

Explicit world-knowledge and distributional semantic representations

ESLLI 2017 Day 2: World-knowledge in the lexicon

Asad Sayeed and Alessandra Zarcone

University of Gothenburg, Saarland University

Lexical vs. world knowledge

- Different theories, roles they attribute to event knowledge and its position in cognitive architecture.
 - Raw nerve: distinction between lexical and world knowledge.
 - Depictions of knowledge:

Linguistic/lexical	World knowledge
Systematic	Situated
Compositional	Culture-dependent
Amenable to generalization	Eluding systematic characterization
Feasible object of analysis	Too chaotic for analysis

Lexical vs. world knowledge

Is this distinction on solid empirical ground?

- Jackendoff [2002] - motivation for such distinction may not be based on solid empirical ground but rather on...

... "lurking fear that general-purpose knowledge are a bottomless pit, and that in order to make the enterprise of semantics manageable it must somehow be restricted. And therefore some distinction must be made so we can stop before drowning in endless detail."

Lexical vs. world knowledge

Hobbs [2009]

“the most common argument in linguistics and related fields for drawing strict boundary between lexicon and world is a kind of despair that a scientific study of world knowledge is possible.”

Lexical vs. world knowledge

Making the argument more concrete:

- Common argument: lexical semantics is impossible because we would need a complete and correct theory of the world
- Counterargument: we don't need a scientifically correct theory of the world, we need to capture people's commonsense theories of the world
- Common argument: syntax is informationally encapsulated, thus amenable to scientific study
- Counterargument: we don't do syntactic analysis of utterances in isolation from world knowledge
and early integration of syntactic and semantic knowledge goes against the idea that language modules are informationally encapsulated

Lexical vs. world knowledge

Hobbs [2009] again

"If the brain makes a distinction between linguistic and world knowledge, it does not appear to be reflected in the temporal course of processing language"

Lexical vs. world knowledge

- Common argument: “a common despair that a scientific study of world knowledge is possible”
- Counterargument: it is possible to identify lexically relevant domains of world knowledge that are accessible to scientific study (e.g. the Generative Lexicon, work on Lexical Conceptual Structures)

**Let's look at a core form of world
knowledge...**

Part 1: generalized event knowledge

World knowledge

Is a systematic study of world knowledge possible?

- **Yes:** let's look at work on generalized event knowledge [McRae and Matsuki, 2009]:
 - Prototypical knowledge of events and their participants
 - Acquired from first- and second-hand experience, i.e., from language too, available in our memory
 - Activated by words in isolation, which cue concepts from typical scenarios (e.g. going to doctor, eating in restaurant).

World knowledge

Is a systematic study of world knowledge possible?

- **Yes:** let's look at work on generalized event knowledge [McRae and Matsuki, 2009]:
 - Words rapidly combine to generate expectations about upcoming input.
 - e.g., Donna used the hose to wash her filthy ... **car**/**hair**
 - **Thematic fit:** the typicality of a filler for a given semantic argument slot.
 - ... e.g., “car” should have a higher thematic fit than “hair” in the above example.

Possible to make predictions and verify hypotheses regarding world knowledge and its role in linguistic processing.

Generalized event knowledge

Two kinds of models:

- ① models that require an architectural distinction between lexicon and world knowledge
 - Have usually predicted that lexical knowledge is available immediately, whereas world knowledge is delayed. (e.g. Katz, 1972; Warren and McConnell, 2007; inter alia)
- ② models that do not have this distinction
 - predict early interaction of lexical and world knowledge to the point of indistinguishability (e.g. Federmeier and Kutas, 1999; McRae and Matsuki, 2009; inter alia)
 - single step models of language interpretation vs. parallel architecture with separation of domains
 - latter still compatible with experimental results that suggest early interaction.

General knowledge cued by the verb

Let's make it more concrete:

- People rapidly integrate various types of semantic and syntactic knowledge
- Verb meaning and situation structure: relations among the entities that commonly participate in the event
- A thematic role is a concept formed through everyday experiences (people learning who and what play specific roles in specific situations)
 - We learn about *accusing* and its agent role from our experiences with people who accuse others and linguistic descriptions of them
- Does reading or hearing a verb result in the immediate computation of information regarding typical agents, patients, instruments and locations?

We're going to do just a little bit of psycholinguistics now, so bear with



Experiments by Ferretti et al. [2001]

Experiment 1: agents and patients

- Priming study at short SOA (250ms) with semantic decision (animacy)
- Exp 1a: 28 related verb-agent pairs (*scrubbing-janitor*) + 28 unrelated (*cheering-janitor*)
- Exp 1b: 28 related verb-patient pairs (*adopting-baby*) + 28 unrelated (*investigating-baby*)
- Fillers have a good thematic fit only in the role they are chosen for (*janitor* is not a good patient for *scrubbing*)

Results: Shorter decision latencies for agent nouns following related verbs than for the same nouns following unrelated verbs

scrubbing → **janitor**

Experiments by Ferretti et al. [2001]

Experiment 2: instruments and locations

- Priming study at short SOA (250ms) with lexical decision
- 26 related verb-instrument pairs (*cut-saw*) + 26 unrelated (*dusted-saw*)
- 26 related verb-location pairs (*skated-arena*) + 26 unrelated (*prayed-arena*)

Results:

- Shorter decision latencies for instrument nouns following related verbs than for the same nouns following unrelated verbs

cut → **saw**

- No priming for location nouns following related verbs than for the same nouns following unrelated verbs

skated ↗ **arena**

Experiments by Ferretti et al. [2001]

Experiment 3: thematic roles as feature-based concepts

- Priming study at short SOA (250ms) with lexical decision
- 20 related verb - patient feature pairs (*tricking-gullible*) + 20 unrelated (*rescuing-gullible*)

Results:

- Shorter decision latencies for patient features following related verbs than for the same features following unrelated verbs

tricking → **gullible**

Experiments by Ferretti et al. [2001]

Experiment 4: interaction of situational knowledge and syntax

- Verbs cue both typical agents and patients
(*arresting* → *cop*, *arresting* → *crook*)
- Cross-modal priming with naming task
- Sentences with good agents and good patients (in congruent vs. incongruent role) and with unrelated verbs
 - She was arrested by the... cop / crook
 - She arrested the... crook / cop
 - She was kissed by the... cop / crook
 - She kissed the... crook / cop

Experiments by Ferretti et al. [2001]

Experiment 4: interaction of situational knowledge and syntax

Results:

- Shorter latencies for good agents and good patients in the congruent role than when presented in the same role with unrelated verbs)

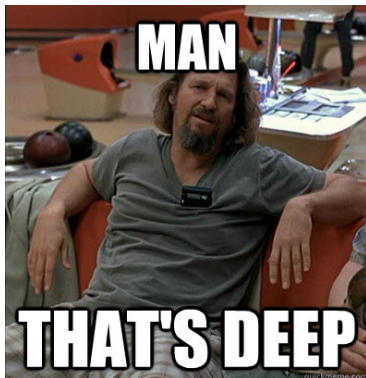
was arrested by the.. → cop

arrested the.. → crook

arrested the.. ↗ cop

was arrested by the.. ↗ crook

The point is that these event expectations have a deep psychological reality.



**So we need a semantics of
knowledge state change.**

Example: the concept of telicity

There is a well-established set of theories about verb “aspect” aka *Aktion-sart*.

A significant concept is “telicity”. Roughly:

- Telic predicate (+telic): action which has a definite end.
Example: break something e.g. glass.
- Atelic predicate (−telic): action for which an end is not necessarily defined.
Example: operate something e.g. machinery.

Semantically characterizing event knowledge

Consider the semantics of change [Pustejovsky, 2013]

Eat as activity (−telic) vs. *eat* as accomplishment (+telic)

- Sam ate ice cream (−telic)
- Sam ate an ice cream cone (+telic)
→ Extent scale (object extent of the theme)

Lengthen as activity (−telic) vs. *eat* as degree achievement (+telic)

- The icicle lengthened (over the course of a week) (−telic)
- The icicle lengthened two inches (+telic)
→ Extent scale (incremental theme verbs)

Gratuitous image of ice cream



Semantically characterizing event knowledge

Consider the semantics of change [Pustejovsky, 2013]

Ascend (directed motion) as $-telic$ vs. $+telic$

- The plane ascended (for 20 minutes) ($-telic$)
- The plane ascended to cruising altitude ($+telic$)
→ Path scale (directed motion verbs)

Proposal: all predication involves measuring an attribute against a scale

The semantics of change

Proposal: all predication involves measuring an attribute against a scale

- measurement is an assignment of a value, relative to an attribute (dimension attributes, physical property attributes, spatial attributes...)
- the nature of the theory interpreting the attributes depends on the constraints we impose (scale)
e.g. nominal scale for binary attributes, e.g. +/– animate

From event types to dynamic event structures

- states are defined as a single frame
- dynamic events are defined as subsequent, concatenated states:
 - a transition (propositional change) is a sequence of frames containing a propositional opposition and a program bringing about the change (mapping from states to states)
 - achievements: require a test to ensure that the change of state is not altered after it's achieved
 - accomplishments: internal process with change of state at the end

**Raises question of where we put
event knowledge: the lexicon?**

Part 2: debates on the contents of the lexicon

Lexical representation

What's in a lexical representation? People use rich information:

- syntactic, semantic, situational, pragmatic
- associated with words
- processed at each point in input stream to build expectations about future input.

Lexical representation

Lexical information includes:

- generalized event knowledge
- verb-specific thematic role filler preferences
 - e.g. the cop arrested vs. the criminal arrested [McRae et al., 1998]
 - What does you predict for each?
- Verb senses driven by thematic role-fillers
 - e.g. The K-mart shopper charged vs. the bull charged vs. the judge charged, etc [Ferretti et al., 2001]

**Question: how much of this should
belong in the lexicon?**

World knowledge and the lexicon

- world knowledge is accessible to non-linguistic cognitive processes (e.g., reasoning, planning, etc.).
 - think of metaphorical uses of cutting, instruments used to cut, etc.
- should we then distinguish between world-knowledge and lexical knowledge?

World knowledge and the lexicon

Three possible solutions [Elman, 2009; 2011]

- ➊ **Parallel architecture** [Jackendoff, 2002]
 - Only systematic info about words which can be generalized across word classes should be part of lexical representation.
 - non-lexical representation should interact fully/bidirectionally with language-specific modules.
- ➋ **Information-rich lexicon** [Langacker, 1987; van Berkum et al., 2005; inter alia]
 - All info exploited during processing should be included (event knowledge, subcat frames, thematic role info)
- ➌ **Empty lexicon** [Fodor and Lepore, 1998]: stripped of all info in (2)

World knowledge and the lexicon

Elman's position: info-rich architecture (2) seems reasonable, but what about duplication?

- Lexicon enriched with info traditionally considered to be world-knowledge
⇒ what about non-linguistic cognitive processes?
- Duplicate info outside lexicon?

Elman: "Can we take the world out of language and put language in the world?"

World knowledge and the lexicon

Elman's solution: lexical knowledge without a lexicon.

- takes information-rich lexical knowledge outside the lexicon
- claims that words are cues to this information rich knowledge
 - (which can also be accessed by other (non-linguistic) cognitive processes)
- unified account of information-rich lexical knowledge accessible both to linguistic and non-linguistic cognitive processes

Core abstract theories

Hobbs [2011]:

- Efforts to encode various aspects of world knowledge formally (commonsense physics and commonsense psychology) but were not designed with language in mind:
ontological distinctions in these models may not be very important in language
- Rather:
 - core theories formalizing linguistically-relevant dimensions (figure-ground, change of state, scale, causality)
 - defeasibility (what matters is how we choose the best interpretation in a context)

A more concrete approach

Pustejovsky's Generative Lexicon:

- focus on semantic composition: combining words and phrases in a meaningful way, recognizing meaningful combinations even if novel
- subtle meaning changes emerging from composition (*fast car, fast typist, fast waltz* or *summarize the book* vs. *dust the book*)
- a theory of lexical semantics should be compositional, generative, creative, constrained, systematic

Making room for productivity and regularity

Getting specific about the Generative Lexicon.

- Language production requires combining words and phrases in a meaningful way
- Language understanding requires recognizing meaningful combinations even if we have never heard them before (e.g. eating turtle soup)

Subtle meaning changes may emerge from composition:

- How do we know what it means to begin a novel, enjoy a burger or finish an essay?
 - begin a novel -> writing / reading
 - enjoy a burger -> eating
 - finish a essay -> writing / reading

Making room for productivity and regularity

- how do we know what it means for a car / a typist / a waltz to be fast?
 - fast car -> moves quickly
 - fast typist -> types fast
 - fast waltz -> has a fast tempo
- what's a book?
 - summarize a book: summarize a text (information)
 - dust a book: dust a physical object

Implicit information

Where do we find this implicit information about lexical items?

Pustejovsky [1991, 1995]: in the lexicon itself

A theory of lexical semantics should be

① compositional:

- should define how simpler semantic elements are combined to form more complex ones

② generative:

- should explain how a finite number of lexical items can be combined to generate an unbounded number of felicitous contexts
- should account for creative uses of language

Implicit information

Where do we find this implicit information about lexical items?

Pustejovsky [1991, 1995]: in the lexicon itself

A theory of lexical semantics should be

- 1 constrained:
 - should define the conditions for the composition operations that constrain how words and phrases combine
 - e.g., can we finish a stone? what's a fast curtain?
- 2 systematic:
 - should account for regular meaning changes and sense alternations.

The Generative Lexicon

A strongly typed semantic system

- lexical entries not as atomistic representations, but information-rich structures
- four levels of representations:
 - ① LEXICAL INHERITANCE STRUCTURE (specifying the relations between lexical structures in the type lattice)
 - ② ARGUMENT STRUCTURE (specifying the number and type of logical arguments and their syntactic realization)
 - ③ EVENT STRUCTURE (specifying the event type and subevental structure of the lexical item)
 - ④ QUALIA STRUCTURE (specifying our “understanding of an object or a relation in the world” (Pustejovsky 1995))

Part 3: qualia and habitats

Qualia and the Generative Lexicon

Qualia: four essential aspects of a word's meaning

- 1 CONSTITUTIVE QUALE
(its relation with its constituents - what a sandwich is made of)
- 2 FORMAL QUALE
(its relation with other objects in a larger domain - what's the place of sandwich in the ontology)
- 3 TELIC QUALE
(its purpose and function - sandwiches are for eating)
- 4 AGENTIVE QUALE
(the factors involved in its origin / creation - sandwiches are artifacts, they are made)

A Generative Lexicon item

$$\left[\begin{array}{l} \alpha \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x \\ \dots \end{array} \right] \\ \text{EVENTSTR} = \left[\begin{array}{l} \text{E1} = e_1 \\ \dots \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{CONST} = \mathbf{\text{what } x \text{ is made of}} \\ \text{FORMAL} = \mathbf{\text{what } x \text{ is}} \\ \text{TELIC} = \mathbf{\text{function of } x} \\ \text{AGENTIVE} = \mathbf{\text{how } x \text{ came into being}} \end{array} \right] \end{array} \right]$$

Qualia structure for “book”

$$\left[\begin{array}{l} \mathbf{book} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = \mathbf{x:information} \\ \text{ARG2} = \mathbf{y:phys_obj} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \mathbf{info \cdot physobj_lcp} \\ \text{FORMAL} = \mathbf{hold(y,x)} \\ \text{TELIC} = \mathbf{read(e,w,x,y)} \\ \text{AGENTIVE} = \mathbf{hold(e',v,x,y)} \end{array} \right] \end{array} \right]$$

From qualia to habitats

Contextualizing qualia:

what are the conditions for a table to be used, or for a chair to be used, for a sandwich to be nourishing?

→ habitats are defined by multidimensional affordances

Are we still in the lexicon?

What do we get from qualia?

We now know that:

- books contain information
- books have physical supports
- books are written
- books are for reading

Qualia structures are a formalization of a very reasonable assumption:

- the idea that lexical items referring to entities are associated with some sort of event knowledge
- this knowledge plays a role in processing and interpretation

What do we get from qualia?

Qualia have the purpose of

- achieving optimal explanatory adequacy within a combinatory and generative semantic system
- enriching lexical information while still keeping it concise and systematic
- maintaining a traditional distinction between linguistic knowledge and world knowledge, without restricting their ability to interact.
- explaining creative uses of language, while establishing its boundaries, constraining the range of sensical sentences

Issues with qualia

Are qualia structures too rigid and limited?

- “Mary enjoyed that rock” \Rightarrow ill-formed, because rocks are not artifacts and therefore lack a telic quale?
- But given sufficient supportive context, the oddness vanishes (“The climber enjoyed that rock”)
 - if you need context, it can't be coming from the lexicon [Fodor and Lepore 1998]
- - Pustejovsky: “exceptions” such as these possible.
 - process still part of lexicon: “enjoy” imposes new telic quale on rock)
- May still need some pragmatic inference or broader context

Can we overcome the rigidity of qualia structure approaches?



Habitats: an expansion of qualia

More recently, Pustejovsky (2012, 2013) expanded the idea of qualia into habitats:

- frames capturing salient aspects of a situation and its affordances
- habitats depict generalizations:
 - about a situation which arise from world knowledge
 - on which compositional process can operate
- multidimensional affordances determine how habitats are deployed and how they modify or augment the context

Habitats: an expansion of qualia

For example:

- in order to use a table, at minimum:
 - the top has to be oriented upward
 - the surface must be accessible
- however, consider the involvement of a chair:
 - a chair must also be oriented up,
 - the seat must be free and accessible
 - it must be able to support the user. . .

Modeling linguistic expressions as cognitive simulations

- (1) a. A car entered the driveway.
b. A woman stepped out.

Habitat composition of these two events:

- event simulations are constructed
- includes a bridging event, statable as a precondition on the second event
- (the defeasible presupposition that the car was not moving when the woman stepped out of it)

Habitats: in the lexicon?

The domain of habitats is closer to perceptual and motor capacities.
Does it belong in the lexicon?

- Cognitive psychology: probably not? it is relevant to other cognitive processes besides language
- Pustejovsky: it helps combinatorics and compositionality so it is “lexical”

What a linguist or a computational linguist calls lexical knowledge may not overlap with what the psycholinguist or the cognitive scientist calls lexical knowledge

Where to go from here?

- Generative Lexicon is one example of a way of integrating traditionally “lexical” and traditionally “non-lexical” data in a productive lexicon.
- But no matter what way you choose, you still need...data.

From somewhere. Which we'll talk about tomorrow. But...

- To be honest: we're far from being able to characterize Generative Lexicon-style knowledge distributionally.
- But we can chew off a piece of it.

**See you tomorrow or maybe just
now if we overspilled into this
lecture, heh.**