Lexicon in Linguistic Theory

James Pustejovsky
Brandeis University

Olga Batiukova
Autonomous University of Madrid

June 24 - July 19, 2019
2019 LSA Linguistics Institute, University of California Davis

Lecture 5: Structure of a Lexical Entry 2: Semantic Typing and Conceptual Structures

Lecture 5: July 8

Structure of a Lexical Entry 2: Semantic Typing and Conceptual Structures

- Semantic typing
- Lexical ambiguity and regular polysemy
- Conceptual structures

Lexical decomposition

- Linguistically relevant part of word meaning can be viewed as consisting of *sublexical* components, which determine what kind of entity a word denotes (its semantic class).
- Models of decomposition we came across so far: lexical-semantic features (in lexical fields, within Qualia Structure), function-argument pairings (Conceptual Semantics), etc.
- Why do we need decomposition?
  - By representing the internal semantic structure of words, we provide a structural explanation for semantic relations (inference, synonymy, antonymy, hyponymy, meronymy, etc.).
    - Semantic type inheritance: *Chalet* < BUILDING < CONSTRUCTION < ARTIFACT < THING.
  - We reduce lexical variability to a list of basic meaning components, which allow explaining lexical acquisition: i.e., how are we able to incorporate > 400,000 lexical entries and their relations in our mental lexicon?
  - We can account for the combinatorial properties of words.
    - The girl$_{[\text{animate}]}$ hates spinach.

Controversial aspects of lexical decomposition

- A word meaning is not merely the sum of the lexical features into which it is decomposed, and it is not clear what else is needed to complete the meaning of a word.
  - [plant]+[small]+[grown for its blossom] = lex. concept *flower* only if the features are meaningfully configured relative to each other.
  - *open*(x,y): CAUSE(x, (BECOME(OPEN(y))))
    → *open* ≠ 'cause something to become open': lexical features are not words themselves.
    → John burnt the meal and somebody else opened the window to air the room: *John caused the window to become open* (OK)/ *John opened the window* (not OK)

Controversial aspects of lexical decomposition

- What features must be considered as primitive?
  - The only full set of primitives put forward so far is Wierzbicka's Natural Semantic Metalanguage, and none of the existing proposals has been unanimously accepted.
  - The ultimate demonstration that a set of features is psychologically real and not just convenient as a representational device must come from psycholinguistics.
  - Showing that a given lexical feature is crucially involved in different modules of grammar suggests that it is linguistically real.
- What procedure/ criteria must be followed in order to discover them, and how far below word level the decomposition process may proceed?
  - We don't know how far lexical decomposition can take us, but this is not necessarily a bad thing.

Strategies for lexical decomposition

- Dictionary definitions exploit the relationships between different words and use words to define other words.
  - **chalet:** a [small] [house] often [within a larger complex] [where people go for vacations].
- This is not decomposition: the defined word is as complex as the elements used in its definition; in decomposition, the number of primitive elements should be smaller than the number of words.
- It can lead to circular definitions since it fails to define basic word meanings.
- It makes it difficult to establish cross-linguistic generalizations.

Simple predicate-based decomposition

- Lexicalized concepts are seen as combinations of more basic features:
  a. **bachelor**: $\text{HUMAN}(x) \wedge \text{MALE}(x) \wedge \text{ADULT}(x) \wedge \neg\text{MARRIED}(x)$
  b. **kill**: $\text{CAUSE}(x, (\text{BECOME}(\neg\text{ALIVE}(y))))$
- This approach does not relate the meaning of the word with the semantic properties that its syntactic context must have (e.g., the semantic typing of the V arguments).

Parametric decomposition

- Additional arguments are added to a predicative expression depending on the inferences one wishes to perform.
  a. **eat**(x,y,**t,l**): '*x* eats *y* at time *t* in location *l*.'
  b. **give**(x,y,z,**t,l**): '*x* gives *y* to *z* at time *t* at location *l*.'

Structured predicate-based decomposition

- Simple predicate-based decomposition + reference to new factors (e.g., semantic types of the arguments), as in parametric decomposition:
  - **kill**(x,y):
    $[\text{ANIMATE}(x) \wedge \text{ANIMATE}(y) \wedge \text{CAUSE}(x, (\text{BECOME}(\neg\text{ALIVE}(y))))]$
  - Conceptual Structure representation:
    [$_{\text{Event}}$ CAUSE ([$_{\text{Object}}$ ANIMATE], [$_{\text{Object}}$ ANIMATE], [$_{\text{Event}}$ INCH ([$_{\text{State}}$ BE ([$_{\text{Object}}$ ANIMATE] [$_{\text{Place}}$ AT ([$_{\text{Property}}$ not alive])])])])]

Structured predicate-based decomposition in Qualia Structure

- In the Generative Lexicon theory, the semantic parameters used to define word meaning include the four Qualia Structure roles: formal (general class, ISA), constitutive (internal constitution), telic (function), and agentive (origin).

$$\left[\begin{array}{ll} \textit{chalet(x)} & \\ \mathrm{QS} & = \left[\begin{array}{lll} \mathrm{F} & = & \mathrm{house(x)} \wedge \mathrm{small(x)} \\ \mathrm{A} & = & \mathrm{be\_built(x)} \\ \mathrm{T} & = & \mathrm{used\_for\_vacation(x)} \\ \mathrm{C} & = & \mathrm{part\_of\_complex(x)} \end{array}\right] \end{array}\right]$$

  a. **chalet**: $[\mathrm{physical\_object}(x) \wedge \mathrm{building}(x) \wedge \mathrm{used\_for\_vacation}(x)]$
  b. **apple**: $\mathrm{physical\_object}(x)$
  c. **fix**: $[\mathrm{fix}(x, y) \wedge \mathrm{animate}(x) \wedge \mathrm{physical\_object}(y) \wedge \mathrm{artifact}(y)]$

- Enriched typing allows for classifying both predicates and their arguments more precisely:
  a. Fix a chalet.
  b. *Fix an apple.

Semantic types

- Words can be classified according to different criteria: phonological, morphosyntactic, semantic.
- Semantic types: the kind of entity denoted by a lexical item.
  - *table* is a kind of FURNITURE → semantic type?
- Ontology: what entities exist and how they can be grouped and related within a hierarchy.

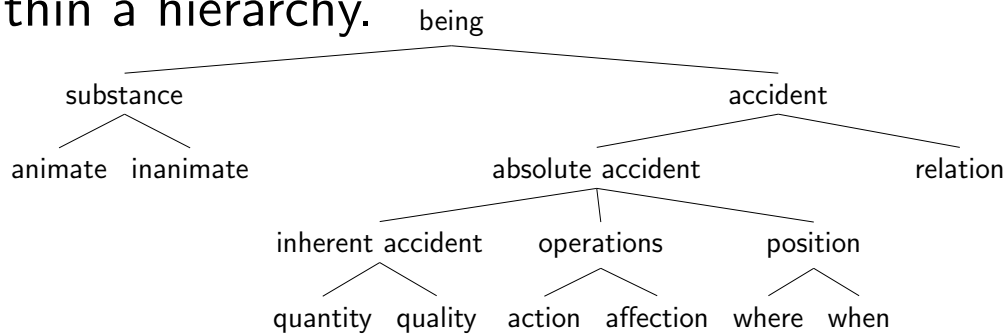


Figure: *Aristotelian ontology*

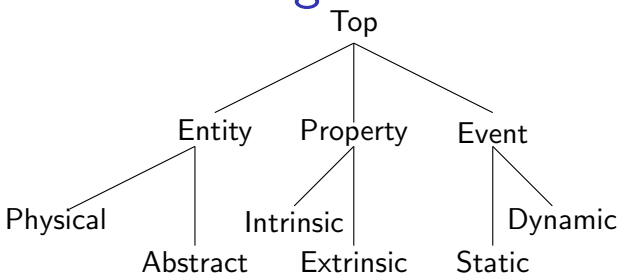


Figure: *Classic Upper Ontology*

Linguistic motivation of semantic types

- Which semantic categories are linguistically relevant?
- A piece of conceptual information can be considered linguistic if it affect other modules of grammar (syntax and morphology).
- Sem. properties systematically shared by arguments of different Vs:
  - ANIMATE
    a. The *kitten* was purring.
    b. The *policemen* cooperated on the case.
  - EVENT
    a. John witnessed {the *robbery/ the bank being robbed*}.
    b. They correctly predicted the financial *collapse*.
  - LOCATION
    a. The tourists reached *the city*.
    b. The tribe inhabited this *valley* for many centuries.
  - TIME
    a. A *decade* has passed since Mary visited Madrid.
    b. John spent the *afternoon* reading.
  - PROPERTY
    a. He seemed *friendly*.
    b. Consider yourself *lucky*!

Linguistic motivation of semantic types

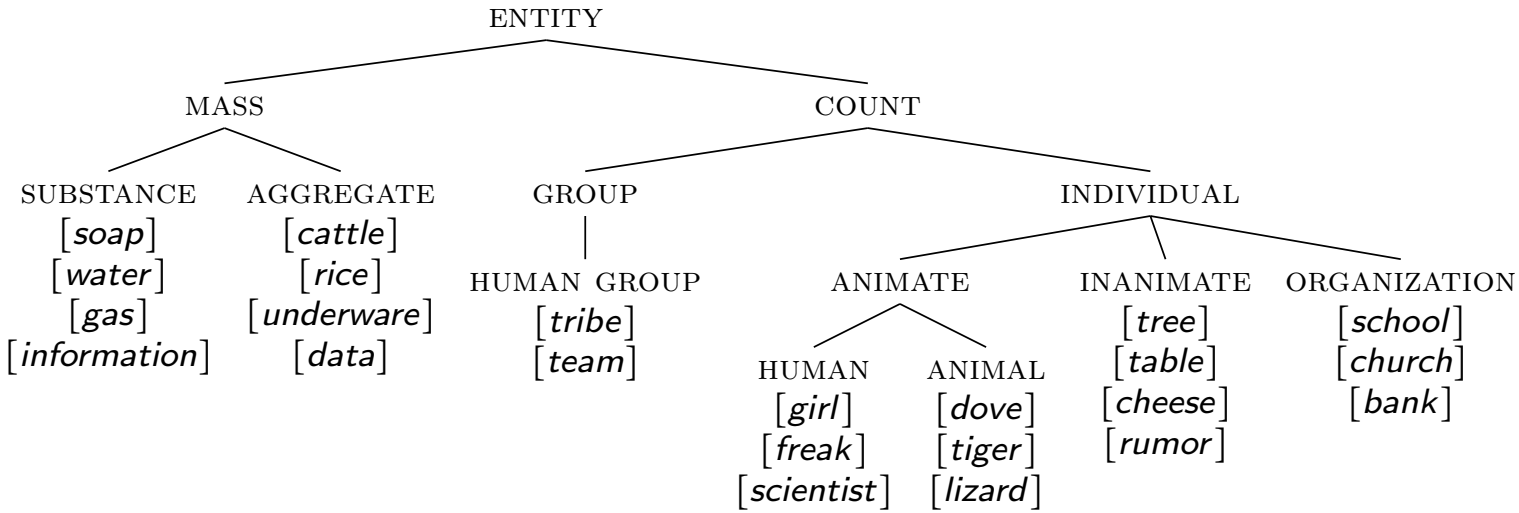
- Can we go any further? Finer-grained classification of PROPERTY
  a. He seemed {*friendly/ intelligent/ respectful*}. IL
  b. He seemed {*tired/ hungry/ depressed*}. SL
  c. I consider him {*friendly/ intelligent/ respectful*}. IL
  d. *I consider him {*tired/ hungry/ depressed*}. SL
- Alternative classification based on inferential properties of the Adj. (Kamp 1975, Kamp and Partee 1995, Amoia and Gardent 2006):
  - INTERSECTIVE: *carnivorous*:
    $||AN|| = ||A|| \cap ||N||$
    a *carnivorous tiger* is both a carnivore and a tiger;
  - SUBSECTIVE: *big*:
    $||AN|| \subseteq ||N||$
    a *big mouse* is big relative to the set of mice, but not absolutely so;
  - INTENSIONAL: privatives (*fake*, *pretend*) and non-subsectives (*alleged*):
    $||AN|| \cap ||N|| = \varnothing$
    a *fake gun* is not a real gun;
    an *alleged* thief is not necessarily a thief, but only suspected as one.

Type list

Semantic type	Example
ENTITY	thing
–MASS	soap, luggage, cattle
–COUNT	cow, table, tribe, bank
EVENT	happening, situation
–STATE	happiness, depression, love, be sick, be German
–DYNAMIC EVENT	demonstration, arrival, learn, build, jump
PROPERTY	
–INDIVIDUAL-LEVEL	tall, intelligent, respectful
–STAGE-LEVEL	hungry, tired, bored
PROPOSITION	(He told me) that you left
–INFORMATION	data/datum, commentary, rumor, message, summary, handout
TIME	tonight, soon, after dark, the day we met
LOCATION	upstairs, world-wide, here, downtown, in the yard
DIRECTION	towards, via, down
QUANTITY	seven, (a) few, (a) little, numerous, great deal, severely
MANNER	fast, happily, cruelly, with joy

ENTITY subtypes

- Only simple inheritance



Qualia-based types: natural and artifactual

- Semantic types can be complex rather than atomic.
  a. {faulty/good/efficient} {umbrella/computer/bicycle}
  b. ??{faulty/good/efficient} {pigeon/tree/oxygen}
  c. John started a {cake/journal/garden}. ORIGIN
  b. John began the {cake/book/trail}. ORIGIN/FUNCTION
  c. ??John {started/began} {a pigeon/a tree/the oxygen}.

a.
$$\begin{bmatrix} \textbf{cake} \\ \text{QS} = \begin{bmatrix} \text{FORMAL} = \textbf{phys\_obj(x)} \\ \text{CONSTITUTIVE} = \textbf{ingredient\_of(v,x), part\_of(w,x),} \\ \text{TELIC} = \textbf{eat(y,x)} \\ \text{AGENTIVE} = \textbf{bake(z,x)} \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textbf{tree} \\ \text{QS} = \begin{bmatrix} \text{FORMAL} = \textbf{phys\_obj(x)} \\ \text{CONSTITUTIVE} = \textbf{trunk\_of(y,x), foliage\_of(z,x)} \\ \text{TELIC} = \textbf{unspecified} \\ \text{AGENTIVE} = \textbf{unspecified} \end{bmatrix} \end{bmatrix}$$

Qualia-based types: natural and artifactual

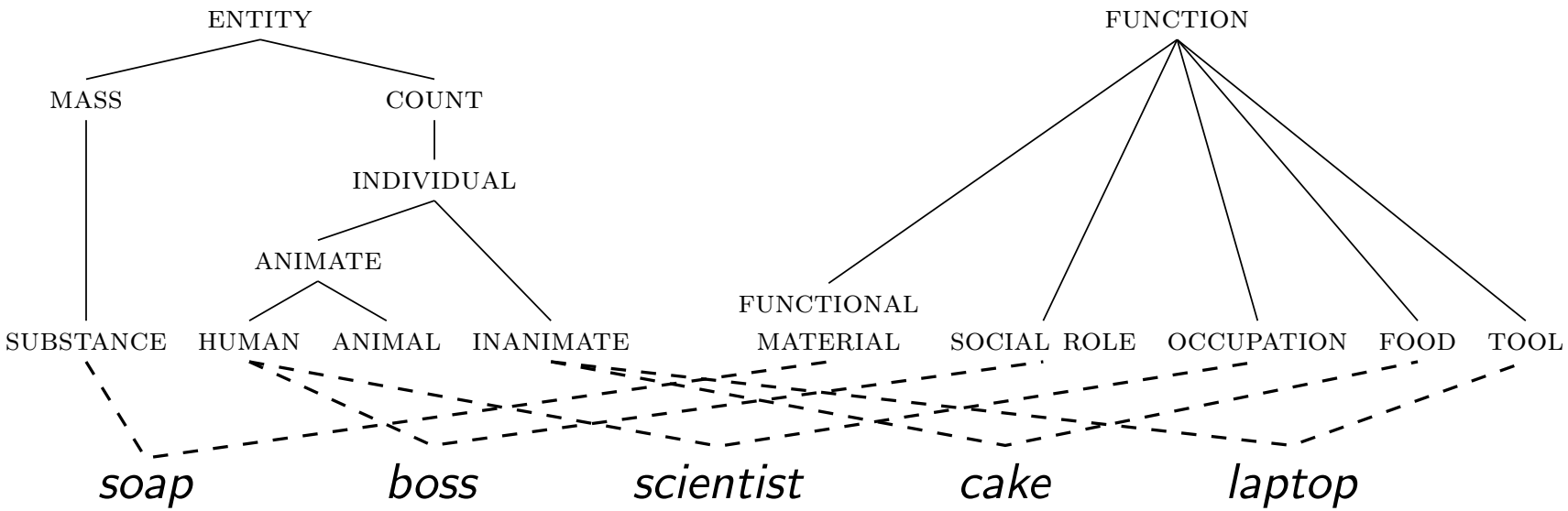
- FORMAL is the head type: an entity can have a function or not but, if it exists, it must 'be something' (physical object, substance, etc.).
- When the property encoded in the FORMAL role persists, other properties may not persist, but if it does not persist no other property persists because the entity no longer exists.
  - *broken camera*: PHYSICAL OBJECT with no function
  - *former boss*: HUMAN INDIVIDUAL with no function
- Type constructor tensor ($\otimes$): introduces types containing other qualia on top of the formal:
  a. *cake*: phys_obj $\otimes_C$ $\{flour, sugar\}$ $\otimes_A$ bake $\otimes_T$ eat
  b. *tree*: phys_obj $\otimes_C$ trunk $\otimes_C$ foliage
  c. *beer*: liquid $\otimes_C$ $\{water, yeast\}$ $\otimes_A$ brew $\otimes_T$ drink

Qualia-based types: natural and artifactual

- Natural/functional adjectives:
  - Natural: *red*, *heavy*, *flat*
  - Functional: *broken*, *faulty*, *efficient*, *useful*, *good*
  - *Easy*: both telic and agentive (*easy interface*, *easy meal*)
- Natural/functional Vs: Vs selecting for natural/artifactual arguments:
  - Natural: *fall* (PHYSICAL ENTITY subject), *die* (ANIMATE subject), *give* (ANIMATE subject)
  - Functional: *spoil*, *fix*

Qualia-based types: natural and artifactual

- Multiple inheritance in artifactual types



Qualia-based types: complex (dot objects)

- Complex types (dot objects) have more than one semantic type specified in their formal role.
- **dot** (•) is the type constructor that creates a complex type $a \bullet b$ from any two types a and b.

a. *novel*: PHYSICAL_OBJECT • INFORMATION

$$\left[\begin{array}{l}\textbf{novel} \\ \text{QS} = \left[\begin{array}{ll}\text{F} = & \textbf{phys\_obj} \\ \text{C} = & \textbf{page, cover...} \\ \text{T} = & \textbf{containing\_information} \\ \text{A} = & \textbf{printing, binding...}\end{array}\right]\end{array}\right] \bullet \left[\begin{array}{l}\textbf{novel} \\ \text{QS} = \left[\begin{array}{ll}\text{F} = & \textbf{information} \\ \text{C} = & \textbf{narrative} \\ \text{T} = & \textbf{entertain} \\ \text{A} = & \textbf{writing}\end{array}\right]\end{array}\right]$$

- For a dot object to be well-formed, there must exist a relation between its contituent types:
  a. *book*: [F = contain(PHYS_OBJ, INFORMATION)]
  b. *Toyota*: [F = produce(PRODUCER, PRODUCT)]
  c. *chicken*: [F = used_as(ANIMAL, FOOD)]
  d. *exam*: [F = ask(EVENT, HUMAN, QUESTION)]

Qualia-based types: complex (dot objects)

- How can we distinguish between different kinds of words involving multiple semantic types?
- Recall *beer*:
  - liquid $\otimes_C$ {*water*, *yeast*} $\otimes_A$ brew $\otimes_T$ drink
  - I have a little time for a quick beer → Is it FOOD•EVENT?
- Dots allow copredication:
  a. The lunch was delicious$_{[Food]}$ and very quick$_{[Event]}$.
  b. ??The beer was delicious$_{[Food]}$ and very quick$_{[Event]}$.
  Beer if FOOD, and it can only be quick only if we relate it to the drinking EVENT.
- Dots are complex lexicalized relational states between two or more participants. They may be built upon all kinds of relations between types, but only conventionalized and productive relations are prominent enough to yield a word with a typical dot behavior.
- → The EVENT meaning ('drinking') is encoded as one of the sublexical features of *beer*, but it is not a part of its head semantic type and hence can only be accessed through coercion.

Qualia-based types: complex (dot objects)

- Verbal dots: those that select for dotted arguments:
  - *Read*/ *write*/ *peruse*/ *scan* + PHYSICAL OBJECT•INFORMATION DO:
    a. read a {book/ article/ report/ chapter}
    b. write a {letter/ essay/ paper/ novel}
    c. peruse the {menu/ catalog/ cookbook/ diary}
    d. scan the {file/ resume/ tag/ photograph}
  - These Vs can coerce their OD into being [PHYSICAL OBJECT•INFORMATION]:
    a. John quickly read the wall.
    [PHYSICAL OBJECT] → [PHYSICAL OBJECT•INFORMATION]
    c. John quickly read the rumor.
    [INFORMATION] → [PHYSICAL OBJECT•INFORMATION]

Kinds of ambiguity

- Lexical ambiguity arises at the level of word meaning, when the word has more than one meaning.
  - Polysemy: a word has distinct but related meanings.
  - Homonymy: two words share the same form but have distinct unrelated meanings (*seal* 'marine mammal'/ 'wax closure').
- Syntactic (structural) ambiguity emerges when the same sequence of lexical items corresponds to two different syntactic structures.
  a. The studio recorded $[_{\text{DP}}$ the singer$]$ $[_{\text{PP}}$ with the microphone$]$.
  b. The studio recorded $[_{\text{DP}}$ the singer $[_{\text{PP}}$ with the microphone$]$ $]$.
- Semantic ambiguity: multiple interpretations are available regardless of the absence of structural or lexical ambiguity.
  a. Every passenger watched a movie.
  b. $\forall x[\text{passenger}(x) \rightarrow \exists y[\text{movie}(y) \wedge \text{watch}(x, y)]]$
  c. $\exists y[\text{movie}(y) \wedge \forall x[\text{passenger}(x) \rightarrow \text{watch}(x, y)]]$

Regular polysemy

- Regular polysemy: when the same kind of semantic relationship is present in more than one pair of senses
- → There are aspects of lexical meaning that are systematic and the processes involved in its construction are very productive
- → Given a regular polysemy pattern, we can predict what senses can be derived from any specific lexical meaning
- Container/Content alternation:
  - **Container**: This is my grandmother's china *bowl*.
    Content: I'd like another *bowl* (of soup).
  - **Container**: We bought two oak *barrels*.
    Content: The entire *barrel* spoiled during the trip.
  - **Container**: A need another paper *bag*.
    Content: The kids ate the whole *bag* (of grapes).
- Two main kinds of regular polysemy:
  - *Inherent polysemy* arises at the level of lexical-semantic structure of words
  - *Selectional polysemy* arises due to contextual influences

Inherent polysemy

- Dot objects are inherently polysemous because of the simultaneous presence of both semantic types
- *book*: PHYSICAL OBJECT • INFORMATION
  a. John bought a book online. / This is a good book.
     PHYSICAL OBJECT and INFORMATION → unresolved ambiguity
  b. This is a heavy book. / Close your book right now!
     PHYSICAL OBJECT
  c. This is a boring book. / Mary disagrees with your book.
     INFORMATION

Selectional polysemy

- The word itself is not polysemous, but in context its basic meaning is further specified or modified.
  a. Mary poured the **coffee** into her cup.
  b. John enjoyed his **coffee**.
  c. The kids devoured the **cake**.
  d. John finished the **cake** just in time for the party.
- The new meanings in (b) and (c) emerge as a result of the meaning adjustment mechanisms (*coercion*) triggered in order to make the selectional requirements of the syntactic predicate compatible with the inherent semantic type of its argument.

Conceptual structures

- A part of lexical meaning cannot be straightforwardly defined as linguistic. It has to do with a more primitive level of cognitive structure, comprising motion, vision, space, object manipulation, and other kinds of basic bodily and perceptual experiences.
- This conceptual content does not lend itself easily to formalization in terms of features, function-argument pairs, or propositional content.
- Most representations of conceptual structures are visual or analogical in nature.

Image schemas

- "A recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience" (Johnson 1987)
- Image schemas represent from simple static elements and spatial settings to complex dynamic relations involving force and abstract, non-spatial constructs (possession, information transfer, social interactions).
- CONTAINMENT-UNCONTAINMENT image schema
  - We experience our bodies as containers into which we put things (food, water) and also as contained objects (clothes, room, car).
  - We know what IN/OUT mean when we move from one location to another or manipulate objects and place them inside other objects.

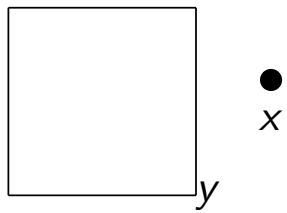


Figure: UNCONTAINMENT image schema

Properties of image schemas

- An image schema may affect how we view a particular situation once we have applied the image schema template to it.
  - CONTAINMENT schema imposes a binary categorization: the content is either inside or outside the container, never 'in the middle'.
    - E.g.: we can decide whether or not something is a fruit, but we cannot distinguish between kinds of fruit that are 'more inside' this category than others.
  - *transitivity*: if X is inside Y and Y is inside Z, than X is inside Z.
    - An *orange* is a *citrus*; a *citrus* is a *fruit* → an *orange* is a *fruit*.
- (UN)CONTAINMENT schema applied to different domains:
  a. The fish is not in the fish tank. [SPATIAL UNCONTAINMENT]
  b. The kids are at the concert. [SPATIAL+PARTICIPATION CONTAINMENT]
  c. The concert is not on Friday. [TEMPORAL UNCONTAINMENT]
  d. John is in denial about his son's needs. [ABSTRACT CONTAINMENT]

Combination of image schemas

- To account for dynamic events, the CONTAINMENT schema must be combined with the PATH schema, which encodes motion and other dynamic relations of the kind INTO/OUT OF
  - Source point
  - End point
  - Force vector relating both points

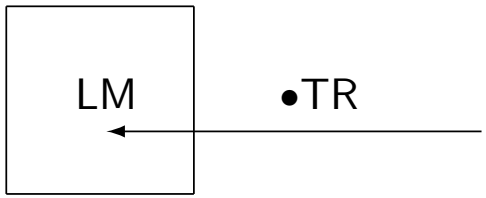


Figure: INTO image schema

a. John ran into the office.
b. The whole town population was evacuated.
c. I really need to blow off some steam.
d. They backed out of the deal.

Frames

- We understand many everyday situations by fitting them into predetermined structured representations of our past experiences and conventional social practices.
- **Semantic frame** (Fillmore): linguistic manifestation of these representations, in particular when the meaning of specific lexical items is defined relative to a background frame or scene.
- The choice of background frame impacts on how we describe a given situation
  - What did Robin Hood do, *steal* or *expropriate*?
- Frame elements in FrameNet:
  - Definition of the situation

    THEFT: situations in which a Perpetrator takes Goods from a Victim or a Source. The Means by which this is accomplished may also be expressed)
  - List of *frame elements* (semantic arguments), core and non-core
  - Relationship between different frames
  - List of lexical items that evoke the frame and share semantic and distributional properties

THEFT frame in FrameNet

CORE FRAME ELEMENTS:	
Goods	Anything (including labor, time, or legal rights) that can be taken away. E.g.: Leslie STOLE *the watch* from Kim.
Perpetrator	The person (or other agent) that takes the goods away. E.g.: *Leslie* STOLE the watch from Kim.
Source	The initial location of the goods, before they change location. E.g.: Leslie STOLE the watch *from the table*.
Victim	The person (or other sentient being or group) that owns the goods before they are taken away by the perpetrator. E.g.: Leslie STOLE the watch *from Kim*.
NON-CORE FRAME ELEMENTS:	
Explanation	The explanation for why a theft occurs. E.g.: Leslie STOLE the watch from Kim *because she'd always wanted a Gucci*.
Instrument	An object used by the Perpetrator in taking possession of the Goods. E.g.: Chis STOLE it *with a long bamboo pole* through the fence.
Manner	The Manner in which the theft occurs. E.g.: Leslie *swiftly* STOLE the watch from Kim.
Purpose	The Purpose for which a theft occurs. E.g.: Leslie STOLE the watch from Kim *in order to sell it on Ebay*.

Frame-frame relations: Inherits from COMMITTING_CRIME, TAKING; Used by ROBBERY
Lexical units: *abscond (with).v, abstract.v, abstraction.n, bag.v, cop.v, cutpurse.n, embezzle.v, embezzlement.n, embezzler.n, filch.v, flog.v, heist.n, larceny.n, lift.v, light-fingered.a*, etc.

Importance of non-core elements in frames

- Both *lift* and *snatch* instantiate the THEFT frame, but only the latter is associated with a specific manner of stealing (sudden and fast).
  a. Ed **snatched** the bag {quickly/ hastily/ *slowly/ *gently/ *carefully}.
  b. Ed **lifted** the bag {quickly/ hastily/ slowly/ gently/ carefully}.

Conventionalized attributes

- Conventionalized attributes: stable conceptual associations that may not be encoded in the lexical entry but which quite often affect the compositional interpretation of those words (Pustejovsky and Jezek 2008, Hanks 2013)
  a. cows **moo**/ dogs **bark**/ frogs **croak**
  b. las vacas **mugen**/ los perros **ladran**/ las ranas **croan**
  c. korovy **myčat**/ sobaki **gavkajut**/ ljaguški **kvakajut**
- How this information can be accessed and exploited in context:
  a. Mary {listened to/ heard} the violin (producing musical sounds).
  Qualia-based exploitation
  b. Mary heard the frogs (croaking).
  Conventionalized attribute
  c. Mary heard the wind (whistling).
  Conventionalized attribute

Amoia, M. and Gardent, C. (2006). Adjective based inference. In *Proceedings of the Workshop KRAQ'06 on Knowledge and Reasoning for Language Processing*, pages 20–27. Association for Computational Linguistics.

Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Mit Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Kamp, H. (1975). *Two theories about adjectives*. Cambridge University Press.

Kamp, H. and Partee, B. (1995). Prototype theory and compositionality. *Cognition*, **57**(2), 129–191.

Pustejovsky, J. and Jezek, E. (2008). Semantic coercion in language: Beyond distributional analysis. *Italian Journal of Linguistics*, **20**(1), 175–208.