CLASSIFICATION AND OTHER WAYS OF KNOWLEDGE ORGANIZATION

Vladimir Marko
Department of Logic and Methodology of Sciences, FiF UK

Analytical methods in social sciences and humanities
APVV-0149-12 · www.amesh.sk
Bratislava, 24.11.2016

Ways of knowledge organization

Three basic assumptions of the lecture:

- There is a need for a precise notion of classification
- “Times have changed” – paradigms of classification are changing
- The outcome is:
- new approaches needs not only some new tools but also willingness for revision of old achievements (reclassification)

Aims of classification can be different

- the **mnemotechnical reasons**
- a problem of **storage information**
- the **effectivity** in knowledge **orientation** and knowledge **reuse**
- more precise dictionary and further development of knowledge:
  - more refined view of **the system of entities**
  - the aim of **searching for unknown** or unseen **interrelations between (or among) classified entities** or data
  - **the ideal**: a way to causal explanation of relationship between entities

A notion of classification - a meaning

- **Hull 1998**: "Any set of entities can be classified in indefinitely many ways…"
- **Hull,** *ibid*. **"**The **ultimate goal for scientific classifications** is to group entities so that these classes function in, or facilitate the formation of, **scientific laws …**
- **Bailey, 1994**: "classification is the general process of grouping entities by similarity;"
- Ibid.: "*the ordering of entities* into *groups or classes* on the basis of their ***similarity***… classification is both **a process** and **an end result**."
- "Complete systems of classification are called typologies" 124-6, Stacey C. Jordan, *Routledge Encyclopedia of Historical Archaeology*, 2002

A notion of classification - a meaning

- Adams, *Classification and Typology*:
  - "Archaeologists often use the terms classification and typology interchangeably, but in this article a distinction will be made.
  - A classification is any set of formal categories into which a particular field of data is partitioned.
  - In contrast, a typology is a particular type of rigorous classification in which a field of data is divided up into categories that are all defined according to the same set of criteria, and that are mutually exclusive.
  - As will be shown, most archaeological classifications of artifacts are typologies, while most classifications of cultures are not."

- Adams & Adams - *Archaeological Typology* 1991, 296-7: "Some participants in the Typological Debate prefer to talk about classification (Linton 1936: 382-400; Rouse 1960; Dunnell 1986), some about typology (Krieger 1944; 1960; J. A. Ford 1954b; Kluckhohn 1960), and some about taxonomy (Brew 1946: 44-66), but to a large extent these terms have been used interchangeably."
- Bailey, 1994, p.4: „Typology is another term for a classification..."

A notion of classification - a meaning

- **Classification is**
- **a analytic (conceptual) method {M}**
    - (as **result** of problem solving activity)

- **an activity {A}**
  - *at the **conceptual (abstract) level***
    - a **problem solving activity** (challenging with some unsatisfactory gap in background knowledge) with the aim of creating way of useful method for knowledge organization is resulting in {**M**} or **M**-hypothesis;

    - activity {**A**} of making {**M**}-plan
      - naming, describing and classifying, rendering of properties;
      - making a M-plan of execution of {**A+M**};
      - thought experiment in testing **M**-hypothesis;

A notion of classification - a meaning

- {**A**} activity *at* ***grounded level***
  - **sorting (or entities)** – *execution of classification plan* (according to *M-hypothesis)*:
    - bounding variables of M-function with concrete data
    - sorting or rendering concrete object according to {M}, for example, books on shelves; trash into appropriate cans

A notion of classification - a meaning

- **A classification list {R}**

[concrete **result of classification activity** {**A**} – of applying a conceptual analytic method {**M**} – *i.e.* outcome of classification: classification model or M-Plan (scheme, table, list, etc.)];

  - **(A+M) → R.**

- **{AG} other forms or activities of grouping or assorting** entities, properties, objects, events, phenomena**,** etc. (with scientific or non-scientific background) that **are or are not outcomes of {M} or connected with {M}:**
  - **naive** (and **non-exhaustive, non-exclusive**) any grouping or activity of sorting, **or**
  - **any other forms of knowledge organization** (sometimes by arbitrary chosen principles or properties; can be either exhaustive or non-exhaustive)
    - **hierarchical** (systematization; partonomy; kinds of polythetic taxonomy)
    - **non-hierarchical** (typology; nomenclature; stratification)

A notion of classification - a meaning

- **Classification** can either be
  - ***unidimensional***, being based solely on **a single dimension or characteristic** or
  - ***multidimensional*****,** being based on **a number of dimensions.**

  - **Structural** – Mendeleyev periodic table…
  - **Historical** – evolution tree; classification of languages…

A notion of classification - a meaning

A brief summary on classification:

- {**M**} **method**
- {**A**} **activity** (a. on *abstract* or *grounded* level; including sorting)
- {**R**} **result** (has to be testable)

- {**AG**} **other kinds** of sorting properties (or things)

Ideal of classification principle

{**M**}, {**A**} & {**R**} are **methods**, **activities** (of applying the method) and **results** (of applying the method) **that are related with a single (unique) classification principle** (or *criterion* of classification).

An alternative to ideal approach

- **Hybrid classification principle:** topological, etiological, operational, ethical, political… (for example, today approach to Dewey's **UDK** (*Universal Decimal Classification*) or **ICD** (*International Statistical Classification of Diseases*);
- Result is a nomenclature.

A classification principle – an illustration

- For example, ***a logical division*** can be of can be rated as a **unique** and **exhaustive** classification principle or method {**M**} (*i.e.* a ***function*** of breaking up of a whole into its component parts, in accordance with their similarities in properties, structure, origin, etc.):
  - division of *sumum genum* to subsets according to *resemblance and difference* in properties of set's members.
  - **A binominal definition**
  - At each level of applying the principle of division we are obtaining definition used for denotation of kind of entity (*genus*) and its specific difference (*differentia specifica*) in respect to other members of the same kind.
  - The principle is applied until all members of the set obtain their unique place.

A classification principle – an illustration

- **Result is**:
  - all entities of the set have own place in hierarchy;
  - definitions are covering by essential properties.
  - It could lead to systematization if the principle of organization can be explanatory

- Some authors emphasize ***the logical division*** as the ***only kind*** *of classification*: it is exhaustive and all entities of set are covered by one principle.

Meronomy or partonomy

- a type of hierarchy that deals with part–whole relationships, in contrast to
- "X" is a meronym of "Y" if an X is a part of a Y
  - Cars have parts: engine, headlight, wheel, ...
  - Engines have parts: crankcase, carburetor, ...
  - Headlights have parts: headlight bulb, reflector, ...
  - ...
- the unit of meronomical classification is **meron**
- Meronomies may be represented in Semantic Web languages such as OWL: „*has-a*“ relation...
- It is hierarchical ordering
- Difference from taxonomy: taxonomy is categorization based on discrete sets not relational features.

Taxonomy

- **Taxonomy** is often employed synonymously with **systematics** (**and/or classification**):
  - Earley, 2005: "Purist definition of a taxonomy – terms have **parent/child relationship**."
  - **Landweh** et al. (1994). “It implicitly embodies a theory of the universe from which those specimens are drawn... "
  - **Gould**, 1989, p.98: "Taxonomy (the science of classification) is ... a fundamental and dynamic science, **dedicated to exploring the causes of relationships and similarities among organisms**. Classifications are theories about the basis of natural order, not dull catalogues compiled only to avoid chaos."
  - Mayr 1942/1982: 6n.1: "The terms systematics and taxonomy are considered by me as **approximately synonymous**...[; i]n America...[,] the term taxonomy seems to be preferred...[; i]n the rest of the world...[,] the term systematics seems to be more widely used

Taxonomy

- **In Liennaeus**:
  - unique binominal nomenclature (not necessary hierarchical systematization);
  - a taxon encompasses all included taxa of lower rank;
  - one property is sufficient for ranking;
  - based on “physical” characteristics.
- **Today**:
  - Taxonomy is used no more for binominal nomenclature but for multidimensional typology!!!,
  - it could lead to classification {M} and systematization
  - measure of distance from another (neighbor) taxon (or taxa);
  - taxon is frequently replaced by the term *clade* (branch) as a basic unit (covering only monophyletic taxa - all descendants of some ancestor – useful to see unit in the scope of systematics, resp. evolutionary systematics).
- **Example - Taxon vs. clade**:
  - Cladistics, another way to classify organisms. It can use data from DNA or RNA sequences, rather than just physical characteristics. It emphasizes the evolutionary relationships between different species.

Typology (simple t. or one representing 'ideal type')

Labeling an entity by one or more (*major* or *minor*) characteristic(s) of selected population

- The whole set of data is an unit of division (100%).
- "unidimensional, monothetic" – **major** distinctive characteristic, disjunctive characteristic, necessary and sufficient
  - (human being: man or woman; man, 50% + woman, 50%)
- "multidimensional, polythetic" – one of **minor but relevant** characteristics, in respect to sorting needs or purpose of sorting (is it woman, Afro-American origin, mother, employed)

A Hypothetical Fourfold Typology

	Motivated	Unmotivated
Intelligent	Success 1	Underachiever 2
Unintelligent	Overachiever 3	Failure 4

Typology (simple t. or one representing 'ideal type')

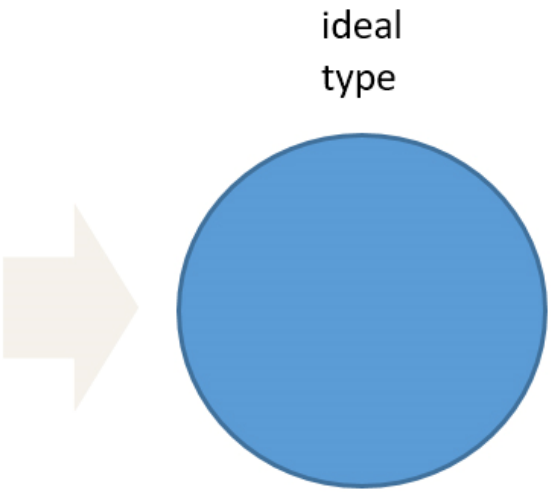
- often based on **arbitrary or ad hoc criteria**
- not essential characteristics or sum of characteristics of an entity.
- **descriptive rather than explanatory or predictive**
- Purposes dictates relevance in type division

Typology (simple t. or one representing 'ideal type')

- Generalized conjecture
- "Ideal type" determines way of data interpretation and data organization
- It is instructive but "never corresponds to reality", so, it is frequently subject to **the problem of reification** (Bailey, 1994) **-** there is no independent verification of its existence.

	a	b	c
A	60%	60%	60%
B	60%	60%	60%
C	60%	60%	60%

1	Aa	60%
2	Ab	60%
3	Ac	60%
4	Ba	60%
5	Bb	60%
6	Bc	60%
7	Ca	60%
8	Cb	60%
9	Cc	60%



Systematics

- **Systematics** (i.e. model), is the study of the pattern of relationships among taxa and their dependencies
- Explanatory power
- One common principle od explanation of patterns and relationship in system observed
- **In biology**, for example, **hierarchical classification** is only one aspect of phylogenetic **systematics**; the principle of explanation is a theory of evolution
- **In physics**, periodical table of atomic elements is ordered **structurally** and explained by physics of elementary particles

Nomenclature

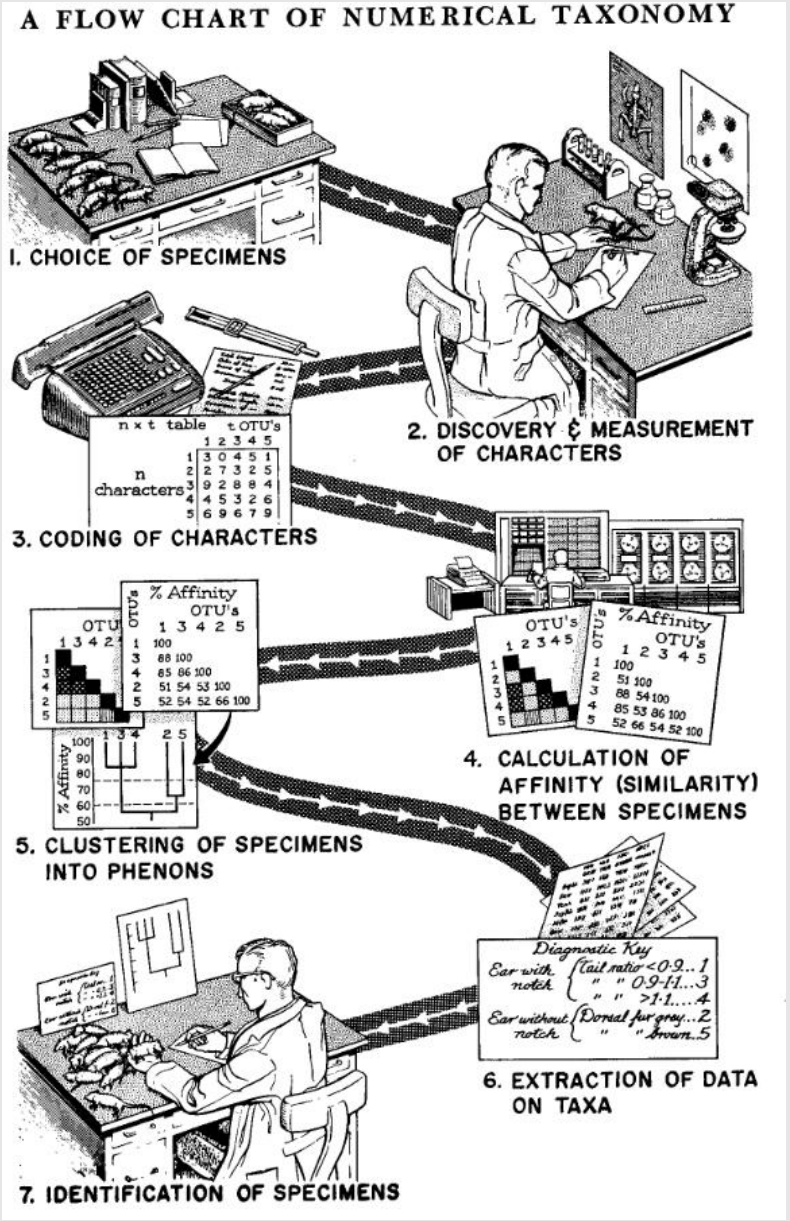
- List of entities ordered by numbers or alphabetically, without respect to their properties

Cluster Analysis

Sokal, 1963

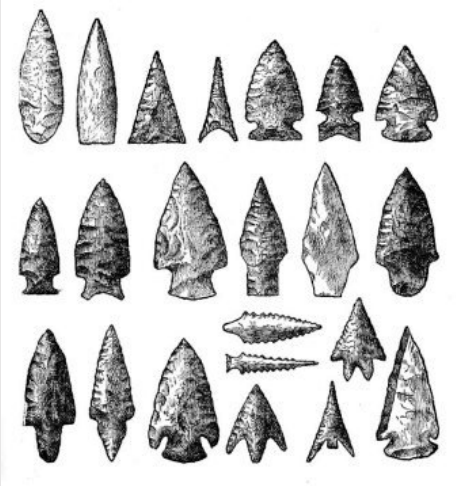
Pheneticism (taximetrics)

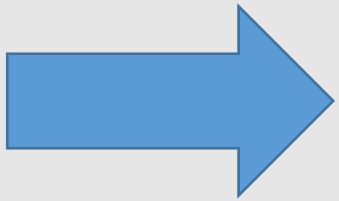
- Today, the notion **taxon** is used for sets of data, that are
- **observable**,
- **measurable** and
- numerically **analyzable** and **representable**.

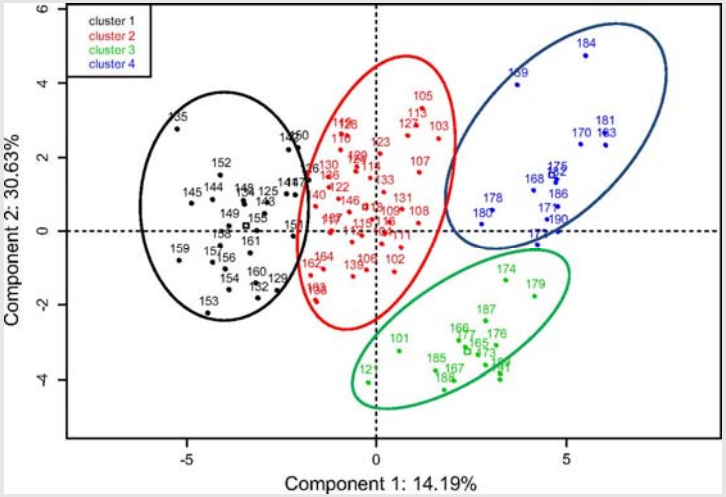


Cluster Analysis

- **Searching for patterns** / familiarities among data
- *qualitative or quantitative* ***distance*** of one or more properties
- Result is **homogenity** or nonhomogenity of types
- CA is (usually) **multidimensional typology**
- Character of CA classification is determined by **resolution of observance** (Mellor)
- New resolution ask for revision of former results
- It **could lead** to classification, and also, at the end, to systematics



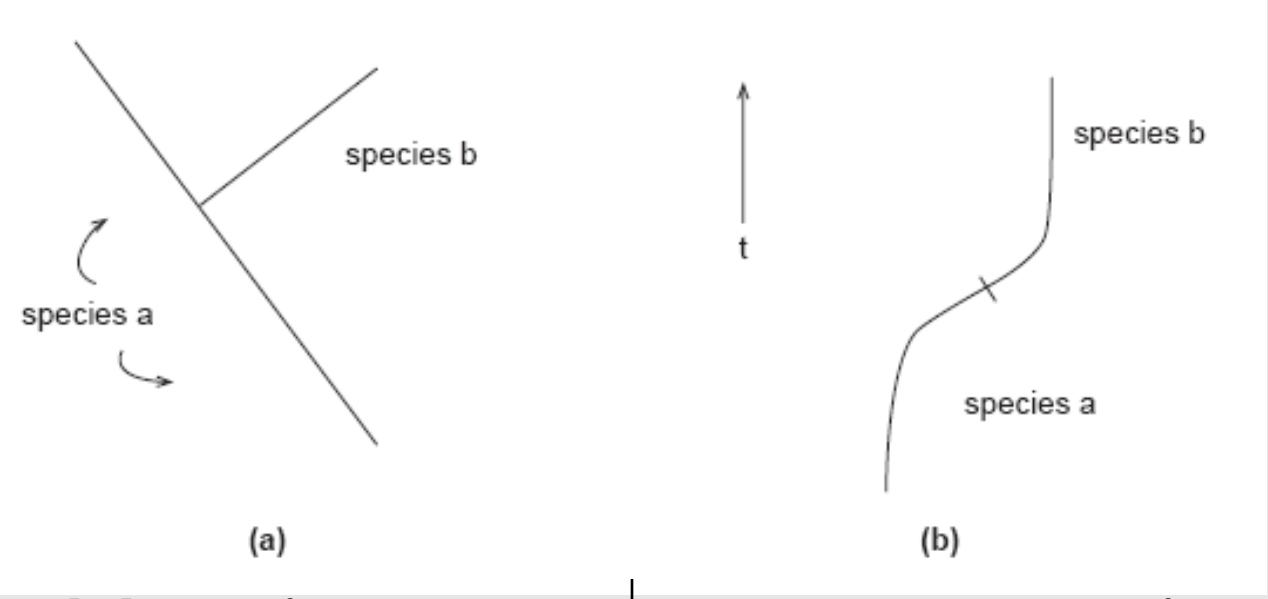




Some problems from modern biology

Vocabulary of biological classification

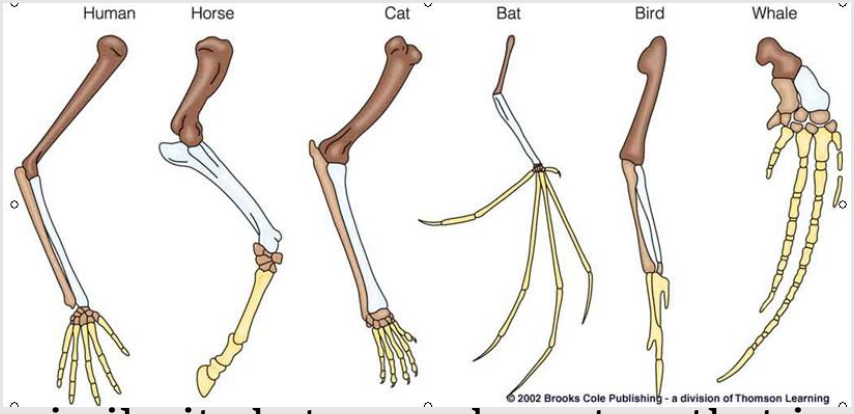
"**evolutionary taxonomists** . . . aim to construct classifications that reflect both of the two major evolutionary processes, branching and divergence (**cladogenesis and anagenesis**)" (Mayr (1981[1994], 290).



cladogenesis	**anagenesis**
A population of species A becomes geographically isolated from the rest of the species. It undergoes a **genetic revolution** and becomes a distinct species, B. In cladogenesis, a species are split in two.	Species A **gradually evolves** until it becomes a new species, B. Evolutionary taxonomists accepts both C & A (Cladists ignores anagenesis divergence – *one reason,* changes could not be causally explained)

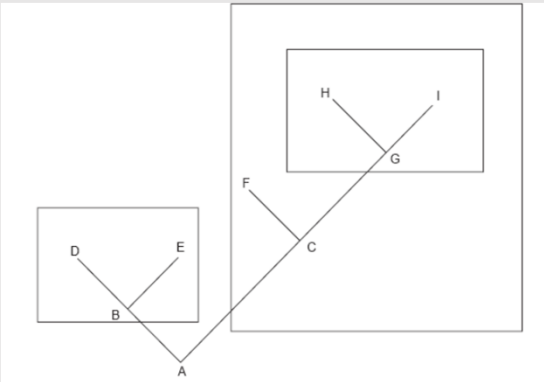
Vocabulary of biological classification

- **Homoplasy** occurs when **characters are similar**, but are **not** **derived from a common ancestor**.



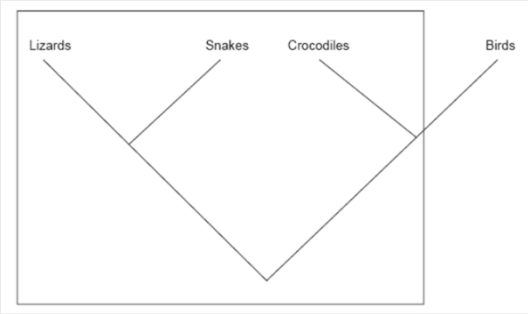
- **Homology** is any similarity between characters that is **due to their shared ancestry**.

Vocabulary of biological classification



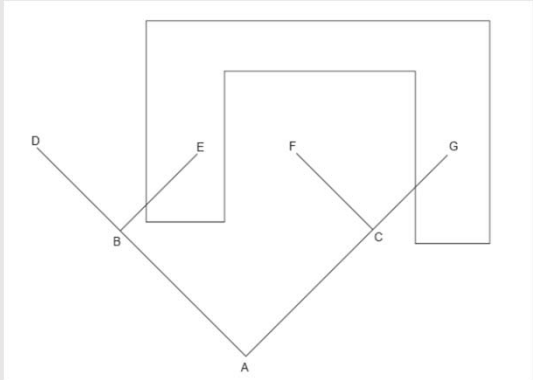
monophyletic taxa

- The taxon consisting of D, E, and B is a monophyletic taxon; so is the taxon containing H, I, and G, and the taxon containing H, I, G, F, and C.
- Each contains an ancestor and all of its descendants.
- One single ancestor (the rule of monophyly)



paraphyletic taxa

- The taxon *Reptilia* contains lizards, snakes, and crocodiles, but does not contain birds. *Reptilia* is a **paraphyletic taxon.**
- it contains an ancestor and some but not all of its descendants.

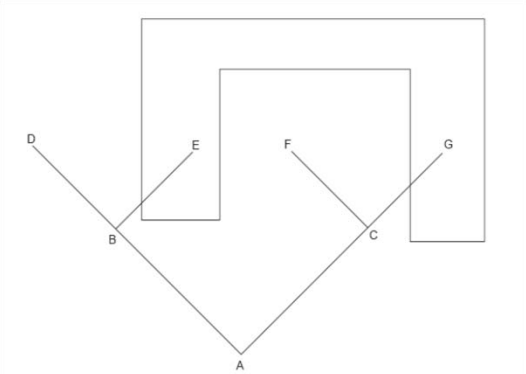


polyphyletic taxa

- The taxon containing only E and G is **polyphyletic**.
- E and G share similar traits (homoplasies) that were not present in their common ancestor, A, but evolved independently in E and G.

- Monophyletic and paraphyletic taxa contain organisms that have *homologies*: traits shared by different lineages that were passed down from a common ancestor.
- A major task of evolutionary taxonomists has been to remove polyphyletic taxa from earlier classifications.

Actual problems

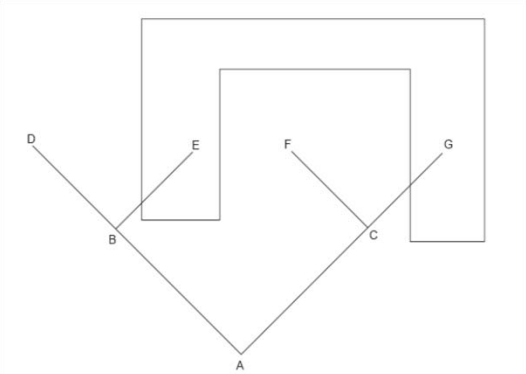


polyphyletic taxa

Essentialistic approach or some other way?

- **Non-gradualism**; species are **discrete entities** (theory of **punctuated equilibria**), not continual;
- Essentialism requires **sharp boundaries** **and precise essences**
- More and less essential - How many "*essential essences*" in clusters?
- Problem - **absence of causes**. Problem with explaining their evolution and transformations
- Kinds with **different pasts** could **have the same essence**! **(**GMO, artificial viruses, etc.**)**

Actual problems



polyphyletic taxa

Some answers…

- **Neither Linnaeus nor *Ideal morphologists* were strict essentialists**
- **Vague essences**? Sober - essentialism is consistent with vague boundaries so long as essences are correspondingly vague.

- **Pheneticists and Pattern cladism (or "transformed cladism"):**
  - the evolutionary framework is inessential, and may be dropped (Patterson).

- **Evolutionary taxonomists and Process cladism (against the ideal morphology):**
  - **the primary relations among organisms are causal ones** such as genealogical and embryological relations (Hennig)

Summary

Classification is:

- **a (analytic / conceptual) method {M} of knowledge ordering**
- **an activity {A}**
- **a result of classification - a classification list {R}; (A+M) → R.**

Other forms of knowledge ordering {AG}

- **hierarchical** (systematization; partonomy; kinds of polythetic taxonomy, CA?)
- **non-hierarchical** (typology; nomenclature; stratification, multidimensional typology - CA)
- **Hybrid classification**, using of several classification principles at once

Thanks for your patience