

II Words and Morphology

The term ‘morphology’ means ‘the study of forms’, from Greek *morphos* (μορφοϛ) ‘form’ and *logos* (λογοϛ) ‘study’. Studying forms is something one can do in the context of many branches of science, not just linguistics. Indeed, there are many scientific fields that feature a subdiscipline referred to as ‘morphology’. ‘In linguistics morphology refers to the mental system involved in word formation or to the branch of linguistics that deals with words, their internal structure, and how they are formed’, as Aronoff & Fudeman (2005:1–2) put it in their introduction to linguistic morphology (hereinafter just ‘morphology’).

II.1 The word and its complexity

The key concept in morphology is the WORD. Intuitively, we have a sense of what is or isn’t a word. For instance, by general consensus *Budapest* and *Hódmezővásárhely* (the names of two towns in Hungary) are single words. *Buda*, *Pest*, *hód* ‘beaver’, *mező* ‘field’, *vásár* ‘market’ and *hely* ‘place’ are also words of Hungarian. So words can themselves consist of other words. Words can get very long; in fact, there is no inherent limit on how long a word can get. So there is no such thing as ‘the longest word of language *L*’. Though *antidisestablishmentarianism* is sometimes cited as ‘the longest word of English’, and Mary Poppins coined *supercalifragilisticexpialidocious* as ‘the biggest word you’ve ever heard’, they can both be made longer still — for instance, by adding *-ness* to the latter, thereby ‘nominalising’ it. Once it has been nominalised with the help of *-ness*, there is not much we can do to make the word even longer — but if we can imagine different types of *supercalifragilistic-expialidociousness*, then we can add the marker *-es* to make it plural, thereby making the word even longer. These two additional markers (*-ness* and plural *-(e)s*) are among a large family of ingredients smaller than words out of which complex words can be constructed, via the morphologically processes of DERIVATION (which substantively changes the host word, for instance by changing its category, as in the case of *-ness*, which attaches to an adjective and turns it into a noun) and INFLECTION (which changes the host word’s syntactic distribution but otherwise leaves it untouched).

Inside the long word *antidisestablishmentarianism*, we easily recognise several smaller-than-the-word ingredients: *anti-* (also found, e.g., in the word *antidote*), *dis-* (as in *disengagement*), *-ment* (as in *disengagement*), *-ary* (found in *documentary*), *-an* (seen in *documentarian*), and *-ism* (rearing its head in words such as *socialism*). Occasionally, we find the same smaller-than-the-word element showing up multiple times in a single complex word — an indication that morphological derivation exhibits what we call RECURSION, the ability to embed a structure of a particular type inside a larger structure of the same type. We see recursion in a rather striking way in Hungarian words such as *jóságosság* and *igazságosság*, derived from the simple adjectives *jó* ‘good’ and *igaz* ‘true’ by first adding one instance of the nominalising element *-ság* ‘-ness’, subsequently turning the resulting noun (*jóság* ‘goodness’, *igazság* ‘truth’) back into an adjective by adding the marker *-os*, and finally nominalising *jóságos* ‘kindly, warm-hearted’ and *igazságos* ‘fair, equitable’ by adding another token of *-ság*. Logically speaking, we should be able to keep repeating the exercise (forming *jóságosságos* and then *jóságosságosságos*, etc. etc.) — but in practice, there would be precious little point in doing so. Still, the naked fact that such complex words as Hungarian *jóságosság* and *igazságosság* exist illustrates within morphology a fundamental property of natural language: recursion.

All the complex words that we have been talking about so far are built on a ROOT, a core that cannot itself be further decomposed. For *jóságosság*, the root is the short and sweet adjective *jó* ‘good’; for *antidisestablishmentarianism*, the root is *establish* — itself quite a large word, but not one that can be further teased apart into component parts that distribute independently. (Although English productively attaches *-ish* to nouns such as *style* to form *stylish* ‘having style’ and to adjectives such as *bad* to form *baddish* ‘somewhat bad’, it should be clear that the last three letters of *establish* do not instantiate either of these uses of *-ish*.) The pieces that are glued on to the left and right of the host word are dependent pieces of complex words: these can build words bigger but cannot stand on their own two feet — they are BOUND forms. Dependent pieces of words that need a host to bind themselves to are called AFFIXES. There are two major subtypes of these: PREFIXES (which stand to the left of their host) and SUFFIXES (which follow their hosts).

A root and an affix together form a STEM, which can be subjected to further affixation. Thus, the root *true* and the nominalising suffix *-th* can form the stem *truth*, to which we can attach either the inflectional plural suffix *-(e)s*, to form *truths*, or the derivational suffix *-ful*, forming *truthful*. To *truths*, we can apply no further morphological operations: this is a full-fledged WORD. But *truthful* can serve as a stem that is input to another round of derivation, this time involving the nominaliser *-ness*, with *truthfulness* as the result. Since *truthfulness*, like *truth*, is a noun, and since nouns can be pluralised, *truthfulness* is eligible in principle for inflection with the suffix *-(e)s*: though we may not often have occasion to use this word, *truthfulnesses* is a well-formed complex word.

Roots and stems can often stand on their own two feet, not requiring the addition of some prefix or suffix to form a well-formed word. This is true, for instance, for English *truth* and for Hungarian *jó*. But now compare *jó* to the form *jav-*, found in many words that are historically related to *jó*, such as the verbs *javít* ‘to improve, repair, correct’ and *javasol* ‘to suggest’. The form *jav-* is a stem (derived from *jó*) that never occurs without having anything attached to it — a BOUND STEM. Roots can also be bound. For English, we can illustrate this with words such as *receive* and *submit*: clearly, these are complex words, formed by attaching the prefixes *re-* (also found in *replay*, *retry*, *return*, *reintroduce*, etc.) and *sub-* (as in *sublet*, *submerge*, *subsystem*) to an underived host, a root; this root, *-ceive* or *-mit*, occurs in combination with other prefixes (*conceive*, *deceive*, *perceive*; *commit*, *permit*, *transmit*) but does not occur on its own, so it is a BOUND ROOT. Neither *-ceive* nor *-mit* is indigenous to English: these bound roots ultimately come from Latin. In the source language, *-cipiō* is itself already a bound root (with *cipiō* ‘take’ being the free form); but the Latin ancestor of English *-mit*, i.e., *mittō* ‘put’, was a free root. Freedom and boundness are not necessarily ‘inherited’ when a root from one language makes its way into a different one; nor is boundness specific to words of foreign origin. That bound roots and stems can belong to the indigenous vocabulary of a language is clear from Hungarian *v*-stems (*jav-*) and from English words such as *uncouth* ‘strange, crude’ and *unkempt* (the forms *couth* and *kempt* are, at best, only usable on their own in jest).

All words have a root; but in view of the existence of bound roots, it is not the case that every root is itself a word. A word is a MINIMAL FREE FORM (where ‘free’ is the antonym of ‘bound’). By ‘minimal’ (or ‘smallest’), the influential early 20th-century American structuralist Leonard Bloomfield (1933:178) understood something smaller than a syntactic phrase. So here we see a definition of the word couched in part in the terminology used by a different subdiscipline of linguistics: syntax (see segment III of this series). The word is a free form (i.e., something that ‘can stand on its own two feet’) that is not as large as a phrase. To see the difference between words and phrases, it may be helpful to consider the contrast between *John is a designer of fashion* and *John is a fashion designer*. In the former, *fashion* forms a phrase that can be replaced with *what* to form a question:

What is John a designer of?. But in the latter, this is impossible: **What is John a designer?* Based on this criterion, we call *designer of fashion* a phrase (which in turn contains *fashion* as another phrase, replaceable with *what*) while we call *fashion designer* a word — albeit a complex one whose component parts, *fashion* and *designer*, are themselves words in their own right. The form *fashion designer* is what we call a COMPOUND WORD (more precisely, a SYNTHETIC COMPOUND: one of the terms of the compound *fashion designer* is itself a derived form, *designer*; in simple compounds, such as *fashion week* or *fashion store*, both terms are underived). All words (whether simple words, derived words, inflected words or compounds) are SYNTACTIC ISLANDS, by which we mean that the syntax cannot ‘break into’ them and manipulate their subparts. The syntactic islandhood of words is one criterion that can be used to differentiate between words and phrases.

Another useful criterion for wordhood, this time leaning on the semantic properties of words, is REFERENTIAL ISLANDHOOD. Just like words are islands for syntactic manipulation, so they are islands for semantic dependencies, such as anaphora. In **John is a sushi chef, and Bill is a pastry one*, we see a failed attempt to have the anaphoric pronoun *one* refer back to *sushi*. We encounter the same problem in **John is a city dweller, and Bill is a village one* — but not in *John is a good chef, and Bill is a bad one*. While *sushi/pastry chef* and *city/village dweller* are compounds (i.e., complex words), *good/bad chef* is a phrase, within which the head *chef* is replaceable with a pronoun, *one*. In our attempt to find an antecedent for the pronoun *one*, we ‘break into’ phrases but not into words. Words are not just islands for syntax, they are islands for semantics as well.

Syntactically and semantically based criteria for wordhood are not always conclusive in deciding whether a given string is a word. The string *idiot doctor*, which can be used to make reference to a doctor who had made lots of medical errors (as in *That idiot doctor amputated the wrong leg!*), is both a syntactic and a semantic island: it is impossible to say **What is he an idiot?*, or **He is an idiot doctor, and she is a genius one*. But nonetheless, *an idiot doctor* is not a (compound) word: the words *idiot* and *doctor* each form their own syntactic phrases. We know this from the behaviour of *idiot doctor* in the realm of STRESS: the stress pattern for *an idiot doctor* features two primary stresses, one on the first syllable of *idiot* and the other on the first syllable of *doctor*. This makes the expression *idiot doctor* behave strikingly differently from the true compound *eye doctor* (i.e., an ophthalmologist), where only *eye* bears primary stress. As a general rule, a word (whether simple or complex) can bear only a single PRIMARY STRESS or WORD STRESS. (This is not to say that there can be only one stress in a complex word: alongside the primary stress, there can be one or more secondary stresses.) Stress gives us a third criterion for wordhood, alongside syntactic and semantic islandhood. By this criterion, the expression *idiot doctor*, with its two primary stresses, is a syntactic phrase, not a compound word.

Primary stress is one of several phonological properties that are helpful in diagnosing words. Another is VOWEL HARMONY — although this phonological criterion is obviously only usable in languages that feature vowel harmony in the first place. English is not such a language, but Hungarian is. In a Hungarian complex word, it is typically the case that the vowels harmonise with one another in certain ways. We will talk more about this in segment V; for now, let us confine ourselves to the extreme case, where the vowels are all exactly the same, as in *köszönöm* ‘I thank You’, *szeretlek* ‘I love you’ and *akarlak* ‘I want you’. We know that the *lek* of *szeretlek* and the *lak* of *akarlak* are not words separate from *szeret* and *akar* because if they were, it would be very difficult indeed to understand why *lek* must be chosen after *szeret* and *lak* after *akar*: plainly, when *szeret* and *akar* are followed by a separate word that happens to begin with the string *l+vowel+k*, this word can have any choice of vowel — thus, *ott szeret lakni* ‘(s)he likes to live there’ and *nem akar lekvárt enni*

‘(s)he doesn’t want to eat jam’ are perfectly fine. But vowel harmony should be handled with care as a criterion for wordhood: although words derived via the addition of a derivational or inflectional affix (such as *-lak/lek*) generally must be harmonic, compound nouns can readily be disharmonic. Thus, consider *Budapest*, the names of the capital of Hungary — a compound consisting of *Buda* and *Pest* (themselves independent cities before they were united), with *a* and *e* living happily side by side, even though they would not dream of cohabiting in **szeretlak* or **akarlek*. From the fact that there is a single primary stress in *Budapest* (falling on *Bu*) and that *Budapest* is a syntactic and semantic island (you cannot say **Buda, she would really like to visit __pest*, and in *She visited Budapest, and really liked it*, the pronoun cannot be interpreted as anaphoric just to *Buda* or *Pest*), we deduce that *Budapest* is certainly a compound word, not a phrase. The fact that the compound is not (fully) harmonic does not subtract from that conclusion; it merely cautions us that not all criteria for wordhood give the same result in all cases.

In the preceding paragraphs, we have seen that the WORD can be defined in terms of criteria that are rooted in syntax, semantics and phonology. A word is a unit that is a syntactic and semantic island and has only a single primary stress. It seems inevitable to define wordhood in terms of the other major subdisciplines of linguistics: phonology, semantics, and syntax. In many other ways, too, morphology is ‘at the mercy of imperialistically minded neighbours’ (i.e., the other subdisciplines of linguistics), making it ‘the Poland of linguistics’ (Spencer & Zwicky 1998:1), often struggling to delimit its own sovereign territory. Spencer & Zwicky (1998:1) eloquently express the quintessential interface nature of morphology as follows: ‘Morphology is at the conceptual centre of linguistics. This is not because it is the dominant subdiscipline, but because morphology is the study of word structure, and words are at the interface between phonology, syntax and semantics.’ By starting our journey around the world of linguistics with words and their internal structure, we are offered glimpses of other parts of the linguistic world *en route*. Of course, you will not be able to fully understand and appreciate all the ingredients of each of these other parts right away — but we will get to them eventually, so at the end of the course you will be able to revisit this discussion and grasp it in full.

With this as background, let us now proceed to examine a number of morphological properties and processes which will help us understand what morphology is about.

II.2 The lexical/functional distinction

Let us begin by comparing the examples in (1a) and (2a) to those in (1b) and (2b).

- | | | | |
|-----|----|---------------------|--|
| (1) | a. | can _{Aux} | (as in <i>Can they sing?</i>) |
| | b. | can _N | (as in <i>They put the peaches in a can</i>) |
| (2) | a. | have _{Aux} | (as in <i>Have they finished yet?</i>) |
| | b. | halve _V | (as in <i>They halve the peaches before serving them</i>) |

When you pronounce *can* in isolation, in its ‘citation form’, the vowel /æ/ comes out the same way regardless of whether by *can* we mean ‘be able to’ (as in (1a)) or ‘tin container’ (as in (1b)). But when *can* is produced in a sentence, and thereby disambiguated towards one or the other of its two meanings, we observe a clear difference in vowel length: in many varieties of English, *can* in (1b) is noticeably longer than it is in (1a). In varieties that pronounce the verb *to halve* with an /æ/, we see the same length effect in (2a) versus (2b), with the vowel having a much longer duration in the latter than in the former.

The duration of the vowel /æ/ is a matter of pronunciation — but what makes the pairs in (1) and (2) of interest in the context of our discussion of morphology is that the question of whether /æ/-lengthening applies is sensitive to information about WORD CLASS. Vowel lengthening happens in LEXICAL (semantically rich, OPEN-CLASS) words but not in FUNCTION(AL) or GRAMMATICAL (semantically not very descriptive, CLOSED-CLASS) words. This information about word class (lexical *vs* functional/grammatical) is morphological information. Phonological processes can be sensitive to such information.

We see this sensitivity of phonological processes to morphological factors also in the context of a vowel lengthening effect in Scots English (part of Aitken’s Law):

- | | | | | | | | |
|-----|----|-------|--------|------|--------|--------|---------|
| (3) | a. | road | [rod] | row | [ro:] | rowed | [ro:d] |
| | b. | brood | [brud] | brew | [bru:] | brewed | [bru:d] |

The vowel is short before [d] in the examples in the leftmost column, but long (marked by :) before the same [d] in the examples in the right-most column. The short vowel in *road* and *brood* shows that Scots English does not have a general rule of vowel lengthening before [d]. What makes the vowel long in *rowed* and *brewed*, then? The [d] of *rowed* and *brewed* is not intrinsically a part of these words: it is the exponent of a grammatical element added to verbs to mark the past tense or the past participle. The hosts for the grammatical suffix realised here as [d] are the verbs in the middle column, where the vowels are long thanks to the fact that they occur at the end of the word. When the verbs *row* and *brew* host the past-tense/participial suffix, the output has a long vowel since the input to past-tense/participle formation already had a long vowel.

II.3 The morpheme

The marker of the past tense or past participle is what morphologists call a MORPHEME (a grammatical or functional morpheme in the case at hand). The roots of the verb forms in the right-most column in (3) (*row* and *brew*) are also morphemes (lexical ones this time). Morphemes are the minimal building blocks or atoms of morphological analysis: elements which cannot be ‘dressed down’ into smaller component parts.

Morphemes are frequently defined as the smallest *meaningful* unit — but it is a matter of debate whether meaning ought to factor into the definition of the morpheme. Bloomfield (1933) called the morpheme ‘a linguistic form which bears no partial phonetic-semantic resemblance to any other form’, thereby giving morphemic analysis access to meaning without necessarily having to attribute a discrete meaning to each individual morpheme itself: it is certainly easier to decide whether or not a particular element ‘bears a semantic resemblance to’ (i.e., ‘means roughly the same thing as’) some other element than it is to decide what exactly its own meaning might be.¹

1 This is a sound strategy, but it may not get us out of the woods entirely. To see this, consider the Dutch word *water-s-nood* ‘water-S-emergency’. Here, we find a linking element between the two semantically meaningful parts of the compound. There can be no doubt that the linking element *-s-* bears a direct phonetic resemblance to the genitive morpheme *-s* (*Jan-s boek* ‘Jan-GEN book, Jan’s book’): indeed, the two are *identical* in phonetic form. But does linking *-s-* bear a (partial) semantic resemblance to genitival *-s* or not? Relevant here is that the word *water-s-nood* occurs alongside *water-nood* ‘water-emergency’, without the linking *-s-*. The two words have opposite meanings: while *watersnood* means ‘inundation, flooding’, *waternood* is used in reference to drought. This semantic difference cannot in any obvious sense be ‘blamed’ on the genitive: the English preposition *of*, which often expresses the same semantic relationship as genitival *'s* (cf. *the building's entrance* and *the entrance of the building*), can be used to express both that one is in need

II.4 When a single morpheme has different surface forms: Allomorphy

The English past-tense and past-participial morphemes, which we encountered in (3), are usually spelled as *-ed* and pronounced as [d] or [t] — the latter generally confined to hosts that end in a voiceless consonant, as in *kissed* and *sipped* (see (5j), below, for an apparent exception). In the examples in the rightmost column in (3), the *e* that appears in the spelling is silent. But when *-ed* is attached to a verb that itself ends in a [d], or in its close relative [t], we cannot pronounce *-ed* as ‘just’ [d]: a sequence of [d] or [t] immediately followed by [d] is not possible within a single (non-compound) word in English. Because words such as *kneaded* or *knitted* cannot be pronounced as *[niːdd] and *[nɪtd] (here and in what follows, the asterisk marks ill-formedness), something needs to be done to rescue these forms. The solution is to actually pronounce the orthographic *e* of *-ed* in these cases, as [ə] (called ‘schwa’, the vowel that you get when you leave your tongue in its neutral position) or as [ɪ] (the same vowel as the one heard in *to knit*); the choice between [əd] and [ɪd] depends on the speaker/dialect. The forms in (4) are alternative realisations of the morpheme spelled as *-ed* — they are, as morphologists say, ALLOMORPHS of this morpheme (where the *allo-* part of *allomorph* is a relative of the *al-* of *alternative*, which gives us a way to remember what allomorphy is about).

- (4)
- | | | |
|----|-----------|--|
| a. | [t] | (as in <i>kissed, sipped, kicked</i>) |
| b. | [d] | (as in <i>buzzed, rubbed, mugged</i>) |
| c. | [əd]/[ɪd] | (as in <i>knitted, kneaded</i>) |

The orthographic form *-ed* is, as was pointed out previously, not actually the representative of one single morpheme — *-ed* can represent the past-tense morpheme (as in *she knitted a sweater*) as well as the past-participial morpheme (as in *she has knitted a sweater*). The past-tense and past-participial morphemes are separate entities in the morphology of English:

(5)

	PAST TENSE	PAST PARTICIPLE
a.	<i>prove</i>	<i>proved</i> or <i>proven</i>
b.	<i>shave</i>	<i>shaved</i> or <i>shaven</i>
c.	<i>slay</i>	<i>slayed</i> or <i>slain</i>
d.	<i>take</i>	<i>taken</i>
e.	<i>ride</i>	<i>ridden</i>
f.	<i>draw</i>	<i>drawn</i>
g.	<i>break</i>	<i>broke</i> or <i>broken</i>

of water and that there is a surfeit *of* water. Indeed, it is difficult to put one’s finger on a constant meaning for the genitival marker of Dutch: *Jans boek* ‘John’s book’ can, like its English translation, express just about any imaginable relationship between *Jan* and *boek* (the book that Jan owns, or has written, or is reading, or has purchased or found, or is reviewing, or hates, or is the protagonist of, etc. etc.). If neither linking *-s-* nor genitival *-s* has a meaning, that makes them semantically alike (in their emptiness) rather than distinct. So then we cannot decide on morphemehood for linking *-s-* based on semantic grounds; and since *-s-* is phonetically identical with genitival *-s*, the surface form of the linking element does not attribute morphemic status to it either. In light of this, the cautious approach will be to call *-s-* a LINKING MORPH (where ‘morph’ simply means ‘form’) rather than a linking morpheme.

In (5a–c), we see three verbs whose past-participial form can employ a suffix different from the one used to form these verbs’ perfectly regular past-tense forms. You would not dream of using the suffix *-en* to form the past tense of *prove*, *shave* or *slay*; but for the past participle of these verbs, *-en* is available, and especially common when the participle is used as an attributive modifier of a noun, as in *a proven method*, *a clean-shaven face* or *a slain dragon*. The *-en* form of the past participle is our only choice with verbs such as *take*, *ride*, *draw* and *break*, as shown in (5d–g). These verbs not only mark their past participles in an ‘unusual’ or ‘irregular’ way (employing *-en* rather than *-ed*) but they do something even more radical in the past tense: here, instead of using a suffix, these verbs change their vowel. The verb *slay* can do this as well (*slew*). And in the case of *break*, the vowel change is found not just in the past tense but also in the past participle (*broken*), the latter combining the suffixation of *-en* with the vowel change strategy seen elsewhere in (5a–f). Such vowel change is called (with a German word) *ABLAUT* or (with a term going back to Greek) *APOPHONY*. *Ablaut* is one illustration of what linguists call *NON-CONCATENATIVE* morphology: instead of concatenating two morphemes in a linear string (as in *knead-ed*), we are combining two morphemes non-linearly, squeezing them together in a form that is no longer than the input string. English uses non-concatenative morphology sparingly in its inflectional system; but Old English and present-day German rely on it a great deal more, and Arabic is particularly famous for its non-concatenative morphology. (We will not explore this further here; but if you are interested in non-concatenative morphology, there is some excellent work on the topic which you will be able to access after completing this course.)

To write an adequate description of the allomorphy of the English past-tense and past-participial morphemes, we have to give cognizance to the forms seen in (5). But before we can go and list all the allomorphs of these morphemes, we also need to give a few moments’ thought to the verbs illustrated in (5h–n), below:

(5)		PAST TENSE	PAST PARTICIPLE
h.	<i>sing</i>	<i>sang</i>	<i>sung</i>
i.	<i>keep</i>	<i>kept</i>	<i>kept</i>
j.	<i>bring</i>	<i>brought</i>	<i>brought</i>
k.	<i>hit</i>	<i>hit</i>	<i>hit</i>
l.	<i>put</i>	<i>put</i>	<i>put</i>
m.	<i>be</i>	<i>was</i>	<i>been</i>
n.	<i>go</i>	<i>went</i>	<i>gone</i>

The verb *sing* is an example of a verb which uses *Ablaut/apophony* not just in the formation of its past-tense form but also to form its past participle, with the particular choice of ablauted vowel being different in the two cases: [æ] versus [ʌ]. So *Ablaut* is not just the prerogative of the past tense in English (indeed, we also find it outside the verbal inflectional system, in the formation of plurals: *man ~ men*, *foot ~ feet*). The case of *keep* and *bring* shows that a combination of *Ablaut* and concatenative morphology is possible in the formation of the past-tense and past-participial forms: the vowel [ɛ] of *kept* and the vowel [ɔ] of *brought* are the result of *Ablaut*, while the final [t] is one of the allomorphs of the regular suffix *-ed*, as seen in (4a). (It may strike you as peculiar that [t] rather than [d] is used in (5j): after all, [t] is generally confined to hosts that end in a voiceless consonant, but in *brought* the [t] is immediately preceded by a vowel in present-day English. The thing to bear in mind, however, is that the *gh* in the orthography of *brought*, while silent in today’s English, used to be pronounced in the same way as the final consonant in the word *loch* — a voiceless consonant

produced far back in the oral tract, also found in the German translation equivalents for past-tense *brought* (*brachte*) and past-participial *brought* (*gebracht*). The realisation of the suffix as [t] was already set in Old English, and has not changed over time, while the *gh* eventually vanished from the pronunciation of *brought*.)

In (5k) and (5l), we see that a morpheme can remain perfectly silent on the surface. The forms in the PAST TENSE and PAST PARTICIPLE columns for *hit* and *put* are morphologically endowed with the past-tense and past-participial morphemes, resp. — but these morphemes have no surface exponent. This tells us that a morpheme can have a SILENT ALLOMORPH, a bit of morphology that has no phonology (‘sound’) associated with it. We can annotate silent allomorphs with the aid of the empty-set symbol: \emptyset . The use of the silent allomorph of the past-tense and past-participial morphemes is an idiosyncratic (‘weird’) quirk specific to certain verbs, not determined by the phonological form of these verbs. This is clear from the fact that alongside *hit* we find *fit*, and alongside *put* we find *putt* (a verb familiar to golfers), the latter forming their past tense and past participle regularly, through the suffixation of *-ed* (*fitted*, *putted*).

Further along the spectrum of idiosyncrasy we find *be* and *go*, two verbs famous for their irregularity. When we look closely at the paradigms for *be* and *go*, what we see is that their past participles are in fact not particularly peculiar: true, they do not employ *-ed*, but they *are* formed via the attachment of a suffix to the stem — the same suffix that we had come across previously in forms such as *shaven* and *taken*. (In addition to suffixation of *-en*, the forms *been* [bi:n] and *gone* [gɔ:n] also feature a (somewhat subtle) vowel change: another illustration of *Ablaut* co-occurring with concatenative morphology.) The irregularity of *be* and *go* really manifests itself in their past-tense forms, which are entirely unrelated to their other forms. There is no way in which we could turn *be* into *was* or *go* into *went* by phonological rule: [b] and [g] never turn into [w] elsewhere in the sound system of English; neither do we find [i] changing into [ɔ] or [əʊ] changing into [ɛ] in the productive phonology of the language. Historically, *was* and *went* are formed from roots different from the ones producing the other forms of the paradigms of *be* and *go*. It is as if we are dealing here with a car that has a round cookie jar replacing one of its regular wheels. The garage apparently ran out of regular wheels and put something else in place of the missing wheel to fix up the car. Similarly, the verbs *be* and *go* fixed up the hole in their paradigm with something completely different. The technical term for such a fix-up is SUPPLETION.

Putting it all together, what we arrive at is the sets of allomorphs for the past-tense and past-participial morphemes of English in (6) and (7), respectively.

- (6) PAST TENSE
- a. [t] (as in *kissed*, *sipped*, *kicked*)
 - b. [d] (as in *buzzed*, *rubbed*, *mugged*)
 - c. [əd]/[ɪd] (as in *knitted*, *kneaded*)
 - d. \emptyset (as in *hit*, *put*)
 - e. ABLAUT (as in *took*, *sang*)
 - f. SUPPLETION (as in *was*, *went*)
 - g. (6a)+(6e) (as in *kept*, *brought*)

- (7) PAST PARTICIPLE
- a. [t] (as in *kissed*, *sipped*, *kicked*)
 - b. [d] (as in *buzzed*, *rubbed*, *mugged*)
 - c. [əd]/[ɪd] (as in *knitted*, *kneaded*)

- d. \emptyset (as in *hit, put*)
- e. *ABLAUT* (as in *sung*)
- f. [n] (as in *shaven, taken, drawn*)
- g. (7a)+(7e) (as in *kept, brought, been, gone*)

The point of (6) and (7) is not only to drive home the fact that there is a great deal of allomorphy in the morphology of the English past tense and past participle (indeed, it may well turn out that there is even more allomorphy here than indicated in (6) and (7), which do not strive for an exhaustive description) but also, and more importantly, to show that the allomorphy of the past-tense morpheme and that of the past-participial morpheme, though very similar, are not identical: (7f) does not occur in the past tense list; and (total) suppletion (6f) is not found in past participles (even the most irregular verbs, *be* and *go*, have non-suppletive participles).

The ‘A B A’ pattern seen in (5m,n), due to the absence of suppletion in the past participle, is actually quite unusual in three-member paradigms: for instance, in the marking of degree on adjectives we never come across such a pattern: *good ~ better ~ best* shows suppletion in an ‘A B B’ pattern; ‘A B A’ *good ~ better ~ *goodest* does not occur. This may lead you to think of the widespread ‘A B B’ pattern for *go* (with *I have went home* instead of *I have gone home*) in a different light — different, that is, from the usual prescriptivist scorn cast on speakers producing this pattern.

II.5 The distribution and acquisition of allomorphy

Adult native speakers of English know about the patterns of allomorphy given in (6) and (7), and use them on a daily basis — although there is variation among speakers in the case of individual verbs, not just in the case of *go* (mentioned in the previous paragraph): think here also about verbs such as *sing* (for which *sung* occurs in the past tense) and *bring* (which in dialects has *brang* in the past tense and *brang* or *brung* as the past participle). These patterns need to be learnt based on exposure to the data (and sometimes ‘corrected’ through explicit instruction). For high-frequency verbs such as *break, bring* and *go*, there is plenty of exposure to the irregular past-tense and past-participial forms. Not surprisingly, therefore, young children growing up in an English-speaking environment will quickly register these irregular forms and produce them themselves in the earliest stages of production. But interestingly, after producing *broke, brought* and *went* very early on, children go through a stage during which they overregularise by applying to these verbs the ‘normal’ past-tense marker *-ed*, saying *breaked, bringed* and *goed*. Fairly soon thereafter, they revert to the irregular forms; but the fact that children pass through this overregularisation stage is an indication that at some point in the acquisition process, ‘the coin drops’: they master the general rule of past-tense formation via the suffixation of *-ed*, and apply it blindly, across the board. What we see illustrated here is that language acquisition starts out with direct imitation of what is heard in the environment, followed by the mastery of the general rule and its application to every possible target, after which the ‘exceptions’ to the rule are learnt on an individual basis and the general rule gives way, on a case-by-case basis, to the various forms of allomorphy discussed above.

What the allomorphy of the English past tense and past participle, and the way children deal with it in the language acquisition process, can tell us is two things:

- (a) morphological processes are rule-governed; and
- (b) more specific rules ‘block’ the application of the general rule

We see (b) at work throughout (5): in each of the examples given there, there are forms which fail to undergo the general rule (which, for the English past tense, is the suffixation of *-ed*, realised in one of the ways given in (4)) because there is a more specific rule of allomorphy that applies; whenever a more specific rule applies, the general rule cannot. This is referred to as SPECIFICITY, or, alternatively, as the ELSEWHERE PRINCIPLE (so called because the general rule applies ‘elsewhere’, i.e., in all cases where the specific rules do not). It is one of the oldest insights in linguistics, dating back all the way to the work of the great Indian linguist known as Pāṇini, whose detailed analyses of Sanskrit morphology mark the origin of linguistic analysis.

It will not come as a surprise to you that some of the more specific ways of past-tense or past-participle formation, inherited from earlier stages of the historical development of the English language, are prone to ‘erosion’, giving way over time to the ‘elsewhere case’. Frequency plays a role in this: verbs that are frequent in everyday use tend to hold on to their irregular paradigms more persistently than verbs are considerably less frequent. Thus, while for *drive* the past tense and past participle are still quite robustly *drove* and *driven*, resp., for the verb *strive*, regular *strived* has made the forms *strove* and *striven* quite uncommon in present-day spoken English.

We see the drive towards regularisation in other pockets of the morphology of English as well. The English plural morpheme for nouns is even richer than the past-tense and past-participle morphemes in its surface allomorphy, as shown in (8). Again, the list is representative but not necessarily exhaustive.

(8)	PLURAL	
a.	[s]	(as in <i>aunts, ducks</i>)
b.	[z]	(as in <i>uncles, dogs</i>)
c.	[əz]/[ɪz]	(as in <i>nieces, horses</i>)
d.	∅	(as in <i>sheep, aircraft</i>)
e.	ABLAUT	(as in <i>mice, geese</i>)
f.	[n]	(as in <i>oxen, children</i>)
<hr/>		
g.	[ə]	(as in <i>criteria, phenomena</i>)
h.	[əj]	(as in <i>foci, loci</i>)
i.	[ej]	(as in <i>larvae</i> — alternate pronunciation: [iː]; see (8j))
j.	[iː]	(as in <i>algae, larvae</i> ; also in <i>tempi</i> , as the plural of <i>tempo</i>)
k.	[iː]+voicing	(as in <i>analyses, (hypo)theses</i>)
l.	[ərə]	(as in <i>corpora, genera</i> ; also in <i>tempora</i> , as the plural of <i>tempus</i>)
m.	[ɪm]	(as in <i>seraphim, cherubim</i>)

Of these various allomorphs of the plural morpheme, only the first six are of indigenous Germanic origin, and only the first three (whose distribution, just as in the case of (6a–c) and (7a–c), is conditioned by rules addressing the phonological properties of the stem) are robustly represented and shared by all native speakers. Especially in the spoken language, (8d) and (8e) show a tendency to give way to one of the forms in (8a–c): thus, *aircrafts* and *oxes* occur frequently; and tellingly, the word *mouse* in its sense of ‘device moved over a flat surface to produce a corresponding movement of the cursor on a computer screen’, which is a recent technological innovation, takes the regular plural suffix [əz]/[ɪz] rather than the archaic *Ablaut* that we find with *mouse* when it refers to a small rodent.

The forms below the dotted line in (8) all belong to the learned vocabulary,² and all involve morphology that was inherited by English from the donor languages (Latin, Greek, Hebrew). English speakers do not necessarily recognise and treat these as plural markers. The fact that (8g), (8l) and (8m) can be found to combine with regular (8b) (as in substandard *criteria*s, *corporas*,³ *seraphims*) is an indication to this effect, as is the fact that words such as *criteria* (the plural of *criterion*), *phenomena* (the plural of *phenomenon*) and *data* (the plural of *datum*), when used as the subject of a finite clause, frequently give rise to the third person singular present-tense inflection *-s* (as in *the criteria is not met*, *the phenomena is rare*, and *the data are/is mixed*⁴). On the other hand, we can also observe a tendency towards overextension of the exogenous plural in (8k): thus, the plural of *bias* (i.e., *biases*) is often pronounced as *bias*[i:z], extending the plural in (8k) to a form that is not derived from a Greek word ending in *-is* (for which (8k) is otherwise reserved); and although *species* (pronounced as [spi:si:z] or [spi:ʃi:z]) is a singular noun, the fact that it ends in [i:z] has led people to treat it as a plural, giving rise to the introduction of singular *specie* in the sense of ‘type, kind’.

All in all, the forms below the dotted line in (8) are nothing but trouble. It is doubtful that they truly belong to the morphology of the English nominal plural. If we set them aside, the pattern of allomorphy seen for the nominal plural marker is remarkably similar to that of the past-tense and past-participial morphemes: all of them have three phonologically conditioned allomorphs plus the zero allomorph plus *Ablaut*. But the [n] form is specific to the plural marker — a niche player in present-day English, but very frequent in medieval English and still today in Dutch and German. And another thing that stands out, when we compare (8) to (6)–(7), is that total suppletion does not appear to be attested for the English nominal plural, nor are there grammatical combinations of *Ablaut* and suffixation.

II.6 When different morphemes sound the same: Homophony

The suffix spelled with an *s* which we encountered in the preceding paragraphs leads an interesting ‘triple life’ in English. It not only forms the plural of nouns, but also serves, still in the nominal domain, as the marker of possessors (spelled with an apostrophe before the *s*, and sometimes referred to as the ‘Saxon genitive’) and, in the verbal domain, as the third person singular present-tense marker. Possessive *'s* and verbal inflectional *-(e)s* both exhibit the same phonologically conditioned allomorphy that we saw in (8a–c). Additionally, both the possessive marker and the third person singular present-tense morpheme have a silent allomorph: the modal auxiliaries of English (*can*, *may*, *must*, *will*, etc.) systematically resist *-(e)s* when they have a third person singular subject; and, again systematically, we do not get *'s* when the possessor is itself a regularly inflected plural (as in *the boys' room*: note that we do not add [əz]/[ɪz] after [boɪz], which is what we would have expected if the marker of possessors had an exponent of its own; the orthography is revealing here, preserving the apostrophe but not spelling the *s* that would otherwise appear after the apostrophe (')).

2 Note, as an aside, that the past-participial morpheme in *learned* is pronounced as [ɪd], not (as one might have expected) as [d]. This shows that the distribution of (7d) is not entirely identical with that of (6d)

3 This is even the standard in the case of *operas*, in the sense of ‘classical theatre pieces involving singing’, for which *opera* is only usable as a singular. Quiz question: what was *opera* originally the plural of? Try to figure this out by comparing *opera* to the forms *corpora* and *genera* and their singulars.

4 The use of singular *is* with *data* can occasionally be observed even in cases in which it is clear that *data* does not make reference to an unindividualised mass, as in *the individual data is included*. This seems less natural, however.

(9)		POSSESSOR MARKING ('Saxon genitive')		3 RD PERSON SINGULAR PRESENT TENSE
	a.	[s] (as in <i>Jack's room</i>)		[s] (as in <i>she laughs</i>)
	b.	[z] (as in <i>John's room</i>)		[z] (as in <i>she sings</i>)
	c.	[əz]/[ɪz] (as in <i>Chris's room</i>)		[əz]/[ɪz] (as in <i>she hisses</i>)
	d.	∅ (as in <i>the boys' room</i>)		∅ (as in the modals)

Though the regular, phonologically conditioned allomorphs of the 'Saxon genitive' and the third person singular present-tense marker are the spitting image of the nominal plural marker, the two morphemes cannot be fully identified because (a) the meanings of the three are quite different from one another, and (b) the range of allomorphy is far greater for PLURAL than for the other two. Instead of collapsing the three morphemes (PLURAL, POSS and 3SG.PRES) into one 'supermorpheme', we will be better advised to take the more cautious route of treating these each as separate morphemes — three morphemes that happen to show a great deal of HOMOPHONY (i.e., they often sound the same).

The English prefix *un-* presents another interesting case of 'double duty', occurring in two different morphological contexts and having two different (though not entirely disjoint) interpretations. When *un-* is attached to verbs, as in (10), it has a reversative meaning: 'undoing something' means to reverse the effect of the act of 'doing something'. But when *un-* is attached to adjectives, as in (11), it has a negative interpretation: 'unhappy' is equivalent to 'not happy'.

(10)	a.	to do	~	to <u>un</u> do	'reverse <i>x</i> '
	b.	to button	~	to <u>un</u> button	
	c.	to cover	~	to <u>un</u> cover	
	d.	to wrap	~	to <u>un</u> wrap	
(11)	a.	happy 'x'	~	<u>un</u> happy	'not <i>x</i> '
	b.	pleasant	~	<u>un</u> pleasant	
	c.	realistic	~	<u>un</u> realistic	
	d.	welcome	~	<u>un</u> welcome	

Are we dealing here with two separate morphemes or can the two instances of *un-* be collapsed into a single morpheme?

Highly relevant with regard to this question is the fact that while the reversative prefix is invariably *un-*, the negative prefix has a second allomorph, *in-* (more accurately, a bunch of allomorphs: *in-* (as in *inaccurate*), *il-* (as in *illegitimate*), *ir-* (as in *irregular*), *im-* (as in *impossible*), and *i[ɪ]-* (as in *incorrect*)), attaching to stems of Latin origin. This recalls the situation regarding the three English markers spelled with an *s* discussed above, which led us not to conflate them into a single morpheme. That negative *un-* and reversative *un-* in all likelihood must be treated as separate morphemes that happen to be homophonous comes from an important generalisation first discovered by the Danish linguist Otto Jespersen: negative *un-* attaches to positive gradable adjectives but not to negative ones:

(12)	a.	unhappy	vs	*unsad
	b.	untrue	vs	*unfalse
	c.	unwise	vs	*unstupid, *unfoolish
	d.	unclean	vs	*undirty
	e.	unhealthy	vs	*unsick

A direct consequence of the fact that *un-* does not combine with negative hosts is that *un-* also does not stack: (13a) is bad because the first instance of *un-* attaches to a negative host created by prefixation of the second *un-* to *happy*. The ungrammaticality of (13a) is not due to a semantic problem: negation certainly can combine with *unhappy* to negate the negation, as in *I am not unhappy*. So the fact that **ununhappy* is impossible reveals an important property specific to *un-*. And importantly, this property is specific to negative *un-*: as (13b) illustrates, reversative *un-* does stack.

- (13) a. *I am ununhappy *_{[NEG un- [_{NEG un- [_{POS happy]]]}]}}
 b. I am trying to un-undo the changes made to this file

The fact that the negative and reversative prefixes behave differently with respect to allomorphy and the polarity of their host leads us to conclude that although negative *un-* and reversative *un-* may be related, and although they have the same phonological form (i.e., they are homophonous), they probably must be included in the lexicon separately, as different morphemes, because they have different distributional properties which cannot straightforwardly be derived on an approach to *un-* in terms of contextual allomorphy. What further strengthens us in the belief that negative and reversative *un-* are different morphemes is the fact that in other Germanic languages, negative *un-* and reversative *un-* are not identical (see (14)).

- | | | NEGATIVE | | REVERSATIVE |
|------|----|----------|--|---|
| (14) | a. | German: | <i>un-</i> (<i>unglücklich</i> ‘unhappy’) | <i>ent-</i> (<i>entdecken</i> ‘un/discover’) |
| | b. | Dutch: | <i>on-</i> (<i>ongelukkig</i> ‘unhappy’) | <i>ont-</i> (<i>ontdekken</i> ‘un/discover’) |

II.7 The hierarchical structure of complex words

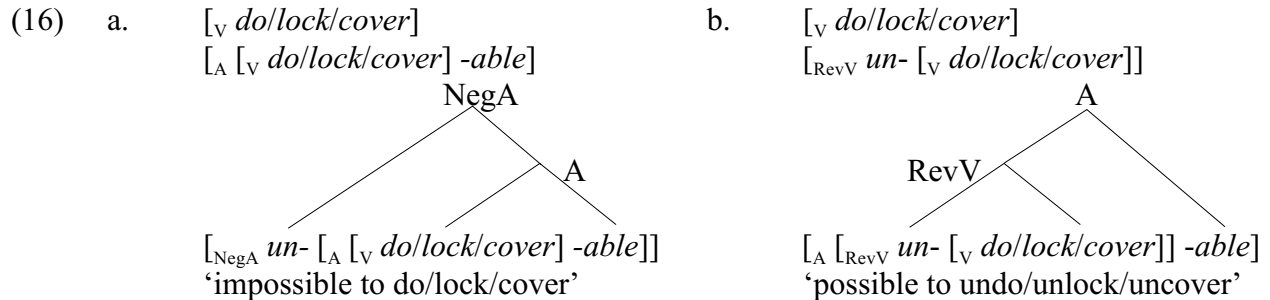
The two morphemes *un-* discussed in the previous paragraphs select hosts of different categories (adjectives for negative *un-*, verbs for reversative *un-*). Therefore, they do not co-occur in a single form, and we cannot substitute one for the other. Apparently contradicting this second statement are forms of the type in (15), which are ambiguous precisely with regard to whether *un-* is interpreted as the negative prefix or the reversative one:

- (15) a. un-do-able ‘impossible to do’ OR ‘possible to undo’
 b. un-lock-able ‘impossible to lock’ OR ‘possible to unlock’
 c. un-cover-able ‘impossible to cover’ OR ‘possible to uncover’

We can appreciate the difference between the two readings of these forms when we place them in a context. In the sentence *We wanted to do the exercise, but it turns out to be undoable* we clearly get the ‘impossible to do’ interpretation; but in *We wanted to undo the changes, and thankfully they turned out to be undoable* we instead find the ‘possible to undo’ reading. What do these forms owe this ambiguity to?

To get to the answer to this question, we first need to recognise that the forms in (15) all feature two affixes: besides the prefix *un-* also the suffix *-able*. The suffix *-able* forms adjectives out of verbs: *do-able* takes the verb *do* as its input and delivers, via suffixation of *-able*, an adjective paraphrasable as *feasible* (itself featuring the *-ible* allomorph, attached to the bound root *feas-*; on bound roots, recall II.1, above). Now, we know that reversative *un-* attaches to verbs and negative

un- takes an adjectival host. The forms in (15) all feature verbs (the bits in between the hyphens); but they also all feature adjectives (thanks to the fact that *-able* is an adjective-forming suffix). The trick to understanding the ambiguity of the forms in (15) is to realise that *-able* in principle has a choice as to when it gets attached: it can attach directly to the underived verb (*do*, *lock*, *cover*), delivering an adjective, when can then serve as the host to negative *un-*; OR the underived verb can first play host to reversative *un-*, after which *-able* turns the *un-+V* combination into an adjective. The two derivations are illustrated graphically in (16). It is via (16a) that we get the ‘impossible to V’ reading, whereas the ‘possible to un-V’ interpretation comes about via (16b).



Tree-like structures of the type given above will continue to serve us well throughout, well beyond morphology. The arboreal representations are equivalent to the labelled bracketings underneath them.

The difference between the two readings of the forms in (15) thus lies in the relative height of attachment of the prefix *un-* and the suffix *-able*. Because these affixes appear on opposite sides of their host, the linear string does not immediately ‘give away’ the order in which these affixes are attached — which, in this particular case, gives us the opportunity to freely vary their relative order of attachment, which, in turn, gives rise to the ambiguity of these complex forms.

The key to the solution to the puzzle presented by the ambiguity of the forms in (15) is that these complex forms have hierarchical structures that are the product of the sequence in which the affixes are combined with their hosts. When *un-* attaches early and is therefore hierarchically lower than *-able*, the host is still a verb at the point at which *un-* attaches, and concomitantly *un-* is read as the reversative marker; but when *un-* attaches after *-able* has already suffixed itself to the verb and is therefore hierarchically higher than *-able*, the host for *un-* is an adjective, and we are dealing with negative *un-*.

II.8 Rule ordering

The discussion of (15) has shown that morphological affixation rules apply in a particular order, and that the order in which they apply has consequences for the hierarchical structure of the complex form and for its interpretation. In the case of (15), the rules of *-able* suffixation and *un-* prefixation apparently apply in either order; but as we have seen, when *un-* is reversative it must attach before *-able* is inserted whereas the opposite is true in the case of negative *un-*. The selectional properties of the two different *un-* prefixes are to blame for the ordering in question.

Another example of selectional restrictions determining the order in which morphological affixation rules apply is presented by the pair of words in (17), once again involving the prefix *un-*.



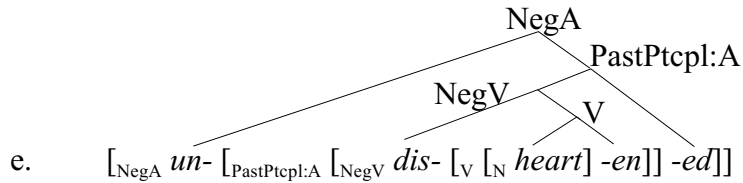
The question that forms of the type in (17) raise is whether the prefix attaches to the root before the suffix does, or the other way around. The surface string does not give us the answer, so we need other criteria to reach a decision. Selection again turns out to be key. Let us consider (17b) first. Here we see negative *un-* co-occurring with nominalising *-ity*. If we attach *un-* before *-ity* is brought in, we have an appropriate adjectival host for *un-* to dock onto. If instead we were to attach *-ity* first, the result is a noun, *grammaticality*, to which *un-* cannot attach itself: nouns just cannot host *un-*; negative *un-* selects specifically for an adjectival host. On this basis, we are led to conclude, therefore, that *un-* must attach to *grammatical*, and that *-ity* subsequently attaches to *ungrammatical*. This derivational history, captured by the bracketed structure given to the right of (17b), happens to also give us the right interpretation: *ungrammaticality* is suitably paraphrased as ‘the state of being ungrammatical’, with ‘state’ (= *-ity*) scoping over all of *un-grammatical*.

In the case of *unhappier* (17a), selection does not pertain to the category of the host: the comparative suffix *-er* does not change the category of its host (*happier* is, like *happy*, an adjective), so categorial selection is unaffected by *-er* suffixation. What matters in the case of (17a) is not what *un-* desires as the category of its host; rather, what is essential here is that *-er* selects for ‘small’ hosts — hosts that have a small number of syllables. In English, suffixation of comparative *-er* is possible to monosyllabic adjectives (i.e., adjectives that are just one syllable in length) and to a subset of disyllabic adjective whose second syllable is unstressed (thus, contrast *nárrow* to *remóte*, the former accepting *-er* and the latter resisting it). (Adjectives that do not meet these requirements can only form their comparative in English with the help of *more*.) The adjective *háppy* falls within the set of adjectives that are eligible for comparative formation via *-er* suffixation — and indeed, *happier* is well-formed. So if we attach *-er* directly to *happy*, and subsequently attach *un-* to *happier*, we get the grammatical output diagrammed to the right of (17a). If instead we attach *un-* to *happy* first, the interim result is a trisyllabic adjective, one to which *-er* should not be attachable.

The bracketing given to the right of (17a) has one drawback: it does not seem to correspond to an intuitive characterisation of the meaning of the complex form *unhappier*. Intuitively, what *unhappier* means is ‘more unhappy’, not ‘not happier’. We can use *unhappier*, for instance, in the following sentence: *John was already unhappy as a child, but as an adult he become even unhappier*, where *even unhappier* means ‘even more unhappy’, with the comparative (‘more’) taking scope over the negation. In the bracketed structure given to the right of (17a), which is the one we were led to based on the restrictions imposed by *-er* on its host, the negation finds itself higher in the structure than the comparative, giving rise to the opposite relative scope of the two: ‘not happier’.

This paradox (i.e., the fact that distributional restrictions lead to one hierarchical structure while our semantic intuitions appear to favour the opposite structure) has plagued the analysis of *unhappier* for a long time. But we can feel reasonably confident that the structure given above is actually the correct one for *unhappier*. This confidence comes from the observation that it is impossible to combine *unhappier* with another instance of *un-*: **unhappier* is ungrammatical. This is indicative of the fact that in *unhappier*, the prefix *un-* is indeed attached outside *happy+er*. Recall from (13a) that negative *un-* does not stack. More precisely put, it is impossible to attach *un-* to a host that itself bears *un-* — for we know from the well-formedness of words such as (18a) that it is actually possible to combine *un-* with a negative host provided that the two negative affixes (here *un-* and *dis-*) are not structurally adjacent. The structure of *un-dis-heart-en-ed* is built up as shown in (18b–d):

- (18) a. undisheartened
 b. $[_V [_N \textit{heart}] \textit{-en}]$
 c. $[_{NegV} \textit{dis-} [_V [_N \textit{heart}] \textit{-en}]]$
 d. $[_{PastPtcl:A} [_{NegV} \textit{dis-} [_V [_N \textit{heart}] \textit{-en}]] \textit{-ed}]$



First *heart* combines with *-en*, then *dis-* is added to form $[\textit{dis-}[\textit{heart-en}]]$, after which *-ed* is added to form $[[\textit{dis-}[\textit{heart-en}]]\textit{-ed}]$, and finally *un-* is brought in. In the structure in (18e), *un-* and *dis-* are linearly adjacent (i.e., next to each other), but structurally they are not: there are two opening square brackets between *un-* and *dis-*, which means that *un-* and *dis-*, the two negative prefixes, cannot ‘see each other’ directly; there is a bracket with the label ‘PastPtcl:A’ (for ‘adjectivally used past participle’) in between the two. What *un-* combines with in (18e) is the form in (18d), which has *-ed* (the past-participial suffix) and the verb *dishearten* as its constituent parts and has the label of *-ed* — a label that is not ‘Neg’, and is therefore a suitable host for *un-*, which can see *disheartened* but not the negative prefix *dis-* embedded inside this participial form. This has the happy consequence that, although *un-* and *dis-* are both negative prefixes, they do not ‘bite one another’ in the case of *undisheartened*: they can be combined thanks to the fact that they are not structurally adjacent.

With the knowledge just gained about (18) in mind, let us return to the structure of *unhappier*, and the observation that this form cannot be combined with another negative prefix (**unhappier*). If *un-* combines with the output of *-er* suffixation, as in the structure given to the right of (17a), this ban on adding another negative prefix to *unhappier* is immediately expected: after all, *un-* here is an immediate constituent of the complex form *unhappier*, and should therefore repel affixation of another negative prefix — $*[_{NegA} \textit{un-} [_{NegA} \textit{un-} [_{ComparA} [_A \textit{happy}] \textit{-er}]]]$ should be just as bad as **unhappy*, which indeed it is. If instead *un-* first combined with *happy*, and then *-er* were brought in, it would be very difficult to rule out $[_{NegA} \textit{un-} [_{ComparA} [_{NegA} \textit{un-} [_A \textit{happy}]] \textit{-er}]]$, whose structure is parallel in relevant respects to the structure of *undisheartened* in (18e).

So the fact that **unhappier* is ill-formed while *undisheartened* is fine now confirms our earlier conclusion, based on the restrictions that comparative *-er* places on its host, that the structure given to the right of (17a) is the correct derivation for *unhappier*. The fact that this structure does not seem to be perfectly in line with our semantic intuitions about the meaning of *unhappier* of course remains a concern. But we know independently that the surface structure does not always correctly reflect the scope of negation: though in the structure of the sentence $[\textit{Everyone} [\textit{did} [\textit{-n't} [\textit{leave}]]]]$, the subject (*everyone*) is hierarchically higher than the negation (*-n't*), it is nonetheless possible to interpret *everyone* in the scope of the negation (i.e., *Everyone didn't leave* can be interpreted in the same way as *Not everyone left*). We will return to mismatches between surface structure and interpretation in our discussion of semantics in segment IV of this series.

For now, the important conclusion emerging from the discussion of (16), (17) and (18) is that that morphological processes of affixation apply in a particular order, giving rise to a hierarchical structure for which evidence is available in the selectional restrictions imposed by the affixes and sometimes also in the semantic interpretation of the complex forms. The peculiar (and complicating) factor involved in the cases in (16), (17) and (18) is that we are dealing with a mix of prefixation and

suffixation, causing the linear string to underdetermine the hierarchical structure. Things are usually a lot simpler when we are dealing with cases multiple affixation in which all the affixes are attached on the same side of the host: the hierarchical structure and the linear string are usually nicely in sync in such cases. As an illustration, consider the Hungarian forms in (19).

- | | | | |
|------|----|---|--|
| (19) | a. | a barát-a-i-é
the friend-POSS-PLUR-POSS.ANAPH | [[[[<i>barát</i>] -a] -i] -é]
‘the one belonging to his friends’ |
| | b. | a barát-já-é-i
the friend-POSS-POSS.ANAPH-PLUR | [[[[<i>barát</i>] -ja] -é] -i]
‘the ones belonging to his friend’ |

In (19a), the possessed plural marker *-i* is both linearly and hierarchically closer to the root than the possessive anaphor *-é* — so we first pluralise ‘his friend’ to ‘his friends’, and then combine that with ‘the one belonging to’ (which is how *-é* is best paraphrased). In (19b), by contrast, the possessive anaphor *-é* is closer to the root than the possessed plural marker *-i*, so we do not apply the plural marker to ‘friend’ but to *-é* instead. The order in which the affixation processes are applied in (19) can be directly read off the surface string: the suffix closest to the root is attached first, after which the suffix to its right is brought in, etc.

II.9 Blocking and the lexicon

A startling fact about morphological derivation is that for particular affixes it sometimes fails in particular words while it works perfectly well in apparently very similar words. The plot thickens when we realise that semantically equivalent affixes may be oblivious to this restriction. The English pair of cases in (20) illustrates this particularly well:

- | | | | | |
|------|-------|---------|---------------|----------------|
| (20) | X-ous | NOUN | + <i>-ity</i> | # <i>-ness</i> |
| | a. | curious | *cury | curiosity |
| | b. | furious | fury | *furiousness |

The words *curious* and *furious* not only sound almost perfectly alike (they differ only in their first consonants), they also belong to the same word class: they are both adjectives. Yet, although both of these adjectives are readily nominalised with the suffix *-ness*, the nominaliser *-ity* can attach only to *curious* — the word **furiousness* does not exist or is extremely unnatural, while *curiosity* is perfectly fine. (When a particular morphologically derived word form is given an asterisk, this should usually be taken to mean that the word in question either does not occur or is extremely rare. It is seldom the case that there is absolute ungrammaticality in the case of morphological derivation.)

The American morphologist Mark Aronoff discovered that the facts in (20a–b) are part of a general pattern, further illustrated in (20c–j). Whenever, for a particular adjective, there exists a simple noun that is semantically related to the adjective, nominalisation of the adjective via suffixation of *-ity* is impossible; but nominalisation via suffixation of *-ness* is perfectly fine across the board. In the concrete case of the forms in (20a) and (20b), we see this generalisation come to the fore very clearly: the adjective *furious* has a corresponding simple noun *fury*, but *curious* does not (**cury*; note that *cure* does exist as a noun, but it is semantically unrelated to *curious*); concomitantly, while *curiosity* is possible, **furiousness* is not.

	X-ous	NOUN	+ity	#-ness
(20) c.	amorous	*amour	amorosity	amorousness
d.	generous	*genery/*genour	generosity	generousness
e.	precious	*precy/*prece	preciosity	preciousness
f.	specious	*specy/*spece	speciosity	speciousness
g.	acrimonious	acrimony	*acrimoniosity	acrimoniousness
h.	arduous	ardour	*arduosity	arduousness
i.	glorious	glory	*gloriosity	gloriousness
j.	spacious	space	*spaciosity	spaciousness

Aronoff's insight was that we are dealing here with a case of BLOCKING: the complex form **furiousity* is blocked by the existence of the simple form *fury*; but for *curiosity* there is no blocker. Blocking gives us a good purchase on the distribution of *-ity* constructs. (You might wonder why *preciosity* is not blocked by *price*, or *speciosity* by *species*. As in the case of *curiosity* and *cure*, the answer is that the simple nouns are not semantically (near-)equivalent to the *-ity* forms. For *monstrosity* and *monster*, this claim is harder to uphold, however.) In essence, the idea is that simple nouns block *-ity* forms because they compete with one another for the privilege of being listed in the LEXICON, which is assumed not to tolerate true SYNONYMY, and therefore not to allow the listing of multiple words that have the same meaning.⁵ In this light, we can understand why the simple noun *fury* blocks *furiousity*, whereas *curiosity* is not blocked. Inevitably, the simple noun *fury* is listed in the lexicon of English, because it is a root, and all roots are lexically listed; and since *fury* is in the English lexicon as a noun, it does not tolerate any derived forms besides it that have the same meaning. The word *furiousity* could only exist in the lexicon of English if it were not synonymous with *fury*; otherwise, *fury* blocks *furiousity*. But in the case of *curiosity*, no blocking could ever jeopardise its lexical listing: there is no semantically equivalent simple noun corresponding to this *-ity* form.

Now that we have a perspective on why certain *-ity* forms do and others do not exist, let us ask why all of the *-ness* forms in (20) are possible. Aronoff's answer to this introduces an important point regarding the LEXICON: not every derived form is listed in it. *-ity* forms have special phonological properties: thus, note the vowel change in *curious~curiosity*, and the stress shift from the first to the third syllable; also note the fact that in *atrocious~atrocidity*, the syllable spelled as *ous* completely disappears in the *-ity* form (a case of 'truncation'). *-ity* forms frequently have specialised meanings as well — meanings that are not directly derivable from the corresponding adjectives. These phonological and semantic properties of *A-ity* forms compel us to perform *-ity* suffixation within the lexicon, so that it can participate in phonological and semantic rules whose application is confined to the lexicon. But *-ness* forms lack these special phonological and semantic properties: in *curiousness*, stress remains on the first syllable, just as in *curious*; in *atrociousness*, we do not witness the loss of the syllable spelled as *ous*; and all *-ness* forms fully preserve the semantics of their adjectival base. To Aronoff, this indicated that *-ness* suffixation does not take place in the lexicon, and that therefore the output of *-ness* suffixation is not listed in the lexicon: *-ness* forms are

5 In connection with this, think of the fact that we do not normally use the word *stealer* (regularly derived by attaching the agentive nominalising suffix *-er* to the verb *to steal*) because we already have at our disposal the simple noun *thief*. But in situations in which a person steals something but is not a thief, we do use *stealer*: thus, in baseball, a person who steals a base is indeed called a stealer (*He is the most successful stealer! *thief of third base in history*). In such situations, the morphologically derived form is not a true synonym of the underived form, hence the two do not compete.

fully productively formed, and no lexical rules apply to them. Cases of fully productive, non-idiomatic affixation do not have their outputs lexically listed. And precisely *because* their outputs are not listed in the lexicon, they cannot compete with elements that *are* listed there. So BLOCKING is expected never to happen in the case of *-ness* suffixation. And this is indeed what we find: any adjective can be nominalised with *-ness*, unrestrictedly. But for *-ity* suffixation, whose output *is* listed in the lexicon, we find the effects of blocking, as illustrated above.

As an aside, let us note that it is sometimes pointed out in the morphosyntax literature that a syntactic phrase can occasionally be ‘blocked’ by a word with the same meaning. Examples of the type in (21a) and (22a) are frequently cited in this connection.

- | | | | | |
|------|----|-----------|---------------|----------------|
| (21) | a. | walk | walked | *did walk |
| | b. | walk | *walkedn’t | didn’t walk |
| (22) | a. | big/fat | bigger/fatter | *more big/fat |
| | b. | corpulent | *corpulenter | more corpulent |

The basic idea here is the same as in the case of (20): if there is a simple form that can be used, it blocks the use of the complex form. But it should be clear that the facts here are far less straightforward than they were in (20): *did walk* does occur under certain circumstances (in particular, when *did* is emphatically stressed, or in contexts of subject–auxiliary inversion: *Did he walk or take the bus?*, *Never did he walk this far before*), and so does *more big/fat* (*He is more big/fat than tall*). Moreover, Aronoff’s rationale for blocking (i.e., competition between two synonymous forms that are both lexically listed) is not applicable in these cases: neither the output of *-ed/-er* suffixation (two completely regular processes) nor that of ‘*do/more*-support’ is lexically listed. To call (21a) and (22a) cases of blocking, therefore, would amount to a very significant stretch of Aronoff’s original concept (although, at a more general, level, it does fit in with the notion of blocking/specificity dating back to the classical Sanskrit grammarian Pāṇini). Although it remains a matter of debate whether treating (21) and (22) in terms of blocking is justified, these data do put their finger on the fact that there are interactions between morphology and syntax — reminders of the fact that morphology has very close ties with the neighbouring disciplines of linguistic enquiry.

II.10 The interaction between morphology and phonology

In the foregoing, we have seen morphology interacting closely with syntax and semantics. In the following paragraphs, we will examine the interaction between morphology and phonology.

Morphological affixation processes can have interesting consequences for the assignment of stress assignment, the pronunciation of the stem, and the pronunciation of the affixes involved. Let us illustrate this with reference to the data in (23) and (24):

- | | | | |
|------|----|----------------|------------------|
| (23) | | to démonstrâre | |
| | a. | démonstrâtable | undemonstrâtable |
| | b. | démónstrable | indémónstrable |
| (24) | | to repâir | |
| | a. | repâirable | unrepâirable |
| | b. | réparable | irrêparable |

The suffix *-able* can attach itself to the verbs *to demonstrate* and *to repair* yielding two discrete outputs. In the a-forms, nothing changes to the verb: the vowels and consonants of the stem are unaffected; primary stress remains on the last syllable of the input verb; in addition to the primary stress on *dem*, the verb *to demonstrate* has a secondary stress on *ate*, and this secondary stress also remains untouched by *-able* suffixation in (23a). When the a-forms are negated, this is done with the negative prefix *un-*, familiar from the preceding discussion. This pronunciation of this prefix is blithely insensitive to the initial consonant of the stem that follows it: it is always [ʌn]

Next consider the forms in (23b) and (24b). Stress here has shifted to the vowel that is two syllables removed from the right edge of the word (the so-called ‘antepenult’): the *o* of *demónstrable* and the *e* of *(ir)réparable*. In concert with this, the vowel of the second syllable of *repair* is pronounced differently (no longer as [ɛə] but instead as [ə]; this is reflected in a change in the spelling of this vowel). Even more dramatically, in (23a) an entire syllable of the input verb disappears — the stem is truncated from *demonstrate* to *demonstr-*; the *ate* part is ‘eaten up’. And when we look at the form of the negative prefix in the b-examples, we see that it is no longer *un-* — instead, we find a prefix that has the vowel *i* [ɪ] followed by *n* (the same consonant we also find in *un-* and, for instance, *inelegant* or *inadequate*) or, in the case of (24b), by *r*. This *r* is the result of ASSIMILATION of the consonant of *in-* to the consonant at the beginning of *reparable*. Such assimilation is obligatory in (24b) (**inreparable* is impossible) but illegal in (24a) (**urrepairable*). All of the ‘special properties’ of the b-forms in (23) and (25) (stress shift, vowel change, truncation, selection of *in-* rather than *un-* as the negative prefix, and consonant assimilation) are connected: you cannot have one without any of the others.

The differences between the a- and b-forms in (23) and (24) are, ultimately, effects of the point in the morphological derivation at which the suffix *-able* is attached. This is schematised in (25). If *-able* is attached at point ①, directly to the verbal stem before the application of the stress rule, it affects the placement of word stress and can even truncate the stem; early attachment of *-able* makes it possible for the *in-* form of the negative prefix to be selected at point ②, which is prior to the application of the consonant assimilation rules, which *in-* can thus undergo. If, by contrast, *-able* is attached at point ②, after the application of the stress rule to the verbal stem, it does not change the placement of main stress and prevents *in-* from being inserted there; the negative prefix *un-* is chosen instead, but it can only be attached at point ③, hence cannot be input to assimilation.

(25)		①	STRESS	②	ASSIMILATION	③			
a.	stem	⇒⇒⇒⇒⇒	✓	⇒⇒	<i>-able</i>	⇒⇒⇒⇒⇒⇒⇒⇒	<i>un-</i>	<i>unréparable</i>	
b.	stem	⇒	<i>-able</i>	⇒	✓	⇒⇒	<i>in-</i>	⇒✓⇒⇒⇒⇒⇒⇒⇒	<i>irréparable</i>

The picture of the interaction between affixation and the rules of word stress and assimilation that suggests itself based on the discussion in the previous paragraph leads us to return to the derivation of the complex word *ungrammaticality*, which we already had occasion to examine before (recall (17b)). From our discussion of Aronovian blocking, we recall that suffixation of *-ity* (unlike *-ness*) sees its output stored in the lexicon. We also know that affixation of *-ity* to *grammatical* causes the placement of word stress to shift: from *grammátical* we get to *grammaticálicity*, with stress on the syllable spelled as *cal*. So this suggests that *-ity* is attached at point ① in (25). From (25), we had already taken away that *un-* is attached at point ③. Putting these two bits of knowledge together, we are led to conclude that in the derivation of the complex word *ungrammaticality*, the suffix *-ity* must be attached before the prefix *un-* is brought in: [*un-* [[*grammatical*] *-ity*]]. But the structure for

grammaticality that we arrived at earlier, and given in (17b), has *-ity* is placed outside *un-*, suggesting that *-ity* is attached after *un-*. This presents a paradox: *-ity* cannot be attached before *and* after *un-*. What to do?

Our earlier conclusion that *ungrammaticality* is structured as shown in (17b) is sound: attaching *-ity* before attaching *un-* would lead to prefixation of *un-* to a host that is a noun (thanks to the previous suffixation of *-ity*), which we know is generally impossible (**unstudent* is ill-formed despite the fact that it could very well be meaningful). So we hold firm on the structure of *ungrammaticality* given in (17b). But this structure, in which *-ity* is attached outside *un-*, entails that the word stress rule (causing primary stress to be placed on *cal*, the penultimate syllable) cannot be assumed to apply exclusively between points ① and ②. If we give the stress rule another chance after point ③, we are in the clear for *ungrammaticality*. Of course, this will make the neat picture sketched in (25) rather more messy, which is an unfortunate sacrifice to make. But we still have the assimilation criterion (remember the difference between *irreparable* and *unrepairable*) with respect to the consonant of the negative prefix. So the sacrifice on stress is not fatal; *not* making it (and insisting instead on systematic application of the word stress rule before point ②) would certainly cause irreparable harm.

For our purposes here, we can conclude that the so-called ‘antepenultimate stress rule’ of English, which applies to words with three or more syllables and assigns primary word stress to the third syllable counting from the end of the word (the ‘antepenult’), applies to the largest word-level construct that meets its structural description, in principle at any point in the derivation. This rule, then, does not interact with word formation in the way envisioned in (25). But the assimilation rule does interact with word formation, being fed by *in-* prefixation but never taking the output of *un-* prefixation as its input.

On this note, we conclude our rollercoaster ride through the world of morphology. Many more questions can be raised about the internal structure of words, but the preceding pages should give you a decent sense of the nature of the problems that morphology is trying to grapple with and the avenues that lead to their solutions.