CHICAGO JOURNALS

Review: [untitled]
Author(s): James D. McCawley
Reviewed work(s):
The Vastness of Natural Languages by D. T. Langendoen ; Paul M. Postal
Source: *International Journal of American Linguistics*, Vol. 53, No. 2 (Apr., 1987), pp. 236-242
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/1265147
Accessed: 13/05/2009 15:55

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ucpress.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *International Journal of American Linguistics.*

http://www.jstor.org

NOTES AND REVIEWS

THE VASTNESS OF NATURAL LANGUAGES. By D. T. Langendoen and Paul M. Postal. Oxford: Blackwell, 1984. Pp. ix + 189.

In McCawley (1982), I introduce a technical sense of the term "red herring": a proposition *p* is a red herring for a scientific community *C* if the members of *C* generally assent to *p* but do not believe *p*. A clear instance of a red herring for the community of transformational grammarians is provided by the proposition that a language is a set of sentences, which few members of the community believe (as evidenced by their practice) but which most present to their students as if it were one of their central tenets. In their book (henceforth *VNL*), Langendoen and Postal (henceforth L & P) argue against the proposition that a language is a set of sentences, but not for the reasons that I would offer, such as that sentences belong to a language only relative to contexts and interpretations rather than absolutely, or that the syntax of a language is not exhausted by the syntax of its sentences. Their conclusion is rather that there are too many sentences in any natural language for the collection of all of its sentences to conform to standardly accepted conceptions of "set"; the sentences of a language form rather a "mega-collection": an object that is structured by member-to-whole and part-to-whole relations the way that sets are but whose size goes beyond the limits that are usually imposed on set theory in order to avoid well-known paradoxes.[1]

L & P arrive at this conclusion by rejecting an assumption that has hitherto been made by everyone (even by them), namely, that the sentences of a language are always finite in length. The point at issue is not the trivial one of whether human beings are capable of uttering sentences that are of infinite length but the considerably more interesting question of whether the practical limitation of utterances to a finite length not only reflects a limit on human endurance but is also built into the linguistic system that a speaker of a given language has acquired. Before going on, I had better point out an ambiguity in the preceding sentence that is of importance in the present context. "Limitation of utterances to a finite length" could mean either (*i*) there is a finite length (say, 100,000 syllables) such that no utterance can exceed that limit or (*ii*) all utterances must be finite in length, though there need not be any finite upper bound to that

---

[1] So as not to beg the question of whether the various "collections" that are considered are sets in the technical sense that is assumed by L & P, I substitute "collection" for "set" in paraphrasing earlier literature in which the possibility of collections that are not sets was not taken into account. Actually, as L & P (pp. 87–88) point out, even "collection" is in some cases not general enough (a notion of "hyper-collection" is introduced to take in objects that exceed even the broad limits that are imposed on collections), but I ignore that technicality.

[*IJAL*, vol. 53, no. 2, April 1987, pp. 236–49]

Permission to reprint a note or review printed in this section may be obtained only from the author(s).

length. Transformational grammarians have, to my knowledge without exception, accepted (*ii*) and rejected (*i*). L & P point out, however, that the acceptance of (*ii*) has rested on little more than a failure to raise seriously the question of infinite linguistic structures (not as objects that an actual utterance would ever correspond to but as objects that conformed to otherwise valid rules for the given language), and that the case that Chomsky has made for rejection of any finite limit on sentence length provides equally good grounds for the rejection of any limit, finite or infinite, on sentence length.[2] Further, they argue that natural languages typically have syntactic constructions that, unless an arbitrary limitation to finiteness of sentence length is imposed, provide for infinitely long sentences, in the sense that there are structures of infinite length that conform to the rules for the given constructions. For example, coordinate structures have no upper bound on the number of conjuncts and would thus take in infinite structures as in (1*a*) and (1*b*), unless infinite conjoining is specifically excluded, and the iteration of complement structures would provide for infinite structures as in (1*c*), unless such iteration is specifically required to terminate:

(1*a*) The positive integers are 1, 2, 3, 4, 5, . . .
(1*b*) 1 is a positive integer, 2 is a positive integer, 3 is a positive integer, 4 is a positive integer, . . .
(1*c*) I think that you think that I think that you think that I think . . .

If one admits infinite syntactic constructions as in (1), the possibility of filling the various slots in them by infinite syntactic constructions, and repeating ad infinitum, yields a collection of sentences that is not merely infinite with a vengeance but is "vast" in the semitechnical sense in which L & P use that word: the bounds of normal set theory, with all its transfinite cardinal numbers, are exceeded.

L & P make an important point about examples like (1) that could easily escape a reader's attention, since it does not appear until the final section of *VNL* and is not hinted at in places earlier in the book where it is relevant, namely, that one can plausibly claim that many infinite structures are fully well-formed syntactically but are ill-formed in respects that are nonsyntactic. For example, L & P point out that an infinitely iterated genitive construction (*his father's father's father's . . . dog*) may be ill-formed because the genitive marker must be suffixed to the last word of the modifying NP, and in the case in question that NP has no last word and thus cannot meet that morphological demand. Likewise (this example is mine, not L & P's), English requires that a terminal intonational contour be imposed on the last primary stress of a sentence; since none of the potential bearers of primary stress in (1) is the last one in its sentence, no constituent would meet the conditions for attachment of the terminal contour, and thus phonological ill-formedness would result.[3] The iterated complement construction of (1*c*) could in addition be held to be semantically ill-formed, in that while each of the occurrences of *think* has as its complement something that is syntactically a sentence, none of those sentences denotes a determinate proposition and thus none specifies what it is that I or you are said to think.

---

[2] While the idea of infinitely long sentences is thoroughly exotic in linguistics, it is not at all outlandish in formal logic, which indeed has a fairly well developed branch ("infinitary logic"; cf. the works by Karp, Barwise, Keisler, and Drake that L & P cite) in which, e.g., the conjoining of infinitely many conjuncts is allowed.

L & P's discussion of the well-formedness of their various infinite examples misses the important point that one can talk about infinite sentences (and even make judgments of their well-formedness in one's language) without uttering, hearing, or reading all the infinitely many words that the sentence consists of. Typographical devices such as the . . . in (1) can be used to get an infinite structure across in a finite amount of time, and as long as one has apprehended the structure (regardless of whether it is finite or infinite), one can make judgments about the respects in which it conforms to one's linguistic knowledge or conflicts with it. Remarks like the following lead me to doubt that L & P appreciate this last point: "Limitations of memory, computational space, etc. preclude the actual mental instantiation even of most finite sentences. . . . Put differently, human beings cannot actually form direct intuitions about most finite sentences just as they cannot about ANY transfinite ones" (p. 121 [emphasis added]). It is not clear to me whether L & P are referring here to the limitations of memory, etc. that would be exceeded in normal comprehension of a spoken or printed sentence. If, as I suspect, they are, their conclusion is a non sequitur: apprehending a sentence by normal means is one thing, and making judgments about a sentence that one has (somehow or other, perhaps by abnormal means) apprehended is quite another thing, and inability to perform the one task does not imply inability to perform the other. Indeed, L & P happily put asterisks and question marks in front of many of the infinite examples that they cite, thus suggesting that they are up to the task of judging the well-formedness even of infinite English sentences, just as long as their notational devices allow them to apprehend the infinite structures without exceeding their memory limitations.

Infinite objects share with finite ones the characteristic that the investigator needs to exercise considerable caution in interpreting an informant's judgments about them, even when he is his own informant, since the informant may not really be judging what the investigator thinks he is judging. For example, putative judgments of grammaticality often turn out to be just reports of whether one has succeeded in finding a plausible context for its use or of whether one has noticed any "grammatical" oddity about it. The latter sort of report is particularly liable to yield false positive responses in the case of infinite objects, where it is often impossible to do an exhaustive search for possible sources of oddity. I would suggest that L & P's acceptance of (2) (and their judgment of it as ambiguous on page 114) is a false positive report:[4]

---

[3] Similarly, a coordinating conjunction in English must be positioned before the last conjunct (and optionally also before earlier conjuncts), and since none of the conjuncts in (1*a*) and (1*b*) is last, none of them meets the condition for placement of the conjunction. Placement of the conjunction is a phenomenon that straddles the border between syntax and morphology; if one chooses to class it as morphology (not what I would do myself, but I see no serious objection to it), one could then regard failure of placement of the conjunction in (1*a*) and (1*b*) as a morphological defect. Even if placement of the conjunction is classed as syntax, the syntactic ill-formedness of (1*a*) and (1*b*) is confined to a restricted realm, and counterparts to such sentences in a language in which coordinating conjunctions are not obligatory (e.g., Chinese) would not be syntactically ill-formed.

(2) *Jack*$_1$ *and his father*$_2$ *and his father's father*$_3$ *and . . . his . . . father*$_{\aleph_0}$ *are visiting relatives.*

The difficulty in (2) is in interpreting the final conjunct (a difficulty that is absent from a corresponding sentence that has no final conjunct), which requires that the genitive construction be iterated to a transfinite depth within a single conjunct. But there are two ways that one could iterate the genitive construction (successively replacing *he* by *his father* at the bottom of the construction or successively adjoining *-'s father* to its top), and L & P's schematic representation of the last conjunct in (2) does not fit either of those understandings of the iteration, since they give both a first and a last word, while the two versions of iteration would respectively yield an NP without a first word and one without a last word:

(3*a*) [[[. . . *'s father*]*'s father*]*'s father*]
(3*b*) *his father's* [*father's* [*father's* . . .]]

Invocations of mathematics by linguists have generally had a polemic function—to demonstrate that the version of linguistics that one favors has some virtue that the versions that one disfavors lack. Not surprisingly, *VNL* contains a fair amount of polemicizing for arc-pair grammar (Johnson and Postal 1980) and against a diverse selection of alternative approaches to syntax, a litany of which is given on pages 73–74. Arc-pair grammar provides easily for the sorts of infinite syntactic structures that L & P argue are syntactically well-formed in natural languages, whereas the alternative approaches are all committed to the proposition that sentences must be finite in length. The trouble with this claim is that the alternative approaches differ enormously in the EXTENT to which they are committed to finiteness of sentences, i.e., how radical a revision of their assumptions would have to be made in order to admit infinitely long sentences. In some cases it would take only a minor revision to allow at least some of the infinite structures that L & P admit (e.g., it is a trivial matter to revise the rules that Montague 1973 gave for two-term coordinate structures so that they would cover structures with any number of conjuncts, finite or infinite).

---

[4] This example illustrates the one important mathematical error that repeatedly occurs in *VNL*, namely, confusion between cardinal and ordinal numbers. L & P's intentions would have been better served here and in similar examples if they had the infinite ordinal number ω where they have instead the infinite cardinal number $\aleph_0$. This error becomes serious on page 171, where they speak of negative infinite cardinal numbers, a notion that is literally incoherent. Ordinal numbers are in fact taken up shortly after that passage, but the location of that material suggests that it was inserted as an afterthought.

Most of the polemics in *VNL*, however, are concerned not specifically with L & P's preferred theory of grammar but with the PLATONIST position as to the nature of a language, which they favor over Chomsky's CONCEPTUALIST position, and with NONCONSTRUCTIVE rather than CONSTRUCTIVE conceptions of the nature of a grammar. In the various versions of conceptualism that L & P distinguish, linguistic units and linguistic systems are psychological in nature, while in the platonist conception they have the same sort of existence as do mathematical objects. L & P are not so clear about what constitutes a constructive grammar. The term is introduced in the statement of their NL (= natural language) Nonconstructivity Theorem: "No NL has any constructive (= proof-theoretic, generative or Turing machine) grammar" (p. 72). The trouble with this list of synonyms is that subsequent discussion makes clear that L & P do not mean "generative" in the technical sense that it has always had in generative grammar (a generative grammar is a system of rules that is interpreted as specifying what belongs and what does not belong to some collection) and that the other two terms are not obviously synonymous either with "generative" or with each other. L & P have shortchanged the readers of *VNL* by failing to provide these terms with adequate elucidations that in a context in which infinite sentences are considered fair game are not question-begging. I suspect that their actual intention is for "generative" to conform to the familiar vulgarizations of generative grammar in which "generate" means "yield as the output of a computational procedure determined by the grammar," for which "constructive" in fact would not be a bad term.[5] They exaggerate when they describe the twenty-seven alternative approaches to syntax that they list on pages 73–74 as "taking NL grammars to be constructive devices"; it would be hard to justify that judgment in the case of Daughter Dependency Grammar and Montague Grammar.

L & P's arguments against the version of the "conceptualist" position that figures in Chomsky's most recent works (e.g., 1981) rest largely on an equivocation over the term "language," an equivocation that can be blamed on the red herring that was alluded to at the beginning of this review—the proposition that a language is a collection of sentences. Chomsky recently has begun speaking in terms that deny to collections of sentences their previous ostensibly central role (or at least that is the most plausible interpretation that I can put on such statements as "Note that the central concept throughout is 'grammar,' not 'language'" [Chomsky 1981:4]). While Chomsky's remarks are not very clearly put, L & P's paraphrases of them ("there are no NL's" [p. 148], "NL's are not real" [p. 149]) are bizarre. Chomsky in this case seems to me to be simply rejecting the red herring of languages as collections of sentences but phrasing his rejection in terms of a technical sense of the word "language" that owes its existence to that very red herring. L & P equivocate between that use of "language" and its ordinary use when they say, e.g., "what sense could there be to the notion of a *linguistic* organ if there are no NL's?" (p. 148). One does not have to recognize "languages" in the above technical sense in order to talk about languages in the ordinary sense, just as language acquisition need not involve acquisition of a language in the technical sense.[6]

---

[5] This understanding of the term "generate" is responsible for the popularity of such locutions as *over-generate* and *How does the grammar generate X?* If *generate* were interpreted in its technical sense (in which it is a stative verb, like *imply*), those locutions would be as incongruous as *over-imply* and *How do Euclid's axioms imply theorem 21?*

Much of L & P's quarrel with Chomsky relates to the proposition that sentences are "mental representations." Their objection to that proposition is that grammars, even as understood by linguists who accept an arbitrary restriction to finite linguistic objects, generate sentences that have never been physically represented in anyone's brain, and indeed sentences that never could be so represented because of the physical limits of the brain. This argument presupposes a conception of "mental representation" that few committed mentalists would accept, namely, one in which a "mental representation" must be some physical state that some brain has actually instantiated. Under other conceptions (e.g., one in which a "mental representation" is the same sort of entity as an image: a structure such as is represented in the brain rather than a physical state such as might sometimes represent it), the argument loses all force.[7]

For L & P, the collection of all learnable languages is a small part of the collection of all languages, and consequently a "language" cannot be the same thing as a "learnable language," as Chomsky and many others take it to be (pp. 149–50). While some of L & P's rhetoric suggests that the collection of all languages is "vast" in their technical sense, I see no respect in which their assumptions would dictate that conclusion.[8] Even if each individual language is "vast," that vastness need not contribute anything to diversity among languages. L & P take up no differences among languages that are not predictable from the finite parts of the languages, and all the examples of infinite syntactic constructions that they cite are infinite only by virtue of generalizations based on finite instances of the constructions. With regard to the actual examples taken up by L & P, the languages could be learned by learning their finite parts, deleting any size limitations that one has (inadvertently or perversely) allowed into one's grammar or one's metatheory, and letting all the infinite sentences that conform to the syntax go along for the ride. They indeed ultimately conjecture that "An account of finite sentences is an account of transfinite sentences" (p. 162). The plural form in the title of *VNL* is thus well chosen: if the title had been *The*

---

[6] This point is made in McCawley (1983). The cutesy technical term "harp" that Hockett (1966) proposed for "set of strings" at least has the virtue of not leading linguists into such an equivocation.

[7] Here I greatly oversimplify the question of the nature of mentally created objects such as images, which is discussed with great insight in Popper (1976:sec. 38).

[8] L & P state that there is no such thing as a "collection of all NL's" on the grounds that (at least in the version of set theory that they assume, that of Fraenkel) sets are the only kind of collections that are allowed to be members of collections (p. 84). But it is misleading for them then to say "The realm of all NL's is thus parallel to the realm of all collections, which also forms no collection." An entity of which all collections are members necessarily creates paradoxes, whereas an entity of which all NL's are members need not in itself lead to any paradox but merely lacks the property that serves in the Fraenkel version of set theory as immunization against set-theoretic paradoxes.

Vastness of Natural Language, it would have vastly overstated L & P's case. This suggests that what importance *VNL* may have is for the philosophy of linguistics rather than the practice of linguistics: one can analyze languages with blithe disregard for infinite linguistic objects and be safe in the knowledge that any misrepresentation that one has thereby engendered can be corrected in a trivial way.

JAMES D. MCCAWLEY, *University of Chicago*

REFERENCES

CHOMSKY, NOAM A. 1981. Lectures on Government and Binding. Dordrecht: Foris.

HOCKETT, CHARLES F. 1966. Language, Mathematics, and Linguistics. The Hague: Mouton.

JOHNSON, DAVID, AND PAUL M. POSTAL. 1980. Arc-Pair Grammar. Princeton, N.J.: Princeton University Press.

MCCAWLEY, JAMES D. 1982. How far can you trust a linguist? Language, Mind, and Brain, ed. T. Simon and R. Scholes, pp. 75-87. Hillsdale, N.J.: Ablex.

________. 1983. Towards plausibility in theories of language acquisition. Communication and Cognition 16:169-83.

MONTAGUE, RICHARD. 1973. The proper treatment of quantification in ordinary English. Approaches to Natural Language, ed. J. Hintikka et al., pp. 221-42. Dordrecht: Reidel. [Reprinted in R. Montague, *Formal Philosophy* (New Haven, Conn.: Yale University Press, 1974), pp. 247-70.]

POPPER, KARL. 1976. Unended Quest. LaSalle, Ill.: Open Court.

A REPLY

Una Canger devotes much space to a detailed critique of my *Analytical Dictionary of Nahuatl* (henceforth *ADN*) in which she raises many questions, some of broad methodological import and others of a very specific nature (*IJAL* 52, no. 2 [April 1986]: 188-96).

Canger is not content with the orthography that I use in the dictionary and suggests her own alternative which would use *v* to represent /w/, *s* to represent /s/, and *h* to represent glottal stop; moreover, she would have the process of alphabetization pass over *h* as invisible. I cannot believe that this would be a convenience for users of a Nahuatl dictionary working with Nahuatl texts. The orthography that J. Richard Andrews adopted for his grammar, *Introduction to Classical Nahuatl*, and which I chose to use in *ADN* to avoid proliferation of orthographies, is very close to that used in most extant Nahuatl texts. It requires a minimum of transliteration (equating *ç* with *z* and—in the oldest texts—*u*/*v* with *hu*/*uh*, putting in syllable-final *h* for glottal stop, paying attention to macrons over vowels) and is Spanish-based (*hu* for /w/, *z* for /s/, *c* for /s/ before front vowels), appropriately for Nahuatl. Words in the work of Andrews and myself look very much as they do in the Nahuatl texts. Canger's proposed orthography, on the other hand, can match no texts, since the *v* she proposes