

Appendix A

Part-of-speech Tagging Program Manual

Tagging program is developed to assign the knowledge of the part of speech and decision classifier to the keywords that should be extracted. Part Of Speech tag set (POS) is derived from ORCHID corpus (Sornlertlamvanich et al., 1997). The tag set contains fourteen categories with forty seven subcategories. To assign decision classifier, we extend three categories of tag set (which is named as “POD”) to the original POS tag set from the ORCHID. These three categories are “criterion”, “alternative” and “other” categories.

Main screen of tagging program is designed to tag Thai discussion data from various sources such as group meeting, online discussion system, webboard and so on. We also designed and created the database for keeping tagging information by using Microsoft Access. Nine tables are included. The relationship of the database design is illustrated in Figure A.1.

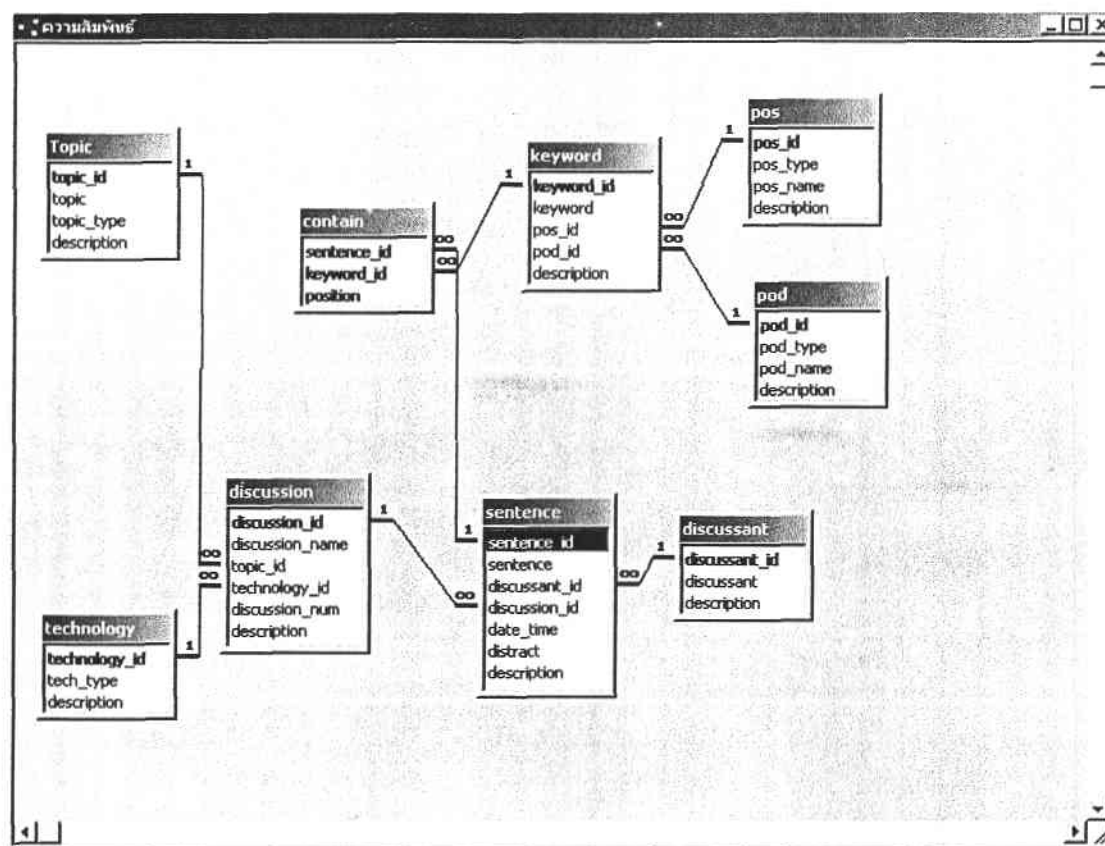



Figure A.1 Database Design of POS Tagging Program

In the tagging process, we develop Visual Basic program to collect the tagging information, which are POS type of words and POD type of word. This program reads sentences from the database and displays them on the user’s screen. Then, the program will segment sentence into a sequence of words for Thai data. After this process, our program allow user to tag type of words with the statistic information for support their decision. The steps of using POS tagging program are shown as follows:

Getting start with POS tagging program

To start POS tagging program:

1. Prepare the database connection by using ODBC (named “postag”).
2. Double click  to start POS tagging program

After these steps, main screen of POS tagging program will be opened as illustrated as in Figure A.2.

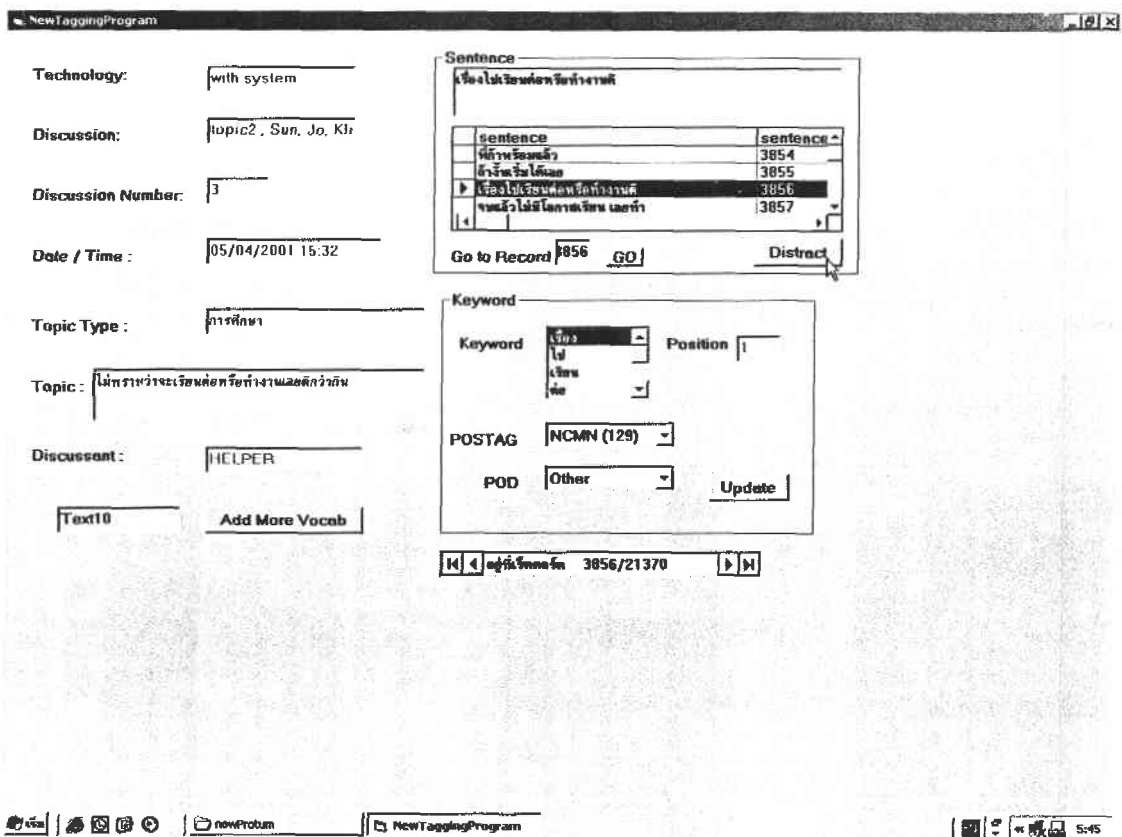


Figure A.2 POS Tagging Program when Connection is Successful

To work with POS tagging program:

After open and connect POS tagging program to database successfully, we can tag type of words by following these steps:

1. Select “**sentence**” that you want to tag.
2. Select “**keyword**” that you want to identify type.
3. Select “**POSTAG**”.

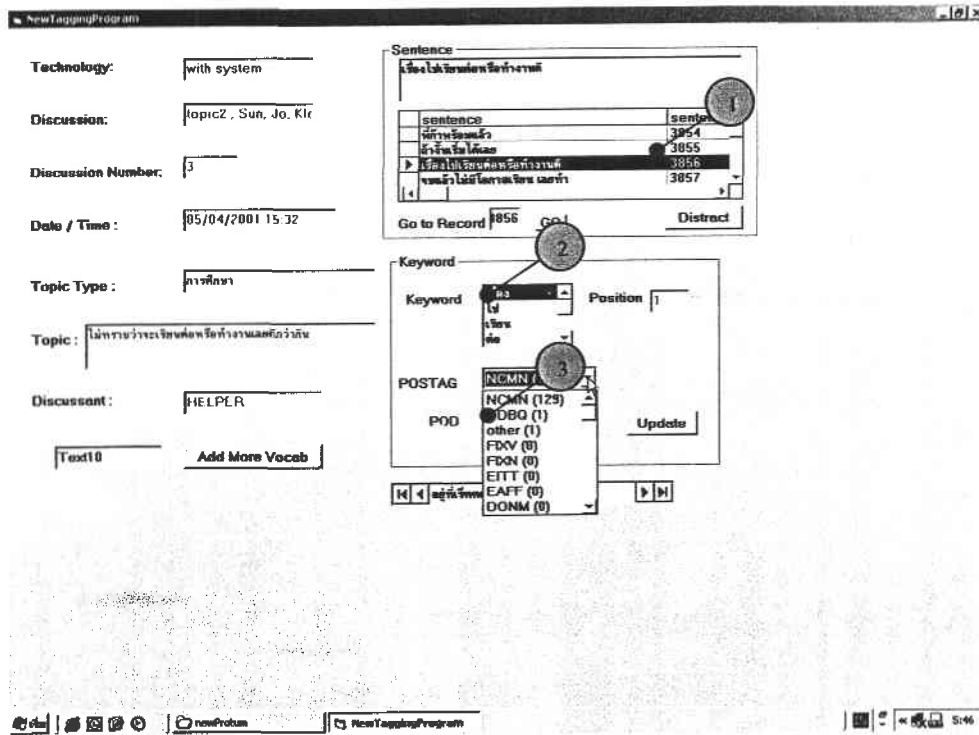



Figure A.3 POS Tagging Program after Selecting Words.

4. Select “**POD**”.
5. Click  to save data into database.

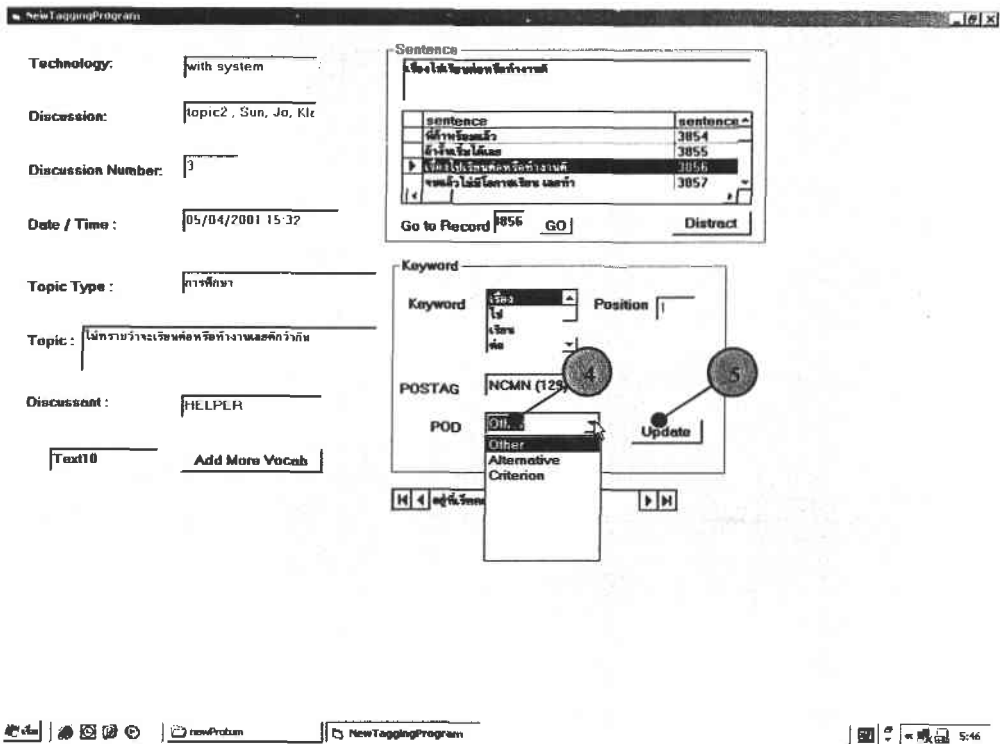


Figure A.4 POS Tagging Program after Selecting POS tag

If the word that we are tagging has already been tagged and does not need any changes, POS tagging program will display the screen as illustrated in Figure A.5. Otherwise, the program will display the successful screen as illustrated in Figure A.6.

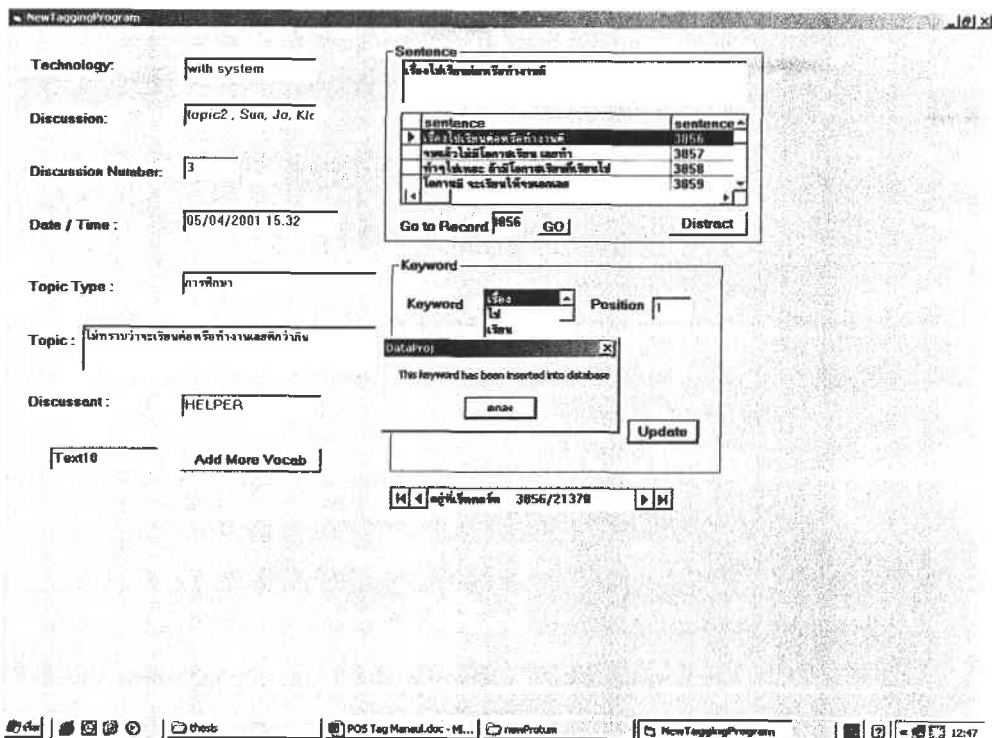


Figure A.5 Warning Screen if Duplication Occurs.

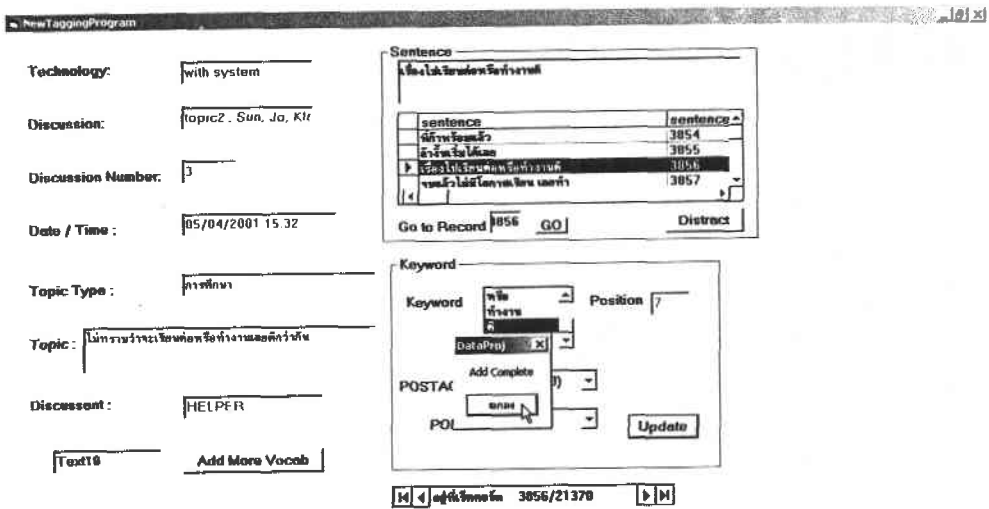


Figure A.6 Confirm Screen if Saving Data is Successful

To add word into dictionary of POS tagging program:

Sometime, our word segmentation algorithm for segmenting sentence into a sequence of words is not correct. Therefore, this program also has a feature of adding words to dictionary. The steps of adding words to dictionary are as followed:

1. If user found incorrect segmentation, go to **Text10**

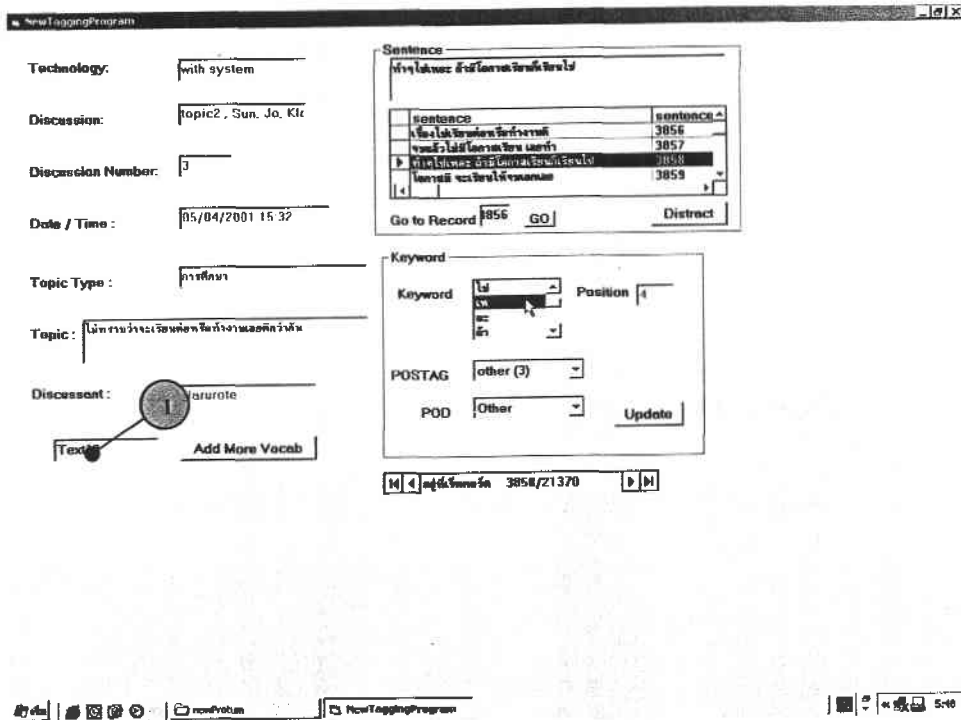


Figure A.7 Example of Incorrect Word.

2. Fill in the correct word and click **Add More Vocab**

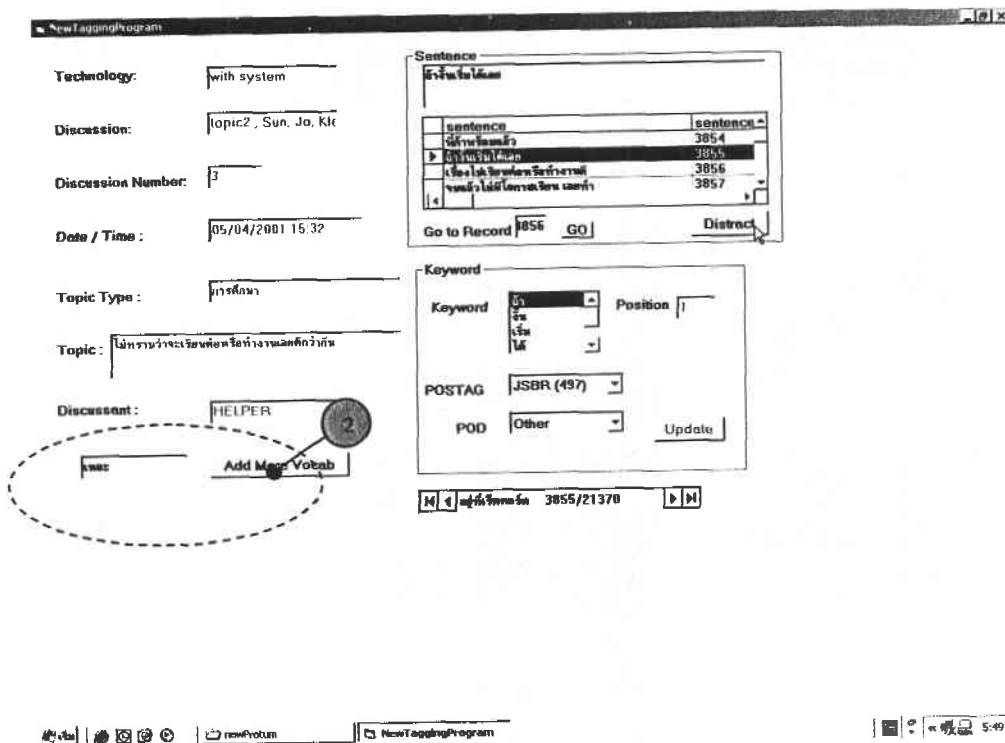


Figure A.8 Example of Filling Correct Word

New word will appear on a sequence of words.

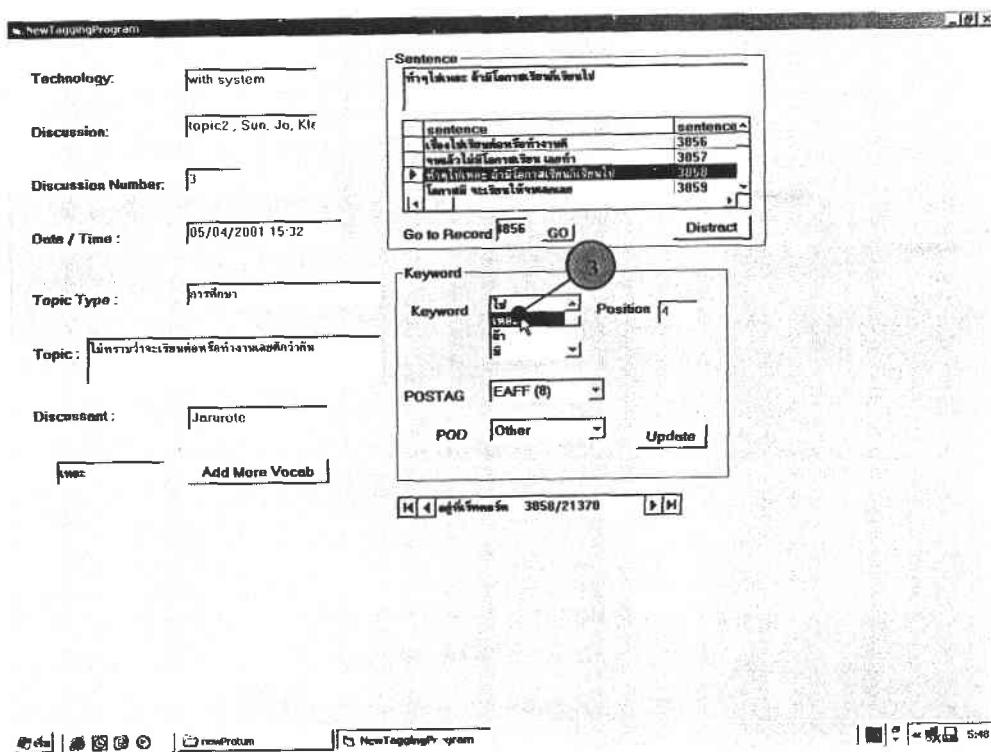


Figure A.9 POS Tagging Screen after Adding Word to Dictionary

To use GO TO of POS tagging program:

1. Fill the sentence number that you want to go to at **Go to Record** | 50 |
2. Click **GO**

The program will bring you to the sentence position.

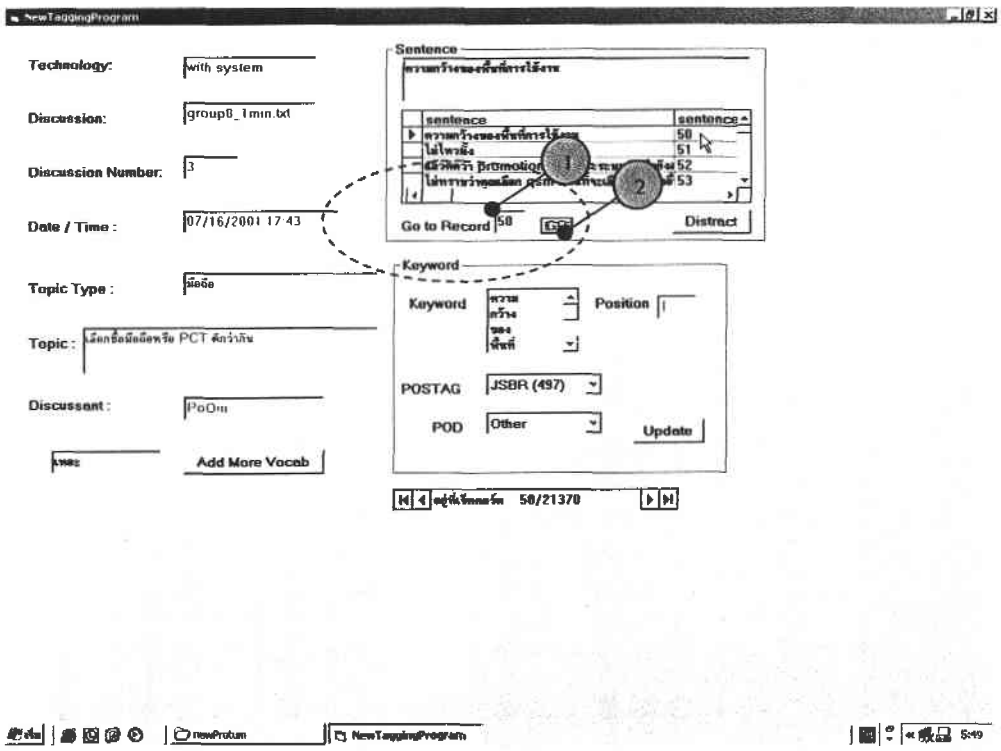


Figure A.10 POS Tagging Screen for “GO TO”