



Language Technology: Research and Development

Science and Research

Joakim Nivre

Uppsala University
Department of Linguistics and Philology
joakim.nivre@lingfil.uu.se



Research and Development

“Research and experimental development (R&D) comprise creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications.” (OECD, 2002)



Research and Development

“Research and experimental development (R&D) comprise creative work undertaken on a systematic basis in order to **increase the stock of knowledge**, including knowledge of man, culture and society, and the **use of this stock of knowledge to devise new applications.**” (OECD, 2002)

- ▶ Research – new knowledge
- ▶ Development – applied knowledge (cf. engineering)



Research and Development

“Research and experimental development (R&D) comprise **creative work** undertaken **on a systematic basis** in order to **increase the stock of knowledge**, including knowledge of man, culture and society, and the **use of this stock of knowledge to devise new applications.**” (OECD, 2002)

- ▶ Research – new knowledge
- ▶ Development – applied knowledge (cf. engineering)



A Very Short History of (Western) Science

- ▶ Philosophy as a precursor of modern science
 - ▶ Antiquity: natural philosophy, Aristotle (600–300 BC)
 - ▶ Middle ages: scholastic philosophy (1100–1500)
- ▶ The scientific revolution (1500–1750)
 - ▶ Copernicus, Kepler, Galileo, Newton
 - ▶ Observation and experimentation
 - ▶ Mathematical models of physical phenomena
- ▶ Modern science (1900–):
 - ▶ Explosion of new scientific disciplines
 - ▶ Natural, social and cultural sciences (arts, humanities)
 - ▶ Computational linguistics (1950s)



Philosophy of Science

- ▶ Study of scientific methods
 - ▶ What distinguishes science from pseudo-science?
 - ▶ What is the nature of scientific reasoning?
 - ▶ What is a scientific explanation?
 - ▶ How does science make progress?
- ▶ Two schools:
 - ▶ Prescriptive – what scientists **should** do
 - ▶ Descriptive – what scientists in fact **do**



Deduction and Induction

- ▶ Deductive inference

All computational linguists are smart.
Ann is a computational linguist.

Therefore, Ann is smart.

- ▶ Conclusion follows logically from premises
- ▶ Characteristic of mathematical proofs

- ▶ Inductive inference

All computational linguists I have met are smart.

Therefore, all computational linguists are smart.

- ▶ Conclusion does **not** follow logically from premises
- ▶ Characteristic of empirical science (and everyday reasoning)



Induction in Science

- ▶ Newton's law of universal gravitation (1686)
 - ▶ Every point mass in the universe attracts every other point mass with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.
- ▶ Fleming's discovery of penicillin (1928)
 - ▶ Penicillium mold kills bacteria.
- ▶ Dürkheim's study of suicide (1897)
 - ▶ Suicide rates are higher in men than women.



Hume's Problem of Induction

- ▶ Induction presupposes “uniformity of nature”
- ▶ How can we rationally justify this assumption?
 - ▶ By deduction – safe but impossible
 - ▶ By induction – more plausible but circular
- ▶ Conclusion:
 - ▶ The principle of induction **cannot** be rationally justified!



David Hume
(1711–1776)



Verification and Falsification

- ▶ Logical empiricism/positivism:
 - ▶ Scientific claims must be **verifiable**
 - ▶ Theories are verified **inductively**
 - ▶ Prefer the **most** probable of competing theories
 - ▶ Observations are **objective** and logically prior to theories
- ▶ Popper's alternative:
 - ▶ Scientific claims must be **falsifiable**
 - ▶ Theories are falsified **deductively**
 - ▶ Prefer the **least** probable of competing theories
 - ▶ Observations are **theory-laden** but must be replicable



Karl Popper
(1902–1994)



The Hypothetico-Deductive Method

- ▶ Universal claims can be falsified (but not verified) deductively:

Bob is a computational linguist.

Bob is **not** smart.

Therefore, **not** all computational linguists are smart.

“No amount of experimentation can ever prove me right;
a single experiment can prove me wrong” (Einstein)

- ▶ Given hypothesis **H** with consequence **C**:
 - ▶ If **C** does **not** agree with observations, **H** is rejected (falsified)
 - ▶ Else **H** is provisionally accepted (corroborated)
- ▶ Science:
 - ▶ Progress through repeated testing, falsification, revision
 - ▶ Knowledge fundamentally uncertain (“current best theory”)



Inference to the Best Explanation (IBE)

- ▶ Another non-deductive inference type

A window has been broken.

A valuable painting is missing.

A thief broke the window and took the painting.

- ▶ Conclusion does **not** follow logically from premises
- ▶ Alternative explanations are possible
- ▶ The principle of parsimony:
 - ▶ Prefer a simpler explanation (theory) over a more complex one
 - ▶ Darwin's theory of evolution
 - ▶ How can this principle be rationally justified?
 - ▶ Is IBE a form of induction (or the other way round)?



Probabilistic Reasoning

- ▶ Laws and theories involving the notion of probability
 - ▶ Every gene has a 50% chance of being inherited (genetics)
 - ▶ Suicide rates are higher in men than women (sociology)
 - ▶ 90% of all lung cancers are caused by smoking (medicine)

- ▶ Inductive inference:

80% of all computational linguists I have met are smart.

Therefore, 80% of all computational linguists are smart.

- ▶ Deductive inference:

80% of all computational linguists are smart.

Ann is a computational linguist.

Therefore, Ann has an 80% chance of being smart.



Scientific Explanation

- ▶ Structured like an argument:
 - ▶ A set of premises (**explanans**)
 - ▶ A conclusion (**explanandum**)

Why did the metal rod expand?

All metal objects expand when their temperature increases.
Fire increases the temperature of objects.
The metal rod was placed in the fire.

Therefore, the rod expanded.

- ▶ Hempel's covering law model of explanation:
 - ▶ Conclusion follows logically from premises (deduction)
 - ▶ Premises are true and include at least one general law



Carl G. Hempel
(1905–1997)



Problems with the Covering Law Model

- ▶ The problem of symmetry

Why is the shadow 5 meters long?

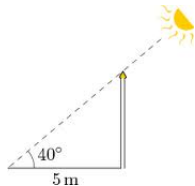
Light travels in straight lines.

Laws of trigonometry.

Flagpole is 4.2 meters high.

Angle of elevation of the sun is 40° .

Therefore, the shadow is 5 meters long.





Problems with the Covering Law Model

- ▶ The problem of symmetry

Why is the flagpole 4.2 meters high?

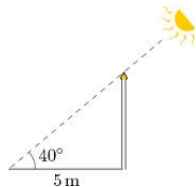
Light travels in straight lines.

Laws of trigonometry.

Shadow is 5 meters long.

Angle of elevation of the sun is 40° .

Therefore, the flagpole is 4.2 meters high.





Problems with the Covering Law Model

- ▶ The problem of irrelevance

Why didn't the man become pregnant?

Anyone who takes birth control pills will not get pregnant.
The man took birth control pills.

Therefore, the man did not get pregnant.

- ▶ The problem of probabilistic laws

Why did the man get lung cancer?

90% of all lung cancers are caused by smoking.
The man was smoking.

Therefore, the man got lung cancer.



Problems with the Covering Law Model

- ▶ The problem of irrelevance

Why didn't the man become pregnant?

Anyone who takes birth control pills will not get pregnant.
The man took birth control pills.

Therefore, the man did not get pregnant.

- ▶ The problem of probabilistic laws

Why did the man get lung cancer?

90% of all lung cancers are caused by smoking.
The man was smoking.

Therefore, his lung cancer was probably caused by smoking.



Scientific Change

- ▶ Traditional view:
 - ▶ Science advances in a cumulative fashion
- ▶ Kuhn's notion of paradigm (normal science)
 - ▶ A set of shared theoretical assumptions
 - ▶ A set of accepted problems and methods ("puzzle solving")
- ▶ Scientific revolutions
 - ▶ Accumulation of anomalies lead to crisis and revolution
 - ▶ Old paradigm abandoned only if new paradigm available
 - ▶ Copernicus, Darwin, Einstein



Thomas Kuhn
(1922–1996)



Beyond Natural Sciences

- ▶ Hermeneutics
 - ▶ Natural sciences seek **explanation**
Why? = What caused it to happen?
 - ▶ Social/human sciences seek **understanding**
Why? = Why did the agents bring it about?
 - ▶ Causality vs. Meaning
- ▶ Design science
 - ▶ Sciences of the artificial
 - ▶ Constructs, models, methods, instantiations
 - ▶ Truth vs. Utility
- ▶ Is there a universal scientific method?



Hans-Georg Gadamer
(1900–2002)



Herbert Simon
(1916–2001)



Research Ethics

- ▶ Traditional view:
 - ▶ Scientific knowledge is neither good nor bad per se
 - ▶ But scientific knowledge can be **used** unethically
 - ▶ Where does the responsibility of scientists begin and end?
- ▶ Ethical considerations in research activities:
 - ▶ Experimentation with humans or animals
 - ▶ Intellectual dishonesty (fabrication of data, plagiarism)
 - ▶ Discrimination and harrassment
 - ▶ Many disciplines have specific ethical guidelines