Academic Journal of Modern Philology

Natalia Pietraszewska
Philological School of Higher Education in Wrocław

e-ISSN 2353-3218
ISSN 2299-7164
Vol. 4 (2015)
73–80

On the Complexity of Creole Languages: The Fractal Approach

Abstract

The current paper aims to compare the complexity of texts translated into English-based creole languages and English. The main motivation for the choice of topic was the growing body of evidence that languages and language phenomena, such as texts, may be regarded as complex adaptable systems of signs. These systems may display some fractal properties, such as self-similarity at different scales. In consequence, texts may be analysed in the same manner as other fractal objects. It is possible, for instance, to estimate their fractal dimensions which, to some extent, reflect the degree of their structural complexity. Such an assumption enables one to calculate and compare fractal dimensions of parallel translations of texts to various languages in order to compare their complexity levels. Methods which enable comparisons of complexity of texts in different languages are particularly important with regard to creole languages, since the complexity of contact languages is still the subject of debate.

In the following study, ten parallel translations of passages from the New Testament were mapped onto time series plots based on the length and the frequency rank of words. The values of Hurst exponent as well as fractal dimension were estimated and it was found that the studied time series did not differ significantly between English and English-based creoles with respect to their fractal dimensions. The results lend support to the idea that creole languages are simply new languages which are merely different from their superstrate language rather than being less complex, at least with regard to their lexical patterns.

Keywords: fractal dimension, language complexity, creole languages, time series.

Semiotic signs, which may take the form of words, sounds or images, are arranged into systems which are embedded in other semiotic systems. In a similar manner, each text is an example of a sign-system embedded in language (Sebeok [1994] 2001: 7–8), which, in turn, is embedded in its cultural setting. Furthermore, texts may be considered to be communicative events embedded in a context that may be

either situational, social or cultural (Chruszczewski 2006). This stems from the fact that text production is governed by sets of cultural practices influencing the signifying system, also known as codes, which help to link signs in interpretative frameworks and thus have a crucial role in text production and interpretation (Hall 1980: 131; Chandler 2002: 147–148).

It is important to acknowledge that texts are not only sets of units, such as phrases or clauses, but also sign systems embedded in higher-level systems and which integrate many sub-systems (Hřebíček 1995: 6–7). Cohesion and coherence are among the most significant standards of textuality (de Beaugrande & Dressler [1972] 1981), and they clearly stem from interactions between the symbolic constituents of a text. The importance of such interactions may lead one to the conclusion that the text may be viewed as a "complex system," *i.e.* according to the science of complexity. This is further supported by the idea, as proposed by the "Five Graces Group" of Santa Fe in 2009, that language is a complex adaptive system (Beckner *et al.* 2009). Such complex systems are postulated to have a number of features which distinguish them from disordered or chaotic structures. First, they are composed of a number of agents, the interactions of which play a vital role in shaping the emergent behavior of the system (Rickles *et al.* 2007). Each complex system is subject to feedback, which means that its components are able to adapt and change, first on a local level of organization and then within the global level of the system, which again modifies the agents. Such complex systems have an intrinsic tendency to self-organize, as local interactions commonly lead to global coordination and to the emergence of global properties stemming from local interactions. The response to modifications is, nevertheless, hardly ever predictable: it is not proportional to causal factors, therefore it is nonlinear (Goldstein 1999). It appears that all of the features of complex systems, including agent-based architecture, robustness, self-organization, and emergence of new properties on a global scale due to interactions between agents, may be found in language (Beckner *et al.* 2009).

If language is indeed a "complex system" as understood by complexity science, this sheds new light on the issue of measuring its complexity. It appears that the methods used to quantify some of the properties of physical systems may also be of use in linguistics. Many complex systems display fractal properties such as self-similarity, which is linked with the issue of self-organization of the system and may be calculated using well-established methods (Kale & Butar Butar 2011; Liebovitch & Scheurle 2000). One of the areas in which such an approach to measuring the complexity of language will be useful is estimating and comparing the complexity of creole languages, since this may help to establish whether language contact leads to simplification or complexification.

Analysis of time series is among the most promising methods used to study the complex behavior of language. It is based on the assumption that languages are dynamical, complex adaptive systems of signs, and that their texts are symbolic entities with a seemingly linear structure which allows their mapping by time series. Nevertheless, as languages are postulated to have fractal organization, texts may be regarded as spatial or temporal fractals, and it is possible to estimate their self-similarity and correlations between statistical properties on various structural levels, or at various timescales.

At this point it is necessary to explain how the term "fractal" is interpreted in linguistics. Fractals may be understood as self-similar patterns at every scale which lead to the emergence of rich and complex spatial or temporal structures (Mandelbrot 1989). These objects make it possible to describe a number of apparently irregular forms observed in nature (Mandelbrot 1989). The main properties of natural fractals include self-similarity, scalability and unique statistical properties, which make it impossible to characterize the data by using well-known statistical measures such as "mean" or "variance." Fractal

objects or processes are commonly characterized by their fractal dimension, which reflects the roughness of the self-similar pattern. Fractal dimension calculations have successfully been used in numerous analyses, and this parameter is effective in establishing the measure of the complexity (Sandau & Kurz 1997). Fractal analysis seems to be rather comprehensive, as it may take into account the lengths of units on all linguistic levels as well as both short-range and long-range correlations between the constituents of a text.

One of the approaches is based on the fact that the distribution of units in texts follows the Menzerath–Altmann law, according to which an increase in the size of a linguistic construct, such as a sentence, is coupled with a decrease in the size of its constituents (Altmann 1980). There is a power-law relation between these parameters that may be a signature of the structural self-similarity that is typical of fractal patterns. As language is composed of units that are arranged in a certain manner, it allows for the creation of more and more complex structures: phonemes are arranged into words, words into sentences, sentences into texts; and as such, complex structures can be analyzed in various scales, so that their fractal dimension may be established (Andres 2010; Andres *et al.* 2012). It is also possible to present the text in the form of time series. Thus far, approaches to the construction of time series have involved using the properties of letters, words and sentences as values on the time axis of the time series (Ausloos 2012). This is followed by an analysis of a plot including the search for long-range correlations, which indicates the fractal nature of texts.

Research method and results

Since words seem to be an example of an easy-to-delimit, and thus accessible, structural unit, they were chosen as the basis for devising the time series. In the following study, ten translations of passages from the Book of Matthew in the New Testament were mapped onto time series plots based on the length of words and on the frequency-rank of the words (as described by Kosmidis *et al.* 2006). Three English translations were studied, namely the English Standard Version, the Common English Bible and the Contemporary English Version, as well as seven translations into English-based creole languages: Belize Kriol, Bislama, Hawaiian Creole, Jamaican Patois, Kriol, Nigerian Pidgin and Tok Pisin. The Book of Matthew was chosen due to the fact that it was the only text to be translated into all English-based creole languages that were of interest at the time. The same fragments of translations of the same text were chosen so that the differences could not be attributed to different register, field or tenor.

The next step involved calculating the Hurst exponent (*H*), which reflects the long-term memory or persistence of time series and is closely linked to the fractal dimension (Weron 2002; Kale & Butar Butar 2011). The idea of long memory refers to the presence of statistical dependencies over long time-scales (Riley *et al.* 2012), *i.e.* when time series exhibit statistical self-similarity. Values of *H* larger than 0.5 are believed to indicate long memory of time series, while values below 0.5 suggest the anti-persistence of a time series, meaning that: "[…] successive changes tend to cancel each other out" (Sandau & Kurz 1997).

The calculations were conducted by using the rescaled range or the *R/S* method based on the formula: $H = \frac{\log(R/S)}{\log(T)}$, where *T* is the time or duration of the sample data and *R/S* is the value of the rescaled range calculated for this duration (for more details, see Kale & Butar Butar 2011). It was assumed that the fractal dimension (*D*) representing surface roughness of a time series equals 2 - *H* (Kale & Butar Butar 2011).

Additional data were obtained from the Ethnologue website in order to observe the possible dependence of the fractal dimension of the language on the number of speakers, the time of creole formation as well as the EGIDS rank reflecting the status of the language among its speakers (see Lewis *et al.* (eds.) 2014). The relevant information for the studied creole languages may be found in Table 1.

Table 1. Selected characteristics of analyzed creole languages according to the Ethnologue website

Language	Belizean Creole (BKE)	Bislama (BIS)	Hawaiian creole (HCE)	Jamaican Patois (JCE)	Kriol	Nigerian Pidgin (NPE)	Tok Pisin (TPI)
Estimated number of L1 speakers	70 000	10 000	600 000	2 670 000	4200	unknown	122 000
Total number of speakers	110 000	210 000	1 100 000	3 205 000	14 200	30 000 000	4 122 000
Approximate time of creole formation	1750s	1880s	1900s	1655	1908	1900s	1860s
EGIDS rank	3	3	5	5	3	3	1

The values of the Hurst exponent *H* as well as fractal dimension *D* (calculated for time series constructed for ten translations of the same Biblical text) are presented in Tables 2, 3 and 4. The texts in various languages are ordered according to their fractal dimension *D*. Abbreviations used are: Belizean Kriol English (BKE), Bislama (BIS), The Common English Bible (CEB), Contemporary English Version (CEV), English Standard Version (ESV), Hawai'i Creole English (HCE), Jamaican Creole English (JCE), Nigerian Pidgin English (NPE), and Tok Pisin (TPI). *N* denotes the number of words in the analyzed version, while *H* denotes the estimated value of the Hurst exponent.

Table 2. Hurst exponent and fractal dimension of word length time series

	BKE	**NPE**	**CEB**	**ESV**	**CEV**	**JCE**	**TPI**	**Kriol**	**HCE**	**BIS**
N	2180	2308	1633	1676	1699	2262	2519	2517	2367	2827
H	0.773	0.758	0.657	0.644	0.629	0.624	0.597	0.559	0.533	0.512
D	1.227	1.242	1.343	1.356	1.371	1.376	1.403	1.441	1.467	1.488

Table 3. Hurst exponent and fractal dimension of word frequency time series with tied ranks

	NPE	**BKE**	**HCE**	**CEB**	**ESV**	**CEV**	**Kriol**	**JCE**	**BIS**	**TPI**
N	2308	2180	2367	1633	1676	1699	2517	2262	2827	2519

	NPE	BKE	HCE	CEB	ESV	CEV	Kriol	JCE	BIS	TPI
H	0.765	0.546	0.536	0.528	0.522	0.516	0.507	0.504	0.497	0.484
D	1.235	1.454	1.464	1.472	1.478	1.484	1.493	1.496	1.503	1.516

Table 4. Hurst exponent and fractal dimension of word frequency time series without tied ranks

	BKE	NPE	CEV	Kriol	CEB	TPI	BIS	ESV	JCE	HCE
N	2180	2308	1699	2517	1633	2519	2827	1676	2262	2367
H	0.706	0.694	0.598	0.592	0.573	0.561	0.545	0.534	0.531	0.520
D	1.294	1.306	1.402	1.408	1.427	1.439	1.455	1.466	1.469	1.480

Conclusions

The aim of this study was to explore the fractality as well as possible complexity of both non-creole and creole languages. If the complexity of contact languages differs much from the complexity of the superstrate language, then a significant difference in their fractal dimensions should be observed.

However, the data calculated here indicate similarities in the Hurst exponents as well as fractal dimensions of texts translated into creole languages and Standard English versions. The data concerning the fractal dimensions estimated for the studied time series do not indicate any significant differences between the fractal dimension of the texts in creole languages and in Standard English versions. This means that, although the values of the fractal dimensions may differ greatly between individual languages (for instance, Tok Pisin and Nigerian Pidgin English have equal word frequency in the time series), there is no observable clustering of contact languages with reference to their fractal dimension. Similarly, it appears that the Common English Bible (CEB), the Contemporary English Version (CEV) and the English Standard Version (ESV) are clustered with respect to their fractal dimension only in word-length time series. However, in this case only two creole languages have a significantly lower fractal dimension, whereas texts translated into five creoles seem to have higher roughness levels than the English translations.

Moreover, in the frequency-based time series without tied ranks (Table 4), the differences between assorted English translations were larger than the differences between the English and creole translations; for example, the fractal dimension of Kriol shows that its roughness lies between the Contemporary English Version and the Common English Bible Version. Similarly, the roughness of the Tok Pisin and Bislama texts is higher than for CEB but lower than in the English Standard Version. Furthermore, the differences in the values of *D* between certain creoles and English are very slight. The assumption that creole languages are less complex than the established and stable languages is therefore not supported by these findings.

The results seem to indicate that creole languages are simply new languages which are merely different from their superstrate language rather than being less complex, at least with regard to their lexical patterns. This lends support to the argument of Holm (2000: 5), who observed that "[i]t is only comparatively recently that linguists have realised that Pidgins and Creoles are not wrong versions of other languages but rather new languages... shaped by the same linguistic forces that shaped English and other 'proper' languages." The values of the Hurst exponent estimated for the time series of the analyzed

texts clearly lend support to such a view. The lack of significant differences between statistical self-similarity in English and the creole texts suggests that contact languages cannot be presented as simplified English. One must, however, bear in mind that the time series constructed on the basis of words allows for an examination of only one aspect of linguistic self-similarity and complexity. Therefore, more research on linguistic units at other levels of language organization is necessary in order to draw any firm conclusions regarding the complexity of languages.

Interestingly, no evidence for correlations between the estimated times of formation of the languages, the numbers of speakers, or the EGIDS ranks and their roughness was found. Despite the fact that the creole languages analyzed here are English-based, and that many of them have quite similar origins and had similar linguistic constraints limiting their development, their fractal dimensions are far from equal. This suggests that in the process of language evolution, even slight differences in initial conditions may eventually lead to the development of large differences in the structure of the system. It also leads to the conclusion that perhaps other factors, such as the number of L2 speakers at the time of creolization or the extent of contact between different communities, played a primary role in shaping the languages in question. Unfortunately, only very limited data concerning such factors are available, which is not surprising since both the degree and type of contact with neighboring speech communities seem to be especially difficult to research and quantify.

The existence of differences between languages may be attributed to the different levels of synthetism of each studied language (Popescu *et al.* 2013), *i.e.* in the case of word-length time series. This stems from the fact that in analytic languages, words tend to be short, while in more synthetic languages the word length is more variable, as it is influenced by affixes. As a result, synthetic languages display more irregular oscillations in word length, and thus lower persistence of time series, thus meaning a lower Hurst exponent and higher roughness. The frequency time series can also provide some information pertaining to the structure of language, or at least to a given text. With high lexical diversity in a text there is a lower chance of observing repetitive sequences and, as a result, much lower values of the Hurst exponent are expected. This may, however, reflect the skills of the translator rather than provide insight into the nature of a studied language.

The method applied in this study is commonly used to estimate the fractal dimension of time series, but one has to remember that in some rare cases random processes (which are not fractal) may also appear to display self-similarity (Sandau & Kurz 1997). It has been assumed that texts are linguistic units which have a fractal nature (Andres 2010; Andres *et al.* 2012), but one has to remain cautious when forming any firm conclusions. Moreover, Ausloos (2012) noted that in the case of texts, multifractal models may prove more successful in the description of their complex dynamics. Other methods of multifractal analysis are still being developed, and if perfected they may become a standard tool for estimating roughness in texts. The level of text fractality could also be determined for a greater number of levels with the use of the method devised by Andres *et al.* (2012), which is focused on estimating the relationships between language constructs on various levels of linguistic organization. The self-similarity of language can be observed with regard to the Menzerath–Altmann law. The method based on defining the statistical relationships between the properties of neighboring linguistic levels could indeed allow an in-depth analysis of the surface text and its statistical self-similarity. The main difficulty, and the reason why such a methodology is not easily applicable to the study of creole languages, is that it involves text segmentation into semantic constructs, clauses, words, syllables and phonemes, and such

segmentation is purely subjective; for instance, because there is great variability in the way speakers pronounce words, there may be a problem concerning the delimitation of syllables. Therefore, the error rate in the analysis of creole languages could be too high for the analysis results and comparisons to be significant and meaningful.

In conclusion, it appears that semiotic systems such as texts (or possibly even languages) can be treated as complex adaptive systems of signs exhibiting self-similarity and structural roughness; and therefore they can be studied by using fractal analysis. Some further research is definitely required, for instance, regarding the influence of situational, social and cultural embedding on the fractal dimension of the texts produced. In the study of creole languages, valuable conclusions could be drawn from the correlations between the fractal dimensions of texts (estimated either with time series analyses or on the basis of MAL) and certain socio-economic factors. Many more texts in creole languages and their superstrate languages need to be analyzed using the same method in order to allow any further comparisons as well as to help establish whether the language of translation, or the skills of the translator, have a larger impact on the fractality of a given text. Nevertheless, it appears that fractal analyses may be of some use in linguistics, and the correlations between various factors and the fractal dimensions of symbolic systems are a promising area of study, which may contribute to a better understanding of language complexity and language evolution.

References

Altmann, Gabriel (1980) "Prolegomena to Menzerath's Law." [In:] *Glottometrika* 2; 1–10.

Andres, Jan (2010) "On a Conjecture about the Fractal Structure of Language." [In:] *Journal of Quantitative Linguistics* 17 (2); 101–122.

Andres, Jan, Martina Benešová, Lubomír Kubáček, Jana Vrbková (2012) "Methodological Note on the Fractal Analysis of Texts." [In:] *Journal of Quantitative Linguistics* 19 (1); 1–31.

Ausloos, Marcel (2012) "Generalized Hurst Exponent and Multifractal Function of Original and Translated Texts Mapped into Frequency and Length Time Series." [In:] *Physical Review* E 86 (3); 1–12.

Beaugrande, Robert de, Wolfgang Dressler ([1972] 1981) *Introduction to Text Linguistics*. London, New York: Longman.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, Tom Schoenemann (The "Five Graces Group") (2009) "Language Is a Complex Adaptive System: Position Paper." Wiley-Blackwell.

Chandler, Daniel (2002) *Semiotics: The Basics*. London: Routledge.

Chruszczewski, Piotr P. (2006) "On the Fractal Nature of Linguistic and Cultural Communication; an Outline Proposal." [In:] Piotr P. Chruszczewski, Michał Garcarz, Tomasz P. Górski (eds.) *At the Crossroads of Linguistic Sciences.* Kraków: Tertium; 23–29 (Język a komunikacja. Vol. 10).

Goldstein, Jeffrey (1999) "Emergence as a Construct: History, and Issues." [In:] *Emergence* 1 (1); 49–72.

Hall, Stuart (1980) "Cultural Studies: Two Paradigms." [In:] *Media, Culture and Society* 2 (1); 57–72.

Holm, John (2000) *An Introduction to Pidgins and Creoles*. Cambridge: Cambridge University Press.

Hřebíček, Luděk (1995) *Text Levels. Language Constructs, Constituents and the Menzerath-Altmann Law.* Trier: Wissenschaftlicher Verlag Trier.

Kale, Malhar, Ferry Butar Butar (2011) "Fractal Analysis of Time Series and Distribution Properties of Hurst Exponent." [In:] *Journal of Mathematical Sciences & Mathematics Education* 5 (1); 8–19.

Kosmidis, Kosmas, Alkiviadis Kalampokis, Panos Argyrakis (2006) "Language Time Series Analysis." [In:] *Physica A: Statistical Mechanics and Its Applications* 370 (2); 808–816.

Lewis, Paul, Gary F. Simons, Charles D. Fennig (eds.) (2014) *Ethnologue: Languages of the World*. 17th ed. Dallas: SIL International. (Online version: http://www.ethnologue.com).

Liebovitch, Larry S., Daniela Scheurle (2000) "Two Lessons from Fractals and Chaos." [In:] *Complexity* 5 (4); 34–43.

Mandelbrot, Benoit (1989) "Fractal Geometry: What Is It, and What Does It Do?" [In:] *Proceedings of the Royal Society* A 423 (1864); 3–16.

Popescu, Ioan-Iovitz, Peter Zörnig, Peter Grzybek, Sven Naumann, Gabriel Altmann (2013) "Some Statistics for Sequential Text Properties." [In:] *Glottometrics* 26; 50–95.

Rickles, Dean, Penelope Hawe, Alan Shiell (2007) "A Simple Guide to Chaos and Complexity." [In:] *Journal of Epidemiology and Community Health* 61 (11); 933–937.

Riley, Michael, Scott Bonnette, Nikita Kuznetsov, Sebastian Wallot, Jianbo Gao (2012) "A Tutorial Introduction to Adaptive Fractal Analysis." [In:] *Frontiers in Physiology* 3; 371.

Sandau, Konrad, Haymo Kurz (1997) "Measuring Fractal Dimension and Complexity – an Alternative Approach with an Application." [In:] *Journal of Microscopy* 186 (2); 164–176.

Sebeok, Thomas ([1994] 2001) *Signs: An Introduction to Semiotics*. Toronto, Buffalo: University of Toronto Press.

Weron, Rafał (2002) "Measuring Long-Range Dependence in Electricity Prices." [In:] Hideki Takayasu (ed.) *Empirical Science of Financial Fluctuations*. Tokyo: Springer; 110–119.