# Terminology, Classification and Ontology in the Biomedical Domain: Past, Present and Future

## Harold Solbrig

## Mayo Clinic

# Outline

- **Terminology – what is it really all about and why should I care?**

- **Terminology in Medicine – where we are now and in the near future**

# Terminology

**No matter what the medium…**

**… spoken word**

**… books and magazines**

**… radio, television and movies**

**… the internet**

**…** *software* **and digital records…**
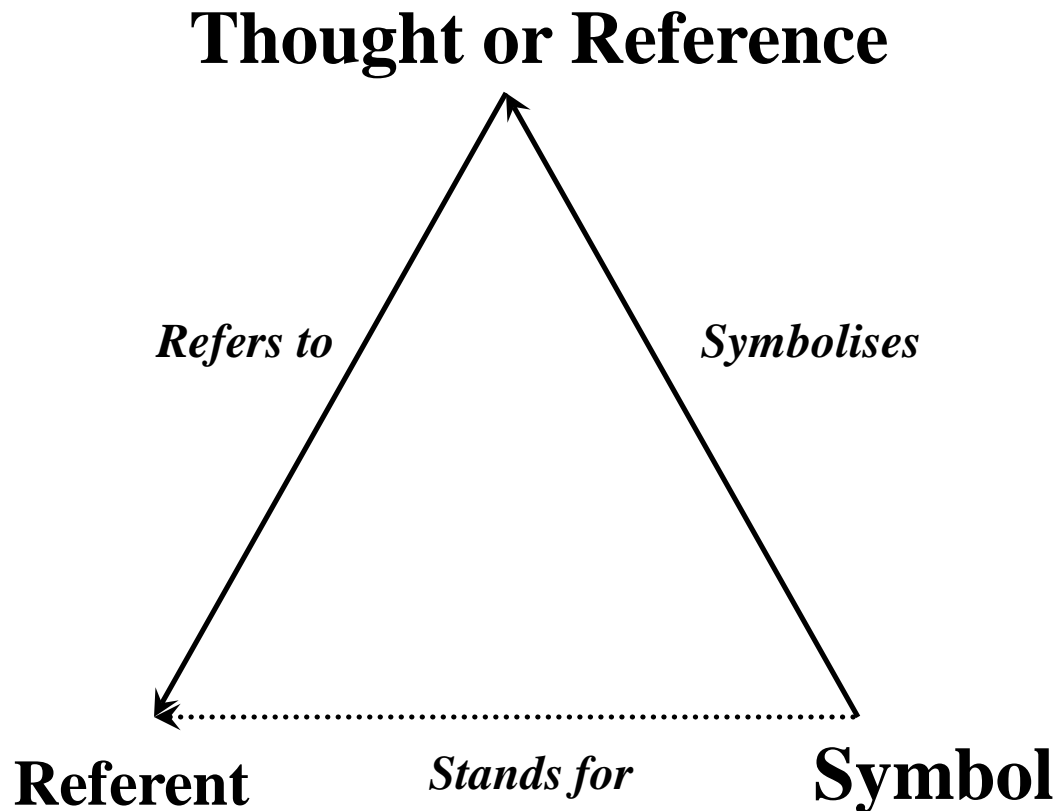
# … the goal is the same:

# Communication

# Communication is about Language

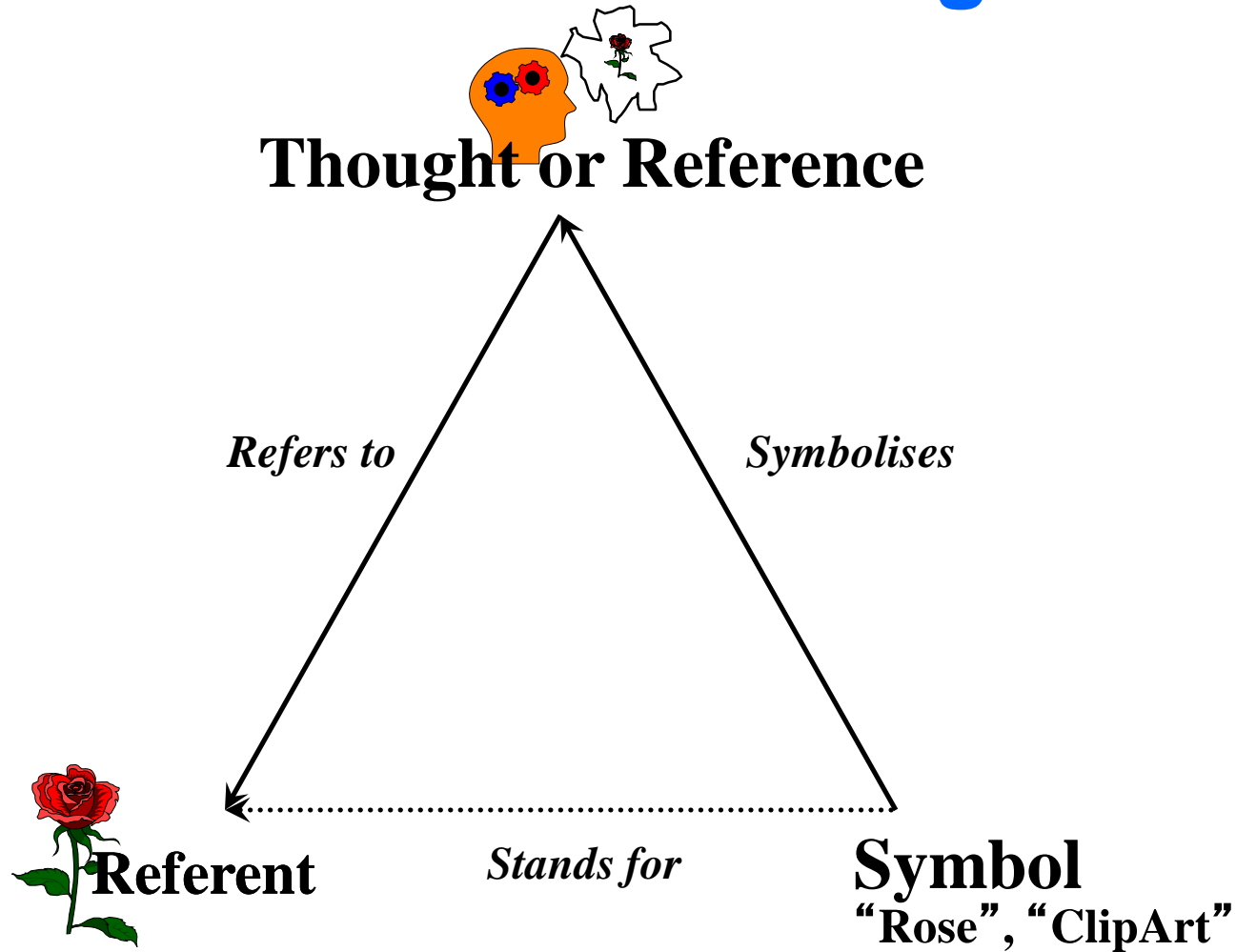**Language - a "specification" that enables communication**

- **<u>Semantics</u> - the association between signs or symbols and their intended "meaning"**

- **<u>Syntax</u> - the rules for ordering and structuring the signs into phrases and sentences**

- **<u>Pragmatics</u> - the relationship between signs and symbols and the recipient. Broadly, the shared *context*.**
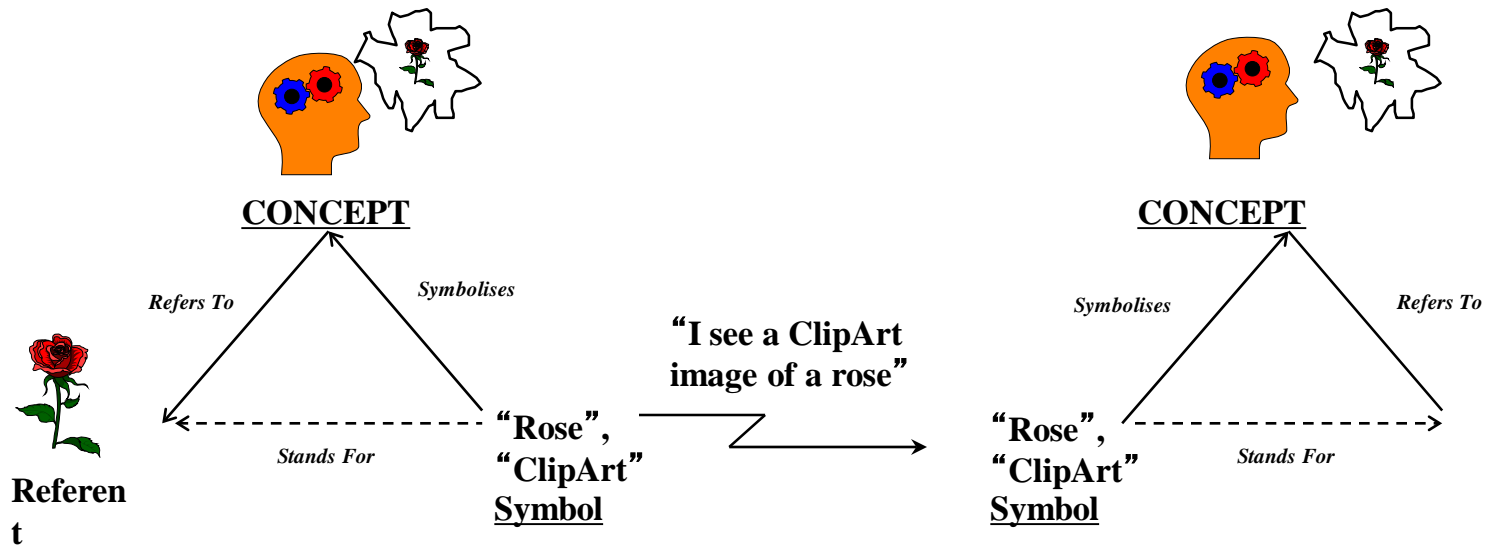
# The Semiotic Triangle

**Thought or Reference**

*Refers to*          *Symbolises*

**Referent**   *Stands for*   **Symbol**

C.K Ogden and I. A. Richards.  *The Meaning of Meaning.*

# The Semiotic Triangle

**Thought or Reference**

*Refers to*

*Symbolises*

**Referent**
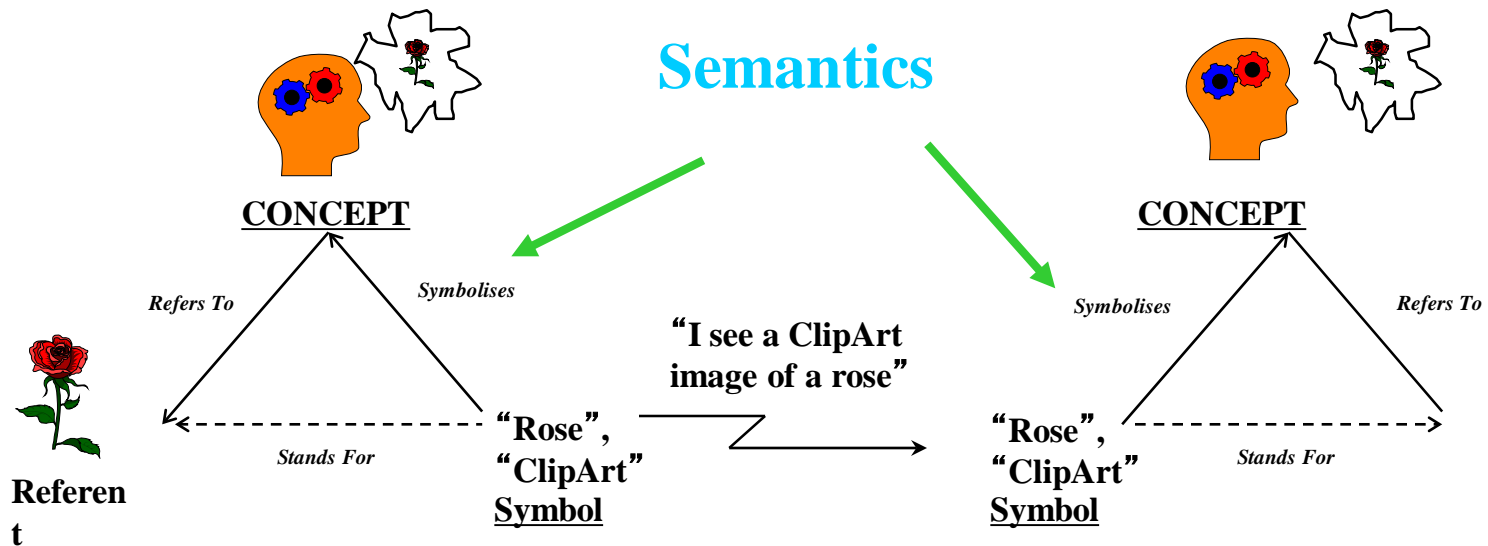
*Stands for*

**Symbol**
"Rose", "ClipArt"

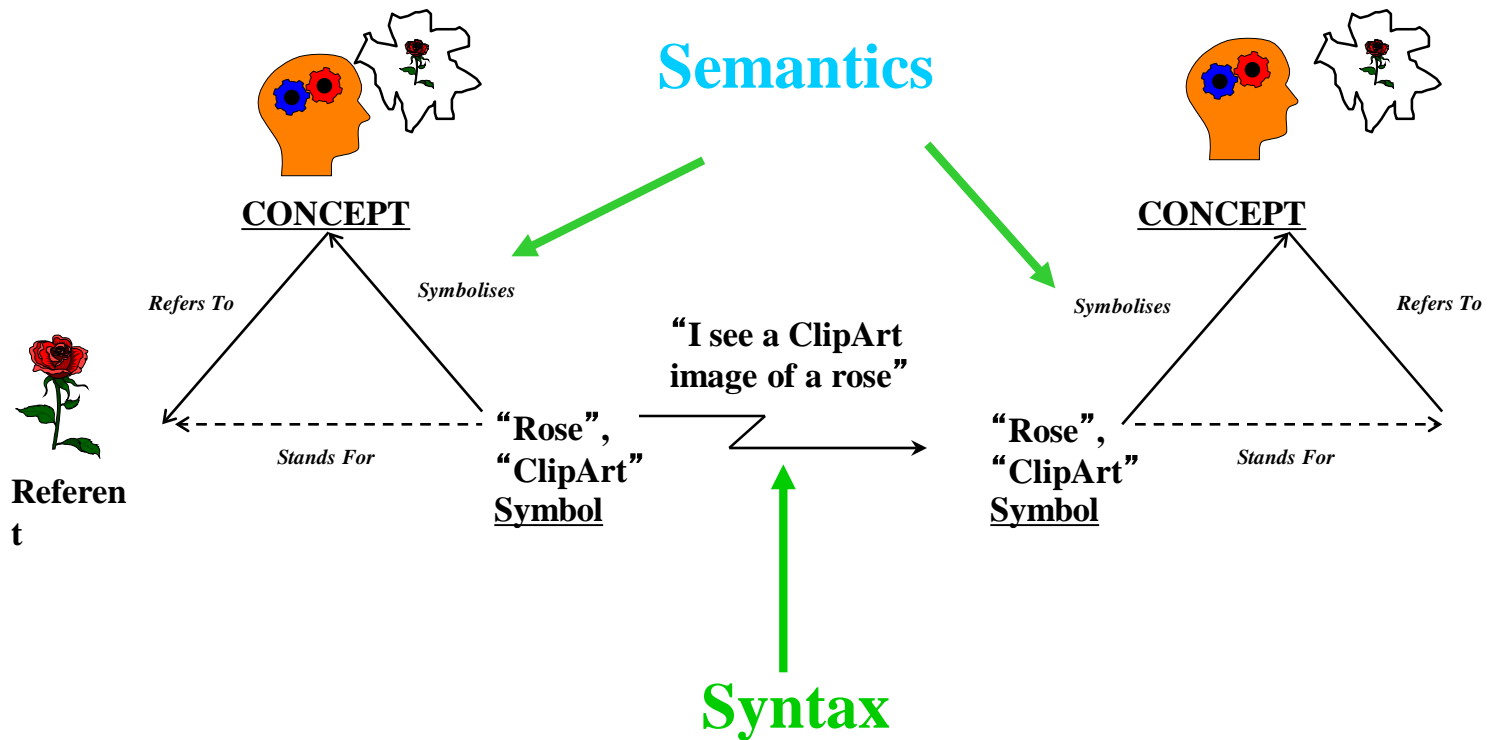C.K Ogden and I. A. Richards. *The Meaning of Meaning.*

# The Communication Process

# The Communication Process

# The Communication Process

# The Communication Process

Terminology in the Biomedical Domain
(c)Mayo Clinic 2013
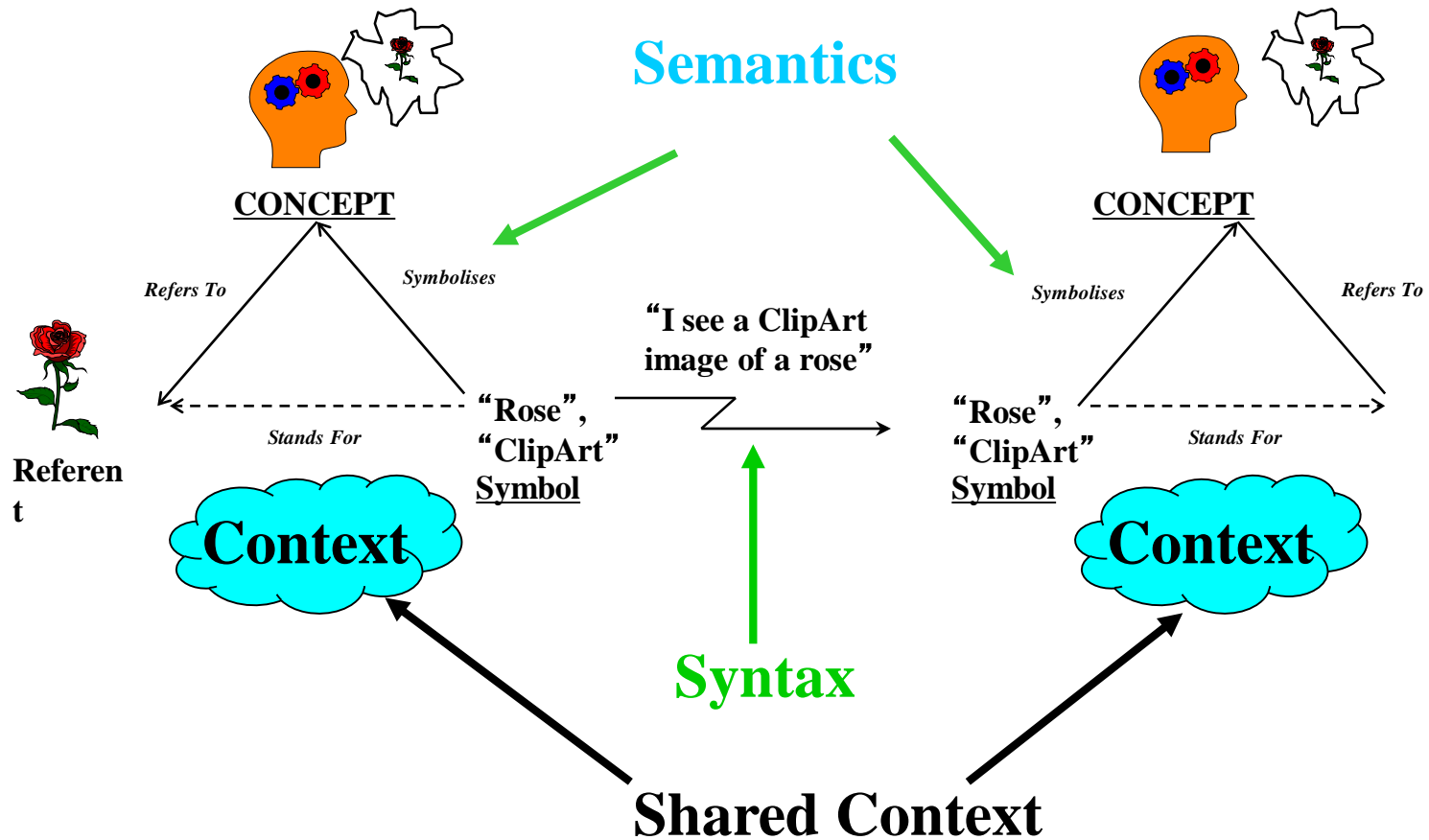
# Shared Context

## Impacts how much information can be contained in a symbol.



Information / Symbol

No Shared Context

Shared Universe

Shared Sun

Shared Planet

Shared Species

Common Language

Common Culture

Similar Education

Common Profession

Common Specialty
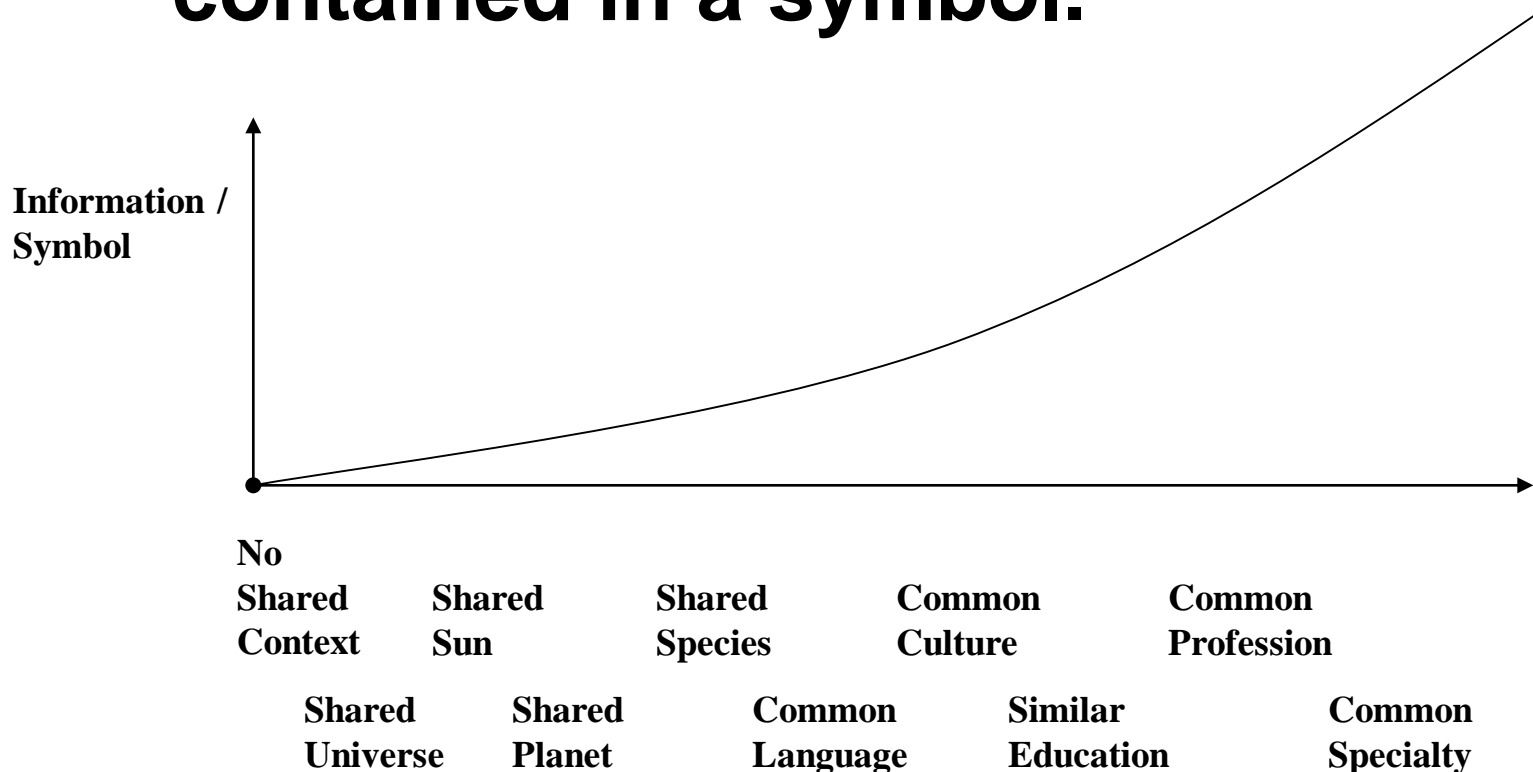
# Minimum Shared Context

**Terminology in the Biomedical Domain**
**(c)Mayo Clinic 2013**
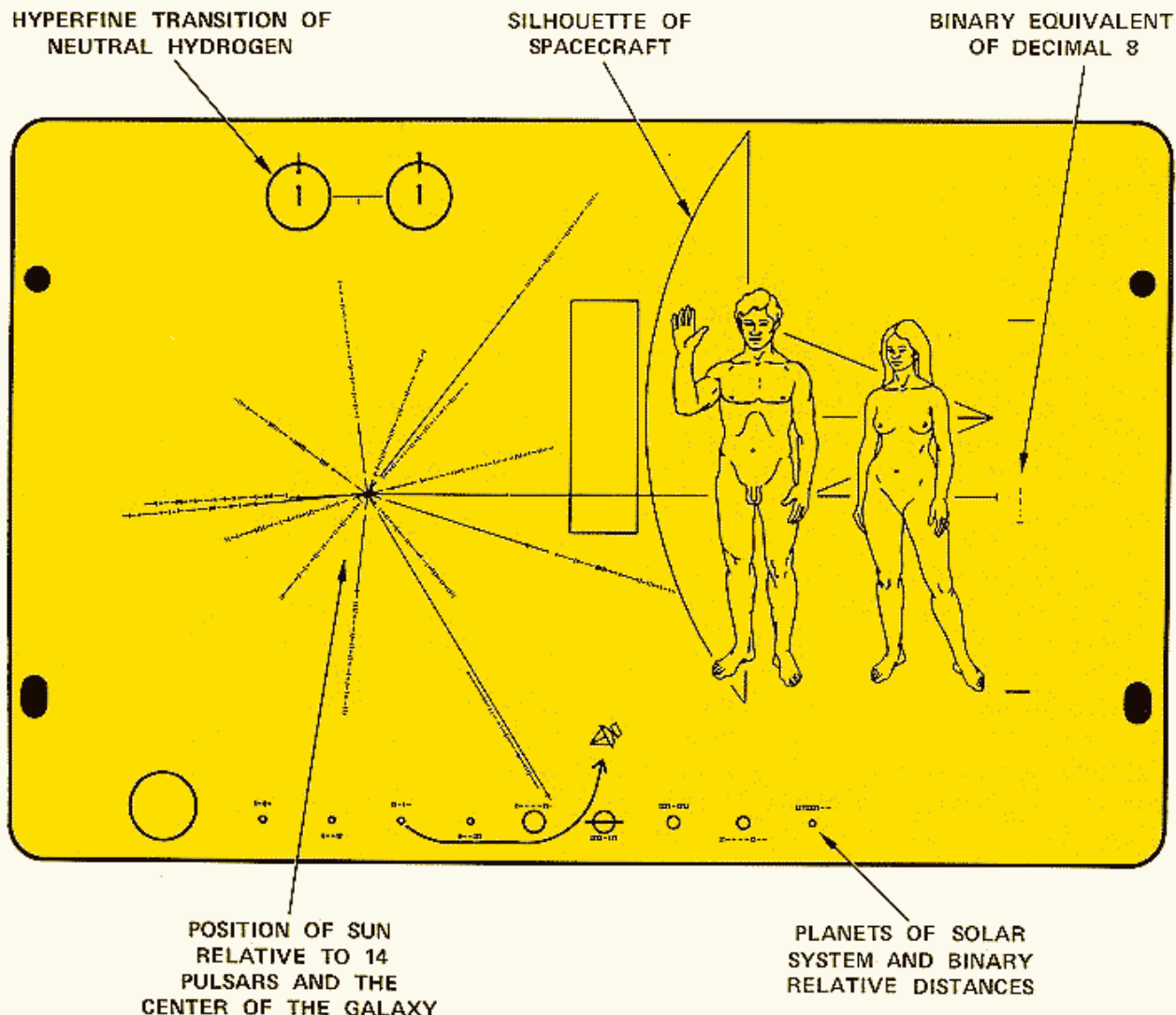
# The impact of context on communication

**Shared context:**

- **Allows information to be communicated in larger, more succinct "chunks".**
  - *Drug*, *analgesic* and *NSAID* are all "chunks", yet differ markedly in conceptual complexity.
- **Enables specialized symbol sets:**
  - Contrast the amount of information contained in the formula $E=MC^2$ versus that contained in this presentation...

# Contextual Formalism

**The degree of formality in a shared context can vary across a wide spectrum:**

- **<u>Tacit context</u>** - context which is assumed

- **<u>Contextual negotiation</u>** - level setting proceeding the actual message

- **<u>Rigorous formal rules and documents</u>** - describing the form and possible meanings behind every message and phrase.

# Factors Effecting the Level of Contextual Formalism

- **Number of participating parties**
  - **Formalism needs to increase as number of participants increase**

- **Geographic, cultural and temporal proximity of communicators**
  - **The further apart communicators are, the less they can assume**

- **Amount of shared context**
  - **The more you have, the more important it becomes to be organized**

# Factors Effecting the Degree Contextual Formalism

- **The cost of imprecise communication**
  - **Poetry and literature - low cost (some may argue actual gain)**
  - **Technical and professional - high to <u>very</u> high cost**
    - **What is the cost of assuming the units of a thrust specification?**
    - **What is the cost of assuming the dose of a prescription?**
    - **What is the cost of assuming the century in which the communication originated?**

# Factors Effecting the Degree Contextual Formalism

## Automation

**If you are going to set computers loose on a block of information, you are going to need to get it right to begin with…**

**… it is difficult enough to reach useful conclusions given precise and accurate inputs…**

**… and to output those conclusions in a useful fashion**

42

# Terminology

**A collection of <u>interrelated</u>, <u>interdependent</u> resources**

- **Code sets**

- **Classifications**

- **Thesauri**

- **Dictionaries**

- **Ontology (with multiple views)**

# This is not a new problem

- **Names and symbolism was the subject of early Greek Philosophy**

- **London Bills of Mortality**
    - **Commissioned 1542 (1598)**
    - **Intended to Track Plague (Black Death)**
    - **~60 disease categories**
    - **Data Table LayoutData Layout**
    - **16th Century Spreadsheet**

# This is not a new problem

- *Samuel Johnson's Dictionary* of the English Language, published in 1755.

- *Roget's Thesaurus* (1805-1852)

- International Classification of Diseases and its Clinical Modifications
  - First published in 1893 by Statistical International Institute
  - Revised every 10+ years
    - ICD8 – 1967 (World Health Organization)
    - ICD9 – 1977 (World Health Organization)
    - IDD1 – 1982 (World Health Organization)

# Weights and Measures

"The nomenclature is of as much importance in this department of inquiry, as weights and measures in the physical sciences, and should be settled without delay."

William Farr, about Cullenian system
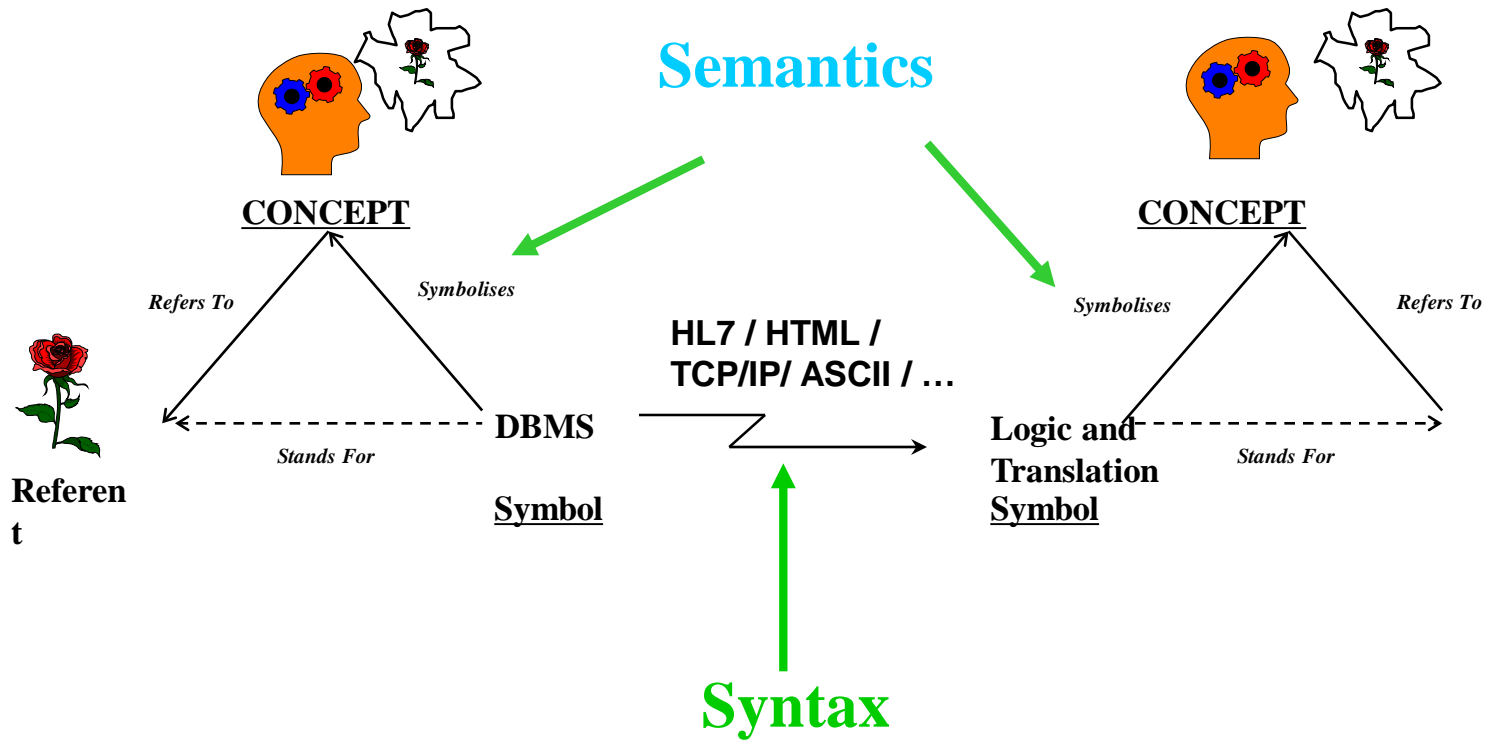
First Annual Report of the Registrar-General of Births,,Deaths, and Marriages in England. London: 1839 p. 99.

# What Has Changed?

The answer, in part is syntax

- Automation has provided a whole set of rules for encoding and exchanging symbols

# Automation



**Semantics**

CONCEPT

CONCEPT

*Refers To*

*Symbolises*

HL7 / HTML /
TCP/IP/ ASCII / …

*Symbolises*

*Refers To*

DBMS

*Stands For*

Logic and
Translation
Symbol

*Stands For*

**Referen
t**

**Symbol**

**Syntax**

# Automation

- **Whole new set of symbols**

- **"Meaning" needs to be shared not just with human creator and human recipients, but with intervening software**

# Centralized Context

**No matter the model or approach, communication depends on *shared meaning* – a *common repository* of symbols, their meaning and rules for their use.**

# Shared Context and Terminology

## Today terminological content is still in the form of:

- **Printed and PDF documents intended for human, not machine consumption**

- **Comma / tab / … separated tables w/ a variety of structures and formats**

- **(Sometimes) services – intended largely for human consumption**

- **RDF / OWL – the Semantic Web**

# The "Semantic Web"

**The Semantic Web**

- **Ontologies, RDF, Linking Open Data**

- **XML and HTML being annotated with RDF**

- **Good step forward, but…**
  - **There is still that pesky issue of symbols and their meaning**

# The Missing Component

**Interchangeable, interoperable models of the semantics themselves**

- **A shared semantics about terminological resources**

- **Syntax(es) (models) for communicating information *about* these resources**

- **Bridge between human / human and human / software for terminology itself**

# Centralized Context

**To share context, one has to have a shared context for sharing context…**

# Medical Terminology Today

- **Systemized Nomenclature of Medicine (SNOMED CT)**

- **Logical Observation Names and Codes (LOINC)**

- **Open Biomedical Ontologies (OBO)**

- **National Center for Biomedical Ontology (NCBO) BioPortal**

- **Unified Medical Language System (UMLS)**

# Medical Terminology Today

**NCI Thesaurus and Metathesaurus**

**ONC Meaningful Use Quality Measures**

**Health Level Seven (HL7)**

# Medical Terminology Near Future

- **W3C Health Care Life Sciences (HCLS)**

- **Clinical Information Modeling Initiative (CIMI)**

- **Common Terminology Services 2 (CTS2)**

- **ICD 11**

- **Genomics / Phenomics / High Throughput Phenotyping (HTP)**

# References

**SNOMED CT**

http://www.ihtsdo.org/snomed-ct/

**LOINC** http://loinc.org/

**OBO** http://www.obofoundry.org/

**NCBO** http://www.bioontology.org/

**UMLS** http://www.nlm.nih.gov/research/umls/

**NCI Thesaurus** http://ncit.nci.nih.gov/

**ONC Meaningful Use**
www.nlm.nih.gov/healthit/meaningful_use.html

# References (continued)

**W3C HCLS**  http://www.w3.org/blog/hcls/

**CIMI**  http://cimiwiki.org/

**CTS2**  http://informatics.mayo.edu/cts2

**ICD 11**

http://www.who.int/classifications/icd/revision/en/

**SHARP HTP**
http://informatics.mayo.edu/sharp/index.php/HTP

**solbrig.harold@mayo.edu**