

# Language & Computation

## Lecture 4.1

Introduction to Language (1):

Words & word-level  
linguistic information

# Me



linguist



Associate Professor Lesley Stirling  
School of Languages and Linguistics  
Room 606, Babel Building  
x 45192  
Email: [lesleyfs@unimelb.edu.au](mailto:lesleyfs@unimelb.edu.au)

# My background

- Linguistics:
  - Syntax (grammar)
  - Semantics (meaning)
  - Discourse analysis
- Cognitive Science



# Kala Lagaw Ya



# Linguistics and computation

linguistics

+

computation

....computational linguistics?



"WHAT IS YOUR LITTLE BROTHER CRYING ABOUT?"  
"OH, 'IM—'E'S A REG'LAR COMP'TATIONAL LINGUIST, 'E IS."

# symbiosis

NLP needs linguists...



"Take me to your stove? ... You idiot!  
Give me that book!"

We gave the monkeys<sub>1</sub> the bananas<sub>2</sub>  
... because they<sub>1</sub> were hungry.  
... because they<sub>2</sub> were ripe.





...And linguists can  
learn a lot from  
doing NLP ...



**Pushing the Green Button**  
(advertisement for the  
8200 copier, c. 1983)

# What *is* linguistics?

Linguistics is the scientific study of human language in all its aspects. In particular,

- it provides a methodology for exploring the structure of particular languages, and
- it investigates what is universal to all human languages and how languages differ.



# Types of linguistic information

**phonetics** = the actual speech sounds employed in utterances  
(their articulation, physical properties & perception).

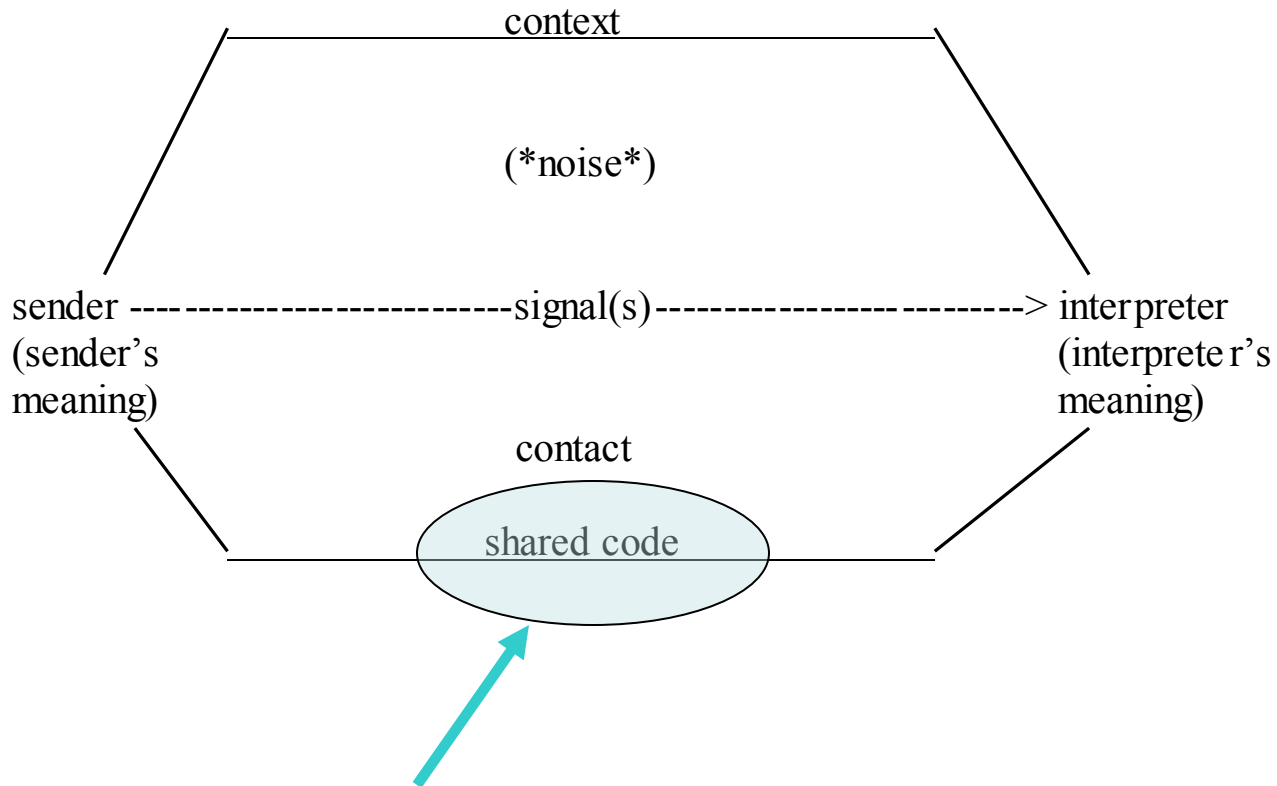
**phonology** = the distinctive sounds of a language  
and how they may combine into syllables or words.

**morphology** = the minimal recurrent meaningful units of the language, (& the lexicon) which either on their own or in combination form words.

**syntax** = how the words are combined into larger, meaningful units such as phrases and sentences.

**semantics** = the meaning of words and of larger expressions.

# What is a language?

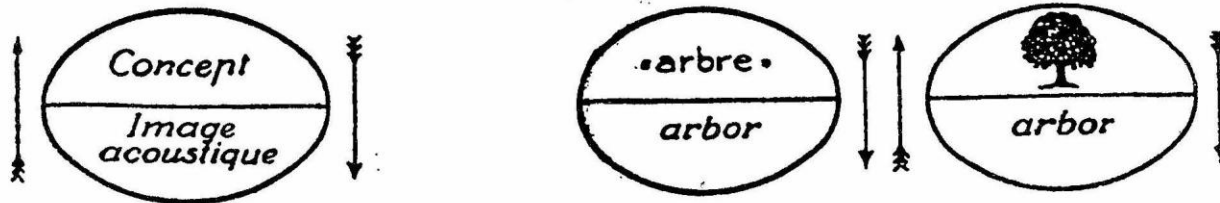


# A “sign”

A sign associates:

a meaning

a signal



# Different kinds of signs

- Most linguistic signs are ‘symbols’: they have the property of **arbitrariness**
- *Arbitrariness*—In general meanings are not predictable from forms and forms are not predictable from meaning [*whale, microorganism*].
- Cf. ‘iconic’ types of sign

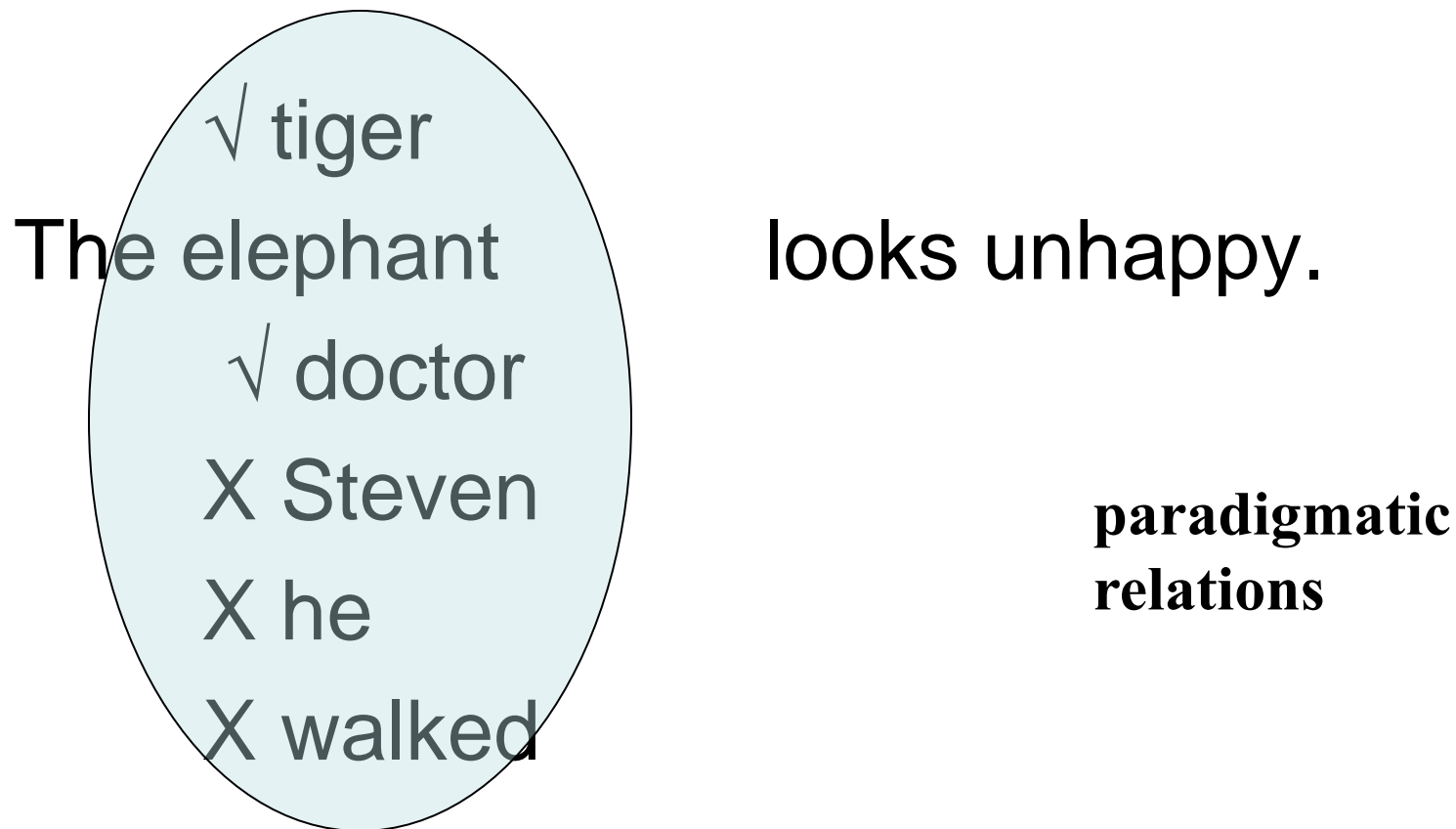
# Signs and structure

- BUT languages are not just sets of signs –
- in communicating in language, the signs are grouped together into sequences which have **structure** ('grammar').
- woman man kiss volkswagon fast  
yesterday today

# Types of structural information in grammar

The elephant looks unhappy.

# Types of structural information in grammar





# Types of structural information in grammar

The + elephant + looks + unhappy.

The + elephants + look + unhappy.

They + look + unhappy.

? The + computer + looks + unhappy.

syntagmatic  
relations

# Grammar

The structural analysis of human language can be stated in terms of:

- (1) discrete units of various kinds
- (2) rules and principles that govern the way these units can be combined and ordered

# Word

- The study of words and their structure comes under the heading of **morphology**.
- The total list of words in the language is called its lexicon.
- What do we know when we know a word?

# Some answers

- Pronunciation (phonological representation)
- Meaning (semantic representation)

**But also:**

- Word class / lexical category / part of speech

# What is lexical category?

- AKA word class, part of speech

**Noun** (a thing, an entity)

*tree, friendship, floor*

**Verb** (an action, a state)

*go, sleep, give*

**Adjective** (a property of a noun)

*red, kind, easy*

**Adverb** (a property of a verb)

*soon, easily, angrily*

**Preposition** (a relation, oft. spatial)

*in, near, with*

# Some major lexical categories

## **‘Open class’:**

Nouns

Verbs

Adjectives

Adverbs

## **‘Closed class’:**

Prepositions

Determiners / Articles

Auxiliaries

Pronouns

Conjunctions

# Words and morphemes

- The **morpheme** is the unit of analysis in studying words.
- A morpheme is the **minimal meaningful unit** in the language.
- Phonemes (distinctive sounds) are smaller but not independently meaningful,
- and morphologically complex words, phrases and sentences are meaningful but larger.



# Words and morphemes

Words can consist of:

- One morpheme (**simple**) *cat, work*
- More than one morpheme (**complex**) *cat-s, work-er*

Complex words usually consist of a **root** and one or more **affixes**

root = *cat*, affix = *-s*

# Roots and affixes

root:

- major component of word meaning
- usually longer than affixes
- often 'free' morpheme: *work*

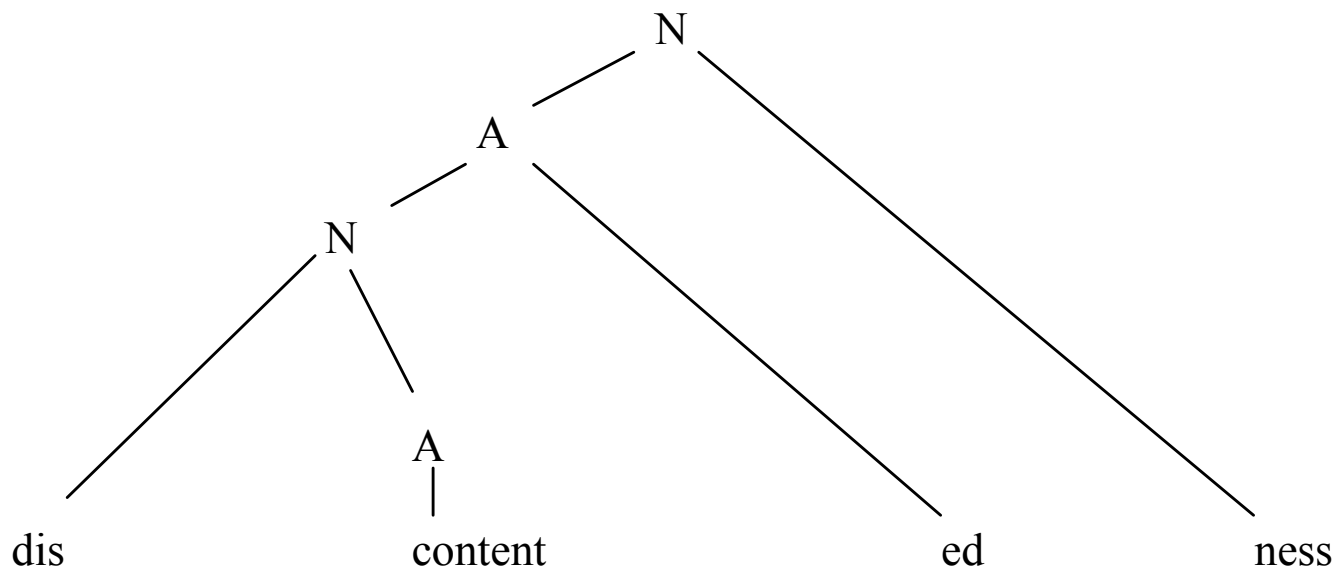
affix:

- supplementary info on word meaning
- by definition, 'bound' morpheme

# An example

[ [ [ dis- [ content ] ] -ed ] -ness ]  
N A N A

[ [ [OPP-TO [ CONTENT ] ] CHARZD-BY ] STATE OF BEING ]



Some languages have more morphemes  
per word than English allows:

*Abanyawoihwarrgahmarneganjginjeng*

‘I cooked the wrong meat for them again’  
(Gun-djeihmi; Arnhem Land)

... & some have fewer (e.g. Korean,  
Mandarin)

# Lexemes

Sets of words conceptually group together into **lexemes** ...

The **lexeme**:

- *cat, cat's, cats, catty, catlike* — 5 words, 3 lexemes
- *(I) work, (I) worked, (I'm) working, worker* — 4 words, 2 lexemes

A lexeme is a 'mental dictionary entry': independent of the particular entities it is applied to.

CAT = *cat, cats, cat's*

WORK = *work, worked, working*

# Another distinction

## Inflectional morphology

- Leaves the lexeme the same
- Does not change the part of speech
- Encodes grammatical meaning specific to that part of speech
  - *sweet*—*sweet-er*
  - *director*—*director-s*
  - *direct*—*direct-ed*

## Derivational morphology

- Changes the lexeme
- May change the part of speech (lexical category)
- Optional
  - *electric*—*electric-ian*
  - *sweet*—*sweet-en*
  - *director*—**ex**-*director*
  - *fortunate*—**un**-*fortunate*



# Lexical category again

- A lexical category is: a set of words which share a common set of linguistic (esp. morphological and syntactic) properties.

# Lexical category (word class) ambiguity

An example from Ch 5:

*They refuse to permit us to obtain the  
refuse permit.*

‘Homonyms’ = An ambiguous word (i.e. a word form corresponding to two separate lexemes), whose senses are far apart and not obviously related, other than through mere accident or coincidence e.g. *mug*, *bank* (cf. homophones: *threw/through*; *rode/rowed*; homographs: *wind/wind*).

- We need to increase productivity.
- We need an increase in productivity.
- Why do you torment me?
- Why do you leave me in torment?
- We might transfer him to another club.
- He's asked for a transfer.
- Are you some kind of pervert?
- Don't try and pervert the course of justice.

- **We need to inCREASE productivity.**
- **We need an INcrease in productivity.**
- **Why do you torMENT me?**
- **Why do you leave me in TORment?**
- **We might tranSFER him to another club.**
- **He's asked for a TRANSfer.**
- **Are you some kind of PERvert?**
- **Don't try and perVERT the course of justice.**

# Lexical category ambiguities

- British left waffles on Falkland Islands
- Lung cancer in women mushrooms
- Clinton wins on budget, but more lies ahead
- Juvenile court to try shooting defendant
- Deer kill 17,000

# Lexical category ambiguities

- British **left** **waffles** on Falkland Islands  
Is *left* a N or a V? Is *waffles* a N or a V?
- Lung cancer in women **mushrooms**
- Clinton wins on budget, but more **lies** ahead
- Juvenile court to try **shooting** defendant
- Deer **kill** 17,000

# Major ways of defining or arguing for lexical syntactic categories:

- **Distributional / syntactic / external evidence**

Where items of this class can occur within a string  
(linear order, and what they can go next to)

- **Morphological / internal evidence**

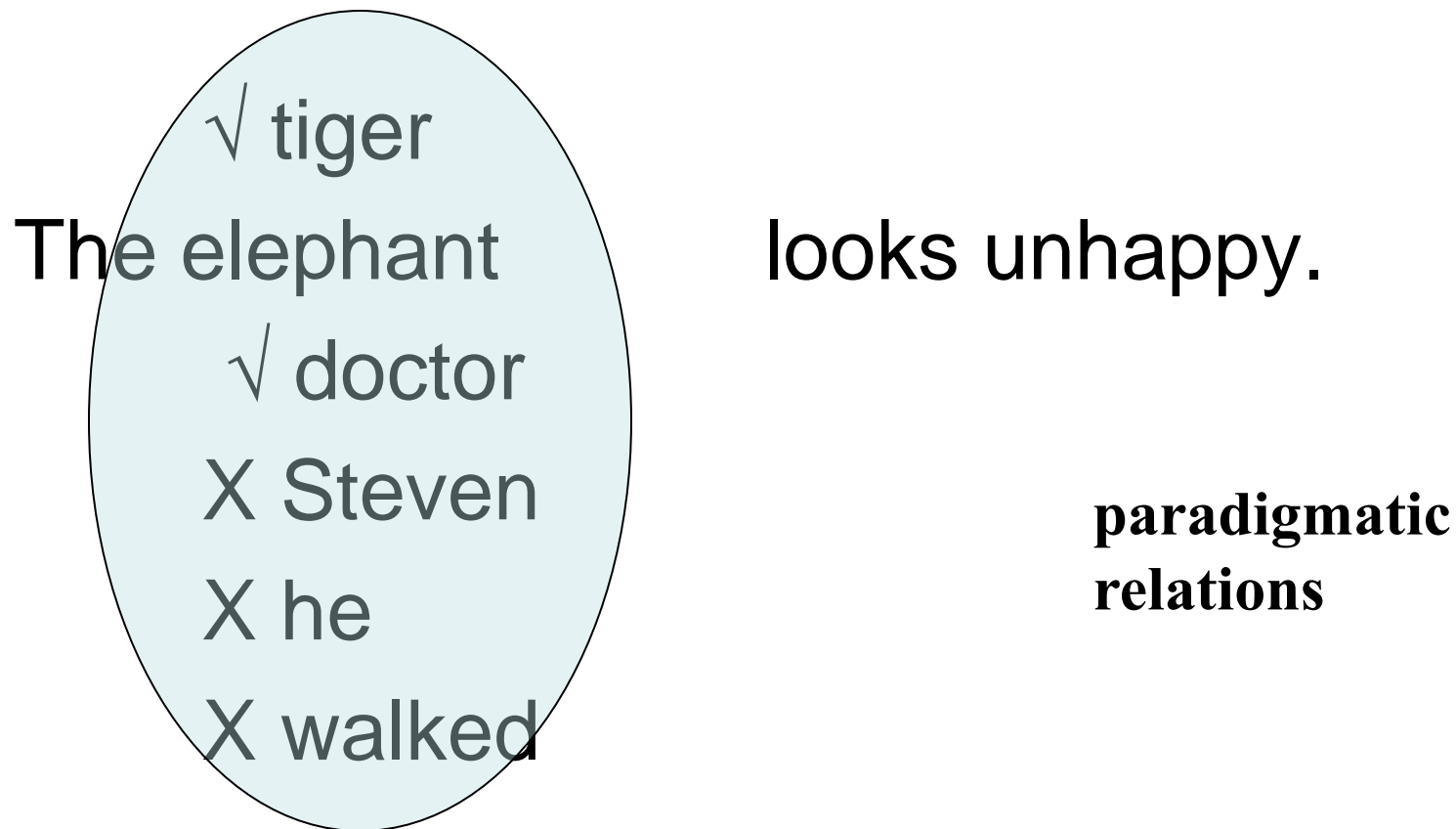
What grammatical categories and actual inflectional forms items of this class can take

e.g. – (common) nouns can usually be marked for number

Thus in English (regular) common nouns can take the –s plural ending



# Types of structural information in grammar



# Example: the class 'Noun' in English

## Distributional criteria:

- Occurs as 'head' of Noun Phrase (NP)
  - Cooccurs with **Determiners** (like *a, the, this, some*)
  - And with Numeral Determiners (*one, two ....*)
  - Cooccurs with **Adjectives** (like *red, special, clever*)
- Within its NP, occurs as subject, object etc. of verb
  - *The elephant grabbed the banana.*

## Inflectional criteria:

- plural morpheme –s (*elephant – elephants*)
- Some affixes characteristically form Nouns (*-ness, -tion, -ity*)

## Other:

- Is associated with other kinds of categories marked in other ways within the NP – e.g. Definiteness, indicated by *a/the*

# And...

Meaning: Refers to things, people, abstract concepts (i.e. not typically events or states). Eg. 'rock', 'mouse', 'truth'.

BUT generally meaning is a poor criterion for lexical category:

- *vaccination* is a noun—but an action
- *bright* is a property, and an adjective;  
*glow* is a property—and a verb
- Mismatches from language to language:  
English *I am hungry* (not *I hunger*)  
Greek verb *pino* (not adjective *pinasmenos*)

# Additional reading

- Linguistic sections of Ch 5 of your textbook, esp. 5.7
- Any introductory linguistics textbook, chapter on words and morphemes, e.g.:

Fromkin, V, R Rodman, N Hyams, P Collins, M Amberber and M Harvey. 2009. *An Introduction to Language* (sixth edition) Melbourne: Cengage Learning.

Finegan, E., Blair, D. and Collins, P. 1997 *Language. Its structure and use*. (Second edition.) Sydney: Harcourt Brace.

[This is the Oz edition: or, Finegan 1997 referenced in your text.]

O'Grady, W., M. Dobrovolsky and F. Katamba. 2004. *Contemporary Linguistics: An Introduction* . New York: Longman.

# An exercise

- How do you detect structure if you don't know the language?
- Or can you ....?
- Have a look at the Mbabarum problem for Wednesday