

词形结构进化与世界语言的多样性

江荻^{①②*}, 康才峻^①, 燕海雄^{①②}

① 上海师范大学语言研究所, 上海 200234;

② 中国社会科学院语言学与计算语言学重点实验室, 北京 100081

* 联系人, E-mail: jiangdi@cass.org.cn

2013-12-19 收稿, 2014-03-30 接受

国家自然科学基金(31271337)和国家社会科学基金(12&ZD174)资助

摘要 世界语言的基本差异源于词形结构(WSS), 而不是区别性的音素. 语言的多样性似乎来自词形结构的演化结果, 取决于音节构造形态类型和词的音节长度. 本文采集世界 179 种语言的 Swadesh 词表来分析词形结构多样性指数的地理分布, 并尝试发现它们在演化过程中的进化位置和深度. 通过建立理想词形结构偏移量模型, 计算每种语言的偏移距离和偏移方向, 根据分析结果把语言划分成 3 组, 每一组代表一种演化类型. 得出词形结构的多样性和偏移量模型代表了人类语言演化的多样性方向和演化过程, 当今每种语言都在一定程度上保留了词形结构最基本的特征, 因此, 词形结构可以视为人类语言“基因”.

关键词

词形结构
语言基因
语言多样性
偏移量模型
地理分布

现代人走出非洲^[1-3], 人类语言也随之扩散到世界各地. 经历数万年, 不同地区的人群形成各自的文明, 也形成了多样性的语言. 语言的多样性反映了演化的复杂性, 但犹如生物自简单而复杂, 语言亦非生而复杂. 历史语言学证明, 几乎所有复杂的词法形态和句法现象都是逐步产生的^[4-6]. 但早期口语一定包含了人类语言作为区别于其他生物交流方式的符号系统所拥有的基本结构, 并清晰地外显. 这种结构经历数万年多样性演化也不会丧失, 而会在现代语言中传承下来, 这就是我们称为语言“基因”的词形结构或词音节结构(word-syllable structure). 词形结构的多样性或差异性决定了语言的类型和面貌, 是观察现代语言分类和追溯语言历史渊源的最重要要素, 也是追踪人类族群迁徙和地理分布的参照线索.

1 什么是词形结构

词形结构是词义的语音形式和形式的结构, 由音节和组成音节的音素构成. 它不是一种元音或辅音这样的个体音素形式, 而是一种结构体. 结构体

的表达是把词的具体的音素(如 a, b, m)抽象为音类 C, V (C 表示辅音, V 表示元音)并以音类符号的排列结构来表达, 例如英语词 sesame [sesəmi] 由 6 个音素构成 3 个音节组织而成, 即 CV·CV·CV. 本文约定, 词形结构有时候可以简称结构, 例如 CV 结构, 形态则指音类符号排序的各种形式. 关于音节, 近年来学者们有一项极为重要的发现, 他们认为 CV 结构是人类语言偏好形式或基本形式^[7-12]. 不过, 尽管这个结论来自心理实验、行为实验、语言习得、生成音系学等多个领域, 从真实词形中抽取的单个 CV 形式却没有语言单位价值, 唯有整体词形才符合词的音义结合的本质. 由此, 我们推论人类语言词形结构的初始形态很可能是{CV}_p (‘p’即 poly-, 表示多音节), 其中当然也包含了单一 CV 词的结构.

假定我们接受人类偏好的 CV 是词形结构的基本构成要素, 也接受{CV}_p 是早期共同语言的词形结构形态, 就可以承认以下推理的合理性: 在语言的演化过程中, 元音或辅音的演化脱落造成了辅音相连或元音连缀的相对复杂的词形结构的出现. 例如英

引用格式: 江荻, 康才峻, 燕海雄. 词形结构进化与世界语言的多样性. 科学通报, 2014, 59: 2084-2090

英文版见: Jiang D, Kang C J, Yan H X. Evolution of word-syllable structures and the diversity of world languages. Chin Sci Bull, 2014, 59, doi: 10.1007/s11434-014-0414-3

语 pluck [plʌk]的词形结构是 CCVC, 有辅音丛和辅音韵尾, 但仅一音节长, 这个单音节词来源于古英语多音节的 plocian, 中古低地德语 plucken, 再比较拉丁语族罗曼斯语构拟 *piluccare^[13] (*表示词形构拟), 可以判断这个词历史上曾发生多处元音脱落。日语词 kudamono(水果)结构单纯, 但是 CV·CV·CV·CV 有四个音节长, 词形结构很接近预设的初始形态。汉语是声调语言, ma⁵⁵(母亲), ma³⁵(苕麻), ma²¹⁴(马), ma⁵¹(骂), 虽然都是 3 个音素(含 1 个声调)和 1 个音节组成, 具有相同的简单结构 CV^(Tone), 但同一个音节声调不同, 有多个意思。汉语乃至东亚和东南亚大多数语言一直被认为是世界语言的特例, 词形长度已达到最简地步, 但正是这一点证明 CV 是词形结构的最基础要素。

人类早期在世界的活动主要呈现为迁徙模式, 随着人口增长逐步扩散到世界各地。如果现代人曾拥有相同的初始语言^[5], 词形结构也应具有一致的初始形态, 正是数万年的迁徙扩散、人群的隔绝和语言的随机演化导致了词形结构的地理区域差异, 造成词形结构在世界范围内呈现出高度复杂性和多样性。

2 词形结构进化的历史深度

Atkinson^[14]曾以音素作为人类语言基因要素探索人类起源于非洲的命题, 这种思想有着深刻的学术背景^[15]。全球人类基因组创始人 Cavalli-Sforza 等人^[16]认为语言进化与基因进化具有平行性, 现存的大约 5000 种语言与今天的民族和土著部落相当吻合。甚至进化论奠基人达尔文也提出“假如我们拥有一个完善的人类系谱, 则人种排列成的系谱将能提供现在整个世界上所说的各种语言的最好分类”^[17]。不过, Atkinson^[14]采用的音素和音素数量并不是语言中稳定的参项, 语言不断发生音素变化, 从一种音素变为另一种音素, 例如 b→β, k→x, 也经常失去原有音素, 或产生新的音素, 音素和音素数量在一代代人之间变化传承, 累积起来足以使一个语言的音素系统面貌全非。历史语言学家 Campbell^[18]指出: 由于跨越如此漫长的年代, 发生如此巨大的变化, 原初语言没有任何东西以任何形式在现代语言里残留下来。Campbell 的观点有点片面, 个体音素的确不停地变化, 词形结构也会变化, 但任何语言都保持着动态稳定的音系、词汇^[19]和基本词形结构, 语言交流功能不会中断。迄今, 语言描写和类型研究告诉我们, 任何

语言都没有所有词形变为 {C}_p 或者 {V}_p 这类结构, 也没有所有词都以 V 起始的语言, 没有所有词都以 C 结束的语言^[20], {CV}_p 结构仍然是人类语言的基本形式。

上文推论人类语言词形结构具多音节性, 其中蕴含了语言起源的单源性假说。这个假说基于语音变化的原理和规则: 所有语言普遍发生语音弱化和脱落, 反之, 增音(音素)则是随机的、临场的和语流音变性的。换句话说, 语言伊始呈现多音节词面貌, 经过脱落, 词长逐渐缩短, 形态复杂度逐渐增加。这个观点跟 20 世纪早期 Jespersen^[21]的观点一致, 他认为语言是从原始多音节词朝单音节词发展。Jackendoff^[5]对人属(包括尼安德特人)的发声音姿单位提出一个有趣的描述: 每个音节都是一个整体发声音姿, 用 10 个这样的音姿, 人们就能建立 100 个双音节原始音节发声串和 1000 个三音节的原始音节发声串, 这种方法很好地实现了开音节结尾。Jackendoff 所说“整体发声音姿”即音节, 他所描述的开音节表明音节形式就是 CV, 然后可以组成双音节词 CVCV 或更多音节的词。

按照音素脱落的模式, 词形结构在漫长的数万年演化过程中产生了复杂的多维度演变, 导致世界语言演化出不同词形形态结果。我们采用递减算法来模拟音素脱落, 以观察语言的演化路径和深度。设定初始词形长度为 3 音节, 演化过程中每次脱落一个音素, 如图 1 所示演化模型。

从演化过程看, 图 1(a)中并非每种结果都实际出现过, 语言凭借自身的自组织系统会选择恰当的路径, 因此, 在某些节点必定存在实际出现的词形结构。深度上, 词形至少经历两次元音脱落才可能转变为单音节词: →CVCVC→CVCC; →CVCCV→CVCC; →CVCCV→CCC; →CCVCV→CCCV, 余则经历更多层次。元音和辅音脱落都带来复杂的词形形态, 例如 CCCV(spray), VCVVC(oriole)等, 因此单个音节可能呈现 CV, V, VC 3 种形态, n 个音节可能产生 3^n-1 种词形结构(至少保留一个元音)。图 1(b)表示词形结构演化过程中可能会出现相同的形态, 但是, 由于音素值的差别, 例如 CV 可能是[ta]或[si]等, 语言中产生的同形词相对还是有限的。

在东亚和东南亚区域, 有上百种语言已发展为单音节词语言, 它们很可能都是这个模型的原型。至于导致语言演变的诸多其他因素, 例如随机变化、地理分布、人口规模、文化接触、区域趋同等, 本文结语有所提及。

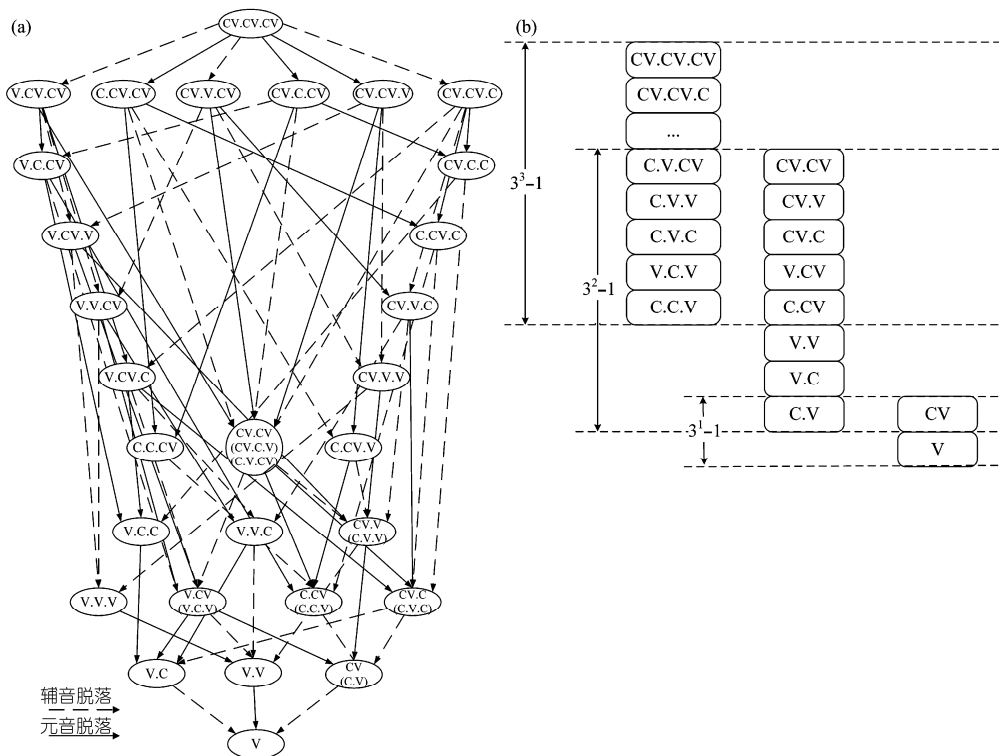


图 1 三音节词形结构的演化过程

3 词形结构多样性的地理分布

选取世界 8 个区域 179 种语言的基本词汇(Swadesh list)数据为样本^[22], 尝试使用 Shannon-Wiener 多样性计算模型对词形结构的类型和数量进行分析^[23,24]

$$H = -\sum P_i \ln P_i. \quad (1)$$

式中 H 代表词形结构的多样性, P_i 表示具有第 i 种词形结构的词汇数占总词汇数的比例.

根据计算结果, 我们编制了表 S1 的词形结构数据, 按语言的分布区域经纬度绘制地理分布图(图 2). 该图以渐进颜色表示词形结构的多样性变化, 深蓝色表示多样性较低一端, 深红色表示多样性丰富一端. 然后按照 8 个语言地理区域(分别是: 东亚, 东南亚, 南亚, 东北亚, 欧洲、西亚和中亚, 太平洋区域, 美洲和非洲)数据给出词形结构多样性的箱体图(图 3).

词形结构的多样性是由演化创新造成的. 最重要的指标是词长变异和形态变异, CVCV 跟 CV, CVCVCV 词长不同, 也跟 CVC, CCV, CVV 形态不同. 图 2 显示出欧洲、西亚等地多样性最丰富, 这主要是该区域语言词的形态变异造成的. 东亚和东南亚多样性最低主要是词长变异造成的, 反映了它们的孤

立语性质^[25]. 非洲、美洲、太平洋区域语言很可能在词长和结构上保留较多早期形式. 显然, 现代语言的多样性聚类跟地理区域分布呈相关关系, 图 3 展现了这样的聚类关系.

4 词形结构的偏移度量

词形结构多样性指数涵盖了一种语言的词形结构的总体丰富度与均匀度信息, 但随机演化结果也可能造成低丰富度和高均匀度系统与高丰富度和低均匀度系统具有相同的多样性指数^[26], 这种情况在图 3 显示为多个区域多样性指数不彰. 因此, 还要对多样性产生的根源, 即词形结构本身进行分析. 那么, 词形结构本身又隐藏着什么样的特征信息呢? 这些特征信息能否确定现代语言处在语言演变过程中的什么位置呢? 上文已经提出词形结构的初始形态为 $\{CV\}_p$, 如果把现代语言中的复杂词形结构看作对词形结构初始形态的偏移, 那么分析现代语言的偏移情况即可度量现代语言的演化程度.

假设词的元音个数为 n_v , 辅音个数为 n_c , 音节长度为 n_s . 根据初始结构的特征, 其元音数与辅音数应等于音节长度 n_s . 以元音数与辅音数构建二维坐标

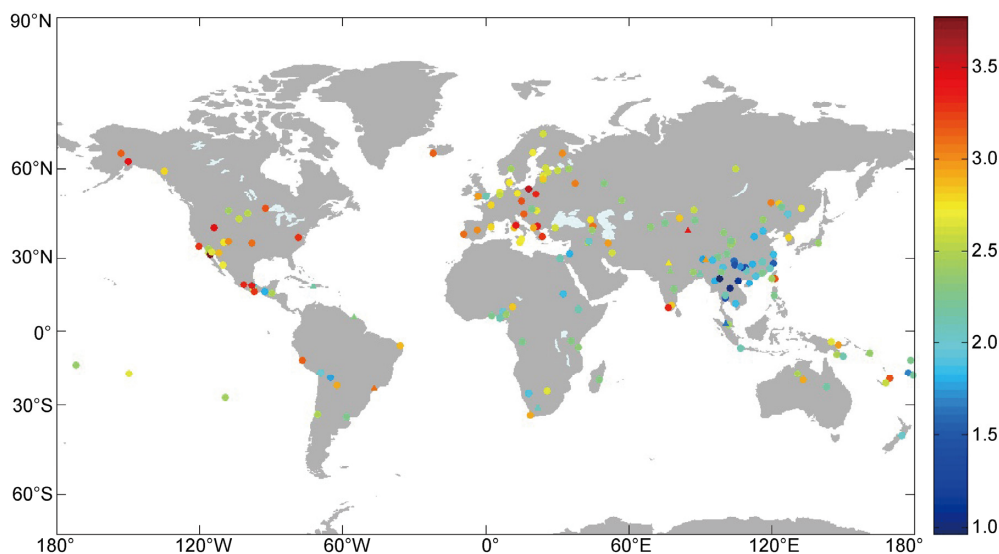


图2 世界语言词形结构多样性分布图

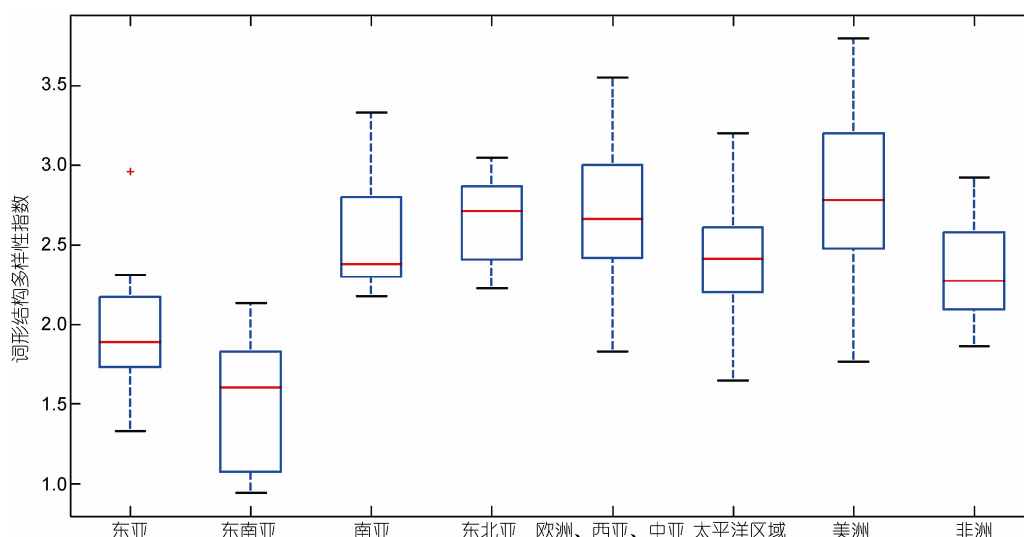


图3 世界语言词形结构多样性箱体图
+古藏语

系,将现有词形与其初始结构分别表示为该坐标系第一象限中的两个向量 $m(n_v, n_c)$ 与 $o(n_s, n_s)$,如图4(a)所示.若这两个向量之差为向量 x ,我们定义 x 的模(即长度)为该词形结构的偏移距离,元音与辅音的脱落数量越多,偏移距离越大.同时,我们将两个向量的夹角 α 定义为该词形结构的偏移角,偏移角为正则词形结构向元音丛倾斜(辅音脱落),反之则向辅音丛倾斜(元音脱落).如果词内辅音和元音脱落数相同,统计上部分词形结构仅偏移距离改变,不发生偏移角变化,以三音节词为例,可能有6种变化,其中仅两种有偏移距离变化(删除线表示脱落): $CV.CV.CV$

$\rightarrow CV.C.V$, $CV.CV.CV \rightarrow C.V.CV$,其余多数偏移角和偏移距离都有变化,例如: $CV.CV.CV \rightarrow C.CV.V$, $CV.CV.CV \rightarrow V.C.CV$.

通过几何计算,可以获得适用的偏移距离计算公式:

$$|x| = \sqrt{(n_s - n_v)^2 + (n_s - n_c)^2}. \quad (2)$$

偏移角的计算公式为

$$\alpha = \frac{\pi}{4} - \arctg \frac{n_c}{n_v}. \quad (3)$$

为直观地分析词形结构的偏移情况,我们将单

个词形结构的偏移数据映射到极坐标系中的一个点上,其极径与极角分别代表偏移距离与偏移角.那么一个语言的偏移情况也就可以通过所有词形结构的偏移距离与偏移角来度量了.本文对 179 种语言词表的词形结构进行计算,得出各种语言的词形结构偏移距离及偏移角数据,以此为基础绘制了世界语言的偏移情况坐标分布图,如图 4(b)所示.

从数据看,全部语言的偏移角在 $[-17.8, 12]$ 弧度范围以内,其中 Chewong 语偏移角 -17.8 弧度,表示该语言的辅音丛最突出.全部语言偏移距离范围是 $[0.148, 2.678]$,其中 Western fijian 最小(0.148),表明最接近 $\{CV\}_p$ 结构,而书面藏语为 2.678,辅音丛尤为丰富.就全局来看,有如下两个显著的特点:(1)偏移角为负的语言要远远多于偏移角为正的语言,说明大多数语言的词形结构呈现辅音丛形态;(2)偏移角为正的语言其偏移距离普遍较短,这说明即使是元音丛形态为主的语言,其元音丛形态相对简单.

对比而言,辅音丛形态为主的语言偏移距离从短到长分布比较均匀,说明辅音丛具有从简单到复杂的多重形态.

具体而言,以各项中值数据看(表 1),欧洲、中亚、西亚和东北亚语言(图 4(b)中圆圈)的偏移距离高于其他区域语言,并且严重朝辅音丛方向偏移,加之词形较长,说明这些语言既有较长的词形,也有丰富的辅音丛,故而归为一个类型.美洲、非洲、太平洋区域语言(图 4(b)中星号)的偏移角很小,词长最大,说明整体偏移不大.实际上,美洲、太平洋区域有部分语言偏移角呈正值,即辅音脱落较多,形成元音丛,这样的现象在世界其他语言不是主流变化.东亚和东南亚语言(图 4(b)中三角形)最为独特,目前采集的 Swadesh 词汇绝大多数是单音节词形式,少量是双音节派生词和复合词,由于历史上可能的 CV 连续脱落,无法从长度追溯初始形态.而且,东亚和东南亚语言词形形态从简单到复杂形成多个等级,藏羌语言辅

表 1 世界语言词形结构形态分类

语言区域	典型词形结构	偏移角(弧度)	偏移距离	词长度(音节数)	词形结构形态特征
欧洲、西亚、中亚、东北亚	CCVCVC	-0.1445	1.0799	1.8792	辅音丛和多音节
美洲、太平洋区域、非洲	CVCVCV	-0.0445	0.8001	2.2994	单辅音和多音节
东亚、东南亚	(C)CV(C) ^(T)	-0.1168	0.6997	1.3596	辅音丛和单音节
汉语	CV ^(T)	-0.0707	0.5653	1.3350	单辅音和单音节

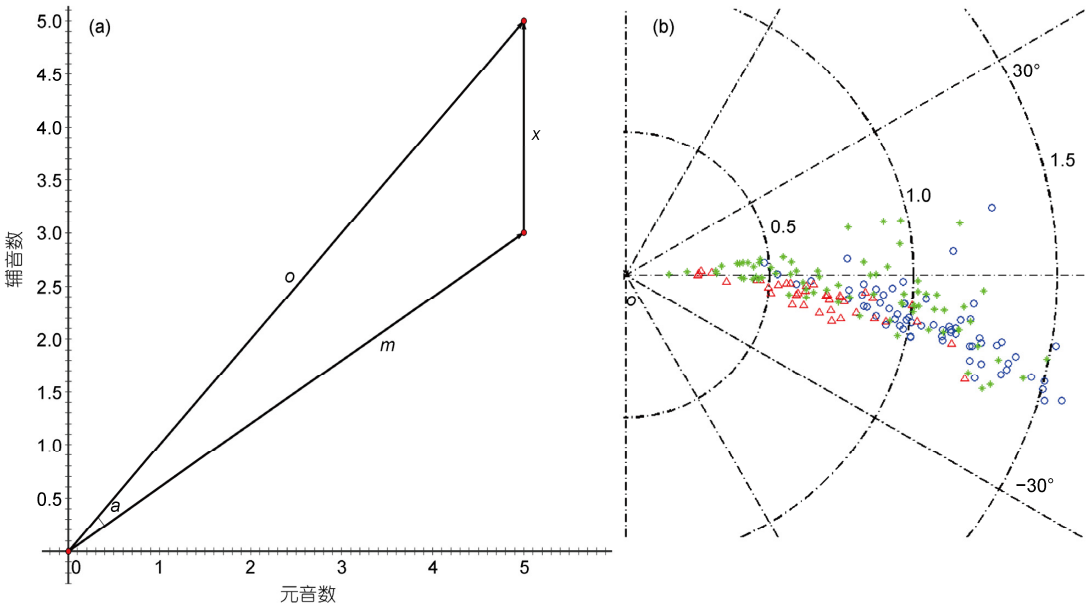


图 4 (网络版彩色)偏移距离和偏移角几何示意图(a)和偏移距离和偏移角极坐标分布图(b)

圆圈表示欧洲、西亚、中亚、东北亚;三角形表示东亚、东南亚;星号表示美洲、非洲、太平洋区域

音丛丰富, 缅彝、景颇、侗台、苗瑶语次之, 汉语最简单, 它的偏移角和偏移距离数据只能从单音节词的辅音丛计算, 偏移距离明显较小。作为对比, 我们单独列出了汉语的数据和形态特征。

据此, 可以根据词形结构的形态将世界语言划分为 3 种类型, 欧洲、西亚、中亚、东北亚语言是辅音丛和多音节并重形态, 美洲、非洲、太平洋区域语言是简单辅音和多音节形态, 东亚和东南亚语言是单音节词以及单音节词内单一辅音与辅音丛交替的形态。南亚语言是多音节词语言向单音节词语言的过渡形态^[27,28], 暂未列入统计数据 and 归类。

5 结语

世界语言词形结构的多样性和偏移量模型代表了人类语言演化的多样性方向和演化过程。数万年来, 随机演化因素造成的元辅音脱落突变并产生新的词形结构形态, 并为不同族群传承下来, 产生了语言地理分布上的多样性。如果说现代人走出非洲至今是人的大历史, 近 1 万年来, 或者新石器中晚期以来只算得小历史。可是这个小历史以数倍于前的演化速率改变了人类, 其中人口规模、人群迁徙、文化

接触、区域趋同、技术发展, 无一不对人的进化和语言的变化施加社会性选择的影响^[14,29,30](与自然选择对应)。我们采用的现代语言数据蕴含了这段小历史作用的结果, 例如, 汉语因为社会状态开放、人口数量较多、外部接触程度高, 在抑扬格词模式演化中转变为单音节词语言^[28,31]。日语是一种倾向词长优势的语言, 虽然地理位置在东亚, 多样性和偏移量等数据更接近美洲、非洲和太平洋区域语言, 这与日语历史上所处封闭的岛屿文明环境必然相关。Chewong 语是西太平洋区域马来西亚的一种南亚的孟高棉语系语言(Austronesian languages)^[32], 因其同时拥有多种词形结构: 单音节词, 一个半音节词; 双音节词, 复辅音声母词, 导致很高的复杂形态。本文的词形结构数据和阐释模型意在解释各语言在人类语言进化过程中位置, 在进一步的微观研究中, 以上人口、地理等数据将逐步加入讨论。

迄今为止, 词形结构的形态演化最复杂的英语和长度演化最简短的汉语都在不同程度上保存着人类语言古老的词形结构面貌。由此可以说, 词形结构是人类语言的“基因”, 是语言演化中保存至今的类似考古学化石型的遗存之一。

参考文献

- 1 Cavalli-Sforza L L, Piazza A, Menozzi P, et al. Reconstruction of human evolution: Bringing together genetic, archeological and linguistic data. *Proc Natl Acad Sci USA*, 1988, 85: 6002-6006
- 2 Leakey R. *The Origin of Humankind*. New York: Basic Books, 1996
- 3 Jin L, Su B. Natives or immigrants: Origin and migrations of modern humans in East Asia. *Nat Rev Genet*, 2000, 1: 126-133
- 4 Bickerton D. *Language and Species*. Chicago: University of Chicago Press, 1999
- 5 Jackendoff R. Possible stages in the evolution of the language capacity. *Trends Cogn Sci*, 1999, 3: 272-279
- 6 Jiang D. The evolutionary consequences of monosyllabic-word languages. *Comm Contemp Anthropol*, 2011, 5: 255-262
- 7 Lowenstamm J. CV as the only syllable type. In: Durand J, Laks B, eds. *Current Trends in Phonology: Models and Methods*. Salford (UK): European Studies Research Institute, University of Salford, 1996. 419-442
- 8 Berces K. What's wrong with vowel-initial syllables? *SOAS Working Papers in Linguistics*, 2006, 14: 15-21
- 9 Levelt C, Schiller N, Levelt W. The acquisition of syllable types. *Lang Acquis*, 2000, 8: 237-264
- 10 Bonatti L L, Penña M, Nespor M, et al. Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychol Sci*, 2005, 16: 451-459
- 11 Capellan A R, Schwartz J L, Laboissière R, et al. Two CV syllables for one pointing gesture as an optimal ratio for jaw-arm coordination in a deictic task: A preliminary study. *Proceedings of the European Cognitive Science Conference*. Taylor & Francis, 2007. 608-613
- 12 Nam H, Goldstein L, Saltzman E. Self-organization of syllable structure: a coupled oscillator model. In: Pellegrino F, Marsico E, Chitoran I, et al, eds. *Approaches to Phonological Complexity*. Berlin & New York: Mouton de Gruyter, 2009. 299-328
- 13 Sykes J B. *The Concise Oxford Dictionary of Current English*. 6th ed. Oxford: Oxford University Press, 1976
- 14 Atkinson Q D. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, 2011, 332: 346-349
- 15 Rulien M. *On the Origin of Languages*. California: Stanford University Press, 1994
- 16 Cavalli-Sforza L L, Cavalli-Sforza F. *The Great Human Diasporas: The History Of Diversity and Evolution*. Boston: Addison-Wesley, 1995

- 17 Charles D. The Origin of Species. London: John Murry, 1929
 - 18 Campbell L. What can we learn about the earliest human language by comparing languages known today? In: Laks B, ed. Origin and Evolution of Languages: Approaches, Models, Paradigms, 2008. London: Equinox Publishing Ltd., 2008. 79–111
 - 19 Liu H T, Cong J. Language clustering with word co-occurrence networks based on parallel texts. Chin Sci Bull, 2013, 58: 1139–1144
 - 20 Jakobson R, Halle M. Fundamentals of Language I. Phonology and Phonetics. Netherlands: Mouton's-Gravenhage, 1956
 - 21 Jespersen O. Language: Its Nature, Development, and Origin. London: George Allen & Unwin, 1922
 - 22 Swadesh M. Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. Proc Am Philos Soc, 1952, 96: 452–463
 - 23 Pielou E C. Mathematical Ecology. New York: Wiley and Sons, 1977
 - 24 Magurran A E. Ecological Diversity and Its Measurement. London: Croom Helm, 1988
 - 25 Schleicher A. A compendium of the comparative grammar of the Indo-European, Sanskrit, Greek and Latin languages(1871). In: Lehmann W P, ed. A reader in nineteenth century historical Indo-European linguistics. Bloomington: Indiana Uni Press, 1967. 87–96
 - 26 Pielou E C. An Introduction to Mathematical Ecology. New York: Wiley-Interscience, 1969
 - 27 Shorto H L. Word and syllable patterns in Palaung. BSOAS, 1960, 23: 544–57
 - 28 Brunelle M, Pittayaporn P. Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia. Diachronica, 2012, 29: 411–433
 - 29 Pagel M, Atkinson Q D, Andrew M. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. Nature, 2007, 449: 717–720
 - 30 Anthony D W. The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World. New Jersey: Princeton University Press, 2007
 - 31 Jiang D. The Phonological Consequences of Evolution of Monosyllabic-word Languages. In: Shi F, Peng G, eds. Eastward Flows the Great River: Festschrift in Honor of Professor William S-Y. WANG on his 80th Birthday. Hong Kong: City University of Hong Kong Press, 2013. 371–389
 - 32 Howell S. Society and Cosmos: Chewong of Peninsular Malaysia. Singapore: Oxford University Press, 1984
-

补充材料

Table S1 Evolution of word-syllable structures and the diversity of world languages

本文以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Supplemental Material 1 for

Table S1 Evolution of word–syllable structures and the diversity of world languages

Lang–area = language area

Long = longitude

Lat = latitude

Distance = offset distance

Angle = offset angle

DI = diversity index

WL = word length

language	Lang–area	Long	Lat	Distance	Angle	DI	WL
Afrikaans	Africa	22.00	–31.00	1.3216	–0.2934	2.0896	1.1401
Akkadian	West Asia	43.13	36.33	0.5254	0.0024	2.0522	2.2581
Albanian	Europe	20.00	41.00	0.9545	–0.1441	2.9187	1.8068
Amis	Oceania	120.00	22.12	0.8639	–0.0758	2.4312	2.5000
Armenian	Europe	44.52	40.17	1.3650	–0.2458	2.3892	1.6473
Aromanian	Europe	21.75	41.67	1.2145	–0.0825	3.3881	2.1602
Ayamara	South America	–69.05	–17.05	0.4017	–0.0245	1.9892	2.5196
Baheng	East Asia	109.42	25.12	0.5973	–0.1169	2.0829	1.3883
Baoan	Central Asia	103.42	36.23	0.7701	0.0746	2.4266	2.3377
Basque	Europe	2.10	41.30	0.8706	–0.0582	2.7723	2.0870
Beijing	East Asia	116.38	39.90	0.5577	–0.0507	1.8370	1.2836
Bengali	South Asia	90.00	24.00	0.6264	–0.0595	2.1712	1.8513
Brazilian Portuguese	South America	–46.62	–23.53	0.9618	–0.1031	3.0938	2.1765
Burmese	Southeast Asia	96.00	21.00	0.6017	–0.1072	1.7876	1.2077
Cahuilla	North America	–116.25	33.50	1.0707	–0.1818	2.4748	1.9669
Chamicuro	South America	–77.02	–12.03	1.0469	–0.0638	3.1497	3.0606
Chewong	Oceania	101.00	3.13	1.3017	–0.3114	1.6422	1.3284
Chumashan	North America	–120.25	34.50	1.3303	–0.2263	3.2829	1.8678
Cocopa	North America	–115.00	32.33	1.2376	–0.2821	2.6567	1.3684
Croatian	Europe	15.97	45.82	0.9231	–0.1268	3.1665	2.0725
Crow	North America	–108.00	47.00	0.5755	0.0758	2.4214	2.1646
Czech	Europe	15.00	50.00	1.2299	–0.2033	3.1900	1.8325
Danish	Europe	10.00	56.00	1.3501	–0.2623	2.7538	1.4755
Daofu	East Asia	101.20	31.63	0.9179	–0.1799	2.2766	1.6205
Dawoer	Northeast Asia	124.20	48.32	1.0053	–0.1663	2.2227	1.4667
Dena'ina	North America	–150.00	62.00	1.1212	–0.1812	3.4321	1.8876
Dinka	Africa	32.60	15.57	1.2097	–0.1366	1.8628	1.5950
Dong	East Asia	108.85	26.55	0.5298	–0.0654	1.4983	1.1421
Dutch	Europe	6.00	52.00	1.5199	–0.2934	2.5772	1.1942
Eastern Fijian	Oceania	179.45	–18.17	0.2116	0.0502	2.2273	2.4497
Eastern Yugu	Central Asia	103.00	35.00	0.7773	–0.1063	2.3811	1.8535
Egyptian–Arabic	West Asia	31.28	30.00	0.9199	–0.1911	2.1357	1.9952
Eloyi	Africa	8.75	7.33	0.5030	0.0469	2.3685	2.2366
English	Europe	0.00	51.50	1.2653	–0.2889	2.1049	1.1304
Estonian	Europe	26.00	59.00	0.8333	–0.1332	2.6279	1.8889
Ewenke	Northeast Asia	119.75	49.55	1.1400	0.0720	3.0441	2.9330
Finnish	Europe	24.93	60.17	0.5943	–0.0553	2.5911	2.5217
Franch	Europe	2.33	48.85	0.9876	–0.1620	2.8054	1.4444
Fula	Africa	11.25	10.25	1.0089	–0.1121	2.8476	2.3383
Fuzhou	East Asia	119.30	26.07	0.7131	–0.1219	2.0653	1.3550
Ge'ez	Africa	38.73	9.17	0.8256	–0.1747	2.1059	2.0821
Georgian	Europe	44.80	41.72	0.9942	–0.1509	3.0910	2.1063
German	Europe	13.33	52.52	1.5034	–0.2693	2.7456	1.5121
Greek	Europe	23.73	38.03	1.0467	–0.0791	3.2824	2.3140
Guangzhou	East Asia	113.27	23.13	0.7044	–0.0970	1.9128	1.2687

language	Lang-area	Long	Lat	Distance	Angle	DI	WL
Guarani	South America	-58.37	-34.60	0.4263	0.0971	2.2662	2.4105
Gurindji	Oceania	130.67	-17.67	0.7426	-0.1284	2.4791	2.3465
Hadza	Africa	35.73	-3.75	0.4521	-0.0153	2.2981	2.2778
Haerbin	East Asia	126.63	45.75	0.5729	-0.0503	1.9402	1.3300
Haitian	North America	-72.33	18.53	0.5704	-0.1202	2.1530	1.5121
Hebrew	West Asia	35.22	31.78	0.9813	-0.1944	1.8281	1.7923
Hezhe	Northeast Asia	132.50	47.65	0.8258	-0.0397	2.7089	2.1798
Hindi	South Asia	77.72	25.00	0.8777	-0.1734	2.3252	1.7729
Hlai	Southeast Asia	110.53	20.05	0.7634	-0.1215	1.8024	1.2455
Hmong	East Asia	104.47	27.45	0.2525	0.0210	1.4462	1.3641
Hopi	North America	-110.00	36.00	0.7012	-0.0751	2.7582	2.4083
Hungarian	Europe	19.25	47.43	1.1175	-0.1934	2.2755	1.5121
Icelandic	Europe	-21.97	64.15	1.5017	-0.2509	3.1646	1.7005
Igbirra	Africa	7.38	8.17	0.9736	0.1927	1.9735	2.4435
Ipai	North America	-116.50	32.17	1.1726	-0.1639	2.5737	2.0398
Italian	Europe	12.00	41.00	0.8228	-0.0860	2.8808	2.3961
Japanese	Northeast Asia	139.73	35.68	0.4814	0.0857	2.3989	2.4879
Jingpo	East Asia	98.97	24.72	0.8120	-0.1628	2.3090	1.5178
Kala Lagaw Ya	Oceania	143.00	-23.00	0.3161	0.0193	2.1667	2.4379
Karelian	Europe	32.00	64.00	0.8985	-0.0776	3.0322	2.9612
Kashubian	Europe	18.00	54.00	1.3845	-0.2086	3.5461	1.8976
Kazakh	Central Asia	57.17	50.28	1.1646	-0.1786	2.4061	1.6550
Ket	Northeast Asia	105.00	60.00	1.3265	-0.1771	2.5507	1.4709
Khmer	Southeast Asia	104.92	11.55	1.2319	-0.3005	1.8497	1.2512
Kiliwa	North America	-115.67	31.33	1.4923	-0.1993	3.7919	1.8864
Kirghiz	Central Asia	75.42	42.75	1.1160	-0.1718	2.2631	1.5800
Korean	Northeast Asia	127.05	37.58	0.8897	-0.1089	2.8005	1.9903
Kurdish	West Asia	53.00	32.00	1.1255	-0.2064	2.6015	1.5314
Kwaio	Oceania	161.00	-8.95	0.4699	0.1140	2.4012	2.5410
Kwamera	Oceania	169.42	-19.58	0.9089	0.0319	3.1864	2.4043
Lakota	North America	-100.00	46.00	0.4713	-0.0176	2.5006	2.2077
Lao	Southeast Asia	102.60	17.97	0.5879	-0.1768	1.0260	1.0663
Latin	Europe	12.50	41.90	1.2077	-0.1297	3.3479	2.2512
Latvian	Europe	24.08	56.88	1.5768	-0.2829	2.8466	1.5411
Lhasa-Tibetan	East Asia	91.17	29.68	0.3508	-0.0668	1.7454	1.6205
Lingala	Africa	15.32	-4.32	0.4423	-0.0336	2.2479	2.3744
Livonian	Europe	24.00	56.83	0.8466	-0.1322	2.3137	1.6976
Loudi	East Asia	112.00	27.73	0.2974	0.0274	1.8257	1.3900
Magori	Oceania	150.12	-10.33	0.4080	0.1017	2.1122	2.3701
Malagasy	Oceania	47.52	-20.00	0.3124	0.0415	2.3086	2.8986
Malay	Oceania	102.00	3.00	1.0050	-0.1281	2.4492	2.2850
Malayalam	South Asia	76.50	10.00	1.0001	-0.1053	3.3210	2.5610
Maltese	Africa	14.42	35.92	1.4258	-0.2562	2.7641	1.7005
Manam	Oceania	145.03	-4.07	0.4660	0.0912	2.6560	2.6963
Manchu	Northeast Asia	126.63	45.75	0.9285	-0.0488	2.9971	2.3757
Maori	Oceania	174.45	-41.28	0.3333	0.0905	2.0395	2.3137
Mapudungun	South America	-70.67	-33.43	0.7816	-0.1014	2.4751	2.0634
Mbula	Oceania	148.10	-5.42	0.6169	-0.0435	2.9677	2.3892
Meizhou	East Asia	116.10	24.27	0.7481	-0.0983	2.2610	1.3950
Mekeo	Oceania	147.40	-9.50	0.7903	0.2101	2.4975	2.4247
Mixtec	North America	-97.58	17.05	0.6654	0.0222	2.5783	2.1523
Mon	Southeast Asia	100.50	14.83	1.1585	-0.2100	2.1286	1.4112
Mongol	Northeast Asia	116.22	43.87	1.1369	-0.1594	2.4075	1.5922
Motuo	East Asia	95.25	29.22	0.7228	-0.1729	1.8613	1.5380

language	Lang-area	Long	Lat	Distance	Angle	DI	WL
Nahuatl	North America	-98.25	19.00	1.8518	-0.0838	4.4892	3.0213
Nama	Africa	18.00	-25.50	0.6316	-0.1230	1.9118	1.3818
Nanchang	East Asia	115.88	28.68	0.6984	-0.1022	2.0913	1.3200
Navajo	North America	-108.00	36.17	1.0787	-0.1132	3.0147	2.1659
Nengone	Oceania	168.00	-21.50	0.5709	0.0036	2.5624	2.2486
North Saami	Europe	24.00	69.00	1.5159	-0.1651	2.6251	2.0216
Norwegian	Europe	10.68	59.93	1.0118	-0.2167	2.4517	1.4058
Nusu	East Asia	99.42	26.42	0.4560	-0.0423	1.9087	1.3057
Ojibwe	North America	-92.50	47.67	0.9652	-0.1077	3.2005	2.6019
O'odham	North America	-112.00	32.00	1.0230	-0.1070	2.8745	1.9561
Oroqun	Northeast Asia	123.15	49.37	0.9636	-0.0249	2.8257	2.1045
Ossetian	Europe	44.00	44.00	1.4543	-0.2479	2.7233	1.7136
Otomi	North America	-97.92	19.08	0.4081	-0.0313	1.9514	1.8204
Persian	West Asia	51.50	35.75	1.3643	-0.2286	2.9798	1.9055
Polish	Europe	21.00	52.22	1.2563	-0.1900	3.3075	1.9852
Portuguese	Europe	-9.13	38.72	0.9747	-0.1029	3.0938	2.1667
P'urépecha	North America	-101.67	19.50	0.8276	0.0531	3.1967	3.1959
Q'eqchi'	North America	-89.83	16.00	1.2487	-0.2000	2.4974	1.6796
Quechua	South America	-65.25	-19.03	0.4611	-0.0338	1.7664	2.0435
Rapanui	Oceania	-109.37	-27.12	0.3980	0.0995	2.4241	2.7041
Romani	Europe	21.50	47.00	0.9680	-0.1834	2.6734	1.8792
Rotuman	Oceania	178.43	-12.05	0.4458	0.0628	2.2516	2.3626
Russian	Europe	37.62	55.75	1.2200	-0.2046	3.0640	1.9034
Sala	East Asia	102.77	36.40	1.0249	-0.1602	2.2715	1.6193
Samoan	Oceania	-171.75	-13.83	0.3895	0.1020	2.3916	2.4526
Sanskrit	South Asia	77.00	28.00	0.7501	-0.1026	2.7287	2.2667
Santali	South Asia	87.00	24.50	0.8618	-0.0931	2.3757	1.7866
SanteeDakota	North America	-103.40	44.13	0.4516	-0.0142	2.5846	2.2850
Shan	Southeast Asia	98.00	22.00	0.7330	-0.2200	0.9400	1.0485
Shoshone	North America	-114.00	41.00	1.0175	-0.0583	3.4456	2.6347
Shui	East Asia	107.00	26.83	0.7624	-0.2003	1.7098	1.1492
Sicilian	Europe	15.00	37.00	0.7777	-0.0655	2.7030	2.4320
Skiri Pawnee	North America	-98.00	35.50	1.1138	-0.1102	3.1125	2.5969
Spanish	Europe	-3.68	40.40	0.9716	-0.1030	3.0217	2.1373
Sranan	South America	-55.17	5.83	0.6288	-0.0787	2.3077	1.8040
Sundanese	Oceania	107.00	-7.00	1.0071	-0.1064	2.1291	2.1176
Suzhou	East Asia	120.60	31.30	0.4956	-0.0907	1.8378	1.3400
Swahili	Africa	39.00	-6.50	0.5219	0.0096	2.3373	2.1981
Swedish	Europe	19.75	64.25	1.2330	-0.2471	2.6781	1.4559
Syriac	Africa	43.00	36.00	1.0136	-0.2146	2.4770	1.8213
Taa	Africa	25.92	-24.75	1.2451	-0.0436	2.6481	1.9333
Tagalog	Oceania	121.00	15.00	0.7374	-0.0470	2.2708	2.1739
Tahitian	Oceania	-149.60	-17.57	0.5507	0.1155	2.6498	2.8792
Taiyuan	East Asia	112.55	37.87	0.4654	-0.0402	1.8067	1.2750
Tamil	South Asia	78.05	11.00	0.8334	-0.0720	2.8171	2.1667
Tatar	Central Asia	49.62	55.75	1.1456	-0.1696	2.2940	1.6150
Telugu	South Asia	79.00	17.60	0.6520	-0.0478	2.2867	2.4638
Thai	Southeast Asia	100.48	13.75	0.6860	-0.1935	1.4115	1.1353
Tlingit	North America	-135.00	59.33	0.9328	-0.1733	2.7785	1.7861
Tocharian	Central Asia	85.00	40.00	1.2505	-0.1784	3.3728	2.0437
Tupinambá	South America	-36.00	-6.00	0.9574	0.1119	2.8721	2.3587
Turkish	Europe	28.97	41.03	1.3137	-0.1896	2.5958	1.8792
Tutelo	North America	-78.75	37.67	1.0815	-0.0639	3.2825	2.2536
Tuwa	Central Asia	87.53	47.12	1.1732	-0.1344	2.5854	1.7475

language	Lang-area	Long	Lat	Distance	Angle	DI	WL
Tzotzil	North America	-92.67	16.50	0.9661	-0.2202	1.7903	1.6548
Uighur	Central Asia	87.63	43.72	1.0836	-0.1635	2.2473	1.6250
Urhobo	Africa	6.00	5.50	1.1751	0.1540	2.1018	2.8397
Uzbek	Central Asia	69.22	41.27	1.1460	-0.1753	2.3296	1.6100
Veps	Europe	35.00	60.00	0.8805	-0.1639	2.4416	1.6860
Vietnamese	Southeast Asia	105.85	21.03	0.5076	-0.1265	1.1215	1.0580
Võro	Europe	24.00	58.00	0.9480	-0.1642	2.6459	1.8019
Votic	Europe	30.00	59.50	0.6426	-0.0318	2.5604	1.9515
Welsh	Europe	-3.18	51.47	1.1612	-0.1603	2.9759	1.6087
Wenzhou	East Asia	120.70	28.00	0.2491	-0.0012	1.5511	1.3800
West Frisian	Europe	6.00	53.00	1.2726	-0.2481	2.6421	1.4155
West Yugu	Central Asia	99.95	39.43	1.0422	-0.1721	2.3561	1.6173
Western fijian	Oceania	177.42	-17.08	0.1489	0.0199	1.6780	2.2606
Wichí	South America	-62.58	-22.50	0.6553	-0.0930	2.9434	2.4221
Written Tibetan	East Asia	92.28	29.37	2.6786	-0.3101	2.9634	1.5578
Xibo	Central Asia	81.20	44.37	1.2924	0.1818	2.8534	2.96
Xide-Yi	East Asia	104.12	29.05	0.2602	0.0575	1.5822	1.6276
Yami	Oceania	121.50	22.00	1.0163	-0.0206	3.1950	2.6702
Yanghuang	East Asia	107.52	26.38	0.6264	-0.1708	1.3325	1.1978
Yaqui	North America	-110.25	27.50	0.5914	0.0239	2.6946	2.6067
Yolngu	Oceania	133.17	-20.17	1.1720	-0.1643	2.9307	2.6398
Yoruba	Africa	2.62	6.48	0.9150	0.2035	2.2701	2.6383
Yupik	Northeast Asia	-153.00	64.00	1.1793	-0.0955	3.1458	2.4830
Zapotec	North America	-97.17	16.50	0.8503	-0.0013	3.2671	2.6931
Zha	East Asia	97.55	30.65	0.6273	-0.0884	2.2495	1.6740
Zulu	Africa	18.58	-33.97	0.8684	0.0033	2.9215	2.4800