

Foreign Language Support

© Bruce Wilcox, gowilcox@gmail.com brilligunderstanding.com

Revision 1/7/2016 cs7.1

Foreign Language Overview

ChatScript comes natively with full English support. If you want to use a different language you need a variety of things. * Pos-tagging support * Spell check support * Concepts in the language * LIVEDATA substitutions appropriate to the language * Patterns in the language * Output in the language

ChatScript has a command line parameter **language=** that tells CS the language you intend. It defaults to **ENGLISH**. The effects of this parameter are several. * If not ENGLISH, internal pos-tagging and parsing are disabled. * The system will use **DICT/language**. * The system will use **LIVEDATA/language** * The script compiler will automatically compile lines marked with the language (see language comments).

ChatScript supports two kinds of conditional compile comments. Single line comments look like this:

```
#ENGLISH this line will compile if the language is English but not if the language is GERMAN
#GERMAN this line will not compile if the language is English but will compile if the language is GERMAN
```

As always, such comments run til end of line. The other comment is the block comment like this:

```
##<<ENGLISH these lines will be compiled under English
until a normal closing block comment ##>>
```

Using conditional compilation, you can make English and other language versions of code sit side by side if you want to.

ChatScript supports UTF8, so making output or patterns in the language is entirely up to you. Ditto for LIVEDATA.

The dictionary file can be just a list of words of the language, one per line. You must list all conjugations of a word because there is no in-built support to figure that out.

If you want the POS values and lemmas (canonical form of a word), you will need a POS-tagger of some sort. While it is possible to hook in an external tagger via a web call, that will be noticeably slower than an in-built system. ChatScript supports in-built TreeTagger system, which supports a number of languages. However, you can only use this if you have a commercial license. You can try it out using ^popen, as is done in the German bot, however it will be slow because it has to reinitialize TreeTagger for every sentence. The in-built system does not.

A license (per language) is about \$1000 and you can contact me if you want to arrange to use it.