Interpreting typological clusters

MICHAEL CYSOUW

Abstract

In Linguistic Diversity in Space and Time *by Johanna Nichols (1992), typology as a discipline is given a new turn. Instead of interpreting typological clusters as an indication of a universal property of human language, Nichols interprets crosslinguistic patterns as indications of particular historical coincidences. In this paper, I criticise the method she uses to interpret her data. It turns out that her main conclusions are still valid, although she reaches them not through but notwithstanding her method. Specifically, the claim that there is a universal opposition between typically head-marking and typically dependent-marking languages cannot be based on her data. In contrast, the large-areal coherence which she observes can be deduced from her data with more detail in a new graphical type of analysis.*

Keywords: *areal distribution, head-dependent marking, macro-areas, methodology, sampling*

1. Introduction

In Johanna Nichols' work (1986, 1992), typology as a sub-discipline of linguistics (Comrie 1989, Croft 1990) is given a new twist. Traditionally, typology has been seen as being strongly related to research into universal properties of human language. In research that compares large sets of different languages, the terms "universal" and "typology" are used almost synonymously:

> Typically, linguists who are interested in language universals from the viewpoint of work on a wide range of languages are also interested in language typology, and it is very often difficult to classify a given piece of work in this area as being specifically on language universals as opposed to language typology. (Comrie 1989: 33)

Linguistic Typology 6 (2002), 69–93

1430–0532/2002/006-069

Typology, seen as the classification of languages into a limited set of types, is one of the major empirical methods of doing research into linguistic universals. Significant correlations between seemingly independent parameters are interpreted as a result of the influence of some universal human property. The strongest correlations are commonly referred to as "universals", following Greenberg (1963).

However, this is not the way typology is interpreted by Nichols (1992). To her, typology is a historical method. She interprets the existence of correlations as resulting from historical coincidences leading to the spread of a certain type over a large area (Nichols 1992: 2):

> Typology will be treated here as a population science, that is, a linguistic counterpart to population biology and population genetics, which analyse variation within and between populations of organisms and use the results to describe evolution. Viewing typology as a population science means shifting typology away from defining "possible human language" and instead pursuing generalisations about the world's languages.

This kind of approach received further backing by the thorough methodological analysis of Maslova (2000). Maslova argues that the typology-as-reflecting-universals approach only holds if the present distribution of the typology has reached a stationary stage (Maslova 2000: 313). Nichols assumed that this was not the case. However, her argumentation for this assumption is quite weak. In this article, I want to outline an approach which is able to show that the current distribution is not stationary. Specifically, I will scrutinise part of the data from Nichols (1992) to highlight some problems with the interpretation of typological data.

I will only consider Nichols' head–dependent typology, which is summarised in Section 2. There is a significant (inverse) correlation between head and dependent marking in Nichols' sample, which could be interpreted as the result of a universal opposing head and dependent marking. Another option is that the significance of this correlation is due to a historical coincidence. In this view, specific linguistic types happen to have had a wider influence than others, resulting in a large geographical spread of these types and thus skewing the typological distribution. The two interpretations ask for different analyses, as discussed in Section 3. In Section 4, I will show that Nichols made some errors concerning the universal interpretation of her data. Because of these errors, useful information is thrown out of her subsequent areal analysis. In Section 5, I will propose a different analysis for the geographical distribution of head–dependent marking, based on the assumption that head and dependent marking are independent parameters. It turns out that Nichols' areal conclusions are still valid, although she reaches them not through but notwithstanding her analysis. The main arguments are summarised in Section 6.

Table 1. *Syntactic constructions used for head–dependent counts, adapted from Nichols (1992: 47)*

Constituent	Dependent	Head
NP	Noun Possessor	Possessed Noun
	Pronoun Possessor	Possessed Noun
	Modifying Adjective	Modified Noun
S	Noun Subject	Verb
	Noun Direct Object	Verb
	Noun Indirect Object	Verb
	Pronoun Subject	Verb
	Pronoun Direct Object	Verb
	Pronoun Indirect Object	Verb

2. The head–dependent typology according to Nichols

The main typological parameter that Nichols uses in her 1992 book is HEAD–DEPENDENT marking. This parameter relies on the assumption that certain syntactic constructions are asymmetric: one part of the construction is the head, the other the dependent; and the typology then rests on the morphological marking being on the head or the dependent.[1] Nichols looks at nine syntactic constructions (see Table 1), and for every language in her sample, she counts the total number of head-marked constructions and the total number of dependent-marked constructions.[2] These counts determine a language's type. Head and dependent marking are interpreted by Nichols as two independent dimensions. If head–dependent marking were taken as one dimension, every construction would be classified as either head marking or dependent marking. In contrast, Nichols does not force a construction to be either one or the other. She also counts constructions that are marked on both head and dependent, and constructions that are not marked at all, neither on the head nor on the dependent. Head and dependent marking points are counted independently of

1. Nichols also describes an in-between form of marking, which she calls "detached" or "floating". As these floating markers, which are neither real head markers nor real dependent markers, occur only sparingly, I have decided to disregard them in my reanalyses rather than to force them into one or the other group. This simplification does not appear to influence the conclusions.
2. Nichols also included adpositional phrases. However, as they are problematic to compare crosslinguistically, she omits them from most of her analyses (Nichols 1992: 60). I follow her in this.

each other. In this way, it becomes an empirical question whether all combinations of head and dependent marking occur or do not occur. In Appendix 2 of her book Nichols summarises this typology (Nichols 1992: 292–301), listing, for instance, Amharic as showing four instances of head marking and nine instances of dependent marking; in other words, Amharic's head–dependent type is 4-9. The data from this appendix form the basis for the analyses in this article.

This typology distinguishes 100 types: both head marking (H) and dependent marking (D) in a language can vary from 0 to 9, a range of 10 possibilities for both dimensions resulting in $10 \cdot 10 = 100$ possible types. In a sample of "only" 172 languages Nichols finds 57 out of these 100 different types. There are strong indications that there is some clustering in the distribution of the 172 languages. Observing that not all 57 types occur with equal frequency, Nichols argues (1992: 76):

> Not all possible combinations of D and H values are attested. When D points are plotted against H points, shown in figure 3 [repeated here as Figure 1 – MC], a tendency towards clustering is revealed. There is a fairly wide high-density region in the lower right, comprising languages with a good deal of dependent-marking morphology and less head-marking morphology, and there is a more compact high-density region in the central to upper left, representing languages with more head-marking than dependent-marking morphology. The middle of the graph is mostly empty, showing that languages cluster in the predominantly head-marking and dependent-marking regions. The upper right and lower left are empty, showing that there is a negative correlation of H with D marking as well as a tendency to avoid complexity that would be entailed.

3. Interpreting the correlation

Some basic statistics indeed show a significant (inverse) correlation between head and dependent marking in the data, as presented in Figure 1.[3] But what does such a significant correlation mean? Nichols herself argues rightly that significance in itself does not mean anything. Rather, it is the phenomenon to be explained (Nichols 1992: 42). The most obvious explanations for a typological correlation are removed beforehand by the design of the sample. The sample is controlled for genetic and known small-size areal bias, ruling out these factors as an explanation of the correlation. There are two possible interpretations left for the significance: either it is due to some universal human constraint or it is due to large-areal coherence in the languages of the world (or of course

3. Pearson χ^2 of 102.1, significant at the .01 level (df = 81), Likelihood Ratio of 112.58, significant at the .001 level (df = 81); with such a high degree of freedom a chi-square is not very meaningful though. Pearson's r-value of −.39997, Spearman Correlation of −.40999.

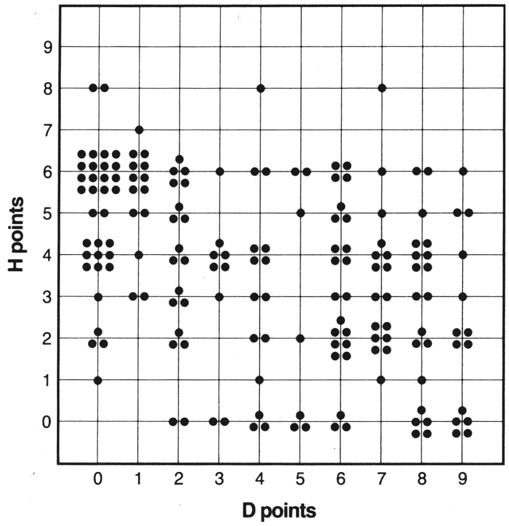

Figure 1. *Plot of D by H points, reproduced from Nichols (1992: 77)*

a combination of both of these factors). Further tests have to be executed to decide on the relative influence of the two interpretations of the data.[4]

An example of a test to check whether a significant correlation is a reflection of a linguistic universal has been developed in the work of Matthew Dryer (1989, 1991, 1992, 1997). He argues that, for there to be a universal, not only should the complete sample show a significant correlation, but a significant number of large areas in the world (each large enough to be linguistically diverse) should also show roughly the same correlation. Nichols does not perform such a test. She interprets the head–dependent clusters as a sign of a universal one-dimensional continuum between predominantly head marking and predominantly dependent marking. However, as will be shown in Section 4, by reducing her data to one dimension she causes the significant correlation to disappear (once the proper tests are performed).

Luckily, Nichols herself does not want to argue for a linguistic universal – she is interested in large-areal coherence in the world's languages. However, there is a touch of circularity in her argumentation. First, she notes that head and dependent marking are correlated. Then she interprets this correlation as a sign of a universal tendency of a language to be either head or dependent mark-

4. The existence of macro-genetic units – not yet discovered or generally accepted in the linguistic community – is a third possible explanation for large-areal coherence. This possibility will not be considered in this article.

ing. Next, she notes that the languages in various macro-areas in the world are restricted to proper subparts of the continuum between head and dependent marking. Finally, she interprets these macro-areas as leftovers from large-areal coherence. But the existence of such macro-areas implies that the original sample was not representative of the world's linguistic diversity, and as a consequence the correlation with which she started is undermined. In the words of Dryer (1989: 267): "If we underestimate the extent to which languages are related (genetically or areally), then we may reach conclusions that are in fact unsupported if not false." The same data cannot be interpreted both as indicating an area-specific pattern and as an area-independent, universal tendency. An analysis of existing areal patterns without this universal interpretation will be put forward in Section 5.

4. The reduction to one dimension

4.1. *Statistical problems*

Nichols collapses the two-dimensional array of head (H) and dependent (D) marking into one dimension with the two gross types: predominantly head marking and predominantly dependent marking, at either extreme. She uses three different transformations to reduce the two dimensions that were originally assumed to one: addition, division, and subtraction. These three transformations will be taken up in turn. I will show that the clustering that is found in the one-dimensional parameters turns out to be, to a large extent, a result of the transformation, not of the data. With the right statistical analysis, two of the transformations do not result in clustering anymore. The third transformation, which does yield clustering, is only occasionally referred to by Nichols in her 1992 book. I will conclude that these reductions to one dimension should not be interpreted as deeper interpretations of the data, but as a coarse summary of them, abstracting away from the interesting properties.

4.2. *The sum: D+H*

The first reduction, ADDITION (D+H), is the most intuitively understandable of these operations. Nichols calls this function the COMPLEXITY of a language. The more marking a language has, be it on the head or on the dependent, the higher the complexity. She argues that high and low complexity are found less frequently than moderate complexity (Nichols 1992: 87–88) (see Figure 2):

> The complexity has a roughly normal distribution, showing that languages avoid the extremes of complexity. Neither zero complexity not the theoretical maximum complexity of [18] points occurs.

There are two problems with this interpretation of the figure, both due to the fact that the numbers in the figure are taken in their absolute value; instead,

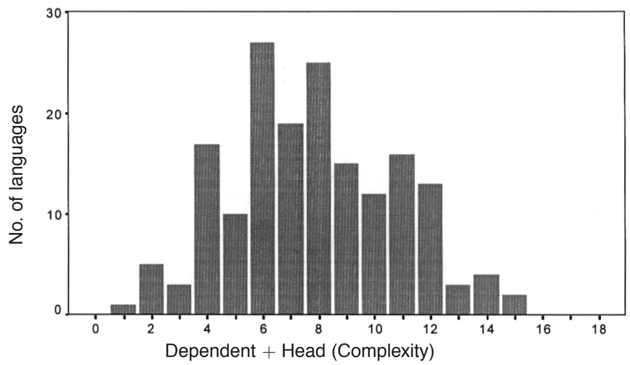

Figure 2. *Frequencies of complexity levels, adapted from Nichols (1992: 88)*

they should be compared to what is statistically expected. The first problem is connected to the fact that the values of head and dependent marking do not vary continuously, and the second problem is that head and dependent marking themselves are not evenly distributed. Both problems will also show up in the other operations.

The first problem stems from the fact that both head and dependent marking are discrete measures: only the integers 0 to 9 show up as possible values. This implies that the distribution found is, to a large extent, explained by mere chance. An extreme complexity of 0 has a low chance, since both head and dependent marking would have to be zero. The same holds for the extreme complexity of 18, where both head and dependent would have to be 9. The chances for a moderate complexity value like 9 are much higher, as this is found in languages with head–dependent type 0-9, 1-8, 2-7, etc. The expected distribution on this chance basis is given in Figure 3, showing a pyramidal distribution roughly comparable to the one found by Nichols in Figure 2.

The second problem is that the head and dependent dimensions in isolation are not evenly distributed in Nichols' sample. The distribution of languages relative to the amount of dependent marking is given in Figure 4. It shows a preference for even numbers, and the interpolation shows that there is a general preference for either a low or a moderately high amount of dependent marking. The distribution of head marking is rather different, as shown in Figure 5. Head marking shows an even stronger preference for even numbers, and the interpolation (with difficulty) shows that there is a preference for a moderate amount of marking. Given these distributions, the statistical expectation of the values for complexity is also influenced. The expected frequency of 0 complexity, for instance, is relatively high because of the high number of languages with

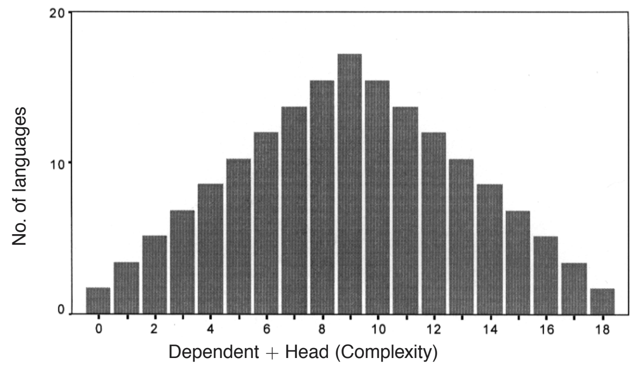

Figure 3. *Chance distribution of complexity levels*

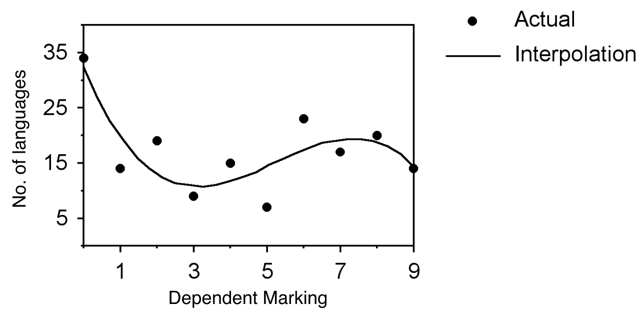

Figure 4. *Distribution of dependent marking (dots) in Nichols' sample with interpolation (line)*

zero dependent marking and the moderate number of languages with zero head marking.

When these two factors are combined, a statistical expectation of complexity can be computed. As an illustration, the computation of a statistical expectation of complexity 5 is explained here in detail. Six different head–dependent combinations have a complexity of 5: 0-5, 1-4, 2-3, 3-2, 4-1 and 5-0. First, the combination 0-5 is considered, which occurs in three cases in Nichols' sample. There are 23 cases with head marking zero and 7 cases with dependent marking five in the total sample of 172 cases (see Figure 1). These frequencies give a chance of $\frac{23}{172} \cdot \frac{7}{172} = 0.0054$ for the combination 0-5. For a sample of 172 cases, this chance gives a statistical expectation of $0.0054 \cdot 172 = 0.93$ cases for this combination. This value has to be compared with the three cases that are actually attested – which shows that the actual value is slightly higher than the

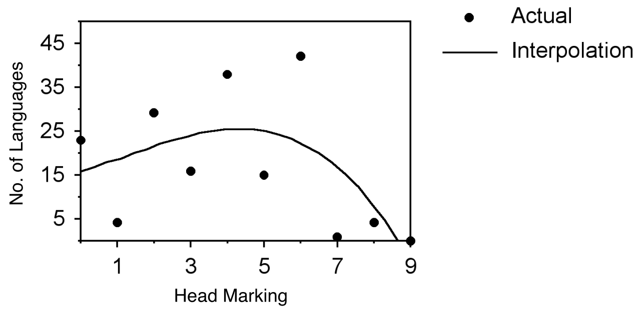

Figure 5. *Distribution of head marking (dots) in Nichols' sample with interpolation (line)*

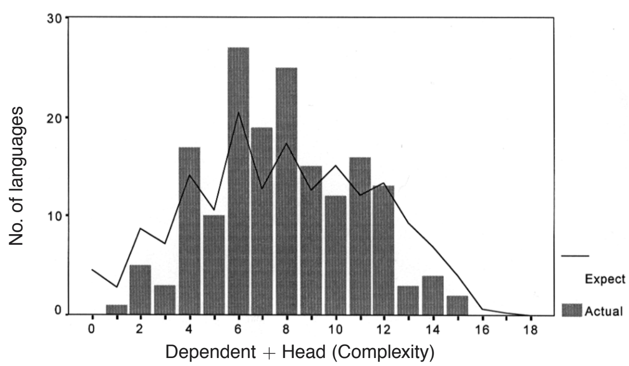

Figure 6. *Expected (line) and actual (bars) distribution of complexity (D+H)*

statistical expectation. The same computation has to be performed for the other five combinations and the results have to be added together to yield the expectation for complexity 5. The result is a statistical expectation of 10.63 cases for complexity 5. Actually, Nichols' sample has 10 cases with complexity 5; so, the statistical expectation almost precisely matches the actual value.

The expected complexity of all different levels of complexity is shown together with the complexity actually found in Figure 6. Although there are slight differences between the two, in general the actual distribution does not seem to differ much from what is expected. The actual distribution is slightly higher in the moderate range of complexity and slightly lower in the extreme ranges, but this tendency towards mid-range complexity is not nearly as strong as might be expected from Figure 2.

4.3. *The ratio: D/H*

The next transformation to be considered is the RATIO of head and dependent marking (D/H). It is not immediately clear what this transformation means linguistically. A possible interpretation is that the ratio D/H gives some idea of the amount of head marking relative to the amount of dependent marking. In this sense, it is a method of placing languages on a one-dimensional continuum between the hypothesised head and dependent poles of marking. Nichols mainly uses the D/H ratio for the argumentation in her 1992 book.[5] She notes that the occurrence of the different ratios is not evenly distributed: certain ratios occur more often than others (Nichols 1992: 73):

> [The distribution of D/H ratios] is bimodal, with the greatest peaks at the extremes of exclusive head marking (ratio of zero since D = 0) and exclusive dependent marking (since H = 0, an actual ratio cannot be computed as it has a zero denominator). [...] The other three frequency peaks suggest that preferred patterns cluster at perceptually simple ratios: two to one, one to one and one to two.

The same two problems that arose in connection with complexity as discussed in Section 4.2 are also relevant here. First, as the head and dependent values are not continuous, certain values of the ratio have a greater chance to occur than others. For instance, the cases 0/1, 0/2, 0/3, etc. are taken together as they all have ratio 0. This makes ratio 0 more likely than ratios like 5/2 or 7/3 that only occur in one instance. Second, the actual non-even distribution of the head and dependent parameters, as shown in Figure 4 and Figure 5, influence the expected occurrences of certain ratios. When both these (a priori) skewings of the data are combined to compute the expected values of the ratios D/H, it turns out that there is no difference between the expected and the actual values. In Figure 7 the bar chart of the ratios, as presented by Nichols (1992: 73), is reproduced showing the seemingly chaotic distribution of ratios. A line is drawn over the chart which indicates the statistically expected values. The expected values turn out to match the actual values almost precisely. The only three notable differences are at the points indicated in the graph: at 2/3, 2/1, and 6/1. These differences are probably just chance effects.

5. Nichols generally uses a transformed ratio D/D+H (which she rounded down to one decimal) to normalise the results of the arithmetic operation between 0 and 1. The ratio D/H is bound towards infinity as the value of H drops to zero, which complicates the calculations. The normalisation from D/H to D/D+H does not change anything on the distribution of the ratios: it only makes it easier to work with the results as infinity is eliminated. For the statistical analyses in her 1992 book, Nichols often uses gross types of these ratios, combining all ratios between .0 and .3 into one type, and likewise all ratios in the range .4 to .6 and .7 to 1 (Nichols 1992: 98). Interestingly, the distribution of these gross types only show a very slight deviation from chance, underlining the current argument.

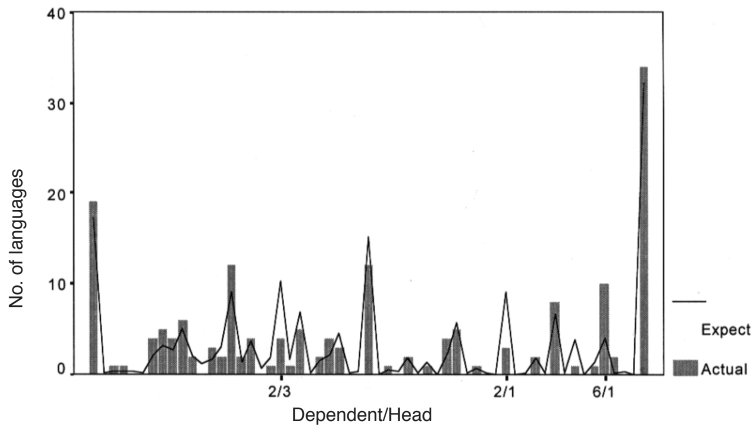

Figure 7. *Expected (line) and actual (bars) distribution of ratios (D/H)*

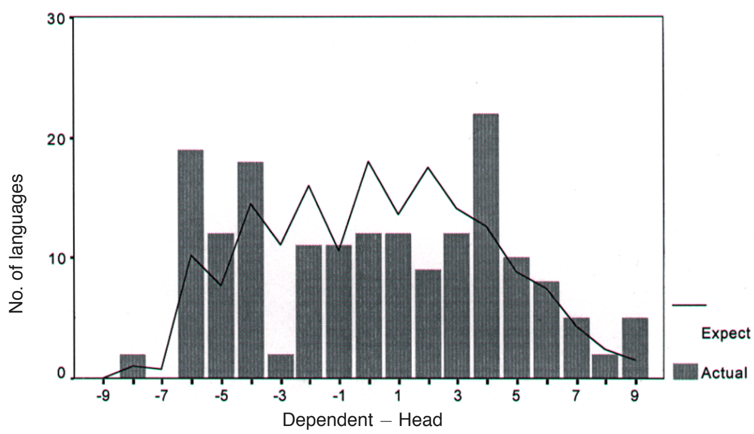

Figure 8. *Expected (line) and actual (bars) distribution of differences (D–H)*

4.4. *The difference: D−H*

The third transformation used by Nichols is the operation of SUBTRACTION (D−H). In Nichols (1986: 74) this measure is used as an argument for clustering, but it vanished almost completely from Nichols (1992). The same kinds of computation as explained above are performed; the actual and expected graphs are shown in Figure 8. This time there is some significant clustering found, roughly around the points where D−H is 4 and −5. In between these two clusters, the actual values are clearly lower than expected. This arithmetic finally seems to substantiate the claim that there is some clustering.

4.5. *Loss of information*

The first two one-dimensional reductions (D+H and D/H) do not support the claim that there is clustering. The third dimension (D−H) shows clustering, but Nichols does not use it for further analyses. Although she shows that there is a slight significant (inverse) correlation between head and dependent marking (see Section 3 above), the reductions to one dimension that she ultimately uses in her book do not reflect this correlation. However, the errors that result from ignoring the statistical expectation of the frequencies are not problematic for the main argumentation in Nichols (1992). Nichols is not primarily interested in the universal interpretation, but in areal phenomena. She devised the transformations to come to grips with the large amount of data for her geographical analyses (Nichols, p.c.). This is sound methodology, as she compares differences between the results of the transformations for different areas, not the results themselves. The criticism of the transformations in this section does not apply if that is the goal for which they are used. The transformations are criticised on the interpretation that they show a universal preference for either head or dependent marking. That interpretation is not supported. Nichols (1992) should therefore not be viewed as substantiating the idea that languages show a universal preference for either head or dependent marking.

Even Nichols herself argues, if only implicitly, for the opposite conclusion: the observed clusters of head and dependent marking are the result of large areal clusters (cf. Nichols 1992: 215–218). However, she is not completely clear as to the interpretation of her data. At times, she hints at a universal interpretation (e.g., Nichols 1992: 76), but that would be a rather strange conclusion given the argument of the rest of the book. If taken seriously, her argumentation would be circular. First she interprets the distribution of head–dependent marking as a sign of universal preference on a one-dimensional scale between either head or dependent marking. Subsequently, she shows (and this is the main part of her book) that there are macro-areal patterns in the distribution of this scale. Yet, if there are macro-areal preferences, then the sample that gave rise to the one-dimensional universal interpretation is biased, because the macro-areas are oversampled.

Dropping the universal interpretation and trying to show the areal patterns without reduction to one dimension can resolve this circularity. The areal results can actually be shown much more clearly if the full fine-grained two-dimensional data set is used instead of the rough reduction to one dimension. A lot of information is excluded by reducing the wealth of the two-dimensional data into a one-dimensional measure because languages with different head and dependent values end up in the same type. For example, head–dependent types 3-4 and 2-5 are of different types, but have the same complexity value, namely 7. Likewise for head–dependent types 6-4 and 3-2: they are of different types, but have the same ratio, namely 1.5. Instead of collapsing the two dimensions into one, it is more revealing to work with the two-dimensional data, interpreting head and dependent marking as two independent dimensions. This idea will be elaborated in the next section.

5. A graphical analysis of areal patterns

5.1. *Areal preferences*

A possible explanation for the existence of clusters in Nichols' head–dependent typology is that they exist due to the overrepresentation of certain macro-areas that show roughly the same head–dependent type. To investigate this I will assume that head and dependent marking are independent variables. This may seem a bit strange because there is a significant correlation as indicated by basic correlation measures (see Section 3). Indeed, the assumption that head and dependent marking are independent is not true. However, I want to argue here that the correlation is (at least to a large extent) the result of the geographical distribution of head-dependent types. Only from the assumption of independence is it possible to decide whether the correlation is or is not due to influence from oversampled macro-areas. The hypothesis that I will attempt to disprove in this section is that head–dependent clusters are independent of areal preferences. If the correlation is due to some universal constraint, then the same clusters are expected to be found everywhere, independently of which region of the world is examined.[6] If this hypothesis is falsified, this indicates that the correlation is at least partly due to oversampling of some coherent macro-areas. [7]

The first step, in Section 5.2, will be to improve on the visual representation of the two-dimensional distributions as shown in Figure 1. The clustering in different geographical regions will be studied in Section 5.3, revealing interesting differences and agreements between regions. The differences between the regional clusters clearly indicate that the hypothesis has to be rejected. This means that areal distribution has at least some influence on the clustering attested. In Section 5.4, I summarise the findings from the various areas and conclude that the hypothesis of areal independence should be rejected.

5.2. *A two-dimensional analysis*

Given the assumption that head and dependent marking are independent, the data have to be investigated without collapsing the two dimensions into one.

---

6. This is so on the premise that the areas are large enough to prevent known biases by common genetic or small-areal influence.
7. This is almost the same test as the one proposed by Dryer (see Section 3), but with the opposite purpose. Instead of showing that a universal is independent of the areal distribution, I want to falsify the claim of areal independence.

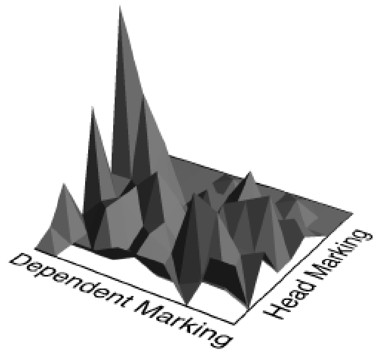

Figure 9. *Head–dependent array depicted as curved plane*

The central idea is to interpret the number of languages at a certain point in the two-dimensional head–dependent array as a value of height. The more languages of a certain head–dependent type are present in the sample, the higher this type will be depicted. The resulting plane is shown in Figure 9. In other words, the height of each junction in this figure is determined by the number of points in Figure 1. The different frequencies in the sample can be clearly seen, and also some clustering is roughly identifiable.

However, this figure is hard to handle. The differences between adjacent points in this diagram change much too abruptly, making it impossible to analyse broader clustering. The peaks and valleys vary so strongly mainly because of the preference for even values of head and dependent marking, as was shown in Section 4.2. A smoothing transformation is executed on the data to abstract away from this preference for even numbers. Instead of taking the actual occurrences of a certain type, the total number of languages of all neighbouring types is added to it. A correction is used to account for a structural deviation on the border of the head–dependent array, which is a result of the definition of the transformation. This process is shown in Figure 10. The three-dimensional graph after the application of the smoothing transformation is shown in Figure 11. The two clusters that Nichols deduced from Figure 1 can be clearly seen in this figure.

A contour plot is made of this curved plane to obtain a better view of these clusters, as shown in Figure 12. The lines in this plot indicate points of equal heights; darker shades indicate greater height. This contour plot should not be interpreted quantitatively anymore. The precise place of the tops in the diagram, the height of the tops and small subtops are strongly influenced by the smoothing transformation and the shading. For that reason, no units are shown on the axes and no quantitative analyses of these contour plots are made. How-

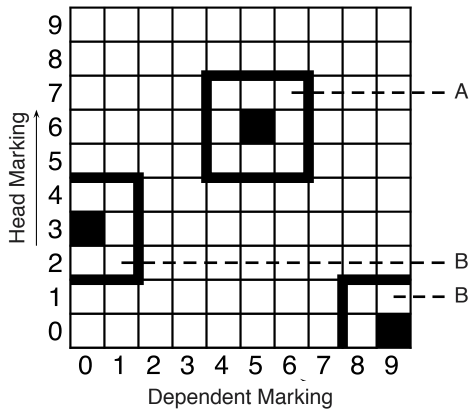

Figure 10. *Smoothing transformation.*
A: All occurrences of the neighbouring types are added to the occurrences of type 6-5 to smoothen over the preference for even values.
B: The values of the border and corner types are corrected because the sums are structurally lower due to less neighbouring types. The total of all occurrences around the "border"-type 3-0 is multiplied by 9⁄6, the total of the "corner"-type 0-9 is multiplied with 9⁄4.

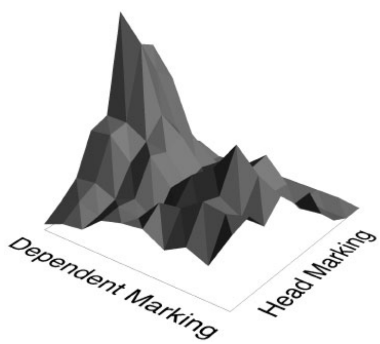

Figure 11. *Head–dependent plane after smoothing*

ever, the figure can be interpreted qualitatively, showing two clusters and a great spread of diversity.[8]

8. The method to represent clusters as outlined in this section is related to the "isopleth" method as developed by Van der Auwera (1998) and in Van der Auwera (ed.) (1998). The method used here is the inverse of the isopleth method, and can thus be called "inverted isopleth". In the isopleth method, LANGUAGES are fixed on a two-dimensional (geographical) map and the height of the plane (to be shown as contours) is determined by the number of features

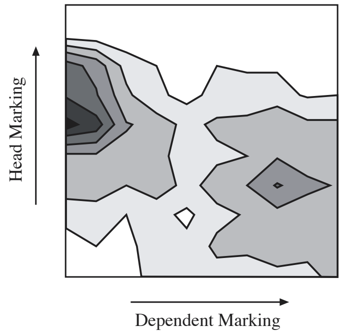

Figure 12. *Contour plot of smoothed head–dependent plane*

The contours in each figure in the next section are placed equidistantly. Within any one figure the height difference between every pair of consecutive contours is identical. However, the height difference changes from one figure to the other, as different intervals were needed to accurately show the three-dimensional form of the planes. This means that the shades of grey that are chosen to show the contours in each figure indicate different heights in the various figures. Because of this, it is possible to compare the form of the tops (e.g., height, amount of spread, steepness) INSIDE any one figure, but it is not possible to compare the form of tops BETWEEN one figure and another. Only the approximate PLACE of the tops in the head–dependent plane will be compared between the different figures.

5.3. *Areal breakdown*

The two-dimensional analysis will now be used to show the clustering in certain areas of the world. Nichols distinguishes three macro-areas divided into ten sample areas as shown in Table 2. Some of these areas do not show clustering if analysed as isolated parts of the world. The languages sampled in these areas are too diverse to bear out any clustering at all, probably because the number of languages in Nichols' sample is too small to represent the linguistic variation in these areas. For this reason, certain sample areas are taken together in the analyses that follow. First, Africa, Ancient Near East, and the western part of Northern Eurasia are lumped together. This is in fact the western part of the

---

from a chosen set that occur in each language. In contrast, in the inverted isopleth method, the FEATURES are fixed and the height of the plane is determined by the number of languages that show a certain combination of features. The isopleth method is useful when there is a single ordered dimension of features (or an unordered set of features); the inverted isopleth method is useful when there are two ordered dimensions of features.

Table 2. *Areal breakdown, adapted from Nichols (1992: 27)*

Macro-area	Sample area
Old World	Africa Ancient Near East Northern Eurasia (East & West) South and Southeast Asia
Pacific	Oceania New Guinea Australia
New World	North America Mesoamerica South America

macro-area "Old World". None of the three areas evidence clustering by themselves, but together they do show one main cluster at the dependent-marking side as shown in Figure 13. Note the great diversity found in this area as shown by the widespread base of the cluster, indicating shallow tops of the curved plane. This area does not have the same clusters as does the analysis of the whole world. The cluster on the head-marking side in Figure 12 is not found here, and the cluster on the dependent-marking side is even more extremely dependent marking. This indicates that the clusters found in the complete sample are different from the clusters found in these restricted areas. This fact challenges the hypothesis that the head–dependent distribution is independent of areal preferences.

Eastwards, the situation gradually starts to differ. In Figure 14 the contour plot of the eastern part of Northern Eurasia is depicted. There are three separate clusters found here. Just as in the previous picture, there are clusters at the

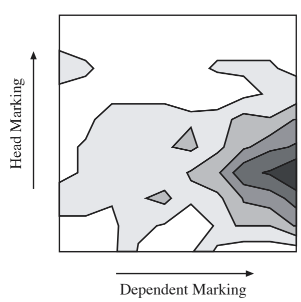

Figure 13. *Africa, Ancient Near East and Northern Eurasia (West)*

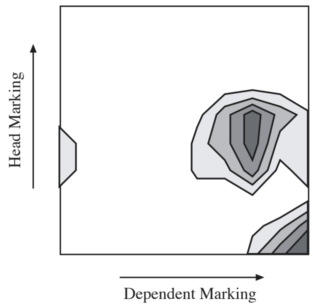

Figure 14. *Northern Eurasia (East)*

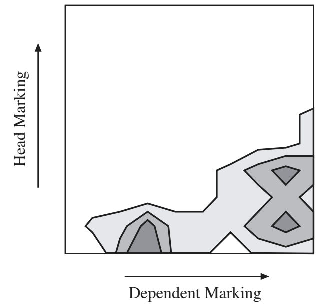

Figure 15. *South and Southeast Asia*

dependent-marking side, but now there are two more or less well distinguished clusters. These different clusters have a geographical correlate. The upper cluster consists of languages from the former Soviet Union (Chukchi, Ket, Nanai, Tuva, Yukaghir, and Nenets) and the lower cluster consists of languages on the southern border of this region (Mongolian, Korean, and Japanese). Note the little cluster on the head-marking side: these are the languages Ainu and Nivkh. Two languages are not enough to amount to a real cluster, but their different behaviour is noteworthy. They roughly fit into the head-marking cluster found in Figure 12. The clusters found in this region are roughly identical to the two clusters found in the world-wide analysis and offer no good reason to falsify the hypothesis of areal independence.

The contour-plot of South and Southeast Asia is shown in Figure 15. There are two clusters found in this region. Again we see a cluster roughly around the same dependent-marking spot as in the other Old World regions. This cluster consists of the more western languages of this area, all found roughly in or around India (Burushaski, Gurung, Kota, Nahali, and Waigali). The other cluster is different: it consists of the Southeast Asian languages (Mandarin, Miao, Temiar, and Thai).[9] These sample languages are rather typical of the region: they show hardly any marking at all. The noteworthy aspect of this cluster is that its position is clearly not a preferred place to form a cluster compared to the whole-world distribution. There is no cluster in the world-wide distribution (as shown in Figure 12) on this lower-left corner in the head–dependent array. But this does not make these languages "exceptions", "unusual types", or languages in a "transitional stage." Mandarin for instance is known to have exhibited this type for a few thousand years. Moreover, the whole region shows this type, and if some more languages from West Africa had been sampled, an-

9. The language Acehnese belongs to this area. However, it does not show up as part of either of the two clusters.

other comparable low-marking cluster would have been found there. It does not seem to be an exceptional or unusual type. The type just happens to be underrepresented in the sample, and therefore does not show up as a cluster in the world-wide analysis. This area provides a clear argument for rejecting the hypothesis of areal independence.

In Figure 16, New Guinea and Oceania are taken together because the clusters of these areas happened to be almost identical. The Oceanic languages tend to be a bit lower on head marking but still fall inside the range of variation found in New Guinea. Two beautiful clusters are found, one a bit more towards the dependent-marking side, the other some more to the head-marking side. There is a slight, but possibly significant difference between the geographical distribution of the languages that form these two clusters. The languages in the dependent-marking cluster are more restricted in their geographical distribution than the languages in the head-marking cluster.[10] The New Guinea dependent-marking languages are all found in the north-eastern region of the island. In contrast, the New Guinea head-marking languages are found all over the island. The Oceanic dependent-marking languages are found more to the south, but as there are only two languages, this is not really significant. In general, this area probably comes closest to repeating the clustering in the whole world: there are two clear clusters, roughly centred around the same types as in Figure 12. The centres of the clusters have been shifted a little towards each other compared to the clusters found in the whole-world analysis. This area does not provide an argument for rejecting the hypothesis of areal independence.

Things are quite different in Australia. As can be seen in Figure 17, the clusters found in Australia do not resemble the whole-world clusters at all. The extreme dependent-marking cluster consists of the Pama-Nyungan languages. They typically do not have any head marking but a lot of dependent marking. Note that this extreme head–dependent type is also found in Northern Eurasia (Mongolian, Korean, and Japanese). It is too far-fetched to propose any direct relationship between these two areas, but the fact that the same type exists in rather distant parts of the world does indicate that it is not an exceptional state for a language to be in, although it does not show up as a cluster in the world-wide distribution. The other cluster in Figure 17, with a fair amount of head marking and varying amounts of dependent marking, consists of languages that are found in Northern Australia (the non-Pama-Nyungan

10. The languages that form the dependent-marking cluster in New Guinea are Abelam, Awtuw, Hua, Kâte, Kewa, Ku Waru, Sentani, Suena, and Waris; in Oceania they are Drehu and West Futuna.

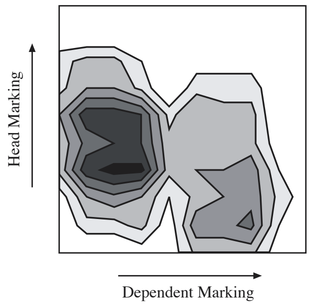

Figure 16. *New Guinea and Oceania*

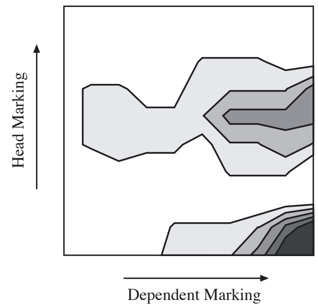

Figure 17. *Australia*

region). This cluster has a tendency towards extreme complexity: these languages show as high an amount of head marking as of dependent marking. Again, this corner of the head–dependent array may seem exceptional as it is clearly more extreme than the dependent cluster found in the whole-world figure. However, just as with the other "exceptions" that were discussed before (Southeast Asia, Pama-Nyungan), languages with high complexity are no exceptions. Besides Northern Australia, there is a group of languages with comparable types south of the Himalayas (cf. the dependent cluster in Figure 15). The clustering in Australia is a strong argument for rejecting the hypothesis of areal independence because it does not resemble the whole-world figure at all.

The last macro-area to be examined is the New World, divided into north, meso, and south by Nichols. The clustering in North America is shown in Figure 18. This clustering strongly resembles the whole-world clustering. The centres of the clusters are roughly identical to the centres in the whole-world picture. The main difference is that the clusters in North America are more clearly separated from each other. This distribution seems to substantiate the hypothesis of areal independence. However, there is still some basis for doubt about areal independence, because the two clusters show a strong areal bias inside the area itself. The languages that form the dependent-marking cluster are geographically strongly restricted near the West Coast of North America.[11] The languages in the head-marking cluster are found all over North America, including the West Coast. Note that, just as in New Guinea, it is the dependent-marking cluster which is geographically more restricted. The head-marking cluster is spread geographically all over North America. The large geographical spread of the head-marking cluster in North America even extends into

11. The languages forming this dependent-marking cluster are Eastern Pomo, Gitksan, Lower Umpqua, Luiseño, Maidu, Southern Paiute, Sahaptin, Squamish, Southern Sierra Miwok, Wappo, Wintu, Yawelmani. The only geographical outlier in this cluster is Tonkawa.

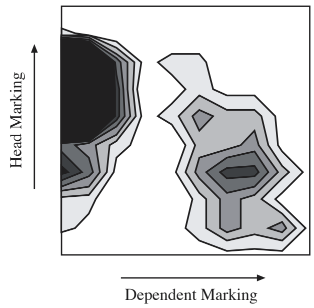

Figure 18. *North America*

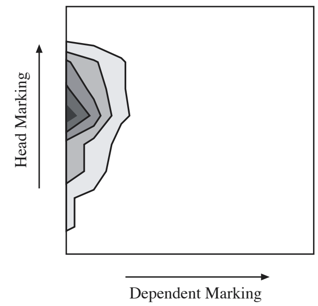

Figure 19. *Mesoamerica*

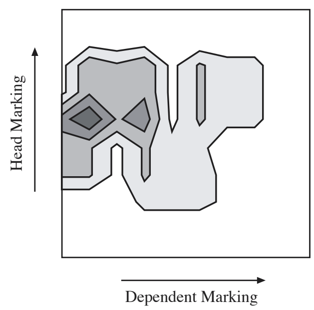

Figure 20. *South America*

Mesoamerica. As can be seen in Figure 19, there is only a head-marking cluster found here.[12]

The last area, South America, is shown in Figure 20. The clustering is hard to interpret as it does not show a regular outline like the others. The languages seem to have a fair amount of head marking but the amount of dependent marking differs strongly. The clusters in this area do not show any resemblance to the clusters world-wide. If this figure is regarded as interpretable, then this region is again a reason to reject the hypothesis of areal independence. [13]

5.4. *Summary of areal clustering*

The areas reviewed here all show clustering, but almost all do it their own way. It is striking to see the strong individuality of clustering in the different areas.

12. Miskito and Tarascan lie outside this cluster. Tarascan could belong to the dependent-marking cluster in North America. The other eight languages in Nichols' sample form the head-marking cluster (Chichimec, Chontal, Huave, Mixe, Mixtec, Pipil, Tepehua, and Tzutujil).
13. The language Cashinahua falls outside the cluster shown.

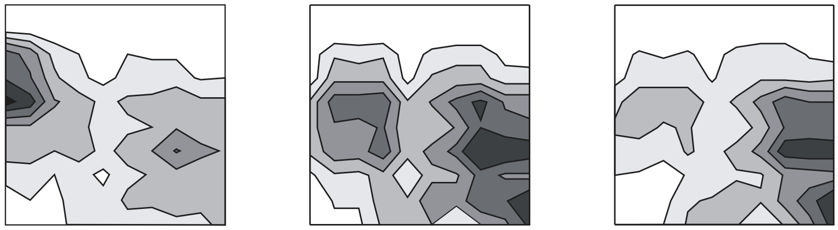

Figure 21. *World (with restrictions). The head-marking cluster vanishes when the most prolific head-marking regions in the world are removed from the sample. Left: All languages in Nichols' sample; middle: without North America and Mesoamerica; right: without North America, Mesoamerica, New Guinea and Oceania.*

Most of them do not resemble the world-wide clustering at all. In the two areas that resemble the whole world to some extent (New Guinea/Oceania and North America/Mesoamerica), the languages that make up the clusters are not evenly distributed over these two regions. In both areas the dependent cluster consists of languages that are found in a rather restricted geographical space. In contrast, the head-marking clusters consist of languages found equally distributed over both areas.

Using a rule of thumb as suggested by Matthew Dryer (p.c.), the influence of these two areas on the world-wide distribution can be investigated by removing the languages in these areas from the sample. If the same clustering as in the world-wide picture remains, this would be a strong counterargument against the view defended here that the world-wide distribution is strongly influenced by regional preferences. The result of this analysis is shown in Figure 21. First, note that the head-marking cluster as found in the world-wide distribution (see the leftmost graph in Figure 21) most strongly resembles the head-marking cluster in North America and Mesoamerica (see Figure 18 and Figure 19). If these languages from North America and Mesoamerica are removed from the sample, there is still a clear head-marking cluster, but it has moved a little (see the middle graph in Figure 21). Now, this head-marking cluster strongly resembles the head-marking cluster as found in New Guinea/Oceania (see Figure 16). If the languages from New Guinea and Oceania are removed from the sample, then the head-marking cluster almost completely disappears (see the rightmost graph in Figure 21). The head-marking cluster as found in the world-wide distribution is thus clearly dependent on these two macro-areas.

The diversity of areal clustering implies that the hypothesis of areal independence has to be rejected: the world-wide skewing in head-dependent marking is (at least to a large extent) due to the non-homogeneous distribution of head-dependent types over the world. This conclusion also means that the world-wide clustering does not unequivocally substantiate a claim for a human universal. In the words of Maslova (2000), there seems to be no stationary distribution. Because the significant (inverse) correlation between head and dependent marking in the complete sample was not very strong, it is even possible that the significance of this correlation is the result of a geographical bias of the sample. There happen to be a lot of languages from North America in Nichols' sample, and they strengthen the extreme head-marking type in the world-wide figure (cf. Dryer 1989: 264–265). On the other hand, there are only a few languages from West Africa and Southeast Asia in the sample, and this fact depresses the occurrence of the low-marking type in the sample.

If it is assumed that the sample of languages used by Nichols is representative of the actual diversity of the world's languages, the non-homogeneous distribution in this sample indicates that the ACTUAL skewing does not necessarily represent POSSIBLE preferences of human language. From the diversity of clusters in different regions it may just as well be concluded that languages can appear just about everywhere in the two-dimensional array of head and dependent marking. They float around through the array in space and time. Clustering world-wide is (at least to a large extent) the result of historical coincidences of these movements. If one would still want to explore the possibility of linguistic universals of head–dependent marking, the way to go would be to estimate the probabilities of a change of type (Maslova 2000: 328–330). In practice, this could be done by comparing the linguistic diversity on the level of linguistic families (as proposed by Dryer 1989, 1992, 2000) with the linguistic diversity within the families. A relatively high within-family diversity would suggest a high probability of change.

6. Conclusion

The analysis of large amounts of data, typical of typological work, is full of pitfalls. One of the more obvious problems is that a pattern in the data may be interpreted as a significant deviation from chance, although, on closer inspection, the pattern is not statistically relevant (cf. Hentschel 2000: 66–69). A rather subtle example of this error has been disentangled in Section 4. In this case, the error occurred because the original parameters were combined into complex parameters by some simple transformations. It turned out that the resulting patterns were the result of the transformations, not of the original data.

A much more difficult problem is the interpretation of a significant correlation, if one is found in a properly stratified sample of the world's languages. At least two possible explanations can be proposed for the significance. First, the significance can be due to the influence of a universal tendency among the world's languages and, second, it can be due to the influence of some large macro-areal typological coherence that was not discounted in the design of the sample (or of course a combination of both explanations). I have argued that the significant correlation that Nichols (1986, 1992) found between head and dependent marking can be explained, at least to a large extent, by large-areal coherence. This is exactly what Nichols herself has argued, but her argumentation has a touch of circularity because she based her analysis on a universal interpretation of the data (see Section 3). The analysis as presented in Section 5 does not use any universalistic interpretation but directly shows the large-areal coherence using a new graphical method.

The influence of large-areal coherence on the interpretations of typological patterns is a topic that deserves more attention. In principle, if large-areal coherence exist, this could have a devastating influence on universalistic interpretations of typological patterns. Typological patterns as observed in the past decades are mostly based on samples that are not stratified according to (eventual) macro-areas. This has not been possible because not much was known about such macro-areas. The least that should be done in the future is a post-hoc analysis of the geographic distribution of a typological pattern. If there are macro-areas to be observed, then any universalistic interpretations of such a pattern should be made with great caution.

Received: 27 May 1998
Revised: 12 November 2001

Zentrum für Allgemeine Sprachwissenschaft

Correspondence address: Zentrum für Allgemeine Sprachwissenschaft, Jägerstraße 10/11, 10117 Berlin, Germany; e-mail: cysouw@zas.gwz-berlin.de

Acknowledgments: I thank Johanna Nichols for comments on and discussion of the content of this article and for her continuous pressure to finally finish it. Further, I thank Matthew Dryer for many detailed questions and comments, which have made it possible for me to substantially improve the article in the course of the many versions that it has gone through. Finally, I thank various anonymous referees for *LT* for in-depth discussion and useful comments. All remaining shortcomings are of course my own responsibility.

References

Comrie, Bernard (1989). *Language Universals and Linguistic Typology*. 2nd edition. Oxford: Blackwell.

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

Dryer, Matthew S. (1989). Large linguistic areas and language sampling. *Studies in Language* 13: 257–292.

— (1991). SVO languages and the OV:VO typology. *Journal of Linguistics* 27: 443–482.

— (1992). The Greenbergian word order correlations. *Language* 68: 80–138.

— (1997). Why statistical universals are better than absolute universals. *Chicago Linguistic Society* 33-2: 123–145.

— (2000). Counting genera vs. counting languages. *Linguistic Typology* 4: 334–350.

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.: MIT Press.

Hentschel, Gerd (2000). On investigating marking variation of noun phrases: Some problems of definition and of quantitative analysis. *Sprachtypologie und Universalienforschung* 53: 62–70.

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

Nichols, Johanna (1986). Head-marking and dependent-marking grammar. *Language* 62: 56–119.

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

Van der Auwera, Johan (1998). Revisiting the Balkan and Meso-American linguistic areas. *Language Sciences* 20: 259–270.

Van der Auwera, Johan (ed.) (1998). *Adverbial Constructions in the Languages of Europe*. Berlin: Mouton de Gruyter.