

## A Lexical Semanticist's Apology<sup>1</sup>

### 1 Lexical Semantics

- Words carve up the world around us into (sometimes) very fine-grained distinctions. Lexical semantics is the study of word meaning, focused on developing a theory of how a lexicon is organized and how word meaning connects to other parts of language and cognition:
  - What is a word meaning? Is it something atomic, or something malleable?
  - If the latter, what substantive word meaning matters in some way? And how do they fit together into larger meanings?
  - Are there limits on word meaning? Possible vs. impossible words?
  - How do word meanings relate to grammar, what do they tell us about how we perceive the world, and how do they happen or are they shaped by deeper cognitive principles?
- A typical methodology for lexical semantic research can be summarized as follows:
  - Identify a semantically coherent set of words
  - Identify some correlating property that you think is interesting or relevant:
    - \* Grammatical context they can (not) appear in
    - \* Interpretation they may (not) have and inferences they (do not) give rise to
    - \* Distributional characteristics across a discourse or larger corpus
    - \* Corresponding non-linguistic behavior
  - Identify a property of word meaning implicated in the behavior, deduce relevant principles linking the two, and integrate







- (14) a. If the verb has an agent argument, it is the subject, else if the verb has an instrument argument, it is the subject, else if the verb has a patient/place argument, it is the subject (i.e. agent-instrument-patient/place for subjecthood)  
 b. If the verb has a patient/place argument that is not the subject, it is the subject.  
 c. Anything not the subject is a prepositional phrase (PP)

- The lexicon can be viewed as a nonlinear rule for linking meaning to syntax (e.g. converting a verb to a function), **or** as a fact about how human lexicalize meaning (as per Dowty 1991). Either way, only denotational types (15a) are possible, not (15b), ruling out the unattested form:

- (15) a.  $[[break]] = [\lambda r a'(\_)]$  (y=agent, x=patient)  
 b.  $*[[break]] = [\lambda r a'(\_)]$  (y=agent, x=patient)

- Now, a key fact of this approach is that participants are **ranked** for subjecthood. That ranking then defines a kind of lexical semantic hierarchy structure that grammar care about.

∴ Word meanings are decomposed into hierarchical semantic pieces, which are structured relative to each other. Rule mapping meaning to grammar are sensitive to

### 3 From Two Verbs to Whole Verb Classes: Regularity and Idiosyncrasy in Verb Meaning

- For every verb that behave the way you'll probably find several that behave just like it.

- (19) a. John hattered/cracked/bent the metal lid (with the hammer).  
 b. The hammer hattered/cracked/bent the metal lid.  
 c. The metal lid hattered/cracked/bent.

- (20) a. John truck/bumped/knocked the windshield (with the hammer).  
 b. The hammer truck/bumped/knocked the windshield.  
 c. \*The windshield truck/bumped/knocked.

- Thus we have really identified two verb **classes**, **change-of-state** and **surface contact**, distinguished by thematic role (the grammatically relevant component of lexical meaning). (Just FYI, but Levin 1993 is an excellent resource on grammatically significant verb classes.)
- But what distinguishes within a class? They are not interchangeable, there must be some meaning that distinguishes them (Dowty 1979, Levin and Rappaport Hovav 1995, 1998, Peacock 1995, Grimshaw 2005, Beaver and Kentz-Garboden 2020, *inter alia*).

- |  |             |
|--|-------------|
| (21) <b>Template</b> (e.g. a thematic role list) | <b>Root</b> |
| a. causer patient                                | BREAK       |
| b. causer patient                                | CRACK       |
| c. causer patient                                | SHATTER     |

- The root affect interpretation and rule out unacceptable sentence even in cases where the grammar is otherwise entirely "correct" according to the template.

- (22) a. #Kim broke the cup. (structural integrity of patient)  
 b. #Kim hit the mirage. (rigid physical extent of place)

- Fillmore has a cute trick for determining this — based on the hearing of the wing, what is the thing doing you think a **twarge** is?

- (23) a. Maxine broke the twarge.  
 b. Maxine folded the twarge.  
 c. Maxine hattered the twarge.

- (24) a. Maxine hit the twarge.  
 b. Maxine lapped the twarge.  
 c. Maxine knocked the twarge.

- So there's two "halves" of a word's meaning. There's some sense in the field that the different semantic distinctions are very lexicographic, and not likely to matter grammatically (as suggested by Fillmore 1970: 129 and Dowty 1979: 32).

- Well, Beaver and Kentz-Garboden (2020) and Beaver et al. (2021) begin to differ! This question will be the topic of my third talk. But let me say a few words here that point to that.

#1 R t them elve clu ter int ubtype (Rappap rt H vav and Levin 1998):

- Rappaport Hovav and Levin (2010) suggested that roots can only have a limited set of meanings, describing a type of (result) state or a type of manner, but never both at the same time:

- (31) a. **Result roots:** break, crack, shatter, burn  
 b. **Manner roots:** run, jig, wim, laugh

- But there is debate about this — Goldberg (2010), Beaver and Kentz-Garboden (2012, 2017, 2020) and Riman (2016) have suggested that there are no limits on how much meaning a root can have (see also Grimshaw 2005 on root meaning complexity).

Word meanings are broken into (a) regular meanings that define word classes and (b) idiosyncratic meanings that distinguish words within a class, providing two axes of variation.

- FYI: What the right analysis is of template (1), and how other factors like truth conditions matter, is the topic of next week's lecture (Beaver 2006, 2010, Beaver and Francez 2012).

#### 4 The Scope of Variation in Verbal Behavior in One Language

- The variation in argument realization above has not been too bad. But it gets gnarlier. Levin and Rappaport Hovav (2005) distinguish two types of variable argument structure.
- What we have seen so far are **argument alternations** where one or more arguments that seem to be part of the verb's lexical content have multiple realization options, with a semantic that's still pretty similar regardless of how that argument is realized:

(32) **The Causative Alternation**

- a. The vase broke.  
 b. Kim broke the vase.

**The Locative Alternation**

- e. Kim painted the wall.  
 f. Kim painted the wall with paint.

**The Dative Alternation**

- i. Kim lent a letter to Sandy.  
 j. Kim lent Sandy a letter.

**The Conative Alternation**

- m. Kim kicked Sandy.  
 n. Kim kicked at Sandy.

**The Object Drop Alternation**

- c. Kim ate a pie.  
 d. Kim ate.

**The Clear Alternation**

- g. Kim cleared the dishes from the table.  
 h. Kim cleared the table of dishes.

**The Benefactive Alternation**

- k. Kim baked a cake for Mary.  
 l. Kim baked Mary a cake.

**The Preposition Drop Alternation**

- . The man climbed the stairs.  
 p. The man climbed up the stairs.

- However, while verbs with similar meanings do have similar alternations, there can be **a lot** of subclauses, and even then there is idiosyncratic variation (see above).
- There are also what Levin and Rappaport Hovav (2005) described as **event composition**, where a verb seems to show up in argument frames that involve participants in the way that be part of their lexical meaning (data drawn from Rappaport Hovav and Levin 1998).

- (33) a. Terry wept. (pure action)  
 b. Terry wept the floor. (action in place)  
 c. Terry wept the crumbs into the corner. (motion toward)  
 d. Terry wept the leaves off the sidewalk. (motion away)  
 e. Terry wept the floor clean. (change-of-state)  
 f. Terry wept the leaves into a pile. (creation)



- (34) a. Kim whistled. (pure undemission)  
 b. Kim whistled at the dog. (directed undemission)  
 c. Kim whistled a tune. (complete performance)  
 d. Kim whistled a warning. (message delivered)  
 e. Kim whistled me a warning. (message delivered to individual)  
 f. Kim whistled her appreciation. (message delivered)  
 g. Kim whistled to the dog to come. (message delivered to individual)  
 h. The bullet whistled through the air. (undplumtion)  
 i. The air whistled with bullets. (undplumtion)
- (35) a. Pat ran. (pure motion)  
 b. Pat ran to the beach. (directed motion)  
 c. Pat ran her self ragged. (change of state of something)  
 d. Pat ran her herself. (change of state of something else)  
 e. Pat ran clear of the falling rock. (change of location and state)  
 f. The coach ran the athlete around the track. (caused motion)  
 (Rappaport Hovav and Levin 1998: 97-98, (1)-(3))

- Can we possibly capture this by positing one verb and a simple set of linking rules?
- **Prima facie**, it seems many verbs can occur in a wide range of argument realization frames with different associated meanings – give them a truly unified analysis.
- At this point one might begin to consider an alternative: could it be that the syntactic constructions here are what are meaningful, and the verb(s) are being mapped to each frame (Goldberg 1995, Fli and Ramchand 2005, Baker 2005, 2013, Harley 2012)?
- This is a significant debate that I won't delve into here, although there are still patterns that delimit the empirical facts that are part of the debate.

#1 The patterns above are representative once again of whole classes of surface contact, undemission, and manner of motion verbs, suggesting again lexical classification.

#2 There is also micro-variation within classes, i.e. **cross-cutting classifications**. Even within undemission verbs there's variation in which allow a causative/inchoative alternation:

- (36) a. The tea kettle whistled  
 b. \*The boiling water whistled the tea kettle.

- (37) a. The teacup clattered.  
 b. I clattered the teacup as I loaded the dishwasher.

(Levin and Rappaport Hovav 2005: 10, (2), (3))

- Surface contact verbs differ in which result expressions they allow:

- (38) a. The sheriff hit/battered the outlaw to death.  
 b. The sheriff hit/\*battered the outlaw dead. (based on Beaver 2008)

#3 Event composition frame doesn't seem applicable to change-of-state verb :

- (39) a. \*Kelly broke.  
b. Kelly broke the table.  
c. \*Cinderella broke her finger to the bone.  
d. \*The clumsy child broke the beauty out of the vase.  
e. \*Kelly broke the dishes off the table.  
f. \*Kelly broke the dishes into a pile.
- (pure action)  
(change of state)

- Across language the event verb type cluster are distinct in their way. For example, *hit* but not *break* verb often all with their object to be realized as a bipartite phrase (i.e. not a canonical subject or direct object), either categorically or at least partially:

(41) a. *M raudi L \*(kau) bau-t-ida.*

M SUBJ L at hit-TA-3S

‘M hit L.’

b. *Aaka bakaka ulni-ki panka (\*kau) bah-t-ida*

this child writing-1S tick at break-TA-3S

‘This kid broke my pen.’ (Ulwa; Andrew Kentz-Garb den’ field notes)

(42) a.

- The intuition is that core transitive causes are reserved for when verb meet the prototypical type (roughly, again, caused change-of-state). Here's the list from (Hopper and Thompson 1980):

(45)		HIGH	LOW
A. PARTICIPANTS	2 or more participant (Agent and Object)	1 participant	
B. KINESIS	action	non-action	
C. ASPECT	telic	atelic	
D. PUNCTUALITY	punctual	non-punctual	
E. VOLITIONALITY	volitional	non-volitional	
F. AFFIRMATION	affirmative	negative	
G. MODE	realis	irrealis	
H. AGENCY	Agent in position	Agent in position	
I. AFFECTEDNESS OF OBJECT	totally affected	not affected	
J. INDIVIDUATION OF OBJECT	highly individuated	non-individuated	
	proper human, animate	common inanimate	
	concrete	abstract	
	singular	plural	
	count	mass	
	referential, definite	non-referential	

- These change-of-state verbs are the most prototypical form a transitivity perspective, everything else. This fits neatly with the event complexity and cross-linguistic variation data above, where *hit* verb (and others) have more grammatical variation than *break* verb.
- Yet language may differ in how much "deviation" the core framework tolerate! English is pretty liberal, using core transitivity for *break* and *hit* verb, but Tibetan clearly isn't this way.

Just a verb meaning vary in way that influence grammatical behavior, the core principle relating meaning to grammar vary a well, another point of variation.

## 6 Moving Beyond Verbs

- So far I've talked mostly about verbs since this is the category where most of the richest subcategories of words have been developed. But there are interesting generalizations in other categories as well. Here I give a quick discussion of adjective.
- Dixon (1982) argued for a classification of adjective that describe -called "property concept" (that can hold something for purely inherent reason, roughly) into 7 basic type:

(46)	<b>dimension</b>	big, small, long, tall, short, wide, deep, etc.
	<b>age</b>	new, young, old, etc.
	<b>value</b>	good, bad, lovely, attractive, perfect, proper, etc.
	<b>color</b>	black, white, red, etc.
	<b>physical</b>	hard, soft, heavy, wet, rough, strong, hot, cold, etc.
	<b>speed</b>	fast, quick, slow, etc.
	<b>human propensity</b>	jealous, happy, kind, clever, generous, cruel, proud, etc.

- He gives numerical diagrams motivating the 7 (Dixon 1982: 17, Table 1 summarize). One

- (47) value dimension physical property speed human property age color
- a.  $\lambda$  white truck,  $\lambda$  tall ladder,  $\lambda$  fat  $\lambda$  un
  - b. ?white  $\lambda$  truck, ?tall  $\lambda$  ladder, ?  $\lambda$  fat  $\lambda$  un

- This suggests again that there are lexical semantic generalizations in the meaning-syntax-grammar

- He then notes that there are two types of language :
  - (52) Large open class adjective (English, Dyirbal)
  - Small(er) closed class adjective (Hausa)
- There are typological generalizations based on these:
  - In / language , all types of property concepts are adjective .
  - In / language , dimension, age, value, and color will go in there.
  - In / language , physical property tends to be verb . **But** if there's a somewhat larger closed class adjective , they'll be adjective .
  - In / language , if physical property is a verb, then speed will be an adverb. If physical property is an adjective, speed will be, too .
  - In / language , human properties tend to be nouns . Even in English and Spanish this occurs (***hunger/hungry, tener hambre***). These are the last going into the adjective class .
- The differences have their manifestations . The relationship between a basic Dixonian state and the indicative that describe the c(a)2.3505(t)-243.95787468(f)-2647cltathhhe 957-0.921(h)-1.0951v

## 7 Conclusion and Notes on Lectures

- To sum up, what we have seen so far is the following general picture:
  - Word meanings are complex objects:
    - \* They are broken down into pieces.
    - \* We can identify these meanings by semantic tests and use in context.
    - \* Some meanings are shared across words (template) and some are not (root).
    - \* The pieces are structured relative to one another.
    - \* There is a logical order putting the pieces together into bigger meanings.
  - Word meanings figure into grammatical generalizations:
    - \* Linking principles describe the relationship.
    - \* Linking principles are sensitive to both the semantic and grammatical structure.
    - \* Not every part of word meanings figures into grammar.
  - Both word meanings and the correspondence to grammatical properties are subject to inter- and intra-linguistic variation, albeit with some systematicity.

The factors point toward the need for a theory of word meanings and its grammatical ramifications in order to have a full understanding of the human language faculty.

- My goal over the next two lectures is to build on the picture we touched upon and to reflect more on where I think the state of the art is and what the open questions are:

**Oct 13:** There are different theories for what a linguistically relevant lexical semantic unit is and its structure is. What are the motivations for these and why, and how much is reducible to the abstract of lexical semantics, especially truth conditional meaning?

- This will be next week's topic, where I'll look at me case studies that I have worked on that have both sought to ground out lexical semantic structure in something that is truly semantic, while also acknowledging that this approach has its limits.

**Oct 20:** The idea that word meanings consist of template and root is widely accepted, but the functional framework has been a template, with less attention paid to roots, which are sometimes even dismissed as uninteresting. Is this justified?

- This will be the topic of the third lecture, where I'll look at case studies showing that template and root meanings are hard to tease apart, forming a continuum. This work has a strikingly typological and cross-linguistic bent, thus showing the value of such work.

- While exploring the picture, there are some caveats I want to lay out right from the get-go.

#1 A noted absence, work in lexical semantics covers a very diverse range of literature from different theoretical traditions and frameworks, maybe more than other fields.

- However, despite the sometimes dramatic variation in the theoretical perspective, there is actually a very surprising amount of convergence on the big ideas.
  - Thematic relevance event structure is a common debate on what verb meanings are.
  - Template vs. root (regular vs. idiosyncratic meaning) is a common distinction.
  - The same tension comes up again and again (causation, change, motion, etc.).
  - The relative role of syntax vs. the lexicon is a common theme.

- Yet it's all my impression that the theoretical difference between the similarities, and blind people to big picture similarities, leading to redundancy and unnecessary debate.
- I'll be drawing on and often bridging differences among a range of approaches, because I







