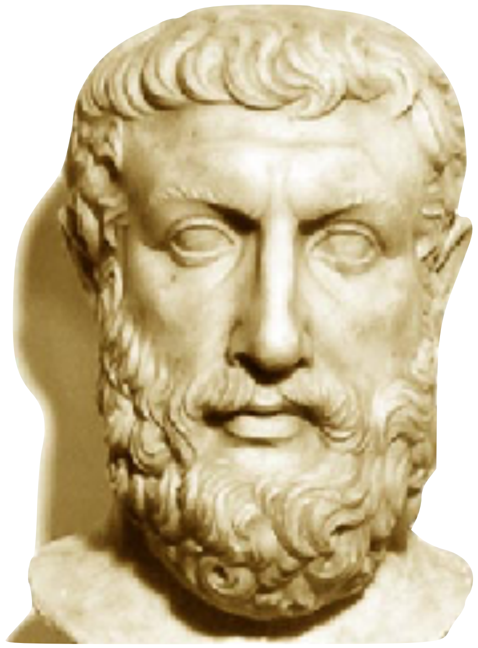
Ontology and Taxonomy

Computational Linguistics
Emory University
Jinho D. Choi

Ontology

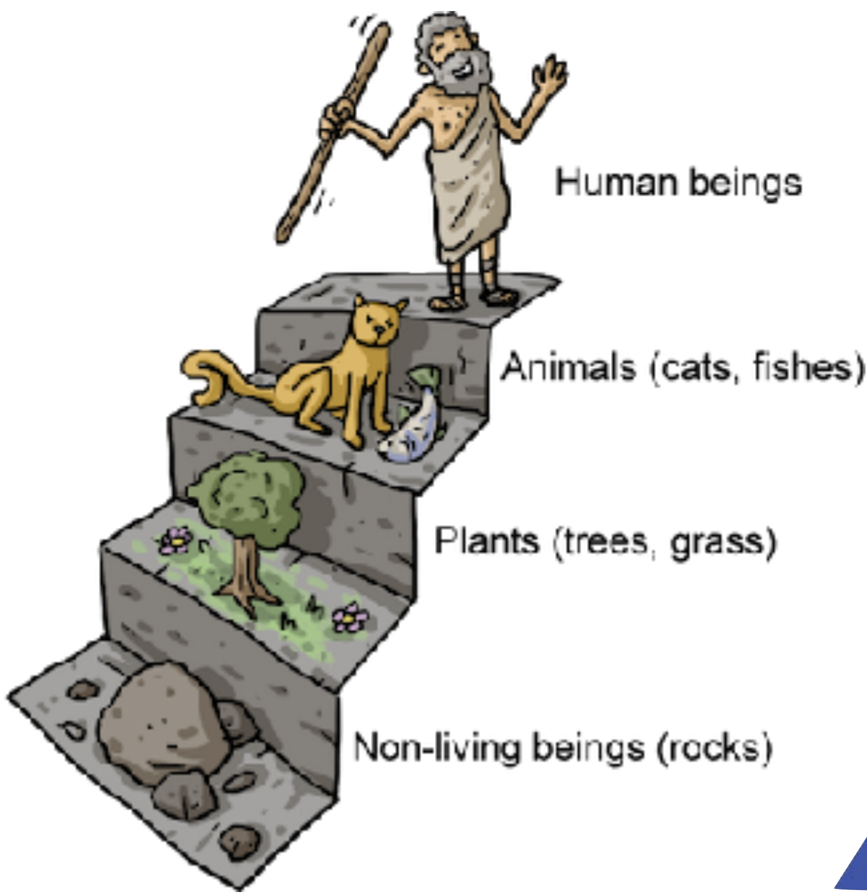


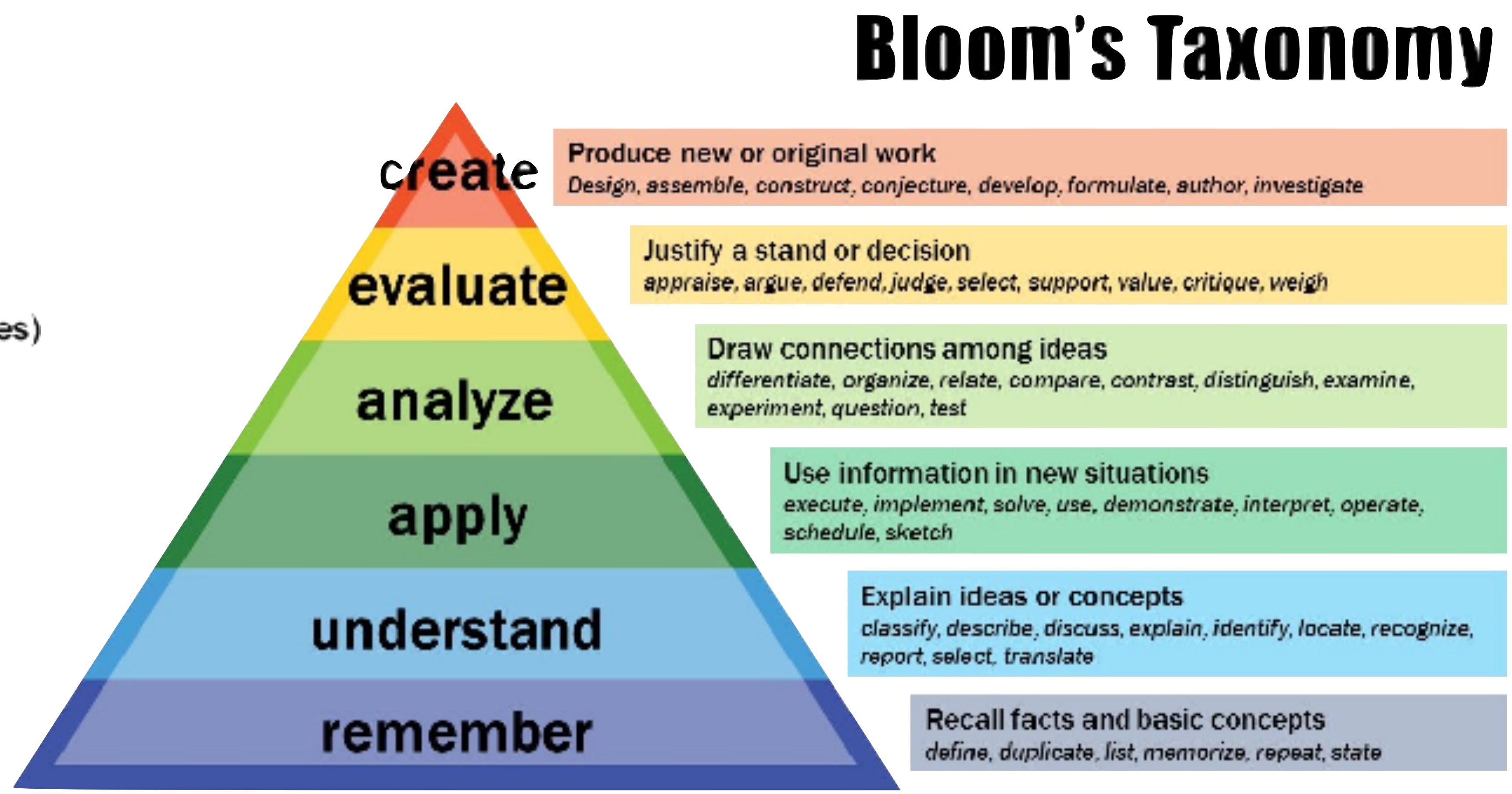
Nature of being, becoming, existence, or reality,
as well as the basic categories of being and their relations.

Types, properties, and interrelationships of the entities
that fundamentally exist for a particular domain of discourse.

Taxonomy

The science of classification according to a pre-determined system, with the resulting catalog used to provide a conceptual framework for discussion, analysis, or information retrieval.





WordNet

A lexical database that groups nouns, verbs, adjectives and adverbs into sets of cognitive synonyms (synsets) interlinked by conceptual-semantic and lexical relations.

Synonymy, Antonymy, Hyponymy, Meronymy

POS	Words	Synsets	Senses
Noun	117,798	82,115	146,312
Verb	11,529	13,767	25,047
Adjective	21,479	18,156	30,002
Adverb	4,481	3,621	5,580
Total	155,287	117,659	206,941

http://wordnet.princeton.edu

Word Sense

A word can have multiple meanings (senses).

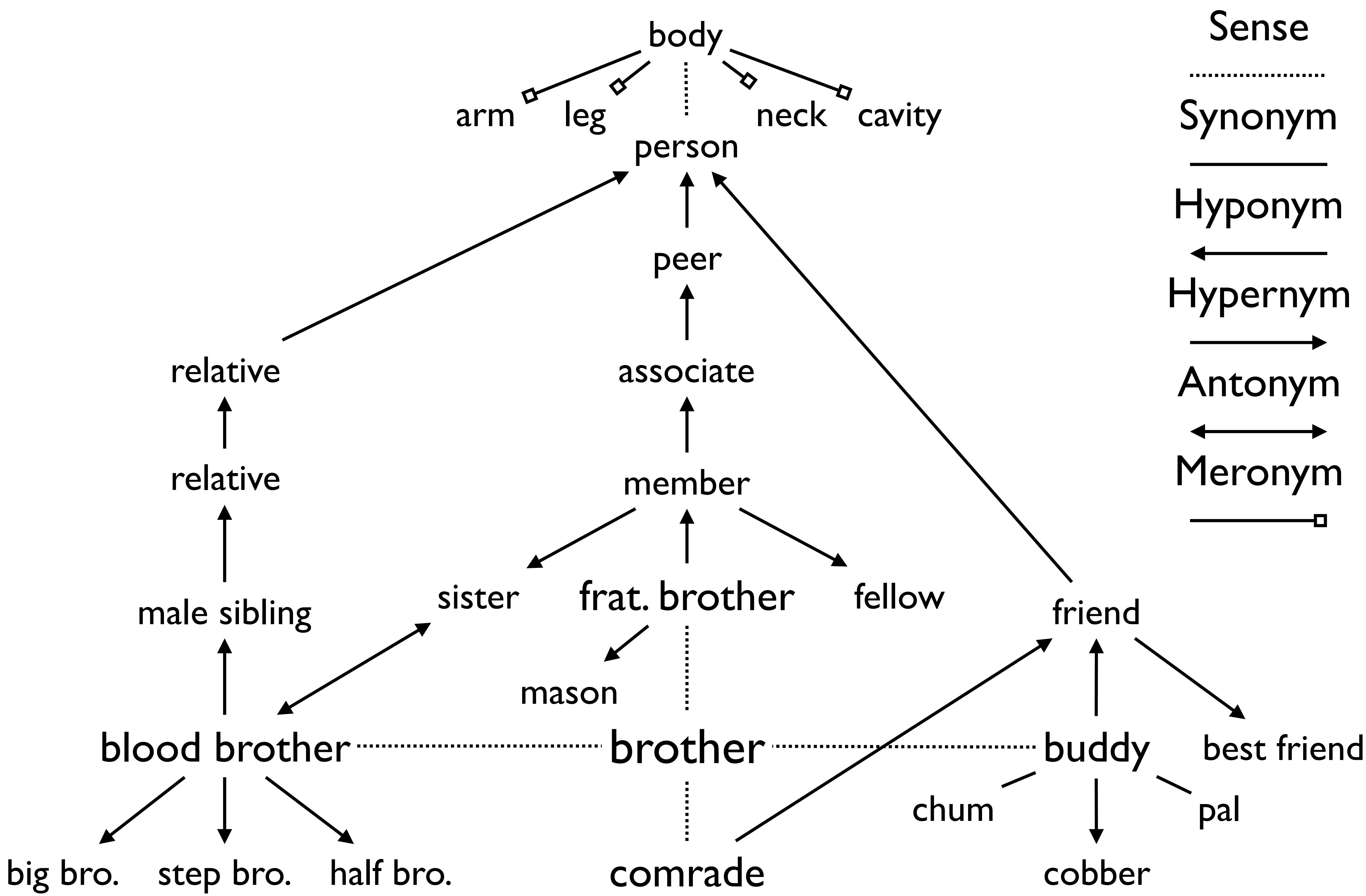
Chair

Noun	Verb
Seat	Act or preside as chair
Professorship	Lead or preside over
President	
Death chair	

How find-grained do word senses need to be?

Automatically distinguish word senses?

Lexical Relation



Entailment

If (V_1 is true), then (V_2 must be true).

If (A is snoring), then (A must be sleeping).

Unless V_1 and V_2 are synonyms, the converse is not true.

If (A is sleeping), then (A must be snoring).

The contradiction is true.

If (A is not sleeping), then (A must not be snoring).

Temporal inclusion

$T(V_1) \subseteq T(V_2)$: If (A is snoring), then (A must be sleeping).

$T(V_1) \supseteq T(V_2)$: If (A bought B), then (A must have paid for B).

$T(V_1) = T(V_2)$: If (A is marching), then (A must be walking).

Hyponym

(To E_1) is a kind of (to E_2).

Noun

A horse is a kind of an animal.

Verb

Ambling is a kind of walking.

Multiple hyponyms

A mule is a kind of a donkey and a horse.

Ambling is a kind of walking and being slow.

Troponym

(To V_1) is (to V_2) in some particular manner.

(To shout) is (to talk) loud.

(To amble) is (to walk) in slow, relaxed manner.

Troponyms → "entailments with temporal inclusions".

(To amble) → (To walk)

(To amble) ⊆ (To walk)

Co-Troponym

Siblings differentiated by their manner.

To walk/run is to move at a pace/fast.

Backward Presupposition

Backward Presupposition

If A failed/succeeded in B, then A must have done B.

If A forgot B, A must have known B"

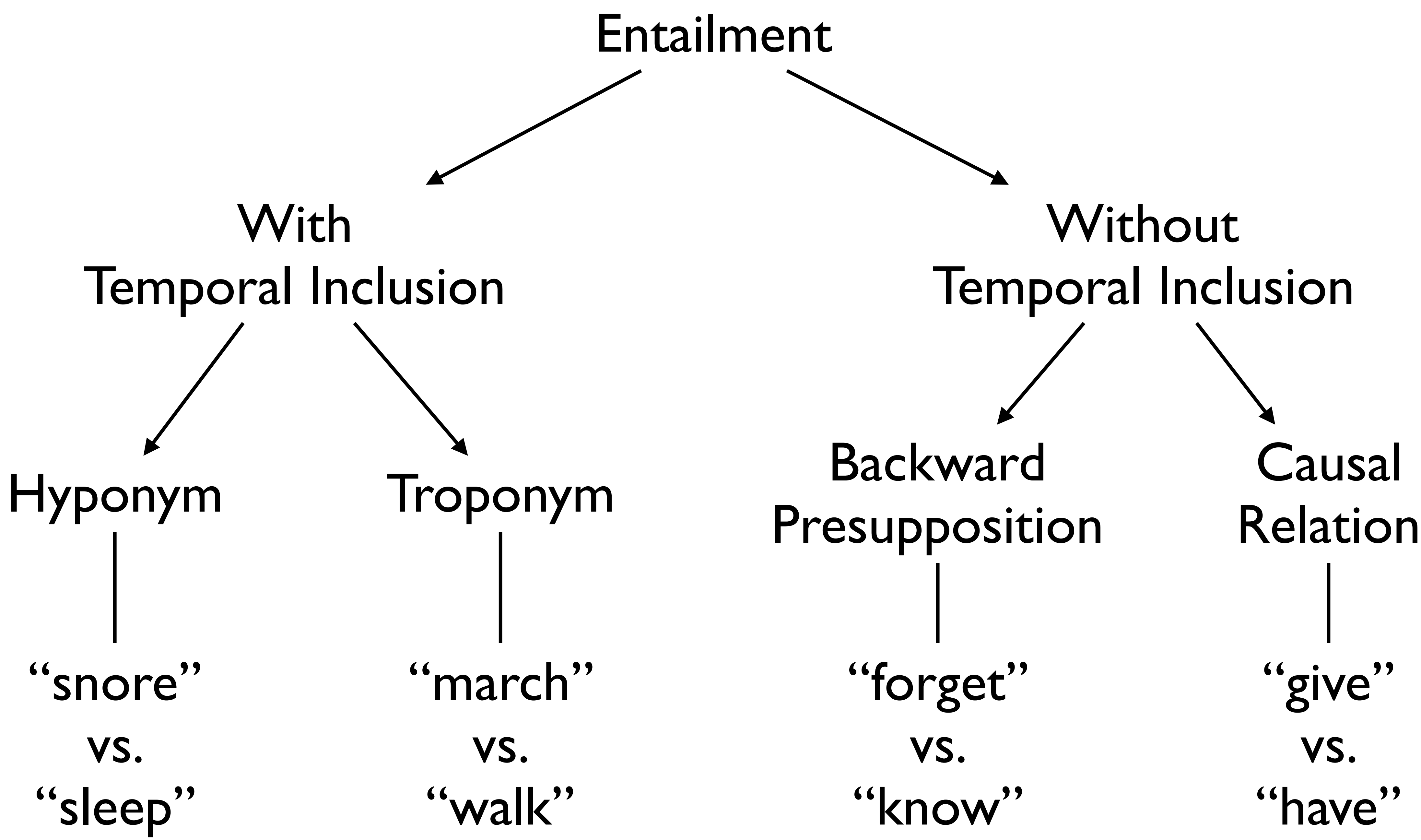
If A is rejected for B, A must have applied for B.

Causative Relations

(V1 causes V2) → (V1 entails V2).

(Give A to B) entails (B have A).

Entailment


Entailment
With Temporal Inclusion
Without Temporal Inclusion
Hyponym
Troponym
Backward Presupposition
Causal Relation
“snore” vs. “sleep”
“march” vs. “walk”
“forget” vs. “know”
“give” vs. “have”

WordNet Similarity

Path Lengths
Wu and Palmer, 1994
Leacock and Chodorow, 1998

Resnik, 1995
Jiang and Conrath, 1997
Lin, 1998

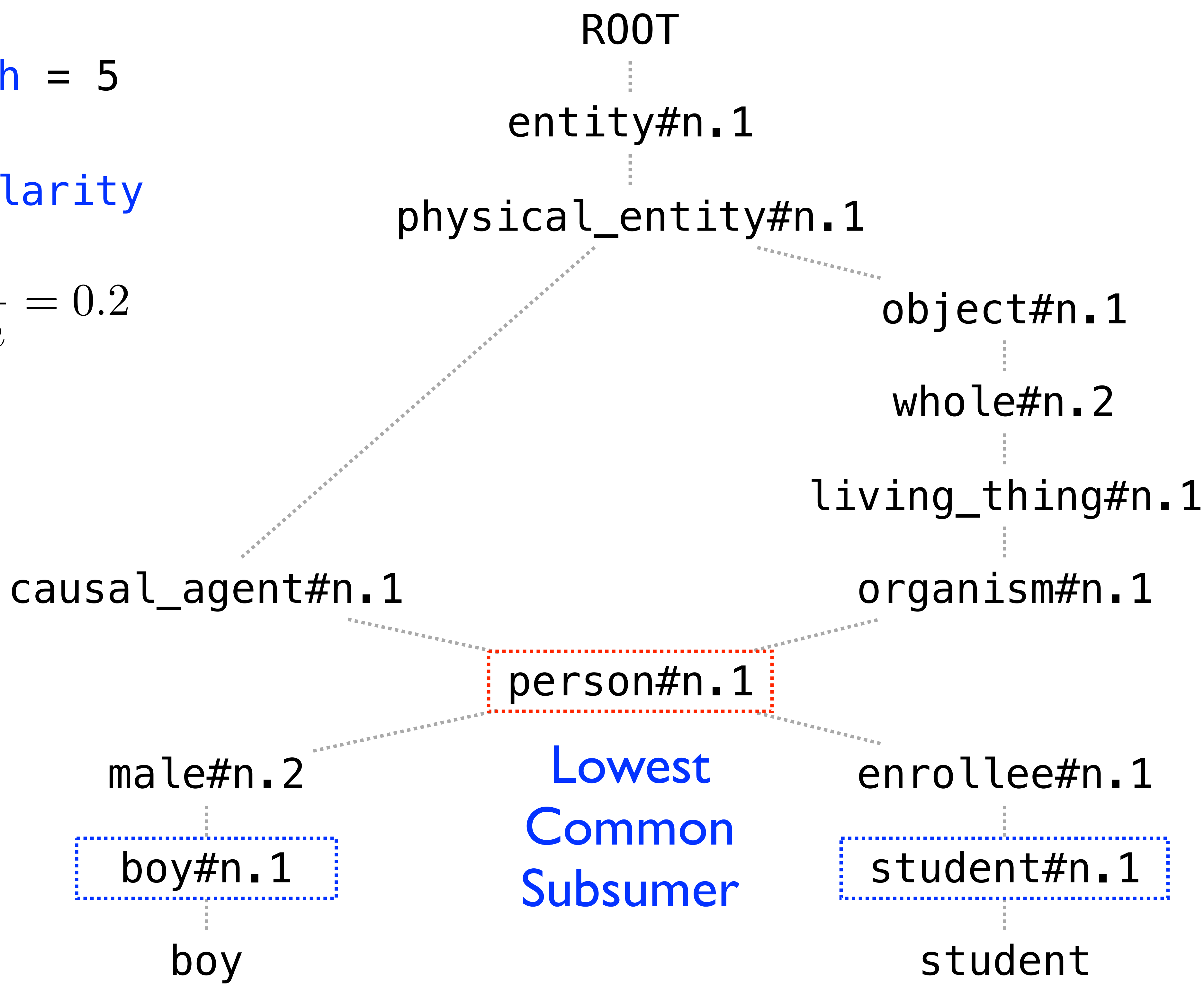
http://ws4jdemo.appspot.com

Path Length

Path = 5

Similarity

$$\frac{1}{path} = 0.2$$

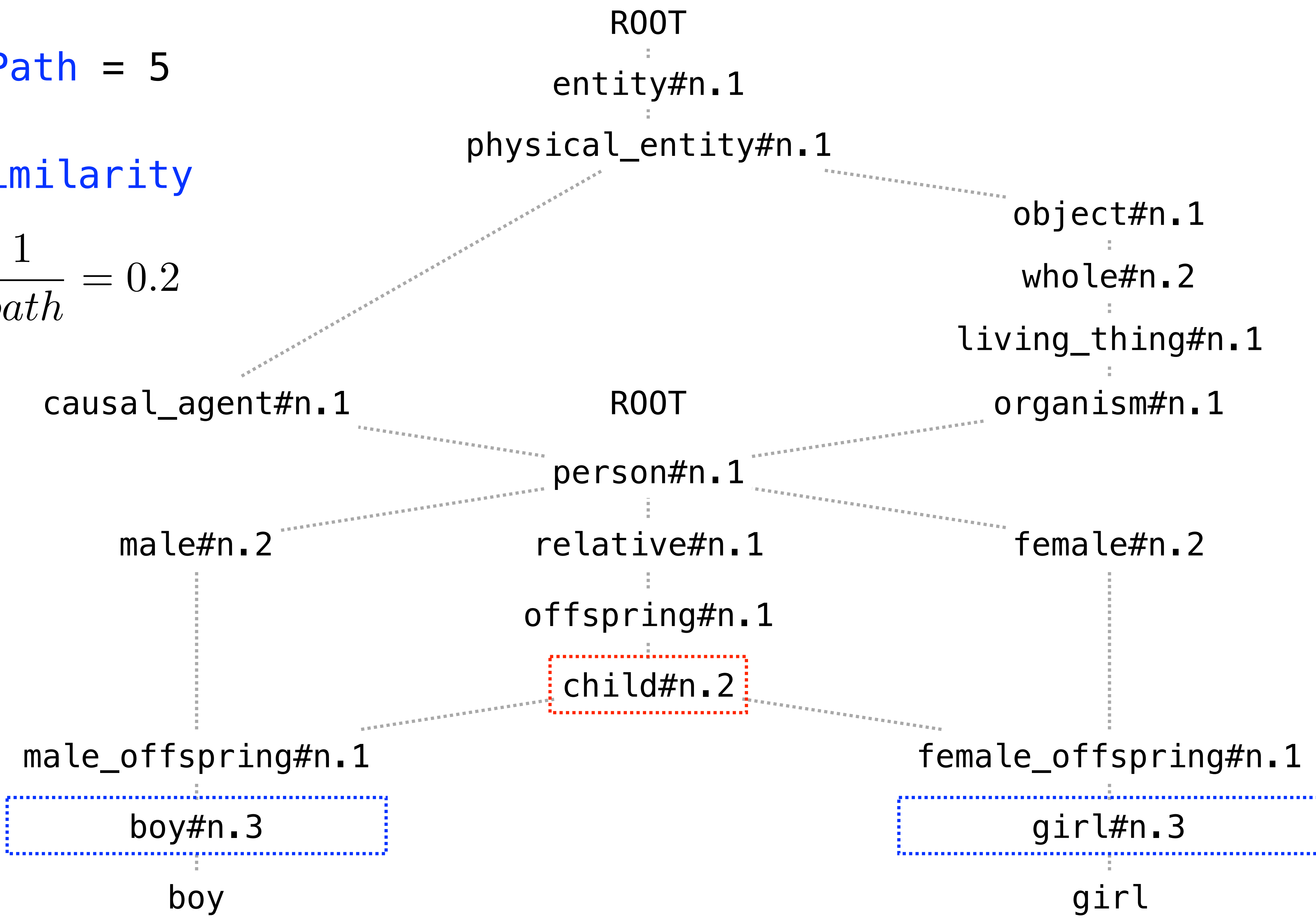


Path Length

Path = 5

Similarity

$$\frac{1}{path} = 0.2$$



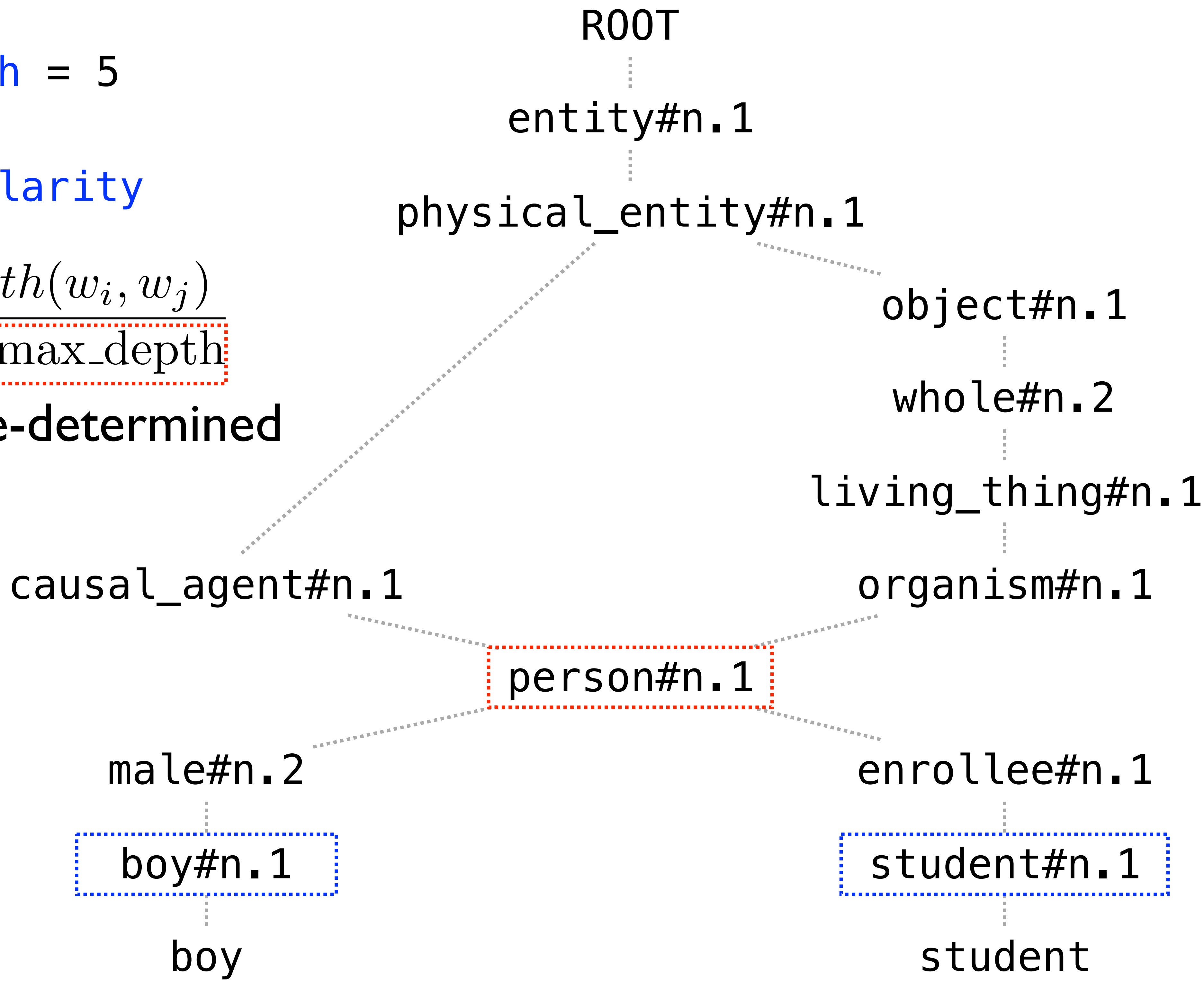
Leacock and Chodorow, 1998

Path = 5

Similarity

$$-\log \frac{path(w_i, w_j)}{2 \cdot \boxed{\text{max\_depth}}}$$

pre-determined



Leacock and Chodorow, 1998

Path = 5

Similarity

$$-\log \frac{path(w_i, w_j)}{2 \cdot \mathrm{max\_depth}}$$

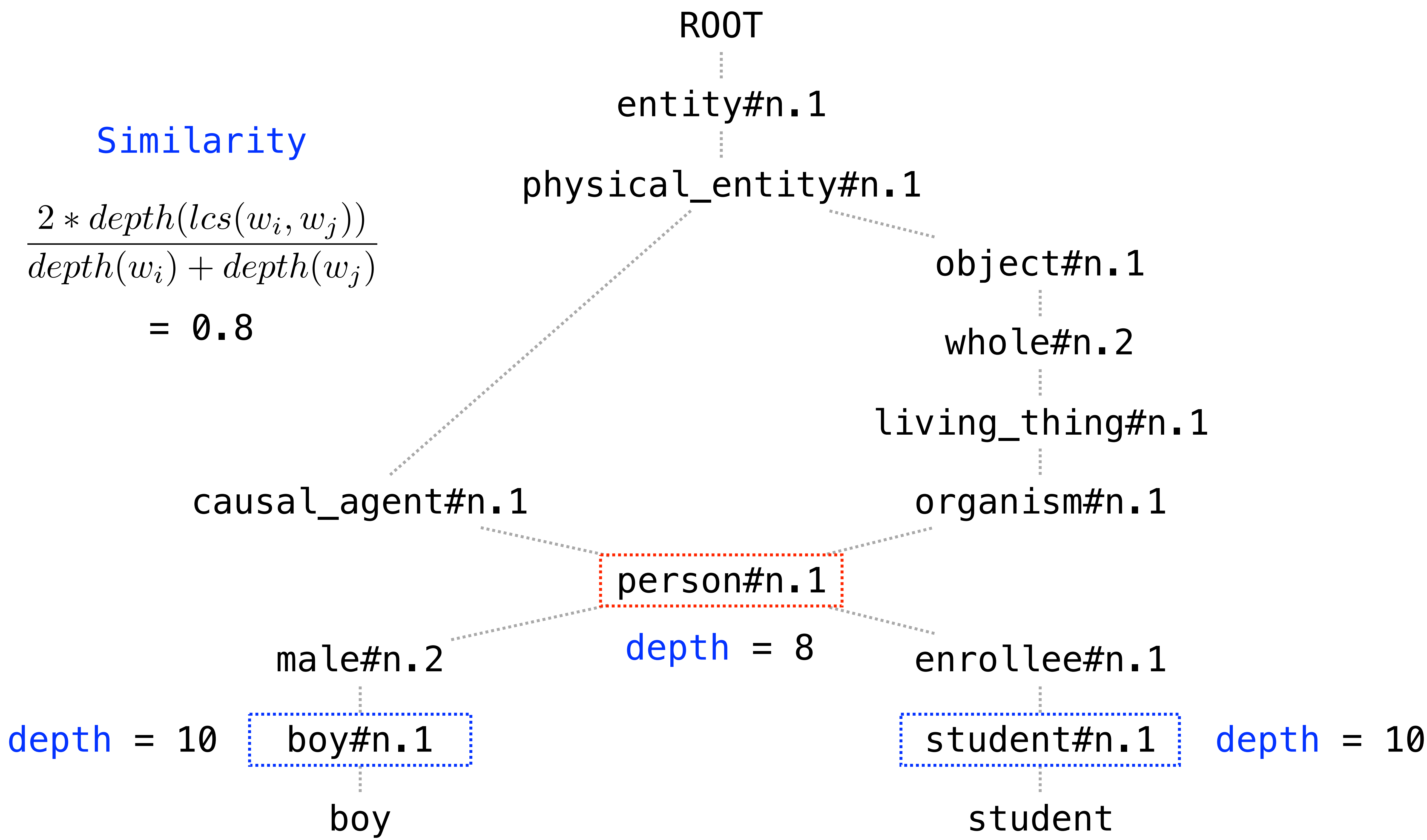
ROOT
entity#n.1
physical_entity#n.1
object#n.1
whole#n.2
living_thing#n.1
causal_agent#n.1
ROOT
organism#n.1
person#n.1
male#n.2
relative#n.1
female#n.2
offspring#n.1
child#n.2
male_offspring#n.1
female_offspring#n.1
boy#n.3
girl#n.3
boy
girl

Wu & Palmer, 1994

Similarity

$$\frac{2 * depth(lcs(w_i, w_j))}{depth(w_i) + depth(w_j)}$$

= 0.8

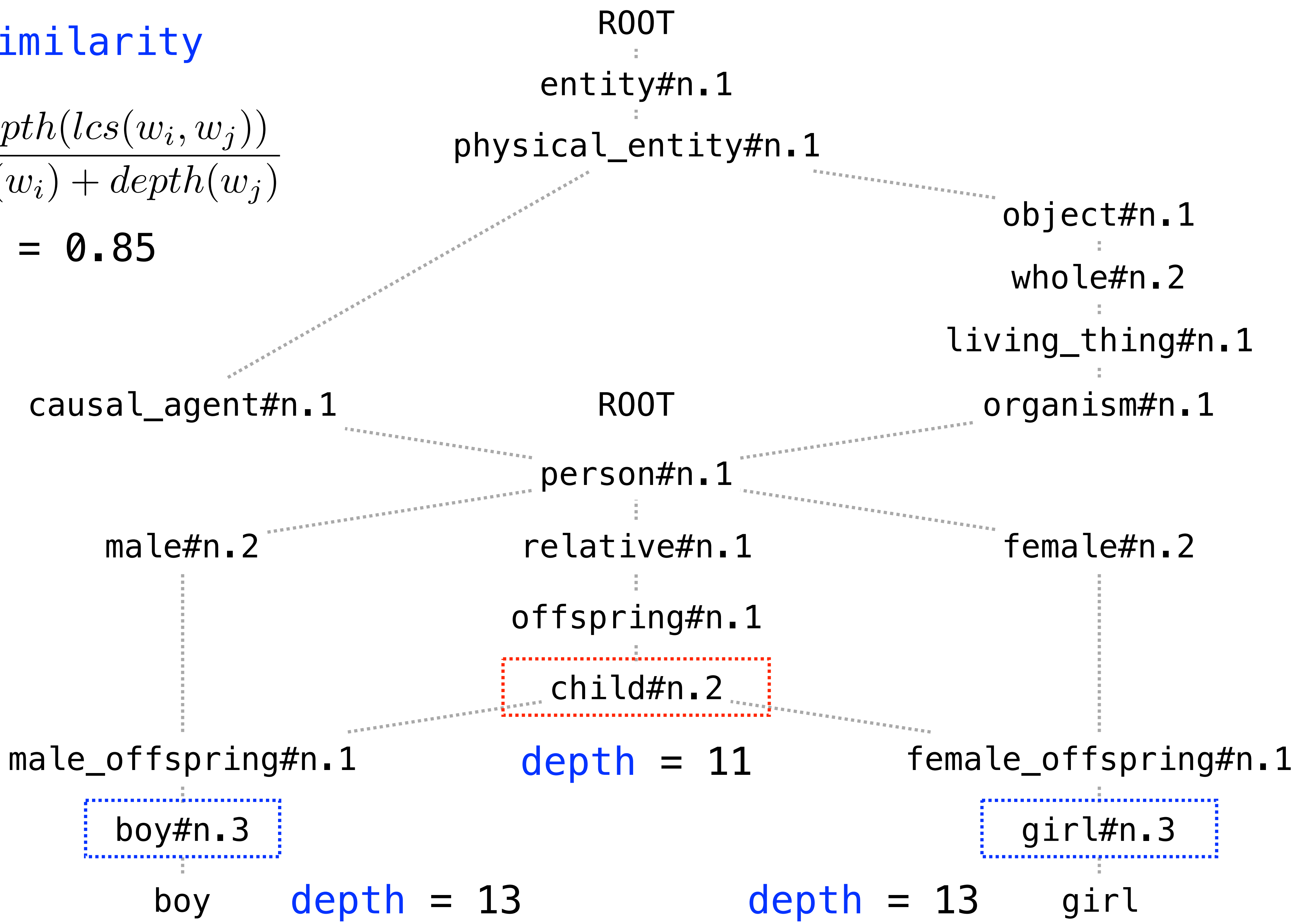


Wu & Palmer, 1994

Similarity

$$\frac{2 * depth(lcs(w_i, w_j))}{depth(w_i) + depth(w_j)}$$

= 0.85



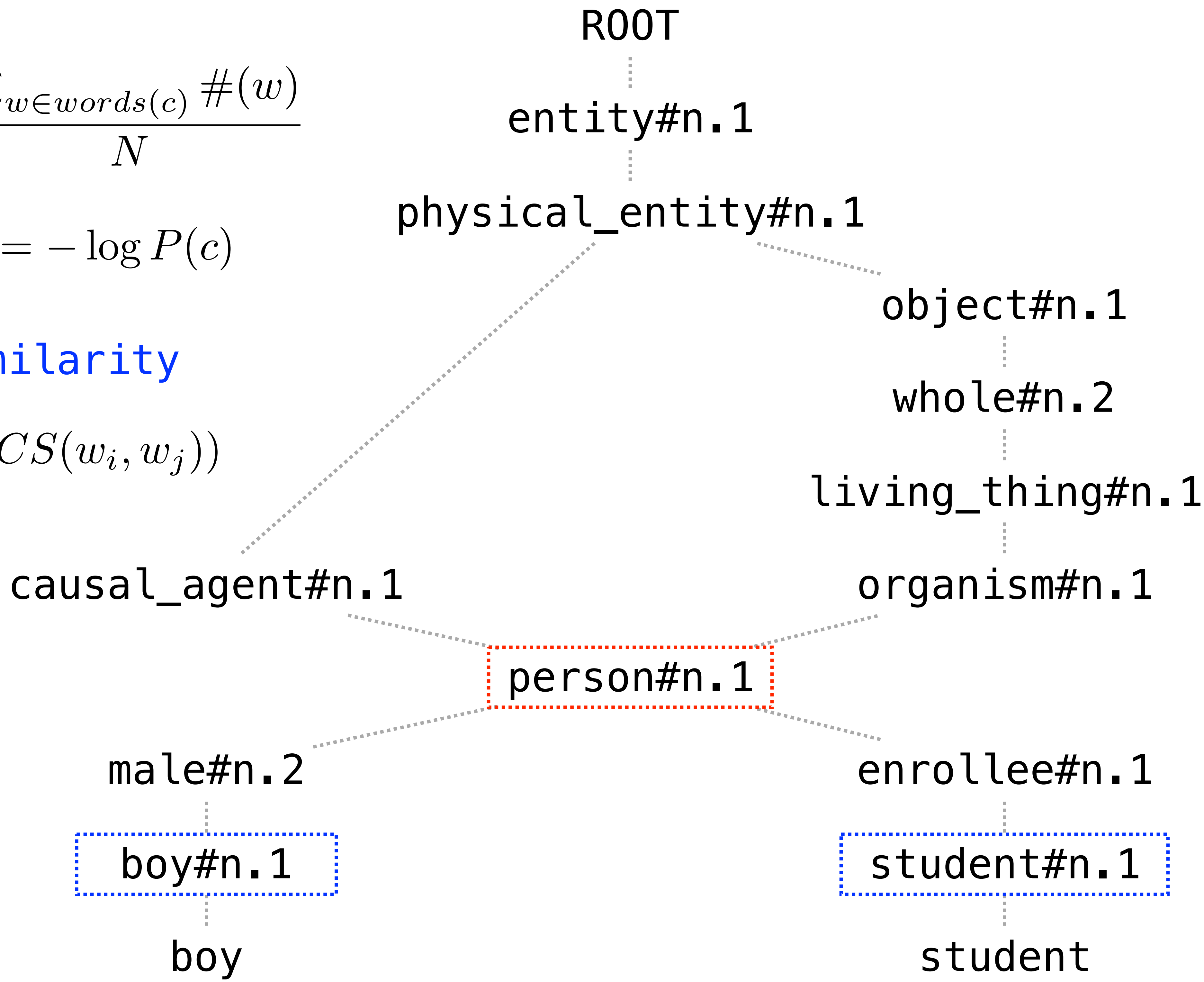
Resnik, 1995

$$P(c) = \frac{\sum_{w \in words(c)} \#(w)}{N}$$

$$IC(c) = -\log P(c)$$

Similarity

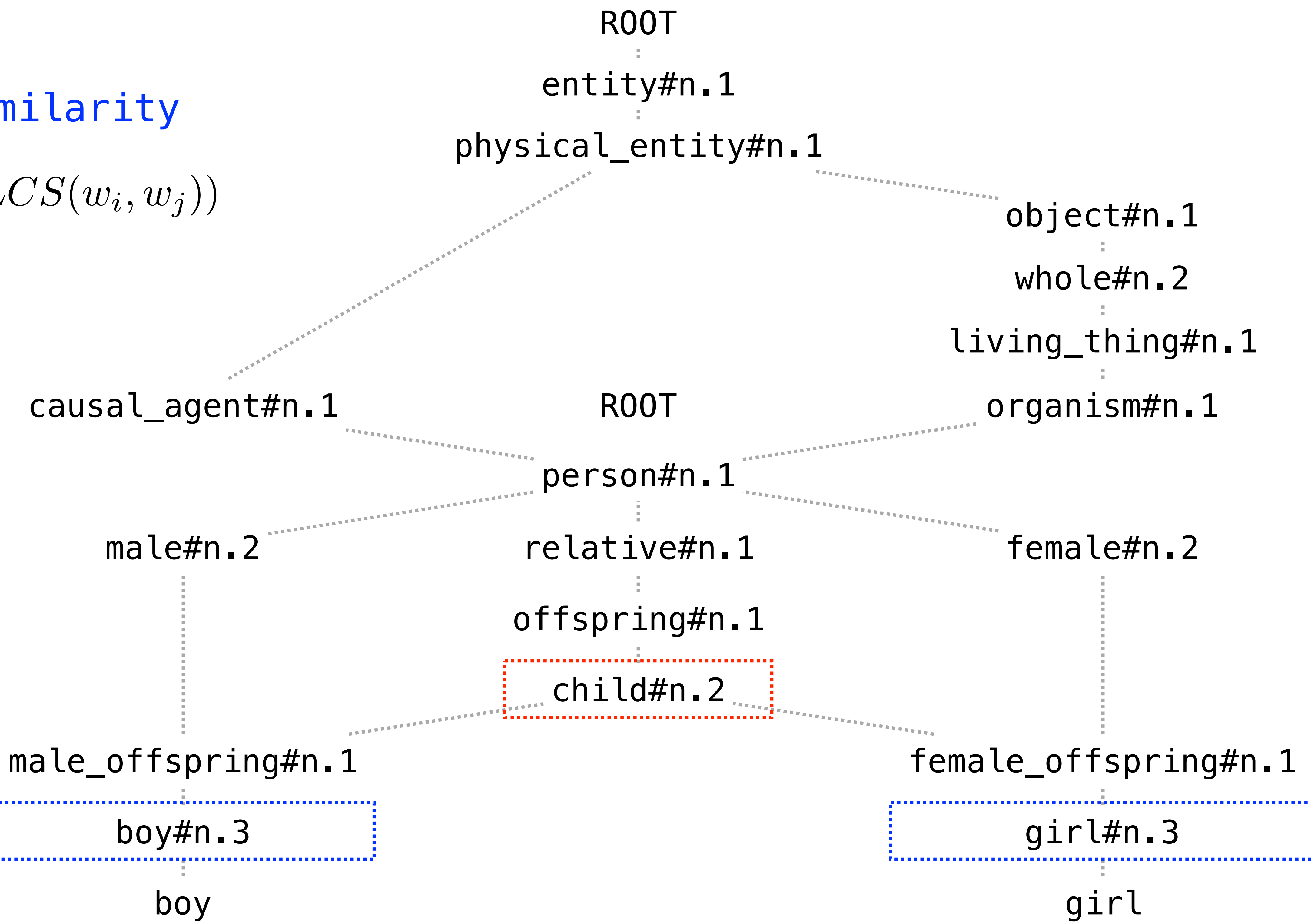
$$IC(LCS(w_i, w_j))$$



Resnik, 1995

Similarity

$IC(LCS(w_i, w_j))$

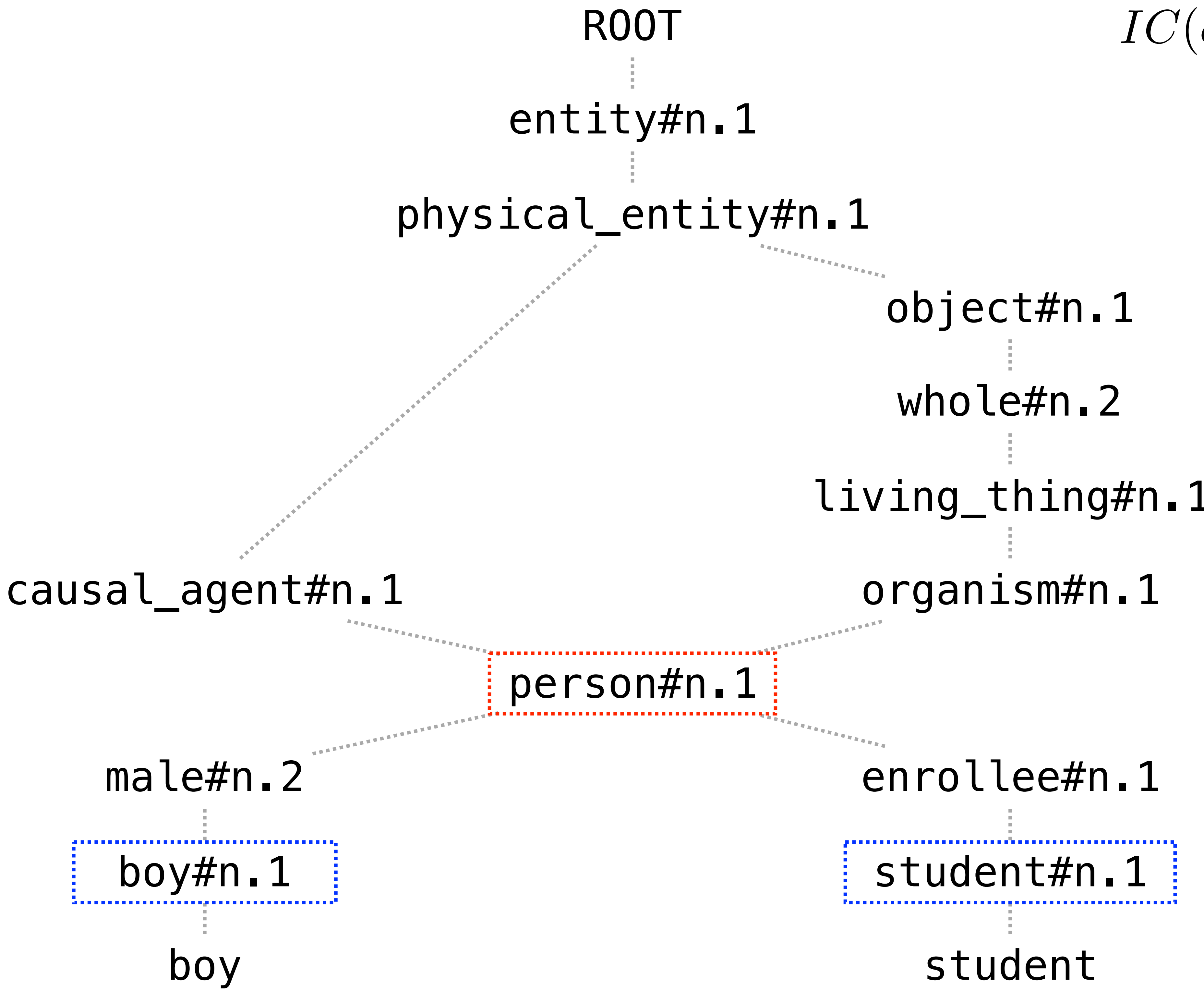


Jiang & Conrath, 1997

$$(IC(c_i) + IC(c_j)) - 2 \cdot IC(LCS(c_i, c_j))$$

Lin, 1998

$$\frac{IC(LCS(w_i, w_j))}{IC(c_i) + IC(c_j)}$$



Jiang & Conrath, 1997

$$(IC(c_i) + IC(c_j)) - 2 \cdot IC(LCS(c_i, c_j))$$

Lin, 1998

$$\frac{IC(LCS(w_i, w_j))}{IC(c_i) + IC(c_j)}$$