CLASSIFICATION OF SERIAL VERB CONSTRUCTIONS IN THAI

Surasa Sookgont[1], Thepchai Supnithi[2] and Rachada Kongkachandra[3]

[1,3] Department of Computer Science, Faculty of Science and Technology, Thammasat University, Klongluang, Pathumthani, 12120, THAILAND.
[2] Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center, Klongluang, Patumthani, 12120, THAILAND.

ABSTRACT

The objective of the research is to classify the serial-verb constructions in Thai automatically by using the word classes from Thai WordNet to classify verbs in the sentence. Due to the Thai language has the extend-to-the-right structure and put the adjective after the noun. Its overall grammar characteristic is the "Subject-Verb-Object" or SVO type. And Thai language can be communicated using one verb after another within the same sentence, that we called "Serial Verb". Today we already have many researches about this serial-verb constructions, but no research is about its automatic classification.

KEYWORDS

Thai WordNet, Serial Verb Contruction, Automatic Classification

1. INTRODUCTION

Thai language is an isolating language. A Thai sentence is constructed by putting words in sequence from left to right. The form of each word is not changed according to tense, case, mood, or voice. Also, Thai words have neither gender nor number. Many words in Thai cannot identify their static meanings, but have to see their context before interpret. Different from English language, all adjective used for clarifying noun are placed after noun instead. The syntactical characteristic of Thai language is the "Subject-Verb-Object". In English language, a sentence should contain only one main verb, but in Thai many verbs in one sentence could be happened. It is called the "Serial Verb" such as "ฉันนั่งอ่านหนังสือ" (I sit reading a book.), "เขาเดินคุยโทรศัพท์" (He walks talking on the phone.), "เขาเด็ดดอกไม้ดม" (He picks a flower to sniff.), and so on.

In Thai language, a sentence is divided into 2 parts, i.e. subject and predicate. Subject Part is an important part used to tell the readers/listeners who the subject of the sentence is. Most of the subject words are the noun and pronoun. Predicate part describes the state or behaviour of its subject, to tell the readers/listeners which state or behaviour that the subject expresses. Most of the predicate part is the verb of the sentence that can be either the transitive verb or intransitive verb.

"Serial verb constructions has been widely interested in Linguistic researches both Thai and other languages such as Chinese and some languages in East Asia."[1] "The sentence structure has 2 serial verbs or more within the same sentence. These verbs use the same subject, and are put altogether without any conjunction between verbs. But the sentence may have or not have the

noun as the object, such as "เขาตะโกนตอบ" (He shouts answering.), "เขายืนร้องเพลง" (He stands singing.), "เขาถูกรถชน" (He gets hit by the car.), etc.

The examples of researches that study serial verb constructions in Thai are [2], [3], [4], and [5]. Factors relating to properties of serial verb constructions absolutely affect translating structure which mean translating a language to another couldn't be perfectly translated in every single word. However, to translate a language to another even though the language that will be translated doesn't have serial verb constructions, after translated better has. Most of the researches tend to focus on the understanding of the constructions of the sentences, consequences, and meaning which semantics will be a tool that play an important role to help emphasize the meaning of the serial verb constructions to be clearer. In general, the conceptual framework to analyse properties of serial verb constructions in Thai language, most of the researches will study the unit of serial verb constructions based on syntax and semantics. Most of the researchers have mutually agreed on these following properties of serial verb constructions [2] (1) No conjunctions between serial verb constructions (2) Every verb in serial verb constructions is mutual in term of aspects, timing, rejection. (3) Verbs in serial verb constructions that are mutual will contain at least one argument (4) Serial Verb Constructions will indicate only a circumstance at a time, but complicated.

With the fact that serial verb constructions have many aspects in term of views according to the study of each researcher. The researches which the author have found, most of them hard to apply linguistically analysis to classify serial verb constructions by programming, except the study conducted by [5] which clearly classified each verb class which can be applied to develop the program and the researcher not found any research that automatically classify the serial verb constructions. Therefore, the author deems the study of Cholthicha [5] can be the beginning of development useful and help in machine translation between Thai and other languages better.

2. RELATED WORKS

The study related research is intended to gathering ideas and theories about SVCs to guide the application of this research follows:

2.1. Serial Verb Constructions in Thai

In the study above describes the grammar and meaning of serial-verb constructions that can be classified to 8 types as below:

2.1.1. Motion Serial Verb Constructions

Consist of the first verb from Motion Verbs such as เดิน (walk), วิ่ง (run), ขับรถ (drive), followed by Deictic Verbs (ไป (go), มา (come)) or Directional Verbs (such as ไป (go), มา (come), ขึ้น (go up), ลง (go down), เข้า (go in), ออก (go out))
1. กานดา เดิน *ออก ไป* ร้องเพลง (Kanda walk *exit go* sing)
2. การดา เดิน *เข้า* โรงเรียน ไป (Kanda walk *enter* school go)

2.1.2. Posture Serial Verb Constructions

Consist of the first verb from Postural Verbs such as ยืน (stand), นั่ง (sit), นอน (sleep), followed by any verb that we call "Open Class Verb".
3. กานดา ยืน *เคาะ* ประตู (Kanda stand *knock* door)
4. กานดา นั่ง *ร้องเพลง* (Kanda sit *sing*)

2.2.3. Take-Serial Verb Constructions

Consist of the first verb which must be, "เอา" or take, followed by any verb (Open Class Verb) which usually the verbs are activity verbs.

5. กานดา <u>เอา</u> ผ้า *ใส่* ตะกร้า (Kanda <u>take</u> cloth *put* basket)

2.2.4. Use-Serial Verb Constructions

Consist of the first verb which must be, "ใช้" or use, followed by any verb (Open Class Verb) which usually the verbs are activity verbs

6. กานดา <u>ใช้</u> มีด *หั่น* ไก่ (Kanda <u>use</u> knife *cut* chicken)

2.2.5. Open Class Serial Verb Constructions

Consist of any verb (Open Class Verb) and verbs in the sentence must not be all the Stative Verb. The example verbs from Stative Verb are "เชื่อ" or believe, "คิด" or think, "ดีใจ" or glad, etc.

7. กานดา <u>หุง</u> ข้าว *กิน* (Kanda <u>cook</u> rice *eat*)

2.2.6. Give-Serial Verb Constructions

Consist of the first verb which must be, "ให้" or give (or let), followed by any verb (Open Class Verb.) But this "ให้", or give, cannot be used with the subject or object that is not the living things because it would be ungrammatical sentence as (9)

8. กานดา <u>ให้</u> จุ่ม *อ่าน* หนังสือ (Kanda <u>give</u> Jum *read* book)
9. นวล *ทาสี* <u>ให้</u> บ้าน (Nuan *paint* <u>give</u> house)

2.2.7. Causative Serial Verb Constructions

Consist of the first verb which must be, "ทำ" or make, followed by any verb (Open Class Verb) that is the Intransitive Verb because if the followed verb is not intransitive, it would be grammatically wrong as (11).

10. กานดา <u>ทำ</u> เด็ก *ร้องไห้* (Kanda <u>make</u> child *cry*)
11. กานดา <u>ทำ</u> เด็ก *อ่าน* หนังสือ (Kanda <u>make</u> child *read* book)

2.2.8. Resultative Serial Verb Constructions

Consist of any verb (Open Class Verb) that the second verb in the sentence is the result from the first verb.

12. กานดา <u>ผลัก</u> วีระ *ล้ม* (Kanda <u>push</u> Weera *fall*)

2.2. Thai WordNet

The researcher use the Thai WordNet (Thai WordNet is a large lexical database of Thai) to assist in classifying words and group the words altogether. Thai WordNet was constructed in 2007 based on the Princeton's word network version 3.0 and the two-language dictionary within the WordNet database. It classifies Thai words into 4 groups: Noun, Verb, Adjective, and Adverb. Each group is classified into separated subgroups as below:

Table 1. Noun class in WordNet

Noun class	Meaning
noun.act	action activity
noun.animal	Animal
noun.artifact	things that human created
noun.attribute	properties
noun.body	body
noun.cognition	knowledge and understanding
noun.communication	communication
noun.event	event or incident
noun.feeling	feeling
noun.food	food
noun.group	group
noun.location	location or place
noun.motive	motive, objective, reason
noun.object	natural phenomenon
noun.person	human person
noun.plant	plant
noun.possession	possession, asset
noun.process	process
noun.quantity	quantity
noun.relation	relationship, cousin
noun.shape	shape, form
noun.substance	substance, matter
noun.time	time

Table 2. Verb class in WordNet

Verb class	Meaning
verb.body	Verb about maintaining and functioning in each part of the body
verb.change	Verb about changing
verb.cognition	Verb about thought
verb.communication	Verb about communication
verb.competition	Verb about competition
verb.consumption	Verb about consumption
verb.contact	Verb about contacting
verb.creation	Verb about creating or making
verb.emotion	Verb about feeling and emotion
verb.motion	Verb about motion
verb.perception	Verb about perception, classified to 5 senses of perception.
verb.possession	Verb about possession
verb.stative	Verb that tells the state/condition.
verb.social	Verb about behavior in society
verb.weather	Verb about weather and climate

2.3. Compound Word in Thai

This work studied about the word construction, type of compound word that came from the verb as the first word to combine.

Compound word is the word that came from 2 words combined together. It may create a new meaning, or the same meaning, or more concise meaning (Compound word = First word + Additional word.) It may use the first word as the verb, and additional word as noun, verb, or adjective. [6][7]

Table 3. Example of compound word with the verb as the first word

Compound word	First word	Additional word
กระจายเสียง (Broadcast)	กระจาย (Distribute)	เสียง (Sound)
อุ้มท้อง (Pregnant)	อุ้ม (Hold)	ท้อง (Pregnant)

3. METHODOLOGY

The researcher has analyzed the word groups compare with the word class in WordNet as below:

3.1. Motion Verb can be compared with verb.motion because the meaning from the WordNet is the verb about motion.

3.2. Postural verb can be compared with verb.body (verb about function of each body part) such as นอน (sleep), แต่งตัว (dress, put on) and verb.contact (verb about contacting) such as ยืน (stand)

3.3. Stative verb can be compared with verb.stative (verb that tells the state/condition), verb,emotion (verb that tells the feeling or emotion), verb.possession (verb that tells the possession), verb.cognition (verb about thought) and verb.perception (verb about 5 senses of perception) because the Stative verb means the verb that describe the state of something that cannot visibly express, just to tell or explain the state or condition that happened.

3.4. Activity Verb. The researcher conclude that the word groups that are not in Stative verb can be included in this group, that are verb.body (verb about maintaining and functioning of each body part), verb.change (verb about changing), verb.communication (verb about communication), verb.competition (verb about competition), verb.consumption (verb about consumption), verb.creation (verb about creating and making), verb.motion (verb about motion), verb.social (verb about behavior in society) and verb.weather (verb about weather and climate)

Table 4. Show comparable between groups of words with WordNet class.

Word group	WordNet Class
Motion verb	verb.motion
Postural verb	verb.body
	verb.contact
Stative verb	verb.stative
	verb.emotion
	verb.possession
	verb.cognition
	verb.perception

Activity verb	verb.body
	verb.change
	verb.communication
	verb.competition
	verb.consumption
	verb.contact
	verb.creation
	verb.motion
	verb.social
	verb.weather

3.5. Mapping the [PPRS] from Orchid to noun.person from WordNet because [PPRS] is personal pronoun

3.6 In case of Give-Serial Verb Constructions that has to use noun as living things (animate subject), the researcher would use the noun class in WordNet as below:

Table 5. Show comparable between an animate with WordNet class

Noun	WordNet Noun Class
Noun in the form of living things	noun.animal
	noun.person

3.7 Determined that Directional Verb (DV) [1] consists of verbs as: ไป (go), มา (come), ขึ้น (up), ลง (down), เข้า (enter), ออก (exit)

Summary of 8 types of serial verb constructions as below:

Table 6. Eight types of serial verb constructions in Thai

Types of serial verb constructions	Verb group
Motion SVCs	Motion verb + directional verb
Posture SVCs	Postural verb + open class verb
Take-SVCs	"เอา" (take) + open class verb (activity verb)
Use-SVCs	"ใช้" (use) + open class verb (activity verb)
Open class SVCs	Open class verb + open class verb *both verbs must not be the stative verb.
Give-SVCs	"ให้" (give) + open class verb (not stative verb) *Required animate subject.
Causative SVCs	"ทำ" (make) + open class verb
Resultative SVCs	Open class verb + open class verb (result)

3.8 Experiment procedure

Gather the sentences from data collecting and save as the .txt-extension file. Then bring this file to segment the sentences in words and identify the word function of Part of Speech (POS) with Orchid. Filter the result to identify the noun, verb, and/or the words that Orchid cannot identify their POS. Do the additional search from Thai WordNet to identify the word class. After finished the data preparation, compare the verb group in the sentence with the 8 types of serial verb constructions.
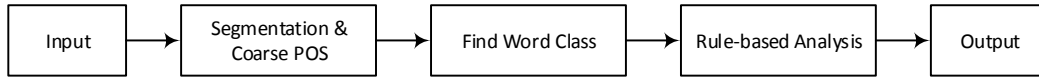
```
┌─────────┐   ┌──────────────┐   ┌──────────────┐   ┌───────────────────┐   ┌─────────┐
│  Input  │──▶│ Segmentation │──▶│ Find Word    │──▶│ Rule-based        │──▶│ Output  │
│         │   │ & Coarse POS │   │ Class        │   │ Analysis          │   │         │
└─────────┘   └──────────────┘   └──────────────┘   └───────────────────┘   └─────────┘
```

Figure 1. Overview of the process

4. EXPERIMENTS

This topic described the result from the experiment that use the type of serial-verb constructions that linguists classified into 8 types in Table.6 to apply to be the automatic sentence-classifying program by using the data from Thai WordNet for verb class identification. This experiment use the sentences from the reference research.

13. เขา@NPRP เดิน@VACT ไป@XVAE ร้องเพลง@VACT
    เขา(noun.person) เดิน(verb.motion) ไป(DV) ร้องเพลง(verb.communication)
    Expect output = Motion SVCs
    Actual Output = Motion SVCs

14. เขา@NPRP นอน@VACT อ่าน@VACT หนังสือ@NCMN
    เขา(noun.person) นอน(verb.body) อ่าน(verb.cognition) หนังสือ(noun.communication)
    Expect output = POSTURE SVCs
    Actual Output = POSTURE SVCs

15. เขา@NPRP เอา@VACT ผ้า@NCMN ใส่@VACT ตะกร้า@CMTR
    เขา(noun.person) เอา(take) ผ้า(noun.artifact) ใส่(verb.body) ตะกร้า(noun.artifact)
    Expect output = Take SVCs
    Actual Output = Take SVCs

16. เขา@NPRP ใช้@VACT มีด@UNK หั่น@VACT ไก่
    เขา(noun.person) ใช้(use) มีด(noun.artifact) หั่น(verb.contact) ไก่(noun.animal)
    Expect output = Use SVCs
    Actual Output = Use SVCs

17. เขา@NPRP หุง@VACT ข้าว@NCMN กิน@VACT
    เขา(noun.person) หุง(verb.creation) ข้าว(noun.food) กิน(verb.change)
    Expect output = Open class SVCs
    Actual Output = Open class SVCs

18. เขา@NPRP ให้@JSBP ฉัน@NPRP อ่าน@VACT หนังสือ@NCMN
    เขา(noun.person)   ให้(give)   ฉัน(noun.person)   อ่าน(verb.cognition)
    หนังสือ(noun.communication)
    Expect output = Give SVCs
    Actual Output = Give SVCs

19. เขา@NPRP ทา@VACT เด็ก@NCMN ร้องไห้@VACT
    เขา(noun.person) ทา(make) เด็ก(noun.person) ร้องไห้(verb.body)
    Expect output = Make SVCs
    Actual Output = Make SVCs

20. เขา@NPRP ผลัก@VACT ฉัน@NPRP ล้ม@VSTA
    เขา(noun.person) ผลัก(verb.contact) ฉัน(noun.person) ล้ม(verb.social)
    Expect output = Resultative SVCs
    Actual Output = Open Class SVCs

5. CONCLUSIONS

During the development of the program for automatically classifying the serial-verb constructions by using the word functions or POS, and word classes in Thai WordNet to assist, we found that verb class form between the Open Class SVCs and Resultative SVCs is similar, so the researcher

cannot classify. And for the Causative SVCs, the researcher think that we should do additional analysis about open class verb because of the issues, such as in the form of "ทำ + open class verb" if the system process the sentence such as "เขาทำเด็กอ่านหนังสือ", the result would be Causative SVCs that is incorrect because the open class verb in Causative SVCs must be intransitive. So the additional analysis of this issue is recommended.

The author believes that Open Class SVCs, Resultative SVCs, and Causative SVCs are verbs that currently classified by using open class verb which is any verb. These verbs should be classified to one of the other verb classes in Thai WordNet in order to replacement of the original pattern, however, sufficient data of serial verb constructions sentences is essential to classify.

REFERENCES

[1]   Wanlee Sutthichatchawanwong, (2006) "A Study of the Translation of Thai Serial Verb Constructions with Directional Verbs and Their Semantic and Syntactic Equivalence in English" Retrieved June 25, 2012, from http://www.paaljapan.org/resources/proceedings/PAAL11/ pdfs/19.pdf

[2]   Thepkanjana Kingkarn, (1986) "Serial Verb Construction in Thai", Ph.D. Dissertation, the University of Michigan.

[3]   Wilawan Supriya, (1993) "A Reanalysis of so-called serial verb constructions in Thai, Khmer, Mandarin, and Yoruba", Ph.D. Dissertation, the University of Hawaii at Manoa.

[4]   Nuttanart Muansuwan, (2002) "Direction Serial Verb Constructions in Thai". Retrieved June 22, 2012, from http://www.stanford.edu/group/cslipublications/cslipublications/HPSG/1/hpsg00 muansuwan.pdf

[5]   Cholthicha Sudmuk, (2006) "The Syntax and Semantics of Serial Verb Constructions in Thai", Ph.D. Dissertation, the University of Texas at Austin.

[6]   http://ict.siit.tu.ac.th/kindml/thainest/index.php?option=com_content&view=article&id=7&Item id=10. Retrieved September 14, 2012

[7]   http://ir.swu.ac.th/xmlui/handle/123456789/956?show=full. Retrieved September 14, 2012

[8]   http://wordnet.princeton.edu/man/lexnames.5WN.html#sect4. Retrieved September 14, 2012

[9]   Ni Luh Windiari, (2012) "Indonesia Serial Verbs with Directive Verb "PERGI" and their translation into English". Retrieved September 14, 2012, from http://www.academia.edu/ 1467923/INDONESIAN_SERIAL_VERBS_WITH_DIRECTIVE_VERB_PERGI_AND_THEIR_T RANSLATION_INTO_ENGLISH

Authors

Surasa Sookgont received a BBA degree from Rajamangala University of Technology Phra Nakhon North Bangkok Campus, Thailand. She is Master student in Computer Science in Thammasat University and work as system analyst in Enterprise Business Operation, ProsperSOF Consulting.

Thepchai Supnithi received the Ph.D. degree in Electrical and Computer Engineering from Osaka University, Osaka Japan. Currently, he works as supervisor of Language and Semantic Technology Laboratory (LST), National Electronics and Computer Technology Center (NECTEC), Thailand. His research interests include Education System, Knowledge Engineering, Natural Language Processing, and Machine Learning.

Rachada Kongakchandra received the Ph.D. degree in Electrical and Computer Engineering from King Mongkut's University of Technology Thonburi, Thailand. Currently, she works as Assistant Professor at the Computer Science department, Faculty of Science and Technology, Thammasat University. Her research interests include Artificial Intelligent, Natural Language Processing, Semantic Processing, and Machine Learning.