This database, compiled by Merritt Ruhlen, contains certain kinds of linguistic and nonlinguistic information for the world’s roughly 5,000 languages. This introduction will discuss the kinds of data that are surveyed. Corrections, addenda, or comments should be sent to ruhlen@santafe.edu.

DATABASE STRUCTURE.

1. Language. This database contains 5,707 records, each representing one language or dialect. The names of the languages and dialects follow the nomenclature given in my Guide to the Languages of the World, Vol. 1: Classification (1991).

2. Alternate Name. No attempt has been made to list every alternate name for every language. For a complete index of all language names one should consult the Ethnologue (2000), published by SIL International and also available on the web at http://ethnologue.org, and the Voegelin’s Classification and Index of the World’s Languages (1977). Alternate names are given here primarily for languages that have two names, both of which are widely used (e.g. Gilyak and Nivkh), or for a name used in a source that differs from that used here.

3. Dialect. Dialect names are given when they were mentioned in a source or where they are needed to distinguish two forms of a language.

4. Geographical Location. I have attempted to provide a geographical location for every language in this database, usually in terms of the country in which it is spoken. However, languages are often spoken in more than one country for the simple reason that there is little correlation between linguistic boundaries and political boundaries in most parts of the world. The Ethnologue provides the definitive answer to these questions, specifying exactly all the countries in which a given language is spoken. I list only a single country for each language, usually the country with the largest number of speakers, but sometimes the country from which the data were taken, which need not be the country with the largest number of speakers.

5. Number of Speakers. There is no more difficult task than ascertaining the number of speakers for many languages. Though I collected all of the population statistics given in the sources I used, these numbers were in many cases already out of date in the 1970’s and are obviously even more so today. The Ethnologue is the definitive source for population statistics for all the world’s languages and dialects. Where possible, I list the number of speakers for a particular dialect in parentheses following the number of speakers for the language itself. For extinct languages the data of extinction is given when it is known.

6. Genetic Classification. The genetic classification given for each language is based on that given in Volume 1 of my Guide, with a few exceptions. The Kusunda language is no longer listed as a Tibeto-Burman language, but rather belongs to the Indo-Pacific family. The Veddah of Sri Lanka speak a dialect of Sinhalese, but their original language, which was definitely not
Indo-European, has been lost, with only traces of it remaining in the Sinhalese dialect they speak. Elamite is treated as an independent language since its putative Dravidian affiliation has been questioned.

I have abbreviated the classification in certain cases, eliminating some intermediate nodes in order to make the classification more readable. Since the entire classification is given in Volume 1 in as much detail as I was able to ascertain, for simple taxonomic questions the reader should consult Volume 1. When sorted on the basis of the genetic classification given in Volume 1 this database will allow one to follow the linguistic topography of the world while at the same time following its genetic topography. In this way one can see how linguistic traits vary in terms of their genealogical history.

The classification given in Volume 1 can obviously be improved and we intend to allow this classification to evolve on the EHL web site, just as the typological data will also evolve with new information and corrections of errors that now exist.

SOURCES

Sources of the data given in this database are divided into four categories: dictionary, grammar, textbook, and other sources. Abbreviations used in the sources are listed at the end of this introduction.

7. Dictionary. This field contains dictionaries or word lists for a given language.

8. Grammar. This field contains grammars or grammatical sketches that deal with the entire language.

9. Textbook. This field contains textbooks for a given language.

10. Other Sources. This field contains all other sources from which data were taken.

PHONOLOGY

Phonology is the study of the sounds of language, primarily but not exclusively consonants and vowels, and it is one of the main foci of this database. No attempt will be made here to give a complete description of how all the possible consonants and vowels are produced or how they are represented in the adaptation of the International Phonetic Alphabet used in this database. There are many excellent books on just this topic. The one I would recommend is *Vowels and Consonants* (2001) by Peter Ladefoged. Lucidly written by one of the world’s leading phoneticians, it also includes a CD ROM that allows one to actually hear all of the various consonants and vowels used by the world’s languages and illustrated in this database by data from over 2,200 languages. The contents of this CD can also be accessed on the web at http://htv.humnet.ucla.edu/departments/linguistics/VowelsAndConsonants. I will, however, give a brief description of how consonants and vowels are produced for those who have no background in phonetics.
CONSONANTS

The symbols for consonants used in this database are shown in Table 1 and are for the most part the same as those used by the International Phonetic Alphabet (IPA). I have, however, substituted different symbols for certain consonants so that, for example, in place of the IPA symbols \( \text{ʃ} \) and \( \text{ʧ} \) I have used \( \text{ʃ} \) and \( \text{ʧ} \), which are the initial consonants in 'she' and 'chop,' respectively. I have also used a dot under a consonant to indicate retroflexion, rather than the special IPA symbols, so that \( \text{t}. \) and \( \text{n}. \) in this database correspond to IPA \( \text{ʈ} \) and \( \text{n} \), respectively. The palatal nasal, \( \text{n} \), in IPA, is here represented as \( \text{ŋ} \), and there are a few other modifications that will be mentioned below. Consonants (and vowels and glides) are enclosed in parentheses in the database to indicate they are marginal in the language concerned. Usually this means that either they are very rare or occur only in loanwords.

Table 1. Consonants

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>1/8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>( p )</td>
<td>( \text{ʧ} )</td>
<td>( t )</td>
<td>( \text{ʧ} )</td>
<td>( c )</td>
<td>( k )</td>
<td>( \text{ kp} )</td>
<td>( q )</td>
<td>( ? )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricates</td>
<td>( p^{\text{ʃ}} )</td>
<td>( \text{ʧ}^{\text{ʃ}} )</td>
<td>( t^{\text{ʃ}} )</td>
<td>( \text{ʧ}^{\text{ʃ}} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( q^{\text{ʃ}} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricatives</td>
<td>( \phi )</td>
<td>( f )</td>
<td>( \theta )</td>
<td>( s )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( s )</td>
<td>( \text{ʃ} )</td>
<td>( \chi )</td>
<td>( h )</td>
<td>( h )</td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td>( \beta )</td>
<td>( \upsilon )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
<td>( \text{ʒ} )</td>
</tr>
<tr>
<td>Nasals</td>
<td>( m )</td>
<td>( \text{ŋ} )</td>
<td>( n )</td>
<td>( \text{ŋ} )</td>
<td>( \text{ŋ} )</td>
<td>( \text{ŋ} )</td>
<td>( \text{ŋ} )</td>
<td>( \text{ŋ} )</td>
<td>( \text{mŋ} )</td>
<td>( n )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laterals</td>
<td>( l )</td>
<td>( l )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
</tr>
<tr>
<td>Trills</td>
<td>( b )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
<td>( r )</td>
</tr>
<tr>
<td>Flaps</td>
<td>( t )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
<td>( l )</td>
</tr>
<tr>
<td>Ejectives</td>
<td>( p^{\prime} )</td>
<td>( \text{ʧ}^{\prime} )</td>
<td>( t^{\prime} )</td>
<td>( s^{\prime} )</td>
<td>( c^{\prime} )</td>
<td>( k^{\prime} )</td>
<td>( k^{\prime} )</td>
<td>( q^{\prime} )</td>
<td>( q^{\prime} )</td>
<td>( q^{\prime} )</td>
<td>( q^{\prime} )</td>
<td>( q^{\prime} )</td>
</tr>
<tr>
<td>Implosives</td>
<td>( b )</td>
<td>( d )</td>
<td>( j )</td>
<td>( g )</td>
<td>( k )</td>
<td>( k )</td>
<td>( k )</td>
<td>( g )</td>
<td>( d )</td>
<td>( d )</td>
<td>( d )</td>
<td>( d )</td>
</tr>
<tr>
<td>Clicks</td>
<td>( \circ )</td>
<td>( / )</td>
<td>( // )</td>
<td>( ! )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
<td>( \text{ʃ} )</td>
</tr>
</tbody>
</table>

\( a \) The numbers at the top of the Table indicate the place of articulation of the consonants as follows: 1. bilabial, 2. labiodental, 3. interdental, 4. dental, 5. alveolar, 6. retroflex, 7. palatal, 8. velar, 1/8. labiodental, 9. uvular, 10. pharyngeal, 11. glottal.

How are consonants produced? There are three basic parameters. First, there must be an airstream mechanism that causes air to move through the mouth. Second, this airstream is modified in certain ways. Third, the modification of the airstream can take place at different places in the vocal tract.

AIRSTREAM MECHANISMS

There are four airstream mechanisms: pulmonic, egressive glottalic, ingressive glottalic, and velaric. The pulmonic airstream is caused by the lungs pushing air through the vocal tract and out of the mouth (see Fig-
ure 1). All languages use this airflow mechanism and for most languages it is the only one used. Consonants produced with a pulmonic airflow are shown in the first block in Table 1 (stops–flaps). The egressive glottalic airflow is produced by closing the vocal cords in the glottis and, with another closure in the vocal tract (e.g. the closure involved in making a k), the glottis is raised so that the air in the mouth is compressed. When the second closure (for k) is released air rushes out of the mouth producing an ejective k’. Implosives are produced with an ingressive glottalic airflow. In this case the closed glottis is lowered and the air in the mouth is rarified. When the second closure (e.g. that involved in making a b) is released air is sucked into the mouth producing an implosive b. Ejectives and implosives together are called glottalized consonants.

The velaric airflow produces clicks (e.g. ‘tsk tsk’ in English). This airflow is produced by (1) raising the back of the tongue to the roof of the mouth (the velum), creating a closure at the back of the oral cavity, (2) a second closure is then made at the front of the oral cavity (say for a b), (3) the body of the tongue is both lowered and drawn backward in the mouth, thereby rarifying the air in the oral cavity (as in the case of implosives), (4) the closure (for b) at the front of the oral cavity is released, allowing air outside the mouth to be sucked in, creating a clicking sound, in this case the click ⊙. This click is sometimes used in English to signify a kiss. While clicks are widely used by themselves in the world’s languages as interjections (as in the English examples just cited) their use as ordinary consonants in words is for the most part restricted to the Khoisan language family and to a few languages which have borrowed clicks from Khoisan languages (e.g. Xhosa and Zulu).

PLACE OF ARTICULATION

Once an airflow is produced it must be modified in some way (manner of articulation) at some place in the vocal tract (place of articulation) in order to make a sound. Possible places of articulation for consonants are indicated in the first row of Table 1 and are shown in Figure 1. Possible manners of articulation are listed in the first column of Table 1.

For most of the places of articulation one method of creating a consonant is by stopping the airflow with a closure somewhere in the vocal tract and then releasing it to make a sound. Consonants produced in this way are called stops (see below). A bilabial stop is made by bringing both lips together (p, b). Dental stops (e.g. t, d) are produced by moving the tip of the tongue to the back of the upper teeth and alveolar stops (t, d) are produced in a similar way except that the tip of the tongue touches the alveolar ridge, just behind the upper teeth. To make a retroflex consonant (e.g. r, l) the tip of the tongue is brought even further back in the mouth, making closure with the front part of the palate. The closure for palatal consonants (c, y) is made by raising the blade of the tongue to the palate. Velar consonants (k, g) and
uvular consonants (q, ṝ) are similar in that the back of the tongue is raised to the velum for the former and to the uvula for the latter. A glottal stop is produced by closing the vocal cords in the glottis and then releasing them, producing a brief moment of silence. We have already seen that this same action is used to produce the glottalic airstream. Labial-Velar stops (kp, gb) involve two simultaneous closures, one at the lips, as with bilabial stops, the other at the velum, as with velar stops. Three places of articulation do not permit complete closure, but only a partial closure that produces fricatives (see below). Labiodental fricatives (f, v) involve raising the lower lip towards the upper teeth. For interdental fricatives (θ, ð) the tip of the tongue is pushed forward, just under the upper teeth. Finally, pharyngeal fricatives (h, ˈ) are produced by retracting the root of the tongue to the back wall of the pharynx.

MANNER OF ARTICULATION

The manner of articulation determines what kind of sound is produced at each place of articulation.

11. Stops. Consonants for which there is a complete closure in the vocal tract are called stops and we saw many examples above.
12. Affricates. Affricates (p', t', t',  č, k') are produced by two movements. First there is a complete stoppage of the airstream, as in the case of stops; second, there is a release of the stop with accompanying friction, as in the case of fricatives.

13. Fricatives. Fricatives (f, s,  s, h) are produced by narrowing the airstream so as to produce friction, but not so much as to actually cut off the airstream completely.

14. Approximants. Approximants are like fricatives except that there is very little friction produced. The glides j and w are often considered approximants, but we will treat them as a separate group of consonants with similarities to both consonants and vowels, as is usually done in the linguistic literature.

15. Nasals. Nasal consonants (m, n, ṇ, ŋ) are made in the same way as the voiced (vocal cords vibrating) stops b, d,  g, except that air is allowed to escape through the nasal cavity as well as the oral cavity.

16. Laterals. For laterals (l, l, ɬ) the tongue makes a complete stoppage of the airstream in the middle of the roof of the mouth, but air is allowed to escape on the sides of the tongue, which is why they are called laterals.

17. Vibrants. Vibrants are a heterogenous class of sounds, listed as trills and flaps in Table 1. They are often represented as different kinds of r’s even though the production of the various kinds of vibrants can be quite different. The two most common varieties of vibrants are the flap r and the trill r. Flap r involves the tongue flapping one time against the roof of the mouth, whereas for trill r the tongue flaps against the roof of the mouth several times in rapid succession. Spanish has both these consonants, flap r in pero ‘but’ and trill r in perro ‘dog.’ What is written r in French is the uvular trill ʀ, and the r in American English is the retroflex approximate ɾ.

18. Clicks. As we saw above, clicks are characterized by their unique airstream mechanism. They can be produced at five different places of articulation, but they are often modified in certain ways such that some languages have several dozen different click consonants.

19. Modified Consonants. In many languages the basic consonants listed in Table 1 can be modified in certain ways to form new, distinct consonants. In fact one such modification is already contained in Table 1, a modification that distinguishes consonants made with the vocal cords vibrating (voiced consonants) and consonants made with the vocal cords at rest (voiceless consonants). In the rows for stops, affricates, and fricatives the top row contains voiceless consonants, the second row voiced consonants. What this means is that the only difference between, say p and b (or t and d, or f and v), is that the vocal cords are vibrating when b (or d or v) is pronounced, whereas there is no vibration for p (or t or f).

Unlike the trait voiceless/voiced, which is indicated by different symbols for the voiceless and voiced pairs, most modifications are usually indicated by diacritics. For example, in many languages, in addition to the regular k,
there is also a k produced with the lips rounded, kʷ. Consonants produced with lip rounding are called labialized and are represented by a superscript w: kʷ. The various diacritics used in this database to represent modifications of basic consonants are the following.

- [j]: palatalized
- [w]: labialized
- [u]: velarized
- [ː]: pharyngealized
- [ʰ]: aspirated
- [Carthy]: glottalized
- [ʷ]: voiced click
- [ᵣ]: syllabic
- [m ɳ]: prenasalized

A palatalized consonant is one in which the basic consonant is accompanied by the simultaneous raising of the front of tongue, as in the case of the glide j and the high front vowel i (see below). A labialized consonant is one in which the basic consonant is accompanied by the rounding of the lips, usually with the simultaneous raising of the back of the tongue. A velarized consonant is is a basic consonant accompanied by the raising of the back of the tongue. In many dialects of American English the l in ‘leak’ is plain, while the l in ‘full’ is velarized. A pharyngealized consonant is one in which the basic consonant is accompanied by the simultaneous retraction of the back of the tongue in the area of the pharynx. Aspirated consonants are followed by a brief puff of air following the plain consonant. In English the p in ‘spy’ is unaspirated, while the p in ‘pie’ is aspirated. Glottalized consonants are produced with the glottal airstream mechanism, discussed above. Clicks may be either voiceless or voiced. A superscript ʰ indicates that the click is voiced. / is a voiceless dental click; /h is a voiced dental click. The subscript ɹ indicates that the consonant is syllabic, i.e. can form a syllable by itself. The n in ‘no’ is non-syllabic; the n in ‘button’ is in many dialects syllabic. All vowels are syllabic. Prenasalized consonants are basic consonants that begin with a brief nasal consonant; the most common are m, n, and ng. The superscript tilde ˜ is used to indicate that a vowel is nasalized, that is, air is allowed to escape through both the oral and nasal cavities. In French, which has nasal vowels as distinctive sounds, this difference in nasalization is the only feature separating beau /bo/ ‘handsome’ from bon /bɔ/ ‘good.’ In some languages there is a contrast between long and plain (short) consonants or vowels. The : following a consonant or vowel indicates length. In Italian, for example, the only difference between fatto ‘fate’ and fatto ‘done’ is that in the latter the t has a longer duration than in the former. Long vowels are even more common than long consonants. Latin, for example, had five plain (short) vowels, i e a o u, and five long vowels, iː eː aː oː uː.

The subscript _ indicates that the consonant is dental, i.e. pronounced with
the tip of the tongue touching the upper teeth. The t in French is dental ʒ, whereas the t in English is alveolar t. As described above, retroflex consonants are produced by the tip of the tongue bending back to touch the front of the palate. Retroflex consonants are indicated by a dot under the consonant: ʒ, ʒ. ɣ. Fortis consonants are indicated by a subscript ˈ. The precise nature of fortis consonants varies from language to language, but in general it involves greater energy in the production of the consonant than that found in the corresponding lenis forms. The subscript ˈ indicates that the consonant or vowel is voiceless. However, since vowels are usually voiced, and for consonants the difference between voiceless and voiced is usually indicated by different symbols, this diacritic is rarely used. Voiceless vowels are rare, but do occur in some languages, such as Comanche. This diacritic is also used sometimes with consonants that are normally voiced, such as nasal consonants. In addition to normal voicing, discussed above, two other kinds of voicing are found in some languages; breathy voice is indicated by a subscript ˌ, while creaky voice is indicated by a subscript ˌ. For breathy voice there is a looser form of vibration of the vocal cords and greater airflow than for normal voicing. Creaky voice has a tighter form of vibration and less airflow than in the case of normal voicing.

In addition to these diacritics three symbols are used to represent groups of consonants: N = nasal consonants (m, n, ɲ); L = liquids (l and r-like sounds, l, ɾ, r); G = glides (j, ɾ, w, ɯ).

20. Glides. Glides, which are also called semi-consonants and semi-vowels because they share properties with both, were listed in Table 1 in the line of approximants: j, ɾ, w, ɯ. They are, however, often considered a set of consonants which behave differently from other consonants and that view is followed here, where the glides are listed separately from the other consonants. The glides j and w, the initial sounds in English ‘yes’ and ‘we,’ are very common in the world’s languages. The other two glides are very rare. In one language, Rumanian, there is an additional glide that is represented as [e], but I have not found this glide in any other language.

VOWELS

Like consonants, vowels are produced by modifying the airstream. However, for vowels the critical factor is the position of the tongue, which varies along two dimensions: high-low and front-back (see Table 2). If the tongue is raised in the mouth toward the roof of the mouth, we speak of high vowels; if the tongue is lowered, we have low vowels. Intermediate levels—lower-high, mid, lower-mid—are also found. The second factor involved is the position of the highest point of the tongue along the front-back parameter; if the highest point of the tongue is in the front of the mouth we speak of front vowels; if it is in the back of the mouth, we have back vowels; inbetween these two positions there are also central vowels. A third factor in vowel production is lip rounding. Normally back vowels are accompanied
by lip rounding, which is more pronounced for high back vowels than for low back vowels. Front vowels and central vowels usually do not involve lip rounding. There are, nonetheless, languages which have front rounded vowels as well as languages with back unrounded vowels, though both are comparatively rare in the world’s languages. French is a language that has front rounded vowels so that the only difference between *ris* ['ri] ‘laugh’ and *rue* ['ry] ‘street’ is that the former has a high front unrounded vowel, while the latter has a high front rounded vowel. In Table 2 unrounded vowels are in the first column under front, central and back, and rounded vowels are in the second. The symbols for vowels are essentially those of the IPA and are illustrated in Table 2.

**Table 2. Vowels**

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i y</td>
<td>i u</td>
<td>u uu</td>
</tr>
<tr>
<td>lower-high</td>
<td>i y</td>
<td>i</td>
<td>u uu</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>ø ø</td>
<td>y ø</td>
</tr>
<tr>
<td>lower-mid</td>
<td>e ø</td>
<td>ø ø</td>
<td>ø ø</td>
</tr>
<tr>
<td>low</td>
<td>æ a</td>
<td>ø ø</td>
<td>ø ø</td>
</tr>
</tbody>
</table>

21. **Front Vowels.** Front vowels are formed in the front of the mouth, as discussed above.

22. **Central Vowels.** Central vowels are formed in the middle of the mouth, as discussed above.

23. **Back Vowels.** Back vowels are formed in the back of the mouth, as discussed above.

24. **Long Vowels.** Long vowels have a longer duration than short vowels, as discussed above.

25. **Nasal Vowels.** Nasal vowels are formed with the air exiting through both the oral and nasal cavities, while for oral vowels the nasal cavity is closed to the exiting airstream.

26. **Modified Vowels.** Modification of the basic vowels was discussed above with regard to modified consonants since in a number of cases the modifications may affect either consonants or vowels (e.g. length, breathy voice, creaky voice).

The vocalic systems of many Mon-Khmer languages of Southeast Asia have what is known as voice register, which means that the vowels are divided into two sets that are differentiated by voice quality. In one set the voice quality is clear (V), in the other, breathy (V′) or, sometimes, laryngealized (V). Gradin (1966) described the difference between the two voice qualities in Jeh as follows: “The deep vowel quality is produced by relaxing the faucal pillars, lowering the larynx, and giving increased pressure from the diaphragm. The result is a deep, somewhat gruff, voice quality. Pitch is usually lower than that of the clear form. Deepness, when occurring with short vowels, changes the vowel height, forcing it up in most instances.”
27. **Diphthongs and Triphthongs.** Diphthongs differ from simple vowels in that the tongue moves from one vowel position to another and thus they can be represented as a sequence of two vowels. In the diphthong *ai* the tongue begins in the position of *a* and then glides upward to the position of *i*. In the literature diphthongs are sometimes represented as a sequence of a vowel and glide so that *ai* may be written as *aj*. I have followed the description of diphthongs as they are given in the sources themselves so that what is described as *ai* in one language is really no different from what is written *aj* in another language. Triphthongs are similar to diphthongs in that there is movement of the tongue from one position to another, but in the case of triphthongs there are two distinct movements, not just one as with diphthongs. Thus in the triphthong *uəi* the tongue begins in the position of *u*, then moves to the position of *ə*, and finally to the position of *i*. Where diphthongs and triphthongs are specified exactly in the source they are listed just that way in this database. In some sources, however, it is clear that there are diphthongs or triphthongs, but it is not clear exactly what they are. In such cases I have indicated this fact by simply listing ‘diphthongs’ or ‘triphthongs.’

**Vowel Harmony**

In some languages vowels are divided into two sets and in any given word only members of one set or the other may occur. The two sets are distinguished by some phonetic trait. In Turkish the eight vowels are divided into two sets of four; one set contains only front vowels, the other only back vowels. There are a number of different phonetic traits that may distinguish the two sets, including the front/back distinction as in Turkish; other traits defining vowel harmony systems are high/low, labialized/non-labialized, advanced tongue root/retracted tongue root, and nasal/non-nasal. Languages reported to have vowel harmony are indicated simply by **vowel harmony** in the field Modified Vowels, but the precise nature of the vowel harmony system is not described.

**Consonant Harmony**

Consonant systems in some languages are, like vowel systems, divided into two sets of consonants, members of only one set being allowed in any given word. The most common type is nasal harmony, in which nasal consonants constitute one class, non-nasal consonants, the other. Languages reported to have nasal harmony are indicated by **nasal harmony** in the field Modified Consonants. In some languages nasal harmony involves both consonants and vowels so that every word contains either nasal consonants and nasal vowels or non-nasal consonants and vowels.

**Phonemes and Allophones**

In addition to the knowledge of how consonants and vowels are produced, sketched above, there are certain fundamental linguistic principles that users of this database who have no linguistic background should be
aware of. The consonants and vowels listed for each language in this
database have the technical name of phonemes in linguistics. A phoneme is
a sound that is capable of distinguishing meaning in a given language. For
example, in English p and b are different phonemes because it is only the
difference between p and b that distinguishes ‘pit’ from ‘bit.’ A phoneme,
however, need not be a single sound, but rather may consist of a set of
different sounds even though speakers of the language hear these sounds,
known as allophones, as being the same sound. In other words the consos-
nants and vowels that one hears depend on the language spoken and not
on the absolute quality of the sound.

An example of a phoneme consisting of two allophones is English p.
Native speakers of English perceive the p in ‘pie’ and ‘spy’ as being identical
even though they are in fact different sounds, the p in ‘pie’ being aspirated
[pʰ] and the p in ‘spy’ unaspirated [p]. You can convince yourself of this fact
by doing a simple experiment. Light a match and hold it in front of your
lips and say ‘spy spy spy pie.’ After each ‘spy’ the match remains lit, but as
soon as ‘pie’ is said the match goes out. The reason for this is that the p in
‘pie’ is aspirated [pʰ] and it is the aspiration, represented by [ʰ] that blows
out the match. Aspiration is just a puff of air. The p in ‘spy,’ however, is
unaspirated—it is not followed by a puff of air—and the match remains
lit. The English phoneme /p/ consists of two allophones: [p] only after s,
and [pʰ] everywhere else. (By convention phonemes are enclosed in slanted
lines /p/, allophones in brackets [p].) It is the fact that these two sounds
never occur in the same environment—and hence can never distinguish
meaning—that leads English speakers to hear them as the same. That /p/
and /b/ are different phonemes in English is demonstrated by the minimal
pair ‘pit’ and ‘bit,’ but there are no comparable pairs of words in English
differentiated only by [p] and [pʰ] because these sounds can never occur in
the same environment. Which variant is used in any particular environment
is automatically determined by the allophonic rule described above. As one
can readily see by examining the consonant systems in different languages
in this database there are in fact many languages in which /p/ and /pʰ/
are different phonemes and can distinguish meaning. In Mandarin Chinese,
for example, the only difference between /pʰeǐ/ ‘to memorize’ and /pʰeɪi/
‘to match’ is that the former begins with an unaspirated p, while the latter
begins with an aspirated pʰ.

The representation of phonemes in this database differs slightly from the
traditional representation. Normally any allophonic variation is ignored in
representing phonemes, as is the precise phonetic nature of the phoneme.
For example, both English and French have a /t/ phoneme which is usually
written just this way in phonemic descriptions of both languages. However,
English t and French t differ in two ways. English t is alveolar, which means
that the tip of the tongue makes contact with the alveolar ridge on the roof
of the mouth, just behind the teeth; and English t is usually aspirated [tʰ]

 unaspirated [t] occurs only after s, exactly like k). French t is, however, dental [t], which means that the tip of the tongue touches the back of the upper teeth, not the alveolar ridge; and French t is almost always unaspirated. It is for these reasons that I have represented the English and French /t/ phonemes differently in this database, English /tʰ/ and French /t/. If someone wants a more traditional representation of phonemes it is easy to eliminate the sign of aspiration [ʰ] in English, and the dental sign [龈] in French, arriving at identical phonemes in both languages: /t/.

There is, however, a further problem of transcription that must be kept in mind. If a language has one /t/ phoneme, which is dental, it is represented as /t/ in this database. This means that the source used described this sound as dental. Similarly, if a source describes /t/ as alveolar, it is represented just this way. The problem that arises is that in many sources the exact place of articulation of /t/ is not mentioned (because it is not phonemic and thus can be ignored). In such cases I have just used /t/, but this means that the /t/ phonemes listed for various languages are a combination of those actually described as alveolar, and those for which the precise place of articulation was unspecified.

There is a final aspect of the representation of t’s that will be apparent to linguists, but could be overlooked by non-linguists. This is the fact that although English /tʰ/ is usually alveolar, and French /t/ dental, there are languages, such as the Tiwi language of northern Australia, that have both dental /t/ and alveolar /t/ as distinct phonemes. This means that the exact meaning of /t/ depends on the language, since the meaning of a phoneme depends on its position in a phonological system.

A similar system of transcription is used for vowels, in which the phonemes are, where possible, represented by their elsewhere allophone. Many languages have two sets of vowels, distinguished by length. As we saw above, Latin had short /i e a o u/ and long /iː eː aː oː/, which are often described in just this manner. This is a very elegant solution, which is in some sense correct, but it conceals the fact that the place of articulation was not really the same for the long and short vowels. The short vowels were really /i e a o u/ and are so represented in this database. There is really no mystery about this; long /iː eː/ is pronounced with a higher tongue position in the mouth precisely because it is long and the tongue has more time to rise higher in the mouth, whereas for short /i/ there is less time for the tongue to rise and it thus never gets as high as /iː/.

There are important implications for such a system, for when a system based on length breaks down, as it did in Latin, the difference in tongue height (which was always present, though allophonic) may replace length as the phonemic factor, as it did in Latin. Thus, Classical Latin /e/ and /eː/, distinguished by length, had become by the time of Vulgar Latin, /e/ and /eː/, respectively, distinguished by tongue height. Analogous situations also occur in other vowel systems which have two sets of vowels distinguished
by one phonetic trait. The voice register systems in Mon-Khmer languages discussed above are such a case. When these systems break down, as they have in many languages, the voice register distinction is replaced by the difference in tongue height which, as in Latin, was always present, but allophonic.

28. Syllable Structure. Syllables are combinations of consonants and vowels. All languages have syllables consisting of a single vowel (V) and a consonant and a vowel (CV) and some languages have only these two syllable types. Most languages, however, also have syllables beginning and ending with a consonant (CVC), or beginning and/or ending with a consonant cluster (CCV, CVCC). In this database possible syllable types are indicated by a schema indicating all possible syllable types. Consonants in parentheses are optional so that (C)(C)V(C)(C) indicates that the language has nine possible syllable types: V, CV, VC, CVC, CCV, CVC, CVCC, and CCVCC. In some cases the kinds of consonants found at certain places in the syllable are limited to a certain set of consonants. Thus, the schema (C)(G)V indicates that the only possible initial syllable cluster is a consonant followed by a glide. In some languages there are also syllables consisting of a single consonant, usually a nasal or liquid. The schema for such syllables involves a diacritic under the consonant; N indicates that nasal consonants can be syllabic by themselves.

29. Tone. Some languages, in addition to consonants and vowels, use tone to differentiate words, that is, the pitch of a syllable can distinguish different words. For example, in Mandarin Chinese there are four tones, so that the syllable ma represents four different words depending on the tone associated with it. With a high-level tone it means ‘mother’; with a high rising tone, ‘hemp’; with a low-falling-rising tone, ‘horse’; and with a high falling tone, ‘scold.’ As can be seen in this database tone languages are particularly common in Africa and Southeast Asia, but they are also found to a lesser degree in many other parts of the world. There are a number of different ways of representing tones—diagrams and numbers are two methods—but by far the most common way in the literature is to describe the pitch just as I have done for Mandarin above. There are five levels of pitch—high, high-mid, mid, low-mid, low—and five forms of pitch movement—level, rising, falling, rising-falling, falling-rising. In addition there is sometimes a difference in voice quality so that some tones may be described as glottalized.

30. Stress. In some languages the position of stress in a word is fixed. In French stress is always on the last syllable of a word; in Hungarian it is always on the first syllable; in Polish stress is on the penultimate syllable; and in Fox it is on the antepenultimate syllable. In many languages, such as English, stress is phonemic, as seen in pairs such as the noun ‘recall’ and the verb ‘recall.’ In other languages stress is non-phonemic, but not determined by a simple rule.
31. Noun Number. In many languages number is marked on nouns by an affix, usually either a suffix (e.g. English ‘cat’ and ‘cat-s’) or a prefix (Bantu mu-ntu ‘man’ and ba-ntu ‘men,’ in which case the singular is also overtly marked by a prefix). Number distinctions in those languages which have them are marked in terms of the following abbreviations: s: singular (one), p: plural (more than one), d: dual (exactly 2), t: trial (exactly 3).

32. Noun Classes. In some languages nouns are divided into classes, sometimes with a semantic basis, sometimes on a seemingly arbitrary basis, and sometimes with a mixture of both. These noun classes affect other grammatical structures, either in the noun phrase or in the entire sentence. The gender classes of Indo-European languages are one example, but we will discuss gender in the following section. The basis for noun classes can be quite varied, but certain categories are more common than others. Noun classes are particularly common in Africa. The Rere language of Sudan has 23 noun classes, some of which have a semantic or grammatical basis—persons, trees, common objects, long objects, large and harmful animals, hollow deep objects, small or domesticated animals, augmentatives, infinitives, liquids, body parts—but the others appear to be an arbitrary collection of nouns. Ndumu, a Bantu language spoken in Gabon has seven noun classes characterized as people, animals, plants, measures, instruments, actions, and abstractions. Bantu languages have both singular and plural noun classes, characterized by different prefixes for the various classes, and these prefixes are found not only on the noun itself, but are attached to every word in the sentence. In Swahili, for example, one noun class has the singular prefix ki- and the plural prefix vi-, as can be seen in the following two sentences:

- *ki-kapu* ki-kubwa *ki-moja* *ki-lianguka* `One large basket fell.'
  - `basket large one fell`

- *vi-kapu* vi-kubwa *vi-tatu* *vi-lianguka* `Three large baskets fell.'
  - `basket-s large three fell`

Caucasian languages also have noun classes. The Bats language has seven classes: masculine, feminine, non-rational beings, nature, things, objects, ideas; the Hunzib language has six classes: masculine, feminine, animate, mixed₁, mixed₂, the word ‘child.’

Languages of the Algonquian family in the United States, including Blackfoot, Cheyenne, Arapaho, and Ojibwa, have two noun classes: animate and inanimate. The Itonama language, spoken in Bolivia, has 17 singular classes and 5 plural classes, including classes for masculine, feminine, animate standing, animate seated, flat-round, oval, planted, liquid, long-winding, cloth, flowing, grains, pots, and canoes.

33. Gender. Gender is a well known characteristic of Indo-Hittite languages. The Anatolian branch of Indo-Hittite, which includes Hittite, had
two classes: common (masculine and feminine) and neuter. In the other branch of Indo-Hittite—Indo-European—the common class was divided into two classes, masculine, feminine, and neuter, and these three classes have been preserved in the Slavic and Germanic branches. While Latin had the same three genders, the modern Romance languages, including Spanish, Italian, French, and Rumanian, have only two genders: masculine and feminine. In the Romance languages the gender of the noun determines which form of adjectives, numbers, and demonstrative pronouns are associated with the noun. For example, in Rumanian un ciine râu ‘a bad dog’ (literally, ‘a dog bad’) for a masculine noun, but o pisică rea ‘a bad cat’ (literally, ‘a cat bad’) for a feminine noun. Rumanian has in fact three noun classes, having innovated a third class, called “mixed.” In this class nouns are masculine in the singular, but feminine in the plural, as seen, for example, in un stat bun ‘one good state,’ but două state bune ‘two good states.’

In Russian gender not only affects the elements of a noun phrase, but also is reflected on the verb: on bil ‘he was,’ ona bila ‘she was,’ ono bilo ‘it was,’ and although gender is not reflected in the first-person pronoun, it is reflected on the verb: ya bil ‘I was,’ if a man is talking, but ya bila ‘I was,’ if a woman is talking.

34. Demonstratives. The two most common demonstrative pronoun systems in the world’s languages involve two terms, as in English ‘this’ and ‘that,’ or three terms, as in Spanish éste ‘this,’ ése ‘that,’ aquel ‘that yonder.’ These two systems are represented in this database as dem:2 and dem:3. Though rare, there are also some systems with a single term: dem:1. All other distinctions, such as ‘that above, that below, that in front, that behind,’ are specified for the language individually. The Awa language of Papua New Guinea has five demonstratives, which are represented in this database as dem:5: [2 + that level over there, that above over there, that below over there]. There is also a demonstrative found in many languages that is translated by phrases such as ‘that being referred to’ or ‘that just mentioned.’ This reference pronoun, which doesn’t involve spatial orientation, is represented in this database by ‘ref.’

35. Articles. Three kinds of articles are indicated for the languages that have them. The first is the indefinite article, such as English ‘a/an,’ identified as ‘indef.’ The second is the definite article, such as English ‘the,’ identified by ‘def.’ The third type, which Greenberg (1990) called a non-generic article, is identified as ‘art.’ Non-generic articles combine both definite and indefinite functions, and often evolve into what Greenberg called Stage III articles, which indicate simply that the following (or preceding) word is a noun. When Stage III articles are inflected for, say, gender this may lead to a system of noun classes based on gender.

36. Pronouns. There is no universally accepted method of representing the typological distinctions found in pronoun systems so I decided to invent
my own, which I have tried to make both simple and transparent. What I consider the basic pronoun system is shown in Table 3. Most (but not all) languages make at least these distinctions, but most languages also embellish the basics in a number of ways. The pronouns surveyed in this work are for the most part independent pronouns like ‘I, you, we.’ However, some languages do not have independent pronouns. In such cases I have used the pronominal affixes found on verbs.

Table 3. The Basic Pronoun System

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Second-Person</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Third-Person</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

In addition to the basic pronouns shown in Table 3 some languages have an indefinite pronoun that might be translated as ‘one.’ French on and German man are two examples of such pronouns. The presence of an indefinite pronoun is indicated in this database by ‘indef.’

GENDER

If we compare the basic pronoun system with the English pronoun system shown in Table 4 we see that English comes relatively close to the basic system, with a couple of complications. A comparison of the two systems shows that 1=I, 2=you, 4=we, and 6=they. The first complication is that there is no 5, which would be the second-person plural pronoun. If we went back to Old English we would find that this complication did not exist because at that time ‘thou’ was the second-person singular pronoun (2) and ‘you’ was the second-person plural pronoun (5). During the past millennium ‘thou’ disappeared from almost all English dialects and was replaced by ‘you,’ which originally was strictly plural. One might say that ‘you’ now serves the function of both 2 and 5, but it is clear that it is today basically singular. Evidence for this is the fact that English speakers feel a need to fill the hole where 5 should be. This is why different dialects have invented new forms of 5—you’s, y’all, you guys—none of which have yet established themselves in the standard language.

The second complication is that there are three forms of 3: masculine ‘he,’ feminine ‘she,’ and neuter ‘it.’ Probably most laymen would assume that a
distinction between at least ‘he’ and ‘she’ would be almost universal since every society has men and women, but as one examines the pronominal systems given in this database one finds just the opposite. Gender is a relatively rare trait in pronom systems and most of the world’s languages do not have different words for ‘he’ and ‘she,’ just one third-person pronoun that has nothing to do with gender. In the pronoun tables given in this database gender is indicated by a letter following the pronoun (m=masculine, f=feminine, n=neuter) so that the English pronoun system is represented as shown in Table 5.

Table 5. The English Pronoun System

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Second-Person</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Third-Person</td>
<td>3mfn</td>
<td>6</td>
</tr>
</tbody>
</table>

NOUN CLASSES

As we saw above, gender is one example of the phenomenon of noun classes, in which nouns are divided into classes distinguished by a certain trait—in this case gender—but other systems of noun classes are based on other traits, and sometimes the classes are arbitrary sets of nouns. In English, gender has been lost in nouns, but still persists in the pronoun system. In Russian, nouns are still divided into three classes (masculine, feminine, neuter) and there are three different third-person pronouns, corresponding to the noun classes: on ‘he,’ ona ‘she,’ ono ‘it.’ Noun classes are indicated for those languages that have them, but details of the various systems are not always given because they are too complex for a work of this nature.

In the Grebo language of Liberia the third-person singular and plural pronouns have two forms, one for humans (h) and one for non-humans (H), as seen in Table 6.

Table 6. The Grebo Pronoun System

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Second-Person</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Third-Person</td>
<td>3hH</td>
<td>6hH</td>
</tr>
</tbody>
</table>

In the Dagbani language of Ghana the third-person singular and plural pronouns also have two forms, one animate (a) and one inanimate (i), as seen in Table 7.
Table 7. The Dagbani Pronoun System

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Second-Person</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Third-Person</td>
<td>3ai</td>
<td>6ai</td>
</tr>
</tbody>
</table>

In the Gadaba language of India the third-person singular and plural pronouns have two forms, one masculine (m) and one non-masculine (M), as seen in Table 8.

Table 8. The Gadaba Pronoun System

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Second-Person</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Third-Person</td>
<td>3mM</td>
<td>6mM</td>
</tr>
</tbody>
</table>

One might note that the pronominal embellishments considered up to this point seem to be associated with third-person pronouns and a perusal of the pronoun systems listed in this database shows that this is often true, but that there are cases in which gender (and similar traits) have spread to the second and first-person pronouns, as we will see below. The classic study of the origin and evolution of gender is Greenberg (1990).

INCLUSIVE-EXCLUSIVE

A different kind of trait is the inclusive-exclusive distinction that is found only in first-person plural pronouns. Many of the world’s languages have two first-person plural pronouns, one called inclusive (i), the other exclusive (e). Inclusive we (4i) means ‘we including the person spoken to,’ whereas exclusive we (4e) means ‘we excluding the person spoken to.’ An example of this distinction is found in the Blackfoot language of Montana, as seen in Table 9.

Table 9. The Blackfoot Pronoun System

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Person</td>
<td>1</td>
<td>4i</td>
</tr>
<tr>
<td>Second-Person</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Third-Person</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

NUMBER

Another way that pronominal systems vary is in the category of number. The systems considered so far have only singular and plural, but many languages also have a dual (d) form to indicate exactly two, and some also have a trial (t) to indicate exactly three. A few languages have a paucal (p), which indicates ‘a few.’ Dual and trial pronouns are special kinds of plural pronouns, representing exactly two in the case of duals, and exactly three
in the case of trials, and this is indicated by the d or t following the normal plural pronoun, 4d 5d 6d 4t 5t 6t, all of which are single pronouns, just as are 1 2 3 4 5 6. The Southern Kiwai language of Papua New Guinea has both dual and trial pronouns, as seen Table 10.

<table>
<thead>
<tr>
<th>Table 10. The Southern Kiwai Pronoun System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>First-Person 1</td>
</tr>
<tr>
<td>Second-Person 2</td>
</tr>
<tr>
<td>Third-Person 3</td>
</tr>
</tbody>
</table>

Up to this point the pronoun systems we have seen involve either traits like gender, or inclusive-exclusive, or number. These are, however, independent traits that often co-occur in pronoun systems. In the Xù language of Angola (also called Kxoe) gender occurs with all three numbers, as seen in Table 11 (c in the dual and trial forms indicates common gender [masculine + feminine]). In this table forms like 4dmfc represent three pronouns, 4dm, 4df, 4dc, just as 3mfn represents the three pronouns 3m, 3f, and 3n. There are thus 24 different pronouns represented in Table 11.

<table>
<thead>
<tr>
<th>Table 11. The Xù Pronoun System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>First-Person 1</td>
</tr>
<tr>
<td>Second-Person 2</td>
</tr>
<tr>
<td>Third-Person 3</td>
</tr>
</tbody>
</table>

In the Qxù language of Namibia (also called !Kung) pronouns combine gender, number, and the inclusive-exclusive distinction, as seen in Table 12. There are in this language six first-person plural pronouns: 4mi, 4me, 4fi, 4fe, 4ci, and 4ce.

<table>
<thead>
<tr>
<th>Table 12. The Qxù Pronoun System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>First-Person 1</td>
</tr>
<tr>
<td>Second-Person 2</td>
</tr>
<tr>
<td>Third-Person 3</td>
</tr>
</tbody>
</table>

Table 13 shows the numbers and letters used in representing pronoun systems. One might note that i may mean either ‘inanimate’ or ‘inclusive.’ This does not lead, however, to any ambiguity since inanimate i almost always occurs with animate a, predominantly in third-person pronouns, whereas inclusive i only occurs in first-person plural pronouns, almost always in conjunction with exclusive e. A few other parameters that are extremely rare are handled individually for the languages concerned.
Table 13. Parameters of Pronoun Systems

1: first-person singular    h: human
2: second-person singular   H: non-human
3: third-person singular    m: masculine
4: first-person plural      M: non-masculine
5: second-person plural     f: feminine
6: third-person plural      F: non-feminine
d: dual                    c: common (masculine + feminine)
t: trial                   n: neuter
p: paucal                  v: vegetable
i: inclusive               P: plain
e: exclusive               s: specific
a: animate                 S: non-specific
i: inanimate

In some parts of the world pronouns may express politeness, social standing or the like. Since there is little cross-linguistic comparability for such forms they are treated individually on a language by language basis. In many languages third-person pronouns are really demonstrative pronouns functioning as independent pronouns. Such forms are indicated by a preceding * (e.g. *3 *6d *6). Forms enclosed in parentheses are rare or marginal in the given language, thus 4(ie) might indicate that an inclusive-exclusive distinction exists, but is seldom used.

37. Word Order. In this database I have surveyed seven kinds of word order: (1) Subject (S)-Verb (V)-Object(O), (2) Adjective (A)-Noun (N), (3) Genitive (G)-Noun, (4) Demonstrative Pronoun (D)-Noun, (5) Number (NUM)-Noun, (6) Possessive Pronoun (POSS)-Noun, and (7) Noun Phrase: Demonstrative-Number-Adjective-Noun. In the first case both subject and object are nouns. In many languages the order of S, O, and V differs when S and O are nouns from that used with pronouns. For example, in French the order with nouns is SVO, *Le père voit son fils* ‘The father sees his son,’ but with pronouns the order is SOV, *Il le voit* ‘He sees him.’

In English the seven categories of word order are represented as:

SVO AN GN/NG DN NUM-N POSS-N D+Num+A+N

GN/NG indicates that English has both constructions (e.g. the boy’s book [GN] vs. the book of the boy [NG]). In some cases a language will have two constructions for a given category, but one of them is clearly dominant. An example is the order of adjectives and nouns in French. Generally the adjective follows the noun, e.g. *le livre noir* ‘the black book,’ but a few adjectives precede the noun, e.g. *un bon livre* ‘a good book.’ In such cases the less common variant is enclosed in parentheses: NA/(AN).

38. Ergative. In Indo-European languages such as Latin the subject of a transitive verb and the subject of an intransitive verb are marked in the
same way, by a nominative suffix, whereas the object of a transitive verb is marked by an accusative suffix. In ergative languages the subject of an intransitive verb and the object of a transitive verb are marked in the same way by an absolutive affix, whereas the subject of a transitive verb is marked by an ergative affix. I have indicated ergative languages by ‘erg.’ However, ergative languages are often not overtly identified in the literature so there are certainly many more ergative languages than those identified in this database. This category will have to be fleshed out in considerably more detail before the exact distribution of ergative and accusative languages becomes clear.

39. Noun Incorporation. Noun incorporation is a process, common in some languages, whereby a noun is incorporated into a verb, forming a new lexical item. It is somewhat similar to compounding, such as English ‘birdwatching.’ Languages reported to have noun incorporation are identified here by ‘noun incorp.’ However, as with the case of ergative, there are many languages with noun incorporation that are not so identified in this database so that further work is needed before the exact distribution of noun incorporation is clear.

40. Prepositions/Postpositions. In English, prepositions precede the noun, as in ‘at that building.’ In other languages, however, such items follow the noun and are therefore called postpositions. An example is Japanese ano biru de ‘that building at.’ Prepositions are indicated by PREP and postpositions by POST.

41. Comments. This field contains comments which apply to individual languages, sometimes clarifying aspects of the language discussed in the other fields.

REFERENCES

ABBREVIATIONS
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Title</th>
<th>Editor(s)</th>
<th>Publisher/Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUDP</td>
<td><em>Cornell University Data Paper</em>. Ithaca, N.Y.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUP</td>
<td><em>Cambridge University Data Paper</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>dialect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFL</td>
<td><em>Daly Family Languages, Australia</em>, D. T. Tryon. 1974. PL C32.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diss</td>
<td>Dissertation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td><em>Encyclopaedia Britannica</em>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSAIL</td>
<td><em>The Grouping of South American Indian Languages</em>, Mary Ritchie Key. Tübingen, West Germany, 1979.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUPLL</td>
<td>[Riley (Socio) &amp; Lado (psycho)] probably Georgetown University Publications in Languages and Linguistics.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFAN</td>
<td>Institut français d’Afrique noire. Dakar, Senegal.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IL  Indian Linguistics. Pune, India.
IU  Indiana Univ.
J  Journal
Lg  Language. Baltimore.
LNA  The Languages of Native America, ed. Lyle Campbell and Marianne Mithun. Austin, Tex., 1979.
LSA  Linguistic Society of America.
LSNBB  Linguistic Survey of the Northern Bantu Borderland, 4 vols. London.
LTS [?check] Languages and their Speakers, ed. Timothy Shopen. Cambridge, Mass., 1979. [Maninka, Malagasy, Guugu Yimidirr; check also LT Status]


MKS Mon-Khmer Studies. Honolulu.


MSILI Miscellaneous Studies in Indonesian and Languages in Indonesia. Jakarta.


NATS Native American Texts Series. Chicago.


OL Oceanic Linguistics. Honolulu.

OLM Oceania Linguistic Monographs. Sydney.

OLSP Oceanic Linguistic Special Publications. Honolulu.


PNG Papua New Guinea


SIL Summer Institute of Linguistics. Arlington, Tex. [Publication series issued at Arlington; individual works published by SIL branches around the world]


SLLR Sierra Leone Language Review. Fourah Bay, Sierra Leone.


SSL South Sea Languages, vol. 2, D. MacDonald. Melbourne, 1891.


SUGIA Sprache und Geschichte in Afrika. Hamburg.


TNL Twelve Nigerian Languages, ed. Elizabeth Dunstan. New York,
1969.


USA   United States of America


ZDMG  *Zeitschrift der Deutschen Morgenländischen Gesellschaft*. 