

Let's preserve some ordering

We can count token pairs, triplets, etc.

- Also known as n-grams
 - 1-grams for tokens
 - 2-grams for token pairs
 - ...

good movie		good movie	movie	did not	a	...
not a good movie		1	1	0	0	...
did not like		1	1	0	1	...
		0	0	1	0	...

- Problems:
 - too many features