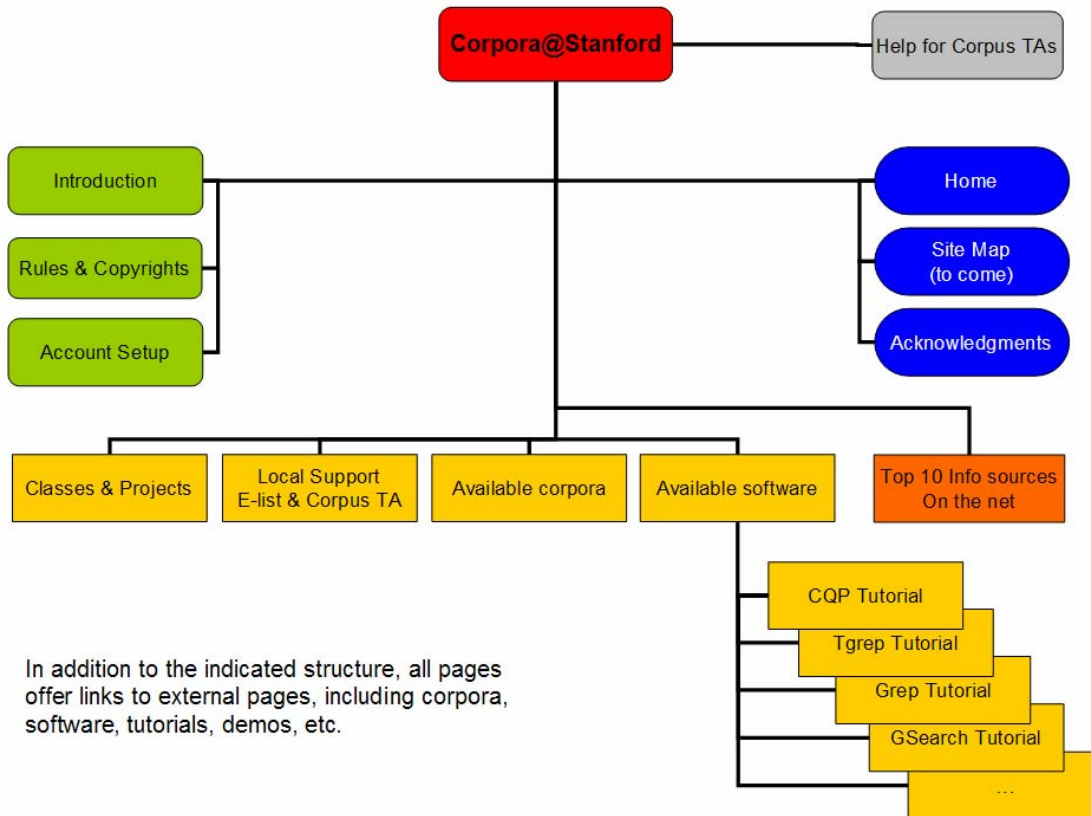


## 1 Introduction

The main purpose of this handout is to give you an overview of the resources available in our department. This handout is intended to be a quick reference to some of the existing tools and the corpora available for them. The number one source of information for Corpora@Stanford is our new website at:

<http://www.stanford.edu/dept/linguistics/corpora/>



## 2 The Corpus Computer

As you may have noticed, our department has a computer which is reserved and specifically configured for doing corpus-based research. It is located in the computer cluster (the computer closest to the printer). Beside the 4 administrator accounts (Allen, Roger, Florian & Susanne), there are two more accounts.

### Accounts on the Corpus Computer

- The *Ling Grads & Faculty* account (password: *ling2003*) is intended for your personal use. Please store any kind of files that you create in the "Shared documents" folder under a directory *YOUR\_NAME*.
- The *Ling Classes* account (password available on request) can be used if you are teaching/TAing a class and you want them to work on corpora. Please store any kind of files that you create in the "Shared documents" folder under a directory *CLASSES/CLASS\_NAME*. Please don't use the *Ling Grads & Faculty* account for this.

**If you want to do a lot of work on the Corpus Computer and you would like to have your own account on it, this can be arranged. Just let the Corpus TA know.**

When you login in, you will see information & guidelines for the use of the corpus computer that we asked you to consider. This will minimize the necessary maintenance.

Most of the **corpora owned by the department** are stored on AFS (a system of shared network drives). A list of the available corpora and their location (i.e. path on AFS) can be found at

<http://www.stanford.edu/dept/linguistics/corpora/cas-corpora.html>

Note that our department owns many more corpora than the ones on AFS. If what you need is not there talk to us - we can always **order corpora** (a lot of them for free due to a membership in LDC).

There are **two ways to get to the corpora on AFS**. First, you can login into *saga* or some other machines using an SSH software (e.g. Samson, Tera Term). This works in every operating systems (Windows, Unix, MacOS). Once you are logged in and have followed the information on:

you can go to the path where most of our installed corpora can be found.

`/afs/ir/data/linguistic-data/`

If you use the corpus PC, there is a **second way to get to AFS for the friends of windows and easy navigation**.

- Enter your account on the corpus PC (e.g. *Ling Grads & Faculty*)
- Log into PC-Leland
- A short time after that, the little lock-symbol at the right bottom corner of the screen will switch from being crossed to being un-crossed. This indicates that you have been authorized to use AFS.
- Click the icon *Corpora on AFS* (a link to the mapped network drives) and you are there! You will see all the directories which are also mentioned on Susanne's list of the available corpora (together with a short description of their content).
- Navigation in this folder works as everything else in windows.
- If you want **to view a file**:
  - o right-click on it with the mouse
  - o choose *Open with UltraEdit*
  - o When you are asked if the line-breaks should be converted into MSDOS mode, click *Yes*.

No matter which of the two ways to access AFS you choose, **don't be afraid** of deleting files or destroying something – although we asked you not to try it, you should not be able to do any harm to the system.

We take it for granted that **by using any of the corpora you agree to the license conditions** – feel free to write us if you have any questions about that. Please also have a look at the our page on license conditions:

<http://www.stanford.edu/dept/linguistics/corpora/cas-rules.html>