

# 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



## Course Outline

- **June 24:** What is the Lexicon and how is it related to Syntax and Semantics?
- **June 27:** Lexicon in Syntactic Frameworks
- **July 1:** Lexicon in Semantic Frameworks
- **July 3:** Structure of a Lexical Entry 1: Argument Structure and Event Structure
- **July 8:** Structure of a Lexical Entry 2: Semantic Typing and Conceptual Structures
- **July 11:** General Architecture of the Lexicon: Syntactic and Semantic Type Hierarchies
- **July 15:** Compositionality 1: The Mapping from Lexicon to Syntax
- **July 18:** Compositionality 2: Coercion, Type Shifting, Constructions, Metaphor

## Lecture 1: June 24

### What is the Lexicon and how is it related to Syntax and Semantics?

- The mental lexicon
- What can we learn from the lexical entries in dictionaries?
  - Orthographic and phonological information
  - Semantics: the definition
  - Syntax: syntactic category
- What do the dictionaries do not account for?
- Syntactically relevant features of lexical items
  - Semantic features and semantic types
  - Countability
  - Interaction of lexical features

What can we learn from the lexical entries in dictionaries?

What do the dictionaries do not account for?

Syntactically relevant features of lexical items

Interaction of lexical features

References

## Mental lexicon

- *Mental lexicon*: the knowledge that speakers and hearers have about lexical expressions (up to 250,000 lexical entries)
- *Lexical competence*: a part of the linguistic competence that enables us to fluently recognize and use words.
- *Active vocabulary*: Can I use the word productively?
- *Passive vocabulary*: Do I understand the word?

## Mental lexicon as a system

- The mental lexicon is not a pile of words but rather a highly organized and complex system. Psycholinguistic evidence:
  - **Lexical decision task:** is this sequence of letters a real word?

Sequence of letters	Expected response
House	Yes (faster response)
Honesty	Yes (slower response)
Halibut	Yes (slower response)
Vouse	No

- It takes us  $< 0.5$  sec. to answer  $\rightarrow$  words cannot be randomly heaped up in our mind.
- Factors known to affect the speed of *lexical access*:
  - 1 *word frequency*: frequently used words are recognized more quickly
  - 2 *imageability* and *concreteness*: words lexicalizing concrete and easy to imagine concepts are recognized more quickly

## Mental lexicon as a system

### Psycholinguistic evidence (cont'd):

- **Priming:** “improvement in speed or accuracy to respond to a stimulus [*target*], when it is preceded by a semantically [morphologically, phonologically, orthographically, etc.] related stimulus [called *prime*] (e.g., *cat-dog*) relative to when it is preceded by a semantically unrelated stimulus (e.g., *table-dog*)” (McNamara 2004).

Prime	Target	Is the target a real word in English?
Cat	Dog	Yes (faster response)
Table	Dog	Yes (slower response)

Lexical knowledge needed to account for this processing difference:

- 1 Are these words related?
- 2 How are they related?

**What can we learn from the lexical entries in dictionaries?**

What do the dictionaries do not account for?

Syntactically relevant features of lexical items

Interaction of lexical features

References

# What can we learn from the lexical entries in dictionaries?

- Orthographic information: *house*, *arrive*, *easy*
- Phonological information: /'haʊs/, /ə'raɪv/, /'i:zi/

## Semantics: the 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.



## Syntax: syntactic category

**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.

- [<sub>DP</sub> the/three [<sub>NP</sub> old *houses*]]

## What do the dictionaries do not account for?

### Fine-grained selectional requirements:

- Why isn't any noun compatible with any adjective?
  - ??*easy house*
- Why isn't any adverb compatible with any verb?
  - ??*The train arrived easy.*
- Being an intransitive verb, what kind of arguments can *arrive* take?
  - \**John arrives the zoo.*
  - \**John arrives at the book.*

# What do the dictionaries do not account for?

## How are different word senses related?

### *Arrive:*

- 1 to reach a destination **SPATIAL**
  - The {guests/ police/ order} *has arrived*.
- 2 to come to be present, happen **TEMPORAL**
  - Ask yourselves if the time *has not arrived* for us to come to our senses.
- 3 to attain a position of success **ABSTRACT**
  - After years of climbing the corporate ladder, he felt he *had* finally *arrived*.

# What do the dictionaries do not account for?

## How are different lexical entries related?

### *Easy:*

- <sup>1</sup>*easy* adjective
  1. requiring no great labor or effort
- <sup>2</sup>*easy* adverb
  1. EASILY

### *House:*

- <sup>1</sup>*house* noun
  1. a building in which people leave
- <sup>2</sup>*house* verb
  1. to put or receive into a house

## Syntactically relevant features of lexical items

- **Syntax** combines lexical items in constructions of varying degrees of complexity according to a set of *well-formedness rules*.
- Syntactic rules are governed by two constraints:
  - *linearity*: word order (Adj < N, *old houses*)
  - *dominance* or *constituent structure*: relation of inclusion ([*DP* the [*NP* old houses]])
- Syntactic rules are *productive* (applied automatically whenever the constraints are met) and *recursive* (can generate infinite word chain).
- They **overgenerate** unless properly constrained by the features encoded in the lexicon.

## Semantic features and semantic types

Selectional requirements are not limited to the syntactic category.

- *arrive*
  - AS:  $\text{arg}_1[\text{cat}=\text{DP}], \text{arg}_2[\text{cat}=\text{PP}]$
  - \*John arrives at the book.
  - AS [rev.]:  $\text{arg}_1[\text{cat}=\text{DP}], \text{arg}_2[\text{cat}=\text{PP}, \text{sem type}=\text{LOCATION}]$
- *hate*
  - AS:  $\text{arg}_1[\text{cat}=\text{DP}], \text{arg}_2[\text{cat}=\text{DP}]$
  - \*The happiness hates John.
  - AS [rev.]:  $\text{arg}_1[\text{cat}=\text{DP}, \text{animacy}=+], \text{arg}_2[\text{cat}=\text{DP}]$
- *easy*
  - AS:  $\text{arg}_1[\text{cat}=\text{DP}]$
  - \**easy house* vs. *easy exam*/ *easy recipe*/ *easy prey*
  - AS [rev.]:  $\text{arg}_1[\text{cat}=\text{DP}, \text{sem type}=\text{EVENT}]$

**Semantic type:** the kind of entity denoted by the lexical item

**Semantic feature:** smaller (*sublexical*) component of word meaning

## Countability

Morphosyntactic classes of nouns with differentiated behavior:

- **Count Nouns** refer to **individuals**: discrete, bounded entities
  - occur in PL:
    - girl-girls, fan-fans, class-classes*
  - usually cannot occur in SG without a determiner:
    - \*Whale is big./ \*I saw {girl/ class}./ \*This is {whale/ girl/ class}.*
  - are quantifiable by cardinal numerals:
    - twenty girls, two classes of '90*
  - quantifiable by 'how many' quantifiers:
    - {A few/ many/ several} fans sneaked backstage.*

# Countability

- **Mass Nouns** refer to **unindividuated 'stuff'**
  - cannot occur in PL, or acquire a derived interpretation:
    - my money - public monies* ('pecuniary sums / budget')
    - beer - Czech beers* ('kinds of beer')
    - The bartender brought our beers* ('{glasses/ cans} of beer')
  - can occur in SG without a determiner:
    - {Natural gas/ gold} is a fossil./ I want water./ This is {gas/ gold}.*
  - are not quantifiable by numerals, or acquire a derived interpretation:
    - \*five moneys/ Two bears, please!/ These are our three top beers.*
  - are quantifiable by 'how much' quantifiers:
    - A little gas leaked from the pipe./ I do not drink much milk.*



## Countability

**Individuation:** process whereby we identify an entity and distinguish it from other entities.

Semantic features related to individuation (Jackendoff 1991):

- [ $\pm$ bounded]
  - [+bounded]: [-divisive, -cumulative]: *girl, fan*
    - 1 girl /  $x \neq 1$  girl
    - 1 girl + 1 girl  $\neq$  1 girl
  - [-bounded]: [+divisive, +cumulative]: *water, gas*
    - water /  $x =$  water
    - water + water = water
- [ $\pm$ internal structure]
  - [+internal structure]: entities composed of separate, distinguishable individuals (*crowd, class; girls, fans*)
  - [-internal structure]: *girl, fan, water, gas*

# Countability

## Individuation types:

- [+b, -i]: individual (*a girl, a fan, a table*)
- [+b, +i]: group (*a class, a committee*)
- [-b, -i]: substance (*gas, water*)
- [-b, +i]: aggregate (*rice, cattle, girls, tables*)

The count-mass distinction seems to be determined by the [ $\pm$ bounded] feature (at least in English):

- [+b] (individuals and groups)  $\rightarrow$  count
- [-b] (substances and aggregates)  $\rightarrow$  mass

## Countability

The syntactic impact of [ $\pm$ internal structure] is more limited but not inexistent.

- Dual agreement with group nouns:
  - *The Dutch team were/ was coached by Louis van Gaal.*
- Ambiguity of *large*:
  - groups: *large family/ community*: big(cardinality)
  - individuals and aggregates: *large apartment/ cattle*: big(size)
- Dedicated unitizers for groups and aggregates:
  - *a grain of rice, a member of the {family/ committee}, a head of {cabbage/ cattle}*

## Countability

Syntactic vs. conceptual basis of the count-mass distinction

- **Syntactic approach**: the count-mass features are not lexically encoded; any noun can be interpreted as mass or count depending on the syntactic structure
- Borer (2005): all Ns are mass by default and the count interpretation emerges due to the functional projections CIP and #P
  - a. *salt*: [<sub>DP</sub> [<sub>NP</sub> salt]]
  - b. *boys*: [<sub>DP</sub> [<sub>CIP</sub> boy [<sub>NP</sub> boy]]]
  - c. *three boys*: [<sub>DP</sub> [<sub>#P</sub> three [<sub>CIP</sub> boy [<sub>NP</sub> boy]]]]
  - d. *a boy*: [<sub>DP</sub> [<sub>#P</sub> a [<sub>CIP</sub> a [<sub>NP</sub> boy]]]]

## Countability

Arguments in favor of this approach:

- Classifier languages, where all Ns must be combined with classifiers in order to be used in SG or PL (Cheng and Sybesma 2005):
  - a. Hufei maishu    qu le  
Hufei buy-book go sentence-final-particle  
'Hufei went to buy a book/books'
  - b. san-ben            shu  
three-CL<sup>volume</sup> book  
'three books'
  - c. yi-wan            tang  
one-CL<sup>bowl</sup> soup  
'a/one bowl of soup'
- Mechanisms that override the count-mass distinction in context:
  - **Universal Grinder**: count → mass
    - There was {blueberry all over the table/ apple in the salad}.
  - **Universal Packager**: mass → count
    - Two {beers/ wines}, please!

## Countability

Counterarguments to the syntactic approach: we cannot really combine any word with any kind of structure

- Djalali *et al.* (2011): **very few N classes are acceptable in grinder constructions** (*There is N all over the {floor/ highway}, There is N in the {bread/ concrete}, {A robot/ a termite} eats N*)
  - Acceptable with: foodstuff (*steak, apple, cracker*), animals (*squirrel, snake, pig*)
  - Rejected by: group terms (*fleet, committee*), artifacts (*computer, toaster, hammer, pencil*), shape names (*tube, sphere, cone*), individual group members (*ship, person*)

# Countability

**Conceptual approach:** count-mass distinctions are determined by how speakers of a given language *conceptualize* the entities

- **Conceptualization** is related to the attributes of real-world entities but it is not determined exclusively by them
  - Count-mass doublets referring to the same kind of entity:  
*leaves-foliage, knives-cutlery, shoes-footwear*
- Factors affecting the count-mass encoding (Wierzbicka 1985):
  - 1 Perceived size: bigger/ easily identifiable → COUNT: [EN] *pea(s) - rice*; [RU] *gorox* 'mass-pea' - *bob(y)* 'beans'
  - 2 Contiguity: things occurring together in space/time → MASS: *hair, scissors; contents, dishes*
  - 3 Standard way of handling and function: heterogeneous collective categories defined by their function → MASS: *mail, furniture, luggage, jewelry*
  - 4 etc.

## Interaction of lexical features

Different lexical features are not independent. Rather, they form an intricate network where they interact with one another.

This presupposes that the lexicon is an interconnected system within the language faculty. → **Session 6**

- Animacy/ semantic role assignment/ predicate aspect
  - The police<sub>ANIMATE</sub> surrounded the park.  
**AGENT** *surround* LOCATION → ACHIEVEMENT
  - The fence<sub>INANIMATE</sub> surrounded the park.  
**THEME** *surround* LOCATION → STATE



## Interaction of lexical features

- Animacy/ countability: scale of “likelihood of participation in the speech event” (Smith-Stark 1974): ANIMATE → COUNT
  - animate < human < rational < kin < addressee < speaker
- Animacy/ countability: the more individuated an entity is, the less likely it is to be viewed as occurring collectively (Grimm 2012)
  - [liquid/substance] < granular aggregate < inanimate collective aggregate < lower animate < higher animate < human

- Borer, H. (2005). *In name only*, volume 1. Oxford University Press.
- Cheng, L. L.-S. and Sybesma, R. (2005). Classifiers in four varieties of chinese. *The Oxford handbook of comparative syntax*, pages 259–292.
- Djalali, A., Grimm, S., Clausen, D., and Levin, B. (2011). What can be ground? noun type, constructions, and the universal grinder. *Berkely Linguistics Society*.
- Grimm, S. (2012). *Number and individuation*. Ph.D. thesis, STANFORD UNIVERSITY.
- Jackendoff, R. (1991). Parts and boundaries. *Cognition*, **41**(1), 9–45.
- McNamara, T. P. (2004). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.
- Smith-Stark, T. C. (1974). The plurality split. In *Chicago Linguistic Society*, volume 10-1, pages 657–672.
- Wierzbicka, A. (1985). Oats and wheat: The fallacy of arbitrariness. *Iconicity in syntax*, pages 311–342.