√ = lexical root <…> = morpheme /…/ = phoneme […] = syntactic constituent

Language & Linguistics

I-language (Chomsky 1986): linguistics as a **cognitive science** (study of an aspect of the **mind**), as opposed to a linguistics as a discipline of the *humanities*

- *I* = *i*nternal to our mind/brain, our **knowledge** of language (ultimately, neurons & synapses), not an *external* phenomenon in the world
- *I* = *i*ntensional *vs.* extensional

  ex. in maths, square numbers can be defined intensionally by the function ('rule') X^2, or extensionally by giving the *infinite* set {1, 4, 9, 16,…}

  ex. in linguistics, plural marking can be defined as the rule √NOUN-*s*, or by the set {*books*, *zucchinis, linguists,…*} (obviously, you need a *list* for exceptions anyway, e.g. √MOUSE-*s* → *mice*,…)

  ➲ An intensional def. is more economical (viz. shorter) than an extensional one!
- *I* = *i*ndividual language ('I') → every human has it (but cf. physical impairments) → truly universal

Nativism: certain skills or abilities (types of knowledge) are hard-wired into the brain at birth (genetic prespecification) → **knowledge of language is innate** → **Universal Grammar** (UG = a *mental* grammar, not a textbook = i-language; cf. Chomsky 1965): *all* humans share this ability!

➲ **I-language** is like an **instinct** → similarly, the ability to walk is an instinct, it need not be taught to a child, but develops automatically → explicit instructions & corrections are ineffective, both for walking and for language!

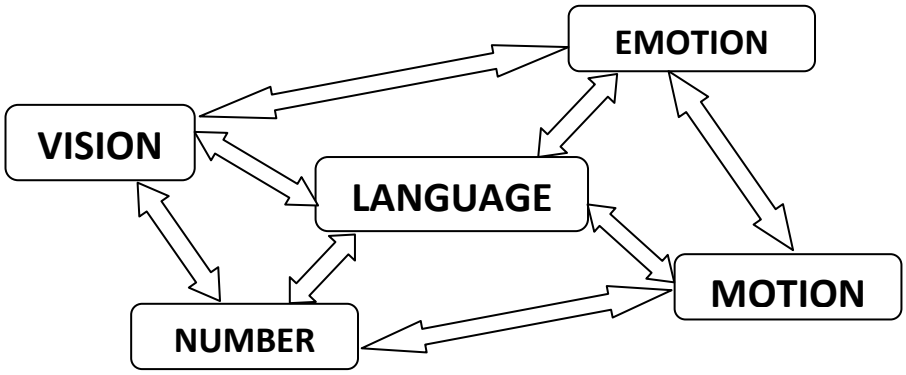
➲ *Specific languages* (English, Italian, KiSwahili,…) are **externalised instances of i-language** (= e-languages) deriving from the interaction of i-language with external linguistic data from the environment → variation among the world's languages is apparent: **for the most part differences pertain to *words*** (whose form is arbitrary; cf. de Saussure), but ***syntactic structure*** is surprisingly regular, varying systematically only along well-defined dimensions → **True chaos is on the surface (words), but at a deeper level, there's regularity (grammar)!**

➲ Knowledge of language is **subconscious**: A speaker of English has the knowledge to **instinctively** recognise a sentence like **Who do you wonder what bought* as **ungrammatical** (not well-formed, marked by '*'), but the average person cannot put her finger on what's wrong with it (i.e. in English, only one and only the leftmost *wh*-element may be fronted; cf. below)!

➲ A note on well-/ill-formedness: While **grammatical ill-formedness** (**ungrammaticality**) refers to violation of *rules of grammar* encoded in UG, there is also *semantic ill-formedness* (marked '#'): cf. the contrast **The rock are hard vs.* #*The rock is pregnant*.

Modularity of mind (Fodor 1983)

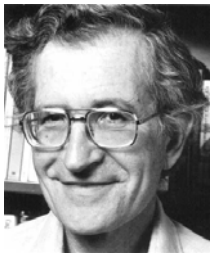
- Cognitive modules are autonomous, encapsulated (▢)
- They communicate via interfaces (⇔)
- They are specialised (domain-specificity)



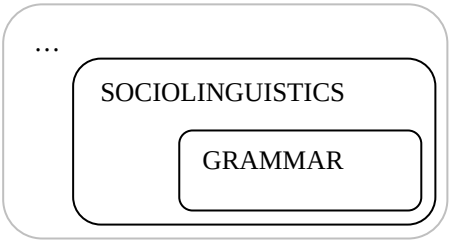
➲ **Mind/brain**: Cartesian dualism made a difference between immaterial substance (mind) and physical matter (brain) → modern cognitive science assumes identity on some level of abstraction, hence 'mind/brain' → modules are not unified chunks of brain matter, but rather networks of neurons and synapses → linguistics is part of **psychology** (*vs.* **neurology**)

[➲ *slides on the Müller-Lyer optical illusion*]

Generative linguistics



- Founded by **Noam Chomsky** (in his 1957 booklet *Syntactic Structures*)
- Theory of **UG** (cf. above) as the **language module** → core linguistics = grammar
  - ➲ Sociolinguistics, etc., is not unimportant, but simply presupposes a theory of grammar, thus is 'peripheral'



Principles & Parameters Theory (PPT; Chomsky 1981)

Principles: properties of UG common to all languages

E.g. *Shortest Move* principle: 'if you have to move an element, take the closest one'. Assume that *wh*-questions (in English) are derived by movement of the *wh*-element:

(1) [John read <u>*which book*</u>] → [<u>*which book*</u> did John read ~~*which book*~~]?

Which book is interpreted both as the direct object of *read* ('John read some book') and the entity being questioned ('For which specific book was it the case that John read it?'). Now, with **multiple wh-questions**, we can only have the **leftmost *wh*-element** in initial position, otherwise we get an ungrammatical (*) sentence (the so-called *Superiority Condition*; Chomsky 1973):

(2) a. [*who* read *what*]? → *[*what did who* read ~~*what*~~?

Parameters: varying dimensions of UG (differences between e-languages)

E.g. *Directionality Parameter*: *heads* (here: adposition/verb) either follow or precede their *complements* (here: pronoun/direct object)

(1) **with** me – *English (preposition)* (2) Ali [elma **yedi**] –*Turkish (SOV)*
benim **le** – *Turkish (postposition)* Ali [**ate** an apple] – *English (SVO)*

In connection with *wh*-movement, there's also a complex parameter (actually, two): Either ① you have to move only the leftmost *wh*-element (e.g. English), ② you multiply move *all wh*-elements (e.g. Bulgarian), or ③ you leave the *wh*-element in place (e.g. Chinese):

(1) **Who** said what? *English*

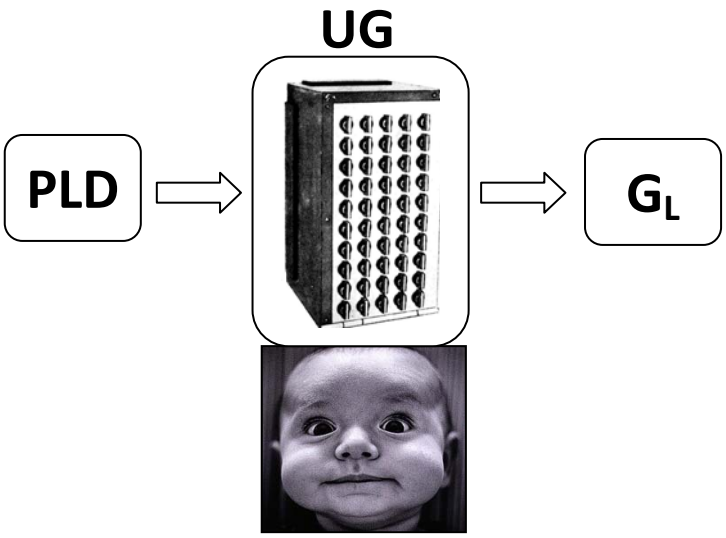
(2) Zhangsan xiang-zhidao [Lisi mai-le **shenme**] *Chinese*
Zhangsan wonder Lisi bought what
'Zhangsan wonders what Lisi bought'

(3) **Koj kakvo** kupuva? *Bulgarian*
who what bought
'Who bought what?'

Child language acquisition (CLA), PPT & the switchboard metaphor:

Acquisition ≠ learning: *acquiring* a language presupposes that there's already something manifested in the child (UG), while learning may imply a blank slate (nothing there specific to language)

*A child experiences **primary linguistic data** (PLD), i.e. human speech, from her environment. The PLD feeds UG, such that the child sets the parametric switches right (e.g. as to [±MOVE-WH], [+MOVE-WH] in an English-speaking environment, [-MOVE-WH] in a Chinese-speaking one). This yields a grammar G of a language L (e.g. of English). From the acquisitionist point of view, UG is also often referred to as **Language Acquistion Device** (LAD).*



➲ **The acquisition of the Lexicon** *is* inductive learning: Given that the lexical sign is **arbitrary**, i.e. the phonological form of a word cannot be predicted from its meaning (and vice versa), words must be stored in the Lexicon **one by one** (i.e. ~50,000 words for an average adult!). A child generalises rules from the words in her as evidenced by (systematic!) *overgeneration* of regular verb forms instead of exceptional ones, e.g. **I goed* instead of *I went* (exceptions must be learned in any case!).

Evidence for innateness & modularity

Language acquistion

a. **Creoles** (Bickerton 1981): The scenario: Slaves and immigrants from different regions of the world share one workplace (originally, plantations). In order to be able to communicate, the different language communities developed an impoverished *lingua franca* – a **pidgin** – primarily based on the locally predominant language (thus, English-based, French-based,…, pidgin), with influences from the other languages present in the community. Characteristics of these pidgins include high word order variability (alternating between SVO, SOV, VSO, etc.), the absence of grammatical morphology (no Tense or agreement suffixes,…), no complex sentences (no relative clauses, no *and-/or-*conjunctions, etc.):

ex. Progressive aspect

He is walk**ing**	*English – inflectional progressive suffix (bound)*
⇩	
He walk	*Hawaiian Pidgin – no progressive marker*

Now, interestingly, within **one generation** (and not two or three,…), these pidgins have frequently been **transformed into a full-fledged language** – a **creole** (i.e. creolisation):

He is walk**ing**	*English – inflectional progressive suffix (bound)*
⇩	
He walk	*Hawaiian Pidgin – no progressive marker*
⇩	
He **stay** walk	*Hawaiian Creole – analytical progressive marker (free)*

*Moreover, different creoles formed from unrelated languages bear **uncanny resemblances** – a rather unexpected fact if the creole were formed independent of grammar-specific knowledge by more general cultural processes! The really astonishing thing is that the 2nd generation children creolised the pidgins into languages with grammatical structure despite there being no corresponding input available to them! A similar case is the following…*

b. **Nicaraguan Sign Language** (ISN: *Idioma de Señas de Nicaragua*) [➲ *film*]: ISN is a signed language spontaneously developed by **deaf children** in western Nicaragua in the 1970s and 1980s. It enabled linguists to witness the **birth of a new language**. In their homes, these deaf kids were only exposed to an impoverished **pidgin-like** signing system (like everyone of us speakers could produce), but when dozens of deaf kids were put together in a special school for deaf kids, they **creolised** their pidgins into a **full-fledged language** (e.g. 'inventing' Subject–verb agreement despite never having experienced it before).

There is a LAD pre-wired into our brain, otherwise the deaf kids couldn't have invented a new language incl. grammatical devices we know as well (Tense, agreement,…) from nothing.

c. **Critical period for language acquisition** [➲ *film*]: In the case of so-called 'wild children' (p.c. **feral children**), children miss the critical period for language acquisition (1 to 12 years of age), they simply cannot acquire language as naturally as at an earlier age any more. The feral child **Genie** (cf. Curtiss 1977), for example, was found in 1970 after 12 years of having been raised in isolation. She has never managed to attain regularly grammatical language. The best she could produce were impoverished, pidgin-like phrases like "Applesauce buy store".

The LAD is only active for the critical period of approx. 12 years, which is also reflected in our having a hard time learning foreign languages, while children can acquire ***multiple languages*** *with astonishing ease – given the appropriate input.*

Language pathology

a. Genetic disorders

❶ So-called ***savants*** (formerly, non-p.c.: *idiots savants*; *Inselbegabte* in German) have an unknown genetic disorder resulting in a 'cognitive dysbalance' – they've got deficits in one particular domain (e.g. social interaction), but are geniuses in another (e.g. language learning) (cf. the character played by Dustin Hoffman in *Rain Man*, which is based on the savant Kim Peek). One famous savant is **Daniel Tammet**, who

- experiences numbers synaesthetically, instinctively *visualising* whether a number is prime or composite (289 is particularly ugly, 333 particularly attractive)
- recited π from memory to *22,514* digits in five hours and nine minutes [➲ *doc*]
- speaks *12* languages

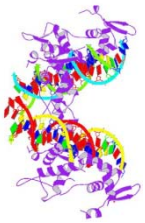
He demonstrated the latter skill by managing to acquire **Icelandic** – a linguistically really quirky European language – in only one week! While Daniel has exceptional skills in other domains than language as well, there exist also specifically **linguistic savants**, such as Christopher (Smith & Tsimpli 1995), whose linguistic abilities are exceptional.

If both Daniel and Christopher have an exceptional ability to acquire languages as if it was nothing, there must be something in their mind/brain exclusively dedicated to language (a module) enabling them to do so.

❷ **Specific language impairment (SLI)**: Only the grammar part of language is affected (not the lexical one, though), other cognitive domains are intact (emotion, etc.).

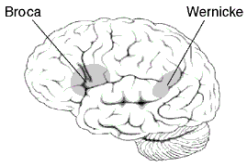
Typical SLI errors

(a) Three cookie.	(Number agreement)
(b) I go yesterday.	(Past Tense)
(c) The boy kiss the girl.	(Verb agreement)
(d) The girl is play with her doll.	(Progressive Aspect)


FOXP2 protein

The **FOXP2 gene** (the "**speech and language gene**") was discovered in 1998, when a geneticist reviewed early 1990s studies of a British family (*the KE family*; cf. Gopnik & Crago 1991), of which 15 members showed ***severe* speech/language disorders** (of the SLI type). Mutations in the gene seem to strongly affect linguistic abilities (among other non-linguistic ones). This lead supporters of the innateness hypothesis to hypothesise that FOXP2 (partially) regulates the 'language module' (i.e. syntax and morphology).

b. **Acquired disorders**



The brain appears to have two "language centres" – **Broca's and Wernicke's Areas** – located in the left hemisphere of the brain. Language impairment acquired by a damage to one of these areas (e.g. by a stroke) is referred to as **aphasia** (a selective impairment):

❶ **Broca's aphasia** (a.k.a. agrammatism) [➲ *film*]: Difficulties with grammatical stuff (inflection such as tense, function words such as auxiliaries), resulting in 'telegraphic speech' ➲ Impairment of Syntax module (morphosyntax):

> Yes... ah... Monday... er... Dad and Peter H... (his own name), and Dad.... er... hospital... and ah... Wednesday... Wednesday, nine o'clock... and oh... Thursday... ten o'clock, ah doctors... two... an' doctors... and er... teeth... yah.

Severe case of Broca's aphasia (cf. Dronkers et al. 1998) [➲ *film*]: Although he's fully capable of *understanding* the meaning of words and simple sentences, his speech is reduction to *one* word (pair): ***tono tono***. Interestingly, he's still able to **count**, the NUMBER module independent of the language module. In addition, he's able to **sing** words (MUSIC module), though only mimicking. Put simply: On the assumption that it is the SYNTAX module that is impaired, the *tono* patient is still able to produce words other than *tono* if MUSIC/NUMBER accesses the LEXICON **directly**, bypassing SYNTAX.

❷ **Wernicke's aphasia** [➲ *film*]: Difficulties with lexical stuff (words), resulting in insertion of wrong words into 'syntactic slots', thus in meaning distortion. [Just like with stuttering, patients are still able to comprehend and re-produce *songs*, knowledge of music being handled by another module than language proper.] ➲ Impairment of the Lexicon:

> I called my mother on the television and did not understand the door. It was too breakfast, but they came from far to near. My mother is not too old for me to be young.

b. **Brain damage**: In 1848, a large iron rod was driven completely through the head of the railroad construction foreman Phineas Gage. As a consequence, he became dysbalanced, egocentric and asocial – the opposite of what he'd been prior to the accident (According to his closer friends). Various scientists have speculated that the rod penetrated the module regulating Gage's **emotion/empathy**?

The structure of language: Grammar

The dual patterning language (Martinet 1957) [Fr. *double articulation* = structuring]: sentences are doubly structured

Phonology	/t, ɔ, ʊ, n, ɪ, l, a, k, s, p, ə, g, e,…/	*no meaning*[#]
Morphophonology	⇩	**1st articulation**
Morphology	<toni>, <spaghetti>, <likes>	*lexical meaning*[+]
Morphosyntax	⇩	**2nd articulation**
Syntax	[SVO Toni likes spaghetti]	*sentence meaning*[*]

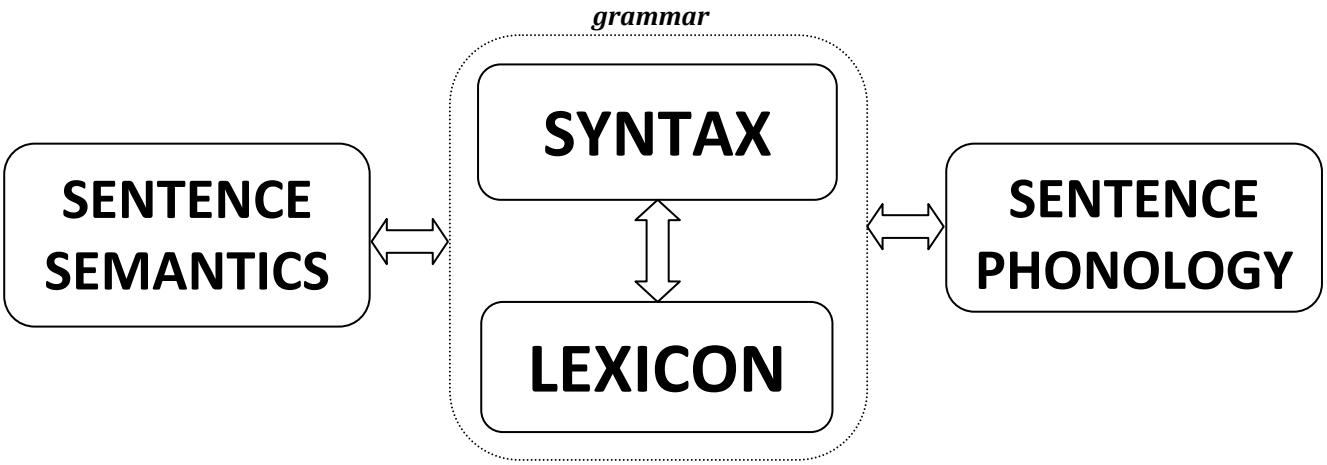
[#] The set of phonemes {/t, ɔ, ʊ, n, ɪ/} makes the word *Toni* different from the word *likes* ({/l, a, ɪ, k, s/}), differentiating their meanings. [More evident in *minimal pairs*: *dog vs. fog*, where the meaning differentiation is performed by the opposition /d/ *vs.* /f/.]

[+] *Toni* = 'an individual named *Toni*'
spaghetti = 'a non-animate, solid, edible,…, object'
like 'an emotional state of fondness of sth./so.'

[*] *Toni likes spaghetti* 'There is a state *x* such that *x* is Toni's liking of spaghetti'

➲ The double articulation is only a *(correct) descriptive observation* of how a sentence is structured, not the model of a *cognitive* process/rule! Incidentally, it may serve as a *speech perception* model (*decoding*: ① recognise phonemes → ② recognise morphemes & lex. meaning → ③ recognise sentence & its meaning), but think about *speech production* (*encoding*: ① take morphemes *incl.* phonemes/lex. meaning → ② construct sentence *incl.* sentence meaning → ③ construct phonological representation to spell out sentence)!

The modular architecture of grammar

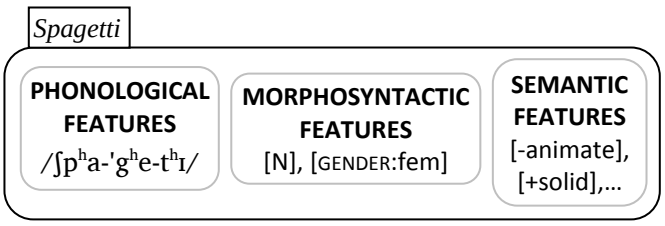


***Syntax* is the really interesting module because it mediates sound and meaning, enabling you to understand your fellow human beings, to understand what they want to say, and (sometimes unfortunately) what they want from you!**

① **Lexicon**: an inventory where *lexical items* (LIs) are stored (~50,000). A LI as a unit of the Lexicon does not necessarily correspond to a 'word', but can also be a simpler, more abstract item like the Tense morpheme <PAST> or <-*ed*> (more below). Does this still qualify as a 'word'? Vice versa, there are complex words consisting of more than one LI (e.g. *light bulb*)…

➲ LIs themselves *are no primitive things*, rather, they consist of **features** (more exactly, **feature bundles**)**: phonological/**phonemic & **semantic** features (➲ held together by an arbitrary link: the Saussurean sign). In addition – and this is a more modern finding – it **also comprises morphosyntactic**/formal features (e.g. grammatical gender in German: <u>*der*</u>$_{[masc]}$ *Tisch* 'the table' is not male, of course!)

ex.



② **Syntax**: *How* lexical items (LIs) from the Lexicon must be combined into more complex structures – 'sentences'.

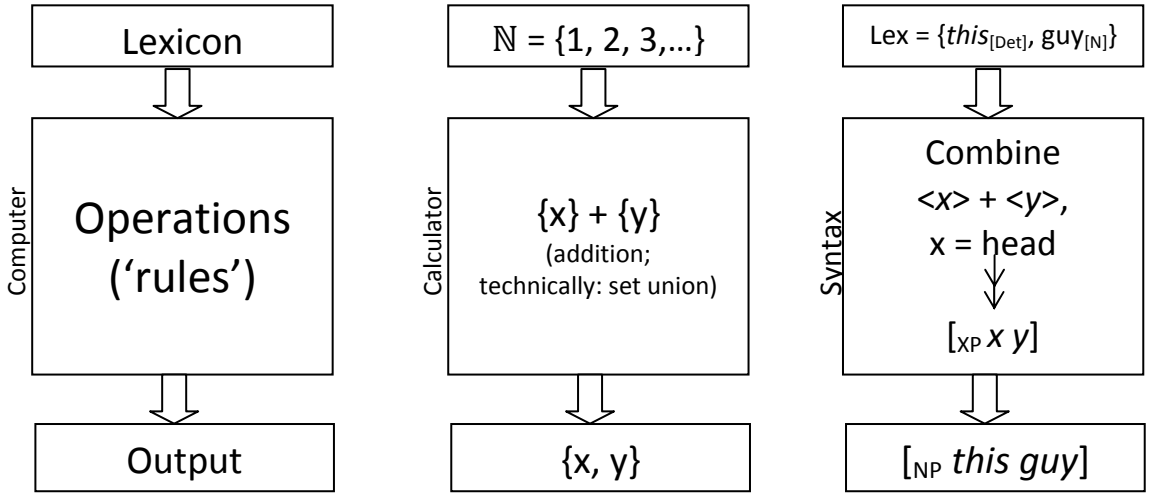
ex. Given a set of lexical items {*Toni*, *eat*, *spaghetti*}

1. Combine [$_{V}$ *eat*] & [$_{N}$ *spaghetti*] → Predicate/VP [$_{VP}$ *eat Spaghetti*]
2. Combine [$_{VP}$ *eat spaghetti*] & [$_{N}$ Toni] → 'Sentence' [$_{S}$ *Toni* [$_{VP}$ *eat spaghetti*]]
3. Establish Subj-V agreement → [$_{S}$ *Toni* [$_{VP}$ *eat*<u>*s*</u> *spaghetti*]]

③ **Sentence Phonology**: This is where *phonological representations* (i.e. phonemes, intonation, stress) are created. These phonological representations are actually *spelled out* as an utterance, or constitute the basis for decoding an utterance.

④ **Sentence Semantics**: Here, the (logical) meaning of syntactic representations are created by compositionally combining the meanings of the single LIs [Frege: 'the meaning of a sentence is a composition of the meaning of its words' ≠ the **'idea/thought'** you want to express, which is the mysterious thing you have prior to forming a sentence, which in turn is the basis for *sentence meaning* (the so-called *logical form*)].

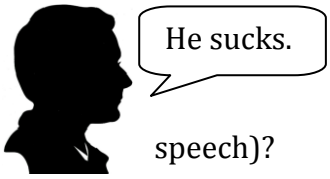
How does the grammar work?



In this seminar, we'll be dealing with the relation Lexicon–Syntax (i.e. ①+②), specifically with that portion of the Lexicon that actually happens in syntax: morphosyntax!

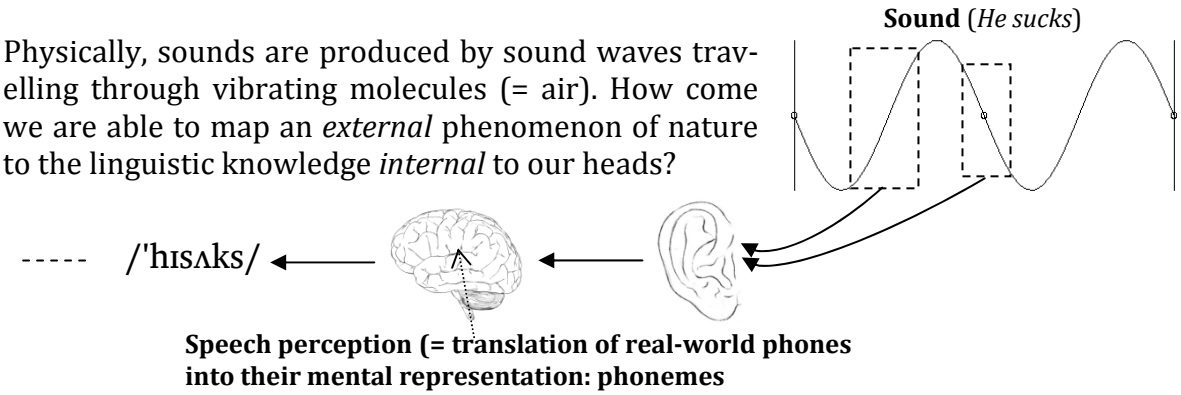
The stuff words are made of: morphology

We have a natural & internal capacity to recognise 'words'

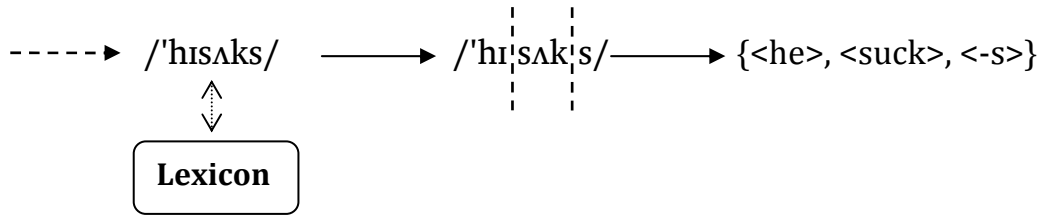


❶ Nature is full of sounds, only a fracture of them pertaining to human language: How are we able to strip this stream of chaos down to just the relevant type of sounds (i.e. human speech)?

❷ Physically, sounds are produced by sound waves travelling through vibrating molecules (= air). How come we are able to map an *external* phenomenon of nature to the linguistic knowledge *internal* to our heads?



❸ Based on lexical knowledge (the Lexicon), we then try to find 'words', i.e. LIs, and try to partition the phonemic string into *meaningful* chunks:



Ok, this gives us an idea of what *phonemes* are, but 'word' remains a rather vague term (as noted above). What *is* a word? Just a construct (like *sentence*). It's more accurate to speak of *morphemes*!

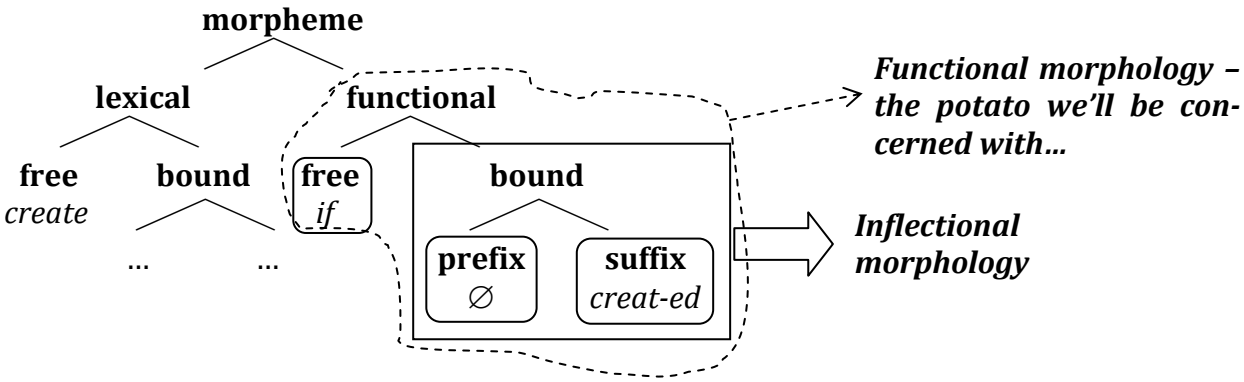
Morphemes

They are the **smallest meaningful units**, out of which 'words' are composed. There're different types of morphemes, **free** and **bound** ones. The former are called lexical **roots** (√), the latter – **affixes** – comprise **prefixes**, **suffixes**, and **infixes** (rare in English → *abso-fucking-lutely*). They can further be divided along a semantic dimension into **lexical** and **functional**.

❶ Lexical/substantive *vs*. grammatical/functional	*content/meaning*
❷ Free *vs*. bound	*morphological structure*
❸ Lexical *vs.* functional	*morphological operation*

Traditionally, ❸ is referred to as 'derivational *vs.* inflectional', but this is incomplete as we will see…

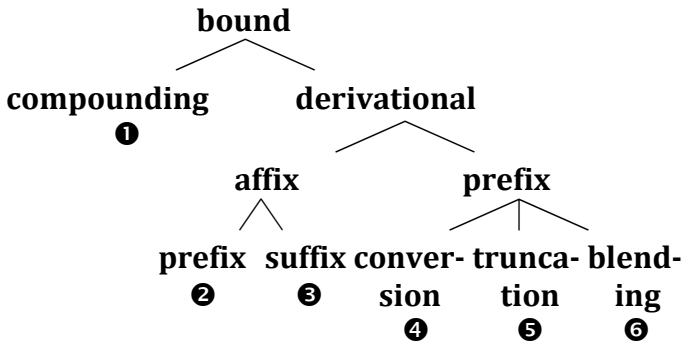
Types of morphemes



Lexical morphology *vs.* inflectional morphology

Lexical morphology happens in the Lexicon! → Unlike functional morphology, it does not care about syntactic structure, we don't care about it…!

Lexical-morphological operations: For completeness, the continuation of the lexical > 'bound branch in the morphological typology tree above is as follows:



❶ Compounding: <$_N$ skin> + <$_N$ head> → <$_N$ skinhead>

❷ Prefixation: <$_{Prefix}$ im-> + <$_A$ precise> → <$_A$ imprecise>

❸ Suffixation: <$_A$ precise> + <$_{Suffix}$ –ion> → <$_N$ precision>

❹ Conversion: change of lexical category, e.g. <$_N$ pepper> → <$_V$ pepper>

❺ Truncation: <$_N$ demonstration> → <$_N$ demo>

❻ Blending: <$_N$ motor> + <$_N$ hotel> → <$_N$ motel>

[❼ Infixation is not very producative in English, and only exists for lexical morphology: <$_{Adv}$ absolutely> + <$_{Infix}$ –fucking-> → <$_{Adv}$ absofuckinglutely>]

All morphological operations are **derivational** (technical term denoting a **step-by-step procedure**), not only 'derivational' ones as just described → all *complex* words have a **derivational history** → the *order* of morphological operations matters:

<inability>

1. adjectival root: <$_A$ √able> 1st stem = root
2. +nominalising suffix <-ity>: *ability* 2nd stem
3. +inverting prefix <in->: *inability*

1'. adjectival root: *able*
2'. +inverting prefix *in-*: **inable* *derivation cancelled*

➲ The derivation can't be right since **inable* doesn't exist (*unable* does, but, in turn, not **unability*)

You see here that a morphological derivation doesn't always run smoothly since *able+ity* doesn't yield *ableity*, but *ability*. As such syllabic tuning is influenced by stress, hence phonology, such processes are referred to as *morphophonology*.

Realisational Morphology

The discussion so far suggests the following picture: The Lexicon not only contains Lis as we know them ('words') but also inflectional morphemes like <PAST>.

dog, *cool*, *go*, *the*, **PAST**,… ⇨ *The cool dog go*+PAST

Why should this be so? The answer is that morphology is sensitive to syntactic configurations – to order. Assume add a verbal element to the sentence. To attach the <PAST> morpheme to the right verb, *syntactic structure must be inspected*: **only the leftmost verbal element is marked for Tense!**

dog, *cool*, *go*, *the*, PAST, ***have***,… ⇨ **The cool dog have **go**+**PAST*
*The cool dog **have**+**PAST** go*+PARTICIPLE

As a last step, the abstract combinations must transformed into the appropriate surface form (phonological form):

*The cool dog **have**+**PAST** go*+PARTICIPLE

⇩

<$_V$ have>+ <*Suffix* -PAST> → *had*
<$_V$ go> + <*Suffix* PARTICIPLE> → *go-*

⇩

The cool dog had gone

Inflectional morphology has characteristics of lexical morphology (it's stored in the Lexicon) *and* of syntax (it cares about syntactic configurations) → the intersection is **morphosyntax**