CS 4650/7650, Lecture 5

Jacob Eisenstein

September 3, 2013

An example

In person she was inferior to both sisters

–*Persuasion*, Jane Austen

In person	*she*		*was*		*inferior*		*to*		*both*		*sisters*	
1-gram	$P(\cdot)$		$P(\cdot)$		$P(\cdot)$		$P(\cdot)$		$P(\cdot)$		$P(\cdot)$	
1	the	0.034	the	0.034	the	0.034	the	0.034	the	0.034	the	0.034
2	to	0.032	to	0.032	to	0.032	**to**	**0.032**	to	0.032	to	0.032
3	and	0.030	and	0.030	and	0.030			and	0.030	and	0.030
4	of	0.029	of	0.029	of	0.029			of	0.029	of	0.029
...												
8	was	0.015	**was**	**0.015**	was	0.015			was	0.015	was	0.015
...												
13	**she**	**0.011**			she	0.011			she	0.011	she	0.011
...												
254					both	0.0005			**both**	**0.0005**	both	0.0005
...												
435					sisters	0.0003					**sisters**	**0.0003**
...												
1701					**inferior**	**0.00005**						

An example

In person she was inferior to both sisters

–*Persuasion*, Jane Austen

2-gram	$P(\cdot\|person)$		$P(\cdot\|she)$		$P(\cdot\|was)$		$P(\cdot\|inferior)$		$P(\cdot\|to)$		$P(\cdot\|both)$	
1	and	0.099	had	0.141	not	0.065	**to**	**0.212**	be	0.111	of	0.066
2	who	0.099	**was**	**0.122**	a	0.052			the	0.057	to	0.041
3	to	0.076			the	0.033			her	0.048	in	0.038
4	in	0.045			to	0.031			have	0.027	and	0.025
...												
23	**she**	**0.009**							Mrs	0.006	she	0.009
...												
41									what	0.004	**sisters**	**0.006**
...												
293									**both**	**0.0004**		
...												
∞					**inferior**	**0**						

An example

In person she was inferior to both sisters

–*Persuasion*, Jane Austen

3-gram	$P(\cdot \mid In,person)$	$P(\cdot \mid person,she)$		$P(\cdot \mid she,was)$		$P(\cdot \mid was,inf.)$	$P(\cdot \mid inferior,to)$		$P(\cdot \mid to,both)$	
1	UNSEEN	did	0.5	not	0.057	UNSEEN	the	0.286	to	0.222
2		**was**	**0.5**	very	0.038		Maria	0.143	Chapter	0.111
3				in	0.030		cherries	0.143	Hour	0.111
4				to	0.026		her	0.143	Twice	0.111
...										
∞				**inferior**	**0**		**both**	**0**	**sisters**	**0**

An example

In person she was inferior to both sisters
–*Persuasion*, Jane Austen

4-gram	$P(\cdot \mid u,I,p)$	$P(\cdot \mid I,p,s)$	$P(\cdot \mid p,s,w)$		$P(\cdot \mid s,w,i)$	$P(\cdot \mid w,i,t)$	$P(\cdot \mid i,t,b)$
1	UNSEEN	UNSEEN	in	1.0	UNSEEN	UNSEEN	UNSEEN
...							
∞			**inferior**	**0**			

Sparsity

New words appear all the time:

- sparsistency; 65,132.14; synaptitude
- New bigrams appear even more often.
- New trigrams, etc – even worse!

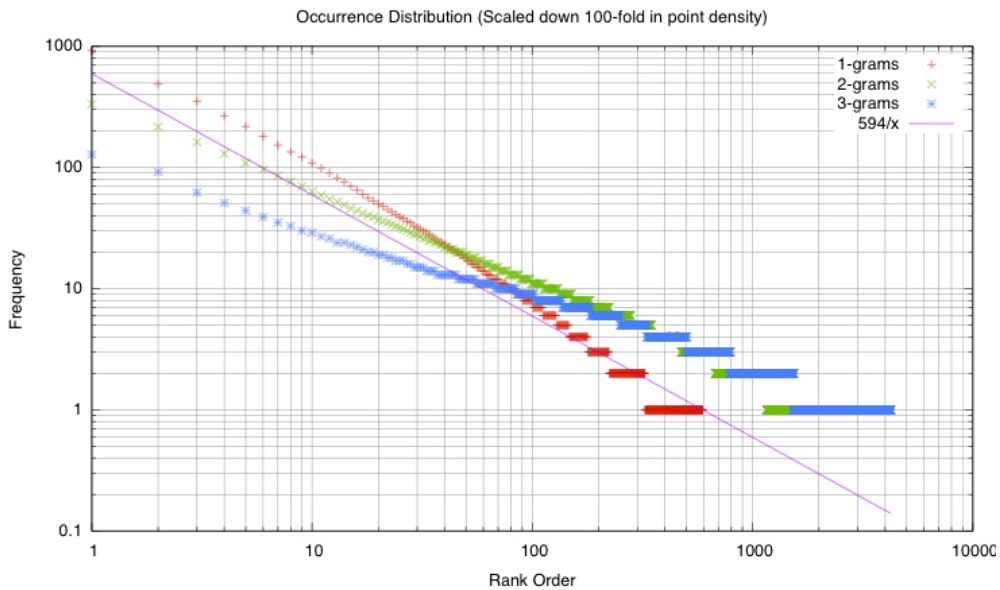
Sparsity

New words appear all the time:

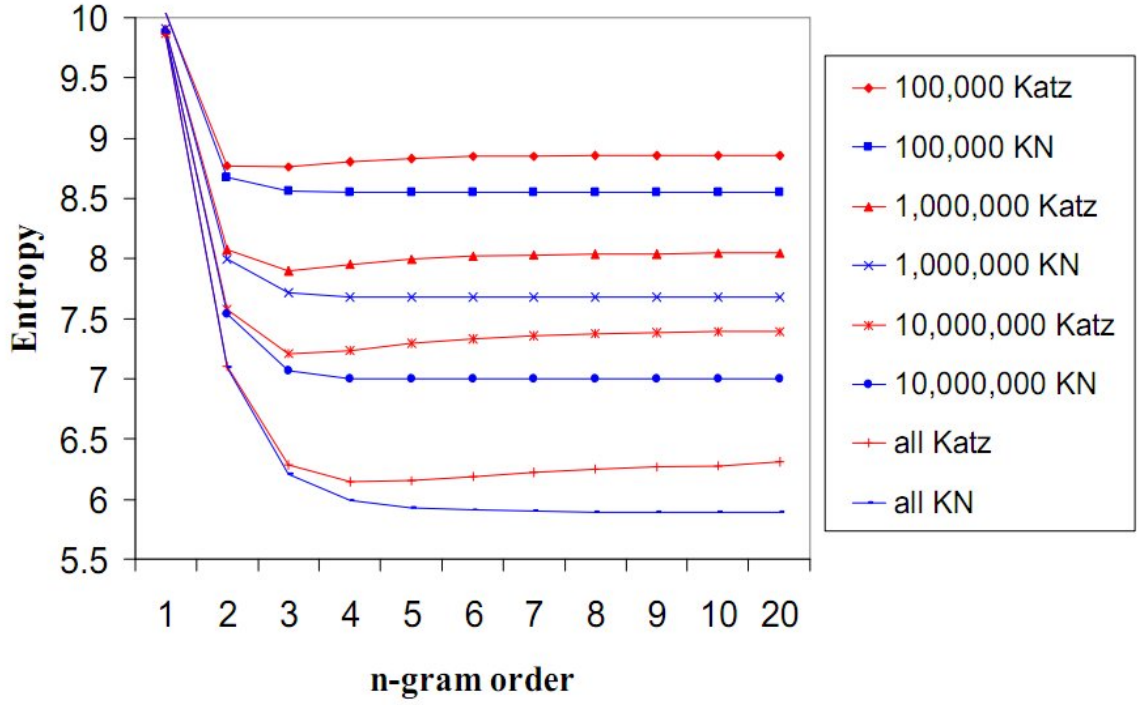
- sparsistency; 65,132.14; synaptitude
- New bigrams appear even more often.
- New trigrams, etc – even worse!

Zipf's law tells us that most word types are rare.

$$freq(\text{word}) \propto \frac{1}{rank(\text{word})}$$



Language models in practice



- Kneser-Ney is very competitive and widely used.
- Use trigrams at least — MT goes to 5-grams and beyond.
- SRILM toolkit makes it easy to play with very advanced LMs.

Language models in practice

- Smoothing controls the variance of higher-order N-gram models.
- But 5-grams are still very difficult to store.
- Recent work uses Bloom Filters to store approximate LM probabilities very efficiently.