

Testing George Carlin's "Soft Language": Cultural Change and Language Change in 20th Century American English

Carey Benom

1. Introduction

This study provides a focused, empirical examination of the frequency of use of certain lexical items in mid-to-late 20th-century American English. With the relationship between cultural change and language change as the larger, contextual background, I use data from both the Corpus of Historical American English (COHA) and Google N-grams to test whether there is reliable quantitative and qualitative evidence to support some of the socially-relevant linguistic observations made by stand-up comedian George Carlin in his 1990 special, "Doin' it Again".

Carlin coined the term "soft language" to describe a certain type of euphemistic vocabulary that he anecdotally observed to have increased in frequency during his lifetime. He argued that "soft language" conceals key aspects of human experience through its indirectness, or, as he put it, that it "takes the life out of life". His first example is the terms *shell shock*, *battle fatigue*, *operational exhaustion*, and *post-traumatic stress disorder*, which all refer to exactly the same condition. The first was replaced with the second, the second with the third, and the third with the fourth, over the course of the 20th century, he claims. The replacements are generally longer, and are constructed from Latinate terms, increasingly dismissing the human element of suffering, which is most deeply and profoundly expressed by the original, simple and plain English term, according to Carlin, who notes that, with the final expression, "the pain is completely buried under jargon."

Other lexical changes he claimed to have witnessed aren't necessarily about hiding pain, but they do share a single, common motivation, according to Carlin. They include: *car crashes* becoming *automobile accidents*, *the dump*

becoming *the landfill*, *sneakers* becoming *running shoes*, and *toilet paper* becoming *bathroom tissue*. “Poor people used to live in slums. Now the economically disadvantaged occupy substandard housing in the inner cities.” And, “(t)he government doesn’t lie, it engages in disinformation.” He noted that “soft language” began to arise more and more frequently in the English lexicon during his lifetime, replacing the more direct versions of various lexical items and phrases - and he claimed to know why. He attributed the rise of “soft language” to the rising power of corporate capitalism in the United States during the mid-to-late 20th century: “Smug, greedy, well-fed white people have invented a language to conceal their sins.” The rise of corporate capitalism is a well-documented cultural change (e.g. Levy 2021, Korten 1999, Perrow and Wealth 2004), and thus, I sought data with the potential to support Carlin’s apparently astute social and linguistic analysis.

The paper will proceed as follows: after giving some background in section 2, I will describe the methods I used to test Carlin’s examples of “soft language”, and my criteria for judging the results, in section 3. I will then present and discuss the results obtained in section 4 before concluding in section 5.

2. Background

Everett (2012) describes language as a cultural tool, used for communication, social cohesion, and to shape the way community members think and behave. In this view, culture is a factor in determining the form of language, and is reflected in the form of language. This idea is difficult to test, as Everett himself discusses (2012:262). After all, language is used to facilitate information exchange within a culture, and for group solidarity, but language also shapes how people in a culture interpret their world.

His proposal leads to questions about the mechanisms by which culture influences language, in addition to the specific areas of language structure that are affected by cultural factors, and precisely how they can be affected. (See Enfield 2013 and Weitzman 2013 for related discussion.) Here, to consider these issues, I will apply the viewpoint of change: When a culture changes, how are these changes reflected or encoded linguistically? Everett does not address any of these questions, which require both more detailed research and a broader range of data (such as historical and typological data) than is currently available.

If we ask *if* cultural change drives language change, in functional and usage-based linguistics the answer is "of course". When a cultural artifact is borrowed or created, a name for that artifact is borrowed or created (e.g. *sushi* and *mouse* (for computers) were borrowed into English and created via metaphorical extension, respectively). When a culture adopts a religion, vocabulary specific to that belief system enters the language (e.g. Sanskrit entering Chinese, and Sanskrit and Chinese entering Japanese through Buddhism, Latin entering English through Christianity, etc.). In fact, if Everett is correct that language is a cultural tool, cultural change *must* drive language change, and we can even explain why - because language is the toolbox by which the culture can realize the new changes. A recently-changed culture requires the tools of language to make use of new categories and express new values. The potential list of examples is long indeed.

But if we ask instead precisely *how* cultural change drives language change, we will find that we lack a larger, generalized explanation. If language is a cultural tool, we can investigate how changes in a culture are related to changes in its tool kit (i.e. how social and cultural changes are reflected or encoded linguistically). For instance, speakers manipulate language both deftly and subtly (often discussed in the literature are the frequency of use of a linguistic variable such as -in' vs. -ing, or the extent of rounding or raising of a vowel) in order to perform culturally relevant roles, as part of using language to express their identity and evoke solidarity. Sociolinguistic work such as Labov et al. 2006, Trudgill 2000, and Eckert 1989, coupled with cognitive sociolinguistic work such as Geeraerts et al. 2010 and Kristiansen and Dirven 2008, has provided many profound insights into the relationship of culture and language, but the complexity of the situation means that a clear and complete answer to this question is not yet available, and thus more data, more analysis, and more discussion are all necessary.

This motivates the topic of the present study. The rise of corporate capitalism, which Carlin argued led to the creation of "soft language", is a well-documented and widely accepted historical development. Whether it truly influenced English in the way he describes is less clear though, and thus requires empirical support. Of course, even if the corpus results support his claims, there is no way to know for certain that it was the rise of corporate capitalism that caused the changes in English - we cannot study causation empirically without creating an appropriate experiment, which is

impossible on this scale, and therefore, the strongest statements we can make could invoke correlation at best.

However, we do know that the deliberate use of euphemisms to influence perception and judgments in a huge variety of spheres of activity, including business, advertising, and the military, is a widely acknowledged practice (e.g. Farrow et al. 2021, Watts 2013). Such euphemisms have proven beneficial to corporate, military, and governmental stakeholders, as they allow them to avoid outright deception and its consequences without the need to state the truth explicitly, and there are countless examples of the creative use of euphemism employed as a business strategy to escape having to make real (and potentially costly) changes (Farrow et al. 2021, La Cour and Kromann, 2011) or as a military strategy to conceal certain issues and frame things in an advantageous way (Kiš 2014, Watts 2013), or in the political arena to camouflage the negative, or for propaganda purposes such as achieving compliance (Qizi 2021).

In fact, little about “soft language” is surprising, since we have been warned about such things since at least George Orwell, who referred to it, or something extremely similar, as “doublespeak”. In addition, if one follows the news, it’s hard not to notice various expressions that seem to be examples of “soft language”, such as the use of *collateral damage* to refer to civilians killed in war (Karam 2011). Therefore, in the absence of another likely cause, if Carlin’s linguistic claims turn out to be supported by the data, we will consider the quantitative evidence as supportive of his narrative - since it fits in with the academic literature, and in particular the attested use (and intended purposes) of euphemism in business, advertising, military, and political propaganda.

3. Methods

The examples Carlin gave of “soft language” in his performance were queried in both the COHA and Google N-grams (which searches Google Books), in order to bolster confidence in the results. COHA is the largest balanced historical corpus of English. Google Books is 3 orders of magnitude larger than COHA, with roughly 360 billion words in Google Books’ English corpus vs. over 400 million words in COHA, and for corpora, size brings some large advantages - especially the chance to observe less-frequent items or combinations of items. But given the fact that Google Books has been

shown to over-represent academic and scientific literature in the later 20th century and to have problems with dates (Nunberg 2010, who nonetheless remains optimistic regarding its utility), employing the cleaner but smaller COHA as well, rather than relying directly and exclusively on N-grams, was deemed essential.

In fact, it has specifically been argued that it is overly simplistic to simply consider trends in word frequencies in Google Books in order to "draw broad conclusions about cultural and linguistic evolution" (Pechenick et al. 2015), and cross-checking all results obtained from N-grams with a more carefully balanced corpus is recommended (Younes and Reips 2019). Therefore, data from the COHA, which is balanced across genres, sub-genres, and decades, was considered to be more trustworthy than that derived from N-grams. When the COHA failed to provide data, I proceeded with extra caution.

In most cases, the corpora produced similar data trends, lending confidence to the results. The exceptions will be discussed below. The results allow a quantitative analysis of frequency changes over time, permitting us to analyze each item (lexeme or phrase) in the light of the historical trends and also to contrastively analyze synonymous items.

In addition to the items noted in Section 1 above, the following additional claims were investigated: *false teeth* becoming *dental appliances*, *medicine* becoming *medication*, *information* becoming *directory assistance*, *motels* becoming *motor lodges*, *partly cloudy* becoming *partly sunny*, *broke* becoming *a negative cash flow position*, *house trailers* becoming *mobile homes*, *used cars* becoming *previously owned transportation*, *room service* becoming *guest room dining*, *hospital* becoming *health maintenance organization or wellness center*, *doctor* becoming *health care delivery professional*, *constipation* becoming *occasional irregularity*, *crippled* becoming *disabled* or *differently abled*, *deaf* becoming *hearing impaired*, *blind* becoming *partially sighted* or *visually impaired*, *stupid people* becoming *people with a learning disorder* or *minimally exceptional*, *ugly people* becoming *people with severe appearance deficits*, *old people* becoming *senior citizens*, and expressions such as *90 years old* becoming *90 years young* and *die* becoming *terminal episode* or *negative patient care outcome* to hospitals and insurance companies, respectively. Finally, he claimed that the pentagon measured nuclear radiation in something called *sunshine units*. While it's not clear if Carlin intended the entire list of examples in all seriousness, to be prudent, all terms were

queried.

However, certain items proved resistant to analysis, due either to their polysemous meanings or their low frequencies. For instance, *information* has been used in English since before 1750, and it's not possible to disentangle this meaning of the term from its use to refer to telephone directory assistance. More than a few items do not appear at all in the COHA, and some, such as *occasional irregularity*, provided insufficient data even using Google N-grams.

I will not interpret Carlin's claims that a certain term "became" another one literally, since words and phrases obviously do not become other words and phrases, and the coexistence and overlap of near synonyms is well-attested (e.g. consider *jail* and *prison*). It is true that terms can replace other terms in use within a speech community over time, but given the context of a comedic performance, Carlin seems to be exaggerating for effect, or possibly referring primarily to some sub-set of contexts that he doesn't mention. In lexical semantics, this is not a black-and-white issue, but one of relative frequency over time. I will operationalize Carlin's claims to mean that the "soft" item came into detectable use and entered into a relationship of lexical competition (MacWhinney 1987) with the original term.

I will focus on testing the terms that appear in Carlin's monologue in the following ways. My concerns will be whether the "soft" version arose at the correct time, from approximately when Carlin was born in 1937 until 1990, the year of his performance, and whether the terms he cites acted as lexical competition during this period. Therefore, I will consider the following types of results to be supportive of his claims a) the creation and emergence of a "soft" item in the correct time period, followed by b) the intensification of competition between the "soft" item and the original item during this time period until c) the "soft" item rivals or surpasses the original in frequency.

Note that it might seem at first that I am being overly generous when I offer my definition of *rival* in (c) above, allowing a "soft" item with a 1:10 frequency vis-à-vis the original term, or a nine percent market share, so to speak, to count. But as an analogy, imagine creating a company with the aim of taking over nine percent of national soft drink sales in a country with 300 million people, for example - it would certainly seem like an uphill task. In much the same way, the emergence of a new term, such as *battle fatigue*, when American English speakers already use the term *shell shock*, would be

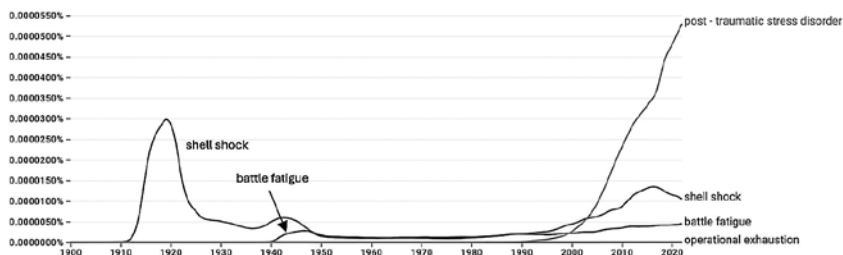
an uphill fight, and for *battle fatigue* to increase in frequency and achieve a 1:10 ratio with *shell shock* would be considered evidence that the terms were in a relationship of lexical competition.

Finally, I will examine the lexical collocates of all items in the COHA, to ascertain if the specific characteristics of "soft language" that Carlin describes can be found in the data. I will observe the strongest collocates for all terms, noting patterns in the data. I will also test various terms relevant to Carlin's argument, such as the affinity of the lexemes *suffer* and *pain* for certain items. To do so, I will obtain what are known as MI ("mutual information") scores (Manning and Schütze 1999). MI scores represent the probability that the two terms occur together as a joint event, as opposed to being a random occurrence. The higher the MI score, the greater the affinity the two lexemes or phrases have for one another, or the more they "belong" together. MI scores of 3 or more are said to be most "linguistically interesting", though what this means precisely is difficult to say. MI scores of 5 and above can be said to reveal a strong attraction.

4. Results and Discussion

I will present and discuss the results for select terms or phrases before summarizing the remaining results at the end of this section. I will present both representative items and those that were closer to being outliers, in order to give a broad picture. While the COHA data was considered primary for the purposes of our analysis, N-gram results will be presented when applicable due to ease of readability and comparison. Our first results are for the synonyms for *shell shock*, which match well in the COHA and Google N-grams, and the latter is presented below.

Figure 1: Contrastive N-gram for Four Synonyms for *Shell Shock*



As can be seen from the graph, *shell shock* was indeed the earliest of the terms, and was used extensively during and immediately following World War 1. It was also the most popular term during World War 2, but at that time *battle fatigue* emerged as an alternative - exactly as Carlin described, supporting his narrative, and it remains so, despite relatively low frequency, to this day. Carlin might have been surprised to know that *shell shock* enjoyed another peak of popularity after his performance - though observation of the recent results in COHA suggests that English speakers now usually employ the term metaphorically, as it collocates with e.g. *economic* and *sexual*, though data are too sparse to draw any firm conclusions. *Operational exhaustion* was infrequent to the extent that not one example was found in the COHA in the 20th century.

Carlin's instincts were strongly supported by the data. *Battle fatigue* did emerge as a strong alternative to *shell shock* during World War 2. N-gram data suggests that *operational exhaustion* also emerged during this period, as seen below.

Figure 2: N-gram for *Operational Exhaustion*

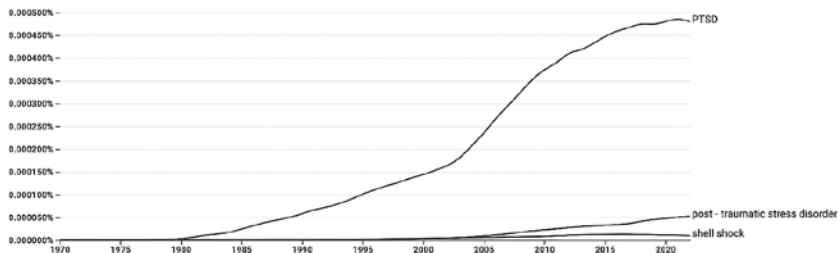


Carlin claimed that it emerged during the Vietnam War, so he wasn't correct about the precise order of the least-used synonym, but his other claims were substantiated. However, a mystery remains: in Figure 1 above, *post-traumatic stress disorder* doesn't seem to arise until after Carlin's 1990 performance. Was he blessed with foresight into future linguistic trends?

It turns out that this is unlikely to be the case. Carlin never claims to list all possible synonyms, and in fact he left out the acronym *PTSD*, which was the most frequent of the synonyms by far in 1990. This can be seen in the N-gram below, where the acronym is contrasted with both the full form and *shell shock*, focusing on a more recent timespan. This would have allowed

him knowledge of the full term at that time.

Figure 3: Contrastive N-gram for Three Synonyms for *Shell Shock*



The COHA data differs somewhat, but generalizing over the data from the two corpora, what we can say is that the acronym is and mostly has been much more frequent than the full form. If Carlin is correct about the cause, purpose, and effects of "soft language", as outlined above, we could see the emergence of the acronym - which appears to be motivated by economy of production - as the grass-roots response to the introduction of military and scientific jargon, intended as propaganda. Dispensing with the Latinate terms entirely could be seen as the people's way of fighting back and reclaiming their language. Alternatively, we could see the spread of the acronym as an issue of mere convenience, in which speakers shortened a long form without acknowledging or revitalizing any of the expression of suffering that was lost through the creation and spread of the jargon in the first place.

An investigation of the collocates of the terms in the COHA reveals that the former story stands a chance of being true, but the latter is debunked. According to the results, and with apologies to Shakespeare, *shell shock* by any other name would smell like suffering. A collocational analysis of the terms reveals that *suffering* is strongly - and nearly equally - attracted to *shell shock* (MI=5.69), *post-traumatic stress disorder* (MI=5.81), *PTSD* (MI=6.83), and *battle fatigue* (MI=5.24) in the COHA. English speakers don't seem to be ignoring the connection between the term and the meaning of suffering, *contra* Carlin. Or, if the term itself hides the suffering, as he claims, then we can say that the surrounding context is providing sufficiently strong clues.

Based on either of these scenarios, it could be the case that suffering, as a core part of life itself, is impossible to hide, and that it matters to us so much

that we are going to discuss it, even if, as Carlin claims, a word or phrase here or there tries to deny it - it will emerge from the larger discourse context. Does buried pain just seep back into the groundwater? We will return to this point.

Next, the results for *car crash* and *automobile accident*, which were surprising in that they turned out to be almost exactly the opposite of what Carlin stated. See Figure 4 below.

Figure 4: Contrastive N-gram for Two Synonyms for *Car Crash*

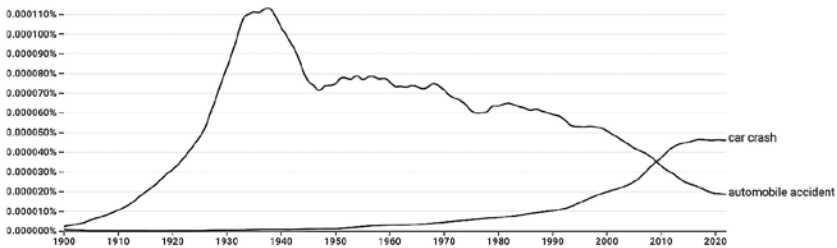
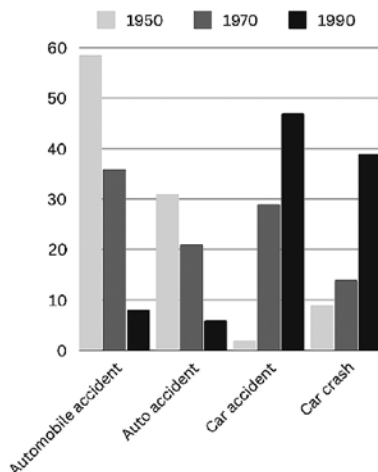


Figure 4 shows that the longer, latin-based phrase was far more frequent and used earlier. It was only replaced in popularity by *car crash* in about 2010, 2 decades after Carlin filmed his special. (Data from the COHA were similar except that *car crash* overtook *automobile accident* earlier, by 1980.) Once again, examining collocations of *car crash* and *automobile accident* in the COHA revealed little or no difference in the extent of suffering expressed in the context, as both had semantically similar collocates such as *fatal*, *killed*, *injured*, *suffered*, *died*, and *hurt* - another aparent case of buried pain seeping into the groundwater.

Additionally, Carlin didn't tell the full tale, as the addition of *car accident* and *auto accident* revealed (*auto crash* and *automobile crash* were nearly nonexistent, and are therefore not included). In the COHA data as well as the N-gram data, until the 1960s, *automobile accident* and *auto accident* were the most frequent, in that order, but then *car accident* took over the number two position by 1970. In the COHA, just 10 years later, *car accident* had achieved the top frequency, and *car crash* was in the number two position (note that this occurred in the N-gram as well, but it took a few more decades). The figure below shows each of the four synonyms at three selected times, and reveals the initial supremacy but subsequent decline of the terms with

automobile and *auto* and the rise of the terms with *car*.

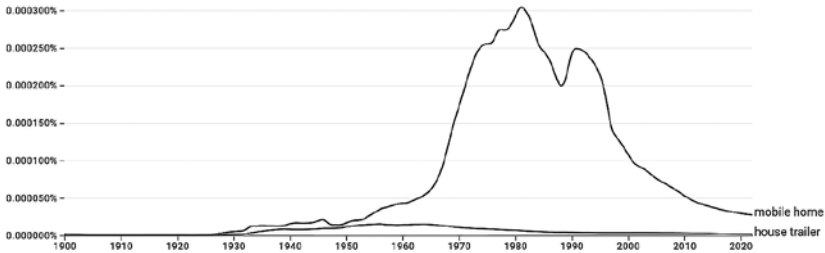
Figure 5: Each of the four synonyms as a percentage of the total of all four, at that time, in the COHA, for three selected times



What we can say, based on the data, is that there was an intense lexical competition between these terms during Carlin's adult lifetime, but the order their ascendancy was the opposite of what he claimed.

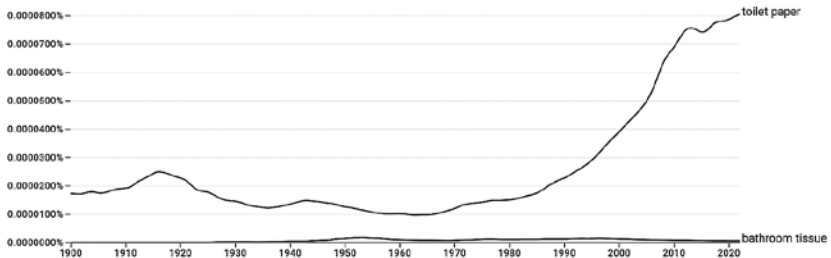
Note that it is conceivable that the domination of *automobile accident* in the early-to-mid 20th century is due to a higher frequency of *automobile* as compared to *car*, at least when referring to automobiles. The corpora don't permit us to easily answer the question of their relative frequency because *car* is polysemous, and originally referred to e.g. train cars. What we can safely say are that Carlin's predictions about the order of the items' ascendancy was not supported, though his contention that they were in competition was supported.

As for *house trailers* becoming *mobile homes*: the "soft" version was always more frequent, and the other was only a reasonable alternative until about 1970, as seen in the contrastive N-gram below:

Figure 6: Contrastive N-gram for Two Synonyms for *Mobile Home*

Based on these results, the terms were in competition from roughly when Carlin was born in 1937 until the 1960s, when the “soft” version exploded and the other faded. So, his claims are supported. A study of the collocates of both terms in the COHA failed to reveal any meaningful differences, probably due to data paucity for *house trailer*.

As for *toilet paper* becoming *bathroom tissue*, the latter is under-represented in both the COHA (with just 6 instances, though it’s worth mentioning that the earliest is in the 1950s) and N-grams, as seen in the N-gram below:

Figure 7: Contrastive N-gram for Two Synonyms for *Toilet Paper*

However, a further query for *bathroom tissue* suggests that it did indeed enjoy a period of relative popularity, as seen below.

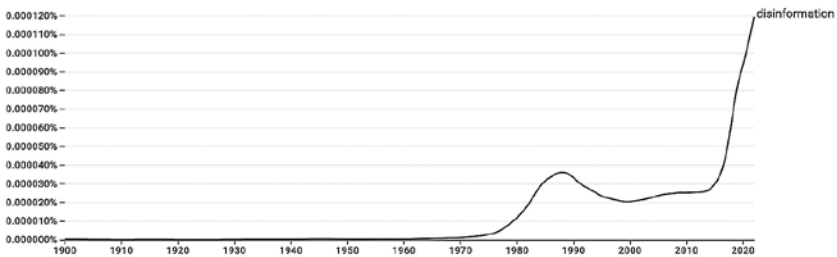
Figure 8: N-gram for *Bathroom Tissue*



Note that when Carlin was a young adult in the 1950s, *bathroom tissue* achieved nearly one-sixth of the frequency of *toilet paper* in Google Books. Thus Carlin's intuitions about the timing of the rise of the "soft" version and the competition between the two were supported.

It was difficult to distinguish the uses of *lie*, but as for *disinformation*, based on the N-gram below (though the COHA lacked data), Carlin seems to have watched it arise and experience an apparent peak in popularity in the 1980s, though he certainly would have been disheartened to know of the explosion in use more recently.

Figure 9: N-gram for *Disinformation*



In any case, *disinformation* basically fits Carlin's predictions. While it proved difficult to study the longer phrases in Carlin's monologue, simplification from a sentence to a noun phrase resulted in some suggestive data, as seen in the N-gram below (the COHA lacks sufficient data).

Figure 10: Contrastive N-gram for Two Synonyms for *government lie*

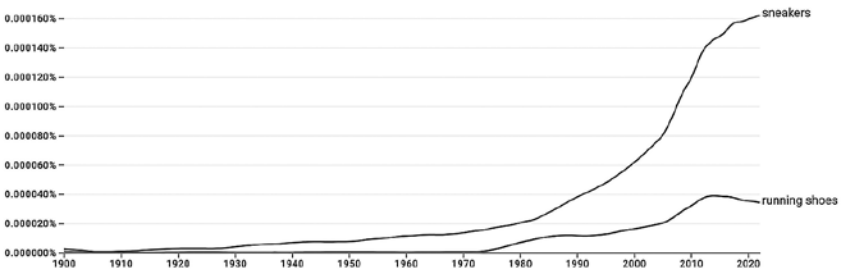


Note that *government disinformation* only arose in about 1980, but by 1990, the two were nearly equal in frequency. Based on this evidence, we can say that Carlin seems to have been correct about the rise in the term *disinformation* as lexical competition for *lie*, and that this was in the context of *government* in particular.

While the COHA lacked copious data for *disinformation*, the differences among collocates for the terms were clear: *lie* collocated with e.g. various personal pronouns, while *disinformation* collocated with *campaign*, *propaganda*, and *spread* - longer, mostly Latinate terms with abstract meaning, contrasting with the physical and tangible reality of *I*, *you*, and *we*. These differences support Carlin's claims.

As for *sneakers* becoming *running shoes*, Carlin seems to have been correct about the rise of the latter, and its legitimacy as a competitor, especially considering that it achieved about half of the frequency of use of the former in the early-to-mid 1980s in the N-gram data.

Figure 11: Contrastive N-gram for Two Synonyms for *Sneakers*



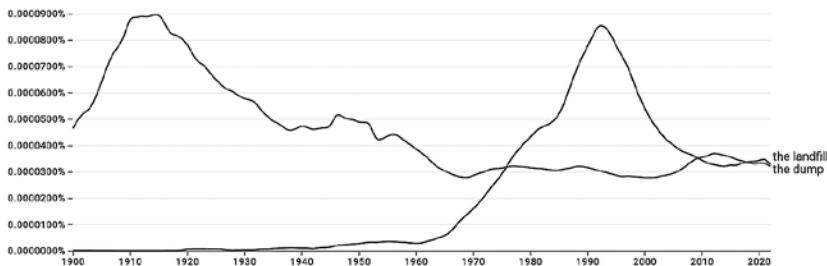
Based on both the N-gram data and the COHA data, which match closely, we can judge that Carlin's linguistic intuition concerning the new competition

between the terms was correct.

An investigation of the top thirty collocates of each in the COHA was revealing as well. *Running shoes* collocates with various brand names, including *Nike*, *Adidas*, *New Balance*, and *Saucony*, but just one color term: *bright orange*, whereas *sneakers* collocates with various color terms including *white*, *blue*, *black*, and *red*, but just one brand name: *Nike*. Additionally, *sneakers* collocates with various sports terms, including *basketball* and *tennis*, whereas *running shoes* did not collocate with any sports terms. *Sneakers* are *dirty*, *worn*, or *scuffed*, whereas *running shoes* are *ragged*, *ratty-looking*, or *expensive*. Carlin's distinction between the old-time, tangible directness and physicality of the non-soft variant and the more modern, more abstract "soft" variant was generally supported by these results.

As for *the dump* becoming *the landfill*, Carlin's claim was strongly supported by data from both corpora. As can be seen in the N-gram below, *the landfill* emerged during Carlin's adult lifetime and temporarily overtook and overwhelmed *the dump* in frequency of use, including in 1990.

Figure 12: Contrastive N-gram for Two Synonyms for *the Dump*



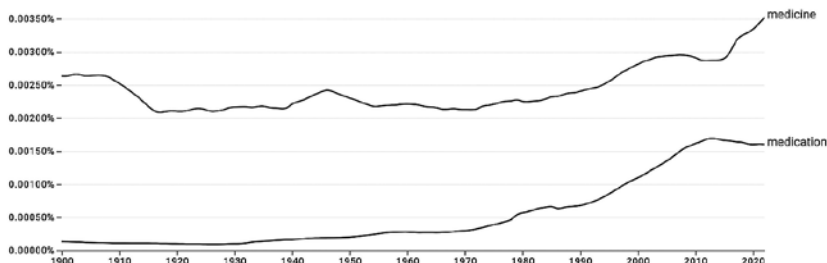
However, using both quantitative data and hindsight, we can say that *the dump* itself has remained quite stable since about 1970, and *the landfill* has now fallen in frequency so that the two are equal. Carlin's intuition was strongly supported here; *the landfill* and *the dump* were lexical competition before 1990, with the former having seemingly come from nowhere and overwhelmed the latter.

An examination of the collocates of both terms in the COHA revealed that *the dump* co-occurred with the pronouns *you*, *me*, and *her*, which was not true for *landfill*, attesting to the concrete nature of the former. Further

investigation showed that there were several different senses of *dump* being used in these tokens, including a) the landfill, b) to sever relations with a romantic or other partner, especially suddenly, c) a metaphorical use of the *landfill* sense that is used to refer to the appearance of people's homes or their physical appearances in a derogatory way. Unlike *the dump*, *the landfill* collocated with terms such as *capacity*, *sanitary*, *transfer*, *station*, and *proposed*. Just as Carlin argued, the physicality of the *dump* contrasted strongly with the sterility of *the landfill*. These results support his larger narrative.

Concerning *medicine* becoming *medication*, the latter doubled in frequency in the 1960s compared to the decade before, then nearly doubled again in the 1970s, and then increased by another 50% in the 1980s, according to the COHA data. By 1990 *medication* had an N-gram frequency more than one-fourth that of *medicine*, as shown below (the frequency in the COHA for 1990 is closer to one-fifth, but the same trend can be seen).

Figure 13: Contrastive N-gram for Two Synonyms for *Medicine*



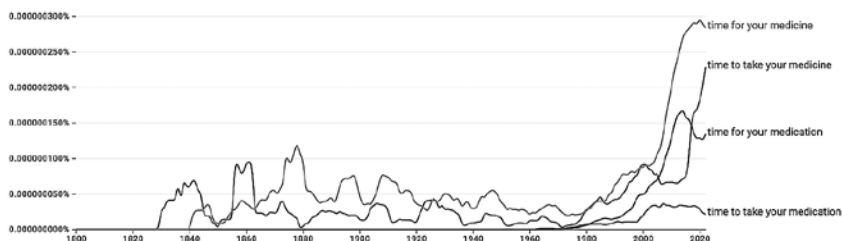
In fact, between 1940 and 1990, *medication* nearly quadrupled in frequency (in words per million) in the N-gram data, whereas *medicine* was essentially unchanged. Thus, we can judge the *medication* to have gone from infrequent in Carlin's lifetime to become a relatively frequent alternative to *medicine* by the time of the performance. Therefore, though *medication* was created earlier than we would expect, I will consider the data to generally support Carlin's claims.

However, a study of the two items' collocates in the COHA revealed something unexpected, based on Carlin's argument: of the two, it was *medicine* that collocated most strongly with more abstract terms, such as

school, *practice*, *university*, *law*, and *journal*, while the collocates of *medication* tended to be physical and to remind one of actual pills, including *take*, *pain* and *oral*. This is because *medicine* is polysemous, referring to both the practice, theory, and study of curing or preventing disease and to a substance used in the treatment of a disease. Both meanings have been around since the 14th century, according to the OED.

Here, it should be noted that we might be in danger of missing Carlin's point: in the context of being given a pill to take, *time to take your medication* is undoubtedly more formal and scientific-sounding than *time for your medicine*. Therefore, to try to get an idea of which synonyms of *medicine* may have been used in such contexts, an N-gram of four synonymous phrases from 1800 to 2022 was obtained, and is presented below.

**Figure 14: Contrastive N-gram for Four Synonymous Phrases
for *Time for your Medicine***



This N-gram supports Carlin's claims in that, in the informal context, *medicine* was the only attested choice until nearly the 1980s, but *medication* had emerged as lexical competition in this context by 1990.

As for *partly cloudy* and *partly sunny*, the latter has never achieved any substantial frequency of use compared to the former, and as such is essentially invisible on a contrastive N-gram of the two. However, an individual N-gram, presented below, suggests that it did experience a rapid rise in popularity in the 1970s, which Carlin seems to have recognized. (In the COHA, there are just 9 instances of *partly sunny*, the earliest in 1960.)

Figure 15: N-gram for *Partly Sunny*

Therefore, we can say that Carlin's claims were partially supported: *partly sunny* appears to have increased in frequency dramatically in the two decades before 1990, but it never became true competition for *partly cloudy*.

A similar story can be observed for *motel* becoming *motor lodge*. The former is and has been the default expression, while the latter emerged mostly in the 1960s, as can be seen in the N-gram below.

Figure 16: N-gram for *Motor Lodge*

It seems that Carlin noted the emergence of *motor lodge* an alternative to *motel*, but it failed as lexical competition - apart from a few years before and after 1940, when it achieved a 1:9 ratio in the N-gram data. Therefore, this evidence is generally consistent with a weak version of Carlin's claims.

To examine Carlin's statement that "*Poor people* used to live in *slums*. Now *the economically disadvantaged* occupy *substandard housing* in the *inner cities*," proved difficult because of the lack of one-to-one correspondence between the terms and expressions in the "before" and "after" versions. Therefore, select contrastive and individual results for these terms will be presented. First, Carlin was correct about the rise of *inner cities* as competition for *slums*, as can be seen in the N-gram data below.

Figure 17: Contrastive N-gram for Four Terms for *Slums*



Inner city and *inner cities* were shown to have emerged as strong potential alternatives to *slum/ slums* by 1990, supporting Carlin's claims. (No clear difference in the collocates of the terms was observed in the COHA data.)

Continuing with the same statement, we find that *economically disadvantaged* did indeed emerge as a viable alternative during the 1960s, and by 1990, remarkably, it was used with more than half the frequency of *poor people*, supporting Carlin's claims.

Figure 18: Contrastive N-gram for Two Synonyms for *Poor People*



Furthermore, an examination of the collocates for *poor people* and *economically disadvantaged* in the COHA reveals that the former co-occurs with terms including *blacks*, *homeless*, *abortions*, *starve*, *dump*, *immigrants*, and *drown* whereas the latter co-occurs with *students*, *children*, *African-American*, *residents*, and *qualifies*. This is exactly the type of pattern that Carlin is describing, with the direct term being more emotionally hard-hitting and starkly conveying the reality including the pain, while the "soft" alternative is used more coldly or distantly. In this case, being *economically disadvantaged* almost seems to bring hope, as you fit within a system, but for *poor people*, the situation seems dire, indeed. These patterns of collocation

strongly support Carlin's general claims about the emotional distance of "soft language". They are quite unlike those we observed above which seemed to be cases of buried pain seeping into groundwater – here, the pain was very much hidden.

Similar to this, though less dramatic, was the term *senior citizens*, which emerged in the 1950s and overtook *old people* before 1990, as Carlin claimed. The collocation patterns of the two in the COHA differ in that the "soft" term collocates with words like *national*, *center*, and *council*, in addition to various prices such as \$12 and \$15, highlighting older people's official governmental and consumer identities, whereas the original term collocates with *people*, *you*, *them*, *us*, and other direct and simple language – just as Carlin argued.

To summarize the remaining results, Carlin's claims were generally all supported, at least in that the "soft" term emerged in the correct time period and entered into a relationship of lexical competition with the original term, though it didn't always outpace the original in frequency. Terms like *90 years young* existed in the 19th century but experienced a rise and peak of popularity in the 20th; *health maintenance organization* and *wellness center* did emerge as lexical competition in the 1970s and 2000s, respectively (Carlin was on the cutting edge with *wellness center*). *Learning disorder* and *minimally exceptional* appeared in the 1970s and 1980s, though they failed to claim a large market share compared to *stupid person*. Still, we can say that the interpretations of his claims taken in this paper were, overall, strongly supported.

Even several claims that this author, as a native American English speaker approximately thirty years younger than Carlin, first judged to be dubious or solely in the spirit of jest turned out to be substantiated by N-grams, if not by the COHA due to data sparsity. They include the emergence of two synonyms for *die*: *terminal episode* in the 1950's and *negative patient care outcome* in the 1980's, as well as the emergence of *negative cash flow position* in the 1970's as a synonym for *broke*, and the emergence of *sunshine units* for units of radiation in the 1950's and 1960's.

Conclusion

This study was an attempt to help shed light on the relationship between shifting culture and linguistic change, focusing on mid-to-late 20th-century

English, and using corpus data to test George Carlin's observations on "soft language." Through analysis of lexical and collocational trends in the COHA and Google N-grams, I investigated whether Carlin's examples reflect the patterns that he argues are tied to the rise of corporate capitalism.

The findings provide empirical support for Carlin's claims, interpreted somewhat generously, as most of the "soft" terms he highlighted a) came into existence in the mid 20th century and b) claimed a stake as lexical competition with the original (non-soft) term during this time period. These frequency changes generally align with the historical rise of corporate and mass media influence in American society, and therefore, his narrative received support from these results.

Whether these shifts in frequency actually downplay human suffering was less clear. With the synonyms for *shell shock* and *car crash*, we saw that suffering was equally present in the linguistic environments all of the synonyms, raising questions about the inevitability of the expression of suffering in human language and whether one "soft" word could effectively suppress the human need to communicate about such a fundamental issue as suffering.

Other collocational evidence, however, did support Carlin's claims about "soft language", such as the physical collocates of the *dump* contrasting strongly with the procedural sterility of those of *the landfill*, *poor people* seeming to be suffering far worse than the *economically disadvantaged*, and *senior citizens* invoking official governmental and consumer identities, making them seem less directly human than *old people*. These results suggest that humanity and directness of experience can indeed be hidden through the use of "soft" terminology, at least in some cases, unlike the results for *shell shock* and *car crash*.

While the evidence does serve to corroborate key aspects of Carlin's linguistic observations, this study also highlights the complexity of attributing linguistic trends to specific cultural forces. Many of Carlin's observations gained empirical support, but we cannot easily claim evidence of a causal relationship on a scale this vast without much more data, and different types of data. Future research could expand on this foundation by examining larger corpora of various languages and exploring cross-linguistic and typological patterns (see Younes and Reips 2019), as well as incorporating sociolinguistic research techniques to better understand the interplay between language,

culture, and power. Ultimately, this study represents one small strand of the converging evidence which is required to understand the connection between cultural change and language change, but it also attests to the fact that even anecdotal insights - including those presented as humor - can inspire meaningful academic inquiry. Finally, the results affirm that seriousness is at the core of at least some humor (Attardo 2014:118).

Bibliography

Attardo, S. (ed.) 2014. *Encyclopedia of Humor Studies*. Los Angeles; Sage.

Eckert, P., 1989. *Jocks and burnouts: Social categories and identity in the high school*. Teachers college press.

Enfield, N.J. 2013. Language, culture, and mind: trends and standards in the latest pendulum swing. *Journal of the Royal Anthropological Institute* (N.S.) 19, 155-169

Everett, Dan. 2012. *Language: The Cultural Tool*. Vintage Books.

Farrow, K., Grolleau, G., and Mzoughi, N. 2021. 'Let's call a spade a spade, not a gardening tool': How euphemisms shape moral judgement in corporate social responsibility domains. *Journal of Business Research*, Volume 131, July 2021, Pages 254-267. Retrieved at <https://www.sciencedirect.com/science/article/abs/pii/S0148296321002381> on 2024 Jan 6.

La Cour, A. and Kromann, J. 2011. Euphemisms and hypocrisy in corporate philanthropy. *Business Ethics, the Environment, and Responsibility*. Special issue: Where Is Business Ethics? July 2011, pp. 267-279.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., and Nagy, N. (2006). Listeners' Sensitivity to the Frequency of Sociolinguistic Variables. *Penn Working Papers in Linguistics 12.2* (Papers from NWAV 34), 105-29.

Levy, J., 2021. *Ages of American Capitalism: A History of the United States*. Random House.

Geeraerts, D., Kristiansen, G. and Peirsman, Y. (eds.). 2010. *Advances in Cognitive Sociolinguistics*. Berlin/New York, De Gruyter Mouton.

Karam, S. 2011. Truths and Euphemisms: How Euphemisms Are Used in the Political Arena. *The Southeast Asian Journal of English Language Studies*, Vol 17(1):5 – 17.

Kiš, M. 2014. Euphemisms and military terminology. *Hieronymus, 1*, pp.123-137.

Korten, D.C., 1999. *The post-corporate world: Life after capitalism*. Berrett-Koehler Publishers.

Kristiansen, G. and R. Dirven (eds.). 2008. *Cognitive sociolinguistics: language variation, cultural models, social systems*. Berlin, Heidelberg, New York: Mouton de Gruyter.

MacWhinney, B. 1987. The competition model. In B. MacWhinney, ed., *Mechanisms of Language Acquisition*, 249-308. Hillsdale, New Jersey: Erlbaum.

- Manning, C. and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.
- Nunberg, G. 2010. Counting on Google Books. *The Chronicle of Higher Education*. December 16, 2010. Accessed December 18 2024 at <https://www.chronicle.com/article/counting-on-google-books/>
- Pechenick, E. A.; Danforth, C. M.; Dodds, P. S.; Barrat, A. (2015-10-07). "Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution". *PLOS One*. 10 (10): e0137041. Accessed Dec 20, 2024 at <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137041>.
- Perrow, C. and Wealth, P. 2004. *Corporate Capitalism*. Princeton: Princeton University Press.
- Qizi, D. N. O. M. 2021. Euphemisms In The Speech Of Politicians And Diplomats In Modern English. *The American Journal of Social Science and Education Innovations*, 3(11), pp.58-64.
- Trudgill, P. 2000. *Sociolinguistics*. London: Penguin.
- Watts, D. 2013. Military Euphemisms in English: Using language as a weapon. 紀要. 言語・文学編, 45, pp.169-198.
- Weitzman, R. 2013. A Review of *Language: The Cultural Tool* by Daniel L. Everett. *Anal Verbal Behav*. 2013;29(1):185–198. Accessed Dec 25, 2024 at <https://pmc.ncbi.nlm.nih.gov/articles/PMC3659492/>
- Younes, N.; Reips, U.D. 2019. "Guideline for improving the reliability of Google Ngram studies: Evidence from religious terms". *PLOS One*. 14 (3): e0213554. Accessed Dec. 20, 2024 at <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213554>.