | \text{COMPS} \\
\{ 1, 2, \ldots, n \} \end{array} \right] \\
\text{NON-HD-DTRS} & \{ \text{SYNSEM} \}, \{ \text{SYNSEM} \}, \ldots, \{ \text{SYNSEM} \} 
\end{cases}
\]

The hd_comp rule, combines a head with its complements, according to the constraint as in (6), was discussed in section 6.1.1.4.

\[
\text{coordination-ph} \Rightarrow \begin{cases} 
\text{CONJUNCT-DTR1} & \left[ \begin{array}{c} \text{phrase} \\ \text{SYNSEM} | \text{LOC} | \text{CAT} \end{array} \right] \\
\text{COORDINATE-DTR} & \{ \text{SYNSEM} \}, \{ \text{SYNSEM} \}, \ldots, \{ \text{SYNSEM} \}, \{ \text{HEAD conj} \} \\
\text{CONJUNCT-DTR2} & \left[ \begin{array}{c} \text{phrase} \\ \text{SYNSEM} | \text{LOC} | \text{CAT} \end{array} \right] 
\end{cases}
\]

This coordination\(^6\) rule says that the two occurrences\(^7\) of elements of the same syntactic category can be conjoined to make a coordinate element of that category (7). This rule requires that all conjuncts of a coordinate structure have identical values for HEAD, SPR and COMPS.

The hd_subj_comp rule allows a lexical head to combine with its subject and complements that it selects to form a phrase of verb-subj-object, in that order, as in (8). The HD-DTR must be an inverted verb and its HEAD value must match the SUBJ value of the SUBJ-DTR. The COMPS values of the HD-DTR must match a sequence of SYNSEM values of NON-HD-DTRS.

---

\(^6\) The coordinate conjunctions are words connecting elements that are, intuitively, of equal status. In English, the paradigmatic coordinate conjunctions are and or or [Sag and Wasow 1999].

\(^7\) The number of the conjuncts of a coordinate structure in the coordination rule proposed by Sag and Wasow [Sag and Wasow 1999] is not limited at two: it could be any number of occurrences of conjuncts. However, the semantics transfer from the conjuncts to the mother is not yet finalized at this time: therefore, our coordination rule considers only two conjuncts for simplicity.
The spr_hd rule (9) says that a phrase can consist of a phrasal head preceded by its specifier, as we discussed extensively in section 6.1.1.4.

In English, adjuncts (modifiers) appear both before and after the objects they modify (HD-DTR), e.g., the ugly duckling, the duckling in the farmyard. Therefore, two rules: adjunct_hd rule and hd_adjunct rule are required to combine the HD-DTR with its modifier as in (10). In Thai, the modifiers must follow the HD-DTR (see section 4.1): thus only the hd_adjunct rule is required. The MODIFIER (MOD) values of the ADJUNCT-DTRs indicate the types of the object that the adjunct modifies. The subtypes pre_mod_synsem and post_mod_synsem specified in the ADJUNCT-DTR restrict which rule will be applied.
Figure 6.11: The lexical entry for *hide*

The grammars we developed have been implemented using the Attribute Logic Engine. Figure 6.12 illustrates the representation of the lexical entry *hide* (its AVM diagram is shown in Figure 6.11) in ALE with some macros. The use of the macro is specified by the @ sign followed with the macro name e.g., @np, @case.
Figure 6.12: The representation of hide in ALE with some macros

Figure 6.13 illustrates the parse (in ALE representation) of the SL in Example 1 and the parse of its TC. The TC is generated by ALMT. The parses are in ALE representations. Both the SL and TC are licensed by our grammars. Their syntax is shown in the dotted boxes. The SL analysis begins by analyzing the word wing and combines it with his by the spr_hd rule. The phrase his head under his wing are licensed by hd_comp and the subj_hd rules since the lexical entry under requires subject and complement as specified lexically in Figure 6.14. Ugly and duckling are combined by the adjunct_hd rule, then the spr_hd rule combines ugly duckling with the. The phrase The ugly duckling hides his head under his wing results from combining The ugly duckling and his head under his wing via the hd_comp and the subj_hd rules. Again, everything is specified lexically; hides requires a subject and two complements. The lexical entry hides is generated by the 3rd-singular verb lexical rule (Figure 6.5).

Example 1:

SL: The ugly duckling hides his head under his wing.

TC: ฉลามเป็ด กี่แวก ที่นั่น แขวน หัว ของเขา ใต้ บุก ของเขา
(ฉลามเป็ด- duckling) (กี่แวก- ugly) (ที่นั่น- clas) (แขวน- the) (หัว- head) (บุก- his) (ของเขา) (ใต้- under) (บุก- wing) (ของเขา- his)
Figure 6.13: The parse syntax of Example 1 and that of its translation
Figure 6.14: The representation of *under* in ALE with some macros

Figure 6.15: The representation of *khīrēe* (ugly) in ALE with some macros
The phrase *The ugly duckling* and *his head under his wing* also obeys the Head Feature Principle in which the HEAD value of the mother (HEAD: verb, Figure 6.13) is identified with that of the HD-DTR. The effect of the semantic principle (CONT) will be discussed in section 6.2.

The TC is analyzed in an analogous fashion, however, different rules are applied due to the syntax differences between languages. For example, ข้าวหน้า (khā̄nkhaw- his) and ปีน (piig-wing) are combined with hd_spr rule to be ปีน ข้าวหน้า (corresponding to *his wing*). The phrase ลูกเป็ด ขี้เร่ (corresponding to *ugly duckling*) results from combining ลูกเป็ด (lūugpēd-duckling) and ขี้เร่ (khīrēe- ugly) via the hd_adjunct rule since ขี้เร่ (khīrēe- ugly) is the type of post_mod_synsem as indicated in the lexical entry shown in Figure 6.15. The phrase ลูกเป็ด ขี้เร่ ตัว หน้า (corresponding to *The ugly duckling*) are formed by the hd_spr rule. The WORDASSO1 value of the classifier ตัว (tua), Figure 6.16, indicates the WordAsso number of the object that compatible with the classifier ตัว (tua).

![Figure 6.16](image)

Figure 6.16: The representation of the classifier ตัว (tua) in ALE

Figure 6.17 illustrates the parse of the SL in Example 2 and the parse of its TC. The SL is licensed by our English grammar, however, there is no parse for the TC since ร่วม (cān-John) and ซื้อเช่า (khāchāw-rent, rental fee) cannot be combined according to our Thai grammar; it is syntactically incorrect. Use of the semantic information of the parse is discussed in the next section.
Example 2:

SL: John rents a room from Joan.


Figure 6.17: The parse syntax of Example 2 and that of its translation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
6.1.2 Semantic Extraction

Once the SL and the TC are parsed successfully, their semantic information will be extracted so that we can cross-examine the meaning of the TC with that of the SL. We first discuss how the semantic information is represented in our grammar. The semantic representation of an expression in GRMT is generally based on the representation in HPSG provided by Pollard and Sag, 1994 and Sag and Wasow, 1999 which will be presented in the next section.

6.1.2.1 Semantic Representation

The features CONT, CONTEXT (CONX) and QUANTIFIER-STORE (QSTORE) hold the semantics of the words. The semantic representation described in CONT (Figure 6.11) corresponds to the situation in which \( \text{the hider} \) hides \( \text{the hid} \) in (or under, from, etc.) \( \text{the hid_place} \). The features SURFACE and WORDASSO indicate the word form and WordAsso number of the lexical entry. These features, SURFACE and WORDASSO, are used in the Repair and Iterate phrase. The feature QSTORE is storage for the quantifiers. The QSTORE value of the in Figure 6.18 is token-identical to the CONT value. The QSTORE value of the mother is passed from its constituents according to the Quantifier Inheritance Principle (QIP) [Pollard and Sag 1994]. QIP says that the RETRIEVED-QUANTIFIERS (RETRIEVED, its

![Figure 6.18: The partial description of the lexical entry the](image_url)
value is a list of quantifiers) value of the mother is a list whose set of elements forms a subset of the union of the \textit{QSTORE}s of the daughters, and is nonempty only if the \textit{CONTENT} of the semantic head is of type \textit{psoa}, and the \textit{QSTORE} value is the relative complement of the \textit{RETRIEVED} value.

The \textit{CONX} feature contains linguistic information that bears on certain context-dependent aspects of semantic interpretation [Pollard and Sag 1994]. The \textit{CONX} value is a feature \textit{BACKGROUND} (\textit{BACKGR}) whose value is a set of \textit{psoas}. Each background \textit{psoa} restricts the possible anchors of indices that correspond to presuppositions or conventional implicature. The \textit{BACKGR} value of \textit{Joan} is illustrated in Figure 6.19. The atomic value \textit{Joan} refers to the name \textit{Joan}, not to an individual named \textit{Joan}. Therefore, its \textit{psoa} corresponds to the presupposition that the referent be identifiable in the utterance context by means of the name \textit{Joan}. The \textit{BACKGR} value of the mother obeys the \textit{Contextual Consistency Principle} in which the \textit{BACKGR} value of the mother is the union of the \textit{BACKGR} values of the daughters as in (12).

$$\text{(12)}$$

$$\begin{align*}
\text{HD-ph} & \Rightarrow [\text{HD-DTR} | \text{SYNSEM} | \text{LOC} | \text{CONX} | \text{BACKGR} | 1 \cup 2 \cup \ldots \cup n]
\end{align*}$$

\text{DTRS} \langle [\text{BACKGR} 1], [\text{BACKGR} 2], \ldots, [\text{BACKGR} n] \rangle

Figure 6.19: The partial description of the lexical entry \textit{Joan}
6.1.2.2 Semantic Extraction Examples

To extract the semantic information, CONT, QSTORE and CONX from the parse, we use a predicate rec/5 provided in ALE as shown in Figure 6.20. However, this predicate rec/5 binds the result (the feature values) to variables in the internal representation of feature structures of ALE. This result will be converted into a Prolog list for use in further procedure. Semantic Comparison.

Figure 6.20 illustrates the extracted internal representation of the semantic information of the SL parse of Example 3. This representation corresponds to the display in Figure 6.21 which illustrates QSTORE, CONT and CONX, the three semantics indicates used in HPSG and explained as follows. This semantic representation corresponds to the situation in which the talker is masculine and his name is John (CONT: TALKER1 masc, CONX: BACKGR: (BEARER1, NAME: john); the TALKER1 and BEARER are unified as indicated by 1). The talked is about the device (spring) which is broken (CONT: TALKED 0, QSTORE: RESTRI: (INSTANCE:0, {broken1, device}); the TALKED and INSTANCES are unified via 0).

Figure 6.22 and Figure 6.23 illustrate the extracted (as above for 6.20 and 6.21) semantic information, in ALE’s internal structure and its corresponding display respectively, of the TC parse of Example 3. This semantic representation corresponds to the situation in which the talker is masculine and his name is John (CONT: TALKER1 masc, CONX: BACKGR: (BEARER1, NAME: John); the TALKER1 and BEARER are unified as indicated by 1)). The talked is about the season (qglulmia - spring, in the sense of the season between winter and summer) which is broken (CONT: TALKED 0, QSTORE: RESTRI: (INSTANCE:0, {broken, season}); the TALKED and INSTANCES are unified via 0). According to this analysis, the TC carries no meaning since season cannot be “broken” either physically and forcibly separated into pieces or cracked or split.

Note that Example 3 is intended to demonstrate the situation in which the SL and the TC convey different meanings. In this example, we forced ALMT to disregard the relationship between the words broken and spring during the word selection process. Therefore, the word งูงูไม่มี (ry’duubajmaajphli- the season) is selected as a translation for spring. As a result, the TC obtained is incorrect. Other possible translations of spring are าาย (sapring- an elastic device), กระดูก (kradood- to move suddenly upward or forward) and น้ำฟู (námphû-place where water comes up naturally from the ground).

In the next section, the semantic representations of the SL and the TC are compared to
determine whether they convey acceptably similar meaning.

Example 3:

SL: John talks about the broken spring.

TC: (caan- John) (phud- talk) (kiaw káb- about)
    (ry’dubajmáajphl- spring, the season)
    (thii- modifying) (hág- broken) (nán- the)

Figure 6.20: The semantic information,
in ALE's internal structure, of
the SL in Example 3

Figure 6.21: The semantic information
of the SL in Example 3
Figure 6.22: The semantic information, in ALE's internal structure, of the TC in Example 3

Figure 6.23: The semantic information of the TC in Example 3
6.2 Semantic Comparison

In comparing the meaning between the SL and its TC, the values of the features CONT, QSTORE and CONX are considered. If the values of these features of the SL parse are the same as those of the TC parse, TCE concludes that the TC does not require repair. If any of these features are different, TCE will provide information of the SL parse which differs from that of the TC parse. This information will be used in the repair process phrase, Repair and Iterate.

The comparison process begins by investigating the CONT value of both SL and TC parses. Figure 6.24 and Figure 6.25 illustrate the semantic information of the SL and that of the TC in Example 1 (section 6.1.1.6). Their simplified displays are illustrated in Figure 6.26 and 6.27. The CONT values of both parses represent the same kind of relation involved, hide2.

```
rec([the, ugly, duckling, hides, his, head, under, his, wing]).


CONX = _G19428-conx(_G19433-e_set)
```

Figure 6.24: The semantic information of the SL in Example 1
and the same type of persons (or things), ref, who/that are participating in this relation (_G18882-ref, _G19053-ref and _G19160-ref in the CONT value of the SL parse: _G21321-ref, _G21470-ref and _G21577-ref in the CONT value of the TC parse), each ref is singular.

The right of a hyphen is a feature value which is bound to the variable on the left of the hyphen.

Once the CONT values of the SL parse and that of the TL parse are recognized as the same. TCE, then cross-examines the QSTORE and the CONX values of the SL parse with those of the TC parse to determine whether each variable (index) in the CONT value is associated to the appropriate object. In Figure 6.24 and 6.25, the first variable in the CONT value of the SL parse is _G18882-ref and that of the TC parse is _G21321-ref. Both variables are associated to a bodypart in the class of 1-4-1-1-1 (in the QSTORE value), it is singular and referred to as head in the SL and as ḫa (hu-a) in the TL. ḫa (hu-a) is the translation of the
Figure 6.26: The semantic information of the parse of the SL in Example 1
Figure 6.27: The semantic information of the parse of the TC in Example 1
word *head* in Thai. The second variables are _G19053-ref (of the SL CONT value) and _G21470-ref (of the TC CONT value). These variables correspond to a *bodypart* in the class of 1-4-1-3-1, it is singular and referred to as *wing* in the SL and ฝลง (piig) in the TL. Again, ฝลง (piig) is a translation of *wing* in Thai. The last variables are _G19160-ref and _G21577-ref. They are associated with an animal in the class of 1-1-1-2-1-2-1 which is *singular* and referred to as *duckling* in the SL and นก (luugpèd) in the TL. นก (luugpèd) is the translation of the word *duckling* in Thai. The CONX feature in this example contains no value. According to the comparison process, TCE found no difference between the CONT, QSTORE and CONX of the SL parse and those of the TC parse. Therefore, the TC is deemed as an appropriate translation for the SL in the Example 1.

The CONT values of the SL and the TC parses in Example 3 (section 6.1.2.2) are not the same. The first variable of the SL CONT (_G8296-ref, Figure 6.20) relates to a *device* in the class of 1-1-2-1-3-1-4 which is referred to as *spring* (indicated in the QSTORE value) whereas its corresponding variable in the TC CONT (_G9224-ref, Figure 6.22) relates to a *season* in the class of 2-4-3-2 which is referred to as งูในมวล (ry’duubajmáaj-phli). งูในมวล (ry’duubajmáaj-phli) is one of the possible translations of the word *spring*. However, it is not an appropriate translation of *spring* in the expression of Example 3. The differences between the SL QSTORE and TC QSTORE indicate that the TC requires a repair.

The CONT values of the SL parse and the TC parse in Example 4 (Figure 6.28 and Figure 6.29) represent the same relation involved, *cover*, and the same types of persons or things who/which are participating in this relation, COVERED, COVERED_PLACE and COVERER. The SURFACE value in the CONT of the SL parse is *cover* and that of the TC is ฝลง (khlum). The value of the feature SURFACE is a word form. ฝลง (khlum) is a translation of *cover*. However, their WORDASSO values are different, *cover* belongs to the class of 2-1-19 whereas ฝลง (khlum) belongs to the class of 2-1-19-2 as illustrated in Figure 6.30. Although *cover* and its translation ฝลง (khlum) carry the same meaning, they are different in language usage. Differences in languages and cultures result in different circumstances of language usage. The word *cover* in *John covers his ears with his hands* corresponds to ฝลง (pid) in Thai. When verifying the meaning of *cover*, the word and its translation are considered acceptably similar in principal features since they belong to the same superclass.
Example 4:

SL: John covers his head with his coat.

TC: จอหนุ คุ้ม หัว ของเขา ด้วย เนื้อคุ้ม ของเขา

(cam- John) (khlum- cover) (hu'a- head) (khā'ankhaw- his) (dāaj- with) (sy'akhlum-coat)(khā'ankhaw- his)

Figure 6.28: The CONT value of the SL parse of Example 4

Figure 6.29: The CONT value of the TC parse of Example 4
Table 6.3: The class of \textit{cover}

<table>
<thead>
<tr>
<th>WordAsso and Class description</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1-19 Cover as to place something on or over to protect or hide.</td>
<td>cover</td>
</tr>
<tr>
<td>2-1-19-1 Cover as to cover eyes, ears.</td>
<td>ปิด (pid)</td>
</tr>
<tr>
<td>2-1-19-2 Envelop</td>
<td>กิ่ง (khum)</td>
</tr>
</tbody>
</table>
Chapter 7

Repair and Iterate

The last phase of GRMT, Repair and Iterate, performs the repair process if the translation candidate carries a different meaning from the SL. The repaired TC is then returned to the Translation Candidate Evaluation module. TCE re-analyzes the repaired TC to determine if additional repair is further required (a different meaning from the source language remains). The repair process repeats until the semantic information (CONT, QSTORE and CONX) of the SL and the TC are acceptably similar.

TC with the semantic differences info. 

Figure 7.1: RI Architecture

RI examines the result of TCE. The TCE output is the TC with the semantic information of the SL which differs from that of the TC (Figure 7.1). With this information, RI is able to

---

1 By acceptably similar we mean that a native TL speaker would find the TL translation represents the meaning of the SL statement to some acceptable degree (this degree may vary from language to language).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
detect the part of the TC which causes the mis-translation. The mis-translated part will be replaced with a more appropriate translation. RI searches the Word Treatment output (generated by ALMT, see section 5.1) for a more appropriate translation based on the information provided by TCE. The TC with the new selection is then put through the Word Ordering module (see section 5.3) to revise its syntax. Once the revision is completed, the repaired TC is returned to TCE.

The semantic information described is represented in terms of CONT, QSTORE and CONX features (see section 6.1.2.1). In the case that the CONT or the QSTORE value of the SL is different from that of the TC, the SL CONT value: SURFACE and WORDASSO features are passed to RI. The SURFACE value indicates the surface form of the word which causes the mis-translation. The WORDASSO value specifies the proper meaning of the word in question in terms of WordAsso number. Therefore, to repair the CONT or QSTORE value of the TC, RI re-selects the corresponding word in the TL for the word specified in the SURFACE value. The re-selection is done by searching the Word Treatment output for the corresponding word which has the same WordAsso number as specified in the WORDASSO value.

In the case that the CONX value of the SL differs from that of the TC, the SL CONX value: the BEARER and NAME features are passed to RI. The BEARER value specifies the index which associates with the certain name specified in the NAME value. RI repairs the TC by associating the right names to the right indices based on the information provided by TCE.

The results of analyzing the SL and the TC in Example 1, below, indicate that the TC requires a repair since it carries a different meaning from the SL (see Example 3 in section 6.1.2.2). The CONT values of the SL and the TC are the same; however, parts of their QSTORE values are different. The QSTORE values of the SL and the TL illustrated in Figure 7.2 indicates that the word spring is mis-translated (SURFACE: spring). According to the QSTORE of the SL, the proper meaning of spring in Example 1 is a device which is classified into the class of 1-1-2-1-1-3-1-4 (WORDASSO: 1-1-2-1-1-3-1-4). Therefore, RI begins the repair process by re-selecting the translation of the word spring. RI searches the Word Treatment output (Table 7.1) for the translation of spring which has WordAsso number 1-1-2-1-1-3-1-4 and thus, the translation aguay (sapring) is selected. The word กายการ์ (ry‘duubajmáajphli) in the TC is then replaced with กายการ์ (sapring). The TC with the new
selection (Figure 7.3) is put through the *Word Ordering* module to revise its syntax. At this step, the classifier ąn (?an- classifier) is added to complete the repaired TC.

Once the repair processes successfully, the repaired TC is analyzed by TCE. Figure 7.4 illustrates the semantic information of the SL parse and that of the repaired TC. Their CONT

Example 1:
SL: John talks about the broken spring.
TC: ąariu tōt tinu qtuulub;u 5 11 vin u`u
(caan- John) (phuud- talk) (kiaw kab- about) (ry’duubajmáajphli- spring, the season) (thii- modifying) (hag- broken) (nán- the)

Figure 7.2 The QSTORE values of the SL and the TC parses of Example 1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 7.1 A part of Word Treatment output of Example 1

<table>
<thead>
<tr>
<th>English</th>
<th>Word Treatment Output</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>spring</td>
<td>ฎ์ฎุฎฎุฎฎ (ry'duubajmáajphli),</td>
<td>the season between winter and summer</td>
</tr>
<tr>
<td></td>
<td>2-4-3-2</td>
<td></td>
</tr>
<tr>
<td>sapring</td>
<td>(sapring),</td>
<td>an elastic device</td>
</tr>
<tr>
<td></td>
<td>1-1-2-1-3-1-4</td>
<td></td>
</tr>
<tr>
<td>kradood</td>
<td>(kradood),</td>
<td>to move suddenly upward or forward</td>
</tr>
<tr>
<td></td>
<td>2-1-5-5-1</td>
<td></td>
</tr>
<tr>
<td>namphu</td>
<td>(námphû),</td>
<td>place where water comes up naturally from the ground</td>
</tr>
<tr>
<td></td>
<td>1-1-2-1-4-1</td>
<td></td>
</tr>
</tbody>
</table>

TC: จำาน (caan- John) ฎู (phûud- talk) แก้ว or (kiâw kâb- about) ฎูไม่นินนิ ฎู (ry'duubajmáajphli- spring, the season) มัน (thii- modifying) มัน (hâg- broken) มัน (nán- the)

TC with the new selection: จำาน (caan- John) ฎู (phûud- talk) แก้ว or (kiâw kâb- about) ฎู (sapring- the device) มัน (thii- modifying) มัน (hâg- broken) มัน (nán- the)

Repaired TC: จำาน (caan- John) ฎู (phûud- talk) แก้ว or (kiâw kâb- about) ฎู (sapring- the device) มัน (thii- modifying) มัน (hâg- broken) มัน (nán- the)

Figure 7.3: The repair of Example 1
Figure 7.4: The semantic information of the SL and that of the repaired TC in Example 1
QSTORE and CONX values are the same. The repaired TC is, then, deemed as an appropriate translation for the SL in Example 1.

In the case that TCE finds no parse for a TC since it is grammatically incorrect according to TL grammar, TCE outputs the TC with the CONT and QSTORE values of the SL. RI begins the repair process by verifying the translation of the word specified in the CONT: SURFACE value of the SL. If the WordAsso number of that translation is the same as the WordAsso number specified in the CONT: WORDASSO value of the SL, then RI proceeds to verify the translation of the word specified in the QSTORE: SURFACE. However, if the WordAsso number of that translation is different from the WordAsso number specified in the CONT: WORDASSO value of the SL, then RI replaces that translation with another translation which has the same WordAsso number specified in the CONT: WORDASSO value of the SL. The TC with the new selection is then put through the Word Ordering module to revise its syntax. If the repaired TC remains syntactically incorrect, then RI verifies the words specified in the QSTORE: SURFACE value.

TCE found no parse for the TC in Example 2 (Figure 7.5) since the TC is grammatically incorrect (see Example 2 in section 6.1.1.6). RI verifies the translation of the word rent, since it is specified in the CONT: SURFACE value of the SL as illustrated in Figure 7.6. The selected translation of rent in the TC is წჰაჰა (khâchâw) with 2-1-3-3-3-2. However, the WordAsso number specified in the CONT: WORDASSO value of the SL is 2-1-3-3-2-2. Therefore, RI searches the Word Treatment Output (Table 7.2) for another translation of rent which has the WordAsso number 2-1-3-3-2-2. RI, then replaces წჰაჰა (khâchâw) with ჰა (châw, Figure 7.7) and revises the syntax of the TC with the new selection. The repaired TC is analyzed and compared to the SL parse. TCE now finds no difference between the SL and the repaired TC parses (Figure 7.8).
Example 2:
SL: John rents a room from Joan.
TC: จอห์น ค่าเช่า ห้อง ห้อง หนึ่ง จาก โจแอน
(common- John) (ประจวบ- rent, rental fee) (ห้อง- room) (ห้อง- clas) (หนึ่ง- a) (จาก- from) (จอแอน- Joan)

Figure 7.5: The TC parse of Example 2

Figure 7.6: Cont value of the SL parse
Table 7.2: A part of Word Treatment output of Example 2

<table>
<thead>
<tr>
<th>English</th>
<th>Word Treatment Output</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rent</td>
<td>คำเช่า (khāchāw), 2-1-3-3-3-2</td>
<td>money paid for the use of another’s property</td>
</tr>
<tr>
<td></td>
<td>ไทยเช่า (hājchāw), 2-1-3-3-2-1</td>
<td>to give possession and use of in return for rent</td>
</tr>
<tr>
<td></td>
<td>เช่า (chāw), 2-1-3-3-2-2</td>
<td>to take and hold under an agreement to pay rent</td>
</tr>
<tr>
<td></td>
<td>เยอะเช่า (rādkhāadh), 2-3-11-1</td>
<td>a large tear</td>
</tr>
<tr>
<td></td>
<td>ธิ (chiig) (^1), 2-1-6-1-3</td>
<td>to divide by force</td>
</tr>
<tr>
<td></td>
<td>ธิ (thī) (^2), 2-1-6-1-2</td>
<td>to pull violently</td>
</tr>
</tbody>
</table>

TC: จอย้น (caan- John) คำเช่า (khāchāw- rent, rental fee) ห้อง (ha¬η- room) ห้อง (ha¬η- clas) หนึ่ง (nyη- a) จาก (cāag- from) จอย้น (joo?en- Joan)

TC with the new selection: จอย้น (caan- John) เช่า (khāchāw- rent, rental fee) ห้อง (ha¬η- room) ห้อง (ha¬η- clas) หนึ่ง (nyη- a) จาก (cāag- from) จอย้น (joo?en- Joan)

Repaired TC: จอย้น (caan- John) เช่า (khāchāw- rent, rental fee) ห้อง (ha¬η- room) ห้อง (ha¬η- clas) หนึ่ง (nyη- a) จาก (cāag- from) จอย้น (joo?en- Joan)

Figure 7.7: The repair of Example 2

\(^1\) The meanings of rent here are a past tense and a participle of rend [Longman 1992].
Figure 7.8: The semantic information of the SL and that of the repaired TC in Example 2
Chapter 8

Machine Translation Evaluation

'There is never a single right translation, and therefore never a solid ground truth against which MT output may be compared'
— John White, 1999

Evaluation plays a crucial role in machine translation for system developers, funding agencies and technology users. The system developers (at least) want to determine whether the system performs in the ways intended. The funding agencies aim at determining whether a particular research project should be funded whereas the technology users want to determine which system best suits their needs and is cost-effective. More importantly, all system developers, funding agencies and technology users want to determine whether the output is acceptable as a translation. Information necessary to make informed judgments is required. Unfortunately, a universally accepted methodology for evaluating MT systems is not yet available, partly because there are different purposes for MT, different interests of the participants in the process, radically different theoretical approaches to MT and different languages. In addition, it is difficult to evaluate the end product of an MT system since "there is never a single right translation, and therefore never a solid ground truth against which MT output may be compared" [quoting White 1999]. Moreover, most evaluations have been done under contracts and often under confidentiality agreements. Consequently relatively little constructive criticism of the techniques employed is in the public domain [King 1990].

GRMT represents an approach to MT that is not unlike what HPSG represents with respect to grammar theory. GRMT has borrowed many ideas from the major translation
paradigms (Direct MT, Interlingual MT, Transfer MT, Example-Based MT, Knowledge-Based MT and Shake-and-Bake MT) and included the notion of adding constraints where applicable, as found in HPSG theory. In this sense GRMT represents a new paradigm for MT systems. Just as it is difficult to evaluate past MT efforts (no generally accepted evaluation methodology), it is difficult to evaluate GRMT precisely. Nonetheless, we outline in this chapter some initial requirements for evaluation and some strategies which we hope to apply to GRMT.

8.1 What needs to be evaluated? (Types of Evaluation)

As in other areas of natural language processing (NLP), three types of evaluation appropriate to three different goals are recognized [Hirschman and Thompson 1996], [Hutchin 1996]:

- **Adequacy Evaluation** is to determine the fitness of MT systems (e.g., will the system do what is required, how well and at what cost) within a specified operational context. Adequacy evaluation is typically performed by funding agencies and by technology users.
- **Diagnostic Evaluation** is to identify limitations, errors and deficiencies, which may be corrected or improved by the researcher or by the developers. Diagnostic evaluation is the concern mainly of researchers and developers.
- **Performance Evaluation** to assess stages of system development or different technical implementations. Performance evaluation may be undertaken by either researchers or developers.

MT evaluations typically include features not present in evaluations of other NLP systems including:

- Quality of the raw translations.
- Usability of facilities for creating and updating dictionaries.
- Extendibility to new language pairs and/or new subject domains.
- Cost-benefit comparisons with human translation performance.

*Quality of the raw translations:* It is extremely difficult to determine what counts as a good translation whether produced by human or machine. "The relative adequacy of
different translations of the same text can only be determined in terms of the extent to which each translation successfully fulfills the purpose for which it was intended” [quoting Nida 1976]. Therefore, intelligibility (or clarity) and fidelity (or accuracy) are the most common criteria in determining a quality of the raw translation [Hutchines and Somers 1992], [Arnold et al. 1994]. However, they remain subjective judgments.

- **Intelligibility** determines whether the translation consists of the correct words in the target language and whether the evaluators can understand the translation with ease. Intelligibility, therefore, can be measured by considering whether the translation is affected by grammatical errors, mistranslations and untranslated words. A common measurement is to assign scores to the translation. ALPAC’s intelligibility scale is a nine-point scale [ALPAC, 1966]. The ALPAC scoring scale ranges from 1, *hopelessly unintelligible*, to 9, *perfectly intelligible*. Another intelligibility scale is proposed by Nagao, Tsujii and Nakamura; this scoring scale ranges from 1, *the meaning of the translation is clear and needs no rewriting*, to 5, *the translation cannot be understood at all* [Nagao et al. 1988]. Generally, only individual sentences are evaluated. However, by isolating sentences from their contexts, such measurements are made even more subjective and uncertain than they might be [Hutchins and Somers 1992].

- **Fidelity** determines whether the translation preserves the meaning of the source language or whether the translation contains the same information as the original. Various tests have been proposed and implemented to provide measurements of fidelity. For example, the evaluators read the translation and judge how much more *informative* was the original. However, this procedure can be criticized as being excessively subjective. Another practical measurement which is feasible for measuring accuracy of translated instruction manuals is to determine whether the person who uses the translated manuals can carry out the instructions as well as the person using the original manual. An alternative measurement is to translate the translation output (generated by MT) back into the original language and then compare the result with the original text. Note that shortcomings are generally magnified by this double translation process. Also note that a translation which is faithful to the original may be difficult to understand while a translation which is easy to read may have distorted the original message [Hutchins and Somers 1992].
Usability of facilities for creating and updating dictionaries: The practical MT system should provide facilities for users to expand and update the dictionaries. Changes in dictionary information may result in improvements. The basic questions to determine the degree of usability are whether the system allows the users to make changes to dictionaries, whether the users can update the dictionaries with ease, and whether there is a limitation on the size of the dictionaries.

Extendibility to new language pairs and/or new subject domains: MT systems designed for one particular subject area are inherently more difficult to extend to other areas, not only because of differences in vocabulary but also because of differences of grammar. The basic question to determine the degree of expansion is whether components developed for one language pair can be applied or easily adapted for another language pair. The extendibility of the system depends highly on the modularity of its components [Hutchins and Somers 1992].

Cost-benefit comparisons with human translation performance: Cost estimates and benefits of the system must be taken into account. Costing of MT systems include the basic cost of software, the cost of installation, upgrading and maintenance, and the cost of pre-editing and post-editing (if any). A comparison will be made with the costs of producing the same quality of output by human translators. Benefits trade-off against costs. Possible benefits include faster production of translations, faster translation of large volumes of text, and simultaneous translation in many languages. Furthermore, the costs of maintaining dictionaries may well diminish with time since the dictionary contains large numbers of lexical entries [Hutchins and Somers 1992].

8.2 MT Evaluation Strategies

Some specialized evaluation strategies have been proposed and implemented. For example, scoring scales for measuring quality of the raw translations (see section 8.1), and error analysis, and the test suite. The error analysis and test suite are briefly discussed in this section. Further discussion on MT evaluation can be found in [Lehrberger and Bourbeau 1988], [King 1990], [King and Falkedal 1990], [Hutchins and Somers 1992] and [Arnold et al. 1994].
8.2.1 Error Analysis

Error analysis strategies focus on the errors the MT system makes. Error counting indicates the amount of work required to correct raw MT output to a standard considered acceptable as a translation [Hutchins and Somers 1992]. During the evaluation, all the errors that occurred in the translation are counted, for example, counting each addition or deletion of a word, or each substitution of one word by another. Different kinds of errors require different correction efforts. Therefore, the errors are classified by relatively difficulty of correction and assigned a weighting factor. For example, the weight factor for incorrect preposition selection in one system is 0.6 while the weight factor for incorrect number agreement (e.g., a computer) is 0.2 (this example is taken from [Arnold 1994]). Then the quality of a translation is determined by calculating the total error score. The total error score for each individual sentence or for the whole text is the sum of all the weighted errors.

Although the error analysis strategy provides more direct information on the quality of the translation, there are problems: differences in what is considered to be an error, different levels of acceptability, and difficulties in classifying errors. The more serious problem is that error types are not independent of each other and it is then difficult to count errors accurately e.g., failure to supply any number inflection for a main verb will often mean that the subject and verb do not agree in number as required [Arnold et al. 1994]. Should this failure be counted as one error or two errors?

8.2.2 Test Suite

A test suite is a useful evaluation tool for NLP participants especially for developers. A test suite is a systematic collection of specially constructed linguistic expressions (e.g., sentences), perhaps with associated annotations and descriptions [Fouvry and Balkan 1996]. Figure 8.1 illustrates a fragment of English test suites (this example is taken from [Arnold, et al. 1994]). This fragment lists combinations of modal verbs (e.g., can, may) with negative not.

The test suite may include ill-formed sentences (grammatically unacceptable sentences e.g., John not run) which the parser component of the MT system should recognize as incorrect.

The test suite can be constructed as follows [Estival et al. 1994], [Balkan, Arnold and Fouvry 1995]:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
by collecting the data (test items) from specific text types, actual texts or knowledge bases. This kind of test suite may be limited to specific problems and a specific vocabulary.

- by collecting from a list of linguistic phenomena which is abstracted from any particular application. This type of test suite is referred to as an artificial test suite. In this test suite, the range of test items and the vocabulary used is unbiased by any particular text type, corpus or knowledge base.

<table>
<thead>
<tr>
<th>John runs.</th>
<th>modal auxiliaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>John will run.</td>
<td></td>
</tr>
<tr>
<td>John can run.</td>
<td></td>
</tr>
<tr>
<td>John may run.</td>
<td></td>
</tr>
<tr>
<td>John should run.</td>
<td></td>
</tr>
<tr>
<td>John does not run.</td>
<td>negation (with do-support)</td>
</tr>
<tr>
<td>John not run.*</td>
<td></td>
</tr>
<tr>
<td>John will not run.</td>
<td>negation and modal auxiliaries</td>
</tr>
<tr>
<td>John may not run.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8.1: A fraction of test suites [Arnold et al. 1994]

Test suites present language phenomena in an exhaustive and systematic way starting with the simplest and increasing in complexity. Furthermore, combinations of phenomena can be generated in a controlled fashion as shown in Figure 8.1.

Test suites can be used for diagnostic evaluation or adequacy evaluation depending on their design [Estival et al. 1994]. With test suites, system developers can determine how their system will perform on a range of controlled examples. A number of test suites have been developed for evaluating MT systems [Estival et al. 1994]. For example:

- The English Eurotra "ET-UK" test suite was developed at Eurotra Essex in 1991. It consists of 955 English test sentences for evaluating the Eurotra translation system. This test suite was designed to test only a restricted number of common syntactic phenomena.
- The English-French and French-English Aerospatiale "AS-NL" test suites were constructed in 1991 to evaluate the commercial MT systems. These test suites contain 1440 English test items, and 346 French test items.
- The French-German SYSTRAN test suite was developed at Stuttgart in 1991 for evaluating the SYSTRAN French-German translation system. This test suite contains 853 French test items.
- The IAI test suite was developed at IAI Saarbrucken in 1993. This test suite was developed for evaluating the LOGOS, German to English and French translation system. The IAI test suite contains 2481 German test items.

The AS-NL, SYSTRAN and IAI test suites were designed for both diagnostic and adequacy evaluations whereas the ET-UK test suite was designed for diagnostic and progress evaluations. According to the Report to LRE 62-089 (D-WP1) [Estival et al. 1994], none of these test suites is claimed to be complete. None of these test suites actually tests the full inflectional morphology of a language systematically and none of them includes testing at a semantic level. The data of the AS-NL and ET-UK test suites are partly derived from sentences occurring in a specific corpus while the data of SYSTRAN and IAI test suites are artificially constructed. Information about the actual use of these test suites is not fully provided.

Methodology and tools are provided by Test Suites for Natural Language Processing (TSNLP) for constructing test suites [Fouvry and Balkan 1996]. The test data constructed by this method are general purpose. They are not specific to a particular domain or text type and they focus on syntactic phenomena. These test data have been validated in a testing phase where a parser, a grammar checker and a controlled language checker have been tested. However, apparently there have not yet been tested on MT systems. Therefore, the problem lies in defining an appropriate test suite for evaluating MT systems.

8.2.3 The ARPA MT Evaluation

The Advanced Research Projects Agency (ARPA) conducted the evaluation of three research MT systems (Candide: IBM Thomas Watson Research Laboratory, Pangloss: Carnegie Mellon University and Lingstat: Dragon Systems), six commercial MT systems.

---

1 The progress evaluation compares successive stages of development of a system over a period of time [Fouvry and Balkan 1996].
2 The Report to LRE 62-089 (D-WP1) presents the characteristics of the existing nine test suites. Four test suites were designed for evaluating MT systems (as listed above) whereas another five test suites were designed for other applications of NLP. This report was written by TSNLP [Estival et al. 1994].

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
(Metal: Sietec, Pivot: NEC, Power Translator: Globalink, SPANAM: Pan American Health Organization, SYSTRAN: SYSTRAN, XLT: SOCATRA) and three human translations [White 1994]. To produce the human translation for this evaluation, the expert translators were prohibited from adding extra value to the text. For example, there are no parenthetical explanations of geographical locations or social customs. The ARPA MT Evaluation methodology consisted of a collection of human judgments on the quality of MT and statistical analyses of those judgments. Each translation was evaluated for adequacy, fluency and informativeness. However, the purposes of each evaluation in ARPA are different from those defined in section 8.1. In the ARPA MT evaluation:

- Adequacy determines to what extent the meaning conveyed in an expert human translation is presented in the translation under evaluation. The measurement is that the evaluators compare the translation under evaluation with human translation of the same passage and assign a score to the translation. The scoring scale ranges from 1, *None of the meaning expressed in human translation is expressed in the MT translation*, to 5, *All meaning expressed in human translation is expressed in the MT translation*.

- Fluency determines the intuitive reactions of evaluators to the well-formness and fluency in context of the translation under evaluation. The measurement is that the evaluators assign a score of 1 to 5, where 1 represents *very poor* and 5 represents *excellent*, to the translations.

- Informativeness determines whether a translation under evaluation provides specific information which the evaluators need. Evaluators must answer a test of six questions that determine an informativeness measurement. These questions are based on the information available in the original text.

### 8.3 GRMT Evaluation

GRMT is in the research and development stage which includes design and development of an advanced prototype system, including compiling dictionaries and grammars. Evaluation is then restricted to the testing of processes without consideration of potential environments. We are unable to make detailed statistical comparison between GRMT and other MT systems. A proper and reliable statistical comparison requires very large dictionaries and corpora, the development of which is part of our future research. In addition, comparison to
other systems permit only a *black box*³ approach to evaluation; we do not generally have access to their internal structures or algorithms. In contrast we want to discuss the various characteristics which we found in GRMT. The evaluation presented here, therefore, was conducted to answer the following questions⁴:

- whether the programs written for the system perform in the ways intended.
- whether the output is acceptable as a translation.
- whether the expansions of dictionaries affect translation performance.
- whether the extendibility to new language pairs is difficult.

To determine whether the programs written for the system perform in the ways intended, we used a *diagnostic trace*. The diagnostic trace is a record of the stages through which a program goes to produce output. A sentence is submitted to the system for processing and the results of each stage are displayed (on screen or in print) to enable us to examine the actual operations taking place. On this basis we can discover whether the program is doing what is intended and if it is not, identify where the mistakes are occurring. We tested each component of the system (ALMT, TCE and RI) separately and together as a complete system (GRMT).

To determine whether the output is acceptable as a translation, the test is performed by running a number of sentences against the system and testing the adequacy of the dictionary information and the grammar. The outputs are then compared to two human translations. One translator is a Thai linguist (HMI) and the other translator is a Thai student (HM2) in graduate school. This evaluation concerns both *intelligibility* and *fidelity*. The evaluator render judgments by answering to the following questions:

- Is the translation is grammatically correct according to Thai grammar?
- Does the translation consist of the correct words in Thai?
- Does the translation convey the meaning presented in the human translation?

³ In the black box evaluation, the evaluators have only input/output(s) pairs to work with while it is assumed that the evaluator has access to all the inner working of the system and can inspect intermediate results in the glass box evaluation [King 1990].

⁴ These questions need to be asked and answered for all MT systems; we answer here for GRMT alone.
The outputs are also evaluated in terms of comprehensibility. In this evaluation, the evaluators are a linguist, a grade 7 student and a grade 4 student. All evaluators are native Thai speakers. The measurement is conducted by reading the translation to the evaluators without providing the source text. The evaluator then determines:

- whether the translation can be understood: if the translation cannot be understood what is the cause of the misunderstanding, e.g., inappropriate word usage, inappropriate word order, or missing word.
- whether it is colloquial Thai.

ALMT was tested to generate translation candidates for 76 sentences from the bedtime story, The ugly duckling, version written by Ronne Randall (1987) based on the original story by H. C. Anderson), and 14 sentences from other sources including, newspapers and articles. The outputs of ALMT (translation candidates) are compared to the two human translations to determine their intelligibility and fidelity. The words used in the TCs of the 23th and 27th sentences (Figure 8.2) are correct. The 23th and 27th TCs are licensed by the Thai grammar and convey the same meaning as presented in the human translations. The 26th sentence carries no meaning since the word where is incorrectly translated, ที่ไหน (thiina'- where), and it results in a grammatically incorrect TC. In the 37th sentence, หนึ่ง (ny'ri- one) is wrongly placed and the translation of the word to, ที่ (kab- to), is incorrect. The misplaced word and incorrect translation render the TC grammatically incorrect. A complete evaluation of all 90 sentences is illustrated in Appendix G. According to the evaluation results shown in Table 8.1: 23 sentences require word re-selection, 15 sentences require word re-arrangement and 47 sentences require no repair. The number of incorrectly translated words in each sentence ranges between 1-3 words. The number of misplaced words which results in a grammatically incorrect translation ranges between 1-3 words.

Figure 8.3 illustrates a fragment of the comprehensible evaluation. A complete evaluation is provided in Appendix H. 57 sentences, Table 8.2, are comprehensible (e.g., sentence 14, Figure 8.3), however, two of them do not convey the intended meaning (e.g., sentence 51, Figure 8.3). The reason the translation does not convey the intended meaning

---

5 The grade 7 and grade 4 students were selected as evaluators because the main part of the source text is a children's bedtime story.
23. SL: The ugly duckling hid in a corner of the farmyard and cried.
   TC: ตุกแก้ว (duckling) ที่ (which) ซ่อน (ugly) ตัว (classifier) นั่ง (the) ใต้ชุด (hid) ใน (in) มุม (corner) มุม (classifier) หนึ่ง (a) ของ (of) ลานดินโฉนด (farmyard) แห่ง (classifier) นั่ง (the) และ (and) ไตร่.TextField (cried)
   HM1: ตุกแก้ว (duckling) ซ่อน (hide) อยู่ (combine clauses) มุม (corner) หนึ่ง (a) ใน (of) ฟาร์ม (farmyard) และ (and) ร้องไห้ (cry)
   HM2: ตุกแก้ว (duckling) ที่ (which) นายเกลอ (ugly) เลนชมองตัว (hid) ที่ (combine clauses) มุม (corner) หนึ่ง (a) ของ (of) ฟาร์ม (farmyard) และ (then) ร้องไห้ (cry)
   No repair required

26. SL: He left the farmyard and ran to the great marsh, where the wild ducks lived.
   TC: เขา (he) ออกไปจาก (left) ลานดินโฉนด (farmyard) แห่ง (classifier) นั่ง (the) และ (and) ได้รับ (ran) ไปที่ (to) หนึ่งมุม (marsh) ใหญ่ (great)
   หนึ่ง (the) ที่ (where) เป็น (duck) บาง (wild) หลาย (plurality) ตัว (classifier) นั่ง (the) ใกล้ที่อยู่ (lived)
   HM1: มัน (it) ออกไปจาก (leave) ฟาร์ม (farmyard) และ (and) วิ่งไปข้าง (ran to) ป่า (marsh) ใหญ่ (great) แห่ง (classifier) หนึ่ง (a) ที่ (where) มี (there are)
   เป็น (duck) บ้าง (wide) อาศัยอยู่ (lived)
   HM2: มัน (it) ออกไปจาก (leave) ฟาร์ม (farmyard) และ (then) วิ่ง (run) ไปที่ (to) หนึ่งมุม (marsh) ใหญ่ (great) ที่ (where) เป็น (duck) บาง (wild) อาศัยอยู่ (live)
   1 incorrectly translated word.

27. SL: “Perhaps they will be kind to me,” he said sadly.
   TC: "บางที (perhaps) เพราะ (they) จะ (will) ให้ (kind) กัน (to) ฉัน (me)," เขา (he) ได้พูด (said) อย่างเศร้าสลด (sadly)
   HM1: "บางที (perhaps) พวกนั้น (that group) อาจจะ (may) ติ่ง (kind) กัน (to) ฉัน (me) ให้ (??)?" มัน (it) พูด (said) กัน (to) ตัวเอง (itself) อย่างเศร้าสลด (sadly)
   HM2: "บางที (perhaps) เพราะ (they) อาจจะ (may) ติ่ง (kind) กัน (to) เราย่ (me) ให้ (??)?" มัน (it) พูด (said) ด้วยความเศร้าสลด (sadly)
   No repair required.

37. SL: One evening, he came to a little cottage.
   TC: หนึ่ง (one) เอน (evening), เขา (he) ไปมา (came) กิน (to) กระหรอม (cottage) เล็ก (little) แห่ง (classifier) หนึ่ง (a)
   HM1: เข้า (evening) วัน (classifier) หนึ่ง (one) ตุกแก้ว (duckling) ซ่อน (ugly) เลน (walk) มา (come) กิน (to) กระหรอม (cottage) เล็ก ๆ (little) แห่ง (little) หนึ่ง (a)
   HM2: เอน (evening) วัน (classifier) หนึ่ง (one) มัน (it) มา (come) กิน (to) กระหรอม (cottage) เล็ก (little) แห่ง (classifier) หนึ่ง (a)
   1 misplaced words, 1 incorrectly translated word.

Figure 8.2: Examples of intelligibility and fidelity evaluation
Table 8.1: The intelligibility and fidelity evaluation result of evaluating ALMT on the 90 sentences test corpus

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Number of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>grammatically correct, correct word usage and convey the original meaning - require no repair</td>
<td>47</td>
</tr>
<tr>
<td>required repair</td>
<td></td>
</tr>
<tr>
<td>grammatically correct, correct word usage but convey different meaning from the original sentence</td>
<td>2</td>
</tr>
<tr>
<td>inappropriate word usage</td>
<td>4</td>
</tr>
<tr>
<td>incorrectly translated word</td>
<td>15</td>
</tr>
<tr>
<td>incorrectly translated words</td>
<td>7</td>
</tr>
<tr>
<td>incorrectly translated words</td>
<td>1</td>
</tr>
<tr>
<td>misplaced word</td>
<td>8</td>
</tr>
<tr>
<td>misplaced words</td>
<td>4</td>
</tr>
<tr>
<td>misplaced words</td>
<td>1</td>
</tr>
<tr>
<td>grammatically incorrect</td>
<td>9</td>
</tr>
</tbody>
</table>

is that some words are incorrectly translated. 33 sentences are incomprehensible. According to the evaluation results (Table 8.2), the reasons the translations are not comprehensible are:

- Some words are missing. These words need to be added to make the sentence complete and correct in terms of meaning and grammar e.g., sentence 6 (Figure 8.3).
- Some words are incorrectly translated. These words may cause grammatical incorrectness which results in a meaningless sentence e.g., sentence 57 (Figure 8.3).
- Some words are misplaced. The word is wrongly placed and it causes the sentence to be meaningless e.g., sentence 30 (Appendix H).

Table 8.2: The comprehensible evaluation result

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Number of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>comprehensible</td>
<td>56</td>
</tr>
<tr>
<td>incomprehensible</td>
<td>34</td>
</tr>
<tr>
<td>not colloquial Thai</td>
<td>9</td>
</tr>
<tr>
<td>does not convey the intend meaning</td>
<td>2</td>
</tr>
</tbody>
</table>
6. SL: The biggest egg hatched last of all
TC: ไข่ (egg) ที่ (which) ใหญ่สุด (biggest) ได้พัก (hatched) ต่ำสุด (last) ของ (of) ทั้งหมด (all)
HM3: incomprehensible, 1 additional word required.
HM4: incomprehensible, 2 additional words required.
HM5: incomprehensible, 2 additional words required.

14. SL: But when they saw the big gray duckling, they began to laugh.
TC: แต่ (but) เมื่อ (when) พ่อแม่ (they) ได้เห็น (saw) ตัวเป็ด (the) สีเทา (gray) ได้ (big) ตัว (classifier) นั้น (the), พ่อแม่ (they) ได้ริเริ่ม (began)
ที่จะ (to) หัวเราะ (laugh)
HM3: comprehensible.
HM4: comprehensible, 1 additional word required.
HM5: comprehensible.

51. SL: He grew more and more sad and more and more tired
TC: เขา (he) ได้เจริญ (grew) ยิ่ง (more) และ (and) เหงา (sad) มากขึ้น (more) และ (and) มệt (tired) มากขึ้น (more)
HM3: comprehensible, does not convey the intended meaning, not colloquial Thai.
HM4: comprehensible, does not convey the intended meaning, not colloquial Thai.
HM5: comprehensible, does not convey the intended meaning.

57. SL: The ugly duckling watched them from his hiding place in the reeds.
TC: ตัวเป็ด (duckling) ที่ (which) รูปร่าง (ugly) ตัว (classifier) นั้น (the) ได้เห็น (watched) พ่อแม่ (them) จาก (from) สถานที่ (place) เก็บตัว (hiding)
ของเขานั้น (his) ใน (in) ต้น (reed) หลาย (plurality) ตัว (classifier) นั้น (the)
HM3: incomprehensible, 1 meaningless word.
HM4: incomprehensible, 1 meaningless word.
HM5: comprehensible

Figure 8.3: Examples of comprehensibility evaluation
6. SL: The biggest egg hatched last of all.
   TC: ไข่ (egg) ที่ (which) ใหญ่ที่สุด (biggest) ได้พัก (hatched) ล่าสุด (last) ของ (of) ทั้งหมด (all)
   Expected TC: ไข่ (egg) ที่ (which) ใหญ่ที่สุด (biggest) ได้พัก (hatched) เป็น (to be) ชนิด (classifier) ล่าสุด (last) ของ (of) ทั้งหมด (all)

10. SL: You are the ugliest duckling I have ever seen.
    TC: คุณ (you) เป็น (are) นกแก้ว (duckling) ที่ (which) ที่สุด (ugliest) เคย (have ever) เห็น (seen)
    Expected TC: คุณ (you) เป็น (are) นกแก้ว (duckling) ที่ (which) นกแก้วที่สุด (ugliest) ที่ (that) เคย (have ever) เห็น (seen)

21. SL: The geese and the turkeys ran after the duckling and tried to chase him away.
    TC: กาชาด (geese) และ (and) ไก่พิราบ (turkey) หลาย (plurality) ตัว (classifier) นั้น (the) ได้ไปตาม (tried) ลูกแก้ว (duckling) ตัว (classifier)
    Expected TC: กาชาด (geese) และ (and) ไก่พิราบ (turkey) หลาย (plurality) ตัว (classifier) นั้น (the) ได้ไปตาม (tried) ลูกแก้ว (duckling)

22. SL: Even his own brothers and sisters joined in and called him names.
    TC: แม่ค้า (even) พี่ชายของชาย (brother) หลาย (plurality) คน (classifier) ของเขา (his) เอง (own) และ (and) พี่สาวของเขา (sister) หลาย (plurality) คน (classifier) เข้ามา (joined_in) และ (and) เข้ามา (called_him_names)
    Expected TC: แม่ค้า (even) พี่ชายของชาย (brother) หลาย (plurality) คน (classifier) ของเขา (his) เอง (own) และ (and) พี่สาวของเขา (sister)

Figure 8.4: Example sentences 6, 10, 21 and 22.
Of the 43 sentences which required repair (Table 8.1), 14 sentences are comprehensible and convey the meaning of the original sentences.

We are interested in evaluating the outputs of GRMT (translations) in an analogous way to the ALMT evaluation just discussed. Of the 43 sentences which required repair, 15 sentences are repairable using the current our HPSG based grammars and lexicons. 7 sentences cannot be repaired by the current RI. They are grammatically incorrect although they contain correctly translated words. To complete the TC, according to the Thai grammar, some words (e.g., classifier or linking word) need to be added while some words (e.g., plurality or classifier) need to be removed. In sentences 10, 21, 22, 58 and 81 (Appendix H), the linking word ใ (thii- that, which, who), ใ (kə-also, too) and ใ (haj- to cause to do something) are required as shown in Figure 8.4. Sentence 6 requires ใ (pen- to be) while the words ใ (la-aj- plurality) and ใ (tua-classifier) in sentence 55 need to be removed. The linking word in Thai can be classified into at least 17 categories [Punmetha 1984] e.g., ใ (thii- that, which, who) is a linking word used to add an instrumental relationship as in “He lives in the house which is big”. ใ (kə-also, too) and ใ (than- and also, together with) are linking words using to combine two sentences which are in agreement as in “If you go then I will also go”. ใ (haj- to cause to do something) is a linking word to indicate the purpose of the preceding action, as in “He give the book to me to make me happy”. There is no explicit rule saying when and which linking word is required.

As we discussed in section 4.1, there is no plural form for nouns in Thai. A quantity of nouns can be expressed either by quantifiers or numbers and classifiers are required. However, adding the plurality ใ (la-aj) and ใ (tua-classifier) to the translation of the word bird in sentence 55 (Figure 8.5) is redundant and grammatically incorrect since flock indicates that there is more than one bird. It is unlike English in which flock requires a plural noun form.

<table>
<thead>
<tr>
<th>a flock of beautiful birds</th>
</tr>
</thead>
<tbody>
<tr>
<td>ใ (bird) ใ (which) ใ (beautiful) ใ (of) ใ (flock) ใ (classifier) ใ (a) ใ (plurality)</td>
</tr>
</tbody>
</table>

Figure 8.5: A fragment of sentence 55.

---

The conjunction ใ (thii- that, which, who), ใ (kə-also, too) and ใ (haj- to cause to do something) often are not translated in English.
We will be expanding the *word addition* (section 5.3.1) rule to cover additional linking words and to add a plurality indicator and classifier (only when necessary) as part of future research.

The remaining 21 sentences of the 90 sentence test corpus require further nontrivial development of the HPSG formalism to account for their semantic representation, for example, sentences containing the logical connections *and, or, but, nor* or the negation *not*. Due to this limitation, the sentences which contain logical connections are separated into linguistic units as illustrated in Figure 8.6 (each unit is underlined). In sentence 26, the word *he* (circled) was added to complete the structure of the second unit. Each linguistic unit is then repaired separately so that we can determine whether the TCE and RI phases improve the quality of the TC. Some example sentences are illustrated in Figure 8.6; the repaired TCs are grammatically correct and convey the meaning of the original sentences.

We will be expanding the lexicon and the language coverage of our HPSG parsers as part of future research and expect that outcomes of that effort to contribute to both the novel MMT paradigm GRMT espouses and to HPSG theory, e.g., the semantics of coordination.

In the first version of ALMT, we developed dictionaries containing 80 English words and 150 Thai words. The dictionaries were expanded to 348 English words and 731 Thai words in the second version of ALMT. There has been no appreciable increase in processing time when running ALMT using the larger dictionaries. This experiment lends positive evidence that a more advanced prototype will scale up reasonably.

Modification to the lexicons and grammars of the SL and TL analyzers will result in improvements. Expanding the lexicons or the coverage of language structure via HPSG formalism is relatively simple. The use of the HPSG formalism provide advantages in terms of comprehensive lexical entries. The lexicons are easy to manage. The use of ALE facilitates lexicon development and grammar writing.

Expansion in terms of adding new languages to the system is manageable since SLs and TLs are treated separately and they are related by using information provided in a SL-TL lexicon. The lexicon and grammar used in the analyzer of each language is developed separately.
26. SL: He left the farmyard and ran to the great marsh where the wild ducks lived.
   Repaired TC: เขา (he) ให้จาก (left) ลานติดโรงเรือน (farmyard) นั้น (the) และ (and) ไปรับ (ran) ไปที่ (to) หุบเขา (great) นั้น (the) ที่นั่น (where) เป็น (duck) น้ำ (wild) หลาย (plurality) ตัว (classifier) นั้น (the) ได้อาศัยอยู่ (lived)
   HM1: มี ออกจาก บ้าน แล้ว ตรงไปบัง บึง ใหญ่ แห้ง หนึ่ง ที่ มี เบ็ด น้ำ อาศัยอยู่
   HM2: มี ออกจาก บ้าน แล้ว รั่งไป ที่ ห่างเกือบทุก ๆ ที่ เบ็ด น้ำ อาศัยอยู่
   No repair required.

39. SL: An old woman lived in the cottage with a fat black cat and a plump brown hen.
   TC: ผู้หญิง (woman) คน (old) คน (classifier) นั้น (an) ได้อาศัยอยู่ (lived) ใน (in) กระท่อม (cottage) นั้น (the) ด้วย (with) แมว (cat) ตัว (classifier) นั้น (the) นัก (fat) ตัว (classifier) นั้น (a) และ (and) ไก่ (hen) สีน้ำตาล (brown) นอ (plump) ตัว (classifier) นั้น (a)
   HM1: มี หญิงสาว อาศัยอยู่ ใน กระท่อม กลาง อยู่ กิน แมว ตัว อวบนี่ สีดำ และ แม่ ไก่ ใหญ่ สีน้ำตาล
   HM2: หญิงสาว อาศัยอยู่ ใน กระท่อม นั่น แมว สีดำ ตัว อวบนี้ และ ไก่ สีน้ำตาล ตัว อวบนี้
   No repair required.

90. SL: The spirit is willing but the flesh is weak.
   Repaired TC: จิตวิญญาณ (spirit) นั้น (the) พร้อม (willing) เดิน (but) ร่างกาย (flesh) นั้น (the) ย่อมแย่ (weak)
   HM1: จิตวิญญาณ นั้น กลางกร่าง หา กาย กลิ่น ย่อมแย่
   HM2: จิตวิญญาณ นั้น พร้อม หา ร่างกาย กลิ่น ย่อมแย่
   No repair required.

Figure 8.6: Example sentences 26, 39 and 90
Chapter 9

Discussion

9.1 Contributions

The central theme of this thesis is to find new solutions that apply when developing an efficient MT and our solution is *Generate and Repair Machine Translation*. GRMT is a novel multi-phase architecture for an accurate machine translation system. The main concept of GRMT was introduced in chapter 3. The architecture of GRMT was designed to take advantage of, and have advantages over, the Direct, Transfer, Interlingual, Non-linguistic and the existing hybrid approaches to MT as illustrated in Table 9.1. GRMT consists of three major components Analysis Lite MT, Translation Candidate Evaluation and Repair and Iteration. ALMT generates the translation candidate, TCE verifies an accuracy of the TC and RI repairs the TC when required.

The main features of ALMT are the use of constraints and the use of the semantic relationship between words. ALMT applies the different sets of constraints to deal with the syntax of the SL and that of the TL in separate modules. ALMT refines the scope of translation choices of each input word by applying the SL constraints. SL constraints apply to the syntax of the SL which is different from that of the TL. The selected translation words are put into grammatical order by applying the TL constraints. TL constraints apply the syntax of the TL which is different from that of the SL. The set of constraints can be applied to languages which share common significant features.

The semantic relationship between words has been developed in the form of the WordAssociation relation. The WordAsso relation is a link between WordAsso numbers. The WordAsso number represents a relation between a word and its class. We classified words according to their meanings and their usages regardless of language which makes it possible to
Table 9.1: Advantages and disadvantages of the existing MT approaches

<table>
<thead>
<tr>
<th>Category</th>
<th>Approach</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three Classic</td>
<td>Direct MT</td>
<td>Simple approach.</td>
<td>Limited accuracy</td>
</tr>
<tr>
<td>Strategies</td>
<td>Interlingual MT</td>
<td>Treat SL and TL separately.</td>
<td>Difficulties in developing interlingual system.</td>
</tr>
<tr>
<td></td>
<td>Transfer MT</td>
<td>Less complexity in analysis and generation.</td>
<td>• Information lost during the transfer process.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Language dependent.</td>
</tr>
<tr>
<td>Nonlinguistic</td>
<td>Example-Based MT</td>
<td>Translations come with scores which might be useful to a post editor.</td>
<td>Relies heavily on the availability of a large and diverse quality corpus which are not available for most languages.</td>
</tr>
<tr>
<td>Information</td>
<td>Statistics-Based MT</td>
<td>Apply statistical techniques in translation process.</td>
<td>• Reasoning is complex.</td>
</tr>
<tr>
<td>Strategies</td>
<td>Knowledge-Based MT</td>
<td>Interpreting the SL by reference to world knowledge.</td>
<td>• More suitable for paraphrase than for translation</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Shake-and-Bake MT</td>
<td>SL and TL are associated via bilingual lexicon without a transfer module or interlingual.</td>
<td>Lack efficiency due to the generation process which tries all possibilities.</td>
</tr>
</tbody>
</table>

to apply these to a multilingual MT system. This word classification plays a crucial role throughout the GRMT translation process.

The use of the WordAsso relation together with the designed (word selection) algorithm allows ALMT to select an appropriate translation for each input word efficiently. The WordAsso relation indicates a relationship between words (e.g., which word can occur with which word in the same expression) including the relationship between nouns and their classifiers. The use of classifiers in Asian languages e.g., Chinese, Japanese and Thai is similar, therefore, with some modifications it is possible to apply our classifier selection to those Asian languages.

GRMT ensures an accuracy of the translation by the TCE and RI phases. TCE analyzes the TC to verify its syntax and semantics. The semantics of the TC is extracted and compared to that of the SL to determine whether the TC conveys the meaning of the SL.

143

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The SL and TL grammars should be developed based on grammar theory which is an integrated theory of natural language syntax and semantics, e.g., the Head-Driven Phrase Structure Grammar. We include the WordAsso number in the semantic representation since the WordAsso number represents the word meaning and the word’s usage. Differences in languages and cultures result in different circumstances of language usage; therefore, the word and its translation are considered acceptably similar in principal features if they belong to the same superclass as illustrated in chapter 6. If TCE detects any dissimilarity between the meaning of the TC and that of the SL, TCE will then provide that (dissimilarity) information to RI.

RI improves the quality of the TC by replacing the mis-translated part with a more appropriate translation according to the information TCE provides.

Each step performed by GRMT is straightforward and simple to implement. GRMT is highly modular and extendible in the following respects. Each component, ALMT, TCE and RI are developed separately. Each module comprises of sub-modules for easy modification and maintenance. The SL and TL are treated separately and related via the SL-TL dictionary. The SL-TL dictionary contains simple information, a SL word form and its possible translations, with the WordAsso number, and nothing else. Other knowledge-bases e.g., the dictionary, constraints, lexicons, and grammars are developed separately, independent of languages. This would seem to be the optimal distribution of information, both from the point of building a larger, possibly multilingual MT system, and for maintaining and extending it. A further note is that GRMT translation process is not limited to a single direction translation. With a few modifications of the SL-TL dictionary, GRMT is reversible.

I hope to have demonstrated the idea of GRMT by constructing the English-Thai MT system. The English-Thai MT system translates isolated sentences (sentence by sentence). This English-Thai translation system has been developed and run under SWI-Prolog 3.2.8. The English and Thai grammars have been developed based on the Head-Driven Phrase Structure Grammar formalism [Pollard and Sag 1987; Pollard and Sag 1994] and implemented on the Attribute Logic Engine (ALE) [Carpenter and Penn 1994].

This English-Thai MT system was evaluated and it performs in the way we intended. ALMT generated acceptable translations (grammatically correct, correct word usage and convey the original meaning) for 47 out of the 90 sentence test corpus without repair. TCE and RI improved 15 sentences using the current our HPSG based grammars and lexicons.
Twenty-one sentences which contain logical connections are first separated into linguistic units before the repair can be performed due to a current inadequacy in HPSG's semantic representation. However, each linguistic unit was then repaired successfully. Seven sentences faced with the problems of adding linking words and classifiers in Thai also require further refinements in order to repair such sentences.

9.2 Further Research

Several issues presented in this thesis require further research. The word addition module can be improved since it currently fails to account properly for some connecting words and some plurality indicators in Thai (see chapter 8). Further research is needed to determine when and which connecting word or plurality indicator is required. With the word classification provided, it is possible to select a proper linking word for each situation and to be able to identify when the plurality indicator is needed.

To develop the WordAsso relations in the future, the semantic concordance which was created at the Princeton Cognitive Science Laboratory can be used as sample set. A semantic concordance is a database combining a text and a lexicon so that words in the text are linked to the appropriate senses in the lexicon [Fellbaum 1998]. We also hope consider how statistical methods or machine learning techniques can be used to help assign the WordAsso numbers to words in an appropriate way.

The lexicon and grammar coverage is limited. When large dictionaries (SL-TL dictionary and SL dictionary) and lexicons are used, more efficient ways of indexing the entries are required. At the moment the dictionaries and the lexicons are held in Prolog's working memory, an approach that is unlikely to satisfy requirements for larger systems.

Additional research is also required to explore the semantic representation of some language structures, e.g., structures containing logical connections and negation. This additional research should be useful not only for GRMT but also for another natural language processing applications e.g., search engines and information retrieval.

I believe the results of this research will contribute to current attempts to develop accurate and reliable machine translation systems and to produce quality translations from one language to another. This accurate and reliable translation methodology should enhance the effectiveness of communication among people.
Nevertheless, I hope to have shown that pursuing further research in this direction is a worthwhile aim, and one likely to result in commercial machine translation systems in the not too distant future.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


154

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


155

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Sag, I. A. and T. Wasow. 1999 Syntactic Theory: A Formal Introduction, CSLI Lecture Notes Number 92, Center for the Study of language and Information, Stanford, California.


Appendix A

WordAssociation Hierarchy

This appendix illustrates the WordAssociation hierarchy, discussed in section 4.2.1. Class and its WordAsso number are shown in the first column. Examples of each are shown in the second column.

A.1 The IS-A Classes

The semantic relation between classes and their superclasses is represented by *is-a*.

<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Thing (Entity)</td>
<td></td>
</tr>
<tr>
<td>1-1 Concrete thing</td>
<td></td>
</tr>
<tr>
<td>1-1-1 Living thing</td>
<td></td>
</tr>
<tr>
<td>1-1-1-1 Human</td>
<td></td>
</tr>
<tr>
<td>1-1-1-1-1 Individual (Person)</td>
<td></td>
</tr>
<tr>
<td>1-1-1-1-1-1 Male</td>
<td>adult child</td>
</tr>
<tr>
<td>1-1-1-1-1-2 Female</td>
<td>boy man</td>
</tr>
<tr>
<td>1-1-1-1-1-3 Relative</td>
<td>cousin aunt</td>
</tr>
<tr>
<td>1-1-1-1-1-3-1 Ancestor</td>
<td>parent mother father grandma แม่</td>
</tr>
<tr>
<td>1-1-1-1-1-3-2 Male sibling</td>
<td>brother พี่ชายน้องชาย</td>
</tr>
<tr>
<td>1-1-1-1-1-3-3 Female sibling</td>
<td>sister ภค.สาวน้องสาว</td>
</tr>
<tr>
<td>1-1-1-1-1-3-4 Spouse</td>
<td>husband wife สามี ภรรยา</td>
</tr>
<tr>
<td>1-1-1-1-1-3-5 Offspring</td>
<td>ลูก ลูกที่หนึ่ง ลูกชาย</td>
</tr>
<tr>
<td>1-1-1-1-1-3-5-1 Male offspring</td>
<td>ลูกชาย</td>
</tr>
<tr>
<td>1-1-1-1-1-3-5-2 Female offspring</td>
<td>ลูกสาว</td>
</tr>
<tr>
<td>1-1-1-1-1-4 Juvenile</td>
<td>หนูน้อย young children เด็กๆ</td>
</tr>
</tbody>
</table>

1 This classification is for associating with the ending in Thai ค่ะ ครับ
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1-1-1-5 Associate</td>
<td>member ต้องการมาจาก  brother</td>
</tr>
<tr>
<td>1-1-1-1-6 Helper</td>
<td>supporter ที่ได้รับสนับสนุน  friend</td>
</tr>
<tr>
<td>1-1-1-1-7 Socialist</td>
<td></td>
</tr>
<tr>
<td>1-1-1-1-7-1 Communist</td>
<td>red คอมมิวนิست</td>
</tr>
<tr>
<td>1-1-1-1-8 Unpleasant person</td>
<td>worm ไม่ดีเค่าน</td>
</tr>
<tr>
<td>1-1-1-2 Animal</td>
<td>สัตว์</td>
</tr>
<tr>
<td>1-1-1-2-1 Vertebrate</td>
<td></td>
</tr>
<tr>
<td>1-1-1-2-1-1 Mammal</td>
<td>cat แมว และ sheep</td>
</tr>
<tr>
<td>1-1-1-2-1-2 Bird</td>
<td>bird สัตว์น้ำ</td>
</tr>
<tr>
<td>1-1-1-2-1-2-1 Aquatic Bird</td>
<td>นกน้ำ นกกระท geliştirก  goos สีน้ำตื้น</td>
</tr>
<tr>
<td>1-1-1-2-1-2-2 Fowl</td>
<td>duck นกที่บิน  goos สีน้ำตื้น</td>
</tr>
<tr>
<td>1-1-1-2-1-3 Aquatic Vertebrate</td>
<td></td>
</tr>
<tr>
<td>1-1-1-2-1-3-1 Fish</td>
<td>ปลาทะเล</td>
</tr>
<tr>
<td>1-1-1-2-1-4 Amphibian</td>
<td>frog งู</td>
</tr>
<tr>
<td>1-1-1-2-1-5 Reptile</td>
<td>งู งูเหล่า</td>
</tr>
<tr>
<td>1-1-1-2-2 Invertebrate</td>
<td>worm ไม่ดีเค่าน</td>
</tr>
<tr>
<td>1-1-1-2-2-1 Arthropod</td>
<td>spider</td>
</tr>
<tr>
<td>1-1-1-2-2-1-1 Insect</td>
<td>butterfly</td>
</tr>
<tr>
<td>1-1-1-2-2-1-2 Crustacean</td>
<td>shrimp, crab เศษ</td>
</tr>
<tr>
<td>1-1-1-2-2-2 Shellfish</td>
<td>mussel</td>
</tr>
<tr>
<td>1-1-1-3 Plant</td>
<td></td>
</tr>
<tr>
<td>1-1-1-3-1 Vascular Plant</td>
<td></td>
</tr>
<tr>
<td>1-1-1-3-1-1 Woody Plant</td>
<td></td>
</tr>
<tr>
<td>1-1-1-3-1-1-1 Tree</td>
<td>maple</td>
</tr>
<tr>
<td>1-1-1-3-1-1-1 Palm</td>
<td>coconut, nipa plam,  ذات</td>
</tr>
<tr>
<td>1-1-1-3-1-1-2 Bush</td>
<td>rose, hydrangea, jasmine ดก</td>
</tr>
<tr>
<td>1-1-1-3-1-2 Herb</td>
<td>lettuce นักกีฬาที่  carrot  taro</td>
</tr>
<tr>
<td>1-1-1-3-1-2-1 Gramineous plant</td>
<td></td>
</tr>
<tr>
<td>1-1-1-3-1-2-1-1 Cereal</td>
<td>cauliflower, broccoli, vegetable ปี</td>
</tr>
<tr>
<td>1-1-1-3-1-3 Vine</td>
<td>reed ต้นก้า</td>
</tr>
<tr>
<td>1-1-1-3-1-3-1 Gourd</td>
<td>wheat ข้าวสาร</td>
</tr>
<tr>
<td>1-1-1-3-2 Fungus</td>
<td>potato ถั่ว, grape</td>
</tr>
<tr>
<td>1-1-1-4 Microorganism</td>
<td>cucumber, ผัก white gourd ผัก</td>
</tr>
<tr>
<td>1-1-2 Non-living thing</td>
<td>mushroom ต้นทุ่ง</td>
</tr>
<tr>
<td>1-1-2-1 Object</td>
<td>seaweed</td>
</tr>
<tr>
<td>1-1-2-1-1 Artifact (Man-made object)</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1 Fabric</td>
<td>duck น้ำผัก</td>
</tr>
<tr>
<td>1-1-2-1-1-2 Construction</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-2-1 Barrier</td>
<td>hatch นกที่บิน  home ประตู</td>
</tr>
<tr>
<td>1-1-2-1-2-1-1 Movable barrier</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-2-2 Housing</td>
<td>cottage บ้าน  home บ้าน</td>
</tr>
<tr>
<td>1-1-2-1-2-3 Room</td>
<td>toilet can ห้องน้ำ</td>
</tr>
<tr>
<td>WordAsso Number and Class</td>
<td>Examples</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>1-1-2-1-2-4 Establishment(^2)</td>
<td>prison, can, cùc</td>
</tr>
<tr>
<td>1-1-2-1-2-4-1 Institution(^3)</td>
<td>school, university, โรงเรียน</td>
</tr>
<tr>
<td>1-1-2-1-2-4-1-1 Penal Institution(^4)</td>
<td>hospital, โรงพยาบาล</td>
</tr>
<tr>
<td>1-1-2-1-2-4-1-2 Educational Insti.</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-2-4-1-3 Medical Insti.</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-3 Instrumentation</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-3-1 Device</td>
<td>รั้ง, ตั้ง</td>
</tr>
<tr>
<td>1-1-2-1-1-3-1-1 Holding device</td>
<td>last, นอกนิ้ว</td>
</tr>
<tr>
<td>1-1-2-1-1-3-1-2 Musical instrument</td>
<td>reed, ฟืน</td>
</tr>
<tr>
<td>1-1-2-1-1-3-1-3 Airfoil</td>
<td>ปีก, ท้าน</td>
</tr>
<tr>
<td>1-1-2-1-1-3-1-4 Elastic device</td>
<td>spring, สายยาง</td>
</tr>
<tr>
<td>1-1-2-1-1-3-1-5 Measuring instrument</td>
<td>watch, นาฬิกา</td>
</tr>
<tr>
<td>1-1-2-1-1-3-2 Container</td>
<td>กล่อง</td>
</tr>
<tr>
<td>1-1-2-1-1-3-2-1 Box</td>
<td>safe, ผ้าห่ม</td>
</tr>
<tr>
<td>1-1-2-1-1-3-3 Implement</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-3-4 Means</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-4 Enclosure</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-5 Covering</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-5-1 Protective covering</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-5-1-1 Blind</td>
<td>curtain, ม่าน</td>
</tr>
<tr>
<td>1-1-2-1-1-5-1-2 Binding</td>
<td>cover, ผ้า</td>
</tr>
<tr>
<td>1-1-2-1-1-5-2 Clothing</td>
<td>coat, เสื้อ</td>
</tr>
<tr>
<td>1-1-2-1-1-5-1-1 Blind</td>
<td>ผ้าห่ม</td>
</tr>
<tr>
<td>1-1-2-1-1-5-1-2 Binding</td>
<td>ผ้า</td>
</tr>
<tr>
<td>1-1-2-1-1-6 Strip</td>
<td>stay</td>
</tr>
<tr>
<td>1-1-2-1-1-7 Creation</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-1-7-1 Representation</td>
<td>reflection, ภาพสะท้อน</td>
</tr>
<tr>
<td>1-1-2-1-1-7-1-1 Picture/Image</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-2 Substance(^5)</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-2-1 Food</td>
<td></td>
</tr>
<tr>
<td>1-1-2-1-2-1-1 Nutriment</td>
<td>กํ้า, นิ้ว</td>
</tr>
<tr>
<td>1-1-2-1-2-1-1-1 Dish</td>
<td>salad, ผัด</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2 Foodstuff(^6)</td>
<td>egg, ไข่</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-1 Vegetable</td>
<td>cucumber, ผัก, แถ่</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-1-1 Root vegetable</td>
<td>carrot, แตง俄国, potato</td>
</tr>
</tbody>
</table>

\(^2\) Establishment is a public or private structure (business or government or educational) including buildings and equipment for business or residence.

\(^3\) Institution is a building or complex of buildings where an organization for the promotion of some cause is situated.

\(^4\) Penal Institution is an institution where persons are confined for punishment and to protect the public.

\(^5\) Substance is that which has mass and occupies space.

\(^6\) A substance that can be used or prepared for use as food.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1-2-1-2-1-2-1-2 Leafy vegetable</td>
<td>lettuce, ผักใบWISE</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-1-3 Crucifer</td>
<td>cauliflower, broccoli</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-2 Fruit</td>
<td>shrimp, ปู, ปลาหมึก</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-3 Seafood</td>
<td>shrimp, ปู, ปลาหมึก</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-4 Meat</td>
<td>duck, hen, ไก่, ไก่, ตุ๊กแย่ง</td>
</tr>
<tr>
<td>1-1-2-1-2-1-2-5 Grain</td>
<td>wheat, ข้าวสาร</td>
</tr>
<tr>
<td>1-1-2-1-2-1-3 Beverage</td>
<td>spirit, แอลกอฮอล์, whisky</td>
</tr>
<tr>
<td>1-1-2-1-2-2 Chemical Element</td>
<td>hydrogen, oxygen</td>
</tr>
<tr>
<td>1-1-2-1-2-2-1 Metal</td>
<td>copper, silver</td>
</tr>
<tr>
<td>1-1-2-1-2-3 Material</td>
<td>soil</td>
</tr>
<tr>
<td>1-1-2-1-2-3-1 Earth</td>
<td>soil</td>
</tr>
<tr>
<td>1-1-2-1-2-3-3 Animal material</td>
<td>hide, หนังสัตว์</td>
</tr>
<tr>
<td>1-1-2-1-2-3-4 Natural fiber</td>
<td>flock, เส้นใย</td>
</tr>
<tr>
<td>1-1-2-1-2-4-1 Liquid</td>
<td>water, น้ำ</td>
</tr>
<tr>
<td>1-1-2-1-2-4-2 Gas</td>
<td>hydrogen, oxygen, air</td>
</tr>
<tr>
<td>1-1-2-1-2-4-6 Solid</td>
<td>ice, น้ำแข็ง</td>
</tr>
<tr>
<td>1-1-2-1-2-6-1 Crystal</td>
<td>ice, น้ำแข็ง</td>
</tr>
<tr>
<td>1-1-2-1-2-7 Body substance</td>
<td>pus, น้ำมูก</td>
</tr>
<tr>
<td>1-1-2-1-3 Location</td>
<td>where, ที่, place, ที่, there, ที่, นี่</td>
</tr>
<tr>
<td>1-1-2-1-3-1-1 Residence</td>
<td>home, ที่อยู่อาศัย</td>
</tr>
<tr>
<td>1-1-2-1-3-1-2 Region</td>
<td>wild, ป่า</td>
</tr>
<tr>
<td>1-1-2-1-3-1-3 Area</td>
<td>กลางนอก, outside, place, สถานที่, ชั้นใน, inside, ภายใน, out</td>
</tr>
<tr>
<td>1-1-2-1-3-1-4 Place</td>
<td>ที่</td>
</tr>
<tr>
<td>1-1-2-1-3-2-1 Geographical area</td>
<td>countryside, ชนบท</td>
</tr>
<tr>
<td>1-1-2-1-3-2-2 Rural area</td>
<td>yard, ลาน</td>
</tr>
<tr>
<td>1-1-2-1-3-2-2-1 Field</td>
<td>yard, สนามหญ้า</td>
</tr>
<tr>
<td>1-1-2-1-3-2-3 Heavenly</td>
<td>sky, ฟ้า</td>
</tr>
<tr>
<td>1-1-2-1-3-2-4 Boundary</td>
<td>edge, ขอบ</td>
</tr>
<tr>
<td>1-1-2-1-4 Natural object</td>
<td>nest, รัง</td>
</tr>
<tr>
<td>1-1-2-1-4-1 Geology</td>
<td>mountain, เขา, spring, พระ, fountain</td>
</tr>
<tr>
<td>1-1-2-1-4-2 Organic Structure</td>
<td>flesh, ร่างกาย, body</td>
</tr>
</tbody>
</table>

7. Element is any of more than a hundred simple substances that consist of atoms of only one kind and that, alone or in combination, make up all substances.

8. Point is the precise location of something.

9. Place is a point located with respect to surface features of some region; this place is nice.

10. Region is a large indefinite location on the surface of the earth.

11. Area is a particular geographical region of indefinite boundary.

12. Space is an area reserved for some particular purpose.

13. Place is a space reserved for sitting (as in a theater or on a train or airplane).
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1-2-1-5-1 Wetland</td>
<td>marsh</td>
</tr>
<tr>
<td>1-1-2-1-6 Body of water</td>
<td>lake</td>
</tr>
<tr>
<td>1-1-2-1-6-2</td>
<td>pond</td>
</tr>
<tr>
<td>1-1-2-1-7 Part/Portion</td>
<td>wing</td>
</tr>
<tr>
<td>1-1-2-1-7-1 Component</td>
<td></td>
</tr>
<tr>
<td>1-2-1-7-1-1 Addition/Improver</td>
<td></td>
</tr>
<tr>
<td>1-2-1 Supernatural thing</td>
<td></td>
</tr>
<tr>
<td>1-2-1-1 Supernatural being</td>
<td></td>
</tr>
<tr>
<td>1-2-1-1-1 Spirit</td>
<td>spirit</td>
</tr>
<tr>
<td>1-2-1-1-1-1 Good spirit</td>
<td>spirit</td>
</tr>
<tr>
<td>1-2-1-1-2 Evil spirit</td>
<td></td>
</tr>
<tr>
<td>1-2-1-2 Imaginary being</td>
<td></td>
</tr>
<tr>
<td>1-2-1-2-1 Imaginary person</td>
<td></td>
</tr>
<tr>
<td>1-2-1-2-2 Imaginary creature</td>
<td></td>
</tr>
<tr>
<td>1-2-2 Vital principle</td>
<td></td>
</tr>
<tr>
<td>1-2-3 Physics forces⁴</td>
<td></td>
</tr>
<tr>
<td>1-2-4 Life Form</td>
<td></td>
</tr>
<tr>
<td>1-3 Cell¹⁵</td>
<td></td>
</tr>
<tr>
<td>1-3-1 Reproductive cell/Germ cell</td>
<td></td>
</tr>
<tr>
<td>1-4 Part</td>
<td></td>
</tr>
<tr>
<td>1-4-1 Body part</td>
<td></td>
</tr>
<tr>
<td>1-4-1-1 External body part</td>
<td></td>
</tr>
<tr>
<td>1-4-1-1-1</td>
<td>head</td>
</tr>
<tr>
<td>1-4-1-1-2</td>
<td>neck</td>
</tr>
<tr>
<td>1-4-1-1-3 External body part of animal</td>
<td></td>
</tr>
<tr>
<td>1-4-1-1-3-1 External body part of bird/ insects</td>
<td></td>
</tr>
<tr>
<td>1-4-1-1-4 External body part of plant</td>
<td></td>
</tr>
<tr>
<td>1-4-1-1-4-1 Plant organ</td>
<td>stem</td>
</tr>
<tr>
<td>1-4-1-1-4-2 Reproductive structure</td>
<td>flower anther fruit</td>
</tr>
<tr>
<td>1-4-1-2 Body covering</td>
<td></td>
</tr>
<tr>
<td>1-4-1-2-1 Growth from the skin of person or animal</td>
<td></td>
</tr>
<tr>
<td>1-4-1-2-2 On the head of human being</td>
<td>hair for</td>
</tr>
<tr>
<td>1-4-1-2-3 body covering of bird</td>
<td>feather</td>
</tr>
<tr>
<td>1-4-1-3 Tissue</td>
<td></td>
</tr>
<tr>
<td>1-4-1-3-1 Animal Tissue</td>
<td>flesh</td>
</tr>
</tbody>
</table>

² Physics is the science concerned with the study of matter and natural forces, such as light, heat, movement, etc.
³ Cell is the basic structural and functional unit of all organisms; A tiny mass of protoplasm that contains a nucleus, is enclosed by a membrane, and forms the fundamental unit of living matter.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1-1 Change</td>
<td>warm up</td>
</tr>
<tr>
<td>2-1-1-1 Change state</td>
<td>ดีน (up) เปลี่ยนเป็นเปลี่ยนไป</td>
</tr>
<tr>
<td>2-1-1-2 Change in size or magnitude</td>
<td>change in size or magnitude</td>
</tr>
<tr>
<td>2-1-1-2-1 Increase</td>
<td>grow, grown, increase</td>
</tr>
<tr>
<td>2-1-1-2-1-1 Natural development</td>
<td>increase in size</td>
</tr>
<tr>
<td>2-1-1-2-1-1-1 Height</td>
<td>grow, grown, become larger</td>
</tr>
<tr>
<td>2-1-1-2-1-2 Length</td>
<td>grow, grown, become larger</td>
</tr>
<tr>
<td>2-1-1-2-1-2 Increase in degree</td>
<td>increase in degree</td>
</tr>
<tr>
<td>2-1-1-2-1-3 Come by degree/Pass into a condition gradually</td>
<td>grow, grown, pass into a condition</td>
</tr>
<tr>
<td>2-1-1-3 Change to the contrary</td>
<td>spread</td>
</tr>
<tr>
<td>2-1-1-4 Change so as to cover or include a greater area</td>
<td>spread</td>
</tr>
<tr>
<td>2-1-1-5 Change posture</td>
<td>rise, rose, arise</td>
</tr>
<tr>
<td>2-1-1-6 Depart (To stop being in or with)</td>
<td>leave</td>
</tr>
<tr>
<td>2-1-1-7 Form</td>
<td>blow</td>
</tr>
<tr>
<td>2-1-1-8 Cook</td>
<td>brown</td>
</tr>
<tr>
<td>2-1-1-9 Correct</td>
<td>fine</td>
</tr>
<tr>
<td>2-1-1-10 Purify</td>
<td>fine</td>
</tr>
<tr>
<td>2-1-1-11 To improve by making or becoming thinner</td>
<td>fine</td>
</tr>
<tr>
<td>2-1-1-12 Break</td>
<td>crack open</td>
</tr>
<tr>
<td>2-1-1-12-1 Break open</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-12-2 Break without dividing</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-12-2-1</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-12-2-2</td>
<td>crack open</td>
</tr>
<tr>
<td>2-1-1-13 To break open</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-13-1 To Crack open</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-13-1-1</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-13-1-2</td>
<td>crack</td>
</tr>
<tr>
<td>2-1-1-14 Terminate</td>
<td>crack</td>
</tr>
</tbody>
</table>

16 Grow is to increase in amount, size e.g., The company has grown rapidly in the last five years.
17 Grow is to increase in size by natural development e.g., Grass grows after rain.
18 Grow is to increase in size by natural development e.g., He's grown six inches.
19 Grow is to increase in size by natural development e.g., She doesn't like her hair short, so she's letting it grow.
20 Grow is to increase in degree e.g., A growing number of people are taking part-time jobs.
21 Grow is to come by degree or to pass into a condition gradually, e.g., He grew more and more sad.
22 Spread is to cause to open, stretch out, e.g., Spread the map out.
23 Spread is to stretch out so as to cover or include a greater area, e.g., The city has spread out rapidly in all directions.
24 uns, to break open (break off with the fingers and thumb), e.g., To crack nuts.
25 man, to break open, e.g., I cracked two eggs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1-1-14-1 Interrupt</td>
<td>school</td>
</tr>
<tr>
<td>2-1-1-15 Improve</td>
<td>school</td>
</tr>
<tr>
<td>2-1-1-15-1 Educate</td>
<td></td>
</tr>
<tr>
<td>2-1-1-15-2 Refine</td>
<td></td>
</tr>
<tr>
<td>2-1-1-16 Be born</td>
<td></td>
</tr>
<tr>
<td>2-1-1-16-1</td>
<td></td>
</tr>
<tr>
<td>2-1-1-16-2</td>
<td></td>
</tr>
<tr>
<td>2-1-1-17 Cause to be/cause to do something</td>
<td></td>
</tr>
<tr>
<td>2-1-1-18 change integrity</td>
<td></td>
</tr>
<tr>
<td>2-1-1-18-1 Break up</td>
<td></td>
</tr>
<tr>
<td>2-1-1-18-2 Liquify</td>
<td></td>
</tr>
<tr>
<td>2-1-2 Come to pass</td>
<td></td>
</tr>
<tr>
<td>2-1-3 Interact</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1 Communicate</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-1 Utter</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-2 Express emotion/feeling</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-3 Ask/Order</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-3-1 Request</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-3-2 Order</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-4 Demand</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-5 Inform</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-5-1 Teach</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-5-2 Let something be known</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-5-3 Narrate/Give a detailed account of</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-6 Telecommunicate</td>
<td></td>
</tr>
<tr>
<td>2-1-3-1-7 Declare</td>
<td></td>
</tr>
<tr>
<td>2-1-3-2 Treat/Interact in the certain way</td>
<td></td>
</tr>
<tr>
<td>2-1-3-2-1 Mock</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3 Transfer</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-1 Give</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-2 Exchange</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-2-1</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-2-2</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-3 Payment</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-3-1 Penalty</td>
<td></td>
</tr>
<tr>
<td>2-1-3-3-3-2 Regular payment</td>
<td></td>
</tr>
<tr>
<td>2-1-3-4 Charge</td>
<td></td>
</tr>
<tr>
<td>2-1-3-4-1 Impose</td>
<td></td>
</tr>
</tbody>
</table>

---

26 Be born is to come into existence through birth.
27 Hatch is (of an egg) to break, letting the young bird out.
28 Make is to put into a certain state, position.. e.g., Eating the unripe apples made him ill.
29 Be e.g., Ex: I lost my wallet, this was during the visit to my parents' house.
30 Want is to wish or demand the presence of e.g., I want you here at noon.
31 Give, e.g., I can’t believe you gave $3000 for that broken-down old car.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1-3-5 Collect in one place</td>
<td>gather รับกลุ่ม</td>
</tr>
<tr>
<td>2-1-4 Support</td>
<td>mother เด็กหญ้า รพ.น้า water</td>
</tr>
<tr>
<td>2-1-4-1 Take care</td>
<td>walk เดิน ไป go wander</td>
</tr>
<tr>
<td>2-1-5 Move</td>
<td>march turn เดิน ไต่ turn</td>
</tr>
<tr>
<td>2-1-5-1 Move so as to change position</td>
<td>turn เดิน</td>
</tr>
<tr>
<td>2-1-5-2 Move so as of a body part</td>
<td>turn หมุน</td>
</tr>
<tr>
<td>2-1-5-2-1 Move round a central or fixed point</td>
<td>turn พัก</td>
</tr>
<tr>
<td>2-1-5-2-2 Move so that a different side faces upwards or outwards</td>
<td>turn หมุน</td>
</tr>
<tr>
<td>2-1-5-3 Move to change a residence, affiliation</td>
<td>move เปลี่ยน</td>
</tr>
<tr>
<td>2-1-5-4 Move thing</td>
<td>จับ carry</td>
</tr>
<tr>
<td>2-1-5-4-1 Move thing with force</td>
<td>poke จับ duck จับกวน</td>
</tr>
<tr>
<td>2-1-5-4-2 Move thing from one place to another</td>
<td>take จับ</td>
</tr>
<tr>
<td>2-1-5-5 Move forward slowly, as if by inches</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-5-1 Move forward as if by jumping</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-6 Move toward something</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-7 Move upward</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-8 Move downward</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-8-1 Submerge</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-9 Move very fast</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-10 Follow</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-11 Put in the certain place</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-12 Drift</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-13 Go across or through</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-14 Cause to move in a circular movement</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-15 Move to (cause to) have influence or effect over a wider area</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-15-1 Diseases</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-15-2 Fire</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-5-16 Stir</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-6 Displace</td>
<td>chase นิ่ง ค้า transplant</td>
</tr>
<tr>
<td>2-1-6-1 Separate</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-6-1-1 Cut</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-6-1-2 Pull violently</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-6-1-3 Tear</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-7 Prod</td>
<td>นิ่ง</td>
</tr>
<tr>
<td>2-1-8 Make/Create</td>
<td>นิ่ง</td>
</tr>
</tbody>
</table>

32 Turn is to move around an axis or a center, e.g., The wheels are turning.
33 Take, e.g., We usually take the children to school in the car.
34 Take, e.g., Don't forget to take your bag when you go.
35 Make is to produce something, e.g., She made a cake.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1-8-1 Reproduce</td>
<td>hatch⁴⁶</td>
</tr>
<tr>
<td>2-1-8-2 Create mentally</td>
<td>hatch</td>
</tr>
<tr>
<td>2-1-8-2-1 Imagine</td>
<td>dream</td>
</tr>
<tr>
<td>2-1-8-3 Produce</td>
<td>ผล</td>
</tr>
<tr>
<td>2-1-8-4 Create verbally</td>
<td>สร้าง</td>
</tr>
<tr>
<td>2-1-8-5 Create from raw material</td>
<td>สร้างจากสิ่ง materia</td>
</tr>
<tr>
<td>2-1-8-6 Create by combining materials and parts</td>
<td>สร้างจากสิ่ง materia</td>
</tr>
<tr>
<td>2-1-9 Think/Cerebrate</td>
<td>think⁴⁹</td>
</tr>
<tr>
<td>2-1-9-1 Think logically</td>
<td>think</td>
</tr>
<tr>
<td>2-1-9-2 Plan</td>
<td>will</td>
</tr>
<tr>
<td>2-1-10 Appear/Come into sight or view</td>
<td>ออก</td>
</tr>
<tr>
<td>2-1-11 Conduct</td>
<td>led</td>
</tr>
<tr>
<td>2-1-12 Cause</td>
<td>led</td>
</tr>
<tr>
<td>2-1-13 Control</td>
<td>led</td>
</tr>
<tr>
<td>2-1-13-1 Manage</td>
<td>ran</td>
</tr>
<tr>
<td>2-1-13-1-1 Direct (be in charge of)</td>
<td>ดำเนิน</td>
</tr>
<tr>
<td>2-1-14 Breathe</td>
<td>ran</td>
</tr>
<tr>
<td>2-1-15 Cause to be perceived</td>
<td>ทำ</td>
</tr>
<tr>
<td>2-1-15-1 Make a certain sound or noise</td>
<td>ทำเสียง</td>
</tr>
<tr>
<td>2-1-16 Cause to open</td>
<td>open</td>
</tr>
<tr>
<td>2-1-17 Cause to start operating</td>
<td>ทำงาน</td>
</tr>
<tr>
<td>2-1-18 Touch</td>
<td>ran</td>
</tr>
<tr>
<td>2-1-18-1 Make physical contact with</td>
<td>ตี</td>
</tr>
<tr>
<td>2-1-18-1-1 Caress</td>
<td>งาน</td>
</tr>
<tr>
<td>2-1-18-1-2 Cause to be in brief contact with</td>
<td>งาน</td>
</tr>
<tr>
<td>2-1-18-1-3 Strike</td>
<td>งาน</td>
</tr>
<tr>
<td>2-1-18-1-3-1 Beat</td>
<td>แรง</td>
</tr>
<tr>
<td>2-1-18-1-4 Handle⁴⁰</td>
<td>แรง</td>
</tr>
<tr>
<td>2-1-18-1-4-1 Operate/Control</td>
<td>ทำงาน</td>
</tr>
<tr>
<td>2-1-18-2 Be in contact with</td>
<td>ป้องกัน</td>
</tr>
<tr>
<td>2-1-19 To place something on or over to protect or hide</td>
<td>ป้องกัน</td>
</tr>
</tbody>
</table>

---

³⁶ Hatch is to (sit) on eggs.
³⁷ Dream is to imagine something, e.g., I never said that! You must have been dreaming; have a daydream.
³⁸ Give, e.g., I hope my son didn't give you any trouble.
³⁹ Think is to use the power of reason; make judgments or careful considerations; use the mind to form ideas and opinion, e.g., She thought long and hard before coming to the decision.
⁴⁰ Handle is to touch, lift or hold with hands.
⁴¹ Cover is to be or lie on or over the surface of something; spread over something e.g., The furniture was covered with dust.
⁴² Cover is to place something upon or over something in order to protect or hide it, e.g., The noise was so loud that she covered her ears with her hands.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1-19-1 Cover as to cover eyes, ears</td>
<td>ปิดตา, ปิดหู</td>
</tr>
<tr>
<td>2-1-19-2 Envelop</td>
<td>ผ้า</td>
</tr>
<tr>
<td>2-1-20 To provide with a covering or caused to be spread</td>
<td>ผ้า, ห่ม</td>
</tr>
<tr>
<td>2-1-21 Close</td>
<td>ปิด, ปิด</td>
</tr>
<tr>
<td>2-1-22 Inactive</td>
<td>นอน, หลับ</td>
</tr>
<tr>
<td>2-1-23 Attempt</td>
<td>พยายาม</td>
</tr>
<tr>
<td>2-1-24 Consume</td>
<td>รับ, ดื่ม</td>
</tr>
<tr>
<td>2-1-24-1 Take in liquids</td>
<td>ดื่ม</td>
</tr>
<tr>
<td>2-1-24-2 Take in solid food</td>
<td>รับ, ดื่ม</td>
</tr>
<tr>
<td>2-1-24-3 Inhale/exhale smoke from cigarettes</td>
<td>กลั่น</td>
</tr>
<tr>
<td>2-1-25 Watch</td>
<td>มอง</td>
</tr>
<tr>
<td>2-2 Activity</td>
<td>ใช้, การทำ</td>
</tr>
<tr>
<td>2-2-1 Human activity</td>
<td></td>
</tr>
<tr>
<td>2-2-1-1 Movement</td>
<td></td>
</tr>
<tr>
<td>2-2-1-1-1 Stoke/A single complete movement</td>
<td>กระทำ, การกระทำ</td>
</tr>
<tr>
<td>2-2-1-1-2 Walking</td>
<td></td>
</tr>
<tr>
<td>2-2-1-1-3 Group action</td>
<td></td>
</tr>
<tr>
<td>2-2-1-1-3-1 Group procession</td>
<td></td>
</tr>
<tr>
<td>2-2-1-1-3-2 Business</td>
<td></td>
</tr>
<tr>
<td>2-2-1-1-4 Play</td>
<td></td>
</tr>
<tr>
<td>2-2-1-2 Stay</td>
<td></td>
</tr>
<tr>
<td>2-2-1-3 Recreation</td>
<td></td>
</tr>
<tr>
<td>2-2-1-3-1 Sport</td>
<td></td>
</tr>
<tr>
<td>2-2-1-4 Work</td>
<td></td>
</tr>
<tr>
<td>2-2-1-4-1 Investigation</td>
<td></td>
</tr>
<tr>
<td>2-2-1-5 Keeping something secret</td>
<td></td>
</tr>
<tr>
<td>2-2-2 Sensory activity</td>
<td></td>
</tr>
<tr>
<td>2-2-2-1 Looking</td>
<td></td>
</tr>
<tr>
<td>2-2-3 Action</td>
<td></td>
</tr>
<tr>
<td>2-2-3-1 Arrival/Reaching</td>
<td></td>
</tr>
<tr>
<td>2-2-3-2 Departure</td>
<td></td>
</tr>
<tr>
<td>2-2-3-3 Rescue</td>
<td></td>
</tr>
<tr>
<td>2-3 Attribute</td>
<td></td>
</tr>
<tr>
<td>2-3-1 Property of living thing</td>
<td></td>
</tr>
<tr>
<td>2-3-1-1 Property of Human</td>
<td></td>
</tr>
<tr>
<td>2-3-1-2 Property of Animal</td>
<td></td>
</tr>
<tr>
<td>2-3-2 Property of non-living thing</td>
<td></td>
</tr>
<tr>
<td>2-3-2-1 of food/thing</td>
<td></td>
</tr>
<tr>
<td>2-3-3 Property of temperature</td>
<td></td>
</tr>
<tr>
<td>2-3-4 Property of emotion</td>
<td></td>
</tr>
</tbody>
</table>

43 e.g., covered her ears, covered her eyes.
44 e.g., covered the body.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-3-5 Property of vision (Visual property)</td>
<td>dark, light, lightness, brightness.</td>
</tr>
<tr>
<td>2-3-6 Property of movement</td>
<td></td>
</tr>
<tr>
<td>2-3-6-1 Come forth</td>
<td>out&lt;sup&gt;45&lt;/sup&gt; ออกมา</td>
</tr>
<tr>
<td>2-3-6-2 Go forth</td>
<td>out&lt;sup&gt;46&lt;/sup&gt; ออกไป</td>
</tr>
<tr>
<td>2-3-6-3 Away from the open air/outside</td>
<td></td>
</tr>
<tr>
<td>2-3-6-4 To or toward the inside</td>
<td>in&lt;sup&gt;47&lt;/sup&gt; เข้าไป</td>
</tr>
<tr>
<td>2-3-7 Property of quality</td>
<td></td>
</tr>
<tr>
<td>2-3-8 Away from a surface or edge</td>
<td></td>
</tr>
<tr>
<td>2-3-8-1 away from a surface or edge</td>
<td>out ออกจาก</td>
</tr>
<tr>
<td>2-3-9 Property of space</td>
<td></td>
</tr>
<tr>
<td>2-3-10 The inner part of a solid object or the faces away from the open air</td>
<td></td>
</tr>
<tr>
<td>2-3-11 Form/Shape</td>
<td></td>
</tr>
<tr>
<td>2-3-11-1 Amorphous shape</td>
<td></td>
</tr>
<tr>
<td>2-3-12 Time</td>
<td>timing</td>
</tr>
<tr>
<td>2-3-13 Property of physically or mentally strong</td>
<td>strength</td>
</tr>
<tr>
<td>2-4 Measure/Quantity</td>
<td></td>
</tr>
<tr>
<td>2-4-1 Number</td>
<td>one, 1, หนึ่ง five ห้า</td>
</tr>
<tr>
<td>2-4-2 Classifier</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1 Classifier of object</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-1 Classifier of living thing</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-1-1 Classifier of person</td>
<td>คน</td>
</tr>
<tr>
<td>2-4-2-1-1-2 Classifier of animal</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-1-3 Classifier of plant</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-2 Classifier of non-living thing</td>
<td></td>
</tr>
<tr>
<td>2-4-2-1-2-1 Classifier of tool</td>
<td>อุปกรณ์</td>
</tr>
<tr>
<td>2-4-2-1-2-2 Classifier of room</td>
<td>ห้อง</td>
</tr>
<tr>
<td>2-4-2-1-2-3 Classifier of external body part</td>
<td>ขา หัว นิ้ว</td>
</tr>
<tr>
<td>2-4-2-1-2-4 Classifier of wing</td>
<td>ปีก</td>
</tr>
<tr>
<td>2-4-2-1-2-5 Classifier of root vegie, crucifer</td>
<td>หัว</td>
</tr>
<tr>
<td>2-4-2-1-2-6 Classifier of place</td>
<td>ที่</td>
</tr>
<tr>
<td>2-4-2-1-2-7 Classifier of marsh</td>
<td>แม่น้ำ</td>
</tr>
<tr>
<td>2-4-2-1-2-8 Classifier of place, lake</td>
<td>ทะเล</td>
</tr>
<tr>
<td>2-4-2-1-2-9 Classifier of marsh</td>
<td>ทะเล</td>
</tr>
<tr>
<td>2-4-2-1-2-10 Classifier of housing</td>
<td>บ้าน</td>
</tr>
<tr>
<td>2-4-2-1-2-11 Classifier of door</td>
<td>ประตู</td>
</tr>
<tr>
<td>2-4-2-2 Classifier for pairs of anything</td>
<td>คู่</td>
</tr>
<tr>
<td>2-4-2-2 Classifier of collection</td>
<td>ชุด, ประเภท</td>
</tr>
</tbody>
</table>

<sup>45</sup> Out is away from the inside e.g., “Open the bag and take the money out”, “Blood poured out from the wound”.

<sup>46</sup> Out is away from the inside e.g., “She opened the cage and let the bird out”.

<sup>47</sup> In is away from the outside e.g., “Open the box and put the money in”.

<sup>48</sup> In is to or toward the inside e.g., “There was a loud explosion and the walls fell in”.

<sup>49</sup> Inside, e.g., He reached into his inside jacket pocket.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4-2-2-1 Classifier of human</td>
<td>กลุ่ม</td>
</tr>
<tr>
<td>2-4-2-2-2 Classifier of animal</td>
<td>สัตว์</td>
</tr>
<tr>
<td>2-4-2-3 Classifier of time</td>
<td>day วัน hour ชั่วโมง</td>
</tr>
<tr>
<td>2-4-2-4 Classifier of measurement</td>
<td>yard นิ้ว</td>
</tr>
<tr>
<td>2-4-2-5 Classifier of frequency</td>
<td>ที่ ครั้ง</td>
</tr>
<tr>
<td>2-4-2-6 Classifier of action</td>
<td>2-4-3 Time</td>
</tr>
<tr>
<td>2-4-3-1 Time period</td>
<td>週末 ตอนเช้า แค่เช้า เข้า morning กลางวัน เวลาเย็น กลางคืน evening ในตอนนี้ then</td>
</tr>
<tr>
<td>2-4-3-1-2 Season</td>
<td>冬 winter 季節 season 春 spring</td>
</tr>
<tr>
<td>2-4-3-3 Calendar month</td>
<td>3月 march เดือนมีนาคม</td>
</tr>
<tr>
<td>2-4-3-4 Time off</td>
<td>ลา ลาออก</td>
</tr>
<tr>
<td>2-4-4 Clock time</td>
<td>6 โมง 50 นาที</td>
</tr>
<tr>
<td>2-4-5 Indefinite quantity</td>
<td>2-4-5-1 Indefinite number more than 2 or 3 but not many (normally followed by clfr.)</td>
</tr>
<tr>
<td>2-4-6 Definite quantity</td>
<td>all ทั้งหมด</td>
</tr>
<tr>
<td>2-4-7 Degree of truth</td>
<td>จริง</td>
</tr>
<tr>
<td>2-4-8 To such a (great) degree</td>
<td>10 อย่างนั้น</td>
</tr>
<tr>
<td>2-4-9 Linear unit</td>
<td>yard นิ้ว</td>
</tr>
<tr>
<td>2-5 Psychological feature (a feature of the mental life of a living organism)</td>
<td>2-5-1 Cognition</td>
</tr>
<tr>
<td>2-5-1-1 Intention</td>
<td>dream ความฝัน</td>
</tr>
<tr>
<td>2-5-1-2 Imaginary</td>
<td>symbol สัญลักษณ์</td>
</tr>
<tr>
<td>2-5-1-3 Representationa process</td>
<td>research การวิจัย</td>
</tr>
<tr>
<td>2-5-1-4 Problem solving</td>
<td>mind51 สมอง</td>
</tr>
<tr>
<td>2-5-1-5 Ability</td>
<td>matter เรื่อง</td>
</tr>
<tr>
<td>2-5-1-6 Content</td>
<td>2-5-1-6-1 Concept</td>
</tr>
<tr>
<td>2-5-1-7 Learning</td>
<td>make เท่า</td>
</tr>
<tr>
<td>2-5-1-7-1 Education</td>
<td>school โรงเรียน</td>
</tr>
<tr>
<td>2-5-2 Motivation</td>
<td>2-5-3 Feeling</td>
</tr>
<tr>
<td>2-5-3-1 Desire</td>
<td>anger ความโกรธ</td>
</tr>
<tr>
<td>2-5-3-2 Emotion</td>
<td></td>
</tr>
</tbody>
</table>

50 All, e.g., We walked all the way.
51 Mind is the ability to think and reason: intellect, e.g., He has a very sharp mind.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-5-4 Remember</td>
<td>think\textsuperscript{52} \text{ recall}</td>
</tr>
<tr>
<td>2-5-5 Confidence</td>
<td>sure \text{ แสบ}</td>
</tr>
<tr>
<td>2-5-6 Mind</td>
<td>\text{ in mind}\textsuperscript{53}</td>
</tr>
<tr>
<td>2-5-6-1</td>
<td>spirit</td>
</tr>
<tr>
<td>2-5-6-2</td>
<td>spirit</td>
</tr>
<tr>
<td>2-6 Perceive (to become aware of through the senses)</td>
<td>\text{ dream}\textsuperscript{54}</td>
</tr>
<tr>
<td>2-6-1 Look</td>
<td>\text{ look \text{ around} \ มองไป\รอบ ๆ}</td>
</tr>
<tr>
<td>2-6-1-1 Look attentively</td>
<td>\text{ watch \ ดู}</td>
</tr>
<tr>
<td>2-6-2 Feel</td>
<td>\text{ anger \ โกรธ}</td>
</tr>
<tr>
<td>2-6-2-1 UNSatisfy</td>
<td>\text{ pleased \ พอใจ \ happy}\textsuperscript{55}</td>
</tr>
<tr>
<td>2-6-2-2 Satisfy</td>
<td>\text{ pleased \ พอใจ}</td>
</tr>
<tr>
<td>2-6-2-3 Pleased</td>
<td>\text{ like \ ชอบ \ love \ รัก}</td>
</tr>
<tr>
<td>2-6-2-4 Like</td>
<td>lonely \ อย่างไร \ เหล่า</td>
</tr>
<tr>
<td>2-6-2-5 Separated</td>
<td>hate</td>
</tr>
<tr>
<td>2-6-2-6 Dislike</td>
<td>mind \ รังเกียจ</td>
</tr>
<tr>
<td>2-6-2-7 Reject/Disapprove</td>
<td>\text{ want \ ต้องการ \ กรราหน้า \ hungry}</td>
</tr>
<tr>
<td>2-6-2-8 Desire\textsuperscript{56}</td>
<td>last \ ผ่าน</td>
</tr>
<tr>
<td>2-7 Continue to exist</td>
<td>last \ สิ้น</td>
</tr>
<tr>
<td>2-8 Remain in good condition</td>
<td>last \ ผ่าน</td>
</tr>
<tr>
<td>2-9 Judge\textsuperscript{57}</td>
<td>think\textsuperscript{58} \ คิด \ believe, consider</td>
</tr>
<tr>
<td>2-9-1 Think</td>
<td>think\textsuperscript{58} \ คิด \ believe, consider</td>
</tr>
<tr>
<td>2-9-2 Expect\textsuperscript{59}</td>
<td>think \ นึก \ ว่า \ must \ จะ \ เชื่อว่า \ believe</td>
</tr>
<tr>
<td>2-9-3 Test/Prove\textsuperscript{60}</td>
<td>test \ prove \ tried \ ลอง \ try</td>
</tr>
<tr>
<td>2-9-4 Accept\textsuperscript{61}</td>
<td>เชื่อ \ believe</td>
</tr>
<tr>
<td>2-10 Keep</td>
<td>\text{ hid \ซ่อน \ hidden \ hide \ hiding}</td>
</tr>
<tr>
<td>2-10-1 To put or keep out of sight</td>
<td>hide \ ซ่อน \ hidden \ ซ่อน หiding</td>
</tr>
<tr>
<td>2-11 Confine/Deprive of freedom</td>
<td>\text{ corner \ จมูก}</td>
</tr>
<tr>
<td>2-12 Decide/Come to decision about something</td>
<td>\text{ decide \ ตัดสินใจ \ ตกลงใจ}</td>
</tr>
<tr>
<td>2-13 Leave</td>
<td>\text{ left \ ออกจาก \ leave}</td>
</tr>
<tr>
<td>2-13-1 Escape</td>
<td>\text{ run away, พังจากไป}</td>
</tr>
<tr>
<td>2-14 Leave behind</td>
<td>\text{ left \ ตั้ง \ leave}</td>
</tr>
<tr>
<td>2-15 Social relation</td>
<td>\text{ social relation}</td>
</tr>
</tbody>
</table>

\textsuperscript{52} Think is to recall knowledge from memory, e.g., I can't think what her last name was.

\textsuperscript{53} Mind is a person's (way of) thinking or feeling: thoughts, e.g., She looks very worried. I wonder what is on her mind.

\textsuperscript{54} Dream, experience while sleeping, e.g., What did you dream about?

\textsuperscript{55} Happy, feeling that something is right or good, satisfied, e.g., The government won't be very happy about the latest unemployment statistics.

\textsuperscript{56} Desire is to feel or have a desire for; want strongly e.g., I want to go home now.

\textsuperscript{57} Judge is to form an opinion of or pass judgement on.

\textsuperscript{58} Think is to believe, consider: have an opinion e.g., I think he is very smart.

\textsuperscript{59} Expect is to regard something as probable or likely, e.g., I didn't think to find her in the kitchen.

\textsuperscript{60} Test/Prove is to put to the test or give experiment.

\textsuperscript{61} Accept is to consider or hold as true. accept as true, e.g., I believed his report.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-15-1 Communication</td>
<td></td>
</tr>
<tr>
<td>2-15-1-1 Written material</td>
<td>will หนังสือ</td>
</tr>
<tr>
<td>2-15-1-2 Content</td>
<td>point, ตรงนี้</td>
</tr>
<tr>
<td>2-15-1-2-1 Significant</td>
<td></td>
</tr>
<tr>
<td>2-15-1-3 Language</td>
<td>Egyptian</td>
</tr>
<tr>
<td>2-15-1-3-1 Natural language</td>
<td>symbol</td>
</tr>
<tr>
<td>2-15-1-4 Signal</td>
<td>traffic lights, สัญญาณไฟจราจร</td>
</tr>
<tr>
<td>2-15-1-4-1 Visual signal</td>
<td>font</td>
</tr>
<tr>
<td>2-15-1-4-2 Written symbol</td>
<td></td>
</tr>
<tr>
<td>2-15-1-5 Visual communication</td>
<td></td>
</tr>
<tr>
<td>2-15-1-5-1 Facial gesture</td>
<td>laugh การหัวเราะ</td>
</tr>
<tr>
<td>2-15-1-6 Auditory communication</td>
<td></td>
</tr>
<tr>
<td>2-15-1-6-1 Utterance</td>
<td>shout การพูด</td>
</tr>
<tr>
<td>2-16 To discover by chance or experience</td>
<td>find หา</td>
</tr>
<tr>
<td>2-17 Encounter</td>
<td>meet พบ</td>
</tr>
<tr>
<td>2-18 Search</td>
<td>look for หา</td>
</tr>
<tr>
<td>2-19 Examine</td>
<td>look ดู</td>
</tr>
<tr>
<td>2-20 Blow</td>
<td>blow พัด</td>
</tr>
<tr>
<td>2-21 Relation</td>
<td></td>
</tr>
<tr>
<td>2-21-1 Part</td>
<td>rest ผู้พัก</td>
</tr>
<tr>
<td>2-21-2 Isolated</td>
<td>alone, อยู่ตัวเดียว</td>
</tr>
<tr>
<td>2-21-3 Spatial relation</td>
<td>room ห้อง</td>
</tr>
<tr>
<td>2-22 Stop</td>
<td>stopped หยุด</td>
</tr>
<tr>
<td>2-23 Not stopping/Continuously</td>
<td>on อยู่</td>
</tr>
<tr>
<td>2-24 Remove</td>
<td></td>
</tr>
<tr>
<td>2-24-1</td>
<td>throw off, shake off หลุด</td>
</tr>
<tr>
<td>2-24-2</td>
<td>spirit</td>
</tr>
<tr>
<td>2-25 Treat/Provide treatment for</td>
<td></td>
</tr>
<tr>
<td>2-25-1 Operate</td>
<td>castrate ยับยั้ง</td>
</tr>
<tr>
<td>2-26 Cause to grow or develop</td>
<td></td>
</tr>
<tr>
<td>2-26-1 Plant/Crops</td>
<td>grow ปลูก</td>
</tr>
<tr>
<td>2-26-2 Hair/Beard</td>
<td>ไว้</td>
</tr>
<tr>
<td>2-27 Let go/Set free</td>
<td>release ปล่อย</td>
</tr>
<tr>
<td>2-27-1 To press so as to allow something to move</td>
<td>release ปล่อย</td>
</tr>
<tr>
<td>2-28 Utilize</td>
<td>use ใช้</td>
</tr>
<tr>
<td>2-29 Be familiar or acquainted with a person or an object, place</td>
<td>know รู้จัก</td>
</tr>
<tr>
<td>2-30 Know how to do or perform something</td>
<td>know รู้</td>
</tr>
</tbody>
</table>

---

62 Search is to try to find, try to locate or discover, e.g., looking for a lost book/ new job.
63 Examine is to observe carefully or inspect; to use the eyes in order to find something, e.g., You could see it if you'd only look.
64 Blow is to be blowing or storming e.g., “The wind blew from the west”.
65 On, e.g., We worked on (and on) all night.
66 All, e.g., He worked hard all last year.
67 Release a button, a handle, etc. that can be pressed to allow part of machine to move.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-31 To have knowledge of something; have information in the mind</td>
<td>know รู้</td>
</tr>
<tr>
<td>2-32 Study</td>
<td>research วิจัย</td>
</tr>
<tr>
<td>2-32-1 Investigate</td>
<td></td>
</tr>
<tr>
<td>2-33 Designate</td>
<td></td>
</tr>
<tr>
<td>2-33-1 Label</td>
<td>call เรียกว่า ตั้งชื่อ name</td>
</tr>
<tr>
<td>2-34 Do/Perform 68</td>
<td>do ทำ make 69</td>
</tr>
<tr>
<td>2-35 Wait</td>
<td>wait_for รอ</td>
</tr>
<tr>
<td>2-36 Begin</td>
<td>begin, start, เริ่ม began</td>
</tr>
<tr>
<td>2-37 Arrive/reach a destination</td>
<td>come มาถึง</td>
</tr>
<tr>
<td>2-38 Pay attention to</td>
<td>mind สนใจ</td>
</tr>
<tr>
<td>2-39 Let/allow</td>
<td></td>
</tr>
<tr>
<td>2-39-1 Give permission</td>
<td>let อนุญาต</td>
</tr>
<tr>
<td>2-39-2 Make it possible through lack of action for something to happen</td>
<td>let ปลอดภัย</td>
</tr>
<tr>
<td>2-40 Join/Get together</td>
<td>joined_in เข้าร่วม</td>
</tr>
<tr>
<td>2-41 have ownership/possession of</td>
<td>owns เจ้าของ</td>
</tr>
<tr>
<td>3 Phenomenon 70</td>
<td>luck โชค fortune</td>
</tr>
<tr>
<td>3-1 Natural Phenomenon</td>
<td></td>
</tr>
<tr>
<td>3-1-1 Physical phenomenon</td>
<td>lightning, reflection การสะท้อน</td>
</tr>
<tr>
<td>3-1-2 Geological phenomenon</td>
<td>flood</td>
</tr>
<tr>
<td>3-2 Chemical phenomenon</td>
<td>fire</td>
</tr>
<tr>
<td>4 Event 71</td>
<td>fire, ไฟไหม้, accident</td>
</tr>
<tr>
<td>4-1 Nature event</td>
<td></td>
</tr>
<tr>
<td>4-1-1 Sound</td>
<td></td>
</tr>
<tr>
<td>4-1-1-1 Human made sound</td>
<td>call เลี่ยงชีวิต</td>
</tr>
<tr>
<td>4-1-1-2 Animal made sound</td>
<td>quack เล่า เท้าเดี๋ยพื้นที่แมว purr</td>
</tr>
<tr>
<td>4-1-2 Movement</td>
<td>peep นิ้ว spread การแพร่กระจาย</td>
</tr>
<tr>
<td>5 State 72</td>
<td>motherhood พ่อแม่ของชาย พ่อแม่ของสาว</td>
</tr>
<tr>
<td>5-1 Relationship</td>
<td>เป็น live มีชีวิต</td>
</tr>
<tr>
<td>5-2 Alive/living</td>
<td>live ดำรงชีวิต</td>
</tr>
<tr>
<td>5-3 Survive</td>
<td></td>
</tr>
<tr>
<td>5-4 Physiological state</td>
<td>cold ไอติม</td>
</tr>
<tr>
<td>5-4-1 Illness</td>
<td>tired เท้าไส้ อาหาร hungry thirsty หิว</td>
</tr>
<tr>
<td>5-4-2 State of body and mind</td>
<td>name ชื่อเสียง</td>
</tr>
<tr>
<td>5-5 Reputation</td>
<td>spirit 73 บรรยากาศ</td>
</tr>
<tr>
<td>5-6 Atmosphere</td>
<td></td>
</tr>
</tbody>
</table>

---

68 Perform is to act or perform an action.
69 Make is to perform an action, e.g., to make a decision.
70 Phenomenon is any state or process known through the senses rather than by intuition or reasoning.
71 Event is something that happens at a given place and time.
72 State is the way something is with respect to its main attributes: “the current state of knowledge”, “in a weak financial state”.
73 Spirit is the general atmosphere of a place or situation, e.g., He came to the party, but didn’t really enter into the spirit of it.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Be&lt;sup&gt;74&lt;/sup&gt;</td>
<td>be อุทัย live, stay&lt;sup&gt;75&lt;/sup&gt;</td>
</tr>
<tr>
<td>6-1 Be upright</td>
<td>stand อยู่</td>
</tr>
<tr>
<td>6-2 Be seated</td>
<td>sit นั่ง sat</td>
</tr>
<tr>
<td>6-3 Be in a horizontal position</td>
<td>lie, lay นอน</td>
</tr>
<tr>
<td>6-4 Cover</td>
<td>cover ครอบคลุม</td>
</tr>
<tr>
<td>6-5 Not move</td>
<td>stay อยู่</td>
</tr>
<tr>
<td>6-6 Stay&lt;sup&gt;77&lt;/sup&gt;</td>
<td>look ดู</td>
</tr>
<tr>
<td>6-7 Seem&lt;sup&gt;78&lt;/sup&gt;</td>
<td>matter สำคัญ</td>
</tr>
<tr>
<td>6-8 count/carry weight</td>
<td>belong เบื้องหน้าของ</td>
</tr>
<tr>
<td>6-9 belong</td>
<td></td>
</tr>
<tr>
<td>7 Reference term for thing</td>
<td></td>
</tr>
<tr>
<td>7-1 Reference term for a living thing</td>
<td>ฉัน คุณ you me</td>
</tr>
<tr>
<td>7-1-1 Reference term for male living thing</td>
<td>he him himself ตัวเขาเอง เขา</td>
</tr>
<tr>
<td>7-1-2 Reference term for female living thing</td>
<td>she ตัวเธอเอง herself her เธอ</td>
</tr>
<tr>
<td>7-1-3 Reference term for a group of living thing</td>
<td>พวกเขากำลัง themพวกเราจะ นั้นเรายังไม่</td>
</tr>
<tr>
<td>7-1-4 Reference term for human</td>
<td>คน one</td>
</tr>
<tr>
<td>7-1-5 Reference term for animal</td>
<td>ตัว one</td>
</tr>
<tr>
<td>7-2 Reference term for a non living thing</td>
<td>ฉัน what จะใด</td>
</tr>
<tr>
<td>7-3 Reference term for a time</td>
<td>when เมื่อไหร่</td>
</tr>
<tr>
<td>8 Group/Grouping</td>
<td></td>
</tr>
<tr>
<td>8-1 Social group</td>
<td></td>
</tr>
<tr>
<td>8-1-1 Organization</td>
<td></td>
</tr>
<tr>
<td>8-1-1-1 Political party</td>
<td>wing พรรค faction bloc</td>
</tr>
<tr>
<td>8-1-1-2 Military Unit</td>
<td>wing กองทัพ</td>
</tr>
<tr>
<td>8-1-1-2-1 Air unit</td>
<td></td>
</tr>
<tr>
<td>8-2 A group of human</td>
<td>flock กลุ่ม</td>
</tr>
<tr>
<td>8-3 A group of animal</td>
<td>กลุ่ม</td>
</tr>
<tr>
<td>8-3-1 A group of fish</td>
<td>school โรง</td>
</tr>
<tr>
<td>8-3-2 A group of birds/sheep/goats</td>
<td>flock</td>
</tr>
<tr>
<td>8-3-3 A group of goats/cattle/elephants</td>
<td>herd</td>
</tr>
<tr>
<td>8-3-4 A group of dogs/wolves</td>
<td>pack</td>
</tr>
<tr>
<td>9 Be able to/have the ability to</td>
<td>สามารถ can could</td>
</tr>
<tr>
<td>10 Get/acquire</td>
<td>ได้ obtain have&lt;sup&gt;79&lt;/sup&gt; ใคร่ has had</td>
</tr>
<tr>
<td>10-1 Accept</td>
<td>take&lt;sup&gt;80&lt;/sup&gt; รับ took</td>
</tr>
<tr>
<td>11 Have</td>
<td>have got, have ถูก possess had</td>
</tr>
<tr>
<td>12 Accumulate/Collect&lt;sup&gt;81&lt;/sup&gt;</td>
<td>gather&lt;sup&gt;82&lt;/sup&gt; รวบรวม</td>
</tr>
</tbody>
</table>

<sup>74</sup> Be is to occupy a certain position or area, be somewhere, e.g., The boy is in the house.
<sup>75</sup> Stay is to stop and remain rather than go on or leave, e.g., Can you stay for dinner?
<sup>76</sup> Cover is to fill; extend over, e.g., The city covers 25 square miles.
<sup>77</sup> Stay is to live in a place for a while as a visitor or guest e.g., My mother is staying with us this week.
<sup>78</sup> To seem by expression or appearance, e.g., You look tired.
<sup>79</sup> Have is to receive or obtain, e.g., We must have your answer by Friday.
<sup>80</sup> Take is to accept, e.g., Do you take traveller's check?
<sup>81</sup> Collect, e.g., collect evidence.
<sup>82</sup> Gather, e.g., He travels about the world gathering facts about little-known diseases.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 Attribute value</td>
<td></td>
</tr>
<tr>
<td>13-1 Name</td>
<td>Johns, Mary, John, Ronne Randall</td>
</tr>
<tr>
<td>13-2 Sex</td>
<td>male, female</td>
</tr>
<tr>
<td>13-3 Title</td>
<td>prof, mr, mrs, miss, dr, dean, chairman, president</td>
</tr>
<tr>
<td>13-4 Job position</td>
<td></td>
</tr>
<tr>
<td>13-5 Occupation/function</td>
<td></td>
</tr>
<tr>
<td>13-5-1 Human’s occupation</td>
<td>faculty, student, nurse, doctor, nun, priest watch</td>
</tr>
<tr>
<td>13-5-2 Animal’s function</td>
<td>working dog, watching dog, hunting dog, toy dog, guard dog, quide dog, carriage dog</td>
</tr>
<tr>
<td>13-6 Kind/Sort</td>
<td></td>
</tr>
<tr>
<td>13-6-1 Person (of color)</td>
<td>Asian, European, Negro, black, white, Egyptian, Thai</td>
</tr>
<tr>
<td>13-6-2 Animal (kind /pedigree)</td>
<td>dalmatian, toy poodle, doberman, alsatian, siamese cat, persian cat, butterflyfish</td>
</tr>
<tr>
<td>13-6-3 Plant (genus)</td>
<td>mountain rose, tea rose, china rose, Mac, PC, Unix</td>
</tr>
<tr>
<td>13-6-4 Equipment</td>
<td>even</td>
</tr>
<tr>
<td>13-6-5 Number</td>
<td></td>
</tr>
<tr>
<td>13-7 Age</td>
<td>10, 20, 30</td>
</tr>
<tr>
<td>13-7-1 Age of living thing</td>
<td>young, old</td>
</tr>
<tr>
<td>13-7-1-1 Age of plant</td>
<td>young, old, green</td>
</tr>
<tr>
<td>13-7-1-1-1 Age of plant part</td>
<td>green</td>
</tr>
<tr>
<td>13-7-1-1-1-1 Age of fruit</td>
<td>green</td>
</tr>
<tr>
<td>13-7-1-2 Age of human</td>
<td></td>
</tr>
<tr>
<td>13-7-1-2-1 Age of male</td>
<td></td>
</tr>
<tr>
<td>13-7-1-2-2 Age of female</td>
<td></td>
</tr>
<tr>
<td>13-7-2 Age of non-living thing</td>
<td>young, old new</td>
</tr>
<tr>
<td>13-8 Appearance</td>
<td></td>
</tr>
<tr>
<td>13-8-1 Size</td>
<td>little</td>
</tr>
<tr>
<td>13-8-1-1 Size of a living thing</td>
<td>biggest</td>
</tr>
<tr>
<td>13-8-1-2 Size of a non-living thing</td>
<td>biggest</td>
</tr>
<tr>
<td>13-8-2 Weight</td>
<td>fat, long, plump</td>
</tr>
<tr>
<td>13-8-3 Length</td>
<td>short, long, ugly</td>
</tr>
<tr>
<td>13-8-4 Beauty</td>
<td>ugly, ugliest, beautiful</td>
</tr>
<tr>
<td>13-8-4-1 Beauty of a living thing</td>
<td>handsome</td>
</tr>
<tr>
<td>13-8-4-2 Beauty of a non-living thing</td>
<td>handsome</td>
</tr>
<tr>
<td>13-8-5 Texture</td>
<td></td>
</tr>
<tr>
<td>13-8-5-1 Covered in fresh grass and leaves</td>
<td>fine</td>
</tr>
</tbody>
</table>

83 Green, e.g., The countryside is very green.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
</table>
| 13-8-5-2 Color           | yellow สีเหลือง สีเหลือง gray black สีดำ  
|                          | ต่าง brown สีน้ำตาล สีขาว white rose สีชมพูคู่ลำคู่ red สีแดง สีเขียว green สีเขียว  
| 13-8-5-2-1               | bright สี  
| 13-8-5-3 Evenness        | even เรียบ  
| 13-8-9 Height            | high  
| 13-8-9-1                 | tall  
| 13-8-9-2                 | animal กรุงเทพ  
| 13-9 Trait               | wild  
| 13-9-1 Wildness (A state of nature) | violent, wild  
| 13-9-2 Behavior          | green กรุงเทพ  
| 13-9-3 Inexperience      | dimka  
| 13-10 Quality            | kind  
| 13-10-1 Kindness         | kind กรุงเทพ  
| 13-10-2 Humaneness       | kind กรุงเทพ  
| 13-10-3 Ability          |  
| 13-10-4 movement and form| graceful กรุงเทพ  
| 13-10-5 Importance       | important กรุงเทพ  
| 13-10-6                  | bright กรุงเทพ  
| 13-10-7                  | fine กรุงเทพ  
| 13-11 Healthy            |  
| 13-12 Degree             | warm  
| 13-12-1 of temperature/weather | cold กรุงเทพ  
| 13-12-2 of brightness    |  
| 13-12-3 of emotionality  | happy กรุงเทพ  
| 13-12-3-1 Happiness      | sad กรุงเทพ  
| 13-12-4 of certainty     |  
| 13-12-5 of truth         | true กรุงเทพ  
| 13-13 Duration           | long  
| 13-14 Timing             | early กรุงเทพ  
| 13-14-1 before the usual/arranged or expected time |  

84 Animal e.g., animal desires.  
85 Green is a young and inexperienced and therefore easily deceived and ready to believe anything.  
86 Bright is showing hope or signs of future success, e.g., You have a bright future ahead of you.  
87 Kindness is tendency to be kind and forgiving.  
88 Humaneness is the quality of compassion or consideration for others (people or animals).  
89 Weak is lacking power.  
90 Bright is clever, e.g., She’s very bright.  
91 สุ่ง is to feel a strong urge, to be having great pleasure.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-14-2 near the beginning of a period</td>
<td>early ตอนแรก</td>
</tr>
<tr>
<td>13-15 Frequency</td>
<td></td>
</tr>
<tr>
<td>13-15-1 at any time</td>
<td>ever เคย, have ever</td>
</tr>
<tr>
<td>13-15-2 not ever</td>
<td>never ไม่เคย</td>
</tr>
<tr>
<td>13-15-3 more</td>
<td>อีก more ยัง</td>
</tr>
<tr>
<td>13-15-4 once again</td>
<td>again อีกครั้ง</td>
</tr>
<tr>
<td>13-15-5 one at a time</td>
<td>one_by_one ลำดับหนึ่ง</td>
</tr>
<tr>
<td>13-16 State</td>
<td></td>
</tr>
<tr>
<td>13-16-1 State of living thing</td>
<td></td>
</tr>
<tr>
<td>13-16-1-1 out of danger</td>
<td>safe ปลอดภัย</td>
</tr>
<tr>
<td>13-16-1-2 danger</td>
<td>dangerous</td>
</tr>
<tr>
<td>13-16-2 State of non-living thing</td>
<td></td>
</tr>
<tr>
<td>13-16-2-1 Not working/not operating</td>
<td>off ยัง</td>
</tr>
<tr>
<td>13-16-3 Readiness</td>
<td>willing</td>
</tr>
<tr>
<td>13-17 Order</td>
<td></td>
</tr>
<tr>
<td>13-17-1 Immediately pass order</td>
<td>ทันเวลา last</td>
</tr>
<tr>
<td>13-17-2 Coming after all others in time or space or degree or being the only one remaining</td>
<td>last ล่าสุด</td>
</tr>
<tr>
<td>13-17-3 Following in time</td>
<td>after หลังจาก</td>
</tr>
<tr>
<td>13-17-4 Immediately following in time</td>
<td>next ห่างๆ</td>
</tr>
<tr>
<td>13-17-5 (of time) before</td>
<td>to ก่อน by before</td>
</tr>
<tr>
<td>13-17-6 at an earlier point in an order than; ahead of</td>
<td>ก่อน before</td>
</tr>
<tr>
<td>13-17-7 Closest in space, order or degree; without anything coming before or between</td>
<td>next ถัดไป</td>
</tr>
<tr>
<td>13-17-8 Just afterward ยิ่ง</td>
<td>next ต่อไป then หลังจากนั้น</td>
</tr>
<tr>
<td>13-17-9 Further in space or time/forward</td>
<td>on ยิ่ง</td>
</tr>
<tr>
<td>13-17-10 Finally</td>
<td>at last, ไปสุด</td>
</tr>
<tr>
<td>13-17-11 A place or appointed time in a fixed order</td>
<td>turn รอบ</td>
</tr>
<tr>
<td>13-18 Likeness</td>
<td></td>
</tr>
<tr>
<td>13-18-1 in the same way as; with the same quality as</td>
<td>like เหมือน</td>
</tr>
<tr>
<td>13-18-2 similar to; almost but not exactly the same</td>
<td>similar คล้าย</td>
</tr>
<tr>
<td>13-19 Distance in space or time</td>
<td>away ห่างไม่</td>
</tr>
<tr>
<td>13-20 Position</td>
<td>left ซ้าย</td>
</tr>
<tr>
<td>13-21 Strength</td>
<td></td>
</tr>
<tr>
<td>13-21-1 physically or mentally strong</td>
<td>weak อ่อน</td>
</tr>
<tr>
<td>13-21-2 having a lot of the material which gives taste</td>
<td>strong แก่ (coffee)</td>
</tr>
<tr>
<td>13-22 Possibility</td>
<td>perhaps บางครั้ง</td>
</tr>
<tr>
<td>13-23 Cleanness</td>
<td>ด่า dirty</td>
</tr>
</tbody>
</table>

---

92 more; again, e.g., I'll repeat the question once more.
93 off; not working; not operating; (of a machine or electrical apparatus), e.g., The TV is off.
94 Just afterward, e.g., First, you heat that fat; next, you add the onions.
95 On, e.g., If you walk on you'll come to the church; It's time to move on.
96 Turn, e.g., You've missed your turn so you'll have to wait; It's my turn to drive next.
<table>
<thead>
<tr>
<th>WordAsso Number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-24 Property of location</td>
<td>lonely&lt;sup&gt;97&lt;/sup&gt; โสดเดี่ยว</td>
</tr>
<tr>
<td>13-25 Shape</td>
<td>bend turn ของโค้ง</td>
</tr>
<tr>
<td>13-26 Of or relating to</td>
<td>algebraic ทางคณิตศาสตร์</td>
</tr>
<tr>
<td>13-27 in or into a state of being disconnected or removed</td>
<td>off&lt;sup&gt;98&lt;/sup&gt; ออก</td>
</tr>
<tr>
<td>13-28 Equality</td>
<td>even เสมอกัน</td>
</tr>
<tr>
<td>13-29 Correctness</td>
<td>right ถูก correct</td>
</tr>
<tr>
<td>13-30</td>
<td>willing ต้อง</td>
</tr>
<tr>
<td>13-31 Aliveness</td>
<td>dead ตาย</td>
</tr>
</tbody>
</table>

<sup>97</sup> Lonely is a (of a building or other object) with no others of the same type near, e.g., a lonely house in the wood.
<sup>98</sup> off, e.g., How do you get this lid off?
A. 2 Other Classes

Words which are used to link phrases with clauses or connect clauses together or provide the basis for a descriptive phrase e.g., prepositions, conjunctions, definite determiner are classified as well in a separate classification.

<table>
<thead>
<tr>
<th>WordAsso number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 be someone or something, being, to be</td>
<td>be คือ เข้า (e.g., He is a teacher. The chairman of the department is Dr. Cercone.)</td>
</tr>
<tr>
<td>R2 too/ together/ in addition</td>
<td>with ด้วย</td>
</tr>
<tr>
<td>R3 in each other's company</td>
<td>with กัน</td>
</tr>
<tr>
<td>R4 by means of using, by means of</td>
<td>with ด้วย in 1 through ผ่าน 2 by ณ 7</td>
</tr>
<tr>
<td>R4-1 Using as a means of traveling</td>
<td>โดย ณ 4</td>
</tr>
<tr>
<td>R5 supported by</td>
<td>ด้วย ณ 5</td>
</tr>
<tr>
<td>R6 a having/possessing</td>
<td>with ซึ่งมี</td>
</tr>
<tr>
<td>R7 belonging to</td>
<td>of ของ animal 6 ของสัตว์</td>
</tr>
<tr>
<td>R7-1 belonging to a living thing</td>
<td>ง ของ ของสัตว์ ผู้</td>
</tr>
<tr>
<td>R7-1-1 belonging to him</td>
<td>his ของเข้า</td>
</tr>
<tr>
<td>R7-1-2 belonging to her</td>
<td>her ของเธอ</td>
</tr>
<tr>
<td>R7-1-3 belonging to animal/plant</td>
<td>its ของมัน</td>
</tr>
<tr>
<td>R8 made from</td>
<td>of ทำจาก</td>
</tr>
<tr>
<td>R9 from the group that includes</td>
<td>of ใน พวกใน</td>
</tr>
<tr>
<td>R10 against what might be expected; on the other hand</td>
<td>but แต่ only 7</td>
</tr>
<tr>
<td>R11 other than, except</td>
<td>but นอกจาก</td>
</tr>
<tr>
<td>R12 introductory various kind of clause</td>
<td>that ว่า</td>
</tr>
<tr>
<td>R13 being the one of two or more people or things that is far or further in time, place thought</td>
<td>that นั้น</td>
</tr>
<tr>
<td>R14 being the person, thing, idea, etc., which is understood or has just been mentioned</td>
<td>that ที่นั้น which</td>
</tr>
<tr>
<td>R15 used to join two things; then, afterwards</td>
<td>and และ</td>
</tr>
<tr>
<td>R16 used before the last of a set of possibilities</td>
<td>or หรือ</td>
</tr>
<tr>
<td>R17 the remaining one of the set; what is left as well as that mentioned; an additional person or thing</td>
<td>other อื่น</td>
</tr>
<tr>
<td>R18 the remaining persons or things of the set; what are left as well as those mentioned; additional persons or things</td>
<td>others อื่นๆ</td>
</tr>
<tr>
<td>R19 so as to be in</td>
<td>into เข้าไป to ที่จะ</td>
</tr>
</tbody>
</table>

---

1 In, e.g., Write it in pencil.
2 Through, e.g., I got this job through an employment agency.
3 On, e.g., A car runs on petrol.
4 On, e.g., on foot, on a ship.
5 On, e.g., He went round the world on the money his aunt gave him.
6 Animal e.g., cooking with animal fats.
7 Only: except that, but; e.g., She wants to go, only she hasn’t got enough money.
<table>
<thead>
<tr>
<th>WordAsso number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>R20 so as to be</td>
<td>into เป็น</td>
</tr>
<tr>
<td>R21 in the direction toward</td>
<td>to ไปที่</td>
</tr>
<tr>
<td>R22 reaching as far as</td>
<td>to ถึง</td>
</tr>
<tr>
<td>R23 until and including</td>
<td>to จนกระทั่ง</td>
</tr>
<tr>
<td>R24 for the attention or possession of/ intended to be given to</td>
<td>to ให้</td>
</tr>
<tr>
<td>R25 show a position</td>
<td>in ใน</td>
</tr>
<tr>
<td>R25-1 show a position/a contained by thing with depth, length and height; within an enclosed space, surrounded by an area</td>
<td>on บน</td>
</tr>
<tr>
<td>R25-2 showing position in relation to a surface or supported by a surface</td>
<td>under ใต้</td>
</tr>
<tr>
<td>R25-3 in or to a lower place than; directly below</td>
<td>in ใน</td>
</tr>
<tr>
<td>R26 being included as part of</td>
<td>in ใน</td>
</tr>
<tr>
<td>R27 wearing</td>
<td>in ใน</td>
</tr>
<tr>
<td>R28 at the time of/during</td>
<td>in ใน on บน</td>
</tr>
<tr>
<td>R29 showing the way something is done or happens</td>
<td>in อย่าง</td>
</tr>
<tr>
<td>R30 with the result that; therefore</td>
<td>so ดังนั้น then และ</td>
</tr>
<tr>
<td>R31 in that case</td>
<td>then ถ้าฉันนี้</td>
</tr>
<tr>
<td>R32 as a result of; because of</td>
<td>เนื่องจาก through ผ่าน for ผ่านสำหรับ</td>
</tr>
<tr>
<td>R33 shows purpose</td>
<td>for ฝากจน เพื่อที่จะ to</td>
</tr>
<tr>
<td>R34 instead of</td>
<td>for แทน</td>
</tr>
<tr>
<td>R35 shows length of time</td>
<td>เป็นเวลา for พัก</td>
</tr>
<tr>
<td>R36 shows distance</td>
<td>เป็นระยะ for เวลา</td>
</tr>
<tr>
<td>R37 as regards or in regard to ???</td>
<td>ในเรื่อง for ต่อ</td>
</tr>
<tr>
<td>R38 on the subject of ???/with regard to</td>
<td>about เกี่ยวกับ on บน</td>
</tr>
<tr>
<td>R39 around; here and there; in all parts of</td>
<td>about ทั่ว</td>
</tr>
<tr>
<td>R40 show name of the person who wrote</td>
<td>ชื่อ by</td>
</tr>
<tr>
<td>R41 show the person or thing that performs an action or causes a result</td>
<td>ชื่อ by</td>
</tr>
<tr>
<td>R42 passing through or along</td>
<td>ผ่าน by</td>
</tr>
<tr>
<td>R43 near; beside</td>
<td>ช่วงๆ by</td>
</tr>
<tr>
<td>R44 not later than</td>
<td>ภายใน by</td>
</tr>
</tbody>
</table>

8 In, e.g., Can you see the mistake in this sentence?
9 In, e.g., A man in uniform.
10 In, e.g., In spring, in January.
11 On, e.g., On the morning of July.
12 In, e.g., Speaking in anger.
13 Through, e.g., The war was lost through bad organization.
14 For, e.g., He was rewarded for his bravery.
15 ด้วยเหตุ e.g., As she has no car, she cannot get there easily.
16 For, e.g., She didn’t answer for several minutes.
17 For, e.g., They ran fast for a mile or two.
18 For, e.g., France is famous for its wines.
19 On, e.g., A book on India.
20 กembali e.g., They came in by the back door. It’s quicker if you go by the main road.
<table>
<thead>
<tr>
<th>WordAsso number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>R45 in accordance with/according to</td>
<td>by ตาม21 under ภายใต้22</td>
</tr>
<tr>
<td>R46 used to show the size of units or groups that follow each other</td>
<td>by ฟีล</td>
</tr>
<tr>
<td>R47 show a point in space</td>
<td>at23 ที่</td>
</tr>
<tr>
<td>R48 at the time that</td>
<td>when24 เมื่อ 25</td>
</tr>
<tr>
<td>R48-I show an exact point in time</td>
<td>at26 ตอน</td>
</tr>
<tr>
<td>R49 show an intended aim or object towards which a thing or action is directed</td>
<td>at27 ใต้</td>
</tr>
<tr>
<td>R50 show the subject or activity in which a judgment about someone's ability is made</td>
<td>at28 ใน</td>
</tr>
<tr>
<td>R51 earlier than the time when</td>
<td>ก่อนที่ before20</td>
</tr>
<tr>
<td>R52 in or into a state of</td>
<td>under ภายใต้20</td>
</tr>
<tr>
<td>R53 considering that</td>
<td>ในขณะ when31</td>
</tr>
<tr>
<td>R54 on condition that; supposing that</td>
<td>if ถ้า</td>
</tr>
<tr>
<td>R55 although</td>
<td>if แม้ว</td>
</tr>
<tr>
<td>R56 exclusive of anyone or anything else</td>
<td>only32 เท่านั้น</td>
</tr>
<tr>
<td>R57 nothing more than</td>
<td>only33 เท่านั้น</td>
</tr>
<tr>
<td>R58 every single one of two or more things or people considered separately.</td>
<td>each แต่ละ</td>
</tr>
<tr>
<td>R59 used in comparison</td>
<td>ดัง เท่า34</td>
</tr>
<tr>
<td>R60 in the way or manner that</td>
<td>ดัง ลักษณะ35</td>
</tr>
<tr>
<td>R61 from one place to another</td>
<td>around ไปที่36</td>
</tr>
<tr>
<td>R62 on all sides; surrounding a centre</td>
<td>around รอบๆ37</td>
</tr>
<tr>
<td>R63 somewhere near; in the area</td>
<td>around แพร่38</td>
</tr>
<tr>
<td>R64 away from inside</td>
<td>ออกไปจาก out_of39</td>
</tr>
<tr>
<td>R65 starting at</td>
<td>from จาก</td>
</tr>
<tr>
<td>R65-I starting at (the stated place, position, or condition)</td>
<td>from จาก</td>
</tr>
</tbody>
</table>

21 ตาม e.g., to play by the rules.
22 Under: according to, e.g., Under the terms of the agreement, you have to pay weakly rent.
23 At e.g., at my house.
24 When: at the time that e.g., Things were different when I was a child.
25 เมื่อ when 25
26 ตอน at 26 ตอน
27 ใต้ at 27 ใต้
28 ใน at 28 ใน
29 ก่อนที่ before 29
30 under ภายใต้ 20 under ภายใต้ 20
31 ในขณะ when 31
32 เท่านั้น only 32 เท่านั้น
33 เท่านั้น only 33 เท่านั้น
34 ดัง เท่า 34
35 ดัง ลักษณะ 35
36 ไปที่ 36
37 รอบๆ around รอบๆ 37
38 แพร่ around แพร่ 38
39 ออกไปจาก out_of ออกไปจาก out_of 39
<table>
<thead>
<tr>
<th>WordAsso number and Class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>R65-2 starting at (the stated time)</td>
<td>from ด้วย</td>
</tr>
<tr>
<td>R66 which is more than might me expected</td>
<td>even แม้แต่</td>
</tr>
<tr>
<td>R67 belonging to oneself and to no one self</td>
<td>own เอง</td>
</tr>
<tr>
<td>R68</td>
<td>to ให้</td>
</tr>
</tbody>
</table>
Appendix B

SL Dictionary

This appendix illustrates a fragment of the English dictionary, discussed in sections 3.3 and 5.1.3, in Prolog representation style as illustrated in (1). Each entry contains the English entry and all its possible syntactic categories.

(1) entry_e (English, [cat1, cat2, ...]).

```prolog
entry_e('Egyptian', [n]).
entry_e('John', [n]).
entry_e('Mrs', [n]).
entry_e('Ronne Randall', [prop]).
entry_e('an', [det]).
entry_e('about', [prep, adv, adj]).
entry_e('again', [adv]).
entry_e('algebraic', [adj]).
entry_e('all', [adj, adv, pron, n, det]).
entry_e('all_by_yourself', [adj]).
entry_e('an', [det]).
entry_e('and', [conj]).
entry_e('anger', [n, v]).
entry_e('animal', [n, adj]).
entry_e('around', [adv, prep, adj]).
entry_e('as', [adv, conj, prep]).
entry_e('ask', [v]).
entry_e('at', [prep]).
entry_e('at_last', [adv]).
entry_e('away', [adv, adj]).
entry_e('be', [v, aux]).
entry_e('beautiful', [adj]).
entry_e('before', [adv, conj, adj, prep]).
entry_e('began', [vpast]).
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
| entry_e(believe,[v])   | entry_e(belong,[v])   | entry_e(big,[adj])   | entry_e(biggest,[adj]) | entry_e(bird,[n])   | entry_e(black,[adj,n]) | entry_e(blow,[v])   | entry_e(born,[vpast,adj]) | entry_e(bright,[adj]) | entry_e(brother,[n]) | entry_e(brown,[adj,v]) | entry_e(but,[conj,prep]) | entry_e(by,[prep,adv]) | entry_e(call,[v,n]) | entry_e(came,[vpast]) | entry_e(came_over,[vpast]) | entry_e(can,[v,n]) | entry_e(cat,[n]) | entry_e(certainly,[adv]) | entry_e(chase,[n,v]) | entry_e(children,[n]) | entry_e(cold,[adj,n,adv]) | entry_e(come,[v]) | entry_e(coming,[v,n]) | entry_e(corner,[n,v]) | entry_e(could,[vpast]) | entry_e(countryside,[n]) | entry_e(cover,[v,n]) | entry_e(crack,[v,n]) | entry_e(crack_open,[v]) | entry_e(cried,[vpast]) | entry_e(cucumber,[n]) | entry_e(curtain,[n]) | entry_e(day,[n]) | entry_e(decide,[v]) | entry_e(do,[v]) | entry_e(door,[n]) | entry_e(dream,[v,n]) | entry_e(duck,[n,v]) | entry_e(duckling,[n]) | entry_e(each,[adj,pron,adv]) | entry_e(early,[adv,adj]) | entry_e(eat,[v]) | entry_e(edge,[n,v]) | entry_e(egg,[v,n]) | entry_e(even,[adj,adv]) | entry_e(evening,[n]) | entry_e(ever,[adv]) | entry_e(farmyard,[n]) | entry_e(fat,[adj,n]) |}

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
entry_e(lonely,[adj]).
entry_e(long,[adj,adv]).
entry_e(look,[v,n]).
entry_e(look_around,[v]).
entry_e(lovely,[adj]).
entry_e(make,[v,n]).
entry_e(make_fun_of,[v]).
entry_e(many,[adj]).
entry_e(march,[v,n]).
entry_e(marsh,[n]).
entry_e(matter,[n,v]).
entry_e(me,[pron]).
entry_e(melt,[v]).
entry_e(mind,[n,v]).
entry_e(peep,[v,n]).
entry_e(ran,[vpast]).
entry_e(ran_after,[vpast]).
entry_e(reed,[n]).
entry_e(reflection,[n]).
entry_e(release,[v,n]).
entry_e(rend,[v]).
entry_e(rent,[v,vpast]).
entry_e(research,[n,v]).
entry_e(rest,[n,v]).
entry_e(room,[n]).
entry_e(rose,[n,vpast,adj]).
entry_e(run_away,[v]).
entry_e(sad,[adj]).
entry_e(sadly,[adv]).
entry_e(safe,[adj,n]).
entry_e(said,[vpast]).
entry_e(salad,[n]).
entry_e(sat,[vpast]).
entry_e(saw,[n,v,vpast]).
entry_e(school,[n,v]).
entry_e(seen,[vpast]).
entry_e(several,[adj]).
entry_e(she,[pron]).
entry_e(should,[aux]).
entry_e(shout,[v,n]).
entry_e(so,[adv,conj]).
entry_e(spirit,[n,v]).
entry_e(spread,[v,n]).
entry_e(spring,[v,n]).
entry_e(stay,[n,v]).
entry_e(stopped,[vpast]).
entry_e(stork,[n]).
entry_e(student,[n]).
entry_e(sunny,[adj]).
entry_e(sure,[adj,adv]).
entry_e(swan,[n]).
entry_e(symbol,[n]).
entry_e(take,
entry_e(talk,
entry_e(that,
entry_e(them,
entry_e(then,
entry_e(there,
entry_e(they,
entry_e(think,
entry_e(thought,
entry_e(through,
entry_e(tired,
entry_e(to,
entry_e(told,
entry_e(traffic_light,
entry_e(tried,
entry_e(trout,
entry_e(true,
entry_e(try,
entry_e(turkey,
entry_e(turn,
entry_e(ugliest,
entry_e(ugly,
entry_e(under,
entry_e(us,
entry_e(use,
entry_e(vegetable,
entry_e(very,
entry_e(wait,
entry_e(wander,
entry_e(want,
entry_e(warm,
entry_e(what,
entry_e(wheat,
entry_e(when,
entry_e(where,
entry_e(white,
entry_e(wild,
entry_e(will,
entry_e(willing,
entry_e(wind,
entry_e(wing,
entry_e(winter,
entry_e(with,
entry_e(woman,
entry_e(worm,
entry_e(would,
entry_e(written,
entry_e(yard,
entry_e(yellow,
entry_e(you,
entry_e(young,
Appendix C

SL-TL Dictionary

This appendix illustrates a fragment of the English-Thai dictionary, discussed in sections 3.3 and 5.1.3, in Prolog representation style as illustrated in (1). Each entry contains the English entry and all its possible corresponding words in Thai. Each Thai word is related to its WordAssociation number.

(1) entry (English, [wordasso ('Thai1', [wordassol, wordasso2,...]), wordasso ('Thai2', [wordassoa, wordassob,...]), ...]).

entryC'Egyptian', [wordassoC'นิวอียิปต์', [13-6-1]], wordassoC'นิวอียิปต์', [2-15-1-31]).
entryC'John', [wordassoC'โจน', [13-1]]).
entryC'Joan', [wordassoC'โจาน', [13-1]]).
entryC'Mrs', [wordassoC'มิส', [13-3]]).
entryC'Ronne_Randall', [wordassoC'โรนน์_แรนเดอล์ล์', [13-1]]).
entryC'a', [wordassoC'ทู', [a1]]).
entryC'algebraic', [wordassoC'สิ่งคณิตศาสตร์', [13-26]])).
entryC'call', [wordassoC'ที่มา', [2-4-6]], wordassoC'มา', [2-4-6-1, 2-23]]).
entryC'call_by_yourself', [wordassoC'ตัวเอง', [2-21-2]])).
entryC'Can', [wordassoC'คาน', [a1]]).
entryC'and', [wordassoC'และ', [r15]]).
entryC'anger', [wordassoC'อารมณ์', [2-41-1]], wordassoC'อารมณ์', [2-5-3-2]])).
entryC'as', [wordassoC'พื้นที่', [r60]], wordassoC'พื้นที่', [r48]]).
wordassoC'ของ', [r59]).
wordassoC'พื้นที่', [r32]]).
entryC'ask', [wordassoC'ถาม', [2-13-1]], wordassoC'ถาม', [2-1-3-1-3-1]], wordassoC'ถาม', [2-1-3-1-4-1]]).
entryC'at', [wordassoC'ที่', [r47]], wordassoC'ที่', [r481]], wordassoC'ที่', [r50]], wordassoC'ที่', [r49]]). entryC'at_last', [wordassoC'ที่สุด', [13-17-9]])).
entry('crack_open', [wordasso('打破了', ['2-1-1-12-2-2-2'])]).
entry('cried', [wordasso('哭了', ['2-1-1-3-1-1'])]).
entry('cucumber', [wordasso('瓜', ['1-1-1-3-1-3-1', '1-1-2-1-2-1-2-1'])]).
entry('day', [wordasso('天', ['2-4-2-3', '2-4-3-1-2'])]).
entry('decided', [wordasso('决定', ['2-1-12'])]).
entry('do', [wordasso('做', ['2-3-4'])]).
entry('door', [wordasso('门', ['2-1-1-2-1-1-1'])]).
entry('dream', [wordasso('做梦', ['2-6'])]).
entry('each', [wordasso('每个', ['2-1-7'])]).
entry('edge', [wordasso('边缘', ['13-1-4-2'])]).
entry('duck', [wordasso('鸭', ['2-1-1-2-1-1', '1-1-2-1-2-1-2-4-1'])]).
wordasso('鸭', ['2-1-5-4-1']).
wordasso('鸭嘴兽', ['2-1-1-2-1-1']).
entry('duckling', [wordasso('鸭', ['1-1-1-2-1-2-1'])]).
entry('each', [wordasso('每个', ['2-1-7'])]).
entry('early', [wordasso('早', ['13-14-1'])]).
entry('egg', [wordasso('蛋', ['1-1-2-1-2-1-2', '1-3-1'])]).
entry('evening', [wordasso('夜晚', ['2-4-3-1-1'])]).
entry('ever', [wordasso('每当', ['13-15-1'])]).
entry('farmyard', [wordasso('农场', ['2-1-2-1-1-1'])]).
entry('fat', [wordasso('肥', ['13-8-2'])]).
entry('feather', [wordasso('羽毛', ['1-1-2-1-2-5'])]).
entry('feeling', [wordasso('感情', ['2-5-3'])]).
entry('find', [wordasso('找到', ['2-16'])]).
entry('fine', [wordasso('细', ['13-10'])]).
wordasso('细', ['13-10-7']).
wordasso('细', ['13-12']).
wordasso('细', ['13-11']).
wordasso('细', ['2-1-3-4-1']).
wordasso('细', ['2-1-3-3-3-1']).
wordasso('细', ['2-1-1-10']).
wordasso('细', ['13-8-5']).
wordasso('细', ['2-1-1-11']).
entry('fish', [wordasso('鱼', ['1-1-1-2-1-3-1', '1-1-2-1-2-1-2-3'])]).
entry('fishing', [wordasso('捕鱼', ['2-2-1-3-1'])]).
entry('flock', [wordasso('群', ['2-1-1-2-1-3-1', '2-1-1-2-1-2-1-2-3'])]).
entry('five', [wordasso('五', ['2-4-1'])]).
entry('flesh', [wordasso('肉', ['1-4-1-3-1'])]).
entry('flew', [wordasso('飞', ['2-1-18-1-4-1', '2-1-5-1'])]).
entry('flock', [wordasso('群', ['8-3-2'])]).
wordasso('群', ['8-2']).
wordasso('群', ['1-1-2-1-2-3-4']).
entry('for', [wordasso('为了', ['r24', 'r33'])]).
wordasso('为了', ['r34']).
wordasso('为了', ['r32']).
wordasso('为了', ['r35']).
wordasso('为了', ['r36']).
wordasso('为了', ['r37']).
entry(friend, [wordasso('พื้น', ['1-1-1-1-1']), \[1-1-1-1-6'])]
entry(from, [wordasso('จาก', ['r65-1'])), wordasso('ลำพู', ['r65-2'])
entry(gather, [wordasso('รวบรวม', ['12'])), wordasso('จับมุม', ['2-1-3-5'])
entry(geese, [wordasso('หาง', ['1-1-1-2-1-2-1'])])
entry(gleaming, [wordasso('เงาสะท้อน', ['13-12-2'])])
entry(go, [wordasso('ไป', ['2-1-5-1'])]), wordasso('ynamoebi', ['2-1-1-1'])
entry(graceful, [wordasso('ภายใต้', ['13-10-4'])])
entry(gray, [wordasso('น้ำตา', ['13-8-5-2'])])
entry(hatch, [wordasso('ฟัก', ['2-1-1-16-2', '2-1-8-1'])]), wordasso('ฟัก', ['2-1-8-2'])
entry(home, [wordasso('เข้า', ['7-1-1'])])
entry(head, [wordasso('หัว', ['1-4-1-1-1'])])
entry(heard, [wordasso('ได้ยิน', ['2-6'])])
entry(hen, [wordasso('ที่น', ['1-1-1-2-1-2-2', '1-1-2-1-2-1-2-4-1'])])
entry(here, [wordasso('ที่นี่', ['1-1-2-1-3-1'])])
entry(herself, [wordasso('ตัวเอง', ['7-1-2'])])
entry(hid, [wordasso('ซ่อน', ['2-10-1']), wordasso('ซ่อน', ['2-10-2'])])
entry(hide, [wordasso('ซ่อน', ['2-10-1']), wordasso('ซ่อน', ['2-10-2'])])
entry(hidden, [wordasso('ซ่อน', ['2-10-1']), wordasso('ซ่อน', ['2-10-2'])])
entry(him, [wordasso('เข้า', ['7-1-1'])])
entry(his, [wordasso('ตัวเอง', ['7-1-1'])])
entry(himself, [wordasso('ตัวเอง', ['7-1-1'])])
entry(home, [wordasso('บ้าน', ['1-1-2-1-1-2-2']), wordasso('ที่อยู่อาศัย', ['1-1-2-1-1-1'])])
entry(hungry, [wordasso('หิว', ['5-4-2']), wordasso('กระหาย', ['2-6-2-8', '5-4-2'])])
entry(l, [wordasso('ที่', ['7-1'])])
entry(l, [wordasso('ที่', ['1-1-2-1-2-6-1'])])
entry(cf, [wordasso('ที่', ['r54']), wordasso('ที่', ['r55'])]
entry(in, [wordasso('ที่', ['r25-1', 'r26', 'r27', 'r28']), wordasso('เข้ามา', ['2-3-6-4'])])
wordasso('เข้ามา', ['2-3-6-3']), wordasso('เข้ามา', ['r29'])
wordasso('เข้ามา', ['2-3-6-4'])
entry(into, [wordasso('เข้ามา', ['r19']), wordasso('เข้าไป', ['r20'])]
entry(it, [wordasso('ที่', ['7-2'])])
entry(its, [wordasso('ที่', ['1-7-3-3'])]
entry(make, [wordasso('หัว', ['2-1-8', '2-3-4']), wordasso('ที่หัว', ['2-1-1-17'])
wordasso('ที่หัว', ['2-5-1-6-1-1'])]
entry('make_fun_of', [wordasso('พาตัว', ['2-1-3-2-1'])])
entry(many, [wordasso('มาก', ['2-4-5']), wordasso('พยายาม', ['2-4-5-1'])]
entry(neck, [wordasso('คอ', ['1-4-1-1-2'])]), wordasso('ลำตู', ['2-1-18-1-1'])
entry(nest, [wordasso('รัง', ['1-1-2-1-4']), wordasso('ที่รัง', ['2-1-8-6'])
entry(never, [wordasso('ไม่เคย', ['13-15-2'])])
entry(new, [wordasso('ใหม่', ['13-7-2'])])
entry(no, [wordasso('ไม่', ['a3']), wordasso('ไม่มี', ['a4'])]
entry(no_one, [wordasso('ไม่มีใคร', ['a5'])])
entry(not, [wordasso('ไม่', ['a2']), wordasso('ไม่นี่', ['a4'])]).
entry(of, [wordasso('พยัคฆ์', ['r7']), wordasso('พยัคฆ์', ['r8']), wordasso('ไม่นี่', ['r9'])]).
entry(open, [wordasso('อยู่', ['2-1-1-1', '2-1-1-16']), wordasso('อยู่', ['2-1-1-17'])]).
entry(or, [wordasso('ที่ทำ', ['r16'])]).
entry(other, [wordasso('อยู่', ['r17'])]).
entry(out, [wordasso('เพิ่ม', ['2-3-6-1']), wordasso('เพิ่ม', ['2-3-6-2'])]).
entry(out_of, [wordasso('เพิ่ม', ['2-3-8-1']), wordasso('เพิ่ม', ['2-1-2-1-3-2-1'])]).
entry(should, [wordasso('อยู่', ['a12'])]).
entry(white, [wordasso('ผักเผย', ['13-8-5-2']), wordasso('ผักเผย', ['13-6-1'])]).
entry(yellow, [wordasso('ผักเผย', ['13-8-5-2']), wordasso('ผักเผย', ['13-6-1'])]).
entry(you, [wordasso('ผักเผย', ['7-1'])]).
### Appendix D

**Types, Their Features and Their Value Types**

This appendix illustrates some types defined in our grammars as discussed in section 6.1.1.2. Their features and their value types are also provided.

<table>
<thead>
<tr>
<th>Type</th>
<th>Features/Type of value</th>
<th>Immediate Supertype</th>
</tr>
</thead>
</table>
| sign          | [SYNSEM synsem  
QSTORE set_quant]                                   | entity              |
| phrase        | ...                                                         | sign                |
| word          | ...                                                         | sign                |
| synsem        | [LOC loc  
NONLOCAL nonlocal]                                 | mod_synsem          |
| pre_mod_synsem|                                                             | synsem              |
| post_mod_synsem|                                                            | synsem              |
| loc           | [CAT cat  
CONT cont  
CONX conx]                                   | entity              |
| conx          | [BACKGR set_psoa]                                          | entity              |
| cat           | [HEAD head  
SPR list_synsem  
SUBJ list_synsem  
COMPS list_synsem  
MARKING marking] | entity              |
| cont          |                                                             | entity              |
| nom_obj       | [MODE mode  
INDEX ind  
RESTR set_psoa]                                    | cont                |

195
<table>
<thead>
<tr>
<th>Type</th>
<th>Features/Type of value</th>
<th>Immediate Supertype</th>
</tr>
</thead>
<tbody>
<tr>
<td>quant</td>
<td>DET sem_det</td>
<td>cont</td>
</tr>
<tr>
<td></td>
<td>RESTIND nom_obj</td>
<td></td>
</tr>
<tr>
<td>forall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>exists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ind</td>
<td>IGENN gend</td>
<td></td>
</tr>
<tr>
<td></td>
<td>INUM num</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IPER per</td>
<td></td>
</tr>
<tr>
<td>it</td>
<td></td>
<td>ind</td>
</tr>
<tr>
<td>there</td>
<td></td>
<td>ind</td>
</tr>
<tr>
<td>ref</td>
<td></td>
<td>ind</td>
</tr>
<tr>
<td>psoa</td>
<td>QUANTS list_quantssem_det</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NUCLEUS qfpsoa</td>
<td></td>
</tr>
<tr>
<td>qfpsoa</td>
<td>SURFACE surface</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WORDASSO wordasso</td>
<td></td>
</tr>
<tr>
<td>property</td>
<td>[ INSTANCE ref ]</td>
<td>qfpsoa</td>
</tr>
<tr>
<td>un_relation</td>
<td></td>
<td>qfpsoa</td>
</tr>
<tr>
<td>bin_relation</td>
<td></td>
<td>qfpsoa</td>
</tr>
<tr>
<td>tri_relation</td>
<td></td>
<td>qfpsoa</td>
</tr>
<tr>
<td>control_qfpsoa</td>
<td></td>
<td>qfpsoa</td>
</tr>
<tr>
<td>lang</td>
<td></td>
<td>entity</td>
</tr>
<tr>
<td>eng</td>
<td></td>
<td>lang</td>
</tr>
<tr>
<td>thai</td>
<td></td>
<td>lang</td>
</tr>
<tr>
<td>head</td>
<td>LANG lang</td>
<td>entity</td>
</tr>
<tr>
<td></td>
<td>AGR agr</td>
<td></td>
</tr>
<tr>
<td>subst</td>
<td>PRD boolean</td>
<td>head</td>
</tr>
<tr>
<td></td>
<td>MOD mod_synsem</td>
<td></td>
</tr>
<tr>
<td>noun</td>
<td>CASE case</td>
<td>subst</td>
</tr>
<tr>
<td></td>
<td>WORDASSO1 wordasso</td>
<td></td>
</tr>
<tr>
<td>verb</td>
<td>VFORM vform</td>
<td>subst</td>
</tr>
<tr>
<td></td>
<td>INV boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AUX boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NEG boolean</td>
<td></td>
</tr>
<tr>
<td>adj</td>
<td></td>
<td>subst</td>
</tr>
<tr>
<td>prep</td>
<td></td>
<td>subst</td>
</tr>
<tr>
<td>reltvz</td>
<td></td>
<td>subst</td>
</tr>
<tr>
<td>clas</td>
<td></td>
<td>subst</td>
</tr>
<tr>
<td>case</td>
<td></td>
<td>entity</td>
</tr>
<tr>
<td>nom</td>
<td></td>
<td>case</td>
</tr>
<tr>
<td>acc</td>
<td></td>
<td>case</td>
</tr>
<tr>
<td>akol</td>
<td></td>
<td>entity</td>
</tr>
<tr>
<td>Type</td>
<td>Features/Type of value</td>
<td>Immediate Supertype</td>
</tr>
<tr>
<td>------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>
| agr  | GENN gend
      | NUM num
      | PER per |
| masc | gend |
| fem  | gend |
| neut | gend |
| sg   | num |
| pl   | num |
| first| per |
| second| per |
| third| per |
| bse  | vform |
| fin  | vform |
| pas  | nform |
| func | [ spec synsem ] |
| mark | head |
| det  | [ count boolean ] |
Appendix E

Lexical Rules

This appendix contains the lexical rules discussed in section 6.1.1.3. Our lexical rule is adapted from the lexical rule which was provided by Gerald Penn [Penn 1993] and Colin Matheson [Matheson 1996]. Lexical rules 1-6 are used for deriving the third singular verb, simple past verb, present participle verb, passive verb and plural noun for English lexical entries. Lexical rule 6 is used to form the simple past verb for Thai.

1. The 3rd-singular verb lexical rule

\[
\begin{align*}
\text{PHON} & \quad X \\
\text{HEAD} & = \begin{bmatrix}
\text{VFORM} & \text{bse} \\
\text{AUX} & \text{[T]} \\
\text{LANG} & \text{eng}
\end{bmatrix} \\
\text{SUBJ} & = \text{NP [nom. sg]}
\end{align*}
\]

\[
\begin{align*}
\text{PHON} & \quad X \rightarrow (X.s) \text{ or } (X.y) \rightarrow (X.i.es) \\
\text{HEAD} & = \begin{bmatrix}
\text{VFORM} & \text{fin} \\
\text{AUX} & \text{[T]} \\
\text{LANG} & \text{eng}
\end{bmatrix} \\
\text{SUBJ} & = \text{NP [nom. sg]}
\end{align*}
\]

2. The simple-past verb lexical rule

\[
\begin{align*}
\text{PHON} & \quad X \\
\text{HEAD} & = \begin{bmatrix}
\text{VFORM} & \text{bse} \\
\text{AUX} & \text{[T]} \\
\text{LANG} & \text{eng}
\end{bmatrix} \\
\text{SUBJ} & = \text{NP [nom. sg]}
\end{align*}
\]

\[
\begin{align*}
\text{PHON} & \quad X \rightarrow (X.ed) \text{ or } (X.y) \rightarrow (X.i.ed) \\
\text{HEAD} & = \begin{bmatrix}
\text{VFORM} & \text{fin} \\
\text{AUX} & \text{[T]} \\
\text{LANG} & \text{eng}
\end{bmatrix} \\
\text{SUBJ} & = \text{NP [nom. sg]}
\end{align*}
\]

or \((X.c) \rightarrow (X.ed)\) or other irregular forms.

198

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
3. The present-participle lexical rule

\[
\begin{align*}
\text{PHON X} & \\
\text{HEAD} & \left[ \begin{array}{c}
\text{VFORM} \\
\text{AUX} \\
\text{LANG}
\end{array} \right] \\
& \text{bse minus eng}
\end{align*}
\Rightarrow
\begin{align*}
\text{PHON X} & \\
\text{HEAD} & \left[ \begin{array}{c}
\text{VFORM} \\
\text{AUX} \\
\text{LANG}
\end{array} \right] \\
& \text{prp minus eng}
\end{align*}
\]

4. The passive lexical rule

\[
\begin{align*}
\text{PHON X} & \\
\text{HEAD} & \left[ \begin{array}{c}
\text{VFORM} \\
\text{AUX} \\
\text{LANG}
\end{array} \right] \\
& \text{bse minus eng}
\end{align*}
\Rightarrow
\begin{align*}
\text{PHON X} & \\
\text{HEAD} & \left[ \begin{array}{c}
\text{VFORM} \\
\text{AUX} \\
\text{LANG}
\end{array} \right] \\
& \text{pas minus eng}
\end{align*}
\]

5. The plural noun formation

\[
\begin{align*}
\text{PHON X} & \\
\text{CAT HEAD} & \left[ \begin{array}{c}
\text{noun} \\
\text{AGR} \\
\text{LANG}
\end{array} \right] \\
& \text{sg minus eng}
\end{align*}
\Rightarrow
\begin{align*}
\text{PHON X} & \\
\text{CAT HEAD} & \left[ \begin{array}{c}
\text{noun} \\
\text{AGR} \\
\text{LANG}
\end{array} \right] \\
& \text{pl minus eng}
\end{align*}
\]

6. The Thai simple_past lexical rule

\[
\begin{align*}
\text{PHON X} & \\
\text{HEAD} & \left[ \begin{array}{c}
\text{VFORM} \\
\text{AUX} \\
\text{LANG}
\end{array} \right] \\
& \text{bse minus thai}
\end{align*}
\Rightarrow
\begin{align*}
\text{PHON X} & \\
\text{HEAD} & \left[ \begin{array}{c}
\text{VFORM} \\
\text{AUX} \\
\text{LANG}
\end{array} \right] \\
& \text{fin minus thai}
\end{align*}
\]
It is noted that the Thai verb does not have an inflection resulting from verb agreement (see section 4.1). The words แ้ (le'ew) and ได้ (dāj) are added to the verb to indicate past tense. แ้ (le'ew) appears immediately after the verb whereas ได้ (dāj) immediately precedes the verb.
Appendix F

SL and TL Grammars

This appendix contains the English and Thai grammars discussed in section 6.1.1.6. Our grammar is a modification to the grammars which were provided by Gerald Penn [Penn 1993] and Colin Matheson [Matheson 1996].

F.1 English Grammar

1. Subject_Head Rule

   subj-hd-ph \[\Rightarrow \begin{align*}
   \text{SUBJ-DTR} & \quad \text{[SYNSEM | LOC | CAT | HEAD \[\text{\#1}\]} \\
   \text{HD-DTR} & \quad \text{[SYNSEM | LOC | CAT | SUBJ \[\text{\#1}\]} 
\end{align*}\]

2. Head_Complement rule

   hd-comp-ph \[\Rightarrow \begin{align*}
   \text{HD-DTR} & \quad \text{[SYNSEM | LOC | CAT | COMP \{\text{\#1, \#2, ..., \#n}\}} \\
   \text{NON-HD-DTRS} & \quad \text{\{SYNSEM \[\text{\#1}\}, [SYNSEM \[\text{\#2}\], ..., [SYNSEM \[\text{\#n}\}\} 
\end{align*}\]
3. Coordination rule

\[ \text{coordination-ph} \Rightarrow \begin{cases} \text{CONJUNCT-DTR1} & \begin{cases} \text{phrase} \\ \text{SYNSEM | LOC | CAT} \end{cases} \\ \text{COORDINATE-DTR} & \begin{cases} \text{SYNSEM | LOC | CAT | HEAD} \end{cases} \\ \text{CONJUNCT-DTR2} & \begin{cases} \text{phrase} \\ \text{SYNSEM | LOC | CAT} \end{cases} \end{cases} \]

4. Head_Subject_Complement Rule

\[ \text{hd-subj-comp-ph} \Rightarrow \begin{cases} \text{HD-DTR} & \begin{cases} \text{word} \\ \text{SYNSEM | LOC | CAT} \end{cases} \\ \text{HEAD} & \begin{cases} \text{INV} \end{cases} \\ \text{SUBJ} & \begin{cases} \text{yes} \end{cases} \\ \text{COMPS} & \begin{cases} 1, 2, ..., 3 \end{cases} \end{cases} \\ \text{SUBJ-DTR} & \begin{cases} \text{SYNSEM | LOC | CAT} \end{cases} \\ \text{SUBJ} & \begin{cases} 4 \end{cases} \end{cases} \\ \text{NON-HD-DTRS} & \begin{cases} \text{SYNSEM} \end{cases} \begin{cases} 0 \end{cases} \end{cases} \end{cases} \]

5. Specifier_head_Rule

\[ \text{spr-hd-ph} \Rightarrow \begin{cases} \text{HD-DTR} & \begin{cases} \text{phrase} \\ \text{SYNSEM | LOC | CAT | SPR} \end{cases} \end{cases} \\ \text{SPEC-DTR} & \begin{cases} \text{SYNSEM | LOC | CAT | HEAD | SPEC} \end{cases} \begin{cases} 2 \end{cases} \end{cases} \]

6. Adjunct_head and Head_adjunct rules

\[ \text{adjunct-hd-ph} \Rightarrow \begin{cases} \text{ADJUNCT-DTR} & \begin{cases} \text{SYNSEM | LOC | CAT | HEAD | MOD} \end{cases} \begin{cases} 2 \end{cases} \begin{cases} \text{pre_mod_synsem} \end{cases} \\ \text{HD-DTR} & \begin{cases} \text{phrase} \end{cases} \end{cases} \]

\[ \text{hd-adjunct-ph} \Rightarrow \begin{cases} \text{HD-DTR} & \begin{cases} \text{phrase} \end{cases} \\ \text{ADJUNCT-DTR} & \begin{cases} \text{SYNSEM | LOC | CAT | HEAD | MOD} \end{cases} \begin{cases} 2 \end{cases} \begin{cases} \text{post_mod_synsem} \end{cases} \end{cases} \]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
F.2 Thai Grammar

1. Subject_Head Rule

\[ \text{subj-hd-ph} \Rightarrow \begin{cases} \text{SUBJ-DTR} & [\text{SYNSEM} | \text{LOC} | \text{CAT} | \text{HEAD} \ 1] \\ \text{HD-DTR} & [\text{SYNSEM} | \text{LOC} | \text{CAT} | \text{SUBJ} \ 1] \end{cases} \]

2. Head_Complement rule

\[ \text{hd-comp-ph} \Rightarrow \begin{cases} \text{HD-DTR} & \text{word} \\ \text{NON-HD-DTRS} & \langle [\text{SYNSEM} 1], [\text{SYNSEM} 2], ..., [\text{SYNSEM} n] \rangle \end{cases} \]

3. Coordination rule

\[ \text{coordination-ph} \Rightarrow \begin{cases} \text{CONJUNCT-DTR1} & \text{phrase} \\ \text{COORDINATE-DTR} & [\text{SYNSEM} | \text{LOC} | \text{CAT} | \text{HEAD} \ conj] \\ \text{CONJUNCT-DTR2} & \text{phrase} \end{cases} \]

4. HeadSpecifier Rule

\[ \text{hd-spr-ph} \Rightarrow \begin{cases} \text{HD-DTR} & 2 \ [\text{phrase} \\ \text{SPEC-DTR} & 1 \ [\text{SYNSEM} | \text{LOC} | \text{CAT} | \text{HEAD} | \text{SPEC} 2] \end{cases} \]
5. Adjunct_head and Head_adjunct rules

adjunct-hd-ph \rightarrow \left[ \text{ADJUNCT-DTR} \quad \text{[SYNSEM|LOC|CAT|HEAD|MOD} \ 3, \quad \text{pre_mod_synsem} \right]
\left[ \text{HD-DTR} \quad 2 \quad \text{[phrase]} \right]

hd-adjunct-ph \rightarrow \left[ \text{HD-DTR} \quad 2 \quad \text{[phrase]} \right]
\left[ \text{ADJUNCT-DTR} \quad \text{[SYNSEM|LOC|CAT|HEAD|MOD} \ 3, \quad \text{post_mod_synsem} \right]
Appendix G

Intelligibility and Fidelity Evaluation

This appendix contains the evaluation of ALMT with respect to *intelligibility* and *fidelity* as outlined in Chapter 8.

**Source Text:**

Sentence 77- 90: Newspapers and Articles.

**Translators:**

HM1: Thai Linguist
HM2: Thai Student (Graduate School)

**Legends:**

☐ Inappropriate word usage: This word does not distort the meaning of the expression but it is inappropriate.

Incorrectly translated word: This word is incorrectly translated. It causes the sentence to be meaningless or the sentence conveys a different meaning from the original.

玕 Misplaced word: The word is wrongly placed and it affects the meaning of the sentence. It should be moved in the direction of the arrow.

§ Misplaced word: The word is wrongly placed and it affects the meaning of the sentence. It should be moved in the direction of the arrow.
1. SL: It was a warm sunny day.

TC: นั้น (it) เป็น (was) วัน (day) ที่ (which) สุดใส (sunny) อบอุ่น (warm) วัน (classifier) หนึ่ง (a)

HM1: นั้น วันที่ อากาศ อบอุ่น แสงแดด จัดจ้านวันหนึ่ง
HM2: วันที่แสงแดดจัดจ้าน

No repair required.

2. SL: A mother duck sat on her nest at the edge of a pond waiting for her eggs to hatch.

TC: แม่ (mother) เป็ด (duck) ตัว (classifier) หนึ่ง (a) ให้นั่ง (sat)บน (on) รั่ง (nest) ของเธอ (her) ที่ (at) ขอบ (edge) ของ (classifier) หนึ่ง (the) ของ (of) แอนั่ง (pond) แอน (classifier) หนึ่ง (a) รอ (waiting) สำหรับ (for) ไข่ (egg) หลาย (plurality) ห่อ (classifier) ของเธอ (her) ที่จะ (to) หัก (hatch)

HM1: แม่ เป็ด ตัวหนึ่ง กอดไข่ ของแม่ อยู่ใน รั่ง ข้างๆ สะบาย แม่ กำลัง รอ ให้ ไข่ หัก
HM2: แม่ เบ็ด หนึ่ง กอดไข่ บน รั่ง ของแม่ ที่ วัน นั่ง เพื่อ รอ ให้ ไข่ หัก

No repair required.

3. SL: one_by_one, the eggs began to crack_open.

TC: วันๆ (one_by_one), ไข่ (egg) หลาย (plurality) ห่อ (classifier) หนึ่ง (the) ใคร่เริ่ม (began) ที่จะ (to) ปริแตกออก (crack_open)

HM1: วันๆ ตัวๆ ไข่ แตก เปิดออก
HM2: วันๆ เริ่มที่ ตัวๆ เปิดออก

I misplaced word.

4. SL: Peep, Peep, said each little yellow duckling as it poked out its head and looked around.

TC: นี่ (peep), นี่ (peep), นู่นเป็ด (duckling) สีเหลือง (yellow) เล็ก (little) แต่ละ (each) ตัว (classifier) นู่น (said) หัว (head) ของแม่ (its) และ (and) มอง (looked) รอบๆ (around)

HM1: เป็นนี่ เป็นนี่ เล็กๆ ที่เล็กๆ สีเหลืองๆ ที่เล็กๆ สีเหลืองๆ หนึ่ง หัว หนึ่ง นู่น เปิดออกใส่ และ มอง ไปรอบๆ
HM2: นี่ๆ เล็กๆ กนธี่ ที่เล็กๆ สีเหลืองๆ แต่ละ ตัว ขณะที่ หนึ่ง หนึ่ง หัว และ มอง ไปรอบๆ

I misplaced word, I incorrectly translated word.

5. SL: Quack, quack, said the mother duck feeling very pleased with herself.

TC: ควัก (quack), ควัก (quack), แม่ (mother) เป็ด (duck) ตัว (classifier) นู่น (the) คุณ (said) ความดี (feeling) พอใจ (pleased) มากๆ (very) กับ (with) ตัวเธอเอง (herself)

HM1: ควักควักแม่เป็ด ร้อง แม่ ดีใจ ที่ ได้เห็น ลูกๆ
6. SL: The biggest egg hatched last of all
TC: ไข่ (egg) ที่ (which) ใหญ่ที่สุด (biggest) ได้พัก (hatched) ล่าสุด (last) ของ (of) ทั้งหมด (all)
HM1: ไข่ ที่ ใหญ่ที่สุด พัก ตัว เป็น ไปสุดท้าย
HM2: ไข่ ใหญ่ที่สุด พัก ตัว เป็น ไปสุดท้าย
Grammatically incorrect.

7. SL: But the duckling that poked out his head was not little and yellow like the others
TC: แต่ (but) ลูกเป็ด (duckling) ตัว (classifier) นั่น (the) ที่ (that) ได้จิ้งจอก (poked) ช้า (out) หัว (head) ของ (his) ไม่ (not) เล็ก (little) และ (and) สีเหลือง (yellow) เหมือน (like) หัวอื่น ๆ (the others)
HM1: แต่ เจ้า ลูกเป็ด ที่ โผล่ หัว ออกมา นั่น ตัว ไม่ เล็ก นัก และ ไม่มี สีเหลือง เหมือน หัวอื่น ๆ
HM2: แต่ ลูกเป็ด ตัว นั่น ไม่ ได้ หัว เล็ก และ สีเหลือง เหมือน หัวอื่น ๆ
1 misplaced word, 2 incorrectly translated words.

8. SL: He was big and gray and ugly
TC: เขา (he) ใหญ่ (big) และ (and) สีเทา (gray) และ (and) น่าเบื่อ (ugly)
HM1: มัน กลับ ตัว ใหญ่ มี สีเทา และ น่าเบื่อ
HM2: มัน กลับ ตัว ใหญ่ สีเทา และ น่าเบื่อ
No repair required.

9. SL: Quack, said the mother duck
TC: แคน (quack), แม่ (mother) เป็ด (duck) ตัว (classifier) นั่น (the) ซุย (said)
HM1: แคน แม่ เป็ด รอง
HM2: แม่ เป็ด รอง ออกมา ท่าน
No repair required.

10. SL: You are the ugliest duckling I have ever seen
TC: คุณ (you) เป็น (are) ลูกเป็ด (duckling) ที่ (which) ที่สุดที่สุด (ugliest) อัน (I) เห็น (have ever) เห็น (seen)
HM1: คุณ เป็น ลูกเป็ด ที่ น่าเบื่อที่สุด เห็น ท่าน เห็น เห็น
HM2: คุณ เป็น ลูกเป็ด ที่ น่าเบื่อที่สุด ที่ ดี เห็น เห็น
11. SL: The mother duck led her ducklings into the water.
TC: มам่า (mother) เลี้ยง (duck) ตัว (classifier) นั้น (the) ได้พา (led) ลูกเป็ด (duckling) หลาย (plurality) ตัว (classifier) ของเธอ (her)

HM1: ผม แตะ พา ลูกๆ ลงไป ใน น้ำ
HM2: ผม เลี้ยง ลูก ลงไป ใน น้ำ

No repair required, inappropriate word usage.

12. SL: The other ducks gathered around.
TC: อีก (duck) ตัวอื่นๆ (the other) หลาย (plurality) ตัว (classifier) ได้จับกลุ่ม (gathered) รอบๆ (around)

HM1: ได้ตัว อื่นๆ ใน สาวย่าง เขามา ใกล้ๆ แน่ๆ และ ลูกๆ
HM2: ตัว อื่นๆ เริ่ม รวมตัว กัน

No repair required.

13: SL: what lovely children you have, they told her.
TC: เด็กๆ (children) ที่ (which) น่ารัก (lovely) อะไร (what) คุณ (you) มี (have), พวกเขา (they) ได้บอก (told) เธอ (her)

HM1: เด็ก มี ลูกๆ ที่ น่ารัก จิ๋วๆ เปิด เพิ่มเติม ชุด งับ แน่ เลี้ยง
HM2: ลูกๆ เลี้ยง น่ารัก ทั้งนั้น เยี่ยง เลี้ยง ตัวอื่น พร้อมกัน บอก เลย

2 misplaced words.

14. SL: But when they saw the big gray duckling, they began to laugh.
TC: แต่ (but) เมื่อ (when)พวกเขา (they) ได้เห็น (saw) ลูกเป็ด (the) สีเทา (gray)โต (big) ตัว (classifier) นั้น (the), พวกเขา (they) ได้เริ่ม (began) ที่จะ (to) หัวเราะ (laugh)

HM1: แต่ เมื่อ เลี้ยง พวกนั้น ได้เห็น ลูกเป็ด สีเทา ตัวใหญ่ พวกนั้น หัวเราะ
HM2: แต่ เมื่อ มองเห็น ลูกเป็ด สีเทา ตัวใหญ่ พวกนั้น หัวเราะ
t
No repair required.

15. SL: That one is not lovely, they said
TC: ว่า (that) หนึ่ง (one) ไม่ (not) น่ารัก (not), พวกเขา (they) ชุด (said)

HM1: เลี้ยง ตัว นี้ ไม่ น่ารัก เลย เลี้ยง ตัวใหญ่ ชุด
HM2: มี ตัว นี้ แหละ ที่ ไม่ น่ารัก
2 incorrectly translated words.

16. SL: what an ugly duckling he is !
   TC: ลูกเป็ด (duckling) ที่ (which) ซ่า เพราะ (ugly) ตัว (classifier) หนึ่ง (an) อะไร (what) เขา (he) เป็น (is) เลย หรือ (!)
   HM1: ลูกเป็ด อะไร นะ น่าเกลียด จริงๆ
   HM2: ลูกเป็ด อะไร นะ น่าเกลียด จริงๆ

2 misplaced words.

17. SL: The mother duck took her ducklings to the farmyard.
   TC: แม่ (mother) เป็ด (duck) ตัว (classifier) หนึ่ง (the) ได้ (took) ลูกเป็ด (duckling) หลาย (plurality) ตัว (classifier) ของเธอ (her) ไปที่ (to) ลานกิจการ (farmyard) แห่ง (classifier) หนึ่ง (the)
   HM1: แม่ เป็ด พวกเธอ น่าเกลียด น่ารำคาญ แท้ หนึ่ง
   HM2: แม่ เป็ด พวกเธอ น่าเกลียด แท้ หนึ่ง

No repair required, 1 inappropriate word usage.

18. SL: When the other animals saw the big gray duckling, they all began to shout.
   TC: เมื่อ (when) สัตว์ (animal) หลาย (plurality) ตัว (classifier) ตัวอื่น (the other) ได้เห็น (saw) ลูกเป็ด (duckling) ที่ (which) สีเทา (gray) ใด (big) ตัว (classifier) หนึ่ง (the), พวกเขานะ (they) หั่นหน้า (all) ได้เริ่ม (began) ที่จะ (to) ตะโกน (shout)
   HM1: ครั้น สัตว์ ตัวอื่น เห็น ลูกเป็ด สีเทา ตัว พวกพวกเธอ พากัน พากัน ตะโกน ตั้ง
   HM2: เมื่อ สัตว์ อื่น เห็น ลูกเป็ด สีเทา ถึง พากัน ห้ามไว้ เลยให้

1 misplaced word.

19. SL: What an ugly duckling he is ! they said.
   TC: ลูกเป็ด (duckling) ที่ (which) ซ่า เพราะ (ugly) ตัว (classifier) หนึ่ง (an) อะไร (what) เขา (he) เป็น (is) เลย หรือ (!) พากัน (they) ได้สุญ (said)
   HM1: ลูกเป็ด อะไรกันนี้ น่าเกลียด จริงๆ พากัน พุ่งชัน
   HM2: ลูกเป็ด อะไรนะ น่าเกลียด จริงๆ

2 misplaced words.

20. SL: He does not belong with us.
   TC: เขา (he) ไม่ (not) ใช่ส่วนหนึ่งของ (belong_with) พวกเรา (us)
   HM1: แล้ว ไม่ใช่ พวกเดียวกัน เขา นะ
21: SL: The geese and the turkeys ran after the duckling and tried to chase him away.
   TC: พบกัน (geese) นั้น (the) และ (and) ไก่งวง (turkey) หลายคน (plurality) ตัว (classifier) นั้น (the) ได้พยายาม (tried) ตูเกิด (duckling) ตัว (classifier) นั้น (the) และ (and) ไก่งวง (tried) ที่จะ (to) ไล่ (chase) นั้น (him) ทำงานไป (away)
   HM1: พบกับ ไก่งวง ทั้งสิ้น ตูเกิด และ พยายาม มัน ออกมา ฝืน
   HM2: พบกับ ไก่งวง ตั้ง บึ้งได้ มัน ไล่ออกมา

   Grammatically incorrect.

22. SL: Even his own brothers and sisters joined in and called him names.
   TC: แม้ (even) น้องชายของเขา (brother) หลายคน (plurality) คน (classifier) ของเขา (his) และ (and) น้องสาวของเธอ (sister) หลายคน (plurality) คน (classifier) เข้ามา (joined in) และ (and) ตั้งคำถามของเธอ (called him names)
   HM1: แม้แต่ น้องชาย ที่ เป็น น้องชายของนั้น เอง อย่างน้อย พ่อแม่ ทั้งสอง และ คู่หู มัน
   HM2: แม้แต่ น้องชาย ของนั้น ที่ถามคำถาม ถ้า ต่างๆ

   Grammatically incorrect.

23. SL: The ugly duckling hid in a corner of the farmyard and cried.
   TC: ตูเกิด (duckling) ที่ (which) ตูเกิด (ugly) ตัว (classifier) นั้น (the) ซ่อน (hid) ใน (in) มุม (corner) ตูเกิด (classifier) หนึ่ง (a) ของ (of)
   ลานพื้นที่ Expert (farmyard) ตูเกิด (classifier) นั้น (the) และ (and) ได้ร้องไห้ (cried)
   HM1: เธอตูเกิด ซ่อน ที่ หนึ่ง ของ พื้นที่ แล้ว เริ่ม ร้องไห้
   HM2: ที่ หนึ่ง ของ พื้นที่ เธอ นั้น เธอ ร้องไห้

   No repair required.

24.1. SL: The next morning,
   TC: เช้า (morning) เวลา (classifier) ตื่นไป (next) นั้น (the)
   HM1: เช้า เช้า วัน วัน
   HM2: เช้า วัน เช้าเช้า

   No repair required.

24.2. SL: The ugly duckling decided to run away
   TC: ตูเกิด (duckling) ที่ (which) ตูเกิด (ugly) ตัว (classifier) นั้น (the) ได้ตัดสินใจ (decided) ที่จะ (to) หนีจากไป (run away)
25. SL: I will find a place where no_one will make_fun_of me, he said to himself.

No repair required.

26. SL: He left the farmyard and ran to the great marsh where the wild ducks lived.

No repair required.

27. SL: Perhaps they will be kind to me, he said sadly.

I incorrectly translated word.

28. SL: But the wild ducks were not kind.

No repair required.

29. SL: You are very ugly, they said.
30. SL: You are certainly not one of us
TC: ศูนย์ (you) เป็น (are) อย่างแน่นอน (certainly) ไม่ (not) หนึ่ง (one) ของ (of) พวกเรา (us)
HM1: แต่ ไม่ใช่ พวกเรา แน่นๆ
HM2: แต่ ไม่ใช่ หนึ่ง ของ พวกเรา แน่นๆ

2 misplaced words.

31. SL: And the wild ducks laughed at the duckling
TC: และ (and) เป็ด (duck) บ้า (wild) หลาย (plurality) ตัว (classifier) นั้น (the) ได้หัวเราะเลย (laughed at) สู่ก้อง (duckling) ตัว (classifier) นั้น (the)
HM1: เ桀รย์ เป็ด บ้าก็ พวกนั้น หัวเราะเลย สู่ก้อง ซึ่งนี้
HM2: เป็ด บ้าก็ พวกนั้น หัวเราะเลย สู่ก้อง

No repair required

32. SL: So the ugly duckling left the marsh.
TC: ดังนั้น (so) สุกเกิด (duckling) ซึ่ง (which) ซึ่งนี้ (ugly) ตัว (classifier) นั้น (the) ใต้ (left) หน้า (marsh) หนั่ง (classifier) นั้น (the)
HM1: สุกเกิด ซึ่งนี้ ซึ่งนี้ ขี้เกียจ ขี้เกียจ หนัง หนัง นั้น
HM2: สุกเกิด ซึ่งนี้ ซึ่งนี้ ขี้เกียจ ขี้เกียจ หนังๆๆ นั้น

No repair required

33. SL: the wind was blowing and it was cold.
TC: ลม (wind) นั้น (the) กำลัง (continuous tense)พุ่ง (blow) และ (and) มัน (it) หน้า (cold)
HM1: มี การพุ่ง พัดมา และ มัน อากาศ เริ่ม เริ่ม เริ่ม ลง
HM2: ลม ที่ แรง และ มัน ที่ หนาว

No repair required

34. SL: Winter was coming
TC: ฤดูหนาว (winter) กำลัง (continuous tense) มา (come)
39. SL: An old woman lived in the cottage with a fat black cat and a plump brown hen.
   TC: หญิง (woman) เด็ก (old) คน (classifier) นั้น (an) ใครอาศัยอยู่ (lived) ใน (in) กระท่อม (cottage) นั้น (the) ด้วย (with) แมว (cat) สีดำ (black) อ้วน (fat) ตัว (classifier) นั้น (a) และ (and) นก (hen) สีน้ำตาล (brown) อ้วน (plump) ตัว (classifier) นั้น (a)
   HMI: มี หญิงสาว อาศัยอยู่ ใน กระท่อม นาง หญิง ด้วย แมว ตัว สีดำ และ นก ตัว สีน้ำตาล
   HM2: หญิงสาว อาศัยอยู่ ใน กระท่อม กัน แมว ตัว สีดำ อ้วน และ นก ตัว สีน้ำตาล อ้วน
   * Incorrectly translated word.

40. SL: The old woman did not mind the ugly duckling but the cat and the hen wanted him to leave.
   TC: หญิง (woman) นั้น (the) ไม่ (not) รังเกียจ (mind) นกเท้า (duckling) นั้น (the) แต่ (but) แมว (cat) ตัว (classifier) นั้น (the) และ (and) นก (hen) ตัว (classifier) นั้น (the) ให้ต้องการ (wanted) เขา (him) กัน (to) จบ (leave)
   HMI: หญิงสาว ไม่ รังเกียจ เจ้า นกเท้า เท่าๆ แต่ แมว และ นก ของนาง ต้องการ ให้ เจ้า นกเท้า ออกไป จาก กระท่อม
   HM2: หญิงสาว ไม่ได้ ว่า อะไรค่ะ แมว และ นก ต้องการ ให้ แมว ออกไป
   * Incorrectly translated word.

41. SL: Can you purr ? asked the cat
   TC: คุณ (you) สามารถ (can) ทำเสียงเพียงบาง (purr) ได้ไหม (?) แมว (cat) ตัว (classifier) นั้น (the) ได้ถาม (asked)
   HMI: แต่ ร้องอย่างบาง ได้ไหม แมวในขณะ
   HM2: เอะ ร้องอย่างบาง ได้ไหม
   No repair required.

42. SL: No, said the ugly duckling.
   TC: ไม่ (no), นกเท้า (duckling) นั้น (the) ซึ่ง เท่า (ugly) ตัว (classifier) นั้น (the) ได้พูด (said)
   HMI: ไม่ได้พูด นกเท้า ซึ่ง เท่า บอก
   HM2: ไม่ได้พูด นกเท้า ซึ่ง เท่า บอก
   No repair required.

43. SL: Can you lay eggs ? asked the hen.
   TC: คุณ (you) สามารถ (can) วาง (lay) ไข่ (egg) หลาย (plurality) ของ (classifier) ได้ไหม (?) ให้ (hen) ตัว (classifier) นั้น (the) ได้ถาม (asked)
   HMI: แค่ วาง ใช้ ได้ไหม แม่ ให้ ถาม บ้าง
   HM2: แค่ อย่า ใช้ ได้ไหม
   No repair required.
35. SL: The ugly duckling was very tired.
   TC: นู (duckling) ที่ (which) ซี่ (ugly) ตัว (classifier) นั้น (the) เหมือน (tired) มาก (very)
   HM1: นู ซี่ เหมือน เหลือ ที่ การ (duckling) ซี่ (ugly) ตัว (classifier) นั้น (the) เหมือน (tired) มาก (very)
   HM2: นู ซี่ เหมือน เหลือ ที่ การ
   No repair required.

36. SL: He had to find a place to stay where he could be warm and safe.
   TC: เขา (he) ต้อง (had_to) หา (find) ที่ (place) ที่ (classifier) นั้น (a) ที่จะ (to) อยู่ (stay) ที่ซึ่ง (where) เขา (he) จะได้ (could_be) อยู่ (warm) และ (and) ปลอดภัย (safe)
   HM1: มัน ต้อง หา ที่ ซิ่ง ที่ (he) อยู่ ที่ (a) ที่จะ อยู่ (stay) ที่ซึ่ง (where) เขา (he) จะได้ (could_be) อยู่ (warm) และ (and) ปลอดภัย (safe)
   HM2: มัน ต้อง หา ที่ ซิ่ง ที่ (he) อยู่ ที่ (a) ที่จะ อยู่ (stay) ที่ซึ่ง (where) เขา (he) จะได้ (could_be) อยู่ (warm) และ (and) ปลอดภัย (safe)
   No repair required.

37. SL: One evening, he came to a little cottage.
   TC: หนึ่ง (one) เย็น (evening), เขา (he) ใครมา (came) กับ (to) กระท่อม (cottage) เล็ก (little) หลัง (classifier) นั้น (a)
   HM1: เย็น หนึ่ง หิ้ง นูเปิด เล็ก มา ถึง กระท่อม เล็กๆ หลัง หนึ่ง
   HM2: เย็น หนึ่ง หิ้ง นูเปิด มา ถึง กระท่อม เล็ก หลัง หนึ่ง
   I misplaced word, I incorrectly translated word.

38. SL: The door was opened and he went inside.
   TC: ประตู (door) เปิด (classifier) นั้น (the) ถูก (passive voice) เปิด (open) และ (and) เขา (he) ไป (went) ข้างใน (inside)
   HM1: ประตู กระท่อม เปิด อยู่ มัน จึง เข้าไป ข้างใน
   HM2: ประตู เปิด อยู่ มัน จึง เข้าไป
   No repair required.
44. SL: No, said the ugly duckling.
   TC: ไม่ (no) ดุ๊กเนี้ย (duckling) ที่ (which) ซึ่งะ (ugly) ตัว (classifier) นั้น (the) ได้พูด (said)
   HM1: ไม่ได้ที่ตัวดุ๊กเนี้ย ดุ๊กเนี้ยซึ่งะ ผีดฝัน
   HM2: ไม่ได้บรรยาย ดุ๊กเนี้ย ซึ่งะ พูด
   No repair required.

45. SL: Then you must go, said the cat and the hen.
   TC: แล้ว (then) แมว (cat) ต้อง (must) ไป (go), แม่ (cat) ตัว (classifier) นั้น (the) และ (and) ไก่ (hen) ตัว (classifier) นั้น (the) ได้พูด (said)
   HM1: แล้วแมว แมวจะออกไป เลยไป เม้าะ และ แม่ไก่ จะออกไปได้
   HM2: ฉันยังคงแมว เลย ต้องไป เม้าะ และ แม่ไก่ ทุก
   Convey different meaning from the original.

46. SL: They chased the ugly duckling out of the cottage and would not let him in again.
   TC: พวกเขารวม (they) ได้ไล่ (chased) ดุ๊กเนี้ย (duckling) ที่ (which) ซึ่งะ (ugly) ตัว (classifier) นั้น (the) ออกจาก (out_of) กระท่อม (cottage) หลัง (classifier) นั้น (the) และ (and) จะ (will) ไม่ (not) ยอมให้ (let) เข้า (him) เข้ามา (in) อีกครั้ง (again)
   HM1: พวกเขามา กลับไป ดุ๊กเนี้ย ซึ่งะออกจากกระท่อม และ ไม่อยู่ ให้ ผีดฝัน เขาไป อีก
   HM2: แม่และแม่ไก่ เลยไล่ ดุ๊กเนี้ย ซึ่งะ ให้ออกไปจากกระท่อมและไม่อยู่ให้เข้ามาอีก
   I inappropriate word usage.

47. SL: The ugly duckling wandered through the countryside for many days.
   TC: ดุ๊กเนี้ย (duckling) ที่ (which) ซึ่งะ (ugly) ตัว (classifier) นั้น (the) ได้เดินไปเดินมา (wandered) ผ่าน (through) ชนบท (countryside) หลัง (classifier) นั้น (the) เป็นเวลา (for) หลาย (many) วัน (day)
   HM1: ดุ๊กเนี้ย ซึ่งะ เดินทางไป จนถึง หลายวัน เป็น เวลา หลาย วัน
   HM2: ดุ๊กเนี้ย ซึ่งะ เดินทาง ขอนมา ตาม ชนบท อยู่ หลาย วัน
   Grammatically incorrect.

48. SL: No one would give him a home.
   TC: ไม่มีใคร (no_one) จะ (would) ให้ (give) เขา (him) บ้าน (home) หลัง (classifier) นั้น (a)
   HM1: ไม่มีใคร ให้ผีดฝัน เขาอยู่ที่บ้าน
   HM2: ไม่มีใคร ยอมให้ อยู่ที่บ้าน ด้วย
   I incorrectly translated word.
49. SL: No one would be his friend.
TC: ไม่มีใคร (no_one) จะ (would) เป็น (be) เพื่อน (friend) ของเขา (his)
HM1: ไม่มีใคร เป็น เพื่อน กับ มัน
HM2: ไม่มีใคร อย่างก็ เป็น เพื่อน กับ คุณ
No repair required.

50. SL: All the birds and animals said that he was ugly
TC: นก (bird) ทั้งหมด (all) มัน (the) และ (and) สัตว์ (animal) หลาย (plurality) ตัว (classifier) ได้ยิน (said) ว่า (that) เขา (he) ซึ้ง (ugly)
HM1: นก และ สัตว์ ทุก ตัว ต่าง พวกเขามัน ๆ รู้ ว่า มัน เป็น สัตว์ ที่ ไม่เกลียด
HM2: นก และ สัตว์ ตัว ๆ พวกเขามัน ๆ บอก ว่า มัน ไม่เกลียด
No repair required.

51. SL: He grew more and more sad and more and more tired
TC: เขา (he) ได้รู้สึกขึ้น (grew) อีก (more) และ (and) เสร็จ (sad) มากขึ้น (more) และ (and) มาหนึ่ง (more) และ (and) เหลือ (tired) มาขึ้น (more)
HM1: คุณเปิด ซึ่ง เขา ต่าง ๆ ตัว เหลือ มัน ครับ เรารู้ชิ้น และ บอก เขามัน ยังกว่า เเตะ
HM2: มัน เรั่ม เรารู้ชิ้นและ เหลือมากขึ้น
Convey different meaning from the original.

52. SL: at last he stopped to rest by a lake that was covered with ice
TC: ที่สุด (at_last) เขา (he) หยุด (stopped) เพื่อที่จะ (to) พัก (rest) โดย (by) ทะเลสาบ (lake) แห้ง (classifier) หนึ่ง (a) ว่า (that) ถูก (passive voice) ปกคลุม (cover) ด้วย (with) น้ำแข็ง (ice)
HM1: แล้ว ที่สุด มัน ถึง หยุด เริ่ม และ หา ที่พัก อยู่ ใกล้ ๆ ทะเลสาบ ที่ ปกคลุม ด้วย น้ำแข็ง
HM2: ที่สุด มัน ถึง หยุดพัก ที่ ทะเลสาบ ที่ ปกคลุม ด้วย น้ำแข็ง เหลือหนึ่ง
2 incorrectly translated word.

53. SL: He stayed there all winter, hidden in the reeds
TC: เขา (he) อยู่ (stayed) ที่นั่น (there) ตลอด (all) ฤดูหนาว (winter), ซ่อน (hidden) ใน (in) ต้นไม้ (reed) หลาย (plurality) ตัน (classifier) นั้น (the)
HM1: มัน อาศัย อยู่ ที่นั่น ตลอด ฤดูหนาว และ ซ่อน ตัว อยู่ ไม่ได้ ต้นไม้
HM2: มัน อาศัย ที่นั่น ตลอดทั้ง หน้าหนาว ซ่อน ตัว ใน หญ้า
No repair required.
54. SL: He was cold and hungry and very lonely.

TC: เขา (he) หนาว (cold) และ (and) หิว (hungry) และ (and) เหนื่อย (lonely) มากๆ (very)

HM1: สูงเกิน งาน ท่าน สิ่ง และ รู้สึก ว่าแหวก
HM2: มัน หนาว หิว และ ร่างกาย มาก

No repair required.

55. One bright morning in early spring, when all the ice had melted, a flock of beautiful birds came to the lake.

TC: หนึ่ง (one) ค่ (which) ว่าง (bright) เขา (morning) ใน (in) ตอนต้น (early) ถูปไปไม่ได้ (spring), เมื่อ (when) น้ำแข็ง (ice) แข็งคง (all)
นั้น (the) ได้ (had) ละลาย (melt) นก (bird) ค่ (which) สวย (beautiful) ของ (of) ฝูง (flock) ตัว (classifier) หนึ่ง (a) หลาย (plurality)
ได้มา (came) ถึง (to) ทะเลสาบ (lake) แห่ง (classifier) มัน (the)

HM1: ฉัน รว่าง เขา ที่ สดน้ำ วัน หนึ่ง ต้น ถูปไปไม่ได้ เมื่อ น้ำแข็ง ใน ทะเลสาบ ละลาย ไป หมด ฝูง นก ที่ ลำานาน ฝูง หนึ่ง ได้ บิน มา
ยัง ทะเลสาบ
HM2: เขา แพะใส วัน หนึ่ง ใน ต้น ถูปไปไม่ได้ เมื่อ น้ำแข็ง ได้ ละลาย หมด แล้ว ฝูง นก ลำานาน ฝูง หนึ่ง บิน มา ยัง ทะเลสาบ

Grammatically incorrect.

56. SL: They had gleaming white feathers and long graceful necks.

TC: พวกเขามี (they) มี (had) ขน (feathers) ค่ (which) สวย (gleaming) สีขาว (white) หลาย (plurality) เต็น (classifier) และ (and)
คอ (neck) ค่ (which) ยาว (long) อ้อมชาย (graceful) หลาย (plurality) คอ (classifier)

HM1: พวกนั้นมี ขน สีขาว แย้มยิ่ง และ ลำคอก ที่ กลาง กลาง
HM2: พาเละเขามี ขน สีขาว แย้มยิ่งและ คอที่ยาว สาง

No repair required.

57. SL: The ugly duckling watched them from his hiding place in the reeds.

TC: ตุ๊กแก (duckling) ค่ (which) ตุ๊กแก (ugly) ตัว (classifier) นั้น (the) ได้เน่าสุ (watched)พวกเขามี (them) จาก (from) สวนที่ (place) การซ่อน (hiding)ของนก (his) ใน (in) ต้นกล (reed) หลาย (plurality) ตัว (classifier) นั้น (the)

HM1: ตุ๊กแก ซ่อน เน่าสุ พวกนี้ จาก สวนที่ ใกล้ ต้นกล
HM2: ตุ๊กแก ซ่อน เน่าสุ พวกนี้ จาก สวนที่ ใกล้ ต้นกล

I incorrectly translated word.

58. SL: They are the most beautiful birds I have ever seen, he thought.

TC:พวกเขามี (they) เป็น (are) นก (bird) ค่ (which) สวย (beautiful) ที่สุด (most) หลาย (plurality) ตัว (classifier) หนึ่ง (1) เบีย (have) เห็น (seen), เขา (he) ได้คิด (thought)
59. SL: If only I could be like them.
   TC: ถ้า (if) เหล่านี้ (only) นั้น (I) สามารถ (could) ชอบ (like) พ่อเขา (them)
   HM1: ถ้า นั้น อย่าง พวกนี้ บ้าง กี่ ครั้ง จะ ถ้า ขี้
   HM2: ถ้า นั้น ได้ เป็น อย่าง พวกเข้า
   I incorrectly translated word.

60. SL: the beautiful birds saw the ugly duckling.
   TC: นก (bird) ที่ (which) สวยงาม (beautiful) หลาย (plurality) ตัว (classifier) นั้น (the) ได้เห็น (saw) ลูกเป็ด (duckling) ที่ (which) ขี้เรื่ย (ugly) ตัว (classifier) นั้น (the)
   HM1: นก ที่ สวยงาม นุ่ง นั้น มองเห็น ลูกเป็ด ขี้เรื่ย
   HM2: นก ที่ สวยงาม นุ่ง มองเห็น ลูกเป็ด ขี้เรื่ย
   No repair required.

61. SL: The biggest one came_over_to him
   TC: ตัว (one) ที่ (which) ใหญ่สุด (biggest) ตามมาที่ (came_over_to) เขา (him)
   HM1: นก ตัว ที่ ใหญ่สุด ไป สู่ บน เขามา หา ผู้อาน
   HM2: นก ตัว ใหญ่สุด เขามา หา
   No repair required.

62. SL: The ugly duckling tried to hide
   TC: ลูกเป็ด (duckling) ที่ (which) ขี้เรื่ย (ugly) ตัว (classifier) นั้น (the) ได้พยายาม (tried) ที่จะ (to) ซ่อน (hide)
   HM1: ลูกเป็ด ขี้เรื่ย พยายาม จะ ซ่อน ตัว
   HM2: ลูกเป็ด ขี้เรื่ย พยายาม จะ หลบ
   No repair required.
63. SL: He was sure that the bird would make fun of him or try to chase him away.
TC: เขา (he) แน่ใจ (sure) ว่า (that) นก (bird) ตัว (classifier) นั้น (the) จะ (would) ทำ (make fun of) เขา (him) หรือ (or) พยายาม (try) ที่ (to) ไป (chase) เขา (him) ทำให้ (away)

HM1: ผม แน่ใจ ว่า นก พาหนะ ต้อง ทำ อะไร หรือ ไม่ ก็ ขับไล่ ใหม่ ถ้า
HM2: ผม แน่ใจว่า นก ตัว ใหม่ ตัว นั้น จะ ต้อง ทำอะไร และ ไม่ มัน ไป ให้ ผ่าน

I incorrectly translated word.

64. SL: But the lovely white bird said, what a fine young swan you are!
TC: แต่ (but) นก (bird) ที่ (which) สีขาว (white) น่ารัก (lovely) ตัว (classifier) นั้น (the) ได้พูด (said), หงส์ (swan) หนุ่ม (young) จังวาม (giggle) ตัว (classifier) หนึ่ง (one) อะไร (what) คุณ (you) เป็น (are) เฉยๆ (!)

HM1: แต่ นก สีขาว ที่ สวยงาม กับ หงส์ บอก ว่า เจ้า เป็น หงส์ น้อย ที่ สวยงาม มาก
HM2: แต่ นก ตัว สว่าง สีขาว หงส์ บอก เสด็จ ข้าว เป็น หงส์ หนุ่มที่ สวย จริงๆ

3 misplaced words.

65. SL: You should not be here all by yourself.
TC: คุณ (you) ไม่ (not) ควร (should) อีก (be) ที่นี่ (here) อยู่ด้วย (all_by_yourself)

HM1: เจ้า ไม่ น่า อยู่ ที่นี่ ตามสั่ง พ่อกำ นะ
HM2: เธอ ไม่ ควร อยู่ ที่นี่ ตามสั่ง

No repair required.

66. SL: The ugly duckling could not believe what he had heard.
TC: ลูกเป็ด (duckling) ที่ (which) ซี่โครง (ugly) ตัว (classifier) นั้น (the) ไม่ (not) สามารถ (could) เชื่อ (believe) อะไร (what) เขา (he) ได้ (had) ได้ยิน (heard)

HM1: ลูกเป็ด ซี่โครง แน่น ไม่ เชื่อ ฟัง ที่ มัน ได้ยิน
HM2: ลูกเป็ด ซี่โครง ไม่ ยิน รู้ ด้วย ซึ่ง ที่ มัน ฟัง ได้ยิน

I incorrectly translated word.

67. SL: He looked at his reflection in the water.
TC: เขา (he) มอง (looked) ไปที่ (at) ตัวเอง (reflection) ของเขา (his) ใน (in) น้ำ (water) นั้น (the)

HM1: ผม มอง หน้า ของ ตัวเอง ใน น้ำ
68. SL: It was true.  
TC: มัน (it) จริง (true)  
HMI: จริงๆ ค่ะ  
HM2: จริงๆ ค่ะ  
No repair required.

69. SL: His feathers were white and gleaming.  
TC: ขน (feathers) หลาย (plurality) เลือด (classifier) ของ (his) สีขาว (white) และ (and) เป็นประกาย (gleaming)  
HMI: ขน ของ (his) สีขาว และ เข็ม (is not in use)  
HM2: ขน (his) เป็น สีขาว และ สะท้อนแสง เล็กน้อย  
No repair required.

70. SL: His neck was long and graceful.  
TC: คอ (neck) ของ (his) ยาว (long) และ (and) สวยงาม (graceful)  
HMI: คอ (his) ยาว และ สวยงาม  
HM2: คอ (his) ยาว และ สังว  
No repair required.

71. SL: He had grown into a handsome swan.  
TC: เขา (he) โต (had) ใบ้ขึ้น (grown) เป็น (into) หงส์ (swan) บุรุษงาม (handsome) ตัว (classifier) หนึ่ง (a)  
HMI: มัน กลายเป็น หงส์ ที่ สวยงาม ตัว หนึ่ง  
HM2: มัน โต เดิมโต เป็น หงส์ หนุ่ม ที่ สวยงาม ตัว หนึ่ง  
I inappropriate word usage.

72. SL: Come with us, said the swan.  
TC: มา (come) กับ (with) พบกับ (us) หงส์ (swan) ตัว (classifier) มัน (the) ได้ฟัง (said)  
HMI: มา กับ พบกับ ที่ หงส์ ตัว หนึ่ง เยอะ ปาก ชวน  
HM2: มา กับ พบกับ ที่
No repair required.

73. SL: The ugly duckling spread his wings and rose into the sky
   TC: ลูกเป็ด (duckling) ที่ (which) นิ่ว (ugly) ตัว (classifier) นิ้ว (the) วาง (spread) ปีก (wing) หลาย (plurality) ปีก (classifier) ของเข่า (his) และ (and) ขึ้นฝู (rose_into) ท้องฟ้า (sky) นั้น (the)
   HM1: ลูกเป็ด ซึ่งมี นิ่ว ของมัน ออก และ ใน ขึ้น ขึ้น ท้องฟ้า
   HM2: ลูกเป็ด ซึ่งมี นิ่ว ของมัน ขึ้น ไป บน ท้องฟ้า
   No repair required.

74. SL: I never dreamed that i could be so happy, he said, and he flew off proudly with his new friends.
   TC: ฉัน (I) ไม่เคย (never) ได้ฝัน (dreamed) ถ้า (that) ฉัน (I) จะได้ (could_be) ตั้งนั่น (so) เป็นสุข (happy), เขา (he) ไวกูศุ (said), และ (and) เขา (he) ได้บิน (flew) ออกไป (off) อย่างภาคภูมิ (proudly) ด้วย (with) เพื่อน (friend) ใหม่ (new) หลาย (plurality) ตัว (classifier) ของเข่า (his)
   HM1: ฉัน ไม่เคย คิด เลย ว่า จะ มี ความสุข อย่างนี้ มัน ดีและ ยิ่ง ไป ใน ท้องฟ้า อย่าง สิ่งของใจ กับ เพื่อน ใหม่ ของมัน
   HM2: ฉัน ไม่เคย คิด มาผัก เลย ว่า จะ มี ความสุข ขนาดนี้ มัน ดีและ ยิ่ง ไป ด้วย ความภูมิใจ พร้อม กับ เข้ามา ๆ ใหม่
   2 incorrectly translated words.

75. SL: The ugly duckling
   TC: ลูกเป็ด (duckling) ที่ (which) นิ่ว (ugly) ตัว (classifier) นิ้ว (the)
   HM1: ลูกเป็ด ซึ่งมี
   HM2: ลูกเป็ด ซึ่งมี
   No repair required

76. SL: Written by Ronne_Randall
   TC: เขียน (written) โดย (by) รอนนี เรนเดล (Ronne_Randall)
   HM1: เขียน โดย รอนนี เรนเดล
   HM2: เขียน โดย รอนนี เรนเดล
   No repair required.

77. SL: The wheat was yellow
   TC: ข้าวสาลี (wheat) นั้น (the) สีเหลือง (yellow)
   HM1: ข้าวสาลี ออก สีเหลืองอารม
   HM2: ข้าวสาลี สีเหลืองอารม
No repair required.

78. SL: Here the stork marched about on his long red legs talking Egyptian.
   TC: นี่ (here) นกกระสา (stork) ตัว (classifier) นั้น (the) เดินไปมา (marched about) บน (on) ขา (leg) สีแดง (red) ยาว (long) หลาย (plurality)
   ขา (classifier) ของเขานะ (his) คุณ (talking) ภาษาอียิปต์ (Egyptian)
   HM1: นกกระสา เยื้อนายาน หลาย ขา สีแดง ยาวๆ ของมัน พวกมัน คุณ ภาษาอียิปต์
   HM2: นกกระสา พาหุ บน ขา ที่ ยาว สีแดง ของมัน คุณ ก็เป็น ภาษาอียิปต์
   2 incorrectly translated words.

79. SL: When I was an ugly duckling, he thought, I never dreamed I could be so happy.
   TC: เมื่อ (when) ฉัน (I) เป็น (was) ตัวเป็ด (duckling) นี่ (which) ซึ่ง (ugly) ตัว (classifier) นั้น (an), เขาก็ (he) ได้คิด (thought), ฉัน (I) ไม่เคย
   (never) ได้ฝัน (dreamed) ฉัน (I) จะได้ (could be) ดังนั้น (so) เป็นสุข (happy),
   HM1: เมื่อครั้งที่ ฉัน เป็น ตัวเป็ด ซึ่ง ฉัน มัน คุณคิด อยู่ในใจ ฉัน ไม่เคย ฉัน ว่า จะ มี ความสุข มาก เลยนี่
   HM2: ตอน ที่ ฉัน เป็น ตัวเป็ด ซึ่ง ฉัน มัน คิด ฉัน ไม่เคย คิด ฉัน เลย ว่า ฉัน จะ มี ความสุข ขนาดนี้
   1 incorrectly translated word.

80. SL: You can take a fish to school but you can not make them think.
   TC: คุณ (you) สามารถ (can) เอามา (take) ปลา (fish) ตัว (classifier) นั้น (a) ไปที่ (to) โรงเรียน (school) แต่ (but) คุณ (you) ไม่ (not) สามารถ (can)
   ทำ (make) พวกเขา (them) คิด (think)
   HM1: คุณ อาจจะ พา หน้า ไป โรงเรียน ได้ แต่ คุณ สอน ให้ เขา คิด ไม่ได้
   HM2: คุณ พา ปลา ตัว หนึ่ง ไป โรงเรียน ได้ แต่ ทำ ให้ แม้ คิด ไม่ได้
   2 incorrectly translated words.

81. SL: Five days before the trout are released.
   TC: ห้า (five) วัน (day) ก่อน (before) ปลาแหวก (trout) ตัว (classifier) นั้น (the) ถูก (passive voice) ปล่อย (release)
   HM1: ห้า วัน ก่อน ที่ ปลาแหวก จะ ถูก ปล่อย
   HM2: ห้า วัน ที่ ปลาแหวก จะ ถูก ปล่อย
   Grammatically incorrect.

82. SL: I rent a room from Mrs Jones.
   TC: ฉัน (I) ค่าเช่า (rent) ห้อง (room) ห้อง (classifier) หนึ่ง (a) จาก (from) นาง (Mrs) โจนส์ (Jones)
   HM1: ฉัน เช่า ห้อง จาก มิสชิล โจนส์
83. SL: She was rending her hair out in anger
TC: เธ่ (she) กำลัง (continuous tense) ยึด (rend) ผม (hair) ออกจาก (her) ออกมานา (out) ใน (in) โทขัย (anger)
HMI: พวกผม กำลัง ที่มี ต่างๆ ความพอใจ
HM2: เธ่ กำลัง ที่มี ต่างๆ ความพอใจ
3 incorrectly translated words.

84. SL: Algebraic symbols are used when you do not know what you are talking about.
TC: ตัวแปร (symbols) ในทางคณิตศาสตร์ (algebraic) หลาย (plurality) คำ (classifier) ถูก (passive voice) ใช้ (use) เมื่อ (when) คุณ (you) ไม่ (not) รู้ (know) อะไร (what) คุณ (you) กำลัง (continuous tense) คุย (talk) เกี่ยวกับ (about)
HMI: ตัวแปร ในทางคณิตศาสตร์ จะถูก เปลี่ยน คุณ ไม่ รู้ ว่า จะ คุย เกี่ยวกับ อะไร
HM2: ตัวแปร ในทางคณิตศาสตร์ ถูก ใช้ เมื่อ คุณ ไม่ รู้ ว่า กำลัง คุย เกี่ยวกับ อะไร อยู่
1 mis placed word.

85. SL: Before he went fishing, John took the worms
TC: ก่อน (before) เขา (he) ไม่ไป (went) ตกปลา (fishing), จอห์น (John) ได้เอานา (took) ใส่คือ (worm) หลาย (plurality) คำ (classifier) นั้น (the)
HMI: ก่อน ที่ เขา จะ ไป ตกปลา จอห์น เอา ด้วนน้ ไป ด้วย
HM2: ก่อน เขา จะ ไป ตกปลา จอห์น เอา หนอน ไป ด้วย
Grammatically incorrect.

86. SL: The ugly duckling hides his head under his wing
TC: ตุ๊กตาเป็ด (duckling) ที่ (which) ซึ่งหรู (ugly) คำ (classifier) นั้น (the) ซ่อน (hide) หัว (head) ของเขา (his) ใต้ (under) ปีก (wing) ของเขา (his)
HMI: ตุ๊กตาเป็ด ซึ่งหรู ซุ้ม หัว ของมัน ไว้ ใต้ ปีก
HM2: ตุ๊กตาเป็ด ซึ่งหรู ซ่อน หัว ของมัน ที่ ใต้ ปีก
No repair required.

87. SL: The beautiful Egyptian talked about her dream
TC: คนอียิปต์ (Egyptian) ที่ (which) สวย (beautiful) คำ (classifier) นั้น (the) ได้พูด (talked) เกี่ยวกับ (about) เธ่ (her) มัน (dream)
88. SL: The countryside is very green in spring.

TC: ชนบท (countryside) แห่ง (classifier) นั้น (the) เขียวชูมิ (green) มากๆ (very) ใน (in) ฤดูใบไม้ผลิ (spring)

HMI: ที่ชนบทนี้ มีเขียวชูมิ ในฤดูใบไม้ผลิ

HM2: ชนบทนั้น เป็นสีเขียวในฤดูใบไม้ผลิ

No repair required.

89. SL: A green salad has only green vegetables such as lettuce and cucumber.

TC: салดิ (salad) สีเขียว (green) หนึ่ง (a) มี (has) เท่านั้น (only) ผัก (vegetable) ที่ (which) สีเขียว (green) หลาย (plurality) อย่าง (classifier) คือ

- ผักกาเสด (lettuce) และ (and) ผักกาเสด (lettuce)

HMI: салดิ สีเขียว มี แต่ ผัก สีเขียว เช่น ผักกาเสด และ ผักกาเสด

HM2: салดิ สีเขียว คือ มี แต่ ผัก สีเขียว เช่น ผักกาเสด และ ผักกาเสด

1 misplaced word.

90. SL: The spirit is willing but the flesh is weak.

TC: วิญญาณ (spirit) นั้น (the) พร้อม (willing) แต่ (but) ว่า (flesh) นั้น (the) ยอมODULE (weak)

HMI: จิตวิญญาณนั้น กล้าแกร่ง หากราย กลับ ยอมแต่

HM2: จิตวิญญาณนั้น พร้อม หากร่างกาย กลับ ยอมแต่

1 incorrectly translated word.
Appendix H

Comprehensibility Evaluation

This appendix contains the evaluation of ALMT with respect to comprehensibility as outlined in Chapter 8.

Source Text:

Sentence 77-90: Newspapers and Articles.

Evaluators:

HM3: Thai Linguist
HM4: Thai Student (Grade 7, Age 12)
HM5: Thai Student (Grade 4, Age 9)

Legends:

▲ Additional word required: The word, when installed, will make the sentence complete and correct in terms of meaning and grammar.
☐ Inappropriate word usage: This word does not distort the meaning of the expression but it is inappropriate.
     Meaningless word: This word renders the sentence meaningless.
✓ Unnecessary word: This word is not needed nor used in spoken language.
⇌ Misplaced word: The word is wrongly placed and it affects the meaning of the sentence. It should be moved in the direction of the arrow.
Misplaced word: The word is wrongly placed and it affects the meaning of the sentence.

It should be moved in the direction of the arrow.
1. SL: It was a warm sunny day.
   TC: มัน (it) เป็น (was) วัน (day) ที่ (which) สดใส (sunny) อบอุ่น (warm) วัน (classifier) หนึ่ง (a)
   HM3: comprehensible
   HM4: comprehensible
   HM5: comprehensible
   comprehensible

2. SL: A mother duck sat on her nest at the edge of a pond waiting for her eggs to hatch.
   TC: มAMA (mother) เป็น (duck) ตัว (classifier) หนึ่ง (a) ใกล้ (sat) บน (on) รั้ง (nest) ของเธอ (her) ที่ (at) ขอบ (edge) ของ (of) ทะเลสาบ (pond) ของ (classifier) หนึ่ง (a) รอ (waiting) สำหรับ (for) ไข่ (egg) หลาย (plurality) ของ (classifier) ของเธอ (her) ที่จะ (to) แตกตัว (hatch)
   HM3: incomprehensible
   HM4: incomprehensible
   HM5: incomprehensible
   incomprehensible

3. SL: one_by_one, the eggs began to crack_open.
   TC: ตัวหนึ่ง (one_by_one), ไข่ (egg) หลาย (plurality) ของ (classifier) หนึ่ง (the) ได้เริ่ม (began) ที่จะ (to) แตกแยกออก (crack_open)
   HM3: comprehensible, inappropriate word order, not colloquial Thai.
   HM4: comprehensible, inappropriate word order, not colloquial Thai.
   HM5: comprehensible, inappropriate word order, not colloquial Thai.
   comprehensible, word order is inappropriate, not colloquial Thai.

4. SL: Peep, Peep, said each little yellow duckling as it poked out its head and looked around.
   TC: ปี๊ป (peep), ปี๊ป (peep), แมวเป็ด (duckling) สีเหลือง (yellow) ตัวเล็ก (little) และ (each) ตัว (classifier) ที่ (as) มัน (it) ได้ทุก (poked) กลับมามา (out) หัว (head) ของมัน (its) และ (and) ได้มอง (looked) รอบๆ (around)
   HM3: incomprehensible, 1 additional word required, 1 meaningless word, 1 word misplaced.
   HM4: incomprehensible, 1 additional word required, 1 inappropriate word usage, 1 meaningless word, 1 word misplaced.
   HM5: comprehensible
   incomprehensible

5. SL: Quack, quack, said the mother duck feeling very pleased with herself.
   TC: แจ็ค (quack), แจ็ค (quack), มAMA (mother) เป็น (duck) ตัว (classifier) หนึ่ง (the) ที่ (as) ความรู้สึก (feeling) พอใจ (pleased) มากๆ (very)
   ที่ (with) ตัวเธอเอง (herself)
   HM3: incomprehensible, 1 meaningless word.
6. SL: The biggest egg hatched last of all
   TC: ไอ (egg) ที่ (which) ใหญ่ที่สุด (biggest) ที่เป็น (hatched) สุดท้าย (last) ของ (of) ทั้งหมด (all)
   HM3: incomprehensible, 1 additional word required.
   HM4: incomprehensible, 2 additional words required.
   HM5: incomprehensible, 2 additional words required.

7. SL: But the duckling that poked out his head was not little and yellow like the others
   TC: แต่ (but) ลูกเป็ด (duckling) ตัว (classifier) นั้น (the) ที่ (that) โอน (poked) ข้างนอก (out) หัว (head) ของเขา (his) ไม่ (not) เล็ก (little)
   และ (and) สีเหลือง (yellow) เหมือน (like) ทั้งหลาย (the others)
   HM3: incomprehensible, 2 meaningless words, 1 word misplaced.
   HM4: incomprehensible, 2 meaningless words, 1 word misplaced, 2 additional words required.
   HM5: comprehensible

8. SL: He was big and gray and ugly
   TC: เขา (he) ใหญ่ (big) แย่ (and) สีเทา (gray) และ (and) ซี่เทา (ugly)
   HM3: incomprehensible, 1 additional word required.
   HM4: incomprehensible, 1 additional word required, 1 unnecessary word.
   HM5: incomprehensible, 1 additional word required.

9. SL: Quack, said the mother duck
   TC: นี่หวี (quack), แม่ (mother) เป็ด (duck) ตัว (classifier) นั้น (the) หญิง (said)
   HM3: comprehensible
   HM4: comprehensible
   HM5: comprehensible

comprehensible
10. SL: You are the ugliest duckling I have ever seen.
   TC: คุณ (you) เป็น (are) ลูกเป็ด (duckling) ที่ (which) คุณที่ตู่สุด (ugliest) มัน (I) เห็น (have ever) เห็น (seen)
   HM3: incomprehensible, 1 additional word required.
   HM4: incomprehensible, 1 additional word required.
   HM5: incomprehensible, 1 additional word required.
   incomprehensible

11. SL: The mother duck led her ducklings into the water.
    TC: แม่ (mother) เป็น (duck) ตัว (classifier) นั้น (the) ได้นำ (led) ลูกเป็ด (duckling) หลาย (plurality) ตัว (classifier) ของเธอ (her)
         เข้าไปใน (into) น้ำ (water) ที่ (the)
   HM3: incomprehensible, 1 unnecessary word.
   HM4: incomprehensible, 1 unnecessary word.
   HM5: incomprehensible, 1 unnecessary word.
   incomprehensible

12. SL: The other ducks gathered around.
    TC: เต่า (duck) ตัวอื่น (the other) หลาย (plurality) ตัว (classifier) ได้เจอกัน (gathered) รอบ ๆ (around)
   HM3: comprehensible
   HM4: comprehensible
   HM5: comprehensible
   comprehensible

13. SL: what lovely children you have, they told her.
    TC: คุณ ๆ (children) ที่ (which) น่ารัก (lovely) อะไร (what) คุณ (you) มี (have),พวกเขานะ (they) ได้บอก (told) เธอ (her)
   HM3: incomprehensible, 1 inappropriate word usage, 1 additional word required, 2 words misordered.
   HM4: incomprehensible, 1 additional word required, 2 words misordered.
   HM5: comprehensible, 1 additional word required, 2 words misordered.
   incomprehensible

14. SL: But when they saw the big gray duckling, they began to laugh.
    TC: แต่ (but) เมื่อ (when) พวกเขานะ (they) ได้เห็น (saw) ลูกเป็ด (the) สีเทา (gray) โต (big) ตัว (classifier) นั้น (the),พวกเขานะ (they) ได้เริ่ม (began) ที่จะ (to)
         หัวเราะ (laugh)
   HM3: comprehensible.
15. SL: That one is not lovely, they said.
TC: ว่า (that) หนึ่ง (one) ไม่ (not) น่ารัก (not), เพราะ (they) คุณ (said)
HM3: incomprehensible, 2 meaningless words.
HM4: incomprehensible, 2 meaningless words.
HM5: incomprehensible, 2 meaningless words.

16. SL: what an ugly duckling he is!
TC: ตัวเป็ด (duckling) ที่ (which) ตัวผุ (ugly) ตัวเป็ด (duckling) 😲 (an) ตัวผุ (what) เข้า (he) เป็น (is) เข้ามั้ย (!)
HM3: incomprehensible, 2 unnecessary words, 2 words misordered.
HM4: incomprehensible, 2 unnecessary words, 2 words misordered.
HM5: incomprehensible, 2 unnecessary words, 2 words misordered.

17. SL: The mother duck took her ducklings to the farmyard.
TC: แม่ (mother) เป็ด (duck) ตัว (classifier) นั้น (the) ไปเก็บ (took) ตัวเป็ด (duckling) หลาย (plurality) ตัว (classifier) ของเธอ (her) ไปที่ (to)
        ลานฟาร์ม (farmyard) แห่ง (classifier) นั้น (the)
HM3: comprehensible.
HM4: comprehensible, 2 unnecessary words, 1 inappropriate word usage.
HM5: comprehensible.

18. SL: When the other animals saw the big gray duckling, they all began to shout.
TC: เมื่อ (when) ตัวอื่น (animal) หลาย (plurality) ตัวเป็ด (duckling) ที่ (which) สีเทา (gray) โต (big) ตัว (classifier) นั้น (the), เพราะ (they) ทั้งหมด (all) ได้เห็น (saw) ตัวเป็ด (duckling) ที่ (which) สีเทา (gray) โต (big) ตัว (classifier) นั้น (the) ตะโกน (shout)
HM3: incomprehensible, 2 unnecessary words.
HM4: incomprehensible, 2 unnecessary words.
HM5: comprehensible.
19. SL: What an ugly duckling he is! they said.

TC:  ดู (duckling) คี่ (duckling) ชีวะ (ugly) คอนโด (classifier) อัง (an) อะไร (what) เข้า (he) เป็น (is) เชนนั้น (I) พวกเขาก (they) ได้เห็น (said)

HM3: incomprehensible, 2 unnecessary words, 2 words misordered.
HM4: incomprehensible, 2 unnecessary words, 2 words misordered.
HM5: incomprehensible, 2 unnecessary words, 2 words misordered.

incomprehensible

20. SL: He does not belong with us.

TC:  เขา (he) ไม่ (not) ใช้ส่วนหนึ่งของ (belong_with) พ่อของ (us)

HM3: comprehensible.
HM4: comprehensible.
HM5: comprehensible.

comprehensible

21. SL: The geese and the turkeys ran after the duckling and tried to chase him away.

TC:  พวกหู (geese) ที่ (the) และ (and) ได้ (turkey) หลาย (plurality) ตัว (classifier) นั้น (the) ไม่ (not) ได้ (tried) ดู (duckling) ตัว (classifier) ตัว (the) และ (and) ให้พยายาม (tried) ที่จะ (to) ไป (chase) เขา (him) ทางไป (away)

HM3: comprehensible, 1 unnecessary word, 1 additional word required.
HM4: incomprehensible, 1 unnecessary word, 1 additional word required.
HM5: comprehensible.

comprehensible

22. SL: Even his own brothers and sisters joined in and called him names.

TC: แม่คน (even) พี่ชายและน้องชาย (brother) หลาย (plurality) คน (classifier) ของเขา (his) ของ (own) และ (and) พี่สาวของสาว (sister) หลาย (plurality) คน (classifier) ที่มา (joined_in) และ (and) พี่สาวของสาวๆ (called_him_names)

HM3: comprehensible, 1 additional word required.
HM4: comprehensible, 1 additional word required.
HM5: comprehensible.

comprehensible
23. SL: The ugly duckling hid in a corner of the farmyard and cried.
   TC: ซุกเปิด (duckling) ที่ (which) ซึ่งหรู (ugly) ด้วย (classifier) นั้น (the) ให้ซ่อน (hid) ใน (in) มุม (corner) มุม (classifier) หนึ่ง (of)
   ลาติดในหน้า (farmyard) แห่ง (classifier) นั้น (the) และ (and) ใครจะให้ (cried)
   HM3: comprehensible.
   HM4: comprehensible, 1 inappropriate word usage.
   HM5: comprehensible.

24.1. SL: The next morning,
   TC: เซา (morning) วัน (classifier) ที่ถัดไป (next) นั้น (the)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.

24.2. SL: The ugly duckling decided to run_away
   TC: ซุกเปิด (duckling) ที่ (which) ซึ่งหรู (ugly) ต้อง (classifier) นั้น (the) ได้ตัดสินใจ (decided) ที่จะ (to) หนีจากไป (run_away)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.

25. SL: I will find a place where no_one will make_fun_of me, he said to himself.
   TC: ฉัน (I) จะ (will) หาที่ไหน (find) ที่ (place) ที่ (classifier) หนึ่ง (a) ที่ซึ่ง (where) ไม่มีใคร (no_one) จะ (will) เข้ามาเยี่ยม (make_fun_of) ฉัน (me), เขา (he) ใจผูก (said) กัน (to) ตัวเขาเอง (himself)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.

26. SL: He left the farmyard and ran to the great marsh where the wild ducks lived.
   TC: เขา (he) ได้จาก (left) ลาติดในหน้า (farmyard) แห่ง (classifier) นั้น (the) และ (and) ใครง (ran) ไปที่ (to) หุบแห่งน้ำ (marsh) ใหญ่หนา (great) หุบแห่ง (classifier) นั้น (the) ที่ไหน (where) เป็น (duck) น้ำ (wild) หลาย (plurality) ด้วย (classifier) นั้น (the) ได้อาศยอยู่ (lived)
   HM3: incomprehensible, 1 meaningless word.
   HM4: incomprehensible, 2 meaningless words.
27. SL: Perhaps they will be kind to me, he said sadly.
   TC: บาง (perhaps) พวกเขา (they) จะ (will) ใจดี (kind) กับ (to) ฉัน (me), เขา (he) ได้พูด (said) อย่างเศร้าสลด (sadly)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

28. SL: But the wild ducks were not kind.
   TC: แต่ (but) เป็ด (duck) พวก (wild) พวก (plurality) นั้น (the) ไม่ (not) ใจดี (kind)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

29. SL: You are very ugly, they said.
   TC: คุณ (you) หรือ (ugly) พวก (very), พวกเขา (they) คุณ (said)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

30. SL: You are certainly not one of us.
   TC:คุณ (you) เป็น (are) อย่างแน่นอน (certainly) ไม่ (not) หนึ่ง (one) ของ (of)พวกเรานะ (us)
   HM3: incomprehensible, 3 words misordered.
   HM4: incomprehensible, 3 words misordered.
   HM5: incomprehensible, 3 words misordered.
   incomprehensible

31. SL: And the wild ducks laughed at the duckling.
   TC: และ (and) เป็ด (duck) พวก (wild) พวก (plurality) นั้น (the) ได้หัวเราะหยาบ (laughed_at) ลูกเป็ด (duckling) นั้น (the)
32. SL: So the ugly duckling left the marsh.
   TC: ตัวบกนี้ (so) ตุ๊กแก่ (duckling) ที่ (which) ซี่เครื่อง (ugly) ตัว (classifier) นั้น (the) ได้จาก (left) ทะเลสาบ (marsh) เหมือง (classifier) นั้น (the)
  HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

33. SL: The wind was blowing and it was cold.
   TC: ลม (wind) ผู้ (the) กำลัง (continuous tense) เดิน (blow) และ (and) มัน (it) หนาว (cold)
   HM3: incomprehensible, 1 unnecessary word, 1 meaningless clause.
   HM4: incomprehensible, 1 unnecessary word, 1 meaningless clause.
   HM5: incomprehensible, 1 meaningless clause.
   incomprehensible

34. SL: Winter was coming
   TC: ฤดูหนาว (winter) กำลัง (continuous tense) มา (come)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

35. SL: The ugly duckling was very tired.
   TC: ตุ๊กแก่ (duckling) ที่ (which) ซี่เครื่อง (ugly) ตัว (classifier) นั้น (the) เหมือง (tired) มาๆ (very)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible
36. SL: He had to find a place to stay where he could be warm and safe.
TC: เขา (he) ต้อง (had_to) หาที่พัก (find) ที่ (place) ที่ (classifier) หนึ่ง (a) ที่จะ (to) อยู่ (stay) ที่ซึ่ง (where) เขา (he) จะได้ (could_be) อยู่ (warm) และ (and) ปลอดภัย (safe)
HM3: comprehensible.
HM4: comprehensible.
HM5: comprehensible.
comprehensible

37. SL: One evening, he came to a little cottage.
TC: หนึ่ง (one) เดิน (evening), เขา (he) ไปมา (came) กับ (to) กระท่อม (cottage) เล็ก (little) หลัง (classifier) หนึ่ง (a)
HM3: incomprehensible, 1 word misplaced, 1 additional word required, 1 meaningless word.
HM4: incomprehensible, 1 word misplaced, 1 additional word required, 1 meaningless word.
HM5: incomprehensible.

38. SL: The door was opened and he went inside.
TC: ประตู (door) เปิด (classifier) คน (the) ผู้ (passive voice) เปิด (open) และ (and) เขา (he) โใบ (went) ข้างใน (inside)
HM3: comprehensible, not colloquial Thai.
HM4: comprehensible, not colloquial Thai.
HM5: comprehensible, not colloquial Thai.
comprehensible, not colloquial Thai.

39. SL: An old woman lived in the cottage with a fat black cat and a plump brown hen.
TC: หญิง (woman) ถ้า (old) คน (classifier) หนึ่ง (a) อาศัยอยู่ (lived) ใน (in) กระท่อม (cottage) หลัง (classifier) หนึ่ง (the) ด้วย (with) แมว (cat) สีดำ (black) อวบน้ำ (fat) ตัว (classifier) หนึ่ง (a) และ (and) ไก่ (hen) สีน้ำตาล (brown) อวบน้ำ (plump) ตัว (classifier) หนึ่ง (a)
HM3: incomprehensible, 1 inappropriate word usage.
HM4: incomprehensible, 1 inappropriate word usage.
HM5: comprehensible, 1 inappropriate word usage.

40. SL: The old woman did not mind the ugly duckling but the cat and the hen wanted him to leave.
TC: หญิง (woman) ถ้า (old) คน (classifier) หนึ่ง (the) ไม่ (not) รังเกียจ (mind) ตัวเป็ด (duckling) ที่ (which) ซึ่ง (ugly) ตัว (classifier) หนึ่ง (the) แต่ (but) แมว (cat) ตัว (classifier) หนึ่ง (the) และ (and) ไก่ (hen) ตัว (classifier) หนึ่ง (the) โค้จำเจียร (wanted) เขา (him) กับ (to) จาก (leave)
HM3: incomprehensible, 1 additional word required, 1 meaningless word.
41. SL: Can you purr? asked the cat.
   TC: คุณ (you) สามารถ (can) ทำเสียงที่คิดว่า (purr) ได้ไหม (?) แมว (cat) ตัว (classifier) นั่น (the) ได้ถาม (asked)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

42. SL: No, said the ugly duckling.
   TC: ไม่ (no), นกเป็ด (duckling) ที่ (which) ซึ่งตำแหน่ง (ugly) ตัว (classifier) นั่น (the) ได้พูด (said)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

43. SL: Can you lay eggs? asked the hen.
   TC: คุณ (you) สามารถ (can) วาง (lay) ไข่ (egg) หลาย (plurality) ตัว (classifier) ได้ไหม (?) นก (hen) ตัว (classifier) นั่น (the) ได้ถาม (asked)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

44. SL: No, said the ugly duckling.
   TC: ไม่ (no), นกเป็ด (duckling) ที่ (which) ซึ่งตำแหน่ง (ugly) ตัว (classifier) นั่น (the) ได้พูด (said)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible

45. SL: Then you must go, said the cat and the hen.
   TC: ด้วย (then) คุณ (cat) ต้อง (must) ไป (go), แมว (cat) ตัว (classifier) นั่น (the) และ (and) นก (hen) ตัว (classifier) นั่น (the) ได้พูด (said)
46. SL: They chased the ugly duckling out of the cottage and would not let him in again.

TC: (they) ไล่ (chased) นกแก้ว (duckling) ที่ (which) ตัว (ugly) ตัว (classifier) นอก (the) ของ (out_of) (cottage) หลัง (classifier) และ (and) จะ (will) ไม่ (not) ปล่อย (let) เขา (him) เข้ามา (in) (again)

HM3: comprehensible, inappropriate word usage.
HM4: comprehensible, inappropriate word usage.
HM5: comprehensible

47. SL: The ugly duckling wandered through the countryside for many days.

TC: นกแก้ว (duckling) ที่ (which) ตัว (ugly) ตัว (classifier) แล้ว (the) ใช้ (wandered) (through) ที่ (countryside) หลาย (many) วัน (day)

HM3: comprehensible.
HM4: comprehensible, 2 inappropriate word usage.
HM5: comprehensible

48. SL: No one would give him a home.

TC: ไม่มีใคร (no_one) จะ (would) ให้ (give) เข้า (him) บ้าน (home) หลัง (classifier) หนึ่ง (a)

HM3: incomprehensible, 1 meaningless word.
HM4: incomprehensible, 1 meaningless word.
HM5: incomprehensible, 1 additional word required.

49. SL: No one would be his friend.

TC: ไม่มีใคร (no_one) จะ (would) เป็น (be) เพื่อน (friend) ของเข้า (his)

HM3: comprehensible.
HM4: comprehensible.
HM5: comprehensible

comprehensible
50. SL: All the birds and animals said that he was ugly.
    TC: นก (bird) ทั้งหมด (all) นั้น (the) และ (and) สัตว์ (animal) หลาย (plurality) ตัว (classifier) ไร่รูป (said) ว่า (that) เขา (he) ซึ่งว่า (ugly)
    HM3: comprehensible, not colloquial Thai.
    HM4: comprehensible, not colloquial Thai.
    HM5: comprehensible, not colloquial Thai.
    comprehensible, not colloquial Thai.

51. SL: he grew more and more sad and more and more tired.
    TC: เขา (he) โปรดชื่น (grew) ถึง (more) และ (and) เศร้า (sad) มากขึ้น (more) และ (and) มาขึ้น (more) และ (and) เที่ยว (tired) มาขึ้น (more)
    HM3: comprehensible, but does not convey the intended meaning, not colloquial Thai.
    HM4: comprehensible, but does not convey the intended meaning, not colloquial Thai.
    HM5: comprehensible, but does not convey the intended meaning.
    comprehensible

52. SL: at last he stopped to rest by a lake that was covered with ice.
    TC: ใดที่สุด (at last) เขา (he) หยุด (stayed) ที่นั่น (there) พัก (rest) โดย (by) ทะเลสาบ (lake) แห้ง (classifier) นั้น (that) ถูก (passive voice) ปกคลุม (cover) ด้วย (with) น้ำแขื่น (ice)
    HM3: incomprehensible, 2 meaningless words.
    HM4: incomprehensible, 2 meaningless words.
    HM5: incomprehensible, 2 additional words required.
    comprehensible

53. SL: He stayed there all winter, hidden in the reeds.
    TC: เขา (he) อยู่ (stayed) ที่นั่น (there) ตลอด (all) ฤดูหนาว (winter), ซ่อน (hidden) ใน (in) ต้น (reed) หลาย (plurality) ต้น (classifier) นั้น (the)
    HM3: comprehensible, not colloquial Thai.
    HM4: comprehensible, not colloquial Thai.
    HM5: comprehensible, not colloquial Thai.
    comprehensible, not colloquial Thai.

54. SL: He was cold and hungry and very lonely.
    TC: เขา (he) หนาว (cold) และ (and) หิว (hungry) และ (and) เหล่า (lonely) มาก (very)
    HM3: comprehensible.
    HM4: comprehensible.
    HM5: comprehensible.
55. SL: One bright morning in early spring, when all the ice had melted, a flock of beautiful birds came to the lake.

TC: หนึ่ง (one) ที่ (which) สว่าง (bright) เวลา (morning) ใน (in) ตอนต้น (early) ของ (part of) ปี (spring), เมื่อ (when) น้ำแข็ง (ice) หลั่นหลุด (all) นั้น (the) ได้ (had) ละลาย (melt) เย็น (bird) ที่ (which) สวยงาม (beautiful) นั้น (of) นั้น (flock) นั้น (classifier) หนึ่ง (a) นั้น (plurality) ได้มา (came) ที่ (to) ทะเลสาบ (lake) แห้ง (classifier) นั้น (the)

HM3: incomprehensible, 2 word misplaced, 1 additional word required, 4 unnecessary words.
HM4: incomprehensible, 2 word misplaced, 1 additional word required, 4 unnecessary words.
HM5: incomprehensible

56. SL: They had gleaming white feathers and long graceful necks.

TC: พวกเขา (they) มี (had)ขน (feathers)ที่ (which) ประกาย (gleaming)สีขาว (white) เย็น (plurality) เย็น (classifier) และ (and) เย็น (neck)ที่ (which)ยาว (long) สวยงาม (graceful) ยาว (plurality) เย็น (classifier)

HM3: incomprehensible, 4 unnecessary words.
HM4: comprehensible, not colloquial Thai.
HM5: comprehensible

57. SL: The ugly duckling watched them from his hiding place in the reeds.

TC: ลูกเป็ด (ducksling)ที่ (which)สิ่งผืน (ugly) เย็น (classifier)นั้น (the)ได้ (watched) การเขา (them) จาก (from) สถานที่ (place) ซ่อน (hiding)ของเขา (his) ใน (in) คลอง (reed) ยาว (plurality) เย็น (classifier)นั้น (the)

HM3: incomprehensible, 1 meaningless word.
HM4: incomprehensible, 1 meaningless word.
HM5: comprehensible

58. SL: They are the most beautiful birds I have ever seen, he thought.

TC:พวกเขา (they) เป็น (are) เย็น (bird)ที่ (which) สวยงาม (beautiful) ที่สุด (most) ยาว (plurality) เย็น (classifier)นั้น (I) เคย (have ever)เห็น (seen), เขา (he)ไม่คิด (thought)

HM3: incomprehensible, 1 additional word required.
HM4: incomprehensible, 1 additional word required.
HM5: incomprehensible, 1 additional word required.

incomprehensible, 1 additional word required.
59. SL: If only I could be like them.
   TC: ต้า (if) เพียงใด (only) นั้น (I) สามารถ (could) ชม (like) พวกเข้า (them)
   HM3: comprehensible, but not in the meaning which it is intented.
   HM4: incomprehensible, 1 meaningless word.
   HM5: comprehensible.
   comprehensible.

60. SL: The beautiful birds saw the ugly duckling.
   TC: นก (bird) ที่ (which) สวย (beautiful) หลาย (plurality) ตัว (classifier) นั้น (the) ได้เห็น (saw) ลูกเป็ด (duckling) ที่ (which) ซึ่งเห็น (ugly) ตัว (classifier) นั้น (the)
   HM3: comprehensible but it is not the way it is spoken in Thai.
   HM4: comprehensible but it is not the way it is spoken in Thai.
   HM5: comprehensible.
   comprehensible, it is not the way it is spoken in Thai.

61. SL: The biggest one came_over_to him
   TC: ตัว (one) ที่ (which) ใหญ่สุด (biggest) ตรงมาที่ (came_over_to) เข้า (him)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible.

62. SL: The ugly duckling tried to hide
   TC: ลูกเป็ด (duckling) ที่ (which) ซึ่งเห็น (ugly) ตัว (classifier) นั้น (the) พยายาม (tried) ที่จะ (to) ซ่อน (hide)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible.

63. SL: He was sure that the bird would make_fun_of him or try to chase him away
   TC: เขา (he) แน่ใจ (sure) ว่า (that) นก (bird) ตัว (classifier) นั้น (the) จะ (would) ทำแย่ง (make_fun_of) เข้า (him) หรือ (or) พยายาม (try) ที่จะ (to) ไล (chase) เข้า (him) ทางหนี (away)
   HM3: incomprehensible, 1 meaningless word, 1 additional word required.
HM4: incomprehensible, 1 meaningless word, 1 additional word required.
HM5: incomprehensible, 1 meaningless word, 1 additional word required.

64. SL: But the lovely white bird said, what a fine young swan you are!
TC: แต่ (but) นก (bird) นี่ (which) สีขาว (white) น่ารัก (lovely) ตัว (classifier) นั้น (the) โค่น (said), งวง (swan) หนุ่ม (young) จงงาม (fine) ตัว (classifier) นี้ (a) อะไร (what) คุณ (you) เป็น (are) หนักนัก (!)
HM3: incomprehensible, 3 words misplaced.
HM4: incomprehensible, 3 words misplaced.
HM5: incomprehensible, 3 words misplaced.

65. SL: You should not be here all_by_yourself.
TC: คุณ (you) ไม่ (not) ควร (should) อยู่ (be) ที่นี่ (here) ได้ (all_by_yourself)
HM3: comprehensible.
HM4: comprehensible.
HM5: comprehensible.

66. SL: The ugly duckling could not believe what he had heard.
TC: นกเป็ด (duckling) นี้ (which) ดูหนึ่ง (ugly) ตัว (classifier) นั้น (the) ไม่ (not) สามารถ (could) เชื่อ (believe) อะไร (what) เขา (he) หัน (had) ไป (heard)
HM3: incomprehensible, 1 meaningless word, 1 additional word required, 1 unnecessary word.
HM4: incomprehensible, 1 meaningless word, 1 additional word required, 1 unnecessary word.
HM5: incomprehensible, 1 additional word required.

67. SL: He looked at his reflection in the water.
TC: เขา (he) ได้ดู (looked) ไปที่ (at) รูปสะท้อน (reflection) ของเขา (his) ใน (in) น้ำ (water) นั้น (the)
HM3: comprehensible, 1 unnecessary word.
HM4: comprehensible, 1 unnecessary word.
HM5: comprehensible.

comprehensible.
68. SL: It was true
   TC: มัน (it) จริง (true)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible.

69. SL: His feathers were white and gleaming.
   TC: ขน (feathers) หลาย (plurality) เส้น (classifier) ของเขา (his) สีขาว (white) และ (and) เบ็ญราบาย (gleaming)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible.

70. SL: His neck was long and graceful
   TC: คอ (neck) ของเขา (his) ยาว (long) และ (and) สวยงาม (graceful)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible.

71. SL: He had grown into a handsome swan
   TC: เขา (he) โต (had) เป็น (into) หงส์ (swan) ดูดงาม (handsome) ตัว (classifier) นั้น (a)
   HM3: comprehensible, 1 inappropriate word usage.
   HM4: comprehensible, 1 inappropriate word usage.
   HM5: comprehensible.
   comprehensible.

72. SL: Come with us, said the swan.
   TC: มา (come) กับ (with) พวกเรา (us), หงส์ (swan) ตัว (classifier) นี้ (the) ได้พูด (said)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.
   comprehensible.
73. SL: The ugly duckling spread his wings and rose into the sky.
   TC: ดูกลังค์ (duckling) ที่ (which) ซึ่งเขา (ugly) ตัว (classifier) นั้น (the) ทรง (spread) ใต (wing) plur (plurality) ประเภท (classifier) ของเขา (his)
   และ (and) นั้น (rose_into) ของ (sky) นั้น (the)
   HM3: comprehensible, 2 unnecessary words.
   HM4: comprehensible, 1 inappropriate word.
   HM5: comprehensible.

74. SL: I never dreamed that I could be so happy, he said, and he flew off proudly with his new friends.
   TC: ฉัน (I) ไม่เคย (never) ได้ฝัน (dreamed) ว่า (that) ฉัน (I) จะได้ (could_be) ดังนั้น (so) เป็นสุข (happy), เขา (he) ได้พูด (said), และ (and) เขา (he)
   ได้บิน (flew) ออกไป (off) อย่างภาคภูมิ (proudly) ด้วย (with) เพื่อน (friend) ใหม่ (new) หลาย (plurality) ตัว (classifier) ของเขา (his)
   HM3: incomprehensible, 2 meaningless words.
   HM4: incomprehensible, 2 meaningless words.
   HM5: incomprehensible.

75. SL: The ugly duckling
   TC: ดูกลังค์ (duckling) ที่ (which) ซึ่งเขา (ugly) ตัว (classifier) นั้น (the)
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.

76. SL: Written by Ronne_Randall
   TC: เขียน (written) โดย (by) Ronne_Randall
   HM3: comprehensible.
   HM4: comprehensible.
   HM5: comprehensible.

77. SL: The wheat was yellow
   TC: ข้าวสาร (wheat) นั้น (the) สีเหลือง (yellow)
   HM3: comprehensible.
78. SL: Here the stork marched about on his long red legs talking Egyptian.

TC: ที่นี่ (here) นกกระสา (stork) ตัว (classifier) นั่น (the) เดินไปมา (marched_about) บน (on) ขา (leg) สีแดง (red) ยาว (long) หลาย (plurality)

( classifier) ของเข่า (his) ทุก (talking) คนอียิปต์ (Egyptian)

HM3: incomprehensible, 2 meaningless words.
HM4: incomprehensible, 2 meaningless words, 2 unnecessary words.
HM5: incomprehensible, 2 meaningless words, 2 additional word required.

79. SL: When I was an ugly duckling, he thought, I never dreamed I could be so happy.

TC: เมื่อ (when) ฉัน (I) เป็น (was) ตุ๊กตา (duckling) นั่น (which) นี่here (ugly) ตัว (classifier) มัน (an), เขา (he) คิด (thought), ฉัน (I) ไม่เคย (never) ได้สัมผัส (dreamed) ฉัน (I) จะสด (could_be) ด้วย (so) เป็นสุข (happy).

HM3: incomprehensible, 1 meaningless word.
HM4: incomprehensible, 1 meaningless word, it is not a colloquial Thai.
HM5: incomprehensible, 1 meaningless word.

80. SL: You can take a fish to school but you cannot make them think.

TC: คุณ (you) สามารถ (can) เอา (take) ปลา (fish) ตัว (classifier) มัน (an) ไปที่ (to) โรงเรียน (school) แต่ (but) คุณ (you) ไม่ (not) สามารถ (can)

ทำ (make) ในหน้า (them) คิด (think)

HM3: incomprehensible, 1 meaningless word, 1 additional word required.
HM4: incomprehensible, 1 meaningless word, 1 additional word required.
HM5: incomprehensible, 1 meaningless word, 1 additional word required.

81. SL: Five days before the trout are released.

TC: ห้า (five) วัน (day) ก่อน (before) ปลาทรูท (trout) ตัว (classifier) นั่น (the) ถูก (passive voice) ปล่อย (release)

HM3: comprehensible, 1 additional word required.
HM4: comprehensible, 1 additional word required.
HM5: comprehensible.

comprehensible.
82. SL: I rent a room from Mrs Jones.

TC: นัน (I) ต่อ ซ่า (rent) ห้อง (room) หญิง (classifier) นั้น (a) จาก (from) มาร์ (Mrs) โจนส์ (Jones)

HM3: incomprehensible, 1 meaningless word, 1 additional word is required.
HM4: incomprehensible, 1 meaningless word, 1 additional word is required.
HM5: incomprehensible, 1 meaningless word, 1 additional word is required.

83. SL: She was rending her hair out in anger

TC: เธอ (she) กำลัง (continuous tense) ริน (rend) ผม (hair) ออก (out) ไว้ (in) โทร์ (anger)

HM3: incomprehensible, 2 meaningless words.
HM4: incomprehensible, 2 meaningless words, 1 inappropriate word usage.
HM5: incomprehensible, 2 meaningless words.

84. SL: Algebraic symbols are used when you do not know what you are talking about.

TC: ตัวแปรจำานวน (symbols) ทางคณิตศาสตร์ (algebraic) หลาย (plurality) ตัว (classifier) ใช้ (use) เมื่อ (when) คุณ (you) ไม่ (not) รู้ (know) เป็น (what) คุณ (you) กำลัง (continuous tense) คุณ (talk) ถึง (about)

HM3: incomprehensible, 1 word misplaced.
HM4: incomprehensible, 1 word misplaced.
HM5: comprehensible.

85. SL: Before he went fishing, John took the worms

TC: นั้น (before) เขา (he) ไป (went) ตกปลาก (fishing), จอห์น (John) ได้อาจ (took) ปลัง (worm) หลาย (plurality) ตัว (classifier) นั้น (the)

HM3: comprehensible, inappropriate word order, not colloquial Thai.
HM4: comprehensible, inappropriate word order, not colloquial Thai.
HM5: comprehensible, inappropriate word order, not colloquial Thai.

86. SL: The ugly duckling hides his head under his wing

TC: บ๊อกซิ่ง (duckling) ซึ่ง (which) ยิ่ง (ugly) ตัว (classifier) นั้น (the) ซ่อน (hide) หัว (head) ของเขามา (his) ใต้ (under) ปีก (wing) ของเขา (his)

HM3: comprehensible, 1 additional word required.
HM4: comprehensible, 1 additional word required.
87. SL: The beautiful Egyptian talked about her dream
TC: นิยิป (Egyptian) ที่ (which) ดาย (beautiful) คณ (classifier) นั้น (the) ได้พูด (talked) เกี่ยวกับ (about) เธอ (her) นั้น (dream)
HM3: incomprehensible, 2 meaningless words.
HM4: incomprehensible, 2 meaningless words.
HM5: incomprehensible, 2 meaningless words.

88. SL: The countryside is very green in spring
TC: ชนบท (countryside) แห่ง (classifier) นั้น (the) เขียวขุ่น (green) มากๆ (very) ใน (in) ฤดูใบไม้ผลิ (spring)
HM3: comprehensible, 1 additional word required.
HM4: comprehensible, 1 additional word required.
HM5: comprehensible.

89. SL: A green salad has only green vegetables such as lettuce and cucumber
TC: ผลัก (salad) สีเขียว (green) หนึ่ง (a) มี (has) เท่าที่ (only) ตัน (vegetable) ที่ (which) สีเขียว (green) หลาย (plurality) อย่าง (classifier)
ตัวอย่างเช่น (such as) ผักสดแก้ว (lettuce) และ (and) แตงวา (lettuce)
HM3: incomprehensible.
HM4: incomprehensible, 1 meaningless word.
HM5: incomprehensible.

90. SL: The spirit is willing but the flesh is weak.
TC: วิญญาณ (spirit) นั้น (the) พร้อม (willing) แต่ (but) ร่างกาย (flesh) นั้น (the) อ่อนแอ (weak)
HM3: incomprehensible.
HM4: incomprehensible.
HM5: comprehensible.

HM5: comprehensible.
comprehensible.