

Hunt Institute for Botanical Documentation
5th Floor, Hunt Library
Carnegie Mellon University
4909 Frew Street
Pittsburgh, PA 15213-3890
Telephone: 412-268-2434
Email: huntinst@andrew.cmu.edu
Web site: www.huntbotanical.org

The Hunt Institute is committed to making its collections accessible for research. We are pleased to offer this digitized item.

Usage guidelines
We have provided this low-resolution, digitized version for research purposes. To inquire about publishing any images from this item, please contact the Institute.

Statement on harmful and offensive content
The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences. Due to the wide range of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions. The Hunt Institute for Botanical Documentation does not endorse the views expressed in these materials, which are inconsistent with our dedication to creating an inclusive, accessible and anti-discriminatory research environment. Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections' creators.

Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do not reflect the views or values of the Hunt Institute for Botanical Documentation or those of Carnegie Mellon University.

About the Institute
The Hunt Institute for Botanical Documentation, a research division of Carnegie Mellon University, specializes in the history of botany and all aspects of plant science and serves the international scientific community through research and documentation. To this end, the Institute acquires and maintains authoritative collections of books, plant images, manuscripts, portraits and data files, and provides publications and other modes of information service. The Institute meets the reference needs of botanists, biologists, historians, conservationists, librarians, bibliographers and the public at large, especially those concerned with any aspect of the North American flora.

Hunt Institute was dedicated in 1961 as the Rachel McMasters Miller Hunt Botanical Library, an international center for bibliographical research and service in the interests of botany and horticulture, as well as a center for the study of all aspects of the history of the plant sciences. By 1971 the Library's activities had so diversified that the name was changed to Hunt Institute for Botanical Documentation. Growth in collections and research projects led to the establishment of four programmatic departments: Archives, Art, Bibliography and the Library.

MON.	WED.	FRI.
	Mar. 23 D Introduction	Mar. 25 Introductory Lab.
Mar. 28 D History of Problems and Objectives	Mar. 30 B Computers - I	Apr. 1 Davis & Heywood Lab.
Apr. 4 B Computers - II	Apr. 6 G Math. - I	Apr. 8 Math. Problems
Apr. 11 G Math. - II	Apr. 13 H How to Describe Plants	Apr. 15 Examine Plants (Uvularia)
Apr. 18 H How to Set Up Characters	Apr. 20 H or D Problems in Finding Similarity Measures	Apr. 22 Code the Study
Apr. 25 G Solutions to Similarity Problems	Apr. 27 D What is a Biological Cluster?	Apr. 29 Test Clusters With Specimens
May 2 G Development of Numerical Method	May 4 G Brief Review of Other Methods	May 6 NO CLASS
May 9 B Role of Computing Machines in Bio Modeling	May 11 D or H How to Interpret Results	May 13 Prepare Data for Input
May 16 D Presentation of Results - I	May 18 D Present Results - II	May 20 Interpret Results
May 23 H or D Future Applications, Implications, etc.	May 25 H or D Recap	May 27 Make Classification

Digitized by the Hunt Institute for Botanical Documentation

ANNOUNCEMENT OF NEW COURSE-SPRING QUARTER, 1967

Botany 170: Taximetrics
Instructor: D. J. Rogers
Prereq: Introd. Biol. (Bot., Zool., Entom., Micro.) one course in systematics (same fields as above)

5 credits
Lectures: 10 MWF
Lab.: 3-5 MW

Text: Principals of Angiosperm Taxonomy. Davis & Heywood (1963).

OUTLINE OF BOTANY 170. TAXIMETRICS

I. Definitions and description of taximetrics (syn. = "numerical taxonomy").
   A. Application of numerical methods to non-numerical data in taxonomy.
   B. The place of taximetrics in the total field of biological taxonomy.
   C. The types of data used in classification process.

II. Historical review of taximetrics.
   A. Michel Adanson, 18th Century French biologist, non-weighted classification, and the subsequent controversies.
   B. The early efforts to quantify taxonomic classifications.
   C. Computers and their influence on taximetrics.

III. How is a computer used in classification?
   A. Simple description of the computer.
   B. How the circuitry of the computer lends itself to classification.
   C. Examples of use of computer in classification. (Psychology, anthropology, biology, social sciences)

IV. Introduction to fundamental mathematical concepts useful in taximetrics.
   A. Arithmetic logic
   B. Sets
   C. Unique determinations
   D. Functions and correspondence.

V. Biological foundations for taxonomy and ecology.
   A. The "synthetic" versus the "analytic" aspects in biology.
   B. The information derived from biological objects--characters.
   C. The role of characters individually and collectively in classification.
   D. Defining taxonomic categories.
   E. Hierarchical versus non-hierarchical classifications.
   F. The formal taxa of biology.

VI. Developing rules and applications of mathematical thought processes appropriate to biological problems (fitting the mathematics to the biology rather than the opposite).

VII. Practical applications using existing methodologies--"doing a classification."

Digitized by the Hunt Institute for Botanical Documentation

10 Weeks 2 Lectures One long lab. or discussion per week

Lecture Number	Subject	Speaker	Contents
1	Introduction	Dave	Purpose of course, how we got into this ourselves. (Anderson's Exp.)
2	History		
3 3/4 4 5	Computers	Bob	Simple exposition of how they work, what they can and cannot do. Examples of uses. (Anthrop. 1st part)
4 5 6	Mathematical concepts fundamental to objective thinking and modeling	Me	Sets, arithemetic, logic. Definitions, unique determination, functions, and correspondences.
7	Review of the history problems and objectives of bio-class	Dave	What is classification for and what have others thought about it. (Davis & Heywood, 1st 4 chapters)
8	How do you describe plants	Hen	Coding and characters keys
9	and what can be done with the subscriptions	Dave	identification information preservation.
10	How do you measure similarity	Me	Discuss similarity measures
11	Problems biological with similarity	Dave Hen	How does similarity measure influence coding desicions. What problems arise.
12	Solutions	Me	How can math solve problems.
13	How should we define a	Dave-	Discuss various definitions
13½	biological cluster	Hen	
14	How can we use these definitions	Me	Discuss efforts on cluster analysis
15	Other measures and methods	Me	
16	The role of the machine in using the method	Bob	Preparation of data formats for printout, etc.
17		Dave	
18	Interpretation of results	Hen	Case studies.
19		Dave	
20	Presentation of results	Hen	Case studies, keys, diagrams.
21	Recaps	Dave	Restatements of objectives and attitudes

Digitized by the Hunt Institute for Botanical Documentation

OUTLINE OF BOTANY 170. TAXIMETRICS

The following is an outline of the course rather than a lecture schedule. This method of presentation is offered rather than the lecture schedule because the course is "experimental" this year. Having no precise prerequisites, and having no idea whether students will be sufficiently well prepared to accept the total offered by this outline, we will most likely have to modify the content as we proceed.

Four staff members will participate in the course. The biological and taxonomic aspects will be taught by D.J. Rogers and H. Fleming, the mathematical aspects by G. Estabrook, and the computer aspects by K. Brill.

The method of presentation will be a combination of lecture, discussion, and practical application. Students will actually do a small, selected classification, run it through the computer, and interpret the results.

We want to keep the class small for the first run-through to allow more flexibility, and to give us more "feed-back" on the value of the presentation.

I. Definitions and description of taximetrics (syn.="numerical taxonomy").

- A. Application of numerical methods to non-numerical data in taxonomy.
- B. The place of taximetrics in the total field of biological taxonomy.
- C. The types of data used in classification process.

II. Historical review of taximetrics.

- A. Michel Adanson, 18th Century French biologist. non-weighted classification, and the subsequent controversies.
- B. The early efforts to quantify taxonomic classifications.
- C. Computers and their influence on taximetrics.

III. How is a computer used in classification?

- A. Simple description of the computer.
- B. How the circuitry of the computer lends itself to classification.
- C. Examples of use of computer in classification. (Psychology, anthropology, biology, social sciences)

Digitized by the Hunt Institute for Botanical Documentation

30 (32)

OUTLINE OF BOTANY 170. TAXIMETRICS
- 2 -

IV. Introduction to fundamental mathematical concepts useful in taximetrics.

A. Arithmetic logic.
B. Sets
C. Unique determinations.
D. Functions and correspondence.

V. Biological foundations for taxonomy and ecology.

A. The "synthetic" versus the "analytic" aspects in biology.
B. The information derived from biological objects -- characters.
C. The role of characters individually and collectively in classification.
D. Defining taxonomic categories.
E. Hierarchical versus non-hierarchical classifications.
F. The formal taxa of biology.

VI. Developing rules and applications of mathematical thought processes appropriate to biological problems (fitting the mathematics to the biology rather than the opposite).

A. This, and the next section, will be student-participation procedures.

VII. Practical applications using existing methodologies -- "doing a classification."

Digitized by the Hunt Institute for Botanical Documentation

For discuss of tax characters -

1. What is the purpose of this data gathering -
Clearly for differentiation

Digitized by the Hunt Institute for Botanical Documentation

The significance of the labs is not to be minimised. They will provide a free form discussion/forum environment for the for the exchange and development of ideas especially as they bear on the relation between theory and practice. Each student will bring to the lab some problem in data synthesis from his own discipline. Lab periods will be spent discussing these several problems and in applying the principles and concepts of the lectures to their organization and solution. It is expected that student data will be analysed with computing machines at times appropriate to the development of the course.

The lectures may be characterized by weeks as follows:

1. Introduction. Purposes of the course. History of the development of the ideas and the need for them.
2 Niave set theory and Boolean algebra.
3. Logic and the propositional calculus. Its analogue with set theory.
4. Functions and rela tions.
5. Definition of a character and rel ted considerations.
6. Elementary combinatorics and probability theory.
7. What is emperical science. What is mathematical modeling in emperical science.
8. Informatuon retreival and computing machines.
9. What is information. How can we measureit . Information content in a character.
10. Interdependence and prediction in characters and classifications.
11. The Graph theory model for classification

Digitized by the Hunt Institute for Botanical Documentation

12. Interpreting the Graph theory model.

13. Other models in classification.

Recall the two main purposes of the labs:

1. To provide an environment for discussion and exchange of ideas between students and experienced emperical scientists.

2. To provide an opportuhity for student to apply the concepts of the course to his own emperical problem in description, characterization, and synthesis of data from his field of interest.

The labs might be characterized by weeks as follows:

1. Introductory discussions developing the history and purpose of the course.

2. Problem and practice sessions in logic and set theory with examples from the represented fields.

3.Delimiting for each student some problem from his field of interest to be maintained as his personal example for the remaining labs.

4. Discussing the related problems of data organization and synthesis.

5. Discussion and construction of descriptive characters for each example study.

6/7 Description of chosen study by means of characters constructed.

8.Analyse character analysis output

9.Analyse character analysis output and redefine characters if necessary

10/11 Analyse graph output

12. Each students presents a formal decsision and repotr

Digitized by the Hunt Institute for Botanical Documentation

to the rest of the class.

13. Reach conclusions and write report.

Detailed content of lectures.

1. Need for principles in biology and social sciences. Early work(Sneath, RogersandTanimoto,Anderson,otehrs). History of the formation and development of the present group.

2/Define set (two ways) operators, equivalence, containment, bays formulae, product. Discuss axiomatic set theory and contradictions.

3. Define propositions and connectors. Truth tables and truth sets. Quantifiers. Equivalence of connectors and opperators. Necessary and sufficient conditions. Logical containment. Schools of proof. Least integer paradox

4. Numbers, binary connectors. Orderings, numbersystems, relations, functions and composition

5.Equivalence relations, partitions, definition of character and classification.

6. Counting techniques and enumeration.Probability space distributions. Independence and conditional probability

7Induction and deduction. Emperical triangle. Analogues.

8. Introdiction to computing machines. The taxir system

9/Quantification of information and its relation to communication and probability ln function and conditional information. Refinement independence and prediction.

10 Causality and prediction. Similarity coefficients Purposes of a classification.

11. Graph theory model (from our papers)

Digitized by the Hunt Institute for Botanical Documentation

12. How to interpret Graph theory model. Iterative procedure.

13. Other Models. Sokal Mitchner, camin Sokal, Wagner et al.

Grading in this course should be based on three factors. 1. The students enthusiasm and effort as a participant in the coirse.

2. The quality of each of two papers describing concepts and their application(midterm and final papers)

3 The excellence of the emperical solution to the sample problem as addressed by techniques learned in the course.

Digitized by the Hunt Institute for Botanical Documentation

3/23/66

We want to generate the ideas as best we can to make a go of it, and this again, of course, will be helped by you if you understand this course, and that your suggestions and contributions can get us off the ground a little bit better. The reason for making these sorts of statements and so forth is that this is really the first time a course like this has ever been offered, so far as I am aware, not only in the United States but anywhere in the world. I may be wrong, but if there are other courses, at least no publications have come out to indicate that this the case. We, in this course, intend to teach taxonomy for the first time. I'll tell you what I mean by that a little bit later. The problem has been in this area of classification for a long time that taxonomists have been trained that the elbows of other taxonomists sort of an apprenticeship, if you will. Where there was no real explanation of the logic involved in the course or there was no logic involved in the process given to the student, there were no sets of procedures that an individual might follow in order to become expert in the area of taxonomy. An English botanist by the name of Cain in about 1960 put it out pretty well. He said that taxonomists are trained like "performing monkeys." As of today, such procedures come to an end, I hope.

Now another difference between this course and any other course that you may have had is that we will try to point out that the kinds of mathematics that apply in this area may or may not coincide with the hallowed grounds of statistics. I'm sure that you are all aware that in the education of biologists on this campus and many other campuses, if there be programs in mathematics that the student must take in order to satisfy the requirements, he may get a course in algebra and then he will very quickly go into the areas of calculus, perhaps he will get some differential equation, but he will almost always end up with a course in statistics, biostatistics as it is frequently called in this case. Don't misunderstand me, there is reasoning for these particular types of math if you know what you are headed for and what you want to do with it. What we hope in this presentation is to discover that indeed

Digitized by the Hunt Institute for Botanical Documentation

there are other types of interesting mathematics which may have more applicability to some areas of biology than do the so-called formal analytic methods in mathematics. We do not, however, intend to make in this course a specific set of equations which you can automatically plug in, spin the wheels, get a regurgitation of an answer and accept that as the God Almighty of an answer which you as a biologist have got to accept. The point is that if you are not aware that sometimes mathematics can be just as wrong as your own biological thinking. This is one of the endeavors that we hope to make here is to point out that you have to be at least aware of what mathematics is, what it does, and how it will apply in your particular areas. So we are not going to make this a course in any particular kind of mathematics per se.

In like manner we don't expect that you will, at the end of this course, be an accomplished programmer. There is an awful lot of garbage put out these days about every biologists should be a programmer. Frankly, programming is a science and art in itself, and the kind of training that most of us as biologists have does not really qualify us as first-class competent programmers. Programming for a computer is much like any structured, very highly structured, language. Therefore, the people who are most likely to be good programmers are those who have intensely investigated the fields of some language more than we do. Our requirements in language for the advanced degrees are, as as you know, just barely enough capability in that language to pass the requirements of the German or French department. Most biologists that I know, including myself, have the minimum requirements to get through reading enough German so that the German professor will say, "Yes, this guy is competent in that language," and we really don't know a dang thing about the language when we get through with it. Let's face it. So if programming be this sort of thing, then let us not anticipate that we will be accomplished programmers at the end hereof. We will, however, I hope, at least be aware of the problems

Digitized by the Hunt Institute for Botanical Documentation

Botany 170 - 1st Period Administration Matters

1. General discussions on time. Would like to have period of open-end time on the last meeting of week--two hours anyway.

2. The schedule. No certainty that we will adhere to this--may go faster, but likely more slowly.

3. Text. Davis & Heywood, "Principles of Angiosperm Taxonomy."

Allied reading: IBM General Information Manual

Computers in Anthropology

Principles of Numerical Taxonomy

Handouts--Monograph of Uvularia

Mimeo sheets

Survey of math concepts, etc.

These are supplements--we will have to write our own book.

4. Presentation. Some lecture, much discussion--participation by all four of us, each interrupting where some pertinent point is necessary or clarification required. One person will organize each session (does not have the option of coming unprepared).

Your points, questions, discussions very important to us. Will ask for, if we do not get your points of view.

5. Grading. Two ways: 1) An informal judgment based on your efforts and discussions. 2) Reports--one due April 29, one due on exam day. First ~~last~~ one will not be more than 10 pages (typed) which is an outline and discussion of one or more topics covered--may be individually presented, or two may work together (if willing to be assigned the same grade).

Digitized by the Hunt Institute for Botanical Documentation

1. Take notes
2. Chapt 4 in D+H, ~~top 107~~ pp 110–127.

Taxometrics - Characters, Mar. 5 '68

1. Where are we - return to flow chart.
2. Characters - general
   1. Problems tax. have had concerning them.
      No clear separation between ~~plants~~ organisms + the way we think.
      Addition of different data -
      1st Morphology -
      Addition of genetic info - at first morph. of c-sameness, then inheritance patterns.
      Finer + finer analytic techniques ⇒ biochem., fine structure anatomy, DNA, etc.
      Each was intended to make possible a "definitive" classification - and molecular biologists now play the game.
   2. Can we shed some light on the problem from a different angle. George's discussions have indicated some ways to think about information generally - we clearly must define our objectives, in classif. then discover what 1) a character must do.
      2) what are the methods of using them.
   3. Definition of a character -
      a. real valued function defined over all objects in the study
      b. Characters are patterns

Note book "Papers on Plant Systematics"

Digitized by the Hunt Institute for Botanical Documentation

c. ~~Patterns~~ Characters, in psychology of learning, are mnemonic devices - memory aids - learning by association - some patterns are related.

3. Characters as we use them in classifications.

1. Character is the definition of the patterns in the study, e.g. leaf shape, c-some number, biochemical property. if all of the same general construction, fails as a useful tool in classification.
if several types of leaf shape in study, then may provide a tool -

2. The attributes, or ~~states~~, of a character are the elements of the character - defined over the study. Keeping in mind the purpose of any classification, to determine an hierarchical set of nested boxes,

Digitized by the Hunt Institute for Botanical Documentation

B 170 - 1st Period (Assign Chapters 1, 2, 3 Davis & Heywood)

I. Welcome to course

First like it anywhere from the standpoint of its biological, mathematical and computer aspects.

Differences--we intend to teach taxonomy for the first time. Taxonomists have formerly been trained (according to an English botanist, Cain) like "performing monkeys".. Such procedures end now, with this.

Another difference--the hallowed grounds of statistics as the sole mathematics type are questioned as the only source of models.

We do not intend to make computer programmers of biologists.

II. Our efforts will be to get across some methods of thinking useful in classification, rather than to train you as some sort of technician.

These processes should be useful in broader scope than in the one biological discipline of taxonomy.

III. How the concepts for this course developed

1. Intrigued by the "taxonomic process"--classification.

Correlation of information.

Discovery of interrelations of higher order.

The simple (or seemingly simple) methods in keys for identification.

Possibility of using such methods for learning in general.

2. The early efforts to classify cultures of tapioca.

Wanted to study evolution.

Classification so poor that had to "straighten out the mess"--grouping the individual varieties into some meaningful relationship.

3. Questions which bothered me in taxonomy.

What do you mean by a "species?"

Digitized by the Hunt Institute for Botanical Documentation

Multitudinous definitions--to fit special situations--Camp (1944) listed some 13-15 special categories of species, but no one could follow what he meant, except for maybe blueberries--a mess.

What do you mean when you say, "this is a primitive species?"

4. The old myth--you can't teach taxonomy--you have to learn it.
5. Can you express in some precise way the procedures used by taxonomists when they make their classifications?
6. Is there some mathematical expression which matches the needs ;of the classification process?

IV. Our objectives

1. To discover something about the taxonomic process.
2. To get some insight into the nature and methods of computers.
3. To take a look at what might be considered the mathematical way of examining the problems in this branch of biology.
4. To put what you learn to work on a taxonomic problem.
5. To give some basis for studying ; your own problems in biology.

V. The flow of work in taxonomy--chart.

A. To understand the processes in taxonomy, this flow chart tells something about the various parts.

1. Two parts to flow chart--General, and Detailed.
2. I. Taxonomy--how do you know you have a problem?

II. How do you start your solution of the problem?

How do you put your solution into a taxonomic framework?

Digitized by the Hunt Institute for Botanical Documentation

BOTANY 431
Spring 1968

A set is any collection of things, concepts, processes, etc.

A set is defined by designating what things (etc.) are in it. There are two ways of defining a set.

1. List: A list of the things in the set to be defined is given. This list is enclosed in wiggly brackets, $\{\ \}$, which are used only to define sets.
e.g. $A = \{a, b, c, d, e\}$.
A is the name given to the set of letters as listed above.

2. Rule: A rule is provided which enables us to test things to see if they belong to the set to be defined. The definition by rule is as follows:

$A = \{x/x$ is one of the first five letters in the alphabet$\}$.
This is read "A is the set of typical elements, x, such that x is one of the first five letters of the alphabet."

The first way is convenient when the set to be defined consists of a few elements, or of heterogeneous elements which do not satisfy a simple rule. The second way is convenient when the set to be defined consists of a large number of elements or when a convenient rule exists.

When X is the name of a set, the sentence, $a \in X$, read "a is an element of the set X," will be true when a is one of the things defined to belong to X.

When X and Y are names of sets the sentence, $X \subseteq Y$, read "X is a subset of Y," will be true when every element of X is also an element of Y; in other words, "$a \in X$ and $a \notin Y$" is false.

By convention ϕ denotes the set with no elements. For any set X, $\phi \subseteq X$.

$A = B$ means A is another name for B.
$A = B$ means A has exactly the same elements as B.
$A = B$ means $A \subseteq B$ and $B \subseteq A$.

A set may have for its elements anything - in particular a set may have other sets for its elements. A set with only one element is called a singleton; $\{a\}$ is read "singleton a."

$$\{a\} \neq a.$$

Digitized by the Hunt Institute for Botanical Documentation

If X is a set, then the symbol, $|X|$, stands for, (is a name for) the number of elements in X. $|X|$ is a number, not a set.

2^X is the name of a set whose elements are subsets of X.

$2^X = \{y/y \text{ is a subset of } X\}$

$|2^X| = 2^{|X|}$.

Closed binary operations on sets.

An operation is a proceedure which converts things to other things. Binary means the operation converts two things to one thing.
Closed on sets means the operation converts two sets to some other set.

Union, $\cup$, is a closed binary operation.
If A and B are sets, then $A \cup B$ is a set defined as follows:

$$A \cup B = \{x/x \in A \text{ or } x \in B\}$$

Intersect, $\cap$, is a closed binary operation. If A and B are sets, then $A \cap B$ is a set.

$$A \cap B = \{x/x \in A \text{ and } x \in B\}.$$

Relative complement, -, is a closed binary operation. If A and B are sets, then $A - B$ is a set and

$$A - B = \{x/x \in A \text{ and } x \notin B\}$$

We may construct complicated names for sets by combining these operations. However, whenever ambiguities as to the order in which operations should take place, parentheses must be used.

$A \cap (B \cup C)$ is unambiguous, $B \cup C$ is the name of the set with which A is intersected.

$(A \cap B) \cup C$ is unambiguous. $A \cap B$ is the name of the set with which C is unioned.

$A \cap B \cup C$ is ambiguous.

Commutativity. $A \cup B = B \cup A$, i.e., the symbol may be turned around without changing its value. Thus the operation, $\cup$, is said to be commutative. Similarly $A \cap B$. $A - B$ is not commutative.

Digitized by the Hunt Institute for Botanical Documentation

Distributivity. In arithmetic we know that whenever x, y, and z are numbers, $x(y + z) = xy + xz$. We say multiplication distributes over addition. We also know that $x + yz \neq (x + y)(x + z)$, and we say that addition does not distribute over multiplication. Addition and multiplication are two closed binary operations on numbers. Intersect and union are two closed binary operations on sets. It is true that

1. Intersect distributes over Union: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. Union distributes over Intersect: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Whenever a set S is designated as the totality of things of interest to us within the context of some consideration, and we wish to work only with sets which are subsets of S, i.e. are members of 2^S, we define $\sim A$ to mean S - A, and we read $\sim A$ as "the complement of A". $\sim(\sim A) = A$.

Bays formulae 1. $\sim(A \cap B) = \sim A \cup \sim B$
2. $\sim(A \cup B) = \sim A \cap \sim B$.

Boolean Algebra is the study of the subsets of a set, including these rules for discovering different names for the same set. E.g.

$\sim(C \cap \sim(B \cap \sim A))$ is the name of a set whenever C, B and A are subsets of some set S. If we read "=" as "is another name for" the following is an example of manipulations in Boolean algebra.

$$\begin{aligned}
\sim(C \cap \sim(B \cap \sim A)) &= \sim(C \cap (\sim B \cup A) \\
&= \sim((C \cap \sim B) \cup \sim(C \cap A)) \\
&= \sim(C \cap \sim B) \cap \sim(C \cap A) \\
&= (\sim C \cup B) \cap (\sim C \cup \sim A) \\
&= \sim C \cup (C \cap \sim A) \\
&= \sim(C \cap \sim(B \cap \sim A))
\end{aligned}$$

It will not be necessary to become expert at this kind of manipulation but some knowledge of the rules which permit it is desired.

Digitized by the Hunt Institute for Botanical Documentation

Numbers: Natural numbers, $\{1, 2, 3, \ldots\} = J$

Integers, $\{\ldots -2, -1, 0, 1, 2, \ldots\} = Z$

Q = Rational numbers = $\{x/x = p/q,\ p \in Z,\ q \in Z,\ q \neq 0\}$

$(A =)$ Algebraic numbers are numbers which when substituted for "x" in sentences such as

$$C_n x^n + C_{n-1} x^{n-1} + \ldots + C_0 x^0 = 0$$

where C_n, C_{n-1} etc. are Rational numbers, make the sentence true.

R = Real numbers = $\{r/r$ is an infinite sequence of integers with one decimal point$\}$

There are many types of numbers; for our purposes, rational numbers will usually suffice. The following inclusions are true:

$$J \subseteq Z \subseteq Q \subseteq A \subseteq R$$

Numbers as names:

Integers and natural numbers will often be used by us as names, or labels or indices. Here we use only the symbolic and order properties of integers. It may be useful, for example, to refer to 5 sets which we might call A_1, A_2, A_3, A_4, A_5.

We may denote these five sets by A_i, $i = 1, 2, \ldots, 5$.

More abstractly we may wish to refer to n sets where n is an integer. In this case we write A_i, $i = 1, 2, \ldots, n$.

This notation is useful in expressing repeated operations, e.g.

$A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5$ may be written $\bigcup_{i=1}^{5} A_i$.

Similarly, $\bigcap_{i=1}^{5} A_i$. We may now express Bays formulae more generally as:

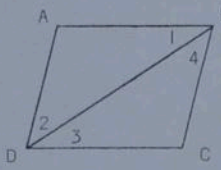
$$\sim \bigcap_{i=1}^{n} A_i = \bigcup_{i=1}^{n} \sim A_i$$

and

$$\sim \bigcup_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \sim A_i.$$

Digitized by the Hunt Institute for Botanical Documentation

We must be careful not to confuse the numerical prpperties of integers with their naming and sequential properties. You may be amused by a "proof" which was once submitted to me by a student.



ABCD is a parallelogram. The numbers 1, 2, 3, 4 are used here as names for angles ABD, ADB, BDC, and DBC, respectively. We want to prove angle ADC is equal in magnitude to angle ABC. The "proof" goes as follows.

$\angle 2 = \angle 4$ Alternate interior angles

$\angle 3 = \angle 1$ Alternate interior angles

$ADC = \angle 2 + \angle 3 = \angle 5$

$ABC = \angle 1 + \angle 4 = \angle 5$

$ADC = \angle 5 = ABC.$

This student's confusion should be clear.

Product: $\times$ Product is a closed binary operation on sets, defined as follows: A and B are sets.

$$A \times B = \{(a,b) / a \in A \text{ and } b \in B\}.$$

This is read "the product of A and B is the set of pairs, (a,b) such that a is an element of A and b is an element of B." The elements of $A \times B$ are ordered pairs of elements. In our discussions (a,b) is not a set. In particular $(a,b) \neq \{a,b\}$.

Digitized by the Hunt Institute for Botanical Documentation

Propositional Calculus

In our examination of naive set theory we concerned ourselves with a class of things called sets, and with some closed operations (unary and binary) on this class. The result was Boolean Algebra. We now wish to examine the class of propositions (sentences which assert something), together with some closed operations on this class.

Examples of propositions:

The sun is shining. Bot 431 is great.

We will concern ourselves with one unary operation and four binary operations. They are:

1. $\sim$ "not" or "it is not the case that" Unary
   It is not the case that the sun is shining, or $\sim$The sun is shining.
2. $\wedge$ and. Binary. The sun is shining and Bot 431 is great.
3. $\vee$ or. Binary. The sun is shining or Bot 431 is great.
4. $\leftrightarrow$ means. Binary. The sun is shining means Bot 431 is great.
5. $\rightarrow$ implies. Binary. The sun is shining implies B 431 is great.

Of the four binary operations, 2, 3, 4, 5, only $\rightarrow$, implies, is not commutative.

It is often convenient to have symbolic names for sentences. Letters are usually used. Let p, q, and r be names of sentences. $\sim p$, $p \wedge q$, $p \vee q$, $p \leftrightarrow q$, $p \rightarrow q$ would then become names for other sentences.

Statements built from other statements by means of these operations are called compound statements. The truth of a compound statement depends on the truth of the statements that compose it. A convention known as truth tables illustrates this well. The operators themselves are defined using truth tables.

A compound statement made up of only one proposition, such as $\sim p$, has a truth table as shown.

Truth of p	Truth of $\sim p$
T	F
F	T

There is a column for p, where is entered the possibilities for the truth of p, from which the truth of not p is determined.

The truth tables of other compound statements involve two or more distinct propositions. We usually head columns in truth tables with "p" instead of

Digitized by the Hunt Institute for Botanical Documentation

"Truth of p" and connectors or operators instead of the whole compound statement. This precedent enables us to write

p	$\wedge$	q
T		T
T		F
F		T
F		F

Here each column is four symbols high to allow for the four possible combinations for the truth of p, and the truth of q. In the center column, headed by $\wedge$, will be entered the truth of the compound statement, $p \wedge q$, as shown below.

p	$\wedge$	q
T	T	T
T	F	F
F	F	T
F	F	F
1	2	1

The columns labeled 1 were done first, and those labeled 2 were done next.

The truth tables for the other connectors are given.

p	$\vee$	q
T	T	T
T	T	F
F	T	T
F	F	F
1	2	1

p	$\leftrightarrow$	q
T	T	T
T	F	F
F	F	T
F	T	F
1	2	1

p	$\rightarrow$	q
T	T	T
T	F	F
F	T	T
F	T	F
1	2	1

Two compound sentences are the same (i.e. different names for the same sentence) if their truth tables are the same.

Truth tables of more complicated statements may be constructed from these basic five. For example, consider the statement,

$$p \wedge \sim(q \vee \sim r).$$

There are eight possible combinations in which the sentences p, q, r could be true or false, so the columns in the truth table should be 8 symbols high. Again the numbers give the order in which the columns must be built.

Digitized by the Hunt Institute for Botanical Documentation

p	∧	~	(q	∨	~	r)
T	F	F	T	T	F	T
T	F	F	T	T	T	F
T	T	T	F	F	F	T
T	F	F	F	T	T	F
F	F	F	T	T	F	T
F	F	F	T	T	T	F
F	F	T	F	F	F	T
F	F	F	F	T	T	F
1	→ 5	← 4	1	→ 3	← 2	← 1

The last column, 5, gives us the truth of the whole statement. $p \wedge \sim(q \vee \sim r)$ is true only if q is false but p and r are true.

Any sentence which is always true no matter what the truth of its constituents is said to be logically true. A sentence which is always false is called logically false. $p \wedge \sim p$ **is logically false. The following are logically true:**

$(p \rightarrow q) \leftrightarrow (q \vee \sim p)$

$(p \leftrightarrow q) \leftrightarrow ((p \wedge q) \vee (\sim p \wedge \sim q)$

$\sim(p \wedge q) \leftrightarrow (\sim p \vee \sim q)$

$\sim(p \vee q) \leftrightarrow (\sim p \wedge \sim q)$

If p is a statement, then there are some states of affairs for which p is true, some for which p is false, and some which have nothing to do with p. If S is the set of states of affairs for which p is true or p is false, then p is said to be truth functional on S. The subset P of S on which p is true is the truth set of p. This concept of truth set establishes an analogy between Boolean algebra and propositional calculus. Let q be truth functional on S as well. If P is the truth set of p and Q is the truth set of q, the truth sets of compound statements are as follows.

Statement	Truth Set
$p \wedge q$	$P \cap Q$
$p \vee q$	$P \cup Q$
$p \rightarrow q$	$\sim P \cup Q$
$p \leftrightarrow q$	$(P \cap Q) \cup (\sim P \cap \sim Q)$
$\sim p$	$\sim P$

Digitized by the Hunt Institute for Botanical Documentation

Necessary and Sufficient Conditions.

If p and q are two statements and p cannot be true unless q is then we say that q is necessary for p. If P and Q are the Truth Sets for p and q, then $P \subseteq Q$.

If p is always true when q is, then we say that q is sufficient for p and we have $Q \subseteq P$. If p and q mean the same thing then q is sufficient for p and q is necessary for p, $Q \subseteq P$, $Q \supseteq P$. Thus $P = Q$. Notice that p is sufficient for q whenever q is necessary for p.

If q is necessary and sufficient for p, q is called a characterization of p or sometimes an equivalent of p.

Digitized by the Hunt Institute for Botanical Documentation

A relation on a set may be thought of in many ways. Formally, a relation, R, on a set A is a subset R of A $\times$ A. In other words, a relation R on the set A is a collection of ordered pairs (a, b) with a $\in$ A, and B $\in$ A. Remember (a, b) $\neq$ (b, a) $\neq \{a, b\}$. Less formally a relation may be thought of as a rule which tells when two elements of A are related.

Example: $A = \{x/x \text{ is an integer}\}$

R = "is divisible by"

6 R 3 may be read: 6 is divisible by 3.

6 R 3 is a true sentence. We say that (6, 3) is in the relation R. If we think of $R \subseteq A \times A$ we may say

$(6, 3) \in R$.

12 R 5 is a false sentence, and we may write $(12, 5) \notin R$.

Example: $A = \{x/x \text{ is a student of B 431}\}$

R = "is taller than"

Bob R Suzan is a true sentence. (Bob, Suzan) $\in$ R.

Properties of Relations

Reflexive: R is reflexive on A if a R a is a true sentence for all $a \in A$.

Symmetric: R is symmetric on A if a R b $\rightarrow$ b R a is a true sentence for all and b in A.

Antisymmetric: R is antisymmetric on A if (a R b $\wedge$ b R a) $\rightarrow$ a = b is a true sentence for all a and b in A.

Transitive: R is transitive on A if (a R b $\wedge$ b R c) $\rightarrow$ a R c is a true sentence for all a, b and c in A.

Partial Order: R is a partial order for A if R is Reflexive, Antisymmetric and transitive on A.

Linear Order: R is a linear order for A if R is a partial order for A and (a, b, in A) $\rightarrow$ (a R b $\vee$ b R a) is a true sentence.

R is an equivalence relation on A if R is reflexive, symmetric and transitive on A. An equivalence relation is a generalization of "=".

A partition for A is a Set of Subsets of A, $\{C_1, C_2, \ldots, C_n\}$ with two properties:

1. $\bigcup_{i=1}^{n} C_i = A$ (exhaustive)

2. $C_i \cap C_j = \emptyset$ if $i \neq j$ (exhaustive)

A partition for A defines an equivalence relation, R, on A as follows: a R b is true whenever a and b belong to the same member of the partition and false otherwise. The members of a partition are called the classes of the partition.

Digitized by the Hunt Institute for Botanical Documentation

Conversely, every equivalence relation R, on A, defines a partition of A. The classes of this partition are called the equivalenceclasses of R. A classification for A is a series of partitions $P_1, P_2, \ldots, P_m$ for A such that if C_i is a class in Partition I then for $j > i$, there is exactly one class C_j in Partition j such that $C_i - C_j = \phi$ and all other classes D_j in Partition j are such that $C_i - D_j = C_i$.

In other words a classification for A is a series of Equivalence Relations for A: $R_1, R_2 \ldots R_m$ such that

$$R_m \supset R_{m-1} \supset \cdots \supset R_2 \supset R_1.$$

A Function is three things,

1. A set called the domain of arguments of the function.
2. A set called the range of values of the function.
3. A rule which associates with each member of the domain a unique member of the range.

Notation. Sometimes the rule is symbolized with a letter, like f, in which case whenever a symbol, like "a", is used for an argument, f(a) stands for the value which the rule f assigns to "a".

A character for a Study S of objects to be classified is a function with

1. Domain S
2. Range a small (less than a dozen or so usually) set of descriptions
3. A rule which assigns to each object in S the value in the range which best describes it.

This very general definition of a character gives you much structural freedom in defining characters. However, this means that you are responsible for developing a good technique to take advantage of this freedom.

Digitized by the Hunt Institute for Botanical Documentation

Functions

I. A relation F is a function if $[(a,b) \in F$ and $(a,c) \in F \Rightarrow b = c]$ is true.

Suppose a relation F is a function.

Then 1. $D = \{x/x$ is the first member of some element of $F\}$ is called the Domain of arguments of F and

2. $R = \{x/x$ is the second member of some element of $F\}$ is called the Range of values of F.

II. Another definition of function is: A function, f, is three things

i. a set D called the Domain of arguments for S

ii. A set R called the Range of values of f

iii. A rule which associates with every element of the Domain some unique element of the Range.

If f is a function and a is an argument then f(a), read "f of a" is a symbolic name for the value which the function f associates with the argument a.

Let $f:D \to R$ be a function. This notation means that D is the Domain, R is the range and f is the symbol for the functional association.

Suppose $A \subseteq D$, by f(A) we mean $\{x/$for some $a \in A$, $f(a) = x\}$. $f(A) \subseteq R$ and is called the image of A under f. f is said to be <u>onto</u> R is f(d) = R. Otherwise f is said to be into R.

Let f be a function in the relation sense. Then f^{-1} is the relation:

$(a,b) \in f^{-1} \Leftrightarrow (b,a) \in f$.

The inverse of f is f^{-1}. Whenever both f and f^{-1} are functions, we say that f is one-to-one. Otherwise we say that f is many-to-one.

Functional Composition.

If $f:D \to R$, and $g:D' \to G'$ and $R \subseteq D'$ we may define a new function $g \circ f:D \to R'$ as follows: $g \circ f(x) = g(f(x))$.

If f is 1 - 1 then $f^{-1} \circ f(x) = f \circ f^{-1}(x) = x$.

Any function, I for which $I(x) = x$ for all arguments x in its domain is called the identity function, $f^{-1} \circ f = I$.

Whenever $B \subseteq R$ and $f:D \to R$ we may define $f^{-1}(B) \subseteq D$ even if f is not 1-1.

$f^{-1}(B) = \{x/f(x) \in B\}$ is called the inverse image of B under f.

If $f : D \to R$ and $R = \{r_1, r_2, \ldots, r_n\}$ then the collection of subsets of D $f^{-1}(\{r_1\})$, $r = 1, 2, \ldots, n$ is a partition for D.

Digitized by the Hunt Institute for Botanical Documentation

STRUCTURE OF A CHARACTER

1. A character is a basis for comparison together with a rule for deciding when any two objects in the study are similar or different with respect to the basis for comparison in question. This rule is an equivalence relation the equivalence classes of which constitute the states of the character.

2. A Character is a classification of the study based on a single criterion.

3. A Character is a function, f, with Domain the study to be described, and with Range a small set of descriptions. The rule f associates with each object in the study the description in the range which best describes it. The respective inverse images of each description in the range constitute the states for the character.

INFORMATION Any system which, at any given time, may exist in one of a number of distinct states contains information proportional to how difficult it is to guess the unknown state of the system without actually observing it. Information quantity is measured in units of yes-no questions. If the guesser asked one who knew about the system in question enough yes-no questions to discover the unknown state, and if the questions were asked in the most clever way possible, the information in the system would be equal in quantity to the average number of such yes-no questions asked.

There are two things which determine this number of yes/no questions (i.e., the quantity of information in a system). These are

1. The number of possible states for the system
2. The probability with which any of these states is known to actually occur.

Any system whose states are known and for which the probability of those states are also known admits a quantification of its information. If K symbolizes such a system H(K) will represent the amount of information in that system.

INFORMATION IN A CHARACTER

In order to measure the information content of a character it is necessary to interpret that character as an information containing system. Think of a character in the function way with the study as the domain and a small set of descriptions as the range. The states of the character become the states of the system when the character is viewed as a system.

Let the character be called K and let K_1 represent the first state, K_2 the second state, etc. The probability of K_1 is $|K_1|/|S|$, the probability of K_2 is $|K_2|/|S|$, etc. We may now express (symbolize) the information in character K as H(K).

Digitized by the Hunt Institute for Botanical Documentation

Dependence of Information Containing System

Assume we have two information containing systems, the first a fair coin with two states, head and tails, and the second a fair die with six states, one dot, two,dots, ..., and six dots. If we flip the coin and roll the die and observe the coin it doesn't help us guess about the die. Similarly observing the die does not help us guess about the coin. In cases such as this for which observing one system does not help us guess about some other system, we say the two systems are independent. They do not influence each other.

Any two systems which are such that observing the state of one helps us to guess the state of the other are called dependent.

Let A and B be systems. Let A have n possible states and let B have m possible states. By the cartesian product of the two systems, written $A \times B$, we mean the system with $m \cdot n$ states each of which is an ordered pair, (a,b) with "a" a state of A and "b" a state of B.

If A and B are independent systems than $H(A) + H(B) = H(A \times B)$. This means that it takes as much information to determine the state of A and the state of B separately as it does to determine both states at once. This is because knowledge of A does not help us guess about B. If A and B were dependent then knowledge of A would help us guess about B. In this case $H(A) + H(B) > H(A \times B)$ because on the left we must discover the states of A and B separately while on the right after discovering the state of A, we now find it easier to discover the state of B. Since, on the right, we still must at least discover the state of A, we know $H(A \times B) \geq H(A)$, and $H(A \times B) \geq H(B)$.

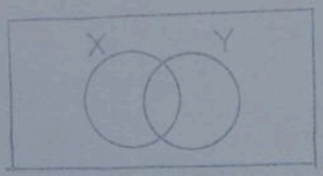
The symbol H(A/B) measures how difficult it is to guess A after observing B. If A and B are independent H(A/B) = H(A). This says that it is just as hard to guess A after observing B as before which is what independent means. In general:

$$H(A/B) = H(A \times B) - H(B).$$

R(A,B) may be thought of as the amount of information in common between A and B. $R(A,B) = H(A) + H(B) - H(A \times B)$. From this it should be apparent that R(A,B) = R(B,A).

We may now envision information in what is called (by me) an heuristic information space. In this space we represent information as subsets. Set X corresponds to the information in A. Set Y corresponds to the information in B.

Digitized by the Hunt Institute for Botanical Documentation



$R(A,B) = X \cap Y$

$H(A/B) = X - Y$

$H(A \times B) = X \cup Y$

As an indication of dependence we define

$$D(A,B) = \frac{H(A/B) + H(B/A)}{H(A \times B)}$$

$D(A,B) = 1$ if A and B are independent.

If $H(A/B) = 0$, then B is a refinement of A. In the heuristic information space, $Y \geq X$.

If $H(A/B) = 0$ and $H(B/A) = 0$, then A and B are identical. In the heuristic information space $X \leq Y$ and $Y \leq X$ and thus $Y = X$.

$D(A,B) = 0$ only if A = B (or A is identical with B).

In general $0 \leq D(A,B) \leq 1$ and this indicates the extent of the dependence.

Digitized by the Hunt Institute for Botanical Documentation

Relations (Continued)

Recall a relation R on S is

1. A subset of $S \times S$, $R \subseteq S \times S$ or
2. A rule "is related to" If a,b are each in S, $(a,b) \in R$ whenever "a is related to b." is true.

The Graph of a Relation

Any relation can be represented by a directed graph. Arrange the elements of S in a convenient way; connect a to b with an arrow from a to b whenever $(a,b) \in R$.

If R is symmetric $(a,b) \in R \rightarrow (b,a) \in R$ in which case the two arrows between a and b may be replaced by a single line.

If R is reflexive then $(a,a) \in R \ \forall \ a \in S$. In this case the looped arrow which corresponds to (a,a) may be omitted.

If $T \subseteq S$ and R a relation on S then we may define a relation R' on T in a natural way. $(a,b) \in R'$ whenever $a \in T$, $b \in T$, and $(a,b) \in R$. T is said to inherit its relation from the relation R. Notice that $R' \subseteq R \subseteq S \times S$ so that R' may be thought of as a relation on S as well.

Let G be the graph of the relation R on S. Let $T \subseteq S$. Let G' be the graph of R' on T where R' is the relation which T inherits from R. Then G' is called a subgraph of G.

Closures of a Relation

Let R be a relation on S. Q is larger than R if $|Q| > |R|$. Notice that if $Q \supset R$ then $|Q| > |R|$.

The reflexive closure of R is the smallest reflexive relation Q such that $R \subseteq Q$.

R fails to be reflexive on S if there is some $a \in S$ such that $(a,a) \notin R$. Thus the reflexive closure of R is R together with just enough additional pairs of the form (a,a) to make this new relation reflexive.

The symmetric closure of R is the smallest symmetric relation Q such that $Q \supseteq R$.
The transitive closure of R is the smallest transitive relation Q such that $Q \supseteq R$.

Digitized by the Hunt Institute for Botanical Documentation

Connected Graphs

Let R be a relation on S. Two elements a,b of S are said to be connected if there is a sequence

$$a = p_1, p_2, p_3, \cdots p_m = b$$

of elements in S such that

$$p_i \, R \, p_{i+1} \quad \text{for} \quad i = 1, 2, \ldots m-1.$$

Such a sequence will be called a chain from a to b.

The graph of a symmetric relation on S is said to be connected if any two elements in S are connected.

Let G be the graph of a relation on S. Let G' be a subgraph of G. Let p be the statement "G is connected" and q be the statement "G' is connected." None of the following statements are true in all cases.

$$p \wedge q$$
$$p \wedge \sim q$$
$$\sim p \wedge q$$
$$\sim p \wedge \sim q$$

Let G' be connected. If every other subgraph H, of G, such that $H \supseteq G'$ is not connected, then G' is a maximal connected subgraph of G. (G' is sometimes called a component of G.)

If G is the graph of a relation R on S then the maximal connected subgraphs of G partition S. In this way each relation R on S corresponds to an equivalence relation on S. This equivalence relation will be the symmetric, reflexive, transitive closure of R, and its equivalence classes will be the maximal connected subgraphs of the graph of R.

Digitized by the Hunt Institute for Botanical Documentation

Bot. 431 Bibliography

GENERAL DISCUSSIONS

R. E. Blackwelder, 1964, Systematics Association Publication #6, pp. 17-28
"Phenetic and Phylogenetic Classification"

E. Mayer, 1965, SSZ 14:2, pp. 73-97
"Numerical Phenetics and Taxonomic Theory"

P. H. A. Sneath, 1956, J. Gen. Microbiol. 17, pp. 184-200
"Some Thoughts on Bacterial Classification"

C. D. Michener and R. R. Sokal, 1957, Evolution 11:2, pp. 130-162
"A Quantitative Approach to a Problem in Classification"

R. R. Sokal, 1962, J. Theor. Biol. 3, pp. 230-267
"Typology and Empiricism in Taxonomy"

G. H. Ball, "Data Analysis in the Social Sciences"

CHARACTERS AND SIMILARITIES

D. C. Eades, 1965, SSZ 14:1, pp 98-99
"The Inappropriateness of the Correlation Coefficient as a Measure of Taxonomic Resemblance."

R. R. Sokal, 1961, SSZ, 10:2, pp. 70-79
"Taxonomic Structure from Randomly and Systematically Scanned Biological Images"

W. R. Lockhart, 1967, Iowa State University
"Hierarchical Characters"

D. W. Goodall, 1966, Biometrics 22:4, pp. 882-907
"A New Similarity Index Based on Probability"

STATISTICAL METHODS (some examples)

J. C. Gower, 1966, Biometrika 53:3 & 4, p 325 seq
"Some distance properties of latent root and vector methods used in multivariate analyses"

F. J. Rolf, 1965, Taxon, 14:8, pp 262-267
"A Randomization Test of the Nonspecificity Hypothesis in Numerical Taxonomy."

R. R. Sokal and C. D. Michener, 1958, U. Kan. Sci. Bull. 38: 22, pp 1409-1438
"A Statistical Method for Evaluating Systematic Relationships"

Digitized by the Hunt Institute for Botanical Documentation

STRUCTURE OF BIOLOGICAL CLASSIFICATION

A. Rescigno and G. A. Maccacaro, 1960. In Information Theory, Colin Chenny, Ed.
"Information Content of a Biological Classification."

R. C. Buck and D. L. Hull, 1966. SSZ 15:2, pp. 97-111
"Logical Structure of Linnean Hierarchy"

J. H. Camin and R. R. Sokal, 1965, Evolution 19, pp. 311-326
"A Method for Deducing Branching Sequences in Phylogeny."

CLASSIFICATION METHODS

J. Rubin, 1965 IBM Corp. Pub.
"Optimal Classification into Groups"

M. Wirth, G. Estabrook, D. J. Rogers, 1966. SSZ, 15:1 pp 59-69
"A Graph Theory Model for Systematic Biology"

P. H. A. Sneath, 1957, J. Gen. Microbiol. 17, pp 201-226
"Application of Computers to Taxonomy"

W. W. Moss, 1967, SSZ 16:3, pp 117-207
"Some New Analytic and Graphic Approaches to Numerical Taxonomy"

R. E. Bonner, 1964, IBM Jour. of Res. and Dev., 8:1, pp22-
"On Some Clustering Techniques"

D. J. Rogers and T. T. Tanimoto, 1960, Science 132:3434, pp 1115-1118
"A Computer Program for Classifying Plants"

L. Silvestri, et al., 1962, Symposium of the Society for General Microbiology
12: pp 333-360

Digitized by the Hunt Institute for Botanical Documentation

Take a look at this Paper

Glen

TAXIMETRICS COURSE OUTLINE

This course is designed for students in the Biological and social sciences. Its purpose is to present to students in these feilds some of the fundamentalx concepts of modern mathematics and to illustrate the power of these concepts for more productive and objective thinking within their chosen feilds.It is not our desire to make mathematicians or even mathemstical technitians of our students. It is our desire to expose tomorrows scientists to the precision, discipline and generality of mathemstical thought. Our higher purpose is less to present the emperical (scientific) discoveries of the past and more to suggest ways of thinking about science in the present and future.

The outline presentdd here is for a five hour semester course of 14 weeks duration of which one week is reserved for testing and papers. The remaining 13 weeks constitute the 13 units in the outline. The 5 hours of credit are broken into 3 hours of lecture and 4 hours of lab per week. The lectures are of a conceptual and methodological nature while the lab is of a practical and applied nature..The analogue between the conceptual and methodological principles of the lectures and their application to the several diciplines represented by the interests of the students will be made in the lab. Thus the lectures should be given by one more mathematically inclined and the labs should be conducted by one more experienced in emperical science

Digitized by the Hunt Institute for Botanical Documentation

involved in programming. What are they? And how do you go about getting together to solve some of these problems? I guess you might say that our efforts in this course are to get across some methods of thinking which may serve you usefully, not only to become a classifier or taxonomist, but such that it will be a useful process no matter what your particular endeavor happens to be. There are some basic procedures, methodologies that occur in the process of classification that we think are extremely valuable no matter what your endeavor turns out to be. It doesn't have to be a scientific endeavor. It can be one in English, it can be one in the arts, or other areas. It is just a thought process that we hope to get across. You might be interested to know how we got into this business. Something of the sort of a background of how it all started.

I started out a number of years ago myself to become a taxonomist just because of the fact that which grabbed me and started me on my graduate program was the fact that there is an awful lot of green stuff out there. Now there is all kinds of ways of studying that green stuff out there, but one of the ways that you might study it is to find out what kinds of green stuff is there out there, and this is the direction that grabbed my fancy very early. The gad fly for me was the fact that the plant materials out there were interestingly different or intriguing alike, either way. And this led me personally into the investigations of taxonomy as a career. But after I had gotten into the field a little bit and was able to consider some of the presentations of things, it soon became clear to me that nebulous thing called a taxonomic process had some intriguing possibilities in it, the business of the correlation of information, pieces of information about the plants that got put together in some way in order to gain more meaning about those plants. I kept asking myself the question, "Well how do you put this information together so that you get some meaning together for classification purposes. Obviously, there isn't any information correlation going on in taxonomy. Furthermore, you can discover that if you

Digitized by the Hunt Institute for Botanical Documentation

have a classification you get some sort of higher order of meaning from the classification if you read the classifications correctly, if you try to understand them. We just don't know how far any individual can go in his interpretation of the classification once it performed. Many/individuals have gained some sort of ability to different carry on a little bit farther the kind of thing that they are interested in doing if they have a classification available to them. You may be a geneticist, for example, or a plant breeder. And if it is interesting to you to know what kooks of organisms you might possibly put together to gain some new genes system or to improve a crop land, or to make a double-flower petunia, and so forth. You might figurr out how to go about it if there be no double-flowered petunia to make one by looking at a classification scheme and discovering what other things are there besides petunia in the petunia group. As a matter of fact, I don't even know the scientific name of petunia right now. But at any rate, this sort of thing you can discover. Some sort of interrelationship of higher order of information that you may wish yourself to take off on. Furthermore, if you look at a classification an output from a classificatory operation, you will discover that a part of it is a sort of a very simple method of how to find a particular piece of information in a whole mess of information, namely, that which is called a ~~dingomistic~~ diagnostic key, which is a nice neat little technique for going through a very large / population of things to get down to an individual. This sort of technique which is used by biologists, some with great success and some;with less success. At least the idea is involved there. So I decided that this was an interesting sort of area that I was interested in.

There is a possibility of using these methods that taxonomists use in the whole process of learning. It seems to me that what happens when a taxomomist is operating is that he is making judgments on the value of pieces of information in a sort of comparative way. How does this fit with this one? What is the likeness in the similarities, if you will, between two particular pieces of information?

Digitized by the Hunt Institute for Botanical Documentation

Some where in here it seemed to me that maybe we could go in and dig out those nuggets of methodologiets by which a person learns in general and by which he might learn better if he had something that we could dig from, this taxonomic process.

Another thing that started me on direction here was my own interest in research. Some of you may know and some of you may not know that the thing that I have stuck with the longest and had the least success with so far is the classification of a group of tropical root crops, or one species of a tropical root crop, a plant which is most commonly known to you perhaps by one of its products, namely, tapioca. The scientific name for tapioca plants is Manihot esculenta. ~~Thxxx~~ This particular species is of fundamental importance as a carbohydrate crop across the tropics of the world. There are no tropical areas except those at high altitudes where at least some tapioca plants are raised. There is an infinite number of variations that these plants have, but there are many, many variations that are possible. A very large number of expressions of small differences between the varieties of these plants in some place or other. My original point of take off for this study was the fact that I was interested in the study of evolution, and I had read somewhere back earlier what Darwin said about the studies of evolution. He said that if you really want to study evolution what you really ought to do is study the plants that man uses, because it is in these groups of organisms where man has become interested in them that you will find the most explosive rates of evolution occurring, and that you might be able to discover from that something about the evolutionary process. So I naively began by selecting Manihot esculenta as my tool for the study of evolution. I soon discovered that I couldn*t make a good study on the evolution of these plants unless I knew something about their organization or classification. I had to have some handles to grab these plants with, and I discovered that not only was tapioca an unknown as far as ~~inxx~~ its classification is concerned, I was also greatly discouraged by the body-taxonomist. I went to Harvard

Digitized by the Hunt Institute for Botanical Documentation

when I was first beginning these studies, and I told a man who is now deceased, (his name shall not, therefore, be brought before you because it isn't fair to him, since he is dead and cannot talk back) I said, "I want to study classification of Manihot esculenta." He said, "Your nuts, your absolutely nuts, you can't do anything with the classification of that bunch of things, particularly a cultivated plant." He couldn't have done a better job to set me on course by waving that kind of flag in my face. But if you stop to think how you are going to study these groups of plants such that you can make some sort of a reasonable classsification of them, you soon learn of the great difficulties, because the variations between plants is so slack from one plant to the next that you have a great deal of difficulty in recalling the total variation that one plant has in relation to the next plant, and then by the time you have gotten through the second, third, or fourth, or fifth, or maybe the hundredth plant, you have forgotten entirely what the content of variation was in the first plant. So to try to make these comparisons, you are in great need of some sort of extension of your arm in order to get into this kind of study. A very factual problem generated some of the ideas which will be going into this course. Along with these ideas the questions of in taxonomy itself that were hallowed words, hallowed ideas, and so forth bugged me for a long time. It seemed when I first got to be a student in taxonomy that the professor somehow knew what a species was, because the professors were guys that were generating new species, they were naming new species all the time, and I always wanted to know, "how in the name of did you know that this thing was a new species, different from anything else that is presently . But I discovered very quickly that some of the professors were quite honest in saying, "Well, I don't know what this new species is, and I don't know whether this thing here is a new species, all I know is that in my study of the groups of organisms I haven't seen anything that looks quite like it. Therefore, it is a new species." How do you grab that definition then? How could you give that to a student and expect that he is

Digitized by the Hunt Institute for Botanical Documentation

going to be willing to accept that? And how is he going to go about it when he gets out away from you? How is he going to find out how to do this thing himself? And most of the answers you get was a bunch of hand waving. Well, you have to learn how. I say, "well teach me. If you can learn it, you must be able to teach it". At which point we got some more hand waving. As a result of that and this sort of questioning kept goading me right along. I couldn't answer it. Neither could anybody else answer this question. If you look through the literature of what is the species you will find a massive set of definitions that individuals have proposed at one time or another. Most of these are definitions which they have generated after having looked at a particular group of organisms, and decided that those other definitions out there don't fly for this particular group of organisms; therefore, I have got to set up a new set of definitions. That strikes me as being sort of a poor way to be a scientist, and that there must be something else, some better way for us to get at it. Incidentally, I don't get definition of a species nor will I present one to you, but I hope that by thinking about these tax input that we are going to do, that maybe somewhere, sometime, some reasonable methodology for discovering how to go about such definitions will be forthcoming.

Furthermore, this sort of thing leads me on to the business of: Can you express some precise way the procedures used by taxonomists when they make their classifications? If you read the output from taxonomists, you see a classification, but you don't see how they got to it. They may, for example, make such statements in their published results that this is a good species. What is meant by good species? Or they may say this is a primitive species in a genus. How do you go about finding out what is primitive and what is not primitive? Are there any ways in which a person can precisely define what he is talking about when he says these words? Probably there are. Another thing that goaded me into this sort of thinking about this kind of a course and something that may grab your fancy as well is if you really stop and think about taxonomic process and the way people talk about taxomomy and the way they operate taxonomy, you may discover that their thought

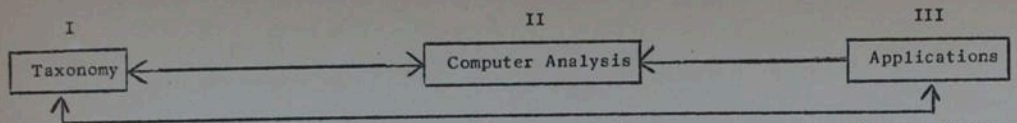
Digitized by the Hunt Institute for Botanical Documentation

processes follow a logic which has some similarity to the kinds of logic and thinking that a mathematician does. Although the terminology and methods of expression are quite different, you may discover that between the two are the most allied thing that you can think about between the discipline of taxonomy and any other discipline is a mathematical procedure of thought. It is not in taxonomy, this procedure is not necessarily as well defined nor is it as rigorous as you can find in mathematics. Because mathematics has a beautifully established procedure for going about setting up their hypotheses, their ideas, if you will, I don't even know what to call these things in mathematics, George will be able to give us better ideas on that, theorists. And how do you go about discovering whether or not there is any possibility of your theorem having any validity. We don't do that in taxonomy, and yet there is something there which says the proof is a different direction. Well it is there. So it seemed to me that this is a valid type of thing is to try to gain some measure of knowledge about what mathematics does, how it works, whether or not you indeed have some similarities, and if you can, can you marry these two disciplines.

Now with these sorts of things in the background, we might say that we have the following objectives in this course. To discover something about the taxonomic process. To get some ~~xixxxixixxxix~~ insight into and methods of use of computers in this field. To take a look at what might be considered the mathematicl way of examining the problems in this branch of biology. To put what you learn to work on a real problem in taxonomy, and to give you some basis for studying your own problems in biology or wherever you are headed for. These are, I suppose, partly objectives.

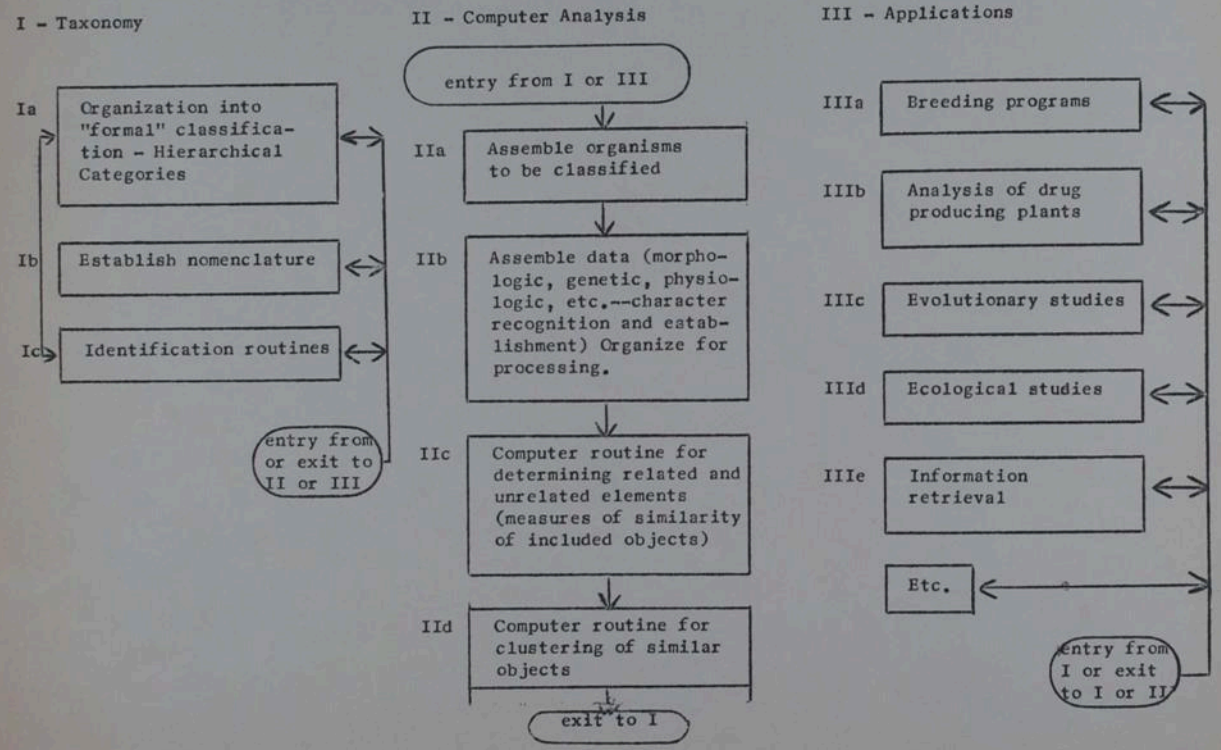
Digitized by the Hunt Institute for Botanical Documentation

GENERAL FLOWCHART OF TAXONOMIC PROCESSES



Note: The procedure begins and ends in one of the taxonomic areas or in one of the applications.

DETAILED FLOWCHARTS



Digitized by the Hunt Institute for Botanical Documentation

Sept © 3/25/66

Discussion on flowchart of taxonomy

last time. I don't know that anybody has ever prepared such a flowchart as this particular one before. At least I haven't seen one published. We have done this as much for our own benefit as for anybody else, to show where we are and where we are going. If you will look at that, we have divided into two parts, there is the taxonomic aspect and the computer analysis and the application. And you will see below it there that these are ~~filled~~ spelled out in some detail and the detailed flowcharts. I might ask you, "What value is this thing? And what does it tell you?" You can probably read it with a little understanding without any . But suppose we look at it . The question is, how do you know when you have a problem in taxonomy, first of all. You can know that from several different ways. You can look, for example, in an unknown animal or plant in your hand, and you wish to identify that animal or plant and you look in the identification routines that you can find in present and available taxonomies, and you discover that this particular organism that you have in front of you is not easily identified. In other words, it may not be present in a present-day classification, and this may then lead you to discover whether or not there may be a better classification scheme so you can start in on that. If that be your entry point, then you can go over from I which we have labeled as formal taxonomy, that is the organization in the formal classification and the categories of it, and a nomenclature for a . You can go from the identification routine back over to the computer analysis and that computer analysis is probably an incorrect terminology. Let's just put it this way, that you are going into some sort of methodology for discovering a classification, and then you follow the flowchart through on your II. There you discover that you have to assemble the organisms relating to the organisms that you do not know, there is a good likelihood that you can probably put almost any organism today into some genus. It is less likely that you

Digitized by the Hunt Institute for Botanical Documentation

will have to discover or describe a new genus of organisms these days than it would have been say a hundred years ago. It is even less likely that you will find that you can't place a particular organism in a particular family of organisms. Therefore, when you are starting out on a thing like this, there is a pretty good chance that the organism that you have for identification will be one that can be assigned at least to a genus. So when you start your solution to a problem of classification, you at least have an advantage of knowing that you already have a group of organisms by a particular name, genetic name let us say. And if you wish to make any new classifications, you may work on a genetic reclassification, or you may wish to work just on the species reclassification o r something of that nature. It depends entirely upon what you find available to you at the moment as to what level of classificatory work that you are going to do, whether you are going to do a monographic piece of work, we are assuming that this is a monograph that you will be working at to reclassify, you have to decide yourself whether the genus needs to be reclassified or just one species needs to be amplified in this classification a little bit. At any rate, if you are able to assemble the organisms that you wish to associate with whatever new organism you may have discovered, then you follow on down ~~through~~ through the processes of assembling the information which will be useful to you for the process of classification. We have listed a few of them here under morphology and genetic, physiologic, and so forth. Then you have the processing right on through from that into the eventual business of discovering clusters. Clusters is another term which does not designate the level of a particular group of organisms but merely says that you have a bunch. It can be a small bunch or it can be a bigger bunch. But at any rate, once that you have got through a level at the bottom of part II there, you have clusters. Then you can go back over from that where it says exit to I and begin to organize these clusters into formal classifications. In other words you assign the hierarchical categories which as you as the individual biologist discover. After you have got the categories, then you have to

Digitized by the Hunt Institute for Botanical Documentation

go and find out some way to hang a handle on those things, in other words, the nomenclature. And usually to be found in outputs from taxonomic work, along with the establishment of the categories and the names therefor, you usually find some identification routine, keys will be provided in many pieces of taxonomic output, most of them, as a matter of fact, And I wouldn't be surprised if this morning we discover that some of you have gotten just about as far as the point of identification. This is what we want you to know, to see where you can locate yourself as to what you are and what you are doing.

Another way that one may discover a problem in classification is suppose he be a geneticist and over there under applications you will see a list of a few kinds of applications of the output of taxonomic work. You will note the several ways that people might become interested in the process of classification itself. And you will note that amongst those things that we have clearly differentiated evolutionary studies from taxonomy. You may make evolutionary studies but you don't do it at the same time you are making a classification. You may be having some sort of a feedback system so that your classification attempt to reflect an evolutionary sequence, but let us clearly differentiate the differences between classification per se and evolutionary studies. That is the purpose of this flowchart is to sort of let people know where they at a particular time and what they are working on. If you have something like this, you can sort of say, "what am I trying to do? Am I lumping two problems into one, or am I really able to attack the problem as I have outlined it?" This is the idea here. We have found this useful for instance to tell us what we are to do next. So let's give you some way to tell you something about where you are as far as classification is concerned. Most of this course will be involved with that which is labeled as IIa, b, c, and d. We will not in this course have time to go into the formal parts of taxonomy. There is just not enough time to

Digitized by the Hunt Institute for Botanical Documentation

do it. This would take another course and hope that maybe we can generate these courses that will be useful adjuncts to this main problem which is the actual process of classification. This is what we really are working at here.

There are other types of thinking processes. There are other types of mathematical models, there are other types of computer methodologies programs which will be useful when we are talking about these other aspects. For example, over here under nomenclature we figure that a massive information retrieval program is required and we are pushing very hard at this moment trying to get the right people and the right places to do the job of information retrieval system for the names of plants and animals. This is, we figure, a critical element and it shouldn't be done at such an institution as Colorado State University because we do not have the library, particularly the library facilities at this institution which are required. As a matter of fact, in this country there is only one really one institution which has the necessary equipment, libraries, and staff for both animals and plants to hack this job, and that is the Smithsonian in Washington. We probably could have done it at New York, but we would have had to combine our efforts with the New York Botanical Garden with those of the American Museum of Natural History, because there is no sense in attempting to do a nomenclature system just for plants or just for animals alone. But I won't go into that one because that would be beyond the scope of what we are up to. This is merely to point out that we will not attack the whole gimmick of taxonomy. If you know that there are other areas is probably a good thing. I would suggest that you can make good use of such charts as this, not only for taxonomy but I suspect that within your own framework of interest whatever it happened to be, that such a flowchart as this will be a useful sort of roadmap for your own directions. I think it has been for us to say only that the chances of this particular chart remaining as is for a long time is not good. Because as we refine our we will probably find out that there are

Digitized by the Hunt Institute for Botanical Documentation

other ways to put down such flowcharts as this, break it apart, or combine pieces of it. In other words this is not a uniquely determined flowchart put together in many different ways. As I remember when we made this up, there were interrelationships double arrows between these boxes that you could get from one of these to the other. I think these were left out in conversion. Yes, this is true. You can add two more sets of arrows, double ended arrows under Ia, b, and c, That is amongst the -- (Bob) between Ia, and Ib. Yes. (Bob) everything Ib and Ic to its double arrow. What is to the right -- that is the arrow from Ia to Ic. Each of those three boxes, you can get from any one of those boxes to any other in any direction. wrong. Yes. No you have to add your own double arrows to Ia and Ib in between I b and Ic. Otherwise there will be no communication between Ib and the other two boxes. (George) Why can't you go out the arrow on the right of Ib and around and up to either of the other two boxes? (Bob) Oh, that is meant to take you out to the edge of the, from the entrance. (George) Those things only turn down. (Dave) yes. (George) It is somewhat ambiguous. (Bob) It is, isn't it? (Dave) It is an ambiguous one statement there if you try to follow those arrows as they were meant to be followed. (George) When you come out of there on the up instead of the down, (Bob) You actually can go up if you are coming from the round box. Your entry from one of the other two.... (George) then you have to go down.... (Bob) (discussion on the flowchart)

The arrows to the left of the boxes give you some clue as to entrance and exit exit points. There is an interesting sideline on this thing. This flowchart was not originally set up this way and I gave it to Bob whose knowledge of flowcharts is much greater than my own, and he reorganized it entirely because what I had written down as a flowchart was confusing as to what it meant. What I intended to do there to point out that there are other disciplines that can give you clues about how to set up your own knowledge better

Digitized by the Hunt Institute for Botanical Documentation

and the person interested in flowcharts, programs is liable to be an individual who is very good at such organization of his information. He has to be in order to make a good computer program, some of which we will get into a little bit later.

I don't labor any more over this flowchart because as it is already pointed out there are errors in where your arrows are and so forth, but it is not intended to be something that you have to memorize or put away, it is just something that you can look at now and again and sort of say, "well where am I right now in this course". And unfortunately right how we are the way the heck off the top of this course. We haven't gotten into it. But this will give you some clue about where we are headed. (Craig Davis) (?) Talking with people like Dr. Kline and Dr. Harrington, it seems to me that they are quite interested ~~inserting~~ in inserting phylogeny in their taxonimic system, and listening to you and reading Davis and Heywood I get the opinion that this (Dave) not necessarily a good idea to taxonomic system./ We would say in this situation, be certain what your thinking processes are at any one state of the development of a classification. If, and you have valid reason for being interested in phylogeny, for no other reason that it is the practice these days, be certain that you know before you start what your data are, what your data tell you, then you can go ahead and if you are certain that you know where you are at any one particular point, you can make investigations of biologeny for the purposes of establishing some order to your classification, some support in some purported evolutionary activity. But be certain that you use pieces of information which will indeed do that which you are expecting it to do. We merely say to you, "If you are going to study phylogeny, show me how. What do you intend? What is your programming? What is your thinking process? What rules are you following?" Rather than to do what has been done in the past, sort of lump all of these various things together without clearly defining in

Digitized by the Hunt Institute for Botanical Documentation

your mind when you are thinking of one thing and when you are thinking of another. Do you see what I mean? (George) This flowchart (20-tape 03/25/66) called a perfect biological cure for this application . Not that its interests are necessarily going to be relegated to the preparatory application. So let's put it in the applications for just a while. Say that we to in column II and come out and get over into column I and come out and get back over into column III and end up in our phylogical box. We can still get out of this same exit back over to column I, cycle around like use the machine results that we find interesting to us, returning these results back into the machine and so forth and cycling it this way. Use this process to develop a phylogy if that is what you are interested in. Without ever assuming to know anything more about phylogy than when you do at the time of the computer machine is actually operating. I think this is sort of the attitude that this flowchart is trying to put forth, rather than relegating

(Dave)
Is that understandable?/ Well it is to me, but it may not be so in the case of the rest of them. Does that answer your question, Craig? (Craig)Pretty much. My main interest here is that for a year now it has been sort of pounded into me that what we are interested in in taxonomy is figuring out evolutionary development of the flora of the world, from starting from the most primitive thing along up through the ages . figured how plant a is related to plant b and judging from this isn't necessarily that important. The main thing we are interested in is figuring out a scheme of identifying plants or arranging plants for whatever purpose you
(Dave)
want, economic purposes, etc./ It is usually pointed out that there is no uniquely determined classification. The classification should serve the purpose of the individual who is doing it, primarily. And all that you would be

Digitized by the Hunt Institute for Botanical Documentation

requested to do there would be to state your purpose and how you went about it. There is nothing that says that if biologeny is not a valid part of your interests but it doesn't necessarily follow that the next guy that comes along is interested in exactly the same thing, is he? I?) I can see that, but isn't it true that phylogical genetic classification would be the ultimate aim for the taxonomist? Figure out how things are related, actually? (Dave) Well, what do you use for input when you say how things are related actually? (?) I don't know whether you can even get the data that you would need to figure out ..... But wouldn't this be the ultimate gain of taxonomy to work out a taxonomic system that is based on phylogeny if we could do it? /I wouldn't say (Dave) that it is the ultimate aim. I refuse to get to those points in this discussion. (Bob) In so far as we at present so can't do the sort the whole question? (Dave) Let us take for example my own particular set of plants, Manihot esculenta and the related species to it. How can I study their phylogeny?

(I stopped typing on 25--the rest of this tape is discussion)

Digitized by the Hunt Institute for Botanical Documentation

tape @ 3/25/66

B 170
March 25, 1966

(Dave) You merely state your case. This is all that I am asking you do. You set up what you mean by relatedness in the first place. (Craig) I follow what you are talking about, this is not only true of working out a phylogenetic relationship but working out any type of relationship you have to know why you make any decision. (Dave) As we proceed through this course, we will attempt to show you how to go about the process of decision making. (Craig) My only question is that in listening to people like Dr. he directed mostly, he gave me the opinion that finally we are getting around to working out a phylogenetic relationships. For instance, of the United States is not a part of the genetic key, and it has been my opinion that Dr. Klein and Dr. Harrington thing that it would be better if it was.

(Dave) Let us put it this way. In Dr. Klein's own particular interest in the genus Oenothera and some species therein. He has made some decisions as a result of his investigations by which he has said that certain of those are the primitive ancestral types in this particular group of Oenothera. Now I can take his data and using another set of concepts turn them around and prove that the primitive ones that he decided upon are the advanced ones. Now what we want to try to get at is: What decision-making process did Dr. Klein use when he decided that this was the case, that he had particular location or an evolutionary sequence. I will not argue with him if you tell me the procedures that he followed in order to establish this, and show me that this is a uniquely determined situation. I am not at all certain that this point gets across in a very large measure. It is only when you begin to investigate these

Digitized by the Hunt Institute for Botanical Documentation

sorts of things as we are that you come to the point that you can make these questions in your mind. Can I turn this thing around just as adequately? I think I can unless I have some procedure for understanding, rules that he followed in order to achieve that. Do you see what I mean?

(Craig) You use the point relationship, now how do you use it? Do you mean a genetic relationship, or do you mean a morphological similarity? (Dave) It depends on you input as to what you mean by it. (Craig) What do you mean by relationship when you use the term? (Dave) Just that I will only define it in those terms, that I will use this relationship to mean that which is determined as a set of clusters using the input data that I gave it to begin with. (Craig) In your genus can you use the term relationship in any more precise way than that? (Dave) Definitely. It is a morphological relationship. (Craig) Morphological, then this could be converted evolution? (Dave) I don't know about that. I have no genetic information which tells me one way or the other what I am talking about. Do you see what I mean? (Craig) That is what I want to get straight. I just wondered how you could do anything besides work out a phylogenetic system if you were using the term relationship to mean a genetic relationship. This would ultimately leave you.....(Dave) The relationship is obviously dependent upon what you use to establish that relationship. The phenotype of the organism may or may not be clues as to its genetic relationship. But at different levels, we are able to use different kinds of information. I cannot use genetic information on Manihot plants, because they are tropical, large whose makes it impossible for me to bring enough of them in to the greenhouses and grow them and discover what genetic complement they have. I can't make cross-breeding experiments with them because it takes too long. I'm not aware of the biological organization in these plants at all. So what I talk about in terms of relationship of the genus Manihot are those purely based upon the phenotype. But I can say that if it later turned out that this

Digitized by the Hunt Institute for Botanical Documentation

became such a gung ho problem that a whole number of people started working on it as they have the Oenothera, then there might be an opportunity for me to take the results of genetic experiments and incorporate those and discover a relationship which may or may not coincide with the morphological. I merely tell you, use your data as the direction which determines your relationship, and tell me how you are going to use it. (Craig) You are looking at a certain type of classification now based on the information you were able to obtain. (Dave) True. (Craig) This is a fine art thing because of your limitations, you can't go any farther than your limitations allow you to. But I disagree with Jack when he says that because it is impossible for us to fill in some of the gaps in history that we should look for another approach. (Jack) This type of approach may not now that we have ;instruments that give us better data than was lacking in the past, and it is going to increase in this direction, this is the data we should use. It is much more realistic than going back into history. (Dave) On this point Jack there is something that always gets us. It seems as we look at the development of particularly what has been called the new systematics, that was a term applied back in about 1940 by Huxley in his book the one that came out by that title, that what was being insisted upon there was that we looked more at the populations of plants than we have in the past, and that we uee as our primary source of information for the taxonomic decision the information derived from cytology and genetics. It turns out that those on pieces of information derived from cytology and genetics are just another type of information useful in the taxonomic process, and are not the ones that uniquely determine the classification. (Jack) I understand that but cytology and genetics areas. You know simply more about morphology than you were able to know in the past. (Dave) And now the latest thing has come on the scene is chemotaxonomy which again, of course, gives you a more precise numbers. Whether or

Digitized by the Hunt Institute for Botanical Documentation

not these precise numbers aid you in the classification or not is a question of how you use it and how you put it together, you see. (Craig) What are we aiming at in the end with chemotaxonomy and cytotaxonomy? (Dave) More information about the plant. (Craig) We are always looking for more information, but why, why are we looking for it? This is my , I'm not getting my point across to you. To me the end product if we could get it, you can't get it perfectly, but if we could get it would be an understanding of how life evolved, plants evolved and the relationships that developed along the way. (Bob) Don't you want to know something about/anatomy and morphology and chemical properties of these plants? (Craig) Yes, we do..... (Bob) Their history their genealogy? (Craig) You do, but that is not necessarily the end product aim in taxonomy, to learn about the .....

(Craig) Right. This is people, individual workers that they each have their own aims and interests, and we may need a classification to fit the special case like you said. (Henry) In the meantime, we are going to have to know what we are talking about. That is one way of putting names down. You mentioned before

11.2

better if it were phylogenetic.

well nigh impossible to use . But they do teach, or do indicate some to you about how the evolved they do that. (Craig) Can't you separate a key used for identification purposes from the taxonomic . (Henry) two keys are given. One key the phylogenetic relationships per se..... (Dave) I would suggest amongst other things that you actually look at the outputs of people in who are phylogenetically inclined. Look at the information given by these people when they write down their descriptions of the taxa which they have accepted.

Digitized by the Hunt Institute for Botanical Documentation

I suggest to you that on the basis of what they put down that their phylogeny is not defined. You just go back and look at any body's key and he can be the most obsessed phylogenist alive. You will discover that when it comes to the actual output that he has not given you any definition by which you can say how he achieved the ordering of his organisms. (Craig) Is this to say that eventually we won't be able to do this? (Dave) I didn't say that. I merely say that these people have not given you any evidence whatsoever for their decisions about the phylogeny. (Craig) Well this individual , most of you seem to be talking about individuals and what they are doing and what they want to do. science of taxonomy, what are we after? Are we after some sort of an understanding of how plants are organized morphologically, or how they are related

(Henry) 14.1 ⓑ 3/26/66

then we went through a period from which phylogenetic trees were found out, because most of them were nonsense. In the meantime, along people started putting genetic problems into it. We have reached about that stage now. I think what has happened to Davis and Heywood is

(Dave)
Who dictates a science? (Craig) No body dictates, but where are we as taxonomists? Are we sitting in here to learn to set up a key so that some of our can go out and say, "Okay, this plant ; is poisonous, this and this one is not". (Dave) This is a valid classification. (Craig) ..... is that really the most important thing. (Dave) I don't know. It is to the farmer using it.

Digitized by the Hunt Institute for Botanical Documentation

(Dave) I will not dictate to you what your interests in taxonomy are, nor your need for classification. We are clearly in this course attempting to get away from that type of need, and to look at the process itself. Then you can use it as you please as long as you understand how you use it. (Craig) Don't get me wrong, I have not been here long enough to decide whether I am interested in phylogeny or what, but it just seems to me that that is what it is all about. related to another and we may never be able to figure out how related to another, but certain things, do you agree that certain things are related?

(discussion by several members of the class - a great deal of interrupting and noise)

(Dave) What we hope to do in this course for you is to make it such that whenever you have made a decision as to what interests you have and what line of approach you will use, you will use a sound process by which you can achieve the ends you have in mind. There is no ultimate goal in taxonomy that I can project for you. I don't know what it is going to be. It may be or it may not be an evolutionary system. I would like merely for you when you have gotten through with this situation, to say to yourself, "Okay, if I wish to make a phylogenetic sequence of the organisms that I think to be a group, I will have a basis for demonstrating how I went about it and being a little bit more artistic about how you got to the results you did. If you want to be interested in phylogeny, fine. I have no objections. I merely want, when you get through here, to be certain that you are aware of the pitfalls in reasoning that you may get into and to try to establish some orderly process for your decision-making procedures.

(More discussion)

Digitized by the Hunt Institute for Botanical Documentation

B 170 Lecture 2
March 28, 1966

HISTORY OF PROBLEMS

I. Not a long hisotry for use of computers in classification.

A. Works first appeared less than 10 years ago.

1. Michener & Sokal (1957) - in Evolution

2. Sneath (1957) - J. Gen'l. Microbiology

3. Rogers & Tanimoto - 1st Publ. In IBM Techn. Infor. Man. (1959).

B. Ideas which generated (found the basis)

1. Sokal & Sneath - statistical

a. Work of such statisticians as Fisher, Pearson, and Wright.

i. Fisher - work on Ag. Expt. Sta. Rothamstead - validity of data, prevention of bias in sampling, size of sample, etc.

ii. Wright - population genetics

b. Work of anthropologists before turn of the century.

c. Work of G. G. Simpson (quantitative methods in paleontology) with aid of Anne Roe (wife) and Ernst Mayer.

2. Edgar Anderson and his influence

a. Read first paragraph or two

b. Introgressive hybridization (1949)

Methods for characters - description on this in Ch. 14 D & H.

3. Tanimoto's use of methods.

Tanimoto & Rogers

Principal idea - learn what the taxonomist is doing, then decide what mathematical approach should be taken

II. Historical background

A. The work of ~~Adanseu~~ Adanson - Fr. biologist

1. Systems of his day (~~Leinu~~ Linnaeus, Tournefort)

2. His work in African tropics

Digitized by the Hunt Institute for Botanical Documentation

3. Classification, character by character, and then correlating each classification.

4. Based ideas on fact that the whole plant could provide information for classification.

B. The contrast with later taxonomists

1. A priori selections of "important" information - the sexual parts

2. Leads into "weighting" and what is meant

III. Summary

The basis for this work is and should be a study of what is required in the methods of thinking.

Digitized by the Hunt Institute for Botanical Documentation

History of Problems

The outline that we have calls for a lecture beginning next time on the computers. It will serve you well if you can at least begin to read for that o available to you between now and then and during the next two sets of lectures on the computers this manual of general information manual here. Now obviously you can't get the whole thing under your belt now do we have enough copies for each of you to get it under your belt, before that begins. So what you should probably do, and I think the way that Bob has organized these lectures, you can read this as backup information for what he will tell you. Now since we have only these copies here, there being five all together, we would ask that if you take these manuals away with you that you not make it more than an overnight, and that you get these things back here and into the laboratory the next morning so that they will be generally available for everybody to have a go at. Again, we are not going to try to make computer experts out of you, these are merely explanations of what we feel are general ~~explana-tions~~ understandings that will be helpful to you. We will not cover this whole book. Absolutely not. This is impossible to do. These are discussions which really do not relate to a particular piece of hardware. We will be so far underneath any, our discussion will be so much below that that it really doesn't matter. We may be able to use one of their pieces of harware merely as a communicative device to Western Data Processing Center in Los Angeles. At the moment we are having trouble with that and we are not getting anywhere, so we are going to Boulder and using the hardware on the Boulder campus. We don't know yet. But that doesn't need to disturb us, what kind of hardware, as long as the stuff flies. Frankly, my attitude about a computer is that it is a microscope and it just so happens that the American Optical Company sells

Digitized by the Hunt Institute for Botanical Documentation

me a piece of hardware that I like, fine. I'll use an AO or B&L or Leizer or what have you. It is the same way with the hardware. I don't care whether we are talking about a Philco machine or a Remmington Rand machine or an IBM machine or a Controlled Data Corporation machine. It doesn't really fly as far as I'm not going to compare my Ford Mustang with your Corvair or anything like that. Do you see what I mean?

The topic calls for a little history of the problems that we've got, and if you stop and think about it a little bit, I should be able to tell you the history of the use of computers and classification in one sentence, but obviously since I have a whole hour it is going to take a few more sentences to fill up the hour about the history of the use of computers and classification.

It isn't really a long history, it can't be because of the fact that the general availability of the hardware really is not more than, not even quite a decade old at the first publications have appeared where someone has said we are using a computer for the purposes of classification, biological classification. Now don't misunderstand me that there are other things that can be called classifications besides what we say, but this is the first time that you will see any biologists talking about it. The first paper that appeared that has any connection with a classification was this paper by Michner and Sokal. Their publication first appeared in Evolution in 1957. You can look at this later if you wish. Michner is a taxonomist of the two in this group, and Michner, of course, you know is interested in classification of bees. That has been his life work. So-called was the bio-statistician who with him. The next paper, and I am not sure whether these two occurred just about the same time or not, but this next one ~~I have it selected~~ is done by an Englishman ~~by the name of~~ Sneath. His initials are P.H.A. And his work also appeared in 1957. The title of it is the "Application of Computers to Taxonomy". Sneath is a microbiologist primarily concerned with some of the organisms and the

Digitized by the Hunt Institute for Botanical Documentation

family Enterobacteriaceae, that is his major biological study. His first paper and there are two papers back to back appeared in the Journal of General Microbiology. The other early contributor so you can now discover that I am an historical figure appeared in an off-beat place worked by Tanimoto and myself. The first appearance of anything like that was this mimeograph output an IBM publication which first showed up in 1959. We were working, I think the only group at the time that was interested in the classification of plants, so you see we have three different interests in biological classification of insects, classification of bacteria, and classification of higher plants. So the three have been the take-off point. Not it is profitable to look at the ideas which form the basis or the groundwork by which we could take off into these things. Obviously, we didn't start in virgin ground and evolve our own methodologies completely out of thin air. There are backgrounds involved in these things. The two other workers mentioned early Sneath and Sokal both form their ideas used the methodologies based upon statistical methods. Statisticians have been at work in biology for a long period of time, right from the very, oh, Bateson I believe was one of the first who really did any statistical analysis and he did his just about the turn of the century, or just a little bit before. But the really big statistical names that will occur and you will hear about in their relationship to these problems over and over again are such men as Ronald Fisher, Pearson, and Sewell Wright, these are people whose contributions were great in certain aspects of biology. There are others obviously, but these people stand out as much as anybody does. Fisher, for example, was an ~~ally and~~ associate ~~and worked~~ at one of the greatest of all agricultural experiment stations in the world and I think it is the oldest agricultural experiment station ~~and~~ that of Rothamstead in England. Rothamstead's experimental operations date back at least 300 years. So they have a fantastic continuous operation.

Digitized by the Hunt Institute for Botanical Documentation

They have had periods of decline, but they have constantly been at work. It is a very fine agricultural experiment station. Working with the people in the agricultural experiment station, Fisher was worried about such problems as the validity of their data, how ~~could~~ good is it, how good a piece of information have you got. He was worried about the prevention of bias ~~in your~~ sampling techniques amongst other things. The size of the sample necessary ~~that you had to have in order~~ to make a significant test. These were the sort of things that were in the back of Fisher's mind and were requirements for him. Sewell Wright, of course, is a man whose contributed ~~xxxx~~ tremendously to the field of population genetics in his statistical methodologies, obviously, valid application when you are talking about populations. This is a very critical and important feature, should be in modern-day taxonomy. It is the study of populations of things, not individual samples.

I think that Sneath was greatly influenced in his work by anthropologists. Now anthropologists have been interested in essentially the clustering of information so that they could get higher meaning from it. And anthropologists working with archaeologists were guys that were always trying to discover some sort of a classificatinn system for the hairy kinds of information that you discover when you are talking about people. You get a lot of qualitative descriptors and anthropologists are always having difficulty in how to use that kind of information. And so the anthropologists were one of the earliest groups that have made any kind of mathematical methodology which ~~they might help, use~~ Leaven to help them in their approach to clustering for higher information from these difficult types of data to deal with. So I think Sneath was really influenced by a set of anthropologists who, as I said, had some of their earliest works even earlier than the biologists did before the turn of the century. A considerable amount of influence has come to this field by the work of George Gaylord Simpson, whom many of you know is the paleontologist of note. Simpson has published many,

Digitized by the Hunt Institute for Botanical Documentation

many books. Among them was a quantitative methods in paleontology, and I haven't gotten the specific reference to this, but obviously paleontologists are always dealing with classification and evolutionary methodologies and so he certainly forms a general background by which biologists have begun to think about these problems. Simpson was a smart many directions. For instance, he himself was not mathematically inclined so he married himself a mathematician, Ann Roe is his wife and she on her own hook, on her own merit was quite a contributor to statistical methodologies in many different areas, not only in biology, paleontology, but also in anthropology, she is quite a worker and contributor in that field. Another individual who has given amount of thought trying to be more objective and less handwaving about taxonomy is the man by the name of Ernst Mayr. Ernst Mayr is a taxonomist of birds and so his work too had considerable background influence on the whole study of populations both from the study of classification purposes and for studies of evolution. It is not quite so obvious perhaps the background what we tried to do, where we tried to take off from, but one of the guys who really has done most of the contribution to the basic biological ideas and thinking that I have had as far as this connection is Edgar Anderson ~~Ed Grant~~ whose long connection with the Missouri Botanical Garden is well known. I think it would be interesting just to read one bit of Anderson's works to show you the kind of things that set me to thinking in these directions. Actually, his ideas were expressed verbally long before they appeared in print. This couple of paragraphs I think does a pretty good job. This is the publication that appeared in a chapter in a book called Statistics and Mathematics in Biology edited by Kempthorne. You can see these publications afterwards to get a more complete reference to them. He starts out in a chapter which is titled "Efficient and Inefficient Methods of Measuring Specific Differences". The following comments, I think, are pertinent.

Digitized by the Hunt Institute for Botanical Documentation

(Read first two paragraphs)

He goes on in this vein, but it was this sort of attitude about classifying your biometry first to discover whether or not that particular thing, whether a particular methodology in biometry was what you wanted as a taxonomist. I think this is one of the things that really has kept me going for a considerable amount of the time. What are the kinds of mathematics which you should employ? One of the things that Dr. Anderson did that has been a real contribution particularly to the input data that we use was first used in the analysis of hybrid swarms to discover if possible whether or not two species were indeed hybridizing and could you detect from specimens, herbarium specimens and from living material, whether indeed you could detect whether there was a problem of hybridization going on without actually making the crosses yourself. In other words he was interested in analysis of natural populations. His work was intended to really discover if possible whether you have species hybridizing and if you could could you locate the individual plant within a hybrid swarm, and could you discover if you didn't have in your hand something about the species parents, the parents of the hybrids. His work is entitled and perhaps is a word that has become very familiar to us called introgressive hybridization. His ~~blue~~ little book appeared in 1949 as a sort of a, actually it was a summary of more than 30 years work by Anderson, but it was so new to most people that they, and his literature up to that point was sort of scattered in various publications and he hadn't brought it all together, and so this little book appeared which is very useful. Part of the methods used in that are discussed by Davis and Heywood in their textbook in Chapter 14, but I found that discussion doesn't really give you the fundamental ideas necessary to understand what is going on. It might be wise to sort of discover just exactly what it was that Andy did that we have used in our methods. We were, of course, interested in classification of all organisms, not necessarily the discovery of hybrids or hybrid swarms, particularly. But we found very early that the methods

Digitized by the Hunt Institute for Botanical Documentation

that Andy used for his hybridization techniques would work particularly well when we were going about setting up an input data for the purposes of classification, classification as contrasted with the discovery of hybrids. Now it is really what Andy would do when he tried to discover hybrids is to try to take pieces of contrasting information from the plant specimens that he employed. For example, he might discover the characters of a plant and population vary according to the amount of ~~infloresscence or~~ pubesence which the plants have, in other words, the amount of hairiness, the amount of foliage or some other part. You might discover that some of his plants had no pubesence whatsoever anywhere, but that as he examines these hybrids he discovers they were ~~the~~ sort of arranged going from none up to a lot of a real fuzzy plant so to speak. So he discovered that he couldn't really very precisely say that a plant in this particular area here had any more or less than some other particular plant, but for his purposes in order to put his plants into packages maybe he could sort of force a decision about the pubescence of a particular plant where he had to sort of have a range. That he could say,"Okay, for my purposes this is the study of hybridization". I can decide that a group of plants which most people can be called a one character, or one attribute of a character of pubescence. That is he looked at these plants and furthermore he discovered that there was some, and this is a very qualitative terminology. They have an intermediate ~~an intermediate~~ amount of pubesence or hair. In other words, it had, they obviously had more than had none, although some of the ones that he may have decided had none had some. nice and mixed up for you. Well this is the way biologists get sort of mixed up so we have to be able to do something . At least he could force some sort of decision where he might decide that this was an attribute where he had some hair and he just labeled whether one, two, and three or whether or not

Digitized by the Hunt Institute for Botanical Documentation

three, two, and one that order of presentation. And where he had gotten out to the plants that were really fuzzy, he usually had no difficulty in deciding the character of state, heavily pubesence or fuzzy, if you will. And he could go through and look at a lot of plants for a lot of different characters and discover that he had reasonably little difficulty in establishing some sort of division of that particular piece of information. For example, he might have a series of

and discover that these population plants had the series of what seemed to him a sort of differentiation on the basis of where it had short, medium, and long which he though would be useful to describing a series of

He might give it a range to which I call a short, medium, long 1-3 cm, another one from 5-10, and one for long from 11 to whatever it turned out to be. This isn't necessar

This sort of methodology some way will go in through and measure a whole lot of space for characters of lost plants. Now Andy was working with hand techniques and/his usual procedure was to take maybe five or six no more than ten pieces of information, but his nice and clear/methodologies which follow from this set, he was able to use a very simple little methodology I don't know whether you would call it statistical or what in order to elucidate the idea of hybrids location within a particular place between two species. I'll come to this sort of thing a little later in another area which will also give you some takeoff point for what we did.

It certainly is profitable for people to go back and read several of the publications of Egar Anderson. His general philosophy about plants, how you look at them, and how you study them are valid and useful kinds of reading. They have certainly been a mine of ideas and a wealth of information for our technique.

Digitized by the Hunt Institute for Botanical Documentation

March 28, 1966

In conversations with this mathematician, Tanimoto, I told him I wanted to use the computer to classify plants. He says, "Okay, you tell me how you classify and I in turn will tell you how to use the computers." This was rather a shock to me that I had to tell him, "Well, here is what I do. I just put my plants in piles, and those that are alike are one thing, those that are different are something else." He says, "Well, how do you know they are alike?" "Well, they look alike." "Yes, but you can't say that to a computer. You don't tell me what you are thinking about when you do this job." And so then I was really set back on my heels and discovered that all of my nice training that had got me my union card and everything had never really put me down to the dirty hard task, thinking: What is it the taxonomist does do? What is he thinking about as he classifies? I had no objectives. I started out with saying, "What do you want to do with your classification? I'd like to study evolution. Yes, but how do you study evolution?" So you see every time I opened my mouth and told him what I wanted to do, he would come back with this sort of question, and say, "How do you do it?" And so from that we got right down to the basic problem. We had to learn what the taxonomist was doing, why he wanted to do it, and then maybe if you learned that, then you could decide what sort of mathematics might be appropriate to be used to follow that set of directions. Clearly, Fisher had a take-off point where he knew what was wanted for his mathematics, and so he could design some sort of a model which would indeed get the kinds of information he was looking for. Just as clearly now but not near so clearly at the beginning, we had no such idea of what was wanted in classification. It is as simple as that. So those are the sort of things that led us to this discussion, this idea of looking basically into our own science first and finding out what it was about and then going ahead and trying to discover what were our methods, what were our thinking procedures, are they logical, are they consistent?

Digitized by the Hunt Institute for Botanical Documentation

And then perhaps we could go ahead and find an appropriate mathematical model. I think if I have done anything in the contribution to this ~~dext~~ field is to keep my mouth shut when it came to the business of a mathematical model and depend upon the competence of a mathematician. This is another point that is a very sore point with me that we have been told so many times that you have got to be a mathematician, biologist, geneticist, and a chemist, and a physicist, and everything before you can be a decent taxonomist. If you stop and think about it, to be a competent mathematician takes a full time job, there is no question about that. To be a competent chemist, takes a full time job. To be a competent taxonomist, takes a full time job. These are critical elements to be considered. This doesn't say that you shouldn't try to learn what it is going on in the mathematics, what is going on in the chemistry, what is going on in the so forth. What you dang well better be sure of your own science first. This is the important thing. Now obviously again we have another bit of contributions that make up the history so to speak behind the use of computers in classification. This one goes way really way back to the start of the time a man by the name of Adanson who was a French giologist who was, I believe, contemporary with Linnaeus. Most of his publications will be a little bit later than those of Linnaeus, but one of his publications came out about 10 years after the publication in 1753 of Linnaeus' Species Plantarum. Adanson, a French biologist had available to him then the classification systems used by Linnaeus proposed by Linnaeus in about 1734 his sexual system. Adanson also had available to him a classification system turned out by a French botanist by the name of Tournefort. In France, then as now, the French thought what the French were doing was better than anybody else, and they tended to look more in favor on Tournefort's classification than they did on Linnaeus'. But it made no difference because Adanson was assigned to do his work in the Senegal, I believe, on the west coast of Africa.

(6)

Digitized by the Hunt Institute for Botanical Documentation

His connection with plants therefore was with a different set of plant materials than he would have found occurring in the temperate parts of Europe. So eitherr Linnaeus' or Tournefort's systems of classification sort of fell flat on their face as when he wanted to make a classification of the plants that he found in the tropics. So when he discovered ~~his~~ that his classification system had to be something different, I can't imagine what background he had, it would be interesting as a historical study to know how Adanson was backed up with ideas from other people. We don't have that available to us now, so we just have to assume that he started from scratch with no other ideas behind him. But basically the idea that ~~Tournefort~~ Adanson had was that ~~any~~ any ~~pa~~ piece of information that you can get from the plant could be used or might be useful in the classification thereof. It didn't matter whether you were looking at the root, stem, leaf, or some part of the reproductive structures, that if there were pieces of information available from them, you should attempt to use that information in making a classification. What he did for himself was to take one piece of information or one character. For example he might of taken the same sorts of characters that ~~add to~~ that Anderson used here, or let us say that he was interested in pubescence. He would try to make a classification of his plants purely on the basis of that character, one character. Then he would take another character of the plant and it might be something about the flowers. Let us say he had a differentiation on the number of stamens in the flower. He would make a classification based upon the stamens, differentiations that he could find. This would be a separate classification from the one ~~that~~ that he made on the basis of pubescence. And he would go through and make classifications based on several characters simultaneously. Then he would go back and try to put these classifications together and see where he got the most content of information. Now this is an ideal idea and, of course, you can see immediately how this is going to get you into

Digitized by the Hunt Institute for Botanical Documentation

hot water, because of the fantastic amount of information that he had to deal with. Remember that he didn't even have any much more than a quill pen available to him to push his data around with. He certainly didn't even have a mechanical desk calculator at his disposal, and so obviously he couldn't really carry through the idea that he had, and some of the classifications that he did make by hand had odd things stuck together. I think probably the reason was that he just didn't have enough ability to manipulate as many pieces of information as he himself probably would like to have done. As a result of this, because he did throw some rather odd-ball things together, his output was looked upon with great disfavor by a large number of other bottanists. Apparently, these other were more "practical people". They said, "Well, we can't hack it with all that information, let's see what we can do with a small amount." And this probably inevitably led them to make the classifications based pretty much on purely sexual characters, that is the reproductive organs of the plants or animals. Now this had a pretty sad effect for a long time, but we did get classifications out of it, some was more useful than others. But you can see the difficulty here that you had a priority selections of information, you are going to say that I have a classification based, must base my classification on sexual characters alone, or some other small set of characters. You can see that this too will lead you just about as much difficulty as if you tried to assume that all pieces of information may have some xxxx validity. Out of such as this kind of historical situation, we have gotten ourselves into some marvelously hairy arguments which really have not had a good solution yet. For example, it has been contended that Adanson used what is called an unweighted system for a use of characters. He did not attempt to assign any particular value more value to one piece of information than to another. Where others have said, "I will use a particular piece of information only if it comes from a particular source, Therefore, I will strongly 'weight this information'". What do we mean

Digitized by the Hunt Institute for Botanical Documentation

when we say "weighting" here? Obviously, there are difficulties that we run into right away. If you say, "weighting" to a statistician, this means one thing. If you say, "weighting" to a taxonomist, he has absolutely no mathematical model that he is thinking about and yet these two words get so mixed up in this particular thing, really we've got a gimmisch as far as what this word weighting versus non-weighting really does mean. little significance in what we are talking about. If you come to me with the problem,"Well isn't this worth more than something else in terms of classification?" I'll say, "That depends on how it does the job of classification. If it does indeed turn out that this one piece of information be more useful in classification, fine. It has value for the purposes of classification." But we must say, obviously in this situation that how are you going to know how much information is contained in a particular character or piece of information until you have made a test of it to see how it does the job. Then after you have made a test of the thing, then you can say, "Oh, sure I'll use this piece of information for classification or at a certain level of classification." Frankly, I don't know of any other real history that has contributed much to this field, except that we do have a whole history of classification which has given us some ideas about how it is indeed done. It is very tough, hard, and dry reading to read other taxonomists output. I don't know of anything that will put a guy to sleep faster than to read a series of species descriptions. Yet, if you can look at it in a different light, you can see that in many of the species descriptions which have been provided, people have attempted to make an interrelationship of many kinds of data for the purposes of classification. Read what they do and what they print, not how they wave their hands. You'll find out that they indeed do have some sort of methodology, although unexpressed. This is really what we have to do, is to consider

Digitized by the Hunt Institute for Botanical Documentation

historically what has happened in classification. As a basis for any take-off that we may have for new methodologies that we may employ. Therefore, it is not my feeling as has been expressed by Paul Erlich, for example, that you should burn all the herbaria, that we should stick a match to all insect collections and start from scratch again. I think this is ridiculous. We do have albeit an imperfect classification, we have a classification it works for large, in large measure. But what we want to try to do is to discover the methodologies in the thinking process and see if we can't make it a little tighter all the way through.

Are there any comments that you may wish to make in the last few minutes left?

Question by student (but I can't get it) 17 on tape.

Answer by Dave: Andy used a lot of nice terms that are difficult really to define. But what he is talking about sense impression ~~itxxxx~~ I think hiss reference to a piece of qualitative information, for example, a sense impression when you are looking at two plants is how big are they, and you aren't using yardsticks or rulers necessary when you ask that question. You can say, a sense impression about these two plants, one is big and one is small. You see? That is a sense impression. And for some purposes, this is all that you need. Or another sense impression is one which deals with the, how hairy is the plant? It is real hairy. You haven't counted the hairs when you say that kind of thing, but you have given a sense impression of it. It has a quality, you see. These are sense impressions upon which we base many of our decisions when we are talking about classifications when we are talking about people, how you describe somebody--well he is a sort of a short pudgy character, or he is a real tall skinny rascal. You see you use a whole mess of sense impressions to try to get across a classification as really a classification. Okay?

Digitized by the Hunt Institute for Botanical Documentation

(Student) Are you suggesting that the basic problem so far in taxonomy is that it is primarily qualitative in nature and not quantitative. That is you say a plant is pubescence, and when you are going through your key you can find the word pubescence but it doesn't really give you any indications in a quantitative sense as to what pubescence means with respect to the specimen you have got. Is that what you are saying? (Dave) That's right. We could take, for the purposes of classification, I always keep these sorts of things in mind, that this, make this statement before you say anything else. For the purposes of classification, how much is gained, for example, when you take flower colors. Now you can make analysis of the pigments and you can make a rather precise analysis to give you the exact concentration of a pigment. But suppose we want to classify plants, we look at the plant, we look at its flowers it has a mixture of colors available on different plants that you have, one is white, one is blue, one is red. How much more is gained by establishing the exact concentration of these pigments in there for the purposes of classification than merely to say, "this plant is blue". That gives you a classificatory statement right there. This, of course, does lead you into difficulty, obviously. When do you stop calling it blue? You see this is the sort of thing, you have to make some sort of a definition of what you mean by blue. You should, yes.

(Student) Well, then is seems to me the problem involves either we don't have, either you have, either in our classification system we can't express the character the plant has because it doesn't have it, or it has it to a degree which is not satisfactory with words alone. (Dave) Well, in these cases we will switch over and use whatever is available to us. In other words a piece of information is only as good as you can make it. I'm not going to say that we can't, that we must define everything as a qualitative. This is wrong.

Digitized by the Hunt Institute for Botanical Documentation

There are times when quantities indeed will be useful. But it is the case that more times our information will be qualitative than quantitative.

(Bob) One of the techniques we hope to present in this course is a way to yield qualified mathematically with some of these qualitative things which all of us previous to our efforts been to this whole problem of taxonomy. There are techniques for handling, handling mathematically/qualitative data. This will be part of the course.....

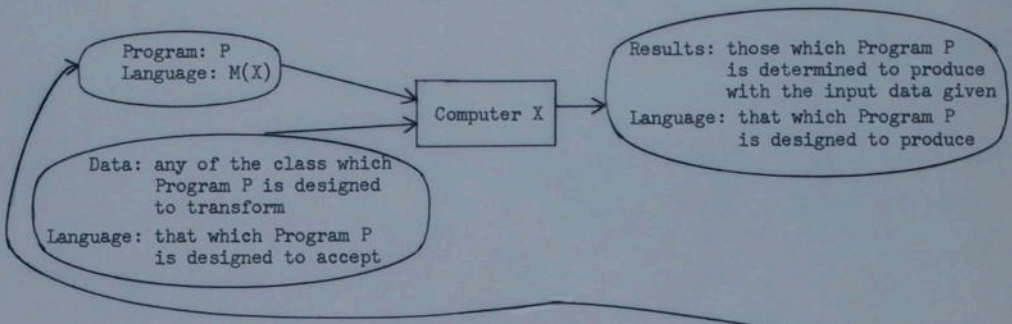
Digitized by the Hunt Institute for Botanical Documentation

PROGRAMMING LANGUAGES

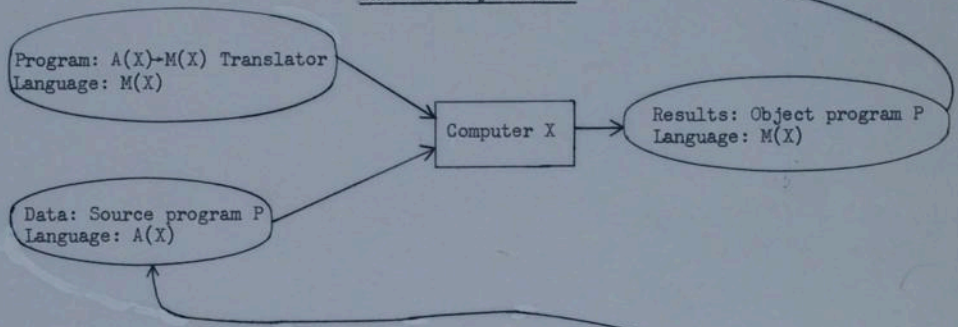
Program P = any program
Computer X = any computer
Source program = input data to a program translator
Object program = output from a program translator

M(X) = machine language of computer X
A(X) = assembly language of computer X
C(X) = compiler language of computer X

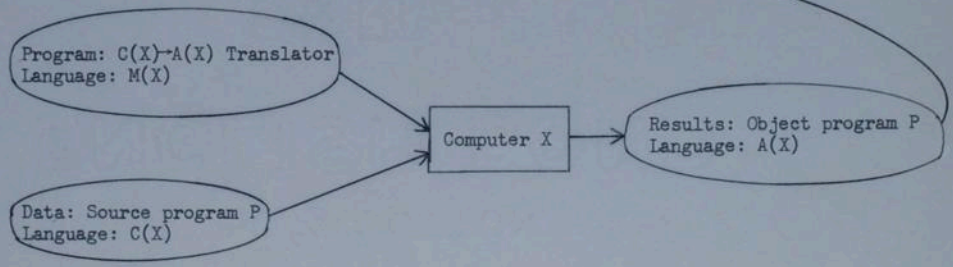
The General Computing Process



The Assembly Process



The Compiling Process

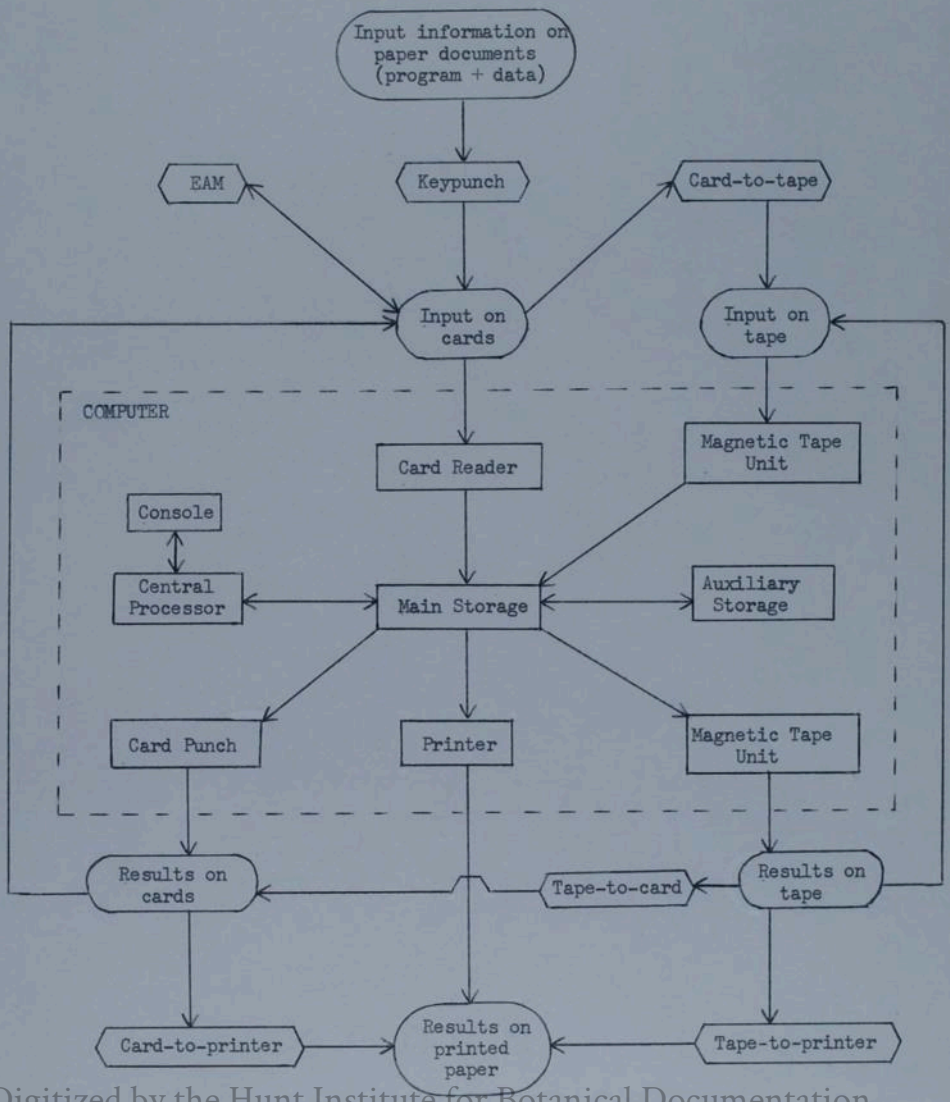


Digitized by the Hunt Institute for Botanical Documentation

INFORMATION FLOW THROUGH A TYPICAL COMPUTER SYSTEM

Key: Information | Computer Components | Peripheral Equipment

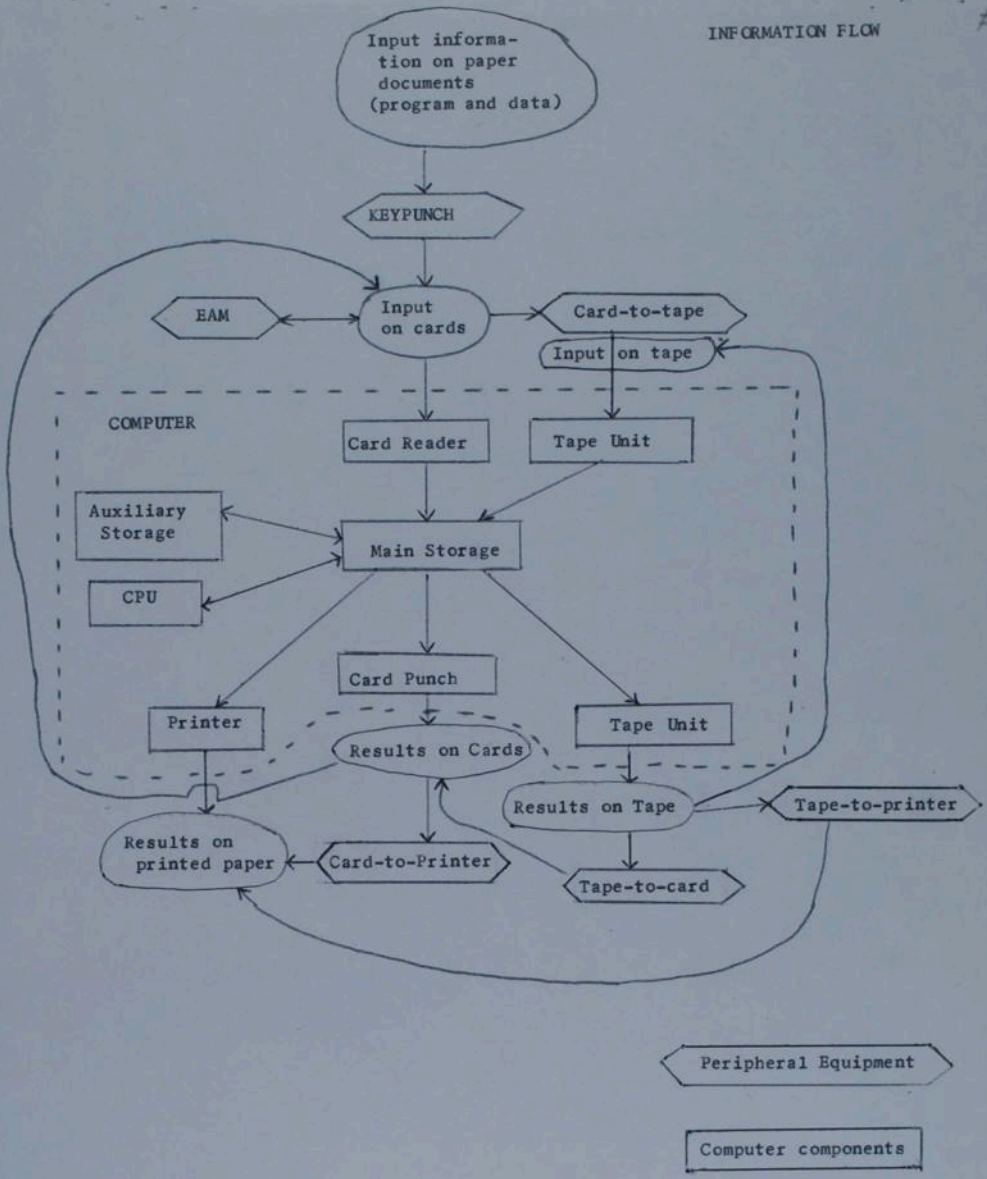
All the components inside the dotted line comprise the hardware of the computer.



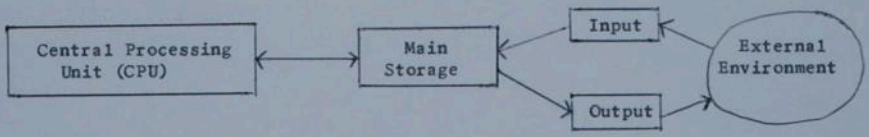
Digitized by the Hunt Institute for Botanical Documentation

#2

INFORMATION FLOW



Digitized by the Hunt Institute for Botanical Documentation



Digitized by the Hunt Institute for Botanical Documentation

Illustrations such as lectures are attached.

B 170
March 30, 1966

Lecture given by Mr. Robert Brill

I'm not going to make any definite assignment in this , but it is quite well worth it. I've read a lot of computer manuals and technical manuals and I've written a couple, and this one is particularly well put together. It manages somehow to go into quite a bit of detail on every subject and at the same time a great deal of generality, it doesn't speak about any specific machine. I'm not going to assign any definite pages to read in it because the topics we are going to talk about today are much more general and less detail than any thing that is in this book. But for anyone who wants to know more detail of what I am going to cover today, this is highly recommended. The essence of the computer is its ability to do things at terrific speed. This is the only economic justifications for its use. There is really nothing mysterious about ;it. Its operates in a fairly straightforward manner. It can't do anything that you can do yourself with a pencil and paper if you would live to be 500 years old and nothing else to do, but solve these problems. So it is only the fact that it can do this stuff at very great speed. You will as a biologist / have certain distinct advantages over this machine. To date one of these machines have been able without the help of human beings to design machines of like nature and none of them have been able to solve problems all by themselves. A man can device an algorithm that is to say a procedure for solving a problem and feed this to a machine, he can solve the problem with an algorithm with pencil and paper without a machine, but again at the present time, and I don't know if this will always be true, but at the present time none of the machines we have have been able to devise their own algorithms. And they have to be fed this algorithm step by step until in very simple-minded way what to do exactly.. Your older computers use electric magnetic relays for the switching circuits and this was fairly slow technique, speed was increased when

Digitized by the Hunt Institute for Botanical Documentation

vacuum tubes came into play instead of relays, and this is increased again in the next generation of machines by the substitution of transistors which are again faster. The current frontier of research in this speed race takes us into the area of cryogenic electronic devices which operate at extreme high speeds when the components are held at very low temperatures, close to absolute zero. We are reaching and probably will reach very soon the theoretical limit for the speed of a computer. This is bounded by the fact that electricity control only a certain ~~through~~ rate through the wire and the distance of the wire from component to another becomes a limiting factor in the speed. Apart from the hardware considerations that I have just mentioned, the biggest breakthrough in speed in the computer came with the advent of the stored program concept. This was the birth of the computer idea. I don't know exactly what the dates are on these things, but ~~thx~~ its all very recent history which is all compressed within about the last 15 years. The stored program concept means that the program which is describes the algorithm which solves the problem is fed directly into the machine as opposed to say the ~~best~~ desk calculator where the algorithm is in the minds of the man who is operating the machine. He describes the algorithm to the ~~best~~ desk calculator by punching buttons in a certain meaningful order, and the speed of his operation is limited by how fast he can punch these buttons ~~thxtxhxxxxxtxxhx~~. He has to be a pretty sharp operator to get any kind of speed out of the thing at all. With the computer the algorithm is described as a series of instructions to the machine step by step which are loaded into the main storage of the machine at the time that the data that is going to be processed is also loaded in. A button is = pressed and the computer then proceeds to execute one instruction after another of this program without the need of any further intervention, and this is very fast indeed. So that even though the human being has an advantage over this machine, and that he is still the master of it, still has the ability to describe this algorithm,

Digitized by the Hunt Institute for Botanical Documentation

The computer has its advantage in that if you decided to have a race and you gave the computer, you fed this algorithm to the computer at the same time as you gave a human being the same algorithm on a piece of paper that he was to follow out with a pencil and paper, the computer would very likely be printing out its results while the man was still sharpening his pencil getting ready to start. This terrific speed that has made this tool so very important to us. It has made possible the attempt at solving certain problems which we couldn't attempt to do before, because we didn't have enough time and energy to solve them. The other advantage that comes with this speed is that the machine is a more accurate than a human being when it comes to doing something which takes such a long time, and man's fatigue tends to mar the accuracy of his results. After a while the computer does make mistakes but it is very rare.

There are two large families of computing machines, the digital computer and the analog computer. The analog computer takes its input in continuous form, that is it takes its inputs, information such as readings from the oscilloscopes or electrocardiographs and instruments of that type. The digital computer takes information in a discrete form as opposed to the continuous form, that is it deals with symbols such as numbers, letters, and various codes, and manipulates this information according to rules of arithmetic, the rules of logic, and any other kind of rules you want to build into your algorithm. And I don't know why and I'm not sure that anyone has analyzed this but for the kind of things we are doing and for most work in taxonomy on evolutionary studies and the like, the digital computer has proved to be more suitable to our needs. I suppose it is because biologists tend to think of their data as being discrete and continuous. It is the point of view of biologists rather than any other reason. Parts of it you will find some use in our application will find some way of using the analog machine. In general/machines the digital are used

Digitized by the Hunt Institute for Botanical Documentation

much more than the analog machines for all applications. The digital machine can be broken up into two families, the binary type computer and the decimal type computer. At bottom all the computers are binary. I'll explain a little more about that in a little while, but the decimal computer using some sort of binary code underneath it all, deals with information in the regular decimal system with a base of 10. The binary computer uses the binary system which has a base of 2 and more details about it. (Bert) Did you say that actually all these digital computers are binary at base? (Bob) Yes, (Bert) The output here is either being xxxxxxx decimal or binary? (Bob) Even on the binary computer that we use, the input and output is in decimal because it is translated for purposes output so that we can read it. The binary language is very hard to read. It is very easy to deal with in terms of but very difficult to read. I don't want to get into this very much because it is not too important, but the decimal machine is binary in a sense that the components the electronic components of the machine have only two states, you can think of these as negative and positive, you can think of them as zero and 1, you can think of them as yes and no, any kind of bistable situation. But the decimal machine actually does its arithmetic on decimal numbers which are coded in terms of series of zeros and ones, with binary at bottom because all of this electronic machinery, all the circuits is made much more simple by doing it this way. But when it gets arithmetic it actually takes a number which represents the decimal number as a series of zeros and ones, but it is a decimal number and add it to another such number and internally operate it into terms of decimal. the terms of the base term. (Bert) You feed in a decimal, it translates it, does the work, and then translates it back into a decimal so you can read it. (Bob) Well the decimal machine is decimal all the way through, except that it is underneath it is coded in binary. This might be confusing for you and it is not too terribly important for our purposes. If you really want to

Digitized by the Hunt Institute for Botanical Documentation

know more about that, I would be glad to take that up with you later so that I won't have to leave out something that is important. We use the binary machine. It is faster than the decimal machine to do arithmetic directly in binary and it is more economical in storage which turns out also to be anything that saves space which we'll explain a little later also tends to save time. It is this time problem that we are always working on. It is this getting more and more speed out of this machine which is its only justification for our using it.

Now I'll tell you something about the hardware itself. I'm not going to tell you about any specific machine, but I'll try to give you a general picture of what any one of these large scale digital binary computers is like. I'll give you a little flowchart to start with. We have something called central processor, central processing unit, sometimes called the IBM, CPU the central processing unit. This is the main calculating unit of the computer. It is connected up with the unit that goes under several names, we call it main storage. When you hear talk about memory of the machine, this is the area of the machine that is being discussed. In recent years, people in the industry have tried to get away from the term because it has anthropomorphic connotations which scare people. Out here we have input and output, and here we have an area which is not the machine which I call the external environment. An external environment is the place we live outside the machine, this is the place where we have problems in taxonomy and this is where we make up our data to go into the machine. When that data gets put into, onto IBM cards, it goes over to this box which is a hardware which is basically a translating device. It doesn't have any power of computation, it merely translates the information that we have out here into a form that we can read and understand to a series of zeros and ones which is handy for the machine to manipulate. None of the information content is lost as it passes from here through here and into here. It is merely

Digitized by the Hunt Institute for Botanical Documentation

transformed into another form. Everything that comes into the machine winds up in here. And this is a sort of repository for information which is of use to the calculation that is going on. The calculation itself takes place in here. There is only a small area upon which the sort of fromm which this data manipulation is done. It is just one or two very special registers which have the power to transform this information and so things are picked up one at a time from here, operated on, put back here, and may be picked up again and used again and transformed, manipulated, put back in here and finally when we are ready to see the results, put out through one of these devices which is the reverse of this, it puts it back into a form mf which we can read and you have got results. The basic medium for getting from here to here is thee use of IBM cards. I think all of you living in this day and age are familiar with IBM cards even if xxxxx you are not too familiar with computers. They have got holes in them. These holes are sensed by machine as they are read into here and information is picked out as a series of holes, but not holes, which corresponded to zeros and ones xkixk that we get internally. (Drawing information flowchart) This is where we start. We start with some taxonomic data on paper documents that says this is information seleced in the field and so on, has been coded by you in a way in which you think it is meaningful to the taxonomic processing. We want to now put this into a computer. We go through a machine called keypunch. This is the basic machine between us humans and these unhuman machines we are dealing with. All data must ultimately go through this. They are bypassing this now. It is not too much in use but there are optical scanning devices which can read documents if they are prepared properly. Any old scribbled note to the milkman can't be read by this thing, but they will read some kind of handwriting. This gives us something called input on cards. Now this flow chart you can draw a key for yourself. I'm drawing these boxes in different shapes and they indicate

Digitized by the Hunt Institute for Botanical Documentation

different things. The oval boxes indicate information as opposed to this system. I want to give this some sort of a, this shape. This indicates that this machine is not part of the computer, but this is part of the peripheral equipment and the computer. This is the card reader and this is the basic input device. I'm not going to talk about all the input devices available, but tell you the main ones that are now in use. You can also go to a piece of peripheral equipment called cards tape. This gets a diamond shape. This gives us the same information we had before but it is now on tape. Goes in a round box. That tape is now a tape unit, it is another input device and these things go into main storage. There is a whole family of other parts of peripheral equipment which may or may not be used in conjuction with this operation, is thee older type of IBM equipment prior to the advent of the computers which go under the name of EAM that is electronic accounting machine. And these include such things as the collater, sorter, the interpreter, the reproducer, and the tab machine, called the accounting machine. These are card machines. When you once have these cards prepared, they you can go through all those sorting and collating, operations necessary before putting them into the machine.

This is what was done before computers. All computers was done with these slow machines. computation going on CPU. We have something else called auxiliary storage about which I'll speak in a moment. This is a part of the computer itself. This will be on your other flowchart too, the earlier one I drew with the double arrow comin g out of main storage. When the computation is finished, this is the blowup of the other one, the other picture I drew. When we have finished our computation we'll take it up here somewhere. We have output coming out in the several possible forms, but the principal forms are again cards that come out on a card punch. Now we are talking about output devices. You see the tape unit serves as both an output and input device, and this gives us printed

Digitized by the Hunt Institute for Botanical Documentation

results that we can read, and the process of must stand here with some sort of a document that can be read. We come out of cards , this thing goes to another piece of peripheral equipment called card-to-printer and this is the same piece of equipment I just spoke of as being in the family of known as IBM equipment the tab machine or the accounting machine. It is an old machine IBM has had for years, basically it just reads cards and prints them out. It can also do things like pick sub-totals and do a certain amount of arithmetic and so on. And the card depends on, will give us printed results, or these cards can be saved for another run so that, this is going to get a little confusing, that the cards come out, I should have a little box in here that just says results on cards that is from this box. I wanted this between and information media. Information media of cards, tape printed. So these results or cards come out of the card punch and these cards can either be fed back in as input input to another run or they can go through this card-to-printer and you get printed results when you are finished. Likewise, when we come out of here we have results on tape which can do the same thing that feeds into this box. That tape can go into its input into another run or it can go through a piece of equipment of card-to-printer and give you printed results. Everything that is in a square box is part of the main computer itself. Everything that is in the triangle boxes is extra equipment that you will see hanging around in any computer installation to facilitate, to manipulating, translating of this information from one form to another. I probably left out a few things.

It will probably be quite considerable time before cards are out-dated, because they still represent the basic link between man and machine, and there is such a precedent for the use of cards in the previous IBM equipment. The problem of compatibility with all their equipment as new equipment comes out. The demand that the cards be available, cards are still the basic medium. Tape is much faster than cards as input-output medium. The problem with it is you

Digitized by the Hunt Institute for Botanical Documentation

can't procure a tape directly on a tape punch, not magnetic tape I'm talking about now, not ticker tape which I am not considering today. As the tape itself gets (student interruption) Magnetic tape which is on this squawk box but it is coated with some kind of iron oxide compound, and magnetized, and the information stored on it that way. Ticker tape has holes punched in it, and this is a fairly minor input-output device. There are other devices I'm not considering. I'm giving much more detail than I should, and I want to get on with other things. Tape is faster than but we'll see what its disadvantages are in a while. I want to talk about the main storage for a little bit.

Tape 2.

The main storage is described as being random access, in other words a better way is to say that it is addressful. And what that means is that the main storage is divided up into a certain number of units depending on the size of the machine. The machine I am using at the moment has 32,000 words, roughly 32,000 words. A word, a computer word is enough space to store one fairly large number or several alphabetic characters. Each word in this storage has an address. On the instructions of the machine, instructions which are being executed here at Central processor, every instruction has some, most instructions have an address in them. That address tells them which word here at main storage we are going to now operate on. When the center processor knows the address of some piece in storage, it can access that piece of information in a very, very rapid time, this is measured in millionths of a second. This is where the real speed of the computer comes in. The input-output devices that I have been talking about are relatively slow compared to computing time, or the time it takes to execute one

Digitized by the Hunt Institute for Botanical Documentation

of these instructions and the time it takes to access the data out of main storage, bring it in here, execute it, store the results back here. This is a millionth of a second input-output thousandth of a second even up to minutes depending on the media and we don't have to go into that now. Tape you see is not adjustable. If you want a piece of information and it is out here on tape, a section of the tape will have to be read into the main storage under the control of the program, the information will have to then be searched and brought in here and operated on, and so on and put back again. It is very slow compared to this sort of operation. We can do directly address and directly access the piece of information from here as fast as any other piece of information in here without some sort of sequential search. We are using a 7094 for taxonomic operations. We don't use any of the machines here at the Stat Lab. They have here a 1620 and a 1401. Both of these are small machines. Our problems are characterized by the use of large amounts of data. The processing and handling of large amounts of data, so the larger the storage we have, this kind of storage we have, the better off we are the faster we can go. So we use the 7094 which is a bigger machine. It sits out in Los Angeles and we talk to it over teleprocessing, telephone lines. Now this auxiliary storage is storage which is used to hold any information that is just plain too big to fit here, and it is characterized by being faster than tape, faster than cards but slower to access than this. It is, however, directly addressable. And the two principal cards in use today are drums and disks. If you want to know about drums and disks, look in here it tells a lot about that. Its not so important for us to go into detail about it. But one characteristic about drums and disks both is that you can turn off the computer when there is information stored on drums and disks, and go home and come back in the morning, turn the machine on and the information will still be stored there for ; you for use. This is true of main storage in the machine. That ;information disappears. The capacity of auxiliary storage is quite huge. This drum generally bigger than the

Digitized by the Hunt Institute for Botanical Documentation

main storage. You will notice that this is not in any way without the device. You can only come in through card-to-tape to here then up to here where it can be stored indefinitely, and then when it is wanted, it is brought through the program control back to the main storage and then used in the CPU as needed. So there is no way of directly putting information into here from the external environment. When they there is something now called the data cell, I don't know so much about it but it promises to be a device of this type. This might eventually makes tapes obsolete. Because tapes are also used as auxiliary storage even though they have the additional capacity of being input-output devices. You will use a tape just to dump some stuff out on while you are processing. You need more room here than you've got, you dumpsome stuff out here on tape and bring it back when you need it. But when this tape is so, gets success in the field, tapes could be because the data cell has the advantages of drums and disks. It is directly addressable, it is quite fast, it is quite huge, and it can be put on, it can be put off from the outside. You can just slip one of ; these data cells in and there it is with information that has been put on by some peripheral device out here. You can just th e thing in and it is already with the information in it ready to be used. I wantt to tell you something about the way information is stored in the machine. In these words of main storage, I'll give a specific example, it is just easier to talk about. I'll talk about the 7094 which has 36 bits. Traditionally, we diagram this thing, I don't know what it looks like inside electronically, but we diagram by drawing a box like this which is divided up into 36, and I'll just draw a few places in which each of these positions a zero or a 1 can be stored. This is actually some kind of ferrous material which is magnetized either in the north or south pole area, and that represents the one with zero. This can really mean anything you want it to mean. There are a

7.86pc.

Digitized by the Hunt Institute for Botanical Documentation

couple of standard codes that the computers build to understand and to interpret. But you can also code this any way you want. You can make every-one of these zeros and ones represent something that is of meaning to you, but, of course, you must write your program that is devise your algorithm in such a way that programmers scanning and looking at this information is built to understand the code which you put in there, and to operate accordingly. The standard way of looking at this information is first of all algorithms are looked at as being a binary system. That is, with the base of 2 as opposed to our regular decimal system with the base 10. I don't, we don't have the time this morning to explain the binary number system, it is if anybody wants to know about the binary system, I'll be glad to tell you afterwards.

(Bert) This 136 unit sequence here, is this one word? (Bob) This is one word of storage, and it has associated here one address. are all sequentially through from zero, they start from zero up to approximately 32,000 in the case of the machine we are dealing with. There are 32,000 words in this type, in this main storgge which is s kind of electronic scratch pad for immediate use, the stuff that you are working on right at the present time. (Student) But it doesn't anywhere near exhaust the number of possible words there that you can have with this system? (BOB) Well, that's all that you can directly access at any one time. If your data takes more than 32,000 words to store, you are going to have to bring in a piece at a time, and process it at a piece at a time. This, of course, is sometimes very, very hairy programming problems. (Student) What I meant was though, with making up words of 36 letters each, you would have infinitely more than 23,000 words it would seem to me. (Bob) I'm sorry I don't understand your question. (Student) Well, (Dave) We are talking about the main storage unit now which is by its manufacturer

Digitized by the Hunt Institute for Botanical Documentation

designed to .....(Bob) Only possible though to hold 23,000 (Dave) 32,000 (Bob) Each one of the words was classed 36 of these bit positions. Now if you're in the binary number system, all of this represents just one number. Even if its that number one, it will still take up 36 bit positions, and it will be a string of 35 zeros and one. But it can hold quite a few numbers. It can hold 2^{36} of numbers. You can also standard code which is called BCD or binary coded decimal, and this is the way, the standard way of representing alphabetic information inside the machine. This is a 6-bit code. In one of these 36-bit words, you can store 6 letters of the alphabet 6 numerical digits. Now these numbers now are not in a binary number system here they are in the decimal number system and they are coded in binary 6-bits. So some figuration of 6-bits, this will give the binary code for every letter of the alphabet, every special punctuation, character, and numerical digits if you want to get into that a little more deeply. It is very program to know this, but this might or whatever You can get 6 of these in a word. for the particular machine we are now talking about. This varies considerably from hardware to hardware.

The important part I want to get across here is that these 36 bits or how many bits it is going to be can stand for anything you want them to stand for. You don't have to use binary numbers, you don't have to use BCD, you can make up any code you want provided you write your program in such a way that it will interpret the meaning according the way you set up your xt rules.

We are in a position now to understand a little bit better what I meant by the stored program concept earlier. I said this was the in speed was when we were able to store the program itself in the machine right along with the data where it/stored here in main storage. In this

Digitized by the Hunt Institute for Botanical Documentation

machine and in most machines but not all, a computer instruction just fills one of these words and looks just like any other piece of data in the machine, It is just a series of zeros and ones. The only way that the machine knows the difference between an instruction and a piece of data is the machine is instructed at the start to begin operating in a certain area of this main storage where the program is kept, can be kept side by side with the data. Then this word is brought to the central processing unit and goes through a registered, usually called the instruction register where the program address register or something of this sort where it is analoged xxxxxhxxxxxx piece by piece for what instruction is, what is to be done, and on the basis of what it sees it then goes ahead and executes the instruction. The is to execute one through storgge. If this is the address, 001, after it finishes executing this instruction it will go on to 002 put up the instruction there, execute it and go to 003 and so on.until it comes to a transfer-type instruction. A transfer-type instruction is an instruction which simply states stop processing this thing sequentially and execute your next instruction from address so and so, follow that instruction to the main address where it tells you to go, where it tells the machine to go to pick up its next instruction, it will go out to the other part of storage and start executing there in a sequential manner until it comes to another such transfer instruction.

14.3

Now because jumble all of the storage that way executing instructions, it can also pick up data from any place in storage, it can treat an instruction as a piece of data because it is sitting there in storage just like any other piece of data and because one of your instructions might be go to a certain location and pick up what you see there and do something to it, add some numberr until it subtracts the number to manipulate in some way and this piece of data

Digitized by the Hunt Institute for Botanical Documentation

might be one of the instructions that the very program is now operating. So this gives the facility to the machine to modify the program while it is running on the basis of intermediate results. This introduces the process known as feedback which is very, very important in computer theory. Feedback is the modification or adjustment of the procedure of the algorithm on the basis of intermediate results. There are all sorts of examples of feedbacks in other applications, engineering, science, and so on. The instruction which makes this possible, extremely critical type instruction, is the transfer instruction known as conditional transfer. The transfer/ I can start here is the unconditional transfer. When the machine comes to this in the program, it just says stop executing sequentially and go somewhere else and start executing without any conditions attached the machine just goes and transfers to some place else. The conditional transfer is a transfer which the machine has, it is given two pieces of data as an example, and it is told to compare them. A certain one is larger than the other, then it transfers to place to get its instructions. If these two pieces of data are equal and algebraical in size, and magnitude, the computer is instructed to go to some other place and maybe execute a different routine. If the one is smaller than the other, then we might go to the third place and so on. This is a conditional transfer. Certain conditions are met or not met and the transfer takes place or doesn't place. This permits branching in the program, This permits the type of flow that ;you see here. This isn't program flow but it is a similar type of thing. You might see something like I will do this in flow chart, though I'm not in program flow, but the program will follow some technique something like this. You ask the question at one point in your program, if a = b, is a equal to b, if yes, well then you might know and compute some value ; if not,

Digitized by the Hunt Institute for Botanical Documentation

you might go to some other value here and this branching can go on and on and it won't and you may want to go back here. You might first change the value of a let's say here, it may make a equal to some other number . You come back here put something down this way

20

. And so on. These programs get filled up into very complex arrangements of this type. ~~The~~ A conditional transfer is what often makes possible this branching or what is/called decision- making in the computer beginning at is exactly what it is in the computer, besides on the basis of some intermediate result what to do next.

Digitized by the Hunt Institute for Botanical Documentation

April 4, 1966
Mr. Robert Brill
B 170 Lecture

PROGRAMMING AND LANGUAGES

I am hoping that there may be some time at the end of this for some questions. I may have to drop a few topics on the agenda to do this. But I have a feeling that some people might fave questions left over from last time and you might have some today. Also feel free to ask the questions as they come up. I'm going to back up just a little bit into last week's lecture and amend the flowchart I gave you, the general flowchart I gave you last week. It will look like this. Follow the various breakdowns in input and output that I include here, but we are not going to be interested in that today. Today we are interested in programming and languages. And there is a piece of hardware on here that wasn't on here last week. You could think of it as being part of the central processing unit, and it is called the console. Today I want to put it separate on the chart because I want to talk about it a little bit. The console is the where in pictures there are people sitting at computers, you see a scientist in front of this device buttons, lights, and switches, but in real life, you rarely see a scientist in front of one of these devices. You rarely see a programmer in front of a console, because what you see there is called a machine operator. He is a person who occasionally wants to take over this control of this machine manually, and that is what the console is for. It is an input and output device, but it does not go directly to a main storage in the automatic input and output devices do. It goes directly to the central processor through which you can control the rest of the machine. You can sit there and work the machine by typing on a kind of keyboard or pressing certain switches in a certain meaningful order and you can key in instructions and cause the machine to do anything that it can do on automatic control. There is an output device that you can read the display either in lights or there is a

Digitized by the Hunt Institute for Botanical Documentation

sort of typewritter device which this starts typing back to you, and you can talk by the means of this typewritter to the computer and it will talk back to you.

When the machine is not under the manual control the console, all operations take place automatically, that is, under the control of the program that you have written to go into the storage. So the program at the beginning of down here in the form of cards around the tape drive waiting to come in. It is brought in automatically by another program, since everything is on the program control, and so the program is already in there and it is called the loader, and, of course, it naturally arises out of the loader here and there and again that is where the console comes into play. There is a button on the console that you press and you get a couple of instructions free that way. That brings in the loader, the loader sits out here some kind of a, that brings in the loader and once your loader is in, programs can be brought in automatically until you shut down the machine for the day. Once the program is in main storage, has control of the machine until the operator decides to interrupt the other console or until the program causes itself to stop. So once the program is in it then reads in the data for itself, does the calculation,necessary, and prints out the results. Now you will notice that the talk about the languages that the program, the data, the results are in. The program must be in the machine language of the computer that it is going into which the computer acts on everything in this red box. Whereas the data and the results can be in any language you want, and this gives the ve ry intersting results. Reasons for this, I have to make a diversion and tell you something about what these instructions actually look like to explain why they must be in the language of the machine. I am going to talk about a hypothetical machine. This machine like other binary machines can see nothing but ones and zeros, and this is again a slight review of what

Digitized by the Hunt Institute for Botanical Documentation

we did last week. Here is a small section of storage in the machine. Now this is a small machine, this is not the 7094 which has 36 bits in every word. Each one of these boxes here is a word ~~xxx~~ the hypothetical machine has 12 bits. As you can see, all of these really look alike. In a minute you see that some of these are data and some of these words contain instructions. Instructions and data look exactly alike inside the machine. And most instructions in most machines and they are in this hypothetical machine at any rate you can divide up the word ~~xxxx~~ in two sections. In this machine the first 6-bits will represent an op code or operation code. Once the machine this in such a way as to say, "Now I've got hold of an instruction." It looks at these first 6-bits and depending on the 6-bits are, what this code is, it knows what instruction its got and goes ahead and executes it. And it executes it upon information found somewhere else in storage at the address named in the second part of the word. Again, this is not a real machine, the are a little bit more complex in their instruction formats than this. So from machine to machine, not only the ~~xxx~~ instruction sets will vary. Most machines can do the same sorts of things but some machines will have some instructions which other machines can't do and vice ~~xx~~ versa. The instruction sets don't coincide exactly. So that when this program comes in it has to be in the language of the computer it is designed for, not only because the instruction sets are different, but because the codes, even when the instructions are identical and do exactly the same thing, the codes will vary from machine to machine. There will be a certain number in here which will mean add on computer X and it will be a different number in here which will mean add on computer Y and the incom- ~~xxxx

using is precisely defined. Such a language is the decimal system. If you happen to be doing some mathematical piece of work, your data might be sitting out here in the form of decimal numbers. You might want to get results and again it will be in decimal numbers. In here you are going to be computing in the binary number system which is the most convenient system for the machine, the computer, and when George takes the floor later in the week, you will hear something about number systems and one of the systems essential for our purposes here can be considered as languages. This program must be able to translate, read in this data, translate it from decimal to binary, compute on it and translate the results from binary back to decimal. But that doesn't mean that is the only language you have to use. This can be the text of Shakespeare, for example. This can be English language, it can come in and you can do some sort of concordance / on it or something like that and put some results again in any kind of language you want, not necessarily English. There has been a certain amount of minimal success with translating kino languages. Russian to English, for example, now runs on a computer but it is fairly crude. The reason probably is that the rules for these languages are not precisely defined, and the correspondence between these languages is not exact. (Student) Would you say then that one of the main functions of the input and output sections is to translate a language? (Bob) Exactly. You can think of these pieces of hardware here as translators also. This takes us backwards, but the data when it is sitting out on paper, you know when you have made up the data, might be in decimal. By the time you put it on cards, it has already been translated into something called Hollerith. That is the language of the holes in cards. And when it goes through the it gets put into storage into something called BCD or binary coded decimal. Its still basically in decimal numbers or whatever lines you have out here. But the form you have been translating by this box. So the hardware input and output are translators and the program itself has to have

Digitized by the Hunt Institute for Botanical Documentation

some translating connected with it, reading in routines and writing out routines. Translation is very, very as computing. And because of this nice ability, we don't have to address the computer in a language like this, this is what the computer must receive. But we can address the computer in languages on a higher level and have the machine translate for us. Now I'm going to run through this example, and we will be able to see a little bit what that is like.

Now these numbers out here represent the addresses of these words in storage. This is the address for this word in storage and so on. The numbers that are in these boxes are the contents of storage at this particular moment in time. Now these last three happen to be instructions, and we'll just draw this conceptional line here to indicate that to show that this is the operation course and this is the address. Now this particular computer, this number here tells the computer that, I have to explain first that we have a special register a very important register here which we call the accumulator. And this says, "Go to location five." And this is the binary code five. This portion here is an address, and it says,"Go to location five and take a number that you find there and stick it in the accumulator." This is location five up here, and it has this number which happens to be a .12. I'll write that in decimal now, but what you are going to get is this bunch of numbers, bunch of ones and zeros up here, stuck in here. Then the machine goes down to the next instruction runs sequentially, picks this one up and says, "Well, that means I'm supposed to find take the number I find in location 6 and add it to what is in here. In location 6 we have a Y. 13 in the accumulator. Most of the arithmetic in the machine is done here in the accumulator. It finally says, "Take what we find in the accumulator and store it back into at location 7." Here is location 7, takes this and puts this 13 in here. (Student) Now am I to understand all of the continuation, I mean the information part of these last three, does it have two addresses?/ Are those numbers on the right also addresses the ones..... (Bob) These numbers here, these last

Digitized by the Hunt Institute for Botanical Documentation

6-bits are addresses. (Student) What are these three in the margin then? Are they also addresses on the left? (Bob) Out here? (Student) Yes. (Bob) yes, these are addresses. This is the address of this word, this is the this word. Each one of these addresses corresponds to a word in memory that it is next to. This doesn't look like this to the machine, but each word in storage does have an address associated with it. These addresses are part of the instruction. (Dave) I think the problem is that the address on the right is the place where you are to go in memory for a particular item. The address on the right tells you to go to the first address, the top of this box. (Bob) This address corresponds to this address. (Dave) You will find something in that address that you can use; whereas, the addresses on the outside are those pieces, where the words are in the machine. (Bob) So just to run this quickly, this instruction tells the machine this lefthand portion says, "Take up what you find at the location named here and stick it in the accumulator." Here is the location named there, this is a piece of data here, the decimal number 12 is stuck here. So these first three are data, the last three are instuctions. (Student) Can ;it run sequentially? (Bob) It runs sequentially until it has a transfer type instruction I talked about, I was just talking about that/last week when we quit. The next instruction here says, "Transfer/to the addresses found here." Then the control will skip down to memory to the address so named, or back up again so you can have loops in the program. And there are certain transfers which will be ;conditional on some, the result of some comparison. It says if you find the, look in/a certain place and if that is a zero, transfer it to the place. If it is not a zero, then just continue with the sequentially operation. And if this type of transfer, particularly the conditional transfer,which permits the to the program. It permits the speedback mechanism to it where you can go back and iterate a procedure

Digitized by the Hunt Institute for Botanical Documentation

putting in different values each time you go through. Do one routine on the basis of the intermediate results or do another if the results are something else. (Student) I"m still not clear how you know that the lower three words are instructions. (Bob) There is no way of knowing but when the load program brings in this program, this program here, control is passed to it, and part of what you have to write in is something that says, "My program begins at location such and such." And then the loader, the last instruction in the loader is a transfer. And a transfer is to that location so named and now your program is in control. We'll do something that says to the loader, "Start my program here in location 8."

(some questions by students which cannot be heard clearly)

This is a very small computer. The addresses only have 6-bits. There are only 64 words in the main storage of this hypothetical computer. If you think this is hard language to write, then the men writing for a real machine like a 7094 which has 36-bits in each word, and each address is 15 bits long, that is why we developed ~~the~~ languages. So the low programming, the ~~fixxt~~ last thing it does is transfer to this address, and now ~~xx~~ the computer and central processor takes up this word and takes it to place where it analyzes it as an instruction. And if you should transfer to another area, you are in trouble, because the machine will just pick up that word and ~~xtxxt~~ try to execute it, and if it is not a valid instruction it will hang up. the assembly language. And this language we have the machine will look something like this. You write this down on a piece of paper and it is punched up into cards, each line here goes on a card. I'll tell you how this language works first I would take this problem and then I'll tell you how this language gets translated. Out here in this column, which you are writing are symbols which stand for these addresses out here, and you should put in something which says, "Origin of my program, will be at location 8, a 5." So this first

Digitized by the Hunt Institute for Botanical Documentation

thing says, "I want to call location 5, I want to call that by the word symbol I, and I want that to be a .12. That gets translated from this right to this, and when the loader gets it it has instructions to put it in location 5, decimal location 5. Similarly, you get a one here and you call it K, and we don't put anything in here, we are going to store our results in here so we can start out with zero, the J's We can start out as a zero and this is where the program begins, so we will call that location begin. We can call that, this is a code in our language which says, " accumulated, has certain value. So that you can remember that that means accumulated. When you unload it, with the contents of location I, and that translates as this and gets put in that location. And now since I has been defined as a five, , the translating program will pick up this, reads it and says no that means accumulated so we have to put it in this series of digits which means to the machine accumulate it. And this is I we have defined over here as being a five so we will put the address five in here. Likewise, this ADD is going to mean to the translating program add and put in the code for add and put in the address symbolically specified. Likewise, you get that instruction and your program willl continue, and the very end you put in a card that says end, this tells the translator that you have finished and that is going to tell the loader, that's going to produce the information to the loader that tells where the programbegins. (Student) What did the last STO stand for? (Bob) Store. Store the result in J which is this location here. This begin tells the loader that I want my program to begin, and location begin will then transfer it into here and start the thing off properly. Now, I would like to have had this out a head of time (drawing on the blackboard) Now because of the facility for translating languages which I mentioned earlier, is a we have a program which translator from this language to that language,

Digitized by the Hunt Institute for Botanical Documentation

So this flowchart corresponds to the one I just erased part of. This is the program in the language of the machine ~~which~~ this M(x) stands for machine language of computer X. At the very beginning somebody had to write the first assembly program in that language. It only had to be done once From then on all of these items could be written in the assembly language. So here it is, this assembly language is in already in machine language and is ready to go in here. And what this program does is translates the assembly language for Computer X to machine language for Computer X. It does this job of translation that we just discussed. And the data for this program is a program in the assembly language and that program is the general terminology as a source program. So this program is a program written in some higher level language which is going to go through this assembly process. And because they and the results can get any language you want we have translations from this language to that one. As a result, it is called an object program. And this program is now ready to run in the machine. So now we have another run where this goes into Computer X, and for its data anything that the data for the particular problem that you want to run this time, in any language that your object program is capable of translating and results in any language that your program is designed to do. Two runs on a computer are required to run a program written in a higher level language. Now this language here, every machine has it. And every machine must have a different one because this is a machine or language. The statements in this language are all basically just symbolic versions of machine coding. In fact this is sometimes called symbolic machine language as it differs from machine language so very little. In the last few years they have facilities for this language and make it fairly fancy, but it still translates almost entirely on a one for one basis . But languages are now more powerful, higher level than this has been developed and most of you have

Digitized by the Hunt Institute for Botanical Documentation

probably heard the word Fortran at one time or another. There are other languages like Fortran. Fortran is the most famous and it is the one that we happen to be using right now for our program. This same program written in Fortran looks like this. And you notice here that there is no reference to the machine at all. Fortran is not a completely machine-free language, it (end of tape)

There are also some better ones than Fortran for that. This says I equals 12 establishing the definition of I and gives you the value of 12 at the same time. Now that corresponds with the statement here. K equals 1 corresponds with this statement here. And J equals I + K says, I K together put the results in place called J and J at the same time. This one statement takes the place of these 4 statements starting with J decimal zero down through this three construction program. When this, we just take this out and just make this L for language for any language, you see, you write this and it gives you back this, and incidentally, goes through this stage in between. It really translates from this to this, and from this to this and does it all in one computer run. In this particular it may not look as though you are making any great decisions, using this language over this language, but that is because this is a very trivial example. In most cases Fortran program will produce a many, a one to many output which will be on the ratio of, it depends entirely on what the program is and how much input and output there is, for it can give you 50 to 1 or 100 to 1 sometimes. Some Fortran statements will only produce one machine instruction, some will produce a great number. (Henry) compare if you don't mind, Bob, that the equal sign does not actually mean equal, it means replaced by a . (Bob) It means that the quantity lying there on the right-hand side of the equation replaces the storage location named J, although once you are in Fortran you can think of this in quantity J or the variable J rather than the storage location J. (Bob) Do I hear a Question? (Jack ?) Not a very smart one I guess, how come J isn't equal to 13? How come you haven't got 13... (BOb) J will equal 13 at the end of this result. (Jack) How come it doesn't equal 13 there, how

Digitized by the Hunt Institute for Botanical Documentation

come that.... (Bob) This is a zero? (Jack) Zero, yes. (Bob) It doesn't matter what it is at the beginning, because we are going to store some result on top of it. We have to define it some how. So we start it as a zero by habit or convention, it doesn't matter what we start it at. (Student) What you store on top of this, is this beginning number lost? (Bob) Yes, it is wiped out. When you store something it wipes out all that was there before. However, when you pick, when you pick up five here to load into I, the value of I still remains in I. When you take something out of storage, you are not really taking out but taking a copy of that, the quantity still stays where it was, but when you something back in, it wipes out what was there. Now twist your heads a little bit more, I'm going to show you how languages can be transferred from one machine to another, if they are general enough, if they are machine independent enough to do. A couple of languages have been translated like that. Fortran is not, when Fortran is written for Computer X and then Computer Y comes out and they want Fortran for Computer Y, the Fortran translate has to be written all over again in the assembly language for Computer Y. And this is because Fortran is not powerful enough to write the Fortran translator. Fortran is designed more ~~rfm~~ for mathematical routines than for the manipulation of and symbols. And these translating programs require a great deal of manipulation of symbols, weeding of input strings and analyzing and translating it. And Fortran translated is such a program, but the language is not powerful enough to write the translator, but it has happened a couple of times in computer ~~hixnx~~ history that languages and developers have been powerful enough to describe themselves, that is describe their own translators and in this way you can get from one machine to another. Language from Xpop has this power and language called Jovial has this power, and both of these techniques have been done. The Xpop procedure still is in the middle of this process. (Writing on the blackboard). We start with a program in the machine language in Computer X.

Digitized by the Hunt Institute for Botanical Documentation

That program is L in mabhine language of X Translator. Now this language is going to be this powerful language which can write so. This can be Xpop, Jovial, or any of the other ones I have mentioned. The data for this is a source program written L which is an L to machine language of Computer Y translator. A base run through one run in Computer X and out comes an object program that runs on Computer X which is L to M(Y) translator. Now you want the same source program through again, that is Xpop written in Xpop, or L written in L. the same source program on the same computer but a different program is running now. This is going to be L to M(Y) translator. So out comes the object program in the machine language of Computer Y which is the L to M(Y) translator. This now run in Computer Y and it will take source programs in L and give you all yourr programs in M(Y) for furthering running on Computer Y. Does that come straight? (Student) Translate this into machine language of/Y, can the computer estimate that? (Bob) No. Computer X can read that if you are thinking of trying to run this object program ~~in~~ here, you might run a computer X. The object program is in that language. (Student) I'd like to ask a question that is kind of off the path here. Sometimes we hear people question whether or not a computer can think, now I would suppose that the answer to that question probably depends largely upon how you define think. So how do you define think in terms of a computer, so far as you are concerned? (Bob) That is a question that has for most people a scary implication, and sometimes I get a little scared myself, on MOndays, Wednesdays, and Fridays. The rest of the time I don't think computers can think. If by think you mean can a computer make decisions, do calculations on the basis of results that it is ~~gaingxtaxpxsdxxx~~ has already produced, change course in mid-stream, and so on, ~~xxx~~ that sort of thing, yes,. Computers can do this, but computers do not, as yet at any rate, design programs for themselves, (Student) But they could? (Bob) We don't really know that. There are people working on a few of the computer program research called artificial intelligence.

Digitized by the Hunt Institute for Botanical Documentation

What they can try to do is to get the computer to design algorithms for itself. I hit you with this word algorithm last time and I got some kick back on it, but this word is not in the vocabulary of most botanists, and it simply means that procedure for solving a problem and it is expressed in the computer in terms of a program. A program describes the algorithm. We are trying to get the machine to describe its own algorithm, and if they suceeed in doing this, I would agree that a computer can think. They have had some very ambiguous success with this. They have a program which one part of the program was called Herman and another part of the program was called teacher, and another part of the program, I think, was some small trivial body of knowledge which Herman was supposed to learn by means of the teacher. And Herman learned. And Herman said we'll do on the computer and he was able to improve his performance with this body of knowledge, I don't remember what it was, something very trivial. And then just to fool around and see what they could do, they took the teacher part of the program out and discovered that Herman did just as well without the computer. That was the last that I have seen about artificial intelligence. A lot has been going on, this has been a long time ago. A lot has been going on since then, I haven't kept up with the field. Nothing terribly come to light or we would probably see it in the newspapers. (Student) I can see how if you define think as making decisions on the basis of what you know is possible in a computer. After all, that is possible in an adding machine. So when you give it something, then you pull the crank intended to make a decision, then it tells you what its decision is. But.... (Bob) It doesn't make its decision. By decision I really mean is that the conditional branch instruction, whereby you might get it. You might get something that says, this is not the program now but this is the flowchart which describes the programming, a = d. You ask this question in the program to the computer. And the computer can decide like take a and compare it with d and decide whether a is equal or not equal to d,

Digitized by the Hunt Institute for Botanical Documentation

and branch accordingly. If it is equal to d, it will just continue with its sequential operation. And it will do a routine called p. If it is not equal to d, it will skip to another part of the program where it will do routine q. This branching can go on indefinitely and can circle back and so on This is what I mean by a decision. (Student) Well now suppose that we say that a computer can't think, or at least we say a computer thinks, and we have defined think as the basis of decision making. Then we further conclude that Computers are not like we are, and the essential difference is that we think in some other way is the word we ;use to say think. It means something else. (Bob) I think it is even more complicated than that. I think there are many kinds of intelligence, and this many kinds of intelligence can be found among different types of people. Some people have very good intelligence for let's say doing mathematics, other people have very good intelligence for being able to say music. There are other things like some people are very good at judging the psychology of other people but can't do mathematics worth a dang, This is another form of intelligence. So if you want to want to give the machine its this say/is the way it represents its intelligence, then it thinks in some sort of primitive way, I suppose it does but from this to the usual popular conception of the thinking machine, thats going to replace its inventor and that sort of thing is a long way off, although at times I wonder if it isn't possible. I think intelligence as required for the machine to become as potent as people fear is this sort of intelligence that artificial intelligence people mean, and that the computer will have the ability to design its own algorithms and/or to improve its performance on the basis of its past performance and the knowledge that it stores up. They have a program now that plays chess, it doesn't play a terribly good game because we don't, well we don't understand the algorithm of the chess, you can do it on a complete trial and error basis, but the number of moves possible are so huge that even with the high speed of a computer, it can't be done in practical time. So we have to

Digitized by the Hunt Institute for Botanical Documentation

have some kind of a short cut algorithm. And we don't have a good short cut algorithm. We don't know what the short cut algorithms are that champion chess player use. IBM brought in Arthur who was at that time U.S. chess champion, and they said,"Arthur, what is that you do when you play chess." And he said, "I don't know." And he worked with them for a while, tried to figure out what he did, but he couldn't figure out what he did. He wasn't as good at doing that as he was at playing chess.

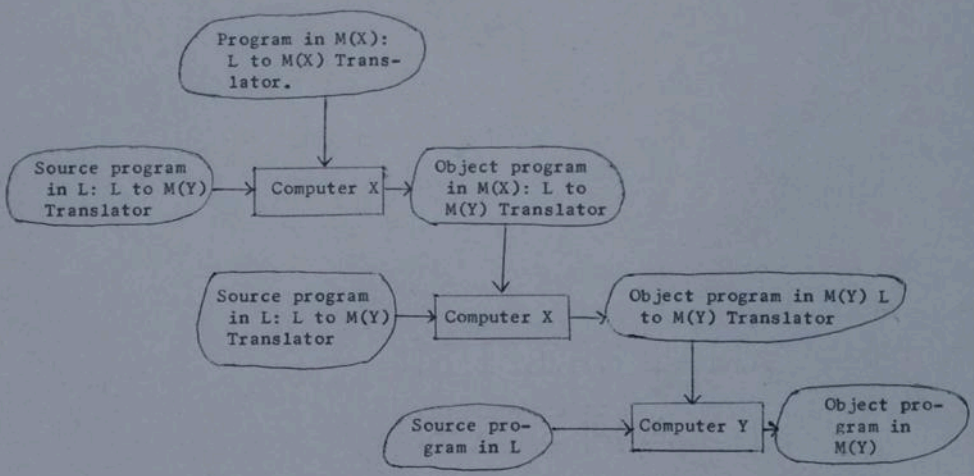
our course here this is what we are trying to do by the taxonomic process. (Henry) What this amounts to is that the computer looks at a certain pattern on the that says, 30 million possibilities here. And the computer goes through the whole 30 million possibilities and then makes a decision on what it is going to do. It takes a long time. Human beings thinks of maybe three possibilities and these are the three salient important possibilities and he makes his decision on three. (Bob) We haven't yet figured out how a human being can stand this mass of material without really looking at it and make decisions on the basis of this. (Dave) There might be another question in this thing. The computer will sit there and not ask itself why I should learn how to play chess. He doesn't tell the computer it is important for you to learn to play chess. (confusion - two or more talking at the same time.) (Student) To me that seems it is possible to think of these kinds of intelligence which he mentioned as being mathematical, artistic, or computational, or some other kind of intelligence that you normally think of could be, I could think of that as a matter of a kind of program, that is, one operating well in a certain kind of language. And if a thought could be put in that kind of a language, the person who operates in that language can handle it. (Henry) It is probably a degree of level that we are talking about. In one sense we could be brought up (much confusion, students apparently leaving class or getting ready to)

Digitized by the Hunt Institute for Botanical Documentation

without thinking I was moving that level. So we can from that venture to say that my nervous system is programmed. But there is an entirely different level going on here with what we are doing in the world. (Student) Aren't we talking about a difference of degree....(Dave) Excuse me must a minute, Bob has a calculus class to go to right now.

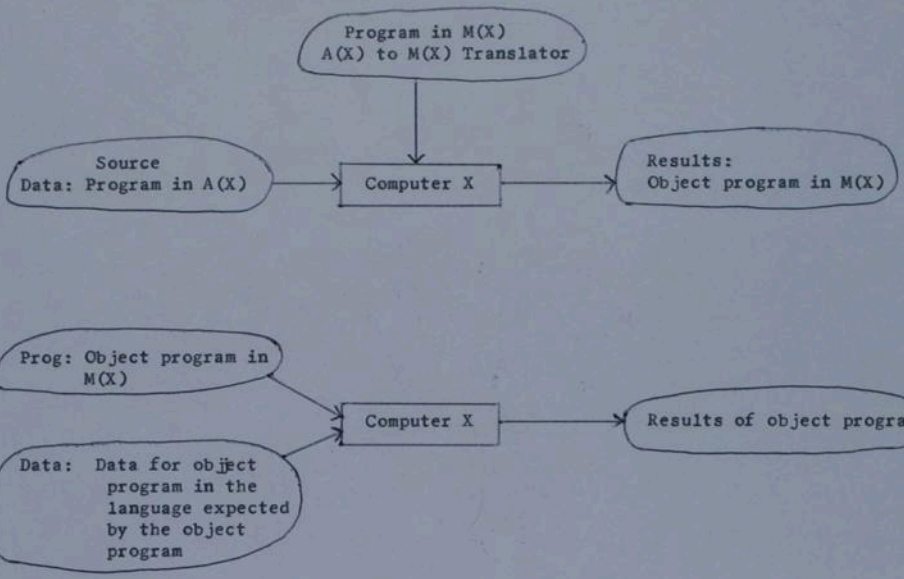
Digitized by the Hunt Institute for Botanical Documentation

BOOTSTRAP PROCEDURE

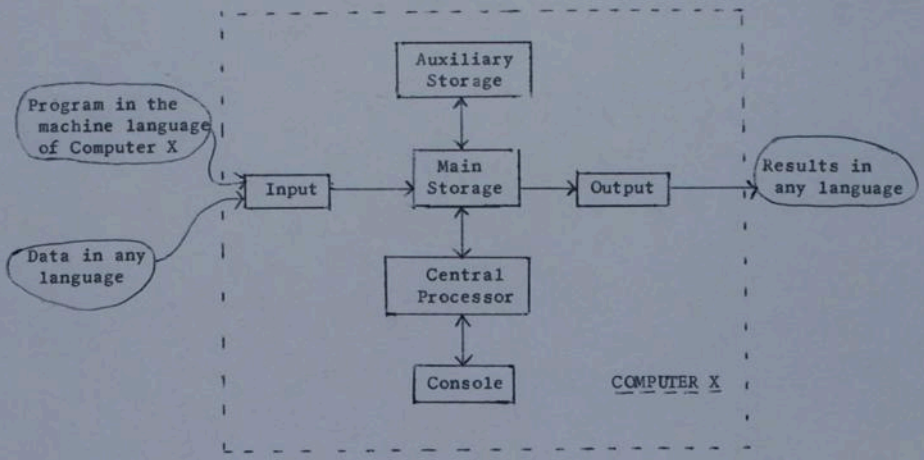


Digitized by the Hunt Institute for Botanical Documentation

ASSEMBLERS AND COMPILERS



Digitized by the Hunt Institute for Botanical Documentation



Digitized by the Hunt Institute for Botanical Documentation

MACHINE LANGUAGE OF COMPUTER X

000101	000000001100
000110	000000000001
000111	000000000000
001000	011101000101
001001	100010000110
001010	110001000111

ASSEMBLY LANGUAGE OF COMPUTER X

I	DEC	12
K	DEC	1
J	DEC	0
BEGIN	CLA	I
	ADD	K
	STO	J
	.	
	.	
	.	
	END	BEGIN

FORTRAN

```
I = 12
K = 1
J = I + K
```

Digitized by the Hunt Institute for Botanical Documentation

Inverse also

B170 Lect Apr 15 – Character

1. Bring up to date – flow chart – where are we?
2. Characters – general –
   1. Can we define? Problems tax. baseball – Not easily separating our thinking from the data
      Defin: A real-valued function, defined over all the objects in the study.
   2. This definition eliminates the arguments about "size of group", "specific or generic" character, etc.
   3. In another sense, and simplification of the definition, characters are patterns, and obviously patterns vary from group to group.
   4. Also, characters, in the sense of, the Gaming Psychology of are mnemonic devices – memory aids – learning by association – we discover that some patterns are related.
   5. Going to the basic determinators of plant + animal form and function, DNA + RNA patterns, or sequences of nucleotides eventually determine the types of materials which an individual plant will have.
      No haphazard throwing things together; rather a precise mechanism for saying "with this material, I can ~~get~~ only get a western yellow pine, or only get a human being."
3. Characters as we use them in taxonomy
   1. A character ~~is a~~ contains a small, finite number of states, or attributes.

Digitized by the Hunt Institute for Botanical Documentation

3. 2. Attributes define a condition of the character.
3. Attributes are mutually exclusive.
4. Best to give an example for meaning.
Ex.: character: leaf shape.
Attributes: ovate
linear
and linear-ovate

More on this subject Monday.

4. The use of herbarium materials in general
1. Herbarium specimens are historical documents.
~~2. Case of these~~

My outline runs for this [illegible] Ans of the herb specimen possible over thousands more herbarium sheets of species under

2. How used:
1. Gross morphology
2. Anatomy — gross & fine (cell size of epid. for evidence of ploidy; pollen size also may provide ploidy evidence; tannins in chemo-tax not modified on drying) indicating that info from dried herb. material limited only by our ingenuity
3. Labels may give extrinsic data: ecology, geography, flowering & fruiting dates, native uses, also may give intrinsic data: flower color, growth habit, ~~height~~, dimensions etc.
4. We have an inheritance from the past, and as is usual, what we want to know today may not have been that required in earlier times.
5. Perhaps we can set up a more scientific way to collect our specimens.

Digitized by the Hunt Institute for Botanical Documentation

5. Data on ~~Protocol with~~ specimens you are to use -
   1. On chart, we have already done the jobs which leads you into part II. particularly part II a.
   2. You will follow through parts II b, c, and d. 1st time for this genus.
   3. Present classif. ~~for~~ Angiosperm, Monocot, Liliales, Liliaceae -
   4. Endemic to N. Am., mostly eastern U.S., southeastern Canada, west to N. Dak., to Oklahoma, none in western U.S.

6. Protocol with specimens.
   1. Handling of sheets, keeping in folder, replacing.
   2. Examination of parts
      1. Extracting dissecting, and preservation
         Restrictions
      2. How to proceed -
         1. Lay out all specs - for over-all view
         2. Attempt a sorting based on over-all aspects -

3. Names covered to prevent influence on your thinking.

Digitized by the Hunt Institute for Botanical Documentation

B 170
Lecture
April 15, 1966

CHARACTERS

I wonder how many of you know where we are now looking back at that general flowchart of taxonomic processes. Are you using that flowchart to sort of keep yourself on track, where your road is. You should, indeed, keep this sort of thing in front of you now and again, because if we don't come and tell you what we are doing at any particular stage of the game, there you can look to see approximately what is coming off. To be sure that you and I concur in where we are, I will point to the places on here that we are beginning to discuss now. Under the detailed flowchart under Roman numeral II Computer Analysis, we have started, and last week we began to talk about the problems dealing with area IIb. Now I listened to part of a tape that was made on Wednesday's lecture, and I thought maybe I had a few more little general comments to make, some of the things that I thought needed to be said.

Needless to say, there are many things to be said, let me just say what I think needs to be said. We were, I believe, beginning to discuss the problem of characters, ~~in~~ the general problems of characters, and it probably would be wise to discover whether we can define characters, and if we get such a definition perhaps then some of the more complicated problems will be more easily answered. Problems like the definitions of characters have bothered taxonomists for a long time, and I think probably the reason they have had difficulty in the definition of characters, is for the same reason that we have had definitions of trying to do anything with taxonomy in a theoretical sense before. That is, they haven't clearly defined or differentiated between their thinking processes about plants and the data they derive from the plants. George's definition that has come to light in the various pieces of our work for a character is as follows: A

Digitized by the Hunt Institute for Botanical Documentation

character is a real value function defined over all the objects in the study. If we look at that definition in its various parts, we can see that definitions like this eliminate many of the arguments that taxonomists have had such as a character in one group may or may not be a character in another group. This is a difficulty, I know that this is a difficulty because I was speaking Wednesday to one of the big stick taxonomists in the country, and he clearly was not thinking in the same way that I was thinking about the definition of character. He wass trying to see all the data that he recalled out there and trying to find some wayy to encompass the definition for all of the stuff that he ran into. If we ;use the definition that we have just said above, we eliminate the arguments about, well, how big a size of group are you talking about. Whether you are talking about a specific character or generic character. Note the definition that I gave, said that it must be defined over all the objects in the study and clearly this is the best way to give you some sense of a definition. Now to amplify a little bit on ;this and let me get a few points across before you interrupt, please, in another sense, and amplification of this definition we can think of characters as patterns, and obviously patterns vary from group to group, but they are indeed still patterns. I shall not get into the problems of psychology here, I think this is a bit a far afield, but we can think also of characters in the sense of the psychology of learning. In this sense a character is some sort of mnemonic device, and I think you all are familiar with the word mnemonic. In our sense we could think of a character as something of a memory aid which is another way of defining mnemonic. Also in this sense, it is learning by association, we discover, for example, that some patterns are related. Now if wee go back to the basic determinators of plant and animal form and function, that iss as far as we know DNA and RNA are the basic determinators of form and function inn plants and animals, we discover that there are patterns or perhaps we may prefer to

Digitized by the Hunt Institute for Botanical Documentation

call them sequences of subunits, nucleotides which eventually determine the types of material which an individual plant will have. For example, there is no haphazard throwing of things together; rather the DNA determines rather precisely the procedures for putting materials together. We will get from any particular quantity of DNA-RNA content in a particular plant or animal only that which was predetermined there anyway. You can get, for example, only a western yellow pine out of certain patters or sets of DNA, or you can get only a human being from a particular set. So again , from the very basic materials in which we have to deal we get some sort of patterns.

Now in taxonomy, we use characters as the basic input data for processes of classification. And again, remember this. Always keep in mind what your purpose is, what characters are for. Now a character in our sense, and I'll give an example a little bit later, contains only a small finite number of states which we most frequently refer to as attributes, that is, that is our terminologyy for character states or the attributes of the character. For our purposes, an attribute defines a particular condition of a character. And again for our purposes, for our purposes here, attributes are mutually exclusive. Now perhaps it is easier for you to get an idea of a character as we use ~~in~~ it in taxonomy by giving the example. For example, a very simple one that you use in botany will be that which we will call leaf shape. Note here that this is in a real sense a mnemonic device, a memory aid for ~~taking~~ talking about particular pieces of information. It focuses your attention down such that you are saying, "All right I will not talk about leaf shape, leaf shape alone." In that sense it is a mnemonic character. Now we have to remember our original definition here of a character: it is a real value function defined over all the objects in the study. There are certain cases where such a character as leaf shape would not apply, not be useful, but remember our conditions here. And in this sense, we get some ideas and feedback on how to make such definitions from what George was telling you the other day. We have to make definitions, these are the arbitrary points

Digitized by the Hunt Institute for Botanical Documentation

upon which we must base such definitions, defined over the study, in any case it is defined over the study. If we are talking about plants without leaves, we certainly wouldn't use a character leaf shape, and so we would again have to confine ourselves to those particular situations. If our particular group does have leaves, then you have shape. After examination of materials in our study, it may be possible for us to assign a small finite number of attributes which specifically tell us what leaf shape a particular object in our study may be defined as. And so you can list your attributes for the study as being one condition of leaves ovate or leaves linear and one will continue to add attributes until all of the objects in the study were able to be fitted into one of these attributes. And we may discover that in some cases the plants in our study may have combinations such that you would have to make these attributes to accommodate the combinations. Suppose, for example, in our study for the character leaf shape we discover that some of our plants had both ovate and linear leaves upon the same organism. Another condition that we would like to use in our methodology is that our attributes should be mutually exclusive. If we wish to continue to use this definition of mutually exclusive and we discover for the character leaf shape that some of the objects in our study have both, we must add another state in order to allow for that mutually exclusive situation and the condition can be merely described as ovate-linear in which we can assign to some member of our study this condition here. Some other members very often will fit in either of the first these two categories. Now with these sets of statements that I have just made here, I am going to stop this generality for characters because we will come back to it again Monday.

Now I am willing to listen to the two questions that you had raised here. Can you recall them, or have they gone by? (Student) You answered one. (Dave) This is usually the case. If one will wait a moment one will hear something else that says it. (Student) Is the function that you were talking

15.9

Digitized by the Hunt Institute for Botanical Documentation

161

about a mathematical function? (Dave) Not sense. (George) The function that Dr. Rogers was mentioning is the same function that I defined for you last time, that is the same idea is involved, its correspondence between two collections of things. The first collection of things being the plants in your study, and the second collection of things being a collection of descriptions which describe the shape of a leaf. And the function is two sets together with a rule which establishes a correspondence between members of the first set and members of the second set. The correspondence is to each plant in your study under consideration some correspondence made between that plant and some description in this collection of descriptions which tells you about ~~the~~ a leaf shape. So that this whole describing process can be thought of as/function, a character per se to be thought of as a function some means by which we correspond with each member of the study some description of some particular interesting basis for comparison for these particular objects one to the other. (Student) Do you need a description for each of these? (George) That is what is meant sort of by saying that these things are mutually exclusive. In order that this be a real function, you say to each object in the first set we must be assigning one and only one object of the second set. So that for each plant we are allowed only to assign one description, so we are not allowed to assign more than one description./ But By means of this rule of this correspondence we are going to associate with each plant some state which describes the shape of its leaves. (Student) Then how could this be defined over all the objects in a study? (George) It has got to be. A function is defined over a set, if the set over which we stated it is defined is the same as its domain for the set of the first of the two. If you look in the handouts that I gave you where it defines function, it talks about two sets together with a rule. Where each member of the first set is put in correspondence with some member of the second set. The first of these two sets is the set over which the function is said to be defined. That is no matter which object in the first set you choose, the rule

Digitized by the Hunt Institute for Botanical Documentation

part of this function enables you to put it in correspondence with something in the second set. And that is what Dr. Rogers means when he says the characters defined over all the objects in the study. For each object in a study my character enables me to associate with that object some particular description from the set of descriptions which place the role of the second set in that definition of function. Are you getting the picture? (Dave) Now you know what a character is. (Student) Mathematically, yes. (Dave) Theoretical. (Student) I have another question. started out by saying that maybe characters can be defined or it seemed to me that ;you were wondering maybe if they can or if they can't. There seems some question about whether they can be defined. Now does this mean that there is some question about whether or not we can think up a good definition of what a character ought to be or if when we look at the plants they are, the characters that we have already thought up are poorly defined in reality? (Dave) The best answer I can give to that is just exactly what I was trying to separate there. The way we think about these things and the data that is out there, and in the past the great difficulty that people have had because they didn't clearly differentiate between these two things, you see. Because they could never find an active description which would take in all of that data as long as they had to include all the exceptions and so forth. (George) It is always entirely possible to define a function in existence to say that I am going to describe plants according to this set of rules. These are my collection of descriptions and this is the role by means of which I am going to establish a correspondence. This is always possible and nobody can contend for a minute that this isn't possible. How useful this particular character that you have defined into existence is going to be for classifying your plants is another question. (Henry) I think perhaps maybe what you were aiming at, the answer that you wanted back is: the character is something which contains comparable properties of the objects that you wish to describe. In otherwords the information is comparable

Digitized by the Hunt Institute for Botanical Documentation

information for particular objects under study. (Dave) In order to allow us to proceed with some work that I want you all to do individually, I have to give you some basic thoughts and processes on herbarium materials. The collections of materials kept by botanists in pigeonholes in cases similar to this are the same thing as a zoologist calls museum specimens, or in our case we call them herbarium specimens. In the last few years or a matter of fact for a very long time, there has been a rather derogatory opinion of anyone who meddled around with scrappy little hunks of junk, squashed out flat, dried and turned brown and plastered down to a sheet of material of paper. I think perhaps that I would like give a little bit of different emphasis on this process here for benefit for those of you who have heard these arguments have concurred with those who said we ought to stick a match to every herbarium in the world and burn the xxx dang thing down. We have to recall that there iss a long history in botany and that as our information about plants grew we changed our concepts about we needed to do at any particular stage of the game. We haven't granted been too precise in definition of what an herbarium ix specimen is supposed to be and do for us. But again, at any era at any particular time these types of things have changed. What do we have to do with it? Personally, I am one of the greatest defenders of a collection of dried herbarium materials. I think this is one of the most marvelous scientific pieces of documents that has ever been invented. Having come from an institution where they figured that xx they had about three million separate sheets of herbarium material, I was always very fascinated that if I need (use two names, the genus and the species name, that I could within 5 minutes go to a particular spot in that collection of three million specimens and discover an object, a particularr object out of all that three million and with that very rapid way of getting to a particular piece of information out of this mass of information, I could

Digitized by the Hunt Institute for Botanical Documentation

discover all kinds of interesting things. Things that interested me perhaps at that particular time. I want to know, for example, what does Quercus alba look like? And with those two words, go and look into that tremendous collection of things and find out something that looked like Quercus alba, or a thing which people call Quercus alba. And I could discover what the thing looked like; I could discover perhaps where the plants representing Quercus alba grew, and all sorts of little goodies and informatinn about them. Now another way of thinking about herbarium specimens is this: that they are a fine set of historical documents. Many of the early collections in this country will bear the label Boston, Massachusetts, and I suspect that if we went back to Boston, Massachusetts right now, we would discover that the place that the man picked up this particular specimen is now completely covered with concrete and steel and yet we have some evidence that this was indeed a plant which did grow in that area before man came along and messed around with the environment. So in a sense it is a, it can be used as a fine record of what we have had in the past, although it is an incomplete record. I also want to give you some feeling about these particular specimens because in your handling of them in the next few weeks, I want you to have the same sense of importance that these particular specimens have such that you will treat them with the respect to which they, you should give them. Now how do we use herbarium specimens? We use them again according to the dictates of the time in which we exist. Primarily, we derive information about the plant materials from herbarium specimens concerning their gross external morphology. But as we have gone along improving our data collection methods and our refinements in what we want to know about plants, we have been able to use these old brown herbarium sheets in many different interesting sorts of ways. For example, although the specimen be dried, it still maintains quantities of cellulosic material and allied substances which do not modify thx tremendously upon drying. For example, you can study quite adequately the anatomical features of certain parts of the plants. If you are ingenuous

Digitized by the Hunt Institute for Botanical Documentation

enough to know how to rehydrate these specimens and go from there, we can discover that we can get some rather indirect evidence sometimes even about the genetics of these plants. For example, we know that cell size is sometimes influenced by the number of chromosomes contained in the cell. That on the average, cells that have a doubled number of chromosomes, in other words the plants are polyploid, the cells will usually be bigger. People have recently been taking some of this plastic material, liquid upon application and dried rapidly, applying that to the surface of leaves on herbarium specimens, peeling off the dried plastic and discovering that they have a nice imprint of the cell sizes, which they can adequately measure and compare to see whether or not there is a range in the variation in the cell sizes of those plants, and they can discover whether the particular group of plants that they are working on are polyploid or diploid. Again, the evidence is indirect, and one would certainly want to check it, but a sufficient number of checks have been made such that they can indeed discover the evidence of polyploidy from an herbarium specimen. There are other ways to do the same thing, sometimes, it doesn't work in all cases but it sometimes will. From an herbarium specimen , one may extract pollen grains, and pollen grains are again are direct evidences of the contents of the nucleus therein whether it be diploid or polyploid. We have been able to discover that some of the information useful for the classification of plants is derived from their chemistry. Some of the chemical products, some of the metabolic products in plants do not modify greatly upon ddesiccation, for example the tannins have been determined to be quite useful evidences for what we now speak of as chemotaxonomy. We can discover from an herbarium specimen 100 years old what sort of tannin that particular plant had in it, and compare that with plants that are as fresh as you can get them. So in a real sense, an herbarium specimen will give us data and we are limited only by our own ingenuity as to how we may extract that

Digitized by the Hunt Institute for Botanical Documentation

information. We have discovered over the last few years that more, and more, and more pieces of useful information for the purposes of classification can be derived from herbarium materials. Furthermore, let us consider another aspect of the herbarium specium. It is a very fine document against which you may check the results of your analysis. So this technique that is essentially a preservation technique used one way by a plant biologist and another way by an animal biologist, lets us test our methodologies in a most precise way, so we have here a tool, a technique for really determining how good a classification we can derive with a computer. I find that this is a marvelous material, methodology developed by people who did not indeed have such things as this type of thing in mind at all. But do not ever discount the value of herbarium materials. It is a real crime to me that any biologist will work and not document the materials that he has studied, whether he be physiologist, geneticist, cytologist, or what, if he but takes the materials which he has been studying and goes with the little extra bother that is required to dry or preserve it in some way, label it, and tell what he did with it, he will have something which anyone in the future may go back and check. And so I say to a physiologist, if you do not tell me the plant materials upon which you worked, I can never discover whether or not you are right or wrong in what you discovered about those plants. I think that we have, as taxonomists, we have sold this particular thing short, this business of herbarium specimens for their value that they maintain for all of us. Many of you will frustrated in the next few days, I am sure, when you look at the documentation that goes along with these herbarium specimens. Why did that collector not putt down all the little nice bits of information that I would like to have had associated with this particular plant? Why did he not say precisely what sort of habitat this plant occupied? Why did he not add something about a precise locality, so that I may go back to that locality and discover whether that particular specimen could still be living there of that particular species. But again

Digitized by the Hunt Institute for Botanical Documentation

recall the historical development of this thing. In Linnaeus' time they were already to begin to make explorations in the New World, but to Linnaeus the New World was just out there somewhere, and it really didn't make much of a difference to him that he described a particular plant as a Polypodium virginianum when indeed if we look at Virginia today there is no such plant as Polypodium virginianum in Virginia but it came perhaps from Canada. But Virginia described to Linnaeus/all he needed to know about the New World. So let us keep these sorts of considerations in mind as you look at these old and battered documents that we have. The types ;of data that one finds upon labels may be that which is associated with the plant but doesn't necessarily come from the plant itself, its ecology, its geography, its native uses may be recorded, and indeed this has turned out to be a very useful sort of recording on the labels of herbarium specimens. Many people have gotten leads as to what sorts of plants might contain some goody for some medicine or other, but the labels may also give some information about the plant itself. It may be,for example, flower color, because as we all know flower ~~xxxx~~ pigments deteriorate rapidly upon drying, and so if he gives you that information, you have ~~xxx~~ gained more than you would otherwise have had. This specimen which represents a fragment of the total perhaps, for example, the specimens of Quercus alba mentioned earlier are only tiny little fragments right up at the tip of one branch of one tree. But perhaps the label may give you something about the girth of the trunk, something about the total height of the ~~xx~~ plant, something about, it may give you something about the branching patterns which that particular plant have. So we can gain both intrinsic and extrinsic types of information from the labelled data. Now that we have more of an idea about what an herbarium specimen should do for us, perhaps we can devise more precise methods ~~xxxx~~ of ~~xxxx~~ collecting these specimens in such a way as to include such types of information as will be useful in the future. I suggest to you, however, that no matter how precisely you may define your methodology

Digitized by the Hunt Institute for Botanical Documentation

for collecting of such specimens, that what you may consider as valuable today may or may not be at all valuable to the generation hence or 10 generations hence. But at least we can ~~xxtxintoxxxxx~~ begin to more precisely define what ~~is~~ a specimen must do for us. I feel all this sort of thing is a very important backup before we begin to actually look at herbarium specimens as we will do in this study here.

We will proceed in the next few weeks with a real problem in taxonomy. It has not been precisely done the way we will do it before. Looking at your flowcharts you will discover that we have done some of the decision-making that you might do upon your own investigations later. We have discovered a ~~crop~~ problem already. We have to shorten it up so ~~xx~~ that you can actually go to work on ~~this~~ problem, assemble the organisms which you are to study. I have writtenn to several herbaria in this country to ~~xxix~~ collect from them ~~xxpxxxxxxx~~ specimens representing the genus Uvularia. ~~Somewhere or other I believe that word~~ is written down, ~~ifxxxt~~ is it not? ~~On you charts. No it is not, yes it is on April 15 it says examined plants Uvularia.~~ Now let me tell you what we have here so that you will have some idea. I'll give you an idea of the classification that is presently given for the genus Uvularia. It is an angiosperm; it is a monocot; it belongs to the order Liliales or there are some other terms some synonyms for this, sometimes for instance you can find for this order Liliflori but we will call it Liliales. It belongs to the family Liliaceae. ~~I forget the subfamily. You can look it up if you want. To which subfamily xxxxxx Liliaceae belongs.~~ And genus Uvularia. I am remiss for not looking up what the word Uvularia means as a Latin term. Perhaps you can find this for yourself. (Some class discussion on the possible meaning of Uvularia). (Dave) You can see the level of our classical education here all of us.

The genus which we will study is endemic to North America. You know what the word ~~xxxxxx~~ endemic means? It means it don't grow naturally no ~~xxxx~~ where

Digitized by the Hunt Institute for Botanical Documentation

else. Mostly the species are from the eastern part of North America, from southern Canada south to Florida. It goes to the West I believe about as far as North Dakota and on the western side and in the south it goes about as far west as Oklahoma to my knowledge, there are no Uvlaria species in Colorado nor into the west from here. I won't tell you how many species you are going to have that is for you to determine. I will at least tell you that the study that we will be working on is that of the size of a small genus.

Digitized by the Hunt Institute for Botanical Documentation

B 170
Lecture
April 15, 1966

CHARACTERS

I wonder how many of you know where we are now looking back at that general flowchart of taxonomic processes. Are you using that flowchart to sort of keep yourself on track, where your road is. You should, indeed, keep this sort of thing in front of you now and again, because if we don't come and tell you what we are doing at any particular stage of the game, there you can look to see approximately what is coming off. To be sure that you and I concur in where we are, I will point to the places on here that we are beginning to discuss now. Under the detailed flowchart under Roman numeral II Computer Analysis, we have started, and last week we began to talk about the problems dealing with area IIb. Now I listened to part of a tape that was made on Wednesday's lecture, and I thought maybe I had a few more little general comments to make, some of the things that I thought needed to be said.

Needless to say, there are many things to be said, let me just say what I think needs to be said. We were, I believe, beginning to discuss the problem of characters, the general problems of characters, and it probably would be wise to discover whether we can define characters, and if we get such a definition perhaps then some of the more complicated problems will be more easily answered. Problems like the definitions of characters have bothered taxonomists for a long time, and I think probably the reason they have had difficulty in the definition of characters, for the same reason that we have had definitions of trying to do anything with taxonomy in a theoretical sense before. That is, they haven't clearly defined or differentiated between their thinking processes about plants and the data they derive from the plants. George's definition that has come to light in the various pieces of our work for a character is as follows: A

Digitized by the Hunt Institute for Botanical Documentation

character is a real value function defined over all the objects in the study. If we look at that definition in its various parts, we can see that definitions like this eliminate many of the arguments that taxonomists have had such as a character in one group may or may not be a character in another group. This is a difficulty, I know that this is a difficulty because I was speaking Wednesday to one of the big stick taxonomists in the country, and he clearly was not thinking in the same way that I was thinking about the definition of character. He wass trying to see all the data that he recalled out there and trying to find some wayy to encompass the definition for all of the stuff that he ran into. If we ;use the definition that we have just said above, we eliminate the arguments about, well, how big a size of group are you talking about. Whether you are talking about a specific character or generic character. Note the definition that I gave, said that it must be defined over all the objects in the study and clearly this is the best way to give you some sense of a definition. Now to amplify a little bit on this and let me get a few points across before you interrupt, please, in another sense, and amplification of this definition we can think of characters as patterns, and obviously patterns vary from group to group, but they are indeed still patterns. I shall not get into the problems of psychology here, I think this is a bit a far afield, but we can think also of characters in the sense of the psychology of learning. In this sense a character is some sort of mnemonic device, and I think you all are familiar with the word mnemonic. In our sense we could think of a character as something of a memory aid which is another way of defining mnemonic. Also in this sense, it is learning by association, we discover, for example, that some patterns are related. Now if wee go back to the basic determinators of plant and animal form and function, that iss as far as we know DNA and RNA are the basic determinators of form and function inn plants and animals, we discover that there are patterns or perhaps we may prefer to

Digitized by the Hunt Institute for Botanical Documentation

call them sequences of subunits nucleotides which eventually determine the types of material which an individual plant will have. For example, there is no haphazard throwing of things together; rather the DNA determines rather precisely the procedures for putting materials together. We will get from any particular quantity of DNA-RNA content in a particular plant or animal only that which was predetermined there anyway. You can get, for example, only a western yellow pine out of certain patters or sets of DNA, or you can get only a human being from a particular set. So again , from the very basic materials in which we have to deal we get some sort of patterns.

Now in taxonomy, we use characters as the basic input data for processes of classification. And again, remember this. Always keep in mind what your purpose is, what characters are for. Now a character in our sense, and I'll give an example a little bit later, contains only a small finite number of states which we most frequently refer to as attributes, that is, that is our terminologyy for character states or the attributes of the character. For our purposes, an attribute defines a particular condition of a character. And again for our purposes, for our purposes here, attributes are mutually exclusive. Now perhaps it is easier for you to get an idea of a character as we use it in taxonomy by giving the example. For example, a very simple one that you use in botany will be that which we will call leaf shape. Note here that this is in a real sense a mnemonic device, a memory aid for talking about particular pieces of information. It focuses your attention down such that you are saying, "All right I will not talk about leaf shape, leaf shape alone." In that sense it is a mnemonic character. Now we have to remember our original definition here of a character: it is a real value function defined over all the objects in the study. There are certain cases where such a character as leaf shape would not apply, not be useful, but remember our conditions here. And in this sense, we get some ideas and feedback on how to make such definitions from what George was telling you the other day. We have to make definitions, these are the arbitrary points

Digitized by the Hunt Institute for Botanical Documentation

upon which we must base such definitions, defined over the study, in any case it is defined over the study. If we are talking about plants without leaves, we certainly wouldn't use a character leaf shape, and so we would again have to confine ourselves to those particular situations. If our particular group does have leaves, then you have shape. After examination of materials in our study, it may be possible for us to assign a small finite number of attributes which specifically tell us what leaf shape a particular object in our study may be defined as. And so you can list your attributes for the study as being one condition of leaves ovate or leaves linear and one ~~that~~ will continue to add attributes until all of the objects in the study were able to be fitted into one of these attributes. And we may discover that in some cases the plants in our study may have combinations such that you would have to make these attributes to accommodate the combinations. Suppose, for example, in our study ~~see~~ for the character leaf shape we discover that some of our plants had both ovate and linear leaves upon the same organism. Another condition that we would like to use in our methodology is that our attributes should be mutually exclusive. If we wish to continue to use this definition of mutually exclusive and we discover for the character leaf shape that some of the objects in our study have both, we must add another state in order to allow for that mutually exclusive situation and the condition can be merely described as ovate-linear / ~~another~~ in which we can assign to some member of our study this condition here. Some other members very often will fit in either of these two ~~characteristic~~ of the first categories. Now with these sets of statements that I have just made here, I am going to stop this generality for characters because we will come back to it again Monday.

Now I am willing to listen to the two questions that you had raised here. Can you recall them, or ~~are~~ have they gone by? (Student) You answered one. (Dave) This is usually the case. If one will wait a moment one will hear something ~~or~~ else that says it. (Student) Is the function that you were talking

Digitized by the Hunt Institute for Botanical Documentation

about a mathematical function? (Dave) Not sense. (George) The function that Dr. Rogers was mentioning is the same function that I defined for you last time, that is the same idea is involved, its correspondence between two collections of things. The first collection of things being the plants in your study, and the second collection of things being a collection of descriptions which describe the shape of a leaf. And the function is two sets together with a rule which establishes a correspondence between members of the first set and members of the second set. The correspondence is to each plant in your study under consideration some correspondence made between that plant and some description in this collection of descriptions which tells you about the leaf shape. So that this whole describing process can be thought of as a function, a character *per se* to be thought of as a function some means by which we correspond with each member of the study some description of some particular interesting basis for comparison for these particular objects one to the other. (Student) Do you need a description for each of these? (George) That is what is meant sort of by saying that these things are mutually exclusive. In order that this be a real function, you say to each object in the first set we must be assigning one and only one object of the second set. So that for each plant we are allowed only to assign one description, so we are not allowed to assign more than one description. But by means of this rule of this correspondence we are going to associate with each plant some state which describes the shape of its leaves. (Student) Then how could this be defined over all the objects in a study? (George) It has got to be. A function is defined over a set, if the set over which we stated it is defined is the same as its domain for the set of the first of the two. If you look in the handouts that I gave you where it defines function, it talks about two sets together with a rule. Where each member of the first set is put in correspondence with some member of the second set. The first of these two sets is the set over which the function is said to be defined. That is no matter which object in the first set you choose, the rule

Digitized by the Hunt Institute for Botanical Documentation

part of this function enables you to put it in correspondence with something in the second set. And that is what Dr. Rogers means when he says the characters defined over all the objects in the study. For each object in a study my character enables me to associate with that object some particular description from the set of descriptions which place the role of the second set in that definition of function. Are you getting the picture? (Dave) Now you know what a character is. (Student) Mathematically, yes. (Dave) Theoretical. (Student) I have another question. started out by saying that maybe characters can be defined or it seemed to me that ;you were wondering maybe if they can or if they can't. There seems some question about whether they can be defined. Now does this mean that there is some question about whether or not we can think up a good definition of what a character ought to be or if when we look at the plants they are, the characters that we have already thought up are poorly defined in reality? (Dave) The best answer I can give to that is just exactly what I was trying to separate there. The way we think about these things and the data that is out there, and in the past the great difficulty that people have had because they didn't clearly differentiate between these two things, you see. Because they could never find an active description which would take in all of that data as long as they had to include all the exceptions and so forth. (George) It is always entirely possible to define a function in existence to say that I am going to describe plants according to this set of rules. These are my collection of descriptions and this is the role by means of which I am going to establish a correspondence. This is always possible and nobody can contend for a minute that this isn't possible. How useful this particular character that you have defined into existence is going to be for classifying your plants is another question. (Henry) I think perhaps maybe what you were aiming at, the answer that you wanted back is: the character is something which contains comparable properties of the objects that you wish to describe. In otherwo | rds the information is comparable

Digitized by the Hunt Institute for Botanical Documentation

information for particular objects under study. (Dave) In order to allow us to proceed with some work that I want you all to do individually, I have to give you some basic thoughts and processes on herbarium materials. The collections of materials kept by botanists in pigeonholes in cases similar to this are the same thing as a zoologist calls museum specimens, or in our case we call them herbarium specimens. In the last few years or a matter of fact for a very long time, there has been a rather derogatory opinion of anyone who meddled around with scrappy little hunks of junk, squashed out flat, dried and turned brown and plastered down to a sheet of material of paper. I think perhaps that I would like give a little bit of different emphasis on this process here for benefit for those of you who have heard these arguments have concurred with those who said we ought to stick a match to every herbarium in the world and burn the ~~xxx~~ dang thing down. We have to recall that there iss a long history in botany and that as our information about plants grew we changed our concepts about we needed to do at any particular stage of the game. We haven't granted been too precise in definition of what an herbarium ~~ix~~ specimen is supposed to be and do for us. But again, at any era at any particular time these types of things have changed. What do we have to do with it? Personally, I am one of the greatest defenders of a collection of dried herbarium materials. I think this is one of the most marvelous scientific pieces of documents that has ever been invented. Having come from an institution where they figured that ~~kk~~ they had about three million separate sheets of herbarium material, I was always very fascinated that if I ~~need~~ will two names, the genus and the species name, that I could within 5 minutes go to a particular spot in that collection of three million specimens and discover an object, a particularr object out of all that three million and with that very rapid way of getting to a particular piece of information out of this mass of information, I could

Digitized by the Hunt Institute for Botanical Documentation

discover all kinds of interesting things. Things that interested me perhaps at that particular time. I want to know, for example, what does Quercus alba look like? And with those two words, go and look into that tremendous collection of things and find out something that looked like Quercus alba, or a thing which people call Quercus alba. And I could discover what the thing looked like; I could discover perhaps where the plants representing Quercus alba grew, and all sorts of little goodies and informatinn about them. Now another way of thinking about herbarium specimens is this: that they are a fine set of historical documents. Many of the early collections in this country will bear the label Boston, Massachusetts, and I suspect that if we went back to Boston, Massachusetts right now, we would discover that the place that the man picked up this particular specimen is now completely covered with concrete and steel and yet we have some evidence that this was indeed a plant which did grow in that area before man came along and messed around with the environment. So in a sense it is a, it can be used as a fine record of what we have had in the past, although it is an incomplete record. I also want to give you some feeling about these particular specimens because in your handling of them in the next few weeks, I want you to have the same sense of importance that these particular specimens have such that you will treat them with the respect to which they, you should give them. Now how do we use herbarium specimens? We use them again according to the dictates of the time in which we exist. Primarily, we derive information about the plant materials from herbarium specimens concerning their gross external morphology. But as we have gone along improving our data collection methods and our refinements in what we want to know about plants, we have been able to use these old brown herbarium sheets in many different interesting sorts of ways. For example, although the specimen be dried, it still maintains quantities of cellulosic material and allied substances which do not modify ~~thx~~ tremendously upon drying. For example, you can study quite adequately the anatomical features of certain parts of the plants. If you are ingenuous

Digitized by the Hunt Institute for Botanical Documentation

enough to know how to rehydrate these specimens and go from there, we can discover that we can get some rather indirect evidence sometimes even about the genetics of these plants. For example, we know that cell size is sometimes influenced by the number of chromosomes contained in the cell. That on the average, cells that have a doubled number of chromosomes, in other words the plants are polyploid, the cells will usually be bigger. People have recently been taking some of this plastic material, liquid upon application and dried rapidly, applying that to the surface of leaves on herbarium specimens, peeling off the dried plastic and discovering that they have a nice imprint of the cell sizes, which they can adequately measure and compare to see whether or not there is a range in the variation in the cell sizes of those plants, and they can discover whether the particular group of plants that they are working on are polyploid or diploid. Again, the evidence is indirect, and one would certainly want to check it, but a sufficient number of checks have been made such that they can indeed discover the evidence of ~~pim~~ polyploidy from an herbarium specimen. There are other ways to do the same thing, sometimes, it doesn't work in all cases but it sometimes will. From an herbarium specimen , one may extract pollen grains, and pollen grains are again are direct evidences of the contents of the nucleus therein whether it be diploid or polyploid. We have ~~ah~~ been able to discover that some of the information useful for the classification of plants is derived from their chemistry. Some of the chemical products, some of the metabolic products in plants do not modify greatly upon ddesiccation, for example the tannins have been determined to be quite useful evidences for what we now speak of as chemotaxonomy. We can discover from an herbarium specimen 100 years old what sort of tannin that particular plant had in it, and compare that with plants that are as fresh as you can get them. So in a real sense, an herbarium specimen will give us data and we are limited only by our own ingenuity as to how we may extract that

Digitized by the Hunt Institute for Botanical Documentation

information. We have discovered over the last few years that more, and more, and more pieces of useful information for the purposes of classification can be derived from herbarium materials. Furthermore, let us consider another aspect of the herbarium specium. It is a very fine document against which you may check the results of your analysis. So this technique that is essentially a preservation technique used one way by a plant biologist and another way by an animal biologist, lets us test our methodologies in a most precise way, so we have here a tool, a technique for really determining how good a classification we can derive with a computer. I find that this is a marvelous material, methodology developed by people who did not indeed have such things as this type of thing in mind at all. But do not ever discount the value of herbarium materials. It is a real crime to me that any biologist will work and not document the materials that he has studied, whether he be physiologist, geneticist, cytologist, or what, if he but takes the materials which he has been studying and goes with the little extra bother that is required to dry or preserve it in some way, label it, and tell what he did with it, he will have something which anyone in the future may go back and check. And so I say to a physiologist, if you do not tell me the plant materials upon which you worked, I can never discover whether or not you are right or wrong in what you discovered about those plants. I think that we have, as taxonomists, we have sold this particular thing short, this business of herbarium specimens for their value that they maintain for all of us. Many of you will frustrated in the next few days, I am sure, when you look at the documentation that goes along with these herbarium specimens. Why did that collector not putt down all the little nice bits of information that I would like to have had associated with this particular plant? Why did he not say precisely what sort of habitat this plant occupied? Why did he not add something about a precise locality, so that I may go back to that locality and discover whether that particular specimen could still be living there of that particular species. But again

Digitized by the Hunt Institute for Botanical Documentation

recall the historical development of this thing. In Linnaeus' time they were already to begin to make explorations in the New World, but to Linnaeus the New World was just out there somewhere, and it really didn't make much of a difference to him that he described a particular plant as a Polypodium virginianum when indeed if we look at Virginia today there is no such plant as Polypodium virginianum in Virginia but it came perhaps from Canada. But Virginia described to Linnaeus/all he needed to know about the New World. So let us keep these sorts of considerations in mind as you look at these old and battered documents that we have. The types ;of data that one finds upon labels may be that which is associated with the plant but doesn't necessarily come from the plant itself, its ecology, its geography, its native uses may be recorded, and indeed this has turned out to be a very useful sort of recording on the labels of herbarium specimens. Many people have gotten leads as to what sorts of plants might contain some goody for some medicine or other, but the labels may also give some information about the plant itself. It may be,for example, flower color, because as we all know flower ~~xxxx~~ pigments deteriorate rapidly upon drying, and so if he gives you that information, you have ~~xxx~~ gained more than you would otherwise have had. This specimen which represents a fragment of the total perhaps, for example, the specimens of Quercus alba mentioned earlier are only tiny little fragments right up at the tip of one branch of one tree. But perhaps the label may give you something about the girth of the trunk, something about the total height of the ~~xx~~ plant, something about, it may give you something about the branching patterns which that particular plant have. So we can gain both intrinsic and extrinsic types of information from the labelled data. Now that we have more of an idea about what an herbarium specimen should do for us, perhaps we can devise more precise methods ~~xxxx~~ of ~~xxxx~~ collecting these specimens in such a way as to include such types of information as will be useful in the future. I suggest to you, however, that no matter how precisely you may define your methodology

Digitized by the Hunt Institute for Botanical Documentation

for collecting of such specimens, that what you may consider as valuable today may or may not be at all valuable to the generation hence or 10 generations hence. But at least we can ~~xxxxxxxxxxx~~ begin to more precisely define what ~~is~~ a specimen must do for us. I feel all this sort of thing is a very important backup before we begin to actually look at herbarium specimens as we will do in this study here.

We will proceed in the next few weeks with a real problem in taxonomy. It has not been precisely done the way we will do it before. Looking at your flowcharts you will discover that we have done some of the decision-making that you might do upon your own investigations later. We have discovered a ~~crop~~ problem already. We have to shorten it up so ~~xx~~ that you can actually go to work on this problem, assemble the organisms which you are to study. I have writtenn to several herbaria in this country to ~~xxix~~ collect from them ~~xxxxxxxxxx~~ specimens representing the genus Uvlaria. ~~Somewhere or other I believe that word is written down, ifxxxx is it not? On you charts. No it is not, yes it is on April 15 it says examined plants Uvlaria.~~ Now let me tell you what we have here so that you will have some idea. I'll give you an idea of the classification that is presently given for the genus Uvlaria. It is an angiosperm; it is a monocot; it belongs to the order Liliales or there are some other terms sometimes for this, sometimes for instance you can find for this order Liliflori but we will call it Liliales. It belongs to the family Liliaceae, ~~I forget the subfamily. You can look it up if you want. To which subfamily xxxxxx Liliaceae belongs.~~ And genus Uvlaria. I am remiss for not looking up what the word Uvlaria means as a Latin term. Perhaps you can find this for yourself. (Some class discussion on the possible meaning of Uvlaria). (Dave) You can see the level of our classical education here all of us.

41.4

The genus which we will study is endemic to North America. You know what the word ~~xxxxxx~~ endemic means? It means it don't grow naturally no ~~xxxx~~ where

Digitized by the Hunt Institute for Botanical Documentation

else. Mostly the species are from the eastern part of North America, from southern Canada south to Florida. It goes to the West I believe about as far as North Dakota and on the western side and in the south it goes about as far west as Oklahoma to my knowledge, there are no Uvlaria species in Colorado nor into the west from here. I won't tell you how many species you are going to have that is for you to determine. I will at least tell you that the study that we will be working on is that of the size of a small genus.

Digitized by the Hunt Institute for Botanical Documentation

B 170 Lect. Apr. 18 Characters + coding.

1. No subst for specialistic knowledge, but a systematizing of effort -
   Can we make the process more efficient
   In this study have to stick to gross morphology
2. Kinds of characters, once a study has been circumscribed
   A. Qualitative
   B. Quantitative
3. Qualitative - a non-numerical descriptive terminology.
   A. Shapes
   B. Coverings (pubsc.)
   C. Colors (if any)
   D. Over all sizes (mixt of qual + quant).
   E. Positions (flowering branches, etc.) also mixt.
4. Quantitative - a measurable item
   A. Numbers of parts (as in flower parts)
   B. Dimensions in general - Lengths + widths,
      Can also be qualitative
      Continue with ratios - as in leaves -
      same ratio for an obovate as well as ovate.
      Angles of divergence (base of leaf blade, etc.
5. How to break up characters into attributes
   A. No. of attributes / charact.
   B. Keep in mind the purpose of chars + attributes
   C. To find some means of establishing -
      Use ~~A~~ bar graphs or other scheme to give →

Digitized by the Hunt Institute for Botanical Documentation

5 C. hints.
D. Essentially, do a classification with each character.

~~6. Ordering of Attributes~~
When?

6. Missing information.
1. Poor specimens.
2. Plants in flower or fruit.

7. Ordering of attributes.

Digitized by the Hunt Institute for Botanical Documentation

B 170 Lecture
Dr. David Rogers
April 18, 1966

CHARACTERS AND CODING

I trust you may have already assimilated all the discussion given in Chapter IV of Davis and Heywood. I think Henry assigned this to you last week on Wednesday, and if you have not, be aware of the fact that it is information that you will find useful to you, and that we will not ourselves really go into what is said, maybe we will reorganize things a little bit but that is about it.

It may occur to you that we are trying to make specialists of you for a particular group of plants, namely, Uvularia. You may also have the uneasy feeling, well gee this is a group of plants that I don't know anything about at all, and before I can adequately study them, I really should spend more time with the plant material. Be assured that our efforts are not to make specialists out of you or expect specialists at the end of this course on this particular group of plants. We know that there is a considerable amount of effort needed by a specialist before he can actually do a thorough job of classification of a group of organisms. He has to have much more familiarity than you will be able to gain in this particular group. What we are really attempting to do is to discover whether we can, and this is the experimental part of this course, actually systematize or make more efficient if you will, the processes than one may, one has to go through in order to get a classification. When you are thinking about this course, it will be very helpful to us if you will indeed consider such problems as whether or not the presentation and ideas that we are giving aid in this process or do not. Then if not aided then why not, this sort of thing interests us to try to get something from you that will give us a better course next year. You can come back and take it then maybe.

Digitized by the Hunt Institute for Botanical Documentation

We all know, and I think we have said this several times already, that really the most difficult problem that is involved with the taxonomic process is the discovery and use of characters. And really all we can do is present some relative generalities here. We can't precisely determine from plant group to plant group what procedure will most adequately apply. If we give you these guide posts, perhaps it will give you a little bit more direction to the method of discovery of what information you want and need for the process of classification. In the study of the organisms that you are working on, we obviously cannot delve deeply into some of their more cryptic characters, if I can use the word cryptic here in the sense that they are characters that would require the use of certain kinds of equipment to discover them, anatomical features or chemical analysis, and so forth, are just beyond us as far as time is concerned. So we stick here to gross morphology, it being understood, however, that some of the same techniques should apply when you are forming information about the organisms in another method of derivation of information, thin layer chromatography as well as the gas chromatography methods as well as fine structure, anatomy, etc. So don't be misled to think that we emphasize gross morphology just because that is a time-honored way of doing things. It is just an efficient way to get into the process, an easier way so to speak. We always, of course, must re-emphasize the business that you have to keep in mind the purpose of a particular piece of information when you are using it as a character. We have already defined characters, and incidentally I made a goof last lecture period in saying that a character is a real value function. George tells me that the word real has all kinds of overtones and implications that I don't mean for it to have here. And a better term is description valued function for esoteric reasons made much more clear by George than by me. Now getting down to the business of characters for our study here, we can divide characters into two broad categories. These categories are

Digitized by the Hunt Institute for Botanical Documentation

not at all exclusive one to the other, and you can merely use this as a sort of guidepost. The characters are either qualitative or quantitative. I'll tell you what I mean by the fact that these things are not mutually exclusive in just a minute. Let us consider first qualitative characters. A qualitative character is some non-numerical descriptive term, and it is the most common way in which we differentiate things, people, objects. Certainly one considers as a qualitative character the shape of an object and just look through the English language and see the richness that we have in adjectives describing shapes. It is an interesting thing to me that this particular aspect of the language is extremely well developed, but it also reflects our need in some essence to communicate about things. One of the anguishing problems that the taxonomist, particularly the botanical taxonomist, have worried about over and over again, how does one describe the shapes of leaves, and of course this can be a very difficult one to discover, just exactly what you mean when you say obovate, ovate, obcordate, cordate, etc. These terms can have definitions all the way across the board. And the definition is, of course, strictly something that derives from a need in the individual. What I may mean by ovate and obovate, may fit into other terms by you. You might for example suggest that an ovate leaf might be broadly elliptic, for example. And it, therefore, behooves you if you use a particular piece of terminology, to discover if you can just exactly what you are saying so that the person to whom you are communicating your information is best informed about that sort of thing. We certainly come to terminology after we have discovered differences in shapes. I for a long time, for example, in working with the *Manihot esculenta* used for the description of the central lobe of the palmate leaf below the leaf linear and obovate as two major categories and then because I was not sure what term was adequate to be used, that is acceptable amongst taxonomists, I called the third state of the thing Mae West, which

Digitized by the Hunt Institute for Botanical Documentation

I think does a pretty good job for this generation. It may not be very useful for the next generation of botanists, but most of us are aware of what we are talking about there. It turns out that this is more adequately defined as pandurate. I still get more of a charge when I think of Mae West than I do pandurate.

Other qualitative types of information are say the coverings or the surface features of organisms. And there is such a multiplicity of types of things that we can only suggest that you go to the literature to discover how many things can be used as descriptive terms for that which is on the surface of objects. One of the things, of course, that comes to mind is the hairy materials. And these, of course, the descriptive terminology is very rich here depending upon which group of organisms that one happens to find such conditions extant in. Again, the early investigations on any particular group of objects will not give you precise terminology unless you have practiced considerable amounts in this field. But you still do not find yourself lacking in the ability to discriminate one surface feature from another even though you not use the proper terminology at the time. Certainly, the colors of things are qualitative pieces of information. We can indeed divide colors by their content of pigments, but how much information have we gained by going to the extra job of collecting or analyzing for the pigments present. However, be you a pigment chemist, it behooves you to discover what kind of pigments there are and how many there are per certain feature. We can use as a qualitative feature the overall sizes of things. It can be big leaves and little leaves. It merely requires that you define what you mean by the smallest big leaf and the biggest small leaf, and you can use their, for the objects that you have under study, some sort of a dimentional limit, up to 10 cm is little. This same sort of thing can apply to the totality of the organism, it is a big tree or it is a small shrub, or it is an herb. Again, these are

Digitized by the Hunt Institute for Botanical Documentation

dimensional sorts of things which you dang well write down. When I say herb or small shrub, how big is a small shrub? And it really doesn't gain you anything to make an overall definition for all the shrubs that appear on the face of the earth. It merely says that you better limit your definition for the group you have under study. However, the terms are useful, very useful immediately in getting your listener or the person to whom you are communicating into the general framework of what you are trying to talk about. It is more adequately to talk about plants as shrubs or trees or something of this nature, as long as you define what you mean. Positions of things on objects can also be qualitatively defined but again with some sort of a quantitative evidence. For example, you could talk about the position of the flowering branch as in the upper portion of the plant, or that branches occur on the base and sequentially upward. Again, this gives an overall word picture of the information, and again probably needs to be defined by some sort of a measuring system. Branching of plants, for example, is a good way to differentiate among certain kinds of them. So I can't define exactly how you must combine the qualitative and quantitative aspects of this. It again is a process ~~that~~ which we could be more definitive about, but I don't think that at this stage in the game we dare be so. Quantitative characters on the other hand are essentially measurable items. In plant taxonomy the number of flowering parts has come in for heavy use as key characters. These are things that Linnaeus used, for example, when he set up his classification system. He counted the numbers of stamins, and all plants that had flowers with certain numbers of stamins got added to the same group. And although he had some odd plants put together in the same grouping, ~~there~~ it was quite evident that if you did indeed count the number of stamins, you could find that particular organism quite nicely using his system.--Numbers of petals, numbers of seeds, numbers of locules, etc. have been great, heavily used, and indeed very useful types of

Digitized by the Hunt Institute for Botanical Documentation

information. Dimensions of plants again, going back to our qualitative as quantitative pieces of information, but we wish to be a bit more precise perhaps for feeling that such precision is the use of the celluoid ruler gives us more accuracy than we have when we say it is a three thumb nail length long, but again is a very useful sort of indicator for those who would like to know about the organisms that you are talking about. One caution I would make with the use of numbers, dimensions, and so forth is that you watch very carefully when you start talking about ratios of things. When they feel, and this is certainly a thing that many taxonomists do over and over again and it has been done in our shop, something is better about a ratio than just an ordinary measurement, one measurement. One taxonomists recently said, or one statistician recently said give a biologist two numbers and he will make you a ratio. It is some sort of response. I don't know exactly what it is. Ratios can obviously be useful if we want to recognize the limitations on them. But in our case, let us consider another problem, that if one measures, for example, a character of the length of the leaf and the width of the leaf, and he also decides to add a character on the ratio of these same measurements, he spent more time and energy and will cause a little bit of extra attention to be given to that particular part, then perhaps it is worthy of its doing so. So that if you use your ratio use or if you prefer, stick to the dimensions. I think that perhaps that I would rather leave ratios alone having had our fingers burned on Manihot in general, Whole series were used which we really could not defend as adequate characters. There are places, of course, where certain numbers are more advantageous. For example, if you wish to adequately describe a differentiating character of the part of insertion where the leaf blade joins the petiole perhaps an angular measurement gives you some more precise definition than to say large angle or small angle, as long as you can adequately and always measure such

Digitized by the Hunt Institute for Botanical Documentation

angles. Up to this point we have been talking in general about characters per se. We said last time that characters are broken up for our purposes into mutually exclusive attributes. Again,emphasizing the fact that we have got to keep in mind what the hect the characters and attributes are to do for us, we can proceed to discover some way to divide a character of leaf shape or the character of the dimension of the leaf into some satisfying set of mutually exclusive attributes. How do you know how many states are attributes adequately reflect your thinking about a particular character? How do you decide that a particular piece of information should be broken into three attributes or four, or five, or six? Again, I wish I had some sort of rule of thumb which would tell you how to know when you have enough attributes. One technique that I use sometimes, I don't defend it as being a very efficient system as merely to make yourself a sort of there is a frequency of occurrence of a particular measurement, for example, and you may find that some measurement here does not adequately break on the objects under study. You may have a nice smooth curve. However, it is frequently the case that one of these categories will be missing, and I say, xk just by looking at this thing there are at least two attributes to this dimensional character. Of course one always worries, has one taken an adequate sample? But with such things as classification processes, you knn never will have an adequate sample. There just ain't such a critter, maybe. Although I think in many cases we can discern that this is not necessarily a bothersome problem. Why not divide these into separate states? I don't know way to tell you which way is the best way to do that. Maybe George can give us a little bit more ideas on this sort of subject. I hope he will sometime. But as far as the biology is concerned, it is a question really what does he anticipate that this particular piece of information will do for him? Will a large number of states divide his group of organisms into too many different types of things than he feels adequately reflects whatever he is going to call his groupings? It is a decision

Digitized by the Hunt Institute for Botanical Documentation

process that cannot be really defined adequately except for the group of organisms that you have there. This, of oourse, is a very subjective sort of judgment. I think I have already said to you that as far as I am concerned there is nothing wrong with subjectivity, and I don't care what we attempt, there is no substitute for the biologist who is doing the job and his knowledge again, of course, is critical in this particular operation. How much do you feel that you really should divide a particular character into? Or you can sort of play a game here. I think this is the thing that Adanson had in mind a long time back. You classify the objects under study one character at a time, and don't worry about any other particular contributing or disturbing factor at the moment that you are playing this particular game. This, of course, is quite a difficult process when you have numbers of features of organisms all there, all staring you in the face simultaneously to try to keep your mind focused on one particular piece of information and not let the influence of something else come in and say, "Well, maybe this would make it do this." But I think that perhaps playinn the game with each character individually is perhaps the most adequate way in which an individual can decide for himself how finely he wishes to break a character up. And, of course, trying to....the first one is easy, the second one is not so hard, but by the time you have done 10 or 12 or 15 characters, those last few are extremely difficult to keep that game because you are indeed influenced considerably in your thinking process by that which you have already done. Again, you have to be quite careful that earlier influences of earlier game-playing with characters are not helping you too much with your decision about the latter characters. There is really nothing wrong I think, however, with our sort of bootstrapping your information up as you go along. Perhaps your first characters will have to be reorganized after you have kept looking for other pieces of information. I think any taxonomist who is worth his salt no matter

Digitized by the Hunt Institute for Botanical Documentation

whether he was an Adansonian taxonomist or a non-Adansonian taxonomist as the majority seem to think they would, indeed were playing this sort of game when they made a classification, that is if you really read a good classical taxonomist output and look at it you will discover that in his species descriptions he mentions things that probably are all sort of working together simultaneously. Many pieces of information are playing role even though he may claim otherwise. So I think one has to look very carefully at classification schemes and discover that indeed thay have been using a multiple number of characters whether or not they wished they were or not. I think that Davis and Heywood addressed this point more adequately than I have, and if you haven't read it, you really should.

Now with these bits of precise definitions that I have given you so far this morning on how to divide characters into attributes, there are some other things that perhaps you would be interested in knowing about on these particular plants that you will be judging. I am sure that if you have examined these plants at all, you will have discovered that there specimens have either flowers or fruit on them. That is you seldom find a specimen which has seldom, never I think, of course if I say that, immediately I am wrong-- flowers on one stem are one specimen, fruit on another. Obviously, in order to describe a group of organisms we have to describe the flowering condition as well as the fruit condition. So we have to worry about specimens where there are nothing but fruits, and the characters that you should be coding talk about flowers. This follows in general under a category of general missing information. Obviously, it is a thing that we can't do anything about. We have to worry about how to most adequately allow characters that can't be studied to be included but not be damaged by making statements that look as though we had something about the information [redacted] on the fruits when the fruits are not there. Now there is a whole host of considerations that we have to make when we are talking about the absence of

Digitized by the Hunt Institute for Botanical Documentation

information, and information that is logically not there. Now I am not going to say anything more about this because Henry and George have been writing a paper on this and they are a heck of a lot closer to this subject than I am at the moment. At least they had better be, otherwise the paper may be a mess. There are other conditions where we have missing information. This occurs most frequently, of course, in any biological material. Lousy specimens, somebody man-handles the material and knocked off the goody that we wanted to measure at that particular time, and for some reason or other we feel duty bound to examine that particular specimen. A good example of what happens here, what we have to do, perhaps you are all familiar with the type method in taxonomy. Are there any of you who are not so familiar this, I think you have probably been through what we mean by type. This is not a morphological type or anything, this is a nomenclatory type. In order to discover what a particular author had in mind when he gave a particular name to an organism , we have to examine the specimen which he had in front of him in which he designated as the main bringing specimen. We would certainly wish to include the specimen in any study that we have. It is very frequently the case, however, particularly with the older types, that they are represented only by extremely small fragments. Again the historical development here can be very useful here in understanding that in the early stage of the game, a guy said, "there is not much use in keeping any more than just a tiny little bit of this plant, because I remember it." Maybe he did indeed remember it, and would remember it for his lifetime, but we are not there to quiz him or query him about what the heck the plant did look like, and so we do get fragmentary information at the beginning. Mostly, when we are making our classification schemes up, we do our very level best to exclude such material as that because we can't really make an adequate classification when there are lots and lots of parts of the plant missing. Maybe bugs chewed the thing up. Somebody didn't do the job

Digitized by the Hunt Institute for Botanical Documentation

of curating the herbarium like they should have. This is frequently the case, You see a nice fruit on the plant and say, "Ha, ha, I can learn all kinds of things about the seed inside it." And you go inside and find out nothing but a whole bunch of sawdust, and the carpet beetles have gotten in and made themselves a home. Now that is a different kind of missing information, obviously, from that where we had a plant either in flower or in fruit. There are many other conditions that we have to deal with under these terms of missing information.

Another problem that you will probably come across as you make your examination of these materials is this: that if you, for example, have decided that a series of linear measurements are meaningful to you, where you have some dimensions and you have broken this for, by your own volition, in categories. And you find that gee whiz when I look at these things, I am forced by the methodology which we are ~~xxxxxxxxxxxx~~ telling you to use here, namely, must have mutually exclusive attributes. That a specimen falling on this line here or a specimen falling on this line here will be complete mismatches if we use completely mutually exclusive states. This is state 1 and this state 2. We say well this doesn't

37

to my ideas about the relationships amongst these organisms. I don't feel comfortable to say the plant here and the plant here are that different. When you can see the kind of box we get ourselves into by any particular methodology or other, we always have to find a way out of it. Now we can do this by allowing a little bit of similarity to be measured between these two, between these and these in any way that you which. We merely call this particular condition where you don't wish to have your mutually exclusive ~~xxxxx~~ attributes to be sudden death or shift your two plants completely apart just by doing what we call order the states and allowing a little bit of similarity to be exhibited between one and the next, and for any number of states that you feel that there may be some judgment

Digitized by the Hunt Institute for Botanical Documentation

in this thing. Of course, the ordering of the attributes is an after the fact business that you have decided that you have mutually exclusive attributes. One can do the ordering, we can indeed instruct the computer to do the ordering for us. We still proceed to set up our input data as though these were indeed completely discrete and not overlapping in their information. The discussion on how we do this is going to be taken up by George and Henry. Are there any other things that I have left out ~~that~~ in this area that I should be putting into this thing? (Henry) Probably not, the logical state is probably best brought up at the same time. (Dave) And the use of the matrix.

Digitized by the Hunt Institute for Botanical Documentation

B 170
Lecture
Mr. Henry Fleming
April 13, 1966

OUTLINE

I. Non-inherited variation

A. Progressive individual variation

1. Age

2. Sesaonal

B. Social variation

C. Ecological variation

1. Habitat - ecophenotypic

2. Host determination variation

3. Density - dependent variation

4. Climatically induced variation

5. Heterogenic variation

6. Neurogenic color variation

D. Traumatic variation

1. Parasite - induced

2. Accidental and teratological variation

II. Inherited variation

A. Sex associated variation

1. Primary sex diff.

2. Secondary sex diff.

3. Alternating generations

B. Non sex-associated variation

1. Continuaous

2. Discontinuous variation (genetic poly)

3. Sex limited variation

Digitized by the Hunt Institute for Botanical Documentation

1. Morphological characters
   a. General external morphology
   b. Special structures
   c. Internal morphology
   d. Embryology
   e. Hargology (and other cytolog. diff)
2. Physiological characters
   a. Metabolic factors
   b. Serological, protein, and other biochemical diff.
   c. Body secretions
   d. Genetic sterility factors
3. Ecological character
   a. Habit and hosts
   b. Food
   c. Seasonal variations
   d. Parasites
   e. Host reactions
4. Ethnological characters
   a. Courtship and other ethnological isolating mechanisms
   b. Other behavior patterns
5. Geographical characters
   a. Distribution patterns
   b. Sympatric - allopatric relationship of population

Digitized by the Hunt Institute for Botanical Documentation

they will equal 1. What often happens when dealing with material, particularly herbarium material, is your lack of information. I have written in here in specimen 3, we could not make any statement about that character to what its property was due to the fact that the plant was damaged or the plant was collected at the wrong time. That is one of the things that I'm to be careful with in making up your coding. You may have no plants that are in flower. On other occasions you may have plants that are in fruit. The plants that are in fruit do not have flowers and vice versa. In that case if character 4 or 5 here and this character dealt with fruit and we had that specimen in flower, we will not be able to make any statement as to what the fruit properties are. In that case we will code in a 0. When we code a 0, the computer will not make a comparison for plant 1 and 3 as regards character 4 which deals with fruit because one of the plants did not have fruit. So this over here is modified when we come to compare plants 1 and 3. There is a number of matches all right, which is 3 but that is not the total number of characters used in a study, it is the total number of characters that both plants possessed that could be compared with one another. So this is due to the fact that we have no comparison possible on this last character. (Student)? They are going to be identical. (Henry) No, a point of fact they may not be identical that is the great danger of lack of information. If we had, let us say, 10 vegetative characters, 10 characters dealing with flowers and 10 characters dealing with fruit, we have a total of 30 characters provided we had full information on all of them. But if one plant has flowers, the other plant has fruit, and conversely they didn't have, you couldn't make a statement on whether the plant with fruit had flowers or not. We would only be making an effective comparison on 10 characters. It would hardly give us a valid classification. This is quite possible, a quite possible condition when we are dealing with tropical herbarium material. If you are dealing

Digitized by the Hunt Institute for Botanical Documentation

with local temperate zone material and you find that you have taxonomic characters dealing with both flower and fruit, you want to make sure that your collection is adequate. In other words, you get collections which are fully documented, that you have the full information on it so that you don't run into a really disastrous situation as I have put it here. That in general is what we are going to be doing. That is how the comparison is made. The significant thing to remember there when you are coding on Uluvaria is that you cannot make a statement in regard to a character that you have set up due to the fact that the specimen was collected at the wrong time or information is just not available on the particular specimen, code in a 0. But give good thought as to whether you want to use that specimen at all in your analysis because it is not giving you adequate information. Now we have one other condition that is important when you are first setting up your characters and that is what we call a logical state. Perhaps the best way to illustrate this is not with a botanical specimen or

14.7

first point this out in publication some time ago. Let us deal with types of hats people wear. Since this publication why quite refined concepts for us. Let us say that is our character. We are going to have the attributes of a man who wears a cap, he wears a Stetson, he wears a Panama, and etc. And we come to a man who doesn't wear any hat at all. The fact then that he wears no hat doesn't mean that we do have information. He just does not wear hats. And yet it is a very real thing to make that statement in that character that he does not wear hats. Sometimes you have to think about the way you form your characters. If I were to, for instance, try to classify these windows here and I use as a character has hardware on the window, I would develop no comparable information because both of them have hardware. There is a latch right here between the two windows. But I reframe that and make the statement that it has a different type of latch on each window is real. If I make the statemnt in to these two windows a character

16.5

Digitized by the Hunt Institute for Botanical Documentation

that one has small panes and the other does not, and then I start dealing with the small-paned window. It is a valid statement to make. It is a logical consideration to make that whenever I am dealing with the characteristics of a big window I code the other windows in as a not logical statement. In other words, they do not have small windows, they do not have hats. (George) You might consider the color of the small window. This comparison could be made very well. (Henry) Do you want to continue with that? I have some different ideas of what (interruption by George) because we are dealing now with a not logical statement, the thing should be pounded on. Would you like to address yourself to it? (George) I was going to say that we could take for example that you have green ones and blue ones, and you are able to differentiate on the basis of the color of the small panes. But if we have a character that says color of small panes meaning three states, small panes green, small panes blue, no small panes. (Student) Is the lack of a character also a character? (George) Not a lack of a character, it is a lack of what seems to be a desirable description for it on the basis of the fact that the observable description can't possibly be present rather than it is not there by default of the collection or default of the season or something like that. (Henery) In other words, you can make the statement while we are dealing with the colors of the windows and the bottom windows do not have any color panes in them, we can make the statement they do not have color panes. (George) Is there an example of this on these herbarium specimens? (Henry) Not on the Halameda. Well we haven't coded out the Ulvularia so we don't know. I can see this condition appearing in the Uluvaria. That is one reason why I am bringing ;it up. It is going to depend upon their coding there though. (George) It will depend upon how these things are observed. (Henry) Perhaps if we pass this out for not logical states. This makes a good example right here. These are coded sheets. These coded sheets by the way were

Digitized by the Hunt Institute for Botanical Documentation

not designed for us. It is one way that a taxonomist may work. This man is director of the botany department at the British Museum and he is working on a group of plants here which are primarily Caribbean plants. If we look down here at a character 4, we have hairs here present and absent. Though we have presence and absence, you can always look almost inevitably for a subsequent character which will contain a state which does not have the property of the plant but is a logical state. If we follow below there, you see that the hairs are short or long. The only hairs that could be short or long are on those objects in which the hairs were present. So therefore when we come to a character called hairs short or long, we must put a logical state in, hairs not present, or at least no statement can be made as to whether they are short or long. In other words if I were to code that up here, we have presence and absence of hairs or character 1. Character 2 has hair short or long. That is object 1 and object 2 being compared present or absent. Object 1 has the hairs present, they are also short hairs. Object 2 has the hairs absent. You cannot say whether they are short or ;long, we put a logical state in which signifies that we cannot make any statment as to whether they are short or long, and if you continue down on these objects, all of those that do not have hairs at all will be coded as similar. I am sure there are some others in here. If you look at 7, you will find the same condition, spines present or absent. Now if they are present, you will notice that the next character ~~x~~ has two attributes in it ~~xpx~~ up to 12 mm less than 12 mm. We will put an additional state in there,logical state spines ~~x~~ are absent. (George) Up to 12 mm means long? (Henry) Long. (George) Up to 12 mm this length of spine does not exceed 12 mm? (Dave) Check. (George) What is the next state? (?) That represents less. (George) Aren't those the same thing? (Dave) It gets/back to the same point that I was talking about the other day in this discussion. ~~Wexxxx~~ Where work has some difference in mind for his own

Digitized by the Hunt Institute for Botanical Documentation

cognizance of that difference he makes some sort of a statement which may or may not be meaningful necessarily to somebody else after the time of this statement being made. He has some difference there that has been stated inaccurately for the reading of the likes of George Estabrook who wishes to say what kind of a nut is this? But I am sure that if you will think of your own kinds of comparisons, the first thing that you make will be some sort of a notation in your mind which says, "I think this is different from that, and I will write something down here which merely says that statement." And that is the case with anybody who is doing this kind of work. They will be very informed about it. Surely if you are going to examine this guy on you are going to find holes in it all the way through. (Student) How would you code this? (Henry) I don't think he means that. I think he means longer than 12 mm. I think that is what he means. It is just the way he has worded it. He didn't do these sheets for us, these are for his own personal sheets. He wouldn't publish it this way. It frankly didn't make too much difference to us because you see he has underlined with red those characters or those properties and attributes for each of his particular specimens. (Dave) Each of these sheets is the actual coding for a specimen of the genus . The red underline is the indication of what that specimen had on it. (Student) You mean even though he didn't tell you what he meant, that you coded it the way he put it down./ (Henry) It goes into the second state. And as I say, these sheets were his own sheets and whatever private way he had.....(Dave) Let us put it this way, he was to make to see/ whether or not our method would hack it with his stuff and it does, it turns out that when we take these coding, take it literally off the specimens there or those sheets, we do the job for him. But his statement about what he means here is not clear obviously. (Henry) But it didn't concern us, we want a word about that. He knew what he was doing so we just put it in second state form for

Digitized by the Hunt Institute for Botanical Documentation

that character. I think you can find some others which are not logical. (Student) Number 11. (Henry) Yes. These decisions, of course, we had to make as to when they were going to use a not logical . (Dave) It is possible on the basis of our operations the way we handle these things statement about mutually exclusive states. It can be possible for one set to have a precisely defined length which is a subset of another which has a less precisely defined length. (Henry) Now those are the essential points to be made, What we will have to do is making pair wise comparisons. The computer will make them for us. It will make pair wise comparisons between all of the objects in the study. In other words, one will be compared with two. One will be compared with object 3. One will be compared with object 4. One will be compared with object 5. Then the computer will go back and take xp object 2 and compare it with 3. One and 2 have alreadyy been compared. Object 2 with 4; object 2 with 5, and so forth. This means a great number of comparisons, one reason for making use of our computer. The number of comparisons that will be made will ba which means then that if we have a 100 specimens we are working with, there will be 100 x 99 divided by 2 which is 4950, you get 4950 comparisons. These problems are too laborious to do by hand, particularly if you have to them in this fashion. And you start making a comparison with 1 down to 100 here you usually get just where you have coded them in your eye missing it, without a mistake. And a computer will get the various degrees that will determine the similarity between each of the comparisons, each of the descriptions we have put in the computer as to how similar each description is. It keeps track of which specimens and we have told the computer which description applies to which object by number the objects. The computer will order these and on Monday George will tell you what we do with these similarity ratios. Some times in your working

Digitized by the Hunt Institute for Botanical Documentation

over your material, remember we have set up the rule, we have mentioned it several times, that we can only code one property to each character. In other words, they must be mutually exclusive. When you some condition or a character the measurement of a leaf, let us say the leaf is 1 to 2 cm, 2 to 3 cm, 3 to 4 cm. If they think the specimen 1 and specimen 2 and the rest of your specimens are all that way, and your 3rd specimen may fall over here, and so forth. And ;you don't want the similarity you expected as a complete mismatch. If we have them mutually exclusive which they must be, and we compare them under the fashion that we have described, then the similarities for this particular character between object 1 and 2 will be 0. We may not like a 0, particularly 2 over here that is also 0. We feel that some how or other these are more closely related than specimen 3 down here which is way out near 5 and 6. George has developed a method for partial matches in which similarities from this specimen and this specimen will be greater for this character than the similarities from this specimen and from this specimen. I won't go into the equation on that today. This is the sort of information that you are apt to arrive at after you have done all your coding and you start looking at your characters and what the significance of your characters are going to be in your classification. You are going to find some characters possibly I don't know that you may not want a complete mismatch. in other words This works pretty much in those areas in which you can order your characters. You sort of feel that along the line, and you want this and this specimen to show higher similarity than those 2 specimens. It is also true for this equation that George has that if we have say states here, you may not want to go across all the states and reflect different degrees of similarity, you might only want neighboring states to reflect the similarity, and that there be a complete mismatch between state 1 here and state 3. We can instruct the computer which is making the comparisons, and that is all we do. your coding goes along

Digitized by the Hunt Institute for Botanical Documentation

identically the same. If you decide you want to order a state you just say Bob Brill that you want character 5 to be ordered and you want it ordered over 2 states, 3 states, 4 states, Tell him what similarity you feel ought to be reflected between the different values. Sometimes we don't want a complete mismatch and yet our material is not well orderable. It doesn't go along nice stringent linear lines the way this did. Up here the similarity may be largely intuitive. We just feel that something is different. Here we can state that between state 1 here and state 2 here ~~xxx~~ there was obviously to 0 and everything that fell in between 1 and 0. No mismatches on the extremes or complete mismatches on the extremes and smaller mismatches in between. Suppose we had ~~xx~~ a condition 1 + - *. We want to make some sort of comparison between those four. There is some similarity between this and this, there is a straight line here similarity between this and this. And these are some lines in here. We can think of these two things as being fairly extreme condition. Let us put it on a matrix amongst ourselves and decide how similar they are. Those two are completely similar. We will put a 1 in, complete similarity. say are completely different. So far as one is concerned, we will have those representing the extremes. Our problem comes in..... of these perhaps arbitrary. We might decide amongst ourselves that this is half like this and it is also half like the other extreme. So you can say that any comparison between two ~~xxxx~~ plants this plant here with one line and the + sign is half similar. and a + with this one which ~~xiii~~ would be the same reason, half similar. This has at least two straight lines similar to that except line, the line here has some additional lines. We may say that the similarity between figure 1 here and this * arbitrary for 25%. We think it is about one-fourth similar. Consequently, on the opposite side it will be 25% similar. And the * and the + sign are much closer to one another than they are to either the 1 or the dash(-). I mean

Digitized by the Hunt Institute for Botanical Documentation

arbitrary (writing on blackboard--I cannot hear it all)

You get this sort of condition. Open up those herbarium sheets in front of you, you see a photograph on them. You want to open them up. (Confusion of moving paper, etc.) You have photographs on each one. You can peek over other people's shoulder there, you find that some of thes plant plants are only branched at the top. I don't know whether you have any specimens there in which they are just branched or completely unbranched. those are the two extremes, and those conditions are met with in regard to these plants. (Dave) You can get up and look around and just by glancing about you can see differences. They are all standardized photographs against that white backboard. (Henry) You can see the differences in the branching pattern going on there. Some of them are very full plants, some only have branches near the top of the plant. We actually determined there was 16 different branching patterns there. But they didn't follow a linear line. It was rather difficult, for instance, to say how close on any linear line a plant that has flowered say a quarter way up from the bottom is to a plant that is a branched yet. A quarter way up from the bottom is to a plant that is branched at the very top and at three quarters. There was no straight linear line to follow. We tried that. But we could code something in that was satisfactory to Dr. Rogers by looking at the plant and asking him, the question, "How similar do you think this branching pattern or this form of the plant is to this form?" And by asking him enough questions, we were able to make up a matrix such as this which is quite a pleasing matrix. But we couldn't have ordered it in any possible way there was no straight ordering line. The line down between the two might be thought of as a complicated branching pattern. And the difficulty is this is not branching of the plant by the way, this is a diagram to show how complicated it would be to say, "What is the relation of that branch to that branch?" How similar is it?" And every once in a while we come across that condition so George Estabrook devised this manner for us to express our

Digitized by the Hunt Institute for Botanical Documentation

intuitive. There may be good logical reasons for making our decisions. We had certain types of reason working in here, but we couldn't have run these particular symbols in a straight line and done it in an ordered fashion. But if you run into this condition in Ulularia, mark the character, take your states and determine how close you ~~xxx~~ think each of those states or attributes are to one another as they express themselves to you. Make up a matrix in this fashion and then just tell Bob Brill put the matrix into the computer. When the computer locks at that particular character, it will put it in a particular similarity that you specify. In other words, if it came in a plus object, ~~o~~l had a plus sign and object 2 had an asterisk, it will panel that character to a 50% resemblance in the comparison of those two objects. If object 1 was a 1 and object 3 were a -, when it sees that ~~thxxx~~ it will say there is no similarity whatsoever.

I had hoped we might be able to look through Uvularia today, but it doesn't look as if we will be able to. Now some of the things you have to watch out for will be expressed in these plants. One of the things as you look across from one to the other, I think you will notice the difference in the lobe of the leaves. Now if any body has some, a flowering branch there, we know from experience that the flowering ~~xxxxxx~~ branches we cannot use. ~~Xfxxy~~ They either have one leaf or they have three leaves, and they are all the same. You don't have the full specimen here. Dr. Rogers in those drawers also has specimens of the roots and specimens of the stems that are associated with each of these specimens. There is another point to be brought out. Characters and properties are where you find them, comparable properties. Here we can use a number of lobes of the leaves. You might also notice that the lobes of the leaves are different shaped. I see a linear one down there. There is a linear leaf. (Dave) You remember the Mae West character? (Henry) There is a Mae West. They also appear to be different lengths. There is another character which ~~xx~~ we would divide into states. Now on the right hand here

Digitized by the Hunt Institute for Botanical Documentation

you will also notice a good deal of data. Much of this data could only be collected in the field on fresh specimens. Colors of the petiole, for instance, are lost in the herbarium specimens without Dr. Rogers taking that information down as he collec ted them. That we can code in because we have a record of it. This is also important for instance in orchid classification. The color of the flowers and the shapes of some of the parts of the flowers are lost in herbarium specimens . You will also find, the business of characters are where you find them and the properties of the plants are where you find them. One would never mistake, for instance, a Rhododendron for a or anything like that. the roots system is entirely different. You may find good valid characters in the roots. A group of insects, mosquitoes that puzzled them over in southern Europe for a long time. In some areas the same mosquito was a vector for malaria and in another area it wasn't. And this kept going on to confuse the medical people and the economic entomologists no end. They finally found out that they could segregate those mosquitoes that were vectors by the way they laid their eggs. One laid them in a and the other laid them . Then they found there was also an ecological difference they found no morphological differences. It makes up what you might call a super species. But in mosquitoes it was unusual to use egg stages for the classification of mosquitoes, and yet it worked in that particular instance.

(Second Tape)

You may have 20 characters, you may have 50 characters. It depends on material and it depends on how, on your own ability to find comparative material.

Digitized by the Hunt Institute for Botanical Documentation

B 170
Lecture
Mr. Henry Fleming
April 13, 1966

OUTLINE

I. Non-inherited variation

A. Progressive individual variation

1. Age

2. Seasonal

B. Social variation

C. Ecological variation

1. Habitat - ecophenotypic

2. Host determination variation

3. Density - dependent variation

4. Climatically induced variation

5. Heterogenic variation

6. Neurogenic color variation

D. Traumatic variation

1. Parasite - induced

2. Accidental and teratological variation

II. Inherited variation

A. Sex associated variation

1. Primary sex diff.

2. Secondary sex diff.

3. Alternating generations

B. Non sex-associated variation

1. Continuaous

2. Discontinuous variation (genetic poly)

3. Sex limited variation

Digitized by the Hunt Institute for Botanical Documentation

1. Morphological characters
   a. General external morphology
   b. Special structures
   c. Internal morphology
   d. Embryology
   e. Hargology (and other cytolog. diff)
2. Physiological characters
   a. Metabolic factors
   b. Serological, protein, and other biochemical diff.
   c. Body secretions
   d. Genetic sterility factors
3. Ecological character
   a. Habit and hosts
   b. Food
   c. Seasonal variations
   d. Parasites
   e. Host reactions
4. Ethnological characters
   a. Courtship and other ethnological isolating mechanisms
   b. Other behavior patterns
5. Geographical characters
   a. Distribution patterns
   b. Sympatric - allopatric relationship of population

Digitized by the Hunt Institute for Botanical Documentation

B 170
Lecture
Mr. Henry Fleming
April 13, 1966

You are doing a revision, aren't you Craig? Are you following the figures more or less in Davis and Heywood, or what model are you following? (Craig) I'm following Dr. Harrington's on the monkey. (Henry) I see. In a formal description of a new species, we are presumably asked to describe the species. Now how do we describe the species? Davis and Heywood, I think, call it the formal description. Let's write this down, I would rather use different terminology. They use such terms as formal description, then he follows that with diagnosis. When we describe a species, or any object for that matter, we do not describe it completely and in complete detail, because we could keep describing this book probably ad infinitum. We could use microscopes on it, electron microscopes, and if you start going into what the constitution of the paper is and all that. We'd never end describing that book. What we have to do is use comparative descriptions. So I'll use a different set of characters for describing a book than I would the ashtrays; I use a different set of describers describing algae than I would higher plants. We use information in other words that is comparable. So right from the beginning, all descriptions consist of a diagnosis. We make a decision as to what characters are important from a comparable point of view. If we want to classify or we want to separate a group of plants, we make a selection of those characters which are pertinent but differentiating them. We take any ordinary tree we can find, well practically all of them for instance, all of the Tsuga for instance will have green needles. There is no point in making a character of green needles for Tsuga. They all have them, it won't differentiate them for us. We have to pick out those characters which are separable and which show some differences, some comparison. In some cases you will find botanical descriptions, some fellow will describe the species and they will go on for two or three pages. Actually, the whole description can be put into one paragraph.

Digitized by the Hunt Institute for Botanical Documentation

He is just not capable of deciding which characters are of comparable value. (Craig) If I go out East and find a new species of Tsuga, don't I have to put in some character description in my description which places that plant in the Tsuga? (Henry) That's right. Well, I'm talking of the species. If we are going into the genetic level, then there are a set of genera characters. They have been pretty well laid down for us. We have got some sort of a cluster of plants that belong in Tsuga. (Craig) My point is, if you collect the plant the first time don't you have to justify placing it in Tsuga instead of Pseudo Tsuga? (Henry) That's right. (Craig) Then I don't understand why you couldn't use a character green needles, something similar that all the other members of Tsuga would have. (Henry) Then we would be in with the genera character not with the species character. (Craig) You differentiate between the description? (Henry) That's right. In the higher up we go. If we want to describe a family completely, we would have to describe every single species and every single genus that we put in that family. That would be a ridiculous description. We would have to read volumes and volumes of information that wouldn't really be pertinent to finding . We select out those characters which are comparable now between families or between genera. (Craig) When you speak of description, are you talking about the classical or Latin description? (Henry) The classical Latin description. And that I would call a diagnosis because we had to diagnose what characters to use. Davis and Heywood call a diagnosis a comparison. I would call that the differential diagnosis where we usually take and try to place a particular species already used for the whatever you are dealing with to a usually more closely related or other related genera of species that one would be apt to use. Sometimes we pick something unrelated just for emphasis. So I would call this the differential diagnosis. In plants or in animals, we have got a number of types of characters and let's deal with those first which, except under unusual conditions, we will probably never

Digitized by the Hunt Institute for Botanical Documentation

want to include in an overall comparison of plants to determine their overall similarity. First of all, we have non-inherited variations and which we can think of - progressive individual variation would be a subtitle there. The age of a plant would be one sort of consideration. If we are taking a group of species, we don't want to compare seedlings with mature plants. The characters that we would find in seedlings, the leaves are apt to be a different size for instance. It will not be in flower and we won't have enough information on that. We could however, if we wanted to make a classification that would be a specialized classification of seedlings, it might have some advantage to a man, for instance, in the nursery business to make a classification just on seedling plants--might be real handy to have such a classification. That would not be a general botanical classification. We would want to compare those characters on plants of all, at least the mature plants somewhat near the same age, We also have seasonal variation. Usually, you get seasonal variation,for instance, in algae. You do in some plants, we can think of for instance deciduous trees outside in the winter not having any leaves on. We would hardly want to then an overall comparison to throw in plants that were without foliage in the winter and compare them with plants with foliage in the summer. In other words we want to collect the plants more or less at the same time so that they look somewhat similar. You run into the same thing in the tropics in the dry season and the wet season. Plants look quite different. Social variation you often get primarily in the insects. If you are doing a classification of ants, you are going to have to have some special biological knowledge in order to make the comparison, because with ants you will have soldier ants, you will have just plain foragers you will have queen ants. You have got to group them together and handle them as one because you have a unit here. You must compare your unit together. If you don't have this specialized knowledge, you are just going to have to use, let us say soldiers, and compare the soldiers for your classification. And your classification will be restricted to one form. And you have got to realize

Digitized by the Hunt Institute for Botanical Documentation

that your one form of, because you have used your one form of insect is going to be restricted and be of value just for that form. You lack a tremendous amount of information on your other forms. Then we have ecological variations. For a general classification again, is a good question whether we should make the habitat part of a character. Habitat is often exclusive information. (from student a question that I can't hear). (Henry's answer) Not actually part of the organism itself. Whether I place this book on this table or that table really has nothing much to do with the book. Whether the book, we might be able to say things because it is placed on different tables, placed on that table is for general use so that everybody in the lab here knows what publications are coming through. We may be able to infer some types of information because of where it is placed. But it is not part of the actual properties of that book as to where it is placed. (Jack) Seems to me like that is one of the most helpful things you can include, though. If this plant, for example, is never found in any other situation but a shady damp area, this should be a character that would be useful. (Henry) It is often the character that we use amplying special considerations to. It may infer, for instance, that this species does not hybridize, or is isolated at least. It is an isolating mechanism. And we use the information in that sense, not in the sense of overall similarity. I could pick out situations or think of situations where I might want to code in that sort of variation. Soil conditions might be important. (Student) It might not be morphologically similar, but it is physiologically similar. (Henry) Then, I've got that down here too. I don't wish to go into that in too much detail at the moment. Physiological information is of great importance, provided you have enough of a , concerns all of the objects that you had to classify. Unfortunately, usually whether we are dealing with plants or animals, some physiologists have worked out in great detail the

Digitized by the Hunt Institute for Botanical Documentation

physiology of a certain phenomenon going on for one species, and there are 100 species in this genus so something like that we don't have any information in regard to the other 99 so it is not comparable information. But we keep that in the back of our minds and we can perhaps whether justifiably or unjustificably infer that maybe the others are the same. (Student) I think to go out on a limb and say that seeds tell me that plant distribution is valid as a key to the identification plant. (Henry) It is, but we are not talking about identification, we are talking about classification now. And it is real handy to know that a certain group of plants only exist in Australia. It limits our searching. We can use that definitely in a key for identification. (Student) But doesn't this fact that they are located only in one place imply that they differ from other plants in some marked way? (Henry) It does imply that. Well, we infer that as true. On some occasions we can take two quite different plants and bring them down to the low lands and may in the ~~Arctic~~ Arctic conditions and let us say that we can grow them under arctic or alpine conditions. Both of those plants may become cushion or dwarf plants, quite different than they are in the low lands, but both of them will react the same way, and I am sure that Craig here can give you a number of examples of that sort of situation on the . They are reacting to their environment. And if you are doing an ecological classification, then such characters are of importance to you.

13.5

Host determination variation, that is primarily one that mycologists get into in particular. And again that is a very good identification procedure, but it is not necessarily valid, unless you can have other ~~inert~~ information if it is a segregating information, then you may be able to use it. But in general, host variation is not important. It is nice that we find that a certain genus of fungi is always on a particular type of plant. It is real helpful for identification purposes, but it is not how we cluster them. It is not how we ~~may~~ determine whether they are a nameable species unit. We also have another type of variation which I have down here as density. Sometimes plants that grow in great masses

Digitized by the Hunt Institute for Botanical Documentation

or you find them all in a small place will take on certain types of variations simply because they are over crowded and won't grow well. And they look aberrant, the leaves will be small, the flowers will be small, or there won't be many flowers and things of that nature will happen to them. And we have got to take that sort of thing into consideration very often in herbarium material. Another thing that happens to herbarium material that I might mention in passing is that if a botanist goes off and he sees a field of certain type plant, he usually picks the best looking plant, has the biggest leaves, has the biggest flowers, It may not be typical of the way the plant normally is, and you ~~xxxx~~ better take that into ~~xxxxx~~ consideration usually in selection of your characters. There is a climatic variation. This has very little again to actually do with 15.5 the overall similarities between a group of objects. It has to be taken into consideration that it is better to handle it as a separate piece of information. As we get into the program later on and we start clustering, very often you might notice that plants go in one after another in their geographical clime or a clime. But it is a separate judgment that you have to make. You have to correlate it yourself. It actually, while it has its affect upon the way the plant looks is not again an intrinsic character of that plant. It is something that is imposed on the plant from the outside. Then, of course, and ~~thxxx~~ sometimes these are ~~diffx~~ difficult to decide on occasion, traumatic variations. You may have a plant for instance that has certain insects eating 15.9 on it that quite radically alter it. Sometimes these form gall or witches brooms and which a tree that has broom on it can look quite different than another type of tree. Insect bites, the way for instance a caterpillar chews on a leaf can help very often in identification. The way certain caterpillars, for instance, leaf often determines what caterpillar you have. I have often purely kidded Dave Rogers that I could classify all of the Passiflora, I can't actually do all of them, but Passiflora in northeastern South America without using a plant character. I can do it on certain genus of butterfly, where they

Digitized by the Hunt Institute for Botanical Documentation

feed on the plant, how they feed on the plant. It wouldn't be a good classification of Passiflora, but nonetheless from an identification procedure and let us say for an entomologist running around and trying to identify eggs or caterpillars on the plant, it ~~xxx~~ would work, but it would not be classifying the genus types of flora. You get into some interesting prob lems even looking at it from that point of view. In northeaster South America, I have never been able to find any of these Heliconius butterflies on the common economic Passiflora edulis. Edulis has been spread pretty well around the world, but I think ~~the~~ originally it was indigenous to Brazil. In south Brazil, you will find one species of Heliconius on P. edulis. It is also a different subspecies. Whether that is nice taxonomic correlation, whether it has anything to do with being a subspecies or not is something else again. But you can go all over into the southern United States and you won't find the Heliconius feeding on it. But down in the indigenous area there is one species. On the hand I can classify probably the butterfly or at least rather not classify but write a key to the butterfly just using the plant characters. And if you can determine the plant characters, you will know what insect is feeding ~~xxxxxxxx~~ on it. But it is not a classification of either the insect, we are using extrinsic information there even though the correlation is very good. And I can say that if you find a caterpillar on the genus Passiflora, it is going to be a Heliconius, but I cannot say if you have got a plant Passiflora with insect damage on that it is necessarily Heliconius. It may be a number of bugs and other things. And from those bug bites you can perhaps identify ~~the~~ your Passiflora, but it is not ~~a~~ part of the speciation of the genus Passiflora. (Student) But it seems to me that all of the extrinsic ~~xxxxxxxx~~ characteristics that we have been discussing here imply something intrinsic about a plant. Maybe not individually but at least the summation of a number of these characters imply something intrinsically different between plants, and it seems that this is what we are after. What is the, maybe it can be said in a philosophical way, what is the essence of the plant, and that is how we make

Digitized by the Hunt Institute for Botanical Documentation

our classification. What is it intrinsically, what are the intrinsic characters of this plant which make all of these plants related and unrelated and everything else? But I can't see how except thats purely arbitrary you can separate the intrinsic from the extrinsic. (Henry) It is in many ways arbitrary, just as it is in trying to classify a collection of books, but there are certain characters which are a part of the genetic makeup of that plant. All of sorts of accidents can happen to the plant otherwise. We do have to at some point make what is largely an arbitrary decision. Many of these things I have mentioned here, for instance, are perfectly good for special classification. If I want an ecological classification, I would use ecological data, but we have to set up some sort of classification for plants that is of use to a great number of different individuals. (Student) And this is not the special? (Henry) And this is not the special. (Student) You define special as being only for one group whereas general botanical classification is..... (Henry) Good for the greates possible number of different uses. In regard to your point you brought up, this is a little bit perhaps off the subject, but there is a certain insect which they were able to feed on certain plants in Czechoslovakia, they brought some of this same insect material to Harvard. At Harvard this insect would never get out of its larval stage. It would reach an adult stage which was non-sexual with many larval characters. There was considerable amount of time spent trying to determine why this insect when brought over to the United States under the same laboratory conditions the same, would not mature become sexually viable. It was finally tied down to the paper that was being used for the food that this insect fed on, and this paper was from Douglas fir. Apparently, all the paper that they tried whether it was magazine paper or anything else as I recall, all had some origin of Douglas fir. And there was something retained from the Douglas fir that was limiting this one insect. The other insects were raised fine in the same genus, just this one had a limiting effect this one insect. This is really not comparable information.

Digitized by the Hunt Institute for Botanical Documentation

It is highly interesting, and it is good in the biological information, but we would hardly use that type information in classifying that insect. We have certain types of inherited variation, sex associated variation. You get primary secondary sexual differences. They are usually true in animals. Alternating generations and higher plants, I don't think that is of any importance really, but if you get down dealing with algae why it becomes critical, and while I have this in the section that we probably wouldn't use it in classifying, we would in the algae, but we would have to get all of our data together on the generations so that at least we weren't trying to compare one stage of an alternating generation with another stage. We would have to have all of our information on all of the stages or else restrict ourselves to one stage, usually. There is a way we can handle that if we only lack some of the information on some of the species or some of the genera or some of the families depending upon what we are handling. But if we lack it on a great number of classification, the final analysis, except for what we can infer, then the inferences will never be sure, will have to be based on those characters in which we are able to supply information to.

Let us go on and deal with what characters we can use in a classification, and these, some of these at least will be handled in the fourth chapter of Davis and Heywood which I suggest you read. While I am mostly familiar with American tropical botany, in American tropical botany youare practically if not solely restricted to external morphology. You know so little else. And that, of course, is primarily all plants and animals basically are classified, just because we lack other types of information. And it is the obvious thing. If you take for instance this Structure of Barnacles here its almost completely morphological information. Do you have other types of information, Craig, in the group you are working with? (Student) I hate to be the dead horse here, but it seems to me that morphology is at least ~~xt2~~ as variable as most of the other things that we have indicated here. (Henry) Let us go on ~~xkit~~ with what I, at least in my own

Digitized by the Hunt Institute for Botanical Documentation

private way and I think you are finally coming to agree with it, does constitute good character. At least they are not varying in the same way, because of external influences or special types of influences. Another type of morphological character that we have that is good are special structures. You will find these in fungi, we find them in higher plants, some groups you are really restricted to separating them out on structures which are rather private to the group you are working with. Don't use another group of plants whatsoever.

On some groups at least, especially if you get up in the temperate zone, we can do a good deal with internal morphology. Embryological characters when we have them can be thought of as morphological. And a good deal of cytological information gets into morphology , and also about as variable as any other character. Of course what we look for in selecting characters is try to find those characters which do not vary greatly, individually at least. But constant variations are good and we are looking for constant variations, of course, between organisms. You have to be just as careful, for instance genetical material as you do looking at leaves. This is often brought out in some of the A very slight change the chromosome sometimes produces tremendous phenotypic effect, sufficient enough so that they wouldn't even trace the let alone in the same species you have been dealing with.

Then we have physiological characters. Here again usually our information is rather limited to a few species that somebody has worked on, and from the few species we make the extension that is probably true of all of the species, but this doesn't by any means necessarily hold. We use metabolic factors, serological in many animals, protein and any sort of biochemical differences. Rutgers set up a bank of serological material. Sometimes the correlation between blood types and established taxonomy is quite good. Just as frequently it didn't, or I say just as frequently, I know some it just simply didn't work out at all. There was so much variation of that any organisms tested that serological material was valueless from a taxonomic point of

Digitized by the Hunt Institute for Botanical Documentation

You could on the other hand if you wanted to set up a classification just on serology, but it would not have been an overall classification, and you still can include it in a taxonomic analysis, but it is not going to correlate in many instances very well with your data. Genetic sterility factors can also be used of course as a physiological character. Now we have ecological characters. You notice before I put them down as one type of variation that couldn't be used, Well under certain conditions, you can use them. It again depends upon your problem. Habit and hosts for instance are really often quite important characters. Type of soil a plant grows on can be an important character. Seasonal variations can be used as separaters very often. There again if you get American tropical plants you are not apt to know very much about that, typically if you are working on herbarium material you don't know anything. Though sometimes if you have enough herbarium you can make inferences from it as to what, you ~~knew~~ notice the dates on the specimen, when it was collected, and if you know about the climate or the area that these plants have been collected in, you can make inferences as to what happens in the seasonal .

Parasites can sometimes be used. They tried using parasites for instance in bird ~~bird~~ taxonomy and as with many characters there is really no . It worked with some groups and didn't work in other groups. Here again on the parasites the taxonomy was set up by the ~~bird~~ morphology, morphological characters seem to be good, they hung together, they hung together with a good number of other things treating the beasts in isolation and just working with the beasts without their relationships, ecological relationships to the host. So in many instances the host thought it was disregarded. But it is nice from an identification point of view if you can pick up a say a mite on a bird and look at it and say, "Oh, well that means this bird"--some identification procedure very often but it is not a classificatory procedure.

Now ethnological characters, of course, work primarily in animals again. I can say for instance on fiddler crabs that if you watch the courtship of two

Digitized by the Hunt Institute for Botanical Documentation

fiddler crab on a beach some place, you can use their name of specie. You can do it with a good ;number of insects; you can do it with a good number of birds; you can do it with a good number of lizards. Unfortuantely, here again usually you don't have comparable information because something is known about something, is known about some of the specie and usually a small number of specie but not known about all the specie. So you bring this information usually in after you have classified. You may readjust your classification on the basis of some of this, but you might just as well call lack of information or specialized knowledge. I cannot think of any botanical example which might involve courtship. Well in a sense, I can too. You can think of isolating mechanisms. Mike Worth who was with us in New York, well it wasn't Mike Worth, actually ;it was Dressler who is not in Brazil who was working on a certain group of orchids in Central America, I think it was Costa Rica. There are two species, well it was on the two species of Oncidium which he had collected a good number of years ago. They were on opposite sides of a mountain. He went down maybe it was four or five years ago and this mountain had been cut down, the forests on this mountain had been cut down and he started finding that at the edge of the forest, on the ridge of the mountain, some orchids that had been undescribed. The more he looked at them, the more he realized that what had happened here was he was getting hybrid between two Oncidium of the east side of this mountain and all of a sudden they were hybridizing. He had never found any bybrids before. The explanation was that two different species of bees, or one species of bees rather, was a pollinator here, and formerly the population at one side of the mountain never got to the other side of the mountain, but with the forest cut down the bees started spreading and one population on one side of the mountain went to the other side and carried the pollin from the orchids on one side over the mountain to the orchids on the other side and that is now you were getting hybrids. That is an explanation of what happened. Again, it can be thought of as a good character.

Digitized by the Hunt Institute for Botanical Documentation

Distribution patternsn can be used as characters. Then you can fit in what you know as characters on their sympatric or allopatric relationships. But in general, with most plants we are going to be restricted, I am afraid, to morphological information, and our general classification will be largely morphological in nature. If we were to classify these top windows in this building, there would hardly be any point in dealing with the size of the windows, they are all about the same size; dealing with the patterns as far as their having one type of glass on the edge because they are all like that. We wouls start by looking at the colors of the glass on the edges, and we could develop a classification from them. Some of them are green, some of them are ;sort of bluish-green....(Student) You have to start with some designation as to relate them from being different from the lower ones. (Henry) Yes, if you want to deal with them all at once, that would be our first step in the classification. I was dealing with the top ones because I saw they are all, they was some slight difference..(Student interruption) (Henry) But by working with the color there and perhaps a number of colors, we could isolate or sort out these windows and classify. There would be some windows for instance with all green edge glass, some mixed with two colors. And I see one over here a different kind of spotty green. And we would classify these, and in this instance what we would be doing is taking comparable information, things that we could compare between these windows, but not characters which are true of all the windows that would not separate the windows out. What could you say about the history of these windows is something else again. You could think of that in perhaps a phylogenetic fashion, and I think it would be just as tenuous as phylogenetic classifications often are. Which are the original windows, and which are the windows that have been repaired. We could probably make a, do it in statistics and say, "Well due to the fact that the glass in that window is different than all the rest, it is likely that has been repaired." But what are we going to do between these green and blue windows? Which is the older? And perhaps they

Digitized by the Hunt Institute for Botanical Documentation

were put that way originally, I don't think so but they were very apt to be, and it is a probability that we have to deal with there. But in dealing with plants, while this is a subcase, dealing with plants is not too much different. Here we are dealing with morphological information in regard to the windows. I think it is getting time to stop the class. Are there any questions?

Digitized by the Hunt Institute for Botanical Documentation

I Non inherited variation

A Progressive indiv variation

1 Age

2 Seasonal

B. Social variation

C. Ecological variation

1. Habitat - ecophenotypic
2. Host det. variation
3. Density - dependent variation
4. Climatically induced variation
5. Heterogenic variation
6. Neurogenic color variation

D Traumatic variation

1. Parasite - induced
2. Accidental & teratological variation

II Inherited variation

A. 1. Sex associated variation

1. Primary sex diff
2. Secondary sex diff.
3. alternating generations
4. Gynandromorphs + Intersexes

B. Non sex - associated variation

1 Continuous
2. Discontinuous variation (genetic poly)
3. Sex limited variation

Digitized by the Hunt Institute for Botanical Documentation

1. Morphological characters
   a. General external morphology
   b. Special structures
   c. Internal morphology
   d. Embryology (+ other cytolog. diff)
   e Karyology

2. Physiological characters
   a Metabolic factors
   b. Serological, protein, + other biochemical diff
   c. body secretions
   d. genic sterility factors

3. Ecological Char.
   a habit + hosts
   b. food
   c. Seasonal variation
   d parasites
   e Host reactions

4. Ethological Characters
   a courtship + other ethological isolating mechanisms
   b. Other behavior patterns

5. Geographical characters
   a. dist patterns
   b. Sympatric - allopatric relationships of population

Digitized by the Hunt Institute for Botanical Documentation

B 170
Lecture
Mr. Henry Fleming
April 13, 1966

You are doing a revision, aren't you Craig? Are you following the figures more or less in Davis and Heywood, or what model are you following? (Craig) I'm following Dr. Harrington's on the monkey. (Henry) I see. In a formal description of a new species, we are presumably asked to describe the species. Now how do we describe the species? Davis and Heywood, I think, call it the formal description. Let's write this down, I would rather use different terminology. They use such terms as formal description, then he follows that with diagnosis. When we describe a species, or any object for that matter, we do not describe it completely and in complete detail, because we could keep describing this book probably ad infinitum. We could use microscopes on it, electron microscopes, and if you start going into what the constitution of the paper is and all that. We'd never end describing that book. What we have to do is use comparative descriptions. So I'll use a different set of characters for describing a book than I would the ashtrays; I use a different set of describers describing algae than I would higher plants. We use information in other words that is comparable. So right from the beginning, all descriptions consist of a diagnosis. We make a decision as to what characters are important from a comparable point of view. If we want to classify or we want to separate a group of plants, we make a selection of those characters which are pertinent but differentiating them. We take any ordinary tree we can find, well practically all of them for instance, all of the Tsuga for instance will have green needles. There is no point in making a character of green needles for Tsuga. They all have them, it won't differentiate them for us. We have to pick out those characters which are separable and which show some differences, some comparison. In some cases you will find botanical descriptions, some fellow will describe the species and they will go on for two or three pages. Actually, the whole description can be put into one paragraph.

Digitized by the Hunt Institute for Botanical Documentation

He is just not capable of deciding which characters are of comparable value. (Craig) If I go out East and find a new species of Tsuga, don't I have to put in some character description in my description which places that plant in the Tsuga? (Henry) That's right. Well, I'm talking of the species. If we are going into the genetic level, then there are a set of genera characters. They have been pretty well laid down for us. We have got some sort of a cluster of plants that belong in Tsuga. (Craig) My point is, if you collect the plant the first time don't you have to justify placing it in Tsuga instead of Pseudo Tsuga? (Henry) That's right. (Craig) Then I don't understand why you couldn't use a character green needles, something similar that all the other members of Tsuga would have. (Henry) Then we would be in with the genera character not with the species character. (Craig) You differentiate between the description? (Henry) That's right. In the higher up we go. If we want to describe a family completely, we would have to describe every single species and every single genus that we put in that family. That would be a ridiculous description. We would have to read volumes and volumes of information that wouldn't really be pertinent to finding . We select out those characters which are comparable now between families or between genera. (Craig) When you speak of description, are you talking about the classical or Latin description? (Henry) The classical Latin description. And that I would call a diagnosis because we had to diagnose what characters to use. Davis and Heywood call a diagnosis a comparison. I would call that the differential diagnosis where we usually take and try to place a particular species already used for the whatever you are dealing with to a usually more closely related or other related genera of species that one would be apt to use. Sometimes we pick something unrelated just for emphasis. So I would call this the differential diagnosis. In plants or in animals, we have got a number of types of characters and let's deal with those first which except under unusual conditions we will probably never

Digitized by the Hunt Institute for Botanical Documentation

want to include in an overall comparison of plants to determine their overall similarity. First of all, we have non-inherited variations and which we can think of - progressive individual variation would be a subtitle there. The age of a plant would be one sort of consideration. If we are taking a group of species, we don't want to compare seedlings with mature plants. The characters that we would find in seedlings, the leaves are apt to be a different size for instance. It will not be in flower and we won't have enough information on that. We could however, if we wanted to make a classification that would be a specialized classification of seedlings, it might have some advantage to a man, for instance, in the nursery business to make a classification just on seedling plants--might be real handy to have such a classification. That would not be a general botanical classification. We would want to compare those characters on plants of all, at least the mature plants somewhat near the same age. We also have seasonal variation. Usually, you get seasonal variation, for instance, in algae. You do in some plants, we can think of for instance deciduous trees outside in the winter not having any leaves on. We would hardly want to then, in an overall comparison to throw in plants that were without foliage in the winter and compare them with plants with foliage in the summer. In other words we want to collect the plants more or less at the same time so that they look somewhat similar. You run into the same thing in the tropics in the dry season and the wet season. Plants look quite different. Social variation you often get primarily in the insects. If you are doing a classification of ants, you are going to have to have some special biological knowledge in order to make the comparison, because with ants you will have soldier ants, you will have just plain foragers you will have queen ants. You have got to group them together and handle them as one because you have a unit here. You must compare your unit together. If you don't have this specialized knowledge, you are just going to have to use, let us say soldiers, and compare the soldiers for your classification. And your classification will be restricted to one form. And you have got to realize

Digitized by the Hunt Institute for Botanical Documentation

that your one form of, because you have used your one form of insect is going to be restricted and be of value just for that form. You lack a tremendous amount of information on your other forms. Then we have ecological variations. For a general classification again, is a good question whether we should make the habitat part of a character. Habitat is often exclusive information. (from student a question that I can't hear). (Henry's answer) Not actually part of the organism itself. Whether I place this book on this table or that table really has nothing much to do with the book. Whether the book, we might be able to say things because it is placed on different tables, placed on that table is for general use so that everybody in the lab here knows what publications are coming through. We may be able to infer some types of information because of where it is placed. But it is not part of the actual properties of that book as to where it is placed. (Jack) Seems to me like that is one of the most helpful things you can include, though. If this plant, for example, is never found in any other situation but a shady damp area, this should be a character that would be useful. (Henry) It is often the character that we use amplying special considerations to. It may infer, for instance, that this species does not hybridize, or is isolated at least. It is an isolating mechanism. And we use the information in that sense, not in the sense of overall similarity. I could pick out situations or think of situations where I might want to code in that sort of variation. Soil conditions might be important. (Student) It might not be morphologically similar, but/physiologically similar. (Henry) Then, I've got that down here too. I don't wish to go into that in too much detail at the moment. Physiological information is of great importance, provided you have enough of a , concerns all of the objects that you had to classify. Unfortunately, usually whether we are dealing with plants or animals, some physiologists have worked out in great detail the

Digitized by the Hunt Institute for Botanical Documentation

physiology of a certain phenomenon going on for one species, and there are 100 species in this genus so something like that we don't have any information in regard to the other 99 so it is not comparable information. But we keep that in the back of our minds and we can perhaps whether justifiably or unjustificably infer that maybe the others are the same. (Student) I think to go out on a limb and say that seeds tell me that plant distribution is valid as a key to the identification plant. (Henry) It is, but we are not talking about identification, we are talking about classification now. And it is real handy to know that a certain group of plants only exist in Australia. It limits our searching. We can use that definitely in a key for identification. (Student) But doesn't this fact that they are located only in one place imply that they differ from other plants in some marked way? (Henry) It does imply that. Well, we infer that as true. On some occasions we can take two quite different plants and bring them down to the low lands and may in the Arctic conditions and let us say that we can grow them under arctic or alpine conditions. Both of those plants may become cushion or dwarf plants, quite different than they are in the low lands, but both of them will react the same way, and I am sure that Craig here can give you a number of examples of that sort of situation on the . They are reacting to their environment. And if you are doing an ecological classification, then such characters are of importance to you.

Host determination variation, that is primarily one that mycologists get into in particular. And again that is a very good identification procedure, but it is not necessarily valid, unless you can have other information if it is a segregating information, then you may be able to use it. But in general, host variation is not important. It is nice that we find that a certain genus of fungi is always on a particular type of plant. It is real helpful for identification purposes, but it is not how we cluster them. It is not how we determine whether they are a nameable species unit. We also have another type of variation which I have down here as density. Sometimes plants that grow in great masses

Digitized by the Hunt Institute for Botanical Documentation

or you find them all in a small place will take on certain types of variations simply because they are over crowded and won't grow well. And they look aberrant, the leaves will be small, the flowers will be small, or there won't be many flowers and things of that nature will happen to them. And we have got to take that sort of thing into consideration very often in herbarium material. Another thing that happens to herbarium material that I might mention in passing is that if a botanist goes off and he sees a field of certain type plant, he usually picks the best looking plant, has the biggest leaves, has the biggest flowers, It may not be typical of the way the plant normally is, and you xxxx better take that into xxxxx consideration usually in selection of your characters.

There is a climatic variation. This has very little again to actually do with the overall similarities between a group of objects. It has to be taken into consideration that it is better to handle it as a separate piece of information. As we get into the program later on and we start clustering, very often you might notice that plants go in one after another in their geographical clime or a clime. But it is a separate judgment that you have to make. You have to correlate it yourself. It actually, while it has its affect upon the way the plant looks is not again an intrinsic character of that plant. It is something that is imposed on the plant from the outside. Then, of course, and xxxx sometimes these are xxxx difficult to decide on occasion, traumatic variations. You may have a plant for instance that has certain insects eating on it that quite radically alter it. Sometimes these form gall or witches brooms and which a tree that has broom on it can look quite different than another type of tree. Insect bites, the way for instance a caterpillar chews on a leaf can help very often in identification. The way certain caterpillars, for instance, leaf often determines what caterpillar you have. I have often purely kidded Dave Rogers that I could classify all of the Passiflora, I can't actually do all of them, but Passiflora in northeastern South America without using a plant character. I can do it on certain genus of butterfly, where they

Digitized by the Hunt Institute for Botanical Documentation

feed on the plant, how they feed on the plant. It wouldn't be a good classification of Passiflora, but nonetheless from an identification procedure and let us say for an entomologist running around and trying to identify eggs or caterpillars on the plant, it ~~xxx~~ would work, but it would not be classifying the genus types of flora. You get into some interesting prob lems even looking at it from that point of view. In northeaster South America, I have never been able to find any of these Heliconius butterflies on the common economic Passiflora edulis. Edulis has been spread pretty well around the world, but I think ~~thx~~ originally it was indigenous to Brazil. In south Brazil, you will find one species of Heliconius on P. edulis. It is also a different subspecies. Whether that is nice taxonomic correlation, whether it has anything to do with being a subspecies or not is something else again. But you can go all over into the southern United States and you won't find the Heliconius feeding on it. But down in the indigenous area there is one species. On the hand I can classify probably the butterfly or at least rather not classify but write a key to the butterfly just using the plant characters. And if you can determine the plant characters, you will know what insect is feeding ~~xxxxxxxx~~ on it. But it is not a classification of either the insect, we are using extrinsic information there even though the correlation is very good. And I can say that if you find a caterpillar on the genus Passiflora, it is going to be a Heliconius, but I cannot say if you have got a plant Passiflora with insect damage on that it is necessarily Heliconius. It may be a number of bugs and other things. And from those bug bites you can perhaps identify ~~thx~~ your Passiflora, but it is not ~~xx~~ part of the speciation of the genus Passiflora. (Student) But it seems to me that all of the extrinsic ~~xxxxxxxx~~ characteristics that we have been discussing here imply something intrinsic about a plant. Maybe not individually but at least the summation of a number of these characters imply something intrinsically different between plants, and it seems that this is what we are after. What is the, maybe it can be said in a philosophical way, what is the essence of the plant, and that is how we make

Digitized by the Hunt Institute for Botanical Documentation

our classification. What is it intrinsically, what are the intrinsic characters of this plant which make all of these plants related and unrelated and everything else? But I can't see how except thats purely arbitrary you can separate the intrinsic from the extrinsic. (Henry) It is in many ways arbitrary, just as it is in trying to classify a collection of books, but there are certain characters which are a part of the genetic makeup of that plant. All of sorts of accidents can happen to the plant otherwise. We do have to at some point make what is largely an arbitrary decision. Many of these things I have mentioned here, for instance, are perfectly good for special classification. If I want an ecological classification, I would use ecological data, but we have to set up some sort of classification for plants that is of use to a great number of different individuals. (Student) And this is not the special? (Henry) And this is not the special. (Student) You define special as being only for one group whereas general botanical classification is..... (Henry) Good for the greates possible number of different uses. In regard to your point you brought up, ~~it's perhaps a little~~ this is a little bit perhaps off the subject, but there is a certain insect which they were able to feed on certain plants in Czechoslovakia, they brought some of this same insect material to Harvard. At Harvard this insect would never get out of its larval stage. It would reach an adult stage which was non-sexual with many larval characters. There was considerable amount of time spent trying to determine why this insect when brought over to the United States under the same laboratory conditions the same, would not mature become sexually viable. It was finally tied down to the paper that was being used for the food that this insect fed on, and this paper was from Douglas fir. Apparently, all the paper that they tried whether it was magazine paper or anything else as I recall, all had some origin of Douglas fir. And there was something retained from the Douglas fir that was limiting this one insect. The other insects were raised fine in the same genus, just this one had a limiting effect this one insect. This is really not comparable information.

Digitized by the Hunt Institute for Botanical Documentation

It is highly interesting, and it is good in the biological information, but we would hardly use that type information in classifying that insect. We have certain types of inherited variation, sex associated variation. You get primary secondary sexual differences. They are usually true in animals. Alternating generations and higher plants, I don't think that is of any importance really, but if you get down dealing with algae why it becomes critical, and while I have this in the section that we probably wouldn't use it in classifying, we would in the algae, but we would have to get all of our data together on the generations so that at least we weren't trying to compare one stage of an alternating generation with another stage. We would have to have all of our information on all of the stages or else restrict ourselves to one stage, usually. There is a way we can handle that if we only lack some of the information on some of the species or some of the genera or some of the families depending upon what we are handling. But if we lack it on a great number of classification, the final analysis, except for what we can infer, then the inferences will never be sure, will have to be based on those characters in which we are able to supply information to.

Let us go on and deal with what characters we can use in a classification, and these, some of these at least will be handled in the fourth chapter of Davis and Heywood which I suggest you read. While I am mostly familiar with American tropical botany, in American tropical botany youare practically if not solely restricted to external morphology. You know so little else. And that, of course, is primarily all plants and animals basically are classified, just because we lack other types of information. And it is the obvious thing, If you take for instance this here its almost completely morphological information. Do you have other types of information, Craig, in the group you are working with? (Student) I hate to be the dead horse here, but it seems to me that morphology is at least xx2 as variable as most of the other things that we have indicated here. (Henry) Let us go on wkix with what I, at least in my own

Digitized by the Hunt Institute for Botanical Documentation

private way and I think you are finally coming to agree with it, does constitute good character. At least they are not varying in the same way, because of external influences or special types of influences. Another type of morphological character that we have that is good are special structures. You will find these in fungi, we find them in higher plants, some groups you are really restricted to separating them out on structures which are rather private to the group you are working with. Don't use another group of plants whatsoever.

On some groups at least, especially if you get up in the temperate zone, we can do a good deal with internal morphology. Embryological characters when we have them can be thought of as morphological. And a good deal of cytological information gets into morphology , and also about as variable as any other character. Of course what we look for in selecting characters is try to find those characters which do not vary greatly, individually at least. But constant variations are good and we are looking for constant variations, of course, between organisms. You have to be just as careful, for instance genetical material as you do looking at leaves. This is often brought out in some of the A very slight change the chromosome sometimes produces tremendous phenotypic effect, sufficient enough so that they wouldn't even trace the let alone in the same species you have been dealing with.

Then we have physiological characters. Here again usually our information is rather limited to a few species that somebody has worked on, and from the few species we make the extension that is probably true of all of the species, but this doesn't by any means necessarily hold. We use metabolic factors, serological in many animals, protein and any sort of biochemical differences. Rutgers set up a bank of serological material. Sometimes the correlation between blood types and established taxonomy is quite good. Just as frequently it didn't, or I say just as frequently, I know some it just simply didn't work out at all. There was so much variation of that any organisms tested that serological material was valueless from a taxonomic point of

Digitized by the Hunt Institute for Botanical Documentation

You could on the other hand if you wanted to set up a classification just on serology, but it would not have been an overall classification, and you still can include it in a taxonomic analysis, but it is not going to correlate in many instances very well with your data. Genetic sterility factors can also be used of course as a physiological character. Now we have ecological characters. You notice before I put them down as one type of variation that couldn't be used, Well under certain conditions, you can use them. It again depends upon your problem. Habit and hosts for instance are really often quite important characters. Type of soil a plant grows on can be an important character. Seasonal variations can be used as separaters very often. There again if you get American tropical plants you are not apt to know very much about that, typically if you are working on herbarium material you don't know anything. Though sometimes if you have enough herbarium you can make inferences from it as to what, you know notice the dates on the specimen, when it was collected, and if you know about the climate or the area that these plants have been collected in, you can make inferences as to what happens in the seasonal .

Parasites can sometimes be used. They tried using parasites for instance in bird taxonomy and as with many characters there is really no . It worked with some groups and didn't work in other groups. Here again on the parasites the taxonomy was set up by the morphology, morphological characters seem to be good, they hung together, they hung together with a good number of other things treating the beasts in isolation and just working with the beasts without their relationships, ecological relationships to the host. So in many instances the host thought it was disregarded. But it is nice from an identification point of view if you can pick up a say a mite on a bird and look at it and say, "Oh, well that means this bird"--some identification procedure very often but it is not a classificatory procedure.

Now ethnological characters, of course, work primarily in animals again. I can say for instance on fiddler crabs that if you watch the courtship of two

Digitized by the Hunt Institute for Botanical Documentation

fiddler crab on a beach some place, you can use their name of specie. You can do it with a good ;number of insects; you can do it with a good number of birds; you can do it with a good number of lizards. Unfortuantely, here again usually you don't have comparable information because something is known about something, is known about some of the specie and usually a small number of specie but not known about all the specie. So you bring this information usually in after you have classified. You may readjust your classification on the basis of some of this, but you might just as well call lack of information or specialized knowledge. I cannot think of any botanical example which might involve courtship. Well in a sense, I can too. You can think of isolating mechanisms. Mike Worth who was with us in New York, well it wasn't Mike Worth, actually ;it was Dressler who is not in Brazil who was working on a certain group of orchids in Central America, I think it was Costa Rica. There are two species, well it was on the two species of Oncidium which he had collected a good number of years ago. They were on opposite sides of a mountain. He went down maybe it was four or five years ago and this mountain had been cut down, the forests on this mountain had been cut down and he started finding that at the edge of the forest, on the ridge of the mountain, some orchids that had been undescribed. The more he looked at them, the more he realized that what had happened here was he was getting hybrid between two Oncidium of the east side of this mountain and all of a sudden they were hybridizing. He had never found any bybrids before. The explanation was that two different species of bees, or one species of bees rather, was a pollinator here, and formerly the population at one side of the mountain never got to the other side of the mountain, but with the forest cut down the bees started spreading and one population on one side of the mountain went to the other side and carried the pollin from the orchids on one side over the mountain to the orchids on the other side and that is now you were getting hybrids. That is an explanation of what happened. Again, it can be thought of as a good character.

Digitized by the Hunt Institute for Botanical Documentation

Distribution patterns can be used as characters. Then you can fit in what you know as characters on their sympatric or allopatric relationships. But in general, with most plants we are going to be restricted, I am afraid, to morphological information, and our general classification will be largely morphological in nature. If we were to classify these top windows in this building, there would hardly be any point in dealing with the size of the windows, they are all about the same size; dealing with the patterns as far as their having one type of glass on the edge because they are all like that. We wouls start by looking at the colors of the glass on the edges, and we could develop a classification from them. Some of them are green, some of them are ;sort of bluish-green....(Student) You have to start with some designation as to relate them from being different from the lower ones. (Henry) Yes, if you want to deal with them all at once, that would be our first step in the classification. I was dealing with the top ones because I saw they are all, they was some slight difference..(Student interruption) (Henry) But by working with the color there and perhaps a number of colors, we could isolate or sort out these windows and classify. There would be some windows for instance with all green edge glass, some mixed with two colors. And I see one over here a different kind of spotty green. And we would classify these, and in this instance what we would be doing is taking comparable information, things that we could compare between these windows, but not characters which are true of all the windows that would not separate the windows out. What could you say about the history of these windows is something else again. You could think of that in perhaps a phylogenetic fashion, and I think it would be just as tenuous as phylogenetic classifications often are. Which are the original windows, and which are the windows that have been repaired. We could probably make a, do it in statistics and say, "Well due to the fact that the glass in that window is different than all the rest, it is likely that has been repaired." But what are we going to do between these green and blue windows? Which is the older? And perhaps they

Digitized by the Hunt Institute for Botanical Documentation

were put that way originally, I don't think so but they were very apt to be, and it is a probability that we have to deal with there. But in dealing with plants, while this is a subcase, dealing with plants is not too much different. Here we are dealing with morphological information in regard to the windows. I think it is getting time to stop the class. Are there any questions?

Digitized by the Hunt Institute for Botanical Documentation

Hippocrates — Hippocrates
Aristotle — 322 BC. genus rules
Brunfels - Bauhin.
John Ray. 1627-1705
Class — identity

Similar things together — presence inf.
predict.

Species not denatural
evolution amphibians - peripatus

peptide ornith, entr. malacology

De Vries Ornithes
Johannes Commor Bem. Selfpol. Hemizygous

Morgan — Drosophila.
Bauer. Autochthonous Drosophila
Fisher.

Now led into a period when many chem disciplines are Biochemical Dep. ecology, cytology physiology, dep. genet.
However some of these disciplines are

Digitized by the Hunt Institute for Botanical Documentation

White Correlated with taxonomic units are extrinsic rather than intrinsic – are also important in respect to the purpose of the classification

general — specific

↓

natural class. reflects the situation as it exist in nature

phylogenies – Theori + hypothesis from above.

Order.
Family
Genus
Spec
↓
specie

Class.
sub class.
Superorder.
sub order.
Super family.
sub family.
↓

Intra

Intrinsic – within to follow – sequi
following inward – lying in the thing itself
real – res – belonging to the very thing
genesis – to bring forth (gignere) spring out
Nature – born – arising from

Digitized by the Hunt Institute for Botanical Documentation

real value artificial value.

Intrinsic value of a book is the price it will get when sold in a regular way.
Extrinsic value, as being the gift of a friend or certain edition or a particular type

Real value is its type & binding – fineness of paper. impropriety in selection of contents

Extraneous, Extrinsic, foreign.

out of hand – not belonging to it

Extrinsic – outward, external

Extraneous is that which forms no necessary or natural part.

Extrinsic is that which forms a part or has a connection, but only in an incidental form. It is not an inherent or component part.

A work is said to contain extraneous matter which contains much matter not necessary belongs to or illustrates the subject

Digitized by the Hunt Institute for Botanical Documentation

A book has extrinsic value when it borrows its value from local circumstances as opposed to intrinsic merit or that which lies in its contents.

Intrinsic Characters are inherited characters
Extrinsic characters are ecological, traumatic geographical
Many extrinsic characters are those that after we have assembled a group, cluster assemblage of specimens, objects, indicate ~~reproductive~~ isolation.

1. morpho + physio diff
2. reproductive iso.
3. ecolog. diff.

Digitized by the Hunt Institute for Botanical Documentation

Biological Clusters

B 170 Lect Apr 27.

1. "Natural" groups – easily ~~defined~~ recognized
   vs
   "Artificial" groups "Widely divergent groups"
   "Good" genera vs narrowly divergent
   vs
   Variable, or unnatural
2. Species raised from subspecific to specific rank, or "reduced" from species → ssp.
3. Grounds for such work –
   Some new biological info, or reinterpretation of older data –
4. These decisions made on grounds of biological info – vs those based on some legal procedure – nomenclatural. Reasons for "reduced to synonymy".
5. The terms morpho species, or taxonomic species, biological species, ecospecies, and many others exist –

Digitized by the Hunt Institute for Botanical Documentation

Historical development of ideas on biological clusters -

1. Early form classifications
   Ex, Clay Herb
   Satisfactory when in certain areas
2. Later developments of families -
   1 Cesalpinia Legumes,
   2 Families ending in ae, (initial of access) indicate "natural", easily delimited groups.
3. Linnaeus - and the "sexual" system, deemphasized the family, making smaller groups - genera - the major emphasis -
4. Following L., French + Swiss Botanists (Jussieu, DeCandolle) Engler Engler Adanson, reemphasized families as most imp. classificatory units.
5. Darwin's evolutionary studies again deemphasized the family unit, and reemphasized genera and lower categories -
6. Influence of genetic studies commencing in early 1900 - eventually brought people down to work on populations, cf - New Systematics.
7. This led to species definitions called "biological species", to differentiate from "guess, taxonomic species -
   — but —
   read Simpsons definition from Science

Digitized by the Hunt Institute for Botanical Documentation

and- show illustration, p. 444
in D+H-

8. Anderson's work, showing that taxonomists from abroad, could take an unknown group, and discover same categories, though disagreed on the level of category -

9. All these have one thing commonly, - they are talking about some sort of grouping with discontinuities between the groups, and the taxonomist searches his data to prove that the gaps exist, and throws out the data that do not exhibit the desired discontinuities.

10. The basic search is for some sort of biological (i.e. part of the organism) feature to do the separation, but very frequently, we use extrinsic information to give us these gaps may discovered upon a geographical basis - N. Am., or S. Am., western vs. eastern, mountainous or plains, or upon ecological, edaphic, geological, moisture, climate, etc.
or on bat history (evolutionary)

Digitized by the Hunt Institute for Botanical Documentation

B 170
Dr. David J. Rogers
April 27, 1966

BIOLOGICAL CLUSTERS

We decided to break George's discussion in half because he didn't get as far as he had hoped he might with the set of notes that were given out Monday. We do want to get this discussion in we are going to give today on biological clusters. George will come back to his discussions on Friday and finish up there.

What we are going to talk about in these biological clusters is perhaps one of the most difficult things that I can think of to really talk about. It is difficult in conceptualization of what we are going to say, and it is difficult because of the fact that historically there are people who have considered biological clusters, talked about them, waved their hands about them, and in ganeral muddied the water something awful. In reading biological or taxonomic output, you can find all kinds of words that if you stop and ask the question of yourself, "What do they mean here?" You have difficulty in finding any definition. The words such as natural, this is a natural group of organisms, come out, versus such terms as, this is an artificial group. For artificial you might read other such terms which seem to have the same meaning as widely divergent groups versus narrowly divergent. You can get such terms as, or such expressions as, this is a good species or this is a variable species. You also run across in the literature that somebody has elevated a subspecies to a species or contrarywise someone has reduced a species to a variety. So these terms are obviously something connected with biological clusters, I think, and our purpose today is not to talk about definitions of species, definitions of genera, various other hierarchical categories, but just to raise some of the problems and lay bare before you some of the difficulties involved.

Now I would like to ask you to play taxonomy with these dots that I want to put on the board. Who wants to play taxonomy with this bunch of dots

Digitized by the Hunt Institute for Botanical Documentation

on the blackboard for me? Thank you, Mr. Taxonomist. I want to assert here that if I ask Jack to go up there, he would disagree to some extent of what Bert has done. He would do something perhaps a bit different. Right? Is that so Jack? You would argue with Bert that you are off your nut. (Jack) No, I..... (Dave) Well, at any rate, what Bert has had no difficulty in doing is making a decision. What sort of a decision has he made? What has he really done when he is working at these sorts of things? If you ask any taxonomist, "Draw me a species, there is a species. That is not the case. He is going to do something like this. Now I can ask you the question because I have had my head battered against the wall on problems that you get into when you ask somebody to draw a species. When Tanimoto and I were first working together on this whole bit, he said, "What do you mean by a species, or what do you mean when you want to cluster things?" I said, "Well, you do this you want to put t ings in a package, like that." See? This is no sweat, everybody wants to do this. When you ask, "What are you going to do?" These are the kind of pictures you want to draw. Now ..... such a person as Tanimoto who is a mathematician and not a biologist, okay this is what Rogers thinks of as a species or groups of some sort. So what can anyone out there who is mathematically oriented assume except that these lines indeed have some sort of a meaning for what a taxonomist wants to do. So in looking at these things here, it would be meaningful would it not to make some sort of a model which indeed discover these nice round or approximately round structures. And it is indeed in the concept of most of us that if we draw things as Bert almost did, he did get away from it a little bit, draw pretty close to ~~xkx~~ circles. It is more circular than other wise, right? So a mathematician would look at this sort of thing and discover some model has to be found which will limit things and put them inside here. And you will probably find some text methodology which will indeed discover whether or not here in this mess or in this mess and you may go to some procedure whereby you will pick out an arbitrary

Digitized by the Hunt Institute for Botanical Documentation

9.1 center of this thing, except that any test that it makes will determine whether or not this object out here comes closer to the center than otherwise. We would not think too much about that if we just hadn't given it a great deal of thought, and indeed Henry and I came banging away at that type of model thinking that that was what we were looking for foraa very long time. And then we were fortunate enough to have to come to work for us a fellow, a taxonomist who said, "Okay this is great for many things, but suppose ;you had in addition to those nice round circles some cigar-shaped structure, thusly, in which the objects here and here should be this cluster not on this object here. Immediately, you see as a biologist, as a taxonomist. We all know, don't we, that clines exist in biological species. This is very simple. This sort of thing can represent a simple cline 11.9 where these objects at the end would indeed from the other but are connected by a series like this such that this object here should clearly be motivated to these objects than something of another cluster here. I think that probably what is meant when we are talking about these terms natural groups versus artificial groups that we have in mind something on this order, but something which, of course, indeed out there in space or those materials you have to deal with in the way of plants or animals are like this and they can be more complicated in having long appendices sticking out in various places like that which may be the case. Now, we will also get into a great deal of trouble.....(Bert) I hate to interrupt you here but I am confused as to what the shapes you are putting on the board really relate to. I see that a circular model implies that there are values or criteria and they are related in such a way that some are closer to the center and some are farther away. But now this is purely an abstraction, but I don't see how this same idea relates to this other shape on the board. You mean in a topographic sense or geographic sense. (Dave) A topographic sense. (Bert) So you could superimpose these on maps somewhere and the closest would tend to be round

Digitized by the Hunt Institute for Botanical Documentation

if the model was good over here and if this model isn't good, they would tend to go like that. (Dave) I will not really at this point differentiate, it can be done, this kind of drawing can be done for a sake of representation whether you are indeed are talking about topographic or geographic, it doesn't really matter. (Student) What does the shape have to do with it? Aren't you saying on that elongated, the one with the lump on the end is still more similar to the ovals than they are to that other group that is above? Is that what you are saying? (Dave) Probably. What I am also saying is this, and in a very generalized sense when we make our envisioning of things our illustrative types of things, this is the sort of lines one draws intuitively. Perhaps these are not illustrations that are worth anything at all, but these types of lines are indeed to be found in many, many taxonomic outputs you see. (Henry) Let me make one comment, it is true I believe in general that if any clusters that are within those circles each of the objects in there are most similar to each other within the circle than they are with anything outside that circle. And this instance over here the extreme objects are that more similar to the center object than it is necessarily to something in a different circle. In other words, some of those objects are futher apart from one another than they are from something that is outside of the cluster. (Dave) Whatever the situation may be, I think that we assume here some sort of a similarity, an undefined similarity is what we are trying to attempt to do here. Perhaps there is something else that goes along with it, too. Now let me back up just a little bit and see how an historical development helps us out to try to explain some of the things that I want to get across. I don't know that this will immediately become apparent in this connection, but I hope that what we will be doing will be laying the ground work for some of our later conceptualizations. In our historcial development, let us start pretty far back by about as far as we have any kind of record of people having made classifications,

Digitized by the Hunt Institute for Botanical Documentation

where within a limited locality such as southern Greece, people made classifications and their classifications would be on life form essentially to start with. That is, some of the earlier classifications divide plants on the basis of trees, shrubs, and herbs. And for that particular locality, such a classification would have some sort of merit. In other words, you could put a plant into some package or other. If you moved from that particular area, as frequently as people might have, and tried to use the same sort of classification, we might discover that we had some discontinuities that would fit into such a picture, but that we had considerable amount of difficulty in moving from one area to another because the objects to be put into the classification would obviously not be the same as the ones as you would have had say in southern Greece. However, that kind of a classificatory system apparently satisfied for quite a long time, if we think mmr of classification purely in terms of taxonomy. There are other kinds of classifications obviously that were made. For instance there were discoveries of food plants which is a kind of categorization. There were discoveries of medicinal plants that had some merit, perhaps there were classifications of poisonous plants that we would be able to use even in those early days, although they were not formalized as classifications. But these types of classifications did indeed exist very early. Only as we moved into the latter parts of the dark ages and so forth and indeed find people beginning to investigate things about them with more freedom, do we change this kind of a classificatory system. One of the very earliest people that was interested ;in classification are, at least made some descriptions which sound to us like normal things, was an Italian by the name of Cesalpino. I don't care whether you remember Mr. Cesalpino's name or not, but one of the things that Mr. Cesalpino did was to discover that there were some plants more like each other because of the characters/they had than that they were like other plants.

Digitized by the Hunt Institute for Botanical Documentation

Mr. Cesalpino was best known amongst taxonomists for the fact that he established a group of organisms which we now know and he used the same term or something similar to it. I don't know whether Mr. Cesalpino used that ending, Leguminosae as we know and they use today, was a group of organisms put together by Mr. Cesalpino on a different basis than had been done in the past. Other early authors were able to recognize such things as the compositae, these are terms I am sure that are in all of our vocabularies. We know approximately what these are. There are other families of plants recognized by us today which were very early established because of some inherent quality of the organisms. In the case of this particular group, it would certainly be the fruiting structure, this particular group would certainly be the total flowering mechanism that is the inflorescence structure, the inflorescence. We, incidentally, may recognize that in an historical development of things that there are in most of the plant families that we use today, the ending on the plant families is different from this. For instance let us take Onagraceae. This ending aceae is most typical of plant family names. What is recognized when we don't use that one on some of these such as the Compositae and Leguminosae is that very early these very striking groups of related organisms were put together, and only later in time did we get this in here and this is an historical development that these people decided they wanted to recognize that somebody earlier had indeed done something useful. And so they have maintained those two forms for the plant families. Now shortly after Cesalpino and such people had discovered the Compositae and the Leguminosae and others of this nature, Linnaeus came along and his first endeavors intended to make some sort of an identifying system, and he used the "sexual system" for his classificatory scheme published in 1734. What this sexual system did was to essentially take counts of things, counts of numbers of stamens, counts of numbers of petals, etc. Wherever he found flowers with the stamen numbers the same, he put them together, and I think he had a total

Digitized by the Hunt Institute for Botanical Documentation

of nine classes according to stamen numbers. Now immediately you do this sort of thing, this sort of sexual system, you will do damage to these, to these. And in order to recognize any kind of a category whatsoever, you will probably have to recognize something a little lower than a category. Again, if you look at Linnaeus' work, you will discover that his major emphasis the whole way through was on the genus level, a lower category than anything that you might have put together in here, in a family category. So if you read Linnaeus', as I read Linnaeus, you will discover he sort of wiped off this family idea completely, that he gets down to a lower level of the genus. Following Linnaeus, other people were not satisfied with that particular system, and a series of botanists like the French botanist Jussieu, I think there were two of them-- father and son. Such Swiss botanists as the De Candolles, the German botanist Eichler, the French botanist Adanson all thought that in so far as classification was concerned that Linnaeus was off track a little bit, and they went back and re-established the idea that the plant family is a major concern for a categorization of things. Now following these people, Darwin comes along with his evolutionary scheme and it seems that people, the taxonomists following Darwin, in efforts to make evolutinnary tracks, again were forced to leave aside the family category and move down into the hierarchical levels a bit, but again in general and/the botanists in the latter half of the 19th century indeed did that emphasize more frequently the genus as a category of classification than the family.

Another influence on the categorical level that we are most primarily concerned with was introduced by the studies of genetic mechanisms and of cytology. They were introduced, of course, by Mendel picked up again by DeVries and two other botanists at just about the turn of the 20th century. These influences in the studies of cytology and so forth obviously could not be worked or be

Digitized by the Hunt Institute for Botanical Documentation

studied as far as taxonomy is concerned on a very high categorical level. So again you come down from even the genus level and people began to have considerable amount of studying, interests im species. Now I am pretty sure as far back as Linnaeus and even earlier there were species considerations but it was never considered to be of particular concern. The category of critical importance would either be at the genus or at the family level than stage. The combination I would guess of the influence of the cytological and genetic studies as far as taxonomy is concerned came with the publication of the book called "The New Systematics" by Huxley about 1940. I think that was about the date. These indeed did focus attention down to a much more "fundamental biological level," and people began to worry from about the time of the beginning of the 20th century on about this category, the species, and the "species problem." has been battered around by groups almost on a clockwork basis. You can figure on there being more critical attention to it, say in one decade than the next, but it is still a real problem. Today we ;have other kinds of definitions for biological species for species, and I just happened to pick up in this last issue of Science a paper by George Gaylord Simpson on "The Biological Nature of Man". And in this paper he gets around to talking about systematics of man and the first thing he has to do is to come along and make a definition of a species. He says: "A biological species is an evolutionary unit composed of tontinuing populations that regularly interchange genes by interbreeding and that do not or cannot have such regular interchange with other species." Those of you who would get this definition may do so. There is a footnote attached to this definition in this article which I think is instructive to us which I shall read. He says in ;the footnote: Age-long agrument on the definition of species is perhaps sufficiently summarized in [illegible]

Digitized by the Hunt Institute for Botanical Documentation

an earlier book by G. G. Simpson. Now along with Simpson who is, I guess, about as modern as we can get in taxonomy, I have another little thing to show you here which I think has something to do with what I am eventually going to try to put across. I don't know how many of you have your textbooks here, but of those of you who do, can look on page 444. On page 444 is reproduced a diagram taken from the interesting work of ~~Kkxxx~~ Clausen, Keck & Hiesey done in 1941. The interesting thing to note there are those drawings first,;and lets not worry about what the informatinn is. It is about a group of Western Compositae species in the genus Layia. But note there the gray-hatched things are indeed drawn as I suggested that most taxonomists will draw things. Imposed upon this, of course, are several other interesting features which deal with the crossability of these organisms. The black heavy lines there in showing this. But note farther that there are a series of dotted lines sort of being radii for the total circle there. These dotted lines being drawn in at points where you can see that the canals between the grayish circles seem to be the smallest canals, and that there are smaller extensions made therein. Now what we are intending here,I think, from these illustrations, the description of a species by Simpson in his paper and what indeed has been the case for all of the botanists, taxonomists, biological systematists throughout all these times is trying to make some discovery which indicates relatedness and not relatedness. That in all of the cases no matter what type of information is employed, the people herein are talking amut some sort of grouping with discontinuities between the groups. For example, in looking back again at Simpson's definition, can we reread that a little bit? "A biological species is an evolutionary unit composed of continuing populations that regularly interchange genes by interbreeding and that do not or cannot have such regular interchange with other species." Those terms do not and cannot obviously may be reread to some extent that there is a gap between this

Digitized by the Hunt Institute for Botanical Documentation

particular group of organisms and some other group of organisms. That if we look at the illustration on page 444 again, that the taxonomist has made a decision therein based upon some discontinuity. The discontinutiy is better informed as far as Layia is concerned because they have attempted to put in information about the shape and the size of the plants, that is the morphological information, and they have attempted as well to incorporate therewith information about the chromosomes of the organism, and not only about the chromosomes or the cytology but also something about their breeding patterns. You will discover that their major discontinuities are between those things that do not have the opportunity to hybridize quite so well, or cross quite so well. This is a very nice pattern that is discovered in this group here, and I would suggest that anybody would be able to discover that if one talked about genus Layia that they would speak of this as being a good taxon, that is thatthey would have no difficulty in the discovery that there is indeed something with a gap between the clusters. The gap may be discovered in different ways, but I assure you that the purpose that the taxonomist has in mind is to discover the kinds of information which will indeed provide him with some sort of a gap, and the chances are good that he will stir around a very long time looking for that kind of information which will indeed give him that type of differentiation between groups of organisms. There is good likelihood that any group of organisms upon which he may wish to settle will have kinds of information in there which indeed will not separate his organisms into groups. It is quite surprising to note that he did not incorporate that data in his taxonomy. And what he is talking about when he says, "I have a good key character here or diagnostic character" means that he has selected a piece of informatinn which will do the job that he wishes for it to do, namely, to give me a bunch in one or a group and that he will not say anything about those other nasty pieces of information

Digitized by the Hunt Institute for Botanical Documentation

that didn't do the job for him. (Student) They didn't do the job all by themselves? (Dave) Right. Now it very frequently is the case that if you will read taxonomic works, you will discover that there is a discussion made after the man has made his classification. These discussions will be the introductory parts, where he will there tell you that it really is not as good and clean as my classification looks like as I have put it out here, but for my purposes this information is not used in the classification. Learn to read what these fellows really say is something that is not quite as easy as you might anticipate. Obviously in this situation the basic search is for some sort of biological feature that is in terms that Henry was using the other day, the intrinsic features of the organism, something about the plant or animal itself. That is, it is nice if we can discover on the plant itself that kind of information. However, when we attack a group of organisms where it is not clear to us which particular piece of information about the plant or animal will make a nice clean separation for you, we will very frequently look outside the plant and there is just nothing wrong with this as far as we recognize what we are doing. That is, you may discover that if you plot your plants out on a map, an outline map, one group for all in North America, another group ;in South America, it is obvious to look amongst those things which fall in the same continent together to discover whether there would be similarities there which would adequately separate these organisms from those in South America. Sometimes we have to go into a different direction in order to establish it. There are known cases, for example, species with disjoint ~~xxxx~~ ranges occurring, a number of plant families have representatives that occur in our own southwest and again do not show up until ~~xxx~~ a similar habitat is found down in northern Argentina in South America. So what we have attempted to do there is to find other basis by which we may or may not separate these organisms. We discover, for example that there are eastern versus western groups of organisms. Uvularia

Digitized by the Hunt Institute for Botanical Documentation

which is almost entirely confined to the eastern part of the United States is going to have something else, isn't it? We can't use the eastern versus western approach to this problem as far as the genus and the species within the genus Uvularia occurs. I am not saying that to throw you a curve on your coding of information, but I sug= gest to you that you will have to discover that if there be differences outside the plants which are useful to you, that you will have to discover something besides geography to do it. It is also obvious that plants have different preferences for growing in certain places which may give the outside type of information that will give us a clue as to how we can separate them. Certainly, soil preferences or perhaps in geological provinces or perhaps moisture, regimes, or total climate aspects, that is whether it be warm, temperate, cool, may have some effect upon how you go about the selectinn of information which will indeed put us out in this busi= ness of separating groups of organisms accirding to their dissimilarities or accord- ing to their similarities. I think perhaps/a that taxonomist gets himself really wrapped up in the problem of phylogeny and evolutinn when has attempted to discover these discontinuities. I may be wrong on this, but I suspect that one of the ways people have attempted to make gaps between things is discovering some evolutionary path- way which clearly will show them discontinuities between groups of organisms. There is no real need xx to say whether or not these groups of organisms that we are talking about are real or are not real. If we choose our information correctly, any group of organisms which we have established as a species may be shown indeed not to be separate from anything else, if we choose our information in the right way. It can be shown to have continuous variation with something, if we choose our information correctly, or incorrectly, I don't know which way to say it here. (Bert) That is if you say that a small bit of information can be used to determine difference? If you allow that? (Dave) Yes. In other words, the information which may show us a gap between plants may, is selected primarily for that particular

Digitized by the Hunt Institute for Botanical Documentation

purpose, to show us the gaps. Whereas another feature of the organism may definitely indicate that there is a relationship t others that makes it such that you can't distinguish them. But when we get down to this sort of level, it is clear what our purposes are in taxonomy, is it not? That for the sake of conveniences and discusssions for me to tell you about these organisms here, that I have to discover some piece of information which indeed will allow us to put them into packages of sizes or clusters, or something of this nature such that we can adequately talk about them. I think that this is primarily our goal in taxonomy. If there is a particular goal, I would suggest that this is about as close to it, the totality of the goal as we can really make for the whole ;thing. I said at the beginning that I have no intention whatsoever of talking about the definitions of different categories of organisms, I think we are a long ways from doing anything like that. Maybe we are not as far away as we think, but at least as far as from where I stand, it looks like we are a long ways away from making an adequate species definition, But all I can clearly say is this: that we must recognize that when we are talking about categories like this, these are man-made constructs. Unless we define them on that basis alone, we are going to continuously find ourselves as people have for the last half century have found themselves, in hot water when they have tried to make a species definition.

Digitized by the Hunt Institute for Botanical Documentation

B 170 Lecture
Dr. David J. Rogers
April 27, 1966

BIOLOGICAL CLUSTERS

We decided to break George's discussion in half because he didn't get as far as he had hoped he might with the set of notes that were given out Monday. We do want to get this discussion in we are going to give today on biological clusters. George will come back to his discussions on Friday and finish up there.

What we are going to talk about in these biological clusters is perhaps one of the most difficult things that I can think of to really talk about. It is difficult in conceptualization of what we are going to say, and it is difficult because of the fact that historically there are people who have considered biological clusters, talked about them, waved their hands about them, and in general muddied the water something awful. In reading biological or taxonomic output, you can find all kinds of words that if you stop and ask the question of yourself, "What do they mean here?" You have difficulty in finding any definition. The words such as natural, this is a natural group of organisms, come out, versus such terms xxxxx as, this is an artificial group. For artificial you might read other such terms which seem to have the same meaning as widely divergent groups versus narrowly divergent. You can get such terms as, or such expressions as, this is a good species or this is a variable species. You also run across in the literature that somebody has elevated a subspecies to a species or contrarywise someone has reduced a species to a variety. So these terms are obviously something connected with biological clusters, I think, and our purpose today is not to talk about definitions of species,definitions of genera, various other hierarchical categories, but just to raise some of the problems and lay bare before you some of the difficulties involved.

Now I would like to ask you to play taxonomy with these dots that I want to put on the board. Who wants to play taxonomy with this bunch of dots

Digitized by the Hunt Institute for Botanical Documentation

on the blackboard for me? Thank you Mr. Taxonomist. I want to assert here that if I ask Jack to go up there, he would disagree to some extent of what Bert has done. He would do something perhaps a bit different. Right? Is that so Jack? You would argue with Bert that you are off your nut. (Jack) No, I..... (Dave) Well at any rate, what Bert has had no difficulty in doing is making a dec ision. What sort of a decisinn has he made? What has he really done when he is working xxxx at these sorts of things? If you ask any taxonomist, "Draw me a species, there is a species. That is not the case. He is going to do something like this. Now I can ask you the question because I have had my head battered against the wall on problems that you get into when yuu xxxx ask somebody to draw a species. When Tanimoto and I were first working together on this whole bit, he said, "What do you mean by a species, or what do you mean when you want to cluster things?" I said, "Well, you do this you want to put things in a package, like that." See? This is no sweat everybody wants to do this. When you ask, " What are you going to do?" These are the kind of pictures you want to draw. Now .... such a person as Tanimoto who is a mathematician and not a biologist, okay this is what Rogers thinks of as a species or groups of some sort. So what can anyone out there who is mathematically oriented assume xxxx except that these lines indeed have some sort of a meaning for what a taxonomist wants to do. So in looking at these things here, it would be meaningful would it not to make some sort of a model which indeed discover these nice round or approximately round structures. And it is indeed in the concept of most of us that if we draw things as Bert almost did, he did get away from it a little bit, draw pretty close to circles. It is more circular than otherwise, right? So a mathematician would look at this sort of thing and discover some model has to be found which will limit things and put them inside here. And you will probably find some text methodology which will indeed discover whether or not here in this mess or in this mess and you may go to some procedure whereby you will pick xx out an arbitrary

Digitized by the Hunt Institute for Botanical Documentation

center of this thing, except that any test that it makes will determine whether or not this object out here comes closer to the center than otherwise. We would not think too much about that if we just hadn't given it a great deal of thought, and indeed Henry and I came banging away at that type of model thinking that that was what we were looking for foraa very long time. And then we were fortunate enough to have to come to work for us a fellow, a taxonomist who said, "Okay this is great for many things, but suppose ;you had in addition to those nice round circles some cigar-shaped structure, thusly, in which the objects here and here should be this cluster not on this object here. Immediately, you see as a biologist, as a taxonomist. We all know, don't we, that clines exist in biological species. This is very simple. This sort of thing can represent a simple cline where these objects at the end would indeed from the other but are connected by a series like this such that this object here should clearly be motivated to these objects than something of another cluster here. I think that probably what is meant when we are talking about these terms natural groups versus artificial groups that we have in mind something on this order, but something which, of course, indeed out there in space or those materials you have to deal with in the way of plants or animals are like this and they can be more complicated in having long appendicles/sticking out in various places like that which may be the case. Now, we will also get into a great deal of trouble.....(Bert) I hate to interrupt you here but I am confused as to what the shapes you are putting on the board really relate to. I see that a circular model implies that there are values or criteria and they are related in such a way that some are closer to the center and some are farther away. But now this is purely an abstraction, but I don't see how this same idea relates to this other shape on the board. You mean in a topographic sense or geographic sense. (Dave) A topographic sense. (Bert) So you could superimpose these on maps somewhere and the closest would tend to be round

Digitized by the Hunt Institute for Botanical Documentation

if the model was good over here and if this model isn't good, they would tend to go like that. (Dave) I will not really at this point differentiate, it can be done, this kind of drawing can be done for a sake of representation whether you are indeed are talking about topographic or geographic, it doesn't really matter. (Student) What does the shape have to do with it? Aren't you saying on that elongated, the one with the lump on the end is still more similar to the ovals than they are to that other group that is above? Is that what you are saying? (Dave) Probably. What I am also saying is this, and in a very generalized sense when we make our envisioning of things our illustrative types of things, this is the sort of lines one draws intuitively. Perhaps these are not illustrations that are worth anything at all, but these types of lines are indeed to be found in many, many taxonomic outputs you see. (Henry) Let me make one comment, it is true I believe in general that if any clusters that are within those circles each of the objects in there are most similar to each other within the circle than they are with anything outside that circle. And this instance over here the extreme objects are that more similar to the center object than it is necessarily to something in a different circle. In other words, some of those objects are futher apart from one another than they are from something that is outside of the cluster. (Dave) Whatever the situation may be, I think that we assume here some sort of a similarity, an undefined similarity is what we are trying to attempt to do here. Perhaps there is something else that goes along with it, too. Now let me back up just a little bit and see how an historical development helps us out to try to explain some of the things that I want to get across. I don't know that this will immediately become apparent in this connection, but I hope that what we will be doing will be laying the ground work for some of our later conceptualizations. In our historcial development, let us start pretty far back by about as far as we have any kind of record of people having made classifications,

Digitized by the Hunt Institute for Botanical Documentation

where within a limited locality such as southern Greece, people made classifications and their classifications would be on life form essentially to start with. That is, some of the earlier classifications divide plants on the basis of trees, shrubs, and herbs. And for that particular locality, such a classification would have some sort of merit. In other words, you could put a plant into some package or other. If you moved from that particular area, as frequently as people might have, and tried to use the same sort of classification, we might discover that we had some discontinuities that would fit into such a picture, but that we had considerable amount of difficulty in moving from one area to another because the objects to be put into the classification would obviously not be the same as the ones as you would have had say in southern Greece. However, that kind of a classificatory system apparently satisfied for quite a long time, if we think of classification purely in terms of taxonomy. There are other kinds of classifications obviously that were made. For instance there were discoveries of food plants which is a kind of categorization. There were discoveries of medicinal plants that had some merit, perhaps there were classifications of poisonous plants that we would be able to use even in those early days, although they were not formalized as classifications. But these types of classifications did indeed exist very early. Only as we moved into the latter parts of the dark ages and so forth and indeed find people beginning to investigate things about them with more freedom, do we change this kind of a classificatory system. One of the very earliest people that was interested in classification are, at least made some descriptions which sound to us like normal things, was an Italian by the name of Cesalpino. I don't care whether you remember Mr. Cesalpino's name or not, but one of the things that Mr. Cesalpino did was to discover that there were some plants more like each other because of the characters that they had than they were like other plants.

Digitized by the Hunt Institute for Botanical Documentation

Mr. Cesalpino was best known amongst taxonomists for the fact that he established a group of organisms which we now know and he used the same term or something similar to it. I don't know whether Mr. Cesalpino used that ending, Leguminosae as we know and they use today, was a group of organisms put together by Mr. Cesalpino on a different basis than had been done in the past. Other early authors were able to recognize such things as the compositae, these are terms I am sure that are in all of our vocabularies. We know approximately what these are. There are other families of plants recognized by us today which were very early established because of some inherent quality of the organisms. In the case of this particular group, it would certainly be the fruiting structure, this particular group would certainly be the total flowering mechanism that is the inflorescence structure, the inflorescence. We, incidentally, may recognize that in an historical development of things that there are in most of the plant families that we use today, the ending on the plant families is different from this. For instance let us take Onagraceae. This ending aceae is most typical of plant family names. What is recognized when we don't use that one on some of these such as the Compositae and Leguminosae is that very early these very striking groups of related organisms were put together, and only later in time did we get this in here and this is an historical development that these people decided they wanted to recognize that somebody earlier had indeed done something useful. And so they have maintained those two forms for the plant families. Now shortly after Cesalpino and such people had discovered the Compositae and the Leguminosae and others of this nature, Linnaeus came along and his first endeavors intended to make some sort of an identifying system, and he used the "sexual system" for his classificatory scheme published in 1734. What this sexual system did was to essentially take counts of things, counts of numbers of stamens, counts of numbers of petals, etc. Wherever he found flowers with the stamen numbers the same, he put them together, and I think he had total

Digitized by the Hunt Institute for Botanical Documentation

of nine classes according to stamen numbers. Now immediately you do this sort of thing, this sort of sexual system, you will do damage to these, to these. And in order to recognize any kind of a category whatsoever, you will probably have to recognize something a little lower than a category. Again, if you look at Linnaeus' work, you will discover that his major emphasis the whole way through was on the genus level, a lower category than anything that you might have put together in here, in a family category. So if you read Linnaeus', as I read Linnaeus, you will discover he sort of wiped off this family idea completely, that he gets down to a lower level of the genus. Following Linnaeus, other people were not satisfied with that particular system, and a series of botanists like the French botanist Jussieu, I think there were two of them--father and son. Such Swiss botanists as the De Candolles, the German botanist Eichler, the French botanist Adanson all thought that in so far as classification was concerned that Linnaeus was off track a little bit, and they went back and re-established the idea that the plant family is a major concern for a categorization of things. Now following these people, Darwin comes along with his evolutionary scheme and it seems that people, the taxonomists following Darwin, in efforts to make evolutinnary tracks, again were forced to leave aside the family category and move down into the hierarchical levels a bit, but again in general and/the botanists in the latter half of the 19th century indeed did that emphasize more frequently the genus as a category of classification than the family.

Another influence on the categorical level that we are most primarily concerned with was introduced by the studies of genetic mechanisms and of cytology. They were introduced, of course, by Mendel picked up again by DeVries and two other botanists at just about the turn of the 20th century. These influences in the studies of cytology and so forth obviously could not be worked or be

Digitized by the Hunt Institute for Botanical Documentation

studied as far as taxonomy is concerned on a very high categorical level. So again you come down from even the genus level and people began to have considerable amount of studying, interests in species. Now I am pretty sure as far back as Linnaeus and even earlier there were species considerations but it was never considered to be of particular concern. The category of critical importance would either be at the genus or at the family level than stage. The combination I would guess of the influence of the cytological and genetic studies as far as taxonomy is concerned came with the publication of the book called "The New Systematics" by Huxley about 1940. I think that was about the date. These indeed did focus attention down to a much more "fundamental biological level," and people began to worry from about the time of the beginning of the 20th century on about this category, the species, and the "species problem." has been battered around by groups almost on a clockwork basis. You can figure on there being more critical attention to it, say in one decade than the next, but it is still a real problem. Today we ;have other kinds of definitions for biological species for species, and I just happened to pick up in this last issue of Science a paper by George Gaylord Simpson on "The Biological Nature of Man". And in this paper he gets around to talking about systematics of man and the first thing he has to do is to come along and make a definition of a species. He says: "A biological species is an evolutionary unit composed of tontinuing populations that regularly interchange genes by interbreeding and that do not or cannot have such regular interchange with other species." Those of you who would get this definition may do so. There is a footnote attached to this definition in this article which I think is instructive to us which I shall read. He says in ;the footnote: Age-long agrument on the definition of species is perhaps sufficiently summarized in

Digitized by the Hunt Institute for Botanical Documentation

an earlier book by G. G. Simpson. Now along with Simpson who is, I guess, about as modern as we can get in taxonomy, I have another little thing to show you here which I think has something to do with what I am eventually going to try to put across. I don't know how many of you have your textbooks here, but of those of you who do, can look on page 444. On page 444 is reproduced a diagram taken from the interesting work of ~~Hiesey~~ Clausen, Keck & Hiesey done in 1941. The interesting thing to note there are those drawings first,;and lets not worry about what the informatinn is. It is about a group of Western Compositae species in the genus Layia. But note there the gray-hatched things are indeed drawn as I suggested that most taxonomists will draw things. Imposed upon this, of course, are several other interesting features which deal with the crossability of these organisms. The black heavy lines there in showing this. But note farther that there are a series of dotted lines sort of being radii for the total circle there. These dotted lines being drawn in at points where you can see that the canals between the grayish circles seem to be the smallest canals, and that there are smaller extensions made therein. Now what we are intending here,I think, from these illustrations, the description of a species by Simpson in his paper and what indeed has been the case for all of the botanists, taxonomists, biological systematists throughout all these times is trying to make some discovery which indicates relatedness and not relatedness. That in all of the cases no matter what type of information is employed, the people herein are taking abut some sort of grouping with discontinuities between the groups. For example, in looking back again at Simpson's definition, can we reread that a little bit? "A biological species is an evolutionary unit composed of continuing populations that regularly interchange genes by interbreeding and that do not or cannot have such regular interchange with other species." Those terms do not and cannot obviously may be reread to some extent that there is a gap between this

Digitized by the Hunt Institute for Botanical Documentation

particular group of organisms and some other group of organisms. That if we look at the illustration on page 444 again, that the taxonomist has made a decision therein based upon some discontinuity. The discontinutiy is better informed as far as Layia is concerned because they have attempted to put in information about the shape and the size of the plants, that is the morphological information, and they have attempted as well to incorporate therewith information about the chromosomes of the organism, and not only about the chromosomes or the cytology but also something about their breeding patterns. You will discover that their major discontinuities are between those things that do not have the opportunity to hybridize quite so well, or cross quite so well. This is a very nice pattern that is discovered in this group here, and I would suggest that anybody would be able to discover that if one talked about genus Layia that they would speak of this as being a good taxon, that is that they would have no difficulty in the discovery that there is indeed something with a gap between the clusters. The gap may be discovered in different ways, but I assure you that the purpose that the taxonomist has in mind is to discover the kinds of information which will indeed provide him with some sort of a gap, and the chances are good that he will stir around a very long time looking for that kind of information which will indeed give him that type of differentiation between groups of organisms. There is good likelihood that any group of organisms upon which he may wish to settle will have kinds of information in there which indeed will not separate his organisms into groups. It is quite surprising to note that he did not incorporate that data in his taxonomy. And what he is talking about when he says, "I have a good key character here or diagnostic character" means that he has selected a piece of informatinn which will do the job that he wishes for it to do, namely, to give me a bunch in one or a group and that he will not say anything about those other nasty pieces of information

Digitized by the Hunt Institute for Botanical Documentation

that didn't do the job for him. (Student) They didn't do the job all by themselves? (Dave) Right. Now it very frequently is the case that if you will read taxonomic works, you will discover that there is a discussion made after the man has made his classification. These discussions will be the introductory parts, where he will there tell you that it really is not as good and clean as my classification looks like as I have put it out here, but for my purposes this information is not used in the classification. Learn to read what these fellows really say is something that is not quite as easy as you might anticipate. Obviously in this situation the basic search is for some sort of biological feature that is in terms that Henry was using the other day, the intrinsic features of the organism, something about the plant or animal itself. That is, it is nice if we can discover on the plant itself that kind of information. However, when we attack a group of organisms where it is not clear to us which particular piece of information about the plant or animal will make a nice clean separation for you, we will very frequently look outside the plant and there is just nothing wrong with this as far as we recognize what we are doing. That is, you may discover that if you plot your plants out on a map, an outline map, one group for all in North America, another group ;in South America, it is obvious to look amongst those things which fall in the same continent together to discover whether there would be similarities there which would adequately separate these organisms from those in South America. Sometimes we have to go into a different direction in order to establish it. There are known cases, for example, species with disjoint xxxx ranges occurring, a number of plant families have representatives that occur in our own southwest and again do not show up until xxx a similar habitat is found down in northern Argentina in South America. So what we have attempted to do there is to find other basis by which we may or may not separate these organisms. We discover, for example that there are eastern versus western groups of organisms. Uvularia

Digitized by the Hunt Institute for Botanical Documentation

which is almost entirely confined to the eastern part of the United States is going to have something else, isn't it? We can't use the eastern versus western approach to this problem as far as the genus and the species within the genus Uvularia occurs. I am not saying that to throw you a curve on your coding of information, but I sug= gest to you that you will have to discover that if there be differences outside the plants which are useful to you, that you will have to discover something besides geography to do it. It is also obvious that plants have different preferences for growing in certain places which may give the outside type of information that will give us a clue as to how we can separate them. Certainly, soil preferences or perhaps in geological provinces or perhaps moisture, regimes, or total climate aspects, that is whether it be warm, temperate, cool, may have some effect upon how you go about the selectinn of information which will indeed put us out in this busi= ness of separating groups of organisms accirding to their dissimilarities or accord- ing to their similarities. I think perhaps/a taxonomist gets himself really wrapped that up in the problem of phylogeny and evolutinn when has attempted to discover these discontinuities. I may be wrong on this, but I suspect that one of the ways people have attempted to make gaps between things is discovering some evolutionary path- way which clearly will show them discontinuities between groups of organisms. There is no real need xx to say whether or not these groups of organisms that we are talking about are real or are not real. If we choose our information correctly, any group of organisms which we have established as a species may be shown indeed not to be separate from anything else, if we choose our information in the right way. It can be shown to have continuous variation with something, if we choose our information correctly, or incorrectly, I don't know which way to say it here. (Bert) That is if you say that a small bit of information can be used to determine difference? If you allow that? (Dave) Yes. In other words, the information which may show us a gap between plants may, is selected primarily for that particular

Digitized by the Hunt Institute for Botanical Documentation

purpose, to show us the gaps. Whereas another feature of the organism may definitely indicate that there is a relationship t others that makes it such that you can't distinguish them. But when we get down to this sort of level, it is clear what our purposes are in taxonomy, is it not? That for the sake of conveniences and discusssions for me to tell you about these organisms here, that I have to discover some piece of information which indeed will allow us to put them into packages of sizes or clusters, or something of this nature such that we can adequately talk about them. I think that this is primarily our goal in taxonomy. If there is a particular goal, I would suggest that this is about as close to it, the totality of the goal as we can really make for the whole ;thing. I said at the beginning that I have no intention whatsoever of talking about the definitions of different categories of organisms, I think we are a long ways from doing anything like that. Maybe we are not as far away as we think, but at least as far as from where I stand, it looks like we are a long ways away from making an adequate species definition, But all I can clearly say is this: that we must recognize that when we are talking about categories like this, these are man-made constructs. Unless we define them on that basis alone, we are going to continuously find ourselves as people have for the last half century have found themselves, in hot water when they have tried to make a species definition.

Digitized by the Hunt Institute for Botanical Documentation