

What is Morphological Inflection?

<u>lemma</u>	<u>feats</u>	<u>infl</u>
walk	V; PST	walked
sit	V; PRS; 3; SG	sits
mouse	N; PL	mice

Background

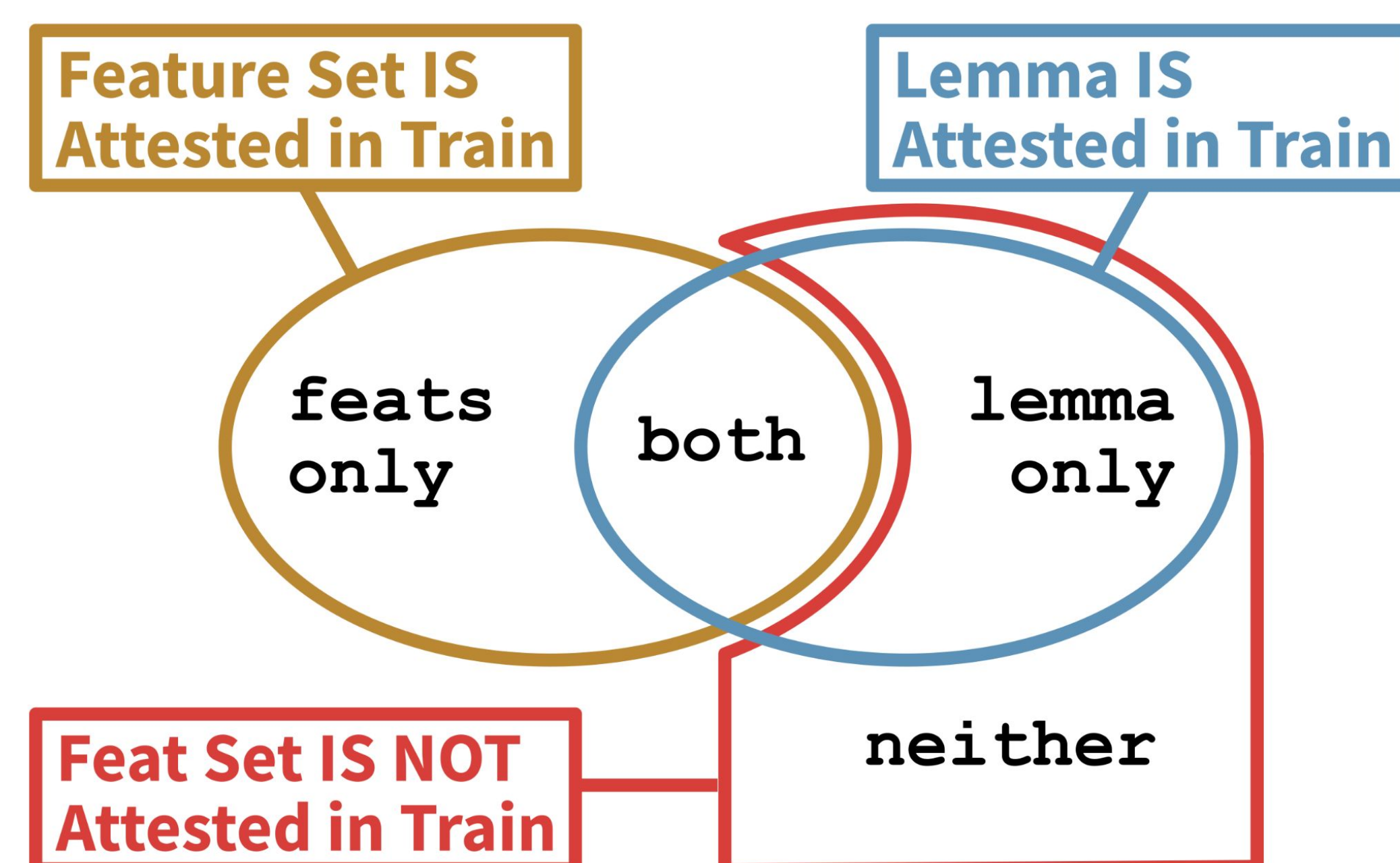
- Popular task in NLP
 - Practical & cognitive applications
- SOTA systems: **highly varying results**
 - Across languages **and** data sets
- **Dataset creation & evaluation** ⇒ variance

We propose *new data sampling and evaluation strategies to improve generalizability and reliability of morphological inflection models*

Data

- **Languages:** English, Arabic, German, Spanish, Swahili, Turkish
 - Morphological features from UniMorph
- **Sampling:** Uniform, weighted, overlap-aware
- **Seeds:** 5

Overlaps



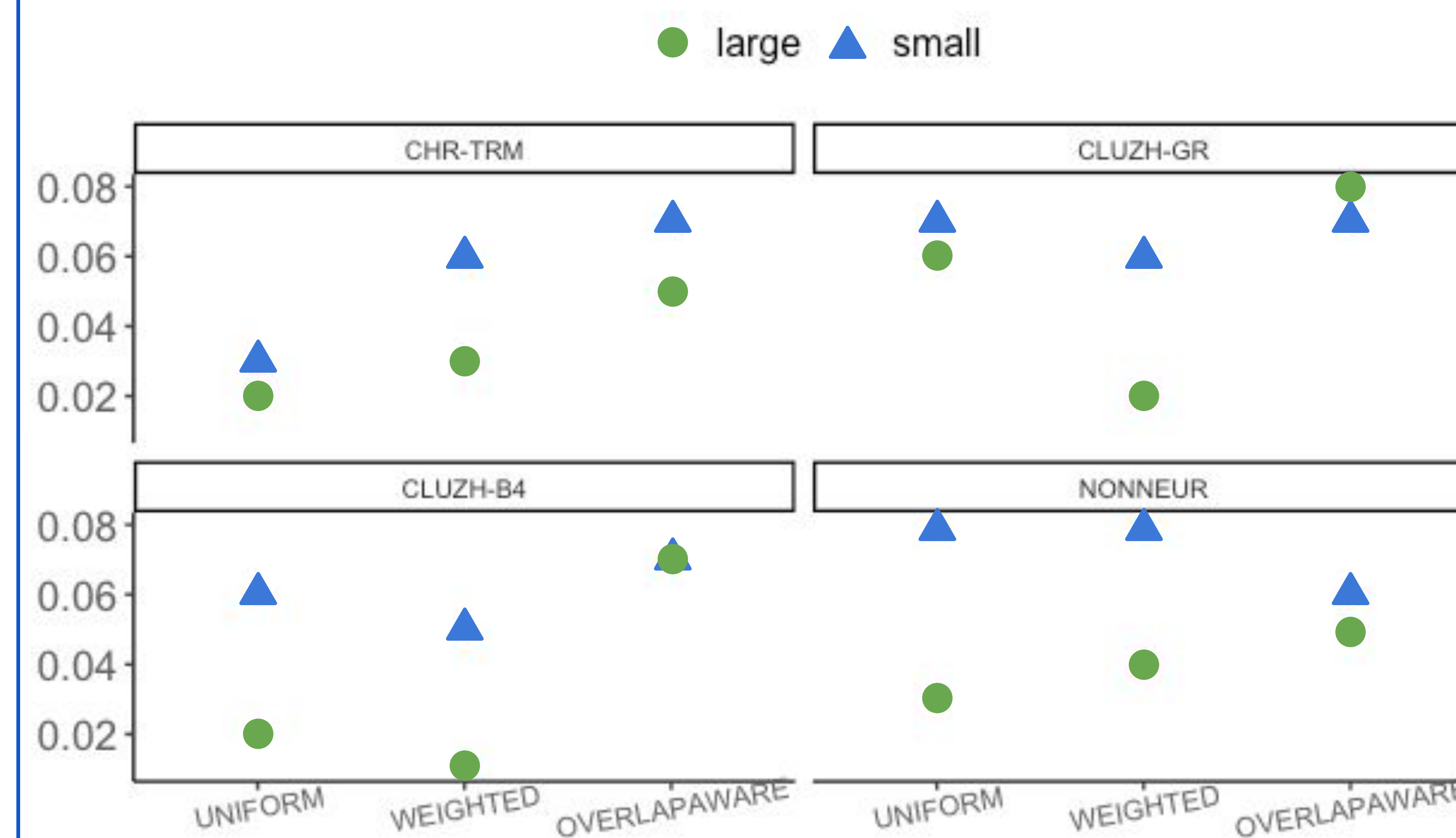
Models

- **CHR-TRM:** character-level transformer
- **CLUZH-GR:** character-level transducer, greedy decoding
- **CLUZH-B4:** character-level transducer, beam decoding size = 4
- **NONNEUR:** majority classifier baseline

Evaluation

- **Effects of:** training size, sampling strategy, overlap type
- **Variation** between random seeds

Results



Test vs S Train	featsAttested	featsNovel
UNIFORM	80.33 (19.50)	19.67 (19.50)
WEIGHTED	90.44 (11.13)	9.56 (11.13)
OVERLAP-AWARE	48.81 (0.98)	51.19 (0.98)
Test vs L Train	featsAttested	featsNovel
UNIFORM	96.17 (5.55)	3.83 (5.55)
WEIGHTED	95.36 (7.28)	4.64 (7.28)
OVERLAP-AWARE	49.92 (0.17)	50.08 (0.17)

Training Size	Sampling Strategy	Score Range	Random Seed Variability
Small	UNIFORM	0.17	0.07
	WEIGHTED	0.09	0.04
	OVERLAP-AWARE	0.20	0.08
Large	UNIFORM	0.12	0.05
	WEIGHTED	0.05	0.02
	OVERLAP-AWARE	0.18	0.07

S.P. gratefully acknowledges the support of the Institute for Advanced Computational Science Graduate Research Fellowship and the National Science Foundation Graduate Research Fellowship Program under Grant No. 2234683. All opinions presented here are those of the authors.

Practices	Current	Proposed
Sampling	Uniform, unweighted	<ul style="list-style-type: none"> • Weighted and ◦ OOV: real distribution ◦ OOV: balanced distribution
Splits	Single	Multiple <i>n</i> splits
Overlap	Uncontrolled	<ul style="list-style-type: none"> • Controlled for combinations of ◦ Lemmas ◦ Features

Test vs S Train	FEATSATTESTED	FEATSNOVEL
UNIFORM	70.74	33.57
WEIGHTED	79.25	22.77
OVERLAP AWARE	79.60	31.13

Test vs L Train	FEATSATTESTED	FEATSNOVEL
UNIFORM	80.00	—
WEIGHTED	85.94	23.74
OVERLAP AWARE	86.22	35.51