

A Non-Parametric Bayesian Approach to Inflectional Morphology

Jason Eisner
Johns Hopkins University

This is joint work with **Markus Dreyer**.

Most of the slides will be from
his recent dissertation defense.

See the dissertation for lots more!
(models -> algorithms -> engineering -> experiments)

Linguistics quiz: Find a morpheme

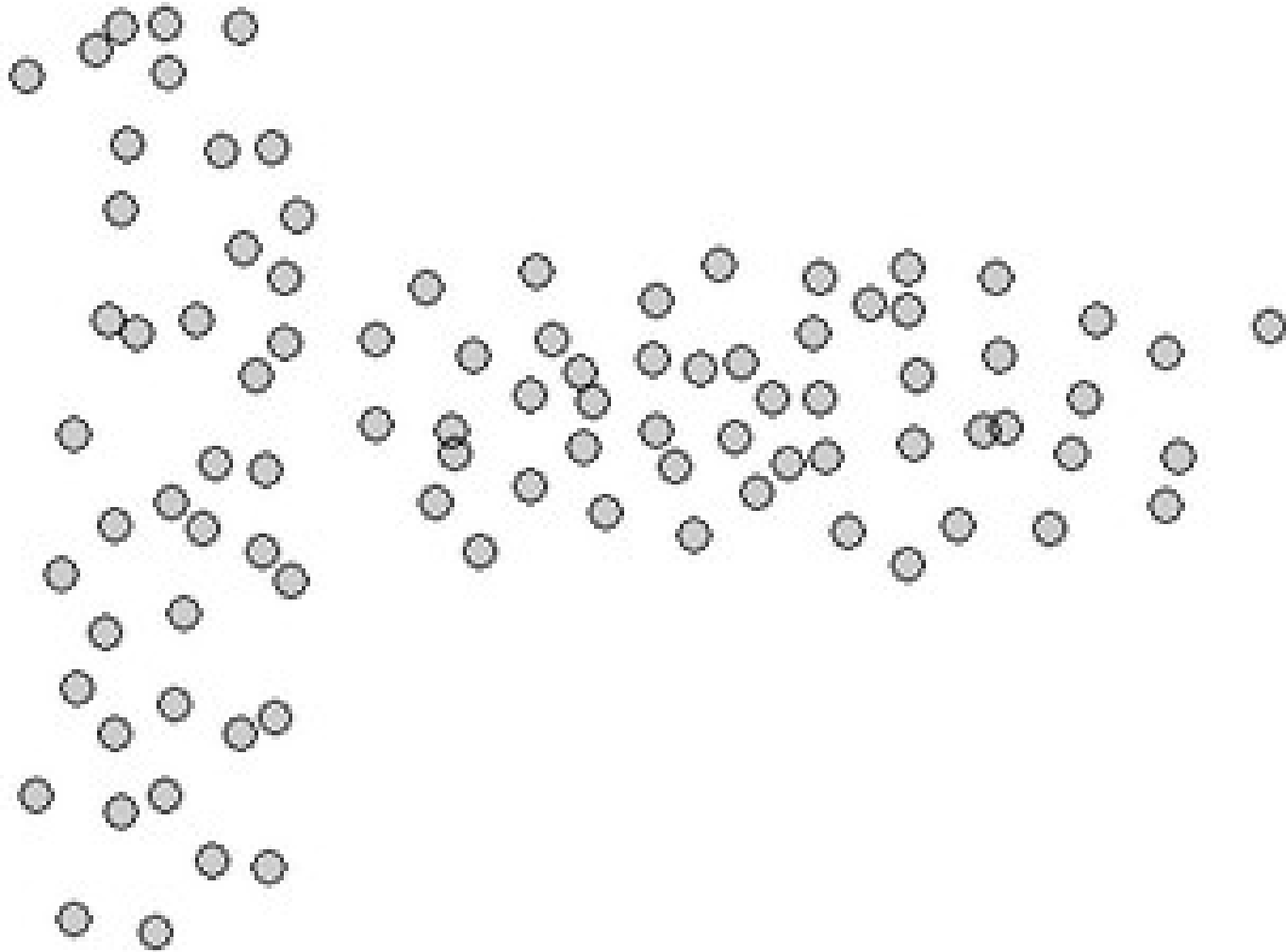
Blah blah blah snozzcumber blah blah blah.

Blah blah blahdy abla blah blah.

Snozzcumpers blah blah blah abla blah.

Blah blah blah snezzcumbri blah blah
snozzcumber.

How is morphology like clustering?



Output of this Work

(all outputs are probabilistic)

- **Token level:** Analysis of each word in a corpus
 - POS + lexeme + inflectional features
- **Type level:** Collection of full inflectional paradigms
 - Including irregular paradigms
 - Including predictions of never-seen forms
- **Grammar level:** Finite-state transducers
 - Analysis and generation of novel forms

Caveats

- Not completely unsupervised
 - Need some paradigms to get started.
- Natural extensions we haven't done yet:
 - Use context to help learning (local correlates, syntax, topic)
 - Use multiple languages at once (comparable or parallel)
 - Reconstruct phonology
 - But the way ahead is clear!

How to combine with MT

- Could hook up to an MT system
 - Candidates for analysis & generation
 - So can be consulted by a factored model
 - Or can just be used as pre-/post-processing
- Better: Integrate with a synchronous MT model
 - Learn morphology jointly with alignment, syntactic refinements, etc.
 - Bitext could be a powerful cue to learning

Modeling First, Algorithms Second

- Get the generative story straight first.
 - What do we actually *believe* about the linguistics?
- Then worry about how to do inference.
 - In principle, it just falls out of the model.
 - In practice, we usually need approximations.
- Advantages:
 - Should act like a reasonable linguist.
 - Approximations are often benign (don't sacrifice whole categories of phenomena) and continuous (we can trade runtime for accuracy, as in pruning).
 - Can naturally extend or combine the model.