A "Graphy" of Words:

A History of English Lexicography and the Changing Shape of the Language

❖

Christopher Bargman
Geography Program / University of Nevada, Reno / May 2018

ABSTRACT:
This research project uses GIS and database management tools to determine qualitative, quantitative, and spatial changes in English lexicography. With an understanding of the quantitative growth in word entries from the 1st English monolingual dictionary to the most recent 3rd *Ed. of the Oxford English Dictionary Online*, a proxy can be made regarding new word entry growth over space and time. For this project, new words to the *OED* from March 2016 - January 2018 can be analyzed to help determine the most recent changes to the shape of language.

KEYWORDS:
lexicography, *Oxford English Dictionary*, *OED*, centroids, etymology, Zipf's Law, spatial analysis, informational content, prescriptivism, descriptivism

PROBLEM/CONTEXT:

The persistence of language echoes in all directions, not dissipating through time, but rather redirecting its course like a boat at sea. As the boat turns and shifts its sails, the ripples in the water delineate its own progress, encoding the boat's existence into waves that continue their path onwards to shore. The first auditory utterances of human language left very similar water-like ripples in the master chronology of life, faintly decipherable, yet nonetheless there. As these ripples of language grew into written forms, the proxy was set for a semantic encoding of the world and its rich, self-referential history. Born within the ontological and topological human condition of man, language and the word marked the dawn of mankind's first information age. The word alone, in its semantic and syntactic singularity, became humanity's most basic and fundamental unit of measure for defining the world.

Focusing in on the English language, there lays a well-developed archival caché of its own youth, upbringing, and continued proliferation. Books, dictionaries, spoken-word transcriptions, archived documents, and a myriad of other textual artifacts serve a dualistic purpose of explaining our historic past while stabilizing the proxy for the language's future changes. The role of the lexicographer is as self-referential as any; through his dictionary he attempts to serve as mankind's most faithful and accurate mirror of humanity's outward sprawl. As James Gleick put it forth in his book, *The Information: A History, A Theory, A Flood*, "the dictionary ratifies the persistence of the word," (66). In doing so it also ratifies humanity's endless quest for information and knowledge. This report will take into account the lexicographic growth of the English language from the origin point of Robert Cawdrey's *A Table Alphabeticall* in 1604 to the present neo-amalgamation of the *Oxford English Dictionary Online*. This period between 1604 and present day will serve as a benchmark to analyze English lexicography by its quantitative growth in words. An extensive analysis of new words coming into the *OED* between March of 2016 and February of 2018 will be used to measure the spatial

influence that the modern English language has on the world, and in doing so will also precisely answer 1.) where these words are coming from, 2.) what types of words are being brought in, and 3.) how many words are associated with different word classification criteria.

The history of English sets a paradigm for the inherent richness and diversity of tongue that our language and dictionaries showcase. This paradigm is simple: our language takes in and borrows from other languages that which it doesn't possess, and this open-sourcing of English syntax and semantics has taken place during all epochs of the language. As one peers across the epochs of English language from Old English to Modern English, a unified theme and process of language is seen: *Immigration, Cultural Infusion, and Language Restructuring.* This same process has carried through to modern-day English where our working language and dictionaries show no distinguishable difference between the linguistic influences made by modern-day Hindi speakers migrating from India to Great Britain, and the linguistic influence left by the marauders and pirates of the Germanic languages as they plundered their way into the English language during the Old English period (500-1100 AD). Both have left their own permanent encoding into the language without barring future linguistic contributors from doing just the same.

As fascinating as this unique characteristic of English may be, it is considerably different from its greatly influential neighbor, the Roman language of the Roman Empire (modern-day Italy). The Roman language and English language share many cognates, as the legacy of Rome's conquest of the British Isles (43 AD - 500 AD) with the continued presence of the Roman language as the *lingua franca* across Europe created a strong diffusion into the English lexicon. However, English and Roman linguistic families are distinctly different in terms of rules of prescribed language usage rules. The Roman language follows a legacy of *prescriptivism,* or the "prescribed" rules in which a word and its spelling, grammar, meaning, and usage must be spoken or written (Lerer, Ch.1). In sharp contrast to the strictness of the Roman language system is the system of *descriptivism*, which the English language and its modern lexicographers use to "describe" and document the language without placing rules on how it must be used. As this concept of prescriptivism and descriptivism sets the tolerances for how a culture assesses their dictionary's contents, the reader must also make an evaluation towards, "What is meaningful information about place, location, culture, and society? What kinds of words stay in a language?" (Lerer, Ch.1). This strategic choosing of "what" and "how" a word gets to be placed inside a dictionary helps resonate its qualitative brilliance; the organic growth of human thought is moving in synchronicity with the organic growth of the lexicon.

As we approach the outermost edge of modern lexicography, the behemoth *Oxford English Dictionary* appears in all directions. The *OED* first came to life in 1895, and it quickly supplanted the "go-to" dictionary of the time, Samuel Johnson's *Dictionary of the English Language* (ca. 1755) (Cowie 230.) Since its founding in 1895, the OED has yet to relinquish its hold as the authority in English lexicography, both in size and in sheer categorical detail. Other dictionaries still coexist and move forward within the *OED*-ruled world of words, not beating the *OED* with sheer volume, but rather contrasting the *OED* with stylistic differentiations. Take for example the idea of the dictionary from an American perspective, and look at Noah Webster's *An American Dictionary of the English Language (ca.* 1828*)*. Aside from containing new words that were strictly American in their origin and usage, Webster's Dictionary consciously chooses to not include highly '*inkhorne*', or intellectual words, while also choosing to remove words that are archaic or no longer in everyday usage. His "working model" of the dictionary gives a weighted precedence to the statistical occurrences of

words in everyday life, while the more historically-minded ***OED*** looks to capture every aspect of the language, including the most obsolete or rare words plus all the statistical commonalities of the Webster model (Cowie 214). While this debate between "what" and "what not" to include in a dictionary grew stronger, it's true implications can be arguably best understood by the people of the modern 21st century. As living participants in the *Information age*, concepts such as *informational content* and *information overload* are indeed real phenomena, and anyone who's dealt with the PowerPoint presentation can understand this.

When the ***Oxford English Dictionary*** moved online in 2000, it had condensed the 2nd Edition from its twenty volumes, 138 pounds, and 22,000 pages with over 400,000 words into a weightless dictionary of the binary and digital domain (Gleick 67). It was now available everywhere and anywhere, at any time it needed to be summoned. The project of digitizing and encoding the entire contents of the ***OED*** began in 1989, with the *Dictionary* finding the first fruits of its success in 1992 with the publishing of ***OED*** in CD-ROM version (*public.oed.com*). From the initial completion date and unveiling of the ***OED Online*** in 2000, the 3rd Ed. of the *Dictionary* was official, but the work was not yet done. The master network of lexicographers working for the ***OED Online*** had also just unlocked the hyper-speed editing capabilities of their current work; the necessity for revision and updating of the O***ED*** now glared off the screen and into their faces more harshly than it ever had before.

Existing on a web-based platform that feeds itself on '*real-time*' data, the language inside the ***OED*** took on a new and living aura, as from year 2000 to modern-day the *Dictionary* has continued its ongoing editing process. In year 2000, the ***OED*** team of lexicographers began dissecting their dictionary from letter "M" and forward, both revising former definitions and sources that were in need of clarification and also adding more new and uncharted words to the lexicon. Each completed revision session would be published online about every three months and would contain all the new word entries and sub-entries. As stated from their official online platform regarding the revision process, it boasts that,

> "At present the dictionary is undergoing its first thoroughgoing revision and update. Around 70 editors, mostly in Oxford and NY, review each word in turn, examining its meaning and history, noting where meanings have changed – or where old definitions no longer suffice – and recraft the entries in the light of the most up-to-date information. The result is the current online edition of the dictionary (in progress)."

By using the ***OED Online*** as a primary data source for this study, it allows access to a pedigree of information that is of the highest quality in its field. Additional information given beyond the definition includes: first documented date of use, examples from literature, etymology and word origin, taxonomic classification types, region of use categories, and context of usage categories. All of these informational class types will be collected and databased for use in the ArcGIS analysis, with a primary focus on the origin-etymology of each word. Understanding where the word originated from or what culture it's being "borrowed" from will help determine the spatial measure of the English language and its sphere of influence.

So, more precisely, "how does a word become elected to be part of the ***OED***?" Well, aside from the 70 editors mentioned in the quote above, there is a team of over 200 lexicographers that assess and cross-reference the language (*oxforddictionaries.com*). This team of lexicographers statistically weights the importance of each potential new word by a.) its total quantified occurrence, b.) its range in different sources of text, and c.) a qualifiable weighting of the word's significance through

time. The age constraint from a word's genesis to its entrance into the *OED* used to be a 2-3 year minimum (*oxforddictionaries.com*). However, within our web-connected world a new word can have an instantaneous influential force so intense that language and the *OED* have no other choice but to acquiesce to it. Today, there is no minimum age constraint on a new word; the selective force lies in the lexicographer and his decision to choose which words will last the test of time (*oxforddictionaries.com*).

Looking deeper into the science of word selection, the *OED Online* has 4 primary sources for its word and language content. The first and foremost source is the *Corpus*, which is "a collection of written or spoken language data in a computer readable format," (*oxforddictionaries.com*). Entire documents sourced largely but not entirely from the web are digitized into the Corpus,

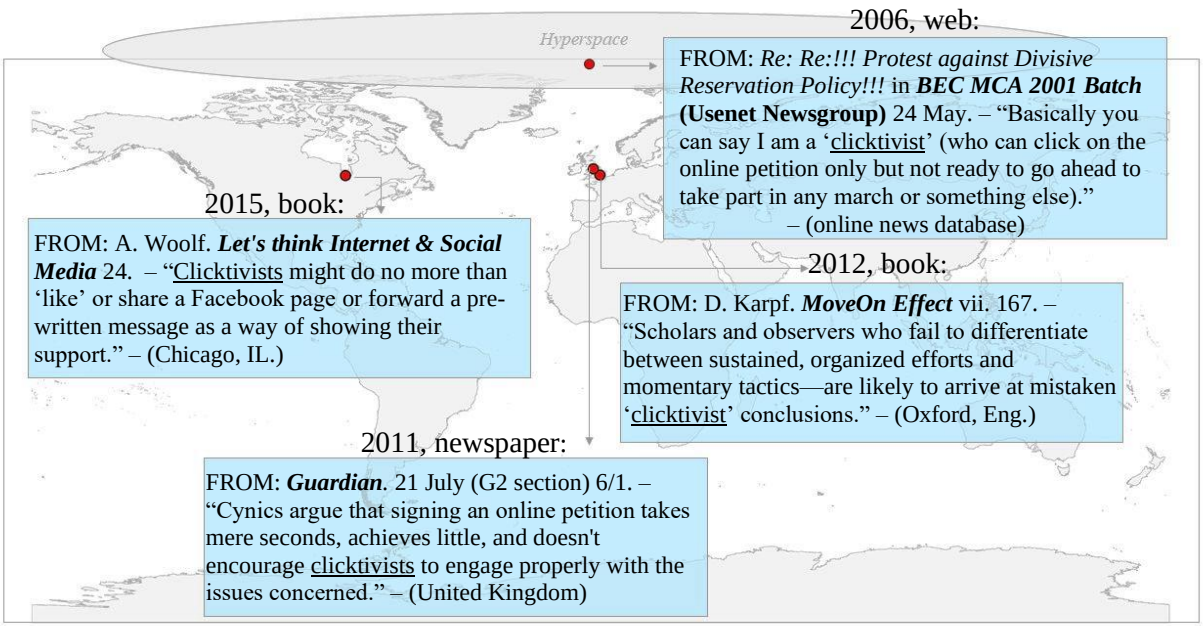

Fig.1: *Cited sources from the OED Online for the word "clicktivist", added in September of 2016. The varied types of sources reflect the word's broad diffusion into the language. The entire collection of sources is not shown here, but can be conceptualized as "links" inside the OED's "Corpus" and other word databases. Just like Google web-page ranking, the OED program language gives strength to the number of word links and diversity of link types when determining a new word.*

which now contains over 10 billion words growing outwards at a pace of 150 million added words per month. The second source for electing a new word is *User-Generated Content*, which is based on community input towards what new language is appearing in the world. Anyone with access to the web can openly contribute to this content. The third source for the *OED* is called the *Reading Programme*, and it collects its data from recruited readers who collect and digitize short extracts from writings ranging from song lyrics to scientific journals. The 4th and last channel for *OED* source content is called *Appeals and Submissions*, and this comes from public participants who send in requests for amendments to existing word definitions and also provide new sources of documentation for these words (*oxforddictionaries.com*). Altogether, the *OED Online* has a robust and varied system of language databases to help standardize their word selection process and maintain a maximum foothold within the universe of language.

GIS PROCESS AND ANALYSIS:

Acquisition and organization of word data-

Engineering the GIS build and analysis required an extensive emphasis on data acquisition and management. Data was acquired in the form of dictionary definitions given directly from the Oxford English Dictionary Online database. Words chosen for the data included all the *new word entries* from the seven most recent OED *New Words Lists*. These seven lists, or datasets, ranged from March 2016 - January 2018 and totaled a number of 1,389 words. Each *New Words List* was rendered into two versions of Excel CSV files: 1.) a CSV file organized by grouping all words with their country of origin in column A, and 2.) a CSV file organized by an alphabetical list of all words in column A.

Fig.1: *OED New Words Lists* included in study:

- March 2016 → 207 words
- June 2016 → 249 words
- Sept. 2016 → 197 words
- March 2017 → 137 words
- June 2017 → 182 words
- Sept. 2017 → 216 words
- Jan. 2018 → 240 words

= 1,389 total words

Grouping the data into two different-styled spreadsheets made it easier to further analyze the data based on classification criteria. One final step for the country of origin CSV file was to copy and paste each country's word data to another CSV file containing all 261 political entities in the world shapefile, along with adding the FID number for each country and adding a new row containing the total number of word entries for each country. Once the initial dataset of March 2016 was completed in CSV form, then each subsequent *New words list* in CSV form was joined to it. The resulting data set structure can be seen to the right, as can the CSV file format below.

Fig.2: *Database structure for the ArcGIS build.*

Dataset 1: Mar. 2016 csv file
Dataset 2: Mar. 2016 + June 2016 csv files
Dataset 3: Mar. 2016 + June 2016 + Sept. 2016 csv files
Dataset 4: 2016 csv files + March 2017 csv file
Dataset 5: 2016 csv files + March 2017 + June 2017 csv files
Dataset 6: 2016 csv files +March 2017 + June 2017 +Sept 2017 csv files
Dataset 7: 2016 csv files + 2017 csv files + Jan. 2018 csv file

Key	Country of Origin	Number of Entries	New OED Entry	Part of Speech	First Date Documented	Etymology	Source: Borrowed (Foreign)	Classification A (if given) (word 1)	Classification B (if given) (word 1)	Cla (if
0	Aruba									
1	Afghanistan	1	boteh	n	1917	Persian	Yes			
2	Angola	1	likembe	n	1948	Lingala	Yes	Music		
3	Anguila									
4	Albania									
5	Aland									
6	Andorra									
7	United Arab	2	Ansar	n	1697	Arabic	Yes	Islam		
8	Argentina	3	botija	n	1588	Spanish	Yes	Hist.	Food	
9	Armenia									

Fig.3: *An extract of the June 2016 CSV file showing database organization structure. Not seen are the subsequent word entries stretching to the right and the entire list of countries stretching down to the last Key value of 260.*

ArcGIS build and analysis-

The goal inside ArcGIS was to create an animated flow map showing the conceptualized idea of words "flowing" into the ***OED*** and English language from 2016-2018. Being that there were 7 total datasets in the study, a total of 7 independent maps were exported from ArcMap to create a final flow-map animation.

First and foremost, the free political world map shapefile from *naturalearthdata.com* needed to be tailor-suited for the study. To do this 'cleaning-up' of the data, it required some basic editing in ArcMap. The first edit was with the 'Split Polygons Tool' in order to separate England and Scotland into two unique polygons. The original shapefile only had one single FID value given to *Great Britain*, but for this study it was very important that England, Scotland, Wales, Ireland, Northern Ireland, and Isle of Man would be treated uniquely. Once England and Scotland were split, then the same procedure was performed for Wales and England, followed by Ireland and Northern Ireland.

After splitting the polygons, it was necessary to use the 'Merge Polygons Tool' to bring multiple land masses into a unitary land mass. This was necessary for the precision of coloration for the choropleth map. The goal was to only give a value to the main political entity and not its entire dominion, as was the case for the scattered small islands all pertaining to Great Britain. From merging polygons, the next course of action was to use the 'Explode Polygons Tool'. This was needed in order to give non-sovereign provinces a unique FID value, and was the case for France and its provincial control of Martinique, Guadeloupe, and French Guiana.

The final polygon editing process was to take the original world shapefile and make a duplicate version without political country borders. This was achieved by using the 'Dissolve Tool' to dissolve countries based on continent type, and the final edited shapefile was used for enhanced visual style in the final flow map. After cleaning up the world shapefile, it was ready to be joined with the CSV files with the word data. A total of 7 unique map datasets were made, with each dataset joining our cleaned-up shapefile with a CSV file for the corresponding *New words list* publishing dates. Once the CSV files were joined and exported to create a new shapefile in the geodatabase, the new map could be give a choropleth color hierarchy based on the attribute table values for "number of word entries". The next step was to collapse the centroids to polygons, which was most smoothly done by running an SQL statement for "Number of Entries ≥ 0", and then exporting and adding this selection as a new layer. From the new layer, 4 new rows were added: x_coord, y_coord, x_start, and y_start. To make the centroids, the x_coord and y_coord rows need to be given a single x and y coordinate value. This was done by using the "Calculate Geometry" tool to find the x and y values for each row, and then running the "Feature to Point" data management tool to create centroids for all countries with at least 1 entry into the ***OED***.

Centroids, which are point data, take the all the x,y values for a polyon's perimeter and average them into one single point with a single x & y value. This step was done so that spatial analyses regarding distance from England to other countries with contributions to the ***OED*** could be measured given an exact value. These measurements of distance were correlated with number of word entries to analyze the diffusion of the English language in terms of new word entries to the ***OED***.

To finalize the GIS build, flow lines were created as a largely visual element to illustrate the amount or "flow" of language coming into the *OED* from surrounding countries. Flow lines were achieved by using the field calculator to give the new rows "x_start" and "y_start" the x & y coordinates for the England centroid (-1.461404, 52.596667). With this information, the "XY to Line" data management tool was executed to create flow lines drawn as *great circle* lines with an ID value corresponding to the "number of word entries" field. The resulting creation were flow lines that were proportionally sized to the number of word entries for each country.

After exporting 7 final flow line maps ranging from Dataset 1 -Dataset 7, the maps were added into Photoshop as a single "script" for the map animation. The ultimate purpose of the animation was to illustrate a change over time that reflected the *OED*'s nearly constant updating process to the English lexicon. The map was saved in GIF format to preserve its animation properties.

Fig.4: *Total Diversity of Languages: (Mar. 2016 - Jan. 2018)*

English = 991	Malay = 10	Indonesian = 2
Latin = 80	Uncertain Origin = 10	Persian = 2
French = 56	Urdu = 9	Old High German = 2
Hindi = 26	Chinese = 9	Romanian = 2
German = 23	Cantonese = 7	Tibetan = 2
Spanish = 22	Afrikaans = 5	Algonquin = 1
Tagalog = 20	Dutch = 4	Anglo-Manx = 1
Sanskrit = 19	Scandinavian = 4	Anuak = 1
Greek = 14	Arabic = 3	Bengali = 1
Japanese = 13	Korean = 3	Choctaw = 1
Italian = 12	Panjabi = 3	Creole = 1
Unknown Origin = 11	Portuguese = 3	.....

..... Danish = 1, Guajarati = 1, Hawaiian = 1, Hebrew = 1, Icelandic = 1, Kimbundu = 1, Ma'di = 1,Low German = 1, Mandarin = 1, Middle Dutch = 1, Thai = 1, Vietnamese = 1 Middle High German = 1, Pali = 1, Romani = 1, Swedish = 1

RESULTS:

Of the 1,389 total new word entries to the *OED* from March 2016 - January 2018, 531 words (38%) were "borrowed", or brought in from other countries. The remaining 858 words (62%) originated from within the English language of mainland England. There was a total of 49 distinct language groups that made their way into the language, with values given to the quasi-language groups of *Unknown Origin* and *Uncertain Origin*. With English as the most obvious influencing force to the *OED's* linguistic arsenal, the 2nd, 3rd and 5th most frequent language groups to be added were Latin, French, and German. This immediate result shows a modern language synthesis that runs parallel with the historical pillars of the English language: Latin, French, and Germanic-Scandinavian language groups (Winchester, 8-11). An analysis of the "first documented date of use" category revealed that of the total new word entries, 9 words first appeared in the language during the Old English period, 40 words first appeared during the Middle English period, and 1,340 words first appeared in the Modern English period.

Looking at the categorical data for "Classification Type", a total of 94 different classification types were associated with the word set. Not every new word was given a classification type by the *OED*, but for those that did the highest proportion was found to be the "Science" classification type with 162 entries. This evidence suggests a relationship between the high number of science-related words added to the *OED* and the ever-growing importance of science in the 21st century. Another high-frequency classification type was "Food", which ranked as the 5th largest field type with 86 entries. When graphed on a scatterplot, the relationship between a country's distance from Oxford, England (x-axis) to its number of food-related words (y-axis) showed a discernable

positive-sloping correlation, with a correlation coefficient of: y= 0.0003x + 2.4932. While this mode of spatial analysis can be suspect to considerable measures of false causality, it still infers to the geographer that a true correlation may reside in the data. The plotted correlation proposes that as distance from Oxford increases, so does the number of food entries for each participating country. Perhaps this correlation has a connection to western society's desire for the exotic, the new, and the distant when looking for food and culinary delight.

For the data regarding word origins, initial results were inconclusive for determining a quantitative spatial diffusion pattern of English language. A spatial analysis was designed to test diffusion of the English language from Oxford, England outwards to the countries who contributed *new word entries.* The method for the analysis was to plot each country on a scatterplot in relation to distance from the Oxford centroid (x-axis) and its corresponding number of word entries (y-axis). While the final flow map did an excellent job at visually illustrating the distribution of countries that contributed words to the ***OED*** between 2016 - 2018, a calculable diffusion of the English language over space and time wasn't achieved. On top of that, using this kind of spatial analysis method can produce a false causality between the two variables being measured.

Altogether, some striking associations can be made with the overall proportions of word contributions by each country. Historical relationships with the diffusion of the English language can be easily viewed on the map; nations of former-British colonial rule have a proportionately higher amount of word entries than their neighbors. Other meaningful associations include regional geographic zones such as the Caribbean and Southeast Asia. The Caribbean contributed a disproportionately large wealth of language added to the ***OED*** in relation to the region's aggregate landmass (165 total entries). This relationship also reflects a strong historical context to the region's culture as a biproduct of European cultural infusion. As for Southeast Asia, the region was another very surprising revelation in the research study, and also provided a disproportionately large amount of new language (83 total entries). Within Southeast Asia lie some unique English varieties that include *Hong Kong English, Singapore English, Malaysian English, and Phillipine English*. This unique region exemplifies a statistical oddity given by lexicographer and university professor Anne Curzan, who states that English is the only language in the world with more people who speak it as a second language than those whose speak it as first language (Curzan, 12). The language of English is truly the *lingua franca* for the world at large.

LIMITATIONS:

First and foremost, the dictionary is a qualitative source of information. While the ***OED*** has taken the art of lexicography and adapted it into a "quasi-science" through statistical weighting of word links, word frequency, and rigorous study of historical and cultural context, the fact still remains the same: a dictionary is a qualitative tool to measure the world. Even a concept as seemingly scientific and taxonomic as creating an alphabetical list is a qualitative measure based on imposed semantic rules. This study used the dictionary as the primary source for qualitative, quantitative, and spatial measure, putting sensitive dependence on the initial conditions provided by the ***OED*** team of lexicographers. On top of that, the study didn't cross-reference the ***OED*** with any other modern dictionaries. While this study was accurate in its explanation of the lexicographic process of putting words into a dictionary, the study failed to really go deeper on the quantitative realm of human language and its spatial interconnectedness. As James Gleick eloquently describes language,

"all words, taken together, form an interlocking structure: interlocking because all words are defined in terms of other words," (*66*). This quotation defines the true matter at hand; language is a *complex system* that can only be understood through the summation of all its individual constituents. This being said, further research must be done.

Other limitations to consider were the "arbitrary" choices made in the process of organizing the data. For example, the ***OED*** gave many words a specific "Origin type" and "Classification Type" that designated a word as pertaining to a certain country or region. A word like *arroz con pollo* contained origin and classification types for *Spanish*, *Latin America*, and the *Caribbean* regions, resulting in the word value being attributed to each political entity in the region. This value allotment can create a map that misleads the reader to think that more unique words are coming from each county, while in reality many countries "share" the same word. Also, the manner which this study drafted boundaries for different world regions may be considerably different than what another study delineates.

Lastly, there was a considerable limitation in the method of spatial analysis used. Plotting two variables on an x,y plane to show a relationship is highly susceptible to create an *ecological fallacy* implying false causality with the two variables. Measuring diffusion of language in this manner puts a tremendous level of uncertainty in the results. To lower this *uncertainty principle*, a better spatial analysis method must be used that approaches a closer reflection of the true organic shape and movement of language. A possible alternative is to give *spatial weights* to various key hubs of English language across the globe, and then assess language diffusion as the product of an interacting spatial network.

FUTURE RESEARCH:

1. Mapping memetic diffusion of language and ideas-

An issue that wasn't deeply addressed in the research study deals with the "vehicles" on which language change takes place. While history and immigration are two obvious and well-known vehicles for language change, there is much to be learned about the informational associations and abstractions of the human mind, and how it processes abstract concepts into information and language-based semantics. As Douglas Hofstadter explains, "the geographical place is merely the

breeding grounds for an ancient set of genes and memes - complexions, body types, hair colors, traditions, words, proverbs, dances, myths, costumes, recipes, and so forth," (273). Understanding the inner workings of human-information processing leads us to the threshold of modern informational dispersion: the worldwide web. Finding ways to chart the elusive spatial elements in the "space-less" web will help benefit the study of language and information science. A tangible way to begin this study is to track "meme vehicles" that can be well defined without the need of web analytics. Slang language is a common meme vehicle as can be seen by the *Happy Days*-inspired meme, "*jump the shark*" (MacGregor). Also, popular culture memes such as bumper sticker and logo associations can be quantified and spatially measured to create an initial geo-referencing of meme domain outside of hyperspace.

Fig.5: *Tracking the meme vehicle.*

2. Statistical frequencies in language and its relation to web analytics-

Web-page ranking and web-based analytics use a detailed frequency measure of language to justify popularity and power of sites on the web. This in turn also serves as a measurement of information content in hyperspace. Analyzing word frequency phenomena from the web can help websites adjust their content load and better target potential viewers. Also, this frequency information can be extremely beneficial for understanding human language trends across space. The statistical frequencies of a certain language in one place may vary from the frequencies recorded somewhere else. A primary method of analysis for this study will be to relate word frequencies to the naturally-occurring fractal proportions seen in Zipf's Law. Zipf's law in association with language asserts that, "in some corpus of natural language utterances, the frequency of any word is inversely proportional to its ran in the frequency table," (Kretzschmar, 83). This self-organizing complex system can be used to help understand language dynamics on human and digital proportions.

3. Language diffusion in the Caribbean archipelago-

As discovered from this research study, the Caribbean region is full of language and celebrates a bountiful linguistic diversity. Its diversity in tongue ranges from the linguistic influence of the European language groups of its colonial past to the distinct forms of *Regional English* that have sprung forth since then. Understanding the spread of language in the Caribbean will require a detailed analysis of each island's lexicon and will also require an advanced system of spatial measure to graphically portray the movement of language across the archipelago. Initial sources for the data acquisition will come from dictionaries that specialize in the Caribbean region, such as *Popular Phrases in Grenada Dialect* by C.W. Francis, *The Dictionary of Jamaican English* by Frederic Cassidy, and *The Dictionary of Bahamian English* by John Holm and Alison Watt Shilling (Cowie, 354-358).

References

Cowie, A.P., editor. *The Oxford History of English Lexicography: Volume I - General Purpose Dictionaries*. Oxford, Clarendon Press, 2009.

Curzan, Anne. *The Secret Life of Words*. Narrated by Anne Curzan, The Great Courses, 2012. Audiobook.

Gleick, James. *The Information: A History, A Theory, A Flood*. New York, Random House, 2011.

Kretzschmar Jr., William A. *Language and Complex Systems*. Cambridge, Cambridge University Press, 2015.

Hofstadter, Douglas. *I Am A Strange Loop*. New York, Basic Books, 2007.

Lerer, Seth. *The History of English*. Narrated by Seth Lerer, The Teaching Company, 1998. Audiobook.

Oxford Dictionaries. How are Dictionaries are Created. 2018. www.oxforddictionaries.com. Accessed 1 May, 2018.

Oxford English Dictionary. More about the *OED*. https://public-oed-com. Accessed 12 April, 2018.

MacGregor, Jeff. "Why 'Happy Days' – and the Fonz – Never Truly 'Jumped the Shark'." *Smithsonian*, September 2017, Accessed Online 1 April, 2018.

Winchester, Simon. *The Meaning of Everything*. Oxford, Oxford University Press, 2003.