Poznań Studies in Contemporary Linguistics 50(2), 2014, pp. 179–195

doi:10.1515/psicl-2014-0013

GLOBAL OPTIMIZATION AND COMPLEXITY TRADE-OFFS

KAIUS SINNEMÄKI
University of Helsinki
kaius.sinnemaki@helsinki.fi

ABSTRACT

It is sometimes assumed in research on language complexity that if languages tended towards some optimal level of global complexity, so that all languages would be roughly equally complex, then local complexity trade-offs should be a general principle in language. Drawing evidence from computer simulations I show that in equally complex systems the proportion of trade-offs (significant negative correlations) is higher than in random systems but far from being a general principle in language. In addition, it may be impossible to determine whether a certain correlation-set comes from random systems or equally complex systems. Based on these results a correlational approach on a handful of typological variables cannot be used to validate, or even falsify, the assumption that all languages are equally complex and, therefore, complexity trade-offs should be kept separate from that assumption. The typological distribution of complexity, drawn from the World Atlas of Language Structures, is further shown to differ from both random systems and equally complex systems.

KEYWORDS: Language complexity; computer simulation; trade-offs; typology.

1. Introduction

It has long been claimed that complexity in one part of a language correlates with simplicity in another (henceforth “the trade-off hypothesis”) so that all languages are roughly at the same level of overall complexity or difficulty (henceforth “the equi-complexity hypothesis”) (Sweet 1899; Hockett 1958; Edwards 1994; Aboh and Smith 2009; among others). These claims have recently been researched in a number of typological studies producing mostly negative results (e.g. Shosted 2006; Nichols 2009; Moran and Blasi, forthcoming). Two types of approaches have been used in these studies, both with their particular problems.

In the first approach, which I call "the summation approach", a number of variables which can be scaled according to some metric of complexity are chosen from a large typological database, such as the *World Atlas of Language Structures* (henceforth WALS; Haspelmath et al. 2005; see Parkvall 2008) or *Autotyp* (Nichols 2009). The complexity values for each variable (possibly proportioned in some way) are summed to derive overall complexity values for the languages or their subdomains. However, while the complexity of typological variables may be more or less plausibly measured, it has been repeatedly argued that complexity values are incommensurable across variables (Miestamo 2008; Deutscher 2009). In other words, it is impossible to weigh the contribution of, say, a two-way tone system or the presence of passive voice, to the overall measure of a language's complexity. This problem (of comparability; see Miestamo 2008) leads the summation approach to a dead end.

The second approach, which I call "the correlational approach", avoids the problem of comparability by measuring the complexity of a set of typological variables, such as consonants, vowels, and syllable complexity, or case, agreement, and rigid order, and conducting correlation tests for each variable-pair in the data instead of summing the complexity values (e.g. Shosted 2006; Maddieson 2006). The basic idea here is that if languages were even roughly equally complex, then complexity trade-offs should be a rather general principle in language. The generality of complexity trade-offs could then be evaluated in a relatively straightforward manner by taking just a few typological variables and studying the presence or absence of significant negative correlation(s) (henceforth "trade-offs") among the variables.

While the correlational approach bypasses the problem of comparability, I argue here that this approach suffers from assuming a too direct relationship between complexity trade-offs and the equi-complexity hypothesis (= global optimization of complexity). For one, the requirement that trade-offs should be a general principle in language is not well-defined. How many correlations among the variables should be negative and significant so that trade-offs could be judged as a general principle in language: 50%, 75% or 95%? Currently there are no criteria for determining this.

What this comes down to is that even if our data provided support for some local complexity trade-offs, we have no principled way of determining when trade-offs were sufficiently general so that languages could be deemed as roughly equally complex. This is bad news for the equi-complexity hypothesis. Limited evidence seems to exist for certain local complexity trade-offs when using typological data (e.g. Sinnemäki, forthcoming), but this evidence is a far cry from showing that trade-offs were a general principle in language – in fact, much more evidence shows the opposite (see Shosted 2006; Maddieson 2006; Nichols 2009; Moran and Blasi, forthcoming; but see Moscoso del Prado Martín, in press, for results using corpus data).

For these reasons, I propose that we should treat complexity trade-offs (the trade-off hypothesis) and the equi-complexity hypothesis separately at the outset. This enables us to evaluate each hypothesis independently of one another. In addition, we should explore complexity trade-offs in different types of data to determine what kind of relationship exists between complexity trade-offs and equal complexity of languages. To come to grips with these questions I simulate and explore different types of data and assess the proportion of significant correlations in each. Three types of data are used: 1. simulated random data, 2. simulated equi-complexity data, and 3. selected data from the WALS. In Section 2, I describe the two computer simulations and their results and in Section 3 I present the results based on the WALS database. Section 4 concludes the paper with brief discussion.

2. Simulation study of the relationship between trade-offs and equi-complexity

To examine the relationship between complexity trade-offs and the equi-complexity hypothesis I conducted two simulation tests. The first test simulated the distribution of complexity in random systems and the second the distribution of complexity in systems with equal overall complexity. In the following, I first describe the rationale of the simulations (Section 2.1) and then present the results (Section 2.2).

2.1. Method and rationale of the simulations

The basic idea of the simulations is to create matrices that simulate the distribution of complexity in different systems and to estimate the degree to which the system's parts correlate positive or negatively in terms of complexity with each other.[1]

Let us first create a data matrix with the wanted number of rows and columns and assign the variables with values chosen from a value-sequence. The

---

[1] For this purpose I wrote a function data.sim in the R programming environment (see Appendix). All computations and graphs were done with R (R Core Development Team 2011).

matrix rows can be thought of as different systems (e.g. languages) and the columns as different variables or parts of the system. The value-sequence determines the lower and upper limits of the "complexity values" assigned to the variables and the distance between each value in the sequence (by default 1; e.g. 1, 2, 3, … 10). The values can be thought of as simulating any type of complexity and any type of linguistic variable (e.g. the number of cases, phoneme inventory size, the presence vs. absence of passive).[2]

Two types of data distributions were simulated. First, in random data (henceforth "unconditioned random data") the probability of each simulated value is the same. In the simulation, this is achieved by sampling from the value-sequence with replacement. If we assume that the data values represent commensurable measurement values, it is possible to sum the complexity values for one system, that is, compute the system's overall complexity given the dataset. According to the central limit theorem, these sums should be normally distributed. This is the case for the row sums of an example matrix in Figure 1, as shown by a standard statistical test (Shapiro-Wilk test, $W = 0.997$, $p = 0.92$).

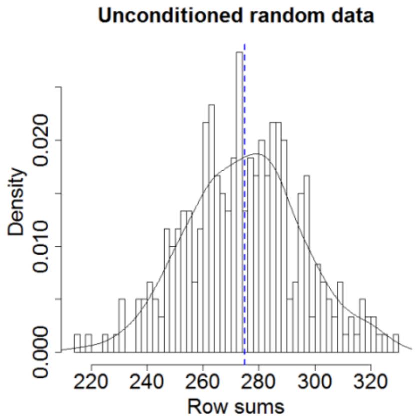

Figure 1. Histogram and densityplot of the row sums for unconditioned random data using 300 rows, 50 columns and number-sequence from 1 to 10. The dashed vertical line indicates the mean of row sums.

[2] In typological databases the ranges of the value-sequences vary across variables, so that some variables may be binomial and encode simply the presence vs. absence of a feature (e.g. passive), while others may encode more detailed distinctions (e.g. the number of tenses). Currently the simulation function does not allow one to create systems with variable value-sequences. However, whether one uses constant or variable value-sequences has no effect on the simulation results.

In the second type of simulated data the systems have equal overall complexity. The data matrices are generated in the same way as for the unconditioned random data: the variable-values are assigned by sampling from the value-sequence with replacement. However, this process is conditioned so that only those systems are chosen in which the row sums match a predetermined constant, assuming that the "complexity values" are commensurable across variables.[3] This type of post-hoc sampling means that the probability of each value is, strictly speaking, not the same.

The question naturally arises how the constant should be determined. One possibility is to use unconditioned random data of the same size as the basis and take, for instance, the mean of row sums for one simulation of such data (275 for the example data in Figure 1). However, this choice is to some extent arbitrary, since choosing a smaller or larger constant has no effect on the results.

The next step in the simulation is to form all possible pairs of variables in the data in order to perform correlation tests for the variable-pairs. The number of variable-pairs f(n) depends on the number of variables (n) and can be calculated with the equation $f(n) = n(n-1)/2$. In the example in Figure 1, there are 50 variables and $50 \times 49/2 = 1225$ variable-pairs, which is also the number of correlation tests to be performed for such data. For the correlation tests I used Kendall's tau correlation coefficient.[4] However, since the data is drawn randomly, the correlation coefficients vary between each simulation of the same size of data. For this reason, I simulated the unconditioned random data and the equicomplexity data 1000 times to get a better understanding of variation in the simulations.

2.2. Results of the simulations

In the simulation tests I used a matrix with 300 rows, 50 variables, and a value-sequence from 1 to 10 (natural numbers). The number of rows is meant to mimic a typical typological sample, which is usually not greater than a few hundred languages.

---

[3] The simulation function allows the specification of deviations from the constant. Allowing deviations makes the data only less extreme and decreases the number of significant negative correlations.

[4] The reason for using Kendall's tau rank correlation instead of linear correlation is that Kendall's tau is a non-parametric test and as such requires fewer assumptions from the data and is overall better suited for typological data (Janssen et al. 2006).

I begin with the unconditioned random data. Conducting Kendall's tau correlation tests for the variable-pairs yields the distribution of coefficients in the left panel of Figure 2 for one simulation. The coefficients are normally distributed, as indicated by the densityplot and shown by a standard statistical test (Shapiro-Wilk test, $W = 0.999$, $p = 0.82$). The mean is also very close to zero ($\bar{x} = 0.00087$). There were 28 negative correlations (2.3%) and 28 positive correlations (2.3%) below the critical significance level (here, $\alpha = 0.05$). In unconditioned random data, the proportion of significant coefficients should be the same as the chosen significance level (here, 5%), as roughly demonstrated in the simulation (here, 4.6%). It is good to keep in mind in quantitative research that in large datasets about 5% of the coefficients may be significant just by chance (given 0.05 as the significance level).

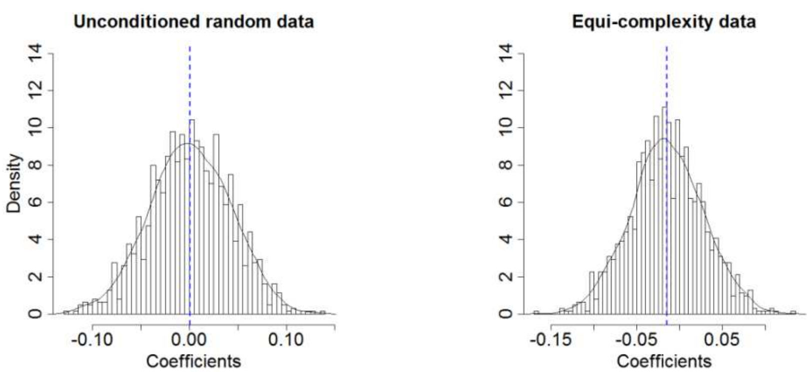

Figure 2. Histogram and densityplot of the Kendall's tau correlation coefficients for all variable-pairs in unconditioned random data (left panel) and equi-complexity data (right panel) for one simulation. The dashed vertical lines indicate the means.

I then simulated the unconditioned random data for 1000 times. The result was an equal average number of significant negative ($n = 30$, 2.48%) and significant positive correlations ($n = 30$, 2.47%). The proportions of significant negative correlations varied between 1.2% and 3.9% and that of positive correlations between 1.1% and 4.2% among the 1000 simulations. There was no marked difference between these distributions as indicated by the box-and-whisker plot in the left panel of Figure 3.

I then conducted the tests for the equi-complexity data. The right panel of Figure 2 shows the distribution of Kendall's tau correlation coefficients for one simulation. The coefficients are normally distributed (Shapiro-Wilk test, $W = 0.9995$, $p = 0.99$), and although the mean is not very far from zero ($\bar{x} = -0.0149$), it is still significantly different from the mean of the unconditioned random data, as shown by the t-test ($t = -9.3$, $df = 2447$, $p < 2.2e^{-16}$). The shape of the distribution is thus similar compared to unconditioned random data, but the coefficients cluster more towards negative correlations. This tendency towards negative correlations is also evident in the number of significant correlations, of which 67 were negative (5.5%) and 11 were positive (0.9%). The rather low proportion of significant negative correlations suggests that equal complexity may be mostly arrived at through small non-significant coefficients.

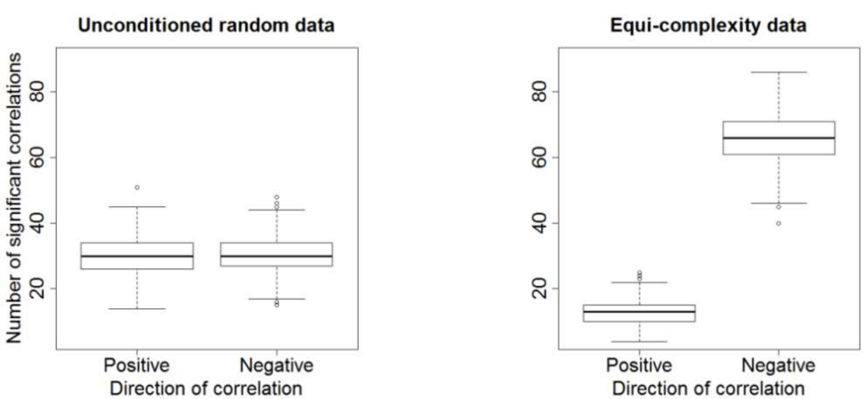

Figure 3. Box-and-whisker plot for the significant positive and negative correlation coefficients in 1000 simulations of unconditioned random data (left panel) and equi-complexity data (right panel).

I then simulated the equi-complexity data for 1000 times. There were on average 66 significant negative correlations (5.4%) and 13 significant positive correlations (1.0%) (the right panel of Figure 3). The proportions of significant negative correlations varied between 3.3% and 7.0% and that of positive correlations between 0.3% and 2.2% among the 1000 simulations. The average number of significant negative correlations was more than five times greater than that of positive correlations, a very strong difference according to the t-test ($t = -223.1$, df = 999, $p < 2.2e^{-16}$). These results show that the correlational profile of the equi-complexity data is markedly different from that of unconditioned random data (see also the box-and-whisker plots in Figure 3).

The results in this section show that there is a bias to negative correlations in equally complex systems. However, the proportion of significant negative correlations (5.4%) is far lower than suggested by the assumption of trade-offs as a general principle. This low proportion of "trade-offs" in the equi-complexity data undermines the proposed close relationship between complexity trade-offs and the equi-complexity hypothesis. Note also that the proportion-range of significant negative correlations in the unconditioned random data (1.2–3.9%) overlaps with that in the equi-complexity data (3.3–7.0%). What these overlapping ranges mean is that if the proportion of trade-offs in the data is, say 3.5%, one cannot tell whether the data come from random systems or equally complex system. All in all, the results of the computer simulations show that the relationship between equal complexity and complexity trade-offs is indirect at best.

3. Typological distribution of complexity

In this section I compare the correlational profile of simulated data to a selection of features from the WALS database. For this purpose I chose 45 features from the WALS that could be scaled according to some measure of complexity (see Table 1 for the features and their complexity scaling).

Most of the selected features were binomial and code the absence vs. presence of a feature (e.g. 10. vowel nasalization, 27. reduplication, and 108. passive constructions; see Table 1). Some of the features concern the number of distinctions made in the particular category, for instance, the size of consonant inventories (feature 1), the number of cases (feature 49), and the number of remoteness distinctions in the past tense (feature 66), while other features may involve the number of coding strategies used for the particular function, such as the presence of head vs. dependent marking in the clause (feature 23) or in the possessive noun phrase (feature 24) or the presence of symmetric and asymmetric strategies in negation (feature 113; see Table 1). Some WALS-chapters have a feature-value "mixed", which cannot be easily interpreted; such features were altogether excluded from the test. In addition, a few selected features included a value "other", but these (rather rare) feature-values were excluded from the data.

Table 1. Selected WALS-features (chapter number and title on the left) and the coding and scaling of their complexity values (Haspelmath et al. 2005).

Feature number and name	Coding and scaling of complexity
1. Consonant inventories	small < moderately small < average < < moderately large < large
2. Vowel qualities	small < average < large
4. Voicing in plosives and fricatives	none < in plosives and/or fricatives
6. Uvular consonants	none < uvular stops and/or continuants
7. Glottalized consonants	none < one < two < three manners of articulation
10. Vowel nasalization	absent < present
11. Front rounded vowels	none < high and/or mid
12. Syllable structure	simple < moderately complex < complex
13. Tone	none < simple tone < complex tone
22. Inflectional synthesis on the verb	0–1 < 2–3 < 4–5 < 6–7 < 8–9 < 10–11 < < 12–13 categories
23. Locus of marking in the clause	none < head or dependent marking < both
24. Locus of marking in the possessive NP	none < head or dependent marking < both
27. Reduplication	none < reduplication
30. Number of genders	0 < 2 < 3 < 4 < 5+
37. Definite articles	no definite article < definite article (definite word ≠ demonstrative, definite word = demonstrative, definite affix)
38. Indefinite articles	no indefinite article < indefinite article (indefinite word ≠ number one, indefinite word = number one, indefinite affix)
39. Inclusive/exclusive distinction	no inclusive/exclusive < only inclusive < inclusive/exclusive
41. Distance contrasts in demonstratives	none < 2-way < 3-way < 4-way < 5+-way
45. Politeness distinctions in pronouns	none < 2 < 3+
47. Intensifiers and reflexive pronouns	identical < differentiated
48. Person marking on adpositions	none < yes (marking on pronouns or both pronouns and nouns)
49. Number of cases	0/borderline < 2 < 3 < 4 < 5 < 6–7 < 8–9 < 10+

55. Numeral classifiers	no < yes (optional or obligatory)
58. Obligatory possessive inflection	absent < exists
59. Possessive classification	none < 2 < 3–5 < more than 5 classes
63. Noun phrase conjunction	‘and’ identical to ‘with’ < ‘and’ different from ‘with’
65. Perfective/imperfective aspect	none < grammatical marking
66. The past tense	no past tense < 0 < 2–3 < 4+ remoteness distinctions
67. The future tense	no inflectional future < inflectional future
68. The perfect	no < yes
70. The morphological imperative	none < number-neutral/singular/plural < < singular and plural
71. The prohibitive	normal imperative and negative < special imperative or negative < special imperative and negative
72. Imperative-hortative systems	neither < minimal and/or maximal system
73. The optative	no inflectional optative < inflectional optative
77. Evidentiality	no evidentials < grammatical evidentials
79. Suppletion according to tense and aspect	no suppletion < suppletion of tense or aspect < suppletion of both tense and aspect
99. Alignment of case marking of pronouns	no case marking < case marking
102. Verbal person marking	none < agent or patient marked < both are marked
107. Passive constructions	absent < present
108. Antipassive constructions	absent < present
109. Applicative constructions	absent < present
111. Nonperiphrastic causative constructions	none < morphological or compound < both
113. Standard negation	symmetric < asymmetric < both
119. Nominal and locational predication	identical < different
120. Zero copula for predicate nominals	possible < impossible

The 45 features chosen formed 990 variable-pairs and for each pair I conducted Kendall’s tau correlation test. There were 69 significant negative correlations (7.0%) and 113 significant positive correlations (11.4%) in the data. Some interesting negative correlations include those between syllable structure (feature 12) and tone (feature 13), between tone (feature 13) and number of cases (feature 49), and between number of genders (feature 30) and numeral classifiers (feature 55) ($p < 0.01$ in all), while some less interesting ones included those between vowel qualities (feature 2) and locus of marking in the clause (feature 23), between front rounded vowels (feature 11) and applicative constructions (feature 109), and between numeral classifiers (feature 55) and future tense (feature 67) ($p < 0.01$ in all). The coefficients were distributed normally (Shapiro-Wilk test, $W = 0.999$, $p = 0.77$), and the mean of the coefficients was slightly positive ($\bar{x} = 0.017$) (see Figure 4).

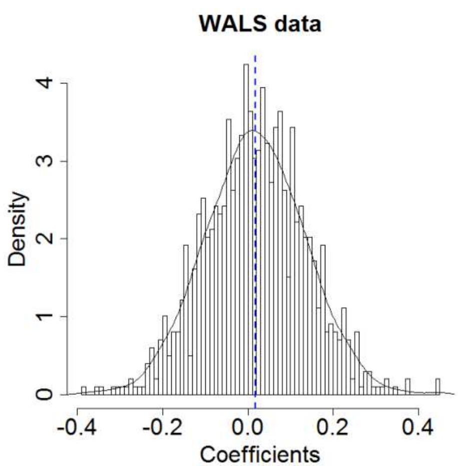

Figure 4. Histogram and densityplot of the Kendall's tau correlation coefficients for all variable-pairs in the selected WALS data. The dashed vertical lines indicate the mean.

The results suggest that the WALS data differs markedly from both the equi-complexity data and the unconditioned random data. As for the significant negative correlations, their proportion in the WALS (7.0%) was even slightly greater than in the simulated equi-complexity data of roughly similar size (5.4%) but almost three times greater than in the unconditioned random data (2.5%). As for the significant positive correlations, however, their proportion in the WALS (11.4%) was more than ten times greater than in the simulated equi-complexity data (1.0%) and more than five times greater than in the unconditioned random data. All in all, the correlational profile of the sampled WALS data is biased towards significant positive correlations but with a notable share of significant negative correlations.

The proportion of all significant correlations in the WALS data (18.4%) is also markedly different from that in random systems of roughly similar size (5%). The interconnectedness between the linguistic variables is almost four times greater than in random data. This result is in line with Bakker (2008), who presents a survey of the WALS to find any significant implicational relationships in the data. He concludes that the proportion of significant implicational relationship is about four times greater than in simulated random data of the same size.

What these results mean is that, on the one hand, the typological distribution of complexity deviates from random data, and on the other hand, that it behaves unlike equi-complexity data. On the contrary, there is a tendency for complexity in one area of grammar to correlate positively with complexity in another area. The WALS data is certainly not fully representative of the distribution of complexity in languages, but still probably one of the most reliable data sets currently available.

4. Discussion

I have presented results on the distribution of complexity in two types of simulated systems and in selected WALS data. The results showed that the correlational profiles in these three types of data are very different from one other. In simulated systems of equal complexity, there are many more significant negative correlations than significant positive correlations but not approaching anything one could describe as a general principle of trade-offs. This suggests that the relationship between complexity trade-offs and equal complexity is quite indirect. In simulated random systems the average proportion of significant positive and negative correlations is roughly equal, as expected for random data. On the contrary, the typological distribution of complexity is biased towards significant positive correlations and, overall, the proportion of significant correlations is much higher than in random systems or in equally complex systems. These results are briefly discussed in this section.

One of the most intriguing questions – at least superficially – is why the relationship between complexity trade-offs and equal complexity is so indirect. Part of the answer lies in the way trade-offs are defined for statistical purposes as significant negative correlations. Although the requirement of statistical significance is a useful heuristic, large systems may arrive at equal complexity by small non-significant correlation coefficients instead of larger and statistically significant ones. In fact the more variables there are in the data, the more likely it becomes that equal complexity is arrived at via small non-significant coefficients, increasingly resembling the distribution of coefficients in random systems. In the simulations in Section 2.2, I used a dataset of 300 rows and 50 variables. A similar dataset using 20 variables (value-sequence from 0 to 1) would yield a greater proportion of significant negative correlations when simulating equally complex systems (14.7% significant negative and 0.5% positive correlations). On the contrary, a dataset using 600 variables (value-sequence from 0 to 1) would yield a much smaller proportion of significant negative correlations (2.8% significant negative and 2.3% positive correlations). Because there are hundreds or thousands of variables in the grammar of each language, using statistical significance as part of the heuristic for complexity trade-offs means that regardless of the set of variables chosen, it may be impossible to determine whether the data comes from a random system or equally complex system. This means that just as the summation approach, the correlational approach to equicomplexity leads to a dead end (cf. Section 1).

As for the typological distribution of complexity, the results raise at least two important issues. First, why was there a bias to positive correlations? At least two issues may be involved here. For one, although I attempted to exclude overlapping features (e.g. number of cases and alignment of case marking of noun phrases), it is possible that some of the sampled features correlate almost by definition. For instance, the presence of voicing for plosives and/or fricatives (feature 4) may affect the size of the consonant inventory (feature 1). A more careful selection of features may have resulted in a lower proportion of positive correlations. On the other hand, a high degree of positive correlations may indicate the functionality requirements of language as a communicative system: the system needs to have a certain level of complexity to function properly. Further research is needed to answer these questions in a more principled manner.

Second, the typological distribution of complexity (see Section 3) suggests that there are local interactions of complexity in languages but no evidence for global optimization of complexity so that all languages would be equally complex. The presence of local interactions and the absence of global optimization are properties usually ascribed to complex adaptive systems (see Beckner et al. 2009 and references there). Complex adaptive systems are characteristically open, have many interacting parts, adapt to the environment and show no signs of global optimization of complexity (among other things). There is a growing body of evidence showing that language is one such system and the results in this paper are in line with this conclusion.

As a conclusion, the current approaches used in typology for researching whether languages are equally complex run into serious methodological problems. Until better methods are devised it is necessary to keep the trade-off hypothesis strictly separate from the equi-complexity hypothesis, whose scientific value remains largely opaque.

REFERENCES

Aboh, E.O. and N. Smith. 2009. "Simplicity, simplification, complexity and complexification: Where have the interfaces gone?" In: Aboh, E.O. and N. Smith (eds.), *Complex processes in new languages*. Amsterdam: John Benjamins. 1–25.

Backer, D. 2008. "LINFER: Inferring implications from the WALS database". *Sprachtypologie und Universalienforschung* (*STUF*) 61(3). 186–198.

Beckner, C., N.C. Ellis, R. Blythe, J. Holland, J. Bybee, J. Ke, M.H. Christiansen, D. Larsen-Freeman, W. Croft and T. Schoenemann. 2009. "Language is a complex adaptive system: Position paper". *Language Learning* 59 (Supplement 1). 1–26.

Deutscher, G. 2009. "'Overall complexity' – a wild goose chase?" In: Sampson, G., D. Gil and P. Trudgill (eds.), *Language complexity as an evolving variable*. Oxford: Oxford University Press. 243–251.

Edwards, J. 1994. *Multilingualism*. London: Penguin.

Haspelmath, M., M.S. Dryer, D. Gil and B. Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

Hockett, C.F. 1958. *A course in modern linguistics*. New York, NY: Macmillan.

Janssen, D., B. Bickel and F. Zúñiga. 2006. "Randomization tests in language typology". *Linguistic Typology* 10(3). 419–440.

Maddieson, I. 2006. "Correlating phonological complexity: Data and validation". *Linguistic Typology* 10(1). 106–123.

Miestamo, M. 2008. "Grammatical complexity in a cross-linguistic perspective". In: Miestamo, M., K. Sinnemäki and F. Karlsson (eds.), *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins. 23–41.

Moran, S. and D. Blasi. Forthcoming. "Crosslinguistic comparison of complexity measures in phonological systems". In: Newmeyer, F.J. and L.B. Preston (eds.), *Measuring grammatical complexity*. Oxford: Oxford University Press.

Moscoso del Prado Martín, F. In press. "The universal 'shape' of human languages: Spectral analysis beyond speech". *PLos ONE*. Available at <http://precedings.nature.com/documents/6097/version/2> (May 26, 2014).

Nichols, J. 2009. "Linguistic complexity: A comprehensive definition and survey". In: Sampson, G., D. Gil and P. Trudgill (eds.), *Language complexity as an evolving variable*. Oxford: Oxford University Press. 110–125.

Parkvall, M. 2008. “The simplicity of creoles in a cross-linguistic perspective”. In: Miestamo, M., K. Sinnemäki and F. Karlsson (eds.), *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins. 265–285.

R Development Core Team. 2011. *R: A language and environment for statistical computing* (R version 2.13.0). Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>

Shosted, R. K. 2006. “Correlating complexity: A typological approach”. *Linguistic Typology* 10(1). 1–40.

Sinnemäki, K. Forthcoming. “Complexity trade-offs: A case study”. In: Newmeyer, F.J. and L.B. Preston (eds.), *Measuring grammatical complexity*. Oxford: Oxford University Press.

Sweet, H. 1899. *The practical study of languages: A guide for teachers and learners*. London: J.M. Dent & Co.

Address correspondence to:
Kaius Sinnemäki
Helsinki Collegium for Advanced Studies
PO Box 4 (Fabianinkatu 24)
University of Helsinki, 00014
Finland
kaius.sinnemaki@helsinki.fi

APPENDIX: R-FUNCTION USED IN THE SIMULATIONS

The simulation was conducted by using an R-function data.sim, which creates the simulated dataset and calculates the correlation coefficients and their p-values for each variable-pair.

```
data.sim <- function(lgs = 100, frs = 10, range = 0:1, equal = FALSE, dev = 0){

    # lgs: the number of systems (or languages) in the simulation
    # frs: the number of features for each system
    # range: the range of values for each feature
    # equal: simulate equal complexity (TRUE) or not (FALSE)
    # dev: how much deviation allowed from equal complexity

    # Create a storage for the data
    mdat = matrix(ncol = frs, nrow = lgs)

    # Is complexity equal
    if (equal == FALSE){
        # Create the simulated dataset by sampling with replacement
        for (i in 1:frs)mdat[,i] <- sample(range, lgs, replace = T)
    }

    else {
        # Set the value for equal complexity
        eq = round(frs*(sum(range)/length(range)))

        # Simulations
        n = 1
        while (n <= nrow(mdat)) {
            # Create a storage for one language
            lg.sim = vector(mode="numeric", length = frs)

            # Create a random dataset for one language
            for (i in 1:frs)lg.sim[i] <- sample(range, 1, replace=T)

            # Choose those where sums equal eq +/- dev
            if(sum(lg.sim) >= (eq - dev) & sum(lg.sim) <= (eq + dev)){
                mdat[n,] <- lg.sim
```

```
            n = n + 1
        }
    }
}

# Create a store for the feature-pairs
var = matrix(ncol = 2,nrow = sum(1:(ncol(mdat) - 1)))

# Perform the correlations for each feature-pair
m = 0
for (i in 1:(ncol(mdat) - 1)){
    for (j in (i + 1) : ncol(mdat)){
        k = m + j - i
        sim.cor <- cor.test(mdat[,i], mdat[,j], method = "k")
        var[k,1] <- sim.cor$estimate
        var[k,2] <- sim.cor$p.value
        }
    m = k
}

# output
return(var)
}
```