Part of Speech Annotator

1. Introduction

Annotator is an annotation tool to facilitate the process of manually tagging the text for part of speech (POS). The tool is supported for windows and Microsoft .NET Framework Version 2.0 (http://www.microsoft.com/downloads/details.aspx?FamilyID=0856EACB-4362-4B0D-8EDD-AAB15C5E04F5&displaylang=en) is required to run it.

2. Annotator Tool

This section describes basic features of Annotator. The Annotator program has three components: *Annotator.exe*, *tagset.txt* and *Memory* folder. Annotator.exe is the main program. The tagset.txt file contains the tagset that will be used to annotate the text. The 1st line of *tagset.txt* contains the total number of tags in the file and remaining each line contains a single tag. Tagged files can be copied to *Memory* folder and these will guide the human annotator while tagging the text.

2.1 Using Annotator

Double click the Annotator.exe to run the program. Click File→Open to open a text file for annotation.

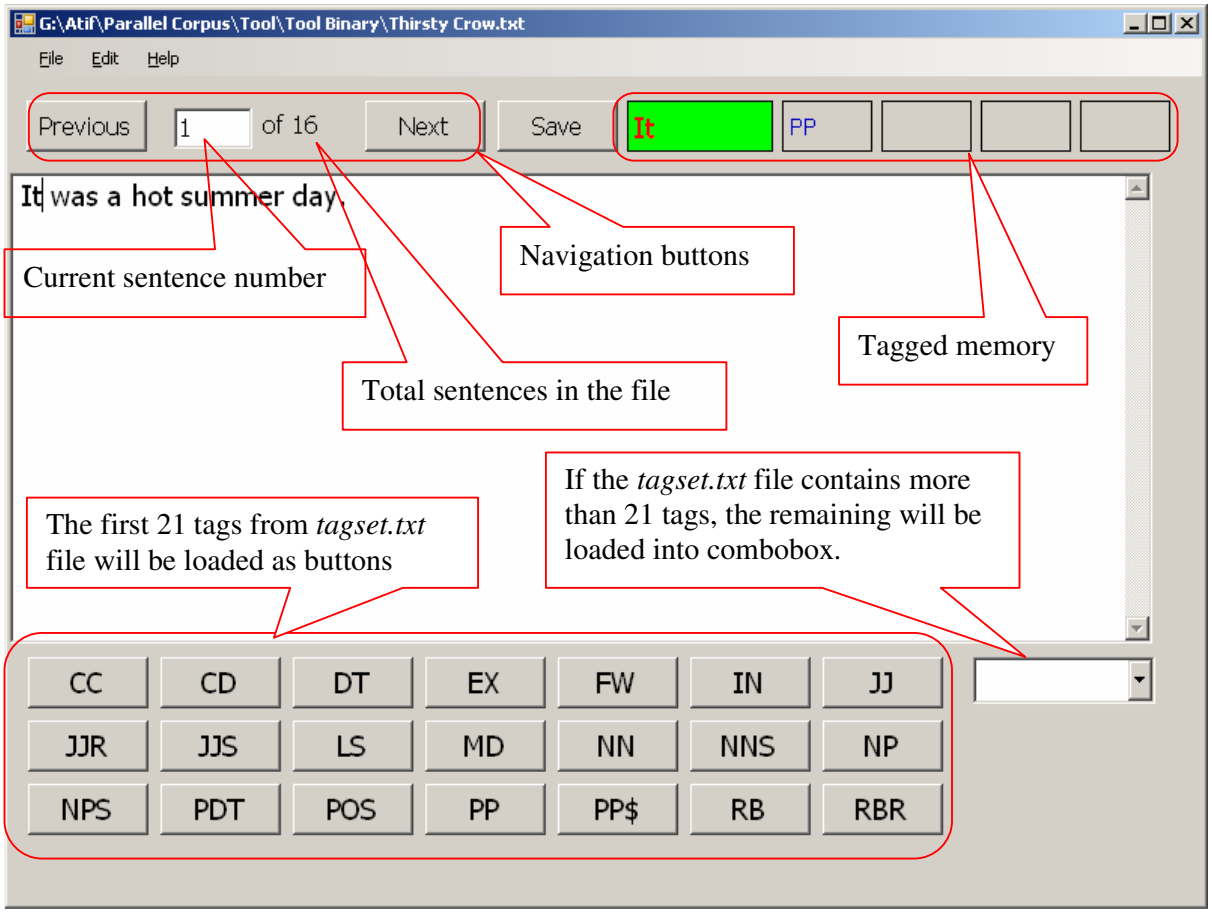

Figure 1: Using Annotator

The Annotator will show the 1st sentence of the opened file as shown in Figure 1. The file can be traversed using navigation buttons: Previous and Next; or by entering the sentence number in the current sentence number text field.

The first 21 tags from *tagset.txt* file will be loaded as buttons and if the *tagset.txt* file contains more than 21 tags the remaining will be loaded into combobox.

2.2 Tagged Memory

Tagged files can be copied to *Memory* folder and these will guide the human annotator while tagging the text. Tagged memory shows the current selected word and the tag/s assigned to it in the *memory* files. Double click on the proposed tag for the current word to see its context in *memory* files as shown in Figure 2.

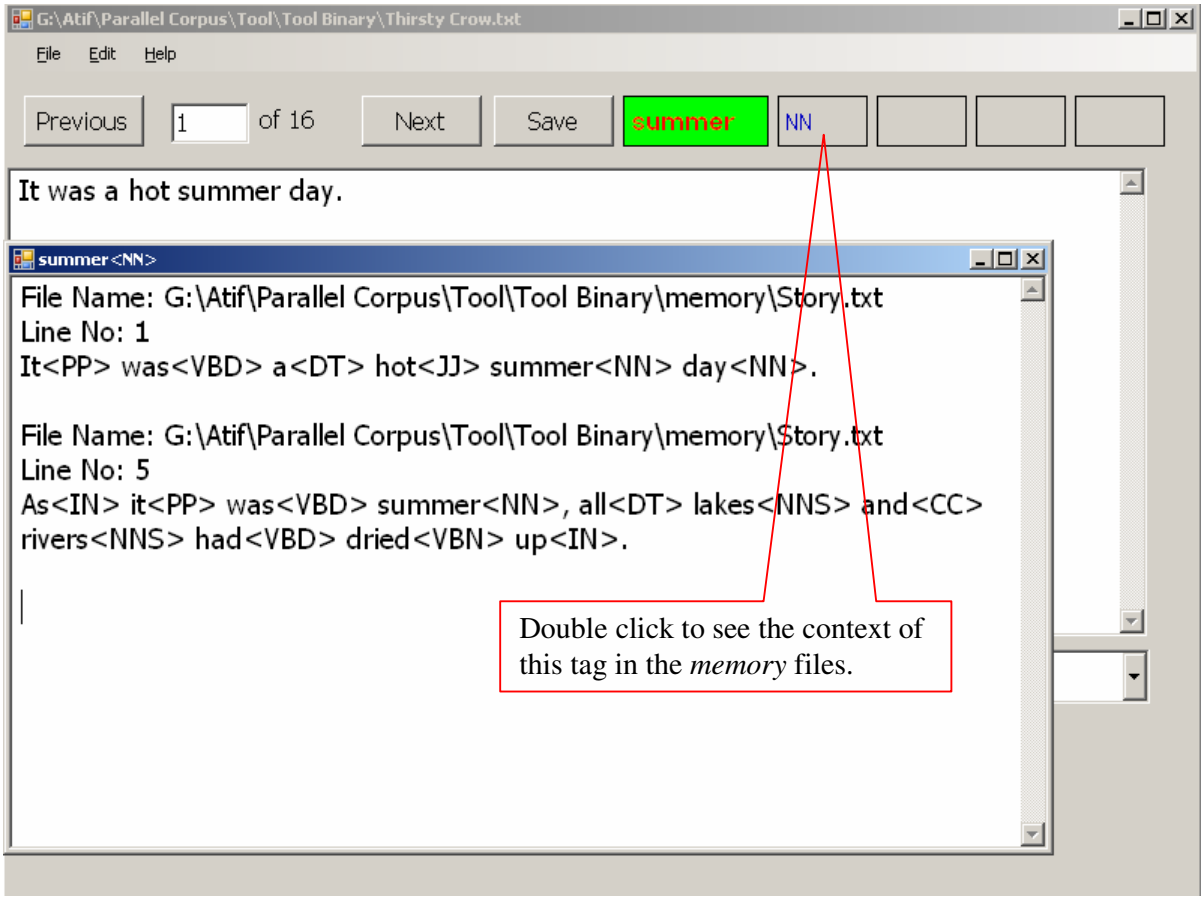

Figure 2: Using tagged memory

Acknowledgement:

The tool is developed under the project of "Parallel Corpus Extension for Urdu and Nepali". The project is funded by Asian Language Resource Network (ALRN) and PAN Localization Project (www.panl10n.net)