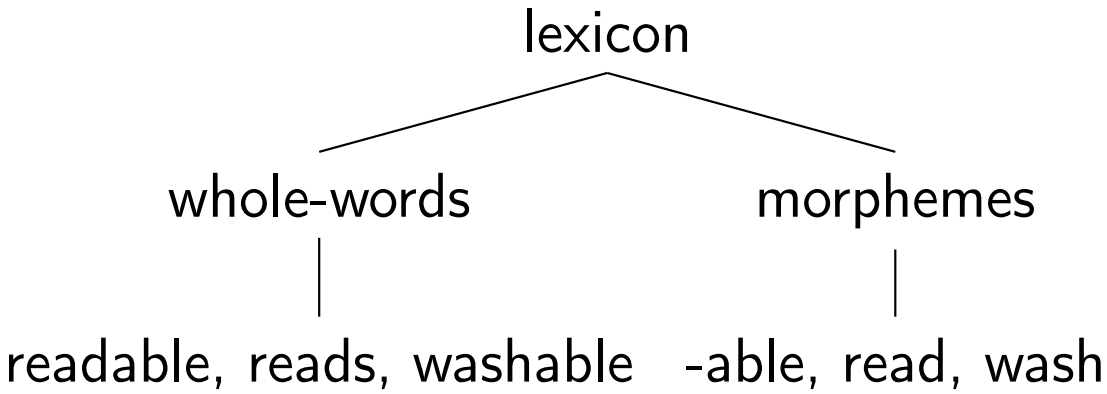
Model of the lexicon

HUL 243: Language and Communication

23rd September 2024

Lexicon

- Words in the language user's vocabulary are stored in the mental *lexicon*
- Words must be **listed** in the lexicon for them to be a part of it
- What is the actual form of this listing?
  - Are words stored whole e.g. *readable, washable, reads*
  - Are words stored as morphemes e.g. *read, wash, -able*



Morpheme

A Morpheme is the smallest linguistic unit with a grammatical function (Aronoff)

- Let's look at the word *reconsideration*
  - It consists of three parts re- + consider + -ation
  - *consider* is the stem i.e. a base unit to which other morphemes may attach
  - re- and -ation are both **affixes**; re- is a **prefix** and -ation is a *suffix*
  - reconsideration consists of 3 morphemes

Lexical access

- **Lexical access** is the process of looking up a word in the lexicon
- If both morphemes and whole words are stored in the lexicon:-
- Two potential ways to look up a given word-
  - **Decomposition** route i.e. access words by breaking them into morphemes
  - **Direct** route i.e. retrieving words without decomposition
- E.g. a word like *insane* is retrieved as *in-* + *sane* or *insane*

- Dual route model suggests both routes are in use: decomposition and direct access
- The winner is whichever method is faster in accessing the information

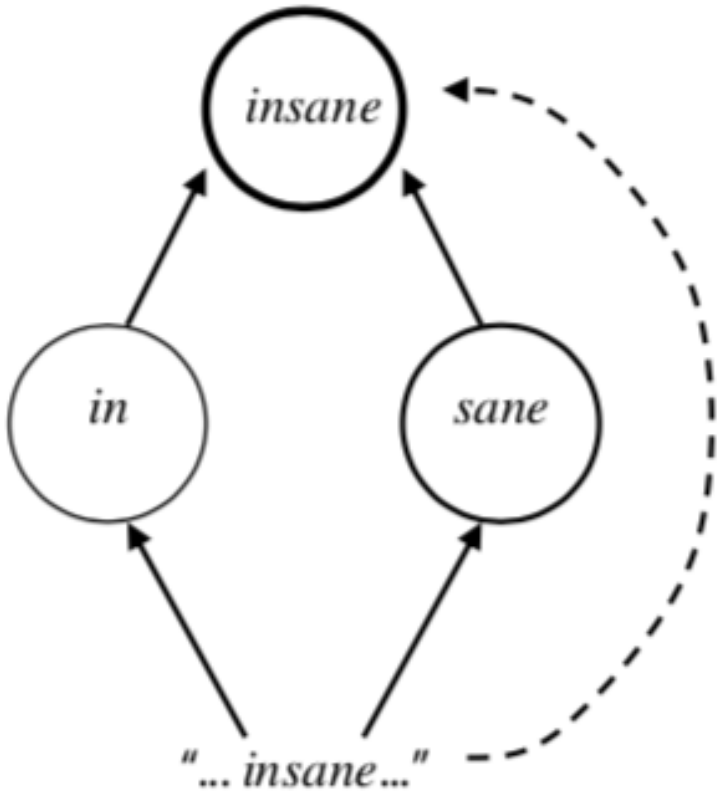


Figure: Schema of the dual-route model, Hay (2001)

Morpheme

A Morpheme is the smallest linguistic unit with a grammatical function (Aronoff)

- Let's look at the word *reconsideration*
  - It consists of three parts re- + consider + -ation
  - *consider* is the stem i.e. a base unit to which other morphemes may attach
  - re- and -ation are both **affixes**; re- is a **prefix** and -ation is a *suffix*
  - reconsideration consists of 3 morphemes

Children's acquisition of words

- First-language learners speak a language without explicit instruction by age 3
- Somehow, children distil out the essential rules of the language based on the finite set of utterances they have heard
- How do children make this leap?
- Children generalize from the set of of linguistic structures they hear

Children's acquisition of words

- First-language learners speak a language without explicit instruction by age 3
- Somehow, children distil out the essential rules of the language based on the finite set of utterances they have heard
- How do children make this leap?
- Children generalize from the set of of linguistic structures they hear

My teacher holded the baby rabbits and we patted them
I finded Renee
Hey I heared a sound

Children's acquisition of words

- Children learn something like 'to form the past tense, add -ed'
- When encountering *find, hear, hold* $\rightarrow$ *found, heard, held*, they **over-regularize**
- Shows that an internal awareness of the past tense formation rule
- Note that children often use *both* forms *held-*holded*, *went-*goed* for some time without realizing it
- At some time, the switch flips– when they realize that *held, went* are `past` (like the -ed forms)

Children's acquisition of words

- This developmental pattern seems to suggest the following:-
  - Regular complex words like *show-showed* are likely to be listed in the mental lexicon as combinations of roots and affixes
  - Perhaps, the idiosyncratic or irregular forms are stored (as these are a minority)
  - Words are formed via computation, structure building
  - This view supports the idea of the lexicon as morpheme-based

Morpheme-based lexicon

- Children eventually learn to build (generate) the rules for existing forms
- They also learn to constrain the rules to **not** generate ungrammatical (non-existent) forms
- When children learn the elements from an inventory of word-forms they **create** via analysis
- A morpheme-based lexicon places emphasis on assembling the pieces of words
- Another way to characterize this would be that it is heavy on **computation**
- Advantage of such a view: elegance/economy of representation

Economy

- Memorizing entire words leads to redundancy:-
- Below is the paradigm for Greek lexeme `filos` 'friend'
- Total of 7 morphemes: fíl-, -os, -i, -on, -u, -us, -o vs. 6 individual word forms

	singular	plural
nominative	fílos	fíli
accusative	fílo	fílus
genitive	fílu	fílon

- But `filos` is part of an inflectional paradigm that includes `kosmos` 'world', `fovos` 'fear', `gamos` 'marriage', `skilos` 'dog'
- 30 word-forms vs. 11 morphemes

Lexical economy

- Even derivational forms can occur in families, `read` and `write` are related in a parallel fashion

read	reader	readable
write	writer	writable

- Morpho-phonological rules of allomorphy are also economical
- E.g. consonant devoicing rule [-z] $\rightarrow$ [-s] before voiceless sounds can be found in plurals and past tense:
  - cat $\rightarrow$ cat[s]; dog $\rightarrow$ dog[z]
  - work $\rightarrow$ worke[t] ; arrive $\rightarrow$ arrive[d]

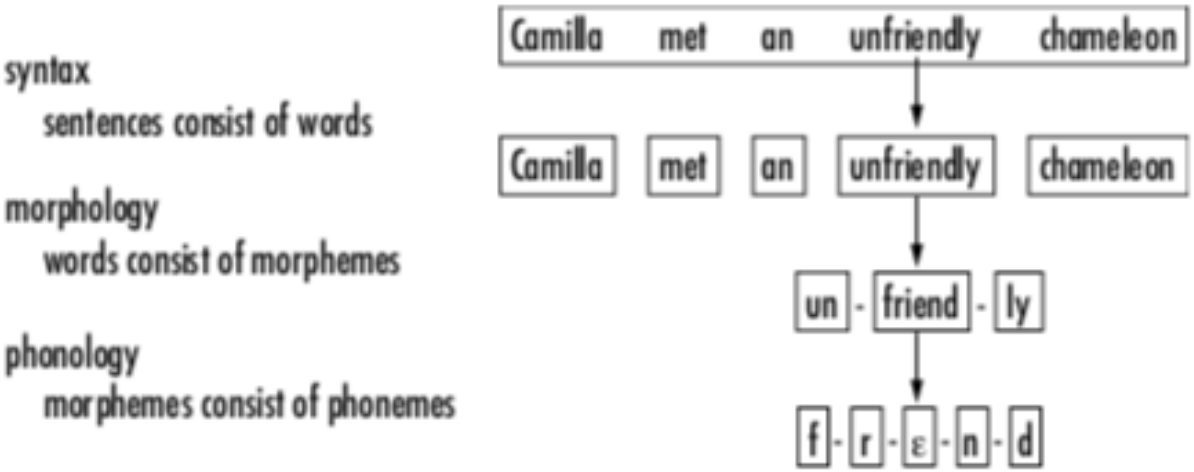
Agglutinative languages

- Agglutinating languages have a high ratio of morphemes to word
- Each word form can occur with many possible affixes
- Impossible to memorize that many forms!

Turkish

(1) oku-ya-ma-yabil-ir-im
read-POT-NEG-POT-AOR-1SG
'I might not be able to read'

Modularity



- At the phonological level, words consist of individual phonemes
- Elements at phonology combine ('assembled') and feed into morphology
- At the morphological level, morphmes combine to form new words
- In syntax, these words combine to form phrases, then sentences

- Computation is cheap– storage is expensive!
- Generate morphological rules from morphemes, phonological rules from phonemes
- No storage in lexicon, unless absolutely needed!
- Lexicon is a space that only has the following:-
  - Monomorphemic forms and affixes
  - Irregular forms (*hold-held, find-found*)
  - Fossilized forms (*ox-oxen, mouse-mice*)
  - Semantic idiosyncrasy (suffix -ee attaching to animate nouns)
- Lexicon is static, the grammar has the moving parts

Challenges for morpheme-based approach

Irregularity

- Irregularity in paradigms is very common
- French paradigm for verb `parler` 'speak'

je parle	I speak	nous parlons	we speak
tu parles	You speak	vous parlez	you(pl) speak
il/elle parle	He/she speaks	ils/elles parlent	they speak

- Irregular `aller` 'go' + many irregular forms of verbs !

je vais	I go	nous allons	We go
tu vas	You go	vous allez	You(pl) go
il/elle va	He/She goes	ils/elles vont	They go

- If irregular forms are stored, economy $\downarrow$

Challenges for morpheme-based approach

Irregularity

- English suppletive forms:-
  - *go* $\rightarrow$ *went* must be stored (go+ past)
  - *bad* $\rightarrow$ *worse* must be stored (bad+ comparative)
- Not a challenge per se, but does imply increased use of lexical storage

Challenges for morpheme-based approach

Predictability of meaning

- Predictability of meaning
  - Composition of forms should yield compositional meaning, but not always!
  - The items in the second column are mono-morphemic, must be stored separately

X + -er	*X*? + -er
paint**er**	cobbler
lectur**er**	butcher
influenc**er**	bursar

- Other meanings may not be predictable e.g. *mailer* 'a pamphlet sent by post', *grinder* 'machine used for grinding'

Challenges for a morpheme based approach

Non-concatenative morphology

- Morphological process not due to affixation
  - Zero expression e.g. *fish (sg)* $\rightarrow$ *fish-*$\emptyset$ *(pl)*
  - Conversion e.g. *repeat (V)* $\rightarrow$ *repeat (N)*
  - Base/stem modification e.g. *sing* $\rightarrow$ sang
  - Reduplication e.g. in Hindi *dheere dheere*

Non-concatenative morphology

- For many of these cases, solutions have been proposed in the literature
- E.g. sing $\rightarrow$ sang have been explained via a phonological rule applying in the environment of a null ($\emptyset$) past-tense morpheme
- Including zero morphemes implies that we still have a economical lexion–
- ...but also different lexical entries:- $\emptyset$ had no form, but meaning

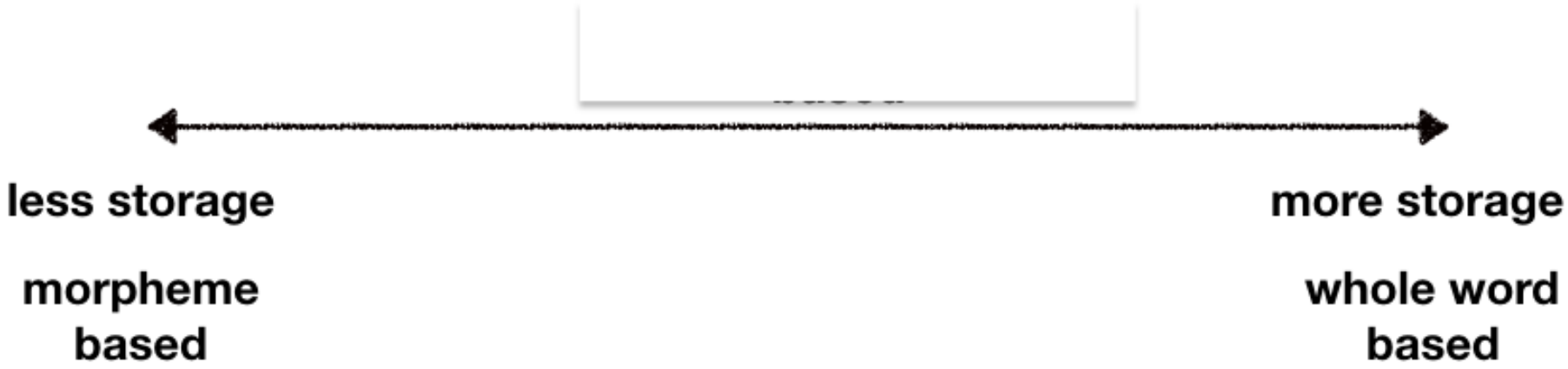
- The morpheme-based approach is represented by a number of morphological theories, each of these may differ in their details
- But in order to follow the 'maximum computation, minimum storage' maxim, the system can become complex Tiv

(Niger-Congo, Nigeria)

Root	Imperative	Gloss
kimbi	kìmbí	pay
de	dé	leave
gba	gbá	fall
va	vá	come

- The imperative meaning is contained within the rule of tone assignment (but not to a particular form)

Two approaches differ based on storage



- Morpheme-based approaches minimize storage, maximize computation
- Whole-word based approaches maximize storage, minimize computation

References

- Haspelmath and Sims, Ch 4
- Steven Pinker 'Why the child holded the baby rabbits: a case study in language acquisition' from 'An invitation to Cognitive Science' eds. Gleitman, Osherson, Liberman