Integrating Prosodics into a Language Model for Spoken Language Understanding of Thai*

Siripong Potisuk

Department of Electrical and Computer Engineering
The Citadel, the Military College of South Carolina
171 Moultrie Street
Charleston, South Carolina 29409 USA
siripong.potisuk@citadel.edu

Abstract. This paper describes a preliminary work on prosody modeling aspect of a spoken language understanding system for Thai. Specifically, the model is designed to integrate prosodics into a language model based on constraint dependency grammar. There are two steps involved, namely the prosodic annotation process and the prosodic disambiguation process. The annotation process uses prosodic information (stress, pause and syllable duration) obtained from the speech signal during low-level acoustic processing to encode each word in the parse graph with a prosodic feature called strength dynamic. The goal of the annotation process is to capture the correspondence between the phonological and phonetic attributes of the prosodic structure of utterances and their intended meanings. The prosodic disambiguation process involves the propagation of prosodic constraints designed to eliminate syntactic ambiguities of the input utterance. It is considered a pruning mechanism for the parser by using prosodic information. The aforementioned process has been tested on a set of ambiguous sentences, which represents various structural ambiguities involving five types of compounds in Thai.

Keywords: Spoken language understanding, Prosody, Thai.

1. Introduction

In the realm of artificial intelligence, there is no denying that the ultimate goal for man-machine communication is to use natural human languages. Humans aspire to build a machine capable of understanding and responding to commands issued in the form of spoken language. Research towards this goal encompasses various research disciplines, and the fruits of such labor include machine translation, speech synthesis and recognition, and spoken language understanding.

Spoken language understanding has been a challenge for scientists and engineers for many decades. Early efforts were primarily aimed at simply recognizing speech, and extensive research has been conducted to improve performance of speech recognition systems. Much of the improvements can be attributed to combining various research disciplines including linguistics, psychology, computer science, and signal processing, as well as improvements in modeling techniques, such as statistical and symbolic pattern recognition. Speech models now incorporate linguistic modeling to account for coarticulation and grammatical constraints to reduce the search

* The author would like to thank the Citadel Foundation for its financial support in the form of a presentation grant.

space for the correct utterance. Also, steady advances in the computational efficiency of computers and hardware have enabled researchers to implement computationally intensive models that were deemed nearly impossible in the past.

With the remarkable advances in speech recognition, the focus has now shifted toward the development of spoken language understanding systems. The goal is to build a machine capable of understanding naturally spoken language by combining together speech recognition and natural language processing technologies. Such systems must be capable of recognizing a large number of words (on the order of tens of thousand), and they also need to make better use of additional information present in the speech signal, such as prosody. It is well known that prosodic information, such as pause, stress, intonation, etc., facilitates human speech communication by helping disambiguate utterances.

Prosody is an important aspect of speech that needs to be fully explored and utilized as a disambiguating knowledge source in spoken language understanding systems. Statistical modeling of prosodic information has the potential to be used as an additional knowledge source for both low-level processing (recognizing) and high-level processing (parsing) of spoken sentences. A knowledge source may be thought of as a module in spoken language understanding system, which contributes to the understanding of a spoken sentence.

Low-level use of prosodic information potentially allows for more accurate recognition at the phonetic and phonological levels. It may also prove to be useful at the stage of lexical access during word hypothesization. The most relevant prosodic information at this level is intra-word and syllable-level stress. A lexicon containing word stress patterns may help the recognizer discard word hypotheses with poorly matched stress patterns. A syllable is considered stressed if it is pronounced more prominently than adjacent syllables. Stressed syllables are usually louder, longer, and higher in pitch.

High-level use of prosodic information can help resolve ambiguities inherent in natural language independent of contextual information since computers do not currently utilize all the knowledge sources that humans do. Relevant prosodic information at this level includes pauses, sentential stress, and intonation contours. Sentential or inter-word stress information can be used in speech recognizers to resolve lexical and syntactic ambiguities. For example, the degree of stress among words in a spoken sentence provides an acoustic cue for distinguishing between content (noun, verb) and function (auxiliary, preposition, etc.) words. Furthermore, the marking of prosodic phrases may improve parsing performance by reducing syntactic ambiguities since prosodic groupings may rule out some of the syntactic groupings for a syntactically ambiguous sentence. Also, the identification of sentence mood (i.e., declarative, interrogative, command, etc.) from intonation contours may reduce syntactic and pragmatic ambiguity.

The overall objective of this research is to study the role of prosody in the implementation of a spoken language understanding system for Thai since prosodic information has the potential to be used as a pruning mechanism at both the low and high levels of spoken language processing. In particular, we specifically examine how salient prosodic features of Thai (e.g., stress) can be integrated with the overall language modeling scheme. Thai is the official language of Thailand, a country in the Southeast Asia region. The language is spoken by approximately 65 million people throughout different parts of the country. The written form is used in school and all official forms of communication.

Language modeling is one of the many important aspects in natural (both written and spoken) language processing by computer. For example, in a spoken language understanding system, a good language model not only improves the accuracy of low-level acoustic models of speech, but also reduces task perplexity (the average number of choices at any decision point) by making better use of high-level knowledge sources including prosodic, syntactic, semantic, and pragmatic knowledge

sources. A language model often consists of a grammar written using some formalism which is applied to a sentence by utilizing some sort of parsing algorithm.

To overcome the difficulties in parsing Thai, we believe that a constraint dependency grammar (CDG) parser proposed by Potisuk (1996) appears to be an attractive choice for analyzing Thai sentences, considering vantage points from both written and spoken language processing aspects of an automatic system. CDG parsers rule out ungrammatical sentences by propagating constraints. Constraints are developed based on a dependency-based representation of syntax. The motivation for our choice of dependency grammar, instead of phrase-structure grammar, stems from the fact that it appears that Thai syntax might be better described by the former representation.

Our work in this paper extends the adopted CDG parsing approach by incorporating prosodic constraints into a grammar for Thai. Prosodic information is specifically used as part of a disambiguation process. Disambiguation is accomplished through the use of prosodic constraints which identify the strength of association between prosodic and syntactic structures of the sentence hypotheses. We next describe the proposed basic framework for a Thai spoken language understanding system and detail the necessary steps for achieving the goal of integrating prosodic information with other aspects of the language model.

2. A Conceptual Model of a Thai Spoken Language Understanding System

The elusive goal of building a machine capable of understanding naturally spoken language that covers a very large vocabulary has not yet been realized. Presently, state-of-the-art spoken language interfaces merely recognize, rather than understand speech, and such systems only achieve high recognition on tasks that have low perplexity. To achieve understanding, we need to improve performance of speech recognition systems as well as combine speech recognition with natural language processing technology.

A spoken language understanding system may generally be thought of as comprising low-level (recognizing) and high-level (parsing) processing components although the detailed configuration may vary from system to system. The low-level processing usually involves acoustic modeling of the speech signal for recognition purposes. At the high level of processing, high-level knowledge sources (such as prosody, syntax, semantics, and pragmatics) are utilized not only for improving recognition performance, but also for analyzing the structure of the sentence hypotheses to obtain the best parse to map to a semantic representation in order to achieve understanding. This is commonly known as the language modeling aspect of the system.

Spoken language understanding should be viewed as a computer-human interaction problem. Understanding how various cues contribute to system performance in the context of spoken language interfaces to task-oriented mixed-initiative systems is crucial in the design of a language model. Such systems are best evaluated and judged in terms of their success in supporting users in accomplishing tasks.

A language model usually consists of a grammar which is applied to a sentence by utilizing a parsing algorithm to account for syntactic representation of the recognized string of words. In addition to syntactic parsing, the language model must be designed to be capable of using other high-level knowledge sources. The goal should be to not only improve the accuracy of low-level acoustic models of speech, but also reduce task perplexity. The choice of the approach to language modeling depends on how easily the model can be integrated with the acoustic model. Of primary concern are the issues of separability, scalability, computational tractability, and inter-module communication. These issues specify the level of interaction between the speech component and the language model, which can be classified into three categories: tightly-coupled, semi-coupled, or loosely-coupled systems.

In this research, we choose a constraint-based system of integration, which can be classified as a loosely-coupled system of integration. It utilizes the language model, which is based on a CDG

parsing algorithm, as a post-processing step to the acoustic model. The high-level knowledge sources are isolated into relatively independent modules which communicate with one another through the use of a uniform framework of constraint propagation. Figure 1 illustrates a conceptual model of a Thai spoken language understanding system based on a constraint-based system of integration.

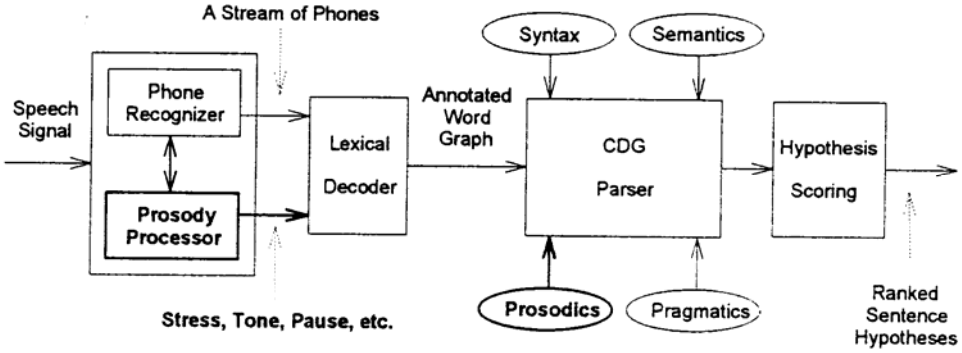


Figure 1: A Conceptual Model of a Thai Spoken Language Understanding System.

It is noted that the prosody processor has been designed to extract the tone feature from the input speech signal. Tone recognition is difficult to incorporate into a phone recognizer (usually HMM-based model), because tone is a property of a syllable, not of an individual phone. As a result, a lexical decoder or word hypothesizer is added to the system as a separate module in order to combine the output of the phone recognizer and that of the prosody processor together.

Because the overall system is loosely-coupled and the language model is based on a constraint dependency parsing algorithm, this approach is a very attractive choice for Thai. The first advantage is that the parser uses a word graph (a directed acyclic graph) augmented with parse-related information. For spoken Thai sentences, a word graph provides a very compact and less-redundant data structure for simultaneously parsing multiple sentence hypotheses generated by a lexical decoder. Secondly, since the knowledge sources are independent, their individual impact is more easily measured. The individual modules can be tested in a stand-alone fashion. Furthermore, the modularity potentially makes the system easier to understand, design, debug, and scale up to larger problems. In addition, the system becomes more versatile since they can easily accommodate more than one task or language by simply replacing individual modules. Thirdly, the system appears to be more computationally tractable because it selectively uses level-appropriate information at every stage of processing. For example, no acoustic decisions are made in the syntactic module.

In the above model, prosodic modeling involves the development of two components (highlighted in bold): the prosody processor and the prosodic knowledge source for CDG parsing. The first module is a part of the acoustic model that deals with automatic detection and classification of the prosodic cues of interest. The second module is a part of the language model that deals with representing and utilizing automatically-detected prosodic cues from the prosody processor to improve the accuracy of the parser by means of prosodic disambiguation. The information is transferred to the parser by annotating the word graph, the central processing structure of the parser. The process of prosodic disambiguation (i.e., choosing a parse or parses with the most likely prosodic patterns from among several candidate parses) is accomplished by propagating prosodic constraints. Prosodic constraints check the agreement between the annotated prosodic structure from the prosody processor and the prosodic structure of each of the competing

sentence hypotheses of the ambiguous utterance, which is generated indirectly from its syntactic structure. A sentence hypothesis is rejected if its prosodic structure is poorly matched by the annotated prosodic structure. This paper deals with the second module only, and in the following sections the use of prosodic information for resolving structural ambiguities in a Thai spoken language understanding system will be described. Two specific issues will be addressed: the prosodic annotation process and the prosodic disambiguation process.

3. CDG Parsing with Prosodic Constraints

Due to the scope of the paper, a description of the basics of CDG parsing of Thai will be briefly described with a parsing example. Interested readers are referred to the paper by Potisuk (1996) for a complete discussion of the basic CDG framework.

3.1. A description of CDG parsing

CDG is defined as a 4-tuple, $G = <\Sigma, R, L, C>$, where Σ represents a finite set of terminal symbols or lexical categories, $L =$ a finite set of labels, {l_1,……., l_q}, $R =$ a finite set of uniquely named roles (or role-ids), {r_1,……., r_p}, and $C =$ a constraint set of unary and binary constraints. Within this grammar, a sentence $s = w_1 w_2 w_3 w_4 .....w_n$ is defined as a word string of finite length *n*. The elements of Σ are the parts of speech of the words in a sentence. Associated with every word *i* in a sentence *s* are all of the *p* roles in *R*. Hence, every sentence contains *n*p* roles. A role can be thought of as a variable which is assigned a role value. A role value is a tuple <*l*, *m*>, where *l* is a label from *L* and *m* is an element of the set of modifiees, {1,2,…….,*n*, nil}. A modifiee is a pointer to another word in the sentence (or to no word if nil). Role values will be denoted in the examples as label-modifiee.

Two types of roles (role-ids) per word are used: *governor* and *need* roles. The governor role indicates the function a word fills in a sentence when it is governed by its head word (e.g., a subject is governed by the main verb). Several need roles (i.e., need1 and need2) may be used to make certain that a head word has all of the constituents it needs to be complete. Each of the need roles keeps track of an item that its word needs in order to be complete (e.g., a verb generally needs a subject for the sentence to be complete).

The function that a word plays within the sentence is indicated by assigning a role value to a role. The label in the role value indicates the function the word fills when it is pointing at the word indexed by its modifiee (e.g., a *subj* label). When a role value is assigned to the governor role of a word, it indicates the function of that word when it modifies its head word. Likewise, when a role value is assigned to a need role of a word, it indicates how that need is being filled.

To parse a sentence using CDG, the constraints (members of *C*) must be specified. A constraint set *C* is a logical formula of the form: (and $C_1 C_2 \ldots C_t$). Each C_i is a constraint represented in the form: (if *Antecedent Consequent*), where *Antecedent* and *Consequent* are either single predicates or a group of predicates joined by the logical connectives (a conjunction or disjunction of predicates). The possible components of each C_i are variables, constants, access functions, predicate symbols and logical connectives. Variables (i.e., x, y, etc) used in the constraints stand for the role values. A constraint involving only one variable is called a unary constraint; two variables, a binary constraint. A maximum of two variables in a constraint allows for sufficient expressivity. That is, using predicate symbols and access functions, unary and binary constraints for a CDG grammar can be constructed. Constants are elements and subsets of $\Sigma \cup R \cup L \cup$ {nil,1,2,……,*n*}, where *n* is the number of words in a sentence. Allowable access functions are *pos* – word position, *rid* – role-id, *lab* – label, *mod* – modifiee position, and *cat* – lexical category. The predicate symbols allowable in

constraints are: *eq* $(x = y)$, *gt* $(x > y)$, *lt* $(x < y)$, and *elt* $(x \in y)$. Lastly, the logical connectives are $\wedge, \vee$, and $\sim$.

Using the predicate symbols and access functions mentioned, unary and binary constraints for a CDG grammar can be constructed. Unary constraints are often used to restrict the role values allowed by a role given its part of speech whereas binary constraints are constructed to describe how the role values assigned to two different roles are interrelated. Examples of unary and binary constraints are given below.

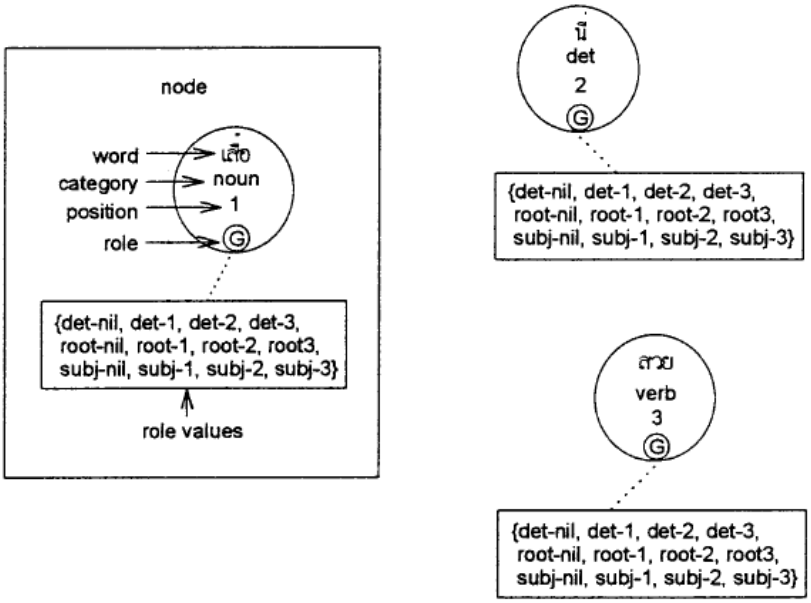
In CDG, a sentence *s* is said to be generated by the grammar if there exists an assignment **A** given a set of constraints *C*. An assignment **A** for the sentence *s* is a function that maps role values to each of the *n*p* roles such that the constraint set *C* is satisfied. There may be more than one assignment which satisfies *C*, in which case there is more than one parse for the sentence.

To illustrate the use of CDG grammars, consider the following simple example grammar *G* used for parsing a Thai sentence, เสื้อนี้สวย.

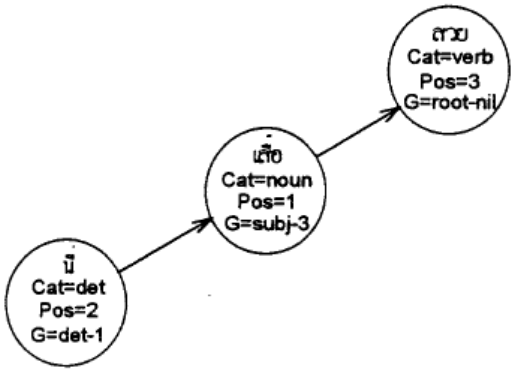
```
Σ = {det, noun, verb}
R = {governor}
L = {det, root, subj}
C =  ∀x, y  (∧
      ;; [U-1] A noun receives the label 'subj' and modifies a word to
         its right.
      (if  (eq (cat  (pos x))  noun)
           (∧ (eq  (lab x)  subj)
              (lt  (pos x)  (mod x))))
      ;; [U-2] A determiner receives the label 'det' and modifies a word
         to its left.
      (if  (eq (cat  (pos x))  det)
           (∧ (eq  (lab x)  det)
              (lt  (mod x)  (pos x))))
      ;; [U-3] A verb receives the label 'root' and modifies no word.
      (if  (eq (cat  (pos x))  verb)
           (∧ (eq  (lab x)  root)
              (eq  (mod x)  nil)))
      ;; [B-1] A subj is governed by a verb.
      (if  (∧ (eq  (lab x)  subj)
              (eq (mod x) (pos y)))
           (eq (cat  (pos y)) verb)))
```

The constraints in this grammar were chosen for simplicity, not to exemplify constraints for a wide coverage grammar. Note that U-1, U-2, U-3 are unary constraints while B-1 is a binary constraint. Also, each word is assumed in this example to have a single lexical category, which is determined by dictionary look-up. For *G* above to generate the sentence, there must be an assignment of a role value to the governor role of each word, and that assignment must simultaneously satisfy each of the constraints in *C*. Figure 2 depicts the initialization of roles for the given sentence and its assignment which satisfies *C* in the form a final parse graph for the constraint network. Details of the constraint network after the propagation of each constraint are omitted for the sake of brevity.

After all the constraints are propagated across the constraint network and filtering is completed, the network provides a compact representation of all possible parses. Syntactic ambiguity is easy to



(a)



(b)

Figure 2: (a) Initialization of roles and (b) the final parse graph for the sentence เสื้อนี่สวย.

spot in the network. If multiple parses exist, then additional constraints, such as semantic constraints, can be propagated to further refine the analysis to the intended meaning of the input sentence. The resulting parse trees are then ready to be prosodically annotated. The annotation process is described next.

3.2. Prosodic Annotation

Prosodic annotation or encoding is a mechanism for passing up prosodic information from low-level acoustic module for high-level processing. It provides to the parser relevant information that adequately captures the essence of the prosodic structure of the input utterance. Prosodic encoding usually involves the process of labeling prosodic patterns in the speech signal. The labeling criteria provide a mechanism for mapping a sequence of acoustic correlates of prosody into abstract prosodic labels. As a result, prosodically-labeled sentences contain information concerning the correspondence between the phonological and phonetic attributes of the prosodic structure of utterances and their intended meanings. Prosodic labels should be chosen to represent abstract

linguistic categories of prosody, such as rhythmic groupings (or phrasing) and prominence. Also, they should be chosen such that they are used consistently within and across human labelers, and they make the automatic labeling process tractable and consistent.

Following the annotation process proposed by Potisuk (2007) for a text-to-speech system for Thai, a brief description of the annotation process will be described. The encoding of the prosodic structure is accomplished by annotating each word candidate in the word graph with a prosodic feature called *strength*. The strength feature is chosen based on the dependency representation of syntax. There are four levels of strength dynamics: strong dependence (SD), dependence (DE), independence (ID), and strong independence (SI). SD describes a strength dynamic at the word boundary within a clitic group, within a compound, between a content and a function word, or between two function words that are interdependent (i.e., both depend on the same governor). DE describes strength dynamic at minor phrase boundaries, i.e., between a subject noun phrase and a verb phrase, between a verb and an object noun phrase, or between two content words. ID describes strength dynamic at major phrase boundaries (intonational phrases). And, SI describes strength dynamic at the sentence boundary. In addition to the strength feature, a word at the end of a phrase or an utterance will receive the feature '*final*' to indicate that it is affected by the final lengthening effect. Final lengthening is always accompanied by a pause. A word with a '*final*' feature also automatically receives a strength dynamic of ID or SI.

The criteria for labeling a parse with the above strength features using the acoustic information are now described. The criteria establish the correspondence between the phonological (strength dynamics) and the phonetic (acoustic correlates) attributes of prosody by minimizing speech disrhythmy while maintaining the congruency with syntax. Since Thai is a stress-timed language, a phonological unit called foot can be used to describe rhythmic groupings within an utterance. A foot is neither a grammatical nor a lexical unit. The domain of a foot extends from a salient (stressed) syllable up to but not including the next salient syllable. A pause is considered a salient syllable, and the beginning of an utterance is always preceded by a pause. The acoustic realization of a rhythmic foot assuming that each rhythmic foot is arbitrarily three units long, regardless of the number of syllables comprising the foot (one through five) can be summarized as follows:

| S |	→	| 3 |	→ | 2 |	if the foot is in an utterance-initial position.
| S |	→	| 3 |	→ | 4 |	if the foot is in an utterance-final position and it does not have a CVS structure.
| S W |	→	| 2 : 1 |	→ | 2 : 2|	if the salient syllable has a CVS structure; or the weak syllable is the first element of a compound that does not have a CVS structure; or both the salient syllable and the weak syllable are function words.
| S W W |	→	| 1½ : ¾ : ¾ |	→ | 1⅔ : 1⅔ : 1⅔ |	if the salient syllable has a CVS structure; or it is in an utterance-initial position; or it is a function word and the two weak syllables are two function words or a function word and a linker syllable.

where S and W indicate salient (stressed) and weak (unstressed) syllables, respectively. It is noted that the four-syllable and five-syllable feet are very rare and are omitted from discussion. Note also that foot boundaries are usually inserted in front of the salient syllables. The above rule can be used to obtain a derived syllable duration information (phonological level) from the corresponding syllable duration (phonetic level) obtained during the low-level acoustic processing.

By using stress, pause, and derived syllable duration information, the input utterance can be divided into feet. Then, the strength dynamics can be assigned as follows. Since we only distinguish

between two classes of stress, the salient syllable immediately after a weak syllable receives a strength dynamic of SD. A word before a pause receives a strength dynamic of DE as well as the '*final*' feature. A word after a pause receives a strength dynamic of SI if it is in the utterance-initial position; otherwise, it receives a strength dynamic of ID.

3.3. Prosodic Disambiguation Process

In this section, we describe our prosodic disambiguation process in CDG parsing. We assumed that the word graph for our CDG parser was perfectly annotated with prosodic information (i.e., strength dynamics). Our method of incorporating prosodics into the CDG formalism is quite similar to the work proposed by Zoltowski et.al (1992) for handling English. To construct prosodic constraints, two additional access functions were introduced:

`(abut x y)`	returns true if x and y abut each other (i.e., x and y are adjacent),
`(strength x y)`	returns the strength dynamic between x and y given that x and y are adjacent.

Given the above access functions, prosodic constraints are constructed in the form of binary constraints (i.e., constraints relating two variables).

The parsing process begins by propagating syntactic constraints to eliminate syntactically ill-formed sentence hypotheses. Then, prosodic constraints are propagated to check the agreement based on the strength dynamics between every pair of word candidates in each of the remaining competing sentence hypotheses of the ambiguous utterance. A sentence hypothesis is rejected if its syntactic structure is incompatible with the prosodic structure encoded via the annotated strength dynamics.

4. Experiments and Results

The above approach of incorporating prosodics into a CDG grammar has been tested on a set of ambiguous sentences, which represents various structural ambiguities involving five types of compounds in Thai: noun-noun, noun-propernoun, noun-verb, non-verb-noun, and verb-noun. There are two test sentences for each type of ambiguity resulting in a total of 10 sentence types for the whole set. These sentences consist of only monosyllabic words in which structural ambiguity in Thai often involves. For a complete list of test sentences, see Potisuk (2007).

To illustrate how the strength dynamic information extracted from the input speech signal is used in a CDG grammar to eliminate prosodically implausible parses, we provide two prosodic constraints to help disambiguate two example sentence hypotheses.

1) $_{SI}$ kɛɛŋ $_{DE}$ p^{h}èt $_{SD}$ mâak $_{DE}$ paj 'The curry is too spicy.'
| subj. | verb | adv. | aux. |

2) $_{SI}$ kɛɛŋ $_{SD}$ p^{h}èt $_{DE}$ mâak $_{DE}$ paj 'There is too much curry.'
| subj. (compound) | verb | aux. |

In this example, structural ambiguity results from the relationship between the first two words, subject-verb vs. a noun-verb compound. To handle this type of ambiguity, the following two prosodic constraints were constructed. The first indicates the requirements for a compound, and the second indicates the requirements for a subject-verb combination of words.

```
;; To have a label root, a verb abutting a modifier of a compound to its
left must have strength of DE.
      (if (∧ (abut (pos x) (pos(y))
             (eq (lab x) root)
             (eq (rid y) governor)
             (gt (pos x) (pos y))
             (~ (eq (mod y) (pos x))))
          (eq (strength x) DE)).

;; To have a label root, a verb abutting a subject modifier to its left
must have strength of DE.
      (if (∧ (abut (pos x) (pos(y))
             (eq (lab x) root)
             (eq (rid y) governor)
             (eq (lab y) subj)
             (gt (pos x) (pos y)))
          (eq (strength x) DE)).
```

Given the strength dynamics as in the first sentence, the word /pet/ must have a label root because it has strength of DE, while the word /maak/ cannot have a label root based on the second constraint. Thus, the compound interpretation is eliminated. On the other hand, given the strength dynamics as in the second sentence, the first constraint only permits the word /pet/ to be labeled as a verb modifier of a compound because it has strength of SD. As a result, the word /maak/ can only become the main verb of the sentence, thus eliminating the subject-verb sequence interpretation.

5. Summary and Discussion

A system combining constraint dependency parsing of Thai spoken sentences with prosodic constraints has been described. Prosodic constraints have been developed to rule out prosodically implausible sentence hypotheses in the face of syntactic ambiguity. These prosodic constraints check for the agreement between every pair of words in each of the competing sentence hypotheses of an ambiguous utterance based on the annotated strength dynamics. A sentence hypothesis is rejected if its syntactic structure is not congruent with the prosodic structure given by the annotated strength dynamics.

Since the algorithm is in a very early stage of development, we are unable to assess its true performance. We are planning to generalize the approach to cover a wider variety of sentences. Specifically, we are modifying our constraint network to tolerate multiple word candidates, which are prevalent in the spoken language domain, both for words in a particular word position (aligned) as well as for word candidates that overlap each other in time (unaligned). The problem of multiple word candidates is likely to occur when we extend our system to account for ambiguity involving polysyllabic words.

6. References

Potisuk, S. and M. P. Harper. 1996. CDG: An Alternative Formalism for Parsing Written and Spoken Thai. *Proceedings of the Fourth International Symposium on Languages and Linguistics*, pp. 1177-1196.

Potisuk, S. 2007. Prosodic Annotation in a Thai Text-to-speech System. *Proceeding of the 21st Pacific Asia Conference on Language, Information and Computation*. pp. 405-414.

Zoltowski, C.B., Harper, M.P., Jamieson, L.H., and Helzerman, R.A. (1992) PARSEC: A Constraint-based Framework for Spoken Language Understanding. *Proceedings of the International Conference on Spoken Language Processing.* pp. 249-252.