

What are the main topics of the State of the Union addresses since 1900 and how do they change over time? By applying the TF-IDF algorithm - it ranks the relevance of the words of a document in a collection of documents - and analyzing the top 20 most relevant words in each decade, it is possible to identify patterns and connect them to the major events of the period.

In the first three decades of the 20th century you'll find words such as **panama**, **canal**, **isthmus**, **freight**, **interstate**, **railroads**, **processes**, **industrial** and **reclamation**. This is related to the Panama Channel, construction of railroads and, in a broader sense, to the Second Industrial Revolution.

From 1910 to 1920 there are also several words associated with the First World War. In the nineteen-tens, there are **battleships**, **submarines** and **german**. Then, in the twenties, you find words correlated to the aftermath of the war: **veterans** (2nd place in the decade), **conference** and **league** (referring to the *League of Nations*).

The thirties is all about the great depression and its consequences. In the top 20 words for the decade you will find: **recovery**, **depression**, **unemployment**, **unemployed**, **banks**, **livelihood**, **overproduction**, **livelihood** and **activities**.

In the forties, almost all the words in the list are about the Second World War. There is **japanese** (the most relevant word in the decade), **hitler**, **fighting**, **planes**, **veterans**, **axis**, **wartime** and, finally, **victory**.

With the end of the Second World War and the increasing polarization between the US and the Soviet Union, the focus of the addresses during the fifties changed to the Cold War. The second most relevant word of the decade is **communist**. There is also **korea** on the list, related to the Korean War. Finally, you can find **atomic**, **weapons**, **deterrent** and **aggression**, that can all be linked to the nuclear expansion.

In the sixties, the cold war is still a big issue, represented by the words **communist**, **soviet** and **nuclear**, but the Vietnam War is on the top of the list with the word **vietnam**.

It is worth noting that there is no word in the lists that correlates to a very relevant issue of the fifties and sixties: the Civil Rights Movement.

The energy crisis of the seventies and its consequences can be linked to some of the words of the decade: **inflation**, **oil**, **jobs** and **spending**. The words **nuclear** and **strategic** also appear on the list and can be related to Non-Proliferation efforts.

In the eighties, the cold war is still present in the words **soviet**, **afghanistan** and **nuclear**, but economic issues gets prominence, linked to the words **spending**, **budget**, **deficits** and jobs.

There is no obvious major event that can be linked to the words for the nineties but the appearance of **parent**, **kids**, **child** and **college** suggests that education was a recurrent theme. Surprisingly, there is no mention of the Gulf War.

Finally, the first decade of the 21st century is monothematic. Practically all the words in the list are related to the War on Terror: **al**, **qaeda**, **homeland**, **terrorists**, **iraq**, **iraqi**, **afghanistan**, **saddam**, **hussein**, **iraqis**, **terror** and **terrorists**.