

# Sobre el factor de la casualidad en la comparación lingüística

IVÁN IGARTUA  
(UPV/EHU)

## Abstract

*In this paper the author describes and compares two different ways of calculating the factor of chance that interferes with the comparative linguistic practice aimed at establishing genetic relationships among languages or language families. The discussion concentrates mainly on the theoretical and methodological sides of the problem, although some attention is also paid to the connection of the mathematical estimations with the results achieved by the comparison.*

*Through the analysis of the probabilistic arguments behind both the "sceptical" (Ringe 1995a) and the "optimistic" (Bengtson & Rublen 1994) views of new (especially distant) genetic relationships, the author of this article tries to show the inadequacy (if not the inability) of such attempts to provide a common mathematical basis for calculating the amount of chance linguistic coincidences which are to be ruled out and, consequently, taken into account by comparativists when they attempt to prove the existence of any language family. Therefore, as the only effective means of guaranteeing certain security in the phonetic/semantic matchings among the languages supposed to be related, there remains the identification of the sets of systematic correspondences (and of those regular, genetically significant commonalities that underlie them).*

## 1. Introducción\*

En los últimos años parece haberse retomado entre los comparatistas la discusión acerca de la importancia estadística que pueden tener las coincidencias (formales al tiempo que semánticas) puramente casuales entre unidades léxicas o morfológicas de lenguas cuyo parentesco se investiga o trata de ser comprobado. El número de cognados que se necesita para probar el parentesco genético de dos o más lenguas no es una cifra exacta y predeterminada (cf. Hymes 1993 [1959]: 25) que se haya de cumplir en cada uno de los intentos comparativos (aunque ciertas versiones del

\*Debo agradecer al ingeniero Enrique Gutiérrez Ibáñez la ayuda prestada en el análisis y la interpretación de los datos numéricos procedentes de la distribución binomial, ámbito en el que soy un perfecto intruso. Agradezco asimismo a Joseba A. Lakarra sus precisiones y sugerencias, que han contribuido a mejorar sustancialmente varios aspectos del presente artículo.

método léxico-estadístico contemplaban un mínimo de entre el 5 y el 8% de coincidencias —más bien semejanzas— entre dos lenguas dentro de un *corpus* restringido de palabras),<sup>1</sup> pero una especie de intuición común indica que unas pocas decenas de formas semejantes o parecidas tanto en la forma como en el significado resultan insuficientes para asegurarse una hipótesis de relación lingüística que conjure y aleje la posibilidad de una coincidencia meramente casual de las formas supuestamente emparentadas (la prueba del parentesco viene determinada más bien por la cualidad y la estructura de esas coincidencias). A. Meillet (1926 [1918]: 72) previno en su día contra determinados vocablos que, pese a su similitud y al hecho de pertenecer a lenguas cuyo parentesco —en algún caso, remoto— se encuentra fuera de toda duda, provienen de étimos distintos y son, por lo tanto, productos de convergencias del todo casuales (ingl. *bad*, persa *bad*, francés *feu*, alemán *Feuer*, etc...). Para Meillet esas coincidencias eran, además, “accidents qui ne peuvent pas ne pas se produire”. V. Pisani (1952), recordado por L. Michelena (1988b [1964]: 58), dividía las semejanzas materiales que se pueden observar entre dos o más lenguas en dos grandes apartados: las similitudes que atestiguan causas históricas (productos del origen común de las lenguas o bien de préstamos interlingüísticos) y las que están desprovistas de ellas, entre las cuales se cuentan las similitudes debidas a la procedencia onomatopéyica, expresiva de algunas palabras (p.ej. vasco *liluratu* ‘deslumbrado, fascinado’ y sardo *allilliradu* ‘indurito, rígido’) y las que son simplemente casuales (cf. thai *faj*, inglés *fire* ‘fuego’, thai *rim*, inglés *rim* ‘borde’, cf. Matisoff 1990: 112, finés *vaara*, tamil *varai* ‘montaña, Jones 1989: 193, o los ejemplos anteriores de Meillet).<sup>2</sup> Es en torno a esta última clase de semejanzas, las debidas a la casualidad, donde se desenvuelve la polémica entre los especialistas cuando el cotejo de las lenguas examinadas entra dentro de lo que se ha dado en llamar *long-range comparison*, *macrocomparison*, o bien, en las versiones ironizantes tanto de Doerfer (1973) como de Matisoff (1990), *Omnicomparatismus* y *megalocomparison*, respectivamente.

En las condiciones propias de las comparaciones de esta naturaleza (cuyo corolario, en ocasiones, es una protolengua o “preprotolengua” —cf. Birnbaum 1977: 51, 1992; Schwink 1994: 10; cf. también Klimov 1991b: 13— de antigüedad notoriamente superior a la que se le supone, por ejemplo, al protoindoeuropeo),<sup>3</sup> las distintas posturas tratan de asignar una función cuantitativa determinada a las coincidencias léxicas casuales como modo de establecer los límites en los cuales puede efectuarse la comparación encaminada a probar el parentesco genético de las lenguas o de los grupos que éstas conforman. Aunque puedan hallarse posiciones

(1) Véase, por ejemplo, Echenique 1987: 28 acerca de los cálculos de A. Tovar referidos al vasco y a distintas lenguas caucásicas. Esas cifras, casi siempre superiores al 5%, podrían representar, no obstante, para decirlo con palabras de Trask (1997: 359), “the proportion of random similarities by which he is prepared to be impressed”. La del 8% es también la barrera que propuso J. H. Greenberg para las coincidencias casuales hace ya cerca de medio siglo (Hymes 1993 [1959]: 27). Para la revisión de los parámetros atribuidos a las coincidencias accidentales en el sistema de Swadesh cf. por ejemplo Fairbanks 1955: 117, Gudschinsky 1956, Fodor 1982 [1966]: 66.

(2) Divisiones parecidas pueden hallarse en Milewski 1960: 265 o Winter 1990: 14.

(3) Para H. Birnbaum son preprotolenguas los sistemas lingüísticos reconstruidos a partir de la comparación de protolenguas.

intermedias (en general, las menos), son dos los puntos de vista enfrentados en este asunto. La visión optimista (vid., por ejemplo, Bengtson y Ruhlen 1994) minimiza la posible incidencia del factor de la casualidad en las semejanzas léxicas, mientras que la concepción pesimista (o escéptica) de parentescos tan remotos (representada, entre otros, por Ringe 1995a) confiere a la coincidencia casual una dimensión estadística capaz de desbaratar cualquier intento, por fundado y sólido que éste aparente ser, de postular y probar relaciones genéticas, al menos a la manera clásica. Hay que señalar que ambas posturas distan de ser simples opiniones o declaraciones de principios: tras ellas se encuentran argumentaciones y conceptos metodológicos casi siempre contrapuestos, cuyo análisis y revisión van a ocupar las páginas que siguen.<sup>4</sup>

## 2. La visión escéptica

D. Ringe (1995a) ensaya en su artículo una comparación entre las situaciones respectivas del indoeuropeo y del nostrático desde el punto de vista de la cantidad y la interrelación de los conjuntos de unidades léxicas y morfológicas reconstruidos para una y otra protolengua. Las fuentes de datos son, en un caso, Pokorny (1959) y, en otro, Illich-Svitych (1971). El método que el investigador aplica al análisis de los datos es la llamada distribución binomial, operación matemática que permite calcular la probabilidad de que un determinado fenómeno se produzca o no se produzca. De esta manera, la probabilidad de que una forma en la lengua A coincida exactamente con otra en la lengua B es interpretada en términos de distribución binomial. Hay que indicar, no obstante, que su aplicación parte de los datos empíricos que proporcionan los diccionarios citados, en concreto, del número de raíces compartido por dos lenguas o grupos de lenguas dentro de la familia lingüística considerada, dato al que se ajusta el cálculo de partida de la distribución binomial. En cualquier caso, los resultados del cálculo probabilístico indicarán desde esta perspectiva las dimensiones de la casualidad como factor que interfiere en la actividad comparativa.

El número de raíces comunes pertenecientes a las lenguas nostráticas se divide en diversos conjuntos, dependiendo de la cantidad de grupos lingüísticos en los que se atestiguan las raíces consideradas en cada caso. Así, son 84 las raíces que unen a dos grupos de lenguas, 70 las que corresponden a tres grupos o 5 las que comparten los seis grupos que conforman la familia nostrática en la concepción de Illich-Svitych. Tras ajustar el índice de probabilidad de éxito (0,4) en seis intentos (los correspondientes a la cantidad de grupos nostráticos) al dato empírico de las raíces compartidas por dos grupos de lenguas, Ringe obtiene el valor numérico que posibilita trazar la curva de distribución binomial para 3, 4, 5 y 6 intentos. Los valores resultantes sólo difieren significativamente de los números de raíces en el dato referente a las unidades compartidas por los seis grupos de lenguas: las raíces del diccionario de Illich-Svitych son 5, en tanto que el resultado que arroja la distribución binomial es

(4) La posibilidad de aplicación de métodos de cálculo exacto a los hechos de parentesco lingüístico, lo que en primer lugar ha supuesto el interés por establecer de forma matemática los límites de las coincidencias casuales, ha sido estudiada y formulada con anterioridad a los trabajos citados en diversas propuestas (Swadesh, Cowan, Bender, Doerfer), la primera de las cuales se debe a Collinder (y data de 1948, cf. también Collinder 1966: 199-200).

1 (Ringe 1995a: 58ss).<sup>5</sup> A pesar de ello, la proximidad general observada entre los datos numéricos de la reconstrucción nostrática y los que proporciona el cálculo probabilístico se interpreta como prueba de que la relación atribuida a las lenguas que supuestamente integran esa macrofamilia se asienta sobre bases en exceso superficiales y vulnerables, puesto que no demuestra mayor proximidad entre las lenguas estudiadas que la que se obtiene calculando el factor de la casualidad en las coincidencias interlingüísticas.

La situación del indoeuropeo es a este respecto notoriamente distinta. El ajuste a los datos empíricos se realiza mediante un coeficiente de probabilidad de 0,2, a partir del cual se obtienen valores que se aproximan a los reales para las raíces o palabras comunes a dos, tres y cuatro grupos. A partir de las cantidades correspondientes a las unidades compartidas por cinco y seis grupos de lenguas las líneas descendentes de un conjunto y otro de datos muestran una evolución dispar en cuanto al ritmo de disminución de los valores numéricos. La disparidad se acentúa entre las grupos 9 a 14 (referidos a otros tantos grupos de lenguas), donde la casualidad no concede ni siquiera la posibilidad de una sola unidad léxica coincidente en ese número de lenguas (cf. Ringe 1995a: 66, lo que confirma de algún modo la primera de las premisas del método probabilístico de Dolgopolsky 1986 [1964]: 29),<sup>6</sup> mientras que, en total, Pokorny documenta sólo en aquellos grupos 271 formas comunes.<sup>7</sup> Este desencuentro entre los datos reales y los valores binomiales —simbolizado por la diversidad de las curvas descendentes a que unos y otros dan lugar— supone, según las coordenadas de Ringe, la confirmación de una similitud lingüística que va más allá de cualquier parecido casual y que, por ello, debe ser atribuida a un origen común de las lenguas entre las que aquélla se manifiesta.

Llama la atención, en todo caso, la significativa presencia que adquiere el factor de la casualidad en estos cálculos, tanto en los que se ajustan a los datos nostráticos

(5) La tabla de datos para el nostrático (Ringe 1995a: 58) es como sigue:

	distribución empírica	distribución binomial
2 grupos	84	84
3 grupos	70	75
4 grupos	37	37
5 grupos	9	10
6 grupos	5	1

(6) Basada, en cualquier caso, en una observación anterior de J. H. Greenberg, según la cual a mayor número de lenguas comparadas menor es la probabilidad de encontrar semejanzas compartidas por cada una de las lenguas que sean producto de la casualidad. I. Hegedùs (1989: 35) refiere este criterio en los términos siguientes: "If 8 out of 100 words coincide in two languages, then only one out of every 25000 words should coincide by chance in the comparison of five languages".

(7) Las raíces comunes a todas las lenguas de la familia son en cualquier caso escasas (sólo 1 para los catorce grupos indoeuropeos), dato que han puesto de relieve Baxter y Manaster Ramer (1996: 380) en su análisis y crítica de Ringe (1992). Desde esta perspectiva, no resultan tan dispares las situaciones del nostrático y del indoeuropeo (e incluso la relación es en este punto favorable al nostrático, que presenta en el diccionario de Illich-Svitych cinco étimos comunes a las seis familias que lo componen. Otra cuestión, pero menos importante, es la que atañe a las características de dichos étimos, que se distinguen, al menos tres de ellos, por su escaso cuerpo fónico, cf. Ringe 1995a: 58; acerca de las formas cortas en las listas de *comparanda*, cf. en general Campbell 1988: 600).

como en los que corresponden al indoeuropeo. Baste recordar que los valores derivados de la distribución binomial aplicada al indoeuropeo suman, en su traducción lingüística, un número de 1.411 raíces que, divididas en los distintos apartados, dependiendo del número de los grupos lingüísticos, configuran el dominio reservado a la incidencia de la casualidad en las similitudes entre unidades de lenguas distintas. La cifra es exactamente el 70,27% de las entradas del diccionario de Pokorny, lo cual lleva a pensar que sólo gracias al restante 29,73% de unidades comunes a varias lenguas se demuestra un parentesco del que nadie duda (esto es, se hace visible la relación genética a través y a pesar del “ruido” estadístico, Ringe 1995a: 70).<sup>8</sup> La correlación de cantidades y la importancia porcentual concedida al puro azar resultan contrarias a toda intuición. Todo ello proviene en última instancia del punto de partida escogido en el cálculo: los valores de probabilidad se obtienen a partir del dato empírico que corresponde a unidades lingüísticas compartidas por dos grupos (84 en el caso del nostrático, 407 en el del indoeuropeo), dato con el que coincide forzosamente —no puede ser de otra manera— la primera cifra derivada del cálculo. Esta circularidad inicial hace que el desarrollo de la aplicación se fundamente sobre un supuesto: si las 84 raíces del nostrático compartidas por dos grupos de lenguas fueran producto de la casualidad, ¿cuántas habrían de ser las raíces coincidentes también por casualidad entre tres, cuatro, cinco o seis grupos? Para decirlo con otras palabras, la distribución binomial afecta no a los propios datos, acerca de los cuales poco puede decir, puesto que depende de ellos, sino a la relación que se establece entre los distintos valores numéricos. El propio autor se ocupa de relativizar los resultados de su aplicación: “These findings do not, of course, prove that the subfamilies of the Nostratic are not related to one another; that should not be surprising, since one cannot prove such a negative empirically... They do not even prove that the similarities between proposed Nostratic cognates arose by chance” (Ringe 1995a: 70-1). Y en cuanto a la posibilidad de una distribución más compleja de los valores casuales en su relación y ajuste con el indoeuropeo, cf. en nota a pie de página: “Strictly speaking, I also haven’t demonstrated that the distribution of cognates in Pokorny (1959) does not approximate some random curve generated more complexly”.<sup>9</sup>

Pero si el cálculo del factor casual afecta exclusivamente a la relación entre los valores, nada impide en principio que ésta varíe cuando son otros los datos sometidos al cotejo de la distribución binomial. No deja de sorprender, por ejemplo, que Ringe emplee solamente los étimos del primer tomo del diccionario de Illich-Svitych y deje de lado los dos siguientes (1976, 1984 [sólo el primer fascículo]). Frente a las

(8) El hecho de que las unidades que integran estos cálculos no sean del mismo valor probatorio con vistas al establecimiento de relaciones genéticas (no es lo mismo una aproximación meramente léxica, aunque pertenezca al vocabulario denominado “básico” (cf., en cualquier caso Haarmann 1990) que una similitud gramatical) no es tomado en cuenta en los cálculos de Ringe, lo cual no quiere decir que carezcan de importancia.

(9) De hecho, la distribución de las formas indoeuropeas no acepta ningún tipo de ajuste o aproximación a su curva mediante índices de probabilidad simples (esto es, únicos). Sin embargo, no es imposible que ajustando el cálculo de base tanto al dato de las raíces compartidas por dos grupos como al de 5 o 6 grupos, momento en que se hace significativa la discrepancia entre los valores numéricos reales y los binomiales, se pueda obtener una curva compleja más aproximada. Cf. los datos indoeuropeos según Ringe (1995: 66):

205 unidades con que opera, el empleo de los otros dos tomos, unido al de los materiales contenidos en Illich-Svitych 1967, le habría proporcionado una cifra en torno a las 600 unidades, con distribución por grupos un tanto diferente.<sup>10</sup> Mientras sigue en elaboración el diccionario nostrático de A. R. Bomhard, el diccionario que ocupa la parte central de la monografía de Bomhard y Kerns (1994) constituye por el momento la fuente más accesible de datos para este tipo de comparaciones. Son 601 las unidades que recoge (en Bomhard 1996 aparecen 652), pertenecientes en su mayoría a siete (y no a seis, como en la versión de la escuela nostrática moscovita)<sup>11</sup> grupos de familias lingüísticas. La distribución de estas formas es la siguiente:

2 grupos	217 formas
3 grupos	216 formas
4 grupos	116 formas
5 grupos	36 formas
6 grupos	14 formas
7 grupos	2 formas

Una distribución binomial basada en un factor de probabilidad 0,4 (como el que se ajustaba a los datos de Illich-Svitych 1971) produce —tras las operaciones necesarias—<sup>12</sup> los siguientes valores:

	distribución empírica	distribución binomial (índice del 0,2)
2 grupos	440	440
3 grupos	407	440
4 grupos	297	303
5 grupos	221	151
6 grupos	151	57
7 grupos	146	16
8 grupos	112	3,5
9 grupos	85	0,5
10 grupos	76	-
11 grupos	36	-
12 grupos	28	-
13 grupos	8	-
14 grupos	1	-

Con respecto a estas dificultades de los métodos matemáticos en su aplicación a conjuntos de datos empíricos, cf. lo que declaran Baxter y Manaster Ramer (1996: 382): "Tests vary in their scope and sensitivity, and no one test can ensure that other, more sensitive tests will find nothing".

Al igual que son sensiblemente diferentes, en cuanto a la aritmética y a las proporciones entre datos, los cómputos que Palmaitis (1977) extrae de los dos primeros volúmenes del diccionario de Illich-Svitych.

(11) En realidad, son más los grupos de lenguas que A. R. Bomhard incorpora —aunque sólo sea de manera tentativa— a la lista nostrática, entre ellos, el grupo chukchi-kamchatkan, la lengua gilyak, el yukaguir (integrado en la familia urálica), el elamita (junto a las lenguas dravídicas) y el etrusco, pero sólo el sumerio (séptimo grupo) comparece con cierta regularidad en las entradas del diccionario (en alguna ocasión aislada lo hace también el elamita, pero por su mismo carácter esporádico no se han tomado en consideración sus datos para el cómputo general).

(12) Esto es, tras multiplicar en cada caso el resultado de la fórmula  $P(z) = \left( \frac{n!}{z! \cdot (n-z)!} \right) \cdot p^z q^{n-z}$ ,

donde n = número total de intentos, z = número de intentos para el que se busca una determinada

2 intentos	217
3 =	241
4 =	160
5 =	64
6 =	14
7 =	1

Como puede observarse, el emparejamiento con los datos empíricos sólo resulta satisfactorio para los grupos de 6 y 7 lenguas que comparten raíces (además del grupo de 2 lenguas, con el que el valor binomial ha de coincidir necesariamente). Si el factor escogido es del 0,3, la aproximación entre la curva descendente de los datos lingüísticos y la de la distribución binomial es algo más deficiente:

2 intentos	217
3 =	155
4 =	66
5 =	17
6 =	2
7 =	0,1

Con 0,2 de factor de probabilidad, la aproximación es prácticamente nula (2: 217, 3: 90, 4: 22, 5: 3,3, 6: 0,28, 7: 0,01), por lo que para ajustar la curva binomial a la real hay que recurrir al empleo de mayor número de decimales. Un coeficiente del 0,35 de probabilidad proporciona valores bastante aproximados, aunque no se produzca un ajuste como el que se obtiene a partir de los datos de Illich-Svitych 1971:

2 intentos	217
3 =	195
4 =	105
5 =	33
6 =	6
7 =	0,46

A la vista de estas comparaciones puede uno concluir que existen al menos dos factores que inciden directamente en los resultados que pueden derivarse del cálculo de la distribución binomial aplicado al número de cognados que se establecen entre distintas lenguas. Por un lado, influye obviamente la cantidad de datos, no sólo la relación que existe entre los diversos valores (incluso puede aventurarse que a mayor cantidad, mayor es también la dificultad de hallar una curva de probabilidad que se le asemeje, dificultad que parece reflejarse precisamente en la relación entre los datos empíricos y la curva binomial del indoeuropeo; esto podrá ir verificándose a medida que avance la elaboración del diccionario definitivo de Bomhard). Por otro, es innegable el influjo del número de grupos de lenguas estudiados (en el correlato binomial: número de intentos para los que se busca una determinada probabilidad

---

probabilidad de éxito,  $p$  = índice de probabilidad,  $q = 1 - p$ , por el valor que se obtiene de la división del número de raíces para dos grupos de lenguas entre el resultado procedente de la aplicación de la fórmula cuando  $z = 2$ .

de éxito) en la relación entre las curvas real y binomial: a partir de un determinado número de lenguas (6 o 7) la disminución de la probabilidad según la distribución binomial es tan pronunciada (con todo tipo de factores de probabilidad) que un mínimo conjunto de formas comunes para esos grupos desfigura de inmediato cualquier aproximación anterior entre ambas series de valores. Aquí puede hallarse la razón de la diferencia cuantitativa que aleja al nostrático del indoeuropeo: al contar con más miembros, este último posee un margen mayor para hacer explícita la divergencia entre datos empíricos y valores binomiales (casi nulos) que demuestra —por encima de la estadística— el parentesco. No hay que olvidar que de las 597 raíces que separan el total de étimos indoeuropeos (2.008) de la suma de los valores numéricos que se obtienen como resultado de la distribución binomial (1.411), 529 —o el 88,6%— corresponden a las unidades compartidas por siete y más (hasta catorce en la clasificación que Ringe toma de Pokorny) grupos de lenguas.

Aun sin saber en qué medida, parece, en cualquier caso, conveniente atenuar el carácter probatorio que supuestamente pudiera tener este tipo de ensayos (cf. ya Pisani 1952: 5). Es, no obstante, indudable el interés que representa la elaboración de métodos de comprobación de parentescos lingüísticos por medio de cálculos que permitan delimitar el alcance de las coincidencias lingüísticas casuales, con independencia de la capacidad heurística y explicativa que cada cual pueda atribuir a esos métodos (cf. Baxter y Manaster Ramer 1996: 383-4). Pero en este caso la comparación entre dos ámbitos de reconstrucción lingüística, basada en criterios sujetos a variables que en modo alguno garantizan la equivalencia real de los parámetros aplicados a uno y otro, no puede constituirse en instrumento fiable para diagnosticar la naturaleza —y cuestionar, en su caso, la validez— de los resultados de la reconstrucción.<sup>13</sup> Por otra parte, la propia experiencia del método comparativo en su aplicación al indoeuropeo debería restar cierto crédito a una hipótesis que confiere tal importancia al factor de la casualidad. Y hay una objeción más, señalada anteriormente: al partir de los datos empíricos, la aplicación de la distribución binomial en busca del valor estadístico de las coincidencias léxicas casuales nada puede demostrar acerca de los datos: éstos no tienen por qué ser sin más producto del azar. Sus conclusiones han de limitarse por tanto a las relaciones entre los datos, y, como se ha podido ver, éstas distan de ser constantes.

### 3. La visión optimista

En el extremo opuesto del cálculo y la valoración de la casualidad como elemento perturbador de la labor comparatista se encuentran concepciones como las expuestas

(13) Existe, además, un problema metodológico crucial: "while a positive result on a properly constructed significance test can be positive evidence for a linguistic relationship, a negative result means only that that particular test failed to find positive evidence" (Baxter y Manaster Ramer 1996: 382). Una comparación basada en listas léxicas entre albanés y galés, o irlandés y bengalí (ejemplo de Vogt 1988 [1971]: 502), por ejemplo, produce sorpresas desagradables para las aplicaciones probabilísticas, ya que pueden conducir, y de hecho conducen (Baxter y Manaster Ramer, loc. cit.), a resultados significativamente inferiores a los que se calculan para las coincidencias o convergencias accidentales. En consecuencia, puede afirmarse que el número de correspondencias efectivas entre lenguas emparentadas y las estimaciones acerca de las coincidencias o semejanzas debidas al azar son conjuntos de datos de cuya comparación no siempre se obtienen resultados que sean de alguna relevancia para la corroboración o la revisión de las aplicaciones del método comparativo.

en Bengtson y Ruhlen (1994), en gran parte deudoras de los principios de *mass comparison* profesados y defendidos por Greenberg (1987). A diferencia de la probabilidad “relativa” o “relacional” que centra el estudio de Ringe (cuyo fin es hallar la relación numérica entre unidades comunes que permita establecer con total garantía un parentesco lingüístico que nunca podría alcanzarse a través de accidentes casuales, por muy numerosos que éstos sean), en el trabajo mencionado ambos investigadores buscan la expresión numérica de la probabilidad absoluta de que se produzcan coincidencias puramente casuales entre dos o más lenguas, independientemente de su grado de parentesco.<sup>14</sup>

La concepción de Bengtson y Ruhlen aparece ilustrada en un ejemplo aparentemente sencillo: entre dos lenguas que comparten un consonantismo y un vocalismo mínimos de 7 y 3 unidades respectivamente la probabilidad de que coincidan exactamente (tanto en la forma como en el significado) entre las dos lenguas dos raíces del tipo CVC es del 1/147 (siendo 147 el número de raíces CVC que se obtiene mediante la combinación de los elementos fonológicos del hipotético sistema, vid. Bengtson y Ruhlen 1994: 282), o lo que es lo mismo, 0,007. Para el significado, si bien parece encontrarse aquí (y no sólo aquí) fuera de todo cálculo, habrá que presumir en este modelo lingüístico ideal la existencia de 147 unidades semánticas distintas (sin concesiones a la polisemia ni a la homonimia). Si la comparación se ampliasa a tres lenguas, la probabilidad de esa misma coincidencia se reduciría al  $(1/147)^2$ , es decir, 0,00049, y así sucesivamente.<sup>15</sup> Teniendo en cuenta que los sistemas fonológicos de las lenguas reales e incluso de las protolenguas son algo más complejos y que la estructura de las raíces presenta en esos casos, cuando menos, cierta variedad, la incidencia de la casualidad a efectos del establecimiento de posibles cognados en función del criterio de la identidad semántico-formal es, en esta visión, prácticamente despreciable. Otra cosa es que, dadas las posibilidades de variación apuntadas, estas estimaciones se ajusten a la realidad en virtud de la noción —siempre vaga— de semejanza o parecido (vid. *infra*). Desde tal perspectiva cualquier identidad formal y semántica entre tres o más lenguas será, por tanto, prueba de parentesco

(14) Pertenece a esta clase de cálculos absolutos el esquema estadístico, no excesivamente sofisticado, propuesto por Doerfer (1973: 69): “Haben wir z. B. zwei Sprachen A und B, beide sagen wir mit 100 Wurzeln und 20.000 Wörtern, so besteht a priori die Wahrscheinlichkeit, daß beide jede 100. Wurzel, d. h. insgesamt 200 Wörter, gemein haben”. Otros ejemplos de cálculos absolutos de la casualidad tanto en lenguas reales como en artificiales pueden hallarse en Fairbanks 1955 y Fodor 1982 [1966].

(15) No puede dejar de señalarse, no obstante, el carácter excesivamente simplificado del modelo propuesto. En él no se toman en consideración, además de las posibles restricciones combinatorias, las diferencias de frecuencia entre las unidades fonológicas (cf. Manaster Ramer y Hitchcock 1996: 607), algo esencial en las sistemas reales, y cuya consecuencia inmediata es la mayor probabilidad de coincidencia (o de correspondencia recurrente entre dos o más unidades distintas pertenecientes a otras tantas lenguas) que caracteriza a los fonemas de rendimiento funcional alto frente a los recesivos. Con respecto a la estructura de las raíces, reducida a una sola en el modelo ideal, lo que, evidentemente, nada tiene que ver con la realidad, existe además el reparo de la interpretación concreta tanto del símbolo C (en posición inicial y final) como del componente V. Harris (1990: 84-87) ha mostrado con gran nitidez las sustanciales diferencias materiales que puede ocultar una identidad estructural como CVC en indoeuropeo y en kartvélico.

genético, a menos que sea consecuencia clara de préstamos interlingüísticos, difícilmente identificables, por lo demás, en los estratos temporales a los que se remonta la comparación de *phyla* o *macrophyla* lingüísticos o simplemente de lenguas de hipotético parentesco remoto.

La práctica imposibilidad de deslindar herencia común y préstamos para las fases de reconstrucción a las que aspira la macrocomparación es una de tantas objeciones a esta práctica lingüística (sobre ello, cf., por ejemplo, Hegedüs 1992). En un artículo de alcance teórico general, Ivanov (1986 [1980]) argumentaba que las semejanzas de carácter gramatical excluyen en gran medida la posibilidad del préstamo (y de hecho, un número significativo de formas comunes reconstruidas para el nostrático son, por ejemplo, de naturaleza gramatical, vid. Aalto 1988: 57-9). Sin embargo, según la observación de Winter (1973: 144),<sup>16</sup> ningún elemento lingüístico es completamente inmune a las consecuencias del contacto de lenguas,<sup>17</sup> y unidades tanto estructurales como correspondientes al vocabulario “básico”, consideradas hasta hace poco estables y sobre las que se fundamentaban métodos de cálculo genético como la glotocronología, se hallan también sujetas —en grados que aún están por determinar— a la posibilidad del préstamo (cf. Haarmann 1990, y antes Teeter 1963: 641-2). Pero aun en el caso de las correspondencias gramaticales entre unidades morfológicas de distintas lenguas, sólo la coincidencia profunda, la que afecta incluso a los detalles, obliga a descartar cualquier otra explicación y a inclinarse por la del parentesco genético (cf. Vine 1991: 21).<sup>18</sup>

Las probabilidades de convergencia fortuita apuntadas hasta este momento se refieren todas ellas a la identidad formal y de significado entre lenguas distintas. La posibilidad de coincidencias casuales aumenta de forma indeterminada cuando en la comparación se admiten parecidos fonológicos entre palabras del mismo significado, lo que conduce al problema de la “accidental phonological *resemblance*”.<sup>19</sup> El factor de la casualidad crece en relevancia también de forma indeterminada en el momento en

(16) “No component of a natural language is totally immune to change under the impression of outside languages”. Acerca de la cuestión de la “unborrowability” de componentes como los pronombres personales, cf. también Matisoff 1990: 113.

(17) El préstamo, en cambio, no puede afectar, en opinión del propio Winter, que coincide en esto con la visión expresada por Ivanov, a sistemas paradigmáticos enteros de una lengua: “On the other hand, chances for a wholesale adoption of morphological features used in grammatical paradigms are very small, probably so small as to be equated with zero for all practical purposes” (Winter 1973: 145). Sobre el cambio morfológico por contacto cf., no obstante, Campbell 1995: 177-181. Vovin (1994b: 96-7) aplica esta misma limitación al vocabulario básico (a pesar de Haarmann 1990), cuando éste muestra correspondencias fonéticas regulares entre dos o más lenguas.

(18) Vine aplica esta regla, con un saldo negativo para la hipótesis nostrática, a las semejanzas entre los tipos nominales heteróclitos —“vaguely similar”— del protoindoeuropeo y del protodravídico.

(19) Suavizar los requisitos formales para que una correspondencia entre dos o más lenguas sea considerada como tal es un recurso que incide directamente sobre el factor casual en la comparación. Cuanto más tolerantes son las estimaciones de proximidad formal, mayor es el porcentaje hipotético de similitudes casuales al que ha de hacer frente el comparatista (cf. las diferencias resultantes de la aplicación del criterio estricto, el ampliado “extended” o el relajado “weakened” a las coincidencias basadas en la lista léxica (de 100 unidades, aunque ligeramente modificada) de Swadesh en Bender 1969: 530).

que al parecido fonológico se une la proximidad semántica. Bengtson y Ruhlen (1994: 281) atribuyen modestamente a la primera variante (“phonological resemblance”) probabilidades (en el esquema mencionado) del 2/147, 3/147, etc..., lo que, en cualquier caso, no constituye un peligro para la comparación multilateral.<sup>20</sup> Parece necesario insistir, sin embargo, en las consecuencias que se derivan de la conjunción de la proximidad fonológica y de la semántica. Con respecto al problema de la “bad semantics”, Ruhlen y Bengtson defienden que las conexiones de significado que ellos proponen “are in fact semantically more conservative than many proposed connections in Pokorny (1959)” (p. 289). Aun en el caso de que esto fuera así (opiniones opuestas, referidas, en general, tanto a la *long-range comparison* como a la comparación multilateral pueden encontrarse en Doerfer 1973, Campbell 1988, Vine 1991, Rankin 1992), resulta claro que no ocurre lo mismo con las aproximaciones fonológicas, que descansan sobre el criterio nunca delimitado de “similitud”. A diferencia de los comparatistas del ámbito estrictamente nostrático, los practicantes de la *mass comparison* renuncian en gran medida, por lo menos en la primera fase de la labor de clasificación de las lenguas, al establecimiento de correspondencias fonéticas entre los sistemas comparados<sup>21</sup> (y, en la generalidad de los casos, también a la reconstrucción de los protosistemas de cada grupo lingüístico, sobre esta cuestión cf. *infra*). Tal vez merezca la pena recordar en este punto, como lo ha hecho recientemente A. Vovin (1994a: 370), y no es el único, el modo en que, mediante el uso del método comparativo tradicional, se determina la relación genética entre lenguas:

I believe that two or more languages are genetically related if: (1) they share a

(20) Teniendo en cuenta esta laxitud material, a la que se une, además, en no pocas ocasiones la semántica, las estimaciones previas sobre los parecidos debidos al azar deberían razonablemente ser un poco más generosas. Para este tipo de operaciones no resulta exagerada la estimación de Fairbanks (1955: 117), que cifraba en torno a un 40% el conjunto de semejanzas casuales entre lenguas de estructura fonológica próxima. La probabilidad de hallar semejanzas de forma y de significado entre raíces o segmentos de estructura CVC pertenecientes a diversas lenguas es *a priori* relativamente alta (cf. el apéndice 2 del presente trabajo, donde se propone un modelo de aproximación a la casualidad a partir de la noción de similitud material), teniendo en cuenta las restricciones de combinación impuestas a los fonemas, el carácter limitado del propio inventario fonológico (lo que da lugar a un conjunto determinado de elementos) y lo extenso del vocabulario, causa inevitable a su vez de la frecuente coincidencia de significados –entre unidades materialmente parecidas– incluso dentro del vocabulario “básico” (cf. Watkins 1990: 292). Dyen (1970: 439), por ejemplo, se siente impresionado por la cantidad de elementos accidentalmente comunes que pueden llegar a mostrar sistemas tan alejados entre sí como el protoindoeuropeo y el protoaustronesio. Por otra parte, no han sido pocos los comparatistas que, desde los tiempos de W. Humboldt y F. Bopp, han pensado en el parentesco de las lenguas indoeuropeas y austronesias (vid., recientemente, Petrov 1997).

(21) Operación que constituye el instrumento básico para que la reconstrucción así como la postulación de cognados puedan ser verificables. Esto es un rasgo esencial que separa a la teoría nostrática de la comparación multilateral de Greenberg y de las etimologías globales (cf. Manaster Ramer, citado por Trask 1996: 390-1). Cf. asimismo Salmons 1992: 223: “Global etymologizing requires dismissing the most central tenets of linguistic reconstruction and makes its results no longer falsifiable”. Sobre el valor de las leyes fonéticas en la comparación cf. Doerfer 1973: 43: “[Der Lautgesetz] allein garantiert eine genügend hohe Zahl möglicher Wurzeln; ohne ihn geraten wir ins Chaos bloßer Wahrscheinlichkeits- und Zufallgesetze”. Greenberg sitúa, no obstante, el establecimiento de correspondencias fonéticas en una fase posterior de la comparación, una vez que se han clasificado las lenguas e identificado los grupos en los que se integran, y cumplido, por lo tanto, el cometido principal de la actividad comparativa que Greenberg preconiza (Greenberg 1990: 8, citado por Bengtson y Ruhlen 1994: 285-6). Cf. también Greenberg 1987: Cap. I.

significant number of common basic vocabulary and basic morphological markers (if there is any morphology) established on the basis of regular phonetic correspondences, and (2) lexical comparisons include a significant number of regular phonetic correspondences... When we try to establish remote genetic relationships (older than, say, IE, Uralic or Semitic, or Proto-Austronesian), it is imperative to work not with actual forms of modern or old languages, but with reconstructions of linguistic groups or families.<sup>22</sup>

Del modo en que se establecen ecuaciones léxicas en el sistema de Bengtson y Ruhlen se deriva, en cambio, una absoluta indeterminación en lo que se refiere a las conexiones formales entre las unidades léxicas o morfológicas consideradas, entre las que en ocasiones es suficiente que coincida o se aproxime un solo elemento de cada unidad (a veces ni siquiera eso) para postular su origen común, elemento que puede perfectamente variar —en principio, sin límite— cuando se comparan otras formas entre las mismas lenguas o grupos de lenguas. Cf. unos cuantos ejemplos: IE. *\*pol* ‘half, side’, Austroasiatic *\*(m)bar* ‘2’; IE. *\*(s)per* ‘to fly’, Austroasiatic: Proto-Munda: *\*apir* ‘to fly’; IE. *\*bhaghu(s)* ‘arm, forearm, elbow’, Austroasiatic (?): Semang *pak* ‘arm’;<sup>23</sup> Nilo-Saharan: Zagawa *sina* ‘nose’, Basque *su-dur* ‘nose’, *sun-da* ‘smell’;<sup>24</sup> Nilo-Saharan: Lotuko *mij* ‘to dream’, Basque *mun* ‘medulla’;<sup>25</sup> pero Nilo-Saharan: Nyangiya *sim-at* ‘hair’, Nandi *sum*, Basque *zam-ar(r)* ‘lock of wood, shock of hair’.<sup>26</sup> Este modo de obrar, ajeno a los procedimientos de la metodología comparativa (véase en el apéndice 1 a este artículo un intento de búsqueda de correspondencias y sus consecuencias) y bautizado por Fodor (1982 [1966]) como “método de las semejanzas”, se encuentra lejos, por lo tanto, de las premisas del modelo teórico e

(22) Estos principios, no obstante, deben ser matizados en conceptos como el del vocabulario básico, no siempre impermeable a la acción del tiempo y del contacto lingüístico (cf. Haarmann 1990: 2: “The “basicness” of certain sectors of the vocabulary cannot be substantiated because the notion is, in itself, a misconception”. Vid., no obstante, Vovin 1994b: 96-7).

(23) La eventual relación entre i.e. *p*, *bb* y austroasiático *b*, *p*, aunque apunta a ser caótica, podría perfectamente encajar en parámetros no del todo extraños. Pero la ausencia de otros elementos de contraste (y de cualquier otra información) no permite concluir nada acerca de ese supuesto vínculo. Para la relación entre segmentos fonológicos con mayor presencia en las etimologías globales vid. el apéndice 1.

(24) Nada permite afirmar que las formas *su-dur* y *sun-da* estén emparentadas (y menos aún si las segmentaciones morfológicas son correctas). Debo esta observación y las dos que siguen a J. A. Lakarra (cf. en cualquier caso Trask 1995a: 62, 1997: 409).

(25) *Mun* (de localización occidental) no puede de ningún modo ser forma antigua en vasco (la sonante inicial *m*- y la forma más extendida *muin* permiten reconstruir *\*buni*).

(26) *Zam-ar(r)* es un claro ejemplo (más) de segmentación morfológica completamente gratuita. La forma con *m* media tampoco puede ser antigua. Trask (1995a: 69, 1997: 409) señala además el error en la definición del significado. No es “lock of wool” (“vedija, mechón de lana”), sino “sheepskin cloak” (“capa de piel de cordero/oveja”), pero la objeción —a juzgar por las acepciones que registran distintos diccionarios (Lhande, Azkue)— parece excesiva. Con respecto a las múltiples correspondencias que cabe construir a partir de semejantes ecuaciones, Doerfer (1995: 254) hace el siguiente comentario (referido en su caso a las correspondencias altaicas de Starostin): “This recalls Voltaire’s words “L’étymologie est une science où les voyelles ne font rien et les consonnes font peu de chose”... It is obvious that when the vowels are *de facto* disregarded the number of available comparisons augments *enormously*. This, however, is no method, but a trick” (*cursiva* del propio autor).

hipotético que minimiza el influjo de la casualidad en las comparaciones interlingüísticas. La cantidad nada desdeñable de concesiones que se hacen a la variación indeterminada, tanto en el aspecto fonológico como en el semántico, de las unidades supuestamente emparentadas, priva a este tipo de aproximaciones de la más mínima garantía, puesto que de la misma manera en que es del todo imprevisible, aunque siempre igualmente válida para el objetivo que se persigue, la cambiante relación entre formas “similares” de dos o más lenguas, es enormemente alta la probabilidad de hallar ecuaciones de ese tipo entre cualesquiera conjuntos de palabras que uno desee cotejar (vid., por ejemplo, las primicias de comparación malayo-altaica con que Doerfer (1995: 261) ilustra esta posibilidad; pero no se trata de un ejemplo aislado: con idéntica facilidad Campbell (1988: 607-8) configura listas léxicas coincidentes en finés y el penutio de Greenberg, Dyen (1970) compara con resultados sorprendentes (incluidas correspondencias fonéticas en apariencia regulares) el protoindoeuropeo y el protoaustronesio y Trask hace lo propio con vasco e inglés (1995b: 196-7) o con vasco y húngaro (1997: 412-4).<sup>27</sup> Nada nos garantiza en estos casos que las similitudes que podamos haber hallado, si operamos con una cantidad imprecisa e indeterminable de variables, no sean casuales.

Se alcanza de esta manera una grave contradicción entre el esquema de base con el que se trata de legimitar —arrinconando el azar con cálculos en principio aceptables— toda posible coincidencia fonético-semántica, y la actividad comparativa que se supone amparada en esos mismos presupuestos. Desde el punto de vista de la coherencia, no parece correcto erigir un modelo que contempla de forma casi exclusiva coincidencias absolutas (tanto en la forma como en el significado), donde la casualidad, consiguientemente, significa bastante poco, y aplicar después sus conclusiones a realidades lingüísticas seleccionadas con criterios mucho más relajados (vid. desde este punto de vista la propuesta de carácter probabilístico y basada en la semejanza de las unidades fónicas que se ofrece en el apéndice 2 de este artículo). Lo

(27) Recordemos, en cualquier caso, que la comparación multilateral va precisamente encaminada a evitar en lo posible el efecto del azar, que tan profunda incidencia tiene en las comparaciones binarias. Si a vasco *ekarri* y *eder* no resulta difícil encontrar paralelos casuales en inglés (*to carry* y *dear* (Trask 1995b: 196), será algo más complicado, de no haber vínculo histórico entre las lenguas, unir terceros miembros a esas ecuaciones, complicación que crece geométricamente si se añaden otras lenguas a la comparación. Atendiendo escrupulosamente a las consecuencias de este modo de obrar, la coincidencia recurrente entre formas de más de dos lenguas alejadas entre sí constituiría, por lo tanto, indicio casi incontestable de parentesco, aun en el caso de que no se percibiesen correspondencias fonéticas regulares entre las unidades léxicas o gramaticales cotejadas (acerca de parentescos casi seguros, pero sin un sistema completo o regular de correspondencias (al menos en una primera fase de aplicación del método comparativo), cf. Hardman de Bautista 1978: 151 y Grace 1990: 166 y 1992; para hallar un marco teórico capaz de dar explicación a estos casos de parentesco “evidente” pero “indemostrable”, vid. ahora Dixon 1997). Pero esta versión ideal del método de la multicomparación se halla lejos, como ha observado Newman (1993: 235), de la práctica que encuentra reflejo en Greenberg 1987 y también —podemos añadir— en las etimologías globales. No existen vocabularios multilaterales que no sean el conjunto incompleto de aproximaciones léxicas entre lenguas que no siempre son las mismas para el mismo “étimo”. En el mejor de los casos bajo las listas de supuestos cognados se oculta una especie de binariedad camuflada por la yuxtaposición de decenas de lenguas que en modo alguno corresponde al procedimiento que limitaría objetivamente el alcance del azar.

que pone de manifiesto esto último es, a fin de cuentas, el escaso valor que resultan tener para los resultados de la *praxis* macrocomparativa las estimaciones matemáticas apriorísticas de Bengtson y Ruhlen acerca de la probabilidad de coincidencias casuales, puesto que el factor con el que en realidad se opera es el de la similitud parcial, no sujeta, además, a reglas ni en el plano semántico ni en el fonológico, entre formas de dos o más lenguas.<sup>28</sup> Éste es precisamente el extremo opuesto al método que guía la labor del “prehistoriador de la lengua”: “The prehistorian has to go through a number of carefully planned and executed steps. His immediate goal is to replace the vague notion of similarity or loose partial agreement (a notion which form the basis of work in lexicostatistics, glottochronology, and large-scale comparison aimed at establishing distant relationships among languages and language groups) by the much more stringent one of regular correspondence (Winter 1990: 15). Acerca del método o más bien la práctica heurística (Campbell 1988: 597) basada en las semejanzas (parciales) en su aplicación a las clasificación de las lenguas africanas cf. Fodor 1982 [1966].

Una de las bases teóricas o justificaciones de la comparación multilateral reside específicamente en la imposibilidad de hallar correspondencias regulares entre formas pertenecientes a lenguas de parentesco remoto.<sup>29</sup> El método comparativo se muestra en estos casos incapaz de demostrar la existencia de tales relaciones, borradas por el paso del tiempo, por lo que se ha de recurrir a otros sistemas (Ruhlen 1991 [1987]: 123), en especial, al método de las semejanzas, detenidamente analizado en Fodor 1982 [1966]: 65-71. Teniendo en cuenta la radical diferencia que separa a la correspondencia fonética de la mera semejanza, en la que por definición no resulta posible establecer principios para enjuiciar la validez de las formas comparadas (esto es, no hay límite que restrinja el alcance de las similitudes ni que controle el componente altamente subjetivo que entraña el propio concepto de similitud, cf. Manessy 1990: 92),<sup>30</sup> todo lo anterior equivale a decir que la relación genética de

(28) Cf. Campbell y Goddard 1990: 25 o Hock y Joseph 1996: 502 sobre la ausencia de un criterio de similitud que explique las operaciones de comparación que se realizan en Greenberg 1987.

(29) Otra de tales premisas la constituyen los obstáculos a veces insalvables a los que ha de hacer frente la comparación tradicional (por ejemplo, las correspondencias múltiples e inmotivadas entre lenguas de parentesco reconocido, cf. los datos de las lenguas swahili, konde y sango en Greenberg 1990 [1969]: 442). La aplicación del método comparativo al estudio de lenguas de historia desconocida (o de notorio alejamiento mutuo) entraña serias dificultades cuando se toman en cuenta, al menos de manera hipotética, los diversos cambios (fonológicos y morfológicos) que pueden haber conducido a la desaparición del contexto que explicaba la variedad de correspondencias entre unos sistemas y otros (acerca del “context-effacing change”, cf. Hamp 1992: 98). La consecuencia de ello es que la eficacia del método resulta efectivamente mermada (vid. para el camito-semítico Chaker 1990); sin embargo, cuando se trata de la imposibilidad de recuperar toda correspondencia fonética, la ausencia de instrumentos alternativos válidos hace que en esos casos, como recuerda F. Schwink (1994: 4), la relación genética, exista o no, resulte sencillamente indemostrable.

(30) Cf., en cualquier caso, en el ámbito de las comparaciones con apoyo matemático el singular procedimiento seguido por A. Jones (1989: 182-4) para buscar una hipotética relación entre las familias australiana, dravídica y urálica. Su criterio de semejanza sí está delimitado en lo que se refiere tanto a la forma como al significado: en cuanto al material fónico, las formas léxicas candidatas a cognados deben satisfacer la exigencia de una coincidencia absoluta (identidad) en la secuencia CV

lenguas entre las que las correspondencias regulares son inexistentes o bien irre recuperables (lo que a fin de cuentas viene a ser lo mismo) es simplemente indemostrable (Schwink 1994: 3-4),<sup>31</sup> al menos mientras los teóricos de la *long-range comparison* no proporcionen un instrumento que permita fundamentar de manera empírica las suposiciones de parentesco lingüístico: la mera semejanza, en cuanto dominio ajeno a definiciones objetivas que permitan fijar la diferencia —o la gradación— que separa a los casos de parecido debidos a un origen común de las coincidencias sin significación genética, no puede constituirse en un criterio de valor probativo equiparable al que posee la correspondencia fonética,<sup>32</sup> cuya identificación, por lo demás, se realiza en función de distribuciones regulares observables y verificables de forma objetiva (cf. Hamp 1992: 96-7) y no implica necesariamente la noción de similitud<sup>33</sup> (pueden existir correspondencias fonéticas entre elementos perfectamente dispares, baste recordar el ejemplo preferido de A. Meillet, arm. (*j*)*erk-*, gr. ai. esl. *dw-* o la correspondencia entre los tonos vietnamitas y ciertas consonantes de las lenguas mon-khmer descubierta por A. G. Haudricourt, vid. Manaster Ramer y Hitchcock 1996: 605).<sup>34</sup> Y puede decirse incluso más: guiarse por la semejanza a la hora de buscar candidatos léxicos a la lista de cognados puede entorpecer el proceso de identificación de verdaderos cognados en lenguas efectivamente emparentadas. Según las estimaciones más favorables de Hock y Joseph (1996: 493), al menos uno de cada dos emparejamientos léxicos basados en la semejanza entre el hindi y el inglés es puramente accidental. En el otro extremo, la aplicación de correspondencias constituye la herramienta que permite poner en relación, pese a las diferencias fonéticas nada triviales, sáns. *akra* con el inglés antiguo *hweogol* “rueda”, o esl. *višī* con gr. *oĩkos* “casa”, incluso avéstico *raox šnā* con latín *luna*, etc... Los resultados que produce un método basado en nociones imprecisas de similitud sólo pueden dar satisfacción material a los mismos criterios por los que aquél se rige, sin que ello signifique en ningún momento que esos resultados posean relevancia genética (Campbell y Goddard 1990: 26).

Conviene no olvidar, en cualquier caso, que las correspondencias se producen con mayor frecuencia —al menos en una primera etapa del análisis comparativo— entre elementos fácilmente discernibles por su similitud. Para K. V. Teeter (1963: 641) “it is true that closely related languages will be likely to have more cognates, and,

---

inicial, independientemente de su segmentación morfológica (para el comentario de estos criterios y la crítica de la elaboración probabilística posterior de los datos, vid. Ringe 1995b).

(31) Circunstancia a todas luces lamentable, pero en la que no hay por qué ver nada especialmente trágico. Sencillamente expresa y recuerda los límites reales del conocimiento empírico.

(32) “Ce n'est pas avec des ressemblances de formes qu'on opère, quand on compare des langues d'une même famille, mais uniquement avec des règles des correspondences” (Meillet 1970 [1925]: 31).

(33) “La parenté n'implique aucune ressemblance actuelle des langues considérées, ni surtout du système général des langues considérées; et inversement il y a beaucoup de ressemblances, soit de structure générale, soit de vocabulaire, qui n'impliquent pas parenté” (Meillet 1926 [1914]: 92).

(34) Más aún, son las correspondencias entre elementos que presentan un cierto grado de diferencias fonéticas las que se consideran pruebas más concluyentes para fundamentar la hipótesis del parentesco. Acerca de la búsqueda (y su significación) de “non-obvious matchings” vid Campbell 1995: 182. Las ecuaciones complicadas corresponden, en la división de Watkins 1990, a la tercera fase etimológica.

since cognate items of closely related languages are usually similar, that such languages will tend to have longer lists of similar items, and of course it is precisely for this reason that lexical similarities are universally recognized as useful for suggesting to an investigator where he might apply the comparative method first".<sup>35</sup> Pero esta transferencia de la relación entre cognados y semejanzas al ámbito de las lenguas cuyo parentesco se desconoce no entraña necesariamente que la similitud entre unidades léxicas constituya por sí sola el único criterio heurístico —y menos aún el factor probatorio— de nuevas relaciones genéticas. La semejanza, si es insistente, en todo caso indica al comparatista, según aclara Teeter, el lugar donde puede dar inicio a sus pesquisas. El parecido léxico superficial, aquel que llama primeramente la atención, pertenece, pues, al dominio de la inspección (cf. Watkins 1990: 292): las expectativas (genéticas) que suscita han de ser adecuadamente contrastadas con el análisis sistemático de los elementos tanto coincidentes como divergentes de las lenguas comparadas.<sup>36</sup>

El conjunto de características de la *mass comparison* señaladas hasta este momento hace pensar en una diferencia fundamental (basada tal vez en la actitud frente al objeto de estudio) con respecto al método comparativo clásico. Éste opera —y no es algo que estemos descubriendo ahora— por eliminación, descartando todos aquellos elementos comunes entre dos o más lenguas que puedan ser explicados por contacto lingüístico, por el carácter fonostésico de las unidades consideradas, o cuya divergencia semántica no se sujete a patrones de variación o de evolución suficientemente conocidos. Sólo entre las unidades restantes, las que no admiten ninguna otra clase de explicación, se establecen, si las hay, las correspondencias fonéticas regulares como paso previo a la proclamación del parentesco. La comparación multilateral, por el contrario, sortea la fase del escrutinio al tiempo que invierte las etapas del proceso entero: todas las semejanzas son válidas y demuestran la relación genética hasta que no se demuestre lo contrario (que es lo que ocurre con sospechosa frecuencia, sobre todo en el caso de los préstamos, cuando investigadores de cada grupo lingüístico afectado por la macrocomparación pasa revista a todo elemento involucrado *malgré lui* en la confección de los étimos, vid. Campbell 1988: 599-600, Kimball 1992: 449).

En la renuncia tácita o expresa a reconstruir los protosistemas de los que se

(35) En términos muy próximos se había expresado con anterioridad A. Meillet (1926 [1914]: 97): "Si donc on peut d'abord constater des ressemblances de vocabulaire entre deux ou plusieurs langues pour indiquer de quel côté il faut chercher, ce n'est pas de là que peut venir une démonstration définitive; le vocabulaire ne peut servir qu'à orienter la recherche; la preuve se trouve ailleurs".

(36) Muy distinta es la práctica —proclamada por Greenberg (1987: 5)— del escrutinio a simple vista de los grados de similitud entre distintas formas. "Is a form A more like B than it is like C?. Given, for example, *pan/fan/ezuk*, who would hesitate?". El problema comienza cuando estas diferencias de grado en el parecido fónico —suponiéndole identidad al significado— adquieren automáticamente relevancia genética en las explicaciones comparatistas. Porque, como recuerda Matisoff (1990: 111), ante ecuaciones igualmente posibles del tipo *er/erku/duo* (mandarín, armenio y latín, respectivamente), ¿no parece más prudente dudar siquiera un instante, antes de extraer consecuencias clasificatorias que pueden resultar excesivamente apresuradas? ¿Es este tipo de semejanzas las que tiene en mente Ruhlen (1991: 124) cuando propone que el comparatista debe contentarse "with "inferences... so conclusive that they can be accepted as proof"?"

derivan las lenguas supuestamente emparentadas<sup>37</sup> late, en resumidas cuentas, el peligro que Doerfer (1973: 69) denominó “casualidad dinámica” (*dynamischer Zufall*), es decir, la posibilidad de convergencias lingüísticas inmotivadas (accidentales) generadas por evolución independiente de cada lengua y cuyo reconocimiento sólo es posible si se atiende a la historia interna (documentada o reconstruida) de las unidades comparadas. El ejemplo del propio Doerfer es como sigue: la raíz árabe *qtl* “matar” puede fonéticamente aproximarse a la forma del islandés antiguo *guðr* “guerra”. Pero se trata de una comparación en la que la sincronía oculta diferencias decisivas: *guðr* procede de la raíz i.e. \**g<sup>u</sup>ben-* (gr. *θενίω*, sáns. *ghnánti*, *ghaná*), la consonante ð es producto de la evolución del grupo *-nd-*, y la sonante final (-*r*), marca de nominativo singular, se remonta a i.e. \**-s*, todo lo cual desfigura la semejanza inicial. Esto es un caso de convergencia fonética debida a la casualidad dinámica, factor al que hay que suponer una incidencia notable en las comparaciones masivas (aunque la naturaleza de esa casualidad sólo sea identificable en aquellos casos en que se posee algún conocimiento de la historia de la lengua, ya sea por medios filológicos o comparativos, o bien gracias a la reconstrucción interna). Existe asimismo la convergencia semántica, como ilustra el caso de la forma *amar* “hablar”, correspondiente a un dialecto esquimal, y de la raíz hebrea *'mr*, de idéntico significado, pero que se remonta a “ver”. La casualidad dinámica, manifestada en este tipo de convergencias independientes (y causa única de coincidencias como la de inglés *bad* y persa *bad*, citadas al inicio de este trabajo y en las que queda representada la mayor parte de las similitudes o coincidencias materiales debidas al azar), no tiene cabida en los cálculos de Bengtson y Ruhlen.

Comparar directamente lenguas de diversas familias (y no los protosistemas obtenidos a partir de los datos comunes de cada grupo lingüístico), práctica habitual en la comparación multilateral, facilita obviamente la tarea de acumular raíces similares en algunas (casi nunca todas ni tampoco las mismas) de las lenguas estudiadas, pero al mismo tiempo multiplica las posibilidades de coincidencias casuales no contempladas *a priori* en el modelo probabilístico que proporciona sustento matemático al método (cf. Ringe 1995a: 60, Manaster Ramer y Hitchcock 1996: 614). Hay que tener en cuenta que ciertos (macro)grupos lingüísticos sometidos a la comparación (en especial, a la global) cuentan con centenares e incluso un millar de lenguas (cf. el grupo Niger-Congo, el austronesio o el amerindio), y que el testimonio de sólo unas cuantas de ellas —el cual, con un alto índice de frecuencia, no provendrá de las

(37) El hecho de que las reconstrucciones varíen notablemente a lo largo del tiempo y en función de quien reconstruye es para Greenberg motivo suficiente para descartarlas como mecanismo conducente a la demostración del parentesco: “Reconstruction becomes neither a necessary nor a sufficient condition for languages to be related” (Greenberg 1995: 212). Sin embargo, hay que recordar que para otro *long ranger*, S. Starostin, la reconstrucción de protolenguas es una operación absolutamente imprescindible para los fines perseguidos por la comparación de lenguas de relación remota y dudosa (Starostin 1992: 76). Sobre la utilidad de la reconstrucción para las hipótesis de parentesco lejano vid. Vovin 1994b: 97). Para Hymes (1993 [1959]) la reconstrucción es la operación mediante la que, una vez probada la relación genética, se lleva a cabo su establecimiento. Vid. su útil división en tres fases del estudio de relaciones genéticas: hipótesis, demostración, establecimiento; y el paralelismo que guarda con ella la división de la actividad etimológica —también en tres fases— propuesta por Watkins (1990: 294).

mismas lenguas de manera constante— resulta en ocasiones suficiente para insertar a todo su grupo en el conjunto lingüístico que da lugar al “étimo global”. El factor de la casualidad depende así no sólo de la proximidad o alejamiento mutuo de los sistemas que se comparan, sino que varía de manera notoria en función del método o los criterios que subyacen a la comparación.

Se ha de señalar, por último, un factor más que contribuye a desbaratar los fundamentos sobre los que se asienta la visión del papel de la casualidad en la comparación lingüística dentro de lo que consideramos concepción optimista de ésta: los errores en las formas comparadas. Errores que afectan en ocasiones a la estructura fonológica general de las palabras, pero que con excesiva frecuencia son debidos a la segmentación y al análisis caprichosos y no sujetos a reglas de los componentes morfológicos de las formas con las que se completan las listas de supuestos cognados, lo que es consecuencia de la manipulación incorrecta e interesada de los datos lingüísticos (vid. algunos ejemplos de ello en Campbell 1988, Salmons 1992 o Hock y Joseph 1996: 498-502).<sup>38</sup> Es aquí de nuevo sintomático que las objeciones a estos conjuntos de formas partan precisamente de especialistas de cada campo específico, críticas a las que no se sustrae la teoría nostrática (vid. las reservas desde el ámbito altaico en Doerfer 1995 y en Shjerbak 1994, desde el urálico en Serebrennikov 1986, desde el indoeuropeo en Vine 1991 o desde el kartvélico en Klimov 1991a, vid. a su vez la respuesta a este último en Manaster Ramer 1995),<sup>39</sup> pero que se multiplican en áreas como la clasificación de lenguas americanas de Greenberg (Chafe 1987, Goddard 1987, Campbell 1988, Adelaar 1989, Rankin 1992, Kimball 1992, Poser 1992, Berman 1992). La insuficiente delimitación de las formas propias y los préstamos en las macrocomparaciones, unida a la presencia demasiado frecuente de vocabulario expresivo en las listas de *comparanda*, es asimismo objeto de crítica (cf., por ejemplo, acerca de los préstamos indoeuropeos en kartvélico Klimov 1994a y 1994b: 472-3; para el vocabulario expresivo Vine 1991: 11-12).<sup>40</sup>

La cantidad de errores y la imposibilidad de computarlos (o de ponerles coto) constituyen las causas que dificultan, en opinión de Ringe (1996: 148), todo intento de aplicación de métodos probabilísticos o estadísticos al corpus de datos amerindios que proporciona el diccionario etimológico amerindio de Greenberg.<sup>41</sup> La introduc-

(38) La segmentación arbitraria de la morfología con vistas a la adecuación de las formas de la lengua A a la estructura de las unidades de la lengua B y viceversa es una práctica con cierta tradición. Sin tener que ir demasiado lejos Michelena (1988a [1950]: 440) la identificó en las comparaciones vasco-caucásicas de K. Bouda, quien descomponía *m-otz* y *ma-ts* de acuerdo con sus necesidades heurísticas.

(39) Cf. incluso la crítica de los métodos de selección de *comparanda*, no suficientemente rigurosos, de Bomhard (1984) por parte de quienes en modo alguno se oponen a la hipótesis del parentesco entre el indoeuropeo y el afro-asiático (Palmaitis 1986, Helimsky 1987).

(40) El problema que constituyen los préstamos (cf. Michelena 1988a [1950]: 441-2) especialmente para quien pretende demostrar una relación genética es a veces sencillamente menoscabado mediante la apelación al factor “probatorio” externo que proporciona la *long-range comparison* (así, vasco *keatu* podría aproximarse con más razón a caucásico \**gäru(jV)* que al latín *cattus*, toda vez que “the Basques may already have had the word before the Romans came”, cf. Bengtson 1995: 92) o bien por medio del recurso a la elaboración de reconstrucciones *ad hoc*, es decir, absolutamente arbitrarias (ejemplos de ambas prácticas pueden hallarse en Bengtson 1995; cf. la respuesta a ellas en Trask 1995b).

ción de una variable más (y en este caso prácticamente incontrolable) en la labor comparativa aumenta ostensiblemente la probabilidad de coincidencias casuales entre lenguas no emparentadas, ya que a las formas reales se unen formas o segmentos léxicos o morfológicos inexistentes como tales en las lenguas respectivas. Vovin 1994b: 103, por ejemplo, detecta por su parte lo que él denomina “lexical ghosts” en el aparato etimológico en que se erige la hipótesis del origen austronesio del japonés según P. Benedict (formas espúreas que aparecen asimismo en los diccionarios de Greenberg, cf. Poser 1992: 218-9).

El concurso de todos los factores enunciados y comentados hace que la prefiguración del modelo teórico probabilístico en que se ampara la visión optimista de parentescos lingüísticos remotos, más próxima en ocasiones a la voluntad de fe que a la práctica científica (Rankin 1992: 346), no tenga la más mínima relación con el influjo real —aunque no sujeto a cálculo (por las dificultadas aducidas)— de la casualidad en las aproximaciones léxicas con que se intenta demostrar el parentesco lingüístico. Se trata de una contradicción en la que resulta necesario insistir (cf. Manaster Ramer y Hitchcock 1996: 616), puesto que el esquema apriorístico construido a partir de coincidencias fonético-semánticas absolutas (Bengtson y Ruhlen 1994: 281-2), en su relación con un método de comparación que, en teoría y aplicado de forma rigurosa, anularía la incidencia del azar, parece ser uno de los escasos asideros —una vez eliminados los criterios de *mass comparison* que sólo satisfacen a sus practicantes— en que se sustenta y al que parece aferrarse en defensa de su existencia la actividad multicomparativa.

#### 4. Algunas conclusiones

El control estadístico de las coincidencias lingüísticas debidas al azar en el marco de los trabajos de comparación encaminados a postular determinados parentescos genéticos entre lenguas o grupos de lenguas, en los que aquéllos no resultan claros (o al menos no convencen a la mayoría de los comparatistas), podría desempeñar la función que la tipología cumple con respecto a las reconstrucciones protolingüísticas.<sup>42</sup> El problema que suscita ese control radica seguramente en la existencia de diversos métodos probabilísticos para tratar de captar la incidencia de la casualidad

(41) Kimball (1992: 448) distingue varios tipos de errores recurrentes tanto en Greenberg 1987 como en los cuadernos de su diccionario amerindio. Entre ellos destaca la transcripción incorrecta de formas, la asignación errónea de unidades léxicas a lenguas que no las conocen (en este punto cf. también Rankin 1992: 342-3) o el uso de la misma unidad léxica, con significados profundamente dispares, en distintas correspondencias etimológicas.

(42) Función consistente en el control, la verificación o incluso el diagnóstico (operaciones que parten del conjunto de datos tanto sincrónicos como diacrónicos procedentes de las lenguas registradas y estudiadas) de los sistemas lingüísticos a los que se accede por medio de la comparación. Esta función es congruente con la idea según la cual las protolenguas han de caracterizarse por los mismos rasgos que muestran las lenguas reales, por lo que “a conflict between the reconstructed state of a language and the general laws which typology discovers makes the reconstruction questionable” (Jakobson 1971 [1958]: 528). El empleo de los datos tipológicos contribuye en esta visión a evitar la reconstrucción de sistemas contrarios a la naturalidad lingüística y a preferir, entre varios sistemas reconstruidos sobre idéntica base comparativa, aquel que resulta tipológicamente más probable (punto éste algo más

en la labor comparativa y, más aún, en la dependencia que esos métodos muestran con respecto a sus propias reglas, difícilmente adecuables, en ocasiones, a los datos empíricos de los que pretenden dar cuenta. En cualquier caso, y a diferencia de la tipología, en la que se ha querido ver un instrumento de signo no sólo negativo (control y revisión de sistemas reconstruidos altamente improbables), sino también positivo (deducción de los estrados y evoluciones lingüísticos más probables, cuando éstos no pueden ser directamente reconstruidos, cf. Schwink 1994: 36, Bomhard 1996: 8), la determinación matemática de la casualidad en las coincidencias o similitudes lingüísticas parece más bien limitada, en el mejor de los casos, a poner en cuestión supuestos parentescos que no rebasan cuantitativamente los parámetros atribuidos a la acción de la casualidad. (Convenimos, pues, aunque parcialmente y a pesar de las diferencias de los métodos probabilísticos que son objeto de examen, con Bender (1969: 530) en que "it is not clear... that mathematical considerations of the sort discussed in this article can accomplish more than ruling out chance convergences").<sup>43</sup> Pero a lo largo de estas líneas se ha podido comprobar que esos mismos parámetros difieren significativamente (hasta el punto de representar extremos opuestos, como los que encarnan Ringe 1995a, por un lado, y Bengtson y Ruhlen 1994, por otro) en función del punto de partida del cálculo así como de la visión teórica en la que éste se integra, circunstancia que delata en la actualidad la inexistencia de un criterio formal único —por objetivo— mediante el que determinar el alcance del factor casual en la comparación lingüística. Pero podría decirse tal vez más: no es claro que la proporción o el número exacto de convergencias debidas únicamente a la casualidad pueda ser en general objeto de determinación matemática. La conclusión de Manaster Ramer y Hitchcock (1996: 617) es, a este respecto, bastante pesimista: "Many problems... are simply not susceptible to precise mathematical treatment. We doubt, for instance, that it will ever be possible to assign a numerically precise value to the probability that two given words from different

---

controvertido, cf., en cualquier caso, Schwink 1994: 36). No obstante, es preciso recordar que el dominio de las relaciones entre tipología lingüística y reconstrucción (especialmente en los niveles fonológico y sintáctico) no es precisamente uno de los campos de investigación que más se distinga por el consenso general entre los especialistas (véanse, por ejemplo, las reparos tempranamente manifestados por Michelena (1995 [1966]: 223) a la propuesta de Jakobson; cf. posiciones encontradas en Gamkrelidze 1981 y Dunkel 1981; aplicaciones concretas de las bases de datos lingüísticos (sincrónicos y diacrónicos) a la reconstrucción puede encontrarse en Vennemann 1986: 65 y en Timberlake 1990; vid. Manaster Ramer 1997 para la tipología del nostrático; acerca del uso y del abuso de la tipología vid. Stevens 1992, Hock 1986, 1992; véase, en general, para las relaciones entre tipología y reconstrucción, Schwink 1994, Comrie 1989: 288 y 1993, Fox 1995: 250-270).

(43) Hay que puntualizar, sin embargo, que esos métodos no siempre descartan de manera fiable un porcentaje necesario de convergencias casuales. Así como para lenguas de comprobado parentesco genético (albanés y armenio, pongamos por caso) los sistemas de cálculo estudiados por Bender pueden llegar a establecer un índice de coincidencias casuales (entre unidades de listas léxicas) que resulte insuperable, aplicados, en cambio, a lenguas de relación mutua equiparable a la que existe entre finés y cakchiquel (lengua maya, Campbell 1988: 596), aquéllos pueden, por el contrario, arrojar resultados por encima del límite asignado a las coincidencias debidas al azar, pese a lo cual nadie (en principio) pensará en emparentarlas.

languages will resemble each other by chance. Assigning numerical values to such possibilities creates a false sense of precision”

Con independencia de estos cálculos, el instrumento material objetivo que permite separar en la práctica las coincidencias casuales de las que no lo son se ha de buscar, por lo tanto, en la correspondencia fonética regular, sean o no similares los segmentos fónicos entre los que se produce esa recurrencia. Dentro del método comparativo clásico, más universal que particular (Hoenigswald 1990: 382), pese a sus limitaciones y a los ataques recientes de que ha sido objeto (cf. Morpurgo Davies 1988: 45-6), la correspondencia fonética, cuya importancia en la comparación es directamente proporcional al significado que para la lingüística general tienen dos postulados básicos, aunque de aceptación dispar (la arbitrariedad del signo lingüístico y la regularidad del cambio fonético,<sup>44</sup> cf. Anttila 1989: 255), puede no ser —y de hecho no es— enteramente suficiente para la demostración de parentescos lingüísticos (vid. Hymes 1993 [1959]: 29), pero su establecimiento resulta una condición necesaria, especialmente desde el punto de vista práctico, para poder deslindar las semejanzas entre formas de dos o más lenguas que atestiguan causa histórica (origen común o préstamo) y las que pueden sencillamente achacarse al efecto del azar, así como para identificar las condiciones en las que se produce la correspondencia y las causas de eventuales desviaciones con respecto a la regla general. Especialmente para el caso de las comparaciones múltiples de carácter léxico (pero en general para toda comparación) puede afirmarse que sin esta operación básica —y a falta de un mecanismo de cálculo que delimite con exactitud y fiabilidad el campo de la casualidad— toda hipótesis de relación genética, no importa si corresponde o no a la realidad, no pasa de ser una afirmación esencialmente gratuita que, por supuesto, no proporciona, pese a las evidencias materiales alternativas en que pretenda cimentarse, soporte suficiente para establecer parentesco lingüístico alguno.

## Bibliografía

- Aalto, P., 1988, “The Problem of “Nostratic””, *AION (Sez. Linguistica)* 10, 49-65.  
 Adelaar, W. F. H., Reseña de Greenberg 1987, *Lingua* 78, 2/3, 249-255.  
 Anttila, R., 1989, *Historical and Comparative Linguistics*, Amsterdam-Philadelphia: John Benjamins.  
 Baldi, Ph. (ed.), 1990, *Linguistic Change and Reconstruction Methodology*, Berlin-New York: Mouton de Gruyter.

(44) Acerca de la hipótesis de la regularidad, generalmente aceptada como premisa válida por los comparatistas, existen, no obstante, objeciones en ocasiones serias (basadas en datos procedentes no sólo de casos especiales como los “patrilectos” australianos estudiados por Johnson 1990 o las lenguas melanesias “aberrantes” de Grace 1990 y 1992, sino también de familias lingüísticas bien conocidas —cf. Lieberman 1990—), que obligan a matizar si no la validez general del método comparativo clásico, sí al menos su eficacia (en lo que se refiere al establecimiento de correspondencias regulares) para determinados casos. A pesar de ello, y en palabras del propio Lieberman (1990: 715): “The common assumption of “exceptionless” regularity... seems to be the only reasonable working hypothesis with which we can approach the study of sound change, even if is not justifiable in the final analysis, and we should always expect a residue of exceptions in practice”.

- Baxter, W. H. y A. Manaster Ramer, 1996, Reseña de Ringe 1992, *Diachronica* 13, 2, 371-384.
- Bender, M. L., 1969, "Chance CVC correspondences in unrelated languages", *Lg* 45, 3, 519-531.
- Bengtson, J. D., 1995, "Basque: An Orphan Forever? A Response to Trask", *Mother Tongue* 1, 84-95.
- , y M. Ruhlen, 1994, "Global Etymologies", in M. Ruhlen, *On the Origin of Languages*, Stanford: Stanford University Press, 277-336.
- Berman, H., 1992, "A comment on the Yurok and Kalapuya data in Greenberg's *Language in the Americas*", *IJAL* 58, 2, 230-233.
- Birnbaum, H., 1977, *Linguistic Reconstruction: Its Potentials and Limitations in New Perspective* (= *Journal of Indo-European Studies* Monograph 2), Washington: Institut for the Study of Man.
- , 1992, "Genetic and Typological Approaches to External Comparison of Languages", in Shevoroshkin (ed.), 10-33.
- Bomhard, A. R., 1984, *Toward Proto-Nostratic: a New Approach to the Comparison of Proto-Indo-European and Proto-Afro-Asiatic*, Amsterdam-Philadelphia: John Benjamins.
- , 1996, *Indo-European and the Nostratic Hypothesis*, Charleston: Signum.
- y J. C. Kerns, 1994, *The Nostratic Macrolanguage. A Study in Distant Linguistic Relationship*, Berlin-New York: Mouton de Gruyter.
- Campbell, L., 1988, Reseña de Greenberg 1987, *Lg* 64, 3, 591-615.
- , 1995, "The Quechumaran Hypothesis and Lessons for Distant Genetic Relationship", *Diachronica* 12, 2, 157-200.
- , e I. Goddard, 1990, "Summary report: American Indian languages and principles of language change", in Ph. Baldi (ed.), 17-32.
- Chafe, W. L., 1987, Reseña de Greenberg 1987, *Current Anthropology* 28, 5, 652-653.
- Chaker, S., 1990, "Comparatisme et reconstruction dans le domaine chamito-sémitique: problèmes de méthodes et de limites", in Ch. Touratier et alii (eds.), *Cercle Linguistique d'Aix-en-Provence. Travaux 8: Linguistique Comparée (Méthode et Résultats)*, Aix-en-Provence: Publications de l'Université de Provence, 161-186.
- Collinder, B., 1966, "Distant Linguistic Affinity", in H. Birnbaum y J. Puhvel (eds.), *Ancient Indo-European Dialects*, Berkeley-Los Angeles, 199-200.
- Comrie, B., 1989, *Universales del lenguaje y tipología lingüística*, Madrid: Gredos.
- , 1993, "Typology and reconstruction", in Ch. Jones (ed.), *Historical Linguistics: Problems and Perspectives*, London: Longman, 74-97.
- Dixon, R. M. W., 1997, *The rise and fall of languages*, Cambridge: Cambridge University Press.
- Doerfer, G., 1973, *Lautgesetz und Zufall. Betrachtungen zum Omnicomparatismus*, Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- , 1995, "The Recent Development of Nostratism", *Indogermanische Forschungen* 100, 252-267.
- Dolgopolsky, A. V., 1986 [1964], "A Probabilistic Hypothesis Concerning the Oldest Relationships Among the Language Families in Northern Eurasia", in V. Shevoroshkin y T. L. Markey (eds.), 27-50.
- Dunkel, G., 1981, "Typology versus Reconstruction", in Y. Arbeitman y A. R. Bomhard (eds.), *Bono Homini Donum. Essays in Memory of J. Alexander Kerns*, Amsterdam: John Benjamins, 559-569.
- Dyen, I., 1970, "Background 'Noise' or 'Evidence' in Comparative Linguistics: The Case of

- the Austronesian-Indo-European Hypothesis", in G. Cardona, H. M. Hoenigswald y A. Senn (eds.), *Indo-European and Indo-Europeans*, Philadelphia, 431-440.
- Echenique, M. T., 1987, *Historia lingüística vasco-románica*, Madrid: Paraninfo.
- Fairbanks, G. H., 1955, "A Note on Glottochronology", *IJAL* 21, 2, 116-120.
- Fodor, I., 1982 [1966], *A Fallacy of Contemporary Linguistics. J. H. Greenberg's Classification of the African Languages and his "Comparative Method"*, Hamburg: Helmut Buske.
- Fox, A., 1995, *Linguistic Reconstruction: An Introduction to Theory and Method*, Oxford: Oxford University Press.
- Gamkrelidze, T., 1981, "Language Typology and Language Universals and their Implications for the Reconstruction of the Indo-European Stop System", in Y. Arbeitman y A. R. Bomhard (eds.), *Bono Homini Donum. Essays in Memory of J. Alexander Kerns*, Amsterdam: John Benjamins, 571-609.
- Goddard, I., 1987, Reseña de Greenberg 1987, *Current Anthropology* 28, 5, 656-657.
- Grace, G. W., 1990, "The "aberrant" (vs. "exemplary") Melanesian languages", in Ph. Baldi (ed.), 155-174.
- , 1992, "How do languages change (more on "aberrant" languages)", *Oceanic Linguistics* 31, 1, 115-130.
- Greenberg, J. H., 1987, *Language in the Americas*, Stanford: Stanford University Press.
- , 1990 [1969], "Review of Fodor's *The Problems in the Classification of the African Languages*", in idem, *On Language: Selected Writings of J. H. Greenberg*, Stanford: Stanford University Press, 439-445.
- , 1990, "The American Indian Language Controversy", *Review of Archaeology* 11, 5-14.
- , 1995, "The Concept of Proof in Genetic Linguistics", *Mother Tongue* 1, 207-216.
- Gudschinsky, S. C., 1956, "Three Disturbing Questions Concerning Lexicostatistics", *IJAL* 22, 3, 212-213.
- Haarmann, H., 1990, "'Basic' Vocabulary and Language Contact: The Disillusion of Glottochronology", *Indogermanische Forschungen* 95, 1-37.
- Hamp, E. C., 1992, "On misusing similarity", in G. W. Davis y G. K. Iverson, *Explanation in Historical Linguistics*, Amsterdam-Philadelphia: John Benjamins, 95-103.
- Hardman de Bautista, M. J., 1978, "Jaqi: The Linguistic Family", *IJAL* 44, 2, 146-153.
- Harris, A. C., 1990, "Kartvelian Contacts with Indoeuropean", in T. L. Markey y J. A. C. Greppin (eds.), *When Worlds Collide: Indo-Europeans and Pre-Indo-Europeans*, Ann Arbor: Karoma, 67-100.
- Hegedüs, I., 1989, "The applicability of exact methods in Nostratic research (preliminary report)", in V. Shevoroshkin (ed.), *Explorations in Language Macrofamilies*, Bochum: Brockmeyer, 30-39.
- , 1992, "On Nostratic", in Shevoroshkin (ed.), 106-110.
- Helimsky, E. A., 1987, "A 'New Approach' to Nostratic Comparison" (reseña de Bomhard 1984), *JAOS* 107, 1, 97-100.
- Hock, H. H., 1986, *Principles of Historical Linguistics*, Berlin: Mouton de Gruyter.
- , 1992, "Reconstruction and Syntactic Typology", in G. W. Davis y G. K. Iverson (eds.), *Explanation in Historical Linguistics*, Amsterdam-Philadelphia: John Benjamins.
- , y B. Joseph, 1996, *Language History, Language Change and Language Relationship*, Berlin-New York: Mouton de Gruyter.
- Hoenigswald, H. M., 1990, "Is the "comparative" method general or family-specific?", in Ph. Baldi (ed.), 375-383.

- Hymes, D. H., 1993 [1959], "Genetic Classification: Retrospect and Prospect", *Anthropological Linguistics* 35, 1, 21-37.
- Illich-Svitych, V. M., 1967, "Materjaly k sravnitel'nomu slovarju nostraticheskix jazykov (indoevropskij, altajskij, ural'skij, dravidijskij, kartvel'skij, semitoxamitskij)", *Ètimologia* 1965, 321-73.
- , 1971, 1976, 1984, *Opyt sravnenia nostraticheskix jazykov*, Moskva: Nauka.
- Ivanov, V. V., 1986 [1980], "Proto-Languages as Objects of Scientific Description", in Shevoroshkin y Markey (eds.), 1-26.
- Jakobson, R., 1971 [1958], "Typological studies and their contribution to historical and comparative linguistics", *Selected Writings*. I. The Hague-Paris: Mouton, 523-532.
- Jones, A. I., 1989, "Australian and the Mana Languages", *Oceanic Linguistics* 28, 1, 181-196.
- Johnson, S., 1990, "Social parameters of linguistic change in an unstratified Aboriginal society", in Ph. Baldi (ed.), 419-434.
- Kimball, G., 1992, "A critique of Muskogean, "Gulf" and Yukian material in *Language in the Americas*", *IJAL* 58, 4, 447-501.
- Klimov, G. A., 1991a, "Some thoughts on Indo-European-Kartvelian relations", *JIES* 19, 3/4, 325-341.
- , 1991b, "On Two Kinds of Reconstruction in Comparative Studies", in W. P. Lehmann y H.-J. Jakusz Hewitt (eds.), *Language Typology 1988. Typological Models in Reconstruction*, Amsterdam-Philadelphia: John Benjamins, 11-15.
- , 1994a, *Drevnejšie indoevropézmy kartvel'skix jazykov*, Moskva: Rossijskaja Akademija Nauk.
- , 1994b, "L'analogie kartvélienne de l'IE \*okētō(u)-", in R. Bielman y R. Stempel (eds.), *Indogermanica et Caucasia (Festschrift für Karl Horst Schmidt zum 65. Geburtstag)*, Berlin-New York: Mouton de Gruyter, 472-478.
- Lieberman, S. J., 1990, "The regularity of sound change: A semitistic perspective", in Ph. Baldi (ed.), 697-721.
- Manaster Ramer, A., 1995, "On 'Some thoughts on Indo-European-Kartvelian relations'", *JIES* 23, 1/2, 195-208.
- , 1997, "Nostratic from a Typological Point of View", *JIES* 25, 1/2, 79-105.
- , y C. Hitchcock, 1996, "Glass Houses: Greenberg, Ringe, and the Mathematics of Comparative Linguistics", *Anthropological Linguistics* 38, 4, 601-619.
- Manessy, G., 1990, "Du bon usage de la méthode comparative historique dans les langues africaines et ailleurs", in C. Touratier et alii (eds.), *Cercle Linguistique d'Aix-en-Provence. Travaux 8: Linguistique Comparée (Méthode et Résultats)*, Aix-en-Provence: Publications de l'Université de Provence, 89-107.
- Matisoff, J. A., 1990, "On megalocomparison", *Lg* 66, 1, 106-120.
- Meillet, A., 1926 [1914], "Le problème de la parenté des langues", in *Linguistique historique et linguistique générale*, 2ème édition, Paris: Champion, 76-101.
- , 1926 [1918], "Convergence de développements linguistiques", in *Linguistique historique et linguistique générale*, 2ème édition, Paris: Champion, 61-75.
- , 1970 [1925], *La méthode comparative en linguistique historique*, Paris: Champion.
- Michelena, L., 1988a [1950], "De etimología vasca", in idem, *Sobre historia de la lengua vasca*. I, San Sebastián: Anejos de ASJU n° 10, 439-444.
- , 1988b [1964], "Sobre el pasado de la lengua vasca", in idem, *Sobre historia de la lengua vasca*. I, San Sebastián: Anejos de ASJU n° 10, 1-73.
- , 1995 [1966], "Problemas generales de la reconstrucción y la posición del grupo de lenguas indoeuropeas del Asia Menor", *FLV* 69, 205-242.

- Morpurgo Davies, A., 1988, "Il metodo comparativo, passato e presente", *AION (Sez. Linguistica)* 10, 27-48.
- Milewski, T., 1960, "Similarities between the Asiatic and American Indian languages", *IJAL* 26, 4, 265-274.
- Newman, P., 1993, "Greenberg's American Indian classification: a report on the controversy", in J. van Marle (ed.), *Historical Linguistics 1991. Papers from the 10th International Conference on Historical Linguistics*, Amsterdam-Philadelphia: John Benjamins, 229-242.
- Palmaitis, M., 1977, "Parent Language — Genetic or Contact Relationship", *Indogermanische Forschungen* 82, 50-60.
- , 1986, "New Contributions to "Proto-Nostratic"", *Indogermanische Forschungen* 91, 305-317.
- Pisani, V., 1952, "Parenté linguistique", *Lingua* 3, 3-16.
- Pokorny, J., 1959, *Indogermanisches etymologisches Wörterbuch (I-II)*, Bern: Francke.
- Poser, W. J., 1992, "The Salinan and Yurumanguí data in *Language in the Americas*", *IJAL* 58, 2, 202-229.
- Rankin, R. L., 1992, *Reseña de Greenberg 1987*, *IJAL* 58, 3, 324-351.
- Ringe, D. A., 1992, *On Calculating the Factor of Chance in Language Comparison (=Transactions of the American Philosophical Society* 82, 1, 1-110), Philadelphia: American Philosophical Society.
- , 1995a, "'Nostratic' and the Factor of Chance", *Diachronica* 12/1, 55-74.
- , 1995b, "The "Mana" Languages and the Three-Language Problem", *Oceanic Linguistics* 34, 1, 99-122.
- , 1996, "The Mathematics of 'Amerind'", *Diachronica* 13, 1, 135-154.
- Ruhlen, M., 1991 [1987], *A Guide to the World's Languages. Volume I: Classification*. Stanford: Stanford University Press.
- Salmons, J., 1992, "A Look at the Data for a Global Etymology: \**rik* 'finger'", in G. W. Davis y G. K. Iverson (eds.), *Explanation in Historical Linguistics*, Amsterdam-Philadelphia: John Benjamins, 207-228.
- Schwink, F. W., 1994, *Linguistic Typology, Universality and the Realism of Reconstruction (=Journal of Indo-European Studies Monograph 12)*, Washington: Institute for the Study of Man.
- Serebrennikov, B. A., 1986 [1982], "On the so-called 'Nostratic' languages", in V. Shevoroshkin y T. L. Markey (eds.), 66-86.
- Shevoroshkin, V. (ed.), 1992, *Nostratic, Dene-Caucasian, Austric and Amerind*, Bochum: Brockmeyer.
- , y T. L. Markey (eds.), 1986, *Typology, Relationship and Time*, Ann Arbor: Karoma.
- Shjerbak, A. M., 1994, "Nostraticheskije issledovania s pozitsii tiurkologa", in *idem*, *Vvedenie v sravnitel'noe izuchenie tiurkskix jazykov*, Sankt-Peterburg: Nauka, 150-163.
- Starostin, S., 1992, "Methodology of Long-Range Comparison", in Shevoroshkin (ed.), 75-79.
- Stevens, C. M., 1992, "The Use and Abuse of Typology in Comparative Linguistics: An Update on the Controversy", *JIES* 20, 1/2, 45-58.
- Teeter, K. V., 1963, "Lexicostatistics and Genetic Relationship", *Lg* 39, 4, 638-648.
- Timberlake, A., 1990, "Typology and phonological history", in W. P. Lehmann (ed.), *Language Typology 1987: Systematic Balance in Language*, Amsterdam-Philadelphia: John Benjamins, 35-56.
- Trask, R. L., 1995a, "Basque and Dene-Caucasian: A Critique from the Basque Side", *Mother Tongue. Journal of the Association for the Study of Language in Prehistory* 1, 3-82.

- , 1995b, "Response to Comments", *Mother Tongue* 1, 172-201.
- , 1996, *Historical Linguistics*, London-New York-Sydney-Auckland: Arnold.
- , 1997, *The History of Basque*, London-New York: Routledge.
- Vennemann, Th., 1986, *Neuere Entwicklungen in der Phonologie*, Berlin-New York-Amsterdam: Mouton de Gruyter.
- Vine, B., 1991, "Indo-European and Nostratic", *Indogermanische Forschungen* 96, 9-35.
- Vogt, H., 1988 [1971], "Indoeuropejskie [faususement pour: Kavkazskie] jazyki i sravnitel'nye metody", in *Linguistique caucasienne et arménienne (=Studia Caucasologica II)*, Oslo: Norwegian University Press, 500-505.
- Vovin, A., 1994a, "Is Japanese related to Austronesian", *Oceanic Linguistics* 33, 2, 369-390.
- , 1994b, "Long Distance Relationships, Reconstruction Methodology, and the Origins of Japanese", *Diachronica* 11, 1, 95-114.
- Watkins, C., 1990, "Etymologies, equations, and comparanda: types and values, and criteria for judgment", in Ph. Baldi (ed.), 289-304.
- Winter, W., 1973, "Areal Linguistics: Some General Considerations", in T. A. Sebeok (ed.), *Current Trends in Linguistics. 11. Diachronic, Areal and Typological Linguistics*, The Hague: Mouton, 135-147.
- , 1990, "Linguistic reconstruction: The scope of historical and comparative linguistics", in E. C. Polomé (ed.), *Research Guide on Language Change*, Berlin-New York: Mouton de Gruyter, 11-21.

**Apéndice 1. Tabla tentativa de correspondencias "globales" para \*k  
(a partir de Bengtson y Ruhlen 1994)**

Étimo "global"	<i>kama</i> 'hold'	<i>kano</i> 'arm'	<i>kati</i> 'bone'	<i>kuan</i> 'dog'	<i>kuna</i> 'woman'
Khoisan	x	x	?	!g, x	-
Nig-Congo	k	k	-	-	-
Afr-Asia	k	?h, k, g	q	k	k
IE	g	k	k	k	g <sup>w</sup>
Ural.	k	k, h	?k	k	-
Dravid.	k	k, g	x, q	-	-k'
Caucas.	k'	gH	k	xH	q <sup>(w)</sup>
Amerind	-	k, g, x	k, k', g, q, c	k, '	kw, g, q, h,
Miao-Yao	ng	-	-	-	-
Austrones.	k	-	-	-	ki
Daic	hn	x	-	?nk	-
Indo-Pacif.	-	g, k	-	-k	?ch, qu, k
Sino-Tibet	-	k	k	qh	-

Intentar trazar posibles correspondencias entre algunos de los elementos que conforman este cuadro —más parecido a un catálogo desordenado de las posibilidades velares y postvelares del aparato fonador— no es ciertamente tarea fácil (y menos aún tratar de dar explicación, habida cuenta de la escasez de datos con que se opera, a las alternancias múltiples que pueden observarse). Las casillas vacías de algunos *phyla* para determinados étimos no hacen sino complicar aún más, si cabe, la situación. Ante panoramas de estas características no resulta excesivamente sorprendente que los practicantes de la *long-range comparison* acaben por denostar el propio concepto (e

instrumento) de correspondencia fonética y proclamen la semejanza superficial, no sometida a restricción metodológica alguna, como único criterio para la comparación. Lo que sí es algo más sorprendente es la aparente facilidad con que se resuelve el problema de la elección del arquetipo fonológico que se adjudica a cada protoforma. Pese a las diferencias que presentan entre sí los segmentos y las relaciones entre ellos, la consonante escogida en los cinco étimos es invariablemente simbolizada mediante el signo *k* (¿hará éste referencia, tal vez, a la naturaleza menos marcada —y por tanto más “reconstruible”— de la oclusiva velar sorda frente a todos los demás segmentos que comparecen en el esquema?).

## Apéndice 2. Para una estimación de la probabilidad de similitudes casuales de forma y sentido entre dos o más lenguas

Atendiendo al esquema de cálculo de las coincidencias accidentales entre formas de dos o más lenguas que aplican Bengtson y Ruhlen (1994) a sistemas fonológicos hipotéticos, es posible ensayar un procedimiento de cálculo de esas coincidencias (que no son estrictas, sino que responden más bien a vagos criterios de similitud formal y semántica) más acorde a la práctica real de los multicomparatistas.

Para estructuras del tipo CVC, tal vez las más frecuentes, dividimos las posibilidades consonánticas en tres grupos (u órdenes locales), a lo que parece autorizar la escasa atención que se presta en este tipo de ecuaciones etimológicas multilaterales al detalle fonético (cf. Adelaar 1989: 253): P es el representante común para todo fonema o sonido labial (p, b, f, m, p', ph, etc...), T engloba a toda consonante dental (t, d, t', th, ð, θ, n) y K a toda consonante velar o incluso postvelar (k, g, x, γ, k', kh, h,'). Las combinaciones máximas posibles de estos tres elementos en dos posiciones —para la vocal V es válida cualquier variación— son exactamente 9 (PVT, TVP, PVK, KVP, TVK, KVT, PVP, TVT, KVK). Abastecer materialmente todo un conjunto de significados, no sujetos a cálculo, a partir de este número de raíces ha de implicar la frecuente coincidencia formal y semántica, basada en la semejanza, de unidades léxicas de distintas lenguas, por lo que una eventual probabilidad del 11% (1/9) para las coincidencias entre dos lenguas, aun siendo bastante alta, no respondería a la realidad (no son nueve las unidades de significación). Este modelo resulta adecuado a gran parte de las etimologías globales, en las que no es difícil hallar juntas formas del tipo *may*, *pek*, *fu'*, *bōx*, entre las que, por lo demás, no pueden detectarse correspondencias regulares. Aunque se introduzcan dos variables más en el modelo (las sonantes *r*, *l*), complicando un tanto el sistema de combinaciones, éstas no irían más allá de las 25, y la posibilidad de convergencias accidentales —dada la vaguedad de las semejanzas tanto en la forma como en el significado— seguiría siendo excesivamente alta como para confiar mínimamente en esta clase de aproximaciones léxicas.

