In search of universals

Article · January 2006

DOI: 10.1017/CBO9780511618215.005

1 author:

Ian Maddieson
University of New Mexico

IN SEARCH OF UNIVERSALS

Ian Maddieson
University of California, Los Angeles

ABSTRACT

The purpose of this paper is to bring together some reflections and a few results from work extending over a number of years on phonetic and phonological universals. In different sections, a growing integration of work on language universals with work on language evolution is noted, the importance of the construction of proper language samples for the empirical discovery of universals is emphasized, and a rough distinction between 'mechanical' and 'ecological' factors accounting for universals is made. Four basic universals are then mentioned: language uses sound, language is oral, language shows sequential variation, and paradigmatic contrast. Other phonetic universals are grouped and briefly discussed as concerning segments, phonetic detail and variability.

1. WHAT ARE UNIVERSALS?

In everyday life we are struck by the *differences* between languages. When we overhear a conversation between people speaking a language we have no knowledge of, we find it completely impenetrable. There is something disturbing, even sinister, about a language you don't understand at all. The fear this engenders is part of what tempts authoritarians everywhere to ban the use of unfamiliar languages, and enforce the use of the familiar. Unfamiliar languages are also frequently looked down on as unstructured and incapable of serving to convey more than simple messages. Linguists confront this fear and contempt, and it is part of the value of the mainstream traditions of linguistics as a humanistic discipline that it insists on the equal worth of every language. Every language is considered equal, in the sense of meriting equally serious consideration, akin to the way that all people are declared to be "equal in dignity and rights" in the Universal Declaration of Human Rights [29]. In contrast to lay emphasis on linguistic differences, for linguists the demonstration of *similarities* between languages functions at the same time as one of the ways by which the proposition of equality is supported and as one of the reasons why linguists hold this view. We feel justified in constructing general theories within which the peculiarities of individual languages can be described, rather than taking it as our task to build quite separate theories for different languages. The nature of some of these similarities, some methodological considerations involved, and some of the factors which may account for similarities will be the topics of this paper. Naturally enough, the focus will be on examples from the domain of phonetics and phonology, but methods and types of explanations may be parallel in other subfields of linguistics.

There are three very different takes on the nature of linguistic similarities that I'm familiar with. In one, the perspective is that the problem of describing all languages can be attacked using similar methodologies, yet the linguist makes no commitment that the similarities are ones of substance rather than of method. Those familiar with work in the tradition of 'prosodic phonology' associated with the name of J. R. Firth may recognize this trait in that work. Its practitioners would surely assert that the phonology of any language can be described using the concepts of phonematic units and prosodies, and phonetic exponents of these categories. However, the nature and content of the categories is so dependent on language-particular interpretations by individual linguists that it makes no sense to compare descriptions of different languages in this model. One of Firth's main concerns was to find a way round what he saw as the excessive inflexibility of competing linguistic models, which threatened to mask differences not only between languages but equally between different subsystems within the same language. By providing flexible descriptive tools he permitted an escape from this rigidity, and accepted the fact that comparability of descriptions only existed at the level of the tools used. Despite being one of Firth's targets, some linguists of the heyday of American structuralism maintained a somewhat similar view that language descriptions were only similar at the level of their methodology, a view jocularly dubbed 'hocus-pocus' at the time.

An opposite extreme of thinking about language similarities is the postulation of an innate Universal Grammar which underlies the grammars of all individual languages and structures the process by which they are acquired in childhood and how they are represented in the adult brain. The existence of such a Universal Grammar is a fundamental assumption within the generative tradition of linguistics, and the discovery of its properties is taken to be one of the main objectives of the science. Universal Grammar in this sense is sometimes discussed as if it was a rather mystical immanent entity — a 'language organ' which appeared somewhat mysteriously at some point in human ancestry with no precursors — but Pinker [23], Deacon [5] and other scholars have shown how the idea can be set comfortably in the mainstream of thinking in evolutionary biology. From this perspective, all languages are similar in certain important ways and cannot be otherwise because humans come 'pre-set' with principles of language design. It is, of course, not necessary to assume that all similarities observed between languages should be attributed to innate adaptation for language. Some shared characteristics could, for example, be due to properties of the human organism that are not specific to language, or to properties of the external environment in which language is used. This point seems sometimes to be missed, by both adherents and opponents of the Universal Grammar hypothesis.

A third perspective is found in the tradition of work that actually bears the title of linguistic universals, or is discussed in terms of universals and typology. An underlying assumption here is certainly the hypothesis that all languages are constructed on similar principles, or at least that the range of their variation is circumscribed within quite definite limits. However, the explanation of any shared properties or limits on language variability is the subject of further specific hypotheses which may be based on a variety of different factors. Although special innate adaptations for language are not excluded, it is clear that many of the linguists working in this tradition would prefer to find their explanations elsewhere. Work in this tradition tends to have a strongly empiricist orientation; universals can (and should) be inferred directly from comparing descriptions of different languages [7, 4]. In order to do this, it has to be assumed that the descriptive terms and constructs used to describe properties of different languages can be sufficiently well equated so that generalizations can be meaningfully made, a quite problematic issue which will not be discussed here.

The meaning of the term 'universal' when used in this context needs to be clarified, since in only relatively few cases is this work focused on properties that are literally shared by all languages. Rather, the interest is in trying to define the limits within which variation among languages is confined, and to study the relative frequency of various traits — a reasonable hypothesis being that a more frequent pattern may represent in some sense a 'better' design feature for a language than one which is significantly less frequent. 'Universals' are thus essentially distributions of individual properties and of patterns of related properties. What justifies the term 'universal' is that these distributions are being studied over the universe of known languages. And studying the universe of known languages offers one way to approach the goal of understanding what the essential nature of human language is, that is, what is to be found in the universe of possible languages.

An interesting convergence between the generativist Universal Grammar tradition and empiricist work on language universals has been apparent in recent years in work on language origins (see Intro to [9]). A combination of several factors, including new lines of thinking about emergent organization, new techniques for studying human brain activity, and greater sophistication in handling experiments with young children and animals, has reinvigorated an area of research which not long ago was largely neglected as sterile if not downright unrespectable. Any innate endowments of modern humans that make language possible and the attributes manifested in language itself cannot be other than different reflections of the evolution of our species. There are many aspects to consider — biomechanical and cognitive development, environmental and social setting, purposive action vs emergent self-organizing patterns and so on. UG theorists may be more likely to emphasize cognitive elements and empirical universalists more likely to emphasize biomechanical, environmental and social factors, but there is a good deal of overlap. Some of the comments below reflect an attempt to place universals research in an evolutionary context: it will be apparent that I am on the side of those [e.g. 28] who consider that it was indeed communication of information (including 'social' information about group solidarity) which was **the** evolutionary advantage of the development of language.

2. HOW DO WE FIND UNIVERSALS?

It follows from what was said above that knowledge of universals comes from surveying what is found in known languages. But, as has often been pointed out before, actually conducting such surveys is laden with both practical and theoretical difficulties. In the first place there is the sheer impracticality of literally surveying the entire universe of known languages, in whatever way that task is conceived. This practical reason is the first thing that drives the universalist to turn to using a **sample** of languages. The laziest way to construct a sample is just to think of the languages you know about, or can find descriptions of in the books and journals in your office. Let us charitably call this an informal sample. It's a part of human nature to generalize from the familiar, and linguists are as prone to do this as anyone. Implicitly or explicitly they may extract 'universals' from such informal samples. Thus, for many linguists Standard Mandarin, being the most familar example of a tone language, comes to serve as the tacit prototype of one, even though wider surveys show that a tone system with only one level tone is typologically very unusual.

Clearly, an explicit sample is preferable to an informal one, since the basis for any conclusions drawn from it can be seen and verified by another linguist. But explicitness alone is not enough, as the sample also needs to be carefully constructed to avoid bias if the results are to be meaningful. Language samples are hard to construct, however. This is not least because of the difficulties of determining and consistently applying appropriate criteria for deciding whether two related speech varieties should be considered two different languages or two dialects of the same language. For example, the 1992 edition of *Ethnologue* [8] distinguishes seven modern Romance languages spoken primarily within the borders of mainland France, whereas Posner [24] is reluctant to recognize more than two, each encompassing more dialectal variation. If you want to compile a sample of fifty languages, how many eligible targets for inclusion are there in France? This makes language sampling very different from conducting a public opinion poll asking 2000 people whether they prefer Pepsi or Coke. Human beings, unlike languages, have rather well-defined boundaries, so they can be enumerated and a random subset easily picked.

The appropriate structure of a sample will naturally vary according to the purpose for which it is designed. A sample to examine the global frequency of some property or pattern of properties must be structured to sample the whole universe of known languages, whereas one designed to examine the strength of association of property *x* and property *y* may be limited to languages in which at least one of these properties occurs. The UPSID sample, originally described in [20], subsequently expanded, and now consultable on the web [25], is designed to study global patterns in the structure of phoneme inventories and the frequency of occurrence of particular segment types or classes in those inventories. It was constructed using a sampling grid based on a genetic classification of the world's languages, and this for a very good reason.

All linguists working on universals, whether their approach is to assume a Universal Grammar or to look for universals empirically, intend to be describing characteristics that are general attributes of being a language. But we know that similarities between languages can have a historical origin, whether due to descent from a common ancestor or from contact; in fact it is precisely the presence of similarities which makes it possible to reconstruct the historical connections among languages. The fact that the word for the warmest season of the year is "summer" is English and that synonyms with similar shapes are found in the other German languages is the result of some arbitrary association of sound and meaning dating back to Proto-Germanic, and carried down to the modern languages with relatively little change. Highly specific similarities, such as here in the shape of particular words, are clearly not candidates for universal status. But what we would consider typological patterns can also also continued down through language families, and some proportion of these patterns may be the result of factors just as arbitrary as the choice of making "summer" mean what it does. For English words like "star", "stone" and "still" we can also find cognates in other Germanic languages. Here we may be tempted not just to note the lexical similarities but to conclude that the initial clusters in these words provide evidence for a generalisation that sibilant fricatives have a special freedom to occur before stops in onset position. Yet the apparently privileged occurrence of sibilants in two-obstruent clusters may be just another inherited peculiarity of Germanic, and more generally of Indo-European languages: other languages give wider freedom to lateral or velar fricatives.

Moreover, it should always be remembered that the selection of languages available for examination today is the cumulative result of all the accidents of the past, which have lead to the extinction of Sumerian, Etruscan, Tocharian, Beothuk, Ningi, Island Carib, Tasmanian, and many other languages, including many thousands of which we know nothing at all, the reduction of some language groups to a handful of survivors or isolates, such as Basque, Burushaski, Ket, Hadza, Albanian, Greek and Armenian, and the explosion of numbers of languages in other groups such as Austronesian or Benue-Congo. Some arbitrary features will thus have had the opportunity to propagate in a large number of daughter languages, while other patterns which existed in the past will be found nowhere today. The raw pattern of relative frequencies is thus not simply a function of how 'natural' (for want of a better word) a particular pattern is, but the outcome of the interaction of naturalness, arbitrariness, and accidents of history. Exactly parallel problems arise in relation to contact. Some of the features spread by contact may be arbitrary ones, and languages have quite unequal chances of participating in contact situations that result in typological-looking change. There is no way to completely avoid the risk of error that these facts create; the principal ones being that of interpreting common patterns as being 'natural' when they may just have been lucky; and of thinking that a typology based on existing languages is exhaustive when other possible patterns just happen not to be around at the moment. However, constructing a sample using **genetic distance** between the languages as a major criterion for inclusion reduces the risk of over-valuing shared inherited features. The majority of carefully drawn-up explicit samples follow this criterion in one way or another. Nonetheless it is still prudent to explicitly examine any putative universal in the light of its distribution across genetic groupings, and to review that the patterns are not plausibly explained by contact.

Sampling is also required for another reason. Many of the observed patterns in universals research consist of relationships between properties. In such a case it is necessary to show that the properties that appear to be related are not just associated by chance. This can best be done by examining the statistical patterns of an explicit sample. I will illustrate this with an intentionally silly example. In the expanded UPSID sample of 451 languages all the languages with clicks in their consonant inventories also have bilabial nasals. Should we conclude that having clicks specifically entails the presence of a bilabial nasal? About 95% of the languages have bilabial nasals; there are only five languages with clicks. Given this frequency of bilabial nasals, we can calculate that for any random subset of five languages from the sample there is about a 77% chance that all five will include a bilabial nasal. The association of clicks and bilabial nasals in inventories is therefore overwhelmingly likely to be due to chance, as any set of five languages — regardless of whether any clicks occur — is far more likely to all have bilabial nasals than not. It is easy to be misled by the co-occurrence of properties in one or more familiar languages into thinking that their co-occurrence is 'natural' but we can only be confident of such an interpretations when the data has been assembled in a way that allows chance association to be ruled out.

We would in any case have been unlikely to consider that the co-occurrence of clicks and bilabial nasals taken as an isolated fact in this way was likely to be of importance. We need a reason to think that any facts we observe have importance. Confidence that a particular phenomenon is reasonably viewed as a universal comes finally from the ability to formulate a theory drawing on factors that are appropriately universal in their scope and which provide a reason for its existence. The particular shape of the English word "summer" may be arbitrary, but the fact that all its component sounds are produced in the mouth — rather than, say, by stamping the feet — is open to a universally general explanation if we can show that oral sounds possess advantages for the construction of a human communication system that foot-stomping and other alternatives don't have.

It is perhaps also worth re-emphasizing here that success in acquiring speakers has nothing to do with 'fitness' as a system of communication. The most widely spoken languages did not achieve their dominance because they are better equipped than others, but because of the political and demographic success of the communities using these languages. English has no greater claim to be representative of 'language' than Tewa, Ogoni or Manam. Forcing the inclusion of all the major languages in a sample, or worst of all constructing a sample that consists of only the most widely spoken languages, risks introducing some serious distortions, particularly if one wishes to count the relative frequency of properties. For example, many of the most widely spoken languages, e.g. English, Arabic, French, and Russian, are very deviant from the most common pattern of segment sequencing, and many of these widely-spoken languages also have particular kinds of comparatively rare sounds in their phonological inventories, such as the [ ] (and its variants) of standard French, and the [ ] of standard American English in words like 'bird'.

Conducting surveys for universals involves other practical considerations besides sample design. Many of the languages in the world today have only been described in quite sketchy fashion. The investigator has the options of considering collecting information with first-hand field-work (usually impractical), or working with information collected by others (which may not include the facts being sought). In practice therefore it is what one might call the 'lowest denominator' theoretical elements that are most readily amenable to broad cross-linguistic study. This is one reason why, in the domain of phonological universals, a great deal of attention has been devoted to analyzing the structure of phoneme inventories and patterns of syllabification, since even the relatively modest publications which are all that is available for most languages usually include a phonemic level of analysis and some information on syllable structure. But even here there are significant gaps; for example, we cannot say how common contrast of tongue root position is in phonemic vowel systems since many descriptions do not consider this possibility, while others use the feature +/- ATR to make up for deficiencies in the features available to distinguish vowel height or other properties. For many other aspects of phonological structure, for example intonation and phrasal structure or patterns of segmental alternation, it can be hard to assemble a sufficiently large database or to find descriptions that are commensurable. For more detailed phonetic aspects of language, such as acoustic variation in different contexts, there is little data of any sort on most languages, and the risk of drawing misleading conclusions from inadequate data is correspondingly great.

3. WHERE DO UNIVERSALS COME FROM?

There are certainly two easily distinguishable kinds of constraints which produce observable universal phonetic and phonological patterns, and perhaps it would be useful to distinguish more. The two I have in mind were labeled 'mechanical' and 'ecological' in an earlier paper [21]. The names are chosen to be suggestive rather than to be strictly construed in narrow senses of these terms, and it is not claimed that there is any strikingly novel insight behind their choice. Rather, this is an attempt to present in a relatively clear and transparent fashion a distinction which it seems important to draw. 'Mechanical' constraints are those which are necessarily so, for example because of physical laws or limits on what human being are capable of doing (which may be ultimately explicable in terms of physical laws too, although it is often convenient to bypass this explication.) 'Ecological' constraints are those which direct selection within the range of what is possible in certain directions rather than others.

If we think about vowels, for instance, it is clear that there are 'mechanical' limits on the range of articulatory and spectral diversity that it is possible to produce. There is a limit to how wide we can make the vocal tract at any point (although the particular distances will vary according to individuals), and if we narrow it beyond a certain degree it becomes no longer possible to produce a vowel. These and other restrictions on the geometry of the vocal tract mean that it is possible to delimit a 'vowel space' with articulatory, acoustic or perceptual coordinates. Vowels will necessarily fall within this space. For example, if we consider a very simplified acoustic representation of vowel space in terms of the two lowest formant frequencies, a given speaker will be able to produce a certain highest first formant frequency and a certain highest second formant frequency. Although it can provide many hours of harmless entertainment for a phonetician in search of distraction to see how close one can get to the impossible, there is no way that these two values can be produced at the same time. Hence, within the rectangle defined by the minimum and maximum values of F1 and F2 shown schematically in Figure 1 there is a smaller space that contains all the possible combinations of F1 and F2 values. This space is represented as bounded by the curved line inside the rectangle. The excluded area in the lower left of the figure (lightly stippled) represents all those combinations of values of high F1 and high F2 that cannot be attained. (There is also an excluded area in the lower right of the figure, but for a quite different reason. By definition, the frequency of F2 cannot be lower than the frequency of F1. This area represents those combinations of values where this definition would be violated.)

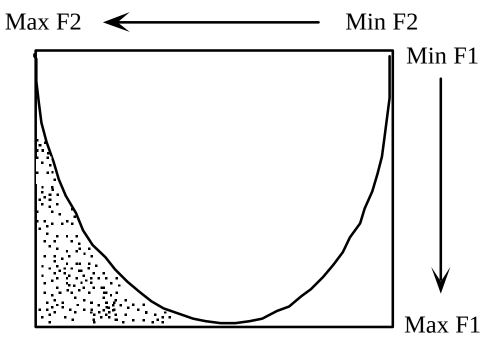

Figure 1. Rectangular space defined by range of first two formants, and interior limits.

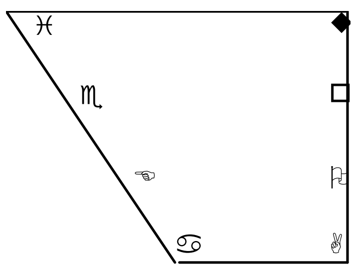

Figure 2. The standard IPA vowel quadrilateral.

The configuration of the standard IPA vowel quadrilateral, shown in Figure 2 with cardinal vowel symbols conventionally placed, represents a similar kind of understanding that there are limits to the vowel space, and that these limits have some interesting assymetries. The shape of the figure implies such constraints as "a vowel with the quality [ ] cannot be assimilated to a fronter vowel without also being raised." This example, of courses, begs many questions about the nature of the parameters which define vowel space, but the details are not crucial to illustrating that there are some 'mechanical' limits to respect.

But there are other factors that are clearly not of this necessary sort. An illustration can be provided by considering the arrangement of vowels in sets of vowel phonemes. Considered cross-linguistically, vowel systems display a number of quite clear trends which have been widely studied [13, 26]. More languages have a system of five distinctive vowels than any other number. But having five vowels is clearly not a necessary pattern as so many other possibilities are also found. The common arrangements of five vowels place them so well-dispersed in phonetic vowel space that it is evidently not a case of running up against limits on human ability to resolve differences. Such 'preferred' arrangements in contrasive sets must therefore be due to a different kind of factor or factors. These are viewed as 'ecological' partly by analogy with, say, a predator's choice of foods within its ecological setting — when times are good and the animal is strong, it will go after its favorite foods, but less desirable foods will serve the purpose of keeping it alive just as well. What the animal actually eats is thus a result of the interaction of variables affecting food supply and the animal's strength. There are some single scales to consider (e.g. food preference), but crucially any outcome requires considering how several factors interrelate. And it is this aspect of considering factors in context that is the essence of an 'ecological' perspective. But the term is not just chosen for the analogy; it seems likely that many of the choices made in language design do result directly from the ecological setting in which language developed or in which it is now used.

The distinction of 'mechanical' versus 'ecological' factors is not as sharp as the foregoing discussion of a couple of prototypes related to vowels suggests, and it may perhaps be better to think of a scale from 'more mechanical' to 'more ecological'. And it is certain that we do not always know where to place effects between these poles. There are many things we still do not know about the mechanisms of speech production, and about the auditory system, and we know less about perception and the representation of linguistic knowledge in the brain. There are surely limits imposed by the 'hard-wiring' of elements in the auditory and perceptual systems, which put bounds on the systems' capacity to resolve difference between, say, two vowel tokens. But we cannot always experimentally see where such physical limits lie, and how they are distinguishable from more functional limitations, shaped by, for example, the learning of some particular language.

4. WHAT ARE SOME UNIVERSALS WE KNOW ABOUT

In the remainder of the paper I will talk briefly about a few selected universals or classes of universals, starting with some which touch on very basic aspects of language. One popular pseudo-universals will also be mentioned, to illustrate that it is also very easy to construct explanations for pseudo-facts.

4.1 Language uses sound

The most basic phonetically-related universal is that all human ethnic groups make use of a language which is based on sound [cf 16]. Another type of language exists which uses our sense of sight, that is, the transmission of a message involves an encoding of the signal in terms of patterns of light (created by movements in space). These languages, like good small children in the English saying, are seen and not heard. I will call them light-based languages to contrast them with sound-based languages, but, of course, they are more ordinarily called signed languages. The light-based languages known from around the world seem invariably to be the preserve of particular subcultures, most often composed of individuals with impaired hearing and their close associates, who form a group within a larger society using a language based on sound.

Why is it the case that sound-based languages are universal across societies in this way? We know that the spatial-visual medium is fully capable of supporting languages adequate to all the demands people make on them, and that such languages are as complex in their structures as sound-based languages. Yet nowhere in the world do we find a family of languages used by hearing people and employing a light-based system for its primary organization, nor do we have evidence to suggest that such a family ever existed This is perhaps surprising in view of the fact that, for humans, sight is considered the most sensitive of the various senses through which we acquire information about the world around us, and our sight is certainly more acute than our sense of hearing. And, in view of the well-established fact that visual cues significantly aid in comprehension of languages transmitted by sound [e.g. 18], we might also have expected to find language families in which there is primarily light-based encoding but with some secondary support from sound.

It seems appropriate to consider the relative balance of advantage between sound and sight in very ecological terms, since it is clear that what is involved is a strong **preference** for sound-based systems, not complete uniformity. Light-based systems have an advantage over sound-based ones in that they can function well in situations where there are other sources of sound present which would interfere with a sound-based system. The world was undoubtedly a much quieter place before the invention of internal combustion engines and amplified music, but there are many natural sources of sound which might have impeded our ancestors' efforts to communicate. With a light-based communication system you can have a conversation in a howling gale, or standing next to Niagara Falls, or while on an egg-stealing expedition to a colony of thousands of screeching seagulls. A slightly similar advantage is that the medium does not interfere with itself. Within a group of people assembled together, several different conversations can be going on at the same time with relatively little interference. A light-based system also does not exclude that fraction of the population with impaired hearing (although of course it does exclude that fraction with impaired vision). Despite these advantages, it is not hard to see where a sound-based system scores over a light-based one. This is precisely because the light-based system requires the presence of light and for the sender of the message to be in the line of sight of the receiver. Using sound, messages can be sent and received in the dark, and between people who are out of line of sight or who just happen not to be looking at each other. Using sound one can communicate in an unlit hut or cave, or outdoors on a night when the moon is new or veiled by clouds. People walking in single file along a path can communicate with each other without the need to turn around. You can call back a child who has wandered away from home, or warn someone that a scorpion is about to sting their foot regardless of where their eyes are turned. The advantages of a sound-based system over a light-based one seem decisive.

Just for completeness, we might briefly consider whether the other senses, apart from hearing and sight, could serve as a basis for a language. We can probably eliminate smell and taste at once since humans cannot generate a range of tastes and smells on demand out of which to create a complex signaling system. Our sense of touch appears sensitive enough to use variation in localization, strength of contact, and movement as well as sequencing to form a language system, but one drawback is the necessity for message-sender and receiver to be within touching range. While this is not always unwelcome, it does limit the distance of possible transmission rather narrowly. Another disadvantage is the inability to use this method to communicate with a number of people at the same time. Touch-based communication systems have found a role in communication for individuals with complex disabilities, but outside this special situation they could not compete with sound-based ones.

4.2 Language is oral

Not only do all human societies have a sound based language, but all these sound-based languages use only what I will call rather loosely 'oral sound' — ones made in and around the mouth and using what we generally call the vocal tract for their production [cf 11]. It's easy enough to generate sound at will in a variety of other areas. We can click our fingers, clap our hands, slap our thighs, stamp our feet, squeeze the air out from under the armpits, and so on. Why is it that our sound-based languages use only oral sound and not other possibilities, either alone or in combination with oral sound? We may speculate that, compared to using the various other possibilities, when the mouth is used it's easier to produce a suitable range of variation in sound with appropriate continuity. But it is hard to conceive how we could persuasively demonstrate these points to be so, since our available experimental population has such extensive prior practice in oral skills, and so much less in, say, distinctive thigh-slapping. A more readily demonstrable advantage of oral sounds is that considerable amplitude can be generated with comparatively little expenditure of energy. This is, as our military brethren might say, thanks to the use of air power. Passing the air we breathe through constrictions which require only very small muscle movements for their creation produces much louder acoustic output than can be produced by movements of similar magnitude elsewhere. Another definite advantage for oral sound is that when communication occurs in close proximity, that is, when the visual channel is used in support of the auditory one, attention can be directed to a limited area, i.e. the face. This is an area which gives a great deal of information about such things as emotional state and direction of visual attention, which can guide and influence the linguistic interaction. Another advantage of using only oral sounds is that this leaves the hands — the most obvious alternative means of producing sound — free to do something else at the same time, for example, to carry a child or tools, or to work or demonstrate a craft. This advantage also counts in comparison with a light-based system using the hands, as those we know all do.

Of course, the oral channel is employed for more limited sound-based communication systems across a very large range of terrestrial vertebrate life forms besides humans. Some general capacity to make use of such signals is evidently very ancient but there is not one single evolutionary continuum involved, and the oral channel can be dispensed with when no longer appropriate, as in marine mammals. This raises some very interesting questions concerning precisely what aspects are directly biologically specified, what aspects are best understood as due to the self-organization which emerges in complex systems without being specified [10, 2], and what aspects remain free. It's obvious that humans inherited some use of vocal signals from our pre-human ancestors, even if in humans this limited inheritance has developed into a very different kind of tool. The average human is born equipped with some clever circuitry to help figure out the connection between speech noises heard and the actions in the vocal tract required to produce a near-equivalent. But does this machinery also include something which determines which oral sounds are possible linguistic sounds and which are not? It is my intuition that some kinds of sounds, such as pure whistles or the Donald Duck kind of noises you can make by squeezing air into the cheek cavity through a lateral constriction, are inadmissable as linguistic sounds, while other sounds that are equally unknown to occur in any actual language — for example, linguo-labial trills — are possible, if unlikely, linguistic sounds. Is this intuition founded on anything more than the ability to describe the second class but not the first in the technical terminology of phonetics? We'll revisit this question again briefly later.

4.3 Language shows sequential variation

To construct a useful signalling system out of sound, there must be some differentiation between different parts of the signal in time. It appears that a basic organization of this differentiation of sound in all (spoken) languages consists of an alternation between louder and quieter levels of sound, with a period not too far from 150-200 ms (by informal survey!). This is what creates the notion of organization into syllables. Although any modulated sound stream will inevitably contain peaks and valleys of amplitude, there are ways in which one can be organized without this particular periodic pattern, for example, with long plateaus or with sequences of steps upwards or downwards. In fact, among the most discussed phonological universals are those which relate to sequence of consonants in syllable or word initial and final positions, and the broadest generalizations drawn are exactly those which organize complex onsets and codas into crescendo and decrescendo patterns respectively [7]. The effort devoted to explaining these sequences perhaps distracts attention from the fact that such clusters are strikingly rare. In a survey of the lexical frequency of syllable types in a sample of 30 languages drawn from around the world, I found that 21 of them (70%) had no consonant clusters or a negligible number of them (frequency of less than 1%). In every one of these language at least 85% of the syllables counted had simple or zero onsets. Although many details are glossed over here, it is clear that this kind of syllable structure is most consistent with a fairly regular wave-like alternation of amplitude peaks and valleys. The occurrence and the timing of this pattern have been suggested to be related to a natural frequency of the jaw [19], which can be approximately equated with a comfortable mastication rate. However, as our parents rightly insisted, this does not mean that we ***were*** meant to talk and eat at the same time.

4.4 There is paradigmatic contrast

If moderately rapid alternation of amplitude peaks and valleys is a natural characteristic of speaking, perhaps as a result of synchronizing to a rhythm of the jaw, a natural segmentation in time emerges between those elements which form the amplitude peaks and those which form amplitude troughs, yielding essentially C and V positions. But unless a variety of qualitatively different sounds are available to fill these positions, it is not possible to create a vocabulary and to build the other components that make it possible to encode in sound a usefully rich enough range of messages [27]. The elements that go into these positions have been studied from several different universal perspectives. Two of these concern overall patterns of segment inventories, and the overall range of possible segments. The universal set of possible segments is more clearly constrained by 'mechanical' considerations (at least in part), but patterns in the structure of segment inventories have usually been viewed as emerging from interactions of factors viewed in an 'ecological' perspective. There are two old favorites in this realm, ease of articulation and ease of discrimination. The joint maximization of these factors, or related ways of expressing the importance of both motoric and auditory/perceptual encoding, have been implemented in models using an adaptive strategy. These fairly successfully predict the preferred patterns in vowel inventories of different sizes as observed in large cross-linguistic surveys [13, 26]. More intriguing still are the successes in replicating these predictions when similar factors are attributed to sets of virtual robots in an interacting population [1, 2]. In these latter experiments, different roles of the factors in acts of 'speaking' and 'listening' can be represented. A general feature of all these models is that they investigate the structure of systems with a given number of elements — for example, given a system of five vowels they will select the particular arrangements of these in vowel space which cause the least offense to articulatory ease and discriminability. They approach the question of the optimal solution for a vocabulary of size n, but not the issue of what is an appropriate size for n, or "How much contrast is enough"?

Total segment inventory size clusters around a mean of 27 segments, but how large a vocabulary of distinct items can be generated in actual languages is in fact relatively little determined by the size of their segment inventory, given how much other variables, such as suprasegmental properties and word structure, contribute. Nonetheless, an apparently widely-believed pseudo-universal holds that segment inventories display internal compensatory strategies to constrain 'excessive' contrast. For example, an idea that seems as durable as the Great Eskimo Vocabulary Hoax is that a language with a small number of distinctive vowel contrasts will necessarily have a large number of consonants (and possibly vice versa although this seems to be less often insisted on). We can re-visit this idea from the vantage point of the expanded UPSID sample, which gives a good idea of how many consonants typically occur and what the range of variation is. As the histogram in figure 3 shows, considerably over half the languages in the entire sample have consonant inventories containing between 15 and 24 segments. The mean falls at 22.5. A curve approximating the normal distribution is superimposed. The range of values shown on the figure is limited to 0-50 for comparability with figure 4, but the percentages and the curve shown are calculated over the full range (only 6 languages in the set have more than 50 consonants).

The most common size of a vowel system is five vowels, so those with fewer than five may be considered to be small. Figure 4 gives the histogram of the consonant inventories of just the 48 languages in the sample with four or fewer vowels. An even higher proportion of these inventories fall within the limits 15-24, and the mean is 22.4. None of these languages has more than 50 consonants. There is absolutely no tendency for languages with few vowels to have larger than average consonant inventories.

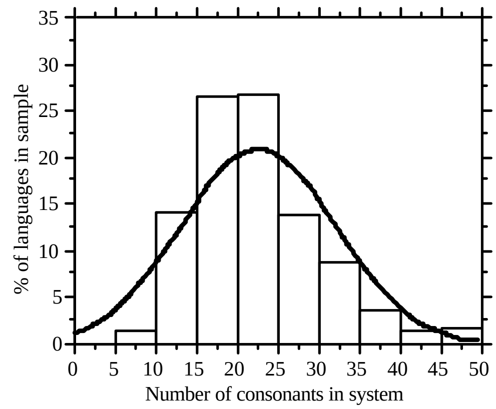

Figure 3. Histogram of number of consonants in expanded UPSID sample. n = 451 languages.

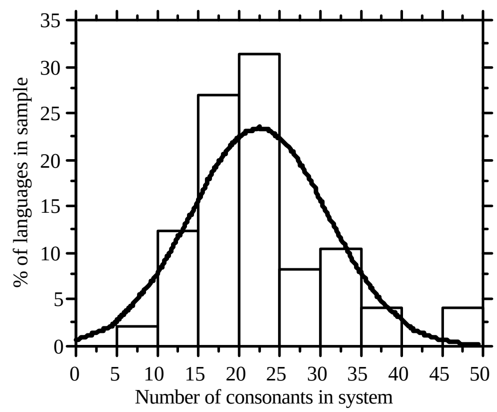

Figure 4. Histogram of number of consonants in languages with less than 5 vowels. n = 48 languages.

4.5 There are constraints on segments

Perhaps the most discussed type of phonetic or phonological language universal relates to segments. Here, there are both universal frameworks to consider, and patterns of relationship between segments. There is space for only very brief commentary here. An attempt to survey all the segmental constrasts known to occur was offered in *The Sounds of the World's Languages* [12], with the overt suggestion that this provided a good approach to the universe of possible contrasts. Phonological feature systems in most cases also assume that there is a something like a universal set of possible segments or contrasts, and aim to restrict the descriptive power of the system so that elements outside this set are excluded. This is particularly clear in the 'feature geometry' model of feature relations [e.g. 3]. Phonetic and phonological approaches may have important differences, particularly in assumptions about the 'categoricity' of the elements concerned, but what is common is that they provide a set of terms or features and a limited number of ways of combining them. Segments that cannot be described using these terms and combinations are assumed not just to be unknown but not to be possible linguistic elements.

The sources of any restrictions noted are obviously diverse, and the interest lies in trying to understand their origins. Some are 'mechanical' (hence, for example, the absence of pharyngeal nasals or velar clicks). Others relate to 'ecological' factors like adequacy of contrast, both in acoustic output and representation as motor programs (e.g. no distinction is made between labio-dentals with the lower lip slightly in front, just below, or slightly behind the upper teeth). Neither type of consideration rules out combining linguo-labial place with lingual trill manner so there is no reason to consider such to be impossible linguistic sounds, even though unattested. But it's not clear that they rule out what I called Donald Duck cheek sounds either. These use pulmonic air squeezed through a narrow lateral aperture so as to generate a vibration somewhat like the voicing vibration of the vocal folds. The fundamental frequency and other characteristics of this vibration can be varied, and the timbre modified by varying the configuration of the lips, the cheeks and so on. This system is quite efficient: the sound output is quite loud and requires neither great effort of the articulators (which are in close to the same position as for ordinary laterals) nor high volume of air flow. In short, here is an apparently good oral source of acoustic energy which human language seems to entirely neglect (except for cartoon voices). Our phonetic classificatory system has no labels for this source (it's not a lateral trill because trills by definition have a low enough rate of vibration that each cycle is heard as a separate beat). And neither is it modeled in any of the computational vocal tract models I have seen. So it is effectively dismissed without evaluation in most studies using an emergent paradigm. Kohler [11] suggests we should understand the limits on possible speech sounds as arising from the fact that "speech is ... an adaptation of the pre-existent anatomy and physiology of the vegetative system" used for "breathing, protection of the breathing system, stabilization of the rib cage, swallowing sucking, chewing" and hence language can use "only those noise productions ... that are related to the basic mechanisms already used for these non-speech functions." My sense is that the sound source I am talking about here is no more remote from these basic vegetative functions than, say, a tongue-tip trill.

Others segmental universals concern relations of segments within systems, including holistic patterns of distribution within subsets of segments, such as vowels systems. But many segmental universals concern relations expressed as 'markedness' between segments or between the attributes of segments [7, 6]. There are some well-known universals of this type concerning voicing in different classes of segments: in obstruent classes voiceless segments occur more frequently in consonant inventories than voiced ones; in sonorant classes voiced segments are more frequent. Moreover, there is a strong implicational relationship such that in general the less frequent type will only be found in an inventory if the more frequent one appears there. Some phonological feature theories reflect this distribution as two different ways of marking the voiced/voiceless contrast: in obstruents as presence or absence of the feature [voice], in sonorants as the absence vs presence of [expanded glottis], the feature used to distinguish aspirated from unaspirated obstruents. Thus both the less frequent types are given a more complex structure with an additional feature present. It is certainly possible that this is a reasonable analog to the way the categorical information on these segments is stored in the brain, but the universals related to voicing are more nuanced than the two-way representation allows. The ratio of voiceless to voiced fricatives is much higher than that of voiceless to voiced plosives, and the ratio of voiced sonorants to voiceless sonorants is higher still. Moreover, a number of significant interactions between voicing and place of articulation have been noted. I'll revisit just one of these. Considering just the three major places of articulation, bilabial, dental/alveolar and velar, in a series of voiced plosives the velar place is more likely to be 'missing' than either of the other two, but in a series of voiceless plosives the bilabial place is more likely to be missing. In the sample of 451 languages in UPSID there are 40 inventories (including Temne, Ket, Thai, Pomo and Huave) which lack / / when it might have been expected (because there are other velar segments and other voiced stops), and 38 (including Yoruba, Ket, Arabic, Warao, and Yareba) which lack /p/ when it might have been expected. For comparison, there are 15 'missing' cases of /d/, 9 of /b/, 3 of /k/ and 1 of /t/. The 'missing / /' and 'missing /p/' cases seems to be frequent enough and sufficiently widespread across languages families and geographical areas that they represent more than a chance pattern. Thus, while voiced plosives are less preferred in general, there seems to be something extra which disfavors / /, and similarly something which makes /p/ less preferable than voiceless plosives at the other principal places.

Analysis of the structure of the speech organs, air-flow patterns in the vocal tract, the functioning of the auditory system, and other components of the overall human organism often provide a basis for understanding the directionality of such preferences [e.g. 22], each considered in isolation. To understand why voiceless fricatives are preferable to voiced ones, we can appeal to the fact that to generate both voicing and frication two separate ratios of air pressures have to be maintained within fairly critical limits (among other requirements). We may conclude that voiced frication is a difficult maneuver to perform, and hence tends to be avoided on 'ease of articulation' grounds. Or we may argue that, given typical human performance limitations, the criticality of the requirements means that intended voiced fricatives will often lose voicing or frication. It matters little, since several lines of reasoning all concord in suggesting that voiced fricatives are indeed an inferior type of linguistic sound to voiceless fricatives. But we are a long way from being able to integrate individual explanations into a larger picture, where the relative importance of inherent effort, performance limitations, and all the other relevant considerations are taken jointly into account. This wonderful theory of the future will explain both why voiced fricatives are more frequent than voiceless sonorants, and 'missing' / / and /p/ are about equally frequent!

4.6 There are patterns in phonetic detail

A good many very strictly phonetic universals pertaining to segments have also been noted. These universals concern details of production, acoustics, etc, having little to do with the basic contrastive structure of phonological systems. These include 'local' patterns of timing and fundamental frequency (e.g., closure duration for bilabial plosives is generally a little longer than for velar ones under matching conditions, and high vowels have a slightly higher F0 than low vowels, other things being equal). Others of the same sort are described in [21]. These universals cannot yet be based on large samples, since the necessary data is not available for many languages. This argues for caution in accepting their universality. It also increases the uncertainty about their appropriate formulation. For example, the two mentioned above are sometimes expressed as general correlations between closure duration and a scale of frontness of stop articulation, or between F0 and a scale of vowel height. A scalar relationship might well be appropriate for the F0/vowel height relationship, although it remains difficult to show, but is more doubtful in the case of closure duration and frontness of articulation, since stops at places in between velar and bilabial have very varied durations.

Some phonetic universals are so well understood that we no longer think of them as empirical observations about languages, but simply as part of an overall phonetic theory. It is a universal that vowel formants are typically lower in the immediate vicinity of a bilabial consonant than elsewhere. Since this follows from general acoustic principles incorporated into our understanding of articulatory/acoustic relations, no-one is likely to find it a good use of their time to conduct a survey of this phenomenon. However, we do not have a good understanding of why plosive closure durations should vary with place, or F0 with vowel height. In such cases surveys are more obviously valuable so that the nature of the pattern can be reliably established before attempts at explanation are made.

Duration and fundamental frequency have important distinctive roles in language, e.g. for contrasts of quantity and tone respectively, and the 'local' effects are superimposed on top of much larger contrastive differences. Superimposition of place-related duration differences on a single/geminate stop distinction is illustrated in Figure 5. Since the mean closure durations for geminates are much longer than for singletons it is obvious that closure durations can be 'voluntarily' adjusted over wide ranges; yet within a quantity they are not adjusted to be equal at different places of articulation. We can note that the **absolute** difference in mean duration between single and geminate pairs at different places is more uniform (70-76 ms) than the **ratio** of the differences (2.09-2.49), meaning that an additive model is better than a multiplicative one for predicting these geminate durations. Whatever causes the place related differences is not perturbed by this addition, suggesting that it might relate to built-in delays in the implementation of motor commands to the different articulators, rather than to, say, differences in their natural frequencies. More and deeper studies on duration and timing would help clarify issues such as these, and where the divisions between universal and language-particular phenomena fall.

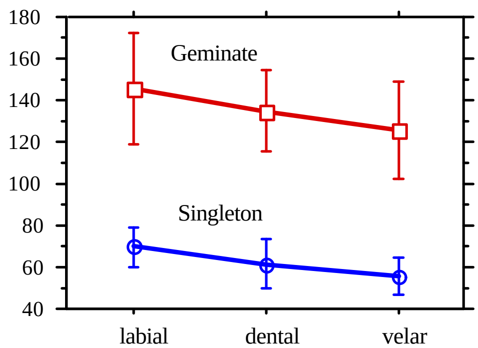

Figure 5. Closure durations of intervocalic plosives in Berber, means of four speakers. See [17] for details.

4.7 There is variability

The final class of phonetic universals to be mentioned here concern the systematic use of variation. Position in prosodic units, intended clarity of speech, and relative redundancy of information conveyed are among factors shown to contribute to controlled variation within repetitions of the same word. Such transformations [15, 14] of the same linguistic material illustrate the adaptive flexibility of speaking. Some particular transforms noted include lengthening in phrase-final position, more energetic articulation in initial position, and more coarticulation in low-information contexts. The fact of plasticity which this reflects is itself an important language universal. However, which particular transforms are truly widespread is something we will only know when more data is in. Increased coarticulation in low-information contexts looks likely to be more general than, say, phrase-final lengthening, since the phrase-final position is in many languages a position of extreme reduction marked by vowel shortening, devoicing or deletion (e.g. CiLuba).

5. FINAL REMARKS

This paper has only lightly scratched the surface of what might be said about phonetic and phonological universals. I have emphasized the importance of basing universals on a sample of a sufficient number of diverse languages, but stressed also that universals are not so much subjects of importance in themselves but are the preface to the formulation of ideas. A fruitful — in fact, necessary — perspective for formulation of many ideas concerning the origin of language universals is the placing of language in its biological setting, considering both the context of its early evolution and the capacities and limits of the human organism as it exists today.

REFERENCES

[1] Berrah, A. R. & Laboissiere, R. 1997. Phonetic Code Emergence in a Society of Speech Robots: Explaining Vowel Systems and the MUAF Principle. *Proceedings Eurospeech97* (Rhodes, Greece).

[2] de Boer, B. To appear. The emergence of sound systems through self-organization. In Hurford, J. R., Knight, C. & Studdert-Kennedy, M. (eds.) *The Emergence of Language*. Cambridge: CUP.

[3] Clements, G. H. & Hume, E. V. 1995. The internal organization of speech sounds. In Goldsmith, J. A. (ed.) *Handbook of Phonological Theory*. Oxford: Blackwell: 245-306.

[4] Croft, W. 1990. *Typology and Universals*. Cambridge: CUP.

[5] Deacon, T. W. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W. W. Norton and Co.

[6] Ferguson, C. A. 1966. Assumptions about nasals: a sample study in phonological universals. In Greenberg, J. H. (ed.)*Universals of Language* Cambridge, MA: MIT Press: 53-60.

[7] Greenberg, J. H. 1966. *Language Universals with Special Reference to Feature Hierarchies*. The Hague Mouton.

[8] Grimes, B. 1992. *Ethnologue: Languages of the World (Twelfth Edition)*. Dallas: Summer Institute of Linguistics.

[9] Hurford, J. R., Studdert-Kennedy, M. & Knight, C. 1998. *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: CUP.

[10] Kauffman, S. 1993. *The Origins of Order*. Oxford: OUP.

[11] Kohler, K. J. 1998. The development of sound systems in human language. In [9]: 265-278.

[12] Ladefoged, P. & Maddieson, I. 1996. *Sounds of the World's Languages*. Oxford: Blackwell.

[13] Lindblom, B. 1986. Phonetic universals in sound systems. In Ohala, J. J. and Jaeger, J. J. (eds.) *Experimental Phonology*. Orlando, FL: Academic Press: 13-44.

[14] Lindblom, B. 1998. Systemic constraints and adaptive change in the formation of sound structure. In [9]: 242-264.

[15] Lindblom, B., Brownlee, S, Davis, B. & Moon S.-J. 1992. Speech transforms. *Speech Communication*. 11: 357-368.

[16] Locke, J. L. 1998. Social sound-making as a precursor to language. In [9]: 190-201.

[17] Louali, N. & Maddieson, I. 1999. Paper presented at the 14th IPhS

[18] MacDonald, J. & McGurk, H. 1978. Visual influences on speech perception process. *Perception and Psychophysics* 24: 253-257

[19] MacNeilage, P. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21: 499-546.

[20] Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: CUP.

[21] Maddieson, I. 1997. Phonetic universals. *The Handbook of Phonetic Sciences*, (ed W. Hardcastle and J. Laver). Oxford: Blackwell. 619-639.

[22] Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. In MacNeilage, P. F. (ed.) *The Production of Speech*. New York: Springer. 189-216.

[23] Pinker, S. 1994. *The Language Instinct*. New York: Morrowo.

[24] Posner, R. 1996. *The Romance Languages*. Cambridge: CUP.

[25] Raulo, D. & Laboissiere, R. 1999. Web-UPSID: A user-friendly interface for the UCLA Phonological Segment Inventory Database. Paper presented at the 14th IPhS.

[26] Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. 1997b. The dispersion-focalization theory of vowel systems. *Journal of Phonetics* 25: 255-286.

[27] Studdert-Kennedy, M. 1998. The particulate origins of language generativity. In [9]: 202-221.

[28] Ulbaek, I. 1998. The origin of language and cognition. In [9]: 30-43.

[29] United Nations. 1948. Universal Declaration of Human Rights. General Assembly Resolution 217 A(III), December 10, 1948.