

LING-L 445: Computation and Linguistic Analysis

Practical 01B

Instructor: Francis M. Tyers

Dante Razo

January 31, 2019

Questions

Why should we split punctuation from the token it goes with?

Should abbreviations with space in them be written as a single token or two tokens?

How about numerals like 134 000?

If you have a case suffix following punctuation, how should it be tokenized ?

Should contractions and clitics be a single token or two (or more) tokens?