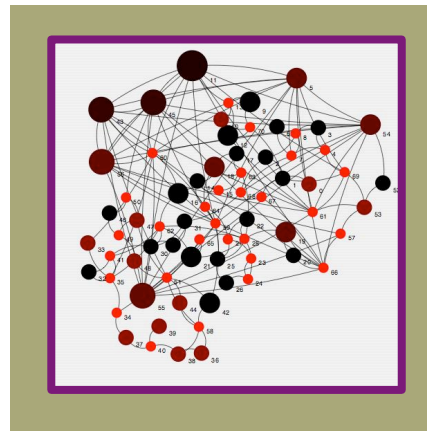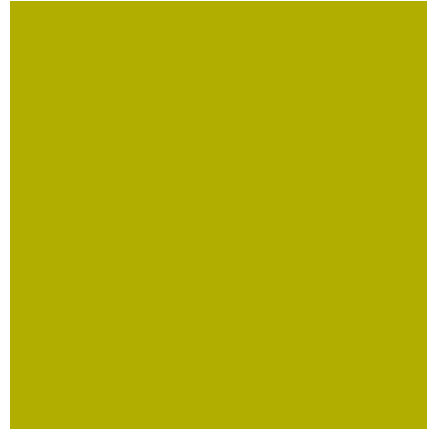# COMP40020
Human Language Technologies

Word-forms, wordforms and word forms
February 2019

Prof. Julie Berndsen
School of Computer Science

Julie.Berndsen@ucd.ie

+ HLT4

Contents:

- The domain of morphology
- Definitions and examples

Aim:

- To give a brief introduction to the most important concepts of morphology. Computational models of morphology will be discussed mainly in HLT 12.

+ Intuitions about Words

Cbnt mnstrs wr wrnd th bf, dry nd pg mt ndstry cld b "wpd t" f Brtn crshs t f th rpn nn, th Sndy ndpndnt cn rvl. Mnstrs wr ls tld n-dl Brxt wld ls "srsly jprds th vblty f lrg sctns" f th cntry's fshng ndstry.

https://www.independent.ie/business/farming/news/farming-news/cabinet-told-nodeal-brexit-will-wipe-out-beef-industry-37777237.html

+ Intuitions about Words

Cabinet minister be warn the beef, dairy and pig meat industry could be wipe out if Britain crash out of the European Union, the Sunday Independent can reveal. Minister be also tell a no-deal Brexit would also "serious jeopardise the viable of large section" of the country fishing industry.

https://www.independent.ie/business/farming/news/farming-news/cabinet-told-nodeal-brexit-will-wipe-out-beef-industry-37777237.html

+ Intuitions about Words

Cabinet ministers warned beef, dairy pig meat industry "wiped " Britain crashes European Union, Sunday Independent reveal. Ministers told -deal Brexit "seriously jeopardise viability large sections" country's fishing industry.

https://www.independent.ie/business/farming/news/farming-news/cabinet-told-nodeal-brexit-will-wipe-out-beef-industry-37777237.html

+ Intuitions about Words

> Cabinet ministers were warned the beef, dairy and pig meat industry could be "wiped out" if Britain crashes out of the European Union, the Sunday Independent can reveal. Ministers were also told a no-deal Brexit would also "seriously jeopardise the viability of large sections" of the country's fishing industry

https://www.independent.ie/business/farming/news/farming-news/cabinet-told-nodeal-brexit-will-wipe-out-beef-industry-37777237.html

+ Intuitions about Words

I have just "invented" the following English word:

mough

What can you say about this word?

- The moughs of the car are in poor condition.

What can you say about this word now? How is it pronounced?

+ Intuitions about Words

Another invented word:

<p style="text-align:center">rough</p>

What can you say about this word?

- He **roughs** every day before he goes to school

What can you say about this word now?

It is pronounced analogously to the actual English word *though*.

Are there problems with this?

+ Morphology

- <u>Morphology</u> is a subfield of linguistics which deals with the structure of words and the relationships between words → it is useful to know about this for Human Language Technologies

- Questions asked in connection with morphology:
  - What are the basic building blocks of words?
  - How are more complex words built up from simpler parts?
  - How is the meaning of a complex word related to the meaning of its parts?
  - What is the relationship between the individual words of a language and other words in the language?

+ The Word

- Most people take for granted that languages contain words. Even illiterate speakers know that their language consists of words.

- Sometimes people do not agree as to what units are to be treated as words. For instance, English speakers might not agree as to whether all right is one word or two

$\rightarrow$ alright or all right

+ The Word

- People can easily recognise words of their language when they see or hear them and native speaker judgements as to what is or is not a word typically coincide.

- For example, English speakers agree that the form brillig in the sentence

   The brillig cat sat on the mat

   is not a word of English - but all the other forms are.

A word is an arbitrary pairing of sound and meaning

+ Morphemes and Morphs

- The term morpheme is used to refer to the smallest unit of <u>meaning</u> or <u>grammatical function</u> which make up words:

  - -er e.g. play-er, teach-er
  - -ness e.g. kind-ness, mad-ness
  - -less e.g. hope-less, help-less
  - ex- e.g. ex-minister, ex-husband
  - pre- e.g. pre-dinner, pre-lecture
  - mis- e.g. mis-conduct, mis-manage

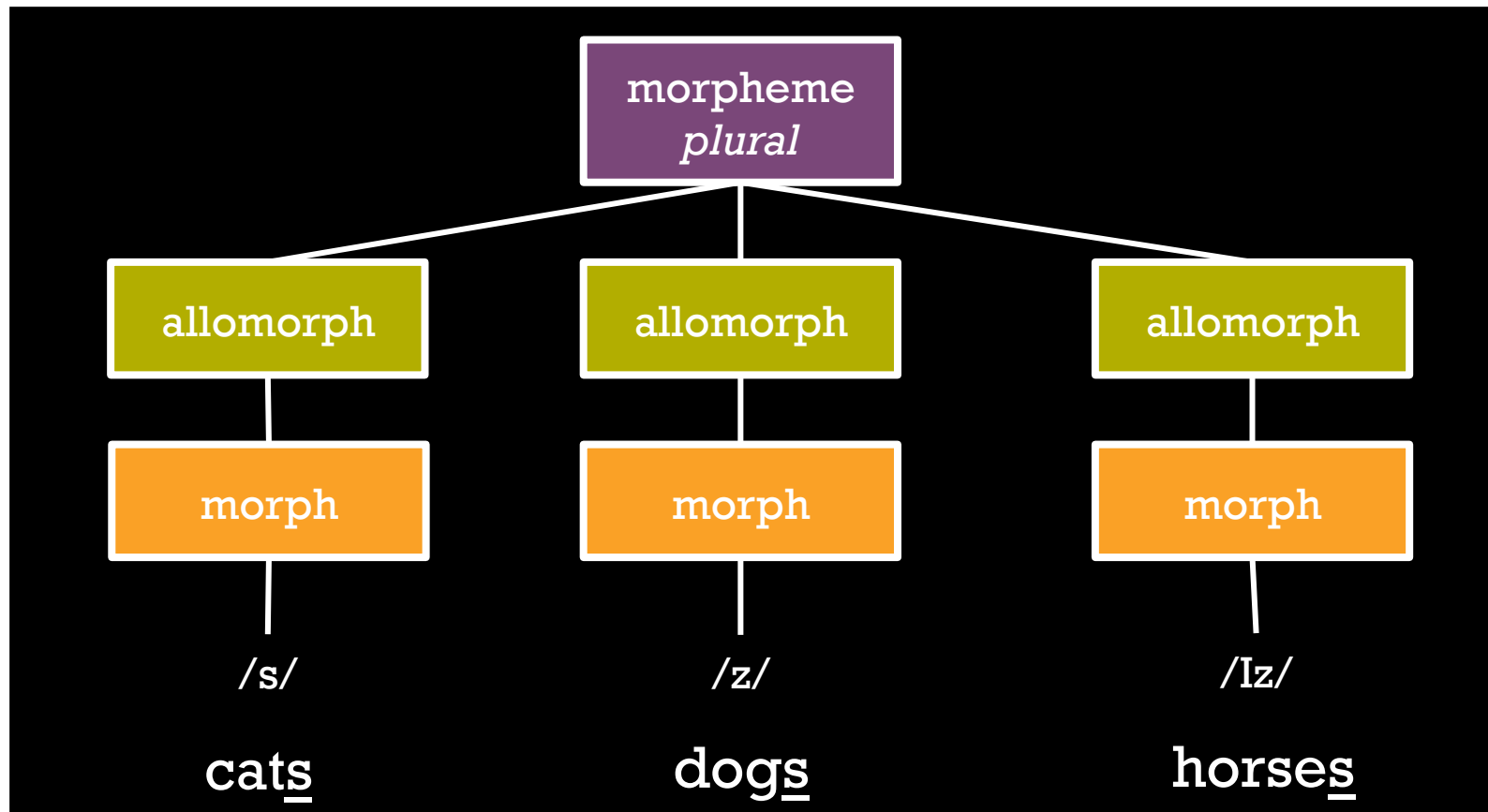# + Morphemes and Morphs

- Words can be analysed into morphemes by the identification of morphs.

- A morph is a physical form representing some morpheme in a language namely a distinctive sound (→ phoneme) or sequence of sounds (→ phonemes).

  - cats
  - dogs
  - horses

+ Morphemes and Morphs

- If different morphs represent the same morpheme, they are grouped together and they are called allomorphs of that morpheme.

```
                    morpheme
                     plural

    allomorph       allomorph       allomorph

      morph           morph           morph

       /s/             /z/            /Iz/

      cats            dogs           horses
```

+ Morphemes and Morphs

What are the morphs?

- He parks the car.
- I park the car.
- We parked the cars
- They mended the boats.
- She speaks slowly.
- He draws roses.

+ Lexemes

- A lexeme is the 'word' in the sense of abstract vocabulary item which can have different realisations.

- Which of the words below belong to the same lexeme?

see	catches	taller	boy	catching
sees	sleeps	woman	catch	saw
tallest	sleeping	boys	sleep	seen
tall	jumped	caught	seeing	jump
women	slept	jumps	jumping	

+ Lexemes

- Word-form:
  - a particular physical realisation of a lexeme
  - → wordform word-form word form

- Grammatical Word:
  - the representation of a lexeme with certain morpho-syntactic properties
    - Usually I cut the bread on the table.
    - Yesterday I cut the bread in the sink.
    - → two grammatical words

+ Word Building

- Word building processes:
  - Inflection
  - Derivation

- Inflectional Morphemes:
  - form new words using affixes which are determined by grammatical information (*see → see*s; *dog → dog*s)

- Derivational Morphemes:
  - form new words by either
    - changing the meaning of the form to which they are attached (*kind → unkind*).
    - by changing the word-class that a form belongs to (*kind → kindly*; adjective → adverb).

+ Word Building

- Are the affixes in these words inflectional or derivational?

  - redo
  - unwind
  - manageable
  - eating
  - winner
  - treats
  - modernise

+ Roots & Stems

- Root
  - A root is the irreducible core of a word, with absolutely nothing else attached to it

- Stem
  - The stem is the part of a word that is in existence before any *inflectional* affixes (i.e. those affixes whose presence is required by the syntax such as markers of singular and plural number in nouns, tense in verbs etc.) have been added.

Noun Stem	Plural
cat	-s
worker	-s

+ Roots & Stems

- Many words contain a root standing on its own. Roots which are capable of standing independently are called free morphemes
  - man, book, tea, sweet, pain
  - cook, bet, very, elephant, walk

- These free morphemes are examples of lexical morphemes (nouns, verbs, adjectives, prepositions or adverbs).

  → carry most of the semantic content of utterances.

+ Roots & Stems

- Many other free morphemes are function words. The function words mainly signal grammatical information or logical relations in a sentence.

    - *articles*: a, the

    - *demonstratives*: this, that, these, those

    - *pronouns*: I, you, we, they, them; my, your, his, hers; who, whom, which, whose etc.

    - *conjunctions*: and, yet, if, but, however, or etc.

+ Roots & Stems

- While only roots can be free morphemes, not all roots are free. Many roots are incapable of occurring in isolation. Such roots are called bound morphemes as e.g. Latinate roots

- -mit ('send, do') as in permit, remit, commit, admit

- -ceive ('take') as in perceive, receive, conceive

- pred- ('prey') as in predator, predatory, predation, depredate

- sed- ('sit') as in sedan, sedate, sedent, sedentary, sediment

UCD
DUBLIN

+ Affixes

- An affix is a morpheme which only occurs when attached to some other morpheme or morphemes such as a root or stem.

- A prefix is an affix attached before a root or stem:
  - re-make, re-read
  - un-kind, un-tidy
  - in-decent, in-accurate

- A suffix is an affix attached after a root or stem:
  - kind-ly, quick-ly
  - wait-er, play-er
  - book-s, mat-s
  - walk-ed, jump-ed

+ Affixes

- An infix is an affix inserted into the root itself.

- Infixes are very common in languages like Arabic and Hebrew but rare in English; however cf. some "contemporary" English examples from Zwicky & Pullum (1987) and Bauer (1983)

  - Kalamazoo   Kalama-goddam-zoo
  - instantiate   in-fuckin-stantiate
  - kangaroo   kanga-bloody-roo
  - impossible   in-fuckin-possible
  - guarantee   guaran-friggin-tee

+ Distribution and Conditioning

- Distribution:
  - impossible, impatient, immovable
  - intolerant, indecent, intangible, inactive, inelegance
  - incomplete, incompatible, ingratitude

- Conditioning:
  - Phonological Conditioning
  - Grammatical Conditioning
  - Lexical Conditioning
  - Suppletion

+ "New" Words

■ Lexicon and Productivity

- necessity for listability of words as opposed to sentences
- lexicon cannot be seen as a static list
  - neologisms
  - borrowings

→ open-endedness of the lexicon

+ "New" Words

- **Blendings**: when a word is formed by combining and contracting two other forms.
  - smog, chunnel

- **Acronyms**: when words are formed from the initial letters of a phrase.
  - NATO, BBC, deejay

- **Clippings**: when two or more syllable words are shortened.
  - advertisement, examination, gymnasium, laboratory

+ "New" Words