Formulating and misformulating language universals

Vladimir Pericliev
Bulgarian Academy of Sciences

Linguists usually define language universals as true statements for all (or almost all) languages. This definition, however, may be misleading since truthfulness alone does not guarantee the linguistic relevance of a posited universal. The article discusses a number of misformulations of universals from the linguistic literature, amends accordingly the concept of ‘universal’, and lists the correct logical form of all universals involving two parameters that can be derived from contingency (tetrachoric) tables.

Keywords: Greenbergian universals, universals definition, universals misformulation, universals discovery

1. Introduction

There are two basic approaches to the study of universals, one advocated by Noam Chomsky, focusing on the study of individual languages in order to reveal properties that are common to all languages, and one pioneered by Joseph Greenberg (1966a, 1966b, 1966c, 1978b) and further developed by other scholars (e.g. Comrie 1981, Croft 1990 and others), focusing on representative samples of the world languages to derive the universal patterns observed in those samples, and potentially valid in all languages. The study of universals has generally been considered a basic goal of linguistic theory, although recently opinions have been voiced (see especially Evans & Levinson (2009), and some of its commentaries; also Lingua Volume 120, Issue 12) that the diversity of natural languages rather should lie at the centre of linguistic research. Whatever the worth of this new proposal and its influence upon the future development of linguistics, however, our considerations in this article, related to how we should conceive the notion
of universal and how to derive universals from data, will be linguistically relevant, as similarities and differences in languages must co-exist, and are but the two sides of the same coin. Thus, in saying that the languages of the world are diverse, we are already committed to the view that they possess properties that are common to all, and which allow us to call them ‘human languages’, that is, we are ultimately assuming the existence of universals. On the other hand, the universalist view, in claiming commonalities between the languages of the world, is committed to the view that these individual languages are distinct, that is, it ultimately assumes the existence of language diversity. Put differently, similarities exist because differences exist, and vice versa, differences exist because similarities exist. We are therefore forced to conclude that both aspects of human language, their commonalities and their diversity, deserve equal attention, and the choice of one perspective over the other is only a matter of research interest.

In this article, I am concerned with universals in the Greenbergian tradition. In the study of universals in this tradition, the linguist faces the task of positing universals, given contingency (i.e. tetrachoric) tables, constructed from the available language samples. The successful accomplishment of this task crucially hinges upon the availability of an adequate understanding of the notion of language universal. Language, or observational, universals (to be referred further on just as ‘universals’) are commonly conceived as propositions that are true for all, or almost all, languages (relative to a sample). Compare, for example, definitions like ‘Language universals are statements that are true of all languages [. . .]’ (Bickel 2010: 77), ‘A universal is a linguistic feature, usually phonological or grammatical, hypothesized as being shared by all human languages [. . .]’ (Hopper 1992: 136), ‘Linguistic universal. Strictly, a property that all languages have, or a statement that holds for all languages.’ (Matthews 1997), ‘Universal (adj./n.). A term used in linguistics, [. . .] referring to a property claimed to be common for all languages.’ (Crystal 2008).

In the article, I will try to show that this conception of a universal, in which the truthfulness (relative to a sample of attested and unattested language types) is conceived as a necessary and sufficient condition for positing a universal, despite its seeming innocence, may be – and in fact has been – linguistically misleading and has resulted in universals misformulations.

Some basic problems with the conception of universals as merely true statements can best be seen in everyday examples. Thus, one can truly assert that tomorrow it will either rain or it will not rain, however no one
will accept this statement as a proper forecast because the statement will always be true, irrespectively of what actually happens (logically, the proposition ‘P or not-P’ is a tautology, i.e. necessarily true by virtue of its logical form). Also, one can truly assert that if one eats some food (P) he will die at some future time (Q), but this true implication P → Q will be misleading in suggesting that the food is the actual cause of the death, whereas the empirical truth of ‘everyone will die or humans are mortal’ is the real reason why the whole statement ‘if one eats some food (P) he will die at some future time (Q)’ will necessarily be true (logically, Q logically implies P → Q).

These illustrations are extreme and conspicuously flawed and may suggest that similar statements will not occur in linguistic (and more generally, scientific) discourse. A more careful look at the literature on universals, however, reveals that they can smuggle in more covert form in positing universals and showing this, and appropriately amending the current definition of ‘universal’ as being just a true statement, is one of the basic goals of this article. This goal is pursued in Sections 2, 3, 4 and 6. Section 5 lists the ‘correct’ formulations of all logically possible universals involving two parameters (i.e. variables) that can arise inspecting contingency (i.e. tetrachoric) tables.

In the following, I will use the standard logical connectives: ¬ (not), ∧ (and), ∨ (or), → (implies), and ↔ (is equivalent to). No logical background is presupposed in the discussion except knowledge of the truth tables for these connectives, which are familiar to linguists in the field (or could be easily checked in any introductory logic textbook).

2. Hawkins and tautologous statements

Our first example comes from research on word order universals. In a familiar book, Hawkins (1983: 85–86), inspecting his sample (Appendix II and the Expanded Sample), notes that in languages with postpositions all logically possible orderings between A(djective) and N(oun) on the one hand, and N and R(elative clause) on the other, are attested, and then positsthe following universal:

(1) Postp → ((AN ∨ NA) ∧ (RelN ∨ NRel))

(Universal XVII: If a language has a Postp(osition) word order, then it has either noun before adjective or adjective before noun, and either relative clause before noun or noun before relative clause.)
Proposition (1) is indeed true relative to the sample as there is no postpositional language in the inspected sample such that it fails to have either noun before adjective or adjective before noun, and either relative clause before noun or noun before relative clause. The problem with this formulation, however, is that in principle there cannot exist such a sample that can falsify (1). The reason for this is that assertion (1) is a tautology. One can readily verify that. Thus, \((AN \lor NA)\) is equivalent to \((AN \lor \neg AN)\), because the parameter adjective-noun order is binary (only ‘basic word order’ is at issue), and \((AN \lor \neg AN)\) will be true for any language; analogously, \((RelN \lor NRel)\), which is equivalent to \((RelN \lor \neg RelN)\), will also be true, and therefore so will be the whole conjunctive consequent, \(((AN \lor NA) \land (RelN \lor NRel))\), of the implication (1). The consequent of (1) being true, the whole implication will also necessarily be true, i.e. the whole implication (1) is a tautology (see next section).

A tautology is a proposition necessarily true simply by virtue of its logical form. The fact that a tautology cannot be false due to its logical form deprives it of any empirical content and makes it uninformative; thus, instead of (1), one could equally truly claim that if a language has a Preposition (rather than Postposition) word order, then it has either noun before adjective or adjective before noun, and either relative clause before noun or noun before relative clause, or, analogously, that the presence or absence of vowels, etc. implies the same. This example thus shows the need to add to our conception of universal an additional requirement: the proposition as a whole, and its sub-expressions, must be contingent: it should be logically possible for them to be either true or false.

I note that tautologous statements, such as (1), should be demarcated from so called ‘definitional universals’, which are also necessarily true statements and have sometimes been labeled as ‘tautologous’ (cf. e.g. Greenberg 1966a: 73) ‘[...] to assert the definitional characteristics themselves is obviously tautologous.’ or Evens & Levinson (2009: 437) who speak about some universals as ‘tautological by being definitional of languagehood’). An example will make the point clear. Thus, the definitional universal ‘Every human language has proper names’ (Hockett 1966: 21) will necessarily be true as a consequence of our (possibly tacit) conception of human language as obligatorily having means to refer to human individuals, the latter being designated by proper names. Such a universal cannot be falsified by further data since if we were to find a human communication system that does not possess proper names, we would deny it the status of a human
language as failing its (possibly tacit) definition. However, on other interpretations of ‘human language’ the universal may be false, which is not the case with our example (1), which is always true, irrespectively from the interpretations of the constituent terms of the universal, but owing entirely to its logical form (as e.g. $P \lor \neg P$).

3. Jakobson and weaker statements

I proceed with my second example, which comes from phonological universals research. It has been a common knowledge that all languages have a stop, whereas some languages have and others lack a fricative. This state of affairs can conveniently be represented by a contingency (tetrachoric) table:

<table>
<thead>
<tr>
<th></th>
<th>Fricatives</th>
<th>No fricatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>No stops</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

In a celebrated book, Roman Jakobson (1941: 51) summarises this common knowledge, claiming that ‘[i]n the linguistic systems of the world, fricatives cannot exist unless stops exist as well’, which amounts to positing the implicational universal:

(2) If a language has a fricative ($P$), then it has a stop ($Q$) (symbolically $P \rightarrow Q$)

Proposition (2) is indeed true relative to the contingency table above. An implication $P \rightarrow Q$ is false in one, and only one, situation, viz. when its antecedent $P$ is true and its consequent $Q$ is false. But there are no languages in the contingency table such that they have a fricative ($P$) and lack a stop ($Q$). Still, there is a logical problem with this formulation. This problem, superficially at least, looks somewhat different from that in the previous example, since (2), unlike (1), is a contingent proposition, which could in principle be false in the situation already indicated.

The problem with (2) is connected with the notion of logical implication, so I turn to this notion. A proposition $X$ is said to logically imply a proposition $Y$ if it is not possible for $X$ to be true while $Y$ is false. ($X$ is said to be the stronger and $Y$ the weaker proposition.) Note that $Q$ (a language has a stop) logically implies (2), viz. $P \rightarrow Q$, since the truthfulness of $Q$ guarantees the truthfulness of $P \rightarrow Q$. This can be verified, for example, by
using truth tables: a proposition logically implies another proposition if in all rows in which the first has ‘True’, the second also has ‘True’. The truth table in Table 1, for instance, shows that proposition Q logically implies the proposition $P \rightarrow Q$. By way of a further illustration of the notion, the table demonstrates that proposition Q also logically implies the disjunctive proposition $P \lor Q$. Thus, Q (second column) is true in the first and third rows, and so are $P \rightarrow Q$ (third column) and $P \lor Q$ (fourth column), therefore Q logically implies both $P \rightarrow Q$ and $P \lor Q$. In contrast, $P \rightarrow Q$, for instance, does not logically imply Q, as it has ‘True’ in the fourth row, while Q has ‘False’ in the same row. Analogously, $P \lor Q$ does not logically imply Q, as it has ‘True’ in the second row, while Q has ‘False’ in the same row.

Now, Jakobson’s positing of the weaker proposition $P \rightarrow Q$ (implication (2)), rather than the stronger one, Q, has several principled shortcomings, as follows.

3.1. Redundancy

Knowing that Q is true, one can logically deduce the truthfulness also of $P \rightarrow Q$, but not vice versa; thus, knowing that all languages have a stop, we easily can infer that if they have a fricative they will have a stop, but knowing that, for all languages, if a language has a fricative it has a stop, we cannot infer that all languages have a stop. Therefore positing the weaker statement $P \rightarrow Q$ as a separate universal is redundant and uninformative. If we were to allow the positing of redundant propositions as separate universals, we will have to face the undesired consequence of having to posit as universals a theoretically unlimited number of redundant propositions. Thus, for instance, $\neg P \rightarrow Q$ (e.g. ‘The absence of fricatives implies stops’), $P \lor Q$ (‘Languages have either fricatives or stops’), $\neg P \lor Q$ (‘Languages either lack fricatives or have stops’), etc. are all logical consequences of Q, which is readily ascertainable, using truth tables (as above) or knowing that when the consequent

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>$P \rightarrow Q$</th>
<th>$P \lor Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
</tr>
</tbody>
</table>

Table 1.
of an implication is true, so is the whole implication; and that when one disjunct of a disjunction is true, so is the whole disjunction. Therefore all these, and an innumerable number of other formulae involving \( Q \) (e.g. such involving disjunctions of arbitrary length, containing \( Q \), such as \( Q \lor P \lor S, Q \lor P \lor S \lor T, Q \lor P \lor S \lor T \lor U \), etc.), as true redundant propositions, should be posited separately, which clearly is an undesired consequence.

### 3.2. Descriptive inadequacy

Frequently, universals are conceived as typologies classifying all languages into allowed and disallowed types. Despite a seeming paradoxicalness, a weaker assertion, true as it is, will nevertheless be descriptively inadequate in the sense that such an assertion does not sanction all and only the attested, or allowed, language types. Table 2, for instance, shows that the proposition \( P \to Q \) (i.e. (2) above) does not, while the stronger proposition \( Q \) (languages have stops), does sanction all and only the attested types in the tetrachoric table. As a further illustration of the descriptive inadequacy of weaker propositions, the table also shows the descriptive problem with another logical consequence of \( Q \), viz. the proposition \( P \lor Q \). Thus, the weaker proposition \( P \to Q \) (third column) is true relative to the contingency table above, presented here as column 5, because the only situation in which it is false (second row) is unattested. Despite its truthfulness, however, \( P \to Q \) predicts languages with no \( P \) and no \( Q \) (fourth row), whereas such languages are actually unattested (see fifth column, fourth row). In contrast, the stronger proposition \( Q \) (second column) is both true and predicts correctly (i.e. everywhere ‘True’ corresponds to ‘attested’ and ‘False’ to ‘unattested’). \( P \lor Q \) (fourth column) similarly is true, since the only situation in which it is false (fourth row) is unattested, but predicts languages with \( P \), but no \( Q \) (second row), which are actually unattested. In the same fashion, one can verify that other weaker propositions than \( Q \) are descript-

<table>
<thead>
<tr>
<th>( P )</th>
<th>( Q )</th>
<th>( P \to Q )</th>
<th>( P \lor Q )</th>
<th>Lg Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>Attested</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>False</td>
<td>True</td>
<td>Unattested</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>True</td>
<td>True</td>
<td>Attested</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>True</td>
<td>False</td>
<td>Unattested</td>
</tr>
</tbody>
</table>
tively inadequate. Weaker claims thus, although strictly speaking true, are overly general claims in that they predict unattested language types.

3.3. Problems with explanations

An important task of universals research is to explain the universals empirically found. This is usually done in terms lying outside the language of observation (representation) – that is, the language in which contingency (tetrachoric) tables are stated – and involves appeal to causal, mechanistic, functional or other reasons. Weaker propositions, however, do not merit such extra-observation-language explanations, for three reasons.

The first reason is that they already have a reliable explanation within the observation language by appealing to a stronger universal; thus, to the question ‘Why is it the case that if a language has a fricative it has a stop?’, the simple answer is ‘Because all languages have stops’.

The second reason is that weaker statements may erroneously suggest correlations that do not actually exist; thus, in (2) a possibly causal correlation between fricatives and stops is suggested, whereas no correlation at all, causal or not causal, exists between the two types of sounds: stops are merely present in all languages anyway, with or without fricatives. Indeed, one could just as easily say truly (and redundantly) that if a language has a speaker named John, then the language has stops, and the existence of a John causes the existence of stops.

The third reason why weaker universal statements may not require explanation is related to the case when universals are conceived as capturing attested and unattested types. As already mentioned above, weaker statements falsely predict the facts about language distribution. Any attempt to explain such weaker statements therefore will commit the logical error of ‘subverted support’ (i.e. the error of providing an explanation for ‘facts’ that do not exist). In our particular example (2), for instance, assuming the fallacious implicational relationship between fricatives and stops, one would have to try to explain why in particular, in accordance with the predictions of an implication, language types should occur that have neither fricatives nor stops, which of course disagrees with the actual facts of language distribution recorded in the contingency table.

Our second example allows us to make the second (and final) amendment to the misleading conception of universal: A universal is the strongest
contingent proposition that is true (relative to a sample). Jakobson's proposal (2) is contingent and true, but not the strongest proposition, and therefore a misformulation in the sense explicated above.

The more logically-minded linguist would note that, strictly speaking, both our examples may be attributed to logical consequence, since a tautology actually logically follows from the empty set (in which case the word “contingent” becomes superfluous in the above definition). It is thus a matter of taste which one of these alternative definitions to use.

It should also be remarked that problems with universals having to do with logical consequence have not remained completely unnoticed in the linguistic literature. Howard (1971) seems to be one of the first to pay attention to the infelicity of posited universals such as “Nasal vowels imply oral vowels” (see Section 6). Greenberg (1978a) cites approvingly Howard, and other linguists have repeatedly quoted this very example. However, previous linguistic discussions have not gone beyond this specific example and have failed to provide a more in-depth logical analysis of the general problem and its various undesired consequences. Neither have they exposed the quite common occurrence of similar misformulations in linguistic writings.

4. Further similar misformulations

There seem to be numerous examples in the linguistic literature which misguidedly posit a weaker implication, $P \rightarrow Q$, instead of the stronger unrestricted universal, $Q$. Jakobson’s example involved the case in which the consequent, $Q$, was a stop, a category known to be universally present in the languages of the world. Further phonological examples, which are simply versions of this misformulation, are given below, alongside with cases where the consequents are a primary fricative or a primary nasal (/n/), also known to be present in all (or almost all) languages.¹

¹ I note that a ‘primary’ consonant is understood here as a consonant with no secondary articulation or other complexities, stressing at the same time that knowing the particular meaning of the terms is, strictly speaking, irrelevant for the following discussion, as the infelicity of the proposed universals is demonstrated purely on the basis of their logical, and not empirical, content.
4.1. Nartey

Nartey (1979) presents a set of universals concerning stops and fricatives. From the set of those involving stops, consider the following:

1. There are at least three primary oral stops (Q).
2. If there are secondary oral stops (P), then there are primary ones (Q) as well (P → Q).
3. If there is a glottal stop (R), then there must be a primary oral stop (Q) (R → Q).

From the previous discussion, it will be clear that proposition (3) logically implies both proposition (4) and proposition (5), and knowing that (3) is true, we can infer the truthfulness of both (4) and (5). Hence, as weaker ones, the latter two propositions are redundant, falsely predict language distribution and require no explanation outside the representation (observation) language. Thus, the right thing to do to describe adequately the data is simply to list (3) as an unrestricted universal and omit both (4) and (5).

We may now look at some examples Nartey gives for fricatives:

1. There is at least one primary fricative (Q).
2. No language has secondary fricatives (P) unless it also has primary fricatives (Q) (If there are secondary fricatives (P), then there will be primary fricatives (Q) (P → Q)).
3. A language cannot have /h/ (R) unless it also has a primary fricative (Q) (If there is an /h/ (R), then there will also be a primary fricative (Q) (R → Q)).

These examples are perfectly analogous to the previous ones, and universal (6) logically implies both (7) and (8) and is hence the only universal that should be posited for the adequate description of the data and eventually explained in terms outside the representation language.

4.2. Greenberg

In a article on glottalic stops, Greenberg (1970) writes that ‘[. . .] unvoiced ejectives never exist without plain stops’ i.e. posits

1. If there are unvoiced ejectives, then there will be plain stops.
Again, the consequent of this implication, viz. the presence of plain stops, as we already know is universally true in languages. Therefore, Greenberg’s statement is a misformation.

4.3. Ferguson

Ferguson (1966) is an influential work on nasals. Ferguson conceives primary nasal consonants as plain voiced sonorant nasals with no secondary articulation or other complexities. He then posits, among others, the following universals:

(10) All languages have at least one primary nasal consonant (Q).

(11) No language has secondary nasal consonants (P) unless it has primary nasal consonants (Q). (If a language has secondary nasal consonants (P), then it has primary nasal consonants (Q) (P → Q)).

(12) No language has nasal vowels (R) unless it has a primary nasal consonant (Q). (If a language has nasal vowels (R), then it has a primary nasal consonant (Q) (R → Q)).

Here, again, the second two statements are logical consequences of the first statement, and are hence superfluous.

4.4. Maddieson

More recently, Maddieson (2009) approvingly cites Ferguson’s universal (10), and notes, correctly, that Ferguson’s universals (11) and (12) above are merely ‘extensions’ of the idea captured by his adequately stated universal (10). Interestingly, however, he writes:

If there are nasals [which is the case by (10)], then a plain front coronal (i.e. a segment representable as /n/) is almost invariably among them.[. . .] Hence there is support for the implication “if /m/, then /n/” which is implicit in F’s [Ferguson’s] observations.

Maddieson’s account amounts to the following two claims:

(13) All languages have /n/.

(14) If /m/, then /n/.

However, obviously the proposed universal (14) is logically implied by universal (13) and Maddieson commits the very mistake he notes in
Ferguson’s formulations mentioned above, as (14) is merely an ‘extension’ of the idea conveyed by (13).

Additionally, the same problem occurs with Ferguson’s claim (10), viz. ‘All languages have at least one primary nasal consonant’, and Maddieson’s claim (13), ‘All languages have /n/’, since the latter unrestricted universal actually implies the former unrestricted universal: from the fact that all languages have /n/ it follows that languages will also have, of necessity, primary nasal consonants, as /n/ is but one particular case of a primary nasal consonant. In sum, the proper description of the whole body of data is just Maddieson’s universal (13); propositions (10), (11), (12), and (14) are merely extensions of the idea already contained in (13), and positing them will have all the undesired consequences of weaker statements we described earlier.

5. Formulating universals from contingency tables

Linguists can read the strongest contingent statements off contingency (tetrachoric) tables. In an important article regarding the formulation of universals, Greenberg (1978a) discusses how linguists should derive universals from tetrachoric tables, but does not treat the matter exhaustively and does not list all logically possible universals comprising two variables (parameters), which are the most common object of study in linguistics. Neither is this done in the other standard references on universals as Comrie (1981), Croft (1990) or any other source known to me.

Table 3 sets out all theoretically possible 16 contingency tables for two binary parameters (2⁴ = 16), alongside with their corresponding strongest contingent propositions (i.e. universals). As previously, a plus indicates that the type is attested and a minus that the type is unattested. In deriving universals from such tables it is essential that all pluses and minuses in each row match observed language distribution.

Below are some appropriate natural-language (English) formulations of the universals in the table, using the same numeration.²

---
² There are of course both alternative ways of verbal formulations and alternative equivalent logical formulae of stating the same things if other logical connectives are used instead of those employed here.
1. (Tautology). A non-contingent statement and hence not a proper universal (in principle, does not exclude the possibility of a ‘linguistic tendency’, sometimes referred to as a ‘statistical universal’, to be statable).

2. $P \lor Q$ (inclusive disjunction). A language has either $P$ or $Q$ or both.

3. $Q \rightarrow P$ (converse implication). If a language has $Q$, then it also has $P$.

4. $P$. A language has $P$.

5. $P \rightarrow Q$ (implication). If a language has $P$, then it also has $Q$.

6. $Q$. A language has $Q$.

7. $P \leftrightarrow Q$ (equivalence). A language either has both $P$ and $Q$ or lacks them both.

8. $P \land Q$ (conjunction). A language has both $P$ and $Q$.

9. $\neg(P \land Q)$ (nand). No language has both $P$ and $Q$.

---

3 ‘nand’ stands for not-and.
10. \(\neg(P \leftrightarrow Q)\) (xor). A language has either P or Q but not both.
11. \(\neg Q\) (negation). No language has Q.
12. \(\neg(P \rightarrow Q)\) (non-implication). If a language has P, then it lacks Q.
13. \(\neg P\) (negation). No language has P.
14. \(\neg(Q \rightarrow P)\) (non-converse implication). If a language has Q, then it lacks P.
15. \(\neg(P \lor Q)\) (nor, non-disjunction). A language lacks both P and Q.
16. (Contradiction). A non-contingent statement and hence not a proper universal (data distribution impossible to occur in practice).

Summarizing, from 16 logically possible cases, two (viz. Nos. 1, a tautology, and 16, a contradiction) are not contingent statements, and hence not proper universals. In some cases (viz. Nos. 4, 6, 11, 13) the proper statements are so-called ‘unrestricted universals’, while in the remaining cases they are different versions of one-way or two-way implications or statements of other form.

This listing would be of practical interest, as it allows the linguist to derive ‘correct’ universals. By way of illustration, the contingency table in Section 3 is identical to the one listed here as No. 6, hence one should correctly state the universal Q. On the other hand, a ‘correct’ universal of form \(P \rightarrow Q\) corresponds to contingency table No. 5, and so on and so forth. Also, if one wants to test the correctness of a universal previously posited, one may check whether it is not the case that it is derivable from a stronger proposition than the one stated. Thus, from the above listing, one can verify\(^5\) that propositions that logically imply an implication \(P \rightarrow Q\), the favourite form of universal in linguistics, are:

\begin{align*}
Q & (i.e. \ Q \text{ logically implies } P \rightarrow Q) \\
P \leftrightarrow Q & (P \leftrightarrow Q \text{ logically implies } P \rightarrow Q) \\
P \land Q & (P \land Q \text{ logically implies } P \rightarrow Q) \\
\neg P & (\neg P \text{ logically implies } P \rightarrow Q) \\
\neg (P \lor Q) & (\neg (P \lor Q) \text{ logically implies } P \rightarrow Q)
\end{align*}

Thus, one could misstate as an implication a language distribution for which quite a number of other logical forms are appropriate. Our examples

\(^4\) ‘xor’ stands for exclusive disjunction.
\(^5\) We recollect that this verification involves checking whether all attested types, or pluses, in the row of the stronger proposition correspond to attested types (pluses) in the row of the weaker proposition.
were exclusively limited to the first shown case, viz. $Q$ logically implies $P \rightarrow Q$, and perhaps this is the most common misformulation in linguistic practice; the last two cases do not seem probable for misformulating an implication, since hardly anyone will suggest an implication between two categories in a situation in which one or both these categories are actually absent, while the other cases are not unlikely and should not be ignored as potential sources of misformulation. By way of illustration of misstating an implication, where an equivalence, or two-way implication, is more appropriate, we can cite the (near) universal $p \rightarrow k$, suggested by Maddieson’s (1984: 13), and the stronger valid (near) universal $p \leftrightarrow k$.

Another useful aspect of the listing is that it makes conspicuous the variety of universals derivable, all of which might be of linguistic interest, though they might have not attracted so far attention in current research, at the expense of unrestricted and implicational universals. Thus, in principle it would be quite interesting to know why a language has either property $P$ or property $Q$ but not both (xor, or exclusive disjunction) or why no language can have both property $P$ and property $Q$ (nand), etc. Our description thus opens new vistas for investigation of universals.

### 6. Croft and some objections

Opinions have been voiced that, contrary to what we suggest in the previous section, universals cannot be mechanically derived from contingency tables. Thus, W. Croft (2003: 57) looks at the tetrachoric (contingency) table, in fact with exactly the same distribution as that above:

<table>
<thead>
<tr>
<th>Oral vowels</th>
<th>Nasal vowels</th>
<th>No nasal vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral vowels</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>No oral vowels</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

Then, Croft writes:

Based on this table, one may formulate an unrestricted universal that states ‘All languages have oral vowels.’ However, this unrestricted universal may not be the best generalization from the perspective of an explanatory account of the cross-linguistic facts. The unattested type with no oral vowels or nasal vowels can be accounted for by the unrestricted universal ‘All languages have vowels,’ and the unattested type with nasal vowels but
no oral vowels can be accounted for by the implicational universal ‘If a language has nasal vowels, then it has oral vowels.’

Other evidence suggests that the alternative hypothesis is the correct one. First, there is additional evidence [having to do with the markedness of nasal vowels with respect to oral vowels] about the relationship between nasal vowels and oral vowels that implies that a dependency holds between the two of the sort described by the implicational universal [. . .]. Second, the unrestricted universal ‘All languages have vowels’ can be explained by the impossibility, or at least extreme difficulty, of articulating speech without vowels, whereas the unrestricted universal ‘All languages have oral vowels’ cannot be accounted for in the same fashion, since a language with only nasal vowels does not have the same articulatory restrictions.

The lesson to be drawn from this example is that unrestricted and implicational universals cannot be mechanically read off tables of attested and unattested language types. Both wider typological patterns and deeper explanations of what is going on must be appealed to in order to construct the best combination of unrestricted and implicational universals to account for the data. Above all, the choice of the correct generalization(s) to account for the constraints on possible language types is determined by the proposed theory behind the relationships between parameters. (italics mine)

Croft’s account amounts to replacing the following universal (that presumably cannot be explained):

(15) A language has oral vowels (Q).

with the following two statements (that presumably can be explained):

(16) A language has vowels.

(17) If a language has nasal vowels (P), then it has oral vowels (Q) (P \rightarrow Q).

Croft’s argument, however, fails to undermine the point that universals can be read off contingency tables, for the following reasons.

First, as we have noted (Valdés-Pérez & Pericliev 1999), his style of argument is misleading, because it blurs the noticing of a universal with its explanation. This distinction is usually respected in other linguistic contexts, and indeed, these are two separate tasks of empirical science (e.g. Simon 1977). Moreover, noticing typically precedes explanation, and these two tasks can be done by two different people, or even by computer, as we have previously shown (Valdés-Pérez & Pericliev 1999, Pericliev 2003, 2008, 2010: Chapters 4, 5, 6). Thus, if noticing universals is at stake, consulting contingency tables is not only a possible, but also a necessary move,
especially if the language samples are large or their statistical significance is computed by certain (e.g. Monte Carlo) methods.

Secondly, even if we grant that only readily explainable universals should be posited (which is apparently an overly optimistic ideal), Croft’s argument does not persuade us that statement (15), unexplainable according to him, should be substituted with (16) and (17), which are presumably explainable. Thus, let us first logically reconstruct Croft’s proposed explanation for (16). An ‘explanation’ in science is usually conceived as a set of premises and a conclusion, which is a logical deduction from the premises. Croft’s explanation for (16) then will look something like (18):

\[
\begin{align*}
\text{(18) Explanation} \\
\text{Premise 1: All languages are such that have sounds that are possible/easy to articulate.} \\
\text{Premise 2: Vowels are sounds that are possible/easy to articulate.} \\
\text{Conclusion: All languages are such that have vowels.}
\end{align*}
\]

Now, if we are to accept (18) as a correct explanation of (16), then we must accept also (19) as correct, since it only adds a further true premise (Premise 3) to (18); thus, Premise 3 is necessarily true since oral vowels are themselves vowels (the more specific, stronger, concept logically implies the more general, weaker, concept):

\[
\begin{align*}
\text{(19) Explanation} \\
\text{Premise 1: All languages are such that have sounds that are possible/easy to articulate.} \\
\text{Premise 2: Vowels are sounds that are possible/easy to articulate.} \\
\text{Premise 3: Oral vowels are vowels and hence are possible/easy to articulate.} \\
\text{Conclusion: All languages are such that have oral vowels.}
\end{align*}
\]

Explanation (19) thus correctly accounts for (15), if explanation (18) correctly accounts for (16), and hence Croft’s proclaimed ground for substituting (15) with (16), viz. the former’s unexplainability, is invalid. In consequence, his whole argument is undermined. Nevertheless, his further proposal for positing the universal (17) deserves attention, as it illustrates another confusion.

Croft’s appeal to markedness as a support to (17) is clearly misguided. I note that a markedness relation may hold only between pairs of type: nasal vowel vs. equivalent oral vowel (e.g. ţ vs. a; ţ vs. o, etc.), and not between a nasal vowel and a non-equivalent arbitrary oral vowel (as e.g. ţ vs. o, etc.).
Put differently, Croft confuses (17) with a markedness universal which should assert that ‘If a language has nasal vowels, then it has the equivalent oral vowels’. That this universal is not equivalent to (17) is seen e.g. in the fact that languages, say, having the nasal vowel ō and the oral vowel a, but not the equivalent oral vowel of the nasal vowel ō, viz. the oral vowel o, will be positive examples to (17), but counterexamples to the ‘markedness universal’. (Though not equivalent, the two statements are not logically independent either: the markedness universal logically implies (17) since if it is true that whenever a language has a nasal vowel it has its equivalent oral vowel, then if this language has a nasal vowel it is guaranteed that it will have an oral vowel.)

In this very context, we may note a comment by Greenberg. Greenberg (1978a: 50–51), inspecting the same contingency table, though agreeing that (17) is a misformulation (see also Howard 1971), argues that statements like (17) can be justified on the basis of the known causal connection between the two types of vowels (diachronically, nasal vowels come from oral vowels, not vice versa). In effect, he proposes the formulation ‘although there are languages without nasal vowels, there are no languages without oral vowels’.

In place of one misformulation (viz. (17)), Greenberg proposes another misformulation, because his claim, viewed as a universal proposition, which it probably is intended to be, amounts to asserting that for all languages (20) holds:

(20) A language either has nasal vowels (P) or does not have nasal vowels (¬P) and has oral vowels (Q) (symbolically (P ∨ ¬P) ∧ Q).

The sub-expression (P ∨ ¬P), however, is clearly a tautology (a non-contingent sub-expression), and hence (20) as a whole is a misstatement. Removing the redundant sub-expression (P ∨ ¬P), (20), in effect, asserts just Q (all languages have oral vowels), no more and no less. It is unclear in what way the new formulation (20) can justify an eventual causal link between oral and nasal vowels, as this would require an implicational statement holding between the two categories. Additionally, it may be observed that Greenberg confuses, just like Croft, proposition (17) with a markedness universal.

It is worth noting that the misconception of statement (17) as a markedness universal, while it is neither a markedness pattern, nor a proper universal seems to persist in the most recent discussions of universals. Thus,
for instance, Evans & Levinson (2009: 438) write ‘Statement (12a) [= (17)] essentially expresses the markedness (or recessive character) of nasal vowels’.

Despite all these infelicities in formulation, there is apparently a grain of truth regarding the existence of causal link, as maintained by Greenberg and Croft. In order to clarify this issue, we need to answer the following questions: (i) if not from Croft’s contingency table, from what data set can we eventually read off the markedness implicational universal nasal vowel $\rightarrow$ respective oral vowel?, and (ii) how can we interpret this implication as indicating a causal connection in the reverse direction, viz. oral vowel causes respective nasal vowel?

The answer to question (i) has already been implicitly given in our formulation of the markedness universal, as holding between a nasal vowel and its equivalent oral vowel (i.e. nasal vowel $\rightarrow$ respective oral vowel). The corresponding contingency table supporting this implicational universal then should contain both categories of sounds and amounts to the following distribution:

<table>
<thead>
<tr>
<th>Equivalent oral vowels</th>
<th>Nasal vowels</th>
<th>No nasal vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>No equivalent oral vowels</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

The distribution is with one unattested type (presence of nasal vowels and absence of equivalent oral vowels); it is captured by a universal of implicational form (as seen from our list of all possible universals between two parameters), and there is no longer the need to misinterpret a distribution corresponding to an unrestricted universal as an implication just in order to capture the previously surmised causal link between the two categories of sounds.

To answer question (ii), we note that, in principle, an implication $P \rightarrow Q$ may be interpreted causally either as: (1) $P$ is the whole cause of $Q$ (i.e. $P$ is a sufficient condition for $Q$), or (2) $Q$ is a part of the cause of $P$ (i.e. $Q$ is a necessary condition for $P$). Apparently, the correct markedness implication nasal vowel $\rightarrow$ respective oral vowel cannot, from a linguistic perspective, be sensibly interpreted in the first way since we know that nasal vowels do not predate their oral equivalents; therefore, we are left with the second alternative. Under this interpretation, things come to their places: oral vowels should always be present (diachronically) for the respective
nasal vowels to occur, but they are only a part of the cause, and there are other causal factors at play, in the particular case the sound $n$ following the oral vowel (i.e. oral vowel + $n$ cause respective nasal vowel), and possibly some other factors. In this context, Greenberg’s (1978a: 51) remark that oral vowels cause (the respective) nasal vowels holds in the sense that oral vowels are a part of the cause for their respective nasal vowels.

7. Conclusion

In the last decades, we have significantly both increased our understanding of the notion of universal and have accumulated a large number of universals in phonology, grammar and semantics (usefully collected e.g. in the Konstanz Universals Archive at http://typo.uni-konstanz.de/archive/intro/). A more careful look at the literature and the collected universals, however, reveals that the idea of universal is not an easy one and needs further logical analysis, if we are to posit linguistically adequate universal statements. This article can be viewed as only one attempt in this direction. Here I proposed some amendments to the definition of a universal, and enumerated all possible ‘correct’ universals holding between two parameters that can be derived from contingency tables.

References


Author’s address:
Institute of Mathematics and Informatics, bl.8
Bulgarian Academy of Sciences
1113 Sofia
Bulgaria
peri@math.bas.bg