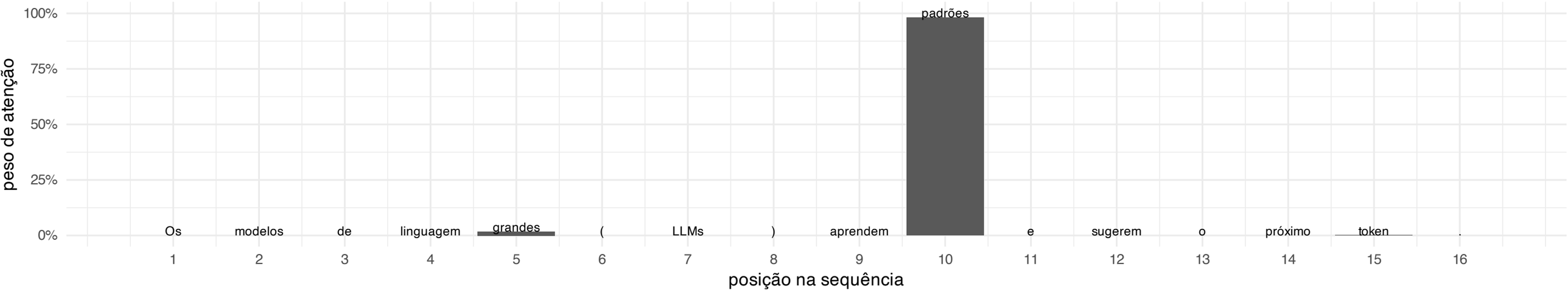
A) Sequência de tokens (após tokenização)

Os	modelos	de	linguagem	grandes	(	LLMs	)	aprendem	padrões	e	sugerem	o	próximo	token	.
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

B) Atenção do último token: "."



C) Próximo token (top-10) – distribuição toy