Words

9.19: Computational Psycholinguistics
18 September 2023
Roger Levy

First, a brief speech perception encore

Incohearent is a hilarious party game that will have all the players laughing halfway through round one. The Judge will turn a card, showing an Incohearent phrase. Players will then read the card aloud and try and decipher what the phrase actually is. Will you be able to hear it before anyone else?

Words and word meanings: two views

1. Within most of philosophy and linguistics, semantics is *referential*. That is, linguistic meaning is analyzed as a relationship between words and the world, and sentence meaning describes a state of affairs that can be mapped to situations in the world...For example, a word like dog has a meaning that allows you to pick out all the dogs in the world.
2. In psychology, the predominant approach to word meaning is that it is a mapping of words onto the conceptual structure...That is, people have concepts that are the building blocks of their world knowledge, and the meaning of a word is essentially a pointer to some subpart of that knowledge. Concepts make the connection between language and the world argued for in philosophical and linguistic approaches to semantics, but in a psychologically plausible way.

(from Lake & Murphy, 2021)

A bit of background in logical semantics

- **Logical** approaches to semantics influential in linguistics, psychology & AI
  - Linguistic expressions have **semantic types**
    - **Nouns**: sets (equiv: functions from entities to truth values, or $e \rightarrow t$)
    - **NPs** (oversimplified): individuals *e*
    - **Intransitive verbs**: sets, or $e \rightarrow t$ functions
  - Syntactic rules have corresponding semantic rules too
    - e.g., set membership check for S→NP VP

Super-brief introduction to logical semantics

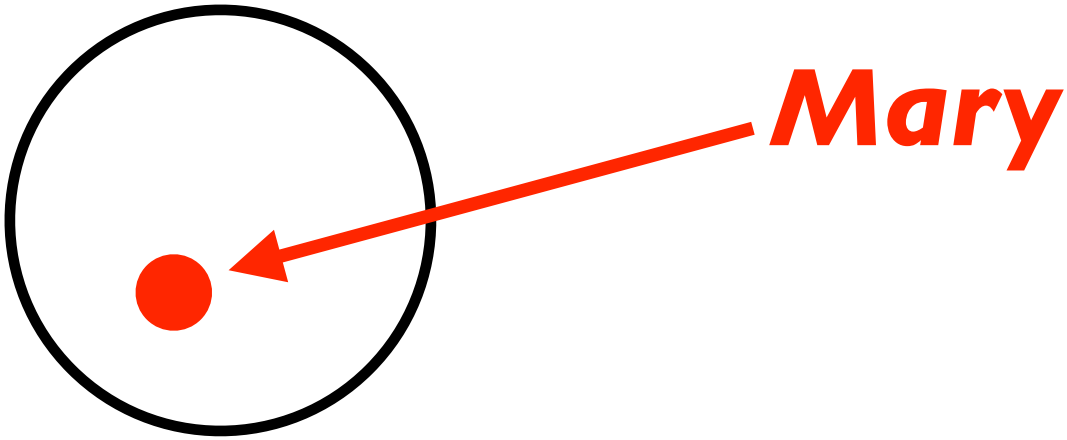
denotation function

$[\![\text{S}]\!] = [\![\text{VP}]\!]([\![\text{NP}]\!])$

NP VP

$[\![\text{Mary sings}]\!] = [\![\text{sings}]\!]([\![\text{Mary}]\!])$

set of entities that sing



Background: logical semantics

Warning!

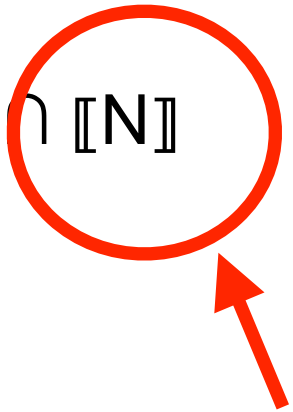
- The previous slide is a serious oversimplification of NP semantics & composition with VP into S, e.g.:
  - *Dogs growl* (dogs is a generic)
  - *Every student studied hard* (quantifiers every need different treatment)
- But, this example hopefully gives a sense of the context in which set-based semantics is put to use

Adjectives: a range of semantic types

- **Intersective**: *living*, *blue*
- **Scalar**:
  - **Relative**: *short*, *expensive*
  - **Absolute**: *dangerous*, *full*
- **Non-intersective**: *possible*, *alleged*
- **Anti-intersective**: *former*, *counterfeit*

Intersective adjectives

- Examples: living, blue
- The criterion imposed by the adjective is independent of the noun
- Formally, ⟦Adj N⟧ = ⟦Adj⟧ ∩ ⟦N⟧



⟦X⟧: "***meaning of X***"

Scalar adjectives

- Examples: *short*, *expensive*, *dangerous*, *full*
- Constrain referent's value on a **scale**
- **Relative adjs** (*short*, *expensive*): scale constraint based on comparison class
  - Highly context-sensitive; compare:

    Bill is a big man!

    Bill is a big mouse!

    Bill is a big elephant!

    Bill is a big basketball player!
- **Absolute adjs** (*dangerous*, *full*, *empty*): constraint is tied to scale boundary
  - Less context sensitive?

    The glass is empty.

    The gas tank is empty.

    The auditorium is empty.

Non-intersective adjectives

- Examples: *alleged*, *possible*
- Adj's meaning contribution depends fundamentally on modified noun
- Adj "releases" referent from constraints on the noun

Anti-intersective adjectives

- Examples: *counterfeit*, *former*
- Adjective's meaning contribution depends fundamentally on modified noun
- Adjective adds commitment that the referent is NOT in noun's denotation

How many words do you know?

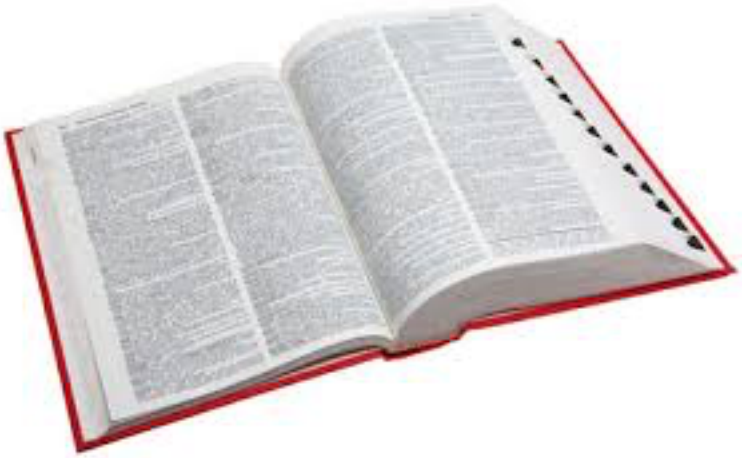
TABLE 3 | Various estimates of the number of English words known by adults (typically first-year university students), together with the way in which "words" were defined and the task used.

Study	Estimate	Definition of "word"	Task
Hartmann (1946)	215,000	All entries from Webster's New International Dictionary	Meaning production
Nusbaum et al. (1984)	14,400	Lemmas present both in Miriam-Webster's Pocket Dictionary and Webster's Seventh Collegiate Dictionary (list of 19,750 words)	Familiarity rating
Goulden et al. (1990)	17,200	Base words (sic) from Webster's Third New International Dictionary, excluding proper nouns, derived words, and compounds.	Indicate whether word is known or not
D'Anna et al. (1991)	17,000	Functionally important lemmas (sic) from the Oxford American Dictionary, with the exception of abbreviations, hyphenated words, affixes, contractions, interjections, letters, multiword entries, slang, capitalized entries, foreign words, alternate spellings, and outdated words.	Subjective estimates of knowledge
Anderson and Nagy (1993)	40,000	Distinct lemmas (sic) from a corpus based on school textbooks; excludes proper nouns and a limited number of very transparent derived words and compounds.	Various tests
Zechmeister et al. (1995)	12,000	Same as in D'Anna et al. (1991)	Multiple choice questions related to the meaning of the words
Milton and Treffers-Daller (2013)	9,800	Same as in Goulden et al. (1990)	Provide synonym or explanation for words known

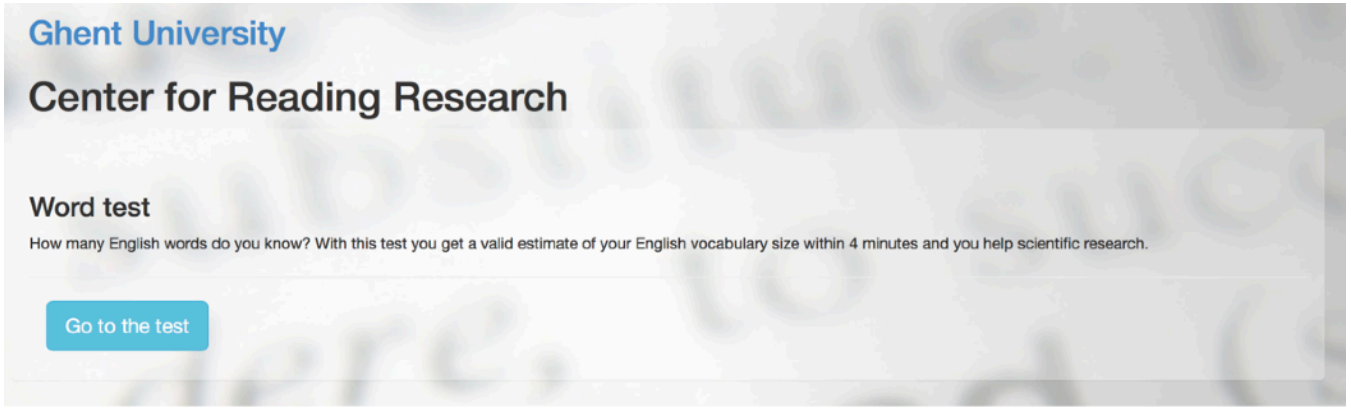
(Brysbaert et al., 2016, Frontiers in Psychology)

How many words do you know?

- The dictionary test…



- A modern, psycholinguistically informed variant (61,800 “worthwhile” lemmas):

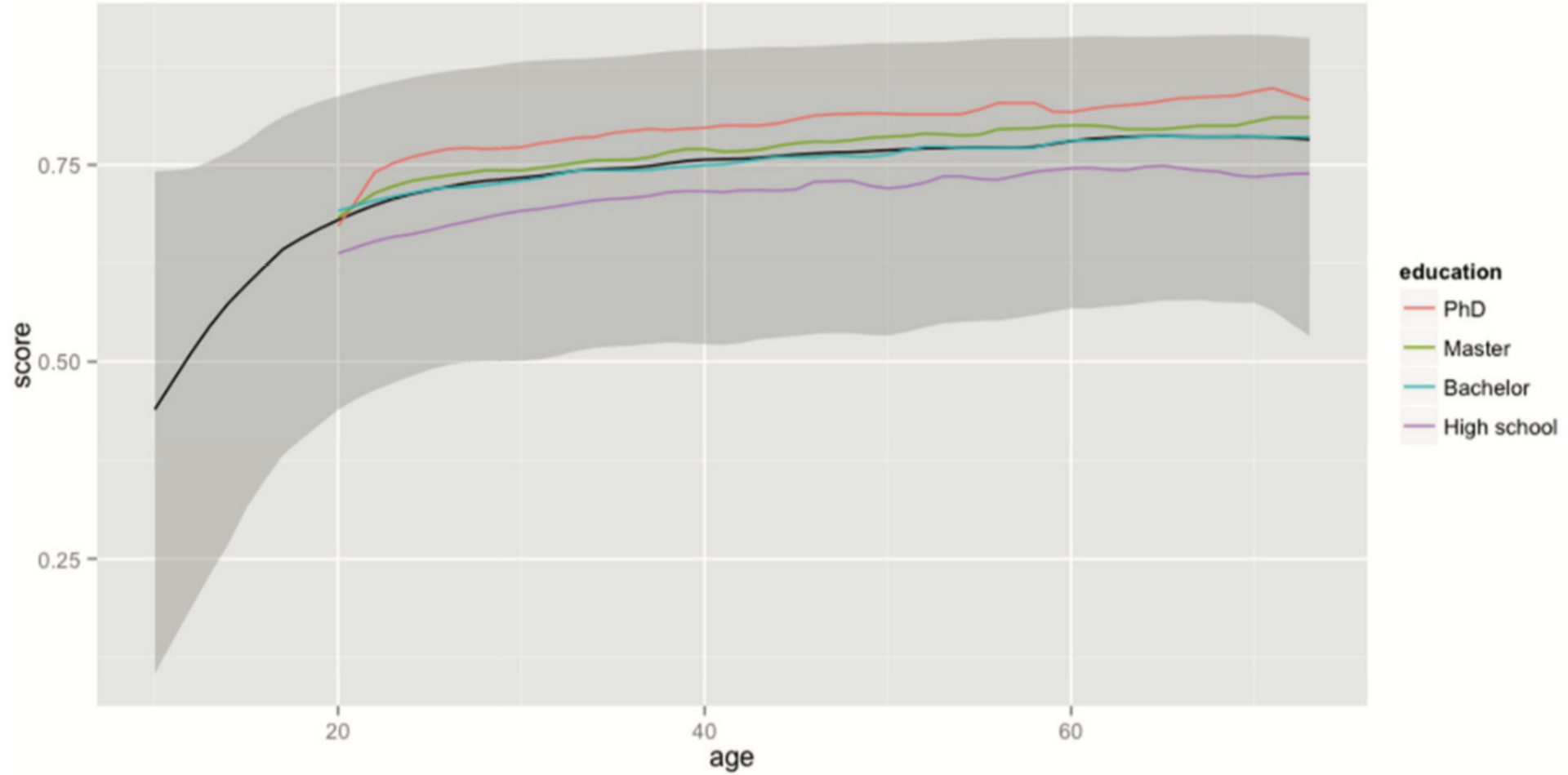


(Brysbaert et al., 2016, Frontiers in Psychology)

How many words do you know?

- Results by age & education level:

61,800



(Brysbaert et al., 2016, Frontiers in Psychology)

How do we learn so many words?

- The average 20-year-old native English speaker knows **42,000 lemmas**
- That is 5.75 lemmas per day, every day!
- The mystery:

The average seventh-grader…must have acquired most of them as a result of reading because (a) the majority of English words are used only in print, (b) she already knew well almost all the words she would have encountered in speech, and (c) she learned less than one word by direct instruction. Studies of children reading grade-school text find that about one word in every 20 paragraphs goes from wrong to right on a vocabulary test. The typical seventh grader would have read less than 50 paragraphs since yesterday, from which she should have learned less than three new words. Apparently, she mastered the meanings of [several] words that she did not encounter.

(Landauer & Dumais, 1997, Psychological Review)