SOME ANALOGIES OF HIERARCHICAL ORDER IN BIOLOGY AND LINGUISTICS*

Martin Zwick
Systems Science Ph.D. Program, Portland State University
Portland, Oregon
Zwickm@pdx.edu
http://www.sysc.pdx.edu/Faculty/Zwick

I. INTRODUCTION

The ubiquity of hierarchical order is obvious, and the obvious is hard to explain, but a number of workers [1] have suggested the possibility of constructing a theory (or cluster of theories), rooted in such disciplines as thermodynamics, information theory, topology, and logic, which might reveal the underlying unity of a wide variety of branching and multi-level systems. It is the purpose of this paper to contribute to both the empirical and theoretical aspects of this discussion, by examining levels of structure and function in molecular biology and linguistics, and by developing, from parallelisms between these two areas, a hierarchical model of possibly greater generality.

We consider first the hierarchy of spoken language [2]: phoneme, morpheme, word, sentence, utterance, discourse; or of written language: letter, syllable, word, sentence, paragraph, section, chapter, book. These lists are straightforward up to and including the "sentence," beyond which they are somewhat arbitrary. The items, “utterance," "paragraph," etc., are meant only to illustrate more complex units and to suggest that linguistic hierarchies consist typically of a small number of levels. This is characteristic of the organizational and/or operational structure of many "concrete" [3] systems; more "abstract" hierarchies, e.g. those which specify gradations of some attribute, often consist of a greater number of elements. For example, the system of cosmological entities, ordered by gravitational forces, ranges from aggregates of galaxies to planetary satellites, meteors, etc., but the hierarchy of the structure of matter extends through additional levels down to elementary particles and the like. Similarly, the organizational or "line" structure of command in the military has fewer levels than might be suggested by existing gradations of rank.

Thus, the following hypothesis: Concrete systems are typically limited to a small number of levels, say five to ten, after which the hierarchy often becomes consolidated in a stable and coherent whole (which may become a base unit for still higher levels). We shall refer to the range between base and terminal levels as a "period." For example, a library consists of two periods: letter to book and book to library. The hierarchy of structures from atom to cell is, in our opinion, also such an interval, the intermediary levels of which include small molecules, polymers, macromolecular aggregates, and the like. We could continue on to tissues, organs, organ systems, organisms, populations, etc., but suggest that the intervals from atom to cell and cell to organism represent natural divisions. In general, there may be some uncertainty about which level should begin and/or end a period, but the range of uncertainty is usually not great, and the principle of this distinction, the idea of periods, is not purely subjective. For example, it would be inappropriate to start the cellular hierarchy at the level of protons and neutrons or quarks, because

*Reprinted from: *Applied General Systems Research: Recent Developments & Trends* edited by George Klir. New York: Plenum Press, 1978, pp. 521-529.

the domain of influence of the cell as a system does not penetrate down to these levels. Or we might extend the biological hierarchy to include tissues, organs, and so on up to the level of the individual organism, but at this point, biological structure most typically passes into social organization, as in families, populations, and communities.

The biological hierarchy we shall actually discuss is the following: atom (or atomic ion); chemical group (or small molecule or molecular ion); amino acid; (monomeric) protein; (multimeric) enzyme; multienzyme complex; organelle (or membrane system); cell. That is, we trace out only one "path" (of many) in the set of structures between single atoms and whole cells, focusing mainly on proteins, the principal dynamic agents in the cell, and especially on those entities of the protein hierarchy most general in function [4]. Thus, for example, above the level of polypeptide chain or monomeric protein, we choose the "spherical" multimer rather than the helical polymer, as the latter typically has a more specialized function. Much of the discussion, which follows, will pertain also to the nucleic acids and to some deeper and more general properties of the cellular period.

II. PARALLELISMS BETWEEN THE BIOLOGICAL AND LINGUISTIC PERIODS

A number of investigators [5] have considered analogies between linguistic entities and the nucleic acid levels of nucleotide, codon, cistron, and operon. Since these nucleic acid units, structurally speaking, vary only in degree of polymerization, and serve as the static repository of genetic information, we have chosen to focus instead on the more dynamic proteins, which functionally express the genetic information, and which exhibit structural levels more clearly differentiated one from another. Most observations on nucleic acid entities as linguistic units can be equivalently stated in terms of protein levels.

Placing our two hierarchies in parallel (Table 1), one observes the following: The smallest unit of linguistic meaning, the morpheme, is built up of more fundamental linguistic entities, just as the typical unit of chemical reactivity in the cell consists usually of more than one atom. That is, the base level entities provide a structural foundation for the period upon which are built up, at the next level, functionally active units. These units, in turn, do not simply participate in a chaotic flux of constant interchange, but are organized into third level entities, functionally at least bivalent, and thus capable of more varied interactions.

Table 1 Biological and Linguistic Hierarchies

8. cell	8. book
7. membrane system, organelle	7. chapter
6. multienzyme complex	6. section
5. (multimeric) enzyme	5. paragraph
4. (monomeric) protein	4. sentence
3. amino acid	3. word
2. chemical group	2. syllable/morpheme
1. atom	1. letter/phoneme

But the most fundamental analogy here is that which associates proteins and sentences. A language is most deeply characterized by the set of sentences which can be generated via its grammar; similarly, a cell may be "defined" by the set of proteins synthesized in it. That is, it is these levels, not higher or lower ones, which represent the essence of the biological and linguistic periods. Proteins and sentences are both linear polymers--of amino acids and words, respectively, and in both cases the set of all possible sequences is so large as to be physically unrealizable. (There are many fewer amino acids than words, but the former are concatenated into longer strings.) At this level of complexity, therefore, a kind of "individuality" first emerges. Of the ensemble of sentence or protein possibilities, only a small subset has ever existed. By contrast, the sets of phonemes or atom types, morphemes or chemical groups, and words or amino acids, which belong to a language or cellular system, are more limited and can be explicitly enumerated.

With the emergence of the possibility of an unlimited variety of expression, elaborate "apparatuses" are needed to guarantee that only those sequences are selected and/or constructed which are functionally meaningful. For the cell, this is accomplished by the transcription, translation, and expression of stored genetic information. For language, some aspects of the operation of a corresponding (but still obscure) sentence-generating system are revealed in the rules and structures of grammar.

Proteins and sentences thus become the principal functional units of the cellular and linguistic hierarchies, and require, as it were, special attention. By contrast, lower and higher levels often exhibit "spontaneous self-assembly." The synthesis of amino acids, and the aggregation of polypeptide chains into enzymes and enzymes into complexes, i.e. the formation of both sub- and supra- protein entities, proceed via local processes of greater simplicity, subject to genetic control only indirectly via the protein level. Similarly, the principles governing the "covalent" interactions of phonemes or morphemes and the "non-covalent" interactions of sentences and higher linguistic units, are less complex and less central to the language than the laws of sentence construction and interpretation.

The significance of the level of protein should not, however, be exaggerated, since it is actually the multimeric enzyme which is the "workhorse" of the cell. The structural complexity of the protein makes possible the catalysis of specific reactions upon which the existence of life fundamentally depends. But each reaction must be coordinated with many others, and this requires the functional sensitivity of the multimer. The next higher level, the multi-enzyme complex, serves the interests of efficiency, but is well beyond the realm of necessity.

Similarly, the basic functional unit of language, the sentence, typically asserts a proposition, i.e. a relation of concepts. But the quintessence of functional potency, which might be called an "idea" or "message" requires greater elaboration. More than one sentence is typically needed to communicate a coherent set of assertions, with adequate force, and suitable context. At still higher levels, we have more sustained linguistic activity, and consolidations of meaning, but less "intensity."

III. PROTEINS AND SENTENCES

Interactions between parts of a protein cause the linear sequence of amino acids (the "primary structure") to fold up into a three-dimensional structure, which performs a specific catalytic function at a locus on the molecule known as the "active site." Similarly, the syntactic or surface structure of a sentence is the result of its primary structure folding up, as it were, in some cognitive space, so as to bring certain words, not neighbors in the linear sequence, into syntactic proximity and perhaps also conferring upon the sentence an active site or principal locus of function. (The sentence, unlike the protein, contains its own substrate.) In Chomsky's linguistic theory [6], a sentence has also a tree-like semantic or deep structure which represents, in terms of its meaning, its prototypical organization. The words which are elements of the deep structure need not be identical with those in the actual sentence, there being transformational rules by which the syntactical organization is derived from its parental (semantic) structure, and by which the word order of the sentence is generated.

One might speak of such a "deep structure" for a protein, embodying the functional essence of the molecule in its most economical form. Despite variations in the sequence and structure of a particular protein across different species and within populations of the same species, some amino acid sites will be essential for the catalytic function and will therefore be invariant, and additional sites may show restricted variability. The set (or family) of such similar proteins (e.g. the hemoglobins or cytochromes) might be regarded as containing variations upon an archetypal or minimal structure. The archetypal sequence and three-dimensional structure may resemble the primeval protein from which the family actually evolved; or, alternatively, the family members might be considered to be "logically" but not historically derived from their archetype by transformational rules.

Considered as separate systems, both proteins and sentences are hierarchical, i.e. have structural levels intermediate between that of the unit which is polymerized (amino acids, words) and the polymer as a totality. Sentences may contain phrases and clauses, i.e. have tree-like syntactic and semantic structures. Similarly, in proteins, portions of the amino acid sequence are often organized into "secondary" (e.g. helices and sheets), "supersecondary," and sometimes still higher-order substructures. These intermediate levels (which might be said to constitute a "micro-hierarchy") are not always present, nor are they generally stable in isolation from the protein whole; their origin and significance is tied to the unique character of the transition to the protein/sentence level.

IV. TERMINAL TRANSITIONS

There is also something special about the terminal transition which completes a period and generates a new base level for further complexification. This is implicit in the notion of a period, and is obvious for the cell: the organization of membrane systems, organelles, multi-enzyme complexes, and all the lower level units into a coherent unity is a structural and functional achievement of magnitude considerably greater than any of the lower transitions. As passage to the level of protein is mediated via the protein synthetic system, so passage to the level of a cell involves the complex apparatus of cell division. We recognize in these two special transitions, which we shall call "secondary" and "primary" respectively, the two basic structural tasks of the cell.

There exists also a primary transition, which completes the period of the linguistic hierarchy, although it is much more obscure. One can multiply sentences into paragraphs, chapters, etc., but the synthesis of these into a true unity can be an accomplishment of a much higher order, and may carry or reflect meaning at a level which might be called "cultural." (We make no particular assertion about the number of levels above that of a sentence needed to sustain such meaning; it is, however, small.) More deeply, if we consider a mental hierarchy of linguistic entities, completion of the period corresponds perhaps to the acquisition of linguistic competence. Mental constructs requiring levels higher than the sentence are implicit in the grammatical system, and are necessary for sentence construction, just as proteins require for their synthesis more complex structures, which contain proteins as subunits.

Primary and secondary transitions are linked. The transition to the level of sentence or protein is conditional upon a higher level transition which has already been accomplished (at least transiently). The protein synthetic machinery presumes, for its context, the milieu of self-reproducing and evolving cells. Similarly, the principles of grammar which govern sentence formation imply the existence of a completed linguistic period, externally in a certain cultural accomplishment, and internally in a mental edifice of sufficient complexity for linguistic function.

This relationship between primary and secondary transitions may offer an interesting perspective on (but no solution to) the "origins problem" which exists in both biological and linguistic areas. Some suggestions on the biological side are sketched in Figure 1. In the primal soup preceding the origin of life, and in the intracellular soup which retains traces of its ancestry, the spontaneous assembly of complex structures, unassisted by the guidance of genetic information, is limited to molecules such as amino acids and nucleotides and, at best, random polymers of such units (Figure l(a)). That is, the secondary transition represents a kind of barrier to the evolutionary tendency. This barrier is not absolute, but permits the random generation of proteins, some of which may be capable of catalytic function, and of nucleic acid polymers. Through an exceptional fluctuation, a system of integrated reactions may come into existence inside some suitable enclosure, which simultaneously bridges primary and secondary barriers (Figure l(b)). Initially, such a cellular precursor would not be likely to exhibit a full range of structural levels, e.g., it would probably not have multi-enzyme complexes or organelles, although it might have a level intermediate between protein and cell, possibly for primitive ribosomes. If viable, such a system may then evolve to populate the (few) additional supra-protein levels which are implicitly possible. Meanwhile, generic control replaces spontaneous polymerization in the generation of proteins, and the rapid evolution of the entire system obliterates its humble origins. (The fully developed system is shown in Figure l(c), which also summarizes several static features of the present hierarchical model: the idea of periods, the two barriers, and the micro-hierarchy.)

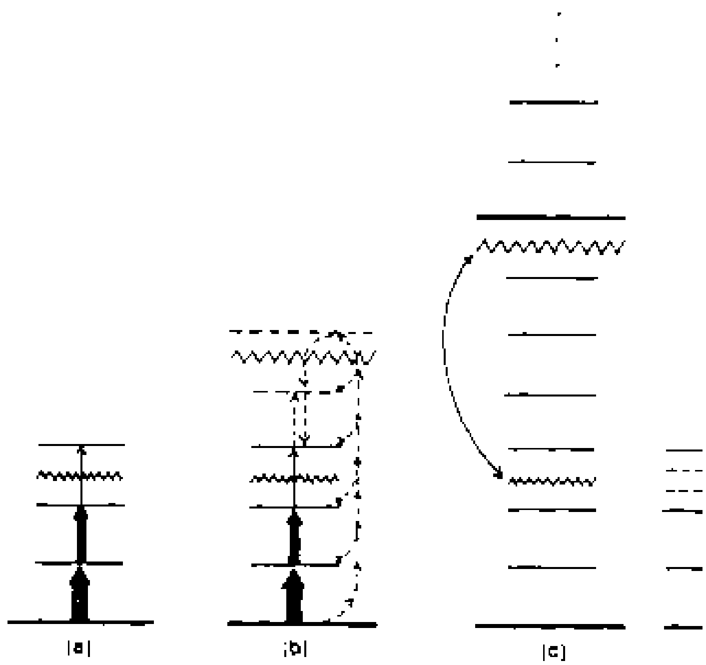

Figure 1 A hierarchical model. (a) Evolution via spontaneous but limited complexification; (b) integration through fluctuation; (c) the resulting structural levels and transitions.

Hierarchical periods thus tend to divide into two domains. In the present case, these correlate roughly with the predominance of covalent versus non-covalent forces. The lower levels exhibit a limited tendency to evolutionary "ascent." The higher levels, by contrast, tend to be organized from above, or require for their existence the completion of the period. In the linkage of primary and secondary transitions, we have a convenient representation of the fact that while the phenomenon of life begins only at the level of the cell, proteins are unquestionably biological entities. We leave for later study the question of whether a similar analysis might cast any light on the linguistic origins problem, i.e. on the "phylogenetic" emergence of language as a capacity of the human species and its "ontogenetic" acquisition by individuals.

The analogies which have been drawn between molecular biological and linguistic structures will be more precisely and fully elaborated in a subsequent paper, and further theoretical and empirical explorations of the proposed hierarchical model are in progress. The parallelisms between linguistics and molecular biology discussed here obviously do not suffice as a justification for this model; they merely introduce it [7].

REFERENCES

1. See, for example, in L. L. Whyte, A. G. Wilson and D. Wilson, eds., *Hierarchical Structures*. American Elsevier, New York, 1969; P. Weiss, ed., *Hierarchically Organized Systems in Theory and Practice*. Hafner, New York, 1971; H. H. Pattee, ed., *Hierarchy Theory: the Challenge of Complex Systems*. G. Braziller, New York, 1973; also the work of Koestler, Laszlo, Leopold, Levins, Mesarovic, Miller, Piaget, Platt, Simon, Thom, von Bertalanffy, Woldenberg, and others.

2. R. Jakobson, "Parts and Wholes in Language." In: *Parts and Wholes*, edited by D. Lerner, Free Press of Glencoe, New York, 1963, pp. 157-162.

3. J. G. Miller, "The Nature of Living Systems." Behavioral Science, 16, No. 4, July 1971, pp. 277-301.

4. R. E. Dickerson and I. Geis, *The Structure and Action of Proteins*. Harper and Row, New York, 1969.

5. See, for example, J. R. Platt, "A 'Book-Model' of Genetic Information Transfer in Cells and Tissues." In: *Horizons in Biochemistry*, edited by M. Kasha and B. Pullman, Academic Press, New York, 1962, pp. 167-187; R. D. Masters, "Genes, Language, and Evolution." Semiotica, 2, No. 4, 1970, pp. 295-320; D. Berlinski, "Philosophical Aspects of Molecular Biology." Journal of Philosophy, 69, 1972, pp. 319-335; R. Jakobson, "Main trends of research in the social and human sciences. Part one: Social Sciences." Offprint, Mouton/Unesco, Paris, 1973; S. Marcus, "Linguistic Structures and Generative Devices in Molecular Genetics." Cahiers de Linguistique Theorique et Appliquee, 11, No. 1, 1974, pp. 77-104.

6. N. Chomsky, *Aspects of the Theory of Syntax,* M.I.T. Press, Cambridge, 1965; J. Lyons, Noam Chomsky. Viking, New York, 1970.

7. The author wishes to thank William C. Wimsatt and R. B. Lees for many helpful discussions. They are, of course, not responsible for any errors in the text.