

LETTER TO THE EDITOR

IMPLICATIONS OF SOCIAL SCIENCE FOR VOCABULARY ARCHITECTURE: WILKINS VS. WORDTREE

Some three centuries ago, during Western Europe's Age of Exploration, there was published a very original and mammoth catalog of "the nature of things." It had been compiled by a Royal Society founder, John Wilkins. It was his pathbreaking *Essay Towards . . . Language* (1668). The essay boldly attempted to specify all variants under each category of thing. For example, hundreds of "Manners" were listed, such as Modesty and Magnanimity (pp. 206-213).

Wilkins's *Essay* is echoed in many later wordbooks (e.g., Roget's *Thesaurus*) and even encyclopedias, although rarely acknowledged.

His arrangement tried to point from a concept toward its best term. That format is now sometimes called a "nomenclator." But in the three subsequent centuries, the more common format has been its inverse, pointing from a word or term to its meaning. The obvious example of such a "semantic" is the alphabetical dictionary. Indeed, little progress has been made in nomenclators since Wilkins.

But today's computerizing world increasingly faces the problem of extracting experience-based wisdom from each apparently novel circumstance. Therefore today's need, especially in the emergent field of artificial intelligence, is far more for nomenclatorial systems than for merely semantic systems.

One such scheme was recently published under the name of *The Wordtree* (Burger 1984). The present author, its editor, was therefore most gratified to find it compared with Wilkins's *Essay* over many paragraphs in this journal by Michael Lesk (1987).

On analysis, however, I find that it contains many basic assumptions of computerization practitioners, but not of social science practitioners (sometimes called "telesizers"). Indeed, Lesk's critique can provide an object lesson in the gap, the chasm, between some computerizers and some telesizers.

The present author has been a computer user since about 1950 and has a doctorate in cultural anthropology.

Editor's note: This letter is in response to a review of *The Wordtree* in Issue 13-1,2 and should have appeared in the following issue, 13-3,4. It was unfortunately omitted from that issue and we only found out about the omission recently.

It would seem valuable to indicate the differences, for they may well be part of computational linguistics' unclaimed turf.

The approach of computational linguistics (CL) to word architecture might, at the risk of oversimplification, be termed mathematical. That of the social scientists may be termed evolutionary and cultural. Anthropology, for instance, must have an inherent interest in CL: symboling is the principal human distinctiveness, and anthropology concentrates human distinctions. Linguistics, a consequence of symboling, is usually considered one of the five branches of anthropology.

WILKINS'S SUBSTANTIVE-PROCESSUAL INTERMINGLING VS. WORDTREE'S BOHRISM

One principal difference concerns the relation of concepts about substances to concepts about processes. Wilkins assumed that "a *Verb* . . . ought to have no distinct place amongst Integrals [=principal words] in a Philosophical [=ideal] Grammar, because it is really no other than [=than] an Adjective . . ." (1668:303). Hence Wilkins constantly intermingled structural adjectives and nouns with processual verbs. Thus, his pages 253-254 leaped from procedures like *encouraging*, *comforting*, and *defending* to substances like *grange*, *fruit tree*, and *tame beast*. Reviewer Lesk comments that, by contrast, the *Wordtree* arrangement "clashes . . . [with] countries, chemical elements, and so on."

True. As *The Wordtree* declares in many places, such as page 28, "we extend [to vocabulary architecture,] Niels Bohr's theory of complementarity: Light is both material (e.g., particle) and process (e.g., wave motion). . . . But physicists find that they can measure little unless they emphasize one or the other analysis. . . ."

Bohr's Nobel-winning concept is central to modern science. It led to the realization that if one specifies the exact location of an object, it must be at rest (= Werner Heisenberg's Principle). And that has produced the system of quantum mechanics.

But this crucial scientific bifurcation does not seem to have penetrated linguistics. Non-*Wordtree* wordbooks routinely intermingle substances (typically, the noun, which the British tradition insightfully terms "substantive" rather than "noun") and process (typically, verbs).

Copyright 1988 by the Association for Computational Linguistics. Permission to copy without fee all or part of this material is granted provided that the copies are not made for direct commercial advantage and the CL reference and this copyright notice are included on the first page. To copy otherwise, or to republish, requires a fee and/or specific permission.

0362-613X/88/0100●-●\$03.00

By contrast, *The Wordtree* has applied to vocabulary, probably for the first time, those principles of hard science first enunciated around the 1910s. *The Wordtree* consequently argued that a substance may ultimately be defined only in terms of other substances, and a process only via other processes. Therefore we endeavored to gradate the entire language's processes purely in terms of a hierarchy of processes.

Lesk went on to test-use *The Wordtree* "for an alternative to [the word] to REP-RESENT. . . .," and found it slower than Roget. Of course. As our page 24 declared, "We do not claim that 'The Wordtree' is a guide to *common* speech. To *describe* the world, use traditional word books. But to maneuver or *change* the world, use 'The Wordtree'!" And to find an alternative for a word is the static task of description rather than change. For that job, Lesk was sound in preferring a thesaurus.

Sometime, however, he may want to find causes, preventives, or effects of a concept; then he should use a branching nomenclator. For instance, "The Wordtree" lists one *tabooing cause* as *sectarianizing*. In lay terms, that means that for one part of the population to be kept off-limits to a certain site or practice, a culture must have been divided into coteries or sects. We do not think that a semantic format can show any such conceptual relationship.

WILKINS'S DISCRETENESS VS. WORDTREE'S CONNECTEDNESS

Another chasm between a computerizer's acceptance of his *Essay* and a brancher concerns conceptual boundaries. Wilkins perceived each of his (more-or-less words) as being discrete. And it sometimes seems as if CL'ers likewise take each word at its face value, rather than considering the unit transformable into neighboring idea possibilities.

By contrast, a branching word system assumes the interrelatedness and indeed the lability of concepts. Our nomenclator does so via geological, biological, and cultural evolution. For example, a vegetable that is motile is, in a generic sense, a kind of bacterium. And so, in *The Wordtree's* language of binary transitive verbs: To *vegetalize* something and to *motorize* it = to *bacterize* it.

But Wilkins was true to his age and church, and implied substantive perpetuity. In fact, his linguistic essay explained how diverse absolute creations could have survived: The good bishop devoted seven pages (162-168) to showing geometrically how Noah's Ark probably accommodated everything from polecats to dung!

How seriously, then, can we take a CL critique that "would rather have [=follow] the older book," because the newer branching system "clashes . . . with traditional and familiar arrangements"?!?

WILKINS'S VOCABULISTIC RESTRAINT VS. WORDTREE'S ETHNOGRAPHY

Wilkins's goal was codification, not heuristics (discovery procedures). Where a genus required uncertain terms, he shyly asterisked the "several words . . . which are yet very questionable" (1668:Reader's Preface).

By contrast, our brancher argued that only a small part of (legitimate) vocabulary has yet been recorded, even in the largest unabridged wordbooks. Processual complexification systems have not been performed previously for an entire language. Therefore our first edition suffers many conceptual gaps, especially in areas like biomechanics. We frequently concede that fact (as on p. 45). The situation resembles the first describer of the color spectrum: he probably could offer only the names of a few basic colors. But gradually the names were found, or invented, for the hundreds of intermediate hues, such as purple.

Our branching system seeks to report those hitherto uncoded names, particularly from the technical lexicon, or *technolect*. We obtained most of them by the anthropological method of "reporting what the natives say." Hence we were able to list an exact site for each of the 24,600 transitives there distinguished. And that first approach represented 30% more transitive morphemes than in the world's hitherto largest dictionary, the *Oxford Unabridged*.

Typically, Lesk disliked to *vorlauf*. But we had evidenced that term to page 207 of a popular-speech reporting journal: One *vorlaufs* (a ski course) to establish its time and conditions. We believe that its very incorporation into the process system of English will henceforth interdigitate *vorlauf* with the norm-settings of inanimate matters, such as to *calibrate*.

Likewise, to *Zip-Code-sequence* (a mailing) was not our invention, but was noted as detected on page 3 of a 1982 official U.S. Postal Service bulletin. We think that revealing it will fill a semantic gap with the many processes of *geocoding* that are now emergent in our era of satellites.

In fact, we specified a source for 100% of our transitives, whether printed or broadcast, by page or by minute. Such documentary thoroughness is an unheard-of rarity in the field of lexicography. These are not, then, "rare or made-up words." They are the sounds of the real world.

The problem of their strangeness is that lexicography has hitherto been dominated by litterateurs gently raised on what field anthropologists sarcastically call the "verandah" of an ivory tower.

WILKINS'S UNILINEARITY VS. WORDTREE'S MULTILINEARITY

Another misunderstanding between the CL use of traditional word arrangements and the social science tradition concerns the path of complexification. Semantic word systems are straight-line. The dictionary moves

from A through Z. And a synonymy, such as Roget's, moves from group 1 through group 1,000. But the world is not merely Aristotelian numerics. Darwin showed over a century ago that complexification proceeds multilinearly (by progenerative branching, "cladistically"). Chimpanzees are our cousins, not our ancestors.

The selection of an evolutionary path is multifactorial. An ecological space, or *lebensraum*, must be vacant. An organic group must be present at its border. That organism must have a supply of appropriate mutations. In sum, "the course of [biological] evolution follows opportunity rather than plan . . .," explained Simpson (1960:160). "Changes occur as they may and not as would be hypothetically best."

That simple observation means that the environment necessarily contains as many factors (and almost surely many times over) as ever can be entered, however automatically, into any computer, however speedy. Computers can, then, never surely predict organic interactions.

But the concept of Darwinism was not present in Wilkins's time. Indeed, *Origin of Species* was not even published until seven years after Roget. Thus, pre-Darwinian hierarchies are mere conglomerates. In that way, Wilkins listed mere "magnitudes" (II:VII) as more complex than "viviparous animals" (II:V). And Roget listed such powered-system characteristics as "excitability (#825) as far more advanced than such human-only characteristics as "book" (#593).

By contrast, using late 20th-century evolutionary concepts, *The Wordtree* shows the gradual complexification from the most primitive terms, such as spatializing. It covered the stages between mere atomization and vegetation by incorporating modern general systems theory. Thus, Marney and Smith (1964:124-127) began with radiant energy. When supplemented with nuclear binding, some of it became nuclei. With atomic binding, and chemical valence, some became molecules. With self-replication, some became polymers. And so forth.

Animality, for example, is currently believed to have required the very basic biogramming of *bipolarizing*, *bacterizing*, *vegetalizing*, etc. And the knowledge explosion in genetic engineering will soon doubtlessly name dozens of intermediate procedures.

WILKINS'S FINITENESS VS. WORDTREE'S OPEN-ENDEDNESS

The static world of 1668 is reflected in Wilkins's limited concept of vocabulary: "There should be little need of other impositions. . . . But . . . for greater eloquence and copiousness of Speech, it should be capable [=allowable] . . . to join the [permissible] words compounded. . . . So the word *idolatry* is [also expressible as] *Idol-worship*, etc." (1668:354).

Alas, the post-1668 emergence of social science shows otherwise. Vocabulary formats evolve with the time and place, or *zeitgeist*. They reflect context and culture. Notorious instances are the Innuits (Eskimos')

multiplicity of *snow* words, and the Yankees' plenitude of *drunk* terms.

Now, only a small part of history involves writing. And only a fraction of humanity has been literate. Yet even the *Oxford Unabridged* neglects the fountainhead of neologism, which is oral speech.

Therefore *The Wordtree* has sought it out, especially by interviewing technicians and professionals. And, as noted earlier, we cite a source for each transitive. Our very first edition shows a quarter-million listings.

Now, Lesk wondered if our binary definitions are "oversimplified": To *fish* something is merely to *catch* it and to *draw* it. So, he asks, wouldn't that apply also to *tempting*, *stealing*, etc.?

Yes, indeed. A language has both general/generic terms and specific terms. *Fishing* may originally have concerned only aquatic creatures. But today it is used broadly. And a branching system defines by the lowest common denominator.

We have found that a culture also creates a precise term for each semantic niche. To fish something out by *scrutinizing*, for example, is to *expiscate* it.

As long as a wordbook forces one term, such as *to fish*, to represent many diverse procedures, automatic translation cannot occur. There are simply too many possibilities. An intermediate step is needed, first to translate which *kind of fishing* is meant. Only then can clear meanings (such perhaps as a defining phrase) appear. And *The Wordtree's* listing at "fish" does define some 15 variant (*superacteme*) forms like *expiscate*.

We do not say that the step must be done by human hands. Content analysis, for example, may assign probability weightings almost instantly. But we say that such a name-specifying brancher has been lacking, and that *The Wordtree* has finally collocated the gradations. Hence CL should rejoice in the open-endedness offered by a branching word system. It should welcome this emphasis on discovering and integrating current technical lexemes with the long-established, abstract ("ink-horn") words.

CONCLUSION: CL NEEDS APPLIED SOCIAL SCIENCE

The continuing lack of social scientists' input both to this periodical and to the CL discipline is not accidental but significant. We have therefore outlined 5 of the fundamentals that applied social science now offers: segregating processual words from substantive words; far more ethnographic reportage of technolact; the connectedness, not discreteness, of terms; the world as branching not unilinear; vocabulary as open-ended not finite; and finally, near-synonyms to be nuanced, not conmingled.

When we note some of the major assumptions present in the latter but not the former, we may understand why much of today's CL is technically brilliant but culturally trivial.

Hence Lesk's discussion of the pre-evolutionary

Wilkins has contributed to CL by revaling the incompleteness of some of its premises. For the real world consists of human interactions. To abstract and analyze will require the contributions and cooperation of many more, especially social, disciplines than currently appear in the typical CL curriculum.

© 1987 by *Henry G. Burger*
University of Missouri
Kansas City, MO 64110-2499

REFERENCES

- Burger, Henry G. 1984 The Wordtree: A Transitive Cladistic for Solving Physical & Social Problems. In *The Wordtree* (1st ed.), Merriam, KS
- Lesk, Michael 1987 Review of *The Wordtree*. *Computational Linguistics* 13: 69-70.
- Marney, M. C. and Smith, N. M. 1964 The Domain of Adaptive Systems. *General Systems* 9: 107-133.
- Simpson, George Gaylord 1949 *The Meaning of Evolution*. New Haven: Yale University Press, New Haven, CT.
- Wilkins, John 1668 *An Essay Towards a Real Character and a Philosophical Language*. Royal Society, London, England. Photographically reprinted by Scolar Press Ltd., Menston, England. 1968.