Ling 555 — Programming for Linguists
Python — Markov Analysis and n-grams

Robert Albert Felty

Speech Research Laboratory
Indiana University

Nov. 3, 2008

Prereqs

While we are waiting

Please download the following files from:
http://robfelty.com/teaching/L555Fall2008/
resources/py/
markovAnalysis.py
```
for file in
{sortDictExcept.py,raising.py,nestedExcept.py};
do curl -o $file
http://robfelty.com/teaching/L555Fall2008/
resources/py/$file; done
```

For Wednesday:

Read Chapter 9 about Magic methods and iterators

Outline

1. Homework questions and comments
   - Type-token
   - questions

2. Some more python tools
   - debugging
   - documentation

3. Markov Analysis and n-grams
   - markov Analysis
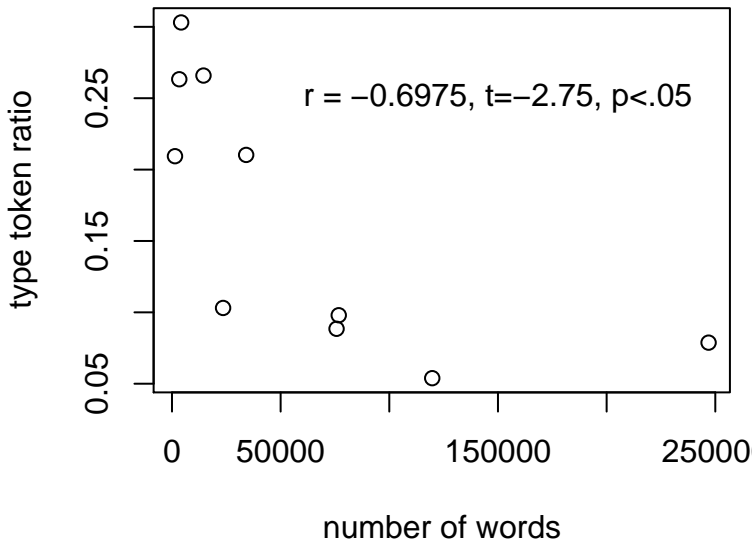   - n-grams
   - example

4. Resources

Type token ratio

r = −0.6975, t=−2.75, p<.05

4

Any questions on homework 8?

Debugging

I recently discovered that python does ship with a command-line debugger - **pdb**

```
python -m pdb myPythonFile.py
```

Documentation

You can create documentation directly from your docstrings using epydoc

```
epydoc --html --docformat
    restructuredtext --noframes
    --show-imports l555.py
```

Markov Analysis

Definition

Markov Analysis basically consists of using prior information to estimate future outcomes.

N-grams

Definition

An n-gram is a sequence of n items. We can split up any string into n-grams.

Example

A bird in the hand is better than two in the bush has the following bi-grams: # a, a bird, bird in, in the, the hand, hand is, is better, better than, than two, two in, in the, the bush, bush #

An example

Let's look at markovAnalysis.py

Resources for you:

Some of the examples we have covered in class today can be found on the website at:
http://robfelty.com/teaching/L555Fall2008/resources/py
The files from today are:

1. markovAnalysis.py