Case marking patterns in diachronic perspective

Elena Maslova & Tatiana Nikitina
SYNTAX OF THE WORLD'S LANGUAGES
Leipzig (Germany), 5-8 August 2004

I. Claim & evidence

Our major claim: a strong universal *preference* for the nominative-accusative case-marking pattern over the ergative case-marking pattern in encoding of lexical NPs (in spite of their roughly equal representation in the language population).

Our database contains 401 languages; it has been designed in such a way as to include

(i) a random sample from the language population.
(ii) a sufficient cross-family and cross-genus coverage.
(iii) a random sample of pairs of (relatively) closely related languages from a variety of relatively large families (genetic groups)

The hypothesis is originally based on analysis of Johanna Nichols' database (1992; an extended database kindly made available by Johanna Nichols).

II. The problem of linguistic interpretation of typological statistical data: preliminaries

A. The hypothesis of stochastic universals

The hypothesis of stochastic universals is, originally, empirical. 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 it simply cannot be so skewed by chance alone and so must have a linguistic reason, hence a meaningful linguistic interpretation (that is, it must reflect something about Language). Yet in order to verify this hypothesis, we need a definition of stochastic universal that would be **independent** of the properties of (any) specific language population; only on this basis we would be able to figure out how to apply statistics to verify the hypothesis – both in general, and with regard to specific parameters of variation.

B. From universals to constants and variables

A **language constant** is a property which is indeed true for all languages; the list would include, along with (absolute) language universals, also "non-linguistic" constants – that is, genuinely constant cognitive, social, physical, biological etc. properties of the environment in which languages exist. **Language variables** can also be both linguistic (those we are genuinely interested in) and non-linguistic.

We suggest defining **language universal** as something (directly or indirectly) determined by language constants (in particular, **linguistic** language constants). A **stochastic** language universal is then a probability distribution function for a linguistic variable (or for a set thereof) determined by its interaction with (linguistic and non-linguistic) language constants. The hypothesis of stochastic universals implies that some of such distributions are non-uniform.

This definition seems to fit well with the typological practice: basically, having established evidence for a non-uniform distribution, a linguist would look for or postulate a language constant (or a set thereof) that might “explain” the phenomenon – i.e. be the possible cause of this phenomenon.

C. States vs. processes

There is only one phenomenon that can provide the “link” through which language constants can influence statistical distributions of linguistic variables among the world’s languages, the process of **language change**:

1. It can be appropriately modeled as a stochastic process – that is, a process whose outcome depends on chance.
2. Some properties of this process (the transition probabilities) can be determined by language constants.
3. This process obviously can change the cross-linguistic distribution of linguistic variables.

Discrete binary parameters:
Variable X = {A, B}; stochastic universal **U**(X) =*P*(A) (*P*(B) = 1 – *P*(A))
Type-shift process:

“Target” state “Source” state	**A**	**B**
A	*s*(A) [**stability**]	*l*(B) = 1-*s*(A)
B	*l*(B) = 1-*s*(B)	*s*(B) [**stability**]

The probability of A-value determined by transition probabilities is it **stationary** probability (if achieved, it will not be further modified by the process of language change in a significant way). The stochastic universal associated with X can then be defined as follows:

(1) $\mathbf{U}(X) = P(A) = l(A)/(l(A)+l(B))$

III. Typology of alignment in case marking systems

Table 1. Basic alignment types

Neutral construction	A-s P-s	widespread	+
Nominative construction	A-s P-p	widespread	+
Ergative construction	A-a P-s	widespread	+
Tripartite construction	A-a P-p	rare	–/+
Double-oblique construction	A-t P-t	almost non-existent	–

Here –s stands for a marker (including null) that can encoded S as well, -a stands for a marker available for A only, -p stands for a marker available for P only, and -t for a marker available for both A and P but not for S.

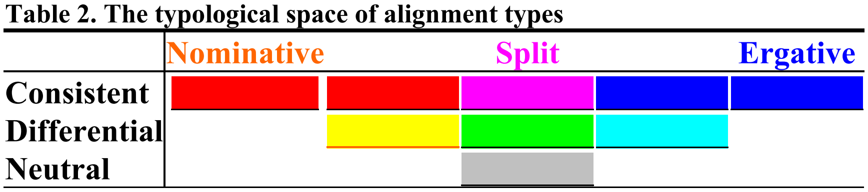
Generalized definitions for transitive constructions

Definition 1. A transitive construction is neutral if A and P are encoded (locally) in the same way (most often, it is zero-marking).

Definition 2. A transitive construction is nominative if A and P are encoded differently and the encoding for A is available for S and/or the encoding for P is not available for S.

Definition 3. A transitive construction is ergative if A and P are encoded differently and the encoding for P is available for S and/or the encoding for A is not available for S.

Table 2. The typological space of alignment types



Three variables:

Nominative = {Consistent, Optional, Absent}
Ergative = {Consistent, Optional, Absent}
Neutral = {Only, Optional, Absent}

Table 3. Distribution in random samples of mutually isolated languages

	Nominative		*Split*		*Ergative*	
Consistent	0.19 / 0.17		0.05 / 0.02		**0.12 / 0.16**	0.36 / 0.35
Differential		**0.24 / 0.10**	0.02 / 0.02	0.02 / 0.02		[+/-Neutral]
Neutral			**0.36 / 0.51**			0.64 / 0.65
	[+Nom]: 0.50 / 0.31			[+Erg] 0.21 / 0.22		

Figures before slash correspond to states defined for lexical and pronominal NPs together; figures after slash, for lexical NPs only; colored figures indicate significant differences.

IV. Estimating the dynamics in distributions of alignment types

A. "Apparent time" in cross-linguistic distributions

Table 4. Distribution in a random sample from the population

	Nominative	*Split*	*Ergative*	***Total***
Consistent	0.21 / **0.22**	0.05 / 0.05	**0.08 / 0.09**	0.34 / 0.36
Differential	**0.39 /** 0.13	0.02 / 0.01	0.02 / 0.02	[+/-Neutral]
Neutral		**0.23** / 0.48		0.66 / 0.64
	[+Nom] 0.67 / 0.41		[+Erg] **0.17 / 0.17**	

Colored figures indicate differences between distributions in family-based (one-language-per-family) samples and in language-based (random) samples from the population that **cannot** be plausibly accounted for by "accidents", but indicate diachronic drifts (effects of language change); red stands for change-determined increase, blue, for change-determined decrease.

Table 5. Family-internal frequencies of uncharacteristic values

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

All variables describe availability of the encoding for lexical NPs (e.g. "Nominative" includes differential nominative and split systems, etc.). The figures for A = nominative are very close, yet hint at a slightly higher probability of acquiring (some) nominativity than of loosing (all) nominativity. The figures for ergativity indicate a significant difference in transition probabilities (a very low frequency of ergativity in predominantly non-ergative families, and a relatively high frequency of non-ergativity in predominantly ergative families): the probability of acquiring ergativity is much lower than the probability of losing it. The situation for neutral encoding is opposite: while the frequency of consistent discrimination in families with higher frequency of neutral encoding is relatively high, yet the frequency of neutral encoding in predominantly discriminating families is even higher.

The difference between one-language-per-family distributions and distributions in random samples from the population can thus be viewed as the typological analogue of what is called "apparent time" in sociolinguistics. While in sociolinguistics a difference in the distributions of sociolinguistic variables in the speech of different generations indicates an on-going language change, so in typology a difference in family-based and language-based distributions indicates a diachronic trend in cross-linguistic distribution (i.e. the difference in total number of transitions).

B. Towards estimates of transition probabilities and stochastic universals

Table 6. Divergence rates for samples with different distributions of parameters

	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

The divergence rate is the frequency of pairs of related languages with different values of the variable; its dependency on the distribution of the variable in the sample indicates a difference in transition probabilities.

Table 7. A (rough) approximation of stochastic universal for (nominal) alignment types

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

Table 7 gives our preliminary (most likely) estimates of the stochastic universal(s) (i.e. stable frequency distributions) for case marking of lexical NPs (we do not have enough data to estimate the distribution within the split/differential ergativity area). It can be easily observed that these estimates conform with conclusions based on the comparison of two type one-level samples: our stochastic universal is indeed very close to the distribution in the population (yet it deviates still somewhat further from the one-language-per-sample distribution). The consistently nominative type is predicted to be much more probable than the consistently ergative type; the difference becomes even more significant if differential marking systems are taken into account. If pronouns are taken into account, the effect on the stochastic universal of alignment types is quite as expected: the optimal type shifts from neutral to differential nominative type (the predicted value for differential nominative encoding is **~0.45**), and (apparently) also from ergative to split systems (although here the effect is more difficult to ascertain due to small numbers).

Appendix:

(1) $f(A) = f'(A)s(A) + (1 - f'(A))l(B)$, where f'(A) is the frequency in the ancestor population

(2) $d(X) = 2f'(A)s(A)l(A) + 2(1 - f'(A))\, s(B)\, l(B)$

(3) $d(X) = af(A) + b$, where $a = 2(l(A)-l(B))$, $b = 2l(B)s(A)$