Pieces for a Global Puzzle

Jan Anward

Pieces for a Global Puzzle
Jan Anward

My official rhetorical position in this paper, that of an ordinary linguist talking to computational linguists, is rapidly becoming obsolete. In a near future, there will be no non-obsolete ordinary linguists who are not also computational linguists, and no non-obsolete computational linguists who are not also ordinary linguists. So, in anticipation of the near future, I will talk as a linguist to other linguists about an exciting possibility that will require some cooperation between those linguists who know about language typology and historical linguistics and those linguists who know about programming and parsing.

1. Language typology and linguistic pre-history

The possibility I want to talk about concerns the use of typological databases to model linguistic (pre)-history and, ultimately, the possible initial state(s) of human language.

Typological databases are of course primarily used to study language typology: We use typological data to chart linguistic resources available to humans, to make inductive generalizations about what is a possible or typical human language, and to construct or support linguistic theories which make sense of the inductive generalizations we have arrived at.

However, through the works of Dryer (1989, 1991, 1992), Maddieson (1991) and Nichols (1992), it has become clear that there is an irreducible AREAL component in language typology. Linguistic diversity does not look the same all over the globe.

This areal component is precisely what allows us to introduce a HISTORICAL component into language typology, as well.

1.1 Nichols

In her important recent book *Linguistic diversity in space and time* (Nichols 1992), Nichols argues persuasively that present-day areal skewings of linguistic diversity can be used as a major source of insights into linguistic pre-history, allowing us to penetrate far beyond the 10 000 years visible to traditional comparative and historical linguistics.

In a survey of four broad structural features and seven grammatical categories in a carefully designed areal-genetic sample of around 170 languages, Nichols shows that there are significant differences in the distribution of these features and categories between three macroareas: the Old World (Africa, Europe, Asia), the New World (the Americas), and the Pacific (Australia, New Guinea, Oceania).

On a global scale, Nichols finds a basic contrast among chiefly head-marking languages, where grammatical relations are signalled by inflections on heads of constructions (e.g. agreement on verbs and nouns), chiefly dependent-marking languages, where grammatical relations are signalled by inflections on dependents (e.g. case on nouns), and double- or split-marking languages, where both methods of signalling grammatical relations are used. However these alternatives are not equally distributed over the globe, as can be seen from table 1: Old World languages are predominantly dependent-marking, while New World languages are predominantly head-marking, and Pacific languages are predominantly double- or split-marking.

Macroarea	Area	Dependent marking %
Africa		70
Eurasia	A N East	60
	N Eurasia	64
	S + SE Asia	74
Oceania	N Guinea	50
	Australia	65
	Oceania	53
America	W North	32
	E North	32
	Meso	19
	South	37

Table 1. Head/dependent marking in macroareas. Based on Nichols (1992).
Head/dependent marking is here measured as the percentage of dependent markings (D) out of all markings of grammatical relations (dependent markings (D) + head markings (H) + detached markings (F)).

Nichols also finds that the contrast between head- and dependent-marking is a good predictor of the distribution of her other structural features: complexity (number of inflections, essentially), alignment (how subjects and objects are marked, through case-marking and/or agreement), and word order. In both language types, moderate morphological complexity, accusative alignment (direct objects have a distinctive marking), and verb-final word order are unmarked, but head- and dependent-marking favor different marked types of complexity, alignment, and word order. Head-marking tends to favor low complexity, stative-active alignment (agents have a distinctive marking) or hierarchical alignment (participants that are high on an animacy hierarchy have a distinctive marking), and verb-initial or free word order, while dependent-marking tends to favor high complexity, ergative alignment (transitive subjects have a distinctive marking), and verb-medial order. As a consequence, the marked types of complexity, alignment, and word order also show significant areal skewings in their global distribution.

The significant contrasts that Nichols finds between the Old World, the New World, and the Pacific (Australia, New Guinea, Oceania) indicate, in her opinion, "long-standing affinities and disparities" (Nichols 1992: 185)

between these areas. Several cluster analyses reveal that inter-area divergence is greatest in the Pacific and that the greatest affinity between areas is between the Pacific and the New World. There is lesser affinity between the Old World and the Pacific and a great divergence between the Old World and the New World. These data, Nichols suggests, support a model of the peopling of the Earth, where the Old World is populated from Africa via the Near East, and then first Australia, second the New World, and finally New Guinea are populated from a center in South East Asia. Relying on archaeological evidence, Nichols dates the colonization of Australia to 50 000 years BP, and the beginning of circum-Pacific colonization to 35 000 years BP.

The mechanisms which Nichols uses to derive present-day linguistic diversity from these migrations are an assumption of initial diversity, and a model, borrowed from population genetics, where initial diversity is stabilized as populations stabilize in colonized areas. A small initial difference with respect to the presence of a feature F, say 60% +F and 40% -F, is eventually stabilized as 100% +F and 0% -F. This would mean that a small initial difference in favor of dependent-marking in the languages of the populations that remained in the Old World would eventually result in 100% dependent-marking languages in the Old World, while a small initial difference in favor of head-marking in the languages of the populations that colonized the New World would eventually result in 100% head-marking languages in the New World. None of the processes would have run their full course, though, due to, for example, insuffICEnt time depth.

1.2. Problems with Nichols' model

Nichols' great merit is to have opened up the fascinating prospect of reading off linguistic pre-history from present-day areal skewings of various linguistic phenomena. However, Nichols' implementation of this prospect is far from satisfying.

Nichols' model of linguistic change on a global scale is essentially a spatio-temporal projection of the statistical differences she finds. As such, it abstracts away from the many local historical processes involved, subsuming them all under the single notion of levelling of initial skewings. However, as soon as we try to spell out levelling in terms of actual historical processes, it becomes clear that Nichol's model is based on a number of questionable assumptions.

Consider the following model case, where we have an area featuring four languages, two of which have case (L1, L2), and two of which have subject agreement (L1, L3). In global terms, L1 is double-marking, L2 is dependent-marking, L3 is head-marking, and L4 is zero-marking. The whole area is double-marking, having 2 instances of case (C) and 2 instances of agreement (A), or, in the measure used in table 1, 50% dependent-markings.

L1. A C	L2. C
L3. A	L4.

Suppose now the area is subject to a population split, and one of these languages 'walks away' to another, previously unpopulated area. The possible outcomes of such a split are shown below.

L2. C L3. A L4.	L1. AC	(a)
L1. A C L3. A L4.	L2. C	(b)
L1. A C L2. C L4.	L3. A	(c)
L1. A C L2. C L3. A	L4.	(d)

As we can see, population splits do not always skew linguistic diversity. When L1 or L4 walks away, as in (a) and (d), respectively, the old area retains its double-marking character, and the new area becomes double-marking, as well. In contrast, when L2 walks away, as in (b), the old area becomes head-marking, and the new area becomes dependent-marking, and when L3 walks away, as in (c), we get the opposite result: the old area becomes dependent-marking, and the new area becomes head-marking.

What might happen to the old area, after the splits in (a) - (d) have taken place? In particular, how might levelling be implemented? Nichols suggests that borrowing plays a vital role in levelling. And borrowing will indeed produce levelling, if we make the further assumption that only areally 'strong' features, i.e. features that are shared by a majority of the languages of certain area, are borrowed. If that is the case, A will spread in the old area in (b) and (d), and C will spread in the old area in (c) and (d), reinforcing the head-marking character of the old area in (b) (from 33% dependent-marking to 25% dependent-marking), as well as the dependent-marking character of the old area in (c) (from 67% to 75%), but retaining the double-marking character of the old area in (a) and (d) (at 50%).

However, the assumption that only areally strong features are borrowed is not an uncontroversial assumption, to say the least. All empirical evidence suggests instead that any linguistic feature is capable of spread, under conditions of political or cultural dominance (Thomason & Kaufman 1988), which means that areal convergence on a certain feature need not reflect an initial skewing in favor of that feature. In our model case, then, A or C may spread in the old area in all four after-split situations, provided that they spread from a politically and/or culturally dominant language.

Another factor which may play a rôle in levelling is grammaticalization, system-internal processes whereby inflections and constructions are formed and disappear. The two standard processes in (1) produce head marking and dependent marking, respectively, in their next to last stages (see Hopper & Traugott 1993 for a review of these processes).

(1)   a.   Pronoun -> Agreement -> Ø
       b.   Noun / Verb -> Adposition -> Case -> Ø

Grammaticalization can also effect levelling, but, as with borrowing, only if it interacts in a crucial way with areal strength. If grammaticalization

produces nothing but further instances of areally strong features, then it may result in A in all of the languages of the old area in (b) and (d), and in C in all of the languages of the old area in (c) and (d).

However, the assumption that grammaticalization produces just further instances of areally strong features is as untenable as the assumption that only areally strong features are borrowed. To take just the most apparent case: The first instances of agreement and case in an area can of course not be further instances of areally strong features. Thus, if the processes in (1) are indeed the only sources for agreement and case, then they must be able to introduce areally weak features.

The consequences of allowing grammaticalization to produce areally weak or even previously absent features are far-reaching. Let us spell out a possible interaction of the processes in (1), in terms of the following assumptions:

(i) Languages start out with only pronouns, nouns, and verbs, and then acquire, and lose, agreement and case through the processes in (1).

(ii) The formation of agreement is faster than the formation of case - there is one more stage involved in the formation of case.

(iii) Loss of agreement or case is much slower than their formation from pronouns and adpositions, respectively - inflections are resistant to erosion.

(iv) Restitution of agreement is about as slow as loss of agreement - a new set of independent pronouns must develop before the process in (1a) can start again).

These assumptions produce a cycle of possible language stages, shown in (2).

(2)   a.   - Agreement   - Case
          b.   + Agreement   - Case
          c.   + Agreement   + Case
          d.   - Agreement   + Case
          a.   - Agreement   - Case

Given an application of the processes in (1) that is constrained only by the assumptions of (i) - (iv), areal convergence on the feature case (stage 2c or 2d) may reflect an initial state in that area without case (stage 2a or 2b) and areal convergence on the feature agreement (stage 2b or 2c) may reflect an initial state in that area without agreement (stage 2d or 2a). Again, with more realistic assumptions about linguistic change, we find that areal convergence on a feature need not reflect an initial skewing in favor of that feature.

1.3. A proposal

I have evaluated Nichols' model of linguistic pre-history by making explicit a number of assumptions about possible linguistic changes. I would now like to suggest that this is the appropriate way forward. We should not be content with simple projections of statistical differences, but we should use what we know about linguistic change to construct precise models,

based on explicit assumptions about population processes and linguistic change under various sociolinguistic conditions, which simulate how present-day diversity may arise from various postulated initial states, and thus arrive at a good guess about which initial state is the most likely one.

Such a model should of course be computational, and it should work in a computational environment, where its predictions can be tested against an actual distribution, as defined by a typological data-base, and where discrepancies between the model and the data-base lead to proposed changes in the model. Moreover, the model should have an interactive graphic interface, which permits instant illustration, on some kind of map of the globe, of actual and theoretical distributions at various times and places. Anyone who is familiar with computer games such as SimCity knows what kind of interactive graphic interface I have in mind.

The desired computational environment of the model is summarized in figure 1 below. I am grateful to Frans Gregersen for suggesting the name SimLing.

```
┌─────────────────────────────────────────────┐
│                  SimLing                      │
│                                               │
│   1. Database -> Actual distribution          │
│                                               │
│   2.  Model -> Theoretical distribution       │
│                                               │
│   3.  Evaluation device:                      │
│       Theoretical distribution                │
│        - Actual distribution                  │
│       = Possible falsification                │
│                                               │
│   4.  Remedial device:                        │
│       Interpretation of falsification         │
│       -> New Model                            │
│                                               │
│   5.  Interactive graphic interface           │
│                                               │
└─────────────────────────────────────────────┘
```
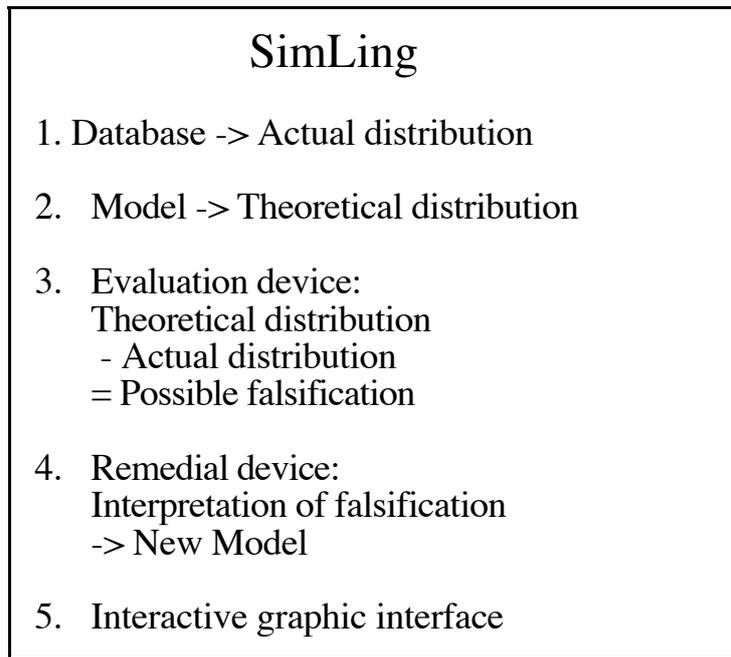
Figure 1. **SimLing**: *desired computational environment of a model of global linguistic diversity.*

2. A model of global linguistic diversity

I will now spell out some possible details of a model of global linguistic diversity, by trying to model the global distribution of two of structural features that Nichols investigates: head/dependent-marking and basic word order. I must emphasize that the specific assumptions I make are very preliminary, and should in no way be taken as established facts. My main aim is to demonstrate that a model of the kind I have in mind is a possible enterprise and to invite other linguists to think along the same lines.

The backbone of a model of global linguistic diversity is the assumption that present-day global linguistic diversity has arisen through a number of population processes which have spread successive versions of an initial state across the globe. As I have already demonstrated, this general picture must be made more precise, by means of a number of explicit assumptions about population processes and linguistic change. In addition, the initial state and its successive versions are constrained by assumptions about which expressive means are available to natural languages, and the successive versions of a particular initial state are constrained by assumptions about which discrepancies between generations can be introduced by language acquisition and language use under various social conditions, in particular the social conditions created by the assumed

population processes. The general outline of a model of this kind is shown in figure 2 below.
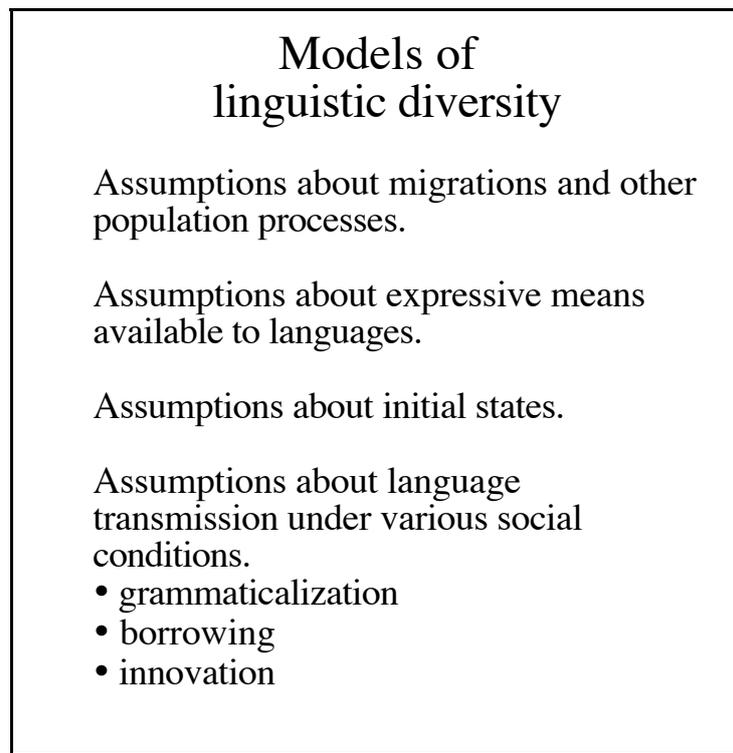
```
                    Models of
                 linguistic diversity

        Assumptions about migrations and other
        population processes.

        Assumptions about expressive means
        available to languages.

        Assumptions about initial states.

        Assumptions about language
        transmission under various social
        conditions.
        • grammaticalization
        • borrowing
        • innovation
```

Figure 2. Components of models of linguistic diversity.

2.1. Population processes

Cavalli-Sforza (1991), summarizing a number of studies of global genetic diversity, suggests that present-day genetic diversity results from two fundamental population splits, which can be surmised to have occurred 60 - 100 thousand years ago. The first split differentiated those who stayed in Africa from those who went on to West Asia, and the second split differentiated those who went to the North, to Europe, Central Asia, and America, from those who went to the South, to South Asia, Southeast Asia, Australia, New Guinea, and Oceania. There are many ways of incorporating these basic splits into a model of linguistic diversity. I would like to propose that the two basic splits first and foremost define a spatial network for global migrations, which is built around two centers. The first of these centers is West Asia, where the two basic splits postulated by Cavalli-Sforza took place: the split between Africa and the rest of the world and the split between the Northward migrants and the Southward migrants. The second center is East Asia, where those who stayed on in Asia were differentiated from those who went on to Australia, New Guinea, and Oceania, on the one hand, and the Americas, on the other hand. East Asia is also the meeting place of the Northward migrants and the Southward migrants. Japanese, for example, which has proved impossible to relate in a simple way to any

language family, might be a very concrete instance of this meeting of North and South. According to Shibatani (1990), the most probable origin of Japanese is an Altaic (Northern) language superimposed on an Austronesian (Southern) language, possibly with a Dravidian (Southern) language sandwiched in between. The meeting of the Southern and the Northern routes may also have resulted in some Southward migrants going on to America, something which would make sense of the strong evidence for a Circumpacific linguistic area that Nichols finds in her data.

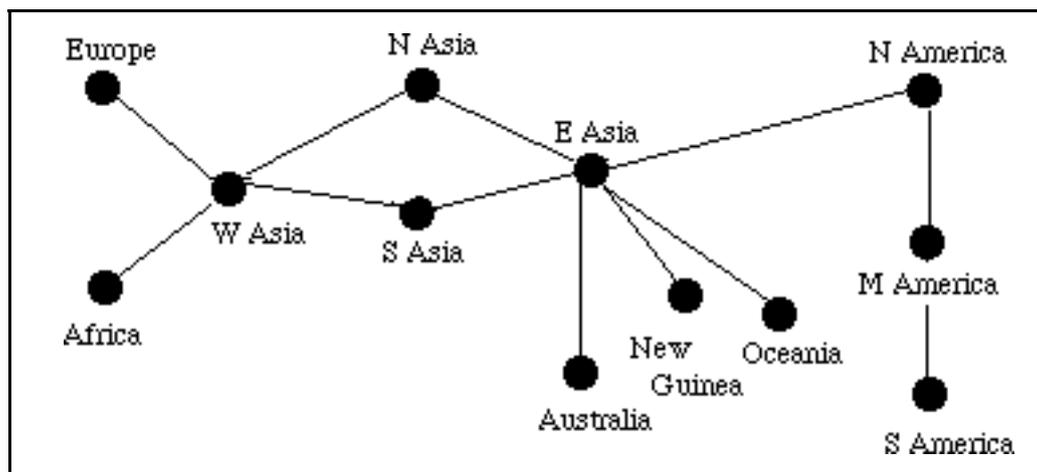This spatial network for global migrations is shown in rough outline in figure 3 below.

Figure 3.Spatial network for global migrations

Renfrew (1992, 1994) has recently proposed a model of the population of the Earth, based on archaeological, genetic, and linguistic evidence. According to Renfrew, populations have spread across the globe mainly through five major waves of migration: 1. *initial colonizations*, before 40 000 BP: colonization by hunter-gatherers of previously unpopulated areas; 2. *circumpolar dispersals*, after 10 000 BP: colonization by hunter-gatherers of areas previously covered by ice; 3. *agricultural dispersals*, after 10 000 BP: expansions, mostly into previously populated areas, in connection with the introduction and spread of agriculture; 4. *élite dominance expansions*, after 10 000 BP: invasions of previously populated areas by a military dominant élite; 5. *late colonial expansions*, after 1500: invasions of previously populated areas by a military dominant élite.

The network in figure 3 can be used to describe the first four of these waves of migration. However, after 1500, communications are reshaped in fundamental ways. First sea routes and then air routes are opened between all contintents, and printing and electronic media enable languages to travel without an accompanying population. The modern linguistic situation can hardly be put on a map. Therefore, I will treat language history up to 1500 only, and will, in this context, ignore both the spread of Indo-European after 1500, and the resulting genocides and glottocides.

Following Ruhlen (1987), we recognize 19, more or less tentative, linguistic macrogroups: In *Africa*:: Khoisan, Niger-Kordofanian, Nilo-

Saharan, and Afroasiatic; in *Eurasia*: Afroasiatic, Indo-European, Uralic-Yukaghir, North Caucasian, Kartvelian, Altaic, Elamo-Dravidian, Sino-Tibetan, Chukchi-Kamchatkan, and Austroasiatic; in *Australasia*:: Austronesian, Papuan, and Australian; and in *America*:: Eskimo-Aleut, Na-Dene, and Amerind.

In Renfrew's model, modified by the assumption of a circumpacific dispersal, these macrogroups have arrived in their present-day places in the following ways (which I will call macrogroup histories):

Initial colonization of Africa:	Khoisan
Initial colonization of SE Asia, from W Asia:	Austric
Initial colonization of America, from W Asia:	Amerind
Initial colonization of C Asia, from W Asia:	N Caucasian
Initial colonization of Australasia, from E Asia:	Australian
	Papuan
Circumpolar disperal to N Eurasia and	
N America, from W Asia:	Uralic
	Chukchi-Kamchatkan
	Na-Dene
	Eskimo-Aleut
Agricultural dispersal in Africa:	Nilo-Saharan,
	Niger-Kordofanian
Agricultural dispersal to S Asia, from W Asia:	Dravidian
Agricultural dispersal to Europe, from W Asia:	Indo-European
Agricultural dispersal in W Asia and to Africa,	
from W Asia:	Afroasiatic
Agricultural dispersal to C Asia, from W Asia:	Kartvelian
	Sino-Tibetan
Agricultural dispersal/circumpacific dispersal	
to Australasia and W America, from E Asia:	Austronesian
	Amerind
Élite dominance expansion to S Asia and C Asia,	
from W Asia and C Asia:	Indo-European
	Sino-Tibetan
	Altaic

Each area in the network of figure 3 can now be assigned a history, which, to simplify matters, is the union of the histories of the macrogroups that presently occupy the area. The history of Africa, for example, is the sum of the histories of Khoisan, Nilo-Saharan, Niger-Kordofanian, and Afroasiatic. Possible components of such areal histories in the model are:

IC(Africa):	Initial colonization in or from Africa before 40 000 BP
IC(W Asia):	Initial colonization from W Asia before 40 000 BP
IC(E Asia):	Initial colonization from E Asia before 40 000 BP
CD(W Asia):	Circumpolar dispersal from W Asia, after 10 000 BP
AD(Africa):	Agricultural dispersal in Africa, after 10 000 BP
AD(W Asia):	Agricultural dispersal in or from W Asia, after 10 000 BP
AD(E Asia):	Agricultural dispersal in or from E Asia, after 10 000 BP
EE(W Asia):	Élite dominance expansion from W Asia, after 10 000 BP

The actual areal histories incorporated in the model are shown in figure 4.

Areal histories	
Africa:	IC(Africa), AD(Africa), AD(W Asia)
West Asia:	IC(Africa), AD(W Asia)
Europe:	IC(W Asia), AD(W Asia)
North Asia:	CD(W Asia)
South Asia:	AD(W Asia), EE(W Asia)
East Asia:	IC(W Asia), AD(E Asia), EE(W Asia)
Australia:	IC(E Asia)
New Guinea:	IC(E Asia)
Oceania:	AD(E Asia)
North America:	IC(W Asia), CD(W Asia), AD(E Asia)
Mesoamerica:	IC(W Asia), AD(E Asia)
South America:	IC(W Asia), AD(E Asia)

Figure 4.Areal histories in the spatial network
IC = Initial Colonization; AD = Agricultural Dispersal;
CD = Circumpolar Dispersal; EE = Élite Expansion

2.2. Expressive means

I want the model being developed to say something interesting about the global distribution of two of Nichols' structural features: head/dependent-marking and basic word order. To begin with, we must decide which expressive means make up these structural features.

Another merit of Nichols (1992) is that she provides a more complete picture of sentence structure options than is normally provided in studies of Universal Grammar. The extension of alignment patterns to include also stative - active alignment and hierarchical alignment and the recognition of both agreement and case-marking as exponents of alignment are necessary steps to achieve a more realistic model of sentence structure options available to natural languages.

Here I will take Nichols' argument one step further. Consider the following story, from Labov (1972):

(3)   This boy punched me
        and I punched him
        and the teacher came in
        and stopped the fight

Punch and *stop* express two-place predicates, and *come in* expresses a one-place predicate. The arguments of these predicates are characterizable in terms of thematic rôles, as in (4), and these thematic rôles form a thematic hierarchy (Jackendoff 1990, kap. 11).

(4)   *come in* (Theme)
        punch (Agent, Goal)
        stop (Agent, Theme)

The arguments are also characterizable along two other dimensions: an animacy dimension, where referents are ranked according to closeness to speech act participants (Silverstein 1976, 1987), and a discourse flow dimension, where referents are ranked according to their topicality in the ongoing discourse. Simple thematic, animacy, and discourse flow hierarchies are shown in (5a), (5b), and (5c), respectively.

(5)   a.   Agent > Goal > Theme
        b.   Ego,Tu > Humans > Animals > Plants > Objects > Abstracts
        c.   Topic > Definite > Indefinite

In (3), animacy ranks the referents as: I/me > this boy, the teacher > the fight. Discourse flow ranking of the referents in (3) is not obvious, but would probably essentially agree with their animacy ranking.

The various kinds of alignment that Nichols recognizes, as well as a few more, can now be explicated in terms of how agreement, case, and word order mark positions on one or more of these hierarchies.

Consider first agreement. If there is one agreement-trigger, then we have as options at least: accusative alignment, where the trigger is the highest argument on the thematic hierarchy (θ-high); ergative or stative-active alignment, where the trigger is the lowest argument on the thematic

hierarchy (θ-low); and hierarchical alignment, where the trigger is the highest argument on the animacy hierarchy (A-high), as well as the highest argument on the thematic hierarchy, unless an inverse marking on the verb shows that the A-high argument is θ-low. If there are two argument-triggers, then the second agreement-marker marks the polar opposite of the first agreement-marker. In accusative agreement, the second agreement-marker often signalizes that the trigger is high on the discourse-flow hierarchy (D-high), as well.

In (3), *this boy*, *I*, *the teacher*, and ø (the null subject of the last sentence) would be primary agreement triggers in an accusative alignment, while *me*, *him*, *the teacher*, and *the fight* would be primary agreement triggers in an ergative alignment, and *me*, *I*, *the teacher*, and ø would be primary agreement triggers in a hierarchical alignment.

A similar story can be told of case. In accusative alignment, θ-low has an overt marking, when it is distinct from θ-high; in ergative and stative-active alignment, θ-high has an overt marking, when it is distinct from θ-low (ergative) or always (stative-active). There is often a component of A-high and/or D-high in accusative case, and a component of A-low in ergative case.

In (3), *this boy* and *I* would have overt case in an ergative alignment, while *me*, *him*, and *the fight* would have overt case in an accusative alignment.

This can be summarized in a simple model, where agreement markers and case markers are taken to signal combinations of θ-high / θ-low, A-high / A-low, and D-high / D-low. And this model can be extended to word order, as well. Position before another argument, and position before or after the head can also be taken to signal such combinations. In a strict SOV-language, for example, where S must precede O, position before the head does not say anything about θ-, A- or D-value, but position before another argument signals θ-high. As is well-known, word order often signals D-value. Word order may also signal A-value.

As Nichols points out, agreement and case (and of course word order) occur in S, NP, and/or PP, with occurrence in PP implying occurrence in NP and occurrence in NP implying occurrence in S.

A rather complete parametric model of the expressive means underlying head / dependent marking and basic word order will thus include the following components:
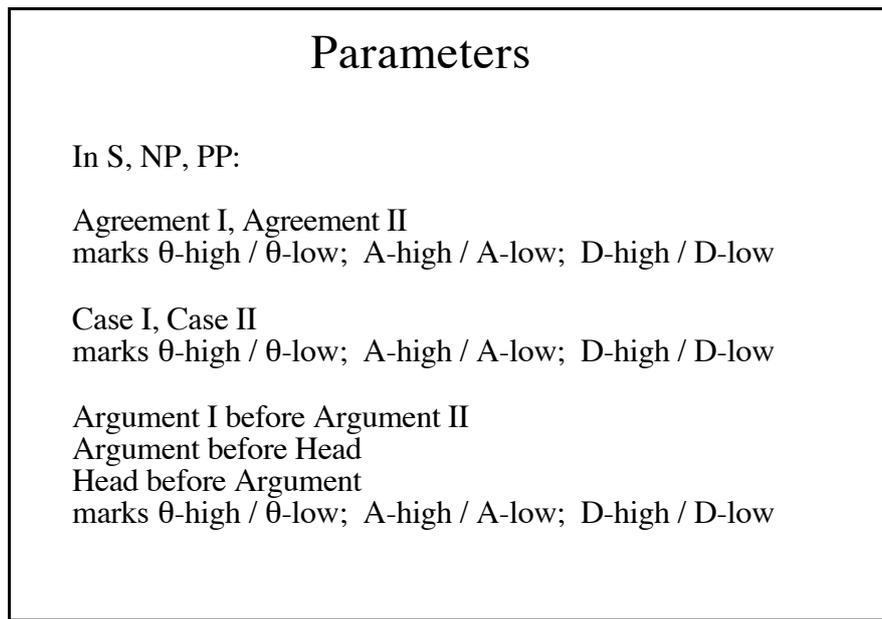
```
┌─────────────────────────────────────────────────────┐
│                   Parameters                         │
│                                                      │
│                                                      │
│   In S, NP, PP:                                      │
│                                                      │
│   Agreement I, Agreement II                          │
│   marks θ-high / θ-low;  A-high / A-low;  D-high / D-low │
│                                                      │
│   Case I, Case II                                    │
│   marks θ-high / θ-low;  A-high / A-low;  D-high / D-low │
│                                                      │
│   Argument I before Argument II                      │
│   Argument before Head                               │
│   Head before Argument                               │
│   marks θ-high / θ-low;  A-high / A-low;  D-high / D-low │
│                                                      │
└─────────────────────────────────────────────────────┘
```

Figure 5. Parameters underlying head/dependent-marking and basic word order

Here, though, I will use an extremely simple parametric model, with only four parameters: Presence (+) or absence (-) of agreement in S, presence (+) or absence (-) of case in S, verb before object (+VO) or object before verb (-VO), and verb before subject (+VS) or subject before verb (-VS). I assume, contrary to fact, that subject always precedes object. [+VO; +VS] then sets basic word order to VSO, [+VO; -VS] sets basic word order to SVO, [-VO; +VS] is excluded, and [-VO; -VS] sets basic word order to SOV. This simplified parametric model is summarized in figure 6.
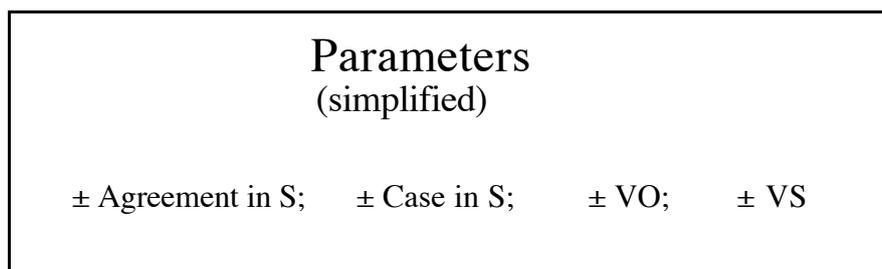
```
┌─────────────────────────────────────────────────────┐
│                   Parameters                         │
│                   (simplified)                       │
│                                                      │
│                                                      │
│    ± Agreement in S;     ± Case in S;      ± VO;      ± VS │
│                                                      │
└─────────────────────────────────────────────────────┘
```

Figure 6. Parameters underlying head/dependent-marking and basic word order
(simplified)

2.3. Global distribution of expressive means

The areal distributions in Nichols' sample of the simple parameter values of figure 6 are shown below in figures 7 and 8. Figure 7 shows the amount of head-marking (agreement) and dependent-marking (case) in S. For each area in the appendix of Nichols (1992), I counted the number of languages with only head-marking in S (H), the number of languages with both head-marking and dependent-marking in S (HD) and the number of languages with only dependent-marking in S (D). As we can see, the result agrees with Nichols' general result: most dependent-marking in the Old World, less dependent-marking in the Pacific, and least dependent-marking in the Americas. Figure 8 shows the distribution of basic VO and VS orders.
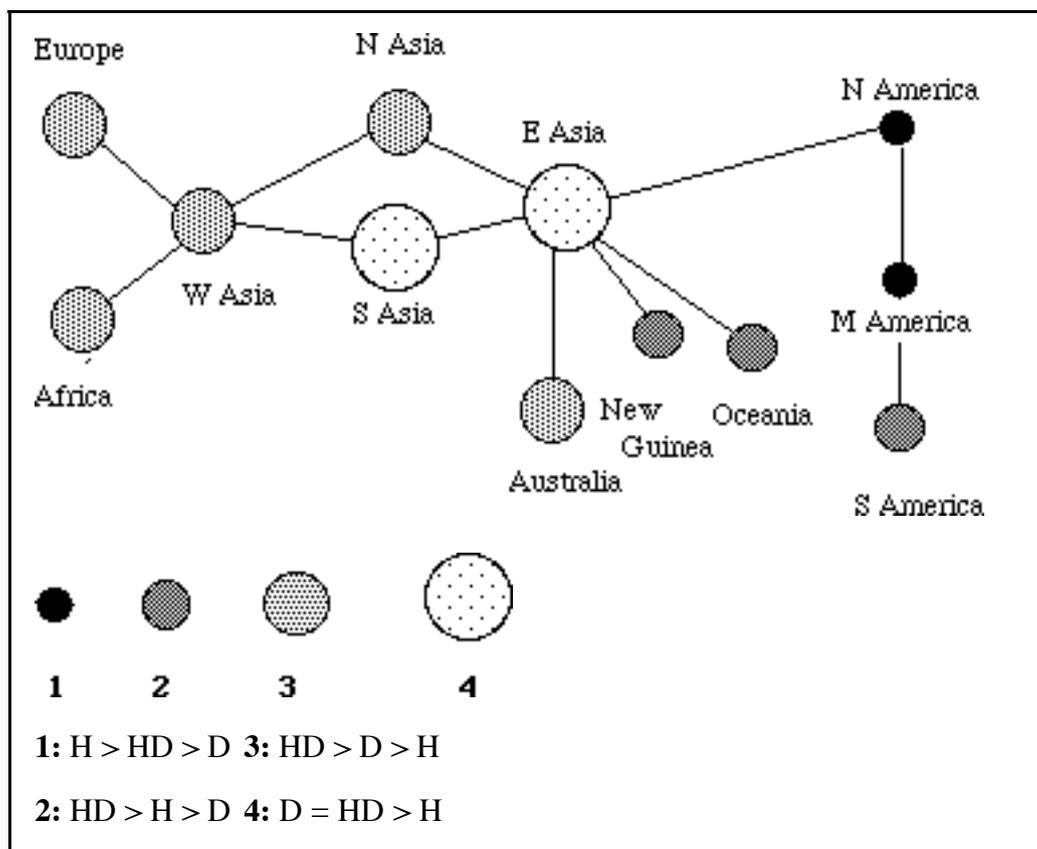
Figure 7. Areal distribution of head-marking and dependent-marking
H = Nr of languages with only head-marking in S; HD = Nr of languages with both head-marking and dependent-marking in S; D = Nr of languages with only dependent-marking in S. Based on Nichols (1992).
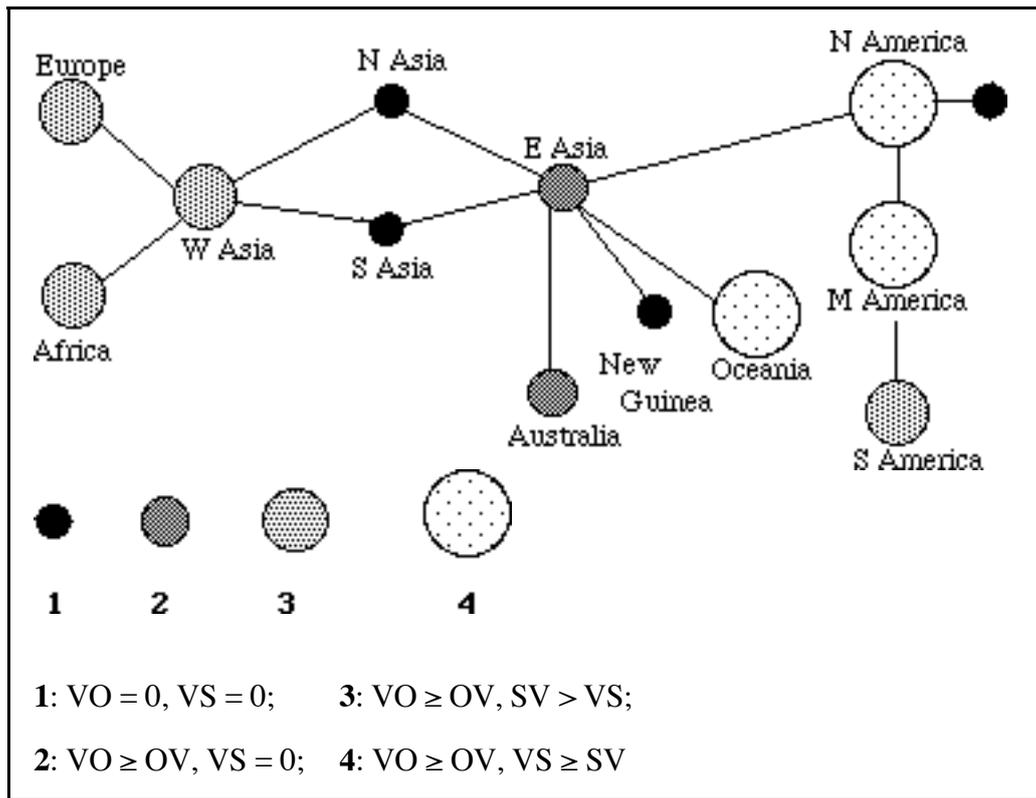
Figure 8. Areal distribution of basic word orders
OV, VO, SV, VS = Nr of languages with OV, VO, SV,
and VS, respectively, as basic order. Note that the figures
for West and East North America are very different.
Based on Nichols (1992).

Inside the figure:

1: VO = 0, VS = 0;   **3**: VO ≥ OV, SV > VS;

2: VO ≥ OV, VS = 0;   **4**: VO ≥ OV, VS ≥ SV

2.4. Initial states

The four combinations of head- and dependent-marking in figure 7 (which I designate as D1, D2, D3, and D4) relate to the cyclic stages of head/dependent-marking in (2) in the following way (since stage a does not appearin figure 7, it is designated as D0):

(6)   a.   - Agreement   - Case: D0
      b.   + Agreement   - Case: D1:   H > HD > D
      c.   + Agreement   + Case: D2:   HD > H > D
                                  D3:   HD > D > H
      d.   - Agreement   + Case: D4:   D = HD > H

Any of these stages can of course be taken as the initial state of global linguistic development, but as far as I know only stage a and stage b have been seriously proposed. Most theories of grammaticalization at least implicate an initial state with only uninflected nouns and verbs, i.e. D0. A minority position is held by Jespersen (1922) and Swadesh (1971), whose suggested initial states are best described as radically head-marking languages, i.e. D1.

As for word order, Givón (1979) has suggested SOV as an initial state, and this suggestion can be taken to motivate a linear model with three stages: [-VO; -VS] - [+VO; -VS] - [+VO; +VS]. The model is linear because there seems to be no way leading from stage c back to stage a. The three stages correspond to the four distributions of VO and VS in figure 8 (which I designate as VO0, VO1, VO2, and VO3) in the following way:

(7)   a.   -VO   -VS: VO0: $VO = 0, VS = 0$
      b.   +VO   -VS: VO1: $VO \geq OV, VS = 0$
      c.   +VO   +VS: VO2: $VO \geq OV, SV > VS$
                                          VO3: $VO \geq OV, VS \geq SV$

Since the model is linear, only VO0 can be an initial state.

This model is hardly the last word on word order change, though. The parameters are too simple, to begin with: neither OV and VO nor SV and VS are necessarily mutually exclusive. And there is no consensus on what is a possible word order change. Thus, the model in (7) should only be taken as an illustrative first approximation.

2.5. Transmission and change

In Indo-European, the changes from D2 to D4/D0 and from VO0 to VO2 seem to have taken around 10 000 years. If we generalize that pace, then the stages in (6) and (7), D0 - D4 and VO0 - VO3, respectively, would each last 5000 years, and the cycle in (6) would take 25 000 years.

With these figures, it is easy to derive predictions about the linguistic history of an area. Take Oceania, for example. Today, Oceania is in D2 and VO3. This means that Oceania would have been in D0 $(2-0) \cdot 5000 = 10$ 000 years ago and in VO0 $(3-0) \cdot 5000 = 15$ 000 years ago. However, since the process in (7) is cyclic, Oceania would also have been in D0 10 000 + 25 000 = 35 000 years ago, 10 000 + 25 000 + 25 000 = 60 000 years ago, and so on.

The general formula for deriving such predictions is given in (9). When a process is linear, $Duration_{cycle} = 0$, by stipulation.

(9) $Stage\ j = Stage\ i + (j - i) \cdot Duration_{stage} + n \cdot Duration_{cycle}$

The predictions computed for each area are given in table 2 below, together with its history.

Area	History	Temporal distance from D0 (thousand years)	Temporal distance from VO0 (thousand years)
Africa:	IC(Africa) AD(Africa) AD(W Asia)	D3 = D0 + 15/40/65/90	VO2 = VO0 + 10
West Asia:	IC(Africa) AD(W Asia)	D3 = D0 + 15/40/65	VO2 = VO0 + 10
Europe:	IC(W Asia) AD(W Asia)	D3 = D0 + 15/40/65	VO2 = VO0 + 10
North Asia:	CD(W Asia)	D3 = D0 + 15/40/65	VO0 = VO0 + 0
South Asia:	AD(W Asia) EE(W Asia)	D4 = D0 + 20/45/70	VO0 = VO0 + 0
East Asia:	IC(W Asia) AD(E Asia) EE(W Asia)	D4 = D0 + 20/45/70	VO1 = VO0 + 5
Australia:	IC(E Asia)	D3 = D0 + 15/40/65	VO1 = VO0 + 10
New Guinea:	IC(E Asia)	D2 = D0 + 10/35/60	VO0 = VO0 + 0
Oceania:	AD(E Asia)	D2 = D0 + 10/35/60	VO3 = VO0 + 15
North America	IC(W Asia) CD(W Asia) AD(E Asia)	D1 = D0 + 5/30/55	VO0 = VO0 + 0 & VO3 = VO0 + 15
Mesoamerica:	IC(W Asia) AD(E Asia)	D1 = D0 + 5/30/55	VO3 = VO0 + 15
South America	IC(W Asia) AD(E Asia)	D2 = D0 + 10/35/60	VO2 = VO0 + 10

Table 2. Temporal distance from initial states

How are we to make sense of these figures? Let me just explore one of several possibilities. Suppose that a population split brings about a discontinuity in the transmission of a linguistic tradition, through which certain aspects of the tradition are lost to a language which 'walks away'. In the case of head/dependent-marking, what would be lost is inflectional morphology - a generalization of a well-known feature of the discontinuity in transmission associated with pidginization and creolization. If we try the hypothesis that this kind of discontinuity is primarily a consequence of initial colonization (including circumpolar or agricultural dispersal into a previously unpopulated area), a hypothesis which is consistent with Nichols' demonstration that head/dependent-marking shows a high degree of genetic stability, then we might, for example, use the data in table 3 to construct a possible scenario.

Area	History	Temporal distance from D0 (thousand years)
Africa:	IC(Africa)	90
West Asia:	IC(Africa)	65
Europe:	IC(W Asia)	40
North Asia:	CD(W Asia)	15
South Asia:	AD(W Asia)	20/45
East Asia:	IC(W Asia)	45
Australia:	IC(E Asia)	40
New Guinea:	IC(E Asia)	35
Oceania:	AD(E Asia)	10
North America	IC(W Asia)	30
Mesoamerica:	IC(W Asia)	30
South America	IC(W Asia)	35

Table 3. Temporal distance from initial D0 in a scenario with initial colonization as trigger

The scenario that follows from table 3 is fairly realistic, if we match it against Renfrew's datings. The split between Africa and the rest of the world would have taken place in Africa 65 000 years ago, the split between North and South would have taken place in West Asia 45 000 years ago, and the splits leading to colonization of Australia, New Guinea, and the Americas would have taken place 40 000, 35 000, and 30 000 - 35 000 years ago, respectively. The date for circumpolar dispersal to North Asia, 15 000 years ago, is a little too early, but the discrepancy is not serious, given the extremely rough calculations on which the model rests. The only serious discrepancy in table 3 concerns South Asia. An agricultural dispersal 20 000 years ago is clearly an entirely unrealistic assumption. However, this discrepancy is easily corrected, if we assume that South Asian languages are the product of a continuous linguistic tradition that goes back to the split between North and South 45 000 years ago, that is, if we introduce IC(W Asia) into the history of South Asia.

What about word order, then? What would be lost here, I suggest, are constraints on word order. Thus, a discontinuity would make it possible to use a non-traditional order for various expressive and communicative purposes. However, this can only happen, I conjecture, when social control is weak, as it would be in agricultural dispersals, when expansion no longer takes place through intact bands of hunter-gatherers, but through a number of step-by-step migrations by smaller family units. In other words, it would take an agricultural, or comparable, dispersal to trigger off the development in (8). This conjecture is consistent with Nichols' demonstration that word order shows a low degree of genetic stability, but a high degree of areal stability.

Consider, against this background, a scenario that be constructed from the data in table 4.

Area	History	Temporal distance from VO0 (thousand years)
Africa:	AD(Africa) AD(W Asia)	10
West Asia:	AD(W Asia)	10
Europe:	AD(W Asia)	10
North Asia:		0
South Asia:	AD(W Asia) EE(W Asia)	0
East Asia:	AD(E Asia) EE(W Asia)	5
Australia:		10
New Guinea:		0
Oceania:	AD(E Asia)	15
North America	AD(E Asia)	15 + 0
Mesoamerica:	AD(E Asia)	15
South America	AD(E Asia)	10

Table 4. Temporal distance from initial VO0 in a scenario with agricultural dispersal as trigger

Fairly compatible with the data in table 4 is a scenario where VO and VS orders result from two independent agricultural, or comparable, dispersals: one from East Asia, starting 15 000 years ago, and spreading to Oceania and the Americas; and one from West Asia, starting 10 000 years ago, and spreading to Africa, Europe, and South Asia. These postulated dispersals may be a little too early, but this can be corrected by adjusting Duration$_{stage}$.

There are three areas that do not fit this scenario at first blush. Word order changes in South and East Asia are too small to match the time depth of the dispersals postulated to affect these areas, and Australia shows word order change without a corresponding dispersal. However, both South Asia and East Asia have been subject to élite dominance expansions, and it is not very far-fetched to assume that these expansions brought along enough social control to arrest word order change in these areas. In Australia, there is evidence of a wide dispersal of one the branches of Australian, Pama-Nyungan, and we may take this dispersal to be responsible for word order change in Australia. AD(Australia) should then be added to the history of Australia.

2.6. Summary

The areal histories in figure 4, complemented by IC(W Asia) in the history of South Asia and AD(Australia) in the history of Australia, the simplified parameters in figure 6, the historical processes in (6) and (7), the assumed initial states of these processes and the stipulated values of Duration$_{stage}$ and Duration$_{cycle}$, and the assumptions that transmission discontinuities with respect to head/dependent-marking are the results of

initial colonizations, while transmission discontinuities with respect to word order are the results of agricultural (or comparable) dispersals, together produce the following scenario to account for the global distributions of head/dependent marking and word order in figures 7 and 8:

A split between Africa and the rest of the world took place in Africa 65 000 years ago, and a further split between North and South took place in West Asia 45 000 years ago. These splits were followed by splits leading to the colonization of Australia, New Guinea, and the Americas, which took place 40 000, 35 000, and 30 000 - 35 000 years ago, respectively. After that, circumpolar dispersal to North Asia, 15 000 years ago, was followed by two independent wide-ranging agricultural, or comparable, dispersals: one from East Asia, starting 15 000 years ago, and spreading to Oceania and the Americas; and one from West Asia, starting 10 000 years ago, and spreading to Africa, Europe, and South Asia. These dispersals were followed by élite dominance expansions into South Asia and East Asia, and were roughly contemporary with a wide-ranging dispersal of Pama-Nyungan in Australia.

This scenario might not be the 'right' one (it is, in fact, unlikely to be the right one, considering the number of corners I have cut), but it allows for a convenient summary of the main points of this paper: 1) It is possible to construct such scenarios from what we know about typology and change; and 2) To do this in an effective way, we should have access to a SimLing environment (figure 1) which produces such scenarios, in a graphically pleasing form, from revisable sets of model assumptions (figure 2), to account for global distributions based on large typological databases.

References

Cavalli-Sforza, L. L. 1991: Genes, Peoples and Languages, *Scientific American*, November 1991, 72-78

Dryer, M. S. 1989: Large Linguistic Areas and Language Sampling, *Studies in Language* 13, 257-292
- 1991: SVO languages and the OV:VO typology, *Journal of Linguistics* 27, 443-482
- 1992: The Greenbergian Word Order Correlations, *Language* 68, 81-138

Givón, T. 1979: *On Understanding Grammar*, Academic Press

Hopper, P. J. & E. Closs Traugott 1993: *Grammaticalization*, Cambridge University Press

Jackendoff, Ray 1990: *Semantic Structures*, MIT Press

Jespersen, O. 1922: *Language. Its nature, development and origin*, Allen & Unwin

Labov, W. 1972: The Transformation of Experience in Narrative Syntax, in *Language in the Inner City*, University of Pennsylvania Press, 354-396

Maddieson, I. 1991: Testing the Universality of Phonological Generalizations with a Phonetically Specified Segment Database: Results and Limitations, *Phonetica*, 193-206

Nichols, J. 1992: *Linguistic Diversity in Space and Time*, University of Chicago Press

Renfrew, C. 1992: Archaeology, genetics and linguistic diversity, *Man* 27, 445-478
- 1994: World Linguistic Diversity, *Scientific American*, January 1994, 104-110

Ruhlen, M. 1987: *A Guide to the World's Languages*, Stanford: Stanford University Press

Shibatani. M. 1990: *The Languages of Japan*, Cambridge University Press

Silverstein, M. 1976: Hierarchy of Features and Ergativity, in R. W. Dixon (ed): *Grammatical Categories in Australian Languages*, Australian Institute of Aboriginal Studies, 112-171
- 1987: Cognitive Implications of a Referential Hierarchy, in M. Hickmann (ed): *Social and Functional Approaches to Language and Thought*, Academic Press, 125-164

Swadesh, M. 1971: *The Origin and Diversification of Language*, Aldine . Atherton

Thomason, S. G. & T. Kaufman 1988: *Language Contact, Creolization and Genetic Linguistics*, University of California Press