Keyword Culture: Transforming Words into Information and the Technical Control of Language

Rex Troumbley, PhD Candidate
Department of Political Science
University of Hawaii at Manoa

Paper for the Western Political Science Association 2014 Conference
April 17, 2014
Keywords are often thought of as simple heuristics, handy compression of expressions into a single word which conveys a bundle of meaning. Keywords allow us to navigate complicated texts, search for specific uses, tag information, or treat language to computational processes. Keywords are also a technology of language governance which regulates the intelligibility of expressions, separates “bad” from “good” language, preconditions automatic filters which sift out taboo language, and is increasingly influencing the way we communicate with each other and with non-human agents. Writing after the digital revolution, as it is practiced by bloggers and spammers, no longer exclusively addresses humans, but also the algorithms of search engines and filters. A new generation of students is learning to adapt their essays according to the criteria of automated grading applications. Computers turn our written communications into data which is then mined for insights about us, our habits, and our digital histories. At the same time, programs are learning how to write for human readers; automatically spamming our emails, commenting on our blog posts, and writing (or rewriting) novels. A few weeks ago, following a small earthquake in southern California, an algorithm designed to write news articles faster than a human reporter automatically published a news story in the Los Angeles Times less than 20 minutes after the quake registered on U.S. Geological Survey instruments.¹ For the first time in history, we human readers are no longer the sole audience of the written text and are no longer the sole writers of texts.

This paper is part of my larger dissertation on the governance of taboo language. Curse words, swearing, obscenity, profanity, and other language taboos were perhaps the first “keywords.” What exactly constitutes a keyword is still the subject of some debate. Rhetoricians consider keywords to be

expressions which reveal the internal structure of an author’s reasoning. Grammarians tend to think of keywords as an essential part of structural composition, reasoning, and comprehension. For cryptographers, especially alphabetic text encryption, a keyword is a cipher of letters which can be used to decode a message. In corpus linguistics, keywords are words which appear in a text more often than we might expect by chance alone. Information retrieval systems treat keywords as terms which capture the essence of the topic of a document and keywords have become an integral part of bibliographic control. In computer science, a keyword is a word that is reserved in programming languages as expressions with special meanings, expressions that cannot be used as variable names, and words which can be commands or parameters for the execution of a program. The keyword has become a ubiquitous feature of language governance, so much so that it often receives little scrutiny today.

The emergence of keywords has been conditioned by several transformations of speech and sound. The first was the transcribing of sounds into the visual format of writing. Writing and print made words appear as discrete units of communication. Later, these discrete units were freed from the sentences and expressions in which they were used to become the objects of context-free language. Writing conditions the development of words with independent meanings and makes possible the identification of “bad” words. Following the transformation of speech into visual formats, dictionaries emerged as a type of written language governance which collected the words freed from the contexts of their use and created stable definitions. As a result, the keyword has become a technology of language governance which operates within computational culture to regulate what is valued, proper, intelligible, and worth thinking about.

Making Sounds Visual and Finding the Words

The conditions which made possible good and “bad” language in the American English tradition are rooted in the conversion of sounds into visual formats. The process of reducing language to a series
of words began with the development and introduction of writing. While there has never been a complete transition from spoken language to a purely written mode of communication, the differences between written English and spoken English are fairly clear. Some writing is intended to be read aloud while other forms of writing are meant to be read silently to oneself. When writing was first invented, it was used primarily to keep financial accounts or historical records. In some cases, official political authority was given to the people (scribes) capable of making and understanding these visual representations of sounds. Still, most information and knowledge was passed on orally, demanding remarkable feats of human memory, and the creation of completed systems of mnemonics. It took many generations for the invention of writing to become interiorized by the cultures which used it and even then, it was often viewed as suspicious and foreign. In Greece, for example, Plato lamented the fact that his dialogues had to be reduced to writing in order to be preserved. Ong summarizes Plato’s objections to writing in the *Phaedrus* and his *Seventh Letter* “as a mechanical, inhuman way of processing knowledge, unresponsive to questions and destructive of memory.” Despite Plato’s objections, it was soon recognized that written methods for storing information and knowledge had advantages over the oral and mnemonic methods of continual repetition. Writing also became a method of communicating with future generations and ensuring knowledge was not lost with the death of the individual who kept it, hence Plato’s reconstruction of Socrates’ dialogues with his fellow Athenians. The translation of the *Iliad* and the *Odyssey* into writing allowed for the ability to compare one oral account with another. For the first time, a poet’s recitation could be checked with a written record and compared for accuracy. Stories like Homer’s, which were subject to constant changes (often requested by audiences), became the object of truth claims and often frustrated Greeks, particularly Plato, because a “written word cannot defend itself as the natural spoken word can: real speech and thought always

---

4 Ong, *Orality and Literacy*, 25.
exist essentially in a context of give-and-take between real persons." The translation of plays, speeches and philosophy into written formats frustrated contemporary people because they could not argue with a text. Similarly the authors of texts, often dead or unreachable, remain at a distance and cannot be made to change their minds. One of the reasons books are burned is because, no matter how much you argue with them, they never change their writing and can never recant.

In other cases, especially where literacy is limited, writing is regarded as dangerous or magical, requiring an expert, priest or guru to control. In the English language tradition, this is most obviously seen by the charms and runes which are still being discovered on stele and hidden in manuscripts from the Middle Ages. Some of these writings “are probably among the oldest lines in the English language” and these magical writings became symbolically powerful even to those who could not read them. The magical power of written and spoken language is most apparent in Middle English words like spell, charm, and grammar. Spell referred to “speech, narrative, discourse,” though spells could be case for curse or protection, largely becoming a symbol of malignance during the Renaissance. Charm, which is usually considered a symbol of protection or good luck in English cultures today, both spell and charm were typically linked to oral performances. Grammar is linked to writing and “came to mean occult or magical lore, and through one Scottish dialectical form has emerged in our present English vocabulary as ‘glamor’ (spell-casting power).” The magical power of writing is preserved today not only in the power often ascribed to taboo language, but also in the popularity tattooing foreign characters, particularly Chinese, as a symbol for luck or for remembering some event.

It was possible to curse before writing, but there was no such thing as a curse-words, swearwords, or keywords before writing. Writing fostered, among other things, the idea that individual

---

5 Ibid., 78.
6 Ibid.
9 Ong, Orality and Literacy, 91.
words had meanings which were independent of their context, their speaker or writer, or their time. Ong points out that before writing, knowledge of skills could not be very abstract. Writing created the conditions for a belief in “objectivity,” particularly the practice of reproducing Learned Latin, which has no purely oral users. Writing also removed facial expressions or gestures from discourse, froze legal codes, and lessened the importance of memorization in education. Writing also allowed for verbatim recordings of performances and the possibility to confront their performance or duplication as “correct” or “accurate.” In Ong’s words, “This is one reason why ‘the book says’ is popularly tantamount to ‘it is true’.” Writing also expanded the concept of an audience from a group of people who literally gather around to listen (the auditory root of audience), to invisible members which could only be imagined by authors. The idea of an audience, outside of the author’s control, also stimulated the need to control content and legislate what could be written for whom. The Master of Revels had been instituted in England to control profanity on the stage, reading through scripts and censoring what could be heard by a play’s audience. When writing could be quickly reproduced, the content became important because the writer and publisher lost control over their audience. A vestige of this shift can be seen in the warning text which precedes many television shows or movies in the form of “For Mature Audiences Only.”

Establishing Good English

When the printing press was first introduced into England, it exaggerated individual and regional differences between written English. However, because most of the early printing presses were in operation in the South-Eastern part of England, print tended to reproduce that dialect at the expense of

10 Ibid., 45.
11 Ibid., 110.
12 Ibid., 41.
13 Ibid., 78.
14 Hughes, Swearing, 121.
others.\textsuperscript{15} It was not long after the invention of the printing press and the democratization of education, primarily vernacular literacy, that people began to call for set rules of writing. Print, and the need for writers to be understood by their readers, encouraged the first English dictionaries to be written. Walter Ong has argued that more than the invention of writing, “print produced exhaustive dictionaries and fostered the desire to legislate for ‘correctness’ in language.”\textsuperscript{16} The process of freeing words from their sentences or expressions, began by the invention of writing, became an even more radical separation with the construction of dictionaries and thesauri. The earliest English dictionaries dealt with translations between English and French, Italian or Latin. In 1592, Richard Mulcaster created the first English language dictionary which listed 8000 English words, but these were not alphabetized and did not always contain definitions.\textsuperscript{17} After it was found profitable to produce reference books, a great number were printed, but with varying degrees of utility or completeness. Many English writers were frustrated by the limitations of these books. The leaders of English literature (Alexander Pope, Daniel Defoe, Jonathan Swift, and others) began calling for the need to fix the language and create reliable depositories of every English word.\textsuperscript{18} It was not until Dr. Samuel Johnson’s \textit{A Dictionary of the English Language} was published in 1755 that dictionary writing required words organized alphabetically or carefully defined. Johnson’s dictionary curiously lacks almost every taboo word a person can think of, despite the fact that Johnson is widely believed to have suffered from Tourette syndrome and certainly new these words existed. Johnson’s dictionary would be the only popularly used dictionary in English for the next 150 years until \textit{Oxford English Dictionary} (OED) was published in small portions from the 1880s onward.

\begin{flushleft}
\textsuperscript{16} Ong, \textit{Orality and Literacy}, 127.
\end{flushleft}
The OED in particular demonstrates the periods’ obsession with counting and defining all the words in the English language. The OED was also a massively democratic undertaking, a precursor to Wikipedia and later crowd sourced projects, and required contributions from thousands of people with personal libraries filled with words. In order to create a definition, the OED editors asked people, using newspaper calls, to volunteer and comb through their libraries looking for first uses of specific words. Like a biography, every word to be included needed a birthday. People would record a word’s first use, along with its surrounding sentence and citation, and several other uses of the word and then mail them to the OED’s editors. The editors would then read these contexts and create a definition, which was supposed to encapsulate the meaning of the word without simply restating the word or a synonym. It is interesting to note that one of the major contributors to the OED was an American ex-army officer who had been confined to a British asylum after killing a man.

What is most important about these dictionaries was their obsession with reproducing written, and not spoken, English as well as their omission of words considered to be taboo. Aside from having Tourette syndrome, Samuel Johnson was also known to use impolite language regularly. When someone asked him what gave him the most pleasure in life, Johnson answered that first was “fucking and second was drinking. And therefore he wondered why there not more drunkards, for all could drink though all could not fuck.” When one woman criticized Johnson for failing to include obscenities to which he replied “No, Madam, I hope I have not daubed my fingers...I find, however, that you have been looking for them.” However, given the obsession with accounting and fixing written English, in many cases words were omitted with the hopes that they would not be reproduced, at least in written form.

Johnson realized it would be impossible to look through everything ever written in English and so

---

19 Ibid., 47.
20 Ibid., 56.
22 Hughes, Swearing, 144.
23 Winchester, The Professor and the Madman, 98.
decided to set limits to his dictionary project. He decided that English writing had peaked with Bacon, Edmund Spencer and Shakespeare. He decided not to include anything written before them or after the works of Sir Philip Sidney who had died in 1586. Within these limits, he declared, could be found all the instances of good, proper English minus an obscenity or two. One of the notable consequences has been that by displaying the various meanings of a word within a text, words are shown to have different meanings, many of which are irrelevant to their current use. The idea of language as a string of standalone words with independent meanings conditions the possibility for “swear words” with a stable taboo meaning. These bad meanings are more than heuristics which precede judgment. Judith Butler has recently pointed out that “if the very definition of the phenomenon involves a description of it as ‘evil,’ then the judgment is built into the definition (we are, in fact, judging before knowing.” Definitions, which have been exchanged with descriptions, are in fact both judgments.

Subjugation and Keywords for Thinking

Benedict Anderson has persuasively shown how nations became objects of collective solidarity through the combination of vernacularization, print, and capitalist production of written texts. The spread of dictionaries, whose original audience was writers, became commodities for middle-class consumption in the late 18th century. Similar to what Michael Shapiro has pointed out regarding landscape paintings and photographs becoming a means by which middle-class consumers could possess the English land and imagine themselves as members of the English nation, owning a dictionary became a way of owning the English language. With the expansion of England and the United States into unfamiliar territories, considerable effort was made to create an English or American literary

24 Ibid., 95.
27 Michael J. Shapiro, Methods and Nations: Cultural Governance and the Indigenous Subject, New edition (Routledge, 2003), 114.
canon which, along with dictionaries, became a cornerstone of schools by the 1870s.\textsuperscript{28} The writing of Elizabethan England, according to Richard Helgerson, was addressed to the issue of “who counts as a member of the nation,”\textsuperscript{29} but during the 18\textsuperscript{th} and 19\textsuperscript{th} century expansion of the British Empire and the United States, the dictionary became a technology of linguistic subjugation. British-English dictionaries became a technology for including and excluding who could be considered a member of the British nation. In America, the dictionary writing practice was headed by Noah Webster who produced several American English dictionaries and insisted on reforming the spelling of British English into the American-style preference.\textsuperscript{30} Dictionaries, and the official form of the English language they contained, included and excluded membership within the American nation. Teaching African American slaves to read and write was expressly forbidden in the United States while at the same time schools were established to teach Native Americans how to speak, read, and write in English.

Keywords have not only historically been a technology for the governance of language, but also have become evidence of malignant forms of thinking. Certain words have become keyed to racism, bigotry, misogyny, and other attitudes which must be governed. Recently, the OED announced that it would be adding four versions of the word 	extit{cunt} to its massive dictionary, a decision which has sparked protest in the U.S. but seems to be largely unimportant in the U.K.\textsuperscript{31} Many have noted that racial slurs have become the most taboo forms of language which can be uttered today.\textsuperscript{32} Recently, dictionaries themselves have become the subject of this regulation. In 1997, the tenth edition of 	extit{Merriam-Webster’s}

\textsuperscript{28} Ibid., 113.
\textsuperscript{30} Winchester, 	extit{The Meaning of Everything}, 67.
Collegiate Dictionary’s definition of nigger sparked a boycott which revolved around the power of a publisher to define the term. Those who supported deletion of the word altogether believed “If the word is not there, you can’t use it.” Merriam-Webster’s publisher responded to the requests to examine their definition of nigger is strikingly similar to the process which Google insists upon when questioned about their results. Merriam-Webster asserted their definitions are produced by technical processes, insisting they “do not invent the words that go into the dictionary, and they don’t decide what meanings they will have” and argued “a dictionary is a scholarly reference, not a political tool.” Others have asserted that these attempts to control dictionary definitions or prevent white kids from using them inappropriately demonstrate too much concern with what white people think. In either event, the idea that deleting a keyword makes its use impossible or that the removal of a word can remove a harmful institution has not been limited to dictionaries.

In 2011, Alan Gribben, a professor of English at Auburn University at Montgomery in Alabama, published Mark Twain’s Adventures of Tom Sawyer and Huckleberry Finn. In this publication, Gribben replaced every instance of the word nigger with the word slave. He also changed “Injun Joe” to “Indian Joe,” and “half-breed” with “half-blood.” For Gribben, Twain’s use of the racially charged terms represented a barrier for teachers and students wanting to read and discuss Twain’s work. Gribben explained his own experience, “For nearly forty years I have led college classes... in detailed discussions of Tom Sawyer and Huckleberry Finn ... and I always recoiled from uttering the racial slurs spoken by numerous characters, including Tom and Huck. I invariably substituted the word ‘slave’ for Twain’s ubiquitous n-word whenever I read any passages aloud. Students and audience members seemed to prefer this expedient, and I could detect a visible sense of relief each time, as though a nagging problem

33 Kennedy, Nigger, 137.
35 Kennedy, Nigger, 170.
with the text had been addressed.”

Missing from Gribben’s preface to his edition of Twain are more subtle critiques he might have used to argue against the text’s use of language. For example, he notes that Twain attempted to accurately depict social attitudes in the 1840s Mississippi, but does not consider whether Twain was attempting to accurately depict black speech patterns when he writes for Jim’s character. He does not discuss Twain’s depiction of Jim as a blackface minstrel or object to this as a white man’s poor depiction of a slave. Rather, Gribben claims he made the changes to language in Twain to protect students and teachers from the need to discuss the politics of racism or racial slurs in Twain’s, or our, time. Gribben’s alterations come in the wake of many debates about the suitability of Twain during the 1990s. His is only one potential solution. Articles in education journals during the 1990s, obviously written for white teachers, suggested alternative strategies for justifying the use of Twain and Twain’s use of the word *nigger*. Among these strategies was the suggestion of consulting with black students before discussing Twain’s books, avoid being seen as an activist or hypersensitive reverse-racist, and most importantly white teachers should invite “black speakers to the class, regardless of their viewpoint.” These strategies, along with Gribben’s substitution of *nigger* with *slave*, are more about protecting white teachers from black students than they are concerned with Twain’s intentions or the political history of the terms.

The process of ascribing meaning to individual words, which began with writing and radicalized by print, resulted in the construction of elaborate grammatical rules and dictionaries. These rules and dictionaries have been incorporated into electronic interfaces as spellchecks, autocorrects or predictive text keyboards. Digital dictionaries programs are responsible for ensuring the intelligibility and “correctness” of what we write into our devices.

37 Ibid., 11.
39 Ibid., 30.
40 Ibid., 29.
Language Governance after the Digital Turn

Print caused the belief that words are things more than writing alone had done. Throughout the Renaissance, writings had been meant to be read aloud. The ornate calligraphy of earlier ages was replaced by clear type, regular lines and boarders. This new form of legibility encouraged rapid and silent reading and shifted the practice of writing as being concerned primarily with production to a concern for consumption.\(^{41}\) Print also created the desire to legislate for “correctness” in language with all deviations from the norms established by dictionaries being viewed as corrupt. Where dictionaries had worked to liberate words from their context by listing words according to an alphabet and carefully defining their meaning, the structuralism which began with Ferdinand de Saussure and Roman Jakobson focused on the factual use of words within the contemporary language, but context was still important to their philosophy. The revolution of semiotics was to argue that language was not itself a system, but was instead a system of signs with a gap between signifier and signified. This allowed linguistics to both apply mathematical formalization and also prevent reducing a text (or a language) to one meaning. The post-structuralist (or deconstruction) movement demonstrated that words can move from one context to another, changing their meaning along the way. The concentration on the words of poetry and literature was only possible after the invention of writing, but “Deconstruction is tied to typography rather than, as its advocates seem often to assume, merely to writing.”\(^{42}\) According to deconstructionists, Jacques Derrida especially, the attempt to fix a normative context (such as writing a definition for a word) is a useless endeavor. This movement of words from contexts is understood to be infinite, making every question about word-meanings unanswerable.\(^{43}\) In a significant way, electronic communications are responding to the deconstructionist critique using methods which McLuhan, Ong, Saussure and Derrida could not have imagined.

\(^{41}\) Ong, *Orality and Literacy*, 120.
\(^{42}\) Ibid., 127.
McLuhan and others have argued that the move toward electronics will be a return to prehistoric communication and thought processes. Ong shares McLuhan’s assumption that electronic cultures will reduce all sensations to visual analogues, though he also argues that “electronic technology has brought us into the age of ‘secondary orality’...it is essentially a more deliberate and self-conscious orality, based permanently on the use of writing and print, which are essential for the manufacture and operation of the equipment and for its use as well.”

McLuhan had argued in the 1960s that this visual print culture would be replaced by electronic media which would change social organizations into a “global village.” Indeed, during this time McLuhan coined several terms which would become popular after the advent of the Internet, including the concept of surfing electronic medias. However, despite the growth of distribution networks for images, sounds and videos since the development of electronics, especially digital electronics, nothing has expanded as rapidly as print and textual literacy. Nearly all writing and certainly all printing is done today with the aid of electronic devices, word processors and electronic networks. Despite the decline in the number of newspapers and magazines over the past few decades, more and more people are interacting via email, social media and SMS text messages.

Numerous reading devices have been created which simulate printed text on e-ink or backlit displays. Massive projects have been initiated to digitize books stored in libraries by public (like Project Gutenberg) and private (such as Google Books) institutions. The most current estimate for the size of the American Wikipedia, counting only text, is at least 50 times the size of the electronic version of the Encyclopedia Britannica and is expected to continue growing. It is certainly many times later than the printed version of any encyclopedia which, even if we do not consider Wikipedia as a reliable knowledge-source, demonstrates at least the massive expansion of writing enabled by online writing

---

44 Ong, Orality and Literacy, 133.
medias. The last several decades has also seen an explosion in the number of companies who set out to
manage, filter, sort and sell this growing body of text. Search engines like Google, Bing, and Yahoo are
how people typically “surf” the Internet and these three engines handle by far most of the searches
done in the United States.\(^47\) Despite what is often argued, electronic devices are not eliminating printed
books, but are actually helping produce more of them. In May 2011, Amazon announced that its sales of
eBooks surpassed its sale of physical books.\(^48\) Since hosts usually do not publish how much text they
work with, one can only guess how much writing is being produced and collected daily by email, text
messages, and social medias like Facebook and Twitter.

**Thinking like a Computer**

The relationship between computers and political thought is as much about how computers
think as it is what they think about. This is not a reference to “thinking machines” or artificial
intelligence. Instead, what interests me here is how computers enact engagement between interfaces
and interference, the concepts of computation as well as its percepts and affects. Deleuze’s
understanding of how thinking happens through film can also be applied to the computer and its
interfaces; “It is at the level of the interference of many practices that things happen, beings, images,
concepts, all kinds of events.”\(^49\) The how of computational thinking becomes obvious when we see the
connections between the concepts and computational forms—interfaces, databases, algorithms,
peripherals, code, and networks. To see the interaction between computational forms and political
concepts, the understanding of language which makes digital filters which screen for keywords or deploy
algorithms to analyze stings of expressions is useful.


\(^49\) Gilles Deleuze, Hugh Tomlinson, and Robert Galeta, *The Time-Image* (Minneapolis: University of Minnesota,
1989), 280.
In computational thinking, language and communication are transformed into data and information. A user interacting with an interface is interacting with objects and properties translated into machine code. This digital code is not simply a reference to the digits of the hand, but references the basic units of computational dichotomies of zeroes and ones which make up digital language. Analogical language, by contrast, is what Deleuze calls “language of relations, which consists of expressive movements, paralinguistic signs, breaths and screams, and so on.”\(^5\) However, it is not enough to say that analogical language “proceeds by resemblance, whereas the digital operates through code, convention, and combinations of conventional units”\(^5\) because, among other things, code allows for the combination of abstract elements (what Alexander Galloway has called “ideology modelling”\(^5\)) and codes can be combined to produce a narrative or a message.

The keyword, originally removed from the context of the expressions in which they originally appeared and defined in dictionaries, becomes something quite different in computational culture. Language becomes a series of “strings,” a data type which treats words as an array of bytes stored in a sequence of characters. In theoretical computer science and mathematical logic, strings are sequences of symbols chosen from an alphabet. The technical treatment of language is less important here than how this technique engenders a different kind of political thinking and political consequences.

One of the first consequences of this shift is to treat people and things as bearers of regular expressions from which strings can be extracted and processed.\(^5\) Natural language is manipulated into something which an algorithm can order and something that a machine can understand. Google Translate, for example, does not transcribe one language into another by searching between dictionaries for a common meaning or reference. Instead, it deploys “statistical machine translation” in

\(^5\) Ibid.
which the computer learns to translate by comparing pairs of human-translated documents and
determining which translations are most often used. The meaning of the expressions is not referential to
objects, events, or linguistic relations but becomes the frequency with which human-produced
translations use a key word or phrase. What is interesting here is what happens in the process of
translation itself, not what the translated materials are about. Another result of treating people and
things as bearers of regular expressions is the possibility of rendering them as datasets amenably to
the technocracies which are playing an increasingly important part in exercising control and surveillance
over populations. Following Ian Hacking, we can see that the bureaucracy of statistics makes possible a
way of “determining classifications in which people must think of themselves and of the actions that are
open to them.”

Another political consequence of keywords in computational culture is the way in which
interfaces become a “point that gathers activity to it.” The recent expansion of social networks, which
was thought of as a reclaiming of the Internet for people rather than products following the dot-com
crash of the 1990s, has become an increasingly pervasive interface of activity. Communication and
language between people in this context is more than simple consumption, but is also a new way to
make time-wasting, idleness, and relaxation into forms of capital production sometimes called
“clickwork” or “micro-labor.” These medias “configure the nervous tensions of the end user into
productive conjunction with new motor skills; the dexterity of thumbs and fingers are valorized for the
production of more informational goods (and bads).” Perhaps more importantly, social media and
digital communications has made possible the indirect shaping of the self through programmatic

55 Fuller and Goffey, Evil Media, 110.
56 Ibid., 136.
57 Ibid., 71.
campaigns which “steer” users towards certain decisions, activities, or modes of thinking rather than attempt to socially engineer their behavior.

The final consequence of thinking through computers is the idea, largely inspired by Norbert Weiner’s theory of cybernetics, that if a device or institution is organized to structurally resemble a mind, it will inevitably begin cogitating and interacting with the world. In political thought, this is similar to the idea that if a sound decision-making structure is created the decisions which it arrives at will be “correct, just, or, at worst, the least bad.” Collective intelligence becomes a means for creating a polity which, in computational culture, is centered around information processing. In *The Control Revolution*, James Beniger noted that in the 1890s, General Electric invented a new information processor, “built of the collective cognitive power of hundreds of individual human beings.” Today, technology industries have succeeded in completing monumental “crowdsourcing” projects with or without users being aware of their digital labors. For example, Google digitized the entire archive of the *New York Times*, thousands of Google Books, and millions of street addresses using “reCAPTCHA.” reCAPTCHA is a program which is designed to the difference between human users and computers spamming websites by having the user enter texts presented in images which cannot be read by a computer. Lately, it has been popular to discuss Google as the realization of a “World Brain” dreamt up by H.G. Wells in the late 1930s. Wells described his world brain as a new, free, permanent and authoritative “World Encyclopaedia” which would help citizens tap into universal information resources and eventually lead to world peace. Less important than the debates about whether or not Google constitutes Wells’ World Brain is how collective decision making via computation treats language. Dumb filters, which simply search for a

---

specific keyword to return or block related content, are slowly giving way to new forms of collective language governance which still treats language as strings of words, but also encourages new kinds of political thought. Facebook and Twitter have become models for thinking about civic engagement and Google, Wikipedia, and iTunes have become interfaces for thinking about the future of politics.63

Codes for Cleaning Language

Increasingly, we practice dialogue with the world using digital technologies. Like all dialogues, that discourse is subject to rules and practices which separate signals from noise and control the intelligibility of statements or questions. The process of ascribing meaning to individual words, which began with writing and was radicalized by print, resulted in the construction of elaborate grammatical rules and dictionaries. These rules and dictionaries have been incorporated into electronic interfaces as spellchecks, autocorrects or predictive text keyboards. These dictionary programs are responsible for ensuring the intelligibility and “correctness” of what we write into our devices. By now, most people experienced the process having their word processing program fix a misspelling or adding a new word to the dictionary’s database. While it might be better to call this a soft form of coercion, there are spellchecks and censorship programs which force or prevent certain words from being entered. A notable example is the project launched by the Special Olympics against the “r-word” or retarded. In their campaign to “spread the word to end the word,”64 the Special Olympics managed to persuade Blizzard Entertainment, the company responsible for the popular massively multiplayer online role-playing game World of Warcraft and the strategic game StarCraft to prevent their players from using retard or retarded in their commenting system. Players who write the offending word and the people who receive the message the text displayed as r******. The “Profanity Filter” which runs continuously along with the game is known for blocking many taboo words, but has recently come under fire for its

---

63 Evgeny Morozov, To Save Everything, Click Here: The Folly of Technological Solutionism (PublicAffairs, 2013), 15.
blocking of the words *homosexual* and *transsexual*, causing many players to argue that Blizzard is prejudiced against queer communities.65

When Google rolled out its Instant search function in 2010, by which Google began searching as soon as a user started typing into the search box, *2600: The Hacker Quarterly* journal discovered a surprising number of words which would not display results unless the user hit enter or clicked the search button.66 Included in the blacklist are words like *anal, big tits, dick, scat, teen,* and *xx*. Google’s failure to search these terms using the Instant search seems rather arbitrary in comparison to terms which it will instantly display including *bastard, breast, erection, gook, oral sex, peep show* and *whore*. Why anything *teen* should be blocked, but *peep show* allowed is difficult to justify algorithmically. A recent update from Apple to the iPhone keyboard made it difficult for users to enter *abortion* or *suicide* into their smartphones. The driving question over debates about the surprising update was whether or not the dictionary loaded into Apple iPhones reflects Apple’s own political stance on issues like abortion and whether the removal of terms like “abortion” from the predictive text or spellcheck represents a form of censorship.67 Hidden in the sourcecode of a recent Android handset update, made by Google, researchers discovered a dictionary of over 165,000 words users would not get help completing from the built-in autopredict or spelling correction function of their devices. Included were terms like *preggers, intercourse, lovemaking, butt, geek, thud, LSAT, pizzle*, and all seven of George Carlin’s dirty words.68 In contrast to Apple, Google gave its handset users the option of turning off the keyword filter,

---


but the options are difficult to find and are enabled by default. Both Apple and Google clearly hope to steer users away from certain content and choices, even if it is simply not helping them automatically complete queries. Rather than try to engineer users, this form of language regulation works to provide a framework around which the user’s attention, enthusiasm, desires, and fears can be cultivated.

Another, less taboo, example of steering users with spellcheck is Google’s dislike for alternate spellings of keywords. For example, searching for the term *fuscia*, an alternate but “correct” spelling for the color, Google shows results the results for *fuschia*, asks “Did you mean: fuschia.” It also gives the option to “Search instead for fuscia” instead. If a user clicks the link to search for *fusica*, Google displays results for an entertainment company, a fine art gallery, a florist and a restaurant which all use the term *fuscia* in their company name. Google also keeps the prompt “Did you mean: fuschia” at the top of every search page for the term *fuscia*. A search for *fuschia* returns a Wikipedia definition, a technology company, and a cosmetics company. Google’s help documentation for the “did you mean” function states “Google uses spell checking software to check queries against common spellings of each word...Please note that our spell check feature is completely automated, and we cannot make manual changes to individual queries.” They do not add that there is no way to disable the “did you mean” option when using Google. How the automated system works is also not described by Google, but it is clearly not based upon number of results. The term *fuscia* returns 14,200,000 results whereas *fuschia* returns 14,000,000. While this might not suggests overt censorship on Google’s part, if a website is unfortunate enough to use the word *fuscia*, or offer products and services with this spelling, users will certainly have a more difficult time finding that information. In mainland China, a more explicit example,

searches for “human rights” often return “Did you mean hunan rice?” A Google representative in China deflected concerns about censorship by stating “It's legitimate to ask if someone was searching for hunan rice. People here like rice, a lot.” Google’s spellcheck, by suggesting one search term over others, attempts to steer user behavior towards pre-determined ends.

Ostensibly, Google relies upon the “network effect” in order to determine the relevance of a search query. While Google insists on making the operations of its algorithms, how the SafeSearch filter makes decisions about what to display, or the identities of those charged with manually manipulating results a closely guarded secret. In addition, Google’s default settings present coercive controls over a user’s ability to opt out in the first place. These default settings, which are constantly moving and changing, are also deceptively simple. When I last access the settings in May 2012, I was offered three settings for SafeSearch; strict, moderate or off. Google described the settings as “Strict filtering filters sexually explicit video and images from Google Search result pages, as well as results that might link to explicit content. Moderate filtering excludes sexually explicit video and images from Google Search result pages, but does not filter results that might link to explicit content. This is the default SafeSearch setting. No filtering, as you've probably figured out, turns off SafeSearch filtering completely.” In a troubling turn, Google recently changed these settings (as well as where these settings could be accessed) to a simple check box next to “Filter explicit results.” Even less description is given about how the filter operates and users’ ability to choose different options has been limited. The coercive power of Google’s filters are made even more obvious in their image search. If a user goes

---

71 Ibid.
to Google and searches for the word *porn*, for example, Google presents a popup box which requires a user to choose “Continue to results” before the search will continue. Perpetuating practices which construct children as in need of protection and to “keep families safe on the web,” Google’s SafeSearch filter seriously complicates central free speech claims. Because Google refuses to reveal how the filter or their central search algorithm works by invoking “trade secret” exceptions to Child Online Protection Act subpoenas, Google is effectively intervening in possible speech. By avoiding free speech claims, Google also demonstrates the shift away from legal or Constitutional understandings of protection against censorship. While three cases in Virginia have recently been concerned with whether or not clicking “Like” on Facebook constitutes free speech75 (it was considered protected), the legal discourse surrounding Google deal with copyright infringement, potential anti-trust actions, and trade secrets. Google is not a free speech engine, though it repeatedly defends the concept of free speech. While Google’s decision to move its Chinese searches to Hong Kong has been framed in a human rights and free speech, Google made the move for economic reasons and cooperates with local governments’ censorship requirements.

Another trend has been the tendency toward localization and the crowding out of alternatives. Location awareness and the tailoring of results to users has the effect of encapsulating people into what Eli Pariser has named a “filter bubble.”76 These bubbles are formed when algorithms selectively guess the information a user would like to see based on information about the user based on their location, their search history and behavior on the Web. These results insulate users from information they may disagree with and they can become isolated from surround discourse. According to Pariser, even when

---


not logged in, Google uses 57 signals to determine what kinds of results to display to a user.77

Additionally, researchers studying language on Twitter have recently found that slang from urban areas is being picked up and reproduced in rural parts of American and in many parts of Great Britain.78 Just as the dialects surrounding the early printing presses tended to absorb and dominate the kinds of English which could be intelligible, these urban slangs are influencing electronic discourse in ways McLuhan and Ong could not have imagined. The increasing prevalence of filter bubbles, especially those which use location awareness, present the possibility of destroying McLuhan’s future image of the global village.

Finally, even without active filters policing the intelligibility of language on the Internet, the restructuring of mental processes according to the idea of keywords has future implications. Google and other searches require that words be finally freed from the rules of grammar. Users become “word curators” who operate “with texts as with word clouds—he or she is interested not in what these texts ‘say,’ but in what words occur in these texts and what words do not.”79 Google has recently taken this a step further, offering reduce the sums of meanings presented with a single answer, by rolling out its semantic search. For example, if a user searches for “renaissance painters,” Google will display on the right side of a search, information about prominent renaissance painters and other information such as their birth and death dates. In terms of emerging issue, we might take this as an example of how search engines are getting better at understanding the meaning of words rather than requiring a radical dissolution of language into series of words. However, Google is still shaping what is considered important about a search; in other words, it assumes the most important thing about Leonardo DaVinci is that he died in 1519. A search for “fuck Renaissance painters” only continues to point to the use the individual words on the Web. Even if Google can find a semantic way to understand a searches’ meaning, Google is still determining which questions are legitimate. This also means that Google’s

77 Ibid., 12.
79 Groys, Google: Words Beyond Grammar, 12.
It is important to consider that the process by which words have been ripped from their context, first with the invention of writing and then with the practice of dictionary writing, were necessary conditions for the creation of search engines. The deconstructionist argument that attempts to fix a normative context are doomed to failure because of the unlimited meaning which can be described to words has been challenged by the development of search engines. These engines operate by rejecting the post-structuralist argument that word meanings are unlimited, in fact they will show you every context in which they are used if entered into their query. Search engines, especially ones which search the Web, display the sum of all uses of a particular keyword searched to show all the contexts in which they are used. Google, as by far the most popular search engine in the English world has considerable influence over how electronic discourse is conducted. By searching for the uses of a particular string of words, Google effectively controls the intelligibility of a query. As Boris Groys has recently argued “One can say that every word becomes characterized by the collection of its meanings—a collection of the context that this word has accumulated during its migration through language.” However, by collecting and displaying the collection of a particular word’s context, this conception of Google assumes a sort of neutrality and equality between meanings. The power of Google’s search algorithm has been its organization of word-meanings by a process they call PageRank named after one of Google’s founders, Larry Page. The PageRank organization promotes or demotes links displayed on a search based upon a large, and continually variable, set of conditions. Google searches tend to privilege contexts which have been clicked most often by other Google users doing similar searches, webpages which are mostly text

80 Ibid., 10.
(it still cannot read Flash or text stored in images), and older sites with content which is frequently updated. Many studies have shown that most Google users do not click links which appear past the first few pages of a search, many never go past the first page or the first few results. This has generated an entire industry of pseudoscientists attempting to predict what changes will privilege sites in a search, the so called Search Engine Optimization consultants.

While Google continually “refines” its searches to better direct users to relevant websites, it also punishes webmasters for trying to exploit the algorithm. For example, in 2011 Google launched a campaign against websites who purchased backlinks (links to the website posted on other websites), which gamed Google’s ranking algorithm by making the site appear relevant and important. Overnight, sites which appeared in the top few organic links were demoted to obscurity. An entire industry has evolved surrounding the pseudoscience of predicting what Google will change next and how to make sure a business or product remains visible. The Search Engine Optimization (SEO) gurus behind this pseudoscience claim to improve a site’s visibility based upon some accurate observations about how Google works. While other search engines are often given some attention, because the vast majority of searches go through Google, most SEO consultants focuses intensely on improving visibility on Google. Because Google closely guards its ranking algorithms, many of the SEO programs market themselves based upon years of experience reorganizing websites the way Google prefers them. At first it seems odd to market oneself as an SEO guru based upon predicting Google based upon past practices which no longer apply and could actually now lead to lower visibility. While it might be that spending years of predicting Google’s secretive algorithm might make a person better at increasing a website’s visibility,

---

and there are simple rules which do not seem to change, the marketing of successful SEO consultants is better understood in terms of understanding how to create website content for Google more than it is concerned with human readers.

Another example of how writing for Google be found in Google Scholar which, using algorithms and keywords, organizes information and displays results in a similar process to Google Search. The problem with Google Scholar is that its understanding of relevance may not be the same understanding a specific discipline or scholarly community applies. Articles which have made it to the top, because they are well written according to Google’s standards or by sheer luck, receive a higher ranking on Google. Being higher, they are read more often and become cited in other scholars’ articles. Because of the network effect, these articles, which may have only been listed higher for an arbitrary reason, are cited and keep their high ranking. The result is Google’s bibleometric understanding of relevance replacing other standards for “impact” or “value” in scholarly communities. Articles Google likes become increasingly relevant to their textual community, shapes what is researched, and contributes to decisions about which research gets funding. In these contexts, Google does not shape discourse through direct manipulation, but works through the interfaces of personal computing to multiply its influence on what is sensed, intelligible, and worth researching.

Conclusion: Irritating and Fucking with Control

Through computational culture, the keyword is increasingly a technology for ordering the seeable, the sayable, the thinkable, and the valuable. Search engines which treat language as a string of words does not imply that readers must become automatons to use them, but it does encourage human readers to automate their reading habits or treat other inputs as data. Fuller and Geoffey point out that the recent popularity of the self-help book Getting Things Done and the recent explosion of

83 Vaidhyanathan, The Googlization of Everything, 188.
84 Fuller and Goffey, Evil Media, 10.
speed reading web applications comes “from the way its techniques enable individuals to deal with—or rather process—much more data, provided that they treat them as data.”\(^8^5\) The Quantified Self movement encourages users to keep track of their daily activities, reducing them to data, so that machines can optimize their behavior later. This is not regulating their words so much as their activities, but the production of data for data-mining purposes is a semioticization of behavior the way search engines semiotize language online.\(^8^6\) These schemes demonstrate that the transcendental mastery of language and behavior is possible, but it is also repressive.\(^8^7\) The discipline societies of the past, with their focus on discipline and biopolitics, are being displaced by what Gilles Deleuze has called “societies of control” which are no longer restrained by enclosure structures like the barracks, school or factory. Instead, we have the freedom to say or do whatever we want, within circumscribed parameters, but members of the control society fall within those parameters without thinking about it. For Deleuze, the computer is the mechanism which best suites the control society. It has free many of us from the factory and allowed us to work from home, for example, but now we are expected to be responsive to the demands of work when away from the factory or office. Additionally, in our control society, we know that we are being tracked but are encouraged not to worry about it. In the control society, the only forms of discourse which might challenge these oppressions fall outside the parameters of the seeable, sayable, and thinkable. Radical indictments of our political system become “terrorist” actions, for example, and our searches online or our tracked routines are monitored for terrorist keywords.

If we are at all interested in escaping or reshaping these mechanisms, Deleuze suggests, “The key thing may be to create vacuoles of noncommunication, circuit breakers, so we can elude control.”\(^8^8\)

\(^8^5\) Ibid., 108.
\(^8^6\) Ibid., 118.
Creating these vacuoles is not easy when computational culture demands that we treat language as primarily communicative. Bad, taboo, and obscene words are useful for demonstrating the limits of transcendental mastery of language. Julia Kristeva noted that obscene words and poetic language create a “permanent contradiction between these two dispositions (semiotic/symbolic), of with the internal setting off of the sign (signifier/signified) is merely a witness, poetic language, in its most disruptive form (unreadable for meaning, dangerous for the subject), shows the constraints of a civilization dominated by transcendental rationality. Consequently, it is a means of overriding this constraint.”

Taboo language and “bad” words can become a means of evading control in systems which make language and action into keywords or data.

Steve Jobs’ famous 1983 superbowl advertisement for the Macintosh computer targeted dealer’s desire to break with IBM’s control of the computer industry. In 1997, Jobs’ “Think Differently” marketing campaign was not encouraging us to think differently, but to identify with the brand of the different. The need for vacuoles of noncommunication where we can think differently—think outside the parameters of control and stumble upon alternatives—becomes more urgent as technology companies work to direct desires by developing what they call “autonomous search” engines. Autonomous search, according to the director of Microsoft’s Bing, will operate “on your behalf—without an explicit request—without you initiating it via a query...Today the trigger is ‘keyword’ plus ‘enter.’ But tomorrow...we’ll be told what we want before we know we want it.” Autonomous search threatens to further limit our capacities for thinking differently. Using the irritating keywords of bad language is one good way to fuck with the computational systems of control.

---

89 Kristeva, *Desire in Language*, 139–140.
References

http://www.bl.uk/learning/langlit/dic/mul/elementarie.html.


http://support.google.com/websearch/bin/answer.py?hl=en&answer=1723.


http://www.lynda.com/Google-Website-Optimizer-tutorials/essential-training/61175-2.html?srchtrk=index%3A9%0Alinktypeid%3A2%0Aq%3Agoogle%0Apage%3A1%0As%3Arelated
case%0Ase%3Atrue%0Aproducttypeid%3A2.


