

An Automated Thematic Role Labeler and Generalizer for Filipino Verb Arguments *

Briane Paul Samson, Bianca Pamela Alcera, Ed Oswald Go, Czarina Meg Gonzales, and
Nathalie Rose Lim

College of Computer Studies, De La Salle University - Manila,
2401 Taft Avenue, Manila 1004, National Capital Region
{bvsam2003, iamfhum, edoswald89, itsyandbitsy}@yahoo.com, nats.lim@delasalle.ph

Abstract. A lexicon is an essential resource in the Natural Language Processing research. It provides the link between the terms of a language and the semantic and syntactic properties they are associated with. For the Filipino language, only bilingual and multilingual lexicons are available electronically. Generally, the only information they contain are the translations of a term from one language to another. They do not have information on thematic roles, which are the relations of verbs and their arguments. These relations are useful because they could allow systems to check whether the required arguments are present in the sentence. To augment manual entries of the thematic roles into the lexicon, automatic learning of thematic roles of verb arguments is explored. This paper presents the resources needed, the processes, and the results.

Keywords: Thematic Roles, Lexicons, Lexicon Constructing Systems, Filipino Language, Natural Language Processing.

1 Introduction

According to Rohwer and Freitag (2004), a lexicon is an essential resource in the Natural Language Processing research area. It provides the link between the terms of a language and the semantic and syntactic properties they are associated with. It can be of use in various tasks such as information extraction, text simplification, and machine translation (Litkowski, 2005).

Ideally, lexicons contain semantic, syntactic, morphological, and phonological information. However, not all of them contain all four of these as they are designed according to the specific needs of their applications. For example, a thesaurus-like lexicon contains information such as synonyms and antonyms, while a bilingual lexicon has translations of a term from one language to another (Litkowski, 2005). Some Natural Language Processing applications, on the other hand, require more complex lexicons – those that keep information on the lexical relations of terms such as thematic roles, which are the relations of verbs and their arguments. Thematic roles are useful as they serve as cues to the senses of the terms (Gildea and Jurafsky, 2002). Furthermore, they could allow systems to check whether the required arguments are present in the sentence.

For the English language, VerbNet – the largest online verb lexicon – keeps track of thematic roles as one of its verb class descriptions (Schuler, 2005). For the Filipino language,

* The work reported in this paper was partially funded by DLSU-URCO. It was done under the supervision and guidance of the following DLSU faculty: Ms. Ethel Ong, Mr. Allan Borra, Mr. Dannel Alcantara, Dr. Ramilito Correa, and Dr. Teresita Fortunato.

most (if not all) currently existing lexicons are simply online bilingual or multilingual lexicons. They do not have essential information such as the thematic roles.

In this research, the authors explored automatic learning of thematic roles to augment manual encoding of entries. The following sections are organized as follows: Section 2 gives an overview of the Filipino sentence structure; Section 3 discusses the resources needed and the processes; Section 4 discusses the results; Section 5 gives the conclusion.

2 Filipino Sentence Structure

In the Filipino language, a sentence is referred to as *pangungusap*. Its subject is called *paksa*, while its predicate is called *panaguri*. It has two forms: *karaniwan* (common) and *di karaniwan* (uncommon). The first one, which is more commonly used by native speakers, entails the predicate to be placed ahead of the subject. It is most likely in the verb-subject-object form. The second one, on the other hand, entails the subject to be ahead of the predicate. In this form, the structural marker *ay* goes in between them (Santiago and Tiangco, 2003). It is similar to the subject-verb forms of the English language.

Shown in Table 1 are some examples of these two forms. The words that are underlined are the *paksa*, while the words that are italicized are the *panaguri*.

Table 1: Filipino Sentence Forms

	Karaniwan	Di karaniwan
Filipino Sentence	<i>Namili sa Divisoria <u>si Kei</u>.</i>	<u>Si Kei</u> ay namili sa Divisoria.
English Equivalent	<i>Shopped in Divisoria was <u>Kei</u>.</i>	<u>Kei</u> shopped in Divisoria.

The Filipino language also follows a free-word order. Thus, sentence structures in the form of object-verb-subject and others may exist. Nevertheless, a change in the position of the words will not necessarily modify the meaning of the sentence (Arndt *et al.*, 2004). A change in the position of the words, however, may alter the form of the verb. If the verb's form is altered, its focus is altered as well.

Verbs are differentiated using their focus. The focus expresses the grammatical role of the subject of the sentence through verbal affixes. These affixes are dependent on the role: actor, object or goal, location, benefactive, or instrument (Ramos and Cena, 1990).

In actor focus, the verbal affixes indicate the actor or doer of the action. In object or goal focus, they indicate the receiver of the action. In locative focus, they indicate the location or direction of the action. In benefactive focus, they indicate the beneficiary of the action. Lastly, in instrumental focus, they indicate anything that is used or acted upon to bring about the action.

Tables 2 to 6 show some examples of these affixes.

Table 2: Actor Focus Affixes

Affix	Sentence	English Translation
-um-	Gumawa ang karpentero ng upuan.	The carpenter made a chair.
	Made the carpenter a chair.	
mag- / nag-	Magpinta ka ng larawan.	(You) paint a picture.
	Paint you a picture.	
ma- / na-	Natulog si Eric sa kama.	Eric slept on the bed.
	Slept Eric on bed.	

Table 3: Object or Goal Focus Affixes

Affix	Sentence	English Translation
-in-	Pinukpok ni Jay ang pako.	The nail was hammered by Jay.
	Hammered by Jay the nail.	
-an / -han	Hinugasan ni Angela ang mga plato.	The plates were washed by Angela.
	Washed by Angela the plates.	

Table 4: Locative Focus Affixes

Affix	Sentence	English Translation
-an / -han	Binilhan ni Bea ng CD ang tindahan.	Bea bought a CD from the store.
	Bought from by Bea a CD the store.	
pag...an / pag...han	Pagsabihan mo ang kaibigan mo.	(You) reprimand your friend.
	Reprimand you the friend you.	

Table 5: Benefactive Focus Affixes

Affix	Sentence	English Translation
-an / -han	Kantahan mo si Ed.	(You) sing for Ed.
	Sing for you Ed.	
i-	Ibinili ni Jin si Pam ng kwintas.	Jin bought a necklace for Pam.
	Bought for by Jin Pam a necklace.	
ipag-	Ipagluto mo si Danica ng kanin.	(You) cook rice for Danica.
	Cook for you Danica a rice.	

Table 6: Instrumental Focus Affix

Affix	Sentence	English Translation
ipang- / ipan-	Ipanlinis mo ng sahig ang basahan.	(You) use the rag to clean the floor.
	Use to clean you the floor the rag.	

3 Automatic Learning of Thematic Roles

To automatically learn thematic roles, the authors created a Java-based system that builds a Filipino lexicon with thematic roles. They named the system fiLex. Its main features and capabilities can be summarized into the following modules: Preprocessor, Subcategorization, Thematic Role Labeling, Subcategory and Thematic Role Learning, and Lexicon Editor.

As illustrated in Figure 1, the process of fiLex starts by accepting text files of unannotated Filipino corpus. The text files are submitted to the Tokenizer, which separates and inserts the words into a vector. During the first pass, wherein the subcategorization frames and the thematic grids are learned, the output from the Tokenizer is tagged by an existing Part-of-speech Tagger (Ciego *et al.*, 2007). The resulting POS-tagged corpus is then passed on to an existing Morphological Analyzer (Aquino *et al.*, 2007) which identifies, extracts, and annotates the root word and affix/es of the words.

During the second pass, wherein the saved annotations and the annotations made on the corpus are tested for its correctness, the output from the Tokenizer is tagged in the Corpus Annotation Module. The annotations used in this module are all obtained from the lexicon of the fiLex system.

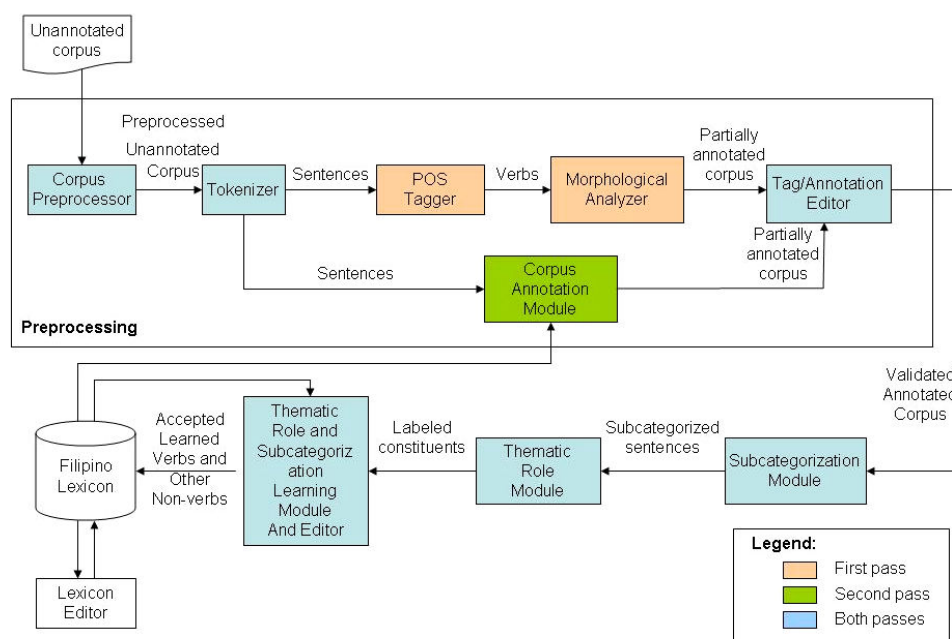


Figure 1: Architectural Design

However, not all automatically generated annotations from either pass are correct and valid.¹ Thus, the output will be passed to the Tag/Annotation Editor which enables Filipino linguists to manually edit the corpus and validate the annotations in the corpus. The Filipino linguists can also add or remove annotations using this facility.

The validated annotated corpus is then passed to the Subcategorization Module, which determines the boundaries of the verbs' arguments by inducing the constituent structure of the sentence (Alcantara, 2008).

Next in the process is the Thematic Role Labeling. Here, the verbs' arguments are tagged with their corresponding thematic roles. The annotated corpus is then passed to the Thematic Role and Subcategorization Learning Module, wherein the verbs' thematic role assignments, structures and arguments are defined and generalized among all samples. In this module as well, Filipino linguists can validate the definitions and generalizations before storing the entries to the Filipino lexicon.

After processing, the words are inserted in the database producing the Filipino Lexicon. Moreover, the linguist can use the Lexicon Editor to edit and validate the entries in the lexicon.

3.1 Resources

Corpus: The fiLex system will take in electronic newspaper articles as input. These articles do not include those in the entertainment section as they tend to be written more informally and have excessive dialogues. The input articles are written in the Filipino language, and are assumed to have neither orthographical nor grammatical errors. Moreover, these sentences are manually checked and transformed by the proponents to be simple, declarative, and of the verb-subject-object (VSO) or verb-object-subject (VOS) form.

Lexicon: If a verb already exists in the lexicon that the fiLex system built, the verb entry will be used during the Subcategory and Thematic Role Learning module for generalizing purposes. It will also be used for automatic annotation during the testing phase.

¹ MAG-Tagalog used 16,540 Tagalog words for testing. 83.84% of them were accurately analyzed.

The lexicon includes a list of Filipino words and their constructs. The verbs in the lexicon contain additional information like subcategorization frame, thematic grid and inflections. The inflections are placed in another table, and they are mapped only to their corresponding root word in the lexicon.

3.2 Thematic Roles

Thematic roles are labels that describe the semantic relations between a verb and its arguments. It marks the role played by the argument with respect to the predicate (Saint-Dizier, 2001).

Table 7 shows the list of thematic roles used by fiLex. Their respective tags and descriptions are also included. fiLex only used these thematic roles because of the limited semantic information that can be gathered.

Table 7: Thematic Roles Used by fiLex

Tag	Thematic Role	Description
AGTR	Agent	the entity that intentionally initiates, makes or originates the action described by the predicate
THMR	Theme	the entity affected by the action or state expressed by the predicate
BENR	Benefactive	the entity that benefits from the action expressed by the predicate
GOAR	Goal	the entity towards which the activity expressed by the predicate is directed
LOCR	Location	the place wherein the activity expressed by the predicate is situated
INSR	Instrument	the entity that is used to do the event

In representing thematic roles, thematic grids are traditionally used. In a thematic grid, each subcategory inside the SUBCAT frame is assigned with a thematic role inside the ROLES frame. There is a one-to-one correspondence between a subcategory and a thematic role. Shown below is an example of this:

binigay: [SUBCAT: <NP₁, NP₂, NP₃>, ROLES<AGTR, THMR, GOAR>]

In the example given, the verb *binigay* (gave) assigned the roles Agent, Theme, and Goal to NP₁, NP₂, and NP₃ respectively.

3.3 Modules

Preprocessor Module: This module prepares the unannotated corpus for analysis. Shown in Table 8 is an example sentence from an unannotated corpus.

Table 8: Example Sentence from an Unannotated Corpus

Filipino Sentence	Binato ni Briane si Meg ng bola.
English Glosses	Threw by Briane Meg a ball.
English Translation	Briane threw a ball at Meg.

The module tokenizes the corpus, and inserts the tokenized words into a vector:

Binato	ni	Briane	si	Meg	ng	bola
--------	----	--------	----	-----	----	------

It then invokes an existing Part-of-speech Tagger and Morphological Analyzer to produce a partially annotated corpus:

Binato <i>VB</i> Stem: <i>bato</i> Affix: <i>in</i> Completed Object-focus	<i>ni</i> <i>DTP</i>	Briane <i>NNP</i>	<i>si</i> <i>DTP</i>	Meg <i>NNP</i>	<i>ng</i> <i>DTC</i>	bola <i>NNC</i>
--	-------------------------	----------------------	-------------------------	-------------------	-------------------------	--------------------

The module also enables Filipino linguists to manually validate the output. Thus, the result of this module will be a validated annotated corpus.

Subcategorization Module: This module determines the boundaries of a verb’s arguments. It receives the vector of words, and processes them per sentence.

It identifies the argument boundaries using PACSI (Alcantara, 2008) to induce the constituent structure of the sentence:

[Binato [ni Briane] [si Meg] [ng bola]] .

The module then assigns a type to each bounded argument: VP for verb phrase, NP for noun phrase, AP for adverb phrase, and ADP for adjective phrase.

[Binato [ni Briane] [si Meg] [ng bola]] .
VP NP NP NP

Thematic Role Labeling Module: This module tags the arguments with their respective thematic roles. Depending on the focus of the verb and the noun marker or pronoun of each argument, a thematic role is assigned.

Table 9: Thematic Role Assignment

Verb Focus	Marker/ Group	Thematic Role	Verb Focus	Marker/ Group	Thematic Role
Actor	NM1	Agent	Locative	NM1	Location
	PN1	Agent		PN1	Location
	<i>ni</i> or <i>nina</i>	Theme		<i>ni</i> or <i>nina</i>	Agent
	<i>ng</i>	Theme		PN2	Agent
	PN2	Theme		first <i>ng</i>	Agent
	NM2 or PN3	Location		“ng”	Theme
	<i>para</i>	Benefactor		NM1	Benefactor
Instrumental	NM1	Instrument	Benefactive	PN1	Benefactor
	PN1	Instrument		<i>ni</i> or <i>nina</i>	Agent
	<i>ni</i>	Agent		PN2	Agent
	PN2	Agent		first <i>ng</i>	Agent
	first <i>ng</i>	Agent		<i>ng</i>	Theme
	<i>ng</i>	Theme		NM2 or PN3	Location
	NM2 of PN3	Goal		only NM2 or PN3 in the sentence	Goal

The approach used by the authors is to identify the focus of the verb and use the markers or pronouns to determine the thematic role. This is based from the verb focus-related rules in the book written by Ramos and Cena (1990). Markers are grouped into NM1 and NM2, while

pronouns are grouped into PN1, PN2, and PN3. For reference, NM1 denotes the following noun markers: *ang, si, and sina*. NM2 denotes the following noun markers: *sa, kay, and kina*. PN1 denotes the following pronouns: *ako, ikaw, ka, kita, kami, kayo, tayo, siya, sila, ito, iyan, iyon*. PN2 denotes the following pronouns: *ko, mo, namin, natin, ninyo, niya, nila, nito, niyan, and noon*. And PN3 denotes the following pronouns: *dito, diyan, and doon*.

Table 9 shows an excerpt of the list of assigned thematic roles to the argument, based on the given verb focus and the marker preceding the argument. Additional rules are applied on some specific cases of markers.

Subcategory and Thematic Role Learning Module: This module learns the thematic grids and subcategorization frames of a particular verb to produce a more generalized thematic grid and subcategorization frame for it.

First, fiLex defines the verb's thematic role assignments and argument structure. Then, it generalizes them by combining all defined thematic grids and subcategorization frames. Here is an example of a generalized subcategorization frame and thematic grid for the verb *bato* using 5 sample sentences having the same root form of the verb and focus:

Generalization, G1:

bato | [SUBCAT: <NP₁, NP₂, NP₃, NP₄, PP₅, PP₆>, ROLES<AGTR, PATR, THMR, INSR, GOAR, LOCR>]

In generalizing, redundant subcategories and their thematic roles will be copied into G1. In the example, NP₁, NP₂ and NP₃ with their thematic roles Agent, Patient and Theme or Instrument respectively are the redundant subcategories and are just copied into G1. In case there are multiple roles assigned to an argument among the sample sentences, all those roles are added in the generalization also. The remaining subcategories, PP₅ and PP₆, which do not co-occur with any other subcategory, are also added. For the order of subcategories in the generalization, a sentence structure for each verb focus would be used as basis.

Table 10: Cooccurrence Table Entry for *bato*

learnedID	subcat	themRole	cooccurrence
1	NP ₁	Agent, AGTR	5
1	NP ₂	Patient, PATR	2
1	NP ₃	Theme, THMR	3
1	NP ₄	Instrument, INSR	2
1	PP ₅	Goal, GOAR	1
1	PP ₆	Location, LOCR	1

In determining the importance of a subcategory, a threshold is maintained to serve as basis. In our example, the threshold would be 3 which is the ceiling of N/2 where N is the number of sample sentences used to derive the generalizations.

Based on Table 10, the subcategory NP₁ with the thematic role Agent has the most number of co-occurrences which makes it the most important subcategory or argument of *bato*. The next important subcategory would be NP₃. It co-occurs with the verb *bato* 3 times. However, the argument NP₄ will not be given an obligatory importance because it only co-occurred with

bato twice. Lastly, the subcategories PP₅ and PP₆ would be the least important among all of them because they both have one co-occurrence with the verb *bato*.

After determining which subcategories and thematic roles are obligatory and optional, additional information would be added into the thematic grid and subcategory frame. A 1 would be put beside a thematic role if it is an obligatory role and 0 if not obligatory or optional.

New Generalization with Importance, G1’:

bato | [SUBCAT: <NP₁, NP₂, NP₃, NP₄, PP₅, PP₆>, ROLES<AGTR:1, PATR:0, THMR:1, INSR:0, GOAR:0, LOCR:0>]

If an entry in the lexicon already exists for *bato*, the learning module will generalize a new thematic grid and subcategory frame to accommodate a previous generalization. Let us say that there is already an entry for *bato* and it has the following thematic grid and subcategory frame:

Lexicon entry for *bato*:

bato | [SUBCAT: <NP₁, NP₂, NP₃>, ROLES<AGTR:1, PATR:1, THMR:0>]

Latest Generalization:

bato | [SUBCAT: <NP₁, NP₂, NP₃, NP₄, PP₅, PP₆>, ROLES<AGTR:1, PATR:1, THMR:1, INSR:0, GOAR:0, LOCR:0>]

In the latest generalization, all similar subcategories and their thematic roles with their importance will be copied. Subcategories NP₁, NP₂ and NP₄ were just copied along with their thematic roles. However, subcategory NP₃'s importance, which is assigned to Theme, was altered from 0 to 1. All remaining subcategories and thematic roles which are not in the previous entry or in G1' will be copied into the latest generalization.

The module allows for the displaying of sample sentences used to derive the verb's thematic role assignment and argument structure. Furthermore, it also enables Filipino linguists to validate and edit the definitions and generalizations of a verb.

Lexicon Editor: This module displays the entries of the lexicon, and enables the Filipino linguists to modify them. It also allows the Filipino linguists to add more word annotations like gender.

4 Results

Four testing methods were employed by the proponents to thoroughly test the capabilities of fiLex. In the Preprocessor Module test, a downward trend in both precision and recall is present in all test corpora. After the first pass, the accuracy of the annotation in the second pass declined. This was due to the numerous entries in the lexicon a word or verb can have after the acquisition of annotations in the first pass.

In the Thematic Role and Subcategorization Learning Module test, for the accuracy of the labeler, the system was able to correctly label phrase types with or without the presence of incorrect tags. Another finding is that the system can label the thematic roles to arguments but not all arguments have a valid or correct role assigned to it. It may label a null role to an argument. Another interesting finding would be that the system cannot proceed whenever the text file or corpus fed through it has at least one sentence without a bracketed argument. For the accuracy of the generalizer, the system able to include all possible argument-role pairs in the generalized subcategorization frames and thematic grids despite the presence of null roles. The fiLex system was also able to tag the correct importance of the role.

In the Non-VSO and non-VOS test, the corpus went through all the modules smoothly and no errors were detected even if the corpus has non-VSO form sentences. The system managed to skip the sentences in non-VSO form and continue with the processing of the sentence in

VSO and VOS form. Generalizations were produced for sentences with VSO and VOS structures.

In the Benchmark test, two online resources were used, namely VerbNet and PropBank since all verbs, whether it is in Filipino or in English, has more or less the same meaning. 300 verbs were then randomly selected from the fiLex system lexicon and their English translations were used to find entries in both VerbNet and PropBank. Four criteria were used in evaluating verbs in the VerbNet Benchmark test while only two were used for the PropBank Benchmark test. Each verb in the VerbNet Benchmark test had 4 scores while each verb in the PropBank Benchmark test had 2 scores.

Table 11: Tally of scores for each criterion under VerbNet Benchmark testing

Criterion	$0 \leq x \leq 25$	$25 < x \leq 50$	$50 < x \leq 75$	$75 < x \leq 100$
Thematic Grid	55	134	62	11
Subcategorization Frame	3	88	139	32
Pairing	74	152	30	6
Position	110	124	23	5

Shown in Table 11 are the numbers of verbs in each score bracket for the VerbNet Benchmark test. The Thematic Grid criterion was the percentage of thematic roles appearing in both thematic grids. The Subcategorization Frame criterion was the percentage of subcategories or arguments present in both subcategorization frames. The Pairing criterion was the percentage of argument-role pairs present in both. Lastly, the Position criterion was the percentage of argument-role pairs appearing in the same position for the subcategorization frames and thematic grids of lexicon entries and VerbNet entries.

Table 12: Tally of Scores for each Criterion under PropBank Benchmark Testing

Criterion	$0 \leq x \leq 25$	$25 < x \leq 50$	$50 < x \leq 75$	$75 < x \leq 100$
Thematic Grid	1	6	1	8
Position	6	5	1	4

Now, for the PropBank Benchmark testing with only 16 verb entries included, most of them in Table 12 fall in the $25 < x \leq 50$ and $75 < x \leq 100$ brackets for the Thematic Grid criteria. The Thematic Grid criterion was the percentage of thematic roles appearing in both thematic grids. The Position criterion was the percentage of argument-role pairs appearing in the same position for the subcategorization frames and thematic grids of lexicon entries and PropBank entries.

5 Conclusion

After going through the different tests, the proponents have concluded that the system is able to generalize accurately. It was also able to annotate the corpus accurately. But for both the annotations and the generalizations to be accurate, the corpus must first be clean and valid. And for the generalizations to be more accurate, the annotations must be accurate. However, when the generalizations of the system are compared to the subcategorization frames and thematic grids of VerbNet and PropBank, the results are not close to being acceptable because it showed very low scores in the different criteria provided by the proponents. The system was also able to proceed with the processing of the corpus despite the existence of non-VSO sentences.

Currently, the proponents were able to gather corpora of roughly 100,000 words from online news sites. Corpora of 44,130 words from the gathered ones were annotated though not all of the annotations were validated. The rest of the roughly 100,000 gathered corpora were not annotated anymore. In the fiLex lexicon, there are already 3853 entries. 938 of those entries are verbs. Almost 300 of those were manually encoded while the rest were acquired by the system after the first pass. Not all of the verb entries in the lexicon have generalized subcategorization frames and thematic grids yet.

Lastly, a new algorithm was developed to generalize subcategorization frames and thematic grids based on the sentence structures for each focus type. And for the thematic role labeling, the proponents developed a new algorithm which used the focus type of the verb and the markers in each argument as basis. The technique designed and implemented can be adapted to other languages that use grammatical markers as cues to determine argument types, even if these languages do not follow the same sentence structure as Filipino. For example, the Filipino sentence “Binigyan ni John si Mary ng mansanas”, the marker *ng* specifies the object *mansanas*. In another Philippine language like Cebuano (“Hinatagan ni John si Mary ng mansanas”), the marker *ng* also specifies the object *mansanas*. Also, in Nihongo, the translated sentence “John-san wa Mary-san ni ringo wo agemashita” has the particle *wo* which specifies the object *ringo*, the translation of *mansanas*.

References

- Alcantara, D. 2008. *Probabilistic Approach to Constituent Structure Induction for Filipino*. Unpublished master’s thesis, De La Salle University – Manila.
- Arndt, K., M. Stebner and M. Stebner. 2004. *Understanding the Language*. Retrieved April 9, 2009 from http://www.lerc.educ.ubc.ca/LERC/courses/489/world-lang/tagalog_ind/Tagalog2/description.htm#grammer
- Aquino, M., E. Fernandez and K. Villanueva. 2007. *MAG-Tagalog: A Rule-Based Tagalog Morphological Analyzer and Generator*. Unpublished undergraduate thesis, De La Salle University – Manila.
- Ciego, R., Z. Huang, G. Navarro and M. Torres. 2007. *HATPOST: A Hybrid Approach to Tagalog Part of Speech Tagging*. Unpublished undergraduate thesis, De La Salle University – Manila.
- Gildea, D. and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3), 245-288.
- Litkowski, K. 2005. Computational Lexicons and Dictionaries. In *Encyclopedia of Language and Linguistics, 2nd edition*. Retrieved August 6, 2008 from <http://www.clres.com/online-papers/ell.doc>
- Ramos, T. and R. Cena. 1990. *Modern Tagalog: Grammatical Explanations and Exercises for Non-native Speakers*. USA: University of Hawaii Press.
- Rohwer, R. and D. Freitag. 2004. Towards Full Automation of Lexicon Construction. *HLT-NAACL 2004: Workshop on Computational Lexical Semantics*, pp.9-16.
- Saint-Dizier, P. 2001. *An Introduction to Lexical Semantics*. France: Universite Paul Sabatier.
- Santiago, A. and N. Tiangco. 2003. *Makabagong Balarilang Filipino*. Quezon City, Philippines: Rex Printing Company, Inc.
- Schuler, K. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.