

# Corpus Assignment Two

Jean Mark Gawron

Linguistics 503  
San Diego State University

January 19, 2013

# Problems

- 1 Problem one: tokenization
- 2 Problem two: split infinitives
- 3 Problem three: passives
- 4 Extra credit Problem X: More passive

## Problem one, slide

Which BNC document contains the following word sequence? [Note: there is exactly one.]

*I said to him, "What's up, Sam?"*

Hoffmann et al. (2008)

## Problem two, slide

Phrases like the following are excluded by prescriptive grammarians.

*To boldly go where no man has gone before...*

Such constructions are called **split infinitives**. The pattern is

*To Adverb bare-verb-form*

Instead of using such constructions, one should (according to said grammarians) write:

*To go boldly where no man has gone before...*

Do a search on both patterns and report on their relative frequencies.  
Hoffmann et al. (2008)

## Problem three, slide 1

### Passives

Do a search that returns sentences with passives in them. Here are some examples.

- i. *John is/was treated badly.*
- ii. *The children are/were seen in the movie theater.*
- iii. *John may have been convicted.*
- iiii. *John may have been unjustly convicted.*
- iv. *John is being cheated.*
- v. *This outrage must be avenged.*

## Problem three, slide II

Adapted from Hoffmann et al. (2008).

- a. What was your query? Are all your result valid? If your search includes some invalid results, try to estimate what per cent of your results are invalid.
- b. Do a distributional analysis.
  1. Is there a difference in the relative frequency of passives in written and spoken texts.
  2. Do a breakdown of passives by genre. Which genre has the greatest relative frequency of passives? Which has the least? Some genres are spoken and some are written. Does this genre breakdown show that finer grained analysis is useful in characterizing the frequency of passives in spoken and written texts? How? What are the complicating factors you observe?

## More passives

Create a new passive search that includes passive examples like the following:

*What book was the prudish principal seen reading?*

That is, in interrogative clauses, we must allow a Noun Phrase to intervene between the auxiliary *was* and the main verb *seen*.

Hoffmann, S., S. Evert, N. Smith, D. Lee, and B. P. Prytz. 2008.  
*Corpus Linguistics with BNCweb: a Practical Guide.*  
New York: Peter Lang.