Available online at www.sciencedirect.com

ScienceDirect

Computers and Mathematics with Applications 55 (2008) 1735–1753

An International Journal
computers & mathematics with applications

www.elsevier.com/locate/camwa

Focus to emphasize tone analysis for prosodic generation

Lalita Narupiyakul[a,c], Vlado Keselj[a], Nick Cercone[b,*], Booncharoen Sirinaovakul[c]

[a] *Dalhousie University, 6050 University Ave., Halifax NS, B3H 1R2 Canada*
[b] *York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3 Canada*
[c] *King Mongkut's University of Technology Thonburi, 91 Pracha-uthit, Thungkru, Bangkok 10140, Thailand*

Received 2 July 2007; accepted 17 July 2007

Abstract

Emphasizing prosody of a sentence at its focus part when producing a speaker's utterance can improve the recognition rate to hearers and reduce its ambiguity. Our objective is to address this challenge by analysing the concept of foci in speech utterances and the relationship of focus, speaker's intention and prosody. Our investigation is aimed at understanding and modelling how a speaker's utterances are influenced by the speaker's intentions. The relationship between speaker's intentions and focus information is used to consider which parts of the sentence serve as the focus parts. We propose using the Focus to Emphasize Tone (FET) analysis, which includes: (i) generating the constraints for foci, speaker's intention and prosodic features, (ii) defining the intonation patterns, (iii) labelling a set of prosodic marks for a sentence. We also design the FET structure to support our analysis and to contain focus, speaker's intention and prosodic components. An implementation of the system is described and the evaluation results on the CMU Communicator (CMU–COM) dataset are presented.

Keywords: Prosodic annotation; Focus analysis; Speaker's intention; Unification-base grammar; HPSG

1. Introduction

A speaker's utterance may convey different meanings to a hearer. Such ambiguities can be resolved by emphasizing accents in different positions. Focus information is needed to select correct positions for accent labels. To determine focus information, speaker's intentions must be revealed. We apply speech act theory to written sentences as our input to determine a speaker's intention. Subsequently our system labels the prosodic marks on the FET structure which are the symbolic representation of speaker's utterances as a result.

We design the information structure for focus. The focus is used to declare what content the speaker wants to emphasize to the listener. When a speaker utters a sentence, the hearer must pay attention to the focus part. The focus part identifies what part of the sentence can be marked with the strong accent or emphasized by a high tone. We explore how focus relates to prosody in this paper. Furthermore, we use the speech act features to classify what are the speaker's intentions. We analyse the relationships of speech acts and tones. We design a set of constraints for these relationships and use them to define the tone marks.

[*] Corresponding author.
E-mail address: ncercone@yorku.ca (N. Cercone).

0898-1221/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2007.07.009

Our paper is composed of eight sections. Section 2 is a review of the existing prosodic generation approaches. In Section 3, the transformation from the Minimal Recursive Semantic (MRS) representation [1] to the Focus Content (FC) structure is explained. In Section 4, the FET analysis is presented, which includes: (i) defining the focus conditions and focus components and (ii) analysing the relationships of speaker's intention and prosody for each focus part. Design of the FET structure to support our analysis is explained in Section 5. An illustrated example of the FET analysis is shown in Section 6. We introduce the FET subgrammar for the Linguist Knowledge Base (LKB) system [2] in the Section 7 and its evaluation in Section 8. The last section presents our conclusion.

2. Background

There are three main methodological approaches to generating prosodic patterns: the learning methods [3,4], which are based on a corpus-based system to find prosodic patterns, the template-based methods [5], and the unification-based methods [6]. The learning method analyses the prosodic features by using the classification techniques. The system builds the prosodic models, scores these models and selects the possible prosodic patterns to fill in the prepared slot. This system is a fully-automatic system and provides the flexibility to apply to any domain that has dataset of sufficiently large size for the learning process. However a limitation of this system is that it has low capability to predict new or unseen information without learning from background information. The template-based approach is frequently used in Natural Language Generation (NLG). In the template-based approach, it is easy to add new template or design new content by hand. However, a disadvantage of a template-based system is that it is domain dependent, i.e. it is not flexible to apply to other domains. Furthermore it is time-costly to build the system and add new templates, because they require significant manual effort.

Two interesting examples of unification-based methods to prosodic annotation are proposed by Klein [6] and Haji [7]. They used a unification-based method to find prosodic constituency from syntactic and semantic structures. There are two main advantages that a unification-based method should provide:

First, we know that one sentence can have several meanings depending on their speaker's utterances. In a unification-based method, the prosodic pattern is created for the utterance based on the analysis of the speaker's intention and focus content. Therefore the ambiguity of speaker's utterance for the same sentence is reduced. The learning and template-based methods generate the possible prosodic patterns based on their templates or a learning corpus, which are both practically too limited to cover all possible utterances for the same sentence. Therefore, the speaker's utterance, generated from the prosodic patterns by learning and template-based methods, may not be able to convey the meaning that speaker want hearer to recognize.

Second, by using the unification-based method implemented in the LKB system, we can adopt the results of theoretical linguistics with minimal reinterpretation, and allow formal theories to be tested and validated on a dataset including prosodic phenomena. Furthermore, our subgrammar can be integrated into the other LKB systems, such as an English parser developed using the LKB system with LinGO English Resource Grammar (ERG) [8].

3. Generating focus content structure from the MRS representation

The MRS representation is a kind of flat semantic representation, which uses the reference numbers to define relations and constraints among objects. It is similar to the conventional predicate calculus representation. The representation is a result of Head-Driven Phrase Structure Grammar (HPSG) parsing by using the LKB system. The LKB system is distributed by CSLI Linguistic Grammars Online (LinGO) Laboratory at Stanford University. The LKB system is a grammar and lexicon environment, to be used with constraint-based linguistic formalisms. For generating the MRS representation, the LKB system requires a particular grammar called ERG. The ERG contains more than 10,000 lexemes and each lexeme consists of the syntactic and semantic information. The LKB system with ERG can parse a sentence, and produce an integrated syntactic and semantic structure of the sentence. The result of this parsing is the syntactic tree and the MRS representation.

In the preprocessing stage, the MRS representation requires a comprehensive structure, which can be in the form of Attribute Value Matrix (AVM). This structure represents clearly the semantic relations of the objects in the sentence, and makes feasible the analysis of focus information and generation of the FC structure. The MRS representation is transformed to an AVM by scanning each feature inside the MRS representation. The reference numbers are kept and mapped to their objects. Every connection that is related to this object and this reference number needs to be recorded.

$$\begin{bmatrix} \text{LTOP}:\text{h1} \\ \text{INDEX}:\text{e2 [e E.MOOD}:\text{INDICATIVE E.TENSE}:\text{PAST E.ASPECT.PERF}:\text{- E.ASPECT.PROGR}:\text{-]} \\ \text{RELS}:\left\langle \begin{array}{l} \begin{bmatrix}\text{prop\_ques\_m\_rel} \\ \text{LBL}:\text{h1} \\ \text{MARG}:\text{h3} \\ \text{PSV}:\text{u5} \\ \text{TPC}:\text{u4}\end{bmatrix}, \begin{bmatrix}\text{a\_q\_rel} \\ \text{LBL}:\text{h6} \\ \text{ARG0}:\text{x7} \\ \text{RSTR}:\text{h9} \\ \text{BODY}:\text{h8}\end{bmatrix}, \begin{bmatrix}\text{"\_young\_a\_rel"} \\ \text{LBL}:\text{h10} \\ \text{ARG0}:\text{e11} \\ \text{ARG1}:\text{x7}\end{bmatrix}, \begin{bmatrix}\text{"\_boy\_n\_rel"} \\ \text{LBL}:\text{h10} \\ \text{ARG0}:\text{x7}\end{bmatrix}, \begin{bmatrix}\text{"\_buy\_v\_rel"} \\ \text{LBL}:\text{h12} \\ \text{ARG0}:\text{e2} \\ \text{ARG1}:\text{x7} \\ \text{ARG2}:\text{x13}\end{bmatrix}, \\ \begin{bmatrix}\text{a\_q\_rel} \\ \text{LBL}:\text{h14} \\ \text{ARG0}:\text{x13} \\ \text{RSTR}:\text{h16} \\ \text{BODY}:\text{h15}\end{bmatrix}, \begin{bmatrix}\text{"\_red\_a\_rel"} \\ \text{LBL}:\text{h17} \\ \text{ARG0}:\text{e18]} \\ \text{ARG1}:\text{x13}\end{bmatrix}, \begin{bmatrix}\text{"\_flower\_n\_rel"} \\ \text{LBL}:\text{h17} \\ \text{ARG0}:\text{x13}\end{bmatrix}, \begin{bmatrix}\text{for\_p\_rel} \\ \text{LBL}:\text{h17} \\ \text{ARG0}:\text{e20} \\ \text{ARG1}:\text{x13} \\ \text{ARG2}:\text{x19}\end{bmatrix}, \begin{bmatrix}\text{def\_explicit\_q\_rel} \\ \text{LBL}:\text{h21} \\ \text{ARG0}:\text{x19} \\ \text{RSTR}:\text{h23} \\ \text{BODY}:\text{h22}\end{bmatrix}, \\ \begin{bmatrix}\text{pro\_poss\_rel} \\ \text{LBL}:\text{h24} \\ \text{ARG0}:\text{i26} \\ \text{ARG1}:\text{x25} \\ \text{ARG2}:\text{x19}\end{bmatrix}, \begin{bmatrix}\text{pronoun\_q\_rel} \\ \text{LBL}:\text{h27} \\ \text{ARG0}:\text{x25} \\ \text{RSTR}:\text{h28} \\ \text{BODY}:\text{h29}\end{bmatrix}, \begin{bmatrix}\text{pron\_rel} \\ \text{LBL}:\text{h30} \\ \text{ARG0}:\text{x25}\end{bmatrix}, \begin{bmatrix}\text{"\_mother\_n\_rel"} \\ \text{LBL}:\text{h24} \\ \text{ARG0}:\text{x19} \\ \text{ARG1}:\text{i31}\end{bmatrix} \end{array} \right\rangle \\ \text{HCONS}:<\text{h3 qeq h12 h9 qeq h10 h16 qeq h17 h23 qeq h24 h28 qeq h30}> \end{bmatrix}$$

Fig. 1. The MRS representation of the sentence “A young boy bought a red flower for his mother”.

```
Line
1 :  prop_ques_m_rel:→
2 :  MARG:go to h3
3 :  LBL = h3:→“_buy_v_rel”:→
4 :                  ARG0:→PAST,
5 :                  ARG1:→3SG,
6 :                              go to h9:ARG0→ ”_boy_n_rel”(x7);
7 :                              go to h9:ARG1→ ”_young_a_rel”(x7); go to h6:ARG0-> _a_q_rel(x7);
8 :                                       ARG2→3SG,
9 :                                                go to h16:ARG1→ ”_red_a_rel”(x13);
10:                                                go to h16:ARG1→ _for_p_rel(x13);
11:                                                          ARG2→3SG,
12:                                                                    go to h23:ARG0→ ”_mother_n_rel”(x19);
13:                                                                    go to h23:ARG2→ pro_poss_rel(x19);
14:                                                                               ARG1→3SG,
15:                                                                                         go to h28:ARG0→ pron_rel(x25);
16:                                                                                         go to h27:ARG0→ pronoun_q_rel(x25);
17:                                                                    go to h21:ARG0→ def_explicit_q_rel(x19);
18:                                                go to h16:ARG0→ “_flower_n_rel”(x13);
19:                                                go to h14:ARG0→ _a_q_rel(x13);
20:  PSV:→u5,
21:  TPC:→u4,
```

Fig. 2. Scanning the MRS representation of the sentence “A young boy bought a red flower for his mother”.

For example, the sentence “A young boy bought a red flower for his mother” is parsed by the LKB system with ERG, and the resulting MRS representation of this sentence is shown in Fig. 1. Focus content scoping begins by scanning this MRS representation and following the reference numbers or indexes. All connections of objects depending on the reference number are recorded. A result of scanning the MRS representation is illustrated in Fig. 2. The indention in this figure represents the scanning into the next level of hierarchy inside the MRS representation.

In the FET analysis, each focus part contains lists of words. Defining the lists of words for the prosodic structure depends on prosodic marking. For example, the words of the sentence “a young boy bought a red flower for his mother.” may be grouped as [a, young, boy], [bought], [a, red, flower], and [for, his, mother], based on the prosodic phenomena [9]. The lists of words in a sentence are marked by prosodic marks sequentially, so in this example it could be represented as follows: [a young boy]$_{t_1}$, [bought], [a, red, flower], [for his mother]; where t_i represents the tone mark at order i. The tone is marked at the actor part and this sentence is focused at “who bought a red flower”. To transform the MRS representation to the FC structure, the hierarchy of MRS structure must be minimized to be a flat tree structure. We design and implement the algorithm called Focus Content Scoping (FCS) to reduce hierarchy for a simple sentence.

3.1. Focus content scoping

The FCS starts from the bottom or leaf nodes of the MRS structure. The *n_rel* is declared as the noun phrase relation, *q_rel* is the quantifier relation, *a_rel* is the adjective relation, and *p_rel* is the preposition phrase relation. The level of *p_rel* is usually lower than *n_rel*'s level of noun phrase modification. We move the level of *p_rel* to upper level of the prosodic structure. The *c_rel* is the conjunction relation, and it is divided to left node index (*L-index*) and right node index (*R-index*). The algorithm of FCS is shown below.

Algorithm 1 Focus Content Scoping (FCS).

```
 1: Input: A tree of MRS structure (T) for the input sentence (S)
 2: Output: Flat Tree
 3: repeat
 4:    start from leaf node
 5:    move a_rel and q_rel to be in the same level of their n_rel
 6:    for each p_rel in T do
 7:       merge the preposition node with a next node
 8:       move the node to upper level
 9:    end for
10:    for each c_rel in T do
11:       merge R-index with conjunction
12:       move both L-index and R-index to the upper level
13:    end for
14: until T is equal to flat tree
```

For instance, in Fig. 3, using the FCS, the MRS representation is transformed into an AVM. The AVM is a standard matrix structure and we use AVMs to represent the relations of the MRS components inside the sentence. This MRS representation can be represented by the tree in Fig. 4(a). Following the FCS's algorithm (line 5), the nodes or leaves of *a_rel*, *q_rel*, and *n_rel* at the same level are combined as shown in Fig. 4(b). The *p-rel* node merges with the next node (line 6 of the algorithm) and moves to the upper level as illustrated in Fig. 4(c). Finally, the hierarchy becomes a flat tree. We can define words or groups of words for actor, act and actee parts. In this example, the actor part is <a, young, boy>, the act part is <bought> and the actee part is <[a red flower], [for his mother]>. The FC structure is shown in Fig. 5. This FC structure is the input of the FET analysis.

4. Focus-to-emphasize-tone analysis

In this analysis, we consider two main parts: (i) focus conditions, and (ii) the relationship between speaker's intention, focus and prosody. We analyse the focus parts (*actor*, *act* and *actee*), focus types (*w-focus* and *s-focus*) and their combinations. Our focus analysis can be classified according to five conditions. These conditions are designed to cover the combinations of focus parts and focus types. We explain these conditions in Section 4.1. In the second part, we investigate the relationships of focus part with speaker's intention and prosody. Based on our analysis, these relationships correspond to the intonation patterns. The focus content can be conveyed to hearer by emphasizing prosody at the focus parts. For our proof-of-concept FET analysis, we consider only three speech act types with the different focus parts. The details are described in Section 4.2.

4.1. Focus conditions

Our FET analysis uses a constraint-based approach. From the FC structure we detect the parts (actor, act, actee or their combinations) that should be in focus. If the focus is marked at a position in a sentence, then the hearer needs to pay attention to the content at that position in the sentence. For example, the speaker could utter the sentence

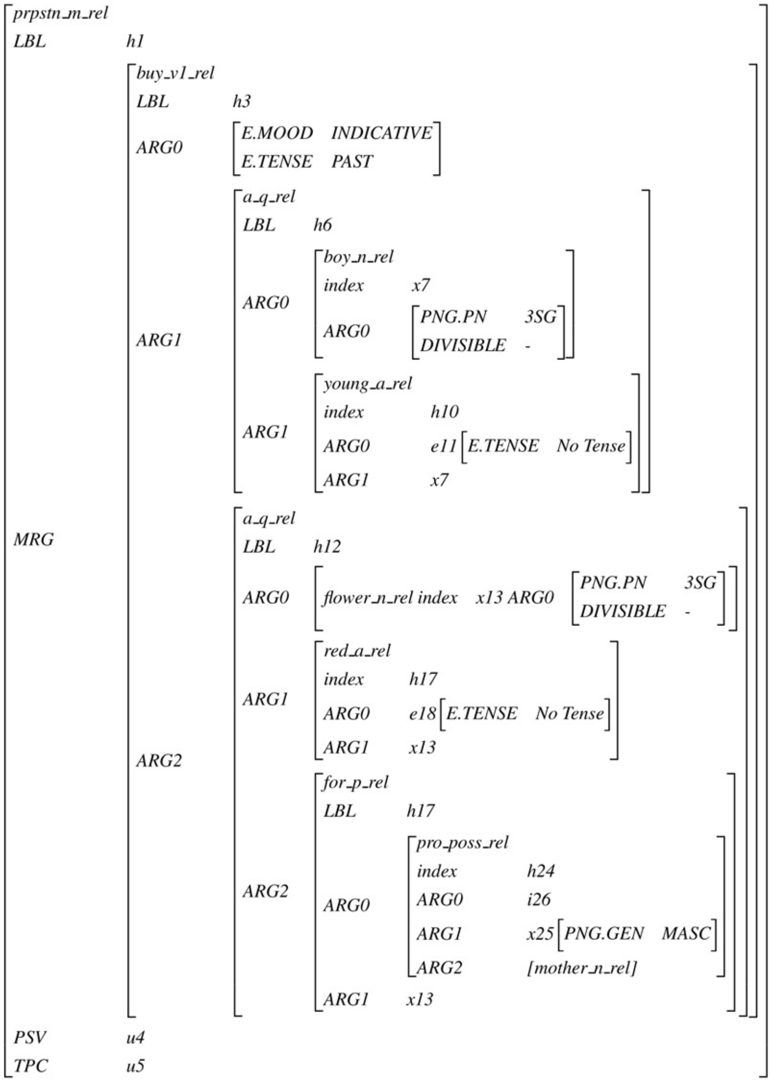

Fig. 3. Transforming the MRS representation to the AVM structure of the sentence “A young boy bought a red flower for his mother”.

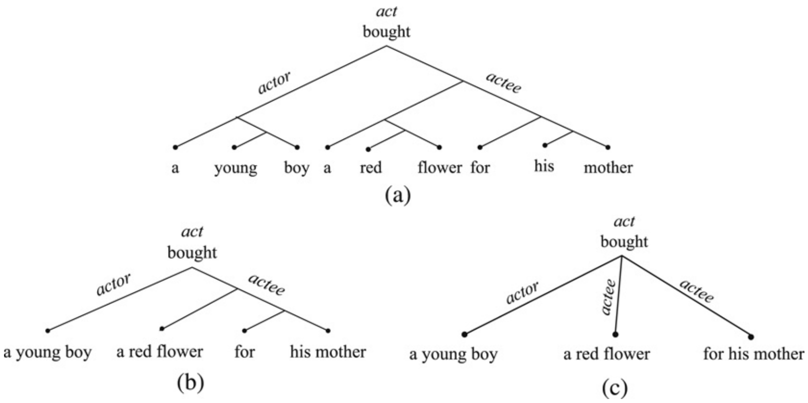

Fig. 4. Tree of MRS representation: (a) The hierarchy of MRS representation for the sentence “A young boy bought a red flower for his mother” (b) Combining the *a_rel*, *q_rel*, and *n_rel* nodes together (c) Collapse the *p_rel* and move to the upper level.

$$content: \left\{ \begin{bmatrix} \textit{Focus–Part} & \textit{actor} \\ \textit{List-obj} & \langle a,\ young,\ boy \rangle \\ \textit{Index} & h9 \\ \textit{RIndex} & h3 \end{bmatrix}, \begin{bmatrix} \textit{Focus–Part} & \textit{actee} \\ \textit{List-obj} & \langle [a,\ red,\ flower], [for,\ his,\ mother] \rangle \\ \textit{Index} & h15 \\ \textit{RIndex} & h3 \end{bmatrix}, \begin{bmatrix} \textit{Focus–Part} & \textit{act} \\ \textit{List-obj} & \langle buy \rangle \\ \textit{Index} & h3 \\ \textit{RIndex} & \{h9, h15\} \end{bmatrix} \right\}$$

Fig. 5. The focus content structure of the sentence “A young boy bought a red flower for his mother”.

Table 1
The different focuses in the sentence

	Focus	Speaker wants to focus on . . .
1	[KIM]$_F$ bought a flower.	Who bought a flower?
2	Kim bought [a FLOWER]$_F$.	What did Kim buy?
3	Kim [BOUGHT a flower]$_F$.	What did Kim do?

Table 2
The focus parts and the focus types

No.	Focus parts	Focus types
A	actor+act+actee	{w-focus(actor),s-focus(act),w-focus(actee)} or undefined
B	actor+act	{w-focus(actor),s-focus(act)}

C	actor+actee	{w-focus(actor),s-focus(actee)}
		or {w-focus(actee),s-focus(actor)}

D	actor	w-focus(actor) or s-focus(actor)
E	act+actee	{s-focus(act),w-focus(actee)}
F	act	s-focus(act)
G	actee	w-focus(actee) or s-focus(actee)
H	∅	undefined

$$\textit{make\_s-focus}: \begin{bmatrix} \textit{s-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle a_1, a_2, \ldots, a_n \rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \end{bmatrix} \rightarrow \begin{bmatrix} \textit{s-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle a_1, a_2, \ldots, a_n \rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \\ \textit{FET-obj} & \langle a_n \rangle \end{bmatrix}$$

Fig. 6. Condition (a) *s-focus* structure: Marking the *s-focus* for the actor or actee parts).

“Kim bought a flower” with emphasis at the different positions in the sentence, as shown Table 1. From the preprocessing step, we transformed the MRS structure to the FC structure. This structure contains “actor” (a person or a thing that acts something in a sentence), “act” (an activity in that sentence), and “actee” (the response of the activity) parts.

Regarding the focus parts, our model incorporates two focus types: *w-focus*, and *s-focus*. The *w-focus* type represents wide focus, which covers a phrase or a word. The *s-focus* type represents single focus, which is placed on a word in the sentence. We assign the actor and actee parts as single or wide focus while the act part is only an *s-focus*. Normally, the focus does not cover only the act part. If the focus covers the act part, then the focus must cover at least one of the related parts (actor or actee). Therefore, we set the focus types following all situations that occur according to a set of rules called focus criteria. Eight focus criteria are shown in Table 2.

The constraints for focus types are defined according to five different cases, which capture a set of simple sentence structures. These cases are described below.

(a) *s-focus on the actor or actee parts*. The last node in the list of objects is defined as the focus position at which the tone is emphasized (*FET-obj*), see Fig. 6.

(b) *w-focus at the actor or actee parts.* The list of objects is *FET-obj* in the sentence as shown in Fig. 7.

$$\textit{make\_w-focus:}\left[\begin{array}{ll}\textit{w-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle a_1, a_2, \ldots, a_n\rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j\end{array}\right] \rightarrow \left[\begin{array}{ll}\textit{w-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle a_1, a_2, \ldots, a_n\rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \\ \textit{FET-obj} & \langle a_1, a_2, \ldots, a_n\rangle\end{array}\right]$$

Fig. 7. Condition (b) *w-focus* structure: Marking the *w-focus* for the actor or actee parts.

$$\textit{merge\_list\_w-focus:}\left[\textit{List-obj}\ \left\langle\begin{array}{l}[a_1, a_2, \ldots, a_n], \\ [b_1, b_2, \ldots, b_n], \ldots, \\ [m_1, m_2, \ldots, m_n]\end{array}\right\rangle\right] \rightarrow \left[\begin{array}{ll}\textit{w-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \left\langle [a_1, a_2, \ldots, a_n],[b_1, b_2, \ldots, b_n], \ldots, [m_1, m_2, \ldots, m_n]\right\rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \\ \textit{FET-obj} & \langle a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n, \ldots, m_1, m_2, \ldots, m_n\rangle\end{array}\right]$$

Fig. 8. Condition (c) merge focus structure: Marking the *w-focus* of multiple lists of objects for the actor or actee parts.

$$\textit{split\_list\_s-focus:}\left[\textit{List-obj}\ \left\langle\begin{array}{l}[a_1, a_2, \ldots, a_n], \\ [b_1, b_2, \ldots, b_n], \ldots, \\ [m_1, m_2, \ldots, m_n]\end{array}\right\rangle\right] \rightarrow \left\{\begin{array}{l}\left[\begin{array}{ll}\textit{s-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle a_1, a_2, \ldots, a_n\rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \\ \textit{FET-obj} & \langle a_n\rangle\end{array}\right] \vee \left[\begin{array}{ll}\textit{s-focus} & \\ \textit{Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle b_1, b_2, \ldots, b_n\rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \\ \textit{FET-obj} & \langle b_n\rangle\end{array}\right] \vee \ldots \vee \\ \left[\begin{array}{ll}\textit{s-focus Relation} & \textit{focus\_part} \\ \textit{List-obj} & \langle m_1, m_2, \ldots, m_n\rangle \\ \textit{Index} & r_i \\ \textit{RIndex} & s_j \\ \textit{FET-obj} & \langle m_n\rangle\end{array}\right]\end{array}\right\}$$

Fig. 9. Condition (d) split focus structure: Marking the *s-focus* of the multiple lists of objects for the actor or actee parts.

(c) *w-focus at actor or actee parts containing the multiple lists of objects.* The lists are merged together to be the *FET-obj* as shown in Fig. 8.
(d) *s-focus at actor or actee parts containing the multiple lists of objects.* If the focus type is an *s-focus* and there are *m* sets of lists of objects (multiple lists of objects), then these lists of objects can be split into the *s-focus* of each list of objects, see Fig. 9.
(e) *focus on the act part.* Two cases of defining the focus types are shown in Fig. 10. The first case, the *s-focus* marks the act part while the *w-focus* marks the actee part. In the second case, the *s-focus* marks the act part and the *w-focus* marks the actor part.

There five conditions cover all possible situations for a group of simple sentences, in which we define focus based on the FC structure.

4.2. Relationship of focus with speech acts and prosody

We define the speech act codes following Ballmer [10]. To mark these codes, we consider the main verb (known as the act part inside the FC structure). These codes define the speech act categories associated with each sentence. A sentence can be marked by more than one code according to speech act classification [10]. We mark the speech act codes for 62 sentences from a part of the CMU–COM dataset [11]. Considering the relationships between speech acts and focus parts, we found some common patterns for marking tones in a sentence. For example, the tone mark `L-L%`, analysed as low phrase tone (`L-`) to low boundary tone (`L%`), is marked at the last word of a sentence for affirmative sentences. The tone marks `H-` (high phrase tone) and `L-` are marked at the last word before conjunction (such as “and”,

$$\textit{make\_act-focus:}\begin{bmatrix}\textit{Relation} & \textit{act}\\ \textit{List-obj} & \langle b_1,b_2,\ldots,b_n\rangle\\ \textit{Index} & s_i\\ \textit{RIndex} & \{r_j,r_k\}\end{bmatrix}\rightarrow\left\{\begin{matrix}\begin{bmatrix}\textit{s–focus} & \\ \textit{Relation} & \textit{act}\\ \textit{List-obj} & \langle b_1,b_2,\ldots,b_n\rangle\\ \textit{Index} & s_i\\ \textit{RIndex} & \{r_j,r_k\}\\ \textit{FET–obj} & \langle b_n\rangle\end{bmatrix},\\ \begin{bmatrix}\textit{w–focus} & \\ \textit{Relation} & \textit{actee}\\ \textit{List-obj} & \langle c_1,c_2,\ldots,c_n\rangle\\ \textit{Index} & r_k\\ \textit{RIndex} & s_i\\ \textit{FET–obj} & \langle c_1,c_2,\ldots,c_n\rangle\end{bmatrix}\end{matrix}\right\}\vee\left\{\begin{matrix}\begin{bmatrix}\textit{w–focus} & \\ \textit{Relation} & \textit{actor}\\ \textit{List-obj} & \langle a_1,a_2,\ldots,a_n\rangle\\ \textit{Index} & r_j\\ \textit{RIndex} & s_i\\ \textit{FET–obj} & \langle a_1,a_2,\ldots,a_n\rangle\end{bmatrix},\\ \begin{bmatrix}\textit{s–focus} & \\ \textit{Relation} & \textit{act}\\ \textit{List-obj} & \langle b_1,b_2,\ldots,b_n\rangle\\ \textit{Index} & s_i\\ \textit{RIndex} & \{r_j,r_k\}\\ \textit{FET–obj} & \langle b_n\rangle\end{bmatrix}\end{matrix}\right\}$$

Fig. 10. Condition (e) merge focus structure: Marking the *w-focus* for the act part.

"or", "but", and so on) or are marked at the last word of the current phrase (following the next phrase). We know that the tone mark H* (high accent tone) is used to emphasize a word or a group of words in a sentence. If we want strong emphasis at a word or a group of words then we use the tone mark L+H* (rising accent tone) instead of H*. The speech acts that we consider include: intending (*EN0ab*), want (*DE8b*), and victory (*KA4a*). Some tone patterns for our FET analysis are include below.

For example, in the speech act code *EN0ab* ("intending"), if the speaker focuses on what the actor wants to do, the actee part is the most important part in the sentence. The components of accent tone (*Accent–Tone*) and boundary tone (*Bound–Tone*) in the actee part can be repeated, as shown in the tone pattern (1).

$$\textit{Actee\_tone} \leftarrow ([\textit{Accent–Tone}] + \textit{Bound–Tone})^n. \tag{1}$$

Note: n is the numbers of phrases, and the variable in square bracket is optional.

The tone patterns are a combination of accent and boundary tones. The accent tone labels for the actee parts are L* or L+H* for affirmative sentences and H* or L+H* for interrogative sentences. For the boundary tone, the possible tone marks are L- or L-L% for affirmative sentences and H- or H-H% for interrogative sentences. The tone patterns of the actee part for *EN0ab* are shown in tone patterns (2) and (3). From the datasets, the tone patterns are found for affirmative and interrogative sentences. The tone patterns for these sentences are repeatable until ending with the tone marks L-L% or H-H%.

For the affirmative sentence

$$\textit{Actor\_tone} \leftarrow (\mathtt{L*} \vee (\mathtt{L} + \mathtt{H*}) + \mathtt{L-})^{n-1} + ([\mathtt{L*} \vee (\mathtt{L} + \mathtt{H*})] + \mathtt{L} - \mathtt{L\%}). \tag{2}$$

For the interrogative sentence

$$\textit{Actor\_tone} \leftarrow (\mathtt{H*} \vee (\mathtt{L} + \mathtt{H*}) + \mathtt{H-})^{n-1} + ([\mathtt{H*} \vee (\mathtt{L} + \mathtt{H*})] + \mathtt{H} - \mathtt{H\%}). \tag{3}$$

Note: n is the number of phrases and the variables in square bracket are optional.

For the speech act code *DE8b* ("want"), the speaker informs the listener that the speaker wants something from listener. Therefore, the actee part is emphasized by tone marks. The relevant tone marks are the accent tone H* and the boundary tones L-L% and H-H%, both occurring at the actee part in the datasets. The tone patterns are shown in tone patterns (4) and (5) and they use simple tone patterns for the affirmative and interrogative sentences.

For the affirmation sentence

$$\textit{Actor\_tone} \leftarrow (\mathtt{H*}) + (\mathtt{L} - \mathtt{L\%}). \tag{4}$$

For the interrogative sentence

$$\textit{Actor\_tone} \leftarrow (\mathtt{H*}) + (\mathtt{H} - \mathtt{H\%}). \tag{5}$$

$$\begin{bmatrix} \textit{List-obj} & \left\langle a_1,a_2,\ldots,a_n \right\rangle \\ \textit{Focus-Info} & \begin{bmatrix} \textit{FET-obj} & \left\langle a_1,a_2,\ldots,a_m \right\rangle \\ \textit{Focus} & \begin{bmatrix} \textit{focus-part} & \\ \textit{Prosody} & [\;] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Fig. 11. A part of the FET structure.

$$\textit{Focus-Info:} \begin{bmatrix} \textit{FET-obj} & \left\langle w_1,w_2,\ldots,w_m \right\rangle \\ \textit{Focus} & \begin{bmatrix} \textit{Focus-Part} & \textit{focus-part} \\ \textit{Focus-Pos} & \textit{focus-pos} \\ \textit{FCGroup} & \textit{fcgroup} \\ \textit{FCType} & \textit{fctype} \\ \textit{Prosody} & [\;] \end{bmatrix} \end{bmatrix}$$

Fig. 12. Information structure.

For the speech act code *KA4a* (“victory”), the focus must be at both the actor and actee parts. The relevant tone marks are the accent tone marks H* and L+H* and the boundary tone marks L-L% and H-H%. The tone patterns are shown in tone pattern (6) for actor part, and the patterns (7) and (8) for the actee part of affirmative and interrogative sentences.

Actor Part

For affirmative and interrogative sentences

$$\textit{Actor\_tone} \leftarrow (\mathtt{H}*) \vee (\mathtt{L}+\mathtt{H}*). \tag{6}$$

Actee Part

For the affirmation sentence

$$\textit{Actor\_tone} \leftarrow ([\mathtt{H}* \vee (\mathtt{L}+\mathtt{H}*)] + \mathtt{L}-)^{n-1} + (\mathtt{L}-\mathtt{L}\%). \tag{7}$$

For the interrogative sentence

$$\textit{Actor\_tone} \leftarrow ([\mathtt{H}* \vee (\mathtt{L}+\mathtt{H}*)] + \mathtt{H}-)^{n-1} + (\mathtt{H}-\mathtt{H}\%). \tag{8}$$

5. Design of the FET structure

The FET structure is designed to contain the feature structures of focus and prosodic information. The prosodic structure is designed to be a subfeature structure of the focus information structure. The FET structure consists of three main features: the list of words (*List-obj*), focus information (*Focus-Info*), and prosodic information (*Prosody*) inside the focus information as shown in Fig. 11.

5.1. Focus structure

The *Focus-Info* structure operates between focus and prosodic features. The *Focus-Info* structure includes the focus part and list of focus words (*FET-obj*) from the FC structure. Furthermore, the *Focus-Info* structure also contains the focus types (*FCType*) and focus criterion (*FCGroup*), see Table 2. The focus part (*Focus-Part*) consists of actor, act or actee parts, while the focus type is *w-focus* or *s-focus*. This feature is analysed following the focus conditions in Section 4.1. Inside the FET structure, the *Prosody* structure is the significant subfeature structure of the *Focus-Info* structure because of the relationships of speech acts and tones depending on the focus parts. These relationships are described in Section 4.2. The coarse *Focus-Info* structure is illustrated in Fig. 12.

$$\begin{bmatrix} \textit{focus-part} & \\ \textit{Prosody} & \begin{bmatrix} \textit{STMood} & \textit{mood} \\ \textit{SPCode} & \textit{code} \\ \textit{Prosodic-Mark} & \begin{bmatrix} \textit{prosody-mark} & \\ \textit{Tone} & [\;] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Fig. 13. The prosodic information structure.

$$\underset{\text{(a)}}{\textit{Tone:} \begin{bmatrix} \textit{Accent-Tone} & \textit{accent} \\ \textit{Bound-Tone} & \textit{bound} \end{bmatrix}} \quad \underset{\text{(b)}}{\textit{Tone:} \begin{bmatrix} \textit{Accent-Tone} & \oslash \\ \textit{Bound-Tone} & \texttt{H-H\%} \end{bmatrix}}$$

Fig. 14. Tone structure (a) Tone structure with variables and (b) An example of tone structure.

$$\underset{\text{(a)}}{\begin{bmatrix} \textit{Prosodic-Mark} & \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \textit{accent} \\ \textit{Bound-Tone} & \textit{bound} \end{bmatrix} \end{bmatrix}} \begin{bmatrix} \textit{hEm\_Lg-break} & \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \texttt{L+H*} \\ \textit{Bound-Tone} & \texttt{L-L\%} \end{bmatrix} \end{bmatrix}_{\text{(b)}}$$

Fig. 15. Prosodic structure (a) Prosodic structure with variables and (b) An example of prosodic structure.

5.2. Prosodic structure

The *Prosody* structure contains features of focus, speaker's intention, and prosody. Based on the relationship among these features, we determine the tone marks. The *Prosody* structure is composed of sentence mood (*STMood*), speech act code (*SPCode*), prosodic mark (*Prosody-Mark*), and tone mark (*Tone*), which includes accent tone (*Accent-Tone*) and boundary tone (*Bound-Tone*). The *STMood* contains the type of a sentence, such as affirmative sentence (*aff*), and is derived from the FC structure. The *SPCode* is adopted from [10], and it denotes the speech act code, such as intending (*EN0ab*). The prosodic information structure is illustrated in Fig. 13.

In the prosodic domain, a word, marked with a tone, is called a prosodic word [7, p. 8], while a word which does not have tone marks is called a leaner [7, p. 6]. Therefore tones are set as marked or unmarked for a word. Tone marks can be separated into two main groups: *Accent-Tone* and *Bound-Tone* following the Tone and Break Indexing (ToBI) system [12]. The accent tone can occur at any part of the sentence except at the end of phrase or sentence and does not include the duration symbol (- or %). The examples of accent tones are H*, L*, and L+H*. The boundary tone occurs at the end of phrase or sentence, such as L-, H-, and L-L%.

The *Tone* structure contains a set of tone marks for the FET structure. The *Tone* structure, as shown Fig. 14(a), is composed of two features: *Accent-Tone* and *Bound-Tone*. These features are used to represent the tone marks for a word or phrase. For example, the phrase "ten red cars" is marked as H-H%. The tone structure for this example is shown in Fig. 14(b).

The tone marks are labelled according to the ToBI coding system. A tone mark can be a combination of *Accent-Tone* and *Bound-Tone*. The tone labels are determined according to the prosodic information, which is classified into types. Each type of prosodic information explains the function of the tone depending on tone position and prosodic phenomena. The prosodic information is separated to into *marked-info* and *unmarked-info*. The *unmarked-info* label means that no prosodic information are to be found at that word. The *marked-info* prosodic label includes additional prosodic information. There are two types of *marked-info*: strong tone (*strong*) and weak tone (*weak*). The strong tone means that words are easy to recognize by the listener such as emphasized marked information (*Em*), and high emphasized marked information (*hEm*). The weak tone represents the words that have low priority or low necessity in the content.

The *Prosodic-Mark* structure contains the prosodic function and tone marks. The prosodic function represents the tone emphasis level, such as high tone emphasis (*hEm*), generated from the combination of accent and boundary tones. The *Prosodic-Mark* structure is shown in Fig. 15(a). For example, the phrase "this black book" is marked with the high emphasized tone at this phrase. The speaker expects the listener to recognize this emphasis by using a strong accent and high tone. Therefore the *Prosodic-Mark* structure for this example is shown as Fig. 15(b).

Table 3
Example of mapping between prosodic marks and accent-boundary tones

No.	*Prosody-Mark*	*Accent–Tone*	*Bound–Tone*
1	*hEm*	`L+H*`	nobound
2	*hEm_EmLg-break*	`L+H*`	`H-H%`
3	*hEm_EmSh-break*	`L+H*`	`H-`
4	*EmLg-break*	noaccent	`H-H%`
5	*no-mark*	noaccent	nobound

$$\mathit{INFO}{:}\begin{bmatrix} \mathit{SPAct} & \begin{bmatrix} \mathit{SPCode} & \mathit{code} \\ \mathit{STMood} & \mathit{mood} \\ \mathit{FCGroup} & \mathit{group} \end{bmatrix} \\ \mathit{Focus\text{-}Info} & \left\{ \begin{bmatrix} \mathit{Focus\text{–}Part} & \mathit{actor} \\ \mathit{FET\text{–}obj} & \langle\rangle \\ \mathit{Prosody} & \langle\rangle \end{bmatrix}, \begin{bmatrix} \mathit{Focus\text{–}Part} & \mathit{actee} \\ \mathit{FET\text{–}obj} & \langle\rangle \\ \mathit{Prosody} & \langle\rangle \end{bmatrix}, \begin{bmatrix} \mathit{Focus\text{–}Part} & \mathit{act} \\ \mathit{FET\text{–}obj} & \langle\rangle \\ \mathit{Prosody} & \langle\rangle \end{bmatrix} \right\} \end{bmatrix}$$

Fig. 16. The FET structure.

$$\underset{\text{(a)}}{\mathit{SPAct}{:}\begin{bmatrix} \mathit{SPCode} & \mathit{code} \\ \mathit{STMood} & \mathit{mood} \\ \mathit{FCGroup} & \mathit{group} \end{bmatrix}} \quad \underset{\text{(b)}}{\mathit{SPAct}{:}\begin{bmatrix} \mathit{SPCode} & \mathit{EN0ab} \\ \mathit{STMood} & \mathit{interrogative} \\ \mathit{FCGroup} & \mathit{G} \end{bmatrix}}$$

Fig. 17. The speech act feature structure: (a) SPAct structure and (b) Example of SPAct structure.

The types of prosodic information (*Prosodic-Mark*) are marked on a word or list of words. These types represent a combination of the accent and boundary tones and are shown in Table 3. For example, if we want to emphasize a word in a sentence and expect that the listener can recognize that word, then we select high tone emphasis (*hEm*) for prosodic information. The high tone emphasis is labeled by the tone mark `L+H*`. The different prosodic phenomena are explained by the different types of prosodic information. An example of prosodic phenomenon is prosody for a Yes–No question, which requires a high tone at the last word of the sentence. We know that this high tone is a type of strong tone emphasis. In the types of prosodic information, the *EmLg-break* is selected for this example to mark the last word of Yes–No question sentence, which is labelled with `H-H%`.

5.3. The structures for the relationships of speech acts and prosodic marks based on focus parts

The FET feature structure constrains the relationships of focus parts with the speaker's intentions and prosodic marks. The FET structure is composed of two main structures: the speech act structure (*SPAct*), and the focus information (*Focus-Info*) structure, as shown in Fig. 16.

The *SPCode*, derived from the speech act classification by Brennenstuhl [10], is defined inside the *SPAct* structure (Fig. 17(a)). Mostly, the *SPCode* is inferred from the main action or main verb of a sentence. Each code identifies a category of the speaker's intention. For example, the speech act code *EN2b*, which means "asking for", is assigned to these verbs: beg, request, require, invite and so on. Another feature in the *SPAct* structure, the sentence moods (*STMood*) is obtained from *E.Mood* feature in the MRS representation and represents the type of sentence such as affirmative sentence. The last feature in the *SPAct* structure is the focus criteria (*FCGroup*). This feature is derived from the Table 2. In the example of *SPAct* structure depicted in Fig. 17(b), that *SPCode* is *EN0ab* ("intending"), *STMood* is interrogative sentence, and *FCGroup* is "G", which indicates that the focus of sentence at actee part can be wide focus (*w-focus*) or single focus (*s-focus*).

The *Focus-Info* structure indicates what focus part (actor, act or actee parts) must be emphasized by tone, and how the prosodic information can be related to focus information. The *Focus-Info* structure contains the focus parts with their focus features, such as *FCType*, and *Prosody* structure. The *Focus-Info* structure is shown in Fig. 18.

Considering the *Focus-Info* structure for each focus part, the *FCType* can be assigned the wide focus (*w-focus*), the single focus (*s-focus*), or no focus. The *List-obj* contains words or lists of words from a sentence, while the

$$
\left[\begin{array}{ll}
\mathit{SPAct} & \left[\begin{array}{ll} \mathit{SPCode} & \mathit{speech\_act\_code} \\ \mathit{STMood} & \mathit{sent\_mood} \\ \mathit{FCGroup} & \mathit{group} \end{array}\right] \\
\mathit{Focus\text{-}Info} & \left\{
\left[\begin{array}{ll}
\mathit{Focus\text{-}Part} & \mathit{actor} \\
\mathit{Focus\text{-}Pos} & \mathit{focus\_position} \\
\mathit{FCType} & \mathit{focus\_type} \\
\mathit{List\text{-}obj} & \boxed{1}\ \mathit{list(obj)} \\
\mathit{Index} & i \\
\mathit{RIndex} & j \\
\mathit{FET\text{-}obj}\ \boxed{1} & \\
\mathit{Prosody} & \left[\begin{array}{ll} \mathit{Prosodic\text{-}Mark} & \mathit{mark} \\ \mathit{Tone} & \left[\begin{array}{ll} \mathit{Accent\text{-}Tone} & \mathit{accent} \\ \mathit{Bound\text{-}Tone} & \mathit{boundary} \end{array}\right] \end{array}\right]
\end{array}\right],
\left[\begin{array}{ll}
\mathit{Focus\text{-}Part} & \mathit{act} \\
\mathit{Focus\text{-}Pos} & \mathit{focus\_position} \\
\mathit{FCType} & \mathit{focus\_type} \\
\mathit{List\text{-}obj} & \boxed{2}\ \mathit{list(obj)} \\
\mathit{Index} & j \\
\mathit{RIndex} & \{i,k\} \\
\mathit{FET\text{-}obj}\ \boxed{2} & \\
\mathit{Prosody} & \left[\begin{array}{ll} \mathit{Prosodic\text{-}Mark} & \mathit{mark} \\ \mathit{Tone} & \left[\begin{array}{ll} \mathit{Accent\text{-}Tone} & \mathit{accent} \\ \mathit{Bound\text{-}Tone} & \mathit{boundary} \end{array}\right] \end{array}\right]
\end{array}\right],
\left[\begin{array}{ll}
\mathit{Focus\text{-}Part} & \mathit{actee} \\
\mathit{Focus\text{-}Pos} & \mathit{focus\_position} \\
\mathit{FCType} & \mathit{focus\_type} \\
\mathit{List\text{-}obj} & \boxed{3}\ \mathit{list(obj)} \\
\mathit{Index} & k \\
\mathit{RIndex} & j \\
\mathit{FET\text{-}obj}\ \boxed{3} & \\
\mathit{Prosody} & \left[\begin{array}{ll} \mathit{Prosodic\text{-}Mark} & \mathit{mark} \\ \mathit{Tone} & \left[\begin{array}{ll} \mathit{Accent\text{-}Tone} & \mathit{accent} \\ \mathit{Bound\text{-}Tone} & \mathit{boundary} \end{array}\right] \end{array}\right]
\end{array}\right]
\right\}
\end{array}\right]
$$

Fig. 18. The focus information structure.

Focus-Part represents what part in a sentence is the focus part (actor, act, or actee parts). The *Index* is used to indicate its structure and *RIndex* indicates the related *Focus-Info* structure of the other focus parts. The *FET-obj* contains the groups of words which must be emphasized by tone. If the focus part is no focus then the *FET-obj* can be the same as *List-obj*.

The *Prosody* structure includes the prosodic information and refers the *FET-obj* from the *Focus-info* structure. The *Focus-Info* structure assigns which positions in a sentence must be labelled by tone marks while the *Prosody* structure declares what tone marks can be labelled at the focus position. The *Prosodic-Mark* structure shows a type of prosodic marks while the *Tone* structure represents ToBI marks as a set of accent and boundary tone marks. These tone marks are labelled for each *FET-obj*.

Mapping between prosodic marks and a set of accent-boundary tone marks is described in Section 5.2 and is shown in Table 3. The *Prosody* structure is designed to map between *Prosodic-Mark* and ToBI mark in *Tone* structures, as illustrated in Fig. 19(a). For the *Focus-Info* structure, the *FET-obj* feature contains the list of words that must be emphasized by tone. Since, the *FET-obj* links to *Prosodic-Mark* and *Tone* structure, then the prosodic marks are selected by analyzing the relationships between tone marks and speech acts code following Section 4.2. The example of this mapping is shown in Fig. 19(b). In this example, the *FET-obj* is <a,book> which is marked by the prosodic mark *Em_Lg-break*. The *Em_Lg-break* represents the emphasized accent tone with long break of boundary tone and

$$\underset{\text{(a)}}{\begin{bmatrix} \textit{FET-obj} & \textit{list(fet)} \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \textit{prosody\_mark} \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \textit{accent} \\ \textit{Bound-Tone} & \textit{boundary} \end{bmatrix} \end{bmatrix} \end{bmatrix}} \quad \underset{\text{(b)}}{\begin{bmatrix} \textit{FET-obj} & \langle \textit{a,book} \rangle \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \textit{Em\_Lg\_break} \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \mathtt{H*} \\ \textit{Bound-Tone} & \mathtt{L\text{-}L\%} \end{bmatrix} \end{bmatrix} \end{bmatrix}}$$

Fig. 19. Mapping feature structure. (a) The structure for mapping between prosodic information and accent-boundary tone marks, and (b) The example of mapping between prosodic information and accent-boundary tone marks.

$$\textit{Prosody-Mapping} \Rightarrow \begin{bmatrix} \textit{FET-obj} & \boxed{1}\ \textit{list(fet)} \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \textit{dEm\_Sh-break} \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \mathtt{L*} \\ \textit{Bound- Tone} & \mathtt{L-} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$\vee \begin{bmatrix} \textit{FET-obj} & \boxed{1}\ \textit{list(fet)} \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \textit{hEm\_Sh-break} \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \mathtt{L+H*} \\ \textit{Bound- Tone} & \mathtt{L-} \end{bmatrix} \end{bmatrix} \end{bmatrix} \vee \ldots \vee \begin{bmatrix} \textit{FET-obj} & \boxed{1}\ \textit{list(fet)} \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \textit{EmLg-break} \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & \oslash \\ \textit{Bound- Tone} & \mathtt{H\text{-}H\%} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Fig. 20. The prosodic mapping structure between prosodic information and accent-boundary tone marks referring from information in Table 3.

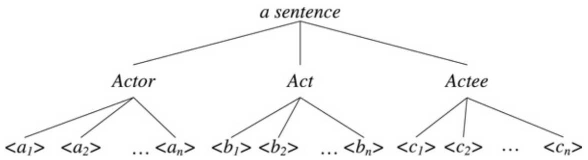

Fig. 21. Several groups of words of actor, act and actee parts.

is mapped to accent tone H* and boundary tone L-L%. In the Fig. 20, the *Prosodic-Mapping* structure represents information in Table 3.

The *FET* structure is designed to represent the relations of focus, prosodic and tone domains. The focus parts have actor, act, and actee parts. Sometimes each part can have more than a group of words as shown in Fig. 21. The groups of words are represented by $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_o$. At each part, the *FET-obj*, which is a list of groups of words, is referred to the *Prosody* structure. Following the prosodic structure in Fig. 20, only a group of words from prosodic information can be mapped to tone information. Therefore, the condition is defined to split a list of groups of words to an individual list of prosodic structures such that each structure contains only a group of words illustrated in Fig. 21. This condition is shown in Fig. 22(a). The *FET-obj* contains a multiple list of words which require splitting the list of words. As a result of splitting, the *Focus-Info* structure includes the list of prosodic structure as illustrated in Fig. 22(b).

6. An illustrative example

In this section, we illustrate our analysis by using an example. The input is the sentence "Mary bought a book" and focus is defined at the actee part. In the preprocessing step, the sentence is parsed by the LKB system with ERG. The result is the MRS representation which is transformed to the FC structure by using the FSC algorithm. The FC structure for this example is shown in Fig. 23.

In this example, the *STMood* is affirmative sentence and the *FCGroup* is *G* for the actee part following Table 2 so that the focus type is *s-focus(actee)*. Based on focus conditions, the first condition "*s*-focus of actor or actee parts" corresponds to this example. In this condition, the *FET-obj* feature is included into the AVM of this *s-focus* structure as shown in Fig. 24.

Following the FC structure, we know that the *SPCode* of this sentence is *EN0ab*. Based on the information that we have obtained, the tone pattern (2) can be chosen for this example and *SPAct* structure can be constructed as shown in Fig. 25. Since we use the *Prosody-Mapping* in Fig. 20, we obtain the *Prosody* structure for the actee part as

$$\begin{bmatrix} \textit{FET-obj} & \left\{ \langle\boxed{1}\rangle, \langle\boxed{2}\rangle, \ldots, \langle\boxed{n}\rangle \right\} \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \left\{ mark_1, mark_2, \ldots, mark_n \right\} \end{bmatrix} \end{bmatrix} \Rightarrow \begin{bmatrix} & \begin{bmatrix} \textit{FET-obj} & \langle\boxed{1}\rangle \\ \textit{Prosodic-Mark} & mark_1 \end{bmatrix}, \\ \textit{Prosody} & \left\langle \begin{bmatrix} \textit{FET-obj} & \langle\boxed{2}\rangle \\ \textit{Prosodic-Mark} & mark_2 \end{bmatrix}, \ldots, \right\rangle \\ & \begin{bmatrix} \textit{FET-obj} & \langle\boxed{n}\rangle \\ \textit{Prosodic-Mark} & mark_n \end{bmatrix} \end{bmatrix}$$

(a)

$$\textit{INFO:} \begin{bmatrix} \textit{FCType} & \textit{focus-type} \\ \textit{List-obj} & \left\langle \boxed{1}, \boxed{2}, \ldots, \boxed{n} \right\rangle \\ \textit{Prosody} & \left\langle \begin{bmatrix} \textit{FET-obj} & \langle\boxed{1}\rangle \\ \textit{Prosodic-Mark} & mark_1 \end{bmatrix}, \begin{bmatrix} \textit{FET-obj} & \langle\boxed{2}\rangle \\ \textit{Prosodic-Mark} & mark_2 \end{bmatrix}, \ldots, \begin{bmatrix} \textit{FET-obj} & \langle\boxed{n}\rangle \\ \textit{Prosodic-Mark} & mark_n \end{bmatrix} \right\rangle \end{bmatrix}$$

(b)

Fig. 22. The conditional structure. (a) The condition to split the list of *FET-obj*, and (b) The complete structure after split the list.

$$\textit{content :} \left\{ \begin{bmatrix} \textit{Focus–Part} & \textit{actor} \\ \textit{List–obj} & \langle \textit{Mary} \rangle \\ \textit{Index} & \textit{h9} \\ \textit{RIndex} & \textit{h3} \end{bmatrix}, \begin{bmatrix} \textit{Focus–Part} & \textit{actee} \\ \textit{List–obj} & \langle \textit{a, book} \rangle \\ \textit{Index} & \textit{h15} \\ \textit{RIndex} & \textit{h3} \end{bmatrix}, \begin{bmatrix} \textit{Focus–Part} & \textit{act} \\ \textit{List–obj} & \langle \textit{want,to ,buy} \rangle \\ \textit{Index} & \textit{h3} \\ \textit{RIndex} & \{\textit{h9,h15}\} \end{bmatrix} \right\}$$

Fig. 23. The focus content structure of the sentence “Mary want to buy a book”.

$$\textit{make\_s-focus:} \begin{bmatrix} \textit{s-focus} & \\ \textit{Focus–Part} & \textit{focus\_part} \\ \textit{List-obj} & \langle \textit{a,book} \rangle \\ \textit{Index} & r_3 \\ \textit{RIndex} & r_2 \end{bmatrix} \rightarrow \begin{bmatrix} \textit{s-focus} & \\ \textit{Focus–Part} & \textit{focus\_part} \\ \textit{List-obj} & \langle \textit{a,book} \rangle \\ \textit{Index} & r_3 \\ \textit{RIndex} & r_2 \\ \textit{FET-obj} & \langle \textit{book} \rangle \end{bmatrix}$$

Fig. 24. Marking the *s-focus* for actee parts.

$$\begin{bmatrix} \textit{SPAct} & \begin{bmatrix} \textit{SPCode} & \textit{EN0ab} \\ \textit{STMood} & \textit{affirmative} \\ \textit{FCGroup} & \textit{G} \end{bmatrix} \end{bmatrix}$$

Fig. 25. Speech act structure.

$$\begin{bmatrix} \textit{FET-obj} & \langle \textit{a,book} \rangle \\ \textit{Prosody} & \begin{bmatrix} \textit{Prosodic-Mark} & \textit{hEm\_Lg\_break} \\ \textit{Tone} & \begin{bmatrix} \textit{Accent-Tone} & H* \\ \textit{Bound-Tone} & L\text{-}L\% \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Fig. 26. Prosodic Structure for the actee part <a, book>.

illustrated in Fig. 26. This prosodic structure is embedded into the FC structure and this embedded structure is called the *Focus-Info*, which is illustrated in Fig. 27.

The *SPAct*, and *Focus-Info* structures are integrated to generate the FET structure including a set of tone marks annotated at the actee part “a book”. The complete FET structure is illustrated in Fig. 28.

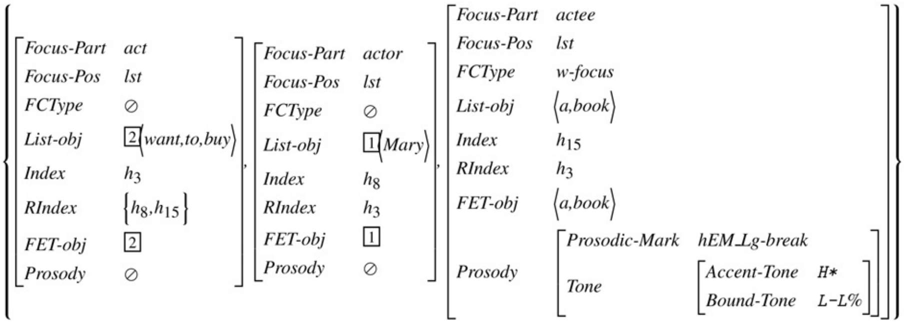

Fig. 27. Focus-Info structure for "Mary want to buy a book".

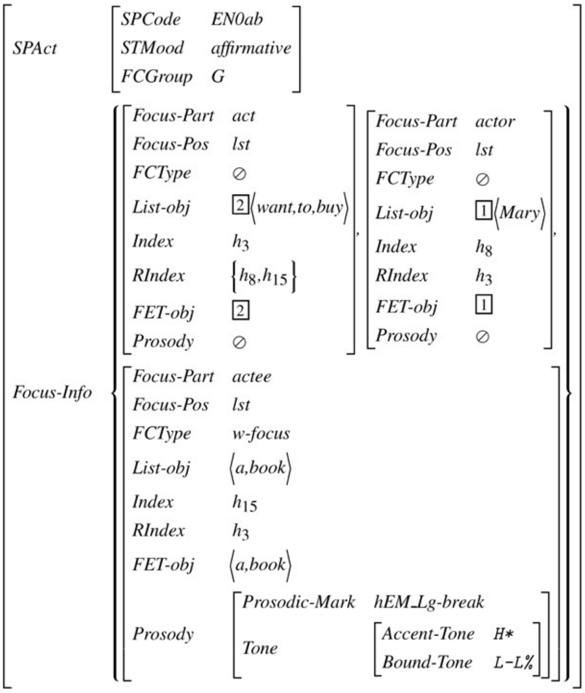

Fig. 28. The information structure of the sentence "Marry want to buy a book".

7. The FET implementation with the LKB system

The FET analysis is implemented by using the LKB system with our FET subgrammar for English language. For our implementation, the FC structure is used to generate a set of focus words. We provide the focus words into the LKB system with the FET subgrammar. This subgrammar consists of the FET type hierarchy, constraints, rules, and feature structures of the focus and prosody described in Section 7.1. Since the LKB system with FET subgrammar can analyse the focus relations corresponding to speech acts and sentence types, the system completes the FET structure by generating the appropriate prosodic structures containing prosodic labels.

7.1. The FET subgrammar for the LKB system

The FET subgrammar is composed of (i) focus words, (ii) the FET typed hierarchy, and (iii) focus and prosodic structures including their constraints and the FET rules. In the preprocessing stage (see Section 3), the MRS representation of the example sentence "Mary wants to buy a book" is transformed to the FC structure which is shown in Fig. 23. For our implementation, this FC structure and its information are used to generate a set of focus words for the LKB system. An example of the focus word (*focus-word*) structure is shown in Fig. 29.

```
book := focus-word &
[ ORTH  "book",
  HEAD  actee-part &
        [ AGRS 1s − actee_G − aff − en0ab],
  SPR   < HEAD actee & [ AGR1 pv − actee_G − aff − en0ab] >,
  COMPS<>].
```

Fig. 29. Focus word "bought".

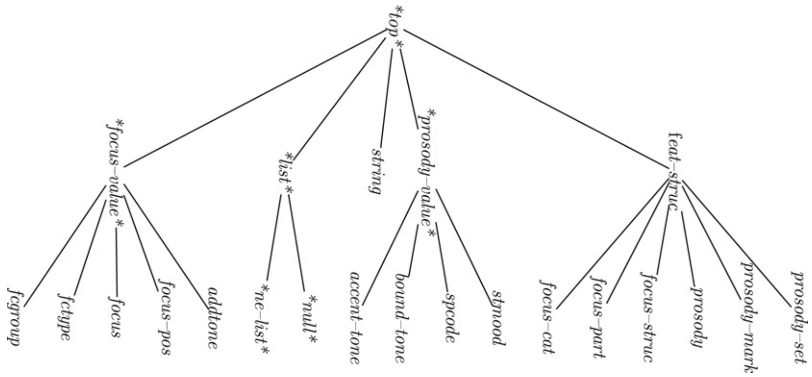

Fig. 30. FET type hierarchy.

Each *focus-word* consists of four main features: ORTH, HEAD, SPR, and COMPS. ORTH represents the orthography of a word. HEAD structure consists of the focus part (*actor-part*, *act-part*, or *actee-part*), and the feature structure ARGS which contains the properties of current word or phrase including speech act code, focus group, and sentence type. SPR structure represents the properties of the previous word while COMPS contains the properties of the following phrase or word.

The FET typed hierarchy

The FET typed hierarchy for the LKB system is illustrated in Fig. 30, we divide the features in our system to three main groups: **focus-value**, **prosodic-value** and *feat-struc*. These features are used to control the focus and prosodic constraints. In the first group, **focus-value** is classified into five subfeatures: focus criterion (*fcgroup*), focus type (*fctype*), focus name (*focus*), and checking whether a tone mark can be marked on the word (*addtone*). In the second group, the **prosody-value** consists of four subfeatures which are sentence mood (*stmood*), speech act code (*spcode*), accent tone (*accent-tone*), and boundary tone (*bound-tone*). In the last group, *feat-struc* is divided to six subfeature structures which are described below:

(1) *focus-cat* is used to constrain the focus components.
(2) *focus-part* is classified into act part, and non-act part (such as actor or actee part).
(3) *focus-struc* is a subfeature of focus word and focus phrase. It contains the focus and prosodic feature structure of the current word and its related words or phrase.
(4) *prosody* contains prosodic components such as prosodic marks.
(5) *prosodic-mark* maps between types of prosody and the combination of accent and boundary tone marks.
(6) *prosodic-set* contains a set of *prosodic-mark* structures.

Focus and prosodic structures

The *focus-phrase* structure, which is illustrated in Fig. 31(a), contains the *focus-struc* subfeature structure and ARGS feature. The ARGS represents the focus properties of a words. In Fig. 31(b), the *focus-word* structure inherits the *focus-struc* structure with *ORTH*. In our FET system, a set of *focus-words* is generated from the FC structure in Section 3. The *focus-struc* is composed of HEAD referred to the *focus-part* structure, *SPR*, and COMPS. The *focus-struc* structure is shown in Fig. 31(c). The *focus-part* contains the focus information, which includes *focus* and *focus-cat* structures as shown in Fig. 31(d). The *focus-cat* is used to constrain focus and prosodic features for each focus part. *focus-cat* is composed of *focus-pos*, *fcgroup*, *fctype* and *addtone* and the *prosody* structure. The *focus-cat* structure is illustrated in Fig. 31(e).

$$\textit{focus-word} := \textit{focus-struc}\ \& \begin{bmatrix}\text{ARGS} & \ast\textit{list}\ast\end{bmatrix} \quad \textit{focus-word} := \textit{focus-struc}\ \& \begin{bmatrix}\text{ORTH} & \textit{string}\end{bmatrix}$$

(a) (b)

$$\textit{focus-struc} := \textit{feat-struc}\ \& \begin{bmatrix}\text{HEAD} & \textit{focus-part}\\ \text{SPR} & \ast\textit{list}\ast\\ \text{COMPS} & \ast\textit{list}\ast\end{bmatrix} \quad \textit{focus-part} := \textit{feat-struc}\ \& \begin{bmatrix}\text{FOCUS} & \textit{focus}\\ \text{ARGS} & \textit{focus-cat}\end{bmatrix}$$

(c) (d)

$$\textit{focus-cat} := \textit{feat-struc}\ \& \begin{bmatrix}\text{FOCUS-POS} & \textit{focus-pos}\\ \text{FCGROUP} & \textit{fcgroup}\\ \text{FCTYPE} & \textit{fctype}\\ \text{ADDTONE} & \textit{addtone}\\ \text{PROSODY} & \textit{prosody}\end{bmatrix}$$

(e)

Fig. 31. Type feature structures of: (a) *focus-phrase*, (b) *focus-word*, (c) *focus-struc*, (d) *focus-part*, and (e) *focus-cat*.

$$\textit{prosody} := \textit{feat-struc}\ \& \begin{bmatrix}\text{STMOOD} & \textit{stmood}\\ \text{SPCODE} & \textit{spcode}\\ \text{PROSODY-MARK1} & \textit{prosody-mark}\\ \text{PROSODY-MARK2} & \textit{prosody-mark}\end{bmatrix}$$

(a)

$$\textit{prosody-mark} := \textit{feat-struc}\ \& \begin{bmatrix}\text{ACCENT-TONE} & \textit{accent-tone}\\ \text{BOUND-TONE} & \textit{bound-tone}\end{bmatrix}$$

(b)

Fig. 32. Type feature structures of: (a) *prosody*, and (b) *prosody-mark*.

```
[ focus-word
  HEAD: actor-part
           FOCUS: ACTOR
           ARGS: [ ls-actor_g-aff-en0ab
                    FOCUS-POS: LAST
                    FCGROUP: G
                    FCTYPE: W-FOCUS
                    PROSODY: [ aff-en0ab-no-prosody
                                  STMOOD: AFF
                                  SPCODE: EN0AB
                                  PROSODY-SET: [ no-prosody
                                                 PROSODY-MARK1: [ no-mark
                                                         ACCENT-TONE: NOACCENT
                                                         BOUND-TONE: NOACCENT ] ]
                                                 PROSODY-MARK2: [ no-mark
                                                         ACCENT-TONE: NOACCENT
                                                         BOUND-TONE: NOACCENT ]
  ] ] ] ] ] ] ] ] ]
  SPR: *NULL*
  COMPS: *NULL*
  ORTH: Kim ]
```

Fig. 33. FET structure of the word "Mary".

The *prosody* structure consists of *stmood*, *spcode*, and *prosody-set,* which contains a set of *prosody-marks* structures. The *prosody* constrains the prosodic features depending on the relationships of speech acts and prosodic patterns in Section 4.2. The *prosody* structure is shown in Fig. 32(a). Each *prosody-mark* structure contains the combinations of *accent-tone* and *bound-tone* as shown in Fig. 32(b).

We define two types of the FET rules, which are head-complement and head-specifier rules. These rules are similar to the HPSG grammar rules, explained in in [13]. Using these rules, the example sentence can be parsed and its result is a complete FET structure, including the focus and prosodic information. For example, the FET structure for the LKB system is shown in Fig. 33.

Table 4
Summary of the evaluation with and without tone mark's alignment

	Section 1		Section 2	
	Sum of words	Percentage	Sum of words	Percentage
Matched tone marks	263	57.80	342	75.16
Unmatched tone marks	192	42.20	113	24.84
Total no. of words	455		455	

Table 5
Comparison the annotations of sentence "Where are you leaving from" between the FET system and the CMU–COM dataset

Sentence	Where	are	you	leaving	from
FET	0	0	L+H*	H*	L-L%
CMU	0	0	0	H*	L-L%

8. Evaluation of prosodic annotation

In this evaluation, the prosodic annotation by the FET system is compared with the annotation dataset from the CMU–COM, which is the reference dataset used to measure performance. The sentences in this dataset are derived from the travel reservation dialogues. One hundred sentences with their ToBI annotations are collected for the evaluation. Parsing the sentences using the LKB system is limited by the number of words in the lexicon and grammar rules of the ERG. In our experiment, sixty-one sentence of one hundred sentences can be parsed by the LKB system using the ERG that provides the MRS representations as the results. The MRS representations are transformed to a set of focus words for each sentence and then these 61 sentences are parsed by the LKB with the FET subgrammar. The results are the sentences with ToBI annotation on each word.

There are two main sections in this evaluation: (i) the evaluation without tone mark's alignment and (ii) the evaluation with the tone mark's alignment. The reason for using tone mark's alignment is to match the tone marks between two systems. Although the tone annotation by the FET system looks different from the annotation of the CMU–COM, by only shifting the accent tone marks to the previous word, both annotations govern the same pitch contour and same boundary, and the sounds after modifying prosody are not different. The result is calculated by dividing the number of words with matching annotations by the total number of words. The total number of words is 455, in the 61 sentences. Summary of this evaluation is shown in Table 4. For the first section, the percentage of matched tone mark is 57.8 percents, comparing between the tone annotation from the CMU–COM dataset and the annotation by the FET system. For the second section, the percentage of matched annotation is equal to 75.16 percents. Considering the unmatched results in Table 4, approximately 25 percents of tone marks are not matched. After examining the errors, we found that most of the unmatched labels occur in a group of prepositions (PP), pronouns (PR), and determiners (DET). Many DET, PR, and PP in the CMU–COM are not marked with any annotation. The FET system has more sensitive tone marking on the PP, PR, and DET than the CMU–COM dataset. For example, the tone annotations of the sentence "Where are you leaving from" are shown in Table 5. A difference is at the word "you". The annotation by FET system has tone mark H* while there is no tone mark in the CMU–COM dataset. Since the FET system considers "you" to be an actee part that needs a focus, "you" is marked with the emphasized tones H* or L+H*. On the other hand, "you" is not considered as the theme of sentence for the CMU–COM dataset so this word has no tone mark.

9. Conclusion

We investigate the relationships between focus inferred from speakers intention, and prosody in speech language generation. Our method begins by transforming the semantic information, derived from the LKB system with ERG, to the FC structure representing the focus information. This focus information is used for focus and prosodic analysis depending on the speaker's intention. After this analysis, the FET structure contains the information of focus and prosody including prosodic marks. In our implementation, the FET subgrammar for the LKB system is constructed as a small parser to annotate the prosodic marks based on focus and speaker's intention information. The FET subgrammar is the core structure and it contains the FET structure, grammar rules, focus words and so on. This subgrammar analyzes the relationships of focus and speaker's intention to find the intonation patterns. Our approach is limited by our use of LKB with ERG. The parser's performance depends on the number of grammar rules and lexicon entries of the ERG. The FET system as the unification-based subgrammar now works separately from the LKB system with ERG. The FET analysis is evaluated using a number of sentences in travel reservation domain from a part of CMU–COM dataset. Our experiment is a domain-dependent implementation such that we control the number of focuses, speaker's intentions and prosodic features. The result of parsing a sentence using the LKB system with the FET subgrammar is a complete FET structure in which the words of the sentence are annotated with the tone marks.

Acknowledgement

This work is supported by NSERC, Canada and Royal Golden Jubilee Ph.D. programme, Thailand Research Fund, Thailand.

References

[1] A. Copestake, D. Flickinger, I. Sag, C. Pollard, Minimal recursion semantics: An introduction, Journal of Research on Language and Computation 3 (2–3) (2005) 281–332.

[2] A. Copestake, Implementing Typed Feature Structure Grammars, CSLI Publisher, Stanford, CA, 2002.

[3] K.R. McKeown, S. Pan, Prosody modeling in concept-to-speech generation: Methodological issues, Philosophical Transactions of The Royal Soceity, Series A: Mathematical, Physical and Engineering Sciences 358 (1769) (1999) 1225–1431.

[4] A.H. Oh, A.I. Rudnicky, Stochastic natural language generation for spoken dialog systems, Computer Speech and Language 16 (2002) 387–407.

[5] M. Theune, Contrast in concept-to-speech generation, Computer Speech and Language 16 (2002) 491–531.

[6] E. Klien, Prosodic constituency in HPSG, in: R. Cann, C. Grover, P. Miller (Eds.), Grammatical Interfaces in HPSG, CSLI Publications, 2000, pp. 169–200.

[7] M. Haji-Abdolhosseini, A constraint-based approach to information structure and prosody correspondence, in: S. Müller (Ed.), Proceedings of The HPSG-2003 Conference, CSLI Publications, 2003, pp. 143–162.

[8] A. Copestake, D. Flickinger, An open-source grammar development environment and broad-coverage English grammar using HPSG, in: Proceedings of the Second Conference on Language Resources and Evaluation, LREC-2000, Athens Greece, 1997.

[9] C. Gussenhoven, The Phonology of Tone and Intonation (Research Surveys in Linguistics), Cambridge University Press, 2004.

[10] T.T. Ballmer, W. Brennenstuhl, Speech Act Classification: A Study of the Lexical Analysis of English Speech Activity Verbs, in: Series in Language and Communication, vol. 8, Springer, Berlin, 1981.

[11] L. T. I. at Carnegie Mellon University, CMU communicator KAL limited domain, http://www.festvox.org, October 2005.

[12] K. Silverman, M.E. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, J. Hirschberg, ToBI: A standard for labelling English prosody, in: Proceedings of the 1992 International Conference on Spoken Language Processing, ICSLP'92, vol. 2, Banff, Canada, 1992, pp. 867–870.

[13] I.A. Sag, T. Wasow, E. Bender, Syntactic Theory: A Formal Introduction, 2nd ed., CSLI Publisher, University of Chicago Press, 2003.