

Generative Lexicon and CoreLex

Lexical semantics

Analysis & representation of word meaning

A generative model of lexical semantics

Representation of word meaning that enables dynamic creation of word meanings ('senses') on demand

An empirical foundation of the generative model

Analysis of sense distribution across a large-scale semantic lexicon

An ontological view of lexical semantics

Reasoning over the ontology enables sense derivation



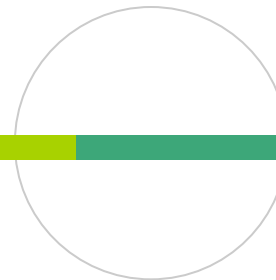
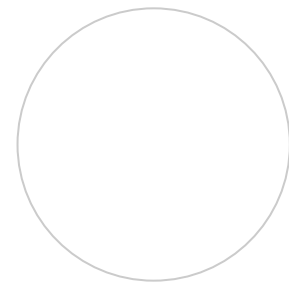
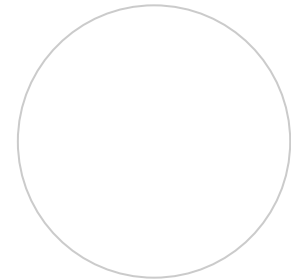
Lexical Semantics

word meaning, senses

lexical semantic ambiguity

systematic polysemy

type coercion, metonymy, bridging



Word Meaning

What is the meaning of 'ball'? (as a noun)

<http://dictionary.reference.com/browse/ball>

ball¹

- 1.a spherical or approximately spherical body or shape. *He rolled the piece of paper into a ball.*
- 2.a round or roundish body, of various sizes and materials, either hollow or solid, for use in games, as baseball, football, tennis, or golf.
- 3.a game played with a ball, esp. baseball: *The boys are out playing ball.*
- 4.Military. a. a solid, usually spherical projectile for a cannon, rifle, pistol, etc., as distinguished from a shell. b. projectiles, esp. bullets, collectively.
- 5.Horticulture. a compact mass of soil covering the roots of an uprooted tree or other plant.
- 6.Literary. a planetary or celestial body, esp. the earth.
- 7.Mathematics. (in a metric space) the set of points whose distance from the zero element is less than, or less than or equal to, a specified number.

ball²

- 1.a large, usually lavish, formal party featuring social dancing and sometimes given for a particular purpose, as to introduce debutantes or benefit a charitable organization.
- 2.Informal. a thoroughly good time: *Have a ball on your vacation!*

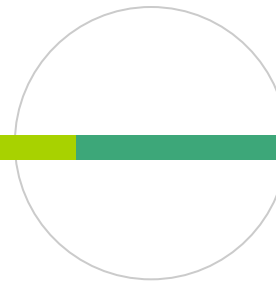
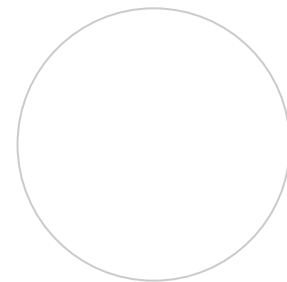
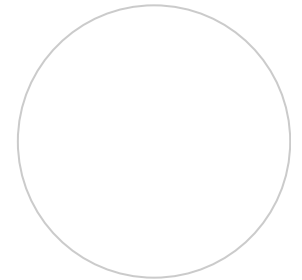
Word Meaning

What is the meaning of 'ball'? (as a noun)

<http://dictionary.reference.com/browse/ball>

spherical body or shape

lavish, formal party featuring social dancing



Lexical Semantic Ambiguity

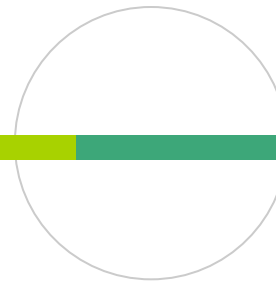
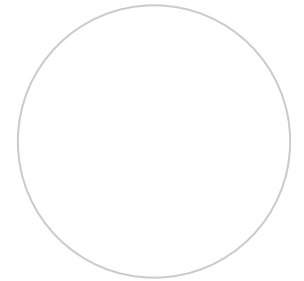
Artifact

The ball went over the fence

Event

The ball went on into the late hours

➤ unrelated senses – “homonymy”



Lexical Semantic Ambiguity

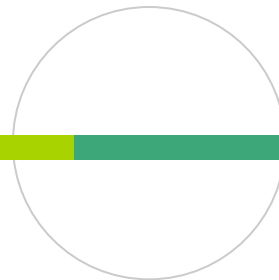
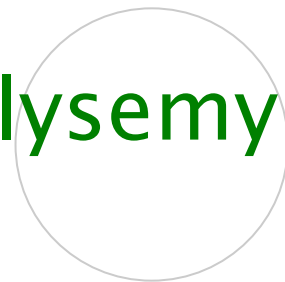
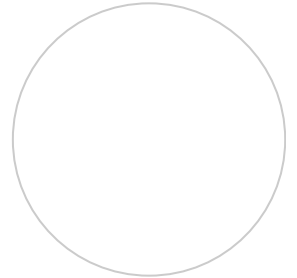
Artifact

The ball went over the fence

Event

The boys are out playing ball

➤ related senses – “systematic polysemy”



Systematic Polysemy

Building

The Boston office has been newly decorated

Organization

The Boston office was founded in 1985

Group-of-People

The Boston office called

➤ related senses – “systematic polysemy”



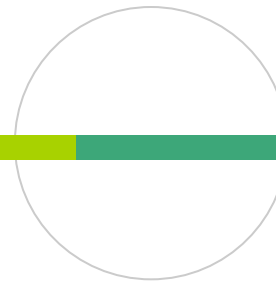
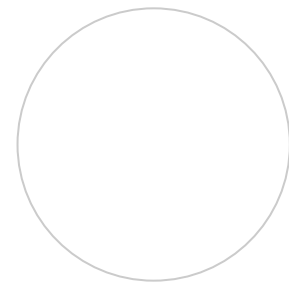
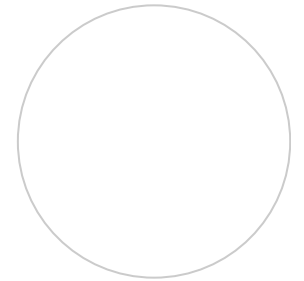
Systematic Polysemy

Referred to in the literature as:

‘regular polysemy’ (Apresjan 1973)

‘logical polysemy’ (Pustejovsky 1991, 1995)

‘systematic polysemy’ (Nunberg & Zaenen 1992)



Systematic Polysemy

Bierwisch, Manfred: 1983, 'Semantische und konzeptuelle Repraesentation lexikalischer Einheiten', in R. Ruzicka and W.Motsch (eds.), *Untersuchungen zur Semantik* (*Studia Grammatika* 22), pp. 61-99. Akademie Verlag, Berlin

A group of people

The school went for an outing

A learning process

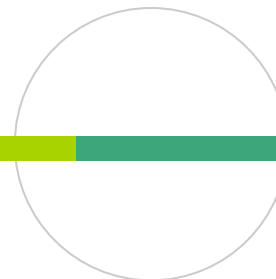
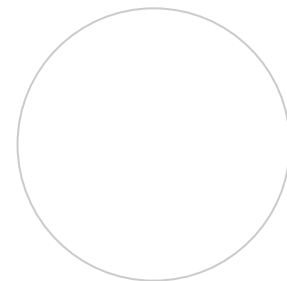
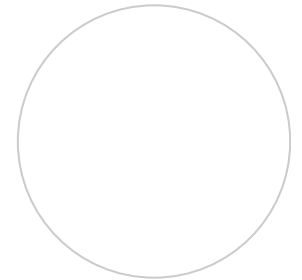
School starts at 8:30

An institution

The school was founded in 1910

A building

The school has a new roof

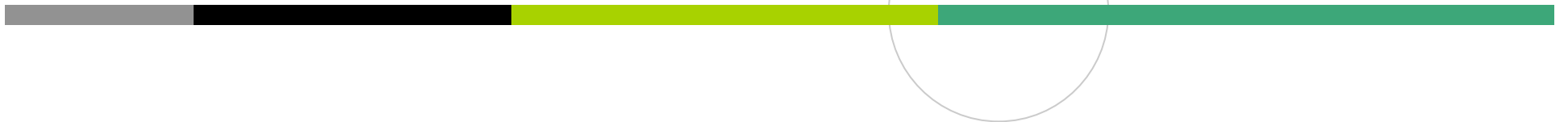


Systematic Polysemy

Hobbs, J. R. (1992). Metaphor and abduction. In A. Ortony, J. Salck, O. Stock (eds.) *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, p35–58. Springer, Berlin

The Boston office called.

office:Organization coerced-into office:Group-of-People



Type Coercion & Metonymy

Pustejovsky (1989) – Logical Polysemy

The Boston office called.

Coerce type of 'office' from Organization into Person

Metonymy – interpret a part as representing the whole

Person works-at Organization (person part-of office)

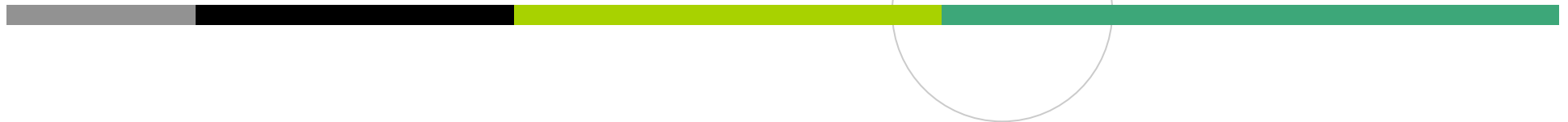


Coercion & Discourse Analysis

The Boston office called. They signed a new contract.

Co-reference resolution between 'office' and 'they'

Coerce referent of 'they' to metonymic **person** of 'office'

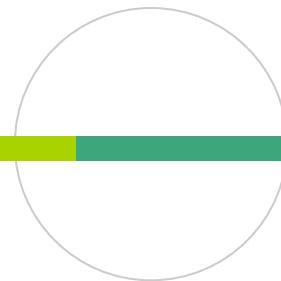
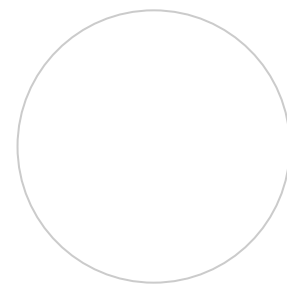
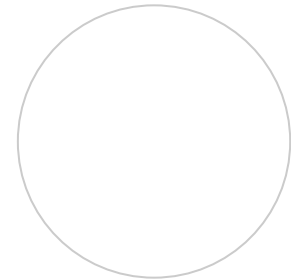


Bridging

Peter bought a car. The engine runs well.

Accommodation of 'the engine' to 'a car'

Lexical semantic inference: *engine part-of car*



Underspecified Discourse Referents

A long book heavily weighted with military technicalities, in this edition it is neither so long nor so technical as it was originally.

*[A long book heavily weighted with military technicalities]*_{NP:event-physical_object-content}

Event

„a long book...“, „it is neither so long...“

> takes long to read – not physical

length

Physical-object

„heavily weighted...“

> the physical weight of the book

Content

„military technicalities...“, „nor so technical...“

> the content is technical

Generative Lexical Semantics

‘Generative Lexicon Theory’

‘Qualia Structure’

Pustejovsky (1991), Pustejovsky and Boguraev (1993),
Pustejovsky (1995), Busa (1996), Bouillon (1997), Buitelaar
(1998), Asher and Pustejovsky (2008),
Pustejovsky (2011), Pustejovsky, Rumshisky, and Hanks (2004)

Type Coercion

I began the book

Type coercion: **direct-object** of ‘begin’ requires an **event**

Infer an event from the lexical semantics of “book” as represented by its ‘Qualia Structure’ (Pustejovsky 1995)

For example: *I began (reading) the book*

“there is a system of relations that characterizes the semantics of nominals very much like the argument structure of a verb ... Essentially the qualia structure of a noun determines its meaning as much as the list of arguments determines a verb’s meaning.” (Pustejovsky 1989)

Qualia Structure for 'book'

Formal (inheritance: is-a, hyponymy)

physical-object, content, ...

Constitutive (modification: part-of, meronymy)

section, ...

Telic (purpose: 'what is the object used for')

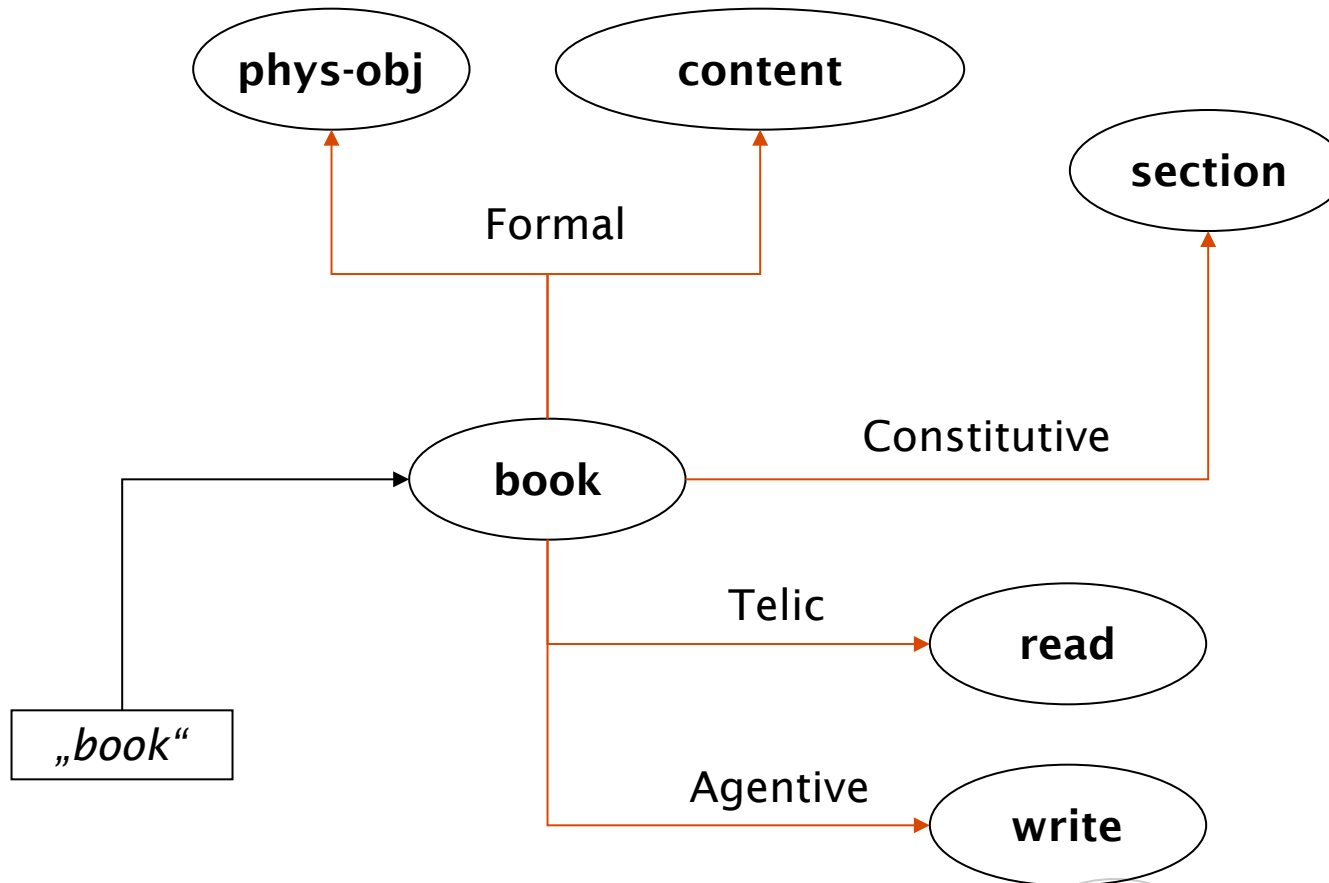
read, ...

Agentive (causality: 'how did the object originate')

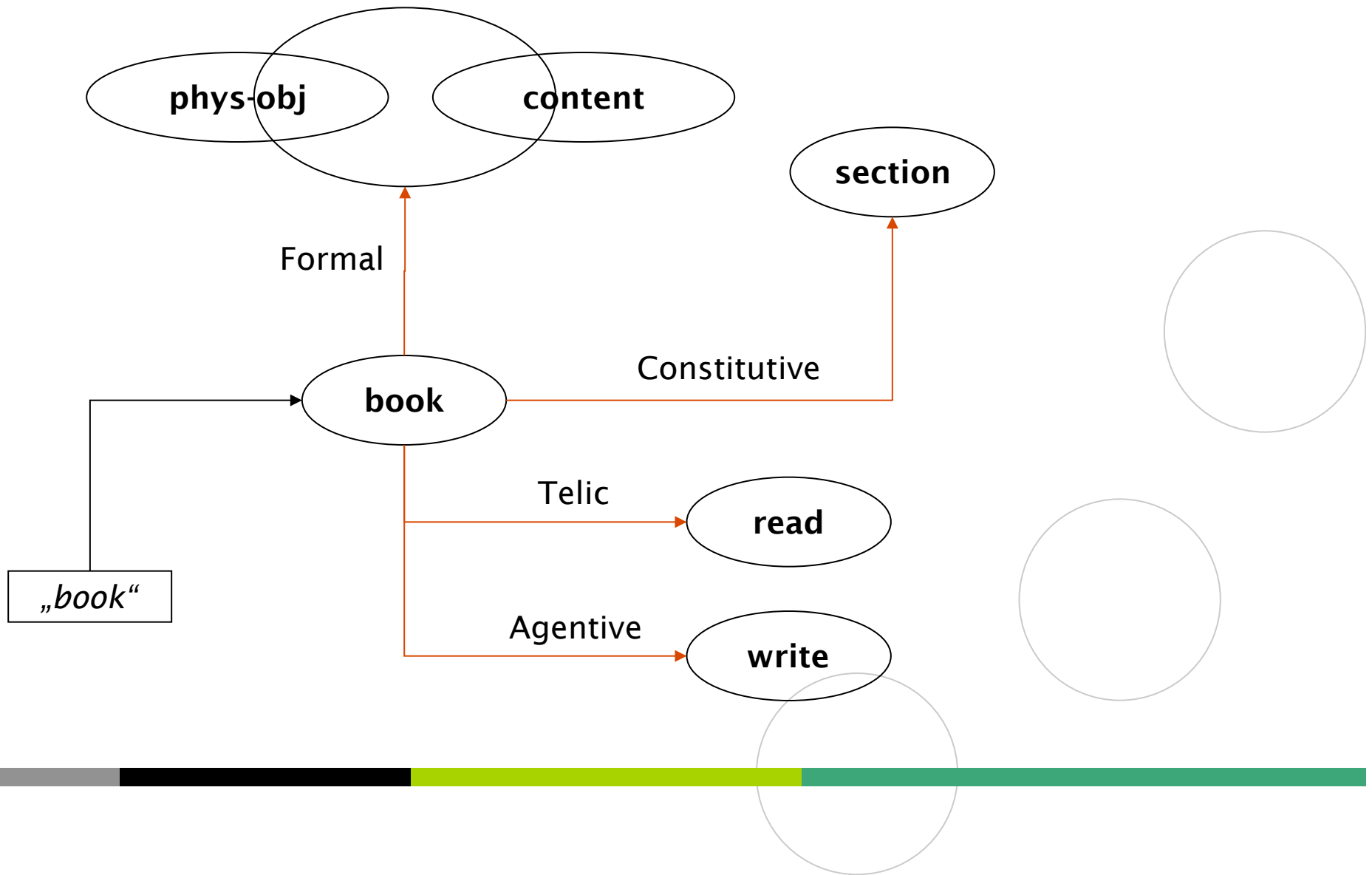
write, ...



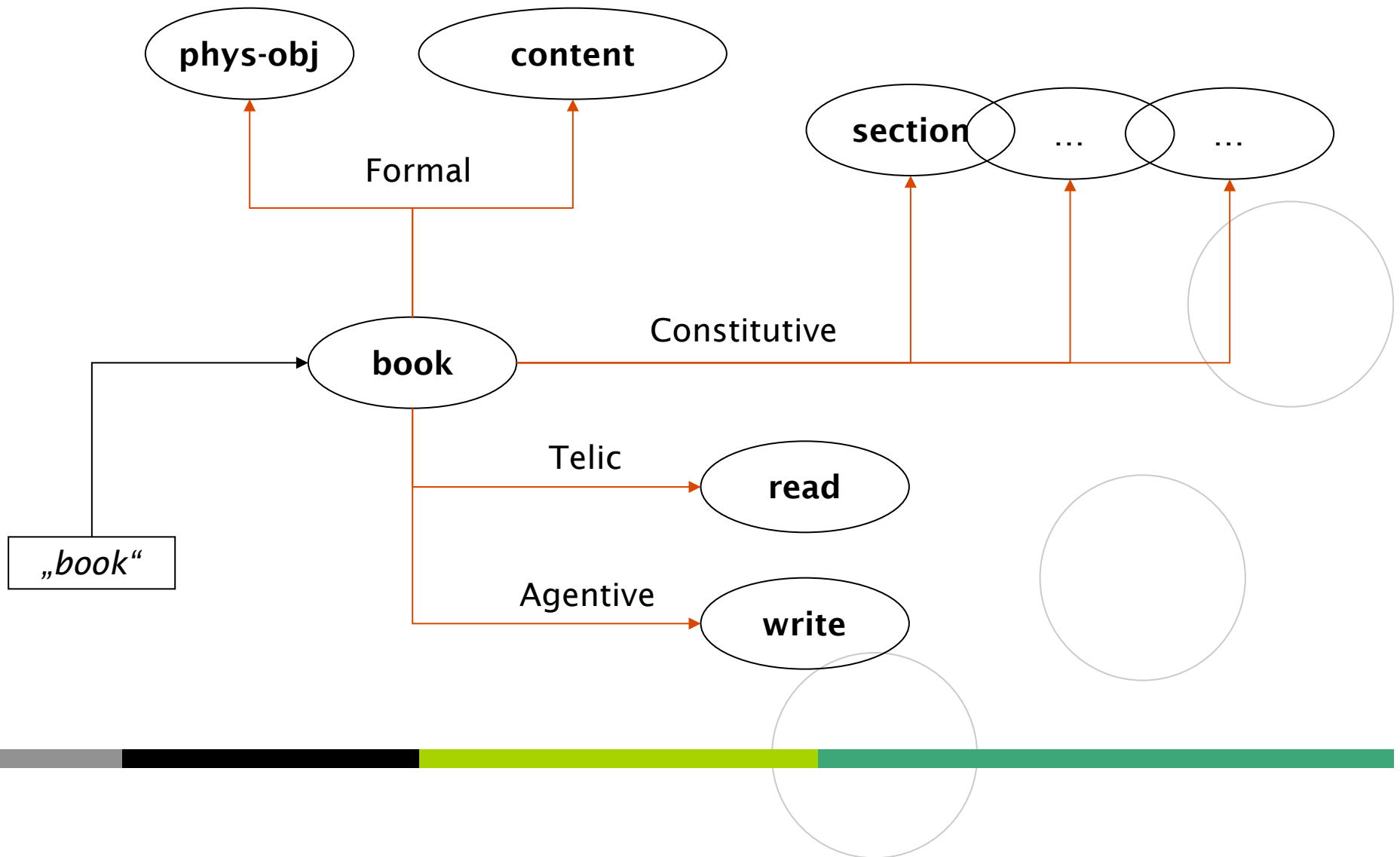
Qualia Structure for 'book'



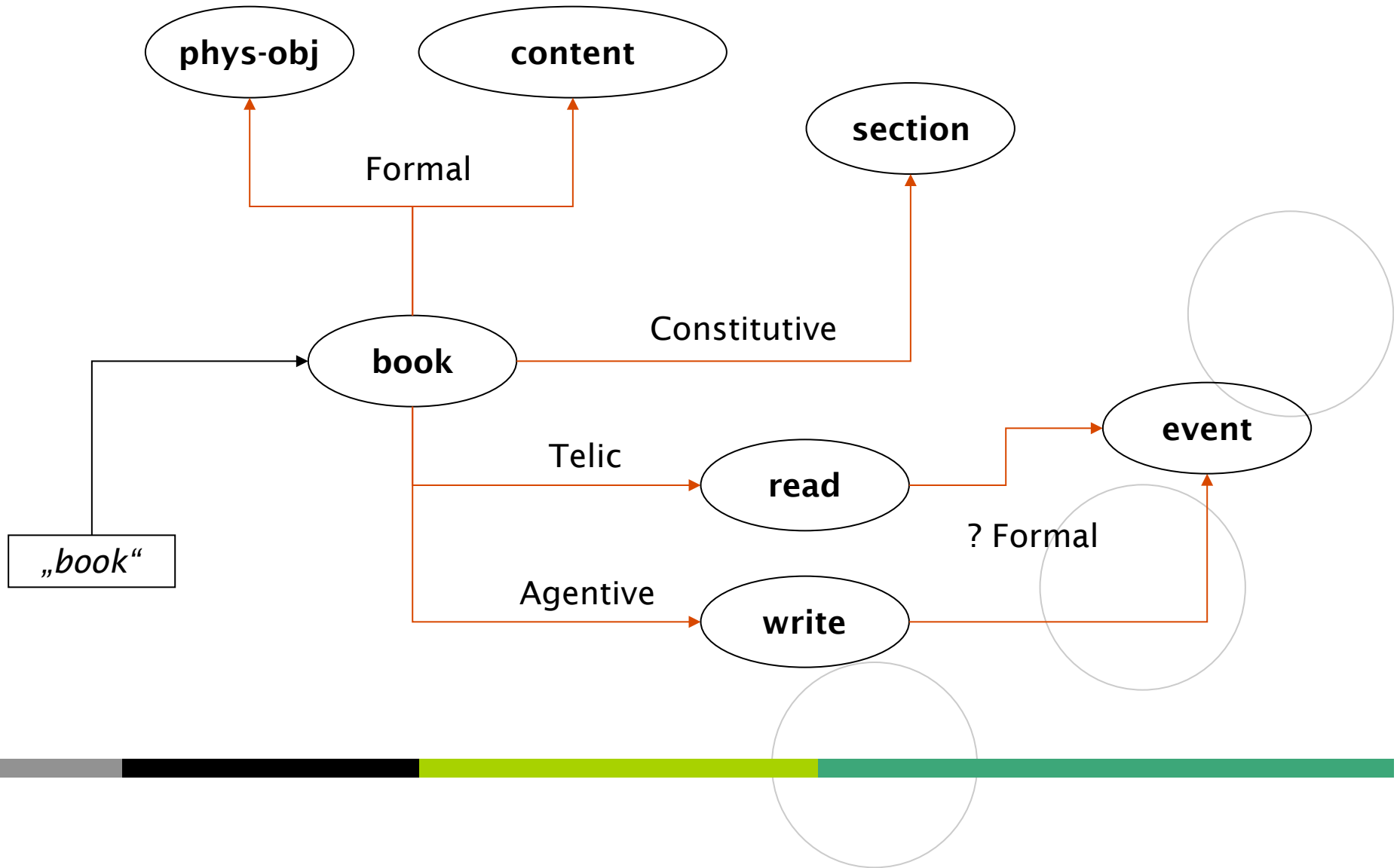
Problematic Issues - Formal



Problematic Issues - Constitutive

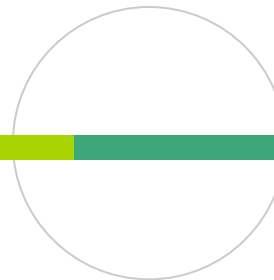
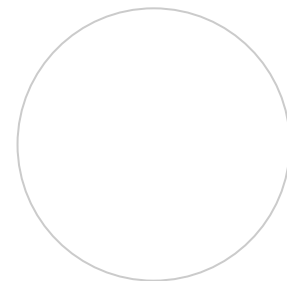
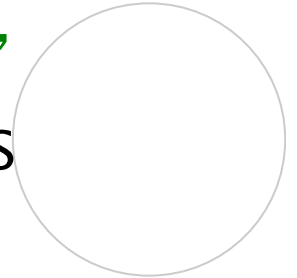


Problematic Issues - Telic/Agentive



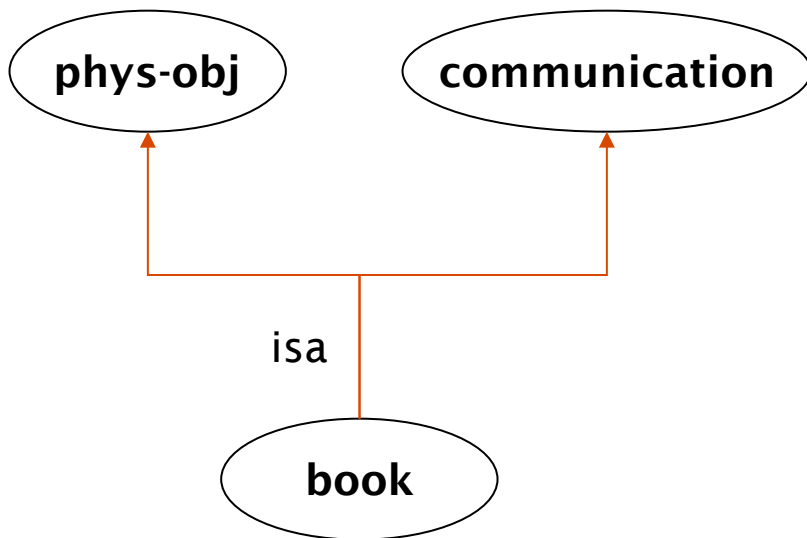
Two Approaches

- Treat QS as a ‘condensed ontology’
 - QS provides a gateway in meaning potential
 - QS roles as ‘shortcuts’ for ontology inference paths
- Condense QS even further into a ‘complex class’
 - Aggregate all types that can be reached through the QS (ontology) into a ‘systematic polysemous class’
 - Each systematic polysemous class introduces a set of underspecified lexical semantic objects
 - CoreLex approach (‘sense clustering’)



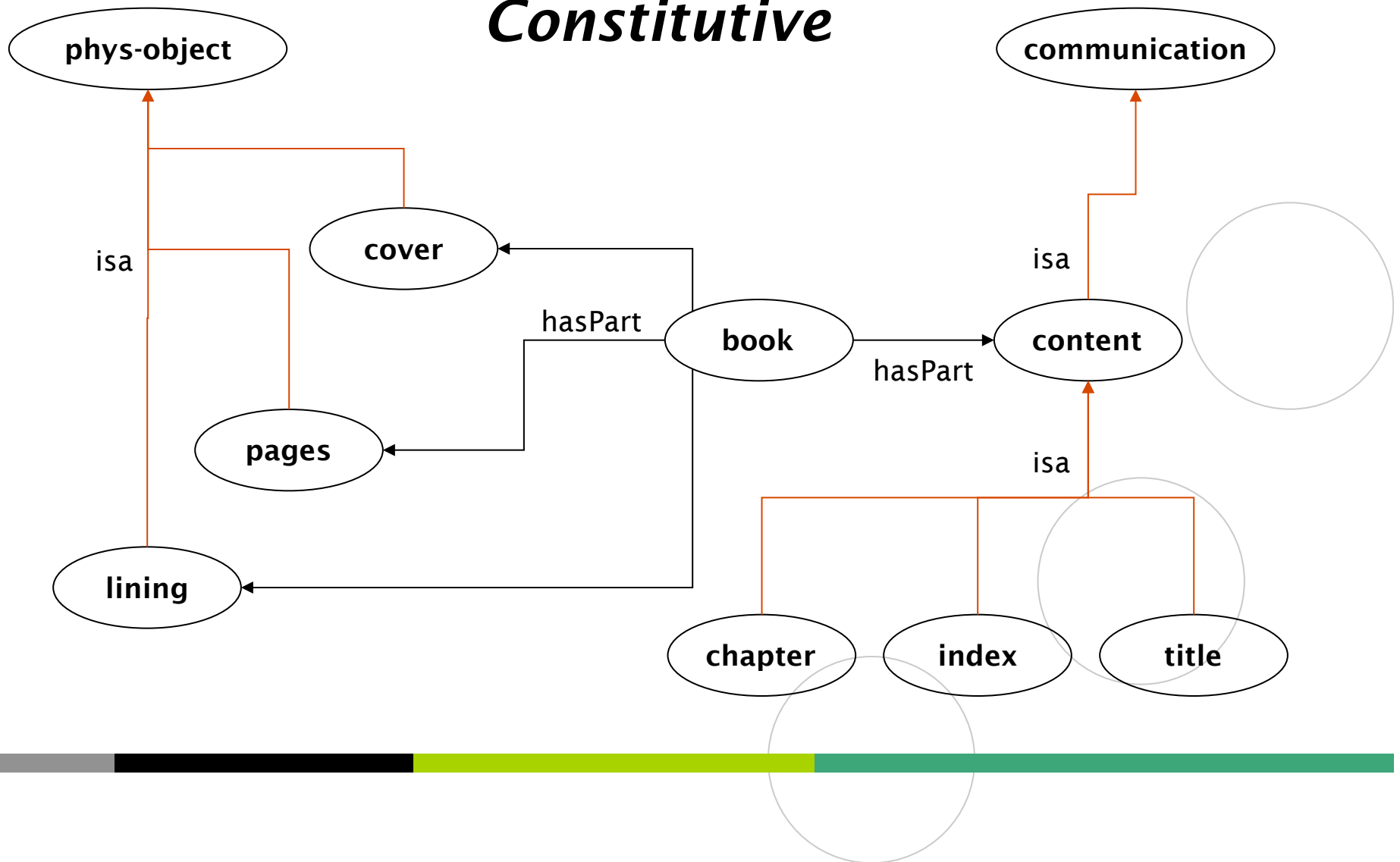
QS as 'Condensed Ontology'

Formal



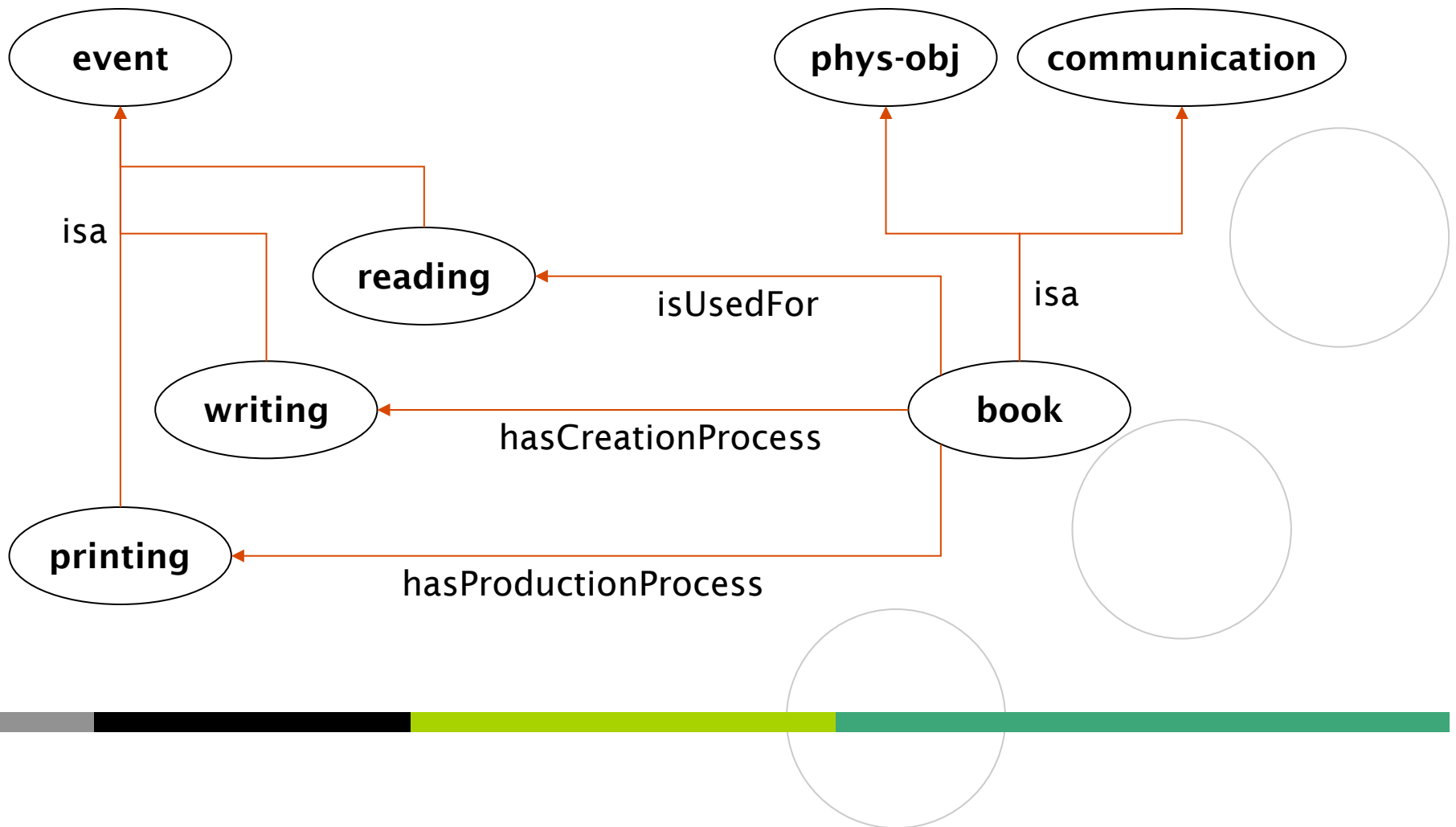
QS as 'Condensed Ontology'

Constitutive



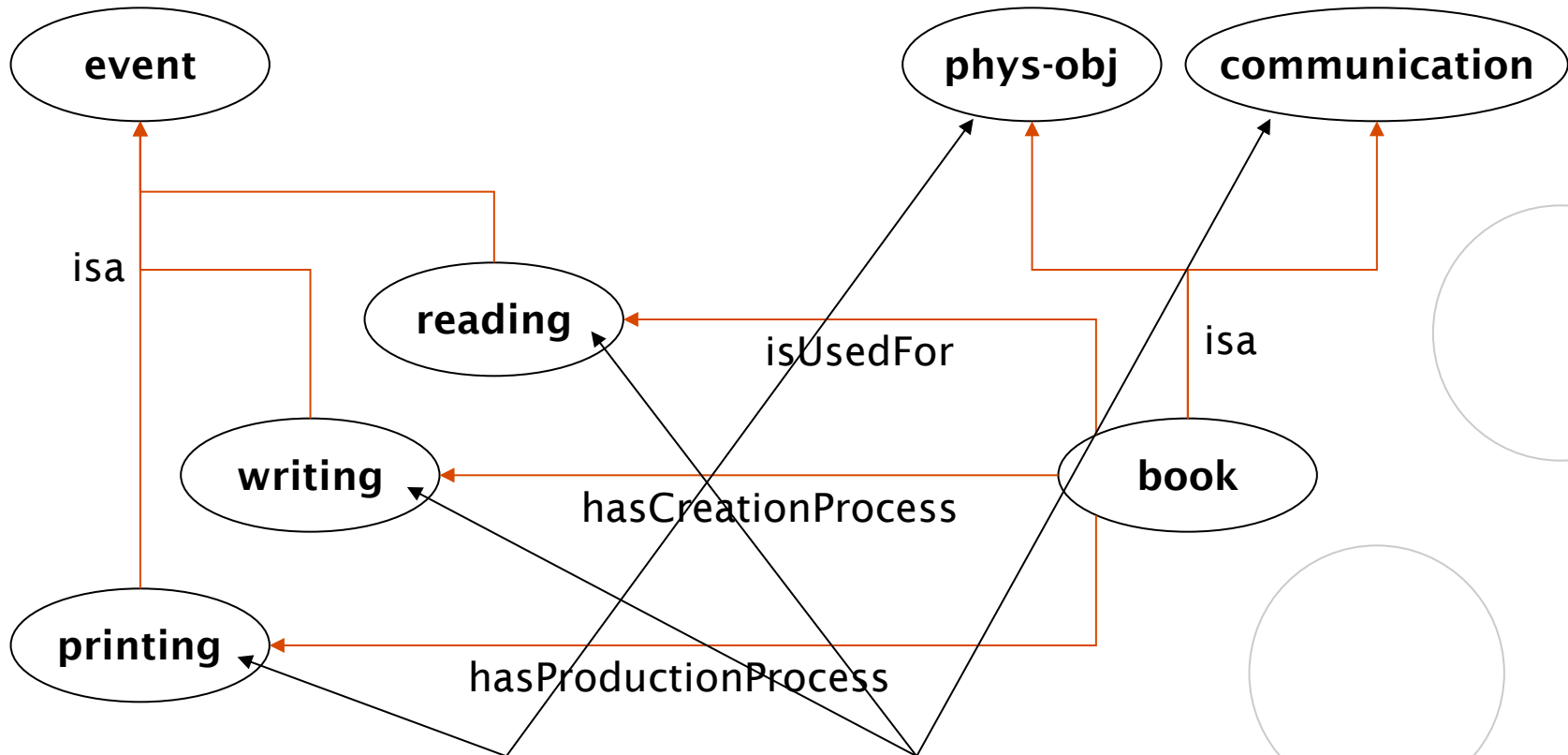
QS as 'Condensed Ontology'

Agentive & Telic



QS as 'Condensed Ontology'

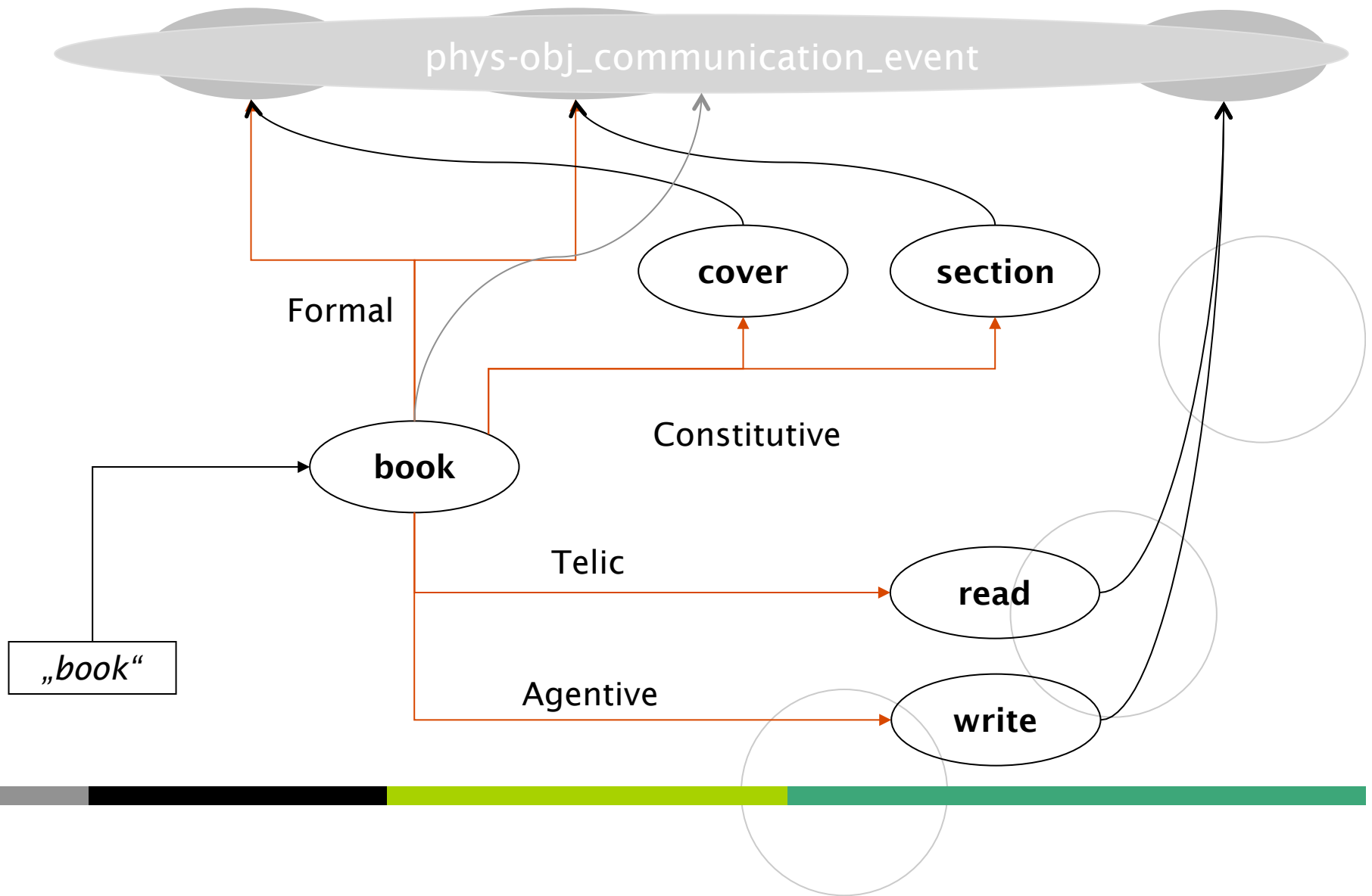
Agentive & Telic



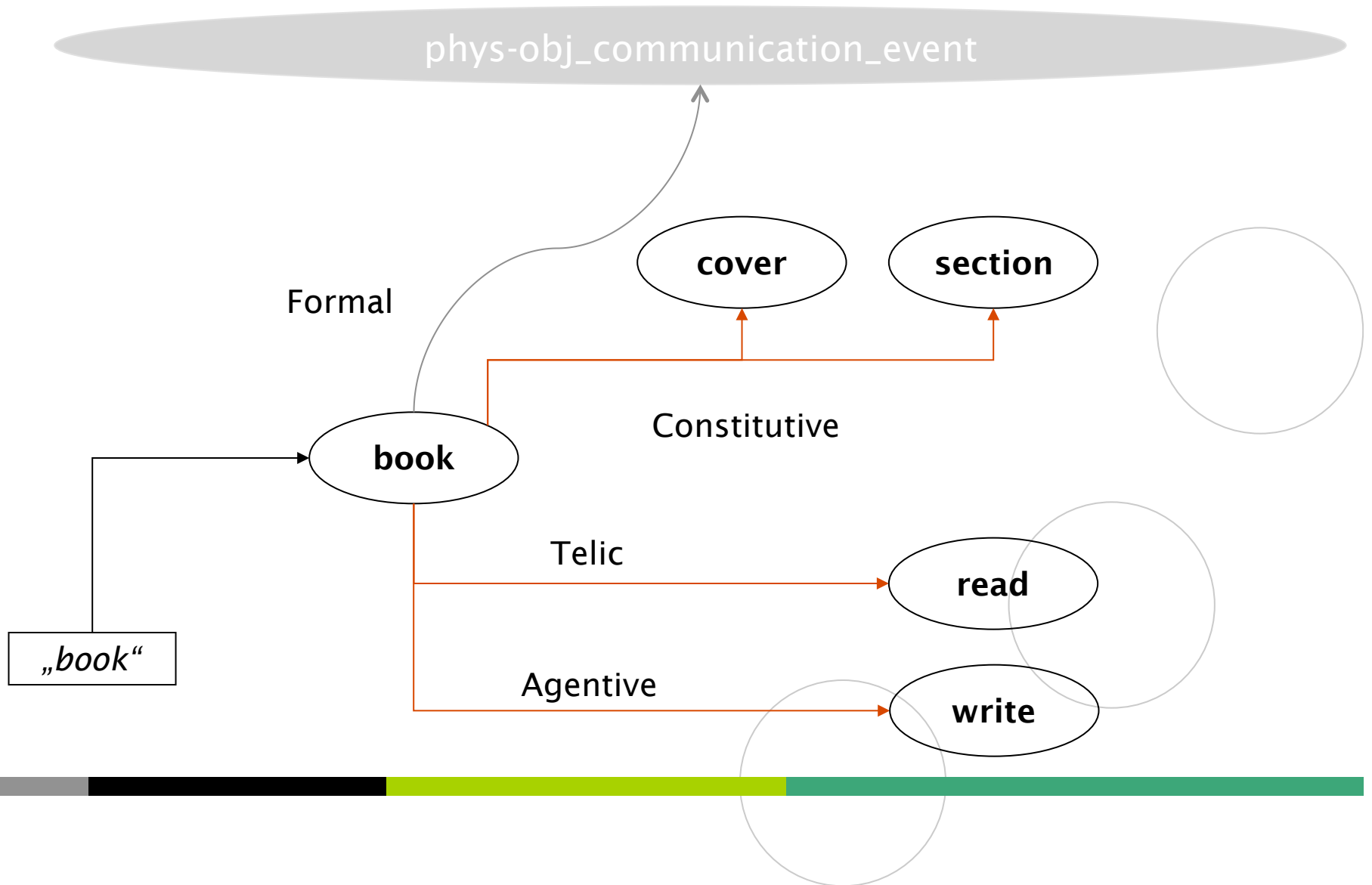
*"They **printed** some very **interesting** books."*



QS as 'Complex Class'



QS as 'Complex Class'

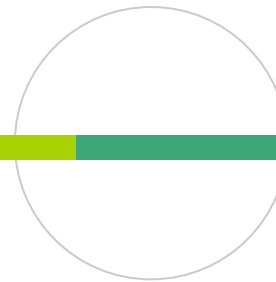
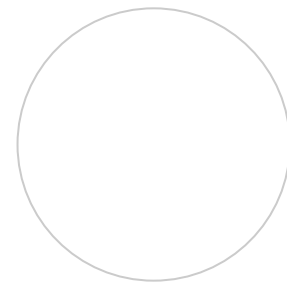
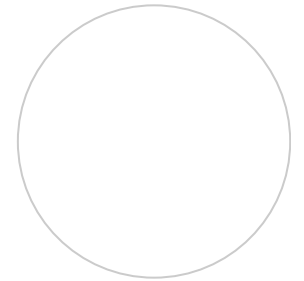


Empirical foundation of generative model

CoreLex, WordNet

Systematic Polysemous Classes

Basic Types



CoreLex (PhD thesis, Buitelaar 1998)

■ Automatic Qualia Structure Acquisition

- CoreLex is an attempt to automatically acquire underspecified lexical semantic representations that reflect systematic polysemy
- These representations can be viewed as shallow Qualia Structures

■ Sense Distribution in WordNet

- Systematic polysemy can be empirically studied in WordNet by observing sense distributions

>> If more than two words share the same sense distribution (i.e. have the same set of senses), then this may indicate a pattern of systematic polysemy (adapted from Apresjan 1973)



WordNet

■ Lexical Semantic Resource

□ Semantic Lexicon

– Maps words to meanings (senses)

□ Lexical Database

– Machine readable (has a formal structure)

■ Freely available

□ <http://wordnet.princeton.edu/>



WordNet - Origins

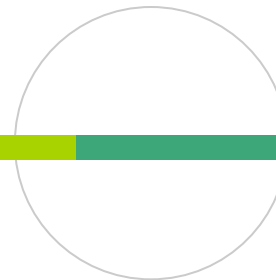
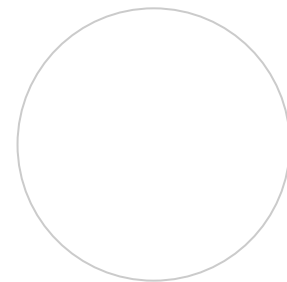
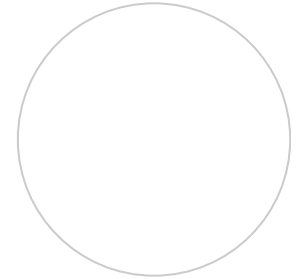
“In 1985 a group of psychologists and linguists at Princeton University undertook to develop a lexical database ... The initial idea was to provide an aid to use in searching dictionaries conceptually, rather than merely alphabetically ... WordNet instantiates hypotheses based on results of psycholinguistic research ... expose such hypotheses to the full range of the common vocabulary”

In anomia, there is a specific inability to name objects. When confronted with an apple, say, patients may be unable to utter “apple,” even though they will reject such suggestions as shoe or banana, and will recognize that apple is correct when it is provided. (Caramazza/Berndt 1978)

Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross and Katherine J. Miller. *Introduction to WordNet: an on-line lexical database*. In: *International Journal of Lexicography* 3 (4), 1990, pp. 235 - 244.

WordNet Synsets

- WordNet is organized around word meaning (not word forms as with traditional lexicons)
 - Word meaning is represented by “synsets”
 - Synset is a “Set of Synonyms”
- Example
 - {*board, plank*}
 - Piece of lumber
 - {*board, committee*}
 - Group of people



Synsets and Senses

■ Synsets represent word meaning

□ Words that occur in several synsets have a corresponding number of meanings (senses)

- **S: (n) board** (a committee having supervisory powers) *"the board has seven members"*
- **S: (n) board** (a flat piece of material designed for a special purpose) *"he nailed boards across the windows"*
- **S: (n) board, plank** (a stout length of sawn timber; made in a wide variety of sizes and used for many purposes)
- **S: (n) display panel, display board, board** (a vertical surface on which information can be displayed to public view)
- **S: (n) board, gameboard** (a flat portable surface (usually rectangular) designed for board games) *"he got out the board and set up the pieces"*
- **S: (n) board, table** (food or meals in general) *"she sets a fine table"; "room and board"*
- **S: (n) control panel, instrument panel, control board, board, panel** (electrical device consisting of a flat insulated surface that contains switches and dials and meters for controlling other electrical devices) *"he checked the instrument panel"; "suddenly the board lit up like a Christmas tree"*
- **S: (n) circuit board, circuit card, board, card, plug-in, add-in** (a printed circuit that can be inserted into expansion slots in a computer to increase the computer's capabilities)
- **S: (n) dining table, board** (a table at which meals are served) *"he helped her clear the dining table"; "a feast was spread upon the board"*

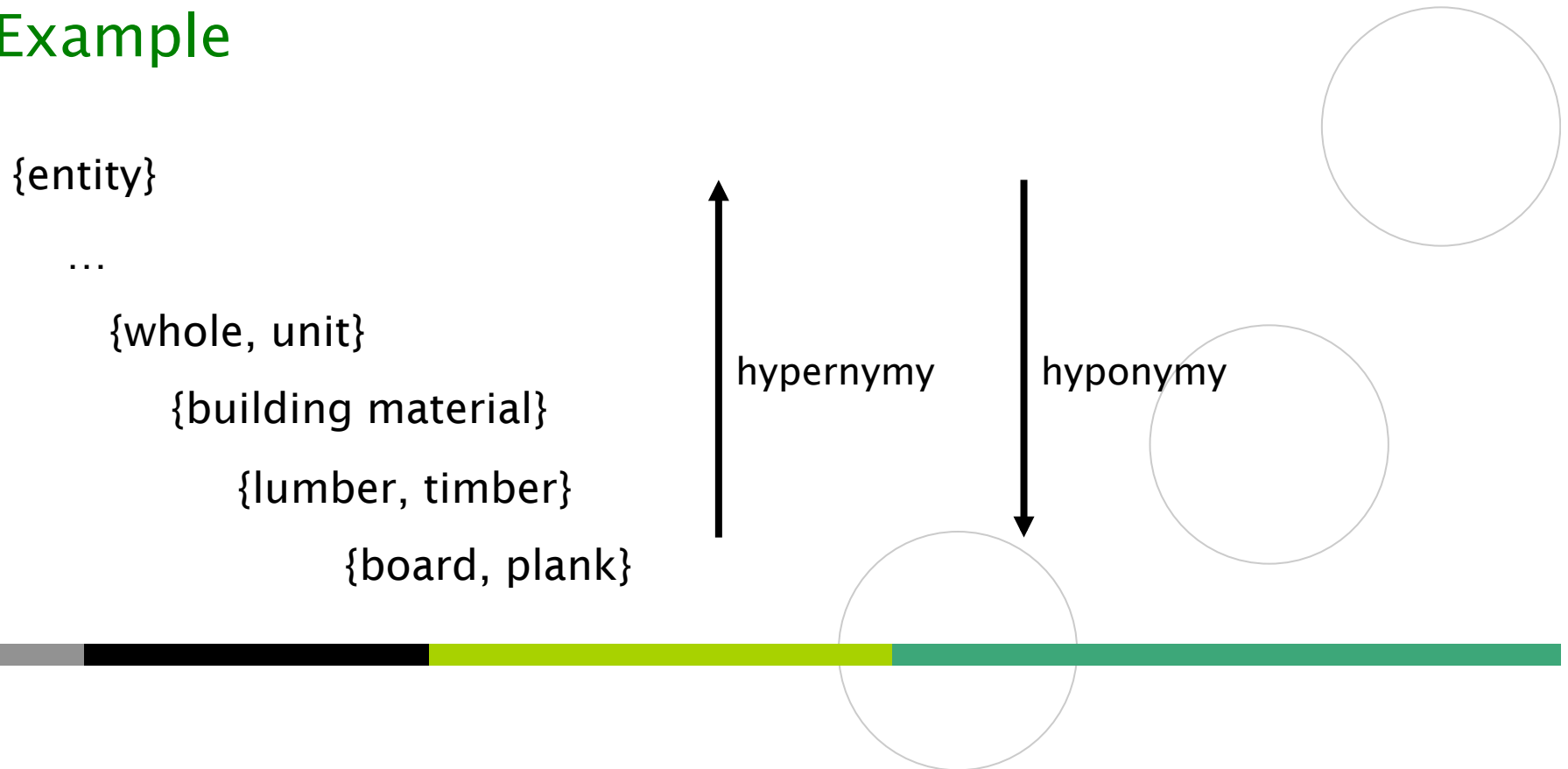
Synset Hierarchy

- Synsets are organized in hierarchies

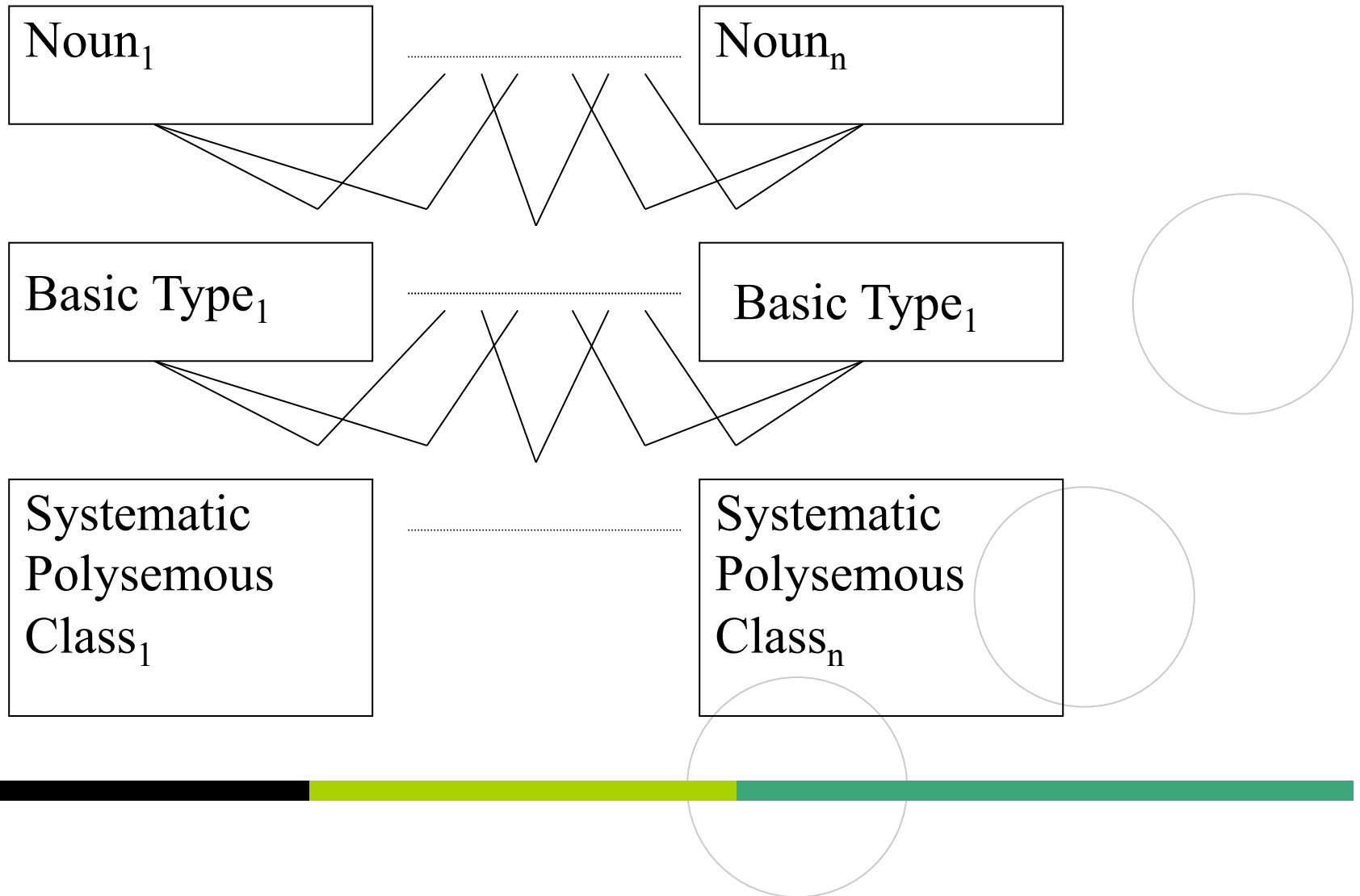
- Defines:

- generalization (hypernymy)
 - specialization (hyponymy)

- Example

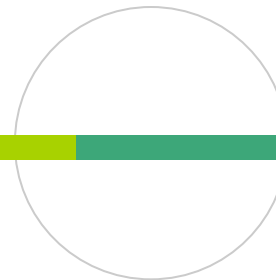
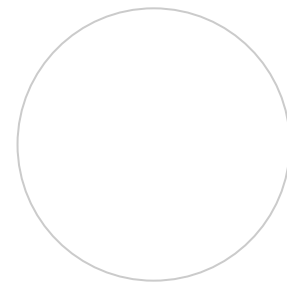
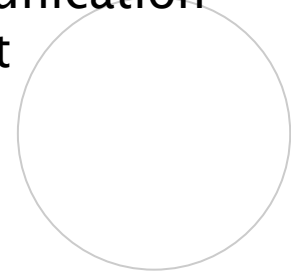


From WordNet to CoreLex



From Synsets to Basic Types

<i>book</i>	1.{publication}	=> artifact
	2.{product, production}	=> artifact
	3.{fact}	=> communication
	4.{dramatic_composition, dramatic_work}	=> communication
	5.{record}	=> communication
	6.{section, subdivision}	=> communication
	7.{journal}	=> artifact



... to Systematic Polysemous Classes

“artifact communication”

*amulet annals armband arrow article ballad bauble beacon bible
birdcall blank blinker boilerplate book bunk cachet canto catalog
catalogue chart chevron clout compact compendium convertible
copperplate copy cordon corker ... guillotine homophony horoscope
indicator journal laurels lay ledger loophole marker memorial
nonsense novel obbligato obelisk obligato overture pamphlet
pastoral paternoster pedal pennant phrase platform portrait
prescription print puzzle radiogram rasp recap riddle rondeau ...
statement stave stripe talisman taw text tocsin token transcription
trophy trumpery wand well whistle wire wrapper yardstick*



Other Examples

animal natural_object

alligator broadtail chamois ermine leopard muskrat ...

natural_object plant

algarroba almond anise baneberry butternut candlenut...

action artifact group_social

artillery assembly band dance gathering institution ...

action attribute event psychological

appearance decision deviation impulse outrage ...

possession quantity_definite

cent centime dividend gross penny real shilling



Problematic Cases

Non-Systematic Classes

action animal artifact

bat drill fly hobby ruff solitaire spat

Partly-Systematic Classes

action geographical_location

bolivia caliphate charleston chicago clearing emirate

michigan prefecture repair ...

Systematic *clearing, repair, wheeling*

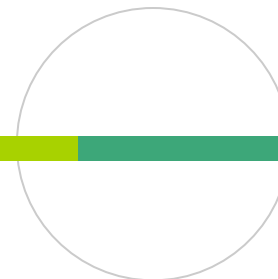
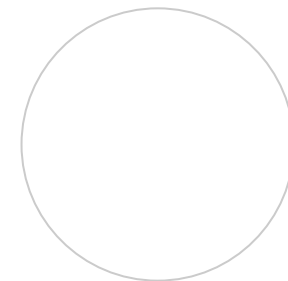
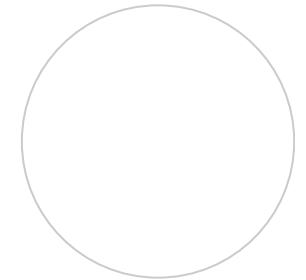
Homonyms *bolivia, charleston, chicago, michigan*

?? *caliphate, emirate, prefecture*



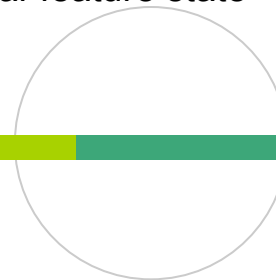
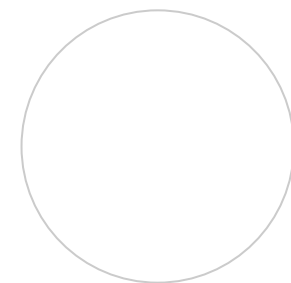
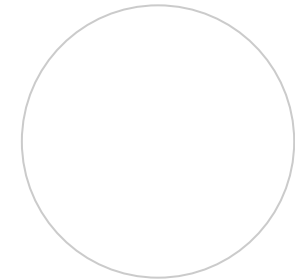
Basic Types

8283	art	artifact artefact
6606	act	act human action human activity
6303	hum	person individual someone mortal human soul
4933	grb	biological group
4137	atr	attribute
3456	psy	psychological feature
3336	com	communication
2703	anm	animal animate being beast brute creature fauna
2311	plt	plant flora plant life
2266	sta	state
1541	fod	food nutrient
1282	log	region geographical location
1277	nat	natural object water land
1189	sub	substance matter
1082	evt	event
992	prt	part piece
940	grs	social group people
777	qud	definite quantity
773	pro	process
699	chm	compound chemical compound chemical element element
628	tme	time period period period of time amount of time time unit unit of time time
624	agt	causal agent cause causal agency
571	pos	possession
567	loc	location any other location
506	rel	relation
420	frm	shape form
345	grp	group grouping any other group
342	phm	phenomenon
295	qui	indefinite quantity
186	pho	object inanimate object physical object
178	mic	microorganism
100	lme	linear measure long measure
61	lfr	life form organism being living thing
57	cel	cell
38	mea	measure quantity amount quantum
28	ent	entity
21	con	consequence effect outcome result upshot
21	spc	space
8	abs	abstraction

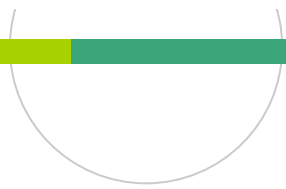
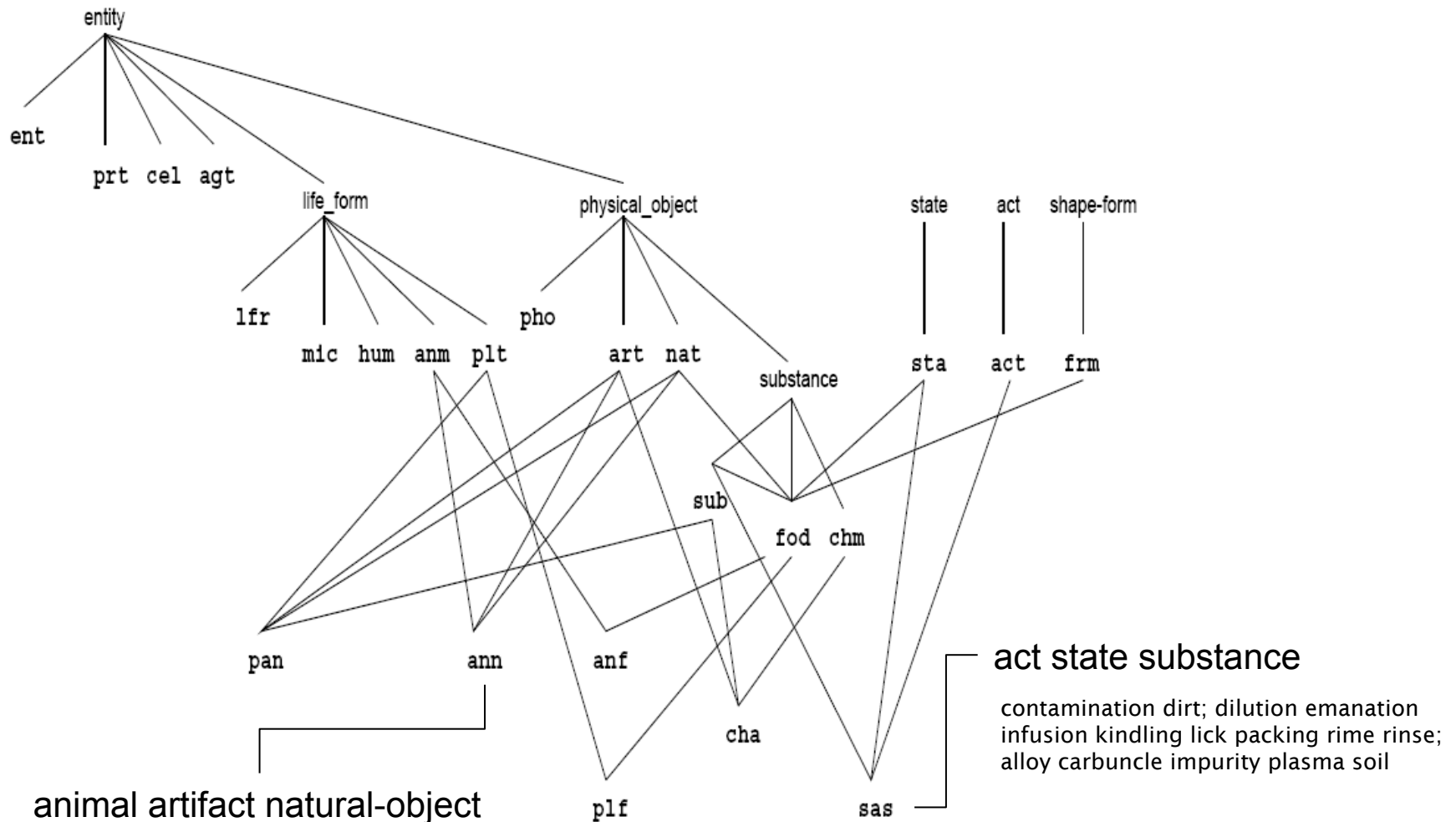


Systematic Polysemous Classes

acp	act attribute process psychological-feature state
acr	act attribute event relation state
acs	act state
aes	act event state
aev	act event
age	act causal agent
agh	causal agent human
agl	causal agent location
age	causal agent animal
agp	causal agent psychological-feature
agt	causal agent
anf	animal food
ann	animal artifact natural-object
anp	animal psychological-feature
aqu	artifact quantity-definite quantity-indefinite
ara	artifact attribute psychological-feature state
arg	artifact group
arh	artifact human
arp	artifact psychological-feature state
art	artifact state
atc	attribute communication phenomenon psychological-feature state
...	



CoreLex vs. WordNet



CoreLex Semantic Lexicon

- CoreLex is available from

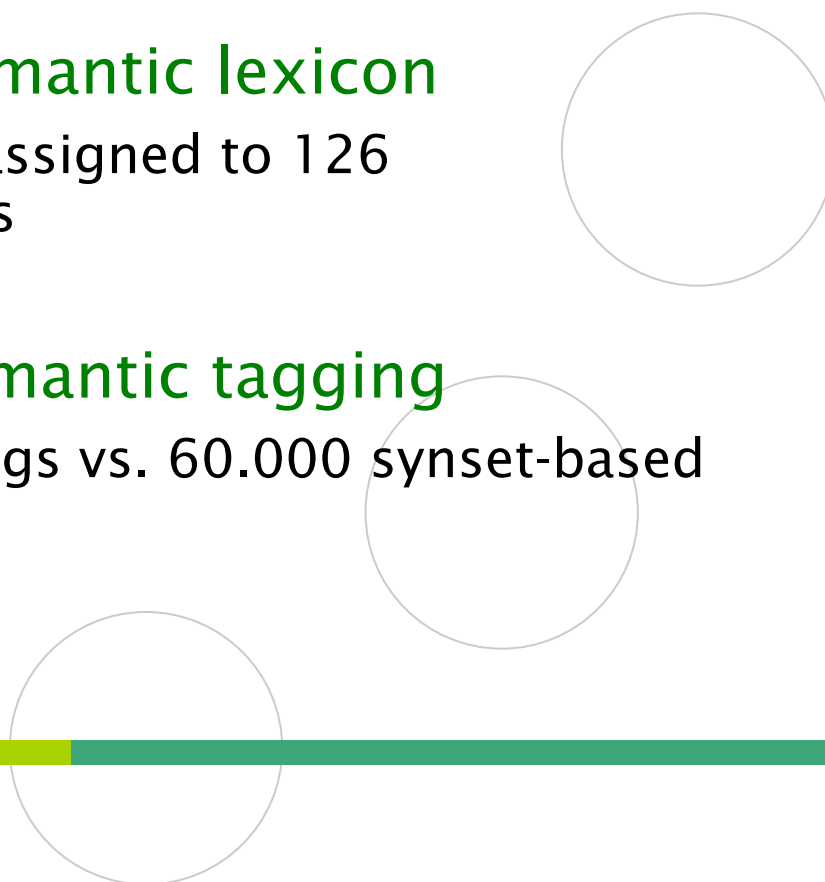
- <http://www.cs.brandeis.edu/~paulb/CoreLex/corelex.html>

- Provides a coarse-grained semantic lexicon

- Covers around 40.000 nouns, assigned to 126 underspecified semantic classes

- Allows for coarse-grained semantic tagging

- 126 underspecified semantic tags vs. 60.000 synset-based senses



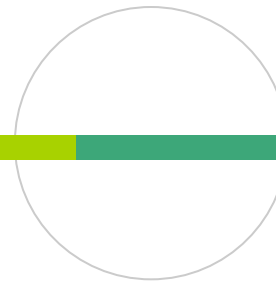
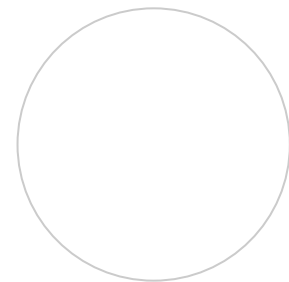
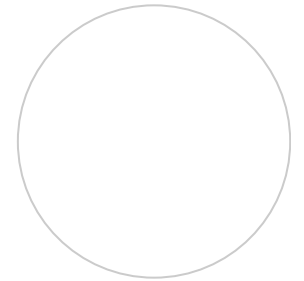
Ontologies and lexical semantics

ontology-driven sense derivation

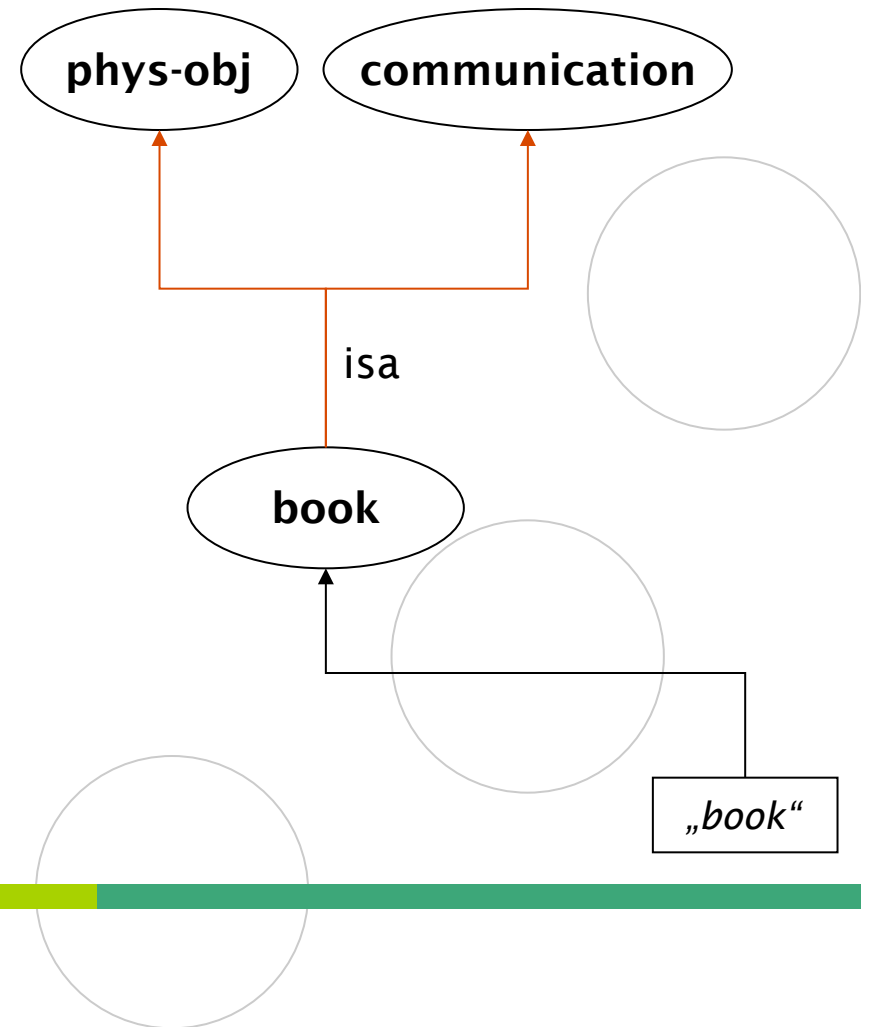
ontological/semantic & linguistic/lexical structure

integration of ontologies and lexicons

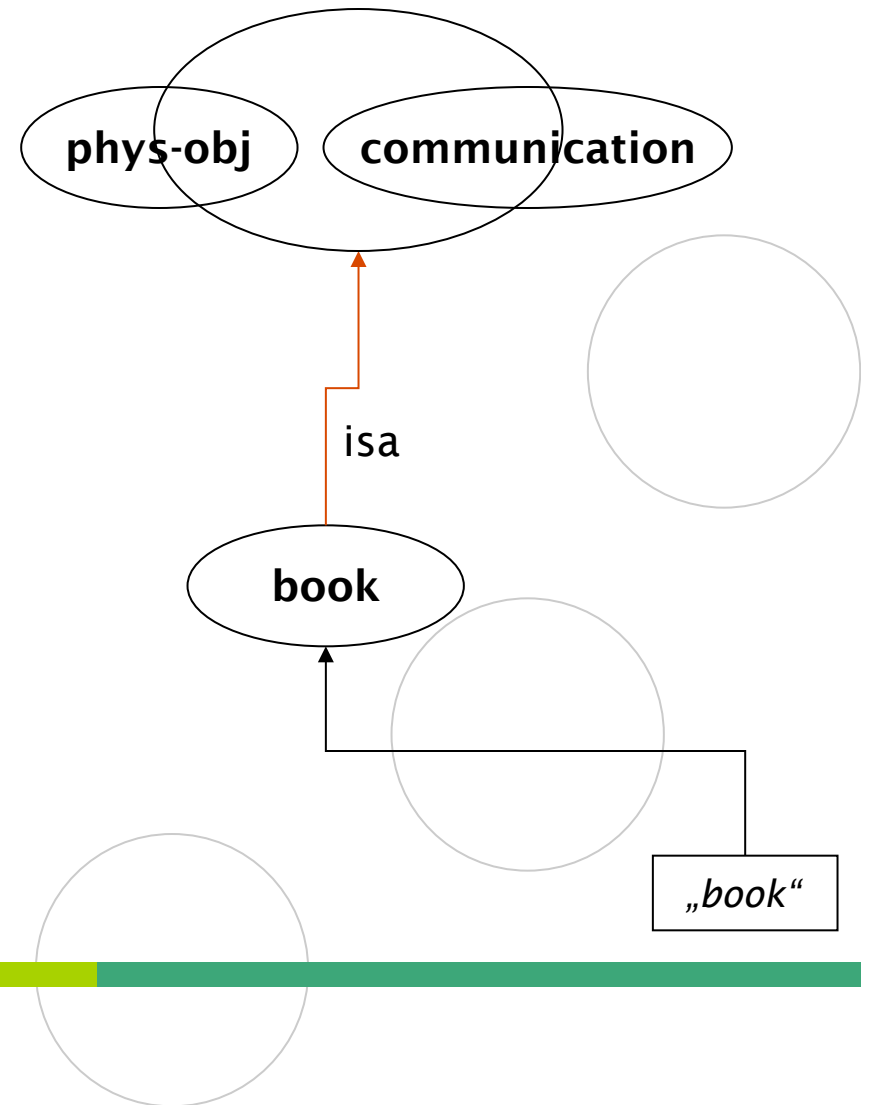
'lemon' : lexicon model for ontologies



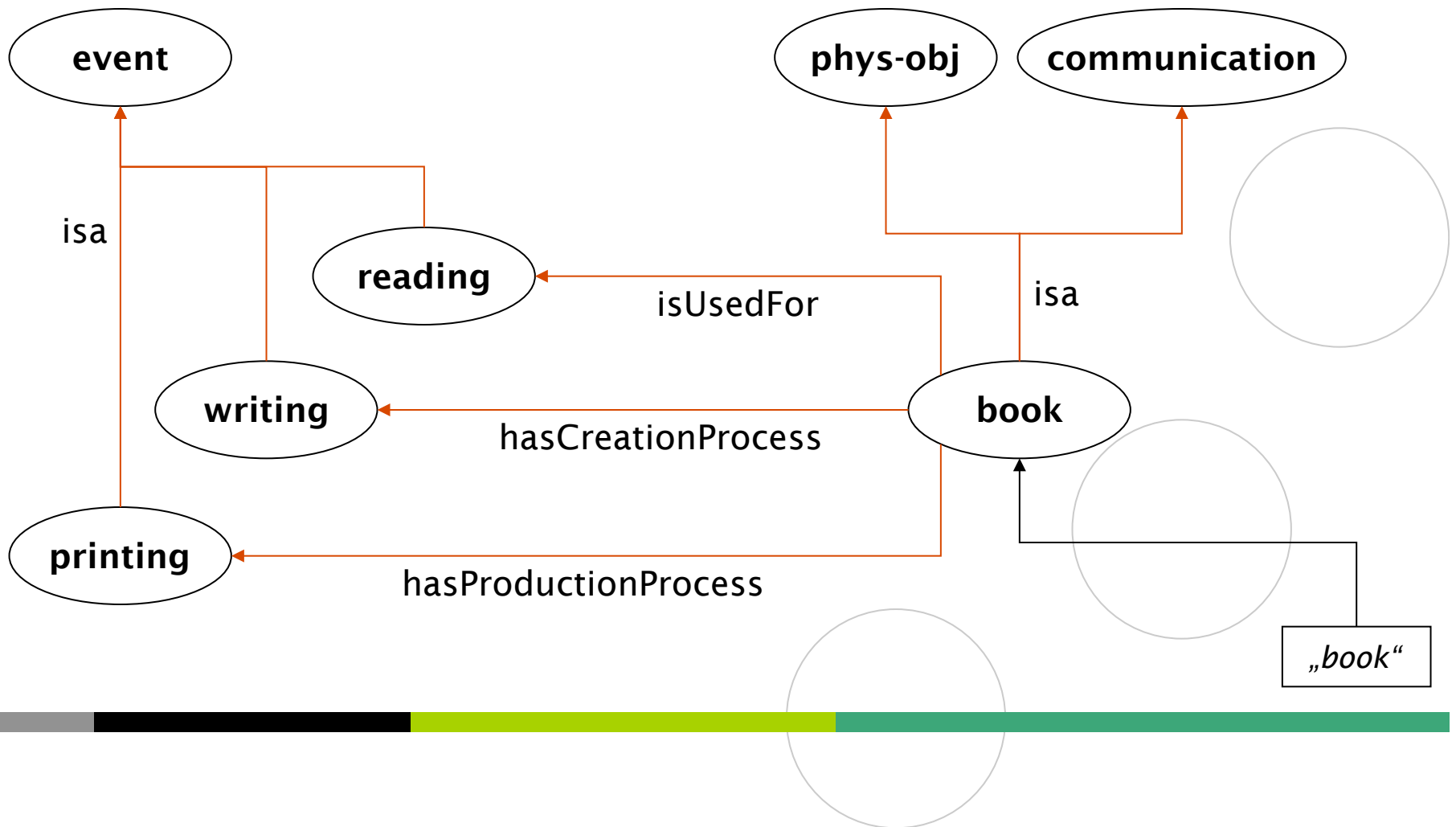
Ontology-driven sense derivation



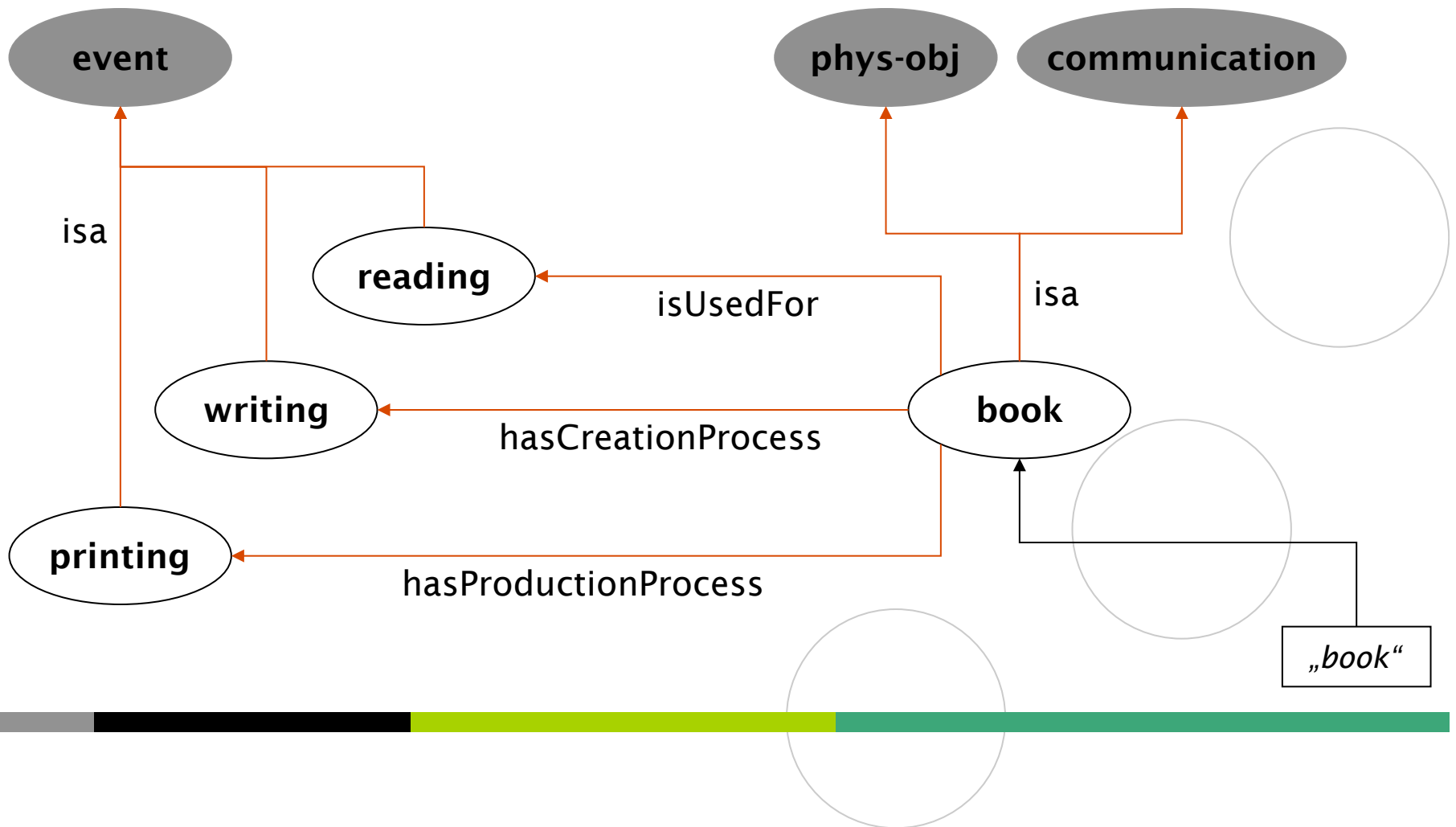
Ontology-driven sense derivation



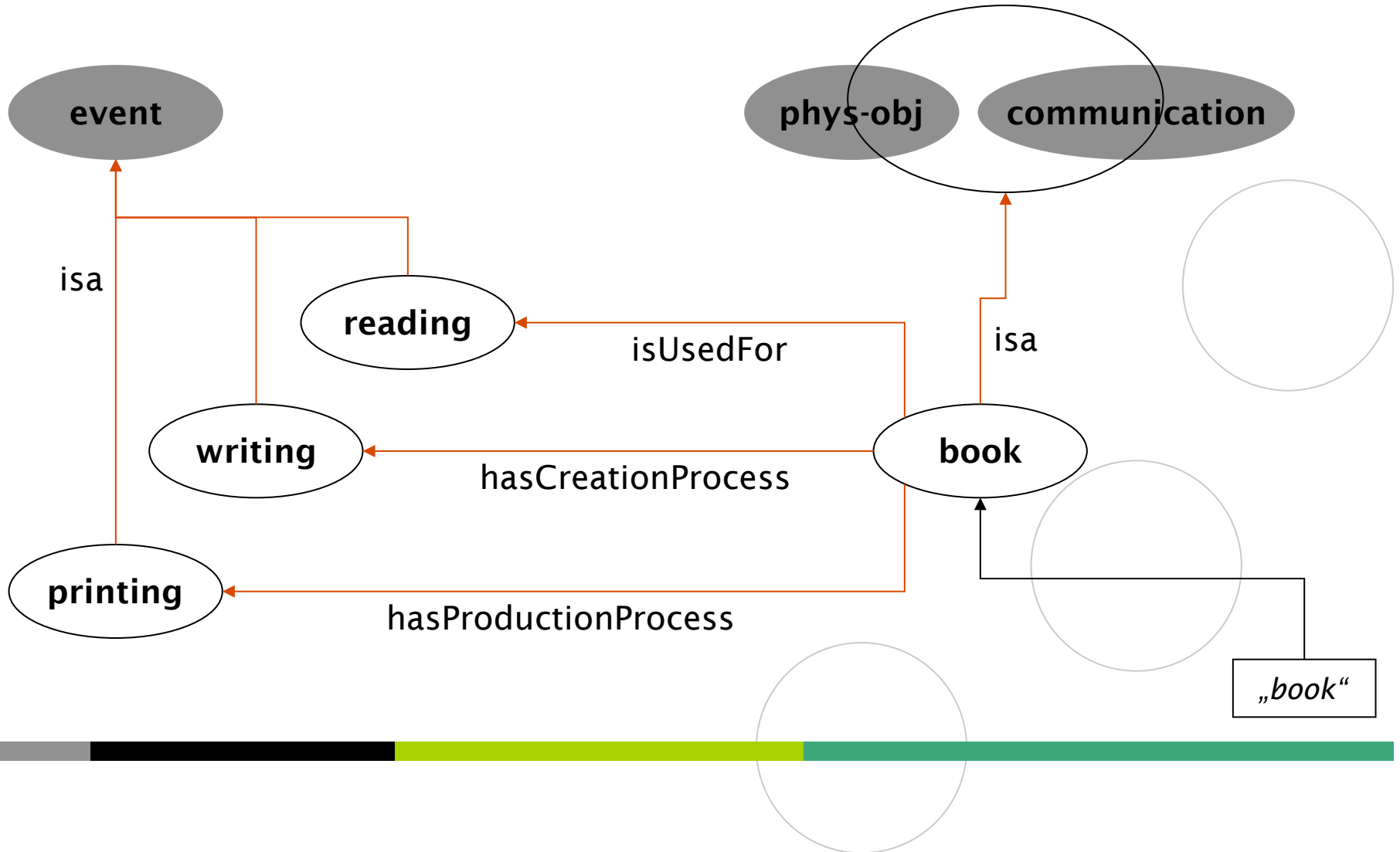
Ontology-driven sense derivation



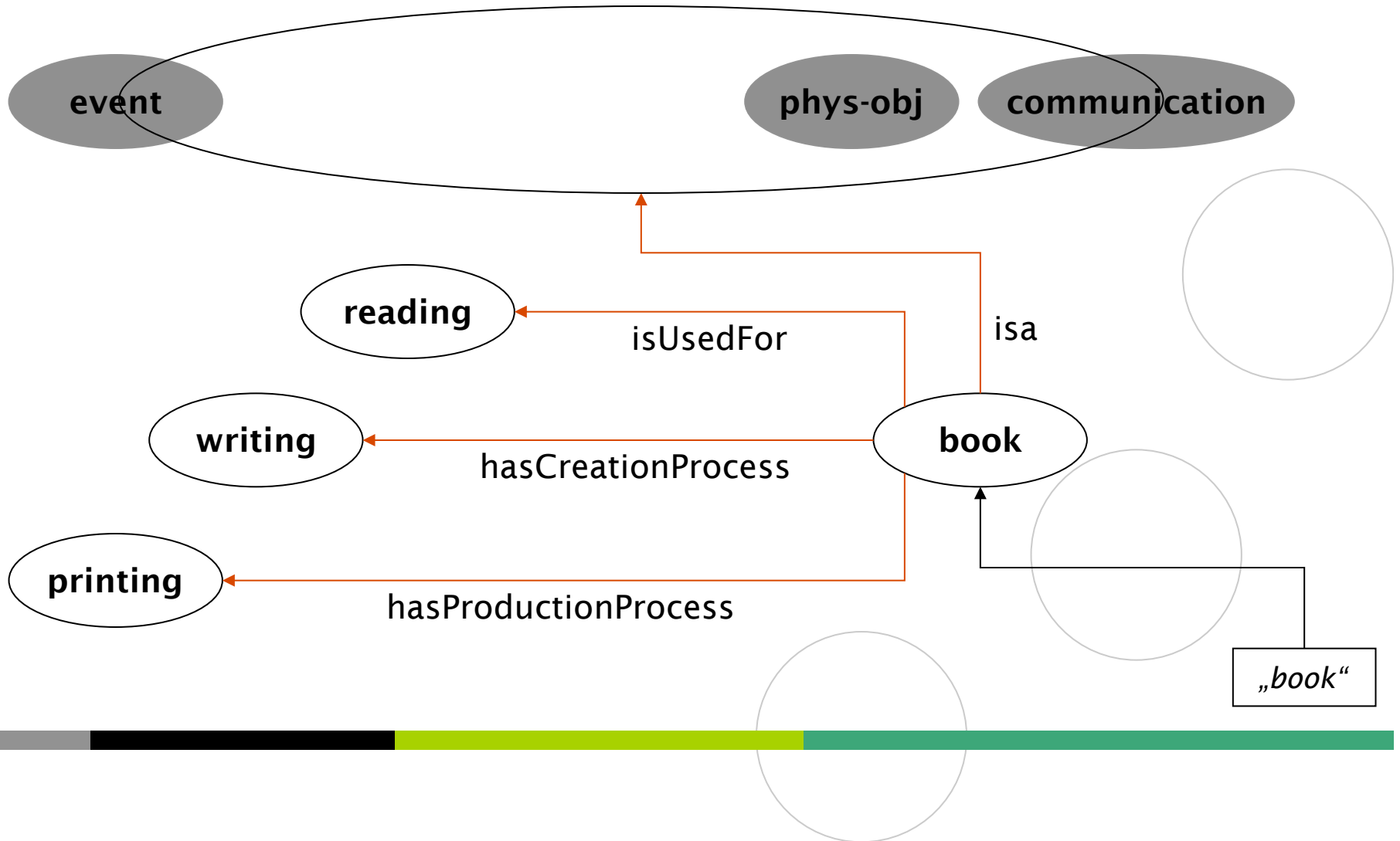
Ontology-driven sense derivation



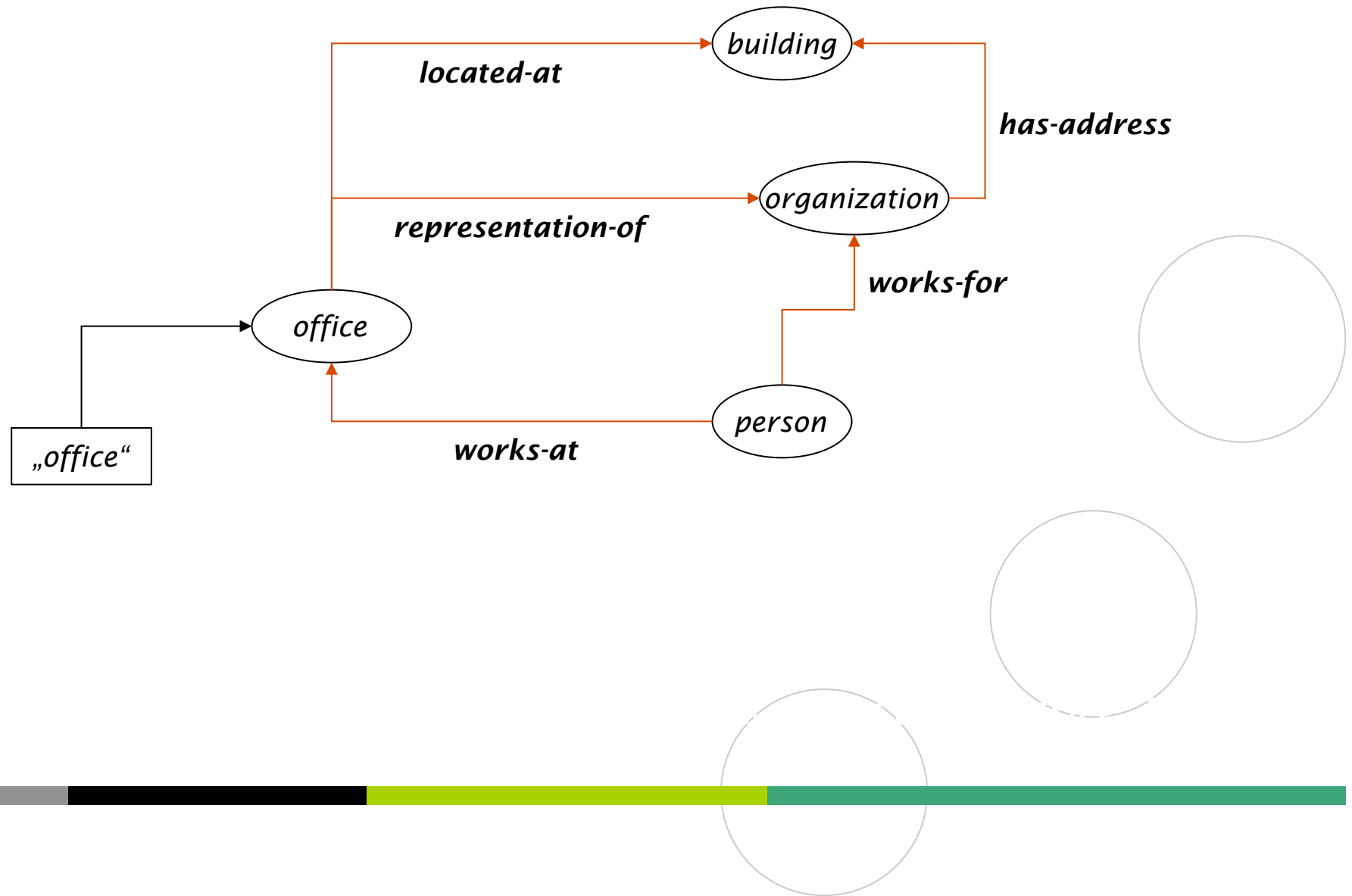
Ontology-driven sense derivation



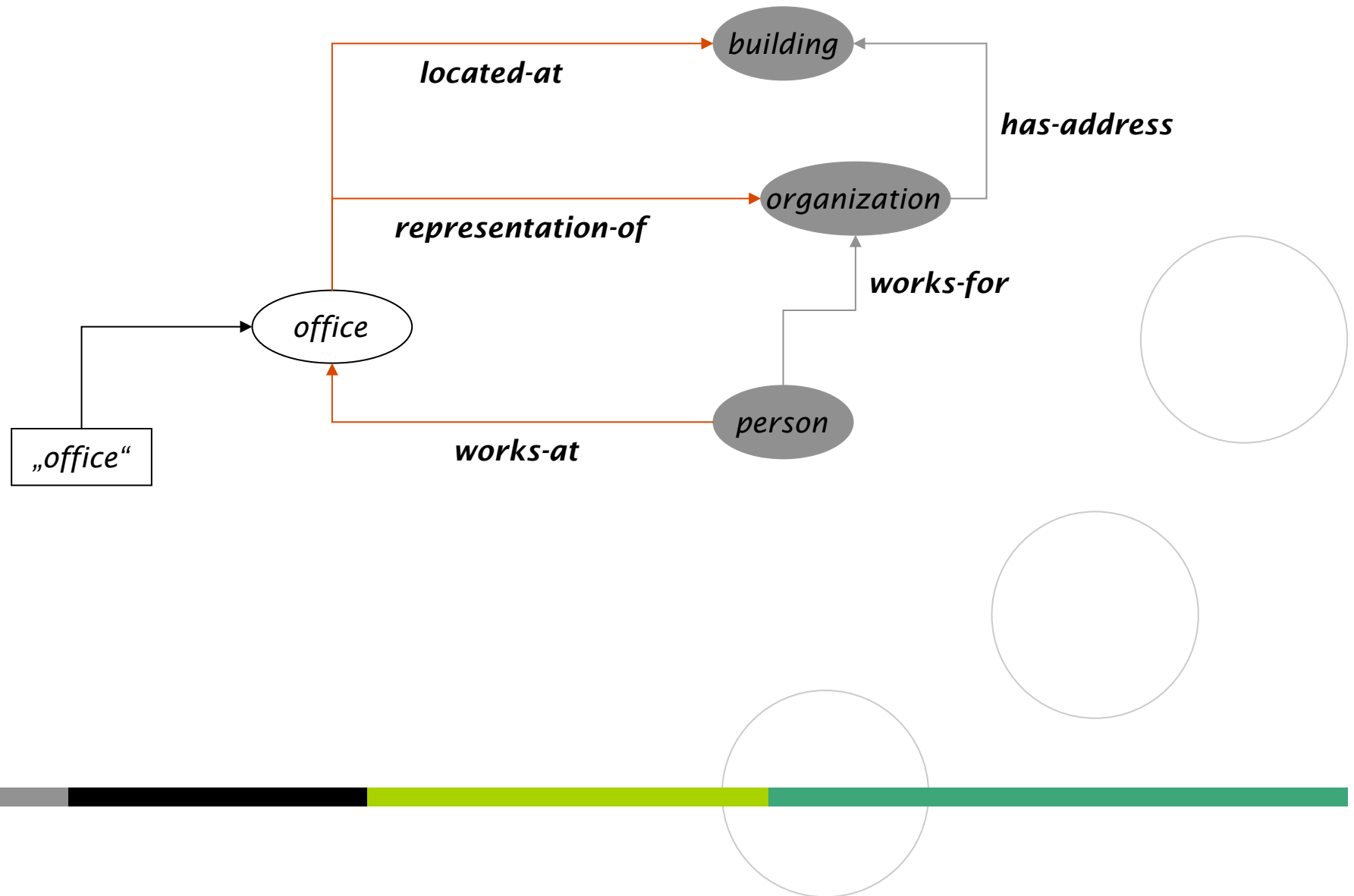
Ontology-driven sense derivation



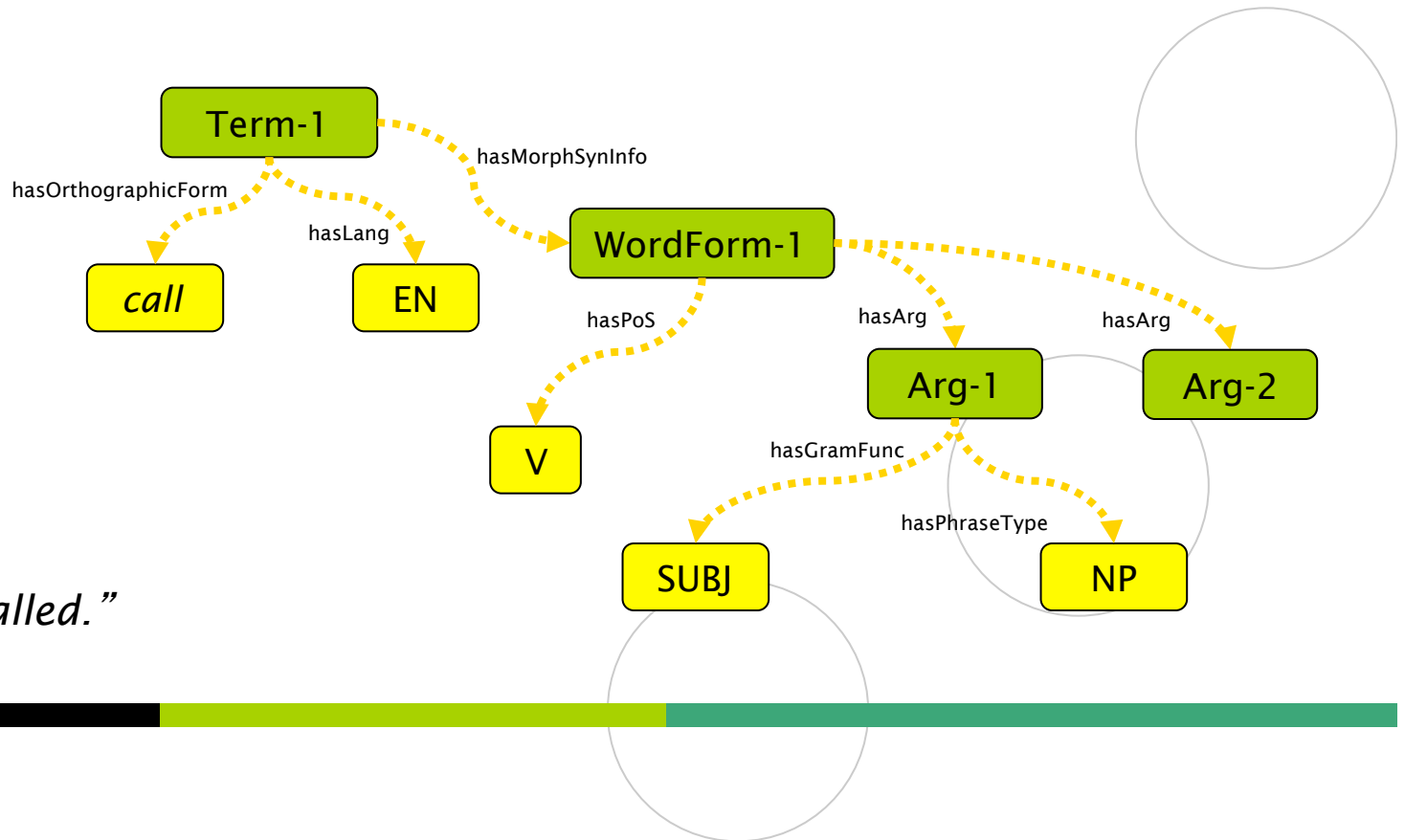
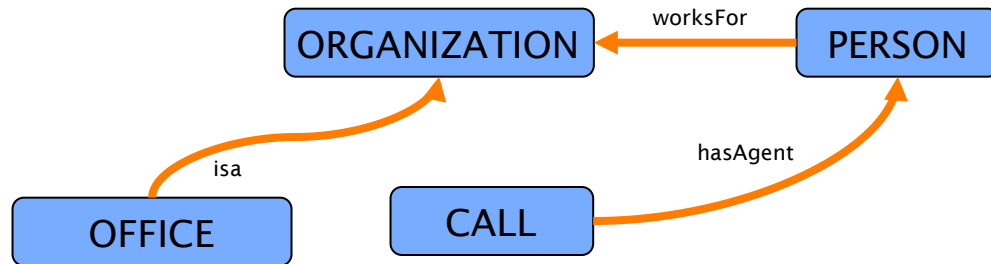
Ontology-driven sense derivation



Ontology-driven sense derivation

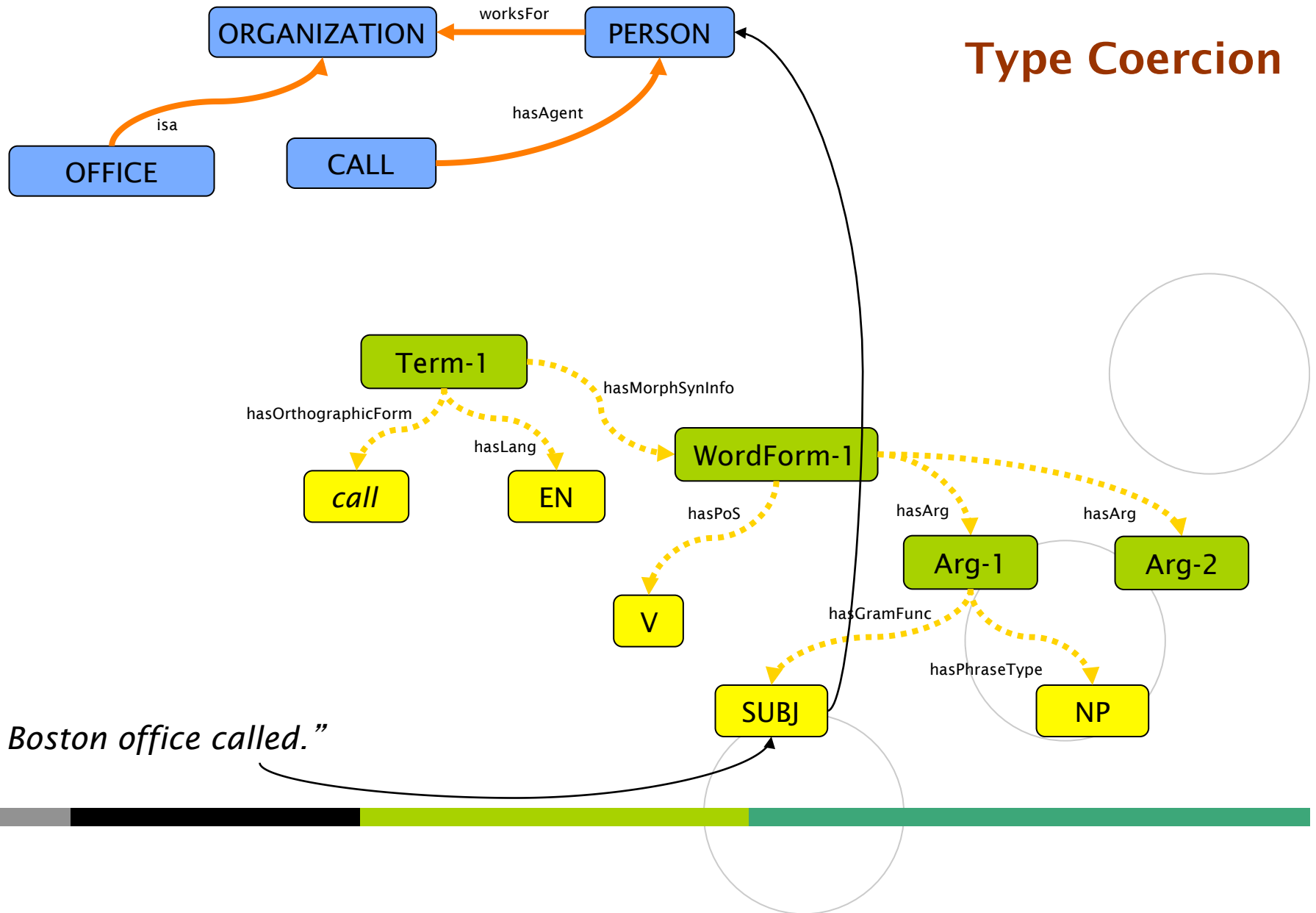


Mapping Lexical to Semantic Structure



"The Boston office called."

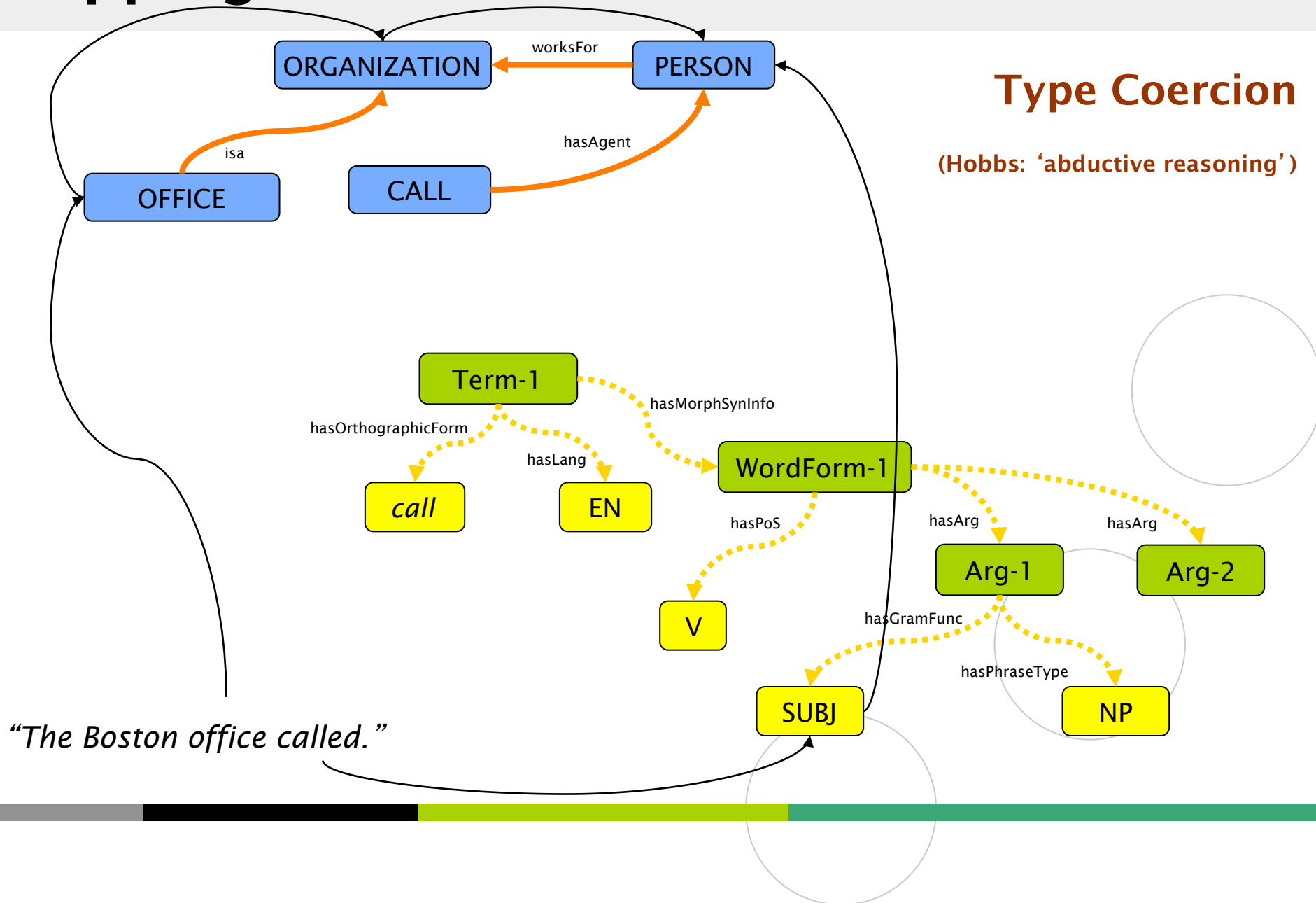
Mapping Lexical to Semantic Structure



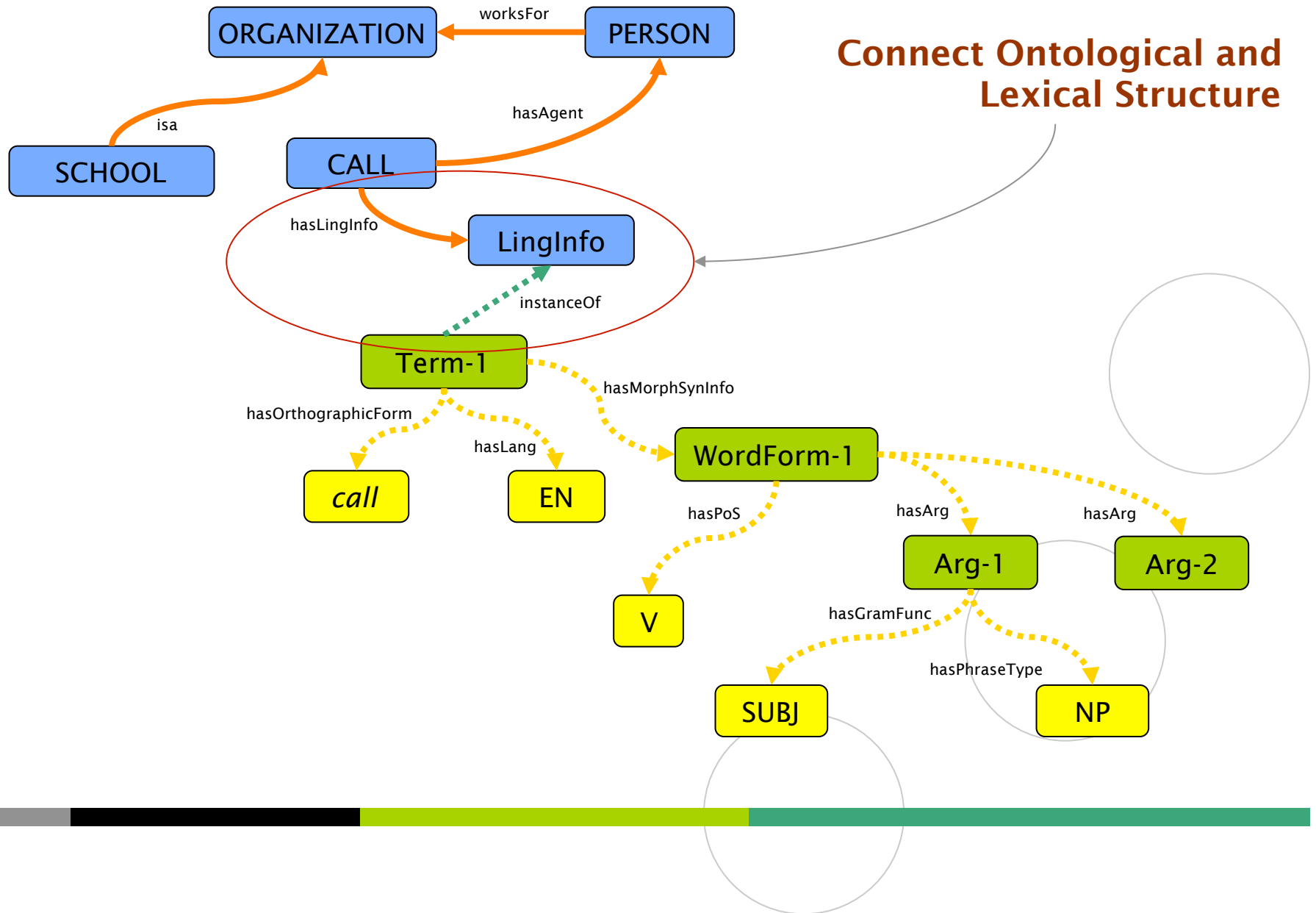
Mapping Lexical to Semantic Structure

Type Coercion

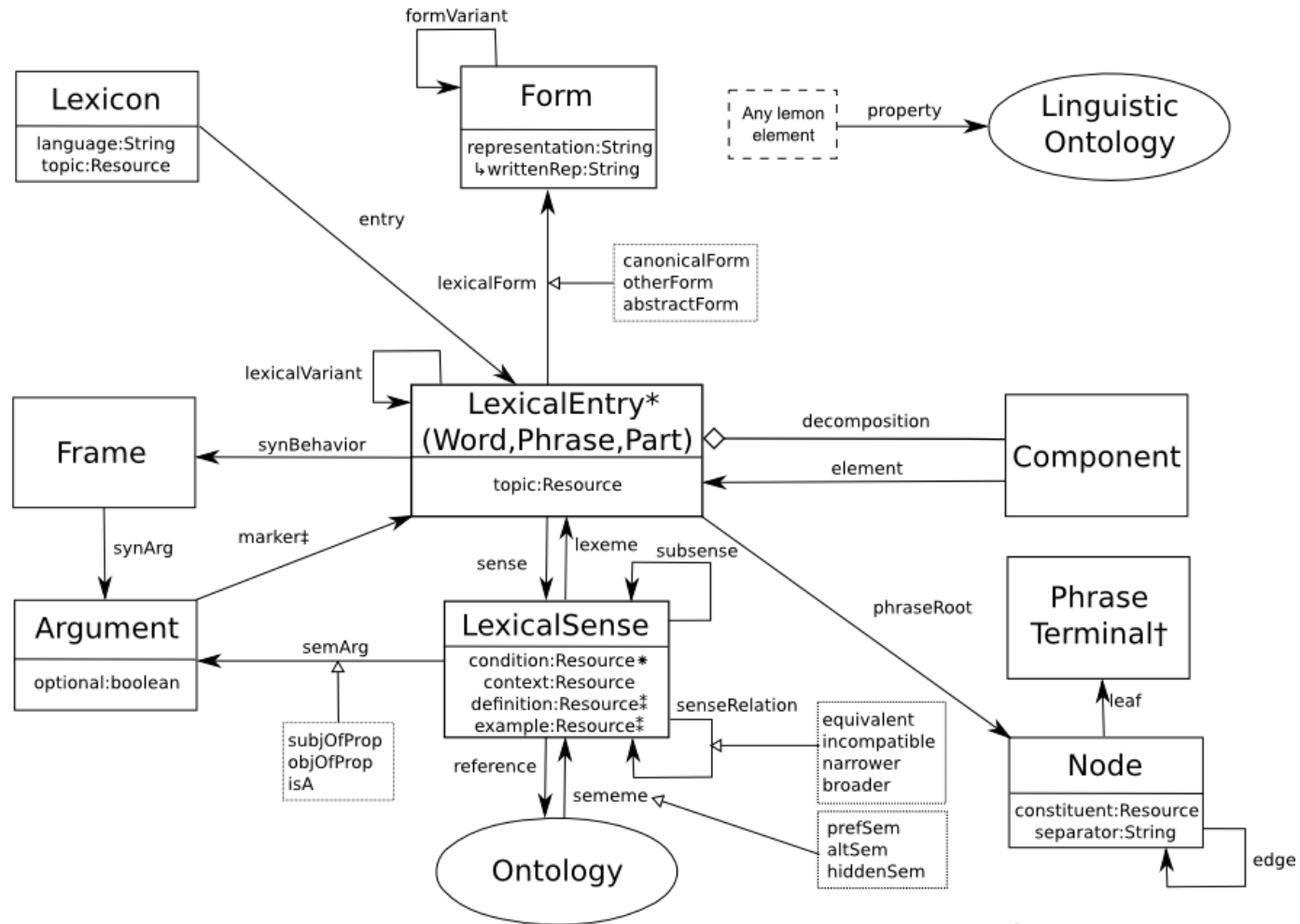
(Hobbs: 'abductive reasoning')



Mapping Lexical to Semantic Structure



Lexicon Model for Ontologies



lemon

Acknowledgements & Further Info

- CoreLex

- <http://pages.cs.brandeis.edu/~paulb/CoreLex/corelex.html>

- lemon (<http://lexinfo.net/>)

- <http://greentacle.techfak.uni-bielefeld.de/drupal/sites/default/files/ontologies/lemon.owl>

- <http://greentacle.techfak.uni-bielefeld.de/drupal/sites/default/files/lemon-cookbook.pdf>

- Slides

- Thanks to Paul Buitelaar for use of his slides!!

