

The origin and evolution of word order

Murray Gell-Mann^{a,1} and Merritt Ruhlen^{b,1}

^aSanta Fe Institute, Santa Fe, NM 87501; and ^bDepartment of Anthropology, Stanford University, Stanford, CA 94305

Contributed by Murray Gell-Mann, August 26, 2011 (sent for review August 19, 2011)

Recent work in comparative linguistics suggests that all, or almost all, attested human languages may derive from a single earlier language. If that is so, then this language—like nearly all extant languages—most likely had a basic ordering of the subject (S), verb (V), and object (O) in a declarative sentence of the type “the man (S) killed (V) the bear (O).” When one compares the distribution of the existing structural types with the putative phylogenetic tree of human languages, four conclusions may be drawn. (i) The word order in the ancestral language was SOV. (ii) Except for cases of diffusion, the direction of syntactic change, when it occurs, has been for the most part SOV > SVO and, beyond that, SVO > VSO/VOS with a subsequent reversion to SVO occurring occasionally. Reversion to SOV occurs only through diffusion. (iii) Diffusion, although important, is not the dominant process in the evolution of word order. (iv) The two extremely rare word orders (OVS and OSV) derive directly from SOV.

Recent work in genetics (1), archeology (2), and linguistics (3) indicates that all behaviorally modern humans share a recent common origin. The date involved is often identified with the sudden appearance, roughly 50,000 y ago, of strikingly modern behavior in the form of more sophisticated tools as well as painting, sculpture, and engraving. This new Upper Paleolithic culture differed dramatically from the Mousterian culture of the anatomically modern humans from whom the behaviorally modern humans emerged. The cause of this abrupt change has been attributed to the appearance of fully modern human language (2, 4), and this is a plausible conjecture. With regard to language, Bengtson and Ruhlen (3) have presented evidence that suggests that all or almost all attested human languages share a common origin. That origin need not necessarily refer all of the way back to the time when behaviorally modern humans emerged and peopled the Old World. There could have been a “bottleneck” effect at a much later time, with a single language spoken then being ancestral to all or most attested languages (5). If that is so, then that ancestral language, like nearly all modern languages, must have had a dominant ordering of the subject (S), verb (V), and object (O) in simple declarative sentences such as “the man (S) killed (V) the bear (O).” One should note that there is great variation in the rigidity of the basic word order in different languages, in part due to the fact that the syntactic functions of subject and object are often marked on the noun, as in Russian, which permits all six possible orders to yield grammatical sentences. Nonetheless, the basic word order of Russian is clearly SVO, and the other orders reflect special emphasis or other pragmatic factors. Australian languages, in particular, are known for their extremely free word order, and it has been claimed that some of those languages have no basic order. Still, as we shall see, the basic word order reported for most Australian languages is normally SOV, although other orders are also found.

Greenberg (6) noted that of the six possible orders, only three are commonly found: SOV, SVO, and VSO. The great insight of Greenberg’s paper, however, was not just an inventory of existing types—which obviously was long overdue—but the recognition that there were strong correlations between what seemed to be unrelated syntactic structures. Thus, for example, an SOV language usually places the genitive before the noun (GN; e.g., “the man’s dog”) and uses postpositions, whereas a VSO language usually places the genitive after the noun (NG; e.g., “the dog of the

man”) and uses prepositions. (Nowadays, these correlations are described in terms of head-first and head-last constructions.) In light of such correlations it is often possible to discern relic traits, such as GN order in a language that has already changed its basic word order from SOV to SVO. Later work (7) has shown that diachronic pathways of grammaticalization often reveal relic “morphotactic states” that are highly correlated with earlier syntactic states. Also, internal reconstruction can be useful in recognizing earlier syntactic states (8). Neither of these lines of investigation is pursued in this paper.

It should be obvious that a language cannot change its basic word order overnight. What is required is a long gradual process during which it is the frequencies of different word orders that change. A language may begin with a high frequency of SOV and a low frequency of SVO. As the language changes, the frequency of SVO may increase at the expense of SOV until there emerges a stage referred to as “free word order,” in which the frequencies of both orders are similar. A final stage may occur when the frequency of SVO becomes high and that of SOV low. It is here that both grammaticalization and internal reconstruction have played and will continue to play a crucial role in further elucidating the precise processes of diachronic change that lead from one state to another.

Research subsequent to Greenberg’s has shown that the other three possible orders—VOS, OVS, and OSV—also occur, but the last two are exceedingly rare (9). We have analyzed the distribution of these six word orders for a sample of 2,135 languages in terms of a presumed phylogeny of the world’s languages (10). The data on which this paper is based are given in *SI Appendix*.

In collecting data on basic word order in the world’s languages there is no doubt that some errors will occur, because most sources do not specify the basic word order and, in languages with relatively free word order, it is not always easy to determine what the basic word order really is. In other cases, different sources give different word orders for the same language. Nonetheless, we do not believe that such errors as may exist will affect our conclusions. We conclude that (i) if there was a language from which all or most attested languages derive, it had the word order SOV [this conclusion supports the conjecture of Givón (11)]; (ii) except in cases of diffusion, the direction of change, when it occurs, has been mostly SOV > SVO > VSO/VOS with occasional reversion to SVO, but not to SOV; (iii) diffusion, although important, is not the dominant process in the evolution of word order; and (iv) the unusual orders OVS and OSV appear to derive directly from SOV.

Of these four conclusions, the second requires further comment. In word order change, the progression SOV > SVO or sometimes VSO seems to have no exceptions apart from cases of diffusion, but the other progression SVO > VSO/VOS has a number of counterexamples. Givón (12) discusses the shift

Author contributions: M.G.-M. and M.R. designed research; M.R. performed research; M.G.-M. and M.R. analyzed data; and M.G.-M. and M.R. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in A Global Linguistic Database, <http://starling.rinet.ru/cgi-bin/main.cgi?flags=eygtnnl>.

¹To whom correspondence may be addressed. E-mail: ruhlen@stanfordalumni.org or mgm@santafe.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1113716108/-DCSupplemental.

from VSO to SVO in Biblical Hebrew and suggests that a similar change appears to have taken place in Luo and Indonesian. England (13) argues for the same change in the Mayan family. A similar shift in the Austronesian family from VSO/VOS to SVO and then back to VSO is discussed below.

In connection with the arrow of time $SOV > SVO > VSO/VOS$, we are discussing two different progressions. One has SOV mutating to SVO (or perhaps occasionally VSO), but not the reverse (back to SOV). According to Givón, “To my knowledge all documented shifts to SOV from $VO \dots$ can be shown to be *contact induced*” (12), a conclusion also arrived at by Tai (14) and Faarlund (15).

The other progression has $SOV > SVO > VSO/VOS$ or sometimes $SOV > VSO > SVO$. Givón has emphasized the latter: “It seems that natural word-order drift follows the paradigm $SOV > VSO > SVO$ as a major typological continuum” (12). Here we disagree with Givón. We find, on the basis of the distribution of word order types, that in natural drift we have $SOV > SVO$ far more often than $SOV > VSO$. There are many known cases, such as English and some Romance languages, where historical records show that SOV has become SVO without an intervening VSO stage.

Fig. 1 illustrates the possible directions of word order change, with the heavy lines indicating the most frequent changes caused by natural drift without diffusion and the other lines indicating other possible changes. These suggested diachronic paths seem to support Dryer’s proposed revision of the traditional view of typology (16). A still more radical simplification would be to drop references to the subject S , in which case we are left with $VO > OV$ only through diffusion, although $OV > VO$ occurs by natural drift as well.

The traditional typology treats the differences between SOV and SVO , between SVO and VSO , and between VSO and VOS on a par. However, the first of these differences is a fundamental one, because they differ in the order of verb and object. The second of these differences is intermediate in importance; they are similar with respect to the important parameter of order of verb and object but they differ with respect to the lesser parameter of order of verb and subject. The third of these differences is the least important, and it is ignored in the typology proposed here because it is not a difference that is predictive of anything else (16).

Vennemann (17) represented possible word order changes as in Fig. 2. According to Vennemann (17), (i) an SOV language can change only to SVO ; (ii) an SVO language can change to VSO or become a free word order language (FWO) in which S and O may be marked by affixes, as in Russian; (iii) a VSO language can sometimes revert to SVO or become an FWO language; and (iv) “a free word order language [may] gradually develop toward the universally preferred SOV type” (18). This last point obviously contradicts Givón’s claim that all shifts to SOV are due to diffusion, and Vennemann gives no examples of such a shift. We will discuss one alleged example of the change $SVO > SOV$ below, but we believe that Givón is basically correct and that the reason there is a large number of languages with SOV word order is not because SOV word order is “universally preferred” but because in many languages it is unchanged from the original order.

In discussing this same question, Harris and Campbell conclude that “Tai’s and Faarlund’s hypothesis that SOV arises in a language only due to contact with other SOV languages is interesting, but clearly overstated. . . . If new SOV languages arose only from contact with older SOV languages, then where did the

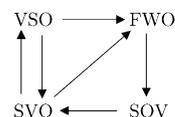


Fig. 2. Possible word order changes (Vennemann).

prior SOV languages come from; and if they too are assumed to be due to contact with SOV languages, then how did the very first SOV language come about?” (19). We suggest that the very first SOV language was in fact the language from which all or most attested languages derive and that most modern languages with SOV word order merely preserve this initial state, except for cases where SOV has been borrowed from neighbors.

It should be noted that our conclusions are at variance with two commonly accepted and seemingly unrelated assumptions: (i) Linguistic evolution is never unidirectional, and (ii) it is difficult, if not impossible, to reconstruct syntax, even for recent and well-studied families such as Indo-Hittite. With respect to the first assumption, Harris and Campbell have claimed that “there is little or no evidence to support hypotheses that languages—or their syntax—are evolving in a single direction through non-renewable changes” (20). With regard to the second assumption, Fox summarizes the current view as follows: “Syntactic reconstruction is a controversial area. . . . Indeed, there is a consensus among many scholars that it is difficult, if not impossible, to carry over into the field of syntax the methods—especially the Comparative Method itself—that have proved so successful in phonology” (21).

Despite these assumptions, Givón (11) noted, 30 y ago, that most of the world’s language families are either predominantly SOV today or derive demonstrably from an earlier stage that was SOV , at least as far back as 7,000 or 8,000 y ago, which at that time was considered the temporal limit of comparative linguistics. Givón further proposed that an original language from which all or most attested languages derive would have necessarily had the word order SOV as a recapitulation of the order found in language acquisition—from single-clause to multipropositional discourse—and he implied that during the unknown interval between the era of this ancestral language and that of 8,000-y-old families the word order would have remained predominantly SOV . Finally, he proposed that syntactic change has been almost exclusively $SOV > VSO > SVO$. Our data fully support all of Givón’s conjectures, with the exception that we find that when SOV changes as a result of drift it usually becomes SVO first, and only then (if at all) VSO . Clearly, the precise diachronic processes that gradually change one word order into another warrant further investigation, particularly from a cross-linguistic perspective.

It should be noted that the concept of free word order is really a misnomer, seemingly implying that a language starts in one state, say SOV , enters a period of free word order—where any order becomes possible—and finishes in a different state, say SVO . However, examination of those languages where two word orders are in serious competition in terms of frequency, or are used in different constructions, shows that not all of the 15 possible combinations occur. In our language sample there were 125 languages with two competing word orders (*SI Appendix*); the number of languages with each combination is given in Table 1. As can be seen, by far the two most common combinations are SOV/SVO , and then SVO/VSO , as expected from the two heavy arrows in Fig. 1. Also important are VSO/VOS , SVO/VOS , SOV/OVS , and SOV/OSV , as expected from the thin arrows in Fig. 1. The remaining five combinations, found in only one or two languages, may be due in part to errors in analysis of these languages. Presumably, the combinations that do occur indicate the changes that are most common and thus support the evolution of word order proposed in Fig. 1.

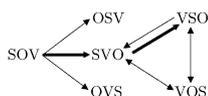


Fig. 1. Evolution of word order.

Table 1. Languages with mixed word order

| | |
|---------|----|
| SOV/SVO | 46 |
| SVO/VSO | 24 |
| VSO/VOS | 17 |
| SVO/VOS | 11 |
| SOV/OVS | 9 |
| SOV/OSV | 6 |
| SVO/OVS | 4 |
| SOV/VOS | 2 |
| SOV/VSO | 2 |
| VOS/OVS | 2 |
| SVO/OSV | 1 |
| VOS/OSV | 1 |

A problem could arise if there were numerous cases of borrowed word order not corresponding to our arrows but leading to mixed word orders. Because our Fig. 1 is not meant to include cases of diffusion, the agreement between Fig. 1 and Table 1 could be spoiled. However, Table 1 looks quite clean in this respect, so we do not seem to have much of a problem.

Because we conclude that the transmission of word order is to a great extent vertical (genetic), as opposed to horizontal (areal), we shall examine the distribution of the six word order types in terms of a tentative phylogenetic tree for languages. See Table 2, where each of the nodes is supported by published evidence. There will no doubt be refinements to this tree, but we do not think that such corrections will affect our conclusions. It is clear from Table 2 that SOV is the most frequent order, followed closely by SVO, with VSO a distant third. The other three word orders (VOS, OVS, OSV) are comparatively rare. Taxonomy, however, is not always democratic, and sheer numbers often count for little. Despite the fact that of the roughly 4,000 recent species of mammals all but 6 give live birth, biologists know that it is the 6 species that lay eggs that preserve the original state simply because the nearest outgroup to mammals—the reptiles—is almost exclusively egg-laying. There are, as we shall see in the discussion below, many cases where word order is essentially uniform in a family (excluding diffusion) and therefore can be presumed to represent the initial state (e.g., Altaic). It is in families with more than one word order (e.g., Indo-Hittite) that outgroup comparison may be used to determine the original word order. Whether the outgroup is the first branch in a family or the nearest relative to a family does not matter taxonomically.

Indo-Hittite

Let us begin with the Indo-Hittite family, the most intensively studied of all families, and one where the original word order is still considered controversial. It is now generally accepted that the Indo-Hittite family consists of two branches, Anatolian and Indo-European. (Here many scholars prefer to use “Indo-European” to mean what we call Indo-Hittite.) The Anatolian word order is strictly SOV, whereas Indo-European shows different orders in different branches: SOV in Tocharian, Indic, Iranian, Italic, and (early) Germanic; SVO in Greek, Armenian, Albanian, and Baltic; and VSO in Celtic and perhaps (early) Slavic. Because Anatolian, the nearest outgroup to Indo-European, is strictly SOV and Indo-European is partially SOV, we may conclude that both Indo-European and Indo-Hittite were originally SOV. This conclusion coincides with that of Lehmann (22), which was based on internal linguistic considerations using Greenberg’s word order correlations, not the taxonomic evidence we are emphasizing here. Lehmann also noted that even before the Anatolian branch was discovered in the early 20th century, scholars such as Brugmann had concluded that Indo-European was originally SOV.

Table 2. Distribution of word order types in the world’s languages

| | |
|--------------------|-------------------------|
| World | 1008-770-164-[40-16-13] |
| Khoisan | 22-11-1-[0-0-0] |
| Congo-Saharan | 61-318-16-[0-1-0] |
| Niger-Kordofanian | 39-279-1-[0-0-0] |
| Nilo-Saharan | 22-39-15-[0-1-0] |
| Indo-Pacific | 223-25-1-[0-0-2] |
| Australian | 59-20-1-[3-1-1] |
| Austriac | 30-220-67-[16-0-2] |
| Austroasiatic | 8-34-0-[1-0-0] |
| Miao-Yao | 0-4-0-[0-0-0] |
| Daic | 1-19-0-[0-0-0] |
| Austronesian | 21-163-67-[15-0-2] |
| Dene-Caucasian | 157-13-0-[0-0-0] |
| Basque | 1-0-0-[0-0-0] |
| Caucasian | 29-0-0-[0-0-0] |
| Burushaski | 1-0-0-[0-0-0] |
| Sino-Tibetan | 84-13-0-[0-0-0] |
| Ket | 1-0-0-[0-0-0] |
| Na-Dene | 41-0-0-[0-0-0] |
| Nostratic-Amerind | 456-163-78-[21-14-8] |
| Afro-Asiatic | 58-37-14-[0-0-0] |
| Nostratic | 182-59-6-[0-0-0] |
| Kartvelian | 4-0-0-[0-0-0] |
| Dravidian | 28-0-0-[0-0-0] |
| Eurasianic | 149-59-6-[0-0-0] |
| Indo-Hittite | 79-47-6-[0-0-0] |
| Uralic | 10-10-0-[0-0-0] |
| Altaic | 50-1-0-[0-0-0] |
| Ainu | 1-0-0-[0-0-0] |
| Gilyak | 1-0-0-[0-0-0] |
| Chukchi-Kamchatkan | 2-1-0-[0-0-0] |
| Eskimo-Aleut | 6-0-0-[0-0-0] |
| Amerind | 216-67-58-[21-14-8] |

The numbers after each family represent the number of languages with SOV, SVO, VSO, VOS, OVS, and OSV orders, given in that order, with the final three word orders in brackets. Note that we have chosen one of the several definitions of Nostratic.

Uralic

The Uralic family has three primary branches—Finno-Ugric, Samoyed, and Yukaghir. Samoyed and Yukaghir are exclusively SOV. Finno-Ugric itself has two primary branches, Ugric and Finnic. Ugric is also SOV, except for Hungarian, which has adopted SVO word order from surrounding Indo-European languages although still maintaining traces of an earlier SOV word order. In Finnic languages one finds both SOV and SVO, although in some cases languages which are today SVO are known to have had an earlier SOV word order (e.g., Estonian). Clearly, the original Uralic word order must have been SOV, as is generally assumed by Uralicists (23).

Nostratic

The Indo-Hittite and Uralic families belong to the Eurasianic macrofamily. The other branches of Eurasianic are Altaic (which includes the Turkic, Mongolian, and Tungusic languages, as well as Korean and Japanese), Chukchi-Kamchatkan, Eskimo-Aleut, Gilyak, and probably Ainu. To these we should most likely add the closely related Kartvelian and Dravidian languages, yielding the current definition of Nostratic. In all of these the word order is SOV, with three exceptions: (i) In Altaic, one Turkic language (Gagauz), spoken in Rumania, has adopted the order SVO under Rumanian influence. (ii) In the Eskimo-Aleut family, Aleut

has a rather rigid SOV word order, whereas in the Eskimo languages SVO word order is fairly common although SOV is the basic order (24). (iii) Chukchi-Kamchatkan also has both SOV and SVO, even in a single language (25). Both Kartvelian and Dravidian are exclusively SOV. We may conclude therefore that Nostratic itself was SOV.

Afro-Asiatic

A close relative of the Nostratic macrofamily is Afro-Asiatic. Together they constitute what Illich-Svitych called “Nostratic” (26). In Afro-Asiatic all three basic word orders are well-attested, but the original order was most probably SOV. Although there is no consensus on the subgrouping of the Afro-Asiatic family, Ehret (27) has proposed, on the basis of both lexical and phonological innovations, the subgrouping shown in Table 3. We have added the characteristic word order for each branch; Ehret did not consider syntax in his analysis. As can be seen, if Ehret’s tree is correct, the original Afro-Asiatic order comes out SOV and the direction of change follows exactly the pattern proposed in this paper.

Amerind

The Amerind macrofamily is one of the few that have languages with all six possible orders. The distribution of word order in this family is given in Table 4, with data on the three rare word orders given in brackets in the order VOS, OVS, OSV. Every branch except Almosan contains at least some SOV languages, and in many branches this order is either the only one found or overwhelmingly predominant (Keresiouan, Hokan, Tanoan, Chibchan, Paezan, Andean, Macro-Tucanoan, Macro-Panoan, Macro-Ge). In addition, Uto-Aztecan is considered to have originally been SOV (28), although both SVO and VSO are found in contemporary languages. Similarly, although most modern languages in the Iroquoian branch of Keresiouan have SVO word order, Rudes (29) reconstructs SOV for Proto-Iroquoian. Given these data, the hypothesis that Proto-Amerind was an SOV language would seem to be the most parsimonious.

Table 4 also suggests that the two rare word orders, OVS and OSV, derive directly from SOV because, for example, in the Paezan, Andean, Macro-Tucanoan, Macro-Carib, and Macro-Ge families almost all of the languages are SOV except for those with OVS or OSV word order. In addition to this external evidence, analysis of individual languages with OVS or OSV word order often shows that SOV is an alternate word order in these languages, sometimes in particular syntactic constructions, sometimes in almost free variation with OVS or OSV (9, 30). (See also Table 1). In the Carib family, for example, Hixkaryana—perhaps the best-known OVS language—has SOV as the only significant variant order and, in the same family, Apalai shows only a slight preference for OVS over SOV, and Bacairi is either an OVS language or an SOV language on the way to becoming OVS. In the Ge family, Chavante is OSV, but other Ge languages are

Table 3. The Afro-Asiatic macrofamily

| | |
|--------------------|-----|
| Afro-Asiatic | SOV |
| Omotiic | SOV |
| Erythraic | SOV |
| Cushitic | SOV |
| Chado-Afro-Asiatic | SVO |
| Chadic | SVO |
| North Afro-Asiatic | VSO |
| Ancient Egyptian | VSO |
| Semitic-Berber | VSO |
| Semitic | VSO |
| Berber | VSO |

Table 4. The Amerind macrofamily

| | |
|----------------|---------------------|
| Amerind | 216-67-58-[21-14-8] |
| Almosan | 0-9-15-[0-0-0] |
| Keresiouan | 14-2-0-[0-0-0] |
| Penutian | 14-10-12-[6-0-0] |
| Hokan | 15-3-2-[1-0-0] |
| Tanoan | 2-0-0-[0-0-0] |
| Uto-Aztecan | 17-6-3-[0-1-0] |
| Oto-Manguean | 1-4-13-[4-0-0] |
| Chibchan | 20-2-0-[1-0-0] |
| Paezan | 12-2-0-[0-0-1] |
| Andean | 10-3-0-[0-1-1] |
| Macro-Tucanoan | 14-1-2-[0-3-3] |
| Equatorial | 29-17-9-[9-2-2] |
| Macro-Carib | 6-1-1-[0-7-0] |
| Macro-Panoan | 50-7-0-[0-0-0] |
| Macro-Ge | 12-0-1-[0-0-1] |

SOV; and in the Tupi family, Urubu is OSV, but has SOV as a principal variant. All of this suggests that the two extremely rare word orders are direct mutations of the SOV word order. Other examples of OSV or OVS found outside of Amerind are similarly associated with SOV.

That VOS is basically a variant of VSO is suggested by the fact that VOS appears only in those branches of Amerind that contain VSO languages (with one exception). We will see the same pattern below with regard to Austronesian.

There is some evidence for a rather close linkage of Afro-Asiatic and Nostratic with Amerind (31, 32), and all three are SOV. There is also evidence of a linkage with Austric and Dene-Caucasian, but here we run into the Austric innovation SVO.

Dene-Caucasian

The Dene-Caucasian macrofamily consists of six branches, three of which are today single languages. As can be seen in Table 2, five of these branches are exclusively SOV. The other branch, Sino-Tibetan, has both SOV and SVO orders, but of the 250 or so Sino-Tibetan languages all have SOV word order with only three exceptions—Chinese, Bai, and Karen, which are SVO. It is usually assumed that these languages borrowed SVO word order from surrounding languages, so the hypothesis that Sino-Tibetan was originally SOV is generally accepted.

Let us turn now to the other five macrofamilies appearing as primary branches in Table 2. They include languages of sub-Saharan Africa, Southeast Asia, and Oceania.

Austric

Of the seven primary nodes in Table 2, Austric shows the least trace of SOV word order; indeed, we will argue that Proto-Austric was SVO and that existing instances of SOV are all later developments. The Austric macrofamily consists of four branches: Austroasiatic, Miao-Yao, Daic, and Austronesian. Austroasiatic consists of two parts, Munda and Mon-Khmer. Munda is strictly SOV, whereas Mon-Khmer is strictly SVO, with two exceptions. On the basis of internal linguistic evidence (similar to that used by Lehmann with regard to Indo-Hittite), Pinnow (33) argued that Proto-Munda was likely SVO, as indicated by the presence of prepositions and the fact that SVO is the normal order for subject and object pronouns. If Pinnow is correct, then Austroasiatic would have originally been SVO, like Mon-Khmer. That Munda should have borrowed SOV word order is highly plausible because the family is located in India, where virtually all languages (of whatever family) are SOV. The second branch of Austric, Miao-Yao, is strictly SVO, as is the Daic branch, with one exception.

The Austronesian family has two kinds of verbal syntax. In the “transitive” type the order is typically SVO, but in the “focus” type the order is either VSO or VOS, with the order of the subject and object apparently free (34). Taxonomic considerations within Austronesian favor VSO/VOS as the original word order, because the Austronesian languages of Taiwan are almost exclusively VSO/VOS (one language has borrowed SVO word order from Chinese) and the other Austronesian languages (Malayo-Polynesian) show this order as well as SVO and SOV, the latter exclusively in languages that have been in contact with Indo-Pacific languages along the coast of New Guinea or on surrounding islands. This conclusion coincides with that of Pawley and Reid: “Verb-initial word order is found in Toba Batak and Merina as well as in Philippine and Formosan languages, and we assume that it was the preferred order in [Proto-Austronesian]. However, verb-initial languages allow or require subjects to be clause-initial in some contexts . . . so that the precondition for a change to S-V-O . . . was no doubt always present” (35). Although Proto-Austronesian seems to have had VSO/VOS as the preferred word order, the Proto-Oceanic subgroup is reconstructed with SVO word order and, within Proto-Oceanic, Proto-Polynesian is reconstructed with VSO word order. There has, thus, been an alternation within the Austronesian family between VSO/VOS and SVO word order, an alternation that perhaps goes back as far as Austric. We conclude that Austric was originally SVO and that only the Austronesian branch has changed this word order to VSO/VOS, and later, in some languages, back to SVO.

Australian

As we noted above, the Australian family is known for its exceptionally free word order, owing to the presence of inflections that identify the subject and object. In some languages this makes all six orders grammatical, as in Russian. In contradistinction to the case of Russian, however, it is not always easy to determine which order is basic, and indeed for some languages it has been claimed that there is no basic order. Whether this is really true is difficult to determine. In any event, notwithstanding the often free word order, the Australian family is generally regarded as having SOV as its most characteristic type (36, 37).

Indo-Pacific

The Indo-Pacific macrofamily (38) has over 700 languages, including almost all of the languages on New Guinea, as well as some on surrounding islands (e.g., Timor). It is a highly diverse macrofamily, but almost all of the languages that have been studied are SOV except for a few along the New Guinea coastline, and on surrounding islands, that have adopted SVO from contact with Austronesian languages. There seems little doubt that Indo-Pacific was originally SOV because virtually all its known languages still are.

Let us turn finally to the three sub-Saharan African families (39).

Niger-Kordofanian

The numbers for this macrofamily in Table 2 indicate a strong preference for SVO word order, but once again consideration of the internal subgrouping of the family suggests that SOV is more likely the primitive state. Table 5 shows the subgrouping and distribution of word order in Niger-Kordofanian. Let us begin with Niger-Congo. Of particular significance is the fact that Mande, which is strictly SOV, is coordinate with all of the rest of Niger-Congo, which is itself partially SOV. In the same way that we can argue for Proto-Mammal being an egg-layer, we can thus conclude that Proto-Niger-Congo was SOV, despite the superficial numbers that indicate otherwise.

Given these data, an original SOV word order seems most likely, with the progression of syntactic change once again following the path SOV > SVO. It is certainly significant that both Givón (40) and Hyman (41) arrived at the same conclusion on the basis of

Table 5. The Niger-Kordofanian macrofamily

| | |
|--------------------|----------|
| Niger-Kordofanian | 39-279-1 |
| Kordofanian | 4-15-1 |
| Niger-Congo | 35-264-0 |
| Mande | 22-0-0 |
| Niger-Congo Proper | 13-264-0 |
| Atlantic | 0-16-0 |
| Kru | 1-3-0 |
| Dogon | 1-0-0 |
| Gur | 8-22-0 |
| Adamawa | 0-16-0 |
| Ubangian | 0-21-0 |
| South Central | 2-52-0 |
| Broad Bantu | 0-16-0 |
| Bantu | 1-118-0 |

The numbers after each family represent the number of languages with SOV, SVO, and VSO word order, given in that order.

internal linguistic evidence similar to that used by Lehmann and Pinnow. According to Givón, “relics of an earlier SOV syntax may be found in all subgroups of Niger-Congo” (42). Because Niger-Congo was originally SOV and Kordofanian is partially SOV, it follows that Niger-Kordofanian also was most likely SOV.

It should be noted, however, that Claudi (43) has proposed that Mande was originally an SVO language that evolved into SOV through a process of grammaticalization. She further argues that Niger-Kordofanian itself also was originally an SVO language, an idea that had already been suggested by Heine (44, 45). If this analysis should turn out to be correct, it would constitute a counterexample to the syntactic arrow of time proposed in this article.

Nilo-Saharan

All three basic orders (SOV, SVO, VSO) are found in Nilo-Saharan, and there is no well-established subgrouping among the dozen or so branches of this diverse macrofamily. Nevertheless, Bender (46) has recently argued that Nilo-Saharan did originally have SOV word order. Bender subdivides Nilo-Saharan into four branches, two of which are SOV (Songhai and Saharan). The fourth branch (Satellite-Core) contains all three basic orders, whereas the languages of the third branch (Kuliak) have, according to Bender, borrowed VSO word order from the Nilotic speakers who surround them and who belong to the VSO section of the fourth branch. According to Bender, “the logical interpretation is that Nilo-Saharan was Type D [SOV] and that an innovation to Type A [SVO] spread through most of Satellite-Core (as will be seen, this agrees well with the spread of morphological innovations). Type C [VSO] is also an innovation, found among neighbouring parts of Surmic, Nilotic [in two of three branches] and also in Kuliak, which, as already noted, is surrounded today mostly by [VSO Nilotic] languages” (46).

Ehret (47) proposes a very different classification of Nilo-Saharan in which the first two branches (Koman and Central Sudanic) are SVO, the next three (Kunama, Saharan, and Fur) SOV, and the final branch (Trans-Sahel) contains all three word orders. If this classification is correct, it would contradict the pattern that we have discerned in virtually all other cases. Ehret does not, however, discuss word order in his book on Nilo-Saharan.

Khoisan

The Khoisan macrofamily consists of a Southern African group and two isolated languages, Hadza (SVO) and Sandawe (SOV). The Southern African group has three branches, Northern [SVO; but Honken (48) suggested that one Northern language,

Žu/hōasi, was originally SOV], Central (SOV), and Southern (SVO). No firm conclusions can be drawn, but the data are not incompatible with an original SOV order.

For sub-Saharan Africa, our conclusion must be that Congo-Saharan (49)—Niger-Kordofanian plus Nilo-Saharan—was in all probability SOV, and Khoisan could well have been SOV. The overall conclusion is that of the seven primary nodes in Table 2, five were originally SOV (Congo-Saharan, Indo-Pacific, Australian, Dene-Caucasian, Nostratic-Amerind), one (Khoisan) may have been SOV or possibly SVO, and one (Austic) was SVO.

Horizontal Versus Vertical Transmission

It is often supposed that horizontal transmission, including areal or Sprachbund effects, is much more important for word order than vertical (genetic) transmission. In addition, it is widely believed that so much time has elapsed since the origin of attested human languages that word order changes have washed back and forth enough to produce a kind of steady state or equilibrium. Our work has indicated, however, that neither of these ideas is correct. There is a very substantial amount of vertical transmission so that, for example, the far-flung Dene-Caucasian macrofamily exhibits a great deal of uniformity. Furthermore, it appears that the ancestor of most attested human languages was spoken recently enough that the whole evolutionary path of word order changes can in most cases still be reconstructed, including the case where the original order SOV has simply been preserved. The changes that do not arise from horizontal transmission seem to be mostly unidirectional, especially SOV > SVO > VSO/VOS.

Conclusion

The distribution of word order types in the world's languages, interpreted in terms of the putative phylogenetic tree of human languages, strongly supports the hypothesis that the original word order in the ancestral language was SOV. Furthermore, in the vast majority of known cases (excluding diffusion), the direction of change has been almost uniformly SOV > SVO and, beyond that, primarily SVO > VSO/VOS. There is also evidence that the two extremely rare word orders, OVS and OSV, derive directly from SOV.

These conclusions cast doubt on the hypothesis of Bickerton that human language originally organized itself in terms of SVO word order. According to Bickerton, "languages that did fail to adopt SVO must surely have died out when the strict-order languages achieved embedding and complex structure" (50). Arguments based on creole languages may be answered by pointing out that they are usually derived from SVO languages. If there ever was a competition between SVO and SOV for world supremacy, our data leave no doubt that it was the SOV group that won. However, we hasten to add that we know of no evidence that SOV, SVO, or any other word order confers any selective advantage in evolution. In any case, the supposedly "universal" character of SVO word order (51) is not supported by the data.

ACKNOWLEDGMENTS. We thank Joan Bybee, Bill Croft, and T. Givón for their criticism of an earlier version of this paper, and the Santa Fe Institute for its support of this research. In addition, the work of the first author was supported by the C.O.U.Q. Foundation, The Bryan J. and June B. Zwan Foundation, and Insight Venture Management, which are gratefully acknowledged. The results of this paper were first presented at a workshop on "Arrows of Time and Founder Effects in Language Evolution" organized by the authors at the Santa Fe Institute in December 1997.

- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The History and Geography of Human Genes* (Princeton Univ Press, Princeton, NJ).
- Klein RG (1999) *The Human Career* (Univ of Chicago Press, Chicago).
- Bengtson JD, Ruhlen M (1994) *On the Origin of Languages*, ed Ruhlen M (Stanford Univ Press, Stanford, CA), pp 277–336.
- Diamond J (1992) *The Third Chimpanzee* (HarperCollins, New York).
- Gell-Mann M, Peiros I, Starostin G (2011) Distant language relationship: The current perspective. *J Lang Rel* 1:13–30.
- Greenberg JH (1963) *Universals of Language*, ed Greenberg JH (MIT Press, Cambridge, MA), pp 73–113.
- Givón T (1971) in *Papers from the 7th Regional Meeting* (Chicago Linguist Soc, Chicago), pp 394–415.
- Givón T (2000) *Reconstructing Grammar*, ed Gildea S (John Benjamins, Amsterdam), pp 107–159.
- Derbyshire DC, Pullum GK (1981) Object-initial languages. *Int J Am Ling* 47:192–214.
- Ruhlen M (1991) *A Guide to the World's Languages* (Stanford Univ Press, Stanford, CA), Vol 1.
- Givón T (1979) *On Understanding Grammar* (Academic, New York), pp 271–309.
- Givón T (1977) *Mechanisms of Syntactic Change*, ed Li C (Univ of Texas Press, Austin), pp 181–254.
- England NC (1991) Changes in basic word order in Mayan languages. *Int J Am Ling* 57:446–486.
- Tai JHY (1976) *Papers from the Parasession on Diachronic Syntax*, eds Steever SB, Walker CA, Mufwene SS (Chicago Linguist Soc, Chicago), pp 291–304.
- Faarlund JT (1990) *Historical Linguistics and Philology*, ed Fisiak J (Mouton de Gruyter, Berlin), pp 165–186.
- Dryer MS (1997) On the 6-way word order typology. *Stud Lg* 21:69–103.
- Vennemann T (1973) *Syntax and Semantics*, ed Kimball JP (Seminar, New York), Vol 2, p 40.
- Vennemann T (1973) *Syntax and Semantics*, ed Kimball JP (Seminar, New York), Vol 2, p 36.
- Harris AC, Campbell L (1995) *Historical Syntax in Cross-Linguistic Perspective* (Cambridge Univ Press, Cambridge, UK), p 405.
- Harris AC, Campbell L (1995) *Historical Syntax in Cross-Linguistic Perspective* (Cambridge Univ Press, Cambridge, UK), p 343.
- Fox A (1995) *Linguistic Reconstruction* (Oxford Univ Press, Oxford), p 104.
- Lehmann WP (1975) *Proto-Indo-European Syntax* (Univ of Texas Press, Austin, TX).
- Janhunen J (1992) *International Encyclopedia of Linguistics*, ed Bright W (Oxford Univ Press, New York), Vol 4, p 208.
- Mithun M (1999) *The Languages of Native North America* (Cambridge Univ Press, Cambridge, UK), p 203.
- Comrie B (1981) *The Languages of the Soviet Union* (Cambridge Univ Press, Cambridge, UK), p 251.
- Illich-Svitych, VM (1971–1984) *Opyt Sravnenija Nostraticheskix Jazykov* [An Attempt at a Comparison of the Nostratic Languages], 3 Vols (Nauka, Moscow) (Russian).
- Ehret C (1995) *Reconstructing Proto-Afroasiatic (Proto-Afrasian)* (Univ of California Press, Berkeley, CA), p 490.
- Steele S (1979) *The Languages of Native America*, eds Campbell L, Mithun M (Univ of Texas Press, Austin, TX), p 459.
- Rudes BA (1984) *Historical Linguistics*, ed Fisiak J (Mouton, Berlin), pp 471–508.
- Gildea S (1998) *On Reconstructing Grammar: Comparative Cariban Morphosyntax* (Oxford Univ Press, New York).
- Ruhlen M (1994) *On the Origin of Languages*, ed Ruhlen M (Stanford Univ Press, Stanford, CA), pp 207–241.
- Greenberg JH (2002) *Indo-European and Its Closest Relatives: The Eurasiatic Language Family* (Stanford Univ Press, Stanford, CA), Vol 2.
- Pinnow HP (1966) *Studies in Comparative Austroasiatic Linguistics*, ed Zide NH (Mouton, The Hague), pp 178–180.
- Clark R (1992) *International Encyclopedia of Linguistics*, ed Bright W (Oxford Univ Press, New York), Vol 1, pp 144–145.
- Pawley A, Reid LA (1980) in *Austronesian Studies*, ed Naylor PB (Center for South and Southeast Asian Studies, Univ of Michigan, Ann Arbor, MI), p 116.
- Dixon RMW (1980) *The Languages of Australia* (Cambridge Univ Press, Cambridge, UK), p 442.
- Blake BJ (1987) *Australian Aboriginal Grammar* (Croom Helm, London), pp 154–163.
- Greenberg JH (1971) *Current Trends in Linguistics*, eds Bowen JD, Dyen I, Grace GW, Wurm SA (Mouton, The Hague), Vol 8, pp 807–871.
- Greenberg JH (1963) *The Languages of Africa* (Indiana Univ Press, Bloomington, IN).
- Givón T (1975) *Word Order and Word Order Change*, ed Li C (Univ of Texas Press, Austin, TX), pp 47–112.
- Hyman L (1975) *Word Order and Word Order Change*, ed Li C (Univ of Texas Press, Austin, TX), pp 113–148.
- Givón T (1975) *Word Order and Word Order Change*, ed Li C (Univ of Texas Press, Austin, TX), p 65.
- Claudi U (1994) *Perspectives on Grammaticalization*, ed Pagliuca W (John Benjamins, Amsterdam), pp 191–231.
- Heine B (1976) *A Typology of African Languages* (Dietrich Reimer, Berlin).
- Heine B (1980) Language typology and linguistic reconstruction: The Niger-Congo case. *J Afr Lg Ling* 2:95–112.
- Bender LM (2000) *African Languages: An Introduction*, eds Heine B, Nurse D (Cambridge Univ Press, Cambridge, UK), p 59.
- Ehret C (2001) *A Historical-Comparative Reconstruction of Nilo-Saharan* (Rüdiger Köppe, Köln, Germany), pp 88–89.
- Honken H (1977) *Bushman and Hottentot Linguistic Studies*, ed Snyman JW (Univ of South Africa, Pretoria), pp 1–10.
- Gregersen EA (1972) Kongo-Saharan. *J Afr Lg* 11:69–89.
- Bickerton D (1981) *Roots of Language* (Karoma, Ann Arbor, MI), p 273.
- Kayne RS (1994) *The Antisymmetry of Syntax* (MIT Press, Cambridge, MA).