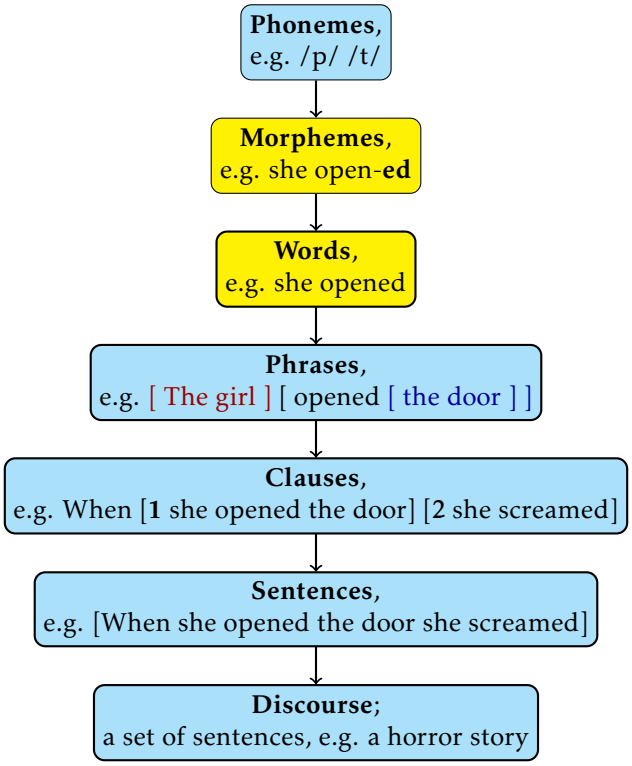
1

The building blocks - words and morphemes

1.1 Introduction

Words and morphemes are the basic building blocks of languages. Figure 1.1 shows where they come in the linguistic hierarchy.

Figure 1.1: The linguistic hierarchy



Words and morphemes can be regarded as the smallest building blocks of syntax, as we put them together to make sentences.

1.2 Definitions of words and morphemes

At first glance it seems very easy to define what a **word** is. A simple definition is that words are separated by spaces when written on the page. But how about a word like *blackbird*? Is that one word or two? We could conceivably write it as *black bird*, so according to our simple definition it can also be two words. While it is difficult to arrive at a complete definition of a word we can identify a couple of important characteristics of words;

1. They can often stand alone (or almost alone) e.g. What did you see? A blackbird

2. They cannot be split e.g. I *saw a black and a white bird* doesn't mean I saw a *blackbird* and a white bird. This is because the meaning of *blackbird*, as in a subspecies of bird, is more than a combination of the meanings of *black* + *bird*.

So a good definition of a word is **the smallest unsplittable unit**. The blackbird example demonstrates that in the case of compound words splitting is bad because we lose the original meaning. The definition of a **morpheme** is slightly more straightforward. A morpheme is **the smallest unit of meaning**. For example, in the word *laugh-ed* there are two units of meaning, one which describes the event, and one which describes when that event took place. However, there are also some cases when it is difficult to decide whether a string of phonemes is a morpheme. An example of this is shown below, but the faint-hearted can skip this!

Food for thought

How many morphemes in *raspberry*? *Berry* is clearly a morpheme referring to the object BERRY. But what does *rasp* mean? An expert berry picker would know that the *rasp* is the green spike left behind when you pick a raspberry (this helps us to distinguish raspberries from underripe blackberries). At the other end of the spectrum a young child may use the word *raspberry* without realising the meaning of *berry*. Most of us lie somewhere in the middle of this spectrum. We probably know that *rasp* specifies the type of berry, but we don't know what it means. So, given that a morpheme is a unit of meaning, an expert berry picker will store it as two morphemes, and a young child may store it as one morpheme. Most of us lie somewhere in the middle, as we recognise *rasp* as a separate morpheme, but do not know what it means. Consequently, it is actually quite difficult to decide whether *rasp* is a morpheme, i.e. a unit of meaning.

1.3 Word Classes

Words are often categorised according to their **word classes**. These are sometimes called **syntactic classes**, **syntactic categories**, and at school you probably referred to them as **parts of speech**. Common word classes are Noun, Verb, Adjective and Adverb. At school you probably defined word classes via their meaning, e.g. "A noun is a doing word, a verb is an action word, an adjective describes a property." These capture some essential truths, but they are not true in all instances. For example, the noun *hide-and-seek* in *we played hide and seek* refers to an activity, so semantically is more verb-like. And the verb *sucks* in *this film sucks* describes a property, so semantically is more adjective-like. Another way to describe word-classes is to look at their form. For example, adjectives often end in *-y*, e.g. *scary*, adverbs can end in *-ly*, e.g. *scarily*, and verbs can end in *-ing*, e.g. *she is laughing* (N.B. these word endings are all examples of morphemes). Nonetheless, these correspondences do not always hold true. For example, -ing, can also be used for adjectives, e.g. *a raging torrent*, and nouns, e.g. *a swimming pool*.

A final way to define word classes is according to position or "distribution". For example, almost all adjectives come in front of the noun, e.g. *a scary film*. Verbs come immediately in front of a noun, e.g.

(1) I really like pizza

(2) *I like really pizza (* *shows that a sentence is ungrammatical)*

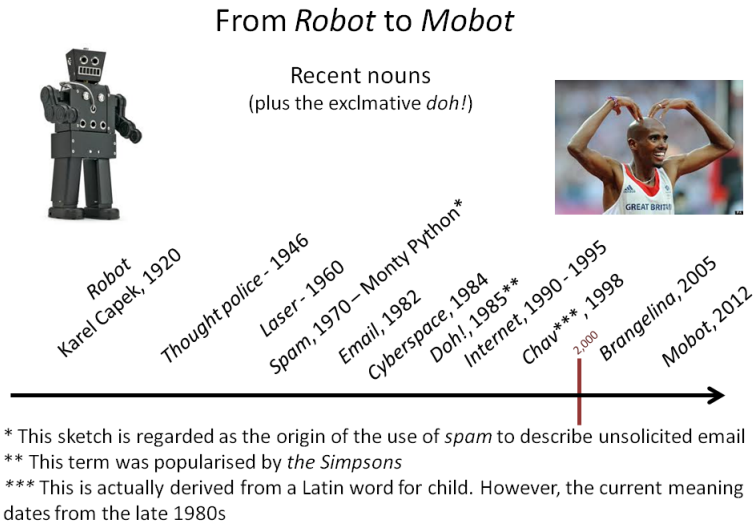
So we have 3 tests of identifying word classes;

1. Meaning
2. Form
3. Distribution

The last test is in fact the most reliable one, but it is also quite difficult to do.

Open and closed-class word classes

Word classes can be divided into open and closed-classes. An **open class** is open to new members, while a **closed class** is closed to new members. Verbs nouns and adjectives are examples of open word classes. The following figure shows recent English nouns.



Examples of closed classes are determiners (*the*, *a*), prepositions (*in*, *on*), pronouns (*he*, *she*, *it*), and auxiliary verbs (*is* in the sentence *She is laughing*).

Open classes contain many members, while closed classes contain very few members. For example, while there are hundreds of thousands of nouns in the English language, there are only a handful of determiners, e.g. *a*, *the*, *this*, *these*.

Open class words are also called **content words** because they express the "content" of the utterance. Closed class words are often called **function words** as they have a syntactic function, but do not express content. For example, if we remove the content words, a sentence is still often interpretable;

(3) The large grey cat pounced on the terrified mouse

(4) Large grey cat pounced terrified mouse

Because open class words carry important content, they are often stressed, while closed class words are often unstressed.

The following table summarizes the difference between open- and closed-class items;

Table 1.1: Summary of the differences between open and closed word classes

Open-class	Closed-class
Admit new members	Do not admit new members
Are numerous	Are few
Express "content"	Do not express "content"
Carry stress	Rarely carry stress
Can be long, e.g. *antidisestablishmentarianism*	Tend to be short, e.g. *the*, *on*

Food for thought
Prepositions, e.g. *in* the box, are generally regarded as closed-class items. However, they are slightly different to other closed class items in that they express "content", e.g. the location of an object. They are also slightly more open to new members. For example, some prepositions, e.g. be**hind**, be**fore**, actually contain open-class words, e.g. the **fore** / **hind** quarters of an animal. This is really obvious for multiword prepositions, e.g. in **front** of. So clearly some prepositions have been created relatively recently using open-class words. This example demonstrates that the distinction between open- and closed-class items is actually blurred, and sometimes open-class words can become closed-class words. This process is called **grammaticalisation**.

1.4 Morphemes

Words can consist of one morpheme, e.g. *dog*, *laugh*, or a number of morphemes, e.g. *dog-s*, *laugh-ed*, *re-consider-ed*. In other words they can be **monomorphemic** or **multimorphemic**. In multimorphemic words there is one morpheme which makes the greatest contribution to meaning; *dog*, *laugh and consider*. These are the **root morphemes**. In English, root morphemes are often grammatical when they stand alone. The morphemes which cannot stand alone; *-s*, *-ed*, *re-*, are called **affixes**. If an affix appears before the word, e.g. *re-*, it is a **prefix**, and if it comes after the word, e.g. *-ed*, it is a **suffix**.

Affixes are **bound morphemes** as they cannot be separated from the root, and cannot exist without the root. Morphemes which are not bound are said to be **free-standing morphemes**.

Another term you might hear is the **stem**. Most of the time the terms root and stem mean the same thing. However, stem is often used to describe a unit that's left when we strip away a single affix, but is not a root, e.g.

un-	break	-able
	root	suffix
prefix	stem	

Note that *-un* is the prefix of *break-able*, but *-able* is not the suffix of *un-break*, as the latter is not a word.

Inflectional and derivational morphemes

There are two types of affixes; **inflectional morphemes** and **derivational morphemes**. Here is a table summarising the differences;

Table 1.2: Summary of the differences between inflectional and derivational morphemes

Inflectional morphemes	Derivational morphemes
make little difference to the meaning of a word	make a big difference to the meaning of a word
do not change the word class of a word	can change the word class of a word
are found across a variety of languages	are fairly language-specific
can only be a suffix in English	can be either a prefix or suffix in English
participate in syntactic processes	do not participate in syntactic processes
are fully productive	are only partially productive

To help you remember the differences, **derivational** morphemes are used to **derive** new words, while **inflectional** morphemes merely modify existing words. In *re-consider-ed*, *re-* makes a big change to the meaning (it means *to do something again*), while *-ed* merely specifies when something took place. *Re-* is derivational, while *-ed* is inflectional. In *care-less*, *-less* makes a big change to the meaning (it means *without*), and also changes the noun *care* into the adjective *careless*, so it is clearly derivational.

With regard to cross-linguistic differences, many languages have a version of the past tense *-ed*, or plural *-s*. However, derivational morphemes do not translate quite so well. For example, English *out*, meaning *to do something better*, e.g. *they out-played the opposition*, has no equivalent in a number of European languages. And languages such as French and Spanish have a diminutive suffix, e.g. *perr-ito = little dog*, which is not found in English.

Inflectional morphemes participate in syntactic processes. For example, if we use plural *-s* on a subject noun, we have to change the verb so that it agrees with the noun;

(5) My dog_ is wonderful

(6) Dog-s are wonderful

This process is syntactic in the sense that it operates across words. And past tense *-ed* is syntactic in the sense that it expresses tense which is a property required by sentences in order to be grammatical.

Finally, inflectional morphemes are fully productive. For example, any verb may take *-ing*, and any regular or novel verb may take *-ed*. All regular nouns can take *-s*. By contrast derivational morphemes are only partially productive. For example while *y* can be added to most nouns to make an adjective, e.g. *horse* → *horsey*, *jazz* → *jazzy*, it cannot be added to *book* as there is already an adjective *bookish*. While *de-* can be added to many verbs, e.g. *defrost*, it cannot be added to *do* (as we have *un-do*), or *agree* (as we have *disagree*).

A typology of morphemes

1.2 shows a typology of morphemes differentiating between free-standing and bound morphemes.

Food for thought
Some morphemes are pronounced differently in different contexts. For example, past tense *-ed*, can be pronounced /d/ if it comes a voiced consonant, e.g. *robbed*, /Id/ if it comes after a vowel, e.g. *headed*, and /t/ if it comes after an unvoiced consonant, e.g. *laughed*. Plural -s also changes in a similar manner (compare *dogs* and *rabbits*). Such phonological changes only characterise **inflectional** morphemes, and do not occur in derivational morphemes. This is another example of a phenomenon which can span subdisciplines; syntax and phonology. Linguists refer to the topic of *morphophonology*, i.e. the phonological realisation of morphemes.

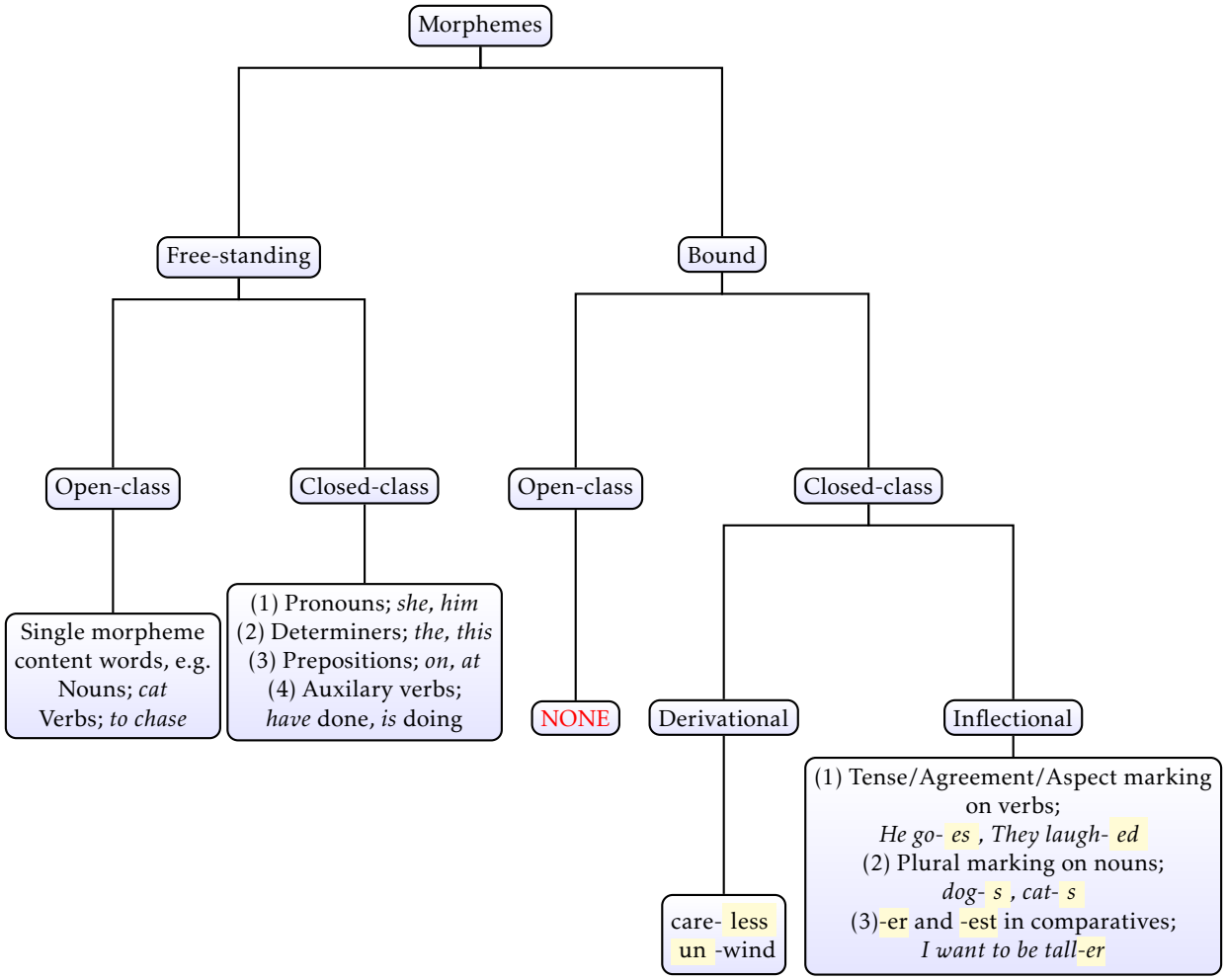
Derivation and word formation

Derivation is often used to coin new words. For example, you may not have heard the word *de-hyphenate*, but you can probably guess that it means to remove the hyphen between words. Another way to coin new words is **compounding**, which is literally the combination of two existing words, e.g.

(7) I don't like cats. I'm more of a **dog man**.

Here we combine the nouns *dog* and *man* to create a new **compound noun**, a *dog man*, i.e. a man who likes dogs. Compounding is a very productive process in English, e.g. *front door*, *book shelf*, *carpet cleaner*, *action film*, *toothpaste*, *table tennis* etc. Both derivation and compounding are examples of word formation processes, which are separate to inflectional processes, as shown in 1.3

Figure 1.2: A typology of morphemes



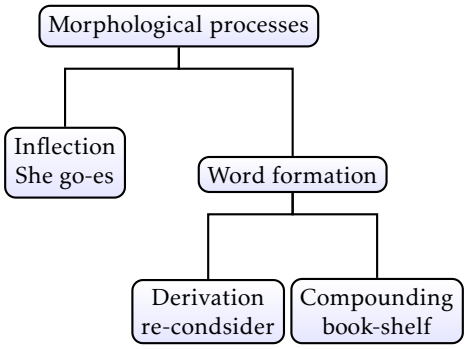
Counting morphemes

Many research papers on language difficulties in both adults and children use a measure called **mean length of utterance** (**MLU**), which is used to determine an individual's approximate language level. To calculate this we take the total number of morphemes in series of utterances and divide them by the number of utterances;

$$\frac{Number\ of\ morphemes}{Number\ of\ utterances}$$

When we do this we normally only count the morphemes which an individual is **using productively**. We therefore count inflectional morphemes, as they are productive. Derivational morphemes are often not used productively, so we don't tend to count these. Compounds are also often unproductive. For example, while *bookshelf* is a shelf for books, we probably do not generate it on the spot, otherwise we would say things like *flower-shelf* or *photo-shelf* (shelves carrying flowers and photos). So when calculating MLU we have to make a judgment call as to whether a compound noun is being used productively. If we think the compound noun has been learnt as a whole, we would only count one morpheme.

Figure 1.3: Morphological processes



Useful Terms

Morpheme, word class (or syntactic class / syntactic category), open class, closed class, content words, function words, monomorphemic, multimorphemic, root morpheme, affix/stem (prefixes and suffixes), bound morpheme, free-standing morpheme, inflectional/derivational morphemes, compounding, Mean Length of Utterance (MLU)