https://doi.org/10.6092/issn.2611-2752/13995
ISSN 2611-2752

Representations of Citizens/hip in 230 Years of American History. A Diachronic Corpus-assisted Approach

Cinzia Bevitori*

Anna Marchi**

University of Bologna (Italy)

Submitted: December 10, 2021 – Revised version: December 10, 2021
Accepted: December 18, 2021 – Published: April 7, 2022

Abstract

The paper examines how American presidents have discursively constructed citizens (and citizenship) over more than two hundred years of American political history from an interdisciplinary perspective. As one deeply contested concept in different political arenas (Wiesner et al 2017), involving aspects of collective identities, citizens/hip has been at the very heart of Western democracies since ancient times, although its significance has gradually increased in modern periods (Marshall 1950, Bayley et al. 2013).

Keywords: Citizenship; US Presidents; American Political History; Corpus Linguistics; Political Discourse.

Cinzia Bevitori: University of Bologna (Italy)

https://orcid.org/0000-0002-1288-4371

cinzia.bevitori@unibo.it
Cinzia Bevitori is Associate Professor at Bologna University's Department of Interpreting and Translation and teaches English language and linguistics in the first cycle degree International and Diplomatic sciences, and Analysis of Political Language in the second cycle degree Mass Media e Politics. Her main research interests involve corpus research methodologies, corpus-assisted (critical) discourse studies, focussing in particular on institutional, political and media discourse, within the framework of systemic functional linguistics. Amongst her recent publications in the field are *Construing justice: Discourses of 'rightness' in the House of Commons: A diachronic corpus-assisted discourse study* (de Gruyter 2020); *'in a world of complex threats...': Discourses of in/security in the State of the Union Address (1790–2014)*, in Critical Approaches to Discourse Analysis across Disciplines, 2017; *Diachronic change from Washington to Obama: the challenges and constraints of corpus-assisted meaning analysis*, (with Paul Bayley, Equinox, 2016).

Anna Marchi: University of Bologna (Italy)

https://orcid.org/0000-0003-1284-6356

anna.marchi@unibo.it
Anna Marchi is a senior assistant professor of English language and linguistics at Bologna University's Department of Interpreting and Translation. Her research interests include corpus-assisted discourse methods, news discourse and political discourse. Among her recent publications: *Self-reflexive Journalism: A Corpus Study of Journalistic Culture and Community in the Guardian* (Routledge 2019), *Corpus Approaches to Discourse: A Critical Review* (Routledge 2018) and "Churchill's inter-subjective special relationship: A corpus-assisted discourse approach" (in Dobson & Marsh eds. *Churchill and the Anglo-American Special Relationship*, Routledge 2017). She studied journalism at the University of Siena and at the University of Texas and earned a PhD in English language and linguistics at Lancaster University. She has worked on international interdisciplinary projects at the universities of Siena, Cardiff, Swansea and Lancaster.

*
cinzia.bevitori@unibo.it

**
anna.marchi@unibo.it

Copyright © 2022 Cinzia Bevitori, Anna Marchi
This work is licensed under the Creative Commons BY License.
https://creativecommons.org/licenses/by/4.0/

1 Introduction

This paper explores how citizens/hip has been discursively constructed by US Presidents over two centuries of American political history. As one deeply contested concept in different political arenas,[1] involving aspects of collective identities, both notions of citizens and citizenship have been at the very heart of Western democracies since ancient times, although their significance has gradually increased in modern periods.[2] As a key component of western political thinking since the classical Greek and Roman periods, historically the notion is seen to enshrine both the ownership of a formal status of membership to a legal and political entity, as well as the rights and obligation within it.[3]

In spite of its classical origins, however, most historians agree that citizenship is an "essentially modern concept," whose meaning has seen several major changes and transformations over time.[4] In particular, while the beginning of the modern era is seen to coincide with the two great revolutions, i.e. the American and French Revolutions in 1776 and 1789 respectively, that process was largely anticipated by revolutionary England in the 1600s, which was clearly influenced by the philosophy of political thinkers such as Locke, Mills, and Rousseau, as well as by their American counterparts Jefferson and Washington, the founders of the liberal democratic ideals of the modern nation-state. What is more, in his seminal study on *Citizenship and Social Class*, Marshall convincingly argues that citizenship is a "developing institution" which "create[s] an image of an ideal citizenship," thus highlighting the inherent struggle in the negotiation of values, ideas and beliefs across and within political cultures in their development.[5] Although Marshall's analysis is focused on citizenship rights in Great Britain, his study undoubtedly marked a turning point in subsequent interpretations of the concept. Not only is the fundamental principle of social equality in membership introduced here for the first time, but what is also emphasized is the close interdependence between ownership of capital (beyond land), with the process of nation-building, national identity and belonging.

In the United States, however, this process was more complex due to a number of factors; i.e. the strong connection with the British cultural heritage, the formation of a new state in the nation-making experience, the system of federalism of the new-born state, alongside the creation of a "*politically* homogeneous people," arriving from different cultural, ethnic, and ideological backgrounds.[6] However, it should be noted that the ambiguity in the usage of terms such as 'people' and 'citizens' is at the very heart of the nation's history since the adoption of the founding documents. According to Wang, the history of American citizenship can be divided into three phases, all of which involve major change in the process of state-building: The Founding Era (1776–1860), the Civil War and Reconstruction (1861–1916), and the New Deal-Great Society Era (1933–1960).[7] Indeed, the 1960s is recognized as the age of rights, and it should be mentioned that the 'Great Society' era is associated with the presidency of Lyndon Johnson. Yet recognizing and tracking the pace of change in the construction of discourses on national citizenry that span long historical cycles of American history may require sophisticated tools of analysis able to bridge the gap between language change and wider historical and socio-political contexts. The aim of this paper is thus two-fold. Firstly, it will try to investigate how US presidents have discursively represented citizenship, and in particular, citizen/s, in more than two

1. Claudia Wiesner, Anna Björk, Hanna-Mari Kivistö, Katja Mäkinen, *Shaping Citizenship. A Political Concept in Theory, Debate and Practice* (London: Routledge, 2018).

2. Thomas H. Marshall, *Citizenship and Social Class and Other Essays* (Cambridge: Cambridge University Press, 1950); Paul Bayley, Giuliani Delphine and Vanessa Serret, "Semantic constructions of citizenship in the British, French and Italian press," in *European Identity. What the Media say*, eds. Paul Bayley and Geoffrey Williams (Oxford: Oxford University Press, 2012), 153–189.

3. Richard Bellamy, "Historical development of citizenship," in *International encyclopedia of the social & behavioral sciences. 2nd edition*, Vol. 3, ed. James D. Wright (Oxford: Elsevier, 2015), 643–649.

4. Bryan S. Turner, *Citizenship and Social Theory* (London: Sage, 1993), VII.

5. Marshall, *Citizenship and Social Class*, 149.

6. Samuel P. Huntington, *Who are We? The Challenges to America's National identity* (New York: Simon Schuster, 2004); Xi Wang, 2010, "Citizenship and Nation-Building in American History and Beyond," *Social and Behavioural Sciences*, 2(2010): 7017–7029.

7. Ibid.

hundred years of American political history; secondly, the paper will attempt to demonstrate how a diachronic corpus-assisted discourse analysis can productively contribute to the study of social change and political representation in long cycle narratives of American political culture.

2 Corpus and Methods

2.1 The State of the Union diachronic Corpus (1790–2020)

As one of the canonical forms through which presidential power takes shape,[8] the State of the Union Address (henceforth SoU) may provide a diachronically relevant and productive territory of investigation. As previous research has suggested, this constitutionally-mandated[9] annual message, or address, delivered to joint sessions of Congress may in fact be regarded as a powerful rhetorical tool that presidents have exploited over time to perform a number of different tasks; i.e. to report and inform on the developments of their administration, to set the political agenda for the year to come, as well as to increasingly engage with Congress and engage with the public. While no clear indication is given in the Constitution as to regularity, the SoU address has been delivered on a yearly-basis since its inception, with a few exceptions in 1841 and 1933, while in 1790, 1953, 1961 and 1981 messages were delivered twice. As regards its delivery mode, it was given orally between 1790 and 1800, whereas for a long period, from 1801 to 1912, SoUs were written texts. Revived orally by President Wilson in 1913, it was broadcasted by Coolidge for the first time in 1923, which was made a permanent feature with FDR in 1934, with some exceptions again (1945, 1946, 1956, 1961, 1973 and 1981). A further technological advance was made with Truman in 1947, delivering for the first time a televised address. Its mediatization, especially in more recent years, is one of the main factors contributing to its transformation over time.[10]

However, while scholars in presidential rhetorical studies have largely conceptualized the history of the presidency and its addresses in dichotomic terms, thus suggesting a traditional/modern divide,[11] others have started questioning the presidency's 'evolutionary shift' from different perspectives.[12] For example, Teten addresses the issue by adopting a broader perspective with the aim of challenging previous narratives of presidential rhetoric. Advocating a more consistent and thorough representation of the presidency, his study expands the timeline back to the founding period in order to shed light on the evolutionary development of this form of address by identifying recurrent patterns of "rhetorical innovation."[13] Indeed, the need to theorize and assess the role of the rhetorical presidency through paying close attention to long term cycles had already been encouraged by a number of scholars in the field.[14] Taking further steps, empirical studies have demonstrated how the use of computational tools can provide a more systematic and objective approach to the investigation of corpora that span long

8. Richard E. Neustadt, *Presidential Power. The Politics of Leadership* (New York: Wiley, 1960).

9. Article II Section 3 of the US Constitution reads that the President: "shall *from time to time* give to the Congress Information of the State of the Union, and recommend to their Consideration such Measures as he shall judge necessary [...]".

10. Paul Bayley, Cinzia Bevitori, "Diachronic change from Washington to Obama: the challenges and constraints of corpus-assisted meaning analysis," in *Systemic Functional Linguistics in the Digital Age*, eds. Sheena Gardner and Sian Alsop (Sheffield and Bristol: Equinox, 2016), 231.

11. Jeffrey K. Tulis, *The Rhetorical Presidency* (Princeton, NJ: Princeton University Press, 1987); Terri Bimes, Stephen Skowronek, "Woodrow Wilson's Critique of Popular Leadership: Reassessing the Modern-Traditional Divide in Presidential History," in *Speaking to the People the Rhetorical Presidency in Historical Perspective*, ed. Richard J. Ellis (Amherst, MA: University of Massachusetts Press, 1998).

12. Leroy G. Dorsey, Introduction to *The Presidency and Rhetorical Leadership, ed.* Leroy G. Dorsey (College Station, TX: Texas A& M. University Press, 2002), 3–19; Ryan Lee Teten, *The Evolutionary Rhetorical Presidency. Tracing Changes in Presidential Address and Power* (New York: Peter Lang, 2011).

13. David Zarefsky,"The Presidency Has always been a place for Rhetorical Leadership," in *The Presidency and Rhetorical Leadership*, ed. Leroy G. Dorsey (College Station, TX: Texas A&M University Press, 2002), 39

14. Vanessa B. Beasley, *You the people: American National Identity in Presidential Rhetoric* (College Station, TX: Texas A&M University Press, 2004); Stephen E. Lucas, "George Washington and the Rhetoric of Presidential Leadership," in *The Presidency and Rhetorical Leadership*, ed. Leroy G. Dorsey (College Station, TX: Texas A&M University Press, 2002).

historical cycles from different methodological and theoretical perspectives.[15] As argued elsewhere,[16] a corpus-assisted methodology can help trace shifts in discursive patterns over long periods of time in this specialized and powerful domain of American political culture.

In order to answer our research questions, a corpus composed of the complete set of 234 transcribed SoU texts between 1790 (President Washington) to 2020 (President Trump), amounting to approximately 1,830,000 running words, will be employed. Figure 1 provides the breakdown of the number of words of each individual president in office from Washington to Trump. As can be observed from the graph (Fig. 1), the sub-corpora are of different size both in terms of number of texts (depending also on the terms in office of each president), as well as word counts; i.e. lengths of texts, which of course, has relevant implications in a comparative approach (see Section 3). While it is evident that the average number of words tends to be higher in the traditional period, which may be partly explained by its written delivery mode, the longest message in presidential history is President Carter's 1981 address, with over 33,000 words, although Theodore Roosevelt's sub-corpus appears as clear spike, as this alone makes up 10% of the entire corpus.

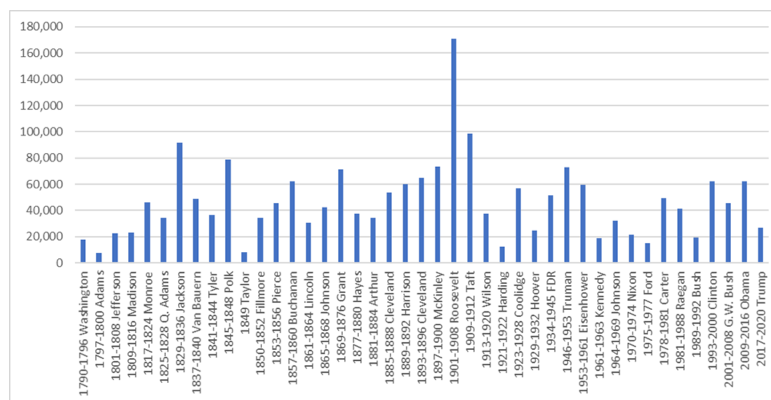

Figure 1. Size (number of words) of each president's sub-corpus.

For this study we set out by experimenting with different temporal segmentations, such as Skowronek's cycles,[17] but we then abandoned theory-informed units of time, to favour a bottom-up approach. We analysed the word *citizens* in each president's sub-corpus and then compared findings to aggregate them on the basis of continuity. In fact, the segmentation of a continuous timeline in discrete periods has a major impact on the results[18] and adopting a data-driven aggregation of variables allows us to then critically examine alternative theory-informed units of time. As will be more thoroughly explained in the following section (2.2), one of the strengths of Corpus-Assisted Discourse Studies (CADS) lies not simply in synergically combining quantitative and qualitative methods of analysis, but also in bringing together different procedures. The paper will attempt to

15. Elvin T. Lim, "Five trends in presidential rhetoric: An analysis of rhetoric from George Washington to Bill Clinton," *Presidential Studies Quarterly*, 32(2002): 328–366; Paul Bayley, Cinzia Bevitori, "Two centuries of 'security': Semantic variation in the State of the Union address (1790–2014)," in *Gentle Obsessions: Literature, Linguistics and Learning in honour of John Morley*, eds. Alison Duguid, Annz Marchi, et.al (Rome: Artemide, 2015), 60–80; Bayley, Bevitori, "Diachronic change from Washington to Obama."; Cinzia Bevitori, "'In a world of complex threats...': Discourses of in/security in the State of the Union Address (1790–2014). A diachronic corpus-assisted study," *Critical Approaches to Discourse Analysis across Disciplines*, 8(2017): 19–36; Alix Rule, et al. "Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014," *Proceedings of the National Academy of Sciences of the United States of America*, 112(2015).

16. Bayley, Bevitori, "Diachronic change from Washington to Obama."

17. Stephen Skowronek, *Presidential Leadership in Political Time: Reprise and Reappraisal* (Lawrence, KS: University Press of Kansas, 2011).

18. Anna Marchi, "Time boxes: epistemological, methodological and practical impact of diachronic segmentation," in *Corpus Approaches to Discourse: A Critical Review*, eds. Charlotte Taylor and Anna Marchi (London: Routledge, 2018), 174–196.

show how through a fine-grained analysis of the selected lemma[19] as a first step, we can empirically map fluctuations of continuity and change of political discussions of citizenship over time.

2.2 Methods and tools

The broad methodological framework adopted for this article is Corpus Linguistics,[20] defined in very general terms as the study of language based on the use of large collections of corpora and on computational tools to interrogate them. More specifically, the analysis here takes a Corpus-Assisted Discourse Studies perspective: what distinguishes CADS[21] from Corpus Linguistics at large is primarily its aim, as CADS is interested in investigating language from a social as well as a structural point of view and explores specific discourse types (in this case SoU addresses) to ask questions that may arise in other (non-linguistic) disciplines. To this aim, CADS combines the quantitative rigor of Corpus Linguistic techniques, with more typical qualitative approaches to discourse analysis, as well as with input from other disciplines, which may provide ideas for the research design and/or guidance for the interpretations of findings. In this sense, CADS is often described as having an 'exploratory', 'omnivorous' and 'intrinsically interdisciplinary' nature.[22]

The analytic process typically moves from general towards particular: from frequencies and distributional information to close reading of concordance lines and individual texts. The concordance is the most fundamental tool of corpus work, by concordance we mean the collection of all the instances containing a target word in the corpus (Fig. 2). The concordance defines the essence of how we look at corpora: while we read texts horizontally, we read concordances vertically and the juxtaposition of fragments, by eclipsing the specificity of individual texts, makes patterns visible and countable. So concordances are the place where quantitative analysis begins, through the analysis of repeated co-occurrences, that is lexical, grammatical and syntactic features (e.g. modal verbs, possessive pronouns) that frequently appear in the same textual environment as our target word (e.g. *citizens*) and thus create and color its meaning. At the same time the broader context and the individual sources are not lost, and the concordance is also the place where the qualitative analysis begins, as it is always possible to zoom into the data and refine the granularity of the analysis, without losing its representativeness and generalizability.

Two concordancing software suites were used in this analysis: Wordsmith 8[23] and SketchEngine.[24] Wordsmith was used to compute frequency lists and track the distribution of lexical items over time and across sub-corpora (i.e. the separate collections of addresses for each president). SketchEngline, which is probably the most sophisticated suite of tools currently available for corpus research, was used for collocation analysis. The concept of collocation is closely related to that of co-occurrence: the repeated co-occurrence of word pairs is termed collocates. The principle of collocation revolves around the notion that the meaning of a word is defined by the relationships that word establishes with other words it "keeps company with:" "collocative meaning consists of the associations a word acquires on account of the meanings of words which tend to occur in its environment."[25] The distinction between a simple co-occurrence and a collocate[26] is that the latter implies the significance of the relationship between the two co-occurring items, significance which can be measured by employing

19. A lemma is a linguistic form that includes all inflections of a specific word in the results; in the case of a noun this means both singular and plural.

20. See for example Tom McEnery, Andrew Hardie, *Corpus Linguistics* (Cambridge: Cambridge University Press, 2012).

21. Alan Partington, "Corpora and discourse, a most congruous beast," in *Corpora and Discourse, eds. Alan Partington et al. (*Bern: Peter Lang, 2004), 11–20

22. Anna Marchi, Nuria Lorenzo-Dus and Steve Marsh, "Churchill's inter-subjective special relationship: a corpus-assisted discourse approach," in *Churchill's special relationship: commemorating the 70th anniversary of the Fulton Iron Curtain speech*, eds. Steve Marsh and Alan P. Dobson (Oxford: Oxford University Press, 2017), 174.

23. Mike Scott, *WordSmith Tools version 8* (Stroud: Lexical Analysis Software, 2020).

24. Adam Kilgarriff, Vit Baisa, et al., "The Sketch Engine: ten years on," *Lexicography*, 1(2014): 7–36.

25. Geoffrey Leech, *Semantics* (London: Penguin, 1974), 20.

26. Michael Hoey, *Patterns of Lexis in Text* (Oxford: Oxford University Press, 1991).

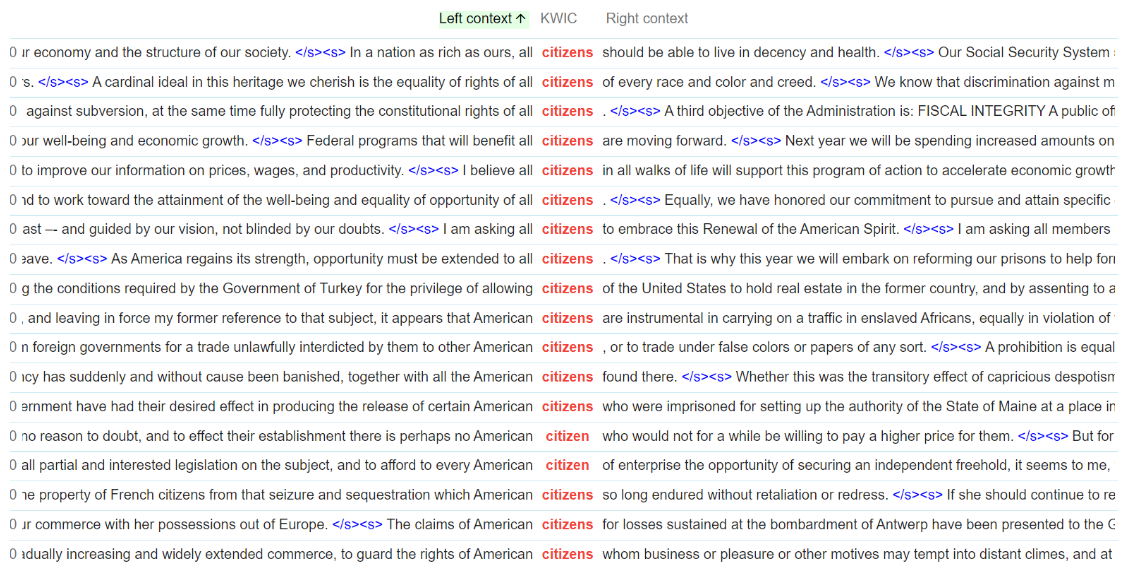

Figure 2. Sample of the concordance of *citizens* in the SoU corpus, sorted alphabetically one word to the left.

different metrics. For this study we intersect two measures: Mutual Information (MI) and T-score. Mutual Information measures the strength of the association, it tends to favour low-frequency words, and it is not affected by corpus size (therefore, it is comparable across sub-corpora), the T-score is statistical significance measure, i.e. it tests the level of confidence that an association between the two words exists, it tends to show high-frequency associations, and it is sensitive to corpus-size. The choice of metrics is followed by the choice of parameters, in this case set at a minimum frequency of 4 co-occurrences in the sub-corpus, a MI of 4 or higher and a T-score of 2 or higher. The calculation of collocates gives us access to the so-called 'profile' of a word. Here it is the emergent recurring patterns of meanings associated with the word *citizens*. SketchEngine is a tool specifically designed to visualize profiles, or, as the software labels them 'sketches' (Fig. 3).

As the analysis that follows will demonstrate, a word sketch is a starting point: a way to enter the data, to then delve into and account for further levels of complexity through the close reading of concordance lines and segments of the texts.

3 Quantitative Survey of *Citizens*

The lemma citizen occurs 2,194 times in the corpus (1.2 times every 1000 words), mostly in the plural form (85% of occurrences). *Citizenship* is, instead, rather uncommon, with just 167 occurrences in the entire corpus. Unsurprisingly the word *citizens* is considerably less frequent than *people*, which occurs approximately twice as much in the dataset and is known to be the second most frequent noun in general English and often the first one in political language corpora,[27] but it is twice as frequent as *Americans* if we consider the corpus in its entirety. What is particularly interesting in comparing these three potentially alternative lexical choices is how their distribution varies over time (see Fig. 4 showing absolute frequencies, and Fig. 5 displaying the trend of frequency relative to corpus size). The data show that, in this particular text type, *people* has not always been the top noun: for the first forty years of American history *citizens* was more frequent and the two words remained more or less even for the first 100 years (almost overlapping Wang's "Founding era"). Even though both Lincoln and Andrew Johnson used the word *people* respectively almost twice and three times as often as *citizens*, it is with Cleveland's first term that *people*'s dominance becomes systematic and monumental. Another significant change occurs towards the other end of our timeline: in the last 60 years of SoU the

27. Bayley, Bevitori, "Diachronic change from Washington to Obama."

modifiers of "citizen"		
fellow (of our fellow citizens)	165	12.3
American (of American citizens)	189	11.0
own (of our own citizens)	54	9.5
individual (the individual citizen)	20	9.2
distinguished (a distinguished citizen of)	16	9.2
patriotic (every patriotic citizen)	16	9.1
good (all good citizens)	47	9.1
private (private citizens)	24	9.1
senior (senior citizens)	9	8.4
injured (our injured citizens)	8	8.3
Mexican (Mexican citizens)	9	8.1
many (many citizens)	15	7.9

verbs with "citizen" as object		
naturalize (naturalized citizens)	27	11.1
become (become citizens of)	33	10.3
protect (to protect our citizens)	24	9.9
engage (of our citizens engaged in)	14	9.6
restrain (restrain our citizens from)	5	8.7
indemnify (indemnifying our citizens)	4	8.5
serve (to serve our citizens)	4	7.9
concern (citizens are concerned)	4	7.8
leave (left our fellow citizens)	4	7.3
be (are citizens)	41	7.1
give (to give our citizens)	4	6.2

verbs with "citizen" as subject		
sustain (sustained by citizens)	10	9.3
suffer (citizens suffer)	10	9.3
resort (citizens who resort)	5	8.5
reside (citizens residing in)	5	8.4
hold (be held by citizens)	6	8.4
do (citizens do)	10	8.1
inhabit (citizens inhabiting)	4	8.1
desire (citizens who desire to)	4	8.1
deserve (citizens deserve)	4	8.0
have (citizens have)	83	7.9
work (citizens working)	4	7.7
make (citizens who have made)	7	7.4

Figure 3. Word-sketch of *citizens* in the SoU corpus.

term *Americans* is employed considerably more frequently than *citizens*. Every president after Lyndon Johnson uses it between three and five times as much as *citizens* (Nixon 19 times more, on account of mentioning *citizens* just two times in his five addresses). In Trump's discourse the use of *people* and *Americans* is numerically even.

If we consider only the trend of *citizens* (the blue line in Fig. 5), the data tell us that the decline of the term started with Harrison, there was an uptake with Truman first and then, more visibly, with Eisenhower and a recent spike which started with Clinton, peaked with George W. Bush, plummeted again with Obama, to then find another high with Donald Trump.

It is essential to consider both relative and absolute frequencies when making any generalizations because, as mentioned in Section 2.1, there are remarkable differences in the size of individual presidents' sub-corpora: Theodore Roosevelt's sub-corpus, is ten times as large as George Washington's and six times as large as Donald Trump's, which makes them impossible to compare in absolute terms. At the same time, because of the limited size of some sub-corpora, high relative frequencies may still correspond to negligible quantities. For example, Adams would have the third highest relative frequency of mentions of *citizens* (Table 1), but this corresponds to just 14 occurrences—and it is, therefore, necessary to always take into account absolute frequencies.

Early presidents, in particular Washington and Jefferson, are the ones in whose addresses the word *citizens* features most saliently, while Jackson and Grant are the ones who use it most frequently in absolute terms (their joint mentions make up for 10% of the total). Polk and Buchanan are the presidents who used the word more times in most of their SoU addresses (three out of four). Only a closer analysis of collocates and concordances (next section) can tell us *how* the word was used, but even mere quantification may offer interesting insights, not only in terms of presence, but also in terms of absence.[28] For instance we may wonder why the word *citizens* featured so prominently in certain addresses delivered in the second half of the nineteenth century, but apparently did not in correspondence with the discussion and ratification of the 14th Amendment (1866 and 1868), which enshrines the rights of citizens in the Constitution.

28. Alison Duguid, Alan Partington, "Absence. You don't know what you're missing. Or do you?," *Corpus Approaches to Discourse*, eds. Paul Baker and Tom McEnery (London & New York: Routledge, 2018), 38–58.

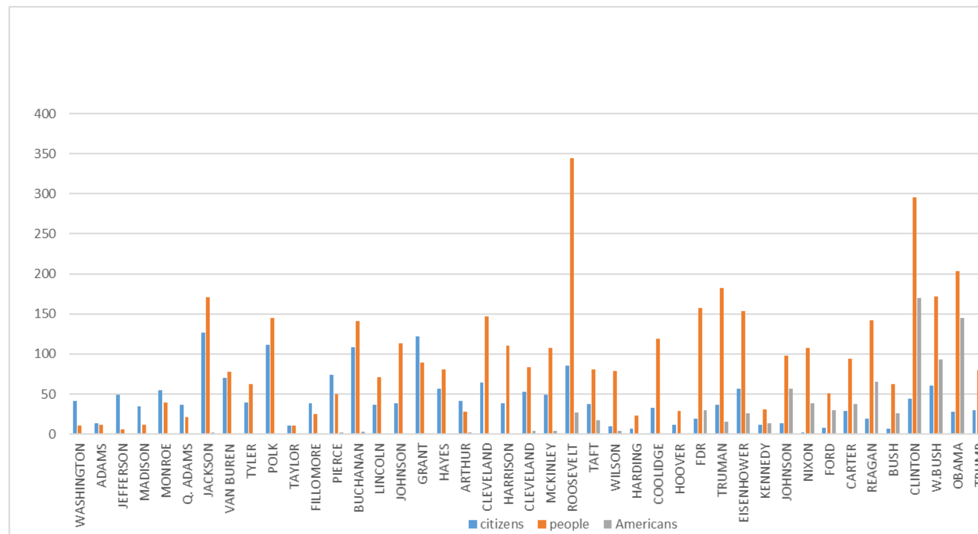

Figure 4. Absolute frequencies of *citizens*, *people* and *Americans* by president.

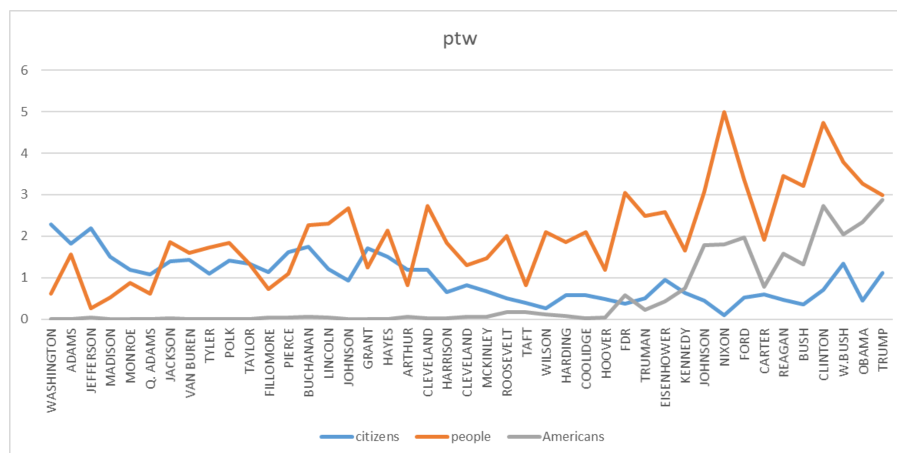

Figure 5. Normalized frequencies (per thousand words) of *citizens*, *people* and *Americans* by president.

PRESIDENT	hits	ptw	PRESIDENT	hits	ptw	PRESIDENT	YEAR	hits	ptw
JACKSON	**127**	1.3	WASHINGTON	41	**2.2**	JACKSON	**1834**	24	1.7
GRANT	**122**	1.7	JEFFERSON	49	**2.1**	VAN BUREN	**1838**	26	2.2
POLK	**111**	1.4	ADAMS	14	**1.8**	POLK	**1845**	26	1.6
BUCHANAN	**109**	1.7	BUCHANAN	109	**1.7**	POLK	**1846**	40	2.2
ROOSEVELT	**86**	0.5	GRANT	122	**1.7**	POLK	**1847**	33	2.1
PIERCE	**74**	1.6	PIERCE	74	**1.6**	PIERCE	**1856**	29	2.7
VAN BUREN	**70**	1.4	HAYES	57	**1.5**	BUCHANAN	**1858**	38	2.3
CLEVELAND	**64**	1.1	MADISON	35	**1.5**	BUCHANAN	**1859**	25	2
G. W. BUSH	**61**	1.3	VAN BUREN	70	**1.4**	BUCHANAN	**1860**	33	2.3
HAYES	**57**	1.5	POLK	111	**1.4**	GRANT	**1875**	22	1.8

Table 1. The different ranking of the top 10 frequency of *citizens* in the sub-corpora, respectively by absolute frequency by president (left column), relative frequency by president (middle column) and number of mentions in individual State of the Union addresses (right column).

A corpus approach allows us to observe language data (texts) in ways that are not available to "the naked eye"[29] and this, in turn, may generate new, unexpected questions, as well as providing anchoring for our findings. The next section will continue the "serendipitous journey"[30] following *citizens* through concordance lines and, now that we have established how much *citizens* are talked about, it will explore how they are discursively represented.

4 Collocation Analysis of *Citizens*

If we look at the collocates of *citizens* taking into consideration the corpus in its entirety, we may get a general idea of the patterns that define the meaning of the word in the context of the SoU. However, because of its uneven distribution these patterns may be quantitatively dominant, but not necessarily pervasive. When analyzing collocational profiles from a diachronic perspective we need to identify which collocates are stable over time, these are known as "consistent collocates,"[31] and which collocates characterize the discourses of particular presidents or periods. Table 2 reports a selection[32] the top collocates of *citizens* in the whole corpus, none of the collocates in this list is shared by all the 45 presidents, as shown graphically in Figure 6.

Collocate	freq	T-score	MI
fellow	155	12.43	9.29
naturalized	28	5.28	8.94
claims	104	10.13	7.31
property	55	7.34	6.52
American	200	13.95	6.19
rights	53	7.14	5.66
become	32	5.53	5.48
United	208	14.09	5.42
our	703	25.89	5.4
protection	29	5.23	5.2
own	57	7.29	4.91
good	42	6.26	4.87
against	47	6.6	4.76

Table 2. Top collocates of *citizens* in the SoU corpus.

Fellow, for example, is the collocate with the strongest association to *citizens* in the corpus as a whole, but it characterizes only early presidents, from Washington to Jackson, and, to a minor extent more recent ones: Reagan, Clinton, George W. Bush, Obama and Trump. This is due to a relative absence of the word *fellow* from extended sections of the corpus, in fact fifteen presidents never utter the word *fellow* in their addresses. Similarly, the word *naturalized*, another statistically significant collocate of citizens (from a synchronic perspective), is anything but consistent from a diachronic point of view: it does not appear until Johnson and it permanently disappears after McKinley, and in fact characterizes the speeches of only four presidents out of 45. The timeline of absences reflected in the empty spaces of Figure 6 demonstrates that the majority of the top collocates of *citizens*, with perhaps the sole exception of the possessive pronoun *our*, are very specific to a limited number of presidents,

29. John M. Sinclair, *Corpus, Concordance, Collocations* (Oxford: Oxford University Press, 1991).

30. Alan Partington, "Evaluating evaluation and some concluding thoughts on CADS," in *Corpus-Assisted Discourse Studies on the Iraq Conflict: Wording the War*, eds. John Morley and Paul Bayley (London and New York: Routledge, 2009), 261–303.

31. Costas Gabrielatos, Paul Baker, "Fleeing, sneaking, flooding: a corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996–2005," *Journal of English Linguistics* 36(2008): 12.

32. The selection is based on the application of the cut-off points for statistical measures illustrated in 2.2, an additional cut-off for frequency to 1 co-occurrence every 100 occurrences of citizens, and the exclusion of some functional (i.e. quantifiers and referents) words following a preliminary review of concordances.

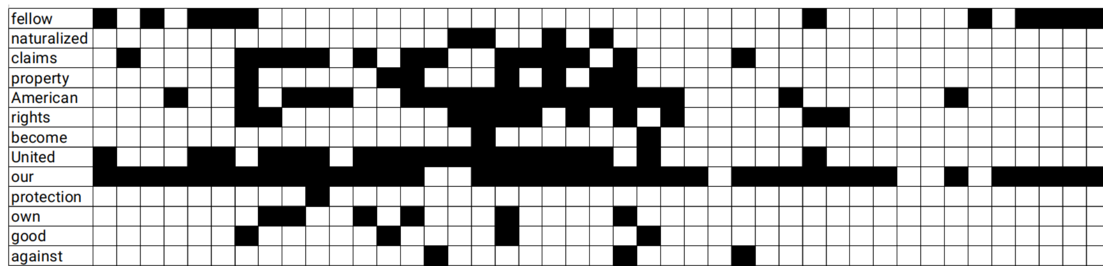

Figure 6. Consistent-collocates on the timeline (each box representing a president's sub-corpus).

which suggests that the discourses surrounding the word may have changed considerably over time. Yet, again as a consequence of distribution, the variation may also simply be an effect of quantity (i.e. different amount of mentions) rather than quality (i.e. different discourses). Only an inspection of the collocational patterns for each president and the close reading of concordance lines may test the hypothesis of diachronic change and allow us to aggregate the data in phases.

Three broad semantic domains emerge from the study of the individual sub-corpora, which overlap, intertwine, and recur over time, but tend to neatly cluster in rather cohesive periods, when each of the following areas of meaning is dominant:

- Protection of citizens: from Washington to Buchanan (1790–1860), and a revival from Cleveland's second term to McKinley (1893–1900), from Taft to Hoover (1909–1932) and finally with George W. Bush and Trump.
- Qualities of citizens: most visible in the Roosevelt era (1901–1908), with a marked comeback with Clinton's and Obama's discourses.
- Rights of citizens: from Lincoln to Harrison (1861–1892), from Franklin Delano Roosevelt to Clinton and, to a minor extent, Obama (1933–2015).

Figure 7 offers an overview of the distribution of categories over the timeline and highlights the overlaps. Since any classification system is necessarily reductive, the segmentation, however, allows us to visualize a progressive move from the context of protection towards a context of qualities and, more markedly, rights. It is, nevertheless, important to highlight that the empty boxes do not imply that meaning is entirely absent, but that it is not dominant.

Figure 7. Dominant semantic domains (each box representing the sub-corpus of a president's discourses).

The discussion that follows will analyze each category in further depth and explain the nuances that a semantic domain may acquire at different times, for example, examining how the concept of protection encapsulates different types of security (economic and physical), or how rights evolve over time, moving from the rights of naturalized citizens (Lincoln to Harrison) to social and civil rights (from FDR onwards).

4.1 Protection of Citizens

Along with security, *protection* may undoubtedly be regarded as a 'cultural keyword' in American political discourse. While the 'protection of citizens' is not mentioned in the Constitution as one of the duties of the president,[33] it is apparent that citizens are construed as the beneficiaries of protection in

33. Art. II Section 1 of the Constitution reads: "preserve, *protect*, and defend the Constitution of the United States."

these presidential messages, particularly in the span of time between the early period and the Civil War. However, on closer inspection, co-occurrence patterns of the items *citizens* and *protection* reveal some nuances about who should be protected, how, and against what. Unsurprisingly, the most frequent pattern relates to the safeguarding of 'lives and property', which the Founding Fathers, notably influenced by the philosophical ideals of John Locke (and other philosophers), considered as one of the cornerstones of a free society. In our material, 'property' emerges as the most frequent and stable associate of 'protection' in the addresses of Jackson, Polk, Pierce, and Buchanan:

> It is sufficient on the present occasion to say that the wanton violation of the rights of **person and property of our citizens** committed by Mexico, her repeated acts of bad faith through a long series of years, and her disregard of solemn treaties stipulating for indemnity to **our injured citizens** not only constituted ample cause of war on our part, but were of such an aggravated character as would have justified us before the whole world in resorting to this extreme remedy (Polk 1847).
>
> Now, if this conditional and contingent power could be constitutionally conferred upon the President in the case of Paraguay, why may it not be conferred for the purpose of **protecting the lives and property of American citizens** in the event that they may be violently and unlawfully attacked in passing over the transit routes to and from California or assailed by the seizure of their vessels in a foreign port? To deny this power is to render the Navy in a great degree useless for the **protection of the lives and property of American citizens** in countries where neither protection nor redress can be otherwise obtained (Buchanan 1859).

This notion reappears later on at the end of the 1800s in the discourses of Hayes and Cleveland, and after disappearing in the early 20th century, re-emerges in the discourse of Coolidge and Hoover between 1927 and 1930 in a period of improved relations between countries:

> Despite the jeopardy to **our citizens and their property** which naturally arises in such circumstances, we have, with the cooperation of the governments concerned, been able to meet all such instances without friction. We have resumed normal relations with the new Governments of Brazil, Argentina, Peru, and Bolivia immediately upon evidence that they were able to give **protection to our citizens and their property**, and that they recognized their international obligations (Hoover 1930).

The protection of citizens and their property from threats and dangers may thus be regarded as one of the rhetorical tools through which presidents tended to legitimize war against foreign countries, and to expand power through trade and commerce abroad as well as in the Territories:

> In the late blockade of Terceira some of the Portuguese fleet captured several of our vessels and committed other excesses, for which reparation was demanded, and I was on the point of dispatching an armed force to prevent any recurrence of a similar violence and **protect our citizens** in the prosecution of their **lawful commerce** when official assurances, on which I relied, made the sailing of the ships unnecessary (Jackson 1831).
>
> Our naval force is intended only for the protection of **our citizens abroad** and of **our commerce**, diffused, as it is, over all the seas of the globe (Pierce 1854).
>
> All attempts at compromise having failed, it becomes the duty of Congress to consider what measures it may be proper to adopt for the **security and protection of our citizens** now inhabiting or who may hereafter inhabit **Oregon**, and for the maintenance of our just title to that **Territory** (Polk 1845).

Albeit rare, instances of protection in the 20th century relate to different semantic motifs, i.e. as safeguard against economic (Truman and Eisenhower) and social insecurity (Clinton). However, the paucity of instances may well indicate that the discursive representation of protection has shifted to

semantically related discourses over time: i.e. construed as discourses of security/insecurity.[34] Interestingly, this semantic motif has resurfaced in more recent times in the discourses of the two latest Republican presidents, Bush and Trump. The former employs it in order to justify an increase of budgetary expenditure after the 9/11 attacks to be used in the "war on terror," which was then reformulated by Trump in his 2017 SoU address as a threat from "radical Islamic terrorism:"

> The next priority of my budget is to do everything possible to **protect our citizens** and strengthen our nation against the ongoing threat of another attack (Bush 2002).
>
> Our law enforcement needs this vital legislation **to protect our citizens**—you need to renew the PATRIOT Act (Bush 2004).
>
> Our obligation is to serve, **protect**, and defend the **citizens** of the United States. We are also taking strong measures to **protect** our Nation from Radical Islamic Terrorism (Trump 2017).

In his latest address to the nation, Trump invokes this motif again in the new context of the coronavirus pandemic, construing citizens as needing protection from a new invisible threat:

> Protecting Americans' health also means fighting infectious diseases. We are coordinating with the Chinese government and working closely together on the coronavirus outbreak in China. My administration will take all necessary steps to **safeguard our citizens** from this threat.

4.2 Qualities of Citizens

While references to the moral character of citizens are to some extent present in the discourse of most presidents (Washington talks about the "zealous support of all good citizens"), a systematic pattern starts to emerge in the second half of the 19th century. Lincoln interestingly makes three references to *disloyal citizens* (and two to *loyal* ones) and loyalty, patriotism and respect of the law represent the enduring hallmark of a good citizen.

> I earnestly appeal to the intelligence and patriotism of all good **citizens** of every part of the country, however much they maybe divided in opinions on other political subjects, to unite in compelling obedience to existing laws aimed at the protection of the right of suffrage (Hayes 1879).

Throughout the period that goes from Lincoln to Roosevelt, the idea of the good, loyal, law-abiding, patriotic citizen intertwines with naturalization, and citizenship is represented as something to be deserved.

> It is due to this Government itself and to the great mass of the naturalized citizens who entirely, both in name and in fact, become citizens of the United States that the high privilege of citizenship of the United States should not be held by fraud or in derogation of the laws and of the good name of every honest **citizen** (Grant 1875).
>
> We need every honest and efficient immigrant fitted to become an American **citizen**, every immigrant who comes here to stay, who brings here a strong body, a stout heart, a good head, and a resolute purpose to do his duty well in every way and to bring up his children as law-abiding and God-fearing members of the community (Roosevelt 1901).

In this period, we also get iconic portrayals of the core American qualities and values: ingenuity, courage, self-reliance, enthusiasm, industriousness and honesty.

34. Cinzia Bevitori, "In a world of complex threats."

> As every patriotic **citizen** rejoices in the constantly increasing pride of our people in American citizenship and in the glory of our national achievements and progress, a sentiment prevails that the leading strings useful to a nation in its infancy may well be to a great extent discarded in the present stage of American ingenuity, courage, and fearless self-reliance; and for the privilege of indulging this sentiment with true American enthusiasm our citizens are quite willing to forego an idle surplus in the public Treasury (Cleveland 1886).

But it is with Theodore Roosevelt that the American citizen's character emerges in earnest from the data. Many of the adjectives Roosevelt associates with citizens are to do with "vigour," for example: *energetic*, *resolute*, *strong*, *stout*. This is, of course, very much in line with his conceptualization of citizenship as being part of the *body politic* and his very frequent use of the physical body metaphor. But it is again honesty that comes to most robustly define a genuine citizen (as opposed to a "so-called" one). This is not immediately visible from the list of collocates (as was the case for Grant), but the concept of honesty is discursively constructed in a variety of ways that can be identified only through the close reading of extended concordance lines.

> Forgeries and perjuries of shameless and flagrant character have been perpetrated, not only in the dense centers of population, but throughout the country; and it is established beyond doubt that very many so-called **citizens** of the United States have no title whatever to that right and are asserting and enjoying the benefits of the same through the grossest frauds. It is never to be forgotten that citizenship is, to quote the words recently used by the Supreme Court of the United States, an "inestimable heritage," whether it proceeds from birth within the country or is obtained by naturalization; and we poison the sources of our national character and strength at the fountain, if the privilege is claimed and exercised without right, and by means of fraud and corruption. The body politic can not be sound and healthy if many of its constituent members claim their standing through the prostitution of the high right and calling of citizenship (Roosevelt 1903).

In Roosevelt's discourses the idea that the right to be (or become) a citizen as a privilege to be earned and the worth of a citizen is measured on their moral standing is reinforced, and in this phase one's belonging is entirely centered on individual qualities. We find a radically different discourse in recent progressive presidents. With Clinton and Obama, being a citizen is strongly associated with the concept of duty, but it is framed as shared duty and responsibility to *our fellow citizens* and the qualities that emerge as most defining are related to cooperation and compassion. Clinton expresses the change of perspective explicitly:

> Self-reliance and teamwork are not opposing virtues; we must have both. I believe our new, smaller government must work in an old-fashioned American way, together with all of our **citizens** (Clinton 1996).
>
> We have to do more to accept responsibility for ourselves and our families, for our communities, and yes, for our fellow **citizens** (Clinton 1995).
>
> Where government provides opportunity and **citizens** honor the responsibility to give something back to their communities (Clinton 1997).

Obama adopts a very similar discourse of mutual responsibility, community and togetherness.

> It is our responsibility as lawmakers and educators to make this system work. But it is the responsibility of every **citizen** to participate in it (Obama 2009).
>
> My fellow Americans, whatever you may believe, whether you prefer one party or no party, whether you supported my agenda or fought as hard as you could against it, our collective futures depend on your willingness to uphold your duties as a **citizen**. To speak out. To stand up for others (Obama 2016).

Both in Clinton and Obama's addresses we find a powerful blend of references to the duty *of* citizens and the duty *to* citizens, this duty to citizens is twofold: on one hand it indicates the duty of the State to educate people to be good citizens (something that was present also in Roosevelt's discourse), and on the other hand it closely relates to the rights of citizens.

4.3 Rights of Citizens

The right of citizens to education, social security, health, employment, opportunity, and overall well-being that we encounter in the addresses of Obama and Clinton find their origin in those of Franklin Delano Roosevelt, Truman and, most prominently Eisenhower, where the guarantee of social rights becomes the most dominant association with being a citizen. It should also be mentioned that security in this period mostly referred to social security in the more generic sense of the term.[35] And this right, alongside the right to bargaining, was established with the New Deal in the 1930s. Before this phase the discussion around rights largely concerned the right to be a citizen and the topic of naturalization, when rights of citizens were discussed, they predominantly referred to economic interests. With Truman and Eisenhower, we see a pervasive insistence in the government's duty to grant citizens their fundamental right to pursue a good life (echoing the *Declaration of Independence*), beyond an exclusively financial stand, as the following short concordance extract exemplifies:

we must make possible greater equality of opportunity to all our	**citizens**	for education (Truman)
all	**citizens**	should be able to live in decency and health (Truman)
We must assure equal rights and equal opportunities to all our	**citizens**	(Truman)
better jobs, better living, better opportunities for every	**citizen**	(Eisenhower)
the creation and preservation of opportunity for every	**citizen**	to lead a more rewarding life (Eisenhower)
every	**citizen**	shall have opportunity to develop to his fullest capacity (Eisenhower)

Figure 8: selected concordance lines of citizen.

This trend continues with Kennedy and Johnson, however the number of occurrences of citizens is too limited to make any sort of generalization in terms of patterns, and with Carter:

> An America with equal rights for all **citizens**—and for women, guaranteed in the United States Constitution—an America with jobs and good health and good education for every **citizen**, an America with a clean and bountiful life in our cities and on our farms, an America that helps to feed the world, an America secure in filling its own energy needs, an America of justice, tolerance, and compassion (Carter 1980).

Even though Kennedy mentions the lemma citizen just twelve times, it is striking that three of these concern the topic of rights and race, for instance:

> And the most precious and powerful right in the world, the **right to vote** in a free American election, must not be denied to any **citizen** on grounds of his **race or color** (Kennedy 1963).

This is not just remarkable because these examples account for a quarter of all concordance lines for citizen in the Kennedy sub-corpus, but also because they account for half of the mentions of citizen having anything to do with race and color in the entire corpus. The overall absence of this discussion is extremely resonant, the only other three exceptions, apart from Kennedy, are one mention in Roosevelt sub-corpus and two mentions in Eisenhower's:

> It is necessary for us firmly to insist upon the rights of our own **citizens without regard** to their creed or **race** (Roosevelt 1904).
>
> Steadily we are moving closer to the goal of fair and equal treatment of **citizens without regard to race or color** (Eisenhower 1941).

35. Bayley, Bevitori, "Two centuries of 'security'."

And this is not simply due to the rare co-occurrence of race with *citizens*—in fact if we concordance the expressions without *regard to race* and *regardless of race* we retrieve just 5 other mentions: one by Roosevelt, one by Eisenhower, one by Clinton and two by Obama.

The path to equal citizenship rights met with many hurdles over time; in spite of the Emancipation Proclamation in the early 1860s that followed the notorious Dred Scott decision and the abolition of slavery following the 13th, and later on, the 14th Amendment, the 'Black' issue was largely silenced by presidents in their institutional messages. The extension of citizenship rights was the result of long negotiations of contested ideas about rights and duties, as clearly expressed in Eisenhower's 1956 address:

> We are proud of the progress our people have made in the field of civil rights. [...] It is disturbing that in some localities allegations persist that **Negro citizens** are being deprived of their right to vote and are likewise being subjected to unwarranted economic pressures (Eisenhower 1956) .

5 Conclusions

Citizenship is a complex phenomenon, and a crucial component of democracy in the United States. However, as Jones aptly argues "for most of its first century, the country neglected to define precisely who was a citizen,"[36] and one of the main things that emerged from this analysis was precisely an element of absence. Absence manifested itself at various levels: quantitative (in the low frequency of references to citizens in some of the sub-corpora) and discursive (in the avoidance specific issues). Nonetheless, as this paper has attempted to show, the notion of citizens/hip is also a 'rhetorical construct'[37] that has been linguistically and discursively articulated by American presidents at distinct points in time, and with different meanings and connotations, in order to address the concerns of diverse audiences. As a crucial site of negotiation of meanings, the analysis of this construct across long cycles of narratives of American history can largely benefit from the theoretical and methodological tools of corpus-assisted discourse analysis. Through the observation of statistically measured linguistic data, which are embedded and need to be properly understood and interpreted within the wider historical and socio-political contexts, we have been able to trace and map both synchronically and diachronically some of the main trends of social change and stability over long cycles of American history. In spite of the breadth of the analysis, however, a few caveats need to be considered. First of all, although citizenship, and being a citizen, is defined as a legal status, its relationship with notions of national identity, sense of community and belonging may at times be complex and differences need to be acknowledged, as highlighted in a number of studies from different theoretical perspectives.[38] A thorough investigation of the concept of American identity would thus need a broader scope encompassing the analysis of a wider range of words and their meanings over time. Moreover, any corpus-assisted analysis is focused on the investigation of the linguistic realizations of concepts and not on their ontology, it necessarily sets off from a lexical operationalization of a concept, which at times cannot be exhaustively covered by a list of terms. Methodologically, combining quantitative and qualitative dimensions of investigation and procedures in any diachronic analysis also implies a great deal of selection and exclusion. However, in spite of these constraints, we believe that in our interdisciplinary study, the adoption of a corpus-assisted approach in aggregating variables, to a certain extent hidden to the 'naked-eyed', has allowed us to take a critical perspective on the investigation of theory-driven temporal units of narratives of American political history.

---

36. Martha S. Jones, "Citizenship," in *The 1619 Project*, ed. Nikole Hannah-Jones (New York: Random House, 2021).

37. Troy A. Murphy, "Romantic democracy and the rhetoric of heroic citizenship," *Communication Quarterly*, 51(2003): 192–208.

38. Orm Överland, *Immigrant Minds, American Identities. Making the United States Home, 1870–1930* (Champaign, IL: University of Illinois Press, 2000); Samuel P. Huntington, *Who are We? The Challenges to America's National identity*; Anna Romagnuolo, "L'America nella retorica presidenziale: identità (con)divise, perdute, ritrovate," *Fictions, Studi sulla Narratività*, XXII(2013): 119–128.