

# **Web Discourses in Nigeria's Democracy: How New Digital Humanities Methodologies Can Be Used to Follow National Language Practices<sup>1</sup>**

*Josef Schmied*

## **1. Introduction**

This contribution follows the Nigerian outsider perspective developed in Schmied (2015) to investigate the extent to which much discourse on the internet can be seen to reflect national language practices and nation-specific usages of current democratic concepts such as *democracy* and *governance*. It tries to illustrate the use of publicly accessible data and tools, so that these methodologies can be transferred to comparative studies inside and outside of Nigeria. Such methodologies are today often summarised under the new buzz word “digital humanities”, although corpus- and sociolinguists have used them for the past fifty years to bring their research more in line with social sciences, rather than traditional humanities. This may have been a strategy to attract more funding or to strengthen the reputation of linguistics as a discipline in a university context, but it may also be an opportunity to expand linguistics' research in terms of new data, new methodologies, and new answers to old questions. New approaches always have to be tested in order to ascertain how much they contribute to current research questions and this is what this contribution tries to do.

Of course, the integration of digital or electronic discourse (also called digitally-/computer-/electronically-mediated communication) into linguistic research is not really new and the development can be followed from Herring (ed. 1996) to Herring/Stein/Virtanen (eds. 2013) and Squires (ed. 2016). The International Conference on *Computer-Mediated Communication and Interaction*, which will be held in Singapore in January 2018, has been running for 20 years. The corpus-linguistic approach followed here is related to other approaches that have proposed new models for new types of data from the World Wide Web. Mair (2013), for instance, has pointed out the possible linguistic effects of the 'super-diversity' characteristic of these virtual communities of practice, and many linguistic studies focus on how features generally associated with spoken discourse are frequently integrated into written electronic discourse for various communicative functions (e.g. Deuber/Sand 2013 on Singapore). Even Digital English in Nigeria has been analysed, but mostly with a focus on Pidgin and other non-standard features (e.g. Heyd 2016). Of course, digital data collection has to be combined with critical analysis (cf. O'Halloran 2017) and this requires considerable involved information as well as considerable detached distance. Finally, it must be made clear that internet practices are no longer seen as a new variety of English, as discussed by Collot (1996),

but as a bundle of different electronic discourse communities that may not necessarily come from one country but are related by common backgrounds and interests.

In this research context, digital methodologies are both more robust and more sensitive than other methodologies: more robust in the sense that they can capture big data, and make use of statistical procedures where the limits of analysis and interpretation are more clearly defined than in hermeneutic methodologies, and more sensitive in the sense that they may uncover “more than meets the eye”, i.e. “invisible” patterns that can only be detected automatically in a large body of data (Koteyko 2010, Schmied 2016). Today, language data are no longer the big problem in a country like Nigeria, which is among the “most digital” nations in Africa and even more in its world-wide diaspora. Of course, access to secondary literature may still be limited, but even there the availability of scholarly information on the web makes “digital” methodologies accessible by digital means. This is why the article prefers to refer readers, especially Nigerian readers, to easily accessible literature that is freely available on the web or through web platforms such as ResearchGate or Academia.edu.

## 2. The Corpus of Global Web-Based English for international comparisons

Over the last few years several on-line corpus tools have been made available at BYU (Brigham Young University in Utah). For our investigations in and comparisons of New Englishes, the Global Web-Based English (henceforth GloWbe) corpus is the most useful, since it includes data from the twenty most important English-speaking countries, covering not only the traditional varieties of “native” Englishes, but also New Englishes, such as varieties in Nigeria, Ghana, Kenya and Tanzania English. (cf. Davis/Fuchs 2015). The self-description of this BYU corpus is as follows:

The corpus of **Global Web-based English** (GloWbE; pronounced "globe") is unique in the way that it allows you to carry out comparisons between different [varieties of English](#). You might also look at the [NOW corpus](#), which supplements GloWbe and which adds about 4-5 million words of data *every day*. GloWbE is also related to many other [corpora of English](#) that we have created, which offer unparalleled insight into [variation in English](#).

GloWbE contains about [1.9 billion words](#) of text from twenty different countries. This makes it about 100 times as large as other corpora such as the [International Corpus of English](#), and it allows for [many types of searches](#) that would not be possible otherwise. In addition to this online interface, you can also download [full-text data](#) from the corpus.

Click on any of the links in the search form to the left for context-sensitive help. You might pay special attention to the [comparisons between countries](#) and the (new) [virtual corpora](#), which allow you to create personalised collections of texts related to a particular area of interest.

**In addition to this online interface, you can also [download full-text data](#) for offline use**

<http://corpus.byu.edu/glowbe/> (last accessed 16/06/07)

When we use GloWbE to compare the absolute and relative use of *governance*, we can clearly see that this is an African topic or issue: Nigerian, Ghanaian and Kenyan websites use it most prominently (followed by Sri Lankan and Tanzanian websites). When we look at the collocates of *governance* in the twenty countries included (see Figure 1 below), the huge difference between native and African usages is visualised. Here we can only take frequent collocates with the highest mutual information score (which calculates a measurement for the link between the two lexemes) and see that predictably *democratic/democracy*, *accountability/accountable* and *transparency/transparent* can be found most frequently in Nigeria and elsewhere.

CORPUS OF GLOBAL WEB-BASED ENGLISH (GLOWBE)		SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]																				COMPARE						
SEARCH	FREQUENCY	CONTEXT	HELP	CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM	ALL	%	MI
1	GOOD	8880	237	231	601	156	276	134	441	939	345	459	172	385	511	70	311	1054	828	665	837	228	2057391	0.43	4.14			
2	CORPORATE	5000	298	337	839	277	322	105	426	152	85	150	235	233	140	366	259	154	225	132	147	118	127826	3.91	7.32			
3	GLOBAL	1872	231	140	412	58	125	118	89	29	39	44	99	23	13	90	78	42	31	128	65	18	335979	0.56	4.51			
4	SYSTEM	1701	140	100	188	96	92	40	144	180	95	78	25	39	28	24	52	65	126	100	50	39	786564	0.22	3.14			
5	DEMOCRATIC	1584	135	53	93	59	41	26	55	174	36	68	7	29	21	5	63	146	320	118	67	68	120521	1.31	5.74			
6	MANAGEMENT	1365	72	126	187	58	87	146	49	41	54	28	32	32	19	40	75	49	68	96	86	20	346990	0.39	4.00			
7	ISSUES	1331	82	72	163	50	80	55	78	76	42	42	32	32	16	29	58	95	90	118	88	33	432031	0.31	3.65			
8	DEVELOPMENT	1242	62	81	112	34	36	35	85	66	33	68	33	22	27	15	53	107	134	121	88	30	584597	0.21	3.11			
9	ECONOMIC	1232	67	40	180	122	62	20	42	50	32	47	38	21	38	28	109	49	92	85	75	35	412021	0.30	3.61			
10	STRUCTURE	1029	80	162	133	49	57	51	59	65	16	31	19	16	9	41	28	45	37	64	36	31	134294	0.77	4.96			
11	DEMOCRACY	967	23	20	58	17	17	14	33	113	23	63	9	15	5	5	59	98	149	119	96	31	115428	0.84	5.09			
12	POOR	837	48	16	93	21	28	16	71	46	71	55	9	24	19	14	45	75	28	88	49	21	312974	0.27	3.45			
13	STRUCTURES	830	46	87	107	85	46	34	27	38	16	10	17	10	9	4	39	39	47	101	45	23	54929	1.51	5.94			
14	EFFECTIVE	798	64	54	109	42	81	30	51	30	25	27	14	12	14	24	37	34	52	45	30	23	203596	0.39	4.00			
15	INTERNET	780	158	45	152	21	47	11	36	12	26	33	10	14	3	19	20	13	34	112	9	5	330438	0.24	3.27			
16	TRANSPARENCY	716	28	23	63	11	27	6	56	52	17	36	19	47	24	4	36	79	74	46	47	21	30310	2.36	6.59			
17	ACCOUNTABILITY	710	16	51	74	31	35	9	26	42	10	38	7	27	14	1	48	71	70	50	66	24	32557	2.18	6.47			
18	LEADERSHIP	701	61	36	53	15	44	50	14	14	13	10	15	22	28	9	51	49	64	77	33	43	166960	0.42	4.10			
19	CORRUPTION	654	22	17	50	9	20	8	61	32	77	44	8	28	17	2	27	81	33	66	46	6	90539	0.72	4.88			
20	GOVERNANCE	630	38	30	51	38	43	20	42	52	20	24	8	26	18	10	36	40	34	48	40	12	57801	1.09	5.47			

Figure 1: Collocates of the search term “governance” in GloWbE (last accessed 15/04/17)

When we use GloWbE to compare the absolute and relative use of “*good governance*”, we can clearly see that this is an African (and Sri Lankan?) issue. Of course, such language practice, i.e. public discourse about a phenomenon, in particular in politically hot election times, does not necessarily reflect objective realities. However, it is noteworthy that the collocation *bad governance* is by far the most widely used combination in Nigeria (168 cases out of 688 for all 20 counties in the GloWbE database).

GloWbE also allows us to look at a KWIC (Key-Word-In-Context) list. Figure 2 below clearly shows that *bad governance* is often used in a sequence with other “bad practices”, such as “misappropriation of state funds, embezzlement of public funds and spending money like water” or “including political repression, corruption and ethnic sectarianism” (Figure 2). Although not all *bad governance* refers to Nigeria (see for example, “... in Kinshasa” in Figure 2), many such utterances

have to be seen in the context of heated political discourse, especially in election times. These word lists have to be seen in contrast to those of *good governance*, which include mostly antonyms of the list on *bad governance*, such as “transparency and accountability”, “quality and selfless services” (cf. Opeibi 2015: 225). This demonstrates again that collocates and wider discourse contexts have to be taken into consideration in all interpretations of the data.

Corpus of Global Web-Based English (GloWbE) [i] [📄] [📥] [🔄] [👤] [?]

SEARCH      FREQUENCY      CONTEXT      HELP

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS), NUMBER (ONE SECTION), OR [CONTEXT] (SELECT) [HELP...]

	<input type="checkbox"/> CONTEXT	ALL	US	CA	GB	IE	AU	NZ	IN	LK	PK	BD	SG	MY	PH	HK	ZA	NG	GH	KE	TZ	JM	ALL	%	Mi
1	<input type="checkbox"/> CORRUPTION	125	1	2	9				6	6	31	7		1	2	1		30	13	10	6		90539	0.14	8.88
2	<input type="checkbox"/> BAD	24	1				3			3			1	1	1	1	1	3	5	4			482002	0.00	4.09
3	<input type="checkbox"/> GOVERNANCE	16				1			1	1	1		1	2				2	4	3			57801	0.03	6.57
4	<input type="checkbox"/> ECONOMIC	14	1		2		1		1	2				5					1			1	412021	0.00	3.54
5	<input type="checkbox"/> POVERTY	13		1	1												1	3	3	2	2		116532	0.01	5.26
6	<input type="checkbox"/> POLITICAL	13	2		1				1	1	2							4	1	1			550801	0.00	3.01
7	<input type="checkbox"/> NIGERIA	12																9	1	2			129344	0.01	4.99
8	<input type="checkbox"/> DUE	12							1	1	4			1				1	1	2		1	441080	0.00	3.22
9	<input type="checkbox"/> LEADERSHIP	11		1		1			2					1				1	5				166960	0.01	4.50
10	<input type="checkbox"/> POOR	11		1			1							1	1			2	1	2	2		312974	0.00	3.59
11	<input type="checkbox"/> MISMANAGEMENT	10			1				1	2	2	1							1	1	1		4694	0.21	9.51
12	<input type="checkbox"/> LACK	10	1	1			1				1					1		2		2		1	242889	0.00	3.82
13	<input type="checkbox"/> UNEMPLOYMENT	9		1					2		3			1				2					63225	0.01	5.61
14	<input type="checkbox"/> RECTIFY	8			1		1			6									5	4			5457	0.15	8.97
15	<input type="checkbox"/> INSTABILITY	8	1		1		1					1						2	1	1			11256	0.07	7.93
16	<input type="checkbox"/> POLICIES	8	1		1				1	1	3										1		154058	0.01	4.15
17	<input type="checkbox"/> OPPRESSION	6														1	1	1	4				17816	0.03	6.85
18	<input type="checkbox"/> LEADING	6									2							2			2		203715	0.00	3.33
19	<input type="checkbox"/> LEADERS	6	1									1		1				2	1				225181	0.00	3.19
20	<input type="checkbox"/> NEPOTISM	5	1							1	1							2					2557	0.20	9.39

Figure 2: KWIC of the search term “*bad governance*” in GloWbE (last accessed 15/04/17)

Using GloWbE to discover particularly high (or low) frequencies is not the only way to investigate culture-specific usage. In many cases, even normal frequencies conceal culture-specific usages, although it is necessary to look at the collocates and even wider discourse contexts to see the polysemy in lexical usage. Culture-specific usages have been described in the usual collections of English world-wide and, of course, with the new with the new digital web-based data collection and analysis methodologies we have today, lexemes from lists from one part of the world can be compared and verified nicely with data from other parts of the world, leading to interesting linguistic insight of culture-specific language usage. The frequency of the lexeme *handout*, for example, is not particularly high in the Nigerian section of GloWbE. It is much more prominent in Malaysia and even in the US and Australia. However, it is a good example of culture-specific usage (which has been mentioned in research on English in Kenya, where the relative frequency of *handouts* is also higher than in Nigeria): Whereas in European academic con-

texts this term refers exclusively to papers that summarise or illustrate the important points of an oral, often complex scholarly text, African usages (and even in other national contexts, like in the US) are always political and refer to (undeserved) “rewards for political favours” more than “necessary state support for the poor”. The following excerpts (see Figure 3 below) illustrate this neatly. The highlighted verbs show the political meaning of the lexeme *handout*. The frequent *depend* signals the political usage of the term especially, although the phrase *students depend on handouts* is equally possible.

but some people still think that disabled person **rely** on **handouts** to get food and money.  
 you are not going to do that by just **giving** out **handouts**, it is okay for me to give cash to the elderly,

I WORK for my money, and **receive** NO **handouts** from the gov't.

Bantustan-style " states " which " governors "**depend** on **handouts** from the centre  
 We are not **asking** Government for **handouts**, we are **asking** for skillful manager who will think about Nigeria first.

will keep the people poor so that they will continually **depend** on them for **handouts** and so that they will not be in a position to oppose them.

# home is where people you used to look up to are now **asking** for **handouts** #  
 the voting electorate -- had a " victim " mentality and were **dependent** on government **handouts**.

to **rely** on their own resources, human and material, not to **depend** on **handouts** and stipends.

O., said he does not believe the state should **depend** on **handouts** to **survive**. #

... has also not bought into the practice of **giving** out' **handouts**' to lawmakers in order to buy their support for his major policies and programmes

discourse after Mitt Romney suggested that 47 percent of Americans who **depend** on " government **handouts** " and who pay no taxes would never vote for him.

Figure 3: Key-word-in-context lists for (verbs associated with) *handouts* in GloWbE

Finally, the *handouts* example may also be interpreted as a case of Americanisation and a euphemism, which obviously plays an important role in Nigerian politics, but not only there (Meutem Kamtchueng 2017).

### 3. The Twitter Corpus collection as a tool to investigate current colloquial usage

Today the standard mass communication tools are Facebook and Twitter. Their special linguistic features have been discussed many times (Schmied 2012 and 2015 or Squires ed. 2016) and they have been used by descriptive linguists to investigate current colloquial usage. Since they are available in digital written form

instantly, they are the ideal input in data analyses of today's language usage, and since they include more oral features than more carefully constructed written language, they allow us to record more immediate language change than was possible until recently. Nowadays, social media are often seen as a double-edged political instrument: they can improve political communication and civic engagement, but they can also lead to national upheaval through hate-speech and mass hysteria. Recent election campaigns in the US and in Europe have demonstrated to everyone the impact of social media, so the importance of social media for research purposes is immediately obvious. This may be an advantage for researchers using it for their investigations.

Twitter today is also the standard mass data-collection tool, since frequent tweets can not only be read and analysed "all the time", but the API interface also makes it possible to feed the results directly into a database, which may be opened as an excel file (as in Figure 5 below).

The corpus- and sociolinguistic issue with Twitter data in a third-world country like Nigeria is whether Nigerians have equal access to it, or whether Twitter language is specific to certain technologically-mined minorities. The issue of how representative Twitter users are for questions of language usage or even of language change always has to be borne in mind, but there is no straightforward solution to representative or even stratified data collection using Twitter, as demographic information on tweeters is difficult to obtain. The wide range of data available, however, allows us to focus our data collection on certain special-interest groups and their topics (such as the famous *#BringBackOurGirls that pleads for the return of the Boko Haram victims in NE Nigeria*).

For the purpose of collecting current Nigerian language data a special Twitter tool with a simple interface was constructed at Chemnitz. The tool TwitCollect is a simple example script for corpus collection from Twitter using Tweepy, based on Twython by Tatjana Scheffler (Albrecht/Schmied 2016) that allowed settings for certain locations (e.g. in Nigeria as in a bounding box in Figure 4) and was left running on the same computer over several weeks to record tweets including the location, date/time and texts (as in Figure 5 below).

If we use the location settings (revealed by a minority of tweeters only!), we can collect data from a wide circle around Nigerian cities, although the high mobility of the Nigerian population cannot exclude (international) "guests" in the area. Unfortunately, no other socio-biographical data can be collected (although all sociolinguistic variables in social media may reflect a self-chosen person rather than a real individual regardless), but for language usage this may be less problematic, because language and particularly English in Nigeria is also chosen in multilingual situations to give the user a special standing well beyond the message itself.

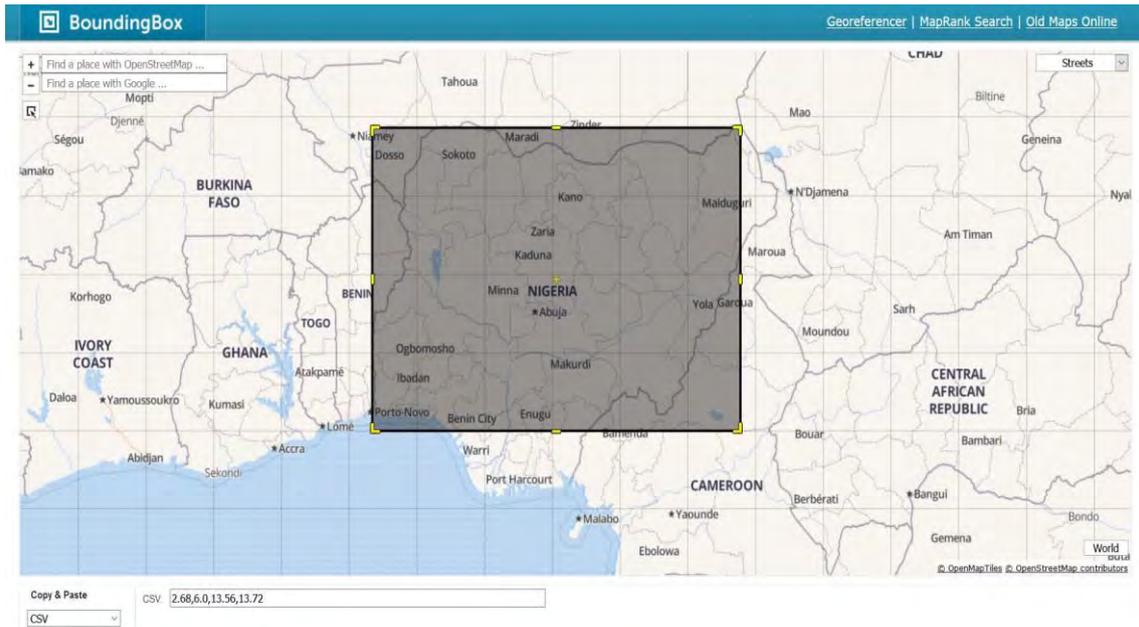


Figure 4: Bounding box for using TwitCollect to compile tweets from (almost) the whole of Nigeria

Thus, we collected Twitter text over several weeks and extracted collocations that indicated positive and negative prosodies of the key word “*governance*” in various short time-spans. Unfortunately, the collected data were not large enough for a detailed quantitative analysis, but in line with the GloWbE above: There were more cases of “good” than of “bad” governance, and the collocates included *effective/efficient* on the positive and *poor/clueless/focusless* on the negative side. In contrast, a search for “*politic\**” (used to cover all nouns and adjective under this lemma, i.e. *politicians, politics* and *political*) revealed slightly more negative (*poor, dirty, violent, ethnic/tribal, and corrupt*) than positive (*good, honest, and decent*) collocates.

Of course, for the interpretation of the data the occurrence of a lexeme does not mean much if we disregard the context. In negated contexts, the opposite of the lexical meaning is discussed, and in some cases, the collocates are the most revealing. In our small Nigeria Twitter corpus, *real, true* and *proper* democracy are more prominent than in other countries. Although they all have a very positive prosody, they are usually seen as an ideal, not a reality, for Nigeria. This can also be deduced from the frequent use of modal auxiliaries in the context: so frequently *politicians will* and *government will* ... Occasionally, a contrast is constructed between *we/us/our* and *them*. Very often, politicians and government are directly addressed as *you* (or, of course, *u*). This is common in this form of “popular” political discourse and line between critical speech and hate speech (or tweets) is often blurry, especially in Nigeria (Opeibi 2011).

Figure 5 below shows the result of a search for governance in the dirty data produced by TwitCollect and shows the expected results: *good governance* occurs first (tweet 1202), as predicted the GloWbE analysis above, and *democratic governance* later, with a gap of 17 seconds in the time stamp (tweet 1230).

1202	14.04.2016 13:24	7,206E+17	If the people of Bayelsa were actually enjoying good governance, I may not be really bothered about the waste in their government!
1203	14.04.2016 13:37	7,2061E+17	Plus qu'une :( [u'type': 'u'Point', u'coordinates': [10.0716273, 5.4599972]]
1204	14.04.2016 13:38	7,2061E+17	#NPN Type of woman @lyanya #freetime @Bluemagic_music
1205	14.04.2016 13:38	7,2061E+17	@MTNFoundation began distributing security patrol vehicles to various states in Nigeria 4 years ago #SecuringLagos
1206	14.04.2016 13:38	7,2061E+17	@Peenqysh @Mss_Yola Diezani pls
1207	14.04.2016 13:38	7,2061E+17	@Tanna Am pretty sure you are uptight as ever Cheers to the numerous endeavors ahead of you
1208	14.04.2016 13:38	7,2061E+17	#OnAir   #Ferari - @yemialadee   #MidDayDrive with @GeorgeNatural1   #NewMusicThursday
1209	14.04.2016 13:38	7,2061E+17	World cup finals goals https://t.co/b00D8tSds
1210	14.04.2016 13:38	7,2061E+17	#NPN: Bend Ova - @LiJlon @Tyga on the #CityJumpOffMix with @Dj_Gunzee
1211	14.04.2016 13:38	7,2061E+17	Lol tbh, men are difficult. If you ask them to fuck off, you're rude. If you air them, you're still rude.
1212	14.04.2016 13:38	7,2061E+17	Realistic goals are for people with realistic work habits. Big goals are for people with big work habits.
1213	14.04.2016 13:38	7,2061E+17	#Rappowerloud# on @RappowerFMBenin with @DeejayHumphrey #np- @TechN9ne ft @2ChainzTruth & @bobat1 - Hood go crazy
1214	14.04.2016 13:38	7,2061E+17	Resource Control or not, if we do not have consequences for mismanagement of public funds by public office holders, we ain't making progress
1215	14.04.2016 13:40	7,2061E+17	BEST NIGERIAN COLLABO IS HERE DOWNLOAD MUSIC ---&gt; https://t.co/XEClqGu4a
1216	14.04.2016 13:40	7,2061E+17	#NP Cambio de Piel - @MarcAnthony on #Thekazbah with @TheKhadeejah #WorldMusicOnSmooth981fm
1217	14.04.2016 13:40	7,2061E+17	1. #Securing [u'type': 'u'Point', u'coordinates': [8.6742, 9.0846]]
1218	14.04.2016 13:40	7,2061E+17	6. #TeeKleff [u'type': 'u'Point', u'coordinates': [8.6742, 9.0846]]
1219	14.04.2016 13:40	7,2061E+17	Is the leader of the Red Brigade heading towards 1 Chancellor Ave or Botswana like they did the last he was graced with a penalty!
1220	14.04.2016 13:40	7,2061E+17	Happy birthc [u'type': 'u'Point', u'coordinates': [7.4949043, 9.07434712]]
1221	14.04.2016 13:40	7,2061E+17	I always knew this would probably be the worst #NASS we would ever have. And from the very head, it's plain.
1222	14.04.2016 13:40	7,2061E+17	#NPN: Run My Race - @burnaboy on the #CityJumpOffMix with @Dj_Gunzee
1223	14.04.2016 13:40	7,2061E+17	Wenger? https://t.co/25O1bmeSOQ
1224	14.04.2016 13:40	7,2061E+17	#NPN- Sugar Baby - @ReekadoBanks on the #RappowerLoungeB wiv Kolade (@yungkolade) x Dandy (@dandy_rocks)
1225	14.04.2016 13:40	7,2061E+17	@SabreNoir @KhidKhorDi Lool that was another giveaway apparently, and she didn't get the hair after.
1226	14.04.2016 13:40	7,2061E+17	Asides the police, agencies such as NAFDAC have also been a beneficiary of the MTNF SSP #SecuringLagos
1227	14.04.2016 13:40	7,2061E+17	The new addition to the current fleet of security patrol vans can go a long way in boosting security in Lagos #SecuringLagos
1228	14.04.2016 13:40	7,2061E+17	#NPN- Gbemisoke - @Joelkennis on the #RappowerLoungeB wiv Kolade (@yungkolade) x Dandy (@dandy_rocks)
1229	14.04.2016 13:41	7,2061E+17	LOL... Still dragging this conversation towards "Our Oil" narrative... You can do better Clarence! @Clarenceonyeks @DIKachii @Oddy4real
1230	14.04.2016 13:41	7,2061E+17	We have been given numerous opportunities to take the position of shining light of how democratic governance should really be

Figure 5: “Dirty” TwitCollect output with search for occurrences of “governance” (first and last line).

The data can also be found in an Antconc analysis, the result of what the “Government should” can be found in Figure 6.



Figure 6: Antconc screenshot of “government should” from our NigeriaTwitterCorpus

#### 4. Further research data options

Even if researchers do not have a special can to create TwitterCorpora, they can use individual tweets, e.g. they can use the archive available in <https://www.tweetarchivist.com/>. Although downloading an archive is a “paid feature” (the Archive also claims to allow “Perform academic research” or “Track a

hashtag campaign”), they can find a number resources, such as the popular <https://nigeriansreportng.blogspot.de/>.

Here, the same results can be found for “governance” as in the Tweets above: The collocations are *good* and *democratic*; a special case is the “role model” from Malaysia below (Figure 7).



Figure 7: Screenshot from Nigerians Report with “governance” (last accessed 19/07/17)

Even a quick (“free trial”) search in [tweetarchivist.com](http://tweetarchivist.com) for *Nigeria* + “*good governance*” retrieves 100 tweets from the last few days that show interesting collocations, such as *responsible and good governance*, *good governance and leadership*, *enhance/struggle for good governance*, *Restructuring Nigeria for National Cohesion and Good Governance*, etc.

It goes without saying that the World Wide Web offers many more opportunities to explore the “digital” connection between Nigeria and *good governance* – interestingly, even Google Scholar (last accessed 18/07/17) lists *corruption* as a “related search item” in relation to this, but it also lists about 498.000 cases of *governance*, 296.000 of with *good* and only 85.00 with *bad*.

## 5. Conclusion

In this contribution, I hope to have shown that using linguistic data from the web and especially from social media, applying digital humanities methodologies can not only be an opportunity for linguists in Nigeria to attract funding, but also a way to find better or more detailed answers to old questions of language usage in the country. With the use of corpus methodology and data-mining digital software, the empirical basis of language analyses is much easier to establish today than it was only a couple of years ago, and the analyses have the potential to be more sophisticated. This allows modern linguistic methodologies to make Linguistics as a discipline more scientific and more respectable, simply by producing relatively easily qualitative and quantitative evidence for crucial word/phrase usage, collocational and idiomatic preferences and even associated thinking. The interpretations of big data, however, has to be conducted as carefully and tentatively as before. O'Halloran (2017) has demonstrated in a whole section (1.1.3) that the “Ethics of digital hospitality” has to be taken into consideration when working with digital data. Digital methodologies can be seen as discovery procedures that allow us to see patterns where constant participatory observation and even deep immersion may not have been enough in the past. Of course, the accessibility of even simple tools discussed in this contribution may not always be easy in Nigeria and more training courses may have to be offered by specialists from Nigeria and beyond. But if we can create and maintain a culture of sharing data, tools and information in Nigeria, we are on the right track. This could lead to a better recognition of linguists' work and their contributions to the documentation of language use, the development of academic learning and related aspects, and research on the influence of the internet (on language usage) in Nigeria, as also demonstrated in this quote:

*“Unlimited Internet access is sacred in Nigeria so sharing your data with a loved one is the ultimate sacrifice”.*

## References

- Albrecht, S., Schmied, J. (2016). *TwitCollect*. English Dept., Chemnitz University of Technology.
- Collot, M., Belmore, N. (1996). Electronic language: A new variety of English. In Herring, S.C. (ed.) *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam: Benjamins, 13-28.
- Davies, M., Fuchs, R. (2015). Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36, 1 – 28.
- Deuber, D., Sand, A. (2013). Computer-mediated communication in Singapore: Spoken language features in weblogs and a discussion forum. In Röder, K, Wischer, I. (eds.) *Anglistentag 2012 Potsdam Proceedings. Trier: WVT Wissenschaftlicher Verlag Trier*, 385-396.
- Heyd, T. (2016). Global varieties of English gone digital: Orthographic and semantic variation in digital Nigerian Pidgin. In Squires, L. (ed.) *English in Computer-Mediated Communication: Variation, Representation, and Meaning*, Topics in English Linguistics [TiEL], 93. Berlin: de Gruyter, 101-122.