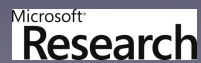


Bugs, au Naturele.

Premkumar Devanbu
DECAL Laboratory
University of California, Davis



```
public class FunctionCall {
    public static void funct1 () {
        System.out.println ("inside funct1");
    }
    public static void main (String[] args) {
        int val;
        System.out.println ("inside main");
        funct1();
        System.out.println ("call funct2");
        val = funct2(5);
        System.out.println ("funct2 returned a value of " + val);
        System.out.println ("call funct1 again");
        val = funct2(-3);
        System.out.println ("funct2 returned a value of " + val);
    }
    public static int funct2 (int param) {
        System.out.println ("inside funct2 with param " + param);
        return param * 2;
    }
}
```

ICSE 2012, "On the Naturalness of Software"



TIGER!!
RUN!!!



Meanwhile, back
in Redmond..
(or Bangalore,
or Shangahi,
or Sunnyvale..)

Why is your
feature
behind
Schedule?

Code, Code,
Code, Code,
Code...

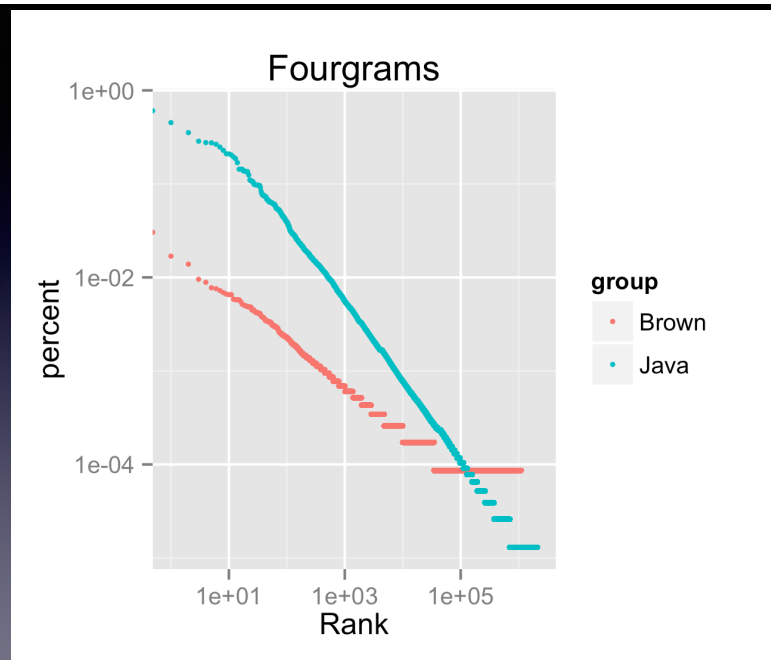
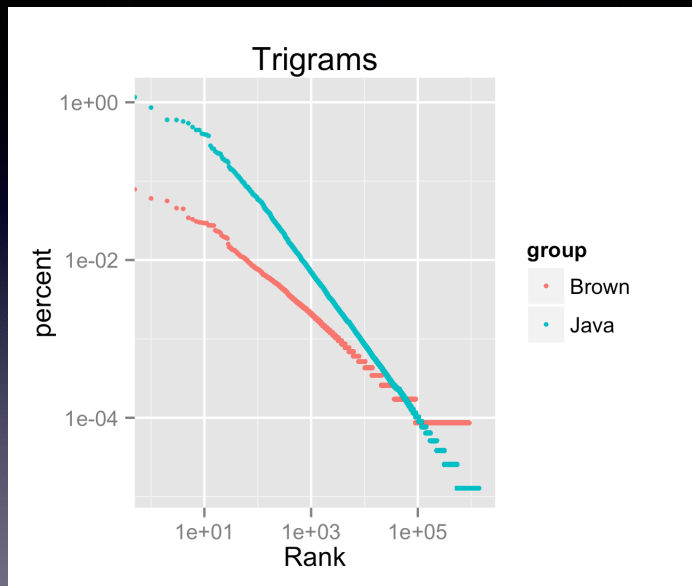
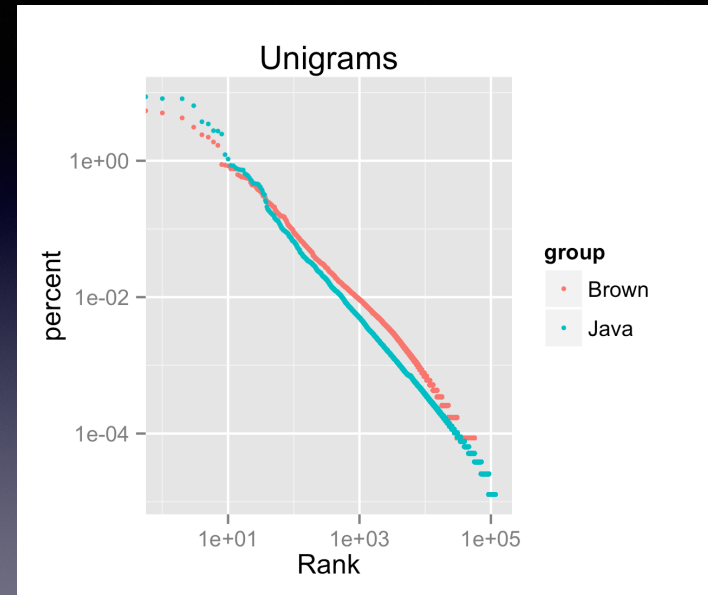


On the Uniqueness of Code (FSE 2010)

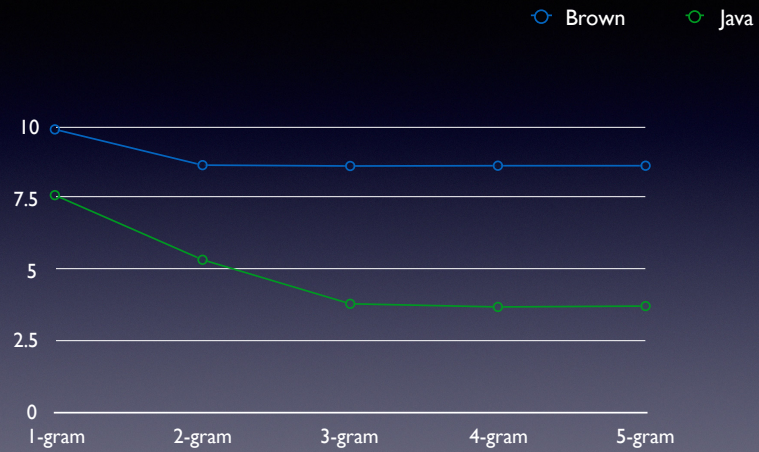
“Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do...”



First, Some Differences

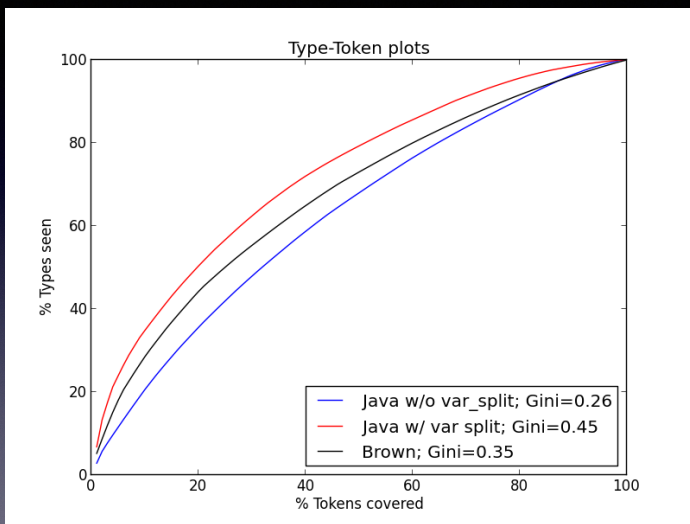


Cross-Entropy per token

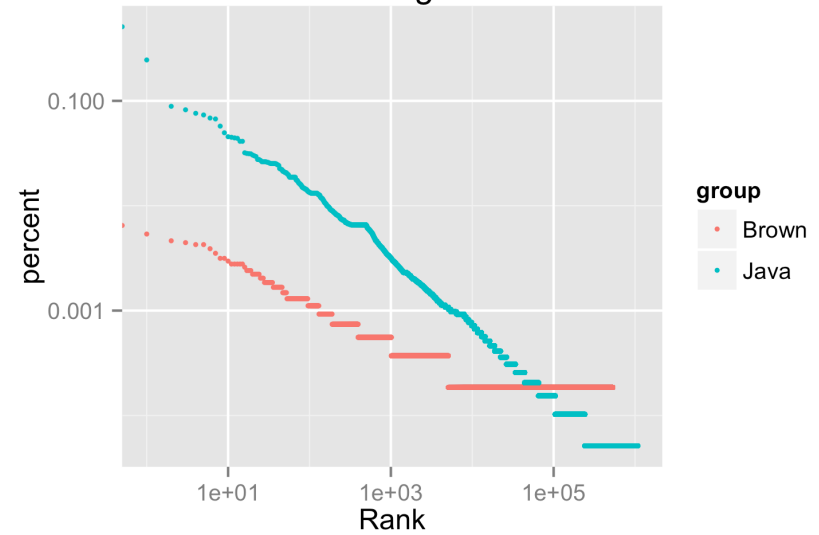


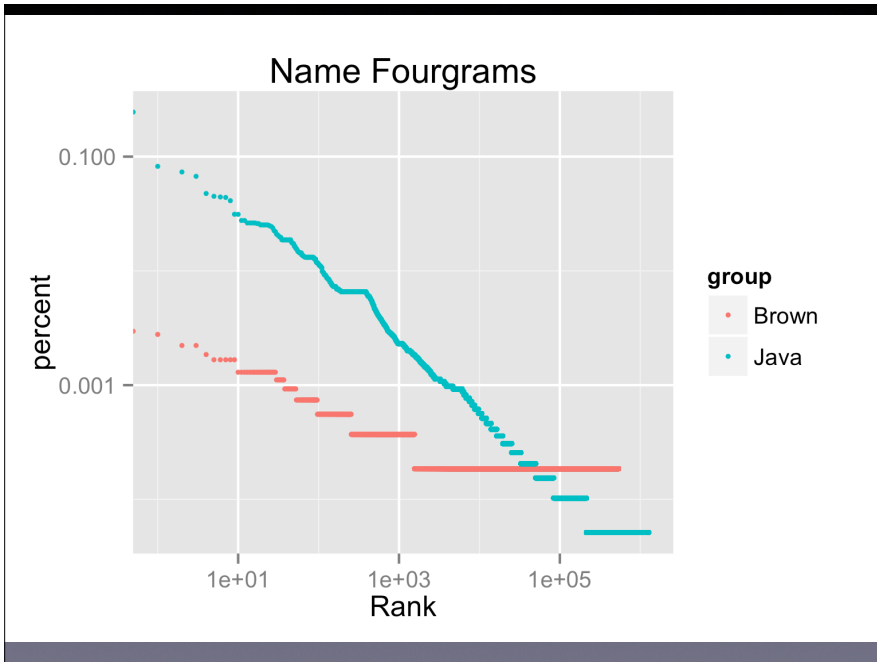
The Skeptic asks..

Is it just that C, Java, Python... are *syntactically* simpler than English?

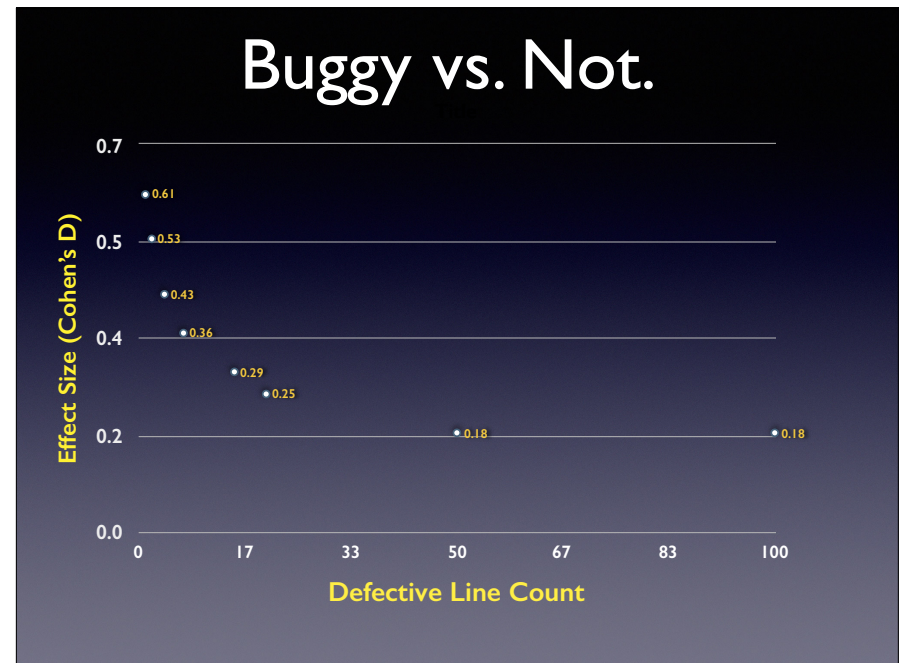
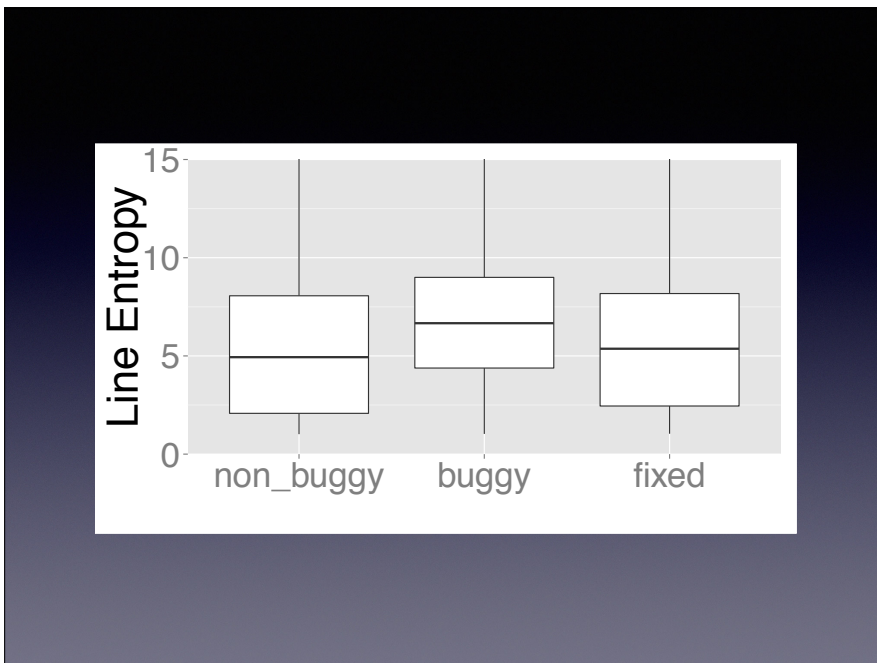


Name Trigrams

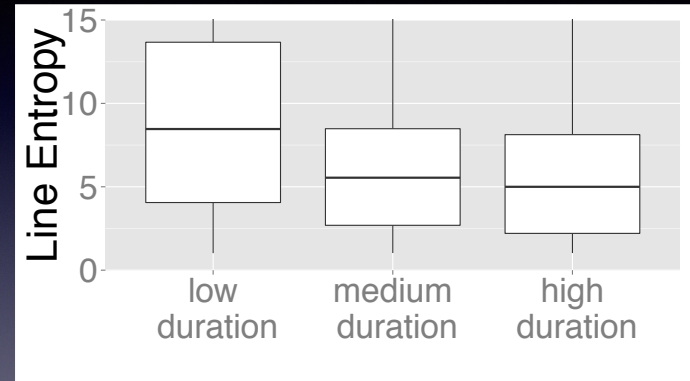
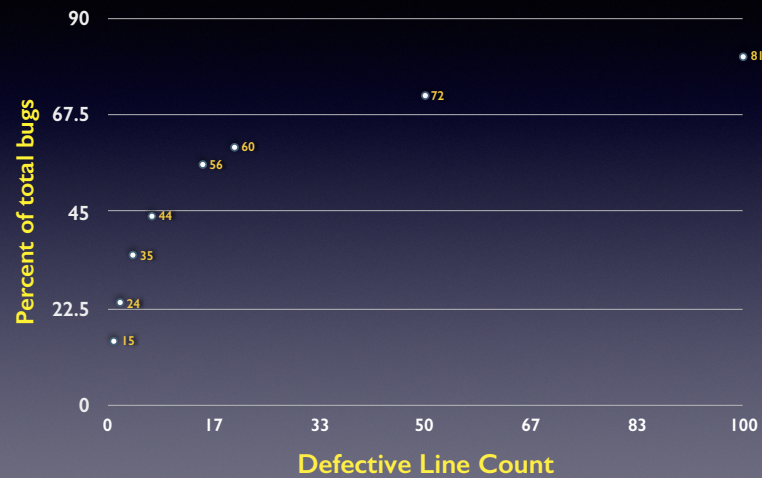




Is buggy code odd?



Defective line count



Does it work? *How to tell?*

- Problem: Line-level!
- Cost-sensitive measures more suitable.
- Comparable to static analysis warnings!

...measured using Cost-effectiveness curve.

Findings Summary

- More cost-effective than logistic regression at 5% inspection budget, but not at 20%.
- Cost-effectiveness similar to FindBugs and PMD.
- Entropy-based ordering improves cost-effectiveness of PMD & FindBugs