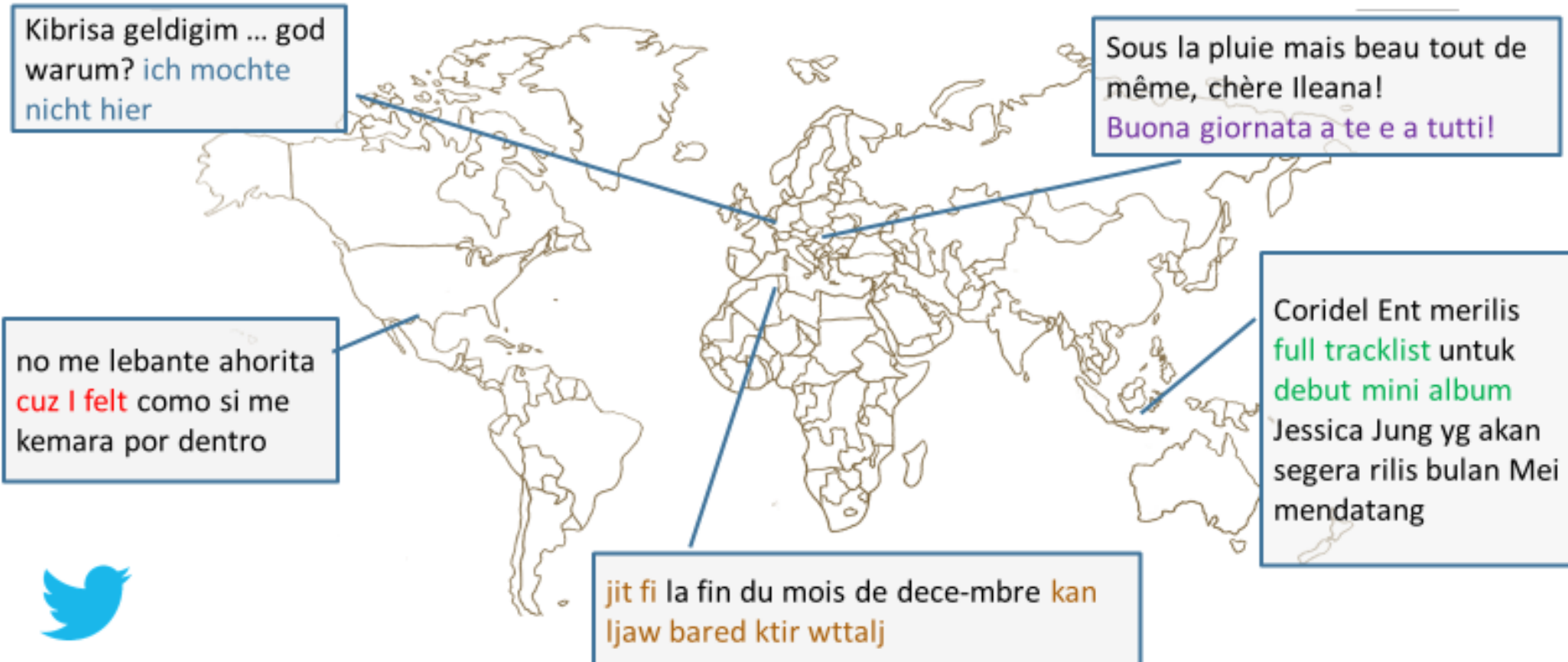


Code-mixing or **Code-Switching** is the mixing of two or more languages in a conversation or even an utterance.



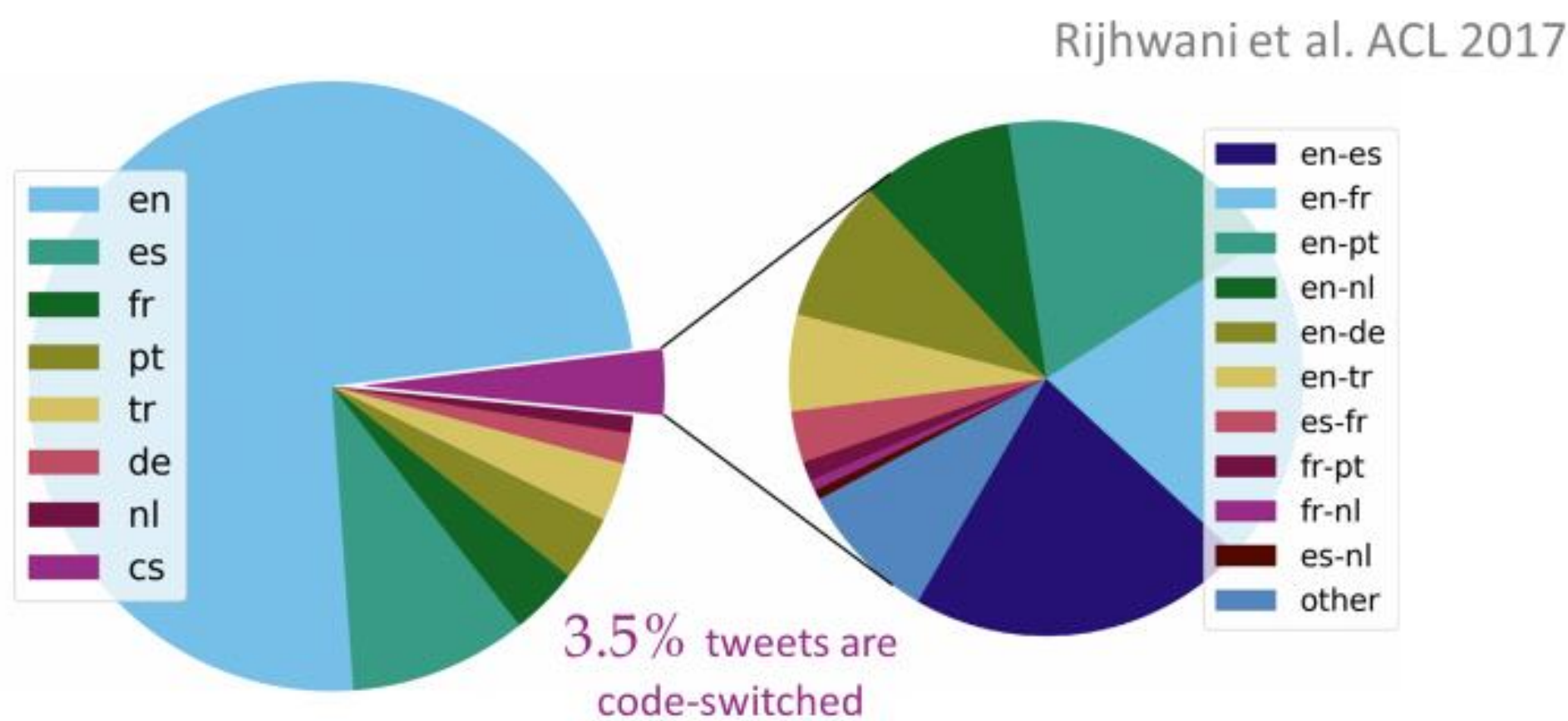
All Speech/NLP Applications break down in presence of Code-switching



Why is it challenging?



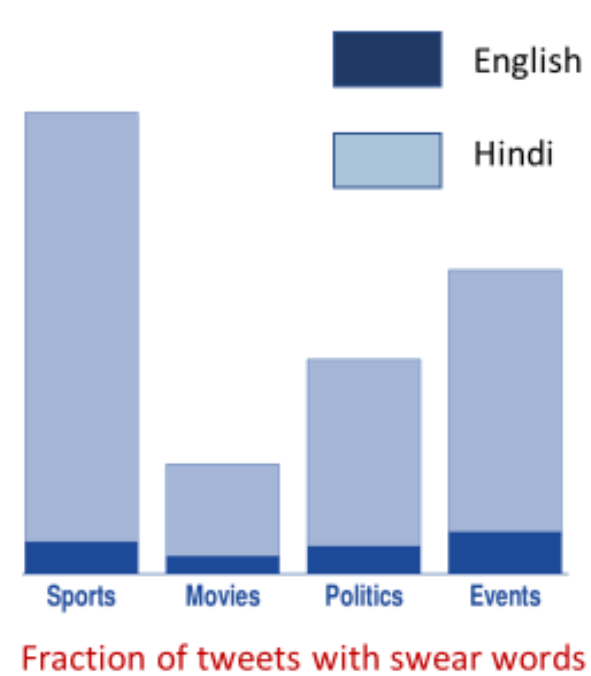
How prevalent is code-switching?



Worldwide language distribution of monolingual and code-switched tweets computed over 50M Tweets (restricted to the 7 languages)

Why and when do people code-switch?

We might praise you in English, but *gaali to Hindi me hi denge!* (Rudra et al., EMNLP 2016)

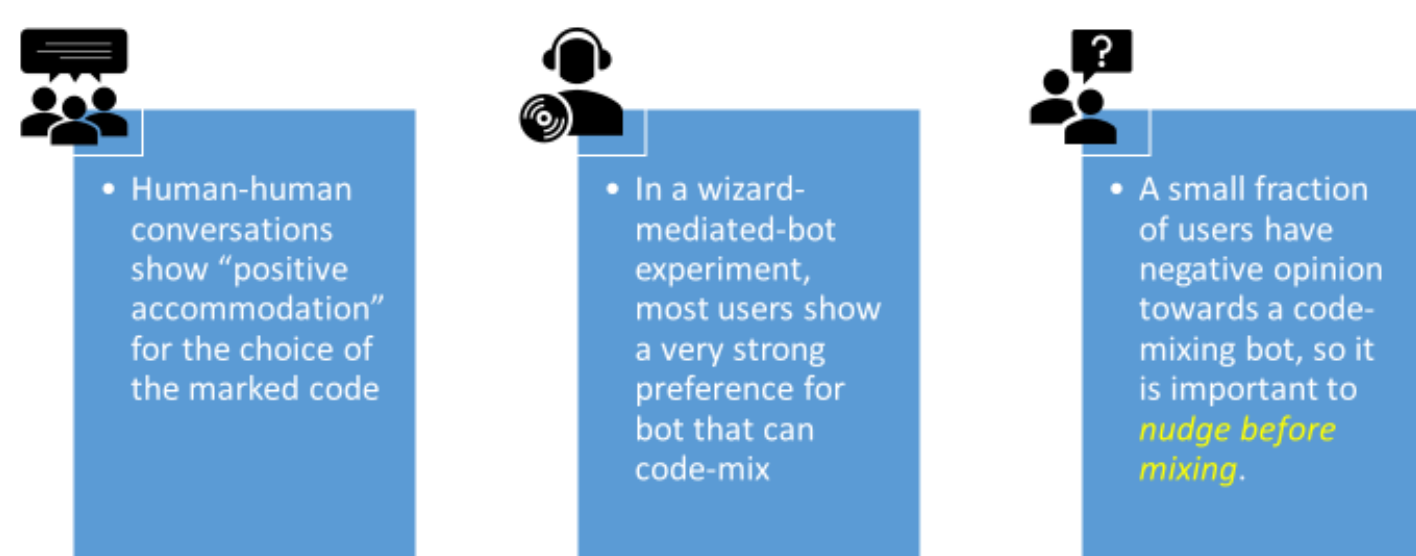


Study of 830K Tweets from Hi-En bilinguals

1. The native language, Hindi, is *strongly preferred* (10 times more) for negativity and swearing
2. English is used far more for positive sentiment than negative
3. Language change often corresponds with changing sentiment

Should chat-bots code-switch?

Nudge, don't push or assume...



How do we solve it?

