Dialect Matters: Cross-Lingual ASR Transfer for Low-Resource Indic Language Varieties

Akriti Dhasmana & Aarohi Srivastava & David Chiang
University of Notre Dame
adhasman@nd.edu & asrivas2@nd.edu & dchiang@nd.edu

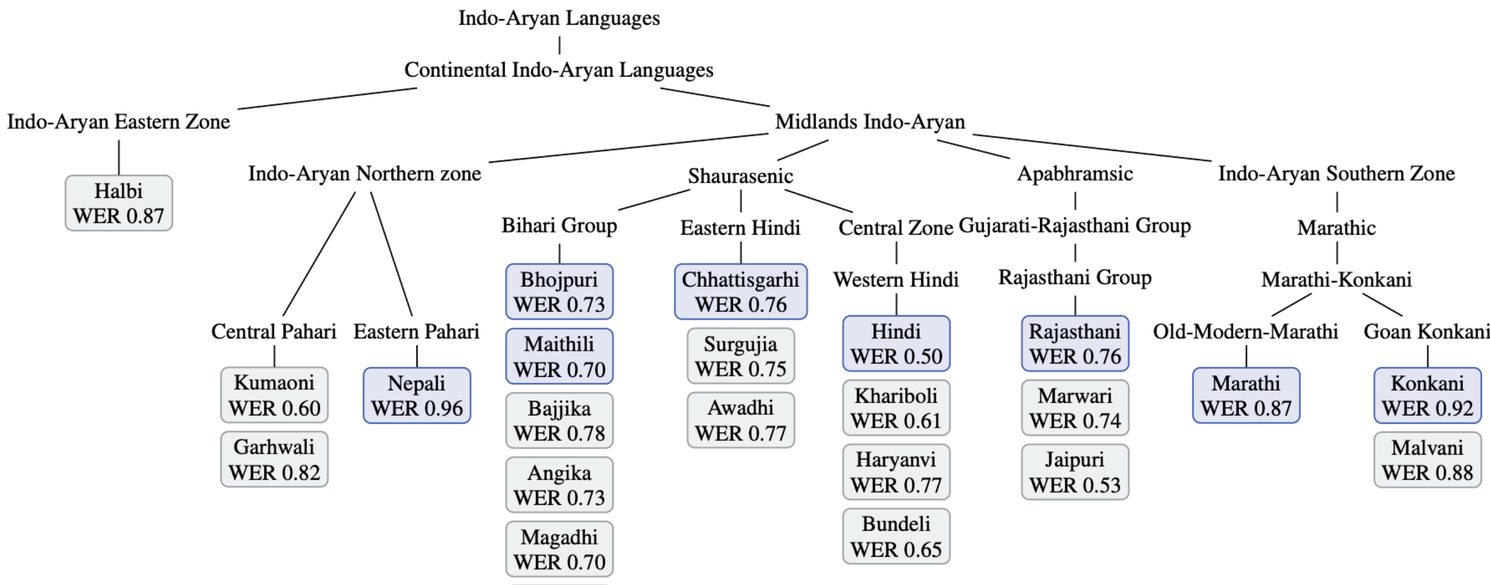

Figure 1: Subset of the Indo-European language family tree showing Devanagari-script Indic languages in the VAANI dataset, based on Glottolog. Languages are annotated with WER for *IndicWav2Vec-Hindi.* Blue indicates a language was seen during pre-training.

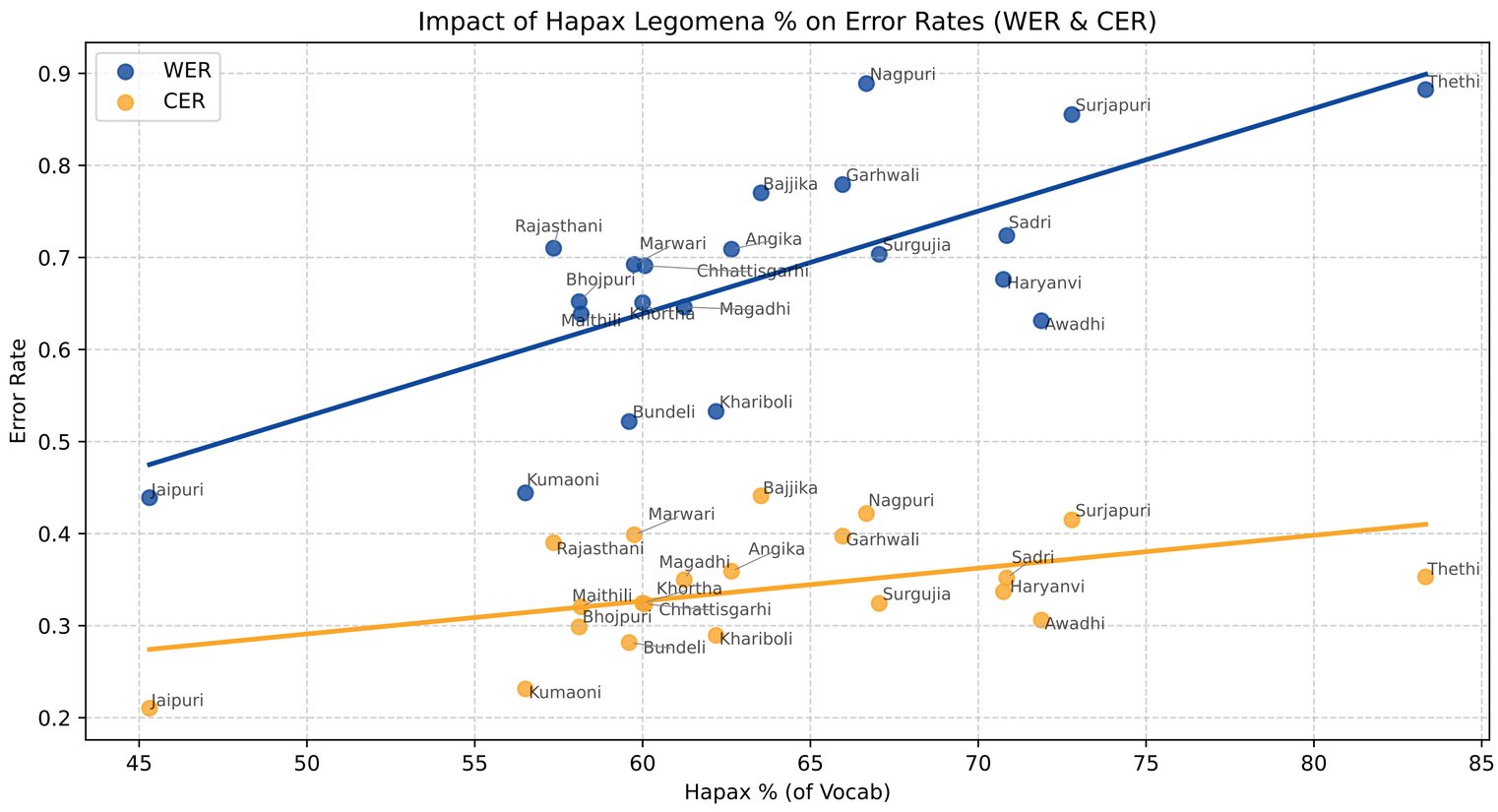

Figure 2: Hapax legomena (once-seen word types) % per language in the test split of VAANI plotted against WER (**Pearson's ρ = 0.705, p = 4 × 10^{-4}**) and CER (*not significant*) using *IndicWav2Vec-Hindi*.

We measure orthographic variation and find a correlation with ASR performance in Indic language varieties **(Pearson's ρ = 0.705)**.

We conduct a **large-scale zero-shot evaluation** on mainstream and dialectal Indic language varieties.

We find that training on a small amount of dialectal speech can yield better performance in dialectal ASR than training on a large amount of mainstream language speech, and is associated with improved ASR performance (**point–biserial correlation r_{pb} = −0.196, p = 5.79 × 10^{-3}**).

We assess the performance of several speech models on Garhwali, a low-resource Pahari language variety, achieving a best case **WER of 0.493 with w2vBERT.**

Our error analysis on Garhwali data shows a **pretraining bias toward Hindi**; models are inclined to change dialectal words to Hindi forms in the generated transcriptions, as shown in Table 1.

Model	# Non-Hi	Correct	To Hi	To Wrong
wav2vec2-BERT	1873	34.8%	23.3%	38.8%
Indicw2v2	2109	4.2%	37.3%	42.9%

Table 1: Comparison of w2vBERT model fine-tuned on Garhwali against IndicWav2Vec-Hindi on Garhwali non-Hindi word handling. Key: Hi = Hindi.

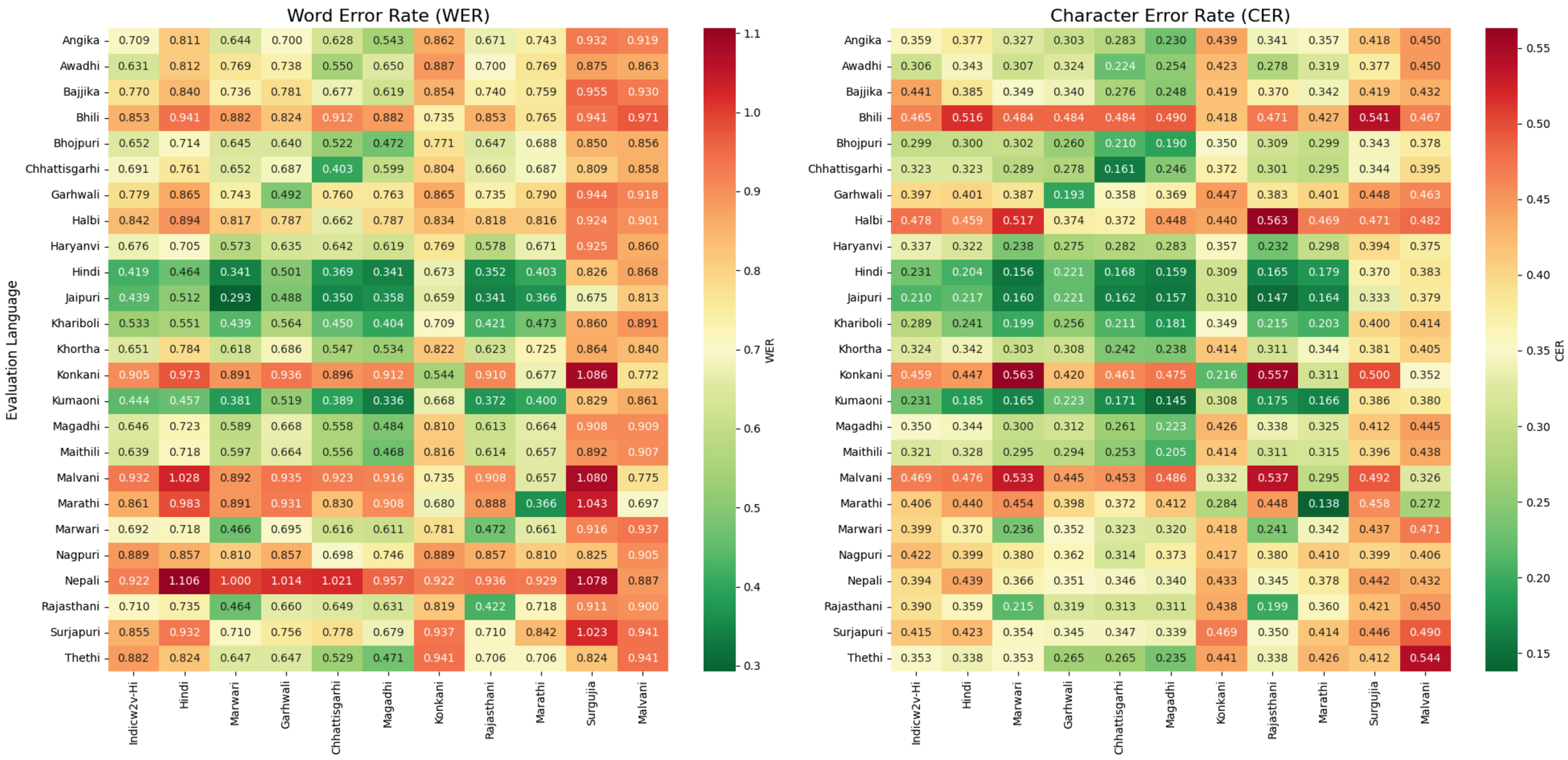

Figure 3: Cross-lingual performance of w2vBERT fine-tuned on 1 to 7 hours of data per language.