Phylogenetic Methods in Historical Linguistics: Greek as a Case Study*

Panayiotis A. Pappas[a] and Arne O. Mooers[b]

[a] Simon Fraser University, Linguistics
papappas@sfu.ca
[b] Simon Fraser University, Biology
amooers@sfu.ca

Abstract

We review and assess the different ways in which research in evolutionary-theory-inspired biology has influenced research in historical linguistics, and then focus on an evolutionary-theory inspired claim for language change made by Pagel et al. (2007). They report that the more Swadesh-list lexemes are used, the less likely they are to change across 87 Indo-European languages, and posit that frequency-of-use of a lexical item is a separate and general mechanism of language change. We test a corollary of this conclusion, namely that current frequency-of-use should predict the amount of change within individual languages through time. We devise a scale of lexical change that recognizes sound change, analogical change and lexical replacement and apply it to cognate pairs on the Swadesh list between Homeric and Modern Greek. Current frequency-of-use only weakly predicts the amount of change within the history of Greek, but amount of change does predict the number of forms across Indo-European. Given that current frequency-of-use and past frequency-of-use may be only weakly correlated for many Swadesh-list lexemes, and given previous research that shows that frequency-of-use can both hinder and facilitate lexical change, we conclude that it is premature to claim that a new mechanism of language change has been discovered. However, we call for more in-depth comparative study of general mechanisms of language change, including further tests of the frequency-of-use hypothesis.

Keywords

lexical change, phylogenetics, frequency effects, evolution

1 Introduction

When the history of a language has been studied as much as the history of Greek has been studied, especially regarding the changes from Ancient Greek,

it can become difficult to see what else can be said about it that is new, or what important insights its study can provide to the field at large. On the other hand, one would expect that the wealth of information that we have amassed for so many periods of the language and the details that we have uncovered about its development should prove very useful to researchers that are trying to understand how language change works, because as Joseph (1987) states "[Greek] offers a 'window' on the nature of language change that few other languages can provide" (see also Joseph 2009). The present paper constitutes such an instance of putting this information to use by testing a proposed mechanism of change against the "Greek Database." Pagel et al. (2007) claim that the frequency-of-use of a lexical item is a separate mechanism of language change, because the more lexemes are used, the less likely they are to change. Their work is part of a new rapprochement between the fields of Historical Linguistics and Evolutionary Biology, in which methodologies that are developed in the latter (primarily in phylogenetics) are applied to data from the former in order to answer controversial questions with what is hoped to be a sufficient amount of scientific vigour. We begin this paper by providing the background for the development of such interdisciplinary studies so that readers can understand the different perspectives that can be applied. In section (2) we review the details of the study by Pagel et al. (2007) and test a corollary hypothesis: the frequency of word use in Standard Modern Greek predicts the amount of change that cognates have undergone from Homeric times until now. There are two significant results to report: First, our results show that frequency of (current) use can only *weakly* predict the amount of change within the history of a single language. Second, and perhaps most interestingly, the amount of change within a single language significantly improves the model proposed by Pagel et al. (2007) for the number of forms across IE.

In our discussion, we evaluate the conclusion of Pagel et al. (2007). We consider frequency of use both as a measure and a process and suggest it is premature to present it as a new mechanism of change. However, we feel the general area of comparative linguistic research exemplified by such studies holds great promise, and we call for more in-depth, statistical comparative study of general mechanisms of language change, including further tests of the frequency-of-use hypothesis.

2 Background

Similarities between the study of change in organisms/species and the study of change in language go back to the beginning of the modern instantiations of both sciences in the middle of the 19th century. As Atkinson et al. (2005) document, the idea of using a tree-like diagram to visualize the relationships between languages on the one hand, or species on the other, arose organically and more or less independently in each field, as a result of researchers' preoccupation with determining whether the existence of the multitude of entities which share features to a lesser or greater degree, can, in either case, be explained by referring to descent through a common lineage.[1] Most notable is the relationship between Darwin's (1859) celebrated *Origin of Species by Means of Natural Selection*, and Schleicher's (1863) *Die Darwinsche Theorie und die Sprachwissenschaft*. Not only do these two works contain the most iconic early trees for their respective fields, one can also discern more than a hint of territorial defensiveness in the linguist's remarks (Schleicher 1869/1983:33) that

> What Darwin now maintains with regard to the variation of the species in the course of time, through which—when it does not reveal itself in all individuals in like manner and to the same extent—one form grows into several distinct other forms by a process of continual repetition, that has been long and generally recognized in its application to the organisms of speech. Such languages as we would call, in the terminology of the botanist or zoologist, the species of a genus, are for us the daughters of one stock-language, whence they proceeded by gradual variation. Where we are sufficiently familiar with any particular family of speech we draw up a genealogical table similar to the one which Darwin attempted for the species of animals and plants.

We note, with some amusement, that almost a century and a half later, and as the two fields appear to be drawing closer together again, similar tensions arise. In response to Atkinson et al. (2008), in which the authors correlate the rate of lexical change with the emergence of new sister languages and conclude that this is evidence that languages evolve in punctuational bursts, Joseph & Mufwene (2008) state:

---

[1] It should be noted here that while most biologists agree that it is possible to construct a tree of life, at least for larger life forms, this question has not been settled for historical linguists and in fact most believe that it will not be possible to construct a single tree of language, at least not with the data that is presently available. It is also worth noting that it may not be possible for biologists to construct a single tree for unicelled organisms due to borrowing. For linguists working on the question of transmission in creole languages, it may be helpful to consider how biologists deal with this problem currently.

> In the Brevia "Languages evolve in punctuational bursts" (1 February, p. 588), Q. D. Atkinson et al. are right that there has yet to be an experimental demonstration of "punctuational bursts" that mark the evolution of language. However, the idea that language evolution proceeds in "bursts" of change alternating with periods of stasis has long been recognized in linguistics.

It seems then that there is some scepticism about the value of such research endeavours. For this reason, we believe it is useful to review the ways in which developments in evolutionary biology have influenced historical linguistics, before delving into the details of our study.

2.1 Four types of language evolution research

In the following subsections we identify four different ways in which evolutionary theory has had an effect on research about linguistic change. We refer to this sort of investigation as language evolution research, but readers should keep in mind that we will not be discussing the question of how language may have evolved as a human trait. We are only talking about understanding how and why languages have been changing in the past ten thousand years, more or less.

2.1.1 Correlating genes and languages

The first type of language evolution research is probably the type that is at the forefront of public imagination given the predominance of genetics and evolutionary biology in our culture today. And that is research that uses biological or genetic information in order to answer questions about membership of a language in a particular family, a concept first expressed in Cavalli-Sforza et al. (1988). A straightforward way of investigating such issues would be to test whether there is a correlation between the genetic affiliation of a set of populations and what we know about the relationship among the languages spoken among these populations. As McMahon & McMahon (2008) note, however, such attempts are unlikely to succeed, especially if they are sampling modern populations that have had significant contact and intermarriage with others in their history. Campbell (2006) further notes that while such studies tend to assume a situation in which there has been negligible admixture in terms of linguistics *and* genetics, linguistic research of language contact situations proves that the opposite (i.e., the existence of *both* linguistic and genetic admixture in the same linguistic community) is quite common. A different problem for such studies, is that they often begin with very tenuous hypotheses about linguistic prehistory, as has been the case with the 100 genetic studies of native American populations (cf. Bolnick et al. 2004), which use the Amerind Hypothesis (Greenberg 1987) as their basic linguistic assumption.

This hypothesis, however, for reasons that Campbell (2006) lays out very clearly, is a very unlikely one.

A different approach is that of Dediu & Ladd (2007) in which they hypothesize that there may be a causal relationship between the population frequency of certain variants (alleles in genetic terminology) of two genes related to microcephaly and the presence of tone in the languages spoken by these populations. However, we need to stress that the authors do not claim to have identified a gene that is the cause behind tonal languages; they observe a correlation and posit a tentative explanation of how the effects of these variants "could involve subtle differences in the organization of the cerebral cortex, with cognitive consequences including linguistic biases in the processing and acquisition of linguistic tone." (*ibid*:10945). Thus, even though such biological approaches to language change that are based on a genes–linguistic features correlation may capture the imagination, they are at the very best in their infancy.

2.1.2 *Evolutionary concepts as metaphors*

The most common way in which evolutionary theory has influenced thinking on language change is by way of metaphor. This is a very understandable transition, because as Atkinson & Gray (2005:514) accurately note, there are more than a few curious parallels between evolution and language change. They provide a list which highlights the parallels between 1) discrete characters in biology and the lexicon, syntax, and phonology in linguistics, 2) homologies and cognates 3) mutation and innovation 4) natural selection and social selection 5) cladogenesis and lineage splits 6) horizontal gene transfer and borrowing 7) plant hybrids and Creoles 8) correlated genotypes/phenotypes and correlated cultural terms 9) geographic clines and dialects/dialect chains 10) fossils and ancient texts 11) extinction and language death. They also note that both fields use the concept of *drift*. Some of the parallels seem apt (e.g. concerning fossils and texts, and extinction or extirpation and language death). Others are more problematic: there are substantial differences between the biology and linguistic correlates, so that the adoption of a concept from the former to the latter must be done after very careful consideration. For example, notice that *drift* in biological evolution is not the same as *drift* in language change. The former is defined (Crow & Kimura 1970) as replication without selection, i.e., change occurring in a population due to random sampling of alleles across generations.

The latter is a much more vague concept first coined by Sapir (1921) that different linguists interpret differently. A succinct and clear history of the term can be found in Keiser (2009:8), which highlights its different permutations.

Suffice it to say that "Sapir himself is imprecise, using 'drift' to describe and account both for the general fact of internally-induced change and for parallel independent developments across related languages." We also note that even internally induced language change is not the same as genetic drift. For example, analogical change is an internally induced change that is not random, for as Kuryłowicz (1947) and Mańczak (1958) have shown, there are some very strong tendencies that influence the outcome of such changes.

Let us also take a closer look at the parallel between cognates and homologies. In biology two or more structures are homologous if they have evolved from a common ancestor structure, for example the wing of a bat and the arm of a primate. Similarly, words in two or more languages are considered cognates if they are derived from the same ancestral word. Crucially, both are hypotheses that need to be supported by evidence. However, there are differences in what this evidence can be. Given the conservative nature of development in biology, surface similarity is an important criterion while function is not (compare a bat's wing and the human arm). In contrast, in order for words to be considered cognates one has to prove that they participate in a complex matrix of correlations between the segments that they comprise (known as correspondences) as well as demonstrate their relationship in terms of meaning. While surface similarity is quite a good indicator of biological homology, it is a very poor indicator of being cognate, as is very well demonstrated in the criticism of Greenberg's work based on multilateral comparison (cf. Campbell 2004:346).

> This approach stops where others begin … inspectional resemblances must be investigated to determine why they are similar … Since multilateral comparison does not do this, its results are controversial and rejected by most mainstream historical linguists.

Nor can we blithely equate horizontal gene transfer to borrowing. There are questions of what, how often, and why. First and perhaps most importantly, borrowing in language highlights the question of the relevant unit of inheritance and the first metaphor regarding discrete units of inheritance: what is the proper analogy to the 'gene'? There is debate within the field of contact linguistics (cf. Winford 2003) as to whether items below the structural level of words can be borrowed. Even if this is possible, it is the case in the minority of instances of borrowing (cf. Nakhleh et al. 2005). The most common occurrence is that segments, affixes and other sublexical items are incorporated into the recipient language only after massive amounts of lexical, phrasal and sentential borrowing: language structure changes only after changes at higher organizational levels. Second, borrowing is quite rare in biological evolution of

multicellular organisms, but ubiquitous in language change. Such a qualitative difference may suggest that languages change in fundamentally different ways from multicellular biological lineages (and is an indication, perhaps, that a formal comparison with unicellular organisms might be illuminating). This difference is likely related in part to mechanism: though subsequent establishment may be a combination of drift and selection, gene transfer itself is random and exceedingly rare. For example, Moran & Jarvik (2010) report that the color variation of the pea aphid (*Acyrthosiphon pisum*) although maintained by selection for camouflage, has its origin in the transfer of a gene from another kingdom (a fungus) to the pea aphid genome. In contrast, the most common mechanism for lexical, phrasal and sentential borrowing is *conscious* need.

As Janda & Joseph (2003) point out, however, problems with drift, homology, and borrowing have not stopped historical linguists from trying to capitalize on these metaphorical connections and treating languages as organisms as in the case of Lass (1997), or as species as in the case of Mufwene (2008; 2007; 2006). Janda & Joseph (2003:12) clearly demonstrate the many pitfalls involved in such an approach, which leads them to conclude that "… it is better for diachronic linguistics if we stand for an embarrassingly long time with our hands stretched out to psychology and sociology than it is for us to embrace the siren song of biological organicism." We discuss the reasons that underlie such failure, after discussing a third category of work in language evolution.

2.1.3 Taking evolutionary concepts literally

The third type of work that brings evolution theory and historical linguistics together is that in which concepts which have been developed in the former field are adopted and/or adapted for use within language change. The work of Ritt (2004) is iconic in this respect, if only because in his title (*Selfish sounds and linguistic evolution*) he pays homage to Dawkins' seminal work *The selfish gene* (Dawkins 1976 [1989, 2006]). Ritt's work is programmatic in nature, paying more attention to the possibility that language can be studied as a complex adaptive system rather than providing answers for what this would mean for a field such as historical linguistics. Ritt is primarily concerned with the nature and existence of linguistic replicators that would then be linguistic instantiations of Dawkins' concept of memes, entities that are capable of transmitting themselves from one brain to another. For Ritt, linguistic replicators are sublexical parts of linguistic structure as he refers to phone-memes, morph-memes, and syntagm-memes, among others.

It is important to emphasize here the literal nature of the instantiation of the evolutionary program into language change that Ritt proposes. This is the

type of approach that Winter-Froemel (2008:217) classifies as the generalized view, in which "both linguistic and biological systems are subcategories of a general type of historically evolved systems." What has made evolutionary biology so successful during the past five decades is the clear identification of the replicator that is involved in the process, namely the gene. For historical linguistics to even begin to approach this success (if it is to adopt this model) it must move beyond the loose parallels between species and languages and try to identify what the linguistic replicators are. As Nettle (2007:484) notes

> When it comes to language, the equivalents of genes are small units of linguistic competence that can survive and replicate, in the sense of persisting and making it into new speakers' competences, or, alternatively, that fail to do so because they are unlearnable, difficult to remember, difficult to hear, or otherwise ineffective at replication. On this view, languages such as English or Hausa are not replicating entities at all but something more like ecosystems[2] or fuzzy-edged pools within which replicators that specialize in particular jobs compete or mutualise.

Nettle, although sympathetic to Ritt's program, is also critical of the research for not following the lead of other work that treats language as a complex adaptive system, and where the evolutionary approach is adopted not only conceptually but also methodologically. A good example of such a complete approach to using concepts from evolutionary biology in the study of language change is work such as that of Croft and his collaborators, where the evolutionary approach is adopted not only conceptually but also methodologically (e.g. Nettle 1999, De Boer 2001, Kirby 1999, Croft 2000; 2008, Blythe & Croft 2009). In Blythe & Croft (2009), the researchers propose two new types of change in addition to neutral evolution (genetic drift) and replicator selection (selection via fitness): this is interactor selection—on the assumption that the interaction between an individual and its environment will cause replication to be differential. They further subdivide this type into neutral interactor selection where the important parameter is the frequency of interactions and weighted interactor selection, in which interactions with certain individuals are more influential than others even if they are not very frequent. This in essence is following Labov's leader or innovator model (Labov 2001). The authors then employ a usage-based mathematical model, called the utterance selection model and look at the formation of New Zealand English from the perspective of each of the different change mechanisms.

---

[2] We must remark that we find even this metaphor somewhat problematic because the analogy of ecosystems to languages and units within languages to replicators bypasses both the individual phenotype and the gene pool or species.

Their results are that the mechanism without selection (neutral interactor selection) accounts very clearly for some aspects of the change (e.g. the frequency effect—majority variants win out), whereas the results for the stronger selective mechanisms are not as clear-cut.

As the name of the model suggests, Blythe and Croft disagree with Nettle (and Ritt) about the nature of linguistic replicators, because for the former these are not sublexical bits of structure but entire utterances, an observation that raises the following question. Must we figure out exactly what the linguistic replicators are before we can proceed with the implementation of phylogenetic methodology in language change? The literature demonstrates that the practical answer (as in what is being practiced) to this is "no". As McMahon & McMahon (2008) note, work in this area is proceeding with different research programs measuring different things. They believe that if there are to be major advances in this area, first there will have to be consensus on what to measure. This does not mean that we have to measure one thing only, but we do need to know what is the best thing to measure in different circumstances.

The same is true in biology: even though there is consensus that the most informative replicator is DNA, there is little consensus on what a 'gene' is, and there may be historical information in other replicators, e.g. histone proteins and chromatin (Chong & Whitelaw 2004). Furthermore, even when the informative replicator is considered to be a stretch of homologous DNA (introns, regulatory elements or classical protein-coding genes), genealogies are often constructed using much more integrated entities—gross morphology, distances among entire genomes, gene orders (Felsenstein 2004)—as well as smaller entities such as homologous positions sampled across stretches of non recombining DNA, such as mitochondria (Cummings et al. 1995). In other words, multiple methods and techniques can be transferred to linguistics even if we are not exactly sure what the replicators are, and even if we have very little certainty as to what memes are or what they do. What is important is that we develop a repertoire of heuristic tools about which we have a clear understanding of what they can help us find or determine.

2.1.4 Phylogenetic techniques applied to linguistic data

The last trend in this type of research is work done by evolutionary biologists who are applying their investigative methods and modelling techniques on linguistic data (cf. Atkinson & Gray 2006ab, Atkinson et al. 2005) or by linguists who are collaborating with computer scientists, mathematicians or physicists and employing similar techniques; for this, see the work of Don Ringe (e.g. Ringe et al. 2002, Nakhleh et al. 2005), the work of McMahon and McMahon (e.g. McMahon & McMahon 1995; 2003; 2008), the work of

Nerbonne (Nerbonne 2009, Nerbonne & Heeringa 2009) or Wichmann et al. (2010).

A good example of this type of work is Gray & Atkinson (2003) on the population expansion from Taiwan into Polynesia. The question they address in this article is the degree of prehistoric contact between the first settlers who moved into the area about 40 thousand years ago and a later wave of colonization that occurred roughly six thousand years ago and expanded throughout the area over the next two millennia. The authors apply phylogenetic methods to linguistic data in order to test a colonization-focused (or "express-train") model, which had been constructed on archaeological evidence, against the evidence from the languages of the area. First they construct a scenario of how the colonization proceeded, with 10 steps: From Taiwan to the Philippines (1); from the Philippines to Borneo (2) and Sulawesi (3); from Sulawesi to central Malayo-Polynesia (4) and west New Guinea (5); from west New Guinea to near Oceania (6); from near Oceania to remote Oceania (7); from remote Oceania to central Polynesia (8); and from central Polynesia to New Zealand (9) and Hawaii (10). Then, using cognate sets for 77 languages from this area, they inferred a maximum parsimony phylogenetic tree[3] whose shape is consistent with a 13-step colonization process. A set of 200 random trees for the same 77 languages yields scenarios with a mean of 49 steps: the correspondence between the language tree and the express-train model is very unlikely to be due to chance, such that the language tree offers statistical support for the archaeological model. With all this as background, we now present the details of our case study which fits in with this last type of language evolution research.

3 Greek as a Case Study

We now turn to the discussion of the present case study, which was inspired by the claim of Pagel et al. (2007) that the frequency-of-use of a lexical item can predict the amount of change that the lexical item has undergone through the course of the radiation of a language family (here, Indo-European). We first present the details of this hypothesis, explain why this is a significant claim, and then discuss the way we tried to test this claim in our work.

---

[3] A maximum parsimony tree is simply the tree that posits the fewest number of changes through time.

Using the 200 cognates in the Swadesh list, Pagel et al. (2007) built a tree of 87 Indo-European (IE) languages and then estimated the rate of change of each of the cognates on that tree (in replacements per word per 10,000 years). This estimate is strongly correlated with the number of distinct forms per cognate as given in the database constructed for Dyen et al. (1992). For example, the database entry for "five" is judged as having only one form, and its rate of change is 0.09,[4] whereas the entry for "dirty" has 46 different forms and a rate of change of 9.27. The strong correlation between the two measures can be seen in Figure 1. These numbers of forms were then correlated with corpus
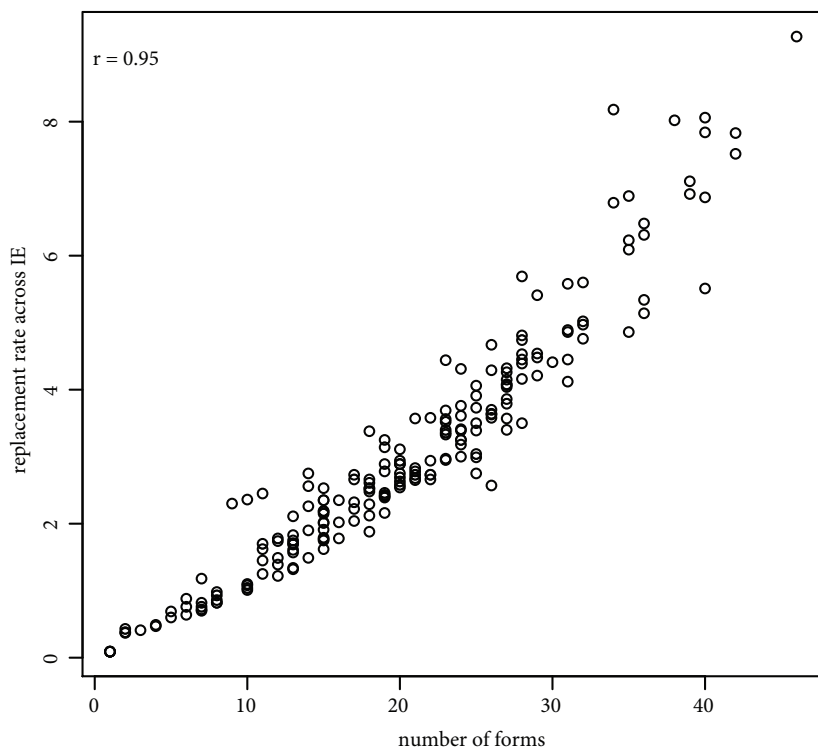
Figure 1. Correlation between number of forms and rate of change across IE.

---

　[4] One reviewer correctly asks "If there is only one form, why isn't the rate of change zero?" The answer to this question is that while it is true that the maximum likelihood estimate of rate for a word that shows no change is 0, Pagel et al. point out in their supplementary methods that such an estimate produces an infinite half-life for the word, which is unreasonable. They therefore use the mean rate from the posterior distribution of rates of their MCMC run: a rate of
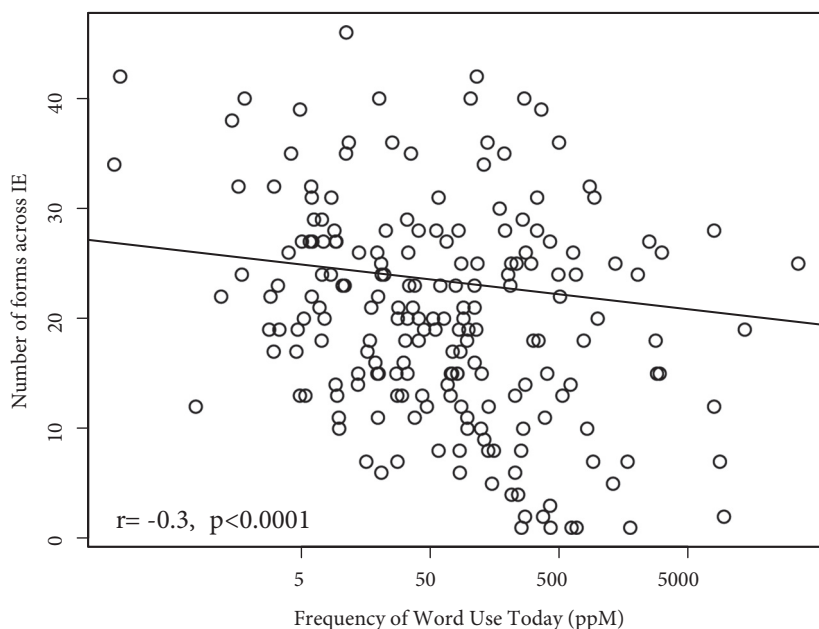
Figure 2. Correlation between frequency-of-use in Greek and number of forms across IE.

data from Greek, Spanish, English and Russian. Rates of change co-varied negatively (r = -0.3) with the mean frequency-of-use (in words per million) across the four modern languages, even when controlling for part of speech. We demonstrate this for Greek in Figure 2, where the number of forms across IE is on the y-axis and frequency-of-use in Modern Greek on the x-axis. The authors draw the conclusion that the more frequently a word is used the less likely it is to change: "the frequency with which specific words are used in everyday language exerts a general and law-like influence on their rates of evolution" (*ibid*: 717).

In effect the authors are proposing that frequency-of-use for lexical items is a mechanism of language change. It is our contention that such a claim is quite remarkable for the field of historical linguistics, because we tend to be very conservative when it comes to proposing mechanisms of language change.

0.09 changes per 10,000 years (the mean of sampled rates for items with only one form on their cognate set) is fully compatible with seeing no change on a tree of length 130,000 years (roughly the total time represented along all the branches of the IE tree).

Currently, we know of only three fairly uncontroversial general mechanisms of language change.

- Sound change: due to the mechanical nature of speech production speakers are not aware of and do not have control over purely phonetic changes which affect, at least at the inception, all relevant contexts in an exception-less fashion. We note that there seems to be conflation in the literature of sound change as a mechanism of mutation and sound change as a mechanism of change. The former, that is the way that phonetic variants may arise in a linguistic variety is well understood, but the latter, that is the question of how or why a particular variant becomes dominant (the *actuation* question according to Weinreich et al. 1968) is still unanswered.
- Analogy: This mechanism is best known for restoring regularity in noun and verb paradigms (levelling) in which a conditioned sound change has introduced irregularity. It is a mechanism of general change, but we note that it does require some specific trigger. It is also posited as the cause of change across paradigms, known as rule levelling or four-part analogy. Despite efforts to understand this mechanism (cf. Kuryłowicz 1947, Mańczak 1958) it is characterized as unpredictable and erratic. For example, in some parts of the U.S. South, the strong form "dove" is gaining ground against the original "dived" (Bernstein 1994) even though the latter represents a ubiquitous pattern; yet this is not the case for all verbs ending in *-ive*; the past tense of "arrive" is still only "arrived". Although we may never fully understand how analogy works, it is clear that it is based on our cognitive tendency to establish transparent associations between form on the one hand and function or meaning on the other. On this assumption we adopt a broader version of this mechanism (*à la* Anttila 1977, Joseph 1998), that includes not only traditional morphological analogy but also phenomena such as phonesthemics, syntactic reanalysis, metaphorical extension, and others.
- Borrowing: the incorporation of linguistic items, mostly lexemes and phrases, from other varieties. The mechanism of both the initial mutation and subsequent change appears to be *need*, although other psycho-social factors such as *prestige* appear to play a role as well.

   The introduction of a new mechanism of language change is a major claim: one need only look at the debate over grammaticalization to realize this. There are those who claim that it is a separate mechanism of language change, and those who maintain that it is an epiphenomenon, the relatively frequent result of other processes (repeated sound changes, semantic and morphological

changes, cf. Campbell 2001 for references for both positions). In a recent review of a book on grammaticalization in Greek (Pappas 2010), the first author wrote that the usefulness of mechanisms of change is in their explanatory power, in that they allow us to fathom how a change may have occurred even if the details are missing; but they also reveal fundamental aspects of the language faculty. In this latter respect, one should be conservative: we do not want a proliferation of fundamental aspects of language.

In the spirit then of testing the claim of Pagel et al. (2007), we constructed the following hypothesis: If the authors are correct that frequency of word-use is a mechanism of diachronic change, then the pattern that they discovered by a cross-linguistic comparison should also be true of change within the development of an individual language. In other words, the frequency of word-use in the modern stage of a language should predict the amount of change of cognates between two different stages in its history. Greek, which is one of the languages used in the Pagel et al. (2007) study but also has a very closely documented history of 3,500 years, presents itself as the ideal candidate for testing this hypothesis.

3.1 Methodology

Our overall plan was to create the Swadesh list (200-word variety) in both Homeric Greek and Modern Greek, devise a method of quantifying the amount of change (AoC) between item pairs, and then see if there is a statistically significant correlation between the amount of change and the frequency-of-use (FoU) numbers provided by the modern corpus in Pagel et al. (2007). There are two changes that we effected: first we use the number of forms across IE (NoF) for each cognate set rather than the inferred rate of change from Pagel et al. (2007): the two measures are strongly correlated (Figure 1) but number of forms is both closer to the raw data and is more normally distributed (Figure 3) and so more amenable for use in parametric statistics.

Second, in order to devise a scale by which to measure change within a particular language, we arranged known types of change in a hierarchy that depicts distance from the original form of the word. The first attempt was a six-point scale as seen in Table 1. The cognate pairs were then checked and corrected by a senior researcher in Greek historical linguistics and, subsequently, two other researchers in the field were asked to apply the scale to the Swadesh list. The inter-scorer correlations are quite robust (all above 0.7), but, on the basis of the scorers' reports that the scale may be too fine-grained, we decided to construct a three point scale, that distinguishes changes on the basis of phonetics (0), morphosyntax (1, 2, and 3) and semantics (4).
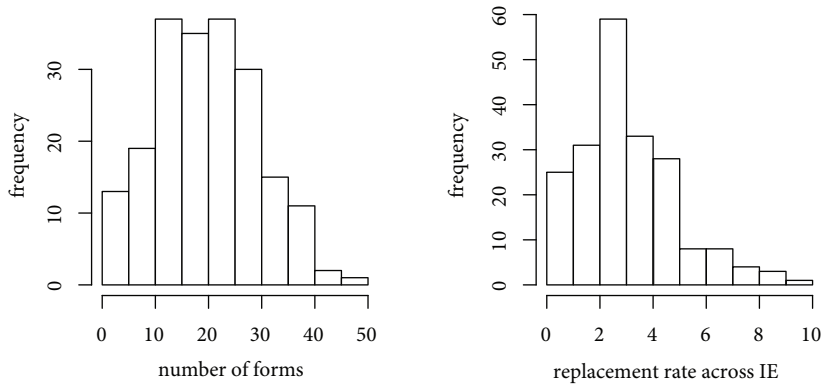
Figure 3. Distribution within NoF (this study) and rate of replacement (Pagel et al. 2007).

Borrowing was eliminated due to the small number of tokens (5). The new scale has a stronger correlation among scorers (>0.8).[5]

Table 1. 6-point scale of change

Value	Type of Change	Example
0	Sound Change	*skʰizo: > skizo* "I tear"
1	Levelling	*megas > meɣalos* "great"
2	Four Part Analogy	*plyno: > pleno* "I wash"
3	Syntactic Reanalysis	*ei* "if" *> an* via *ei an* "if only"
4	Semantic Change	*he:par* "liver" *> sikoti* via *siko,* "fig"
5	Borrowing	*melas > mavros* via Latin *maurus* "black"

---

[5] Some aspects of the scoring process merit further explanation. For example, how should one treat synonyms? Modern Greek has the following synonyms for "squeeze": *stivo, zulo, piezo, sfigo,* while in Homeric Greek we only find *piezo:*. How does one score this change? Do we score on the basis of this lexical item or one of the other three? In such cases we decided to take the most conservative approach, and since *piezo* is still used colloquially in the modern language to score this as sound change. By doing this we are also following the approach Dyen et al. (1992), who assign cognation levels on the basis of the highest degree of cognation between any of the lexical forms that express a concept in the respective languages. On the other hand, the Greek word for "year" is given as both *etos* and *xronos*. The use of the latter reflects semantic change (it originally meant "time") while the first one is a reflex of the original Homeric Greek lexeme. Our judgment is that *etos* is used mainly in higher registers, so in this instance we have semantic change and score it as such. We took into consideration examples such as *poson eton iste* "how

3.2 Results

Having thus assigned a score for 196 of the 201 words, we performed a multi-level logistic regression with amount of change as the y-variable and frequency-of-use as the *x*-variable. This returned a value of $p = 0.066$ ($t = -1.513$) which is only suggestive evidence that the frequency-of-use of a lexical item in Modern Greek can predict how much it has changed from Homeric Greek. In other words we have found weak support for the results of Pagel et al. (2007), but not necessarily for their mechanism. We return to this issue later when we discuss the results.

The question that arises naturally from this initial result is whether our measurement has any predictive power. In order to determine this, we tested the amount of change in Greek against number of forms across Indo-European, which returned a value of $r = 0.215$ ($t = 3.0683$, $p = 0.00246$). This value is lower than the one for frequency-of-use in the Pagel et al. (2007) study ($r = -0.3$) but it is still comparable. The correlation, however, is positive: the more a word has changed in the history of Greek, the more likely it is to have more, not fewer, forms across Indo-European.

The next question we ask is whether amount of change within Greek and frequency-of-use are independent predictors of number of forms across Indo-European. We ran Generalized Linear Models predicting the number of forms across Indo-European with amount of change and log (number of forms) as continuous predictor variables. We compared models using the Akaike information criterion (AIC), which compares the fit of different models to the data while correcting for the number of parameters; lower numbers describe better models, and models that are more than 10 AIC units above the best model are not preferred. As can be seen in Table 2, the model for number of forms across Indo-European predicted by both frequency-of-use *and* amount of change is by far preferred over both single-variable models.[6]

----

old are you?" which is an appropriate construction whereas *poson eton ise* which combines the familiar *ise* with *etos* is infelicitous for most speakers; furthermore *eftixismeno to neo etos* is appropriate for wishing someone "happy new year," because *eftixismeno* is a high register word, whereas *xarumeno to neo etos* is not because *xarumeno* is not. Our rankings are informed by such sociocultural considerations as much as this is possible. It should also be noted that although the interscorer correlation is high (r=0.8 using the 3-point scale), some entries were more difficult than others, and the scale can likely be refined in future work.

[6] The distribution across words of frequency-of-use is very strongly skewed, with most words used relatively rarely, and a few words used very often (see Pagel et al., 2007, Figure 2). Such distributions are expected, and are the norm, in Nature. Log-transformation produces a more normal distribution, and more normal distributions on input data produce more normally

Table 2. Linear models for number of lexical forms across 87 Indo-European languages (n=196 lexical items).

Model	Linear Model R^2	AIC	ΔAIC
NoF ~ AoC	0.04	1442.2	15.9
NoF ~ FoU	0.08	1460.9	34.6
NoF ~ AoC + FoU	0.12	1426.3	0

Schematically, we can depict our results as in Figure 4, the idea here being that the correlation between FoU and NoF is not mediated by AoC, such that the correlation between FoU to NoF is not actually due to two separate correlations with NoF (the bottom path). Both FoU and AoC might be considered independent predictors of NoF.
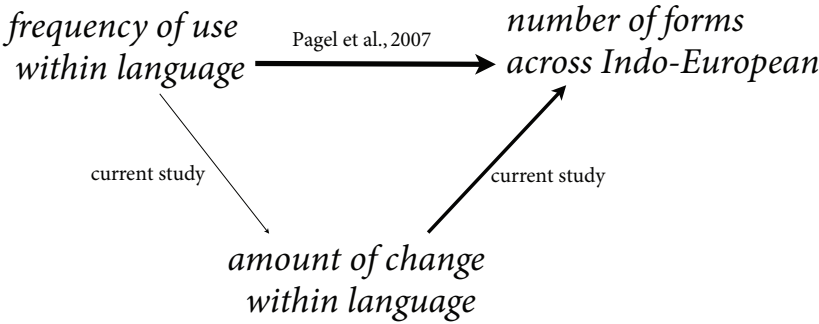
Figure 4. FoU and AoC are independent predictors of number of forms across IE.

In order to test the robustness of some of these patterns, we looked at our strongest correlation (that between amount of change and number of forms in IE, *r=0.215* , n=196) in two additional datasets. The first is a restricted version of the Swadesh list, which contains only 100 lexical items (for which we could score 88) and which is supposed to be even more stable through time than the 201 item list (Campbell 2004). The restricted Swadesh datasets shows a similar correlation (*r=0.221*). We also considered the alternative cognation classification of Nakhleh et al. (2005), which includes 259 lexical items across 24 (instead of 87) languages, with many missing entries. Here we could score 117 items, and the correlation is weaker (*r=0.141*).

---

distributed errors, which in turn are required for the type of model-fitting we do here (indeed, for all parametric testing). The interpretation is that a percent change in FoU (say a 10% increase) predicts a change in the number of forms across IE.

4 Discussion

Had the results of our study confirmed the hypothesis that frequency-of-use also predicts the amount of lexical change within the history of an individual language, this would have provided strong corroboration for the claim by Pagel et al. (2007) that frequency-of-use is a separate mechanism of change. Even though our results do not directly refute this claim, we contend that when taken together with other considerations they do call into question the usefulness of this model. In addition to the fact that we do not find strong support for the hypothesis that frequency-of-use predicts the amount of change in the history of Greek, our results also indicate that these two metrics are independent predictors implying that the co-variation we discovered may not be due to a common underlying mechanism. This raises the concern for us that we may not have a firm understanding of what these predictors may be capturing.

This concern is compounded by the fact that we see no direct reason why the frequency-of-use that we determine based on modern day corpora, and, in the case of Greek, a corpus that relies heavily on newspaper articles, should be affecting the rate of change as this was occurring hundreds or even thousands of years ago. Pagel et al. (2007) find a high degree (close to .9) of pairwise correlation between the frequency-of-use for the four languages. Perhaps they see this as evidence that the 200 words in the Swadesh list would have the same frequency rates through time, or at least similar enough to explain the changes. But that is not a given. For example, the word for "bark" is one of the least frequent items in all four languages, but it is hard to imagine that this was also the case a thousand or even 500 years ago. The same can be said about words such as "belly", "woods", "wing" and many others. Beyond this and as Campbell (2004) documents, the whole notion of a core or universal vocabulary, even when you pare the list down to 100 items, is problematic. The two main reasons are that borrowing does indeed occur to a lesser or greater extent even within this list,[7] and that it is not always easy to establish a one-to-one match between the different forms. We can see an example of this last concern in the word for "flower" in Greek. The form that is listed in the Dyen et al. (1992) database (and consequently the one that Pagel et al. (2007) use) is the Ancient Greek form (*anθos*) while the modern form is *luluδi*, a borrowed form either from Albanian *lʲulʲe* or Latin *lilium*. However, due to Greece's diglossic culture, which has existed in one form or another since Hellenistic times, both

---

[7] As one anonymous reviewer notes, Holman et al. (2008) calculate an average of 8.5% borrowing for 99 of these 100 items.

forms are available, in different registers. In a cursory search of the Hellenic Corpus (Goutsos 2003), which has a smaller number of words but from a more evenly balanced sample of sources, the word *anθos* returned 102 results while the word for *luluδi* returned 506. In the Pagel et al. (2007) study, Greek has the lowest frequency for this entry—17.17 vs. 76 for English, 109.56 for Spanish and 129.13 for Russian—but one wonders if the results would have been the same if both forms were counted; and we imagine that similar questions can be raised for many other forms in each of these languages.

Both of these issues must be taken into consideration along with the fact that frequency effects have been known to linguists for some time, especially in morphological change, but they appear to both hinder and facilitate change. For example we know that frequent categories are less likely to undergo syncretism, more likely to exhibit suppletive allomorphy, (cf. Lieberman et al. 2007) and more likely to have cross-cutting categories than rare categories. As Haspelmath (2002:44-45) states "[a frequency effect] is a general feature of animal (including human) cognition: the more often a cognitive stimulus occurs, the more easily it is remembered."

However, he goes on to point out that this does not mean that word forms that are frequent are necessarily always preserved.

> … regular inflected forms may undergo an idiosyncratic sound change and thus become irregular. For instance, the past-tense form of *have*, *say* and *make* used to be perfectly regular in Earlier English (*haved*, *sayed*, *maked*). Since these are among the most frequently occurring verbs of the English language, they were vulnerable to an idiosyncratic shortening.

Bybee (2006) notes that there are three ways in which frequency can influence the processing and storage of linguistic information. According to her research, while high frequency words are more likely to maintain their morpho-syntactic structure, they are also more likely to undergo *phonetic reduction*, and extreme high-frequency words are more prone to *autonomize*, i.e. to become more opaque in terms of their internal structure, and more likely candidates for grammaticalization. Due to these considerations, we do not share the certainty of Pagel et al. (2007) that a new mechanism of change has been discovered. The linguistic evidence indicates that it will not suffice to make a general statement that "the frequency with which specific words are used in everyday language exerts a general and law-like influence on their rates of evolution." Instead, further research is necessary in order to understand the different effects of frequency on different aspects of lexical organization. Moreover, since this frequency effect most likely stems from a general feature of human cognition (memory), we predict that such patterns of correlation will eventually be subsumed under a future and broader model, perhaps along

the lines of exemplar theory, as Bybee (2006) suggests. At the same time, we find that research such as that of Pagel et al. (2007) and Lieberman et al. (2007) is very valuable because it can help us quantify the various levels of frequency and understand the different aspects of a mechanism of change that is based on general cognitive features.

Our project has revealed that adapting the concepts and methods of evolutionary biology to the study of language change can lead us in interesting directions, but it can also present us with several challenges both in terms of methodology and in terms of interpretation. Our opinion is that the benefits outweigh the costs and that this pursuit is indeed a worthwhile one. But we also agree with more or less every other scholar who has written on this subject that we are still at the infancy stage of this effort.

The biggest issue for linguists will be constructing the types of databases that will allow researchers to test a variety of hypotheses (cf. Dryer & Haspelmath 2011 and Wichmann et al. 2011). These databases will have to include a wide range of information on languages and dialects because we are not exactly sure what we are supposed to be counting yet, and because the more data points we have, the easier certain statistical comparisons will be. The problem here is that there is still disagreement as to how to classify certain structures, especially in syntax. And if there is anything that our case study has shown, it is that this enterprise cannot be carried out on word lists alone, as convenient as this may be. If and once an adequate amount of information is gathered, then statistical analyses should inform us about which linguistic features or combinations of linguistic features can be measured and examined with the most confidence. It may be that different features or different combinations are suitable for different questions, but that will not constitute a problem. As long as we know what we are measuring, how to measure it, and why we are measuring it, we will end up learning something about language change each step of the way. Future work might consider the effect of frequency on the lexical development within English (from Old to Modern) and Spanish (from Latin) to see how those results compare to Greek. Frequency effects on change might also be tested on an expanded list of lexical items that covers several different semantic fields.

References

Anttila, Raimo. 1977. *Analogy*. The Hague: Mouton.
Atkinson, Quentin D. & Russell D. Gray. 2005. Curious parallels and curious connections–phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, 54, 4: 513–526.

Atkinson, Quentin D. & Russell D. Gray. 2006a. Are accurate dates an intractable problem for historical linguistics? In C. Lippo, M. O'Brien, S. Shennan & M. Collard (eds.), *Mapping our Ancestry: Phylogenetic Methods in Anthropology and Prehistory*, 269–296. Chicago: Aldine.

Atkinson, Quentin D. & Russell D. Gray. 2006b. How old is the Indo-European language family? Illumination or more moths to the flame? In J. Clackson, P. Forster & C. Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 91–109. Cambridge: MacDonald Institute.

Atkinson, Quentin D., Andrew M. Meade, Chris Venditti, Simon J. Greenhill & Mark Pagel. 2008. Languages evolve in punctuational bursts. *Science*, 319, 5863: 588.

Atkinson, Quentin D., Geoff Nicholls, David Welch & Russell D. Gray. 2005. From words to dates: Water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society*, 103, 2: 93–219.

Bernstein, Cynthia. 1994. Drug usage among high-school students in Silsbee, Texas: A study of the preterite. *Publication of the American Dialect Society*, 78: 138–143.

Blythe, Richard A. & William A. Croft. 2009. The speech community in evolutionary language dynamics. *Language Learning*, 59, 1: 47–63.

Bolnick, Deborah A. (Weiss), Beth A. (Schultz) Shook, Lyle Campbell & Ives Goddard. 2004. Problematic use of Greenberg's linguistic classification of the Americas in studies of Native American genetic variation. *American Journal of Human Genetics*, 75, 3: 519–523.

Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language*, 82, 4: 711–733.

Campbell, Lyle. 2001. What's wrong with grammaticalization? *Language Sciences*, 23, 2-3: 113–161.

Campbell, Lyle. 2004. *Historical linguistics: an introduction*. Edinburgh: Edinburgh University Press, 2nd edn.

Campbell, Lyle. 2006. Languages and genes in collaboration: some practical matters. Paper presented at L*anguage and Genes: an interdisciplinary conference*, University of California, Santa Barbara, Sept 8-10, 2006. http://www.humis.utah.edu/humis/docs/organization 919 1166141662.pdf (Acessed March 20, 2011).

Cavalli-Sforza, Luigi Luca, Alberto Piazza, Paolo Menozzi & Joanna Mountain. 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistics data. *Proceedings of the National Academy of Sciences USA*, 85: 6002–6006.

Chong, Suyinn & Emma Whitelaw. 2004. Epigenetic germline inheritance. *Current Opinion in Genetics & Development*, 14, 6: 692–696.

Croft, William. 2000. *Explaining language change: an evolutionary approach*. Harlow, England; New York: Longman.

Croft, William. 2008. Evolutionary linguistics. *Annual Review of Anthropology*, 37: 219–234.

Crow, James F. & Motoo Kimura. 1970. *An introduction to population genetics theory*. New York: Harper and Row.

Cummings, Michael P., Sarah P. Otto & John Wakeley. 1995. Sampling properties of dna-sequence data in phylogenetic analysis. *Molecular Biology and Evolution*, 12, 5: 814–822.

Darwin, Charles. 1859. *The origin of species by means of natural selection*. London: John Murray.

Dawkins, Richard. 1976 [1989, 2006]. *The Selfish Gene*. New York, New York; Oxford UK: Oxford University Press.

De Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Dediu, Dan & D. Robert Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, 104, 26: 10944–10949.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2011. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/ (Accessed May 18, 2011).

Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82, 5: 1–132.

Felsenstein, Joseph. 2004. *Inferring Phylogenies*. Sunderland, Mass.: Sinauer Press.

Goutsos, Dionysis. 2003. Hellenic Corpus: design and implementation (Soma Elinikon Kimenon: sxediasmos ke ilopiisi). In Georgia Katsimali, Alexis Kalokerinos, Elena Anagnostopoulou & Ioanna Kappa (eds.), *Proceedings of the 6th International Conference of Greek Linguistics*, University of Crete, September 18-21 2003. Http://www.philology.uoc .gr/conferences/6thICGL/gr.htm (Accessed Feb 15, 2011).

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426: 435–9.

Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.

Haspelmath, Martin. 2002. *Understanding morphology*. New York: Arnold.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller & Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica*, 42, 2: 331–354.

Janda, Richard D. & Brian D. Joseph. 2003. On Language, Change, and Language Change - Or, Of History, Linguistics, and Historical Linguistics. In Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, 3–180. Malden, MA: Blackwell.

Joseph, Brian D. 1987. Greek. In Bernard Comrie (ed.), *The World's Major Languages*, 410–439, London: Croom Helm Publishers [Second edition, 2009, Routledge (Oxford/New York)].

Joseph, Brian D. 1998. Diachronic Morphology. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 351–373, Oxford, UK ; Malden, Mass.: Blackwell.

Joseph, Brian D. 2009. Why Greek is one of The World's Major Languages. *Journal of Greek Linguistics*, 9: 187–194.

Joseph, Brian D. & Salikoko S. Mufwene. 2008. Parsing the evolution of language. *Science*, 320, 5875: 446.

Keiser, Steven Hartman. 2009. When 'speech islands' aren't islands. Parallel independent development, drift, and minimal levels of contact for diffusion. *Diachronica*, 26, 1: 1–35.

Kirby, Simon. 1999. *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Kuryłowicz, Jerzy. 1947. La nature des procès dits analogiques. *Acta Linguistica*, 5: 15–37.

Labov, William. 2001. *Principles of Linguistic change. Volume II: Social Factors*. Oxford: Blackwell.

Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature*, 449, 7163: 713–771.

Mańczak, Witold. 1958. Tendences générales des changements analogiques. *Lingua*, 7: 298–325 and 387–420.

McMahon, April & Robert McMahon. 2003. Finding families: Quantitative methods in language classification. *Transactions of the Philological Society*, 101, 1: 7–55.

McMahon, April & Robert McMahon. 2008. Genetics, Historical Linguistics and Language Variation. *Language and Linguistics Compass*, 2, 2: 264–288.

McMahon, April M. S. & Robert McMahon. 1995. Linguistics, Genetics and Archaeology: Internal and external evidence in the Amerind controversy. *Transactions of the Philological Society*, 93, 2: 125–225.

Moran, Nancy A. & Tyler Jarvik. 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, 328, 5978: 624–627.

Mufwene, Salikoko S. 2006. Language evolution: The population genetics way. *Marges Linguistiques*, 11, May: 243–260.

Mufwene, Salikoko S. 2007. Population movements and contacts in language evolution. *Journal of Language Contact*, 1: 63–92.

Mufwene, Salikoko S. 2008. *Language evolution: contact, competition and change*. London, New York: Continuum,.

Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language*, 81, 2: 382–420.

Nerbonne, John. 2009. Data-driven dialectology. *Language and Linguistics Compass*, 3, 1: 175–198.

Nerbonne, John & Wilbert Heeringa. 2009. Measuring dialect differences. In Jürgen Erich Schmidt & Peter Auer (eds.), *Language and Space: Theories and Methods*, 550–567. Berlin: Mouton De Gruyter.

Nettle, Daniel. 1999. *Linguistic diversity*. Oxford: Oxford University Press.

Nettle, Daniel. 2007. Review of Ritt, Nikolaus. 2004. *Selfish Sounds and Linguistic Evolution: A Darwinian Approach to Language Change*. Cambridge, UK: Cambridge University Press. *Journal of Linguistics*, 43, 2: 482–486.

Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449: 717–721.

Pappas, Panayiotis A. 2010. Review of Markopoulos, Theodore. 2009. *The Future in Greek: From Ancient to Medieval*. Oxford; New York: Oxford University Press. *Linguistics*, 48, 4: 951–954.

Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society*, 100: 59–129.

Ritt, Nikolaus. 2004. *Selfish Sounds and Linguistic Evolution: a Darwinian Approach to Language Change*. Cambridge, UK: Cambridge University Press.

Schleicher, August. 1863. *Die Darwinsche Theorie und die Sprachwissenschaft*. Weimar: Hermann Bohlau.

Schleicher, August. 1869/1983. Darwin Tested by the Science of Language. Alex V. W. Bikkers (trans). In Konrad Koerner (ed.), *Linguistics and evolutionary theory: three essays by August Schleicher, Ernst Haeckel, and Wilhelm Bleek; with an introduction by J. Peter Maher*, 13–72. Amsterdam: John Benjamins.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehman & Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.

Wichmann, Søren, Eric W. Holman, Dik Bakker & Cecil H. Brown. 2010. Evaluating linguistic distance measures. *Physica A*, 389: 3632–3638.

Wichmann, Søren, André Müller, Viveka Velupillai, Cecil H. Brown, Eric W. Holman, Pamela Brown, Sebastian Sauppe, Oleg Belyaev, Matthias Urban, Zarina Molochieva, Annkathrin Wett, Dik Bakker, Johann-Mattis List, Dmitry Egorov, Robert Mailhammer, David Beck & Helen Geyer. 2011. *The ASJP Database (version 13)*. http://email.eva.mpg.de/ wichmann/languages.htm (accessed May 19, 2011).

Winford, Donald. 2003. *An introduction to contact linguistics*. Malden, MA: Blackwell.

Winter-Froemel, Esme. 2008. Towards a comprehensive view of language change: Three recent evolutionary approaches. In Ulrich Detges & Richard Waltereit (eds.), *The Paradox of Grammatical Change: Perspectives from Romance*, 215–250. Amsterdam: John Benjamins.