Stochastic universals and dynamics of cross-linguistic distributions: the case of alignment types

Elena Maslova & Tatiana Nikitina

1. Introduction

This paper has two goals. The first is to describe a novel approach to statistical analysis and interpretation of cross-linguistic typological distributions; more specifically, we describe two methods for detecting systematic differences in probabilities of shifts along parameters of typological variation on the basis of synchronic cross-linguistic data. Statistical evidence about such differences (or lack thereof) gives a straightforward criterion for answering one of the fundamental methodological questions of empirical typology, namely, whether an attested statistical pattern reflects historical accidents or probabilistic ("soft") language universals. Furthermore, the suggested methods provide estimates of typological distributions that would be determined solely by systematic differences in transition probabilities and free of accidental effects. Informally, these methods are based on comparison of cross- and intra-genetic distributions; this idea goes back to Joseph Greenberg (1978; 1995: 146-153). The challenge was only to transform it into specific methods of analysis.[1]

We introduce the methods by describing a case study, an analysis of one of the most well-studied typologies, the typology of alignment systems. This linguistic topic is chosen precisely because it has been extensively studied both typologically and theoretically, so that the linguistic phenomena under discussion are familiar to most linguists and relatively well understood. Moreover, there is a general understanding of how alignment types are distributed among the world's languages (Comrie 1989: 124-126; Comrie 2005; Nichols 1992: 69). More specifically, cross-linguistic studies of alignment types appear to show no significant difference between the frequencies of nominative-accusative and ergative-absolutive alignment patterns in the domain of case marking of full NPs (as opposed to personal pronouns, cross-reference markers on the verb and behavioural syntactic properties, which are predominantly nominative). Insofar as statistical cross-linguistic distributions can be considered linguistically meaningful, this seems to suggest that these alignment patterns provide nearly equally optimal compromises between conflicting constraints involved in local morphological encoding of core participants (Comrie 1978: 330-334; Comrie 1989: 124-125; see also Jäger forthcoming for a more mathematically explicit version of the same idea). In other terms, there are no language universals that would strongly favour one case-marking pattern over the other. The second goal of the present paper is to present evidence against this conclusion; more specifically, we contend that it is highly probable that there exists a stronger universal preference for the nominative-accusative case-marking pattern over the ergative one than implied by their synchronic frequencies, so that the linguistically motivated probability of nominative alignment is at least three times higher than the probability of ergative alignment. To put it in other terms, the expected life time of nominative construction is considerably longer than that of ergative construction (see Hawkins 1983: 256ff; Maslova 2000 on equivalence of these statements).

This hypothesis first emerged on the basis of Johanna Nichols' cross-linguistic database (1992; Johanna Nichols kindly shared with us a more recent, expanded, version of this database), yet it turned out to be not quite sufficient to verify the hypothesis, primarily because it was not designed with this goal in mind. The study reported here is based on a database of 400 languages (see Appendix 1). It was designed in such a way as to (i) contain a random sample from the language population, (ii) represent a sufficient number of distinct genetic stocks, and (iii) a sufficient number of pairs of (relatively) closely related languages. The reasons for these requirements will become clear as we describe our methods of statistical analysis.

The paper is organized as follows. We begin by introducing our version of typology of alignment (Section 2). Section 3 discusses the concept of stochastic (or statistical) universal and its intrinsic relation to the existence of systematic differences between transition probabilities. In a nutshell, it outlines the theoretical foundation for the empirical methods introduced in Section 4, where these methods are applied to analysis of alignment typology. The conclusion summarizes the findings.

2. An overview of alignment typology

The typology of case alignment systems explored in this paper is a fairly traditional one (Comrie 1989: 124-126; Dixon 1994: 6-18) and is based on the widely used concept of three types of core participants, the sole (S) participant of an intransitive clause, and the agent-like (A) and patient-like (P) participant of a transitive clause. We begin with three clear-cut and most broadly cross-linguistically represented types, namely nominative, ergative and neutral:

- In a nominative system, S and A are encoded identically, and this encoding differs from that of P.
- In an ergative system, S is encoded in the same way as P, and this encoding differs from that of A.
- In a neutral system, all core participants are encoded identically.

Most languages prove to fit into one of these three major types in a straightforward fashion. With one exception to be discussed below, other, less consistent, case-marking systems, can be plausibly represented as different "mixtures" of these three types. The most frequent mixed type is the so-called DIFFERENTIAL case marking (Silverstein 1976; Bossong 1991; Aissen 2003), which combines the neutral encoding and one of the two major types of overt marking (for example, the accusative marker can be optional, or its presence can be determined by grammatical context). The second type of mixture subsumes various SPLIT systems, which combine ergative and nominative mechanisms (depending on various properties of the grammatical context, such as properties of noun phrases, tense/aspect, or the semantics of intransitive verb). And, finally, all three case-marking strategies can be combined within a single language.

These considerations straightforwardly define a typological "plane" with two dimensions, "nominative – ergative" and "neutral – marked". Each language can be located on this plane within what can be referred to as ALIGNMENT TRIANGLE, with apices corresponding to three "pure" types (consistently nominative, consistently ergative, and neutral), three domains along the edges corresponding to the mixed types (split, differential nominative, differential ergative), and the "inner" domain for systems combining some sort of nominative/ergative split and the neutral encoding (see Table 1 for a visual representation of this triangle). The explicit inclusion of "mixed" types into the typology resolves most problems associated with assigning a specific language to an alignment type (see (Comrie 2005) for the most recent overview of these issues). Of the classification problems listed by Comrie, only one is relevant for our version of alignment typology, namely, identification of a construction type as "basic" or "marked" in languages with non-canonical voice-like paradigms, so that one construction can be analysed, for instance, as either "passive" or "ergative" (the former solution would assign the language to the nominative type, and the latter, to the split type). In the study reported here, dilemmas of this sort were generally resolved in favour of treating a construction as "basic" whenever the issue was mentioned in the sources as a non-trivial and/or controversial one. There are two considerations behind this approach. First, it seems that a construction (and thus the associated coding mechanism) must play an important role in its language to pose a descriptive challenge of this sort. Secondly, this approach seems to make the alignment triangle more diachronically meaningful, in the sense that type shifts are possible only between neighbouring domains. For instance, a "leap" from consistent nominativity to consistent ergativity in case marking seems to be logically possible only under the assumption that two constructions can change their "basicness" status within the same brief time interval: what used to be "passive" is reanalysed by all speakers simultaneously as "basic ergative", and what used to be "basic active" becomes an "antipassive" at the same time (Harris & Campbell 1995: 243-251). Even if one is willing to reify such descriptive labels, it still seems necessary to assume a period of "double analysis" and/or a period of propagation of the innovative analysis through the language community, which means that "basicness" must be a matter of degree (for at least some periods in the history of language), rather than a discrete binary variable (Timberlake 1977; Kroch 1989; Harris & Campbell 1995: 70, 77-79; Croft 2000: 166-189). It seems likely that this is what would lead to descriptive problems and controversies for some languages.

On the other hand, there are case marking systems which do not seem to fit into our “alignment triangle”: these are rare systems in which S is encoded differently from both A and P. This subsumes so-called “tripartite” and “double oblique” encoding (Payne 1979: 443). Such systems are so rare cross-linguistically that it does not really matter (in the essentially statistical context of this paper) how this problem is resolved. Our solution is based on the following considerations. The alignment triangle should be viewed as a projection of a multidimensional typological space to a plane “anchored” by three well-defined and widely represented “points” (nominative, accusative, neutral). This means that all other domains of this triangle subsume a variety of genuinely different mixtures of coding mechanisms (e.g. so-called “active” systems, systems based on various NP classifications, systems with splits along tense/aspect paradigms, and so on). The choice of this particular projection is justified primarily by cross-linguistic salience of the three major types, which define two clear typological “dimensions” with a straightforward theoretical interpretation. Although both major marking mechanisms are characterized by one coding identity (A=S and P=S) and one coding distinction (P≠S and A≠S respectively), we decided that, if these criteria happen to disagree, we would take the distinction (and not the identity) as the type-defining criterion. As the problematic tripartite and double oblique systems exhibit both nominative-like and ergative-like distinctions, they fall into the “split” domain of this projection of the overall typological space. The major rationale behind this decision is that it keeps the “anchor” points of our triangle well-defined and linguistically homogeneous, and only slightly increases the hidden typological heterogeneity of the other domains.

Finally, it is well-known that the assignment of a language to an alignment type strongly depends on whether one takes into account only lexical NPs or personal pronouns as well (Silverstein 1976; Dixon 1994: 83-96; Nichols 1992: 69-70; Comrie 2005). Although we analysed both classifications, this paper focuses on the typology based on full NPs only, for two reasons. First, a cross-linguistic preference for nominative encoding of personal pronouns is well established in the literature, and our findings simply corroborate this tendency; secondly, the distinction between free personal pronouns and bound verbal cross-reference affixes is often controversial, so our data for pronouns is somewhat less reliable. However, some results for personal pronouns are mentioned in the conclusion.

As outlined in the introduction, the starting point for our investigation is given by the widely received assumption that nominativity and ergativity have a roughly equal cross-linguistic representation. This assumption is supported by statistical data based on samples of genetically mutually isolated languages (see Table 1); the table gives percentages (rather than absolute figures), because it represents mean values for several sub-samples of our database, each containing a single randomly selected language from each genetic stock (the database contains languages from 67 stocks). This sampling procedure was chosen as a starting point because it represents, oversimplifying the matter to some degree, what is widely considered the "ideal" approach to "probabilistic" typological sampling, which produces data least distorted by the effects of historical accidents. We will refer to such samples as I-samples below ("I" is intended as mnemonic for "isolated" or "independent").

Table 1. Distribution in random samples of mutually isolated languages

	Nominative		*Split*		*Ergative*	
Consistent	0.17		0.02		0.16	0.35
Differential		0.10	0.02	0.02		[+/-Neu]
Neutral			0.5			0.64
	[+Nom]: 0.31				[+Erg] 0.22	

Whereas the consistent nominative and consistent ergative types do indeed have a roughly equal representation in these samples, this is not the case for nominativity and ergativity in general, since the differential object marking (i.e. differential nominative) seems to be considerably more common than the differential subject marking (i.e. differential ergative). As a result, the nominative coding mechanism appears to be deployed more frequently than the ergative one. This fact is summarized in two figures in the bottom line of the table: [+Nom] corresponds to the typological variable "weak nominativity" (i.e. the presence of nominative-accusative encoding, possibly along with neutral and ergative encoding), [+Erg], to the similarly defined variable "weak ergativity". Further, the neutral encoding seems to be the most widely represented option:

it is the only possible encoding in ca. 50% of languages and one of alternative options in ca. 65% of languages (in I-samples). Note that this result is very similar to that presented by Comrie (2005: 399), based on a 190-languages sample, if one takes into account that his typology is defined in such a way that his "nominative" and "ergative" are much closer (albeit not identical) to our "weak" types than to our "consistent" types: ca. 52% of languages in Comrie's sample are neutral, ca. 27% are (weak) nominative, and ca. 24% are (weak) ergative. Apart from the slight differences in the definitions of types, another source of some divergence in figures might be a difference in sampling procedures: Comrie does not describe his sample in any detail, but judging from the general WALS guidelines (Comrie et al. 2005: 4) and the sample size, one can assume that the sampling procedure was also designed in such a way as to increase the "genetic distance" between languages, yet there was no strict one-language-per-stock constraint. Contrary to the generally received assumptions, such a sample may in fact give more linguistically relevant statistical evidence than an I-sample, unless some sampling decisions were made based on some a priori knowledge of individual alignment systems and/or other non-random considerations. We will return to this issue in Section 3.3. For now, it is important to stress a rather remarkable agreement between the results of these two absolutely independent "typological experiments".

3. Stochastic universals and language change

3.1. The hypothesis of stochastic universals

Linguistic typology has extended the concept of empirical language universal in such a way as to include so-called statistical, or stochastic, universals (or "linguistic preferences") (Greenberg 1963; Hawkins 1983; Comrie 1989: 19-22; Croft 1995; Dryer 1998; *inter alia*). The hypothesis of stochastic universals is, originally, a purely empirical one. It is grounded in the observable properties of the language population – in effect, in the observation that the distribution of languages along some parameters of variation is so uneven that (as our intuition tells us) it simply cannot be so skewed by chance alone and so must have a linguistic cause (Comrie 1989: 20). Yet the very concept of stochastic universal implies a very important theoretical hypothesis: namely, that Language (as a universal phenomenon) has certain probabilistic properties; i.e. that at least for some parameters of cross-linguistic variation there exist PROBABILITY DISTRIBUTIONS that are, in some sense, linguistically meaningful. For example, the data presented in Table 1 above might be interpreted as an indication that there exist some universal linguistic pressures against case-marking splits and/or for the presence of neutral encoding as a possible option (at least in some contexts). Indeed, possible linguistic reasons for the attested statistical patterns readily come to mind: one can imagine that the former class of constraints might be associated with avoidance of excessive paradigmatic complexity, and the latter, with avoidance of excessive structural markedness. As a matter of fact, if the hypothesis of stochastic universals is accepted, a uniform distribution can be taken to be just as linguistically significant: for instance, the same data set would tell us that the hypothesized probabilistic ("soft") universal pressures are as it were completely indifferent to the very existence of overt case markers for core participants in a language, since the consistently neutral alignment type (i.e. the typological state with no such case markers) is attested in ca. 50% of the languages.

It must be acknowledged from the very beginning that this empirical foundation for so crucial a hypothesis is a shaky one; indeed, it is by far easier to challenge it than to defend it. To begin with, we might ask ourselves, what kind of typological distributions would we expect to find if there was nothing probabilistic in the nature of Language, just some universals (genuinely obligatory properties) and some parameters of (limited) variation, with each value "doing" equally well and being equally probable? Could we realistically expect that any typological parameter defined by any linguist would have a roughly even distribution (in the language population as a whole or in any sample thereof)? If this had been the case, then that, indeed, would have been a sign of a divine intervention in linguistic affairs, for at least three independent reasons. First, any actual parameter of variation can be defined in a variety of different ways, resulting in different typologies and thus, inevitably, in different cross-linguistic distributions. To give the simplest example, some well-defined "types" can sometimes be justifiably split into two or more "types" depending on one's theoretical goals, and we certainly cannot expect the representation of all types in the population to be roughly equal in both cases. The differences in definitions of “nominative” and “ergative” in the present study and in (Comrie 2005), outlined in Section 2, constitute another case in point. Secondly, we know little about how random processes work in the language population; the point is, the randomness of underlying processes does not necessarily entail a uniform distribution, and there are no reasons to assume that this is the case for cross-linguistic distributions (Maslova 2000; Maslova forthcoming). And finally, even if uniform distributions were expected, for statistical reasons, in absence of universal probabilistic pressures, this would mean that such distributions would be observed for the MAJORITY of randomly selected parameters of variation, yet not for ALL OF THEM. Even assuming that significant deviations are unlikely (e.g. they occur with a probability around 0.05), it is still to be expected that the more parameters the typological community explores, the more likely it is to find some “skewed” distributions. Moreover, parameters for large-scale statistical typological studies can by no means be said to be selected randomly; rather, such a study is more likely to be undertaken if something interesting (that is, a significantly skewed distribution) is expected for a specific typological parameter, based on data available prior to the study. Accordingly, the total number of skewed distributions found so far is likely to be much higher than it would have been in any representative sample of typological parameters (however this concept is defined). The bottom line is that a fair number of skewed typological distributions were bound to be attested, quite independently of whether or not languages have interesting probabilistic properties.

This does not mean, to be sure, that the hypothesis of stochastic universals is false. It just means that we need some new ways to explore it. And to begin with, we need to divorce the theoretical hypothesis from its original empirical source – if only to be able to verify it by empirical data of the same sort. In other words, we need a definition of stochastic universal that would be, on the one hand, INDEPENDENT of the properties of the specific language population, and, on the other hand, sufficiently explicit and formalized to “interact” with statistical tests in a meaningful way. Only on this basis would we be able to figure out how to apply statistics to verify (or falsify) the hypothesis, both in general and with regard to specific parameters of variation. In a sense, this approach is opposite to the current typological practice, which seeks to “subtract” the properties of the population that are known to have nothing to do with language universals, that is, to "construct" a sample free of such non-linguistic effects as, say, the size of language families (this is what we did in Section 2 by constructing I-samples, i.e. giving all genetic stocks equal representation in each sample). The existing methods of statistical typological analysis more or less explicitly DEFINE a stochastic universal as something arrived at by means of application of these same methods (Dryer 1989; Perkins 1989, 2000); certainly, under such a definition, these methods are bound to be "correct", yet it means very little in terms of the relationship between their results and the universals of Language. In methodological discussions, we usually encounter some arguments why gathering statistics without suggested manipulations cannot give valid results, but hardly any as to why these manipulations can lead to results that are more so (if only because there is no independent explicit definition of what we actually want to achieve, and that is what we suggest to begin with). On the other hand, the data-independent definition we are going to suggest in the next section seems to conform to what is usually meant by statistical, or distributional, universals. In this sense, it does follow the common typological practice.

3.2. Language constants and language change

The concept of language universal is based on the notion that all human languages are instances of essentially the same phenomenon (Language with the capital "L") – in effect, the same "experiment", repeated by the history over and over again. This notion is particularly important for the concept of stochastic universals since their very manifestation depends on multiplicity of these experiments and thus on the assumption of the identity – in relevant respects – of the circumstances under which these experiments take place. It follows that a definition of stochastic universal should invoke LANGUAGE CONSTANTS, i.e. all aspects in which these experiments have been indeed identical. Roughly speaking, a language constant is a property which is true and must be true for each language; the list would include, along with absolute language universals (such as, for example, the existence of distinct (morpho)syntax and phonology), "non-linguistic" constants, that is, genuinely constant cognitive, social, physical, biological etc. properties of the environments in which languages exist and are transmitted from one generation to another; the failure of a language-like phenomenon to satisfy these properties entails, for a linguist, that this phenomenon should be excluded from a typological study, or at least treated carefully; e.g. pidgins or non-native languages (as spoken by adult learners) may be relevant examples. There may be a hierarchy among language constants, some of them being derivable from others; some constants might be considered theoretically irrelevant (not interesting) for some linguists (e.g. what is sometimes referred to as "performance pressures" might be disregarded by those only interested in "competence"), but this need not concern us here. We suggest to define LANGUAGE UNIVERSAL as a property which is directly or indirectly derivable from language constants, including but not limited to linguistic constants. A STOCHASTIC UNIVERSAL is, then, a probability distribution for a typological variable DETERMINED BY LANGUAGE CONSTANTS (or a joint probability distribution for several mutually dependent variables). A particular case of such universal, under this definition, would be a uniform distribution, which corresponds to a situation when the effect of language constants on linguistic variables amounts to limiting the range of possible values, without non-trivial probabilistic properties. The hypothesis of stochastic universals implies, then, that language constants have non-trivial (stochastic, non-deterministic) effects on some typological variables. This definition seems to conform with the actual typological practice: basically, having established some statistical irregularity (some sort of skewing in distribution), a typologist would look for, or postulate, a language constant (or a set thereof) that might explain the phenomenon, that is, constitute the possible cause of this phenomenon (cf. Hawkins 1990: 96).

The hypothesis of stochastic universals entails that a typological state (such as, for example, the state of having a consistently nominative-accusative alignment) has a certain probability of occurrence determined by language constants (i.e. the probability of a language being in this state). Moreover, this probability has to be manifested in the distribution of the type in the language population (this wording is intended to include but not to be limited to the frequency of the type in the population or a subset of the population). In other words, this is a property that is supposed to be "visible", at any given time, only because there are multiple languages in different typological states. That is, empirically, stochastic universals are visible at the level of language population, not at the level of any individual language. On the other hand, the loci of possible language constants are specific languages, i.e. individual speakers of each language and individual language communities. How, then, can these constants influence the statistical properties of the language population?

To begin with, some non-deterministic effects are apparently present both at the level of individual speakers (e.g. the choice of expression is not always fully determined by the intended meaning and its context, etc.; see (Bod et al. 2003) for a recent overview) and at the level of language community (e.g. there is a certain degree of randomness in how an innovation may or may not be propagated through the community (Labov 1994: 1-35)). At these levels, language constants interact with the individual properties of the specific language, including the current values of typological variables. Thus, the language behaviour of individual speakers and its effects on other members of the community can result in a change of the value of a typological variable. We can plausibly hypothesize that the likelihood of a language shifting to each possible "target" state is affected by language constants and by the current ("source") state of the language. There is a lot of unknown and controversial about how these processes might work, and a further discussion of the matter is beyond the scope of this paper. Two facts are essential in the present context: on the one hand, if language constants indeed determine systematic differences between transition probabilities for different logically possible pairs of "source" and "target" values of typological variables, this dependency provides a causal link between language constants and cross-linguistic statistical distributions, as implied by the hypothesis of stochastic universals. On the other hand, this is also the only logically possible link: there are simply no other ways in which language constants might affect statistical cross-linguistic distributions. In other words, the hypothesis of stochastic universals is, in fact, the hypothesis of existence of systematic differences in transition probabilities determined by language constants. It follows that, in order to decide whether a certain statistical pattern observed in the language population represents a stochastic universal we have to check whether or not it is determined by systematic differences in probabilities of transitions between typological states.

This idea is, of course, not new (cf. Greenberg 1978; 1995); the question is, how it can be implemented. The rest of the paper is intended to demonstrate that this can be done based on synchronic typological evidence, combined with information about genetic relationships between languages, but without specific assumptions on their prior typological states. However, although we try to follow the typological tradition in our understanding of what has to be done to establish a stochastic universal, our conclusions about how this has to be done in actual practice are quite different from the accepted typological wisdom.

3.3. "Apparent time" in linguistic typology

Let us begin by comparing the distribution of alignment types in I-samples (Table 1) and in a random sample (below, R-sample) from the language population (Table 2). Generally speaking, there is one major difference, namely, a shift along the horizontal ("nominative – ergative") dimension of our typological plane. For the sake of comparison, we repeat the figures from Table 1 (the frequencies attested in I-samples) in parentheses and show all significant differences in boldface.

Table 2. Distribution in a random sample from the language population

	Nominative		*Split*		*Ergative*	
Consistent	**0.22** (0.17)		0.05 (0.02)		**0.09** (0.16)	0.36 (0.35)
Differential		0.13 (0.10)	0.01 (0.02)	0.02 (0.02)		[+/-Neu]
Neutral			0.48 (0.5)			0.64 (0.65)
	[+Nom] **0.41** (0.31)				[+Erg] **0.17** (0.22)	

Frequencies from samples of mutually isolated languages are given in parentheses for comparison; significant differences are highlighted by boldface; for absolute numbers, see Table 6.

The general typological wisdom is to consider the distributions observed in I-samples as more linguistically meaningful than those in R-samples. The reasoning behind this approach is that the size of family is, from the linguistic point of view, an accidental property; and since it is highly probable that all or most members of a family exhibit the inherited value, giving the family a fair representation in the sampling procedure would unfairly increase the frequency of this inherited value (Dryer 1989: 258; Whaley 1997: 39). According to this logic, the higher frequency of nominative languages in the R-sample distribution is *a priori* attributed to a "conspiracy" of historical accidents resulting in a more rapid growth of "nominative" language families. Hence, it is considered more reasonable to give a single "slot" in the sample to each family and thus to reduce the potential effect of historical accidents. In our example, then, we would have to conclude that the right thing to do is to draw linguistic inferences from roughly equal representation of consistently nominative and consistently ergative local encoding, as observed in I-samples, and not from the significantly higher frequency of nominative encoding in the random sample.

This reasoning, however intuitively plausible, is seriously flawed. To begin with, the probability of the "birth-and-death" process (a.k.a. "historical accidents") producing significant differences in frequencies of typological states is very low in a large language population – to the extent that, statistically, we can consider it negligible (the relevant estimates are described in (Maslova 2000)). As a matter of fact, this is why the idea of "conspiracy" of historical accidents is commonly invoked to account for differences like those described above. The problem is, of course, that historical accidents cannot and do not conspire, and that's what statistics is all about. The real question is, if historical accidents cannot account for the observed differences, then what can? Statistically, the most likely answer is the general tendencies of language change, that is, systematic differences between transition probabilities.

Consider, for example, two interrelated differences between the I-sample distribution and the R-sample distribution, the increase in frequency of consistently nominative alignment and the decrease in frequency of consistently ergative alignment. Apparently, both types were represented in the population of ancestors of the modern genetic stocks. If both types are stable enough for most languages to have retained the inherited value (which is why I-samples are preferred in the first place), then an I-sample is most likely to contain a language with the inherited value from each stock (simply because there are more such languages in each or almost each stock). Yet if, say, the ergative alignment type is LESS STABLE than the nominative alignment, i.e. if there are systematic differences in transition probabilities, then there will be more languages that will have changed their alignment type among the descendants of ergative ancestors than among the descendants of nominative ancestors. As a result of this difference, the frequency of ergative languages in the modern language population (and, accordingly, in the R-sample) will have increased (which is what we actually observe). Thus, while an I-sample is most likely to represent a genetic stock with the inherited value, a R-sample would, as a rule, contain a higher percentage of more stable values. In other words, the I-sample distribution is very likely to be closer to the distribution in the ancestor population than the R-sample distribution, and the differences between them are likely to be determined primarily by the effects of language change during the time separating the ancestor population from the modern language population.[2]

There are no reasons to believe that the typological distribution in the ancestor population represents stochastic universals “better” that the corresponding distribution in the modern language population. On the contrary, it is very likely to be less linguistically meaningful, since the language population was not always large enough for the effects of historical accidents (i.e. of the birth-and-death process) to be insignificant. In a small population, this process has a good chance to bring about very strong effects (Maslova 2000), which means that, by the time when the language population became large enough for the law of large numbers to counteract the birth-and-death effects, its typological distributions used to reflect primarily the effects of these prehistoric accidents. Only after a large size had been achieved by the language population could the processes of language change begin to gradually shift these early distributions in a linguistically meaningful direction.

To sum up, the differences between I-sample and R-sample distributions are likely to reveal the diachronic dimension in synchronic cross-linguistic distributions, a typological analogue of the “apparent time” in sociolinguistics (Tillery et al. 1991; Bailey 2002; Labov 1994: 75-78). Like in sociolinguistics differences in the distributions of sociolinguistic variables in the speech of different generations are likely to indicate an on-going language change, so in typology differences between I-sample and R-sample distributions indicate an on-going shift in cross-linguistic frequencies of language types.

Contrary to what is usually assumed, R-sample distributions are likely to be more strongly affected by statistical regularities of language change and less strongly affected by "historical accidents" than the corresponding I-sample distributions.

This does not mean, of course, that we can draw linguistic inferences from R-sample distributions without further ado, nor that diachronic shifts in frequencies, like the shift along the "nominative – ergative dimension" described in this section, can be straightforwardly interpreted as linguistic preferences. As we try to show in Section 4, the key to establishing stochastic universals lies in combining synchronic and diachronic evidence. Before we turn to this problem, however, it seems necessary to discuss another commonly invoked argument against the validity of R-samples, namely, the argument from "non-independence" of genetically related languages (Bell 1978; Perkins 1989; Dryer 1989). What is usually meant by this is that many related languages often represent the same inherited value of a typological parameter, that is, a single event of change toward this value by the ancestor language. It would seem that, if we are interested in probabilities of change, we must take precautions against counting a single event of change multiple times, and an I-sample is the ultimate method of avoiding this trap. What this argument misses is that these languages also represent multiple events of RETAINING the typological value. If a family is large, then presumably a long time has passed since the time of the original language split; many linguistic things have changed – otherwise, we would not consider the languages as distinct. Yet the value of our parameter has not changed in most languages, which gives us statistical evidence of a fairly high stability of this value. It is this crucial piece of evidence that is lost in I-samples. One can say that the history obligingly stages multiple experiments, and we prefer to disregard them because we do not quite know how to interpret their results, and thus view blessings as methodological problems. The next section describes how the observable results of such historical experiments can be used to establish stochastic universals.

4. Establishing stochastic universals for alignment typology

4.1. Evidence from family-internal distributions

Although the general estimates of potential effects of the birth-and-death process referred to in the previous section strongly suggest that the differences between the I-sample distribution and the R-sample distribution cannot be accounted for by this process and thus must be due to the processes of language change, they cannot, strictly speaking, PROVE it for these particular distributions. After all, what is statistically unlikely can still occur in some cases (see Section 2.1). We can, however, also test this hypothesis by comparing intra-family distributions. As mentioned above, we generally assume that type-shifts are rare, and thus it is likely that the majority of languages in a family retain the inherited value. However, if one value (A) is even less likely to change than the opposite value (B), then this majority will be more significant in families that inherited the A-value than in families that inherited the B-value. In other words, if there exists a systematic difference in transition probabilities, we expect that the intra-family frequency of uncharacteristic ("minority") value will depend on which value is predominant for that family (a similar measure is used by Nichols (1992: 163-168) for a slightly different purpose). That is, in our hypothetical example, there will be, on average, more A-languages in B-families than B-languages in A-families.

These frequencies can by no means be taken as estimates of transition probabilities, because the effects of birth-and-death process within a single family can be very strong (since a single family can be thought of as a small population, at least at the first stages of its existence, see Section 4.1). However, they can provide some idea of whether there has been a significant difference between the transition probabilities over the time period separating the ancestor population from the modern population.

Table 3. Family-internal frequencies of uncharacteristic values

A =	*Nominative*	*Ergative*	*Neutral*
Frequency of B-languages in A-families	0.14	0.18	0.17
Frequency of A-languages in B-families	0.17	0.03	0.25

Table 3 represents our estimates of the family-internal frequencies of uncharacteristic values for three "weak" binary variables, [+Nom], [+Erg], and [+Neu]. The figures for [+Erg] indicate a significant difference in transition probabilities (a very low frequency of ergativity in predominantly non-ergative families is opposed to a relatively high frequency of non-ergativity in predominantly ergative families): the probability of acquiring an ergative encoding mechanism has apparently been much lower than the probability of losing such a mechanism, so that the overall effect of language change must have been a decrease in the frequency of ergative encoding. Hence, this contrast between the I-sample and R-sample distributions indeed cannot be attributed solely to the birth-and-death process; it is determined by a considerably higher diachronic stability of non-ergativity (as opposed to ergativity).

For [+Nom], on the other hand, there seems to be no significant dependency on the predominant type, that is, no systematic difference between transition probabilities detected by this rough statistics. This might seem to contradict our interpretation of the differences between the I-sample and R-sample distributions as indicating an increase in frequency of [+Nom]-languages due to language change. However, this is not the case. In order to demonstrate this, it will be convenient to represent the hypothesized difference between transition probabilities toward and from each language type in terms their ratio (α). Let us assume that, as evidence from family-internal distributions suggests, there is no difference between transition probabilities toward and from [+Nom], i.e. $\alpha(+\text{Nom}) = 1$. Assume, further, that the synchronic frequency of [+Nom] in the language population at some point in history was 0.3 (a figure close to the frequency of [+Nom] in the I-sample distribution). Now, a certain fraction of all languages in the population change their value of this variable within a certain time interval following this point in history. The question is how the frequency of [+Nom] will change as a result, i.e. whether there will be a drift to increase the frequency of nominativity, a drift in the opposite direction, or no change at all. It might seem that no significant shift is possible, because, according to our assumptions, there is approximately one transition to [-Nom] for each transition to [+NOM]. However, this would have been the case only if the initial frequencies of both types had been roughly equal; as it is, there were apparently more [-Nom] languages in the ancestor population, and, accordingly, more changes toward [+Nom], hence the diachronic drift towards nominativity. In sum, a diachronic change in cross-linguistic frequency is not only possible for a parameter with equiprobable transitions, but unavoidable if the existing frequency is not close to 50%.

The last variable, [+Neu], demonstrates the opposite situation: the frequency of consistent overt discrimination of participants in predominantly neutral families is relatively high, yet the frequency of neutral encoding in predominantly non-neutral families is considerably higher. This seems to indicate a systematic difference in transition probabilities in favour of [+Neu]. Let us assume, for the sake of argument, that α(+Neu) = 2, i.e. it is twice more likely for a language without neutral encoding option to acquire this option than for a [+Neu]-language to lose this option. Further, assume that the frequency of [+Neu]-languages is ca. 2/3 (which is very close to what we actually observe both in the I-sample distribution and in the R-sample distribution). What is likely to happen, under these assumptions, after a certain period of time, when some languages will have changed their value of this variable? The answer is, the distribution will have remained unchanged, since there will be roughly the same number of transitions in both directions. This is demonstrated by the following simple formula:

(1) $f'(+\text{Neu}) \approx f(+\text{Neu}) - 1/3c \cdot f(+\text{Neu}) + 2/3c \cdot f(-\text{Neu})$,

where c denotes the overall frequency of transitions along this parameter; since the likelihood of transition towards [+Neu] is twice higher than that of the reverse transition, the frequency of such transitions is ca. 2/3c, and the frequency of reverse transitions, 1/3c. The first term in the second part of the near-equation is the initial frequency, the second term corresponds to languages that will have lost their neutral mechanism, and the third term, to languages that will have acquired it. It can be easily observed that if f (+Neu) is 2/3 (and, accordingly, f(-Neu) is 1/3), then the last two terms cancel each other, so that there can be no shift in these frequencies due to language change: the synchronic frequencies are in the state of equilibrium determined by the ratio of transition probabilities. This means that evidence from family-internal distributions does not contradict, but rather supports our interpretation of the differences between I-sample and R-sample distributions: given that the actual frequency of [+Neu] is close to 2/3, we would not expect the type-shift processes to have changed this value if the probability of transition toward [+Neu] is approximately twice higher than the probability of reverse transition; accordingly, given the evidence about systematic differences in transition probabilities from family-internal distribution, we would expect no significant difference in frequency of +Neu between I-sample and R-sample distribution, and this is what we actually find.

In discussion of these data, we have established two important general points. First, a diachronic shift alone does not, by itself, provide evidence for systematic difference in transition probabilities; nor does the absence of a diachronic shift along some parameter demonstrate that there are no such differences. We have to take into account synchronic differences as well, for the simple reason that the total number of certain transitions depends not only on the probability of such a transition, but also on the number of languages in the appropriate source state. Secondly, a state of equilibrium between a synchronic distribution and diachronic tendencies can be achieved, so that the synchronic frequency of a type is not likely to be changed by further type-shift processes.

4.2. Stochastic universals as limiting distributions

The formula in (1) describes the expected change in synchronic frequency of [+Neu]. It can be easily generalized. As above, $\alpha(+X)$ is the ratio of transition probabilities, and c, the overall frequency of transitions along the same parameter within a certain time interval:

(2) $f'(+X) \approx f(+X) - (1 - \beta(+X))\cdot c\cdot f(+X) + \beta(+X)\cdot c\cdot f(-X)$,

where

(3) $\beta(+X) = \alpha(+X)/(1 + \alpha(+X))$.

It can be observed that if the current frequency of [+X] equals $\beta(+X)$, then the last two terms in (2) cancel each other, i.e. there are approximately equal number of transitions in both directions. Once achieved, this frequency would remain constant (disregarding slight statistical fluctuations). If the current frequency happens to be lower than $\beta(+X)$, then the processes of language change would gradually increase it until it reaches this value; if it happens to be higher than $\beta(+X)$, these processes would gradually decrease it. In other words, $\beta(+X)$ is the LIMITING FREQUENCY of +X: metaphorically speaking, it is the “goal” of the processes of language change with the ratio $\alpha(+X)$ of transition probabilities. After it is achieved, the synchronic distribution is in the state of equilibrium: if it accidentally shifts from this state, it will be soon “pushed” back by processes of language change. This is the unique distribution that is determined solely by systematic differences in transition probabilities, and thus the only possible candidate for the role of “stochastic universal” associated with a linguistic variable (see also Maslova 2000).

What is important is that processes of language change cannot really fail to modify a cross-linguistic distribution if the state of equilibrium is not achieved. Accordingly, if there has been no diachronic shift in frequencies over a long enough period of time, this strongly suggests that these frequencies approximate the limiting distribution, where “long enough” means simply that there have been some transitions from one value of the variable to the other, and vice versa. So far, we have identified one alignment-related variable that has apparently achieved the limiting distribution, [+Neu]: approximately two thirds of languages in both the I-sample and the R-sample distributions have a neutral encoding option; on the other hand, evidence from family-internal distributions suggests that quite a lot of transitions along this parameter have happened since the time of the ancestor population. If the actual frequency of [+Neu] approximates its limiting frequency, as suggested by this evidence, then we can also estimate the ratio of transition probabilities α(+Neu) (see the formula in (3)) as approximately two, that is, a language without a neutral option is twice more likely to acquire it than a language with a neutral option to lose it. Interestingly, the frequency of consistently neutral alignment seems to have remained roughly constant as well (approximately half of all languages in both samples). The corresponding estimate for the ratio of transition probabilities is one, i.e. transitions from and to this state are equiprobable.

What can we say about a typological variable if a diachronic drift in its distribution is attested, that is, there is no evidence that the limiting distribution is achieved? Some inferences can be drawn from the fact that such drifts would increase frequencies that are lower than their limiting values and decrease frequencies that are higher than their limiting values. For example, the frequency of [+Nom] is ca. 0.4 at the present time, and it has been increasing, which means that its limiting frequency, β (+Nom), cannot be lower than that. Then, the formula in (3) gives us an estimate of the LOWER BOUND for the ratio of transition probabilities, namely, α(+Nom) must be equal to or higher than ca. 2/3 (0.4 divided by 06). That is, if we could observe an equal number of languages with and without nominativity over the same period of time, there would be two or more shifts towards nominativity for every three losses of the nominative mechanism. Note that this estimate of the lower bound for α(+Nom) also agrees with the evidence from intra-family distributions, which do not demonstrate any significant differences in transition probabilities.

For [+Erg], the drift has been in the opposite direction: the frequency of languages with ergative encoding option decreased to ca. 0.17. This entails that the limiting frequency β(+Erg) cannot be higher than this value, which gives us an UPPER BOUND of ca. 1/5 for α(+Erg). That is, an ergative language has been at least five times more likely to lose its ergativity than a language without ergativity to develop an ergative case marker. Thus, even though we still do not know the exact values of the ratios of transition probabilities for these two variables, we can establish a rather significant probabilistic difference between [+Nom] and [+Erg]:

(4) $\alpha(+\text{ERG}) \leq 1/5$; $\alpha(+\text{NOM}) \geq 2/3$.

This means that, if we could observe the limiting distribution, we would be likely to find that nominative encoding is at least twice more probable than ergative encoding. In other words, language constants seem to favour, in this sense, morphological nominativity over morphological ergativity.

To conclude this section, an interesting question is why the language population apparently achieved the limiting distribution along the "neutral – marked" dimension some time ago, whereas the similar process for the "nominative – ergative" dimension drags behind. The most likely reason for this is that the former parameter is more mobile, i.e. the overall rate of change along this dimension has been consistently higher. Accordingly, it has taken less time for the processes of language change to obliterate the strong random effects of prehistoric accidents and to bring about the limiting distribution determined solely by the ratios of transition probabilities. This hypothesis is also supported by evidence from intra-family distributions (see Table 3): the frequencies of uncharacteristic values are higher for [+Neu], which indicates a higher probability of change along this dimension. Linguistically, this hypothesis seems plausible as well: it must be easier for a language to acquire or lose a single case marker than to change from one overt marking mechanism to the other (which would involve at least two different case markers).

4.3. Evidence from divergence rates

As shown above, evidence from comparison between I-samples and R-samples, supported by evidence from family-internal distributions, gives us an estimate of stochastic universal only if it turns out that the existing synchronic cross-linguistic distribution is close to the state of equilibrium with the corresponding type-shift processes: in this case, the synchronic frequencies can be taken as an approximation of the limiting frequencies determined by the ratios of transition probabilities. If a diachronic shift is detected (as in the case of the "nominative – ergative" dimension), this means that the language population is likely to be still drifting towards the limiting distribution. In such situations, a comparison between I-sample and R-sample distributions can only give us upper or lower bounds for the ratios of transition probabilities, depending on the direction in which the frequency is changing.

In order to obtain some estimates of these ratios in cases like this, we use, following (Maslova 2004), a new kind of typological statistics, called DIVERGENCE RATE.[3] The divergence rate is measured for a sample of PAIRS of related languages with a relatively small time depth and corresponds to the frequency of pairs that exhibit DIFFERENT values of this variable. The idea of this method is to measure divergence rates for at least two different samples with different synchronic distributions of the variable under investigation, and thus to detect a dependency between the frequency of each value and the corresponding divergence rate.

The rationale behind this method can be informally described as follows. Assume, for the sake of argument, that we know which value of the typological variable was exhibited by the ancestor language of each pair. Then, we can split the whole sample of such pairs into "A-pairs" and "B-pairs" depending on which value is inherited (as before, A and B denote the opposed values of a binary variable). If the A-value is more likely to change, then the first sub-sample will exhibit a higher divergence rate. Now, the same would be true even if the first sub-sample contained not only A-pairs, but just a higher percentage of A-pairs than of B-pairs: since there were more A-languages, there have been, on average, more changes. On the other hand, since changes are relatively rare events in any case, the first sub-sample would also exhibit a higher frequency of A-languages. These observations give us an opportunity to estimate the ratio of transition probabilities even if we do not know the ancestor values for our pairs. We can just select samples of pairs with different current synchronic distributions: since the time depth of pairs is relatively low, the difference in current frequencies is very likely to indicate a difference in the frequencies that existed a short while ago in the same sub-population of languages. In order to obtain samples with different synchronic distributions, we take one sample of pairs from predominantly A-families and the other sample, from predominantly B-families. Once such samples are obtained, we can estimate transition probabilities on the basis of synchronic frequencies and divergence rates in these samples, because both are determined by the initial frequencies and the transition probabilities (the relevant equations are given in Appendix 2). Note that this procedure actually does not involve any assumptions about the "ancestor" values; such assumptions were invoked here only to describe the essence of the method in informal terms. For a more detailed description of the method, see (Maslova 2004).

Table 4. Divergence rates for samples with different distributions

	Neutral		*Nominative*		*Ergative*	
	Frequency	*Divergence*	*Frequency*	*Divergence*	*Frequency*	*Divergence*
I.	0.85	0.20	0.45	0.26	0.62	0.56
II.	0.11	0.20	0.05	0.13	0.2	0.05

Consider, for example, the neutral alignment type (in this case, we discuss the neutral alignment in the strong sense, that is, the absence of any overt distinctions). The sub-sample from predominantly neutral families contains ca. 85% of neutral languages, and the subsample from predominantly non-neutral families, ca 11% of neutral languages. Yet the divergence rate turns out to be exactly the same in both cases (0.20), which indicates that the probability of change along this parameter does not depend on the current value (i.e. transition probabilities are roughly equal); see Table 4. Note that this conclusion conforms with our previous observations: both in the I-sample distribution and in the R-sample distribution, the frequency of neutral alignment is around ca. 50%. Thus, evidence from divergence rates supports our previous conclusion that this is indeed the limiting frequency. In other words, we can confirm the existence of a stochastic universal stating that the probability of neutral alignment is ca. 0.5.

If we repeat the same procedure for another variable, the consistently nominative encoding, we get a drastically different picture; as shown in the second pair of columns of Table 4, the divergence rate is 0.26 for a sample with ca. 45% nominative languages, and

0.13 for a sample with ca. 5% of such languages. In other words, there is a rather strong dependency: the more nominative languages, the higher the probability of change. The maximum likelihood estimate of the ratio of transition probabilities based on this data is 0.3 (that is, it is more than three times more likely for a consistently nominative language to lose this consistency in one or another way than for a language of a different type to acquire consistently nominative case marking). The corresponding estimate for the limiting frequency of consistently nominative encoding of full NPs is 0.23, which is only slightly higher than the corresponding actual frequency, as attested in our R-sample. For the consistently ergative encoding, we observe a similar direction of dependency (the higher the frequency of ergative languages, the higher the divergence rate), yet this dependency is even stronger (the divergence rate for a sample with a higher frequency of ergativity is 0.56), and the corresponding estimate for the ratio transition probabilities is, accordingly, even lower (ca. 0.08). The predicted limiting frequency is ca. 0.07 (which is slightly lower than the corresponding actual frequency in the R-sample).

Table 5. An estimate for the limiting distribution for alignment types

	Nominative		*Split*	*Ergative*	
Consistent	0.23		0.05	0.07	0.3~0.35
Differential		0.15			[+/-Neu]
Neutral			0.50		0.65~0.7
	[+Nom] 0.45			[+Erg] 0.12	

Table 5 summarizes our preliminary estimate of the stochastic universal (i.e. the limiting frequency distribution) for case marking of full NPs (we do not have enough data to estimate the distribution within the split/differential ergativity domain). It can be easily observed that these estimates confirm the conclusions based on the comparison between I-sample and R-sample distributions: our stochastic universal is indeed very close to the R-sample distribution, yet deviates still somewhat further from the I-sample distribution. The consistently nominative type is predicted to be at least three times more probable than the consistently ergative type; the difference becomes even more significant if differential marking systems are taken into account: the probability of a language having a nominative-accusative construction (possibly along with other coding options) is almost four times as high as the probability of having an ergative construction.

4.4. Summary

Our conclusions about the stochastic universals are based on three independent types of evidence (or “data points”):

a) The overall distribution of alignment types in the modern language population, as estimated on the basis of R-sample (synchronic distributions).
b) Intra-genetic distributions in genetic stocks with different predominant types and the resulting difference between the I-sample and R-sample distribution (major diachronic drifts on the time scale associated with the temporal distance between the modern languages and the ancestors of genetic stocks).
c) Divergence rates (transition probabilities for relatively short time intervals, corresponding to time depths of our pairs of closely related languages).

The presentation above may give an impression of non-independence of these data points, for two reasons. First, each method of analysis employed makes use of two types of data simultaneously. Secondly, within the accepted stochastic model of type shifts, all statistical measures used here depend on the value of a single parameter, the ratio of transition probabilities (that is why these measures are used in the first place). However, they are still independent if viewed as data points. This means that if our model were grossly wrong,[4] i.e. there was no “real” counterpart for the hypothesized consistent (i.e. temporally uniform) ratio of transition probabilities for each parameter (cf. (Croft 1990: 204; Newmeyer 1998: 320-325)), conclusions drawn from different data points would have been extremely unlikely to corroborate one another and to converge, as they did, on very similar estimates for the ratios of transition probabilities and the corresponding limiting distribution.

The most striking convergence is that between the R-sample distribution and the estimate for the limiting distribution based on divergence rates (cf. Table 2 and Table 5): there are virtually no statistically significant differences between these distributions. To be more precise, if we take our predictions based on divergence rates as a hypothesis about the actual distribution of alignment types in the modern language population and use our R-sample to test this hypothesis, it will or will not be rejected depending on the selected significance level, i.e. on the acceptable probability of rejecting a true hypothesis (for example, the χ^2-test will reject the hypothesis if the significance level is 0.05 and will fail to reject it at the significance level of 0.01; see Table 6). It seems, therefore, that the actual distribution is very close to the limiting distribution determined by the ratios of transition probabilities, as estimated for the most recent historical period. Since this historical period alone would not be long enough to bring about the limiting distribution, this convergence strongly suggests that the same systematic differences in transition probabilities have been at work for a much longer period of time (possibly for as long as the language population exists). As described above, this conclusion is also corroborated by the evidence from family-internal distributions and from the contrast between the I-sample and R-sample distributions.

Table 6. Testing the hypothesis of limiting distribution in the modern language population

	Nominative	*Nom. Diff.*	*Ergative*	*Split & Erg.Diff.*	*Neutral*
Expected	92	60	28	20	200
Actual	88	52	36	32	192
	$\chi^2 = 11.05$, $v = 4$, $p = 0.03$				

5. Conclusion

We hope to have shown that linguistically meaningful stochastic universals can only be discovered on the basis of statistical evidence about the dynamics of cross-linguistic distributions, and, furthermore, that such evidence can be obtained by analysis of synchronic distributions if we do not confine our analyses to samples of genetically isolated languages. As suggested by Greenberg (1978; 1995), this evidence is hidden in differences between cross-linguistic and intra-genetic distributions, which, if analysed properly, can reveal systematic differences between transition probabilities for parameters of typological variation. An important point is that a stochastic universal does not reveal itself in synchronic frequencies or diachronic trends taken separately: a synchronic distribution can retain some traces of prehistoric random effects (rather than being determined by language constants); on the other hand, a higher total number of changes in one direction can reflect a higher synchronic frequency of the corresponding source type (rather than a systematic differences in transition probabilities). The key to establishing stochastic universals lies in comparison between these two types of evidence, which makes it possible to find out how the synchronic frequency of a type differs from its limiting frequency determined by the ratio of transition probabilities.

We have discussed two different statistical approaches to the problem. One is relatively low-cost and is based on a comparison between an I-sample distribution and a R-sample distribution. This method provides a criterion for comparison between the existing cross-linguistic distribution and the hypothesized stochastic universal, i.e. it shows whether the language population has achieved the limiting distribution for the parameter under investigation. If yes, then the stochastic universal is established; in our specific case study, this happened to be the case for the neutral alignment type and associated linguistic variables. If not, this method will give only lower or upper bounds for linguistically meaningful typological probabilities, depending on the established direction of change. For many linguistic inferences, this is likely to be sufficient. Otherwise, the more time- and effort-consuming method based on divergence rates can be used. It requires a rather large "two-level" language sample, i.e. a sample of pairs of related languages from different language families, which could be split into at least two sub-samples with as different synchronic distributions of the variables under investigation as possible (see Appendix 2). Alternatively, the second level of sampling can be areal (rather than family-based), i.e. different samples of pairs can be drawn from different linguistic areas (Maslova 2004).

In the case of alignment typology, these methods give two major results. First, the existing cross-linguistic distribution along the "neutral – marked" dimension can be taken as a stochastic universal: linguistic constants apparently work in such a way that the probability of consistently neutral encoding is close to 1/2, and the probability of a language having a neutral encoding option is close to 2/3. This is another way of saying that the transitions are equiprobable for consistently neutral encoding, whereas the rise of neutral encoding as a grammatical option is twice as probable as its loss. Secondly, the cross-linguistic distribution along the "nominative – ergative" dimension is also rather close to the limiting distribution, but this is so only for the distribution in the modern language population as a whole (or random sample thereof), not for I-sample distributions. A distribution in a sample of genetically mutually isolated languages would reflect an earlier stage in the history of language population, when the limiting distribution had not yet been achieved. The most striking difference between the two is in the relative frequency of nominative and ergative languages: the I-sample distribution gives the impression of a roughly equal representation, whereas the limiting probability of nominative encoding is more than three times higher than the limiting probability of ergative encoding. To put it in slightly different terms, the nominative alignment is more diachronically stable than the ergative alignment, i.e. the expected life-time of nominative construction is considerably longer than that of ergative construction. The question of why language constants might work in such a way as to make morphological ergativity less stable than nominativity is beyond the scope of this paper; we would like to mention just one possible factor, namely, personal pronouns. It is well known that pronouns are much more likely to exhibit nominative encoding than full NPs (Silverstein 1976; Nichols 1992: 90-91; Comrie 2005: 400); this is corroborated by our study: if pronouns are taken into account, the predicted limiting distribution shifts towards the nominative apex of the alignment triangle. This means that nominative languages almost invariably have a single mechanism of discriminating between core participants for NPs and pronouns, whereas ergative languages are much more likely to have two different mechanisms. This heterogeneity can well be one of the factors that make the overall case-marking system less diachronically stable: it seems plausible to hypothesize that a shift to another alignment type is easier and therefore more likely if this type is already present in the case-marking system in some form (Harris & Campbell 1995: 255-263).

Finally, the results of our study allow for a rather optimistic conclusion for statistical cross-linguistic studies in general. Indeed, the alignment typology seems to be comprised of relatively stable, slow-changing typological parameters (Nichols 1992: 163-183), and the time needed to achieve the limiting distribution is determined primarily by the mobility of parameters. This entails that if, as our analysis suggests, the limiting distribution (or something very close to it) has been achieved for the alignment typology, it is very likely to have been achieved for all more diachronically mobile parameters as well, which means that their distribution in the modern language population can indeed be used for linguistic inferences. Thus, the working assumptions of statistically informed typological studies prove to be more plausible than they might have seemed.

Appendix 1. Database

!Kung (!Xu)	Neu
Abaza	Neu
Abkhaz	Neu
Achinese	Neu
Adyge	Erg
Afrikaans	Neu
Agul	Erg
Akan (1)	Neu
Akan (2)	Neu
Albanian	NomDiff
Aleut	ErgDiff
Altay	Nom
Alutor	Erg
Amharic	NomDiff
Amis (Nataoran)	Split
Andi (1)	Erg
Andi (2)	Erg
Andi lges	Erg
Arabana	Split
Arabic	Nom
Argobba	NomDiff
Armenian	NomDiff
Arosi	Nom
Assamese	Nom
Assiniboine	Neu
Assyrian	NomDiff
Asu	Neu
Avar	Erg
Avestan	Nom
Aymara	Neu
Azerbaydzhani	Nom
Bahnaric lges	Neu
Balangao	Neu
Balinese	Neu
Balochi	NomDiff
Baluchi (Beludzh)	Split
Bambara	Neu
Basaa	Neu
Bashkir	Nom
Basque	Erg
Batak (Toba)	Neu
Bats	Erg
Belorussian	NomDiff
Bemba	Neu
Benga	Neu
Bengali (banla)	NomDiff
Berber (KYL)	Nom
Berber (TZM)	Neu
Bete	Neu
Bidiya	Neu
Bikol	Split
Blackfoot	Neu
Bongo	Neu
Bontoc Igorot	Neu
Boso	Neu
Brahui	Nom
Breton	Neu
Bribri	Neu
Buginese	Neu
Bulgarian	Neu
Burmese	Nom
Burushaski	Erg
Buryat	NomDiff
Cabecar	ErgDiff
Cajun French	Neu
Cambodian (Khmer)	Neu
Carib	Neu
Catalan	Neu
Cebuano	Split
Chai (Suri)	ErgDiff
Cham	Neu
Chamorro	Neu
Chechen	Erg
Cherokee	Neu
Cheyenne	Neu
Chinese, Standard	Neu
Choctaw	Neu
Chukchi	Erg
Chuvash	NomDiff
Coptic	Neu

Cornish	Neu	German	NomDiff
Cree	Neu	Gikuyu	Neu
Crow	Neu	Godie	Neu
Czech	NomDiff	Gondi	Nom
Dagaare	Neu	Gorontalo	Neu
Dakota	SplitDiff	Gothic	Nom
Dan	Neu	Grebo	Neu
Dangaleat	Neu	Greek, Modern	NomDiff
Danish	Neu	Guarani	Nom
Dargva	Erg	Gujarati	Split
Degema	Neu	Gunwinggu	Neu
Dewoin	Neu	Gurenne	Neu
Dinka	Neu	Haida	Neu
Djingili	Erg	Haitian Creole	Neu
Douala	Neu	Hausa	Neu
Dumbea	Neu	Hawaiian	Neu
Dungan	Neu	Hebrew	NomDiff
Dutch	Neu	Hindi	Split
Dyirbal	Erg	Ho	Neu
Efik	Neu	Hopi	Nom
Enets	NomDiff	Hungarian	Nom
Engenni	Neu	Ibibio	Neu
English	Neu	Icelandic	Nom
Estonian	NomDiff	Idoma	Neu
Ethiopic	Nom	Iduna	Neu
Even	Nom	Igbo	Neu
Evenki	Nom	Ila	Neu
Ewe	Neu	Ilokano	Neu
Faeroese	Nom	Indonesian	Neu
Fe'fe'	Neu	Ingrian	NomDiff
Fijian	Neu	Ingush	Erg
Finnish	NomDiff	Inuit	Erg
French	Neu	Inuktitut	Erg
Fulani (FUB)	Neu	Irish	Neu
Fulani (FUH)	Neu	Irula	NomDiff
Gade	Neu	Ishkashim	NomDiff
Gagauz	Nom	Italian	Neu
Garawa	Erg	Itelmen	ErgDiff
Garo	Nom	Ivrit	NomDiff
Georgian	Split	Japanese	Nom

\

Javanese	Neu	Lakota	Neu
Juang	Neu	Lango	Neu
Kabard-Cherkes	Erg	Lao	Neu
Kabyle	NomDiff	Lappish	Nom
Kachin	Neu	Latin	Nom
Kalkatungu	Erg	Latvian	Nom
Kalmyk	Nom	Laz	Erg
Kannada	Nom	Lele	Neu
Kara	Neu	Lese	Nom
Karachay-Balkar	Nom	Lezgi	Erg
Karaim	Nom	Lhomi	Erg
Karakalpak	Nom	Lingala	Neu
Karelian	NomDiff	Lithuanian	Nom
Karen	Neu	Logo	Neu
Kasem	Neu	Loma	Neu
Kazakh	Nom	Lozi	Neu
Kedang	Neu	Lusatian	NomDiff
Kerek	Erg	Lyele	Neu
Ket	Neu	Maasai	Neu
Khakas	NomDiff	Macassarese	Neu
Khanty	Neu	Macedonian	NomDiff
Kharia	Neu	Madurese	Neu
Khasi	NomDiff	Malagasy	SplitDiff
Kirgiz	Nom	Malayalam	Nom
Kisi	Neu	Maltese	Neu
Komi	Nom	Mam	Neu
Komi-Zyryan	NomDiff	Mamvu	NomDiff
Korana	Neu	Manchu	Nom
Korean	Nom	Maninka	Neu
Koryak	Erg	Mano	Neu
Kpelle	Nom	Mansi	Neu
Kumyk	Nom	Manx	Neu
Kurdish	NomDiff	Maori	Nom
Kurmanji	Split	Mapudungu	Neu
Kwaio	Neu	Marathi	Split
Kwakiutl	Neu	Margi	Neu
Kwegu	Neu	Mari (MAL)	Nom
Ladakhi	Erg	Mari (MRJ)	Nom
Lahnda	Split	Marshallese	Neu
Lak	Split	Masalit	NomDiff

Maya	Erg	Orok	Nom
Mbara	Neu	Oromo (GAX)	Nom
Mende	Neu	Oromo (HAE)	Nom
Menomini	Neu	Osetin (Iron)	NomDiff
Minangkabau	Neu	Ossete	Nom
Mingrelian	Split	Palaung	Neu
Miskito	NomDiff	Pali	Nom
Mixtec (Jicaltepec)	Neu	Pangasinan	Split
Modo	Neu	Panjabi	Erg
Mokilese	Neu	Pashto	Split
Mon	Neu	Pero	Neu
Mongolian (KHK)	Nom	Persian	Nom
Mongolian (MVF)	Nom	Pitjantjatjara	Erg
Moore	Neu	Pitta-Pitta	Split
Mordva	Neu	Polabian	Nom
Motu	ErgDiff	Polish	NomDiff
Mundari	Neu	Portuguese	Neu
Mungaka	Neu	Pulaar	Neu
Murle	NomDiff	Quechua	Nom
Musgu	Neu	Quiche	Neu
Nahuatl	Neu	Romanian	Neu
Nama	Neu	Romany (Baltic)	NomDiff
Nanay	Nom	Rukai	Nom
Nandi	Neu	Runga	NomDiff
Negidal (1)	NomDiff	Russian	NomDiff
Negidal (2)	Nom	Saami, Kildin	Nom
Nenets	Nom	Sama/Bajaw	Neu
Nepali	Split	Samaritan	Neu
Nganasan	Nom	Samoan	Neu
Ngbaka Ma'bo	Neu	Sanskrit	Nom
Ngombe	Neu	Santali	Neu
Nicobarese	Split	Sarikoli	NomDiff
Nivkh	Nom	Saurashtra	NomDiff
Nogai	Nom	Scottish Gaelic	Nom
Norwegian	Neu	Selkup	Nom
Nubian	Neu	Seneca	Neu
Occitan	Neu	Serbo-Croat	Nom
Onondaga	Neu	Sherbro	Neu
Oriya	Nom	Shilluk	Neu
Oroch	Nom	Shughni (Bartangi)	NomDiff

Shughni (Rushani)	Neu	Tswana	Neu
Sicilian	Neu	Tupi	Neu
Sinaugoro	ErgDiff	Turkana	NomDiff
Sindhi	Neu	Turkish (Anatolian)	Nom
Sinhalese (Sinhala)	NomDiff	Turkmen	NomDiff
Slovak	Nom	Tuvinian	Nom
Slovene	Nom	Ubykh	Erg
Somali	Nom	Udege (OAC)	Nom
Soninke	Neu	Udege (UDE)	Nom
Sorbian (WEE)	NomDiff	Udmurt	Nom
Sorbian (WEN)	NomDiff	Ukrainian	NomDiff
Spanish	Neu	Ulch	Nom
Sundanese	Neu	Urak Lawoi'	Neu
Svan	Split	Urali	Nom
Swahili	Neu	Uygur	Nom
Swedish	Neu	Uzbek	Nom
Syriac	Neu	Veps	NomDiff
Tabassaran	Erg	Vietnamese	Neu
Tadzhik	Nom	Vot	NomDiff
Tahitian	Nom	Wakhi	Nom
Talysh	Split	Walbiri	Erg
Tamazight	NomDiff	Wambaya	Erg
Tamil	Nom	Wangkumara	Split
Tangale	Neu	Welsh	Neu
Tat (Muslim)	NomDiff	Wolof	Neu
Tatar	NomDiff	Xaracuu	Neu
Tawala	Neu	Yagnob	NomDiff
Telugu	Nom	Yakan	Erg
Tennet	Nom	Yakut	NomDiff
Tetun	Neu	Yala	Neu
Thai	Neu	Yanyala	Erg
Tharaka	Neu	Yazgulyam	Split
Thargari	Split	Yi (Lolo)	Neu
Tibetan	Split	Yoruba	Neu
Tigre	Neu	Yue	Neu
Tigrinya	NomDiff	Yukaghir (YKG)	NomDiff
Tiwi	Neu	Yukaghir (YUX)	NomDiff
Tlingit	Neu	Yupik, Sirenik	Erg
Tongan	Split	Zapotec	Neu
Trukese	Neu	Zuni	NomDiff

Appendix 2. Divergence rate

Let f_a denote the frequency of a language type in the set of ancestors of all language pairs in a sample; l denotes the probability of transition toward this type, and s, the probability of transition from this type. Then the probability P of a randomly selected language from this sample belonging to this type can be expressed as follows:

(5) $P = f_a \cdot (1 - s) + (1 - f_a) \cdot l$

The first term corresponds to languages retaining this type, and the second, to languages acquiring it.

The probability D of a randomly selected pair of related languages belonging to different types is:

(6) $D = 2f_a \cdot s \cdot (1 - s) + 2(1 - f_a) \cdot l \cdot (1 - l)$

The first term corresponds to pairs with common ancestors belonging to this type, the second term, to all other pairs. A divergent pair arises if exactly one language changes its type, hence the probability of transition must be multiplied by the probability of retaining the type (for the other language). The factor of 2 is needed because we make no assumptions about which language in each divergent pair exhibits the inherited value, i.e. our pairs are not ordered.

Now we can exclude the unknown value of f_a and obtain the following equation, which expresses D as a linear function of P:

(7) $D = aP + b$, where $a = 2(s - l)$, $b = 2l \cdot (1 - s)$

Using frequencies and divergence rates in two or more samples as estimates for P and D respectively for different (unknown) values of f_a, we can find the most likely values of the coefficients a and b, and, accordingly, the most likely value of the ratio $\alpha = l/s$ of transition probabilities.

Notes

1. For previous proposals on possible implementations of this idea, see (Maslova 2000; 2004); for a somewhat different approach, see (Nichols 1992).
2. Note that if the typological parameter under investigation happens to be very mobile, i.e. transitions between types are relatively frequent events (Hawkins 1983: 92-94), then the reason for the preference for an I-sample simply disappears, since the likelihood of preserving the inherited value is low (we will return to this issue in Section 4.1).
3. Please refer to the downloadable handout of (Maslova 2000) for an earlier English-language version.
4. The most likely reason why this model might be wrong is, of course, the influence of language contacts. A detailed discussion of how the role of this factor of language change can be tested for and/or taken into account is beyond the scope of this paper (see Maslova 2004). Suffice it to note that systematic, linguistically motivated differences between transition probabilities can co-exist with contact-induced changes, so our model does not imply that all type shifts must be internally motivated.

References

Aissen, Judith (2003). Differential object marking: iconicity and economy. *Natural Language and Linguistic Theory* 21/3: 435-83.

Bailey, Guy (2002). Real and apparent time. In J. K Chambers, Peter Trudgill & Natalie Schilling-Estes, eds. *The Handbook of Language Variation and Change*, 312-32. Malden, Mass: Blackwell Publishers.

Bell, Alan (1978). Language sampling. In (Greenberg et al. 1978), 125-56.

Bod, Rens, Jennifer Hay & Stefanie Jannedy, eds. (2003). *Probabilistic Linguistics*. Cambridge, Mass: MIT Press.

Bossong, Georg (1991). Differential object marking in Romance and beyond. In Dieter Wanner & Douglas A Kibbee, eds. *New Analyses in Romance Linguistics: Selected Papers From the Xviii Linguistic Symposium on Romance Languages, Urbana-Champaign, April 7-9, 1988*, Amsterdam, Philadelphia: John Benjamins.

Comrie, Bernard (1978). Ergativity. In Winfred P. Lehmann, ed. *Syntactic Typology : Studies in the Phenomenology of Language*, 329-294. Hassocks: Harvester Press.

— (1989). *Language Universals and Linguistic Typology : Syntax and Morphology*. 2nd ed. Oxford: Blackwell.

— (2005). Alignment of case marking. In (Haspelmath at al. 2005), 398-405.

Comrie, Bernard, Matthew S. Dryer, David Gil, and Martin Haspelmath (2005). "Introduction." In (Haspelmath at al. 2005), 1-8.

Croft, William (1990). *Typology and Universals*. Cambridge; New York: Cambridge University Press.

— (1995). Modern syntactic typology. In (Shibatani & Bynon 1995), 85-142.

— (2000). *Explaining Language Change : An Evolutionary Approach*. Harlow; New

York: Longman.

Dixon, Robert M. W. (1994). *Ergativity*. Cambridge, New York: Cambridge University Press.

Dryer, Matthew S. (1989). Large linguistic areas and language sampling. *Studies in Language* 13: 257-92.

— (1998). Why statistical universals are better than absolute universals. In *Papers From the 33rd Annual Meeting of the Chicago Linguistic Society*, 123-45.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, ed. *Universals of Language*, 73-113. Cambridge, Mass: M.I.T. Press.

— (1978). Diachrony, synchrony and language universals. In (Greenberg et al. 1978), 61-91.

— (1995). The diachronic typological approach to language." In (Shibatani & Bynon 1995), 143-66.

Greenberg, Joseph H., Charles Albert Ferguson, and Edith A Moravcsik, eds. (1978). *Universals of Human Language.* Stanford: Stanford University Press, 1978.

Harris, Alice C. & Lyle Campbell (1995). *Historical Syntax in Cross-Linguistic Perspective*. Cambridge, New York: Cambridge University Press.

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie, eds. (2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press, 2005.

Hawkins, John A. (1983) *Word Order Universals*, New York: Academic Press, 1983.

— (1990). Seeking motives for change in typological variation. In: William, Croft, Keith Denning and Suzanne Kemmer, eds. *Studies in Typology and Diachrony. Papers* *presented to Joseph H. Greenberg on his 75th birthday*, 95-128. Amsterdam, Philadelphia: John Benjamins.

Jäger, Gerhard. (forthcoming) Evolutionary game theory and typology. A case study. *Language*.

Kroch, Anthony S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1 (1989): 199-244.

Labov, William. (1994). *Principles of Linguistic Change*. Oxford, Cambridge [Mass.]: Blackwell.

Maslova, Elena (2000). A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4-3: 307-333.

— (2002). *Distributional universals and the rate of type shifts: towards a dynamic approach to "probability sampling".* Lecture given at the 3rd Winter Typological School, Moscow [www.stanford.edu/~emaslova/Publications/Sampling.pdf].

— (2004). Dinamika tipologičeskih raspredelenij i stabil'nost' jazykovyx tipov. [Dynamics of typological distributions and stability of language types]. *Voprosy jazykoznanija* 5: 3-16.

— (forthcoming). Meta-typological distributions. *Sprachtypologie und Universalien-vorshung*

Newmeyer, Frederick J. (1998). *Language Form and Language Function*. Cambridge, Mass: MIT Press.

Nichols, Johanna (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Payne, J.R. (1979). Transitivity and intransitivity in the Iranian languages of the USSR.

In Paul R Clyne, William F Hanks, and Carol L Hofbauer, eds. *The Elements, a Parasession on Linguistic Units and Levels, April 20-21, 1979 : Including Papers From the Conference on Non-Slavic Languages of the Ussr, April 18, 1979,* 436-47. Chicago: Chicago Linguistic Society.

Perkins, Revere D. (1989). Statistical techniques for determining language sample size. *Studies in Language* 13: 293-315.

Perkins, Revere D. (2001). Sampling procedures and statistical methods. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible, eds. *Language Typology and Language Universals: An International Handbook*, 419-34. 2001.

Shibatani, Masayoshi & Theodora Bynon (1995). *Approaches to Language Typology*, Oxford, New York: Clarendon Press, Oxford University Press.

Silverstein, Michael (1976). Hierarchy of features and ergativity. In Robert M. W. Dixon, ed. *Grammatical Categories in Australian Languages*, 112-71. Canberra: Australian Institute of Aboriginal Studies Humanities Press.

Tillery, Jan, Wikle, Tom, Bailey, Guy & Sand, Lory (1991). The apparent time construct. *Language Variation and Change* 3/3: 241-64.

Timberlake, Alan (1977). Reanalysis and actualization in syntactic change. In Charles N. Li, ed. *Mechanisms of Syntactic Change*, 141-77. Austin: University of Texas Press.

Whaley, Lindsay J. (1997). *Introduction to Typology : The Unity and Diversity of Language*. Thousand Oaks, Calif: Sage Publications.