



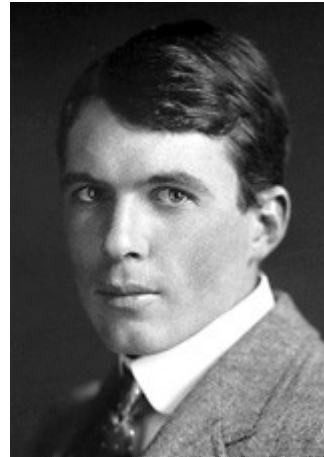
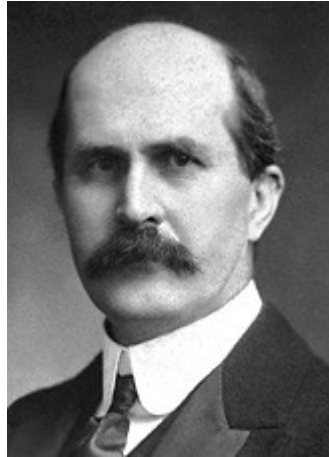
# Semantics in the Age of Corpora

Dirk Geeraerts



University of Leuven  
RU Quantitative Lexicology and Variational Linguistics

# Nobel prize Physics 1915



William Henry Bragg & William Lawrence Bragg

pioneers in using x-rays for analyzing molecular structure: W.L. led the lab at which Watson & Crick discovered the double helical structure of deoxyribonucleic acid

# Types of scientific development

## concept-driven revolutions

Newtonian physics

Darwinian evolution theory

Einstein's relativity theory

→ T Kuhn 1962, The Structure of Scientific Revolutions

# Types of scientific development

## concept-driven revolutions

Newtonian physics

Darwinian evolution theory

Einstein's relativity theory

→ T Kuhn 1962, The Structure of Scientific Revolutions

## tool-driven revolutions

Galilean cosmology, based on the telescope, based on new lens technology

molecular biology, based on x-ray technology

→ P Galison 1997, Image and Logic



# Types of scientific development in linguistics

## concept-driven developments

the phoneme

constituent structure

conversational maxims

ICMs

## tool-driven developments

not so obvious:

the notion of an algorithmic rule system, perhaps

or corpus linguistics ?

# Types of scientific development in linguistics

can the advanced quantitative approach to the growing amount of corpus data as illustrated by (e.g.)

Glynn, Dylan and Fischer, Kerstin (eds.): Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches. Mouton de Gruyter, Berlin/New York, 2010

be considered a technology-driven paradigm shift in (lexical) semantics ?

# A corpus-driven paradigm shift ?

there are at least three reasons why quantitative corpus linguistics contributes to a major shift in the study of meaning

① re **corpus**:

the growing popularity of usage-based conceptions of language cannot be understood without the availability of large amounts of actual usage data

i.e. usage-based approaches are the result, not the cause of the availability of corpora

# A corpus-driven paradigm shift ?

there are at least three reasons why quantitative corpus linguistics contributes to a major shift in the study of meaning

② re **quantitative**:

the quantitative, statistical approach allows for true hypothesis testing in linguistics, i.e. linguistics can implement the empirical cycle model of scientific research (going from theory to hypothesis to empirical falsification etc.)

# A corpus-driven paradigm shift ?

there are at least three reasons why quantitative corpus linguistics contributes to a major shift in the study of meaning

## ③ re *linguistics*:

the quantitative, statistical approach forces linguists to rethink the vocabulary we inherited from structuralist linguistics: concepts (like polysemy) that were initially defined as binary, categorical variables have to be reanalysed as continuous, fuzzy, probabilistic ones

# A corpus-driven paradigm shift ?

however, a closer look at the history of lexical semantics reveals that the quantitative corpus analysis of meaning has various links with previous developments in the history of lexical semantics



# (Lexical) Semantics in the Age of Corpora (and Before)

Dirk Geeraerts



University of Leuven  
RU Quantitative Lexicology and Variational Linguistics

# Read all about it !

the story sketched here is treated in full detail in

D. Geeraerts 2010

Theories of Lexical Semantics

Oxford University Press



# Trajectory

the **theoretical** development of lexical semantics

the **empirical** development of lexical semantics

the **methodological** development of lexical semantics



# Trajectory

the **theoretical** development of lexical semantics

what are the successive theories that dominate the historical development of lexical semantics, and how are they related?

the **empirical** development of lexical semantics

what have the successive approaches contributed to the descriptive scope of lexical semantics?

the **methodological** development of lexical semantics

what materials and what methods of analysis have the successive approaches used?

# Trajectory

the **theoretical** development of lexical semantics

the **empirical** development of lexical semantics

the **methodological** development of lexical semantics



# Trajectory

the **theoretical** development of lexical semantics  
is characterized by an almost circular succession of  
contextualizing, decontextualizing  
and then recontextualizing theoretical movements

the **empirical** development of lexical semantics

the **methodological** development of lexical semantics



# Theoretical development

stages:

- historical-philological semantics
- structuralist semantics
- neostructuralist and generativist semantics
- cognitive semantics

# Historical-philological semantics

- **period:** 1850-1930
- **people:** Bréal, Paul, Darmesteter, Nyrop, Stern etc. etc.
- **problem:** change of meaning
- **perspective:** predominantly semasiological, with a psychological conception of meaning (meaning as a mental phenomenon)
- **practice:** conceptual analysis applied to philologically documented cases of semantic change
- **product:** elaborate classifications of semantic change

# Structuralist semantics

- **period:** 1930-1970
- **people:** Weisgerber, Trier, Porzig, Coseriu, Pottier, Lyons etc.
- **problem:** the structure of the vocabulary
- **perspective:** onomasiological,  
with a system-based conception of meaning  
(language as an autonomous structure)
- **practice:** conceptual analysis applied to introspectively  
accessed words
- **product:** a conceptual apparatus for describing relations  
among words (lexical fields and semantic relations)

# Generativist semantics

- **period:** 1960-1970
- **people:** Katz
- **problem:** the formalization of lexical meaning in a generative model of the grammar
- **perspective:** semasiological, with a generative conception of language (language as a rule system with cognitive reality)
- **practice:** conceptual analysis applied to introspectively accessed words
- **product:** examples of formal dictionary entries ('bachelor')



# Neostructuralist semantics

- **period:** 1970 →
- **people:** Jackendoff, Pustejovsky, Wierzbicka, Mel'cuk a.o.
- **problem:** meaning within the 'first cognitive revolution'
- **perspective:** mostly semasiological, with a reductionist conception of the relation between language and cognition
- **practice:** formal and semi-formal description applied to introspectively accessed words
- **product:** models for formal and semi-formal description; lexicons

# Neostructuralist semantics

- period: 1970 →
  - people: Jackendoff, Pust
  - problem: meaning with
  - perspective: mostly s  
with a **reductionist** e  
language and cognition
  - practice: formal and se  
introspectively access
  - product: models for for  
lexicons
- maintaining a strict distinction between linguistic system and (encyclopedic) cognition and/or
  - relying on a restrictive vocabulary of 'primitives' (meaning components or semantic relations) and/or
  - couching semantic description in a formal language

# Cognitive semantics

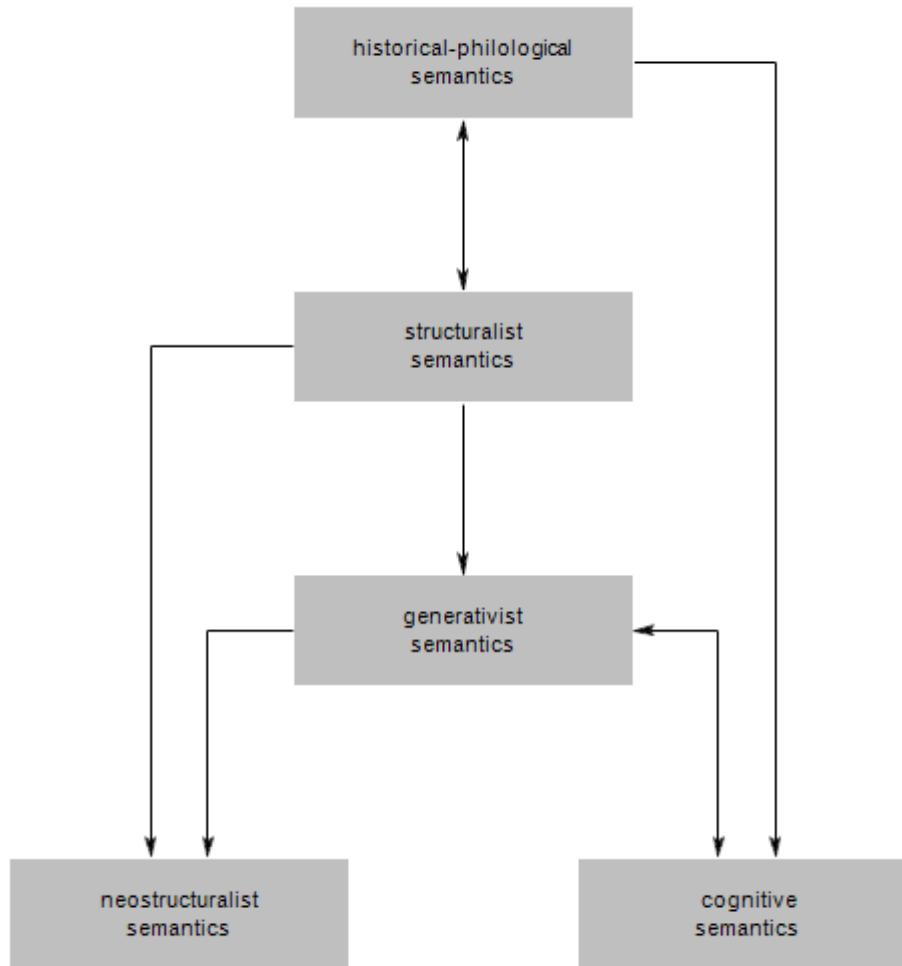
- **period:** 1980 →
- **people:** us
- **problem:** meaning within the ‘second cognitive revolution’
- **perspective:** semasiological and onomasiological, with a non-reductionist conception of the relation between language and cognition
- **practice:** from conceptual analysis to (quantitative) corpus analysis and experimental research
- **product:** concepts and models like prototypicality, conceptual metaphor, conceptual metonymy, image schemas etc.

# Cognitive semantics

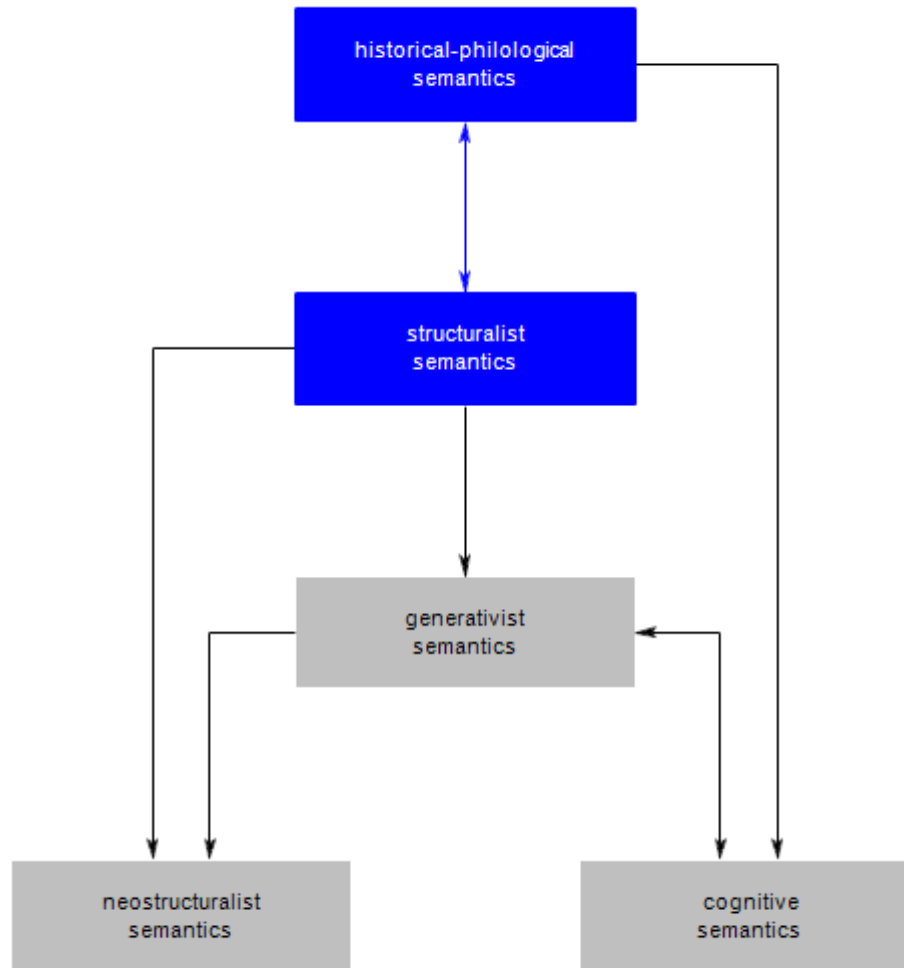
- period: 1980 →
- people: us
- problem: meaning within the ‘second cognitive revolution’
- perspective: semasiological and onomasiological, with a non-reductionist conception of the relationship between language and cognition
- practice: from conceptual analysis and experimental
- product: concepts, metaphor, conceptual

- schema-based vs rule-based
- probabilistic processing vs symbol-manipulating algorithms
- embodied vs symbolic
- socially situated vs cognition-internal

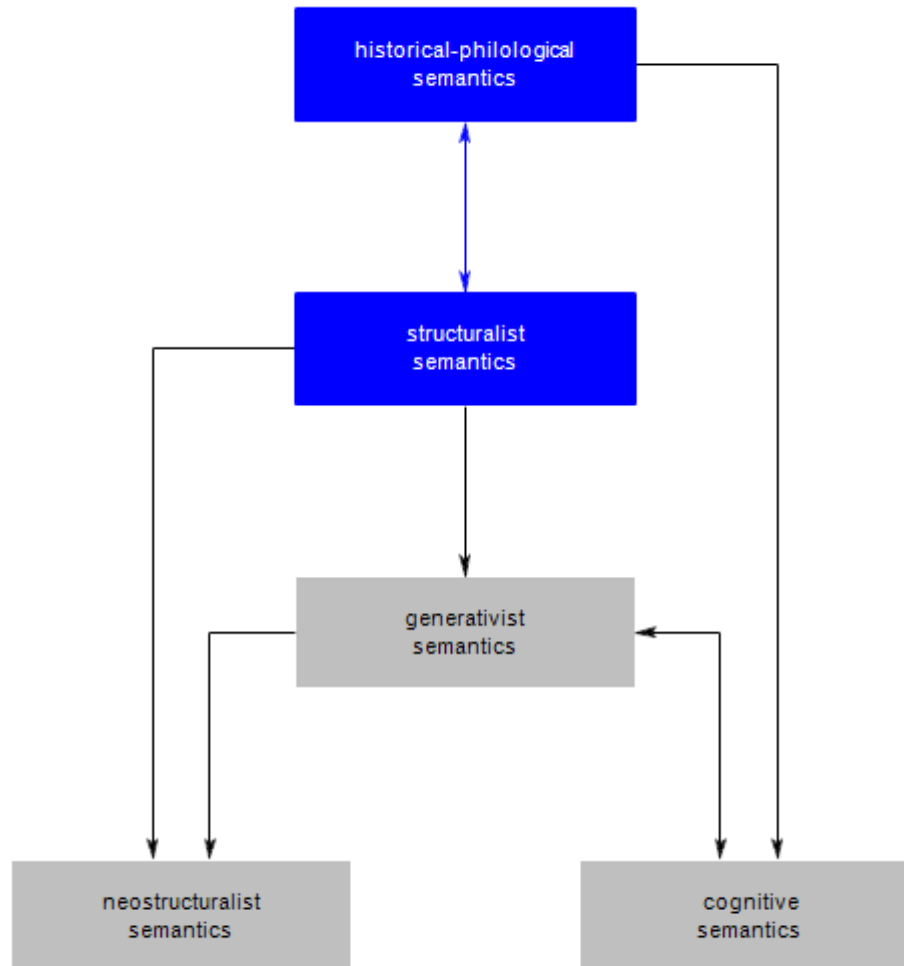
# Theoretical links



# Theoretical links



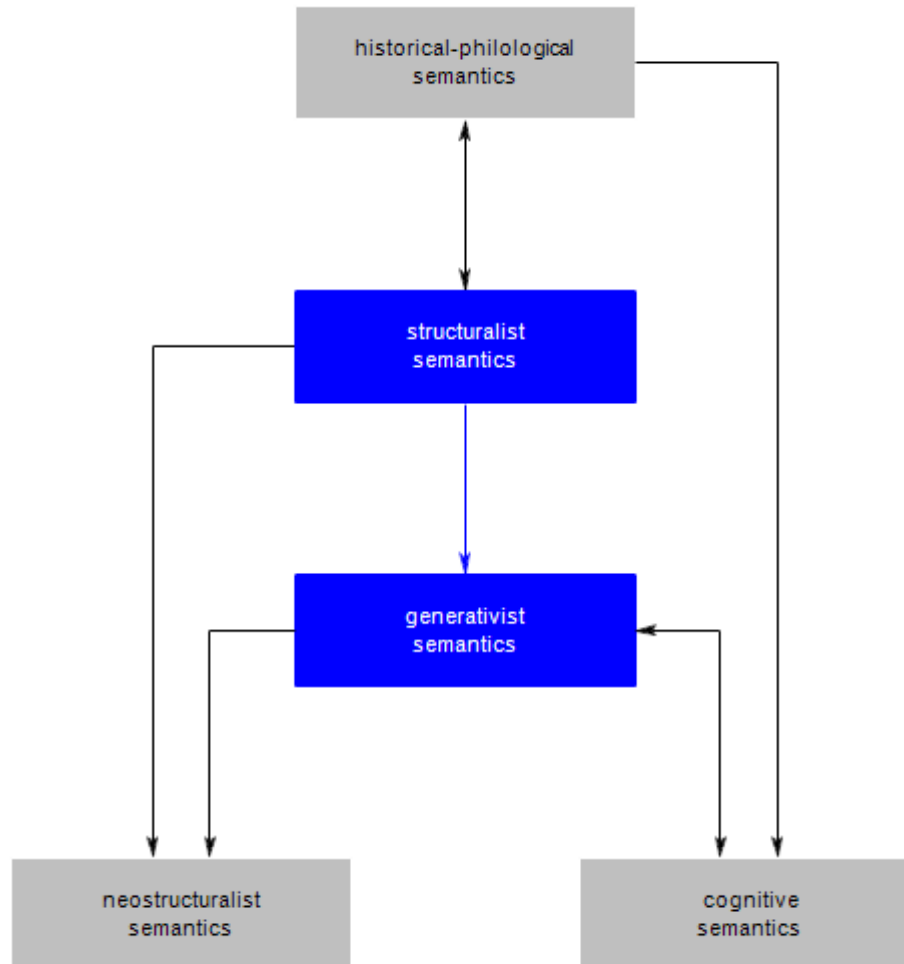
# Theoretical links



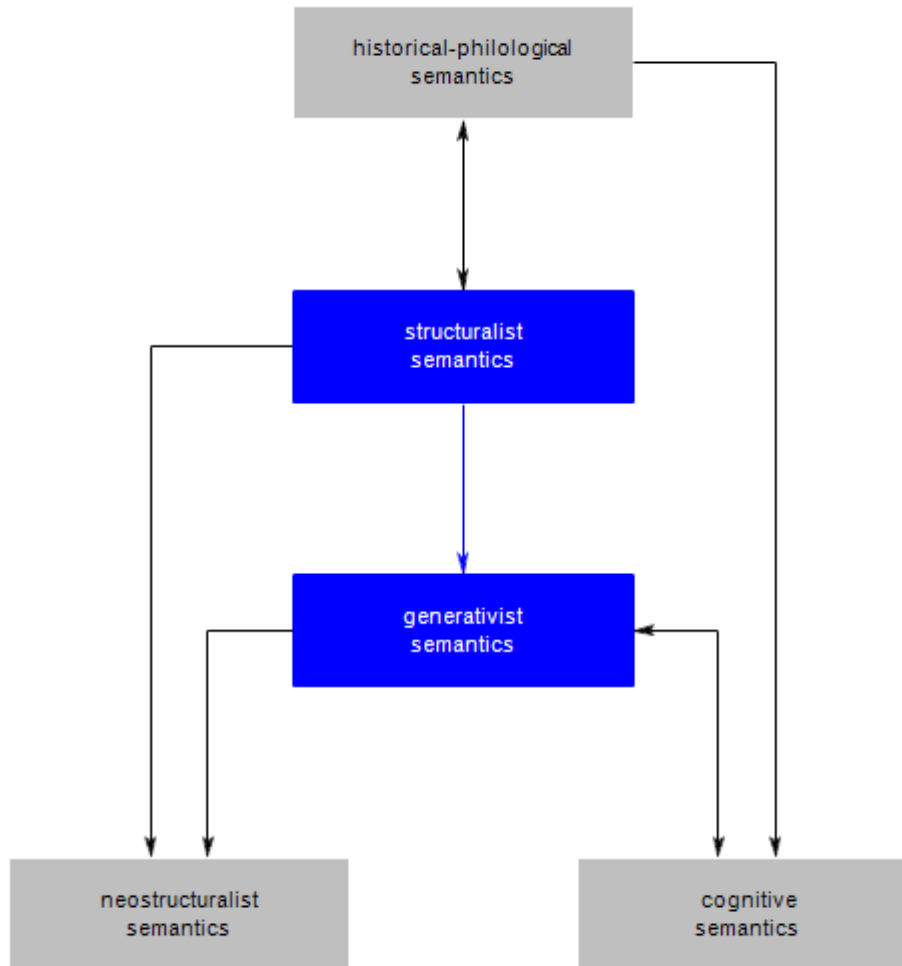
a step of opposition:

- introducing the notion of an autonomous system
- introducing the notion of 'primitive' concepts and relations

# Theoretical links

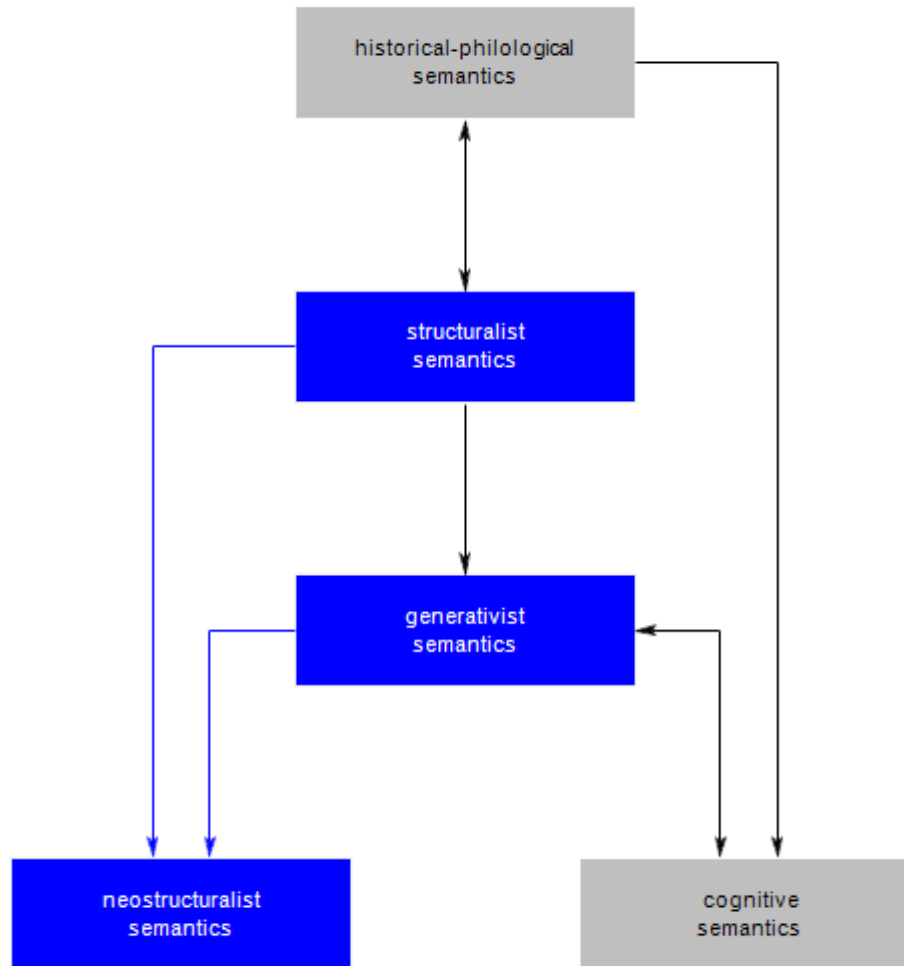


# Theoretical links

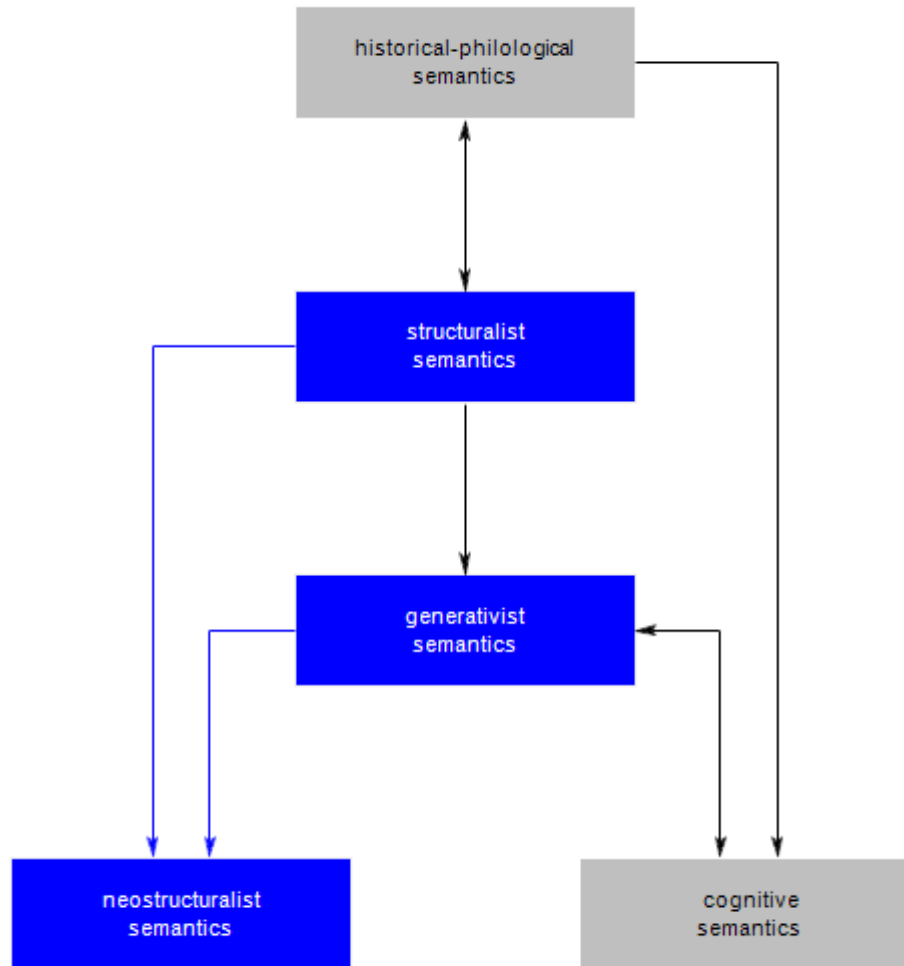


a step of development:  
adding the notion of a  
formalized description

# Theoretical links

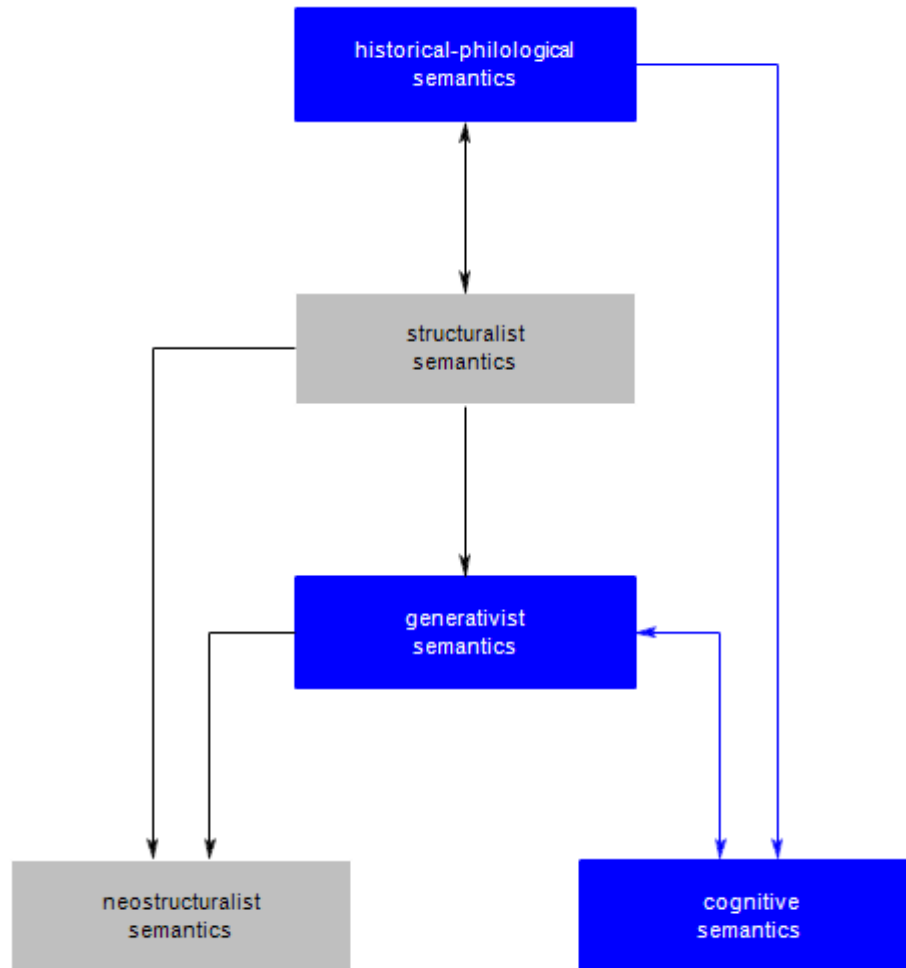


# Theoretical links

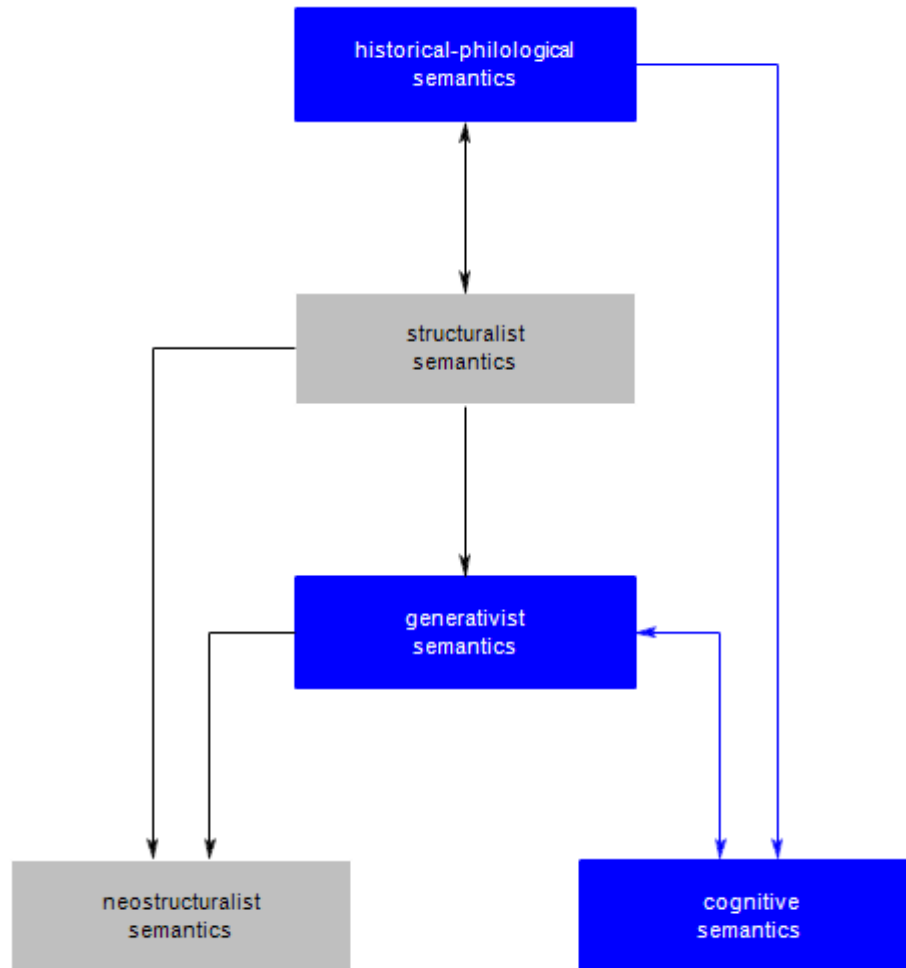


a step of development:  
further developing the idea  
of a systemic, reductionist  
description

# Theoretical links

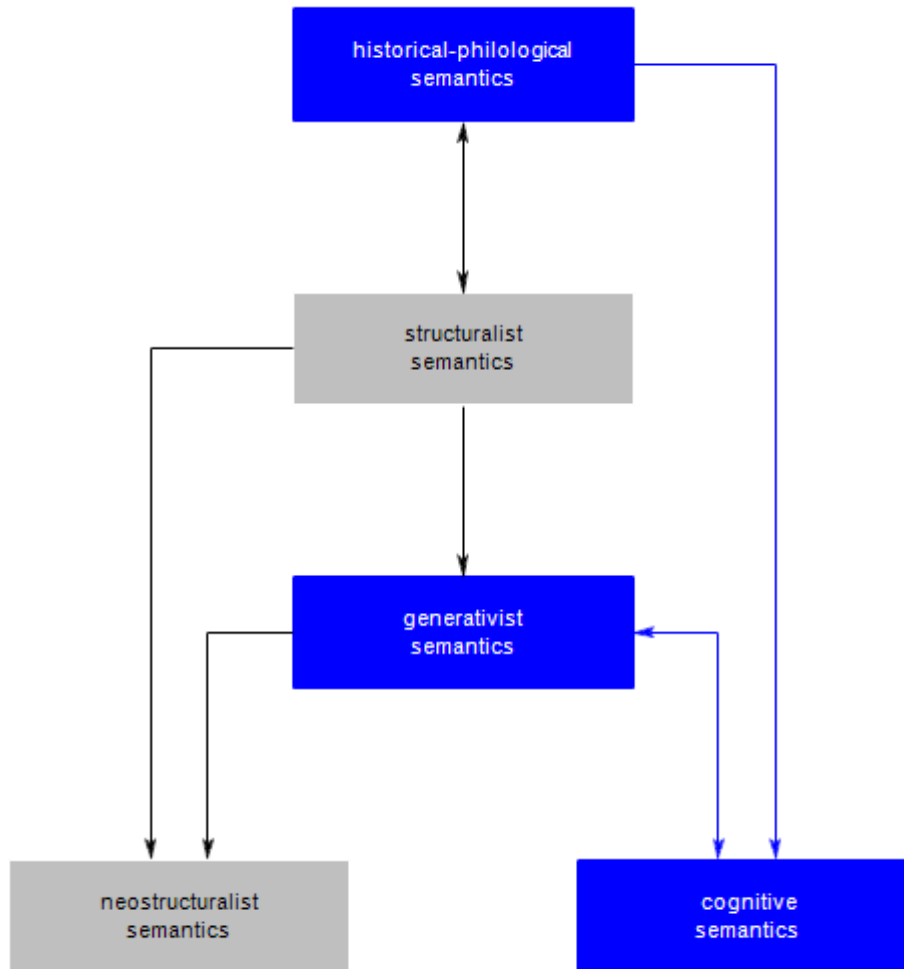


# Theoretical links



a step of opposition:  
rejecting a reductionist  
approach to meaning  
and

# Theoretical links



a step of development:  
returning to the concerns of  
the prestructuralist tradition

- flexibility of meaning and the relation between usage and change
- metaphor, metonymy etc.
- contextual and experiential grounding

# Trajectory

the **theoretical** development of lexical semantics

the **empirical** development of lexical semantics

the **methodological** development of lexical semantics



# Trajectory

the **theoretical** development of lexical semantics

the **empirical** development of lexical semantics  
is characterized by a systematic broadening of the  
descriptive range of the discipline

the **methodological** development of lexical semantics

# Empirical expansion

the empirical development of lexical semantics may be charted on two levels:

- globally, looking at the distinction between semasiology and onomasiology and the distinction between qualitative and quantitative approaches
- specifically, by zooming in on the types of relations that play a role in the qualitative approach

it will become clear that all approaches have contributed significantly to the evolution



# Empirical expansion

	semasiology	onomasiology
investigating qualitative structure: elements and relations	senses and semantic links among senses (metaphor, metonymy etc.)	semantic relations among lexical items (fields, taxonomies, lexical relations etc.)
investigating use and its quantitative effects: salience and fuzziness	prototype effects within senses and polysemous items	salience effects between categories, pragmatic onomasiology

# Empirical expansion

	semasiology	onomasiology
investigating qualitative structure: elements and relations	senses and semantic links among senses (metaphor, metonymy etc.) ↑ historical-philological semantics	semantic relations among lexical items (fields, taxonomies, lexical relations etc.)
investigating use and its quantitative effects: salience and fuzziness	prototype effects within senses and polysemous items	salience effects between categories, pragmatic onomasiology

# Empirical expansion

	semasiology	onomasiology
investigating qualitative structure: elements and relations	<p>senses and semantic links among senses (metaphor, metonymy etc.) ↑ historical-philological semantics</p>	<p>semantic relations among lexical items (fields, taxonomies, lexical relations etc.) ↑ (neo)structuralist semantics</p>
investigating use and its quantitative effects: salience and fuzziness	<p>prototype effects within senses and polysemous items</p>	<p>salience effects between categories, pragmatic onomasiology</p>

# Empirical expansion

	semasiology	onomasiology
investigating qualitative structure: elements and relations	senses and semantic links among senses (metaphor, metonymy etc.) ↑ historical-philological semantics	semantic relations among lexical items (fields, taxonomies, lexical relations etc.) ↑ (neo)structuralist semantics
investigating use and its quantitative effects: salience and fuzziness	prototype effects within senses and polysemous items ↑ cognitive semantics	salience effects between categories, pragmatic onomasiology

# Empirical expansion

	semasiology	onomasiology
investigating qualitative structure: elements and relations	<p>senses and semantic links among senses (metaphor, metonymy etc.) ↑ historical-philological semantics</p>	<p>semantic relations among lexical items (fields, taxonomies, lexical relations etc.) ↑ (neo)structuralist semantics</p>
investigating use and its quantitative effects: salience and fuzziness	<p>prototype effects within senses and polysemous items ↑ cognitive semantics</p>	<p>salience effects between categories, pragmatic onomasiology ↑ cognitive semantics</p>

# Empirical expansion

	semasiology	onomasiology
investigating qualitative structure: elements and relations	senses and semantic links among senses (metaphor, metonymy etc.) ↑ historical-philological semantics	semantic relations among lexical items (fields, taxonomies, lexical relations etc.) ↑ (neo)structuralist semantics
investigating use and its quantitative effects: salience and fuzziness	prototype effects within senses and polysemous items ↑ cognitive semantics	salience effects between categories, pragmatic onomasiology ↑ cognitive semantics

# Empirical expansion

	semasiological relations	onomasiological relations
inclusion	specialization/generalization	taxonomies
literal similarity	family resemblances	lexical fields
figurative similarity	lexical metaphor	conceptual metaphor
contiguity	metonymy	frames

# Empirical expansion

	semasiological relations	onomasiological relations
inclusion	specialization/generalization ↑ historical-philological semantics	taxonomies
literal similarity	family resemblances	lexical fields
figurative similarity	lexical metaphor ↑ historical-philological semantics	conceptual metaphor
contiguity	metonymy ↑ historical-philological semantics	frames

# Empirical expansion

	semasiological relations	onomasiological relations
inclusion	specialization/generalization ↑ historical-philological semantics	taxonomies ↑ structuralist semantics
literal similarity	family resemblances	lexical fields ↑ structuralist semantics
figurative similarity	lexical metaphor ↑ historical-philological semantics	conceptual metaphor
contiguity	metonymy ↑ historical-philological semantics	frames

# Empirical expansion

	semasiological relations	onomasiological relations
inclusion	specialization/generalization ↑ historical-philological semantics	taxonomies ↑ structuralist semantics
literal similarity	family resemblances ↑ cognitive semantics	lexical fields ↑ structuralist semantics
figurative similarity	lexical metaphor ↑ historical-philological semantics	conceptual metaphor ↑ cognitive semantics
contiguity	metonymy ↑ historical-philological semantics	frames ↑ cognitive semantics

# Empirical expansion

	semasiological relations	onomasiological relations
inclusion	specialization/generalization ↑ historical-philological semantics	taxonomies ↑ structuralist semantics
literal similarity	family resemblances ↑ cognitive semantics	lexical fields ↑ structuralist semantics
figurative similarity	lexical metaphor ↑ historical-philological semantics	conceptual metaphor ↑ cognitive semantics
contiguity	metonymy ↑ historical-philological semantics	frames ↑ cognitive semantics

# Trajectory

the **theoretical** development of lexical semantics

the **empirical** development of lexical semantics

the **methodological** development of lexical semantics



# Methodological continuity

there are at least three reasons why quantitative corpus linguistics continues earlier developments in lexical semantics

- ① the use of quantitative corpus evidence constitutes a **theoretical continuity with historical-philological semantics**

the cognitive, usage-based conception of language is a return to the concerns of the prestructuralist, historical-philological tradition

# Methodological continuity

there are at least three reasons why quantitative corpus linguistics continues earlier developments in lexical semantics

- ② the use of quantitative corpus evidence constitutes a **methodological continuity with historical-philological semantics**

textual materials are the oldest kind of empirical evidence in the development of lexical semantics: you cannot do diachronic semantics unless you have textual evidence

other kinds of evidence (referential, experimental) came later



# Methodological continuity

there are at least three reasons why quantitative corpus linguistics continues earlier developments in lexical semantics

- ③ the use of quantitative corpus evidence constitutes a **methodological continuity with (some forms of) structuralist semantics**

the lineage of quantitative corpus semantics:

# The lineage of corpus analysis

step 1

structuralist semantics basically studies paradigmatic relations  
(fields, components, lexical relations);

less conspicuously, syntagmatic relations are included:

Porzig 'wesenhafte Bedeutungsbeziehungen'

Coseriu 'lexikalische Solidaritäten'

Firth 'you shall know a word by the company it keeps'

# The lineage of corpus analysis

step 2

in his work on the Cobuild dictionary, John Sinclair develops the Firthian idea into a theory of collocations and collocational behaviour as a window onto the meaning of items;

statistical measures of collocability are developed for studying collocational behaviour in corpora (Church & Hanks)

# The lineage of corpus analysis

step 3

quantitative analysis of corpus distributions as an inroad to semantics is – largely independently – further developed in two fields:

computational linguistics, using vector space models

cognitive linguistics, using collocation analysis – and further developments of statistics

(current work at QLVL, viz. by Kris Heylen and Yves Peirsman, aims at merging both traditions)



# Conclusion

1

the development of lexical semantics is characterized by

- a theoretical oscillation between reductionist and maximalist approaches
- a systematic descriptive expansion of the field, in which all the approaches participate
- a growing importance of empirical methods, with specifically, a continuity between recent corpus approaches and older forms of investigation

# Conclusion

2

would it be so different in physics ?

the development of x-ray technology which revolutionized molecular biology was not a coincidence, but was itself theoretically inspired

hence, we need a version of the empirical cycle that not only leads from theories to **hypotheses** and then to experiments and analysis

but also from theories to **tools** and then to experiments and analysis



# Conclusion

SO ...

will quantitative corpus analysis discover  
the DNA of meaning?