

The Mental Lexicon

Eva Maria Luef

SS 2025

Project class

- Conceptualize psycholinguistic/ lexical project
 - Formulate hypothesis
 - Gather data (corpus, record yourself)
 - Analyze data
 - Write up project report
-
- **Class Moodle**

How many words does a person know?

- In the first language
- In second and other languages (L3, L4...)
- Question important for various disciplines related to language processing, development, acquisition, and education
- Vocabulary size estimates depend on how words are defined
 - For L1: range between less than 10.000 to over 200.000 words (Brysbaert et al., 2016)

Define “word”

What is a word?

- Oxford Dictionary
- “a single distinct meaningful element of speech or writing, used with others (or sometimes alone) to form a sentence and typically shown with a space on either side when written or printed”
- Wikipedia
- “despite the fact that language speakers often have an intuitive grasp of what a word is, there is no consensus among linguistics on its definition and numerous attempts to find specific criteria of the concept remain controversial”

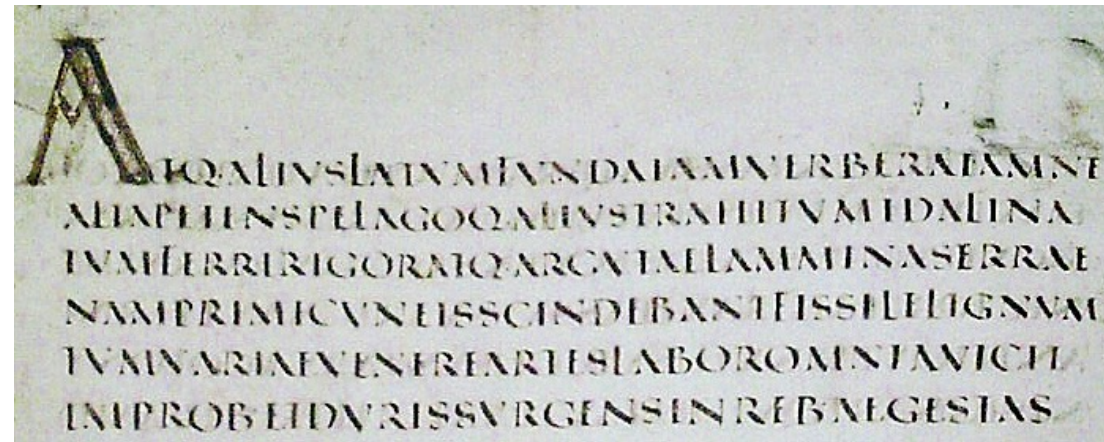
The morphosyntax of words

- General assumption that morphosyntactic word is a fundamental and universal category of language structure
- Morphology and syntax both defined in terms of words
 - Morphology deals with composition of words
 - Syntax deals with combination of words
- Haspelmath (2011): “words cannot readily be identified”
- How do we tell a word from an affix or a syntactic phrase?
 - Separation issue: pauses in speech (*but* liaison, connected speech), blank spaces in written language

Evidence for wordhood

- words are real but we haven't found good definition yet
- speakers have clear intuitions about what is and what isn't a word (Langacker, 1972)
 - Education, literacy has taught us what a word is
- the idea of **universality of words** is due to the bias towards written language and the strong influence of the habit of word separation by spaces in Western languages that has been with us for about a thousand years

Orthography



- Orthographies of languages based on Greek, Cyrillic, and Latin alphabets use spaces between words
 - Other languages may not: Chinese, Standard Japanese writing, modern Thai
- Until around 1000, European languages used *scriptio continua* (“continuous writing”, see Saenger, 1997)
 - Spaces probably facilitated readability
- Modern use of spaces is certainly guided by language structure, but in many languages there are obvious inconsistencies in the spelling rules
 - British *healthcare* vs. American *health care*
 - Web site → website
 - German infinitive marker “to”: ... *zu gehen*....., *wegzugehen*

Semantic non-compositionality

- Sapir, 1921: words are “the smallest, completely satisfying bits of isolated meaning into which the sentence resolves itself”
 - essentially a morph
 - Smallest meaningful unit
- Dixon & Aikhenvald, 2002: “words have a conventionalised coherence and meaning”
- However, many complex words are semantically compositional, and many phrases are non-compositional (“fat cat” wealthy, privileged person, “spill the beans” reveal secret)
 - Meaning alone is difficult to use as definitional criterion

Phonology

- Phonological features can discriminate words
 - Phonetic/ phonological features that indicate word beginnings
 - German glottal stop “arm” [ʔa:m] vs. “warm” [wa:m]
 - And word endings
 - Final devoicing
 - E.g., German *Tag* ['ta:k]
 - Segmental prolongation at the end of words: [s] longer in “ice” [aɪs:] than in “sun” [sʌn]
- Phonological word = a string of sounds that behaves as a unit for certain kinds of phonological processes, especially stress or accent
 - Phonological words are domains of syllabification → “they hate us” syllabified as “ha-tus” [heɪ-tʌs]
 - Last syllables [tʌs] on lexical boundary between verb and pronoun
- “the phonological word does not always coincide with the morphological word” (Bresnan & Mchombo, 1995)
- Clearly, phonological criteria alone cannot define words (Bloomfield, 1933; Hockett, 1958)

Morphosyntax

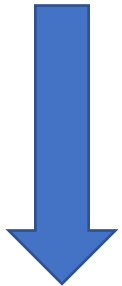
- Some morphosyntactic criteria and problematic cases
 - *Free occurrence*
 - Free morphemes, bound morphemes
 - But: transitive verbs, articles, etc. cannot occur on their own
 - * “We discussed in the video” → requires object: “different topics”
 - If an element can occur independently it must be minimally a morphosyntactic word, not an affix
 - *Mobility*
 - Affixes are bound, but words can occur in different position
 - Most words have fixed position with respect to other words
 - Few languages have truly free word order

Non-universal definition

- Best solution may be to define words language-specific
 - “what we call ‘words’ in one language may be units of a different kind from the ‘words’ in another language “ (Lyons, 1968)
 - “there may be clear criteria for wordhood in individual languages, but we have no clear-cut set of criteria that can be applied to the totality of the world’s languages” (Spencer, 2006)
- Words are fuzzy concepts
 - Notion of half-words (or semi-words, e.g., articles)
- But talking about words is deeply entrenched in linguists’ habits

What is word knowledge?

- What does it mean to *know* a word?
- Continuum of word knowledge



- No knowledge; don't even know it exists
 - Awareness that such a word exists, but don't know what it means
 - Vague notion of what the word means, in a particular context
 - Rich understanding; know the word well and can use it
-
- Passive vs. active vocabulary
 - Guessing: “*roundhouse*”
 - Instances of never having seen the word but ‘knowing’ it instantly
 - morphological compounding

Collections of words

- Mental lexicon
 - dictionary-like structure
 - contains the knowledge about words acquired over a lifetime
 - and in various languages
- Human repository of lexical knowledge (Oldfield, 1966)
 - can be seen as a dynamical memory system supporting linguistic processing that is continuously adapting
- A typical educated American adult knows the meaning of about 40,000 words (de Deyne et al., 2017)
- At age 5, children know about 3000 words (Aitchison, 2012)

The mental lexicon

- organizes words according to their various properties
 - semantic properties (meaning)
 - syntactic properties (e.g. part-of-speech)
 - perceptual characteristics (e.g. pronunciation)
 - pragmatic ones (e.g. appropriate usage)
- Representation and processing of words in the mental lexicon holds information crucial for
 - theories on language acquisition and development
 - cognitive principles underlying human lexical storage
- Abstract concept
 - best conceptualized as an ideal, abstract notion, rather than a concrete list of word knowledge (Aitchison, 2012)

Memory and search

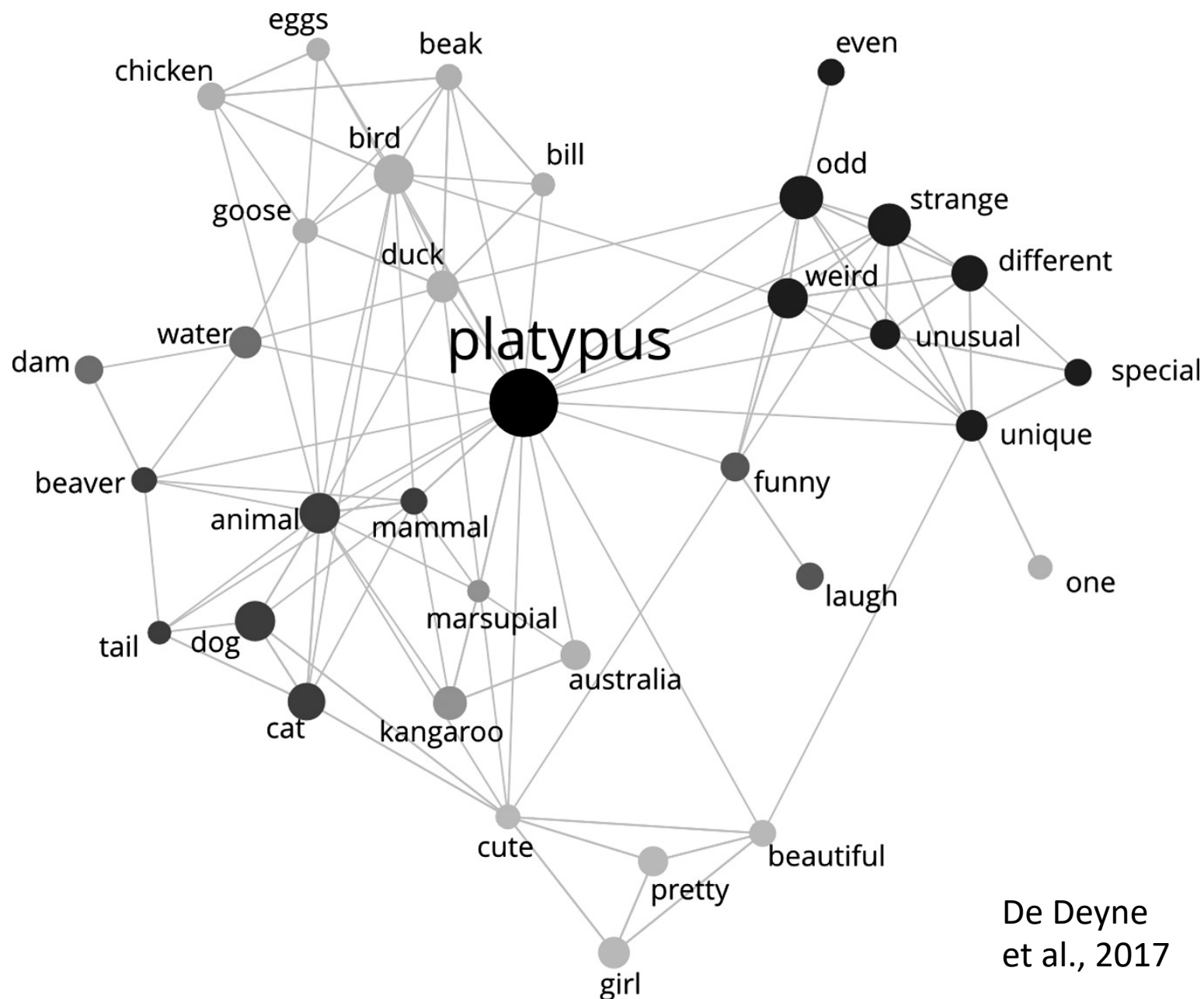
- ML is unlike a real dictionary
- rather than provide explicit definition for words, ML represents meanings in terms of patterns of word use and the connections between words (Elman, 2009)
- “lexical processing” refers to retrieving words from the mental storage in perception or production
 - Production and perception processes different
- lexical processing experiments can give insights into how the ML is organized and structured and what factors play a role for perception or production retrieval

Which words do you associate
with the word
platypus?

Organization

Linked entities

- Figure shows an associative network around the word “platypus”
- Association experiment:
 - Name 5 words that come to your mind when you hear “platypus”
 - Based on human relatedness judgement



De Deyne
et al., 2017

Word associations

- Perceptual similarities
- Semantic
 - Psychological associations (--> *cat* and *dog*)
 - both are common pets, words remind of each another
 - Shared features
 - share at least one semantic feature (*four-legged, furry, animal*)
 - Taxonomic relationship
 - classification, e.g., subordination: *animal* – *dog*
 - Co-occurrences
 - *bread + butter, parking + lot*
- Phonological
 - Maximal phonological similarity, not more than one segment difference
 - phone – tone → foun - toun
 - cat – mat → kæt - mæt

Cognitive net of words

Aitchison (2012):

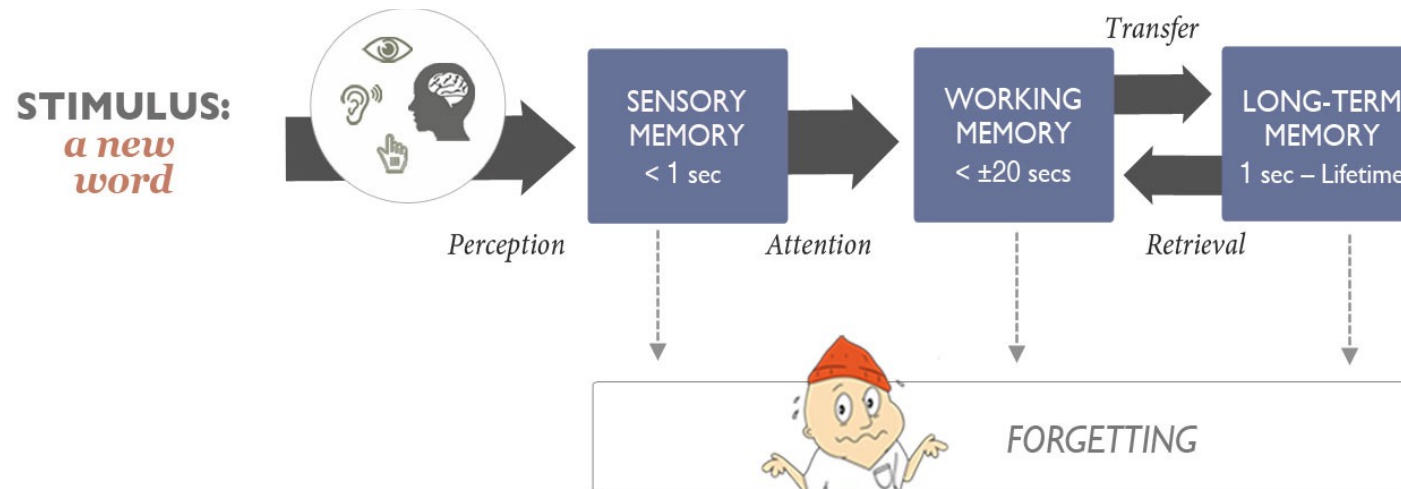
a large structure of word relations
exist in the mental lexicon
the mental lexicon can be seen as a
“gigantic multidimensional
cobweb”

Organized according to perceptual
similarity



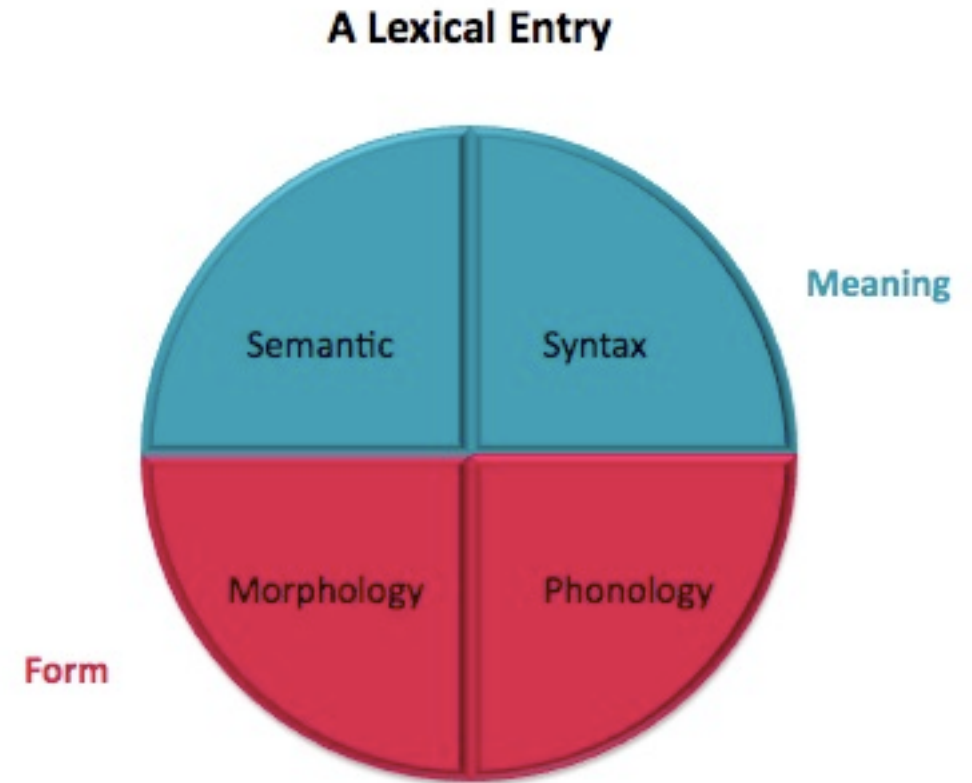
Long-term memory

- Mental lexicon is part of the long term memory
- Mental lexicon is the part that stores lexical items
 - including information that goes with them (morphology, meaning)
- From functional perspective, long-term memory component of the mental lexicon is comprised of three levels:
 - concept, lemma, phonology



Lexical access

- Lexical access occurs in speaking when the pronounceable word forms are found which encode the concepts the speaker wishes to communicate
- and it occurs in listening when the meaning is found which is expressed in the word forms the listener has heard
- Dimensions of words

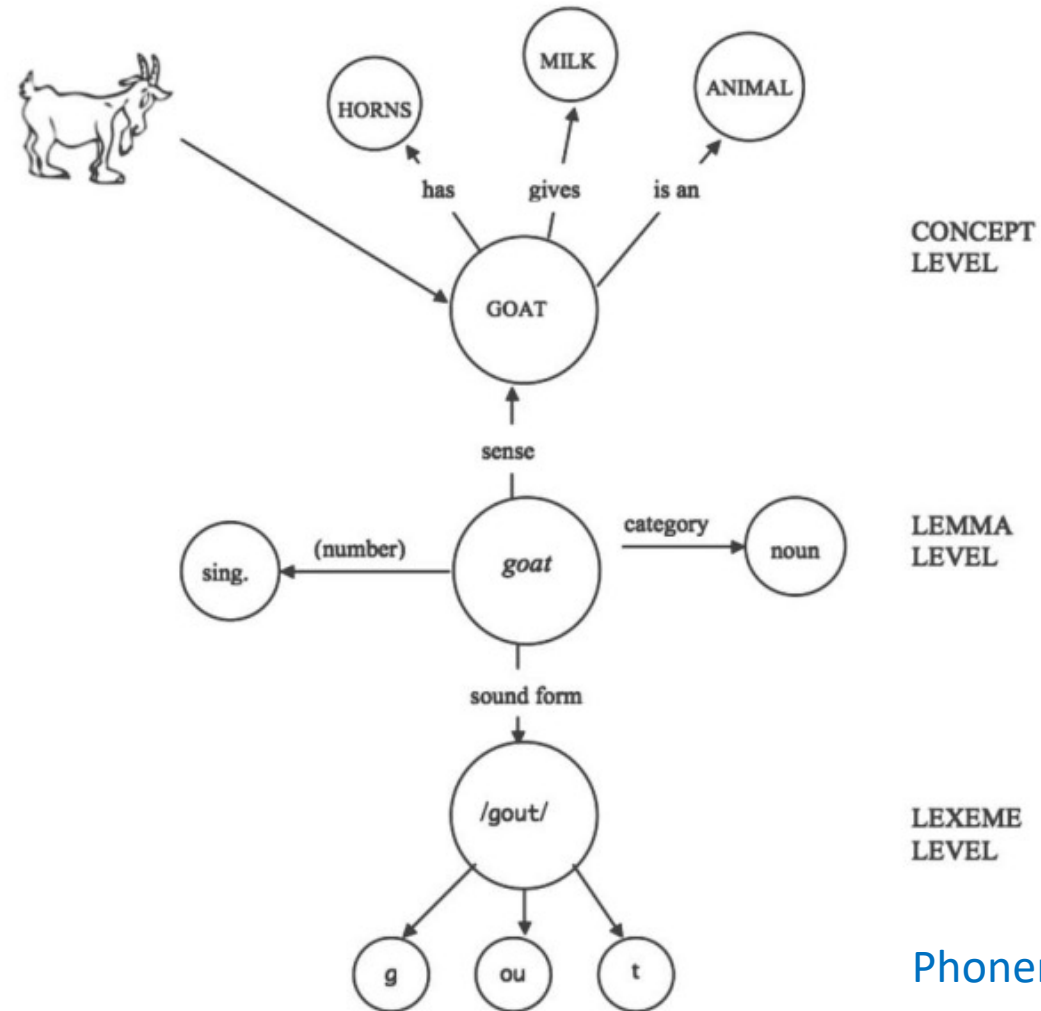


Levelt, 1989

The existence of lexical entries

- Working definition of word: “mapping sound to meaning”
 - Arbitrary, conventional
 - [noʊz] ‘nose’
 - Presumably non-arbitrary: onomatopoeia “quack”,
 - French “coin”, Danish “rap”
 - Homophony [aɪ] ‘eye’ or ‘I’
- Mental representations of words can be defined and categorized in many ways (Brown et al., 2020)
- One approach to defining words relates to the different in their forms (McLean, 2018)
 - Word **types** consist of each unique word form
 - *record, records, unrecorded*
 - **Lemmas** are made up of a headword and its inflections
 - *record* = head, +infl.
 - Part of speech: noun(*record, records, recording*), verb(*record, recording, recorded*)
 - **Flemmas** are larger classifications of lemmas
 - one headword “record” + all inflections and derivations (*pre-record, recorder...*)

Levels of representation

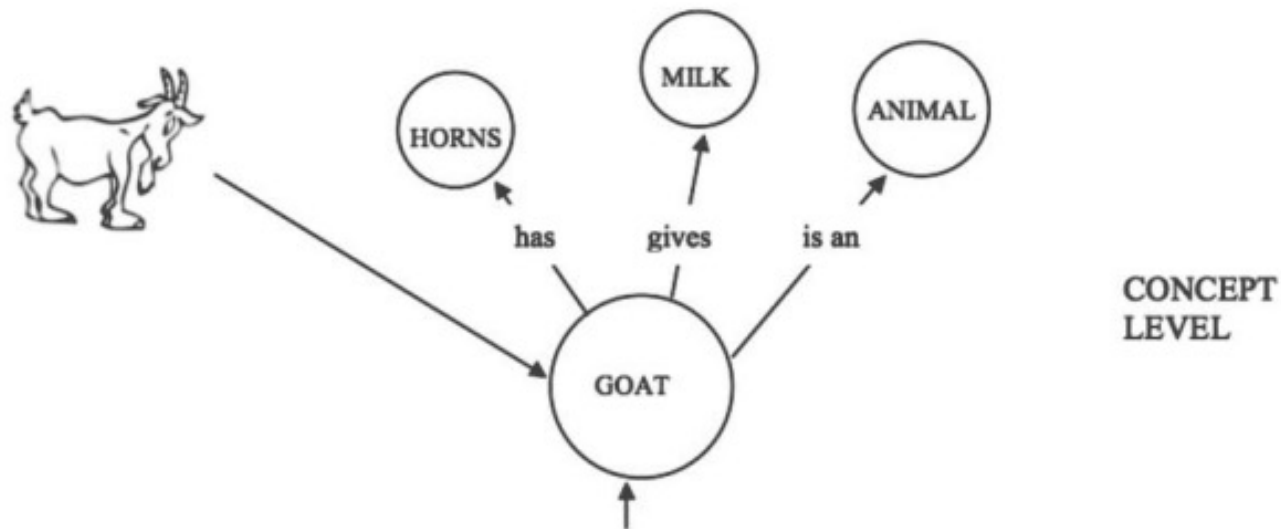


Bock & Levelt, 1994

Phonemes/
segments

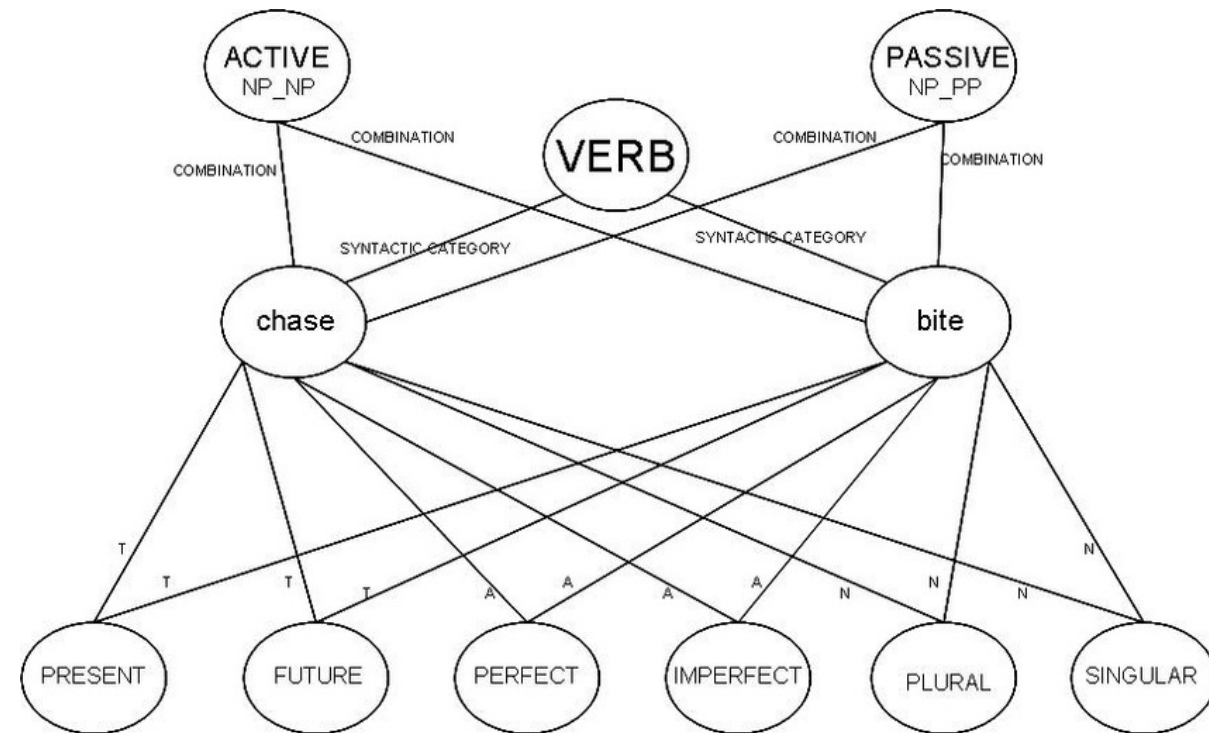
Conceptual level

- at the conceptual level, knowing a word involves knowing its meaning
 - about a goat we know it is a kind of domestic animal that produces milk, etc.
 - and also that it typically **de-selects** certain other words such as *think* or *smile* typically reserved for humans
 - inhibition



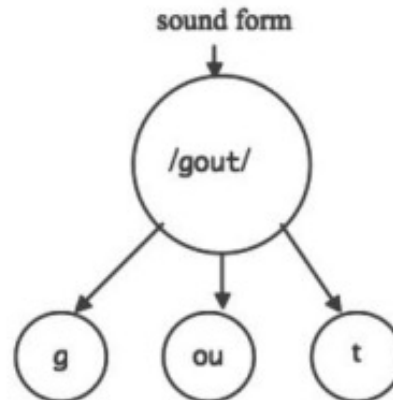
Lemma level

- at the lemma level, a word has syntactic properties, a bundle of grammatical features, including also combinatorial information (Sells 2001; Kim & Sells 2008), which place it in its syntactic frame
 - the English word goat is a noun
 - Its Italian equivalent *capra* is also a noun, but in addition it has feminine syntactic gender
 - *Nouns*: grammatical gender, number
 - *Verbs*: person, number, tense, mood
 - ...



Lexeme level

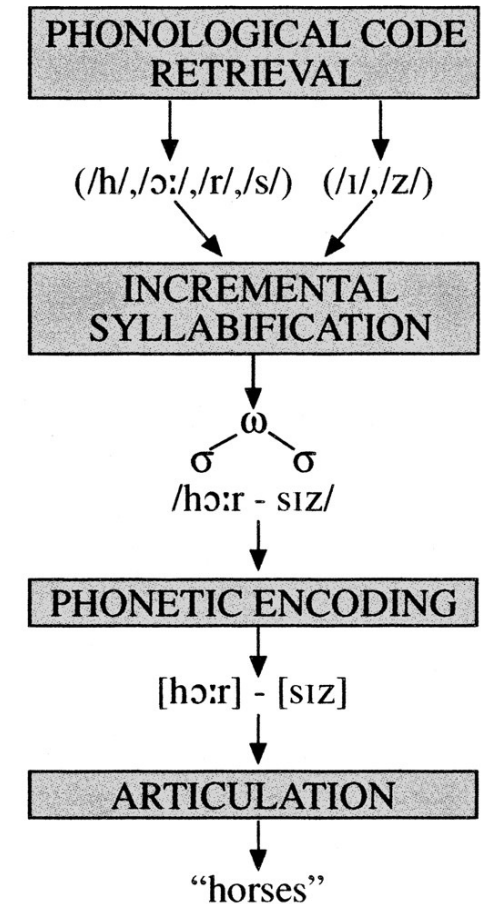
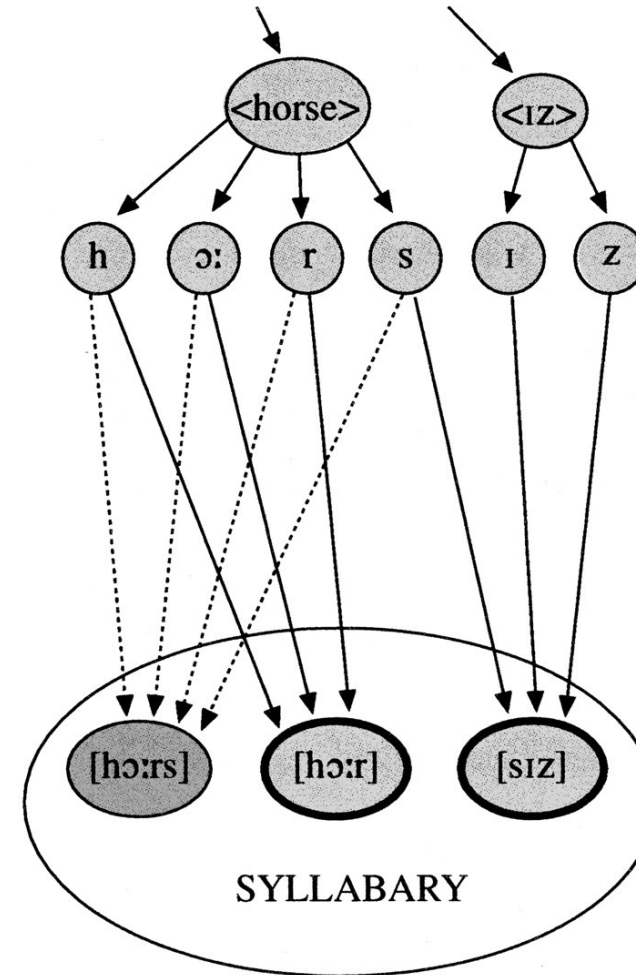
- at the lexeme level, knowing a word means knowing its formal properties that is, its morphological and phonological shape
 - the word goat is monomorphemic and consists of three phonological segments: /g/, /ou/, and /t/
 - whereas the Italian word *capra* consists of two morphemes, a stem (capr-) and a suffix (-a), and five phonological segments: /k/, /a/, /p/, /r/, and /a/



LEXEME
LEVEL

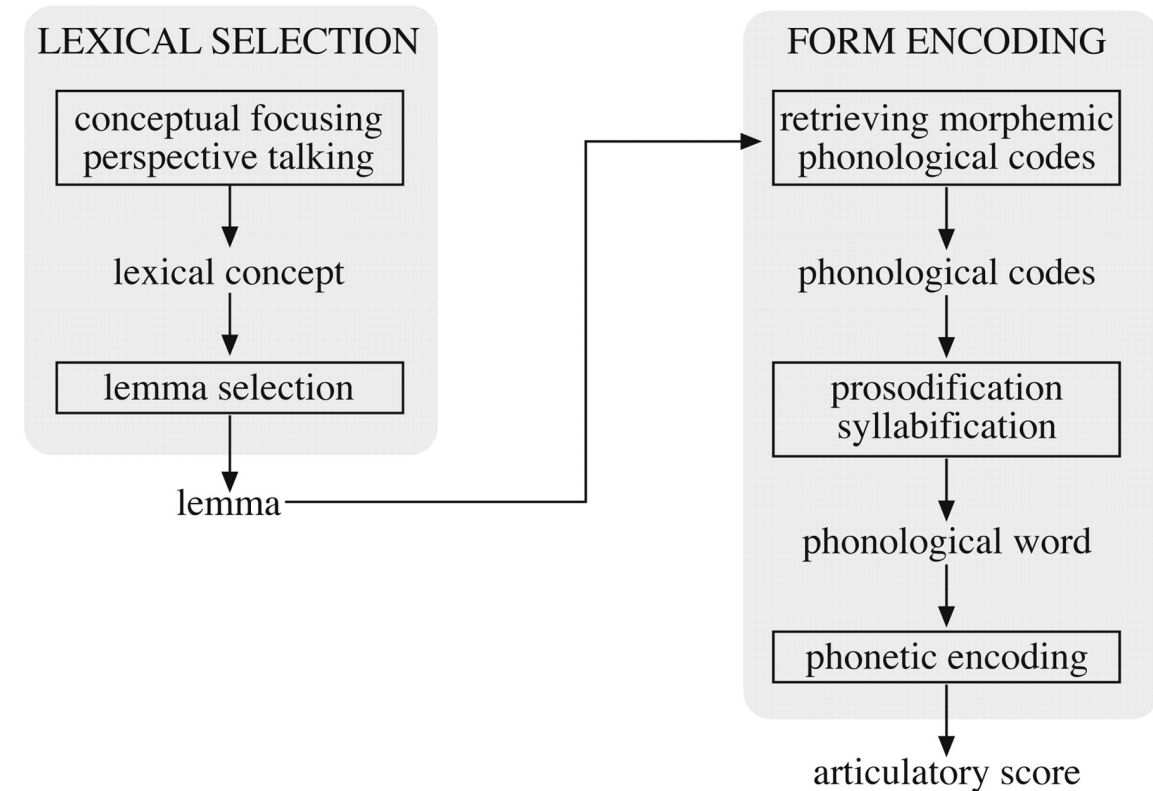
Segment level

- Phonological encoding
 - assembling words from lexical entries
 - prepare articulation in speech production
 - constructing the phonological form of a target token before articulatory gestures can be prepared in spoken word production (Dell, 1986; Levelt et al., 1999; Caramazza et al., 2001)



Phonological pre-activation

- phonological pre-activation or phonological preparation (Li et al., 2017) is initiated at the start of the target word
 - Dutch word generation: faster responses in a homogeneous context (e.g. *hut*, *heks*, *hiel*—hut, witch, heel) relative to a heterogeneous context
- phonological segments of target words are sequentially activated (Oppermann et al., 2010)
 - segment = level of preparation in Germanic languages
 - onset segment(s) attributed a crucial role in the process (Meyer, 1991; O’Séaghdha et al., 2010)
 - implicit phonological prime
 - loner, local, lotus → preparation faster and more accurate
 - as opposed to “loner, beacon, major”



‘Proximate units’

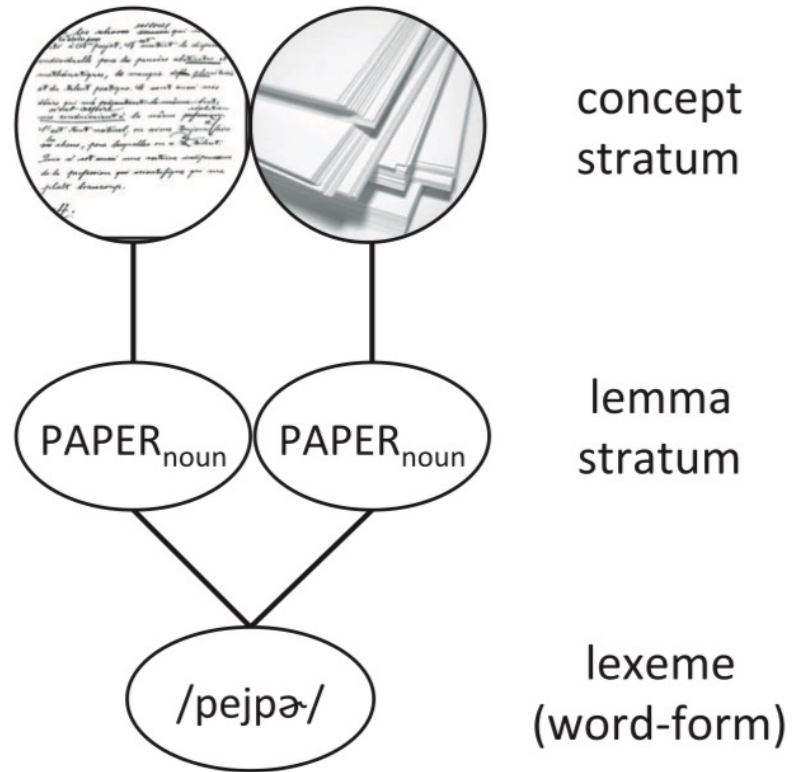
- not all languages reap word-activation benefits from the preparation of the **onset** phoneme
 - in Japanese phonological encoding begins at the level of the mora (Kureta et al., 2006)
 - in Mandarin Chinese at the syllable (Chen et al., 2016)
- concept of ‘proximate unit’ (O’Séaghdha, 2015)
 - different types of phonological units that serve as activating mechanisms for encoding
 - defined as the “first selectable phonological unit below the level of the word or morpheme” (O’Séaghdha et al., 2010: p. 286)
 - in Germanic languages = segment
- proximate units are influenced by frequency effects (Stemberger and Macwhitney, 1986; Levelt and Wheeldon, 1994)
 - high-frequency onsets and syllables generally facilitating phonological construction of a target word (Aichert and Ziegler, 2004; Luef and Resnik, 2023; Macizo and Van Petten, 2007)

Special cases of words: Homophony, polysemy

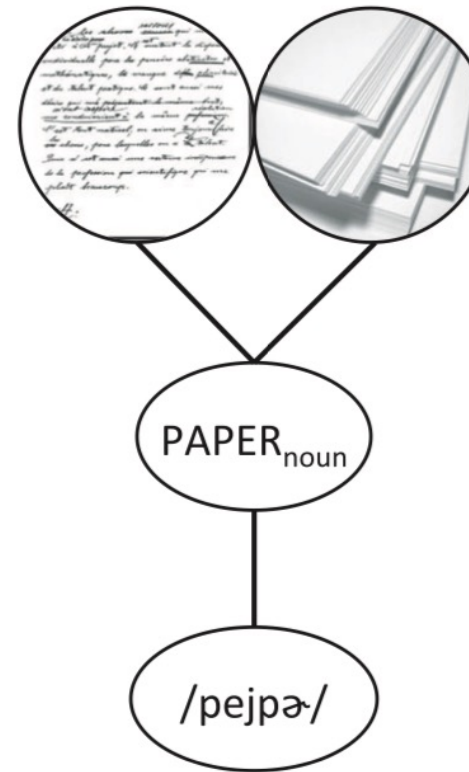
- ubiquitous across languages (Antilla 1989)
- how they are created is an important question in psycholinguistics (Wedel, Kaplan, and Jackson 2013)
 - Can arise from sound change (/k/ dropped in „knight“ – homophonous to „night“; see Lutz, 1988)
 - Polysemy (*mouse*: rodent, computer device)
 - Continuum of semantic relationship
 - Zero conversion: to water sth – water, sheep - sheep
- two lemmata linked to one phonological and phonetic form
 - bank (\$), bank (river)
 - cut (V), cut (N)

Polysemy/ homophones

A Separate-storage view



B Core-lexical view

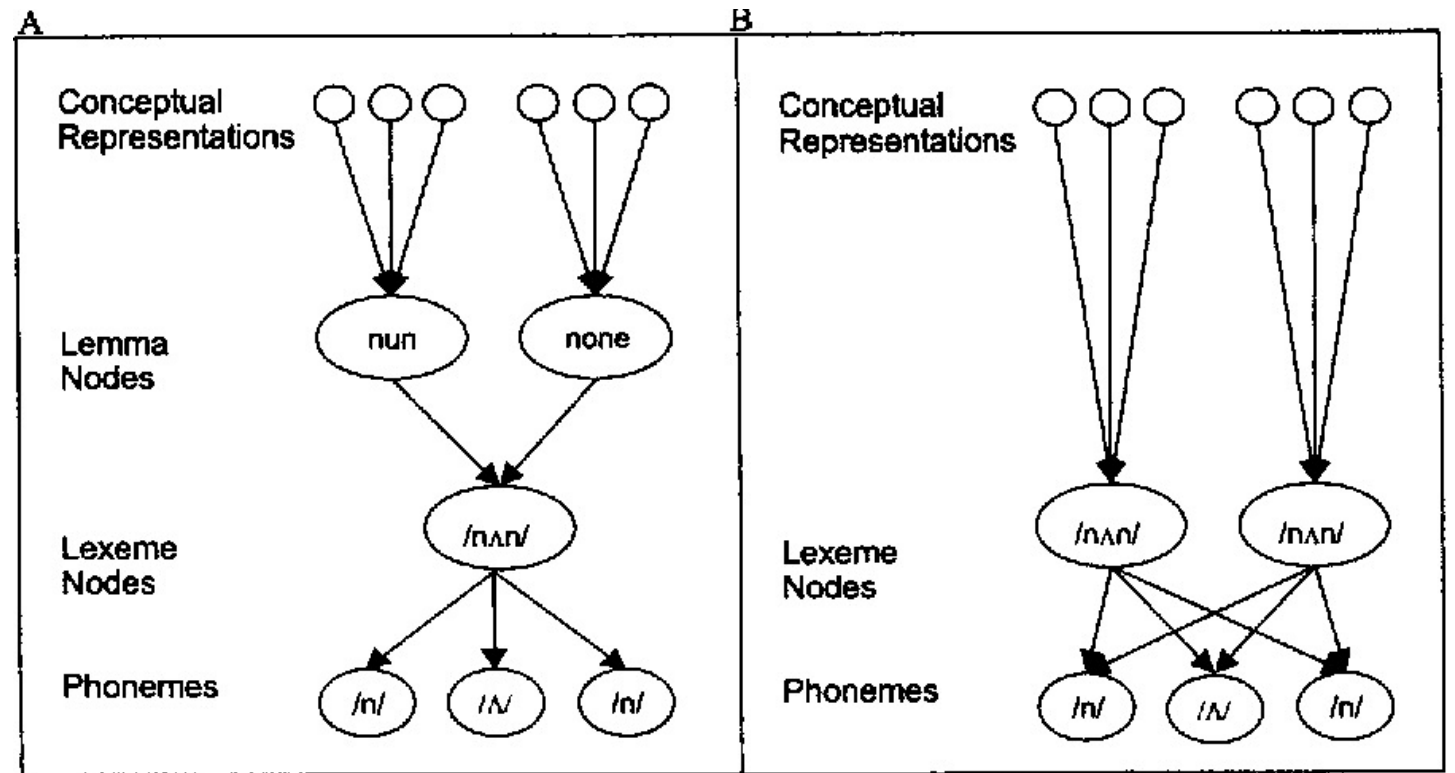


What if the two
lemmas showed
different word forms?

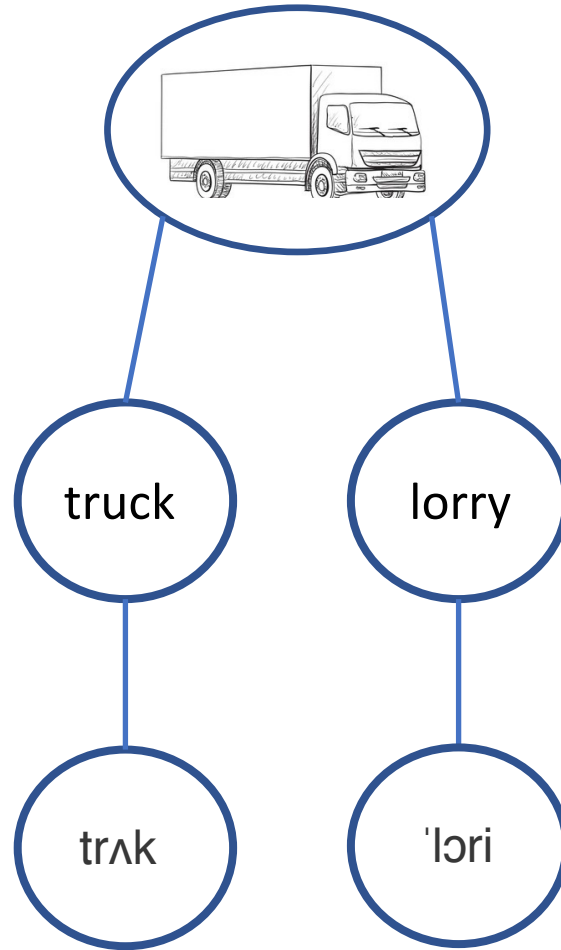
Frequency inheritance

- Frequency effects reduce a word phonetically

- e.g., *time* – *thyme* (Gahl, 2008; Lohmann, 2017)
 - HF “time” is shorter
- HF “none”
- LF “nun”
 - Phonetic realizations are not identical
 - Lemma frequency inheritance → phonetic realizations should be identical

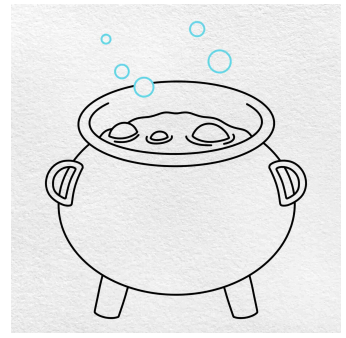


Special cases of words: (Full) synonyms



One concept
Two lemmata
Two lexemes

No “cauldron of lexical soup”



- Word memory is organized
- Words are linked to one another according to certain principles
 - Phonological similarity (*nun - none*)
 - Semantic similarity (*sheep – goat*)
- Interconnections exist
 - Influence priming

References

- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. Wiley-Blackwell.
- Bloomfield, L. (1933). *Language*. New York: Holt.
- Bresnan, J., & Mchombo, S. M. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory*, 13, 181-254.
- De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. (2017). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Frontiers of cognitive psychology: Big data in cognitive science* (pp. 174-202). New York: Routledge/ Taylor & Francis Group.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547-582.
- Hockett, C. (1958). *A course in modern linguistics*. New York: MacMillan.
- Haspelmath, Martin. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45(1), 31-80.
- Lyons, J. (1968). *An introduction to theoretical linguistics*. Cambridge, UK: Cambridge University Press.
- Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology*, 18, 340-353.
- Saenger, P. (1997). *Space between words: The origins of silent reading*. Stanford: Stanford University Press.
- Sapir, B. (1921). *Language: An introduction to the study of speech*. New York: Harcourt Brace.
- Spencer, A. (2006). Morphological universals. In R. Mairal & J. Gil (Ed.), *Linguistic universals* (pp. 101-129). Cambridge: Cambridge University Press.