Beyond Valence: Deeper Semantic Representations in the Lexicon

James Pustejovsky
Brandeis University

DSNA 24
University of Colorado
Boulder, CO
June 3, 2023

Outline

- Argument Structure in Verbs
  - Syntactic Type and Context
  - Valence and Semantic Type
- Nominal Argument Structure
  - Relational Nominalizations
  - Relational Nominals
- Qualia Structure for Nouns
  - Richer Semantics for Non-verbal Categories
  - Qualia Structure as Valence for Nouns and Adjectives

Dictionary Lexical Entry: Syntactic Information

Syntactic category: a group of lexical items presenting morphological, syntactic, and semantic similarities.

- *house* noun
- *arrive* (intransitive) verb
- *easy* adjective/ adverb

Knowing the syntactic category helps us determining the syntactic structure that the word can project when it is the *head* of a phrase.

- $[_{DP}$ the/three $[_{NP}$ old *houses*]]

Dictionary Lexical Entry: Definition

Definition: explanation of the word meaning.

- *Genus*: larger class in which the word is included → hypernym
- *Differentiae*: distinguishing features that differentiate the concept from the other members of the same class
  - *house* 1: a **building** in which people leave.
  - *arrive* 1: to **reach** a destination.
- Synonyms
  - *easy*: easygoing, relaxed.
  - *arrive*: reach or attain.
- Antonyms
  - *easy* 2: not severe; not steep or abrupt.

Valence and Function-Argument Behavior

Familiar concept from Algebra

- Both mathematical relations and natural language sentences can be seen as **saturated** expressions having two parts: a relational component and its valence (the list of arguments).
  - $5 < 7$
  - $\frac{1}{2} = .5$
- 'Saturated' refers to the fact that there are no argument values missing. Together, they make up a proposition that is either true or false.
  - Karen likes chocolate.
    like(Karen, chocolate)
  - Peter attended MIT.
    attend(Peter, MIT)

Semantic Features and Semantic Types

Selectional requirements are not limited to the syntactic category.

- *arrive*
  - AS: arg_1[cat=DP], arg_2[cat=PP]
  - *John arrives at the book.
  - AS [rev.]: arg_1[cat=DP], arg_2[cat=PP, sem type=LOCATION]
- *hate*
  - AS: arg_1[cat=DP], arg_2[cat=DP]
  - *The happiness hates John.
  - AS [rev.]: arg_1[cat=DP, animacy=+], arg_2[cat=DP]
- *easy*
  - AS: arg_1[cat=DP]
  - **easy house* vs. *easy exam/ easy recipe/ easy prey*
  - AS [rev.]: arg_1[cat=DP, sem type=EVENT]

Semantic type: the kind of entity denoted by the lexical item
Semantic feature: smaller (*sublexical*) component of word meaning

Defining Argument Structure

- Argument structure is the specification of both the number and type of logical arguments to a predicate.

(1)
$$\begin{bmatrix} \textbf{smile} \\ \text{AS} = \begin{bmatrix} \text{ARG}_1 = \begin{bmatrix} x = \begin{bmatrix} \text{CAT} = \textbf{DP} \\ \text{SEM TYPE} = \textbf{animate} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(2) smile(x:{DP,animate})

Unexpressed Arguments

- IMPLICIT ARGUMENTS: Both object in (a) and subject in (b):
  - Kro stroit etot dom? - Ja stroju.
    Who builds this house? - I build.
    'Who is building this house? - I am.'
  - No lo quiere.
    Not it wants.
    'He/she does not want it.'
- DEFAULT ARGUMENTS:
  - Mary ate (dinner) hastily.
  - John built a table (out of wood).
- SHADOW ARGUMENTS:
  - They lived a *(happy) life.
  - The couple danced a *(beautiful) dance.
  - She buttered her toast with *(organic) butter.

Multiple Lexical Structures - Generative Lexicon

- Argument structure: specification of number and type of the predicate's arguments.
  - Rich argument typology, including unexpressed arguments.
- Event structure: overall event type of the predicate, its parts (*subevents*), their relative ordering and prominence.
  - *Dynamic Event Structure*: scalar properties of events.
- Qualia structure: decomposed representation of lexical meaning in terms of four dimensions
  a. FORMAL (F): basic semantic typing (**is_a** relation: *fence* is a kind of 'barrier'); features that distinguish the object within a larger domain
  b. AGENTIVE (A): factors involved in the origin of an object ('build' for *fence*).
  c. TELIC (T): purpose/function of the object ('separate/ prevent from entering or leaving' for *fence*).
  d. CONSTITUTIVE (C): relation between an object and its proper parts ('wood/metal' for *fence*), or what it is a part of.

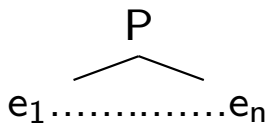
Subatomic Lexical Structure

$$\begin{bmatrix} \alpha \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = x \\ \dots \end{bmatrix} \\ \text{EVENTSTR} = \begin{bmatrix} \text{EVENT1} = e1 \\ \text{EVENT2} = e2 \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{CONST} = \textbf{what } x \textbf{ is made of} \\ \text{FORMAL} = \textbf{what } x \textbf{ is} \\ \text{TELIC} = e_2\textbf{: function of } x \\ \text{AGENTIVE} = e_1\textbf{: how } x \textbf{ came into being} \end{bmatrix} \end{bmatrix}$$

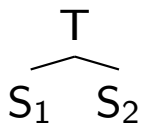
An independent level of Event Structure

- ES as a level of representation that is distinct from the representation of other lexical properties (AS), although it is related to them.
- Event typology (Pustejovsky (1995, 2011)
  a. STATE: simple event evaluated without referring to other events: *be sick*, *love*, *know*

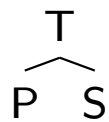
  b. PROCESS: sequence of events identifying the same semantic expression: *run*, *push*, *drag*



  c. TRANSITION: event identifying a semantic expression evaluated with respect to its opposition: *give*, *open*; *build*, *destroy*
  Two-state transition (ACH):



  Extended transition (ACC):



Argument Structure in Nouns

SORTAL_NOUN(arg_{ref})

(3) a. Jaco is a *dog*.
DOG(*Jaco*)

b. Mike is a *professor*.
PROFESSOR(*Mike*)

c. Some professor is American.
$\exists x[\text{PROFESSOR}(x) \wedge \text{AMERICAN}(x)]$

1. INHERENTLY RELATIONAL NOUNS, which establish an association between two or more arguments by virtue of their inherent lexical meaning.
2. QUALIA-BASED RELATIONAL NOUNS, which establish an association between two or more arguments by virtue of aspects of their qualia structure.
3. NOUNS DERIVED FROM VERBS AND ADJECTIVES, which inherit their arguments from their bases.

Argument Structure in Nouns

RELATIONAL_NOUN(arg_{ref}, arg_{rel})

(4) a. John is Mary's father/ John is the father of Mary
FATHER(John, Mary)

b. John is Mary's friend/ John is a friend of Mary's
FRIEND(John, Mary)

c. John is Mary's student/ John is a student of Mary's
STUDENT(John, Mary)

- **Kinship terms**, such as *father* and *sister*;
- **social relations**, such as *student* and *friend*, take as their argument an animate individual related to another individual by this relation.

Argument Structure in Nouns

PART_OF(arg_{ref},arg_{rel})

(5) a. John's hair
HAIR(arg_{ref}, John)

b. the bank of the Charles River
BANK(arg_{ref}, Charles)

c. the roof of the hotel
ROOF(arg_{ref}, the_hotel)

- These nouns encode both the referential argument (entity referred to as the PART) and the relational argument (the WHOLE);
- *hair* takes as its argument the animate entity that has it;
- *bank* presupposes it is a part of a river or a lake;
- *roof* is part of a building of some sort.

Argument Structure in Nouns

ATTRIBUTE_OF(arg_{ref}) = arg_{val}

(6) a. the color of the car
COLOR(arg_{val}, the_car)

b. the shape of the table
SHAPE(arg_{val}, the_table)

- The nominal attribute refers to a specific value for the noun;
- COLOR_OF(the_car) = green;
- SHAPE_OF(the_table) = square.

Argument Structure in Nouns

QUALIA_RELATIONAL(arg_{ref}, arg_{Q})

(7) a. Pythagoras' theorem
THEOREM(arg_{ref}, Pythagoras)

b. Gaudí's house
HOUSE(arg_{ref}, Gaudí)

c. Mary's report
REPORT(arg_{ref}, Mary)

- These nouns exhibit a relation that is associated with the AGENTIVE qualia role, rather than the FORMAL role.

Argument Structure in Nouns

REPRESENTATION_OF(arg_{ref},arg_{rel})

(8) a. Da Vinci's painting
PAINTING(arg_{ref}, Da Vinci)

b. Rodin's statue
STATUE(arg_{ref}, Rodin)

- Unlike artifactual nouns such as *house* representational artifacts are actually inherently relational.
- arg_{ref} is the depiction of the relational argument in both examples: the painting is a two-dimensional image of la joconde, while the statue is a three-dimensional model of David.

(9) a. painting of la joconde
PAINTING(arg_{ref}, la_joconde)

b. statue of David
STATUE(arg_{ref}, David)

- DEVERBAL NOUNS can refer to the event or state encoded by the base verb, or to one of its participants.
- They very frequently saturate the Agent argument (as in (12b)), and also sometimes the Instrument argument, as in (12d).

(10) a. The *translation* of the ancient manuscript took two years.
b. I read the impeccable *translation* of the ancient manuscript.

(11) a. John_{EXP} loves $\text{Mary}_{\text{CAUSE}}$ / $\text{John's}_{\text{EXP}}$ love for $\text{Mary}_{\text{CAUSE}}$.
b. John_{EXP} fears the $\text{dentist}_{\text{CAUSE}}$ / $\text{John's}_{\text{EXP}}$ fear of the $\text{dentist}_{\text{CAUSE}}$.

(12) a. John manages the Department.
MANAGE(Agent,Theme)
b. manager of the Department
MANAGER_OF(Theme)
c. The machine calculates taxes.
CALCULATE(Agent,Theme)
d. tax calculator
CALCULATOR_OF(Theme)

Qualia Structure

- Formal (F): encoding taxonomic information about the lexical item (the `is-a` relation);
- Constitutive (C): encoding information on the parts and constitution of an object (`part-of` or `made-of` relation);
- Telic (T): encoding information on purpose and function (the *used-for* or `functions-as` relation);
- Agentive (A): encoding information about the origin of the object (the `created-by` relation).

What is a *Quale*?

- A *Quale* (singular of *Qualia*) indicates **a single aspect of a word's meaning**, defined on the basis of the relation between the concept expressed by the word and another concept that the word evokes.
- Among the conceptual relations that a word may activate Qualia relations as defined in GL are those that are exploited in our understanding of linguistic expressions.
- *fresh bread* = "bread which has been baked recently."

Qualia Structure

$$\begin{bmatrix} \alpha \\ \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \textbf{what } x \textbf{ is} \\ \text{C} = \textbf{what } x \textbf{ is made of} \\ \text{T} = \textbf{function of } x \\ \text{A} = \textbf{origin of } x \end{bmatrix} \end{bmatrix}$$

Qualia Structure

$$\begin{bmatrix} \textit{car} \\ \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \textbf{vehicle} \\ \text{C} = \textbf{engine, door, wheels,...} \\ \text{T} = \textbf{drive} \\ \text{A} = \textbf{build} \end{bmatrix} \end{bmatrix}$$

Linguistic phenomena motivating Qualia relations

Contextual modulations of meaning, due to selecting predicate.

- This car weighs over 2,000 lbs.
- We buy vehicles such as cars and buses.
- John started the car.
- You should warm your car up in winter.
- Did you lock the car?
- The car screeched down the road.

Linguistic phenomena motivating Qualia relations

- Inference of implicit predicates from particular constructions.
  - Verb-Noun combination: finish the beer/cake.
  - Adjective-Noun combination: comfortable chair/shoes.
  - Noun-Noun compounding: dinner dress/table.
- Light verbs specifications.
  - take a tablet.
  - take a train.
- Noun-to-Verb transformations.
  - fax a document.
  - microwave the chicken.
  - lace the shoes.

Methodology for identifying Qualia Values

Pustejovsky and Jezek (2012)

- Linguistic evidence determines what information is stated to be lexically associated with the Qualia Structure of a word.
- If a piece of knowledge is regularly exploited in our understanding of linguistic expressions, then it is likely to be part of lexical information and it is included in QS as default value for a Quale.

Identifying Qualia Values

- The rock shattered the window.
  - ([C = pane])
- Wooden windows are prone to rotting.
  - ([C = frame])
- John was going to paint his room.
  - ([C = walls])
- She has swept the room.
  - ([C = floor])
- The knife cut his finger.
  - ([C = blade])
- sharp knife.
  - ([C = blade])

Unspecified roles

- Not all lexical items carry a value for each Q-role.
- Some values are left unspecified, while others are populated with more than one value.
- Nouns denoting natural kinds (e.g., *rock*, *fish*, *air*, *sea*) typically do not have a value for the A.

The Formal

- The Formal (F) encodes the relation between the entity denoted by the word and the category it belongs to.
- This relation enables one to grasp the nature of an entity by discriminating it from other kinds.
- What type of entity is x denoting? *rock* denotes a natural kind, *table* denotes an artifact, *car* denotes a vehicle, *park* denotes a location, *water* denotes a liquid, *plant* denotes a living thing, *fish* denotes an animal, *hand* denotes a body part, *glass* denotes a container, and so on.

The Formal

- More classifications are possible for the same type of object: for example, a *knife* can denote both a weapon or a kitchenware.
- Classifications at different levels of generalization are available for reference:
- a liquid such as water.
- fluids such as water or air.
- substances such as fluids, salts, glucose and carbon dioxide.

Formal Factors for Nouns Denoting Concrete Entities

- Spatial characteristics, intrinsic orientation.
- Size and dimensional properties.
- Shape and form.
- Color.
- Position.
- Surface.

Values for Formal Factors

- a red car
  - (Color_F)
- a long dress
  - (Dimension_F)
- a round table
  - (Shape_F)
- a red pen
  - (Color_F) or T/C (depending on contextual interpretation)
- a flat screen
  - (Shape_F)
- a thick sweater
  - (Dimension_F)

Values for Formal Factors

- the length of the table
  - (Dimension$_F$)
- the facade of the building
  - (Orientation$_F$)
- wipe the floor
  - (Surface$_F$)
- a large round table
  - (Size$_F$) (Shape$_F$)
- *a round and square table
  - (Shape$_F$) (Shape$_F$)

The Constitutive Quale

- The Constitutive (C) role encodes information about what is "inside" the object denoted by the word, particularly the material the object is made of (i.e., its stuff), and the parts it consists of.
- There is a fundamental distinction between inherently individuated things, such as humans, tigers, and trees, and inherently undifferentiated stuff, such as water, air, and sand.
- This is reflected in the count vs. mass distinction in linguistics.

Count Nouns and Mass nouns in GL

COUNT NOUN: (where $\alpha \neq \beta$).

$$\begin{bmatrix} \textbf{N} \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \alpha \\ \text{C} = \beta \end{bmatrix} \end{bmatrix}$$

MASS NOUN:

$$\begin{bmatrix} \textbf{N} \\ \text{QUALIA} = \begin{bmatrix} \text{F} = \alpha \\ \text{C} = \alpha \end{bmatrix} \end{bmatrix}$$

The Formal-Constitutive Equivalence Constraint for mass nouns in GL

$$\left[\begin{array}{l} \textit{water} \\ \textsc{qualia} = \left[\begin{array}{l} \textsc{f}/\textsc{c} = \textbf{liquid} \end{array}\right] \end{array}\right]$$

$$\left[\begin{array}{l} \textit{rock} \\ \textsc{qualia} = \left[\begin{array}{l} \textsc{f}/\textsc{c} = \textbf{solid\_substance} \end{array}\right] \end{array}\right]$$

but cf.

$$\left[\begin{array}{l} \textit{rock} \\ \textsc{qualia} = \left[\begin{array}{l} \textsc{f} = \textbf{solid\_object} \\ \textsc{c} = \textbf{solid\_substance} \end{array}\right] \end{array}\right]$$

MADE-OF RELATION Introduced in Composition

- Nominal compounding
  - plastic bag
  - paper cup
  - leather shoes
  - gold watch
  - milk chocolate
- Adjective-Noun constructions
  - a golden ring
  - a wooden floor
  - a metallic paint

Hidden Events and the Telic Role

- The Telic relation (T) encodes information about the intended use or function of an object.
- It expresses the relation that allows us to grasp what an entity is by knowing what it is used for.
- It encodes a potential activity of the object.
- First systematic mention of Telic in Pustejovsky and Anick (1988) as hidden event.

Default Values for Telic Role

- We skipped the cake and settled for another coffee.
  - ([T = eat]) ... ([T = drink])
- the next customer.
  - ([T = take_care_of])
- the next slide.
  - ([T = project])
- This is a difficult problem.
  - ([T = solve])
- This is a difficult question.
  - ([T = answer])

The Agentive Role

- The Agentive quale (A) encodes information about the origin of the object denoted by N.
- It provides a mechanism for discriminating those objects that present themselves to us (occurring naturally) from the various artifacts that we create through our own activities and intentional behavior.
- Differently from the value of T, A introduces an existentially bound or existentially quantified event, that precedes the existence of the object.

Agentive-Telic Pairing

- Inherent in this is an association between the Agentive and Telic of the object, i.e., the object is made for a purpose (Agentive-Telic pairing).
- Natural kinds lack this association, as they do not encode an Agentive value.

Default values for Agentive

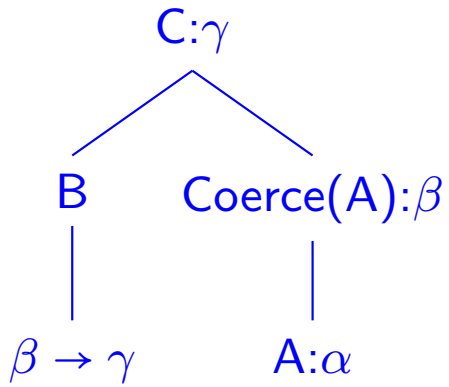
- He just finished and published his first novel.
  - ([A = write])
- Wes Anderson has started a new movie.
  - ([A = direct, film])
- John began a large oil painting yesterday.
  - ([A = paint])
- Mary made a cake.
  - ([A = bake)
- Her mother made her a dress.
  - ([A = sew])
- fresh coffee
  - ([A = brew])

General Theory of Type Coercion

Pustejovsky (2011)

- A dynamic operation in the composition of a phrase, converting an argument to the type expected by its predicate,
  - where otherwise their combination would result in a type conflict.
- Any predicate can potentially coerce its argument;
- Languages vary in the degree to which coercion applies in the grammar.

(13) COERCION: In the structure $[B\ A]_C$, where A is of type α, and B is a function of type, $\beta \rightarrow \gamma$, a coercion function, *Coerce*, can apply to A, changing its type to that expected by B: i.e., *Coerce*(A) is of type β. Now, the functional type, $\beta \rightarrow \gamma$, applies to *Coerce*(A), to return γ, the type of the phrase C.



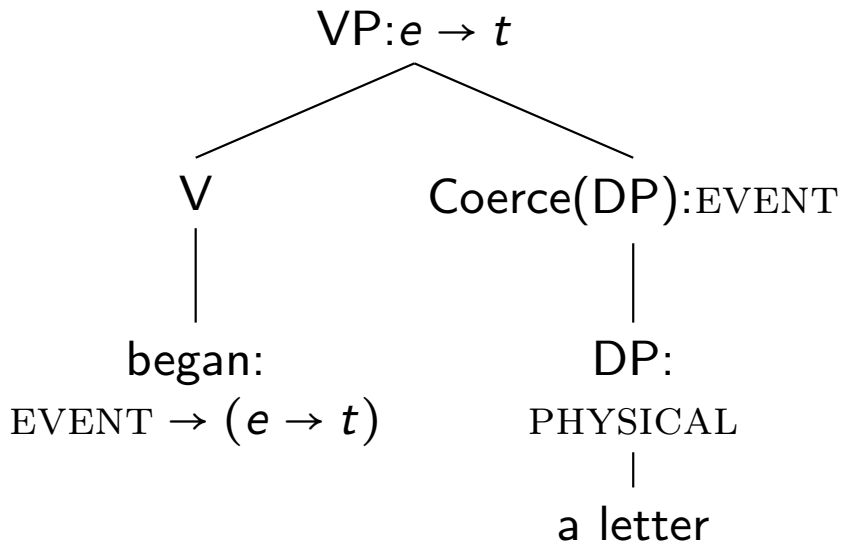
(14) a. Mary began [to write a letter to her mother]$_{\text{EVENT}}$.
SELECTION: EVENT

b. Mary began [a letter to her mother]$_{\text{PHYSICAL}}$.
COERCION: ENTITY $\Longrightarrow$ EVENT

(15) a. The spokesperson denied [the statement]$_{\text{PROPOSITION}}$.
SELECTION: PROPOSITION

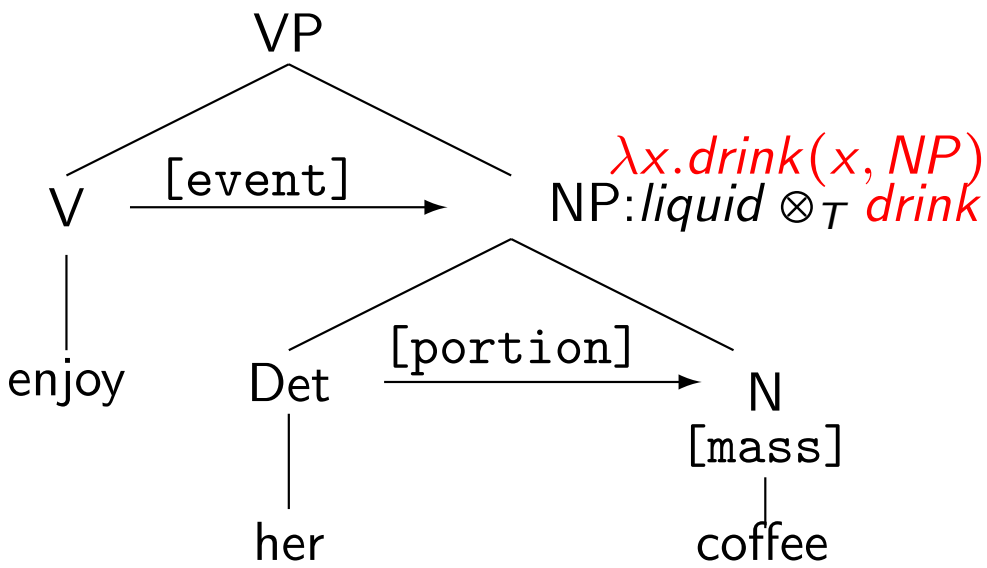
b. The spokesperson denied [the attack]$_{\text{EVENT}}$.
COERCION: EVENT $\Longrightarrow$ PROPOSITION

(16)



Type Coercion: Qualia Exploitation

1. Mary enjoyed her coffee.



Qualia Structure at Work in Composition

Type Coercion

- The verb *begin*, which is typed as a relation between events and individuals, EVENT $\rightarrow (e \rightarrow t)$, applies to the result of the coerced DP, which has undergone a coercion from PHYSICAL to EVENT. The result is a conventional VP type, $e \rightarrow t$.
- in (14b) the EVENT information associated with a *letter* comes from its AGENTIVE qualia role, i.e., *writing*.

(17) a. John enjoyed the movie. (TELIC = "watching")
b. Mary finished the coffee before we left. (TELIC = "drinking")
c. Mary began her dessert. (TELIC = "eating")

Conclusion

- Semantic representations beyond predicate-based valence structures are important;
- Qualia structure provides a systematic and structured encoding of meaning for for nouns and adjectives in language;
- Qualia values are referenced in semantic composition to provide masked or implicit semantic information.