The Scourge of Sparsity

In the real world, **sparsity** hurts us:

- Low vocabulary coverage in training $\rightarrow$ high OOV rate in applications (poor generalization)
- "one-hot" representations $\rightarrow$ *huge* parameter counts
- information about one word *completely unutilized* for another!

Thus, a good representation must:

- reduce parameter space,
- improve generalization, and
- somehow "transfer" or "share" knowledge between words