referring to a considerably weaker form of alarm than *panic*), an onomasiological analysis would start with a general concept, FRIGHTEN, and list *all* of these verbs as its possible realizations. The difference between the two approaches corresponds to the difference between a dictionary and a thesaurus. As a semasiological tool, a dictionary is a list of words, and one accesses meanings through words. A thesaurus, on the other hand, is a list of concepts: for a particular concept, the thesaurus gives access to the different words through which the concept could be expressed.

Semasiological and onomasiological analysis are in no way exclusive: the semasiological approach emphasizes differences between lexemes, the onomasiological one similarities. Furthermore, both are necessary to a full description of the processes underlying communication. A complete description of linguistic performance will show how a speaker achieves the mapping between the concept or meaning she wishes to express and the word forms actually chosen: given the need to express the concept or meaning FRIGHTEN, for example, what are the onomasiological principles according to which one of the possible verbs listed above is chosen? For the hearer, however, a semasiological approach is called for. Hearing or reading the word *frighten* in a particular context, what is the meaning which the hearer will assign to this verb?

2.2 The units of meaning

Any attempt to associate meanings and forms needs to ask what the minimal meaning-bearing units of language are. Individual lexemes like *spider*, *crazy* or *elongate*, are, quite clearly, the best examples of units with individually describable meanings. But as we will see, we need to recognize meanings both above and below the word level, and ambiguities about the level of grammatical structure to which meaning is correctly attributed are not infrequent.

2.2.1 Words and morphemes

How can we determine what counts as a lexeme (word) in a language? Without a secure criterion of wordhood, it will be hard to decide – especially in unfamiliar languages – what units we should be trying to attribute meanings to. For European languages with a well-established tradition of literacy, this question usually does not arise: words are the units surrounded by spaces in standard orthography. This definition of 'word' will not take us very far, however, for two reasons. The first is that languages which have only recently been written down often have a very fluid practice of word-division. A meaning-bearing unit considered by one speaker as only *part* of a word will not infrequently be written as a separate word by another speaker (see Dixon and Aikhenvald 2002: 7–9

for details). Speakers of Northern Sotho (Niger-Congo; South Africa), for instance, show two ways of writing the sentence meaning 'we shall skin it with his knife' (Dixon & Aikhenvald 2002: 8, quoting Van Wyk 1967). The first is to put spaces between each *morpheme*:

(11)	*re*	*tlo*	*e*	*bua*	*ka*	*thipa*	*ya*	*gagwe.*
	1PLS	FUT	3SGO	skin	INST	knife	9	his

The second is to recognize three distinct orthographic words:

(12) *retloebua kathipa yagagwe.*

These differing practices may sometimes become conventionalized in such a way that closely related, typologically similar languages may adopt differing orthographic conventions. For example, Northern Sotho's relations, Southern Sotho (Niger-Congo, South Africa) and Tswana (Niger-Congo, Botswana) are usually written according to the convention in (11), while Zulu (Niger-Congo, South Africa) and Xhosa (Niger-Congo, South Africa), whose morphological structure is entirely equivalent to that of the other group, typically follow the convention in (12).

The second reason to be suspicious of writing as an indicator of wordhood is that orthographic practice itself is not even stable within longstanding traditions of literacy. An unbroken tradition of literacy links Modern and Ancient Greek. Yet Ancient Greek was written without any word-division, whereas modern Greek observes the norms familiar from languages like English. We would obviously not want to say, however, that Ancient Greek did not have words. Similarly, the reform of German spelling rules made standard (for a trial period) in German schools since 1998 resulted in strikingly different word divisions, as can be seen from the following list:

(13)	Old (pre-1998) spelling	Current spelling	
	eislaufen	*Eis laufen*	'skate'
	aufsein	*auf sein*	'to be up'
	gefangenhalten	*gefangen halten*	'keep prisoner'
	wieviel	*wie viel*	'how much'

Linguists have advanced many criteria for the demarcation of the word as an isolable linguistic unit. One common criterion is that of 'potential pause': words are units before and/or after which pauses can be found in spoken language. For languages like Chinese, which lack complex morphology, this criterion may be workable. But for languages which show even a small degree of morphological complexity, like English, it is clearly unsatisfactory. Thus, Dixon and Aikhenvald (2002: 11) point out that one may well pause at morpheme boundaries within a single word, for example '*it's very un-* <pause, perhaps including *um*> *suitable*.' (Similarly, expletives in English can be inserted within what we normally consider a single word: *abso-bloody-lutely*.) Bloomfield's famous definition of 'word' (1933: 178), as 'a *minimum free form*', i.e. the minimal unit which may appear on its own without any additional grammatical material, is clearly

insufficient: many canonical words like *the, of* or *my* do not usually appear alone, but must presumably be considered as fully fledged words.

In order to introduce some clarity into the confusion over wordhood, it seems necessary to distinguish two different levels on which words may be defined. The first is the phonological level. Here, divisions between words are determined according to the domain of application of phonological rules and processes. In Dagbani (Gur, Northern Ghana), for instance, the clearest description of stress can be given by assuming the existence of a unit – the **phonological word** – each example of which bears only one main stress, normally on the penultimate syllable (Olawsky 2002: 206). In order, therefore, to determine whether a given phonetic string is a phonological word in Dagbani, one need only count the number of main stresses: if the unit in question has more than one main stress, then it is more than one phonological word. Furthermore, the fact that the penultimate syllable is typically the tonic (accent-bearing syllable) allows us to determine *where* the word boundaries lie. Many languages are like Dagbani in calculating stress on the basis of phonological words: as a result, stress is typically a useful indicator of the phonological word. Other indicators are also found, however. Like Dagbani, Bare (Northern Arawak, Brazil) shows penultimate stress (Aikhenvald 1996: 494). But this language possesses an additional marker of phonological wordhood: aspirated consonants can only be found in word-initial position (Aikhenvald 1996: 494): as a result, given a string with *n* aspirated consonants, one is guaranteed of the presence of at least *n* phonological words.

The phonological level alone will often not be enough to demarcate word-boundaries. Thus, stress in Dagbani and Bare is only *mostly* on the penultimate syllable: exceptions are possible, and this can lead to ambiguity in word division. As a result, the *grammatical level* of wordhood must also be considered. Dixon and Aikhenvald (2002: 19) propose three criteria for the recognition of those linguistic units which are independent **grammatical words**: *cohesiveness, fixed order* and *conventionalized coherence and meaning*. The last criterion 'indicates that the speakers of a language think of a word as having its own coherence and meaning. That is, they may talk about a word (but are unlikely to talk about a morpheme)' (Dixon and Aikhenvald 2002: 20).

Question How reliable a criterion of grammatical wordhood is this? Do speakers ever talk about the meaning of morphemes?

What about the first two criteria, cohesiveness and fixed order? Ancient Greek (Indo-European, Eastern Mediterranean) provides a clear illustration of both. Ancient Greek verbs were obligatorily multi-morphemic, consisting of at least the elements root + inflection, as in the verb meaning 'cure', *therapeu-ō*:

(14) *therapeu-ō* 'I am curing'
cure-1SG.PRES.INDIC
therapeu-ete 'You (pl.) are curing'
cure-2PL.PRES.INDIC
therapeu-ousi 'They are curing'
cure-3PL.PRES.INDIC

These elements must co-occur: the verb root *therapeu-* cannot occur without an inflectional suffix, and the suffix cannot occur without a verb root. The combination of verb root and inflectional affix thus constitutes a word on the criterion of cohesiveness. These forms also illustrate fixed order, in that one cannot invert the order root-suffix: the inflectional markers are suffixes, not prefixes. As a result, the combination verb root + inflection constitutes an unambiguous grammatical word in Ancient Greek.

Mismatches between grammatical and phonological words

The criteria of grammatical and phonological word do not necessarily coincide, as can be shown by compound nouns in Georgian (Kartvelian, Georgia). Consider for example the following compound, constructed from the noun roots *t'ól* 'person of the same age group' and *amxánag* 'comrade':

(15) *t'ól-amxánag-i*
person.of.same.age.group-comrade-NOM
'comrades of the same age'

Based on considerations of cohesiveness and fixed order, this is a single grammatical word. Neither *t'ól* nor the suffix *-i* may occur on its own. Thus, the suffix *-i* is obligatorily an affix, and the root *t'ól* requires its own inflectional suffixes when it appears independently as a fully fledged noun. Similarly, the order of the elements of the word is fixed: the meaning 'comrades of the same age' is expressed by the form *t'ól-amxánag-i,* not (for example) **i-t'ól-amxánag* or **t'ól-i-amxánag*. *T'ól-amxánag-i* thus conforms to the criteria of cohesiveness and fixed order and constitutes a grammatical word. From the point of view of stress-assignment, however, (15) is *two* phonological words: Georgian phonological words take just a single primary stress per word (Harris 2002: 232–233), whereas *t'ól-amxánag-i* has preserved the stress of both of its original noun elements. Such mismatches between grammatical and phonological words are by no means the norm in the languages of the world. Nevertheless, their existence illustrates the problematic nature of the category 'word', which seems at first glance to be an entirely intuitive and straightforward concept.

If words are the clearest type of meaning-bearing unit in a language, they are certainly not the only ones: the domain of meaningfulness extends both above and below the threshold of the individual word. Below word level, morphemes, by definition, have meanings. Given the definition of a morpheme as the 'minimal meaning-bearing unit' of language, it is clearly impossible to conceive of a morpheme without a meaning – even if it is often hard to specify exactly what this meaning is. Quite often

in linguistic analysis, it proves surprisingly difficult to come up with a settled analysis of the meaning of a given morpheme. This is the case, for instance, with the meanings of the possessive suffix *-s* and of many morphemes involved in the verbal tense/aspect system in English (see 9.2): semanticists agree that these morphemes *have* meanings, but disagree about exactly what they are.

Above the level of the individual word, phrasal verbs and compounds are two clear cases where a single meaning is associated with a combination of lexemes. Phrasal verbs consist of one or sometimes two 'full' verbs followed by one or more particles, as in (16):

(16) *dispose of*, *touch down*, *play around*, *call off*, *set up*, *break down*, *put up with*, *get on with*, *look down on*, *make do with*…

Compounds are most clearly illustrated by noun compounds, which consist of two or more nouns conjoined into a single conventionalized semantic unit:

(17) *tree house*, *tennis match*, *instruction book*, *computer problem*, *space age*, *ink jet printer*, *car insurance contract*, *pedestrian underpass*, *junk food*, *garbage collection, zebra crossing, box office, hit man, getaway car*, *bullet train*, *knuckle sandwich*…

QUESTION Noun compounding is an extremely frequent means of word-formation in English, and shows many different types of meaning relation between the compounded elements: *a tree house* is a type of house in a tree, but a *lighthouse* is a type of 'house' which contains a light, and a *poorhouse* was an institution for the accommodation of the poor. A *computer problem* is a problem *with* a computer, and a *zebra crossing* is a crossing that is striped *like* a zebra. Find twenty examples of noun compounds from a newspaper, and describe the semantic relationships between the constituent parts. Can you discern any regularities?

Idioms, discussed in 1.4.3 in relation to *throw in the towel*, also demonstrate the existence of units of meaning associated with several words simultaneously, and we will consider the question of the meaning of grammatical constructions in a later chapter (10.3). Thus, although we most often think of meaning as something belonging to individual words, we must actually recognize that words are only the most obvious of a number of meaning-bearing units.

2.2.2 Meanings below the morpheme: sound symbolism

The question of what level of grammatical structure a meaning should be attributed to may often be problematic, and boundary cases, where meanings seem to straddle several different grammatical units, occur quite frequently. One such boundary case is **sound symbolism**, (also known as **ideophony** or **onomatopoeia**). This is the existence of semi-systematic correspondences between certain sounds and certain meanings, usually within the domain of the individual morpheme, such as English *clash*,

clang, *clatter,* etc. Such associations may sometimes have a clear imitative basis, as with English *click*, *thwack*, *meow*, etc. Sound symbolism is by no means limited to English, of course. In Ilocano (Cordilleran, Philippines), for instance, a high front vowel is often used in words denoting high pitched sounds, as in (18):

(18) *singgit* 'high pitched voice'; *sing-i* 'sobbing (of a child)'; *sultip* 'whistle'; *riri* 'whimper' (Rubino 2001: 304).

Here the choice of vowel imitates the characteristic timbre of the sound referred to. Similarly, the alveolar fricative is often found in words representing rustling sounds or the sound of water:

(19) *karasakas* 'rustling sound of leaves'; *karasikis* 'rustling sound of bamboo'; *kiras* 'sound of slippers'; *saraisi* 'sound of rippling water'; *barasábas* 'sound of heavy rain, downpour'; *barasíbis* 'sound of light drizzle, drizzle'; *dissuor* 'waves breaking' (Rubino 2001: 305)

A possible connection might be discerned here between the acoustic quality of the fricative and the irregular, 'perturbed' sound of the referent. But the imitative basis of such associations is often less obvious, at least to English speakers. Egbokhare (2001: 90–91), for example, documents the fact that many words indicating 'smallness' contain *kp* in Emai (Niger-Congo, Nigeria):

(20) *kpúkú* 'pointed/protruding'; small, compact and round, short
kpútú 'stumpy'; small, compact and round, disproportional
kpúshú 'stubby'; small, compact and round, rough
kpódó 'round'; small, circular and supple, proportional
kpúdú 'pellet-like'; small, compact and round, proportional
kpédé 'proportionate'; small-sized, firm, proportional
kpéké 'petit'; small, thin, short.

In all these cases we have a sound-meaning correspondence which exists *below* the level of the individual morpheme. Neither the high front vowel nor the alveolar fricative in Ilocano, nor *kp* in Emai can, formally, be considered as individual morphemes, since one cannot remove them from the ideophonic words in (19)–(20) and retain possible roots to which other morphemes could attach. Yet the correspondence is widespread: although not every *s* in Ilocano is used in words referring to rustling sounds (cf. *sarotsot* 'quick succession', Rubino 2001: 315), the correspondence is systematic enough to allow a hearer who is unfamiliar with *karasakas*, for instance, to infer that the word probably refers to some sort of sound. Reference to a rustling sound can therefore be considered as, in some way, a semi-predictable part of the meaning of a unit which is neither a word nor a morpheme. Yet it is only in the words in which they occur that this meaning exists: in describing sound symbolism in Emai, it is necessary to specify that there are many words containing *kp* which do *not* refer to small objects (e.g. *úkpun* 'cloth'; *ókpósó* 'woman'; Schaefer 2001: 344). Sound symbolism can therefore be considered simultaneously as a

property of a word and of the relevant submorphemic unit, and the description of sound symbolism in these languages must invoke both lexical and submorphemic units: the reference to sound is conveyed by a particular segment or sequence of segments, but only in certain words.

2.2.3 Meanings above the word level: idioms

Idioms constitute another boundary case where it is not clear what the correct level is for the characterization of meaning. We defined idioms in 1.4.3 as non-compositional phrases – phrases like *throw in the towel* whose overall meaning is not the same as the combined meaning of the individual parts. However, it is often possible to advance an interpretation of the individual words of an idiom which removes its idiomatic or non-compositional character. For example, the English idiom *to scoop the pool*, which means something like 'to win or gain everything', seems on the face of it to lack any connection whatsoever with either pools or scooping: a speaker simply associates the meaning 'win or gain everything' with the entire unit *scoop the pool*, without trying to break the phrase down further. Nevertheless, if we imagine *scoop* as having a meaning like 'quickly gather up a large quantity of something in a single movement', and *pool* as meaning 'the entire set of available items' (cf. *car-pool*, *pool of credits*, etc.), then the arbitrariness and non-compositionality of the expression is reduced, and the interpretation 'win or gain everything' can follow unproblematically from the combined meanings of the expression's elements. For an empirical inquiry, everything hangs on the question of whether speakers *do in fact* interpret *scoop the pool* compositionally or non-compositionally, and there is doubtless no single answer to this question. Thus, some English speakers will analyse it completely into its constituent parts in the way just mentioned, others will interpret it as a single, non-compositional idiom, and still others will interpret it as partly compositional: the 'quickly gather up' interpretation of *scoop*, for instance, might be 'active' for some English-speakers, while *pool* will not receive any compositional interpretation. The fact that a *variety* of possible interpretations is available for each component of the idiom, with consequent differences in the overall interpretation of the expression, only adds to the ambiguity. Thus, other speakers of English might associate *scoop* with a *scoop* in journalism (a news story obtained exclusively by a single journalist), while others might analyse *pool* as in some way referring to a body of water.

As we have been using the term, an idiom is a non-compositional combination of *words*. But if we define an idiom as a non-compositional combination of *morphemes*, then idioms can also exist on the sublexical level. The English suffix *-able* is a case in point. Usually this suffix has its historical meaning, 'able to be V-ed': *fillable* 'able to be filled', *emailable* 'able to be emailed', *movable* 'able to be moved'. In words like *considerable* and *fashionable*, however, this meaning is not present, and the entire word needs to be given a different analysis. Sublexical idioms are often found in many American languages, which are characterized by a large degree of noun-incorporation, a process in which independent noun stems may

be compounded with a verb stem in order to produce a larger, derived verb. In the following example from Lakota (Siouan, Mississippi Valley; Rankin *et al.* 2002: 181–182), a noun stem meaning 'heart' is compounded with the verb stem meaning 'be good'; the meaning of the resulting compound, 'I made him/her angry', is in no way simply the combination of the individual meanings of its component morphemes:

(21) *blučhą́lwaxtešni*
Ø- *b-* *yu-* *čhą́t-waxte=šni*
3OBJ- 1ACTR-BY.HAND heart-be.good=NEG
'I made him/her angry'

Not all noun-incorporation is as semantically opaque or idiomatic as this, but there are many less extreme examples. An interesting one comes from another American language, Comanche (Uto-Aztecan, Oklahoma). Thus, the composed meaning of the noun-verb compound in (22) is something like 'throw paper by force'. This verb can only be used, however, to refer to the type of paper-throwing that one does when playing cards: the meaning of the incorporated noun *wana* is 'paper', but in the verb in question it only designates playing cards. As a result, the compound means 'to gamble' (Mithun 1984: 855):

(22) *wana-roh-peti-*
paper-by.force-throw
'to gamble'

2.2.4 Contextual modulation of meaning

The examples of noun-incorporation we have just seen show the meaning of words and other morphemes varying according to their **collocation**, the immediate linguistic context in which they occur. This sort of variation is found throughout language. We can see a similar phenomenon in English, where the meanings of verbs seem to vary slightly depending on the noun which they govern. If I *cut my foot*, for example, I am doing something that is rather different from what I am doing when I *cut the grass*, or when I *cut a cake, cut someone's hair, cut the wood, cut a diamond, cut a deck of cards, cut a disc* or *cut a notch*. The nature of the event, the means by which it is accomplished, its typical object, and the extent to which it is deliberate may all vary in these different uses. Despite this variation, we have the strong sense that essentially the 'same' meaning of *cut* is involved in all those cases (in other words, we do not usually think of this verb as polysemous; see 5.3). Cruse (1986: 52) refers to this phenomenon as the **contextual modulation** of meaning. The degree of semantic 'distance' gets even greater if we consider more 'extended' meanings, like *cut a deal*, *cut corners*, *cut a paragraph* or *cut prices.*

This type of phenomenon poses an interesting descriptive and theoretical problem: do the differences in meaning of the different collocations arise compositionally or not? Are the meanings of the collocations just the results of the combinations of the meanings of their parts, or are the