Bridging the gap: Thai - Thai Sign Machine Translation

Srisavakon Dangsaart
Institute for Innovation and Development of Learning Process,
Mahidol University, Bangkok, Thailand
sri_dang@hotmail.com

Kanlaya Naruedomkul
Department of Mathematics
Mahidol University, Bangkok, Thailand
scknr@mahidol.ac.th

Nick Cercone
Department of Science and Engineering
York University, Toronto, Canada
ncercone@yorku.ca

Booncharoen Sirinaovakul
Department of Computer Engineering,
King Mongkut's University of Technology
Thonburi, Bangkok, Thailand
boon@kmutt.ac.th

Abstract

We propose a multi-phase approach, Thai-Thai Sign Machine Translation (TTSMT), to translate Thai text into Thai Sign language. Thai Sign language is the language of the deaf in Thailand.

TTSMT begins the translation process by segmenting the input sentence since Thai is a non-word boundary language, converting the segmented sentence into simple sentence forms since most Thai Sign are expressed in a sequence of such form, and then generating the intermediate sign codes which link a Thai word to its corresponding Thai Sign. The most appropriate sign codes will be selected and rearranged in the spatial grammatical order for generating the Sign language with pictures. The distinction between the Thai text and Thai Sign Language in both grammar and vocabulary are concerned in each processing step to ensure the accuracy of translation.

We have developed the Thai-Thai Sign translation system based on TTSMT. The developed system was implemented and tested to translate Thai sentences used in everyday life. The test results have proved that TTSMT can provide accurate translations.

1 Introduction

Our world is fully loaded with all kinds of information; parts of it appear in text form. To fully utilize such information, the hearing impaired and the deaf should have an ability to effectively read written text. In addition, many accessibility aids for the hearing impaired and the deaf, like television closed captioning or teletype telephones, computer interfaces, Internet information and services, assume that the users have strong language literacy skills.

Most of the hearing impaired and the deaf experience difficulties in mastering their reading skills, and are often faced with comprehension problems when they read written text. Although hearing impaired and deaf students are taught the written language, their reading skills are lower than what they are supposed to be due to their inability to hear the spoken language. As concluded in the research at the Gallaudet Research Institute (1996), "Young deaf and hearing impaired adults aged 17 and 18 had the same average reading score as average ten-year-old hearing children".

Thus for assisting the hearing impaired and the deaf to overcome such problems and to be able to communicate in a world of spoken languages, an interpreter is needed. However, having a human interpreter at their side at all times is not the answer because of the cost and the limited availability of trained personnel. Therefore, it is not surprising that a Sign Language Automated Machine Translation System is in great demand to make more information and services accessible to the hearing impaired and the deaf

Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics, pages 191–199.

on a more economical basis when text material is too complex or an interpreter is unavailable.

A number of Sign language translation systems have been developed around the world, e.g. TESSA system (Bangham and Cox, 2000) is the direct approach and restricted domain. Weather Reports Generate System (Angus and Smith, 1999) is the transfer approach and using the limited domain. ViSiCAST Translator (Safar and Marshall, 2000) is semantically driven used HPSG semantic feature structures and Discourse Representation Structures to represent the internal structure of linguistic object that is rule-based and specific domain. TEAM Project (Zhao and Kipper, 2000) is the transfer approach that uses synchronized tree adjoining grammar. ZARDOZ system (Veale and Collins, 1998) is the interlingual systems. South African Sign Language Machine Translation System (Zijl and Barker, 2003) is transfer approach and rule-based. ASL Workbench (Armond and Speers, 2001), TGT System-Polish Text into Sign Language (Suszczanska and Szmal, 2002) are the transfer approach. Spatial and Planning Models of ASL Classifier Predicates for MT and American Sign Language Generation: Multimodal NLG with Multiple linguistic Channels (Huenerfauth, 2004,2005) are the hybrid models that combined the transfer and interlingual approaches. This work focuses in particular on models for classifier predicates. An example-based approach to translating Sign language (Morrissey and Way, 2005) applied example-based methods for automatic translation base on one of the ECHO corpora which is robust for sentences already seen in training, but has problems with unseen word and phrase combinations. Morpho-Syntax Based Statistical Methods for Automatic Sign Language Translation (Stein and Bungeroth, 2006) is statistical-based.

Most of them are English-to-American Sign Language, although there have been a few other language pairs. In Thailand, Thai Sign Language Recognition Systems and Thai Sign Language dictionaries have been developed. However, no Thai into Thai Sign language automated translation system which is able to generate reliable translation results is yet available.

To bridge the communication gap between the hearing impaired and hearing worlds in Thailand, the Thai-Thai Sign Machine Translation (TTSMT) approach is proposed. TTSMT was designed to translate Thai text into Thai Sign Language.

2 Thai and Thai Sign Language: Linguistic Issues

The Sign language uses gestures instead of sound to convey the meaning. Gestures can be characterized by manual and non-manual parameters. Manual parameters include hand-shape, hand-orientation, position, and motion while non-manual parameters include posture of the upper torso, head-orientation, facial expressions, lip movement and position (Sutton-Spence and Woll, 1999; Stokoe, 1978). Although Sign language is the native language for the hearing impaired and the deaf, it is not a universal language. Regionally, different Sign languages have evolved such as the American Sign Language (ASL) used in the United States, the British Sign Language (BSL) used in the United Kingdom and the Thai Sign Language (TSL) used in Thailand.

TSL is different from the Thai language, featuring its own lexicon and grammar rules. The sentence type in the TSL is normally simple but it corresponds to a complex or compound sentence in the Thai language. Moreover some ordering of a compound sentence in the Thai language is unlike the ordering of simple sentences in the TSL as shown in Table 1. The structure of a Thai sentence contains Subject (S), Verb (V) and Object (O) in sequence which differs from the TSL. The TSL contains O, S and V respectively and there is a Classifier[1] (CL) in some positions as shown in Table 2. A Thai phrase represents the different word sequences from a TSL phrase as shown in Table 3. In addition, the gesture of verb in the TSL depends on subject and/or object whereas in the Thai language, it does not.

Table 1. Compare some sentence ordering.

No.	Thai Sentence Ordering	TSL Sentence Ordering
1	$S = S_1 + \text{and} + S_2$	$S = S_1 + S_2$
2	$S = S_1 + \text{or} + S_2$	$S = S_1 + S_2$
3	$S = S_1 + \text{because} + S_2$	$S = S_2 + S_1$

Table 2. Some different syntax.

No.	Thai Syntax	TSL Syntax
Active voice		
1	S+V	S+V
2	S+Neg+V	S+V+Neg
3	S+V+O	O+CL+S+V
4	S+Neg+V+O	O+CL+S+V+Neg

[1] Classifier is the relationship between the symbols

No.	Thai Syntax	TSL Syntax
5	S+V+diO+indO	indO+CL+diO+CL+S+V
6	S+Neg+V+diO+indO	indO+CL+diO+CL+S+V +Neg
Passive voice		
7	O+S+V	O+CL+S+V
8	O+Neg+S+V	O+CL+S+V+Neg

Table 3. Compare some word sequence in a phrase.

No.	Thai Phrase Ordering	TSL Phrase Ordering
1	VP = Neg+V	VP = V+Neg
2	NP = Neg+(Adj,Adv)	NP = (Adj, Adv)+Neg
3	PP = P+NP	PP = NP+P
4	$NP = NP_1+P+NP_2$	$NP = NP_2+CL+NP_1+P$
5	$NP = NP_1+P_1+ NP_2+P_2+NP_3$	$NP = NP_3+CL+NP_2+ P_2+CL+NP_1+P_1$

3 TTSMT Architecture

The architecture of Thai-Thai Sign Machine Translation (TTSMT), an approach to translate Thai Text into Thai Sign, is illustrated in Figure 1. TTSMT comprises five modules: Sentence Treatment, Word Treatment, Sign-Code Selection, Sign-Code Ordering and Image Mapping.

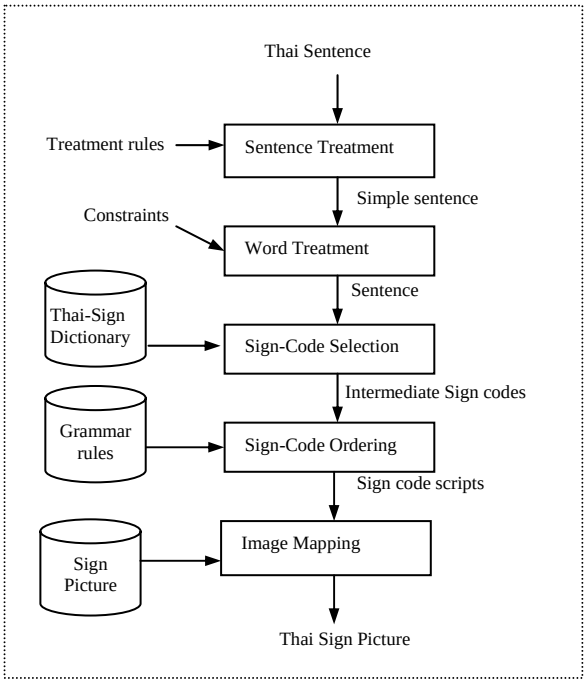

Figure 1. TTSMT architecture.

3.1 Sentence Treatment

The first module consists of two steps*: Word Segmentation* and *Sentence Segmentation*. The *Word Segmentation* step is required since the Thai language has no word boundary marker in its writing system. An automatic Thai word segmentation tool, (SWATH), was applied to segment words in TTSMT. Table 4 shows an example of word segmentation.

The treatment rules are utilized, since most TSL are not expressed in a complex or compound form. Therefore, any input string which contains a conjunction, e.g. "และ (and)", "หรือ (or)", "แต่ (but)", "เพราะ (because)" will be converted into a sequence of simple sentences in the *Sentence Segmentation* step. Table 5 illustrates an example of a segmented sentence.

Table 4. Example of a segmented word.

Segmented Input	Segmented Word
พ่อกินก๋วยเตี๋ยวแต่แม่กินข้าว (The father eats noodle, but the mother eats rice)	พ่อ-กิน-ก๋วยเตี๋ยว-แต่-แม่-กิน-ข้าว (The father eats noodle, but the mother eats rice)
[พ่อ-father],[กิน-eat],[ก๋วยเตี๋ยว-noodle],[แต่-but],[แม่-mother], [กิน-eat],[ข้าว-rice]	

Table 5. Example of a segmented sentence.

Segmented Input	Segmented Sentence
พ่อ-กิน-ก๋วยเตี๋ยว-แต่-แม่-กิน-ข้าว (The father eats noodle, but the mother eats rice)	1. (พ่อ-กิน-ก๋วยเตี๋ยว) (The father eats noodle) 2. (แต่) (but) 3. (แม่-กิน-ข้าว) (the mother eats rice)
[พ่อ-father],[กิน-eat],[ก๋วยเตี๋ยว-noodle],[แต่-but],[แม่-mother], [กิน-eat],[ข้าว-rice]	

3.2 Word Treatment

This module generates the intermediate sign code (ISC) from segment input, which will be used as a keyword in the *Sign-Code Ordering* module, in three steps: *Word Constraint, Word Addition, and Dictionary Lookup*. An ISC links a Thai word to its corresponding Thai Signs. Thai language constraints, TSL constraints, the Thai-Sign dictionary, the Thai dictionary and the Sign dictionaries are used in this module.

The Word Constraint is applied to simplify the structure of a Thai input sentence by removing some Thai words which are not required to be translated into the TSL due to the differences between Thai and TSL grammars. Such words include synonyms, conjunctions and units of countable nouns[2]. If there are synonyms in the sentence and one locates directly next to the

[2] In this research, the word "classifier" is avoided to be used as a unit of countable nouns.

others in the sentence, only the first word is kept while the others are removed. In Table 6, the word "*มากมาย* (many)" is a synonym of the word "*เยอะแยะ* (many)". Therefore, the word "*เยอะแยะ* (many)" is removed.

Table 6. Example of word constrain.

Segmented Sentences	Constrained Phrases
เด็ก-มี-ตุ๊กตา-*มากมาย-เยอะแยะ*	เด็ก-มี-ตุ๊กตา-*มากมาย*
(A children has many many dolls)	(A children has many dolls)
[เด็ก-children],[มี-has],[ตุ๊กตา-doll],[*มากมาย*-many],[*เยอะแยะ*-many]	

After applying all related constraints, any word needed to complete the string (according to the TSL grammar) and to retain the meaning of the input sentence will be added in the *Word Addition* step. In the following example (Table 7), the word "บ้าน (home)" was added to specify the place where "I" and "grandma" live in.

Table 7. Example of Word Addition.

Constrained Phrases	Word Added Phrases
ฉัน-อาศัยอยู่-กับ-ยาย	ฉัน-อาศัยอยู่-*บ้าน*-กับ-ยาย
(I live with grandma)	(I live at *home* with grandma)
[ฉัน-I],[อาศัยอยู่-live],[กับ-with],[ยาย-grandma],[บ้าน-home]	

In *Dictionary Lookup*, each input word will be mapped to all possible corresponding ISC. Some Thai words may correspond to more than one ISC, e.g. "กิน (eat)" and "เดิน (walk)" correspond to at least seven and three different ISCs respectively (Table 8), these results in different ISCs from the input sentence. The most appropriate ISC will be selected in the Sign-Code Selection.

Table 8. Examples of words and its corresponding to ISCs.

Word	Possible Corresponding ISC
กิน	- ISC of human-eat-rice
(eat)	- ISC of human-eat- noodle
	- ISC of mammals-four legs-eat (except elephant)
	- ISC of elephant-eat
	- ISC of birds-eat
	- ISC of aquatic-birds-eat
	- ISC of poultry-eat
เดิน	- ISC of human-walk
(walk)	- ISC of mammals-four-legs-walk
	- ISC of animal-two-legs-walk

3.3 Sign-Code Selection

The ISC's codes were created by combined some advantages of *Word Association Number* (WordAsso) (Naruedomkul and Cercone, 1999) with our new idea. The WordAsso is adapted from the concept classification, so words are classified into groups. The ISC's code property keeps the relation of ISC that can determine the involvement with other ISC codes.

The most appropriate ISC of each input word is selected in this module (if there are more than one) based on the *semantic* and *syntactic* relationship.

The Semantic relationship refers to the relation between the words in close proximity which can be determined by their context. If there are more than one ISC correspond with one input word. To select the correct ISC, each ISC's code will be examined with the left and right ISC's codes to identify which ISC's code match with the left and right ISC's codes. For example, the word "ข้าว (rice)" in the sentence "คนกินข้าว (human eats rice)" refers to "เมล็ดข้าว (rice grain)" while it refers to "ต้นข้าว (rice plant)" in the sentence "ข้าวอยู่ในนา (rice is in a field of rice)". Therefore, different ISCs will be selected for each "ข้าว (rice)" which corresponds with the semantic.

Selecting the sign gesture for a verb is quite complicated, since it must agree with subject and /or object and voice of sentences (active or passive voice).

The syntactic relationship refers to the agreement of verb, subject and object gestures that determined by the conformance of *verb* with *subject* or *verb with both subject and object*. If a verb word corresponds with many ISCs. The ISC's codes are considered with the subject and/or ISC's code of object which ISC's code relate with subject and object. For example, the verb "กิน (eat)" in the sentence "คนกินข้าว (human eats rice)" and in the sentence "คนกินก๋วยเตี๋ยว (human eats noodle)" were mapped to different ISCs since the objects of each sentence are different. In the sentence "ไก่กินข้าว (A hen eats rice)" and "หมากินข้าว (A dog eats rice)", the verb "กิน (eat)" is mapped to another ISC according to its subject. For active voice and passive voice, both sentence "ไก่กินหนอน (A hen eats worm)" and "หนอนถูกไก่กิน (The worm was eaten by a hen)", the hen eating gesture (peak) is selected, not worm eating gesture as shown in Table 9.

Table 9. Examples of selected ISC.

Phrase	Selected ISC	Sign (verb only)
Verbs depend on both subject and object		
คนกินข้าว (A human eats rice)	คน-คนกินข้าว-ข้าว (human- use a spoon to eat - rice)	 spoon
คนกินก๋วยเตี๋ยว (A human eats noo-dle)	คน-คนกินก๋วยเตี๋ยว-ก๋วยเตี๋ยว (human- use a chopstick to eat- noodle)	 chopstick
[คน-human],[กิน-eat],[ข้าว-rice], [ก๋วยเตี๋ยว-noodle], [คนกินข้าว-use a spoon to eat],[คนกินก๋วยเตี๋ยว-use a chopstick to eat]		
Verbs depend on subject		
ไก่กินข้าว (A hen eats rice)	ไก่-จิก-ข้าว (hen- peck -rice)	 peck
หมากินข้าว (A dog eats rice)	หมา-เลีย-ข้าว (dog- lap -rice)	 lap
[ไก่-hen],[กิน-eat],[ข้าว-rice],[จิก-peck][หมา-dog],[เลีย-lap]		
Active voice and passive voice		
ไก่กินหนอน (A hen eats worm)	ไก่-จิก-หนอน (A hen pecks worm)	 peck
หนอนถูกไก่กิน (A worm is eaten by hen)	หนอน-ไก่-จิก (worm-hen-peaks)	 peck
[ไก่-hen],[กิน-eat],[หนอน-worm],[จิก-peck]		

3.4 Sign-Code Ordering

Owing to the structures of the TSL phrases, sentences and compound sentences are different from the Thai language. Therefore, the ISCs selected in the previous step must be rearranged. To perform the correct grammatical ordering, three categories of word ordering rules were created from the syntactic level to order the intermediate sign codes (ISC) into sign code script (SCS). First, the phrase structure rules line up words in a phrase. Second, the sentence structure rules order phrases in a sentence, and the multi-sentence structure rules rearrange sentences in a compound sentence. Those rules correspond with the TSL grammar as shown in Table 10.

Table 10. Some structure rules.

No.	TSL Phrase Structure Rules
1	Neg = (V, Adj, Adv) + Neg
2	P1 = NP + P
3	P2 = NP_2+ CL+NP_1+ P
4	P3 = NP_3+ CL+ NP_2+ P_2+CL+ NP_1+ P_1
	TSL Sentence Structure Rules
5	V1 = S + V
6	V2 = S + V + Neg
7	V3 = O + CL + S + V
8	V4 = O + CL + S + V + Neg
9	V5 = indO + CL+ diO + CL+ S + V
	TSL Sentence Ordering rule
10	เพราะ [Because] = S_2 + S_1

Table 10, row 1 shows the rule for ordering a negation. Row 2 represents a preposition phrase (PP) ordering rule. Row 3 presents the rule for ordering a noun phrase (NP) and a PP while the rule for ordering NPs and PPs is in row 4. Row 5 is a intransitive verb ordering rule. Row 6 and 8 display the rule for ordering a negative sentence. Row 7 represents the rule for ordering a transitive verb while the rule in row 9 is the ditransitive verb. The rule for ordering a compound sentence indicates that in row 10.

Sign-Code Ordering rearranges the ISC string in TSL grammatical order and can order accurately both active voice and passive voice sentences. This module consists of two steps: *Ordering* and *Sentence integration steps*.

The Ordering step orders the ISC to SCS according to the TSL grammatical order. There are two steps in the Ordering step: extraction and ordering steps.

The Extraction step composes of searching and extracting processes. After the ISC was selected successfully, the searching process traverses each code in the ISC to search the head word (*Verb, Prep* and *Neg*). The head word is identified as a root of tree. The extracting process extracts the ISC to the sub- ISCs (NP_{ln}, NP_{rn}, *R*) as shown in Figure 2.

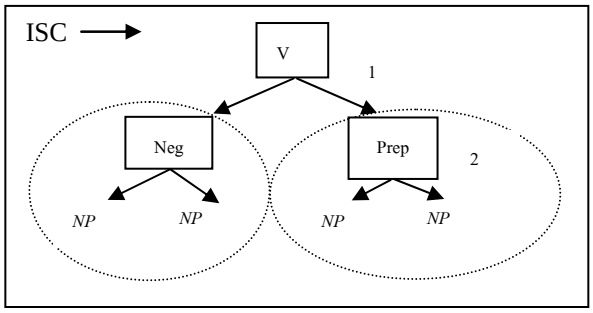

Figure 2. Example of extraction.

The Ordering step is an operation to order the ISC to the SCS. There are three processes in this step: mapping, ordering and concatenating. The mapping process maps *NP* from the lowest level of the left node to the right node groups with the

word ordering rule by considering the word category of the root of a sub-ISC. The ordering process rearranges the nodes in the *NP* which corresponds to the word ordering rule. After rearrange the word order, the concatenating process joins *NP* from the lower to the higher levels as shown in Table 11.

Table 11. Examples of Sign code ordering.

Thai sentence = แมว-ไม่-น่ารัก-กิน-ปลา-เล็ก-ใน-จาน-ใหญ่
(The cat is not beautiful eat small fish on the big dish)
ISC = {[แมว-cat],[ไม่-not],[น่ารัก-beautiful],[กิน-eat],[ปลา-fish], [เล็ก-small],[ใน-on],[จาน-dish],[ใหญ่-big]}
The extraction step:
Level 1
ISC = {[แมว-cat],[ไม่-not],[น่ารัก-beautiful],[กิน-eat],[ปลา-fish], [เล็ก-small],[ใน-on],[จาน-dish],[ใหญ่-big]}
R = {[กิน-eat]}
NP_{l1} = {[แมว-cat],[ไม่-not],[น่ารัก-beautiful]}
NP_{r1} = {[ปลา-fish],[เล็ก-small],[ใน-on],[จาน-dish],[ใหญ่-big]}
Level2L
NP= {[แมว-cat],[ไม่-not],[น่ารัก-beautiful]}
R = {[ไม่-not]}
NP_{ll2} = {[แมว-cat]}
NP_{lr2} = {[น่ารัก-beautiful]}
Level2R
NP= {[ปลา-fish],[เล็ก-small],[ใน-on],[จาน-dish],[ใหญ่-big]}
R = {[ใน-on]}
NP_{rl2} = {[ปลา-fish],[เล็ก-small]}
NP_{rr2} = {[จาน-dish],[ใหญ่-big]}
The reordering step:
Level2R
NP= {[ปลา-fish],[เล็ก-small]},{[ใน-on]},{[จาน-dish],[ใหญ่-big]}
Mapped *NP* with the ordering rule $NP = NP_2+CL+ NP_1+ P$
Reorder *SCS* = {[จาน-dish],[ใหญ่-big],CL,[ปลา-fish], [เล็ก-small],[ใน-on]}
Level2L
NP= {[แมว-cat],[ไม่-not],[น่ารัก-beautiful]}
Mapped *NP* with the ordering rule $Neg = NP_1+ NP_2+ Neg$
Reorder *SCS* = {[แมว-cat],[น่ารัก-beautiful],[ไม่-not]}
Level1
NP= {[แมว-cat],[น่ารัก-beautiful],[ไม่-not]},{[กิน-eat]}, {[จาน-dish],[ใหญ่-big],CL,[ปลา-fish],[เล็ก-small],[ใน-on]}
Mapped *NP* with the ordering rule $V = O + CL+S + V$
Reorder *SCS* = {[จาน-dish],[ใหญ่-big],CL,[ปลา-fish],[เล็ก-small], [ใน-on],CL,[แมว-cat],[น่ารัก-beautiful],[ไม่-not] [กิน-eat]}
Generated sign code script
SCS = {[จาน-dish],[ใหญ่-big],CL,[ปลา-fish],[เล็ก-small], [ใน-on],CL,[แมว-cat],[น่ารัก-beautiful],[ไม่-not],[กิน-eat]}

Figure 2 and Table 11 show the extracting and ordering of the ISC "แมว-ไม่-น่ารัก-กิน-ปลา-เล็ก-ใน-จาน-ใหญ่ (The cat is not beautiful eat small fish on the big dish)".

The last step, *Sentence Integration*, is activated only if the input sentence was divided into sub-sentences. The sentence integration combines all sub-sentences together. As previously stated ISCs have been ordered to SCSs separately. To complete the ordering, all of SCSs are integrated into one SCS. There are two processes in this step: mapping and integration processes. The mapping process matches the conjunction with the multi-sentence rule to integrate the SCSs. The integration process combines the SCSs which correspond to the rule as shown in Table 12.

3.5 Image Mapping

The final module generates TSL by mapping a list of SCS with the pictures of the TSL.

Table 12. Examples of Sentence Integration.

Thai Sentence	Sentence Integration
1. พ่อ-ได้-เงิน-มาก (The father gets a lot of money) 2. เพราะ (because) 3. พ่อ-ทำงาน-หนัก (The father works hard)	[พ่อ-father],[ทำงาน-work], [หนัก-hard],CL,[เงิน-money], [มาก-a lot of],[พ่อ-father], [ได้-get]
[พ่อ-father],[ได้-get],[เงิน-money],[มาก-a lot of],[เพราะ-because], [ทำงาน-work],[หนัก-hard]	

4 Knowledge-Based Representation

The essential Thai and TSL knowledge-bases required for TTSMT include: Treatment, Constraint, Dictionary, Grammar and Lexicon. The knowledge-bases are developed separately for easy modification, extendibility and management.

Treatment is the set of specific characteristics of the Thai language that is not available in the TSL e.g. "และ (and)", "หรือ (or)", "แต่ (but)", "เพราะ (because)".

Constraints are the different characteristics between the Thai language and the TSL. There are two sets of constraints: Thai language and TSL constraints.

- Thai language constraints are used to simplify the structure of the Thai language and to narrow the scope of possible TSL words which correspond to each Thai language word such as the synonym "สวย (beautiful)", "น่ารัก (pretty)", "มีเสน่ห์ (charming)", conjunctions "และ (and)" and units of countable nouns "อัน (aun)", "ชิ้น (chein)".

- TSL constraints are required to retain the meaning of the Thai language and to make them grammatically correct. For example the words "พักอยู่(stay)", "อาศัย (live)", "อยู่อาศัย (reside)" are the intransitive verb in the Thai language, but these

must remain in a place in the TSL. To specify the place, the word "บ้าน (home)" must be added.

Dictionary, There are three sets of dictionary: Thai dictionary, Sign dictionary and Thai-Sign dictionary. Each entry in dictionary contains semantic and syntactic categories.

Grammars and Lexicons of the Thai language and the TSL have been developed base on the HPSG (Pollard and Sag, 1994). This grammar formalism is well suited to the TSL because it operates on feature structures allowing access and modification of various levels of linguistic specification. Each lexical entry contains morphological, syntactic and semantic information.

5 Experiment

The Thai to Thai Sign MT system has been developed using TTSMT approach and run under SWI Prolog 5.2. The grammar using in this system was developed using Attribute Logic Engine version 3.2.1. With the proposed design, TTSMT does not need a human operator to intervene between pipeline stages to correct the linguistic analysis prior to final surface realization.

The developed system was used to translate phrases/sentences. Some translation results are shown in Table 13.

Table 13. Examples of translation results.

Thai Sentence	Intermediate Sign Code	Sign Code Script
ลูกเป็ด-ไม่-น่ารัก (The duckling is not cute.)	[ลูกเป็ด-duckling], [ไม่-not], [น่ารัก-cute]	[ลูกเป็ด-duckling], [น่ารัก-cute], [ไม่-not]
บ้าน-ไม่-พัง (The home does not fall down)	[บ้าน-home],[ไม่-not], [พัง-fall down]	[บ้าน-home], [พัง-fall down], [ไม่-not]
หมาป่า-เจ้าเล่ห์-เหิ่น-หมูน้อย-ใน-บ้าน (The sly wolf see the little pig in the home)	[หมาป่า-wolf], [เจ้าเล่ห์-sly], [เหิ่น-see], [หมูน้อย-little pig], [ใน-in],[บ้าน-home]	[บ้าน-home],CL, [หมูน้อย-little pig], [ใน-in],CL, [หมาป่า-wolf], [เจ้าเล่ห์-sly], [เหิ่น-see]
ชาวนาใจดีให้ฟางทั้งหมดกับหมูน้อย (The kind farmer give all straw to the little pig)	[ชาวนา-farmer], [ใจดี-kind],[ให้-give], [ฟาง-straw], [ทั้งหมด-all],[กับ-to], [หมูน้อย-little pig]	[หมูน้อย-little pig], CL,[ฟาง-straw], [ทั้งหมด-all],CL, [ชาวนา-farmer], [ใจดี-kind], [ให้-give]

Table 13, the negation in the phrase is translated in row 1. Row 2 shows the translation of the negation in the sentence. Row 3 represents the translation of the preposition in the sentence. The translation of ditransitive verb is in row 4.

6 TTSMT Evaluation

TTSMT evaluation was restricted to test the translated processes without considering the potential environments. The 297 test sample sentences /phrases were collected from different sources: textbooks (affirmative, negative, interrogative, order, compound sentences, active and passive voice), the bedtime story (The three little pigs), cartoons, newspapers and the public labels.

The accuracy, precision, recall and F-measure were used to determine the translation in terms of intelligibility and fidelity by five testers including two linguistic teachers (hearing and deaf), two deaf students and one Thai sign interpreter. All testers are from Ratchasuda College, a college for persons with disabilities in Bangkok, Thailand. Here, the intelligibility and fidelity are evaluated from the following 7 measures. The evaluation results are shown in Table 14.

1 = correct grammar,
2 = correct word usage,
3 = inappropriate word usage,
4 = incorrectly translated word,
5 = correct CL,
6 = convey the original meaning,
7 = convey the different meaning but can understand the original meaning.

Table14. The intelligibility and fidelity evaluation results.

Measurement							Accu-racy (%)	Preci-sion (%)	Recall (%)	F (%)
1	2	3	4	5	6	7				
*	*			*	*		94	100	94	96.9
*		*		*	*		95	100	95	97.4
*			*			*	94	100	94	96.9
	*			*		*	93	100	93	96.3
*	*				*		94	100	94	96.9

In Table 14, row 1 displays the grammar, word usage and CL that are correct and convey the original meaning. Correct grammar, inappropriate word usage, correct CL and convey the original meaning are shown in row 2. Row 3 presents correct grammar, incorrectly translated

word, convey the different meaning but can understand the original meaning. In the same way correct word usage, correct CL, convey the different meaning but can understand the original meaning are represented in row 4. Correct grammar, correct word usage and convey the original meaning are in row 5.

The first step, TTSMT can be augmented into the Thai-Thai Sign educational application by mapping the SCS with pictures of TSL gestures. The application was developed to satisfy the needs of users, so the user interface was designed to support users from the beginning level and deaf, hearing impaired students whose first language is the TSL as shown in Figure 3.

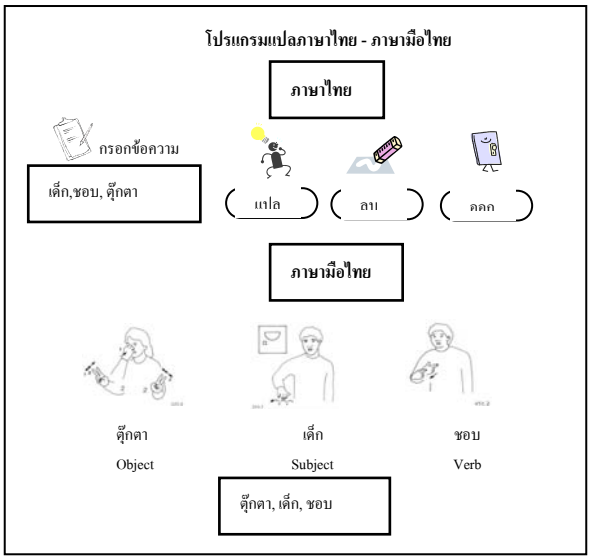

Figure 3. TTSMT user interface.

The questionnaires were used to collect data to evaluate users' satisfaction in terms of simple and convenient, good advantage and need of user. The users' satisfaction evaluation results as shown in Table 15.

There are five levels of satisfaction: 5 means excellent; 4 means good; 3 means fair; 2 means poor; 1 means very poor.

Table 15. Users satisfaction evaluation.

	Simple Convenient	Good advantage	Need of user
Mean	4.03	4.12	4.07
S.D.	0.66	0.65	0.64
Fair (%)	20.2	16.1	17.2
Good (%)	56.3	55.8	58.6
Excellent (%)	23.5	28.1	24.2

The results in Table 15 show that users satisfy in Thai-Thai Sign education application.

7 Concluding Remarks

We present TTSMT, an alternative approach to translate the Thai language into the Thai Sign language. TTSMT is able to generate an accurate translation result since it takes into account the differences between the Thai language and the TSL in terms of both lexicons and syntax. TTSMT exploits the advantage of semantic and syntactic relationship between words in close proximity to generate an appropriate list of sign code script which best represents input Thai sentences.

In our initial implementation, TTSMT was applied to develop an automated translation system for assisting hearing impaired and deaf people to enhance their abilities in learning, to read and to communicate with other people.

In addition, with some educational applications, this system can turn into a Thai Sign Language (tutor) training tool for students (both hearing and hearing impaired) to learn the Thai language and the TSL faster and easier.

References

Angus B. and Smith G. 1999. English to American Sign Language machine translation of weather reports. *Proceedings of the second High desert Student Conference in Linguistics, Albuquerque, NM.*

Armond D. and Speers L. 2001. *Representation of American Sign Language for Machine Translation*. PhD dissertation, Department of linguistics, Georgetown University.

Bangham J. A. and Cox S. J. 2000. Signing for the deaf using virtual humans. *In IEEE Colloquium on speech and Language Processing for the Disabled and Elderly People, London, England.*

Gallaudet Research Institute. 1996. *Stanford Achievement test,* 9th Edition, Form S, Norms Booklet for Deaf and Hard-of-Hearing Student. Gallaudet University, Washington.

Huenerfauth M. 2004. Spatial and Planning Models of ASL Classifier Predicates for Machine Translation. *In The 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2004)* Baltimore, MD., pp.65–74.

Huenerfauth M. 2005. American Sign Language Generation: Multimodal NLG with Multiple Linguistic Channels. *In Proceedings of the ACL Student Re-*

search Workshop (ACL 2005) Ann Arbor,MI., pp.37–42.

Morrissey S. and Way A. 2005. An example-based approach to translating sign language. *In Workshop Example-Based Machine Translation (MT X–05)*, pp. 109–116, Phuket, Thailand, September.

Naruedomkul K. and Cercone N. 1999. The Role for Word Association Numbers in Machine Translation. *Proceedings of the Conference Pacific Association for Computational Linguistics (PACLING'99)*, Waterloo, Ontario, Canada, pp.379-392.

National Association of the Deaf in Thailand. 2000. *Thai Sign Language Book*. Baphit Press: Bangkok (Thailand).

Pollard C. and Sag A. I. 1994. *Head-driven Phrase Structure Grammar*. Center for the Study of language and Information, Stanford, The University of Chicago Press, Chicago and London.

Safar E. and Marshall I. 2000. The architecture of an English-Text-to-Sign-Language translation system. *In G. Angelova, editor, Recent Advances in Natural Language Processing (RANLP)*, pp. 223-228, Tzigov Chark, Bulgaria.

Stokoe W. C. 1978. *Sign language structure*. (2nd ed.), Silver Spring, MD:Linstok Press.

Suszczanska N. and Szmal P. 2002. Translating Polish Text into Sign Language in the TGT System. *Proceeding of the 20th IASTED International Multi-Conference Applied Informatics AI 2002*, Innsbruck, Austria 2002, pp. 282-287.

Sutton-Spence R. and Woll B. 1999. *The linguistics of British Sign language, An Introduction*. University Press, Cambridge.

SWATH. Smart Word Analysis for Thai., *http://www.links.nectec.or.th/download.php*, as of 10/01/07.

Veale T. and Collins B. 1998. The challenges of cross-modal translation: English to sign language translation in the ZARDOZ system. *In Machine Translation 13*, pp. 81-106.

Zhao L. and Kipper K. 2000. A Machine Translation System from English to American Sign Language. *Association for Machine Translation in the Americans.*

Zijl L. V. and Barker D. 2003. A Machine Translation System for South African Sign Language. *In ACM SIGGRAPH Conference Proceedings Series*, Cape Town South Africa, February, 2003.