

regularize, don't construct embeddings using spellings

usage \perp spelling | embedding
(duality of patterning \sim conditional independence)

The connection between the *signifier* and the *signified* is **arbitrary**.
 spelling \uparrow meaning \uparrow
 – de Saussure, 1916, translated

The meaningful elements in any language—"words" in everyday parlance [...]—[...] are represented by [a] small stock of distinguishable *sounds* which are in themselves wholly meaningless.
 \uparrow characters
 – Hockett, 1960

Meaning is **not fully predictable** from spellings.
 (neither **silly** nor **folly** is an adverb, though they both end in -ly!)

"Meaningless" character composition should be separate from "meaningful" word composition!

\uparrow "construction" models (e.g., $e(\text{caged}) := \text{CNN}(\text{c a g e d})$) ignore this!

\uparrow character-level models ignore this! (and they're slow as hell)

Allowing any pairing regardless of spelling, but using spellings as a prior / as regularization allows for *idiosyncratic* embeddings!

Only use a word's spelling only to *define* it – not to later *use* it!

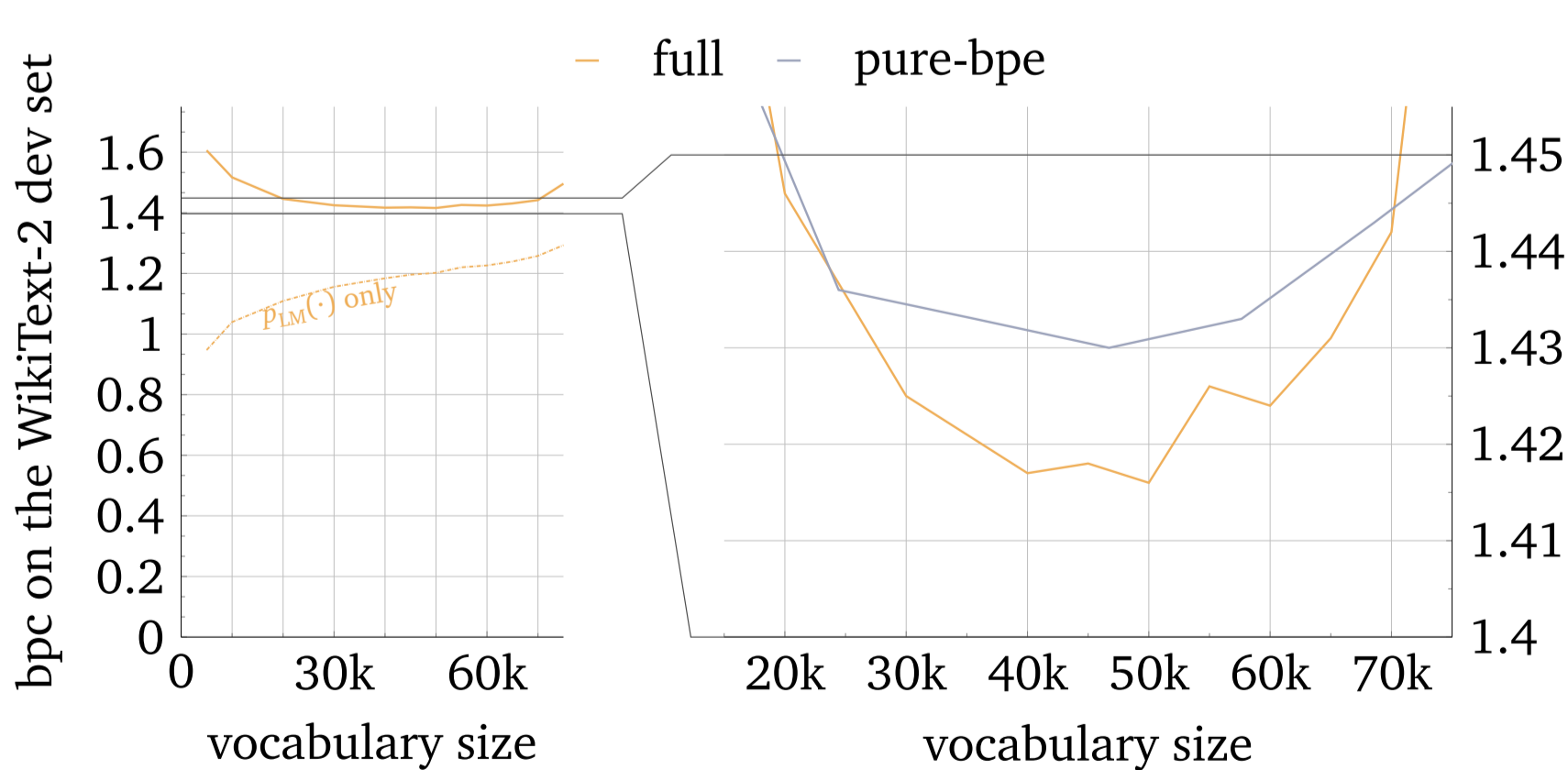
- \Rightarrow Irregular words have uncommon spellings ...yet we use them like regular words! (Example: `children`)
- \Rightarrow Function words have uncommon spellings ...yet we use them all the time without feeling weird! (Examples: `the`, `of`)

How to evaluate fairly?

Evaluation results

- Report likelihood bits per character $p(\text{held-out text})$ as perplexity (\downarrow lower is better)
 - \triangle no final UNKs allowed!
- \Rightarrow predict everything, regardless of a tunable "vocabulary size," that merely decides what is temporarily UNK:

WikiText-2 (Merity et al., 2017) 2.5 million tokenized words from the English Wikipedia	on dev data			test all
	novel words	rare words	frequent words	
t h e _ c a t _ c h a s e d _	3.89	2.08	1.38	1.775
HCLM + cache previous SOTA (Kawakami et al., 2017)	–	–	–	1.500
BPE: t h e c a t t c h a s e d	4.01	1.70	1.08	1.468
our full model: Spell Once, Summon Anywhere	4.00	1.64	1.10	1.455



...and plenty more baselines, ablations, datasets, and questions answered in the paper!

Sampled text from our model:

Following the death of Edward McCartney in 1060, the new definition was transferred to the **WDIC** of **Fullett**.
 \leftarrow novel word
 with contextually appropriate spelling

known spelling \rightsquigarrow novel spelling
sampled from its embedding

grounded	\rightsquigarrow stipped
differ	\rightsquigarrow coronate
Clive	\rightsquigarrow Dickey
Southport	\rightsquigarrow Strigger
Carl	\rightsquigarrow Wuly
Chants	\rightsquigarrow Tranquels
valuables	\rightsquigarrow migrations

How does this relate to...

Code & arXiv

	closed-vocab	open-vocab
(pure) words	Mikolov et al. (2010), Sundermeyer et al. (2012)	-impossible-
words + chars	Kim et al. (2016), Ling et al. (2015)	Kawakami et al. (2017), Hwang and Sung (2017), \star
(pure) chars	-impossible-	Sutskever et al. (2011)

<https://github.com/sjmielke/spell-once>
<https://arxiv.org/abs/1804.08205>

Twitter: @sjmielke
 Email: sjmielke@jhu.edu