SINO-PLATONIC PAPERS

Number 202 May, 2010

An “ABC” Exercise in Old Sinitic Lexical Statistics

by
David McCraw

Victor H. Mair, Editor-in-Chief
Sino-Platonic Papers
Department of East Asian Languages and Civilizations
University of Pennsylvania
Philadelphia, PA 19104-6305 USA
vmair@sas.upenn.edu
www.sino-platonic.org

SINO-PLATONIC PAPERS

FOUNDED 1986

SINO-PLATONIC PAPERS is an occasional series dedicated to making available to specialists and the interested public the results of research that, because of its unconventional or controversial nature, might otherwise go unpublished. The editor-in-chief actively encourages younger, not yet well established, scholars and independent authors to submit manuscripts for consideration. Contributions in any of the major scholarly languages of the world, including romanized modern standard Mandarin (MSM) and Japanese, are acceptable. In special circumstances, papers written in one of the Sinitic topolects (*fangyan*) may be considered for publication.

Although the chief focus of *Sino-Platonic Papers* is on the intercultural relations of China with other peoples, challenging and creative studies on a wide variety of philological subjects will be entertained. This series is **not** the place for safe, sober, and stodgy presentations. *Sino-Platonic Papers* prefers lively work that, while taking reasonable risks to advance the field, capitalizes on brilliant new insights into the development of civilization.

Submissions are regularly sent out to be refereed, and extensive editorial suggestions for revision may be offered.

Sino-Platonic Papers emphasizes substance over form. We do, however, strongly recommend that prospective authors consult our style guidelines at www.sino-platonic.org/stylesheet.doc. Manuscripts should be submitted as electronic files, preferably in Microsoft Word format. You may wish to use our sample document template, available here: www.sino-platonic.org/spp.dot.

Beginning with issue no. 171, *Sino-Platonic Papers* has been published electronically on the Web at www.sino-platonic.org. Issues 1–170, however, will continue to be sold as paper copies until our stock runs out, after which they too will be made available on the Web.

Please note: When the editor goes on an expedition or research trip, all operations (including filling orders) may temporarily cease for up to three months at a time. In such circumstances, those who wish to purchase various issues of *SPP* are requested to wait patiently until he returns. If issues are urgently needed while the editor is away, they may be requested through Interlibrary Loan. You should also check our Web site at www.sino-platonic.org, as back issues are regularly rereleased for free as PDF editions.

An "ABC" Exercise in Old Sinitic Lexical Statistics*

(in which we build a gizmo that uses etymological roots to shed light on the roots of Old Sinitic, the wanderings of peoples who helped form the Sinosphere, and the affiliations of Old Chinese texts)

This article explores a new resource. Recent research — the fruit of many long years Axel Schuessler has spent gathering words — reveals an astonishing number of very old Southeastern words in the Old Sinitic lexicon.[1] Schuessler has, in his words, uncovered "the multiple origins of the Chinese lexicon";[2] as Schuessler remarks, amazedly, "When pursuing OC and TB/ST etyma down to their roots, one often seems to hit AA bedrock, that is, a root shared with AA."[3] When the germ idea of this paper first occurred to us, we must admit to having dismissed it as farfetched. But later, on one of those occasions when insomnia decides to spend the night, we found ourselves somehow convinced that you could use these old "Southeastern" words to investigate our earliest Chinese texts.[4] ABC's insights enable a new tool that can cluster

---

* This article owes much to many experts in linguistics and statistics. At the Univ. of Hawaii, Benjamin Bergen, Patricia Donegan, Leon Serafim, and Alexander Vovin all gave sage advice. Abroad, Wolfgang Behr and Victor Mair offered suggestions about an inchoate draft and also sent valuable scholarly reprints. George van Driem also kindly sent copies of his recent work. Edward Shaughnessy and Jonathan Smith sent invaluable datasets from early inscriptional literature. Paul Sidwell and Axel Schuessler both made trenchant and helpful criticisms. With all this advice, we should have emerged with a perfect gizmo. Its flaws and omissions remain, alas, the responsibility of its author.

1 Axel Schuessler, *ABC Etymological Dictionary of Old Chinese* (University of Hawai'i Press, 2007); henceforth, "ABC." Here, let OC stand for "Old Sinitic," TB for "Tibeto-Burman," and AA for "Austro-Asiatic."

2 Axel Schuessler, "Multiple Origins of the Old Chinese Lexicon." *Journal of Chinese Linguistics* 31.1 (2005),1–71.

3 ABC 2007:4.

4 For the moment, we shall not worry much about whether our "Southeastern" words come from Austro-Asiatic

texts by their differential use of words from either Tibeto-Burman or "Southeastern" origins. A bit more precisely, we shall rely on the comparanda in ABC to separate western sheep from southeastern goats.[5] Moreover, you can use this gizmo to help shed light on longstanding controversies about "early China." For example:

1. Did the Shang and Zhou share a common language? Can we meaningfully speak of "Chinese" — and at what time?
2. Do the early (Zhou) chapters of the *Documents* 書經 really belong, roughly, to the era of the Zhou conquest?
3. When — and how — do these "Southeastern" words surface in the Sinitic lexicon?
4. How do unusual texts like *Yijing, Laozi*, and the early *Chuci* fit in the Warring States [WS] textual system? And what organizes that system, anyway?

Incredibly, our gizmo — a kind of etymological root-sorter — works, and can usefully comment on these questions. Yet before we show you how it works, we need to clear some ground and set a proper foundation. Unless we understand a little how these words got to (and sometimes from) the East Asian Heartland (EAH — the Yellow River and Wei River valleys and adjoining areas; for convenience, we shall sometimes refer to the "Chinese heartland" or "Central Valley"[6]), and who brought them, and when and how, we won't manage to figure out why this gizmo works.

---

sources, or from Austronesian, proto-Thai (Tai-Kadai), Miao-Yao (Hmong-Mien), or even Japonic sources. We will address this issue later.

5 We shall not much concern ourselves with the "cognate" issue, or whether some of these "Southeastern" languages have genetic links to Old Sinitic. At a sufficient time-depth — and Schuessler's careful to indicate later loans that we ignore here — shared words indicate at least close contact, and our gizmo will find such contacts "close enough for jazz." See below, pp. 15–6, for further explanation.

6 See Victor H. Mair, "The North(west)ern Peoples and the Recurrent Origins of the 'Chinese' State," in *The Teleology of the Modern Nation-State: Japan and China*, ed. Joshua A. Fogel (Philadelphia: University of Pennsylvania Press, 2005), pp. 46–84.

Readers need to take a short detour, while we sketch, as well as current states of ignorance allow, the events and peoples who can situate our gizmo and our outlandish claims for it [for obscure places and cultures mentioned, see Map, below].

Let's keep a few grand outlining themes in mind. First, the EAH and nearby lands really were a "*Central* Kingdom." Whether you hazarded the long, dry trek skirting the Taklamakan Desert from the northwest, or labored over the forbidding mountains from the southwest, or hopped up the sandy Eastern Coast northward, exhausting the shellfish of one estuary after another, this Central Kingdom offered a treasure-house of resources. Its deep loess deposits enabled a number of promising cereal-grasses to grow, its forests abounded with nuts and wildlife, and the Great River (with its tributaries) ensured an enduring supply of fish and water.

Modern readers will want to remember that, thousands of years ago, the warmer wetter climate and prelapsarian state of the EAH — before we had deforested and eroded it — made

northern China a different place.[7] Let one telling detail suffice for a hundred: in the early Neolithic, elephants roamed the forests around Beijing.[8]

How many times did hopeful, hungry foragers, hunters, gatherers, and (eventually) would-be farmers make the arduous trek to this promised land? How often did the successful survivors dispatch members back to tell their kin of this new land?

We do not know, anymore than we can attest how much they suffered, how many died, and what transformations these journeys wrought. The archaeological record, too, has its thorny thickets and morasses, its trackless wastes and forbidding passes. History has left very few traces to guess by. But this we can surmise — Old Sinitic, as nearly all agree, has its genetic taproot in the Tibeto-Burman (TB) family.[9] TB languages have their center of maximum diversity in the Central Himalayan highlands. From there, TB speakers apparently fanned out southwest into northern India, southeast into Burma, and northeasterly, either into Tibet proper, or along the great river valleys eastward into, no doubt, Yunnan and Sichuan.[10] For our purposes, let's focus

---

7 For a convenient review of paleoclimatological data, see Mark Elvin, *Sediments of Time: Environment and Society in Chinese History* (Cambridge University Press, 1998).

8 Mark Elvin, *The Retreat of the Elephants: an Environmental History of China* (Yale University Press, 2004), Chapter 1.

9 We shall use "Tibeto-Burman" vaguely, without committing ourselves to any particular phylogeny. Indeed, the experts now incline to a similar agnosticism about this family's internal organization. For state-of-the art analysis, see esp. the writings of George van Driem, such as *Languages of the Himalayas: an Ethnolinguistic Handbook of the Greater Himalayan Region* (2V: Leiden, 2001); George van Driem, "Tibeto-Burman vs. Indo-Chinese: Implications for Population Geneticists, Archaeologists and Prehistorians," in Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, eds. *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*.(London, 2005), pp. 81–106; George van Driem, "Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory," in Yogendra Yadava,ed., *Contemporary Issues in Nepalese Linguistics* (Kathmandu, 2005), pp. 285–338; George van Driem: "The Diversity of the Tibeto-Burman Language Family and the Linguistic Ancestry of Chinese," *Bulletin of Chinese Linguistics*, 1 (2007): 211–270 [henceforth, cited as van Driem 2007]; George van Driem, "To Which Language Family does Chinese Belong, or What's in a Name?" in Alicia Sanchez-Mazas, et al., eds, *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics* (London, 2008), pp. 219–253 [henceforth, cited as Sanchez-Mazas 2008]

10 Observe that both camps, the Benedict/Matisoff "S-T" side, and the van Driem "TB" camp, now endorse the idea

on that last route. Granted, some hardy trekkers might have taken the deadly track up the Tang La pass, through the frozen plateaus of Qinghai, and thence to the headwaters of the Yellow River. But a better-documented route descended the steep flanks of the Himalayan massif, past Markam, skirted Gongga Shan, and followed the steep track down past Kangding to Chengdu; another ran along the "Southern Silk Route" from Mandalay, crossing the headwaters of the Salween, and trekking over to Dali.[11]

Readers will note that we subscribe to a "pull" theory of EAH peopling, rather than a "push" theory that either envisions first farmers moving from the EAH to Tibet, or else imagines a grand "demic diffusion" of TB speakers who overwhelmed indigenous EAH populations.[12] Of

---

of a Central Himalayan homeland. See esp. James A. Matisoff, *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction* (University of California Press: Berkeley, 2003); van Driem 2007: 40.

11 For Sichuan as the staging area for TB explorations into the "Chinese heartland," see the writings of George van Driem, for example van Driem 2007:40. For fragmentary evidence about the trans-Himalayan route down into Sichuan, see also Steven Sage, *Ancient Sichuan and the Unification of China* (State University of New York Press: 1992): Chapter 1; Jay Xu in Robert Bagley, ed., *Ancient Sichuan: Treasures from a Lost Civilization* (Princeton UP: 2001):22; Rob Hudson, "The Archaeology of the Early Buddhist Kingdoms of Thailand," in Peter Bellwood et al., eds., *Southeast Asia: from Prehistory to History* (London, 2004): 149–176: see esp. pp. 149–157. Explorer Zhang Qian made the earliest reference to India-Burma-Yunnan-Sichuan trade routes in *Shiji* 116.2995–6; cf. 123.3166. Compare the review in Sun Laichen, "Chinese Historical Sources on Burma: A Bibliography of Primary and Secondary Works," *The Journal of Burma Studies*, Volume 2 (1997):1–116. For a recent, trenchant critique of Sun and other Chinese writings on the "Southern Silk Road," see Yang Bin, "Horses, Silver, and Cowries: Yunnan in Global Perspective," *Journal of World History* 15.3 (2004): 281–322. Scholars await more complete findings from the southwest!

12 For the former view, see Su Bing, et al., "Y Chromosome Haplotypes Reveal Prehistorical Migrations to the Himalayas," *Human Genetics* 107 (2000), esp. pp. 588–9 (or the unsupported assertion in David Bradley et al., eds., *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff* [Australian National University, 2003]:3–4), and the appropriate criticisms by Roger Blench in Sanchez-Mazas 2008:110–2. For the latter view, see the Bellwood/Renfrew model of "farming dispersion" in Peter Bellwood and Colin Renfrew, eds., *Examining the Farming/Language Dispersal Hypothesis* (Cambridge UP, 2002), which several participants in that volume criticize cogently. To simplify a complex set of issues, a grand demic "farming dispersion" theory works well with the Austronesians who fanned out and peopled or overwhelmed thinly populated islands across the Pacific; it has more trouble accounting for agricultural expansions and migrations in complex,

course, the Central Kingdom's magnetic appeal provides not only an explanation, but also a problem. If many diverse tribes came seeking the good life, how can we separate out the complex strands of diversity and distinguish who spoke what? Well, we can't; as they say, "pots don't speak," and neither do haplotypes.[13] It seems fair to imagine that a great number of different ethnolinguistic groups may have crossed through the EAH from, roughly, 10000BP to 4000BP. But observe this does support the inference that alleged speakers of "Old Sinitic" (if the term has any meaning back so far) had no monopoly on this territory.[14] In fact, the closer you look, the less likely a special Separate Destiny for Sinitic and "the Chinese" looks. Modern reconstructions of Old Sinitic make it resemble, as linguists have noted, a "run-of-the mill TB language."[15] Genetic studies find modern Northern Han Chinese closer to Altaic populations than to Southern Chinese, who in turn have closest genetic ties to other Southeast Asians speaking a variety of Southeastern tongues.[16] And archaeology has uncovered a bewildering variety of tribes and cultural horizons; you need seek no farther than the recent Wikipedia timeline of Chinese archaeology to turn the hunt for a "Han ancestor" tribe into a bemusing wild-Han chase.

Let's close our discussion of the "western route" with one illustrative detail. Among the many common Chinese words with no known etymology, "mountain" 山 (**sran**) surely stands

---

populated mainlands like the Near East (see the criticisms by Zvelebil in Bellwood 2002) or Asia (see, inter alia, the criticisms in van Driem 2005).

13 Well, not yet, not so that we can understand. For suggestive evidence of a Himalayan-to-South China movement, see H. Shi et al., "Y-Chromosome Evidence of Southern Origin of the East Asian–Specific Haplogroup O3-M122," *The American Journal of Human Genetics*, 77.3 [2002], 408–419.

14 An inference drawn elegantly — inter alia — by Roger Blench, in Sanchez-Mazas 2008:112. cf. note 14, below.

15 See for example, the quoted passage from van Driem 2007:19.

16 A result frequently obtained by population geneticists, among them, L. Luca Cavalli-Sforza et al**.**, eds., *The History and Geography of Human Genes* (Princeton UP, 1994), 225. Intriguingly, scholars more and more incline to envisioning the earlier YRV settlers as speakers of something other than "Chinese." For the thesis that Altaic speakers produced Yangshao culture (on grounds of geographical probability and "cultural vocabulary"), see, among others, Ilia Peiros and Sergei Starostin, "Altaic Loans in Old Chinese, Sanchez-Mazas 2008: 254–262. While probably no more likely than a "Chinese" hypothesis, this at least demonstrates we have no firm ground to assume some proto-Sinitic language has roots reaching "all the way down."

out. TB speakers, passing along the roof of the world, hardly lacked words for mountains! But their native words, though they found such modest places in the Chinese lexicon as "ridge" 岡 (WT:**sgang**) and "mound" 陵 (PTB:**m-rang**), yielded to the bewildering **sran**, for which no origin and no close relative emerge. You'd imagine that either native mountain-cults — and surely most ancient peoples recognized mountains as gods — didn't travel well, or perhaps the mana of "Sinitic" oreads and their name(s) eclipsed native magic. But neither any TB orothearch nor any known corresponding Chinese alpine deity has left its name. Instead, we have the mysterious **sran**, to stand for all those lost tribes whose passage to and from the EAH region left no other trace.

Such complexity refuses to end with waves of TB settlers and their backwash(es). We must also consider how southern and eastern languages could have left such numerous traces on Old Sinitic. People have long guessed that the central Yangzi region, which we may for convenience call Chu (though the name would prove anachronistic projected back much before 800 BCE), provided a route to the EAH. It appears that contacts between northern Yangshao and southern Daxi (and other) cultures find documentation back at least 7000 years.[17] The occasional proto-Thai and Miao-Yao words surfacing in the *Chuci* alone would suggest a channel for those interactions.[18] Instead, we shall turn to a greater mystery: how could a language family like Austro-Asiatic (AA) contribute so much to the formation of the Old Chinese lexicon?[19] Don't today's AA speakers belong to Southern Asia and, at best, the far southern reaches of China?

It turns out that recent Chinese archaeological efforts can shed a good deal of light. The central Yangzi Daxi 大溪 culture, for example, spread its connections as far south as the Pearl

---

17 See Li Shaolian 李紹連, "Shilun zhongyuan he Jianghan diqu xinshiqi shidai wenhua de guanxi" 試論中原和江漢地區新石器時代文化的關係, 考古學集刊 1984.4; K.C. Chang et al., eds., *The Formation of Chinese Civilization: an Archaeological Perspective* (Yale UP, 2005), 56 [n.43, p.315]; Lin Xiang 林向, "Daxi wenhua yu Wushan yizhi" 大溪文化與巫山遺址, 中國考古學會第二次年會論文集 (1980) pp.128–30.

18 For an enlightening recent review of relevant Chu issues, see Wolfgang Behr's post at WSW [http://www.umass.edu/wsw], 1-29-2006.

19 Below we will examine statistics; for now, suffice it to say that AA words constitute the majority of ABC's Southeastern comparanda.

River in Guangdong; meanwhile, their artifacts turn up in Shandong![20] For later Neolithic times down to the Eastern Zhou, we're fortunate to have the excellent critical literature review and accounts in Luo 1999.[21] Luo describes how an assemblage she calls the "coastal culture" came to dominate the Eastern Seaboard from Liaodong to Guangdong and play a pivotal role in the formation of the "Central Valley" culture we now think of as China's "cradle of civilization." She relates how numerous regional subcultures more or less coalesced to form what modern Chinese archaeologists would, oversimplifying, call Yi 夷 culture (think Dawenkou 大汶口 and Longshan 龍山, among others), focused in coastal Shandong; she also relates how numerous regional subcultures more or less coalesced to form what modern Chinese archaeologists would, oversimplifying, call 百越 culture (think Hemudu 河姆渡, Liangzhu 良渚, and their heirs, among others), centered in the lower Yangzi and Zhe river valleys. Cultures from these two areas, moreover, kept up an early, historically intensifying, and persistent intercourse that allows us to envisage an overarching "coastal culture" stretching (with regional subvariations) from the Gulf of Bohai to the Pearl River estuary.

Would these diverse yet similar cultures have spoken a common language? Well, no one really knows, but it seems most unlikely.[22] The shifting patchwork of settlements over this vast area, during the last several thousand years, would allow room for Laurent Sagart's hypothetical original Austronesians, perhaps originally speaking a sister-tongue of Old Sinitic's ancestors, to spread southeast over to Taiwan, and thence across the Pacific.[23] It would allow room for wave

---

20 See n.14.

21 Luo Chia-lin, *Coastal Culture and Religion in Early China* (Indiana UP: 1999). For more on Yi culture and how it influenced Shang culture, see esp. Wang Xun 王迅, *Dong yi wenhua yu Huai yi wenhua yanjiu* 东夷文化与淮夷文化研究 (Beijing, 1994), Chapter 1. For recent coverage of Baiyue 百越 cultures, see esp. Lin Weiwen 林蔚文, *Zhongguo baiyue minzu jingjishi* 中国百越民族经济史 (廈门大学:2003), Ch.1; Zhou Youtao 周幼涛, "Lun diyu wenhua shiye zhong de Yue wenhua yanjiu" 论地域文化视野中的越文化研究, in 南方文物 2007.4, esp.196–7.

22 For acute comments on this issue, see Paul Sidwell's review of ABC, "Was There Ever an Austroasiatic Substrate in Central or Eastern China?" SOAS (Australian National UP), 2008 [henceforth, cited as Sidwell 2008].

23 For the latest edition of Sagart's "Sino-Austronesian" thesis, see his essay in Sagart/Blench, *The Peopling of East Asia,* 2005.

after wave of hopeful immigrants, speaking perhaps unknown tongues, washing up on this shore or another between modern Shanghai and Tianjin, making (with luck) inroads and, perhaps, eventually disappearing without a notable trace. And it affords plenty of room for speakers of AA languages to venture north, mingle perhaps with intermediaries and, directly or indirectly, influence the development of the "Old Chinese lexicon" (and, of course, carry its influences back down south).

Thus, well before the grand Han expansion that expelled earlier cultures from southern China (during the mere last 2000+ years), extensive EAH and SE coastal contacts find plausible explanation.

Let's explore just one significant detail, to illustrate what oddities may have occurred. Jade — nephrite in particular — we tend to imagine as a northwestern resource. The famous jade-fields of Khotan and the Altai would suggest a homeland for this most peculiar word 玉 (**ŋok**). And a very peculiar word it seems — in its splendid isolation (Karlgren family #1216, with only a distant descendant 琟 to keep it company) and complete lack of TB cognates.[24] Oddly, its only close phonetic neighbor, 嶽 (**ŋrok**) "marchmount," comes from AA **ŋok** (ABC:p.596)! Our minds naturally associate jade with its alpine sources; fascinating to speculate that "jade," against all odds, may come not from the northwest but from the southeast! Intriguingly, it turns out that the Liaodong peninsula, home of the famous Hongshan 紅山 culture, and the hills of Zhejiang, close by the equally famous Liangzhu 良渚 culture, produce jade — nephrite, to get specific. Not only did both cultures win renown for their excellent jadework, they traded with the intervening cultures. Jade traveled — and traveled widely.[25] In the face of

---

24 See Bernhard Karlgren, *Grammata Serica Recensa* (Stockholm: Museum of Far Eastern Antiquities, 1950) and Schuessler 2005 (for the "minimal Old Chinese" reconstruction we adopt). Inexplicably, ABC omits this word, which does find its way into Axel Schuessler, *Minimal Old Chinese* (Hawaii UP, 2008), p. 157.

25 On jades from Hongshan and Liangzhu and their sources, see esp. Luan Bing'ao 欒秉璈, "Shiqian guyu yuzhi ji yuliao laiyuan wenti yanjiu," 史前古玉玉質及玉料來源問題研究; 南陽師範學院學報 4.2 [2005]:113–6). (Thanks to W. Behr for providing me with a copy!) On the relevant directions of the jade trade, see esp. Liu Xinru, *The Silk Road: Overland Trade and Cultural Interactions in Eurasia* (Oxford UP, 1998). Just to give one example, ancient bifurcated scepters occur from Shaanxi to Shandong, from Vietnam to Sichuan, as noted in Deng Cong 鄧聰, et al.,

persistent failure to find "northwestern" etymologies for jade,[26] we must wonder if the word and its magic did not reveal yet another southeastern contribution….

We do not want to mislead readers into thinking "non-TB" equals "AA." Far from it. To redress this potential misunderstanding, we offer another common, mysterious member of the OC lexicon, "wind" 風 **prjəm**. The old TB word didn't survive in China; of the Shang names for the winds of the four quarters (compare the Greek *Anemoi*), only fragments survive as curiosities in the miscellaneous lore of *Huainan zi* or *Shanhai jing*.[27] Instead, the Old Korean word for wind 風 **palləm** remains as our outcrop from the vanished linguistic past of this area.[28] Why? Well, in our own cultural memories, the lingering traces of Boreas, Notus, Zephyrus, and Eurus have dimmed with the fading of their cults. Perhaps in the Sinitic case strong taboos on the names of the winds — see Japanese scholars, who for some reason have proven acutely sensitive to the power of "divine winds" — resulted in a swerve to less charged appellations. Alternately, the power of a Korean wind-god left a palpable Sinitic trace, among the fathomless migrations of peoples to and fro. In short, for every trace that chance and "natural process" have left us, overlayering and erosion have obscured, say, seventeen other lost words.

Observe that we shy away from trying to identify any specific ethnolinguistic groups with any particular prehistoric cultural horizons. Remember, pots don't speak, and archaeogenetic tools as of today lack the necessary resolution to illumine the details of ancient tribal interactions.

---

eds., *Nan Zhongguo ji linjin diqu guwenhua yanjiu: qingzhu Zheng Dekun jiaoshou congshi xueshu huodong liushi zhounian lunwenji* 南中國及鄰近地區古文化研究: 慶祝鄭德坤教授從事學術活動六十週年論文集 (Hong Kong, 1994), passim; cf. Lothar von Falkenhausen, "The External Connections of Sanxingdui," *Journal of East Asian Archaeology* 5 (2003), 205–6.

26 Alexander Vovin, p.c., 4/2009. We await proof to the contrary from interested parties!

27 See, inter alia, the excellent treatment in Akatsuka Kiyoshi, *Chūgoku kodai no shūkyō to bunka–In ōchō no saishi* (Tōkyō, 1977, 425–440; cf. the review in Kuriyama Shigehisa, "The Imagination of Winds and the Development of the Chinese Conception of the Body," in Angela Zito and Tani Barlow, eds., *Body, Subject and Power in China* (Chicago: University of Chicago Press, 1994), pp. 23–41. See also Chapter 2 of Wolfgang Behr's invaluable 1997 dissertation, *Reimende Bronzeinschriften und die Entstehung der Chineschen Endreimdichtung* (Goethe Univ. PhD diss., 1997), esp. pp. 73–77, for updated sources.

28 ABC:238, confirmed by Alexander Vovin in seminar — Nov., 2006

We’ll simply have to remain content, for now, with a rough sketch. If we could make one grand generalization, we would liken these peoplings of what we now call northern China as a landscape painted by the tides or, if you prefer, a watercolor built from innumerable layers of wash. When you observe a beach, most of the waves leave no lasting trace onshore; only the most powerful leave a high watermark and some flotsam we can examine and analyze. Similarly, only the uppermost and most strongly colored layers of a watercolor will still register to your eye. Regrettably, the great Han expansion of historical times has erased a great many early traces, and much linguistic diversity. But the triple window afforded by historical linguistics, archaeology, and archaeogenetics can let our imaginations envision some broad outlines for how Old Sinitic might have formed.[29] We must keep multiple origins and the inland-coastline divide in mind, in order to understand why our outlandish gizmo can work at all.

29 “Triple window” borrows from William S-Y Wang, "Three Windows on the Past," in Victor H. Mair, ed., *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia* (Journal of Indo-European Studies Monograph No.26: 2V, Pennsylvania UP), 2.508–534.

Part 2: The gizmo

ABC contains roughly 5,500 Old Sinitic words; its index includes roughly 2,300 English equivalents (nearly 2.4 Chinese examples per English word fits our experience with building the gizmo). We sampled 15 pages (3 from roughly each 100 pages) from its nearly 500-page body to estimate where ABC's words came from.[30] Here's what we found:

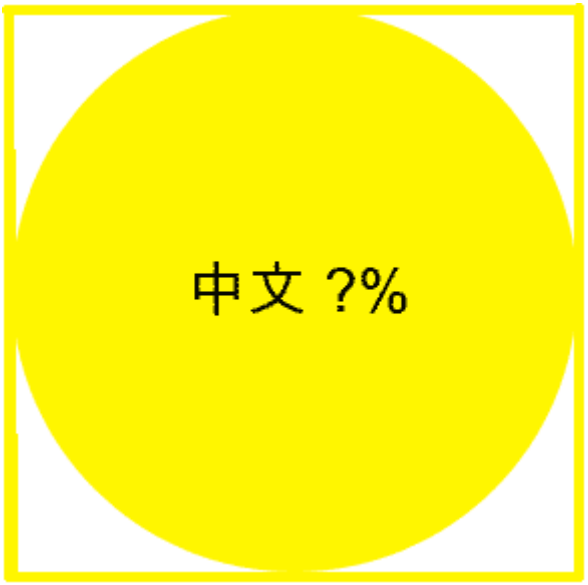
range	sample	# words	TB	AA (MK)	p-Thai	MY	AN	other	?	2 source	area
p.149–246	208–10	35	13	6	1			1	13	1	1
p.247–344	294–6	33	12	12					10	3	
p.345–442	346–8	29	16	3	2	1			7	1	1
p.443–540	455–7	27	9	12	2				4	1	2
p.541–638	545–7	38	21	6	1				11	1	1
Total	15pp.	162	71	39	6	1	0	1	45	7	5

(Note: multi-source and area words increase the total count beyond the "number of sampled words" (from column 3).

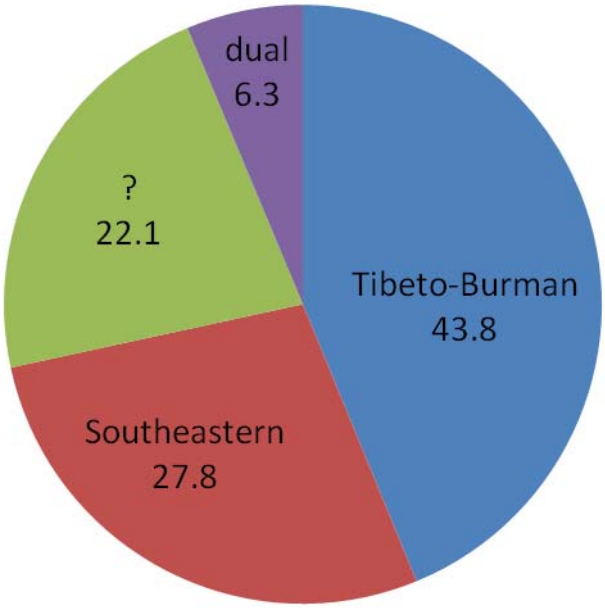
The following pie-charts summarize what we imagine readers might expect, and ABC's unexpected discovery:

30 We repeatedly threw dice to get the three pages from each of ABC's five hectosectors.

Chinese etymologies



Etymologies by %



It may surprise readers, but this ABC sample suggests that, for any substantial number of Old Chinese entries, TB comparanda will account for roughly 43.8% of entries, AA-affiliated words for 24.1% (and "southeastern" ones for 27.8%), unknown etymologies for 27.8%, and

words with two plausible alternative roots and area words will account for 8% (again, figures add up to slightly more than 100%, because of the last two categories). Given roughly 5,500 words in ABC, we would estimate he found about 2,420 TB comparanda, 1,323 AA, 204 Thai, maybe 34 MY, a meager handful of AN, 34 "others," about 1,527 unknown etymologies, and perhaps 440 area-words and words with two or more plausible etymologies.[31]

Fine, but what can you do with this knowledge? Well, here's what: take a standard word-list (we have used a slightly modified Swadesh 200-wordlist), mine ABC and its index to find relevant Old Sinitic "equivalents," list those, look up their etymologies, and, for each ancient text or chunk of text you want to examine, determine which words occur. Adding one or two refinements, you get a chart that looks like the file:Aswordlist; see Appendix, below.

Observe that it includes about 275 English words, partly because some of the Swadesh-200 demanded more than one translation when fitting Chinese, partly because we added some 45 terms having to do with basic actions, basic features of the Chinese landscape (who can imagine a basic Chinese wordlist without a word for "wall/city" 城?), a few tools (musical and martial, in particular; who would accept a basic Chinese wordlist without a word for "knife/chopper" 刀?), and some basic clan-fief-cult terms (a basic Chinese wordlist without "ancestor"祖 simply wouldn't do). These 275 words generated some 1,098 Chinese words, nearly all monosyllabic (we worked very hard to find 3.95 Sinographs per English word; we used many rare alternatives, nearly every one in ABC; if our wordlist had included fewer common English words, our average would have dropped closer to 2.4). Of these, 588 had TB comparanda or roots, 260 had

31 But note the strong methodological criticisms in Paul Sidwell, "Was there Ever an Austroasiatic Substrate in Central or Eastern China?" (Sidwell 2008). Prof. Sidwell cautions that Schuessler greatly overestimated AA matches and underestimated AN ones; he found what he looked for hardest — Paul Sidwell, pc: 4/18/09. Axel Schuessler notes that his ABC comparanda run a gamut from near-certain cognate, through likely ancient loans, to interesting speculations (Axel Schuessler, pc: 4/20/10). We have not attempted the fine distinctions among these subcategories that Prof. Schuessler has discerned, which would lead to different, lower figures. But though our gizmo does not make over-fine distinctions, we have counted strong connections as "+/-2" and weak ones as "+/-1" (see chart, below). Thus, we're not completely blind to the subtle distinctions ABC makes.

"southeastern" leanings,[32] and 250 had uncertain roots or seemed like "area words." This means 53.6% of basic vocabulary had TB roots, 23.7% had SE inclinations, and 22.8% fell into the null category (unknown, disputed, or area). Observe that TB roots proved more numerous in basic vocabulary than in our sampling, an expected result. Before we show you how our gizmo can use this data, let's anticipate an important methodological criticism: how do we know ABC got its etymologies right? A crucial question! First, many scholars have compiled TB-Chinese comparanda wordlists; Schuessler used his erudition and judgment to select from among them. Few would find more than an occasional dubious case among his choices. Second, for AA and proto-Thai, for example, he also had reference works of somewhat less reliability to work with. Let's take the AA example, because it has the greatest influence on results. A gloomsayer would argue that proto-AA hasn't gotten a convincing reconstruction; therefore, "there are no justified proto-Austroasiatic reconstructions." [33] Technically, one might say so, but that doesn't mean all of Harry Shorto's reconstructions thereby fall into unreliability. Granted, they vary in reliability;[34] still, they should mean *something*.[35] We will rely on force of numbers to drown out the few unreliable reconstructions and make them statistically meaningless. As Baxter (2000) has astutely observed, even a crude device will provide surprisingly reliable results if you insure correct statistical procedure.[36]

Let's also address a second, no less important, concern. Old Sinitic and old AA forms largely consist of "C-V-C"; can we assure comparing such simple syllabic structures won't result in tons of look-alikes? Has ABC accounted for this? ABC has not, and Sidwell 2008 rightly

---

32 We will shortly break down how many AA, how many Thai, etc.

33 Blench, in Sanchez-Mazas 2008:119. Cf. n.27.

34 So says Shorto's editor, Paul Sidwell (Sidwell 2008:3). See Harry Shorto [with the help of Paul Sidwell et al., ed.], *A Mon-Khmer Comparative Dictionary* (Australian National UP, 2006), p.3.

35 AA expert Patricia Donegan finds Blench's worries excessive — pc 4/06/09.

36 William H. Baxter and Alexis Manaster Ramer, "Beyond Lumping and Splitting: Probabilistic Issues in Historical Linguistics," in Colin Renfrew, April McMahon, and Larry Trask, eds., *Time Depth in Historical Linguistics* (Cambridge, UK: 2000): 167–188.

criticized Schuessler for not considering the "null hypothesis" that his Sinitic-AA likenesses could result from chance. Sidwell produced some intriguing calculations to estimate the probabilities of matches for a large dictionary. Unfortunately, he both overestimated the number of terms in ABC and also made a math mistake.[37] We have redone his math with the correct figures. For 5,500 words (using Sidwell's criteria), you would expect roughly 43 OC-AA random matches; Schuessler has found more than 1,320! Thus, we can discount the possibility that these matches arose due to pure chance.

Next, we assembled small databases, each roughly 2,000 Sinographs. We divided our list of 1,098 OC words into categories according to our appended *ASwordlist*. We then performed a mini corpus linguistics experiment, searching each database in turn for each word (using Word searches with wild-card functions).[38] First, we selected sixteen mini-corpora (asterisked); later, as data became available, we extended our gizmo's scope to the most ancient texts. Surprising results then suggested more texts to analyze. Below, we summarize data from months of searches.

---

37 Sidwell 2008: "Was there Ever an Austroasiatic Substrate in Central or Eastern China?", 8. We concur with Sidwell: ABC has undoubtedly overstated the number of OC-AA matches — but hardly by a factor of 30!

38 We would have preferred a strong concordance program, but because programmers have set these up for token-searches using large databases, they don't suit small-base searches for types, like ours. For the distinction, see standard introductions to corpus linguistics, such as Douglas Biber, et al., *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge UP:1998); and Tony McEnery and Andrew Wilson, eds., *Corpus Linguistics: an Introduction* (Edinburgh UP, 2001). Thanks to Jonathan Smith from the University of Pennsylvania for suggesting a Word macron program, which arrived too late for us to apply.

txt/affil>	+3	+2	+1	0	-1	-2	-3	hits/total	+%	sampl 字	Ind. x^2
商	0	35	3	11	1	8	1	53/61	.87	~2,500	**5.16**
金文 1	7	94	23	41	11	22	5	162/203	.80	2,848	**9.78**
書經 1	3	74	20	50	14	23	4	153/188	.81	4,673	**10.3**
*詩 4 周頌	3	50	7	17	8	19	2	64/104	.62	1,687	0.161
書 2	1	60	18	35	14	28	2	65/158	.41	2,912	4.42
易經	2	44	14	34	9	29	3	32/135	.24	2,732	20.04
*詩 3 大雅	3	51	16	44	8	23	6	54/151	.36	1,972	4.778
*詩 5 商魯頌	4	59	15	44	8	24	7	68/160	.425	2,011	2.042
*詩 1	3	40	12	29	9	21	1	47/115	.41	1,650	2.156
*詩 2	4	81	14	49	9	34	6	93/198	.47	3,731	0.988
金文 2	1	46	12	35	12	14	1	64/121	.53	1,350	0.17
*郭店		37	8	36	4	8		62/93	.67	1,932	0.589
*老		51	16	31	8	8		92/107	.87	1,929	**4.483**
*楚 1	1	61	17	26	11	31	2	63/149	.42	2,005	2.033
*論語		54	6	37	9	7		75/121	.62	3,660	0.206
*左 1	2	49	12	37	11	19	1	64/131	.49	2,027	0.505
左 mid	3	65	19	42	17	25	2	85/173	.49	2,945	1.04
*左 2		83	11	33	13	21	1	119/152	.78	1,991	**3.112**
墨子		43	9	3	10	3		79/66	1.20	2,583	**20.26**
墨子 2		65	20	37	20	26	1	75/169	.44	3,642	**2.86**
*莊子 1		76	10	43	17	10	1	125/157	.78	3,155	**3.507**
*禮記 1	2	74	24	46	16	25	2	106/189	.56	2,469	0.002
*孟		71	20	33	10	21	3	101/160	.63	2,447	0.343
荀子	2	68	16	43	14	17	1	107/161	.66	2,658	1.58

txt/affil>	**+3**	**+2**	**+1**	**0**	**-1**	**-2**	**-3**	**hits/total**	**+%**	**sampl 字**	**Ind. x^2**
韓非	2	95	16	35	18	27	2	134/195	.69	2,825	**2.792**
*莊 29		79	10	44	18	17		116/167	.69	3,100	1.255
商君	0	83	15	28	19	15	0	132/160	.825	3,405	**9.64**
*呂氏春秋	2	69	8	38	8	37	3	61/165	.37	3,747	**4.372**
16 texts				(587)		(82)	(145)	1,310/2,319	(.56)	39,513	

The x^2 results in our farthest-right column should astonish readers. You would expect that most pre-Han texts would represent a more-or-less uniform language, or "koine" *yayan* 雅言. Yet here our gizmo has dissected 28 text samples and found no fewer than 14 significantly deviant slices. 14 have markedly TB leanings, and 9 incline strongly toward SE words, at levels of confidence ranging from 91% to 99.9%! Whatever could explain such irregularities within our early "koine"?

Before exploring possible explanations, a few chart details demand an account. First, the rows across feature all words on our list found in any text (chunk); we do not keep track of repetitions (tokens), only "hits" (types). +3 means TB and rare; +2 means definitely TB; +1 means probably a TB root. 0 means unclear (see below). -1 through -3 do for SE links what +1 to +3 did for TB, in reverse. "Hits/total" simply adds the "positive" weighted figures from columns 2–7 and subtracts "negative" weighted figures. What exactly do we mean by "SE"? Well, as our chart above suggests, you might find words with an AA, AN, Thai (Tai-Kadai), Miao-Yao, or even proto-Japonic root/comparandum. Here we have lumped them all together. But, for readers curious about SE "splits," we sampled our first 100 "SE" words. Here's what we found:

> SE (1st 100 words) AA~68 [M-Kh 62 Munda 3 Vt 5]; AN 2; p-Tai 28; M-Y 6; others(?) 9–10.

Much as in our chart above, AA words strongly predominate, with a healthy minority of proto-Thai roots and links.

What does the "0" class mean?. Well, we might not know where the root comes from, or ABC might give alternate plausible etymologies, or it may suppose an "area word" with wide distribution. Here's how ABC's first 100 "0 words" break down: area=11; dual etymologies=22; unknown=67. As you can see, unclear etymologies account for 2/3 of our "0 words."

Readers may wonder why some words rate a +3 or -3, while others clearly affiliated with either TB or SE languages only rate a +-2. We decided to give extra weight to words that occur only rarely in OC. A statistics expert advised us against mixing "informativeness" and rarity in

one chart, but we finally decided that rareness does inform us in a significant way.[39] Let me illustrate. In *Song* #10.3 (Mao edition, of course), the stanza begins: 魴魚**赬**尾,王室如**燬**

Both "**ruddy**" and "**ablaze**" have well-known TB roots, but both scarcely occur in any other old Chinese text. To our minds, such strikingly conservative examples of "TB" vocabulary deserve special weight and attention.

Finally, the right-hand column gives an x^2 value. In our experiment, with "degree of freedom=1," any x^2 value of 3.84 or greater has significance at the 95% confidence level. Let's make sure we understand what that means. Given a text that has a notably higher (or lower) profile, with an x^2 value of 3.84, the chance of a random result *does not exceed* 5%. But who said we need 95% confidence? Any x^2 value over 2.71 gives a 90% confidence level; we decided (in advance) that 90% sure sounds awfully good for Old Chinese studies. After all, how significant an outlier do you need before you recognize that it stands apart from the baseline of which it forms a 1/16 part, in most cases? Thus, when the tail end of *Zuo zhuan* turns out to have an x^2 value of 3.11, we decided that a 7.7% chance of a random result would not faze us; we'll figure, for the moment, that this result seems likely to have some significance. We're not testing new medicines here; we're trying to explore the dark side of an ancient Sinitic moon, hoping to distinguish *mons* from *mare*.

How did those AA speakers get to Shandong? Perhaps they didn't. Words for important tools and cultural innovations can permeate tribal borders without any significant demic migration. Or, individual tribes can intermarry and push the words another 100 kilometers up the coast. With time, those words might travel even without a direct OC-AA interaction. Maybe coastal peoples developed a patois from many different languages and language families, after complex interactions over time. Perhaps many waves of boatgoing coastal people, speaking who knows what kinds of pidgin, washed up in Shandong and gradually forged Longshan culture from multiple sources. At present, we cannot guess which scenario best fits the past. But we can surmise that Sundadonts did not leave traces in Longshan Shandong, and most known AA

39 Thanks to Benjamin Bergen of the Univ. of Hawaii, who offered sage statistical advice we did not always take.

speakers from the far south of Asia are Sundadonts, not Sinodonts.[40] Pulleyblank (1996) has also found traces of AA vocabulary in two Shandong toponyms. Perhaps this reflects a wave of AA speakers; or, perhaps, the word represented by Zou 陬 and related graphs also travelled farther than its ancestors.[41]

What kinds of words reached the EAH? Well, first we need to point out that OC-AA comparanda don't come with directional guarantees. Presumably, someone borrowed *something*, but often we don't know in what direction, or even at what time, though we can guess. But, for example, several status terms for women — but not for men — have deep AA roots. This would fit with a picture that has aggressive "Chinese" speakers reaching out and grabbing "indigenous" women as they grabbed land. A classic scenario … so we get words like "wife" 婦,妻, and "geisha/courtesan"妓 (which means a "base woman" in old M-K).[42] Surprisingly, four of the five classic Sinitic "feudal ranks" 公侯子男 (but not 伯) made it from OC to AA, or vice versa. Here, it seems more intuitive to imagine the old Sinitic set of ranks percolating into very old AA (note that missing fifth rank!) than it does to envision AA speakers furnishing Chinese with the trappings of social stratification — but who knows for sure? (and what would "Chinese" mean back this far, anyway?)

But, most of all, AA words emerge as colors, place-names/geographical features, clan-political-religious terms, animal names, meteorological phenomena, and adverbs. Basic terms like pronouns, body parts, fundamental verbal actions, questions/negatives, and numbers (which

---

40 See Matsumura H, Oxenham MF, Nguyen LC, Nguyen KT, Nguyen KD, Huffer D, Dodo Y, Domett K, and Yamagata M., "Morphometric Affinity of the Late Neolithic Human Remains from Man Bac, Ninh Binh Province, Vietnam: Key Skeletons with Which to Debate the 'Two Layer' Hypothesis," *Anthropological Science*, 10.1537 (2008); also Nwe Nwe Aung et al., "Dental Traits among Five Tribes in Myanmar," *Journal of Oral Biosciences* Vol. 47 (2005) , No. 3, pp. 272–279.

41 Edwin Pulleyblank, "Zou and Lu and the Sinification of Shandong," in Philip J. Ivanhoe, ed., *Chinese Language, Thought, and Culture: Nivison and his Critics* (Chicago, IL, 1996), 43–44. For a classic, seminal study of Sinitic-AA relations, see Tsu-lin Mei and Jerry Norman, "The Austroasiatics in Ancient South China: Some Lexical Evidence," *Monumenta Serica* 32 (1972): 274–301. At a deeper time-level, ultimately, practically *all* inhabitants of the EAH came from Southeast Asia, as Victor Mair kindly reminds me.

42 ABC:297.

we mostly didn't even bother to list) rank lowest. A classic distribution reflecting significant interaction without genetic affiliation.[43]

What do our statistical results really tell us about the relevant Old Chinese texts? What significance do these results have? Do they reflect temporal distinctions? Spatial provenance? Or perhaps differentials in stratified language, where certain texts use more "formal," high-class vocabulary, where others use more words from a less prestigious substrate?

Discussion:

We anticipated a few possible principal component candidates:

1. *Time*. Offhand, we might expect a clear early-to-late pro-TB to pro-SE profile. Such a finding, while hardly exciting, would at least support a core-periphery model of early Sinitic development.
2. *Place*. We would have expected a clear west-to-east cline from predominantly TB to mostly SE vocabulary, a very useful and suggestive picture that would argue for a clear model of centripetal influence.
3. *Social register*. Perhaps we can perceive some substratum-superstratum relation; can we

43 This seems like the place to add two more pebbles to the pile of proof in favor of AA and/or AN connections for the coastal cultures (cf. ABC and the Shandong toponyms in Pulleyblank 1996:272–9). Consider the toponym 瑯琊 **langyaa** (attested in *Mengzi*, ca.-300), a central port for early coastal cultures (and, for a while, a Viet kingdom capital: see Eric Henry, "The Submerged History of Yuè," *Sino-Platonic Papers* 176 [2007], 10–11). It greatly resembles Old Khmer **lngyaa**: "sesame," a classic SE *wanderwort* (our reconstructions assume Old Sinitic regularly construed ancient AA sesquisyllables as binomes). Or consider the renowned ancient Go 吳 kingdom sword 鏌鋣 *maʔzyaa* (attested in *Zhuangzi*, ca.-300; see John Cikoski's reconstruction in his draft dictionary, *Notes for a Lexicon of Classical Chinese* [2008; http://www.gkarin.com/cikoski/]). *maʔzyaa* rather resembles old SE words for "serpent," like ancient AA *msyaň* or, even better Proto-Austronesian **b/uh/aja** (thanks to Paul Sidwell for pointing this out). Given the snaky patterns on the 鏌鋣 sword and the likelihood that an ancient SE swordsmith might seize on "toothed biter" as a name for his masterpiece (Viet kings — unlike ones from the EAH — bestowed titles upon swords; see Henry,19), it seems fair to consider these as further evidence for an early SE presence along the "China" coast. For our AA reconstructions, see Harry Shorto, *A Mon-Khmer Comparative Dictionary* (Australian National UP, 2006), Items #34, #937. Scholars interested in adding more pebbles to this pile may do well to consult the *Yuejue shu* 越絕書; Henry 2007.

detect evidence of diglossia, of "father tongue" overlying "mother tongue," or some similar linguistic stratification?[44]

Observe that we do not stress semantic categories as a likely principal component of differentia. You might have expected, for example, that a preponderance of TB or SE vocabulary would gather in texts preoccupied with agriculture, or sacrifice. But consider: the *Yijing* holds many words devoted to sacrifice and has a notably "SE" profile. But the Shang oracle bones and early Zhou inscriptions have a no less intense focus on sacrifice, yet exhibit markedly pro- TB leanings. Different sections of the *Songs* 詩經 focus on both agriculture and sacrifice, yet they display very different etymological profiles.

Examining our evidence, it helps to group our texts into some (non-exclusive) clusters. First, the "founder tongue" group:

商	.87
金文 1	.80
書經 1	.81
*詩 4 周頌	.62

These results strongly support the view that the "authentic" (Zhou 周誥) sections of *Shujing* 書經 belong with texts closely associated with the Zhou conquest. Most of our early texts look very "conservative," that is, pro-TB; but the *Yijing* 易經 makes a stunning exception. Additionally, most sections of the *Songs* 詩 (including, surprisingly, the "Greater Royal Odes"大雅) blur that picture considerably. *Zhuangzi* Chapter 29, perhaps our latest piece analyzed, has a pretty conservative profile, while the last years in *Zuozhuan* 左傳 prove significantly more conservative than the first (with only .016 likelihood of a chance result)!

Second, the "small-type" group, reflecting a limited vocabulary that seems fundamental and strongly pro-TB:

商	.87
周頌	.62
郭店	.67
老	.87

44 For the linguistic application of these terms to E. Asian linguistics, see esp. van Driem 2007:24.

墨子	1.20

The founders and "small-types" afford one very important observation; a pro-TB "Chinese language" extends back very far indeed. In both these groups — particularly in the first — who's the extreme outlier? The *Yijing* 易經, of course. That leads us to group 3, the "dead poet society," apparently reflecting the "SE warp effect" of formulaic verse composition.[45]

詩 1	.41
詩 2	.47
詩 3 大雅	.36
易經	.24
詩 5 商魯頌	.425
楚 1	.42

Now we can revisit a "diglossia thesis." Intriguingly, early texts associated with formal discourse to ancestors — the OBI and early bronze inscriptions, and only the "Zhou Ancestral Hymns"周頌 among *Songs* — all have very high, "conservative" profiles. By contrast, early *Songs* otherwise have surprisingly "low," pro-SE values. Just possibly, ancestor cults from the Shang and early Zhou kings may represent a linguistic superstratum imposed on a populace whose "mother tongue" surfaces robustly only in song. On the other hand, we must recognize that the "Shang-Lu Ancestral Hymns"商魯頌 do not fit this picture. Moreover, some later, presumably highly formal texts like the late *Shujing* 書經 chapters have anomalously "low" values. *Zhuangzi* and *Laozi* 老子, both chock-full of verse, also do not fit this "father language" versus "mother language" paradigm. Still, it remains suggestive; why does the presumably ancient *Yijing* have such a "low" profile? Perhaps because it represents the vocabulary of early diviners, plausibly a group that would use less formal speech. Moreover, a substratum-superstratum dynamic might help explain the puzzlingly low profile of the "Greater Royal Odes";

45 The *Yijing* 易經 also contains verse, much of it closely related to Songs verse. For a classic demonstration, see Edward Shaughnessy, "Marriage, Divorce and Revolution: Reading Between the Lines of the Book of Changes," collected in *Before Confucius: Studies in the Creation of the Chinese Classics* (Albany, NY: SUNY Press, 1997),13–30. Recall that the 易經 hexagrams rhyme extensively.

as those Zhou conquerors spread east — "Huai River aliens come seek us out 淮夷來求" and so forth — they would have picked up many SE words from their captives and booty. All in all, it seems fair to consider substratum-superstratum relations as a relevant issue, but not one that leads to a clear and compelling picture of our data. After all, superstrate relations usually describe an expansion of vocabulary (Norman on top of Anglo-Saxon, for example), while our data show an early pro-TB vocabulary, with SE vocabulary gradually emerging later on. Not a classic "father tongue" on top of "mother tongue" scenario, though here we must wonder about quality of evidence.

Observe that we do not necessarily consider the *Chuci* a case of direct Southern influence. Our previous rhyming study demonstrated the incoherence of a "Chu rhyme" category and the early *Chuci*'s fantastically close adherence to *Songs* categories.[46] Rather, we suspect that, again, the *Nine Songs* 九歌 and *Lisao* 離騷 simply follow their poetic models. But we must mention two important examples countering this pro-SE "verse society": *Laozi* and, to some extent, *Zhuangzi*, feature lots of verse, but little SE vocabulary. Of course, you can't, precisely, call these "poems"詩. These early "pro-SE" texts suggest that SE words surface en masse in the Old Sinitic vocabulary all during the W. Zhou and Springs-Autumns 春秋 periods; too bad our data do not let us get more precise. Neither do they illuminate whether this vocabulary got imported, or whether it sprang from some unknown substrate.

Fourth, a "Tory" group, consisting of texts associated with the Shang successor-state Song 宋 (above all — but also with Qin, at least) and also, to some extent, Lu texts, which form a more moderate "Orthodox" subset:

46 david McCraw, *Stratifying Zhuangzi: Rhyme and Other Quantitative Evidence* (Taibei: Academica Sinica, 2010), Chapter one.

*郭店	.67
*老	.87
*左 2	.78
墨子	1.20
莊子 1	.78
莊子 29	.69
*論語	.62
孟	.63
荀	.66
商君	.825
韓非	.69

This offers a good opportunity to review factors of place. Instead of our anticipated west-to-east pro-SE cline, we find a much murkier picture. True, early texts associated with the Western Zhou founding and consolidation (early *Shujing* chapters, Part One from our Bronze Inscription corpus, the "Zhou Ancestral Hymns") cluster gratifyingly around high pro-TB values. But Shang OBI, a more "eastern" corpus, looks even more conservative. Early *Mozi* 墨子 chapters, allegedly from the Shang successor-state of Song 宋, have the most conservative profile; *Zhuangzi* [Part]1, also supposedly from Song, has also a surprisingly high "pro-TB" rating. The *Mozi* finding surprises us most; given a likelihood that it comes from a lower-class origin, we would have expected a "substrate" phenomenon. Perhaps Mohists strove to emulate "higher-class" discourse? Observe also that early Mohist chapters feature relatively "archaic" deictic features like *zi* 玆(5 times) and *zhi* 之(18 times), quite like *Zhuangzi*.[47] It appears that Song conservatism outweighed other plausible social features. Perhaps we really can ascribe some validity to traditional notions that Song retained many antique features from its Shang forbears! On the other hand, geographical proximity to "coastal cultures" in Shandong would lead us to expect very "low," pro-SE values for Lu texts. Yet our core *Analects* and *Mengzi*

47 *Zhuangzi* features deictic 之 9 times, 4 in Chapter.1; pronominal 斯 occurs 9 times, 2 in Chapter 1). See the not-always-reliable statistics in Redoaune Djamouri, "Pronouns demonstratifs en Chinois Bas-archaiques," R. Djamouri, ed., *Collected Essays in Ancient Chinese Grammar* (l'École des hautes études en sciences sociales: Paris, 2001), 161–175. We have, of course, rechecked and retabulated his numbers.

passages weigh in at somewhat higher-than-average values, while early *Liji* 禮記 selections, usually pegged as "Lu texts,"[48] strike exactly dead-center median in late WS orthodox vocabulary profiles. Apparently, we can't wax too dogmatic about the importance of geographic origin. One might speculate that "marginal" eastern writers felt a particular need to hew to a "conservative" vocabulary; after all, some think Kongzi came from a Yi 夷 background![49] On the other hand, by late WS times one could also think of Lu as a conservative bastion. This issue seems less than clear-cut; geographic provenance does not emerge as compellingly as we would have imagined.[50] Observe that *Xunzi* 荀子 and *Hanfeizi* 韓非子, neither from Lu-Song (though note Xunzi's longstanding connection with Qi 齊), form outsiders here.

Finally, we come to late pro-SE "Whigs," who comprise something of a "leftover" category:

書 2	.41
金文 2	.53
*左 1	.49
左 mid	.49
墨子 2	.44
呂氏春秋	.37

Though most apparently belong to a "later Northern formal school" of writing, this

48 See Wang Qiming 汪啓明, *Xianqin Lianghan Qiyu yanjiu* 先秦兩漢齊語研究 (成都, 2001).

49 Robert Eno, "The Background of the Kong Family of Lu and the Origins of Ruism," *Early China* 28 (2003), 1–41. Cf. Wang Li et al., "Genetic Structure of a 2,500-Year-Old Human Population in China and Its Spatiotemporal Changes,"*Molecular Biology and Evolution* 17 (2000), 1396–1400. Wang and co. found that people living in Shandong 2500 years ago had notably different mitochondrial DNA profiles than more recent inhabitants. Unfortunately, Wang and co. did not include Southeastern populations in their baselines, leading them to wild speculations about "Europoid" populations in early Shandong. What would they have found if they looked southeasterly? Thanks to the indefatigable Wolfgang Behr for calling my attention to this article.

50 Frankly, our preconception had envisioned something like the Eastern Coastal Barbarian-Western [proto-Tibetan?] Barbarian 東夷西戎 thesis adumbrated by early scholars; see, for example, Fu Sinian's famous *Yi Xia dong xi shuo* 夷夏東西說, in 傅斯年全集 (7V: Taibei, 1980) 3.86–157. But cf. our ensuing discussion….

groups proves more innovative than the "Tories," with their Song-Lu connections. Perhaps we could imagine a geographical division of textual lineages, rather reminiscent of the "Six States'" writing systems.[51] To test this, we further examined texts like *Xunzi, Han Fei*, and the *Shangjun shu.* According to last-minute data, *Xunzi*, *Han Fei*, and *Shangjun* 商君 all look like Tories, complicating our picture of any "late northern "Whig" cabal. Considering the *Shangjun*'s striking pro-TB profile, we must seriously consider whether Qin does not constitute another "conservative" stronghold (Of course, what's left, after Qin, Lu-Song, etc., have gone? Only "leftovers," sequels, and late anthologies). For now we will rest for a bit from our search labors, content to digest and reexamine these intriguing data afforded by our etymological word processor.

We would not want to imply that our gizmo has fully matured. For example, the current state of our data, the gizmo, and its operator forbid a rigorous statistical explanation of "principal components." We can, however, offer a purely descriptive "principal component analysis." Time, for example, plays a role, though less a role than anticipated. Let's call it a 1/3 factor. Place and semantic category, surprisingly, prove nearly non-starters, explaining scarcely more than 5% of variance. Social register plays the leading role, and we would estimate its importance at least above 50%. This does not, however, say as much as it might seem. Levels of social difference can involve many subcategories, and a rigorous analysis would want to distinguish the relative importance of formal vs. informal discourse, superstratum vs. substratum relations, court vs. "folk" features, and so on. At the moment, our device lacks the sophistication to separate these diverse factors.

Observe that even an immature gizmo can help illuminate a few more dark corners. For example, a long controversy has simmered over the OBI: do they faithfully reflect some variety of Old Sinitic, or could the Shang have used them to write a different language?[52] The narrow

---

51 Magisterially described in Qiu Xigui 裘錫圭, [Transl. Gilbert L. Mattos and Jerry Norman] *Chinese Writing* (Early China Special Monograph Series No. 4. Berkeley: The Society for the Study of Early China and the Institute of East Asian Studies, University of California, Berkeley, 2000), pp.78–89.

52 Professional Sinologists like Boltz, Keightley, and Pulleyblank argue the former. See, for example, Edwin Pulleyblank "Early Contacts between Indo-Europeans and Chinese," *International Review of Chinese Linguistics* 1.1

scope, high percentage of unknown graphs, and low percentage of secure phonetic elements have allowed room for doubt. But our data strongly suggest the Shang and early Zhou shared a common language. Why? Well, if the Zhou had simply borrowed a Shang writing system to write a different language, it strains credibility to imagine they would come up with almost the same vocabulary profile. If a typical American, for example, had to borrow an Arabic writing system to write, he would likely include fewer words like "mosque," "muezzin," "mullah," "date-palm," and so forth, and include more words like "hot dog," "pick-up," "Christian," etc. But the profile-match between Shang and Zhou inscriptions makes a random similarity a less than 1% chance; we can safely guess, based on current data, that the language Shang kings — or at least their scribes — transcribed into OBI largely matched that of early Western Zhou inscriptions.[53]

Another consideration: People often argue about *Zuo zhuan* — a stratified text? One largely composed by some historian? Consider that its first 2,100 graphs and nearly 3,000 pulled from its midsection (Xiang 10–12) have *exactly* the same profile. Considering how different the profiles from even adjacent parts of text have looked, it seems obvious that this result could hardly have arisen randomly. Perhaps they *do* come from the same brush!? Of course, then we must confront *Zuo zhuan*'s final 3,000+ graphs, which have a much more conservative (pro-TB) profile. How to explain this? Well, a different brush, for sure, and possibly a composer/compiler from a more conservative neck of the Old Chinese woods. Song, for example? Alternatively, one could imagine our hypothetical compiler adopting a conservative style to impart a "loftily antique"高古 modal effect. Clearly, we cannot assume too much, not without more intensive examination.

A bottom line: we will need to examine the entrails of more texts before we can hope to

---

(1996):1–24 (an argument from intelligent design — vulnerable to counterarguments based on exaptation). Outsiders like Paul Benedict and George van Driem (2007:36–7) often raise questions, based on inherent improbability (vulnerable to deep-time considerations, when the improbable can become quite possible).

53 On linguistic variety within Shang OBI, see esp. Takashima Ken-ichi and Anne Y. Hashimoto [Yu Aiqin], "Evidence of Possible Dialect Mixture in Oracle-Bone Inscriptions," in Ding Bangxin, Yu Aiqin, eds., *Yuyan bianhua yu Hanyu fangyan* 語言變化與漢語方言(Li Fanggui xiansheng jinian lunwenji: Taibei, 2000): 1–52., esp. pp. 20–23. Axel Schuessler (pc: 4/20/10) remarks that OBI display many of the same loan-graphs as later Zhou inscriptions and texts, a very strong corroboration of close linguistic similarity.

divine some of our gizmo’s oracular pronouncements. These preliminary explorations have uncovered some suggestive textual relationships, but our gizmo must — at the very least — await corrections and further improvements from Sinology’s basement tinkerers. Let’s face it; our useful gizmo has uncovered some marvelous buried data, but what exactly do they mean? We haven’t found anything like a “Danelaw” line, either in space or time. The observed mix of TB and SE vocabulary must have undergone some long and complex “stirring.” This supports notions of emerging Sinitic as a “Mischsprache.”[54] To adopt a “Chinese” metaphor, ancient varieties of Sinitic look more and more like “stir-fries” or, less anachronistically, different cauldrons simmering with various styles of millet stew. Their TB, SE, and sundry lexical elements got thrown in the pot, with hints and soupçons of words whose origins now escape us. The relatively crude distinctions afforded by our gizmo cannot fully satisfy, but they can encourage more sophisticated mechanisms and — one dares to hope — savvier connoisseurs who can appreciate these stew-pot “flavors” more keenly.

We end with an irony; we designed the gizmo to analyze ancient texts and inscriptions. It's telling us something important about them, but our existing models can't interpret the message competently. Data outstrips hypothesis, as when an ancient inscription remains only half-deciphered. Sinology awaits the more sophisticated analyses and new hypotheses that can fully decode our gizmo's cryptic call.

---

54 See Wolfgang Behr (with Robert Gassmann), *Antikchinesisch*: *Ein Lehrbuch in drei Teilen* (3V; Bern: Peter Lang, 2005), 3.387–8. The idea that Sinitic resulted from the combination of elements drawn from two or more language groups has long been propounded by Victor H. Mair. See, inter alia, "Language and Script, pp. 19–57 (esp. pp. 21–23) in Mair, ed., *The Columbia History of Chinese Literature* (New York and London: Columbia University Press, 2001).

Appendix 1

Here it seems meet to address another potential bug in our gizmo involving considerations that would have seemed arcane earlier on. Clearly, vocabulary expands (roughly) with time, so that later texts mostly exhibit a higher ratio of "types to tokens" than earlier ones.[55] But could different "type-token" ratios among later texts introduce an uncontrolled experimental artifact that skews our data? Here we shall begin a preliminary examination based, for convenience, on some whole-text statistics. Ideally, we would have precise numbers for each text-chunk (we plan to install a concordance program on a later version of the gizmo; cf our previous chart, which *does* provide targeted vocabulary numbers for each piece).

Texts	Types	Tokens	t-T ratio	Pro-TB ratio
書	1,627	25,700	15.80	(1) .81
易經	868	4,161	4.74	.24
*詩	2,826	39,200	13.87	(1).41
*論語	~1,400	~24,000	17.14	.62
*左	3,290	~176,000	53.5	(1).49
墨子	2,609	117,480	45.03	(1)1.20
*老	720	5,060	7.03	.87
孟	1,913	35,417	18.51	.63
莊子	2,933	64,926 (Mc)	22.13	(1).78
荀子	2,835	78,815	27.80	.66
商君				.825
韓非	2,754	107,128	38.90	.69
*楚 1	2,376	28,057	11.81	(1).42
*禮記	3,628	97,973	27.00	.56
呂氏春秋	3,115	101,411	32.56	.37

On the surface, we can't discern any obvious bias associated with large or small t-T ratios. Referring back to the "total hits" figures in our main chart (p. 6, column 9), we see that the only

55 Of course, type-token ratios refer to words, while our gizmo's chewing Sinographs. For the moment, we shall assume the ratio of disyllabic and polysyllabic words in different late WS texts does not differ enough to invalidate using the phrase. We await demonstrations to the contrary; meanwhile, doubters can take comfort in our "scare-quotes."

salient patterns involve low-vocabulary with highly pro-TB profiles (only the 詩 1 feature a low-vocabulary/low profile combination); perusal of our "Tories" and "innovators" shows that higher vocabulary figures have no necessary correlation with lower profiles. Since we consider our "small-type" class a fundamental category rather than an experimental bias (note that these feature limited *vocabulary*, *not* necessarily small *samples*), we shall assume (for the moment, pending further investigation) it does not matter much.

Appendix 2:

Gizmo innards. It consists of:

a. a souped-up 200+-word Swadesh wordlist (245 English words) — see attachment below;
b. a 1,100 Sinograph searchlist, generated by matching a) with ABC's English–Chinese index — see attachment below;
c. the handchecked results generated by searching a) with b), via Word wild-card searches (results exist only in my study). We have tabulated those results in the main chart above.
d. coverage. To get exact:

Text Sinographs	parts included	
商=	甲骨文合集 *Chant* PDG files 001 to 0015 (董 Period I)	~2500
金文 1=	6 西周金文 inscriptions, from Ed Shaughnessy	2848
書經 1=	洛誥, 酒誥, 康誥, 大誥	4673
*詩 4 周頌=	詩經 #266–296	1687
書 2=	禹貢, 洪範	2912
易經=	卦 1–32	2732
*詩 3 大雅=	詩經 #235–245	1972
*詩 5 商魯頌=	詩經 #297–304	2011
*詩 1=	詩經 #1–25	1650
*詩 2=	詩經 #26–64	3731
金文 2=	4 春秋 and 1 戰國 inscription, from Ed Shaughnessy	1350
*郭店=	complete	1932
*老=	all 篇 1–81, *except for* Chapters included in 郭店	1929
*楚 1=	九歌離騷	2005
*論語=	Chapters 2–7	3660
*左 1=	隱公, first 2027 字	2027
左 mid=	襄公 10–12	2945
*左 2=	哀公, last 1991 字	1991
墨子=	尙賢,尙同,兼愛 (上)	2583
墨子 2=	魯問, 公輸	3642
*莊子 1=	逍遙遊 (except its excrescent ending), first half of 齊物論	3155
*禮記 1=	曲禮, first 2469 字	2469
荀子	Ch.17, 1st part of Ch.21	2658
商君	Ch.1–3	3945
韓非	Ch.1,12	2825
*孟=	Chapter 1A, first 2447 字	2447
*莊子 29=	complete	3100
*呂氏春秋=	(random selection) 慎行論 p.1515>貴直論 p.1577	3747

Appendix 2: our wordlist, with English vocabulary (left column) and Chinese rough equivalents (arranged on a continuum from rarest TB words to rarest SE words at the right).

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
I		吾我			予余		
you		汝如而爾					
s/he			厥	其	渠		
this		是時		此伊			
that		夫彼				爾	
here							
there							
who		誰孰					
what		何奚				某	
where				惡安			
when							
how							
not		無			不		
all		皆盡俱舉凡悉			僉	咸兼	
many		諸庶衆		烝		多儕數	林
some[one]			有或			某	
few		寡		鮮尠	少	幾	
other		殊				他	
on[c]e		一				每	
paired		二雙			兩		
begin[ning]		元先端	作	始初			
end[ing]		後終		末已屆			
big	毗朏荒芒厖	大阜巨膴博		宏洪世			
small	瑣[剉]487	細		小渺藐眇			
long			曼長	永延修悠			
short		短					
wide		侈	扈	廣	博溥		
narrow		扁		厄隘		狹	
thick		敦篤		厚			
thin		薄細纖					

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
glad/ joy			樂	喜欣			
angry		撓惱疾堲	慍	懠懥憤			
sad/grieve				悲哀	憂	愁	騷慅踧
red	霞赮赬赭	赤	倩	紫	紅彤	丹朱	黮
green/blue		蒼青		綠		藍	
yellow		黃					
white	皤	白				皓素縞	暵鶴
black	慝	黑幽墨		玄緇	盧	黔	黮紺
empty		空沖	虛		唐		
full		盈	飫		厭	湛添	
new		新鮮					
old		古耆	老	陳		舊	
good		藏嘉	好	善			
bad	慝	惡					
rotten		腐		陳			
dirty		髒		妝			
clean		濯		淨			
straight		値挺亭	正	尹頡		縮	
round		桓	團	囘圓		環	
sharp	略剡	利厲廉鋒					鑱
dull		鈍					
smooth		滑					
wet		濕					
dry	歖	乾涸灘	暴			曬	
correct			正當	義			
near		邇匿				近	
far		彼	悠	遠	桌趠	迥	逖洵坰浻迥浻
high	磊峈嵩崧	陵危岩巖	嵬巍	高喬憍騎厲		隆	崇卬昂
low		低			卑		
middle		中				央	
right			右				
left						左	
lucky [bless]		福吉嘉	慶	祥瑞	祈禱祝		
unlucky [curse]		災	凶	禍咎			
woman		婆	女			妓婦	

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
male		夫		男			
person		民				人	
child		孺子				兒	
wife		嬪	女			婦妻	
husband		夫					
animal		物畜	獸				
fish		魚			鯉		
bird		鳥	隹			禽	
dog		犬				狗	獀
sheep	羭	羯羖	犢	羊墳			
pig	豝	豭	豬			豗豚豕	彖
tiger						虎	於兔騶虞
dragon						龍	
phoenix				鳳			
turtle						龜	
louse		蟣虱					
monkey	[寓]	猴		猿猶			猱蜼
worm		蟲蜎			蚯蚓螾		
tree		薪	木		樹		
root		本				根	
forest		叢		森林		麓	
stick		梃	干	杖			
fruit				果實			
orange						橘	
peach						桃	
plum				李		梅	
pine…	蒸			柏		松	
seed		種子					
leaf	箬	葉					
bark	[�currency]	皮					
flower		葩瓣	榮	英	花華		
grass	荊	莽	蘇薦	草		芻	
body		體躬形		身尸			
skin	[禚]	皮鞹		膚			
meat	隋]膴	臛	脢		肉		
blood		血衊			脅		盍
bone			骨弧				

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
fat	臊	雋		膏脂肪	脅		
egg		卵					
horn	觡	角			觤		
tail		尾					
feather			侯	翮羽			
hair	剃(Li221)	髮毛 [彪]*					丣[彪]*
head		元首			頭		
eye		目			眼		
ear		耳				朵	
mouth		口	嘴				
nose		鼻				喙	
tongue			舌				
tooth			齒			牙	
finger		指		丑			
hand		把		手			
foot/stop		止		腳	足		骹
leg		脛	腓			股	
knee		節鞠	膝				
wing	豳	翼	羽	翮			
belly	[蚹] 漢莊子		腹	肚胃			
gut		腸					
neck	脰	項領頸		脖頑[ ]			
back		旅				背	
breast		乳	膺		胸		
heart		仁		心			
liver		肝				膽	
excretion		屎				尿	
drink		侑飲					
eat		食參	啜餔哺饗	吃享饗		啖飯茹吞	
bite		咬噬		呷含		啀	
suck		漱吮	噆				
spit		吐噴	唾				
vomit		吐		嘔			[呎]548
blow		弗				吹	擤
breathe		呼吸息			噓吁呴		
laugh	顟唲	矧哂		笑猶			

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
see	冒	見覽		睹		瞻	覘佔
hear	耳			聞聽			
smell		腥香馨		聞聽臭嗅			
think		義議思惟	意憶想	念			
know		見悉	識			曉	白
fear	蘇索	赫		憱忪聳悚兇畏懼	懍	慴驚懾	
sleep		臥綏寢眠寐		睡			
sick/hurt		疾			病	傷癘厲恙	
live		生蒼				活	
die		死	崩	薨			
kill	戡	劉殊戮	殺	祖殪弒	貪惏		
fight		鬨爭戰伐		鬨			
hunt		田狩		獵			
hit	搭	觸搗撞		擊扣考中			
cut		刺切斫辟	析	剃刻刲剔剜刓	砍	割	劃釗
split	扐	離破別坼劈理披				擘	副
stab		刺		刲		戳	擉
scratch		搔					
knife/ sword					刀	劍	
bow/ arrow		矢射		箭		弓	
axe		斧					鉞戊
spear				戈矛			
drum	咢			鞉		鼓	
flute	筵			簫笛簫[葥]			
bell		甬		鐘鈴南			

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
				鐃			
zither				琴瑟			
rope		繩率	縰纚紉縶			索	纆
dig		突		挖		掘	
come	儀假迨	至來格蒞		戾到	逮	迪	
walk	假踖尢			邁	行		
sit			坐				
lie		伏臥					
stand		立起定			待樹		
turn		轉般鈞還桓	反	回			
fall		落廢隕蔽跌		顛躓	委	墜	零
give	畀	奏奉乞購稅授賦	承蒸賞賜舍施	獻休	與予	捨	
get		奉受稟	承	得			
hold		秉容將執挾					
squeeze		撚					
rub	卹	揉					磋
wash	盥沃	濯	浴		釋	澡漱	洮
wipe	[謐]0						
pull		曳拙	引	臾	牽		
push		搗擠	推				
carry	舁	荷擧任		保載		負背擔	
throw		投	播	擲抛			
tie		結		繫			
sew				縫			
jump				跳踰		踊	
lead		將帶領率先		迪		牽	
follow	遹	隨從遂順後		由	沿緣送		
flee		奔脫				洮遁	
return		反復		回		歸	
count		量歷	算			數	
say		言語談譚話曰	云	謂	道		

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
sing		唱歌			和		
play		戲					
fly	翻豳弁			飛翔			
swim		游				泳	
float		汎氾流游颺		飄		浮	
flow		流游		淳漂泫湲	汩		
freeze	硬			沍淩冽		冰	
swell		張		封阜墳			
sun/day		日					
moon		月		魄			娥朏
star		星辰					
rainbow						虹	蝀蝀
rain		雨				零	凍
night		夜夕默				冥	
day		日晝					
dawn	霞	旦		煦辰旭		朝亮昧	
year		年	載	歲期			稔
warm	燠尋燄	暖溫炎		煦			
cold		冷		淒涼清寒凔			
Spring						春	
Summer					夏		
Fall				秋			
Winter		冬					
cloud					氛雲		
fog/dew		霧		煙		[露]	
sky				粉天			
wind	弗			風嵐颯		颸	
snow						粉雪	
ice	剛強					冰	澤
water				水			
river		川		河		江漢汝	溪
lake			池	湖			塘潭
sea						海	夷,蓬萊
salt	[嵯]	鹽				鹵	鹽
stone		磊			墝磽	石	

Engl	+3 (rare)	+2 (S-T)	+1(ST!?)	0 (?)	-1 (Aa!?)	-2(Aa…)	-3(rare)
sand	[磨]chu			沙			
dust	[磨]	炭塵		粉			
earth	畬				土	地	壤
smoke		熏		煙			
fire	燬	火熱	炎			烙	
ash		灰					
burn		焚熱然燃焦型煇燔	炎				
road		徑			道路略	行	
mountain		崗岑陵墟		山		嶽岳墩丘阜	嶠
wilds				野			
at/in		于		於		即	
with		與以					
and		與及				即	
if		如若				使	
because					爲		
name		名	姓				
house/home		家室宗				宮屋公	廛宅廟
grave		冢喪				墳墓	
spirit		神宗帝天		祖			靈
demon		魑		魅	鬼		
sacrifice	侑禘雩	用祐	殺	祭祀* 示供登榮	社祈禱祝	禳	禜歃*
ancestor		宗帝妣		祖			
wall		城里堲	屏	牆垣			
city		城里邑			都*	村邨市	
lord		后		王禮		公侯伯男	
Alien (n)						狄	
Alien (w)				戎			
Alien (e)						夷	
Alien (s)				蠻			
subtotal	80	378	84	218	73	142	67
T=+522	(+1080)						(-558)

Thus, in total, we have 1043 Old Sinitic words. Their TB subtotal weighted bias amounts to +1080, and their SE subtotal weighted bias amounts to -558. Adding these and dividing by the total gives you the average "TB pull" for any word in the list: 522/1043 (total Chinese words in chart) = .50.

Since June 2006, all new issues of *Sino-Platonic Papers* have been published electronically on the Web and are accessible to readers at no charge. Back issues are also being released periodically in e-editions, also free. For a complete catalog of *Sino-Platonic Papers*, with links to free issues, visit the *SPP* Web site.

www.sino-platonic.org