

Part-of-Speech Tagging



A Canonical Finite-State Task

The Tagging Task

Input: the lead paint is unsafe

Output: the/Det lead/N paint/N is/V unsafe/Adj

- **Uses:**
 - text-to-speech (how do we pronounce “lead”?)
 - can write regexps like (Det) Adj* N+ over the output
 - preprocessing to speed up parser (but a little dangerous)
 - if you know the tag, you can back off to it in other tasks

Why Do We Care?

Input: the lead paint is unsafe

Output: the/Det lead/N paint/N is/V unsafe/Adj

- The first statistical NLP task
- Been done to death by different methods
- Easy to evaluate (how many tags are correct?)
- Canonical finite-state task
 - Can be done well with methods that look at local context
 - Though should “really” do it by parsing!