

## *What do you say?*

What do the candidates say most?

Actually, what the candidates say most can be a little boring ... that is, most of the words that they use (like all of use) are little words like *the*, *and*, *of*, *a*, and *in*.

To be sure, we can learn **A LOT** about people from those words ... and many blogs here show that point very well.

But maybe we'd still like to have a better idea of what candidates say most that is a little more than *and* and *the* and *of*? Well, let's take a look ....

One way to assess what candidates say most is to use *clustering bi-grams*.

A bi-gram is two words occurring next to each other in the text. So, in the sentence *God, bless you ...* we have two bigrams: *God,\_bless* and *bless\_you*.

But bi-grams **may not** tell us very much either. We are left with *going\_to*, *and\_I*, *I\_think*, and *of\_the*. So what if we take the most common bi-grams – treat those bi-grams as single words, and then find the most common bi-grams again? This is called clustering. Each time we cluster, we bind the previously most commonly occurring clusters into bigger and bigger clusters.

What we find through this kind of clustering is that the most common fragments of speech emerge to identify who the candidate is and what the candidate's message is.

I wrote a little software to find out what the most common clusters are for the leading three candidates as of today: Clinton, Obama, and McCain.

In the table below, you can see how those clusters grow and grow ... from single words to very identifiable fragments. I have listed the table as Candidate 1, 2 and 3. See if you can tell who is who (Clinton, Obama, and McCain) ... the answers are written below the table.

Candidate 1				
The	going to	Going_to be	well, i_think	well, i_think
To	i think	the_american people	part_of_the reason	let me
That	that we	we're going_to	To_make_sure that_we	well, first_of_all,
And	of the	and i_think	all_across_the country.	ability_to_bring people
I	we have	Going_to have_to	and_i_think that	about_how_we're_going_to solve
Of	have to	that_we are	as_president_of_the united_states.	and_a_nation_of immigrants.
We	to be	to_make_sure_that	going_to_be able_to	as_careful getting_out
A	in the	one_of_the	It_is_important for_us_to	as_we_were careless_getting
Is	and i	for_us to	the_fact_of_the_matter	at_any_time since
In	want to	that_we can	well, first_of_all,	but the_fact_of_the_matter

<b>Candidate 2</b>				
The	You know,	and i think	well, i	well, i
To	i think	i think that	and i think that	at the end of the day,
And	and i	we're going to	we're going to have to	well, i am
I	of the	one of the	don't ask/don't tell	well, i think it's important
That	going to	we have to	balanced budget and a	well, wolf,
Of	to be	a lot of	... i want to	and, you know,
We	we have	i want to	for 35 years.	and it is clear that he
A	have to	Going to be	i voted against	back in '93 and
Have	in the	that we have	weapons of mass	balanced budget and a surplus.
Is	that we	to have a	well, i'm	based on my assessment

<b>Candidate 3</b>				
The	of the	we have to	secure the borders first.	california pollution
And	and i	we've got to	california have	a cleaner world.
To	in the	a lot of	get out of control.	ahead of us.
Of	and we	one of the	president of the united states,	and there's no such thing
I	Have to	we're going to	a chance to	as president of the united states,
That	we have	and we will	and by the way,	department of homeland security,
We	and the	president of the	as president of the	education and training
In	going to	we need to	but the fact is	end up in the
A	i think	out of control.	i'm proud of my	follow us home.
Have	that we	and we have to	i'm the only one on this stage	get out of control. We

**Candidate 1: Obama**  
**Candidate 2: Clinton**  
**Candidate 3: McCain**