

# Treebank To XML

## Reference Manual

*Radboud University Nijmegen,  
Centre for Language Studies  
Erwin R. Komen  
Version 1.0 – September 30, 2011*

## Contents

1	Introduction .....	1
2	Installation and setup.....	1
2.1	Requirements .....	1
2.2	Installation of TreebankToXml .....	1
2.3	Using TreebankToXml for the first time .....	2
3	Batch conversion .....	2

## 1 Introduction

The “TreebankToXml” program is a simple batch converting program. The program “Cesax” is able to import “psd” files (treebank files in bracketed labelling format), and save them as “psdx” files (a TEI-P5 compliant variant of xml). The program “TreebankToXml” basically does the same thing, but then as a batch converter.

## 2 Installation and setup

### 2.1 Requirements

The current version of TreebankToXml takes the following assumptions:

- You are working on a computer with the Windows XP, Vista or Windows7 operating system. TreebankToXml has been developed on an XP computer, so if you have any problems under Vista or Windows7, try and see if these problems disappear when you run TreebankToXml on an XP computer.
- You are working with **psd** files from one of the following corpora as input: YCOE (Old English), PPCME2 (Middle English), PPCEME or PCEEC (Early Modern English), PPCMBE (Late Modern English).
- Alternatively, TreebankToXml has also been tested on converting files from the Wall Street Journal (WSL) corpus.

### 2.2 Installation of TreebankToXml

The TreebankToXml program is freely available from the homepage of Cesax, which you can reach through the [software](#) page of the Radboud University’s “Language in Time and Space” sub programme. Installation and setup is as follows:

- a) Go to the [software page](#) of the Radboud University’s LiTS sub programme.
- b) Navigate to the Cesax homepage from there.
- c) Choose “Install TreebankToXml”, which will lead you to the actual “publish” page.
- d) For the first time, choose “Install”. Otherwise, choose “launch”.
- e) Follow the instructions in the setup program
- f) The software will be available under Start/ProgramFiles/RU-English

You now need to adjust the input and output directories.

### 2.3 Using TreebankToXml for the first time

The TreebankToXml program needs to be set up correctly for it to work to your satisfaction. When you start up TreebankToXml for the first time, the program will automatically produce a “settings” file. You don’t need to worry about this file normally—it will be installed automatically. But you should not delete it afterwards, because if you do, you may lose some of your settings.

- A settings file called TreebankToXmlSettings.xml is put in a new “TreebankToXml” subdirectory of your “ApplicationData” folder (probably on your C-drive).
- One entry is added into the registry of your Windows computer, which indicates the location of the TreebankToXmlSettings.xml file.

What is kept in the TreebankToXmlSettings.xml file basically is the last directories you have been using.

### 3 Batch conversion

The batch conversion goes this way:

- Define the input and output extensions you are using. For the parsed corpora of English the input extension should be “.psd”. Don’t forget to put the **dot** before the extension!!
- Define the input directory in which the “.psd” files (or other files) are located:
- If you are converting Wall Street Journal, check that box.
- Define the output directory where you want your converted files to show up.
- Select **T**ools/**C**onvert or press F3 for the conversion to start.