

VARD 2.5 Basic Instructions

VARD Website: <http://ucrel.lancs.ac.uk/vard>

VARD User guide: <http://ucrel.lancs.ac.uk/vard/userguide>

VARD is available for academic use, and can be requested at:

<http://ucrel.lancs.ac.uk/vard/availability>

Starting VARD 2.5

1. In download location provided by the instructors, follow "VARD 2.5.4" to download the latest version of VARD. *VARD2.5.4.zip* will be downloaded to your computer. This zip file should be copied to an easily accessible location on your system and unzipped, a folder named *VARD2.5.4* will be created.
2. To start VARD 2.5 please open *run.bat* (Windows), *run.command* (Mac OS) or *run.sh* (Unix). If none of those work, you can open "gui.jar" directly (although VARD may run more slowly). Alternatively, open a command prompt / terminal window, locate the *VARD2.5* folder and enter the command:

`java -Xms256M -Xmx512M -jar gui.jar`

Manually normalising a single text

1. Open a text to process by clicking on *File>Open* and choose a text from the *Lampeter corpus* (in the corpora folder) or from your own corpus if you wish.
2. Detected spelling variants will be highlighted in yellow. Choose one of the variants highlighted which you agree should be considered a spelling variant, *right click* this word. Observe the normalisation options given (if any) and the confidence scores also given. If present, select the correct normalisation and *Normalise instance*, or *Normalise all* if you believe all instances of this variant should be normalised to the selected form.
3. Select a highlighted word which you disagree to being a variant, *right click* this word. Select *Mark all as Not variant* and if you agree the word should never be marked as a variant you can choose *Yes* to adding the word to the dictionary.
4. Select a non-highlighted word which you believe should be marked as a variant, *right click* this word. Select *Mark all as variant*. If you believe this word should always be marked as a variant you can choose *Yes* to removing the word from the dictionary.
5. The entire text can be processed in this manner. You can choose to normalise types using the *Types List* or use *Step Complete* to sequentially process the text or types. You can also *Auto Normalise* the text with a threshold set which each variant's top normalisation suggestion's confidence score must reach to be normalised.
6. Save the text and training data using *File>Save All*, select *Save changes*. Select *Save with XML tags* and select a suitable location to save the normalised text (it is not advisable to overwrite the original file).

Change Training

1. Some training for the Lampeter texts on the VARD download page. This is only a small amount of training from manually normalising a small sample of Lampeter. Download and unzip in your VARD2.5 folder.
2. If you've done a substantial amount of processing you'll probably want to make sure you've saved your changes. You will be prompted to do this when changing the setup folder.
3. In the *Setup* sidebar, select *Options*, in *Training Folder* use *browse* to locate the "lampetertraining" folder you've just downloaded (or just drag the folder to the textbox).
4. Click 'Save' (at the bottom of the *Setup* sidebar) to save the changes to the setup.
5. VARD will now use this training, notice how the method weights have changed in the top right table. VARD should now find correct suggestions more frequently.

Batch processing multiple texts

1. Select *Batch* in the sidebar.
2. Use *Add folder(s)* to add the whole *Lampeter* corpus to be processed.
3. Choose a suitable threshold, 70% for example. (Please refer to evaluation of threshold effect presented).
4. Select *Changes Tagged*.
5. Select a suitable output folder for the normalised texts.
6. Click *Process list*. The texts will be processed and output to the selected folder.

XML Output

1. Use an appropriate text editor to view the output from your manual and automatic normalisation.
2. Note the different XML tags used by VARD to mark normalisations, non-variants when marked by the user and variants marked by the user when not originally marked as variants by VARD.

Change setup - Twitter

1. Clear any text you have opened with *File->New*.
2. Download the "Twitter setup (untrained)..." zip from the VARD download page into your VARD2.5 folder, and unzip.
3. In 'Setup' (top left of VARD), either drag the "twitter-setup" folder to the dropdown box, or browse for the folder with the folder-find button.
4. The new setup will load, you can observe the setup properties in the *Setup* sidebar.
5. Download (from corpora folder), and open the "lancaster_tweets.txt" file by dragging it to the text area (or using *File->Open*). Notice which strings are "ignored" by VARD.
6. Try dealing with some of the spelling variation found (as described above).