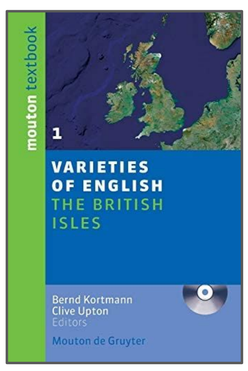
Multi-VALUE: A Framework for Cross-Dialectal English NLP

Caleb Ziems[1*], **William Held**[2*], **Jingfeng Yang**[3], **Jwala Dhamala**[3], **Rahul Gupta**[3], **Diyi Yang**[1]

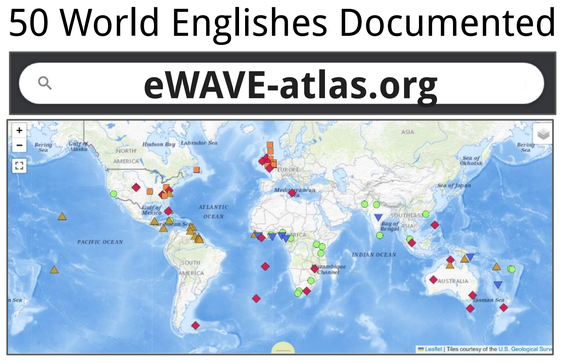
[1]Stanford University, [2]Georgia Institute of Technology, [3]Amazon

value-nlp.org/

Linguistic Foundations



50 World Englishes Documented → 235 Features



No.	Feature Name
1.	She/her used for inanimate referents
2.	He/him used for inanimate referents
3.	Alternative forms for referential (non-dummy) it
…	……

189 Perturbations

```
def referential_thing(self):
    # feature 3
    replace = "the thing"
    for token in self.tokens:
        if token.dep_ != "expl" and \
            token.lower_=="it":
            self.set_rule( token,
                    replace )
```

1. Introduction

Problem: Dialect Disparity

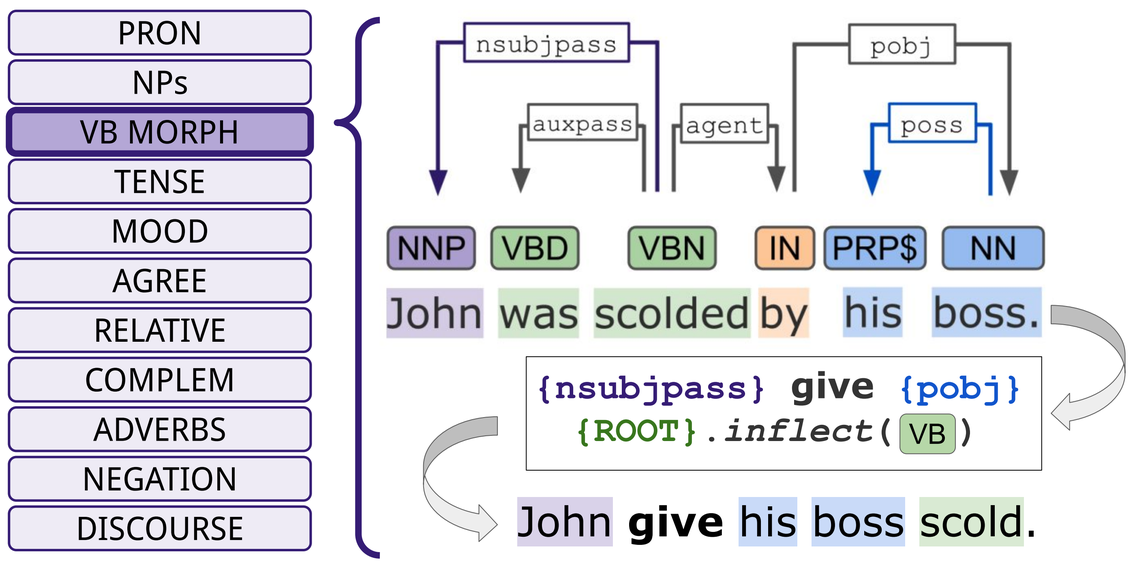
- **Inequitable NLP:** systems may struggle with *language variation* caused by regional, social, and economic factors
  - **Empirical Understanding:** cross-dialectal disparities have not been measured systematically
  - **Public Awareness:** without measurement, there is less research and public attention on this important issue
- **Low-Resource NLP:** with very limited dialectal corpus data, we need to use resources strategically to reduce dialect disparity

Proposed Solution: *Multi*-Dialectal *V*ern*A*cular *L*anguage *U*nderstanding and *E*valuation (*Multi-VALUE)*

- **Rule-based Translation:** inject 189 *dialect features* into text
  - ↪ **Stress Tests:** scale up empirical understanding to 50 English dialects
  - ↪ **Augmented Training:** close performance gaps for low-resourced dialects
- **Gold Standard Benchmarks:** Chicano + Indian English CoQA
- **Dialect-Robust Models:** hosted on huggingface hub

2. Dialect Perturbations

Morphosyntactic Re-writes: POS tags, inflection, dependencies



3. Scope + Reliability of Multi-VALUE

Scope: 189 Features; 50 Dialects; each 80-94% Fully-Implemented

Aboriginal, **Appalachian**, Australian, **Bahamian**, Black South African, **Cameroon**, Cape Flats, **Channel Islands**, Chicano, **Colloquial American**, East Anglican, **Falkland Islands**, Fiji, Ghanaian, **Hong Kong,** Indian, Irish, **Jamaican, Kenyan**, Liberian, **Malaysian**, Maltese, **New Zealand**, Newfoundland, **Orkney and Shetland**, Ozark, **Philippine**, Pakistani, **Scottish**, South African, **Sri Lankan,** St. Helena, **Tanzanian**, Tristan da Cunha, **Urban African American**, Ugandan, **Welsh**, Zimbabwean

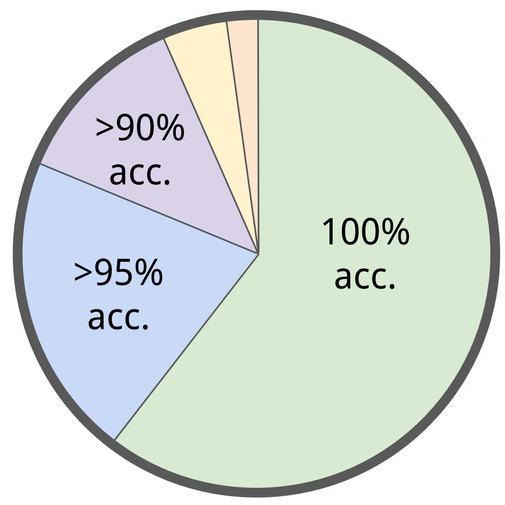
4. Reliability of Multi-VALUE

Reliability: 92 perturbations validated by 72 native speakers

- **Validation Goal:** confirm that our rules are aligned with real speakers' grammars
- **Gold Standard:** if the transformation is unacceptable, annotators provide an alternative "translation"

Sentence (1): can't nothing good happen
Sentence (2): nothing good can happen

- 100% accuracy: 55 features (60.4%)
- [95%, 100%) acc: 19 features (20.9%)
- [90%, 95%) acc: 11 features (12.1%)
- [85%, 90%) acc: 4 features (4.4%)
- [80%, 85%) acc: 2 features (2.2%)

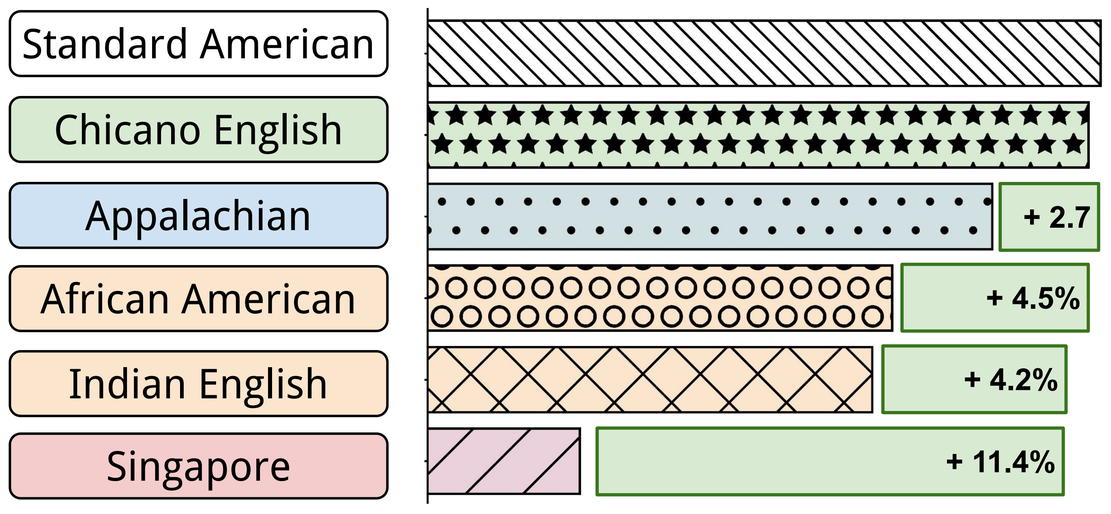


5. Using Multi-VALUE

Stress-Test Domains:

Conversational QA | Semantic Parsing | Machine Translation

Augmented training closes dialect performance gap:



6. Multi-VALUE Benefits

1. **Interpretable** (not **black-box**)
2. **Flexible** (tunable **feature-density**)
3. **Scalable** (**mix + match** datasets)
4. **Responsible** (**speaker-validated**)
5. **Generalizable** (truly **cross-dialectal** findings)