

## LEXICOSTATISTICS HAS NOT YET ATTAINED THE STATUS OF A SCIENCE

EPHRAIM CROSS

In 1950 Morris Swadesh approached me with a request that I test a certain suggested measurement system by applying it to the Romance Language subfamily of the Indo-European languages. The technique included the selection of a relatively small standard and current vocabulary composed of so-called "basic" or "universal" terms, as concrete as obtainable within the framework of natural languages. The determination of what is "basic" or "universal" is, pending much further scientific accumulation of data to establish the definitive fact, a purely empirical foundation compounded of a not too subjective, but none the less relatively limited knowledge of human language, present and past.

In anticipation of this exposition it seems essential to note that in the realm of the Romance languages the lexeme "to be" is basic and universal because it is found in the parent speech and has persisted in all of the daughter languages for a period of altogether two thousand years.

However, the usefulness or significance of the inclusion of such a vocabulary element might be questionable for general Indo-European, would be confusing for a late stage of Indo-European and would certainly be inappropriate for Proto-Indo-European in any comparison with Semitic or Finno-Ugric and quite untenable for American-Indian and indigenous African languages, whether for internal or external comparisons.

The suggested method of comparison and calculation was a very ingenious one. It derived, as so much of our current linguistic terminology and methodology, from the sciences of physics, chemistry, biology, and mathematics. I omit engineering, which is likewise drawn upon. Specifically, it was the structuring of an analogue to the use under parallel conditions of carbon 14 in dating archeological and geological finds.

After using several longer trial lists with variant results I finally constituted a vocabulary of 241 words:

above - over (adv.), alive, all, animal, arm, ashes, aunt (pat.), ax, back (n. - of man), bad, bark (of tree), before - in front (adv.), belly, below - under, bend (trans. - twig), berry, big, bird, bite, bitter, black, blood, blow (v. - with mouth), blue (like sky), bone, branch (of tree), break (trans. - stick), breathe, brother (older, of man), burn (intrans.), call out - shout, (I) can, carry (in arms), child (not infant), claw, close - shut, cloud, cold (weather), cry - weep,

dark, daughter, day, dead, dig, dog, drink, dry, dull (blade), dust, ear, earth (soil), eat, egg, eight, eleven, empty (container), eye, fall (e.g. tree), far, fat (of animal), father (of man), fear (v.), feather (plume), feel (with hands), find, fire, fish, five, flat, flesh, flow, flower, fly (v.), fog (mist), foot, four, front, full, gather up (acorns), give, go (afoot), good, grandfather (paternal), grandson, grass, green, guts (entrails), hair, half, hammer, hand, hard (not soft), he, head, hear, heart, heavy, hide - conceal, hill, hit - strike (with hand), hold (in hand), hot (heat - weather), how much?, hurt - pain (v. intrans.), husband, I, ice, inside (adv.), kill, knee, knife, know (facts), lake, leaf, lefthand (at or to left), leg (of man), lie (down), lift - raise, light (n.), light (of weight), lip, little - small (adj.), little (adv.), long, lose, louse, love - like (a person - not necessarily sexually), man (human), man (male), middle, moon, more (adv. quant.), mother, mouth, move (v. intr.), much, name, near (adv.), neck, new, night, nine, nose, not, oil, old, one, open, other, outside (adv.), part (n.), play, pull - draw, put - place, quick, rain, red, right (at, to the), river, road - path, root (e.g. tree), rub, run, salt, sand, scratch, sea, see, seven, shade, sharp (e.g. blade), shine, short, sister (elder), sit, six, skin, sky, sleep, slow, smell (v. trans.), smoke, snow, soft (to touch), son, speak - talk, spit, squeeze, stand, star, stone, straight, stretch, strong, suck, sun, swallow, sweat, sweet, swim, tail, take (in hands), tear (v.), tear (from weeping), teat, ten, that, thick, thin, this, three, throw, thunder, tongue, tooth, tree, turn (v. intrans.), twelve, twist, two, uncle (paternal), vomit, wash (objects), water, we, weak (not strong), wet, what?, where?, white (like milk), who?, wife, wind, wing, wish - want, woman, wood, worm, year, you (int. sing.), young.

By means of this vocabulary I examined and correlated the corresponding lexica of Latin, French, Italian, Romanian, Catalan, Spanish, Portuguese, Rhaeto-Romance (specifically, Upper and Lower Engadinian), and Sardinian (specifically, Southern and Northern Logudorese). My approach was entirely free of bias, with an open-minded, enthusiastic wait-and-see attitude and, as can be ascertained, without any intention of forcing answers.

I could have limited my list to one hundred items and thus facilitated the entailed mathematical calculations. This facility I rejected, despite the unfavorable laboratory conditions under which a teacher and researcher in my university is constrained to operate. During the twelve years of this investigation there were available no computers, nor human assistants to furnish relief to the drudgery of preliminary and elementary chores that have been a drag on the aspirations, exertions, and accomplishments of energetic teachers. I admit that not only for Romance, but no doubt for more general modern Indo-European application, several items of my list could well be excluded.

I gave the project the unambitious title of "Correspondences and Linguistic Proximity" and presented the results to an annual meeting of the Linguistic Society of America (1950). My report was limited to various correspondences and the conclusions as to chronological and geographic proximity to be deduced therefrom. This information could at all points be checked against our historically established record of the time and area distribution of the Latin and Romance Languages.

Present at the publication of my findings, E. H. Sturtevant, while not being in a position to challenge the accuracy of my interrelational percentages, nevertheless, voicing the hostile reaction of vested interests of that day, remarked forcibly that it was