

Background.

The theoretical status of inflection and derivation have been the subject of extensive debate within linguistics. While many general tendencies of the two categories exist, these criteria often conflict, making a consistent definition elusive.

This has led some linguists to conclude that the distinction is gradient, or even irrelevant. Nevertheless, these concepts are frequently operationalized within theoretical work, linguistic corpora, and NLP.

These debates within theoretical linguistics have lacked a grounding in quantitative evidence from a wide range of languages, which we aim to provide.

Our measures.

We operationalize the intuition that derivations create larger and more variable changes to words, in terms of form and distribution. We use orthographic form as a proxy and edit distance as a distance metric. Distribution is used as a combined measure of semantic and syntactic similarity, with distance operationalized as Euclidean distance between FastText embeddings.

$$\begin{array}{ll} \text{Magnitude} & \text{Variability} \\ \text{form} & \|\Delta_{\text{form}}\| & \text{var}(\Delta_{\text{form}}) \\ \text{distribution} & \|\Delta_{\text{embed}}\| & \text{var}(\Delta_{\text{embed}}) \end{array}$$

These measures are computed between base and constructed forms and averaged over individual morphological constructions to compute 4 features for each construction.



Corpus-based measures discriminate inflection and derivation cross-linguistically.

With 90% accuracy across 26 languages!

Coleman Haley Edoardo Ponti Sharon Goldwater
coleman.haley@ed.ac.uk eponti@ed.ac.uk sgwater@ed.ac.uk

How do you quantify this?

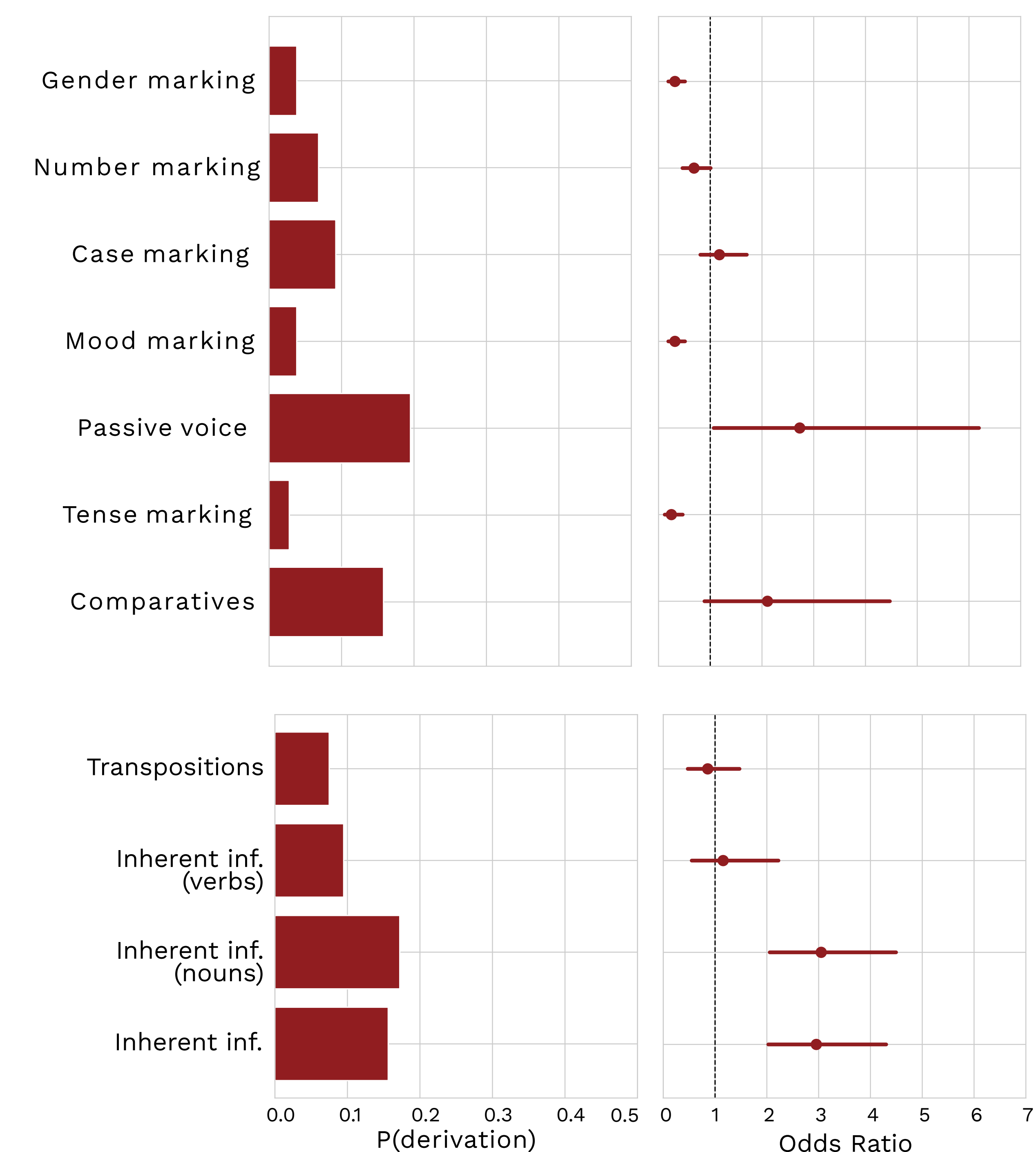
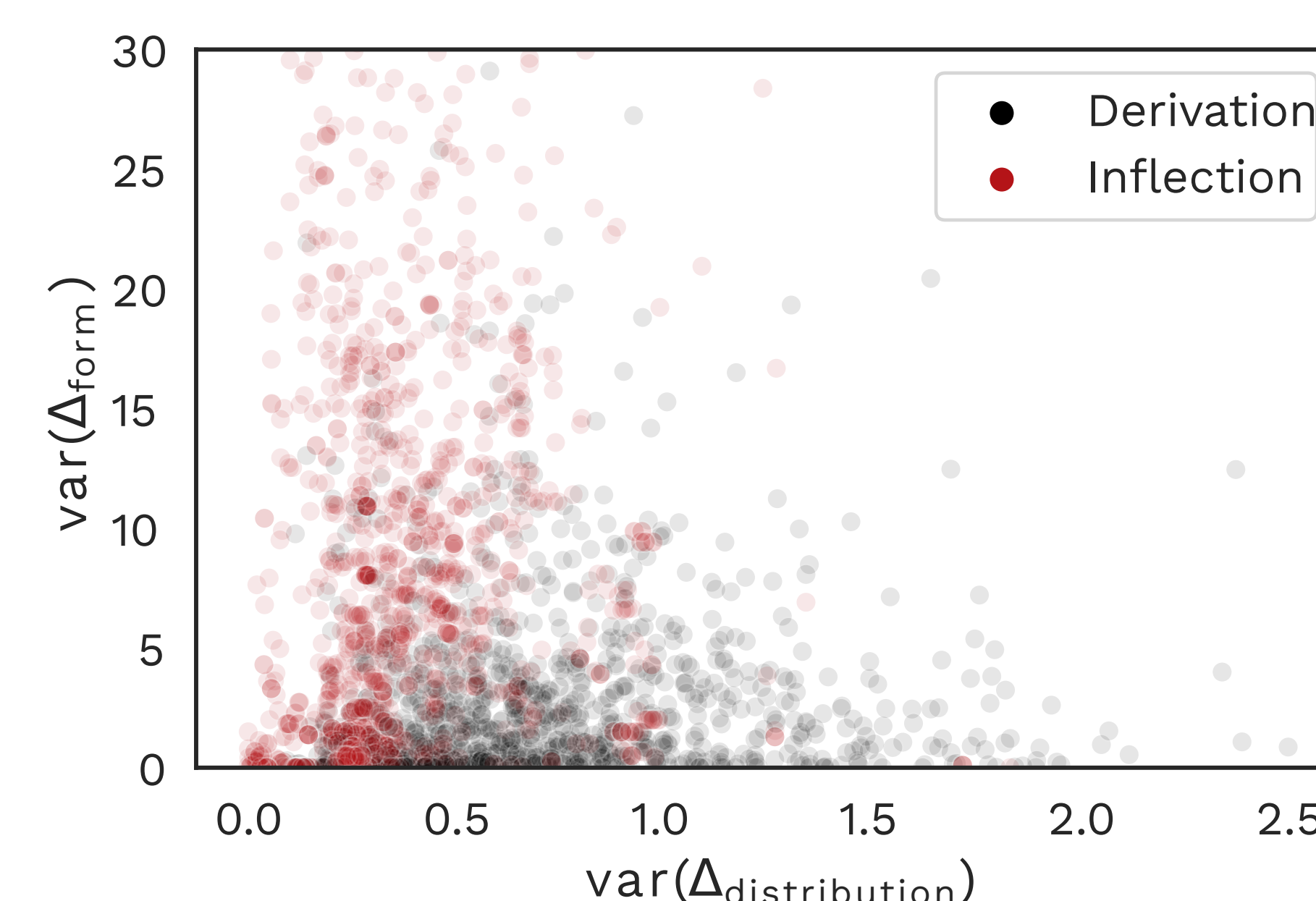
To identify how consistently the terms inflection and derivation are applied in Unimorph, we train logistic and MLP classifiers on our features and predict on held out constructions. We find using all 4 features performs best, achieving a classification accuracy of 90%, across 26 languages with no language specific features!

What about non-IE languages?

We have 5 non-Indo-European languages in our dataset. We achieve a classification accuracy of 81%, indicating a degree of generalization, but also perhaps less consistency for non-IE languages. For example, Turkish is typically considered to be exceptionally inflectionally rich, but our classifier identified many of its inflections as derivations.

Is the distinction gradient?

While inflection and derivation are highly consistently assigned in terms of our measures, the distinction appears more gradient than categorical under our measures, with the measures not forming clear clusters.



Is there a pattern to the errors?

We looked at several semantic and syntactic categories of inflection. No category did our model more frequently classify as derivation than inflection, however some categories were classified as derivations more frequently than others. We identify inherent inflections, especially inherent nominal inflections, as especially non-canonical inflections, but find this effect driven by nominal categories.