Chapter 2

Words

All languages have **words**, and most of us feel that we know what a word is. A wide variety of modern written languages mark word boundaries by means of spacing. But even English is not consistent in this respect. Should one, for instance, write *tax payer, tax-payer* or *taxpayer?* Why, according to the *Oxford English Dictionary*, is *road map* written as two words, while *roadwork* is written as one, and *road-horse* with a hyphen? And why are *today* and *tonight* written in one word, while *to hand* and *to work* (meaning 'at work' in colloquial American English) are written in two words? Such idiosyncracies of English orthography prompt us to ask: What is a word?

One possible answer is that a word is a meaningful unit with a relatively high degree of autonomy. Thus, *happier* is one word and *more happy* two, because *-er* is less autonomous than *more*, although in this context they have the same meaning. In other words, the degree of cohesion between different elements within a word is much higher than that between different words within a phrase.

The degree of autonomy of the word, as well as the degree of cohesion within the word, is reflected both in its grammatical and phonological properties. While *happier* is inseparable (nothing may be inserted between *happi-* and *-er*), *more happy* is not, as is clear from sentences like *After my divorce, I have become more, not less happy*. And while *redneck* has one main stress, *red neck* has two. Exactly how word boundaries affect grammar and phonology varies from language to language.

The term 'word' is useful in describing most languages in the world. The famous linguist Edward Sapir taught a speaker of the American Indian language Nootka to write his own language. The Indian had no difficulty in determining the words, and in the hundreds of pages of Nootka text that he wrote down, word boundaries are virtually always placed exactly where a linguist would place them.

The centrality of the word, however, varies a lot across the world. Many languages do not have a non-technical word for 'word'. Chinese is one of them, and written Chinese does not mark word boundaries. Even when the language is written with the Roman alphabet, it is much more common to insert a space between every syllable than between every word. There is much disagreement even among specialists on how to parse a Chinese sentence into words.

2.1 Word meaning

We noted above that meanings are mental concepts. The meaning of the word *ball*, for instance, is a mental idea of what it takes to be a ball, defined by the Oxford

English Dictionary as a 'globular body'. Of course, such concepts are assumed to have a reference, to refer to some kind of external reality, either in the physical world or in other mental or even spiritual realms. (What is, for instance, the reference of the word *God*?) Meaning and reference, however, are not the same. The meaning of the word *ball* is not the actual balls that happen to roll around (or fail to do so) in our physical world (or even in our fantasy or in some spiritual realm), but, again, our concept or idea of what it implies to be a ball, to be a globular body.

Many concepts are clearly **socially constructed**, formed by the child (or, sometimes, the adult) in the process of learning how to relate words and grammar to reality. To form a concept of a computer, for instance, one needs a certain experience or, at least, extensive explanation. In the case of *ball,* however, some linguists would argue that the concept of a globular body is not, or at least not only, socially constructed, but **innate**. While most of us have hardly ever seen a perfectly globular body, all of us have a concept of what it means for something to be one. Is that because we are born with this idea? Some would say yes and insist that

Is a non-globular ball a ball?

the basic elements of semantics are universal and innate. Others would say no and insist that even if the concept of a ball should turn out to be universal, that is not because of some innate idea, but because the experience of (nearly) globular bodies is common to mankind all over the world. Whether or not some basic concepts are innate, however, the complex webs of meaning in natural languages are undoubtedly socially constructed. An innate idea of a globular body will not help a speaker of English to understand how the word *ball* can be used to refer to an American football or an English rugby ball, none of which is particularly globular.

All concepts belong to one or more semantic **domains**, to larger areas of linguistic meaning. For instance, seconds, minutes, days, weeks, and years belong to the the *temporal domain*; red, green, blue, and yellow belong to the *colour domain*, offside and penalty kicks belong to the *football domain*, and ski wax and pistes belong to the *skiing domain*. Some domains, like the temporal domain and the colour domain, are universal, while others, like the football domain and the skiing domain, are only found in certain cultures. Domains are of many different kinds; some are quite general, while others are highly specific.

Dictionary definitions are often based on such domains. In the *Oxford English Dictionary*, the meaning of the noun *milk* is explained as follows:

The meaning of milk

A whitish fluid, rich in fat and protein, secreted by the mammary glands of female mammals (including humans) for the nourishment of their young, and taken from cows, sheep, etc., as an article of the human diet.

Among the domains that this definition is based upon are those of *color*, *anatomy*, *zoology*, *alimentation*, and *dairying*. Without some basic knowledge about these domains, we cannot understand all the aspects of this definition. Of course, not all speakers of English have such a broad understanding of what milk is. As children of 4

or 5 years we may only know that milk is a white liquid that is kept in the refrigerator and that our parents give us to drink in a cup.

Another way of studying concepts is by analyzing them into smaller components, often referred to as **semantic features** (or components or properties). Such analysis builds on the similarities and differences between various concepts. For instance, the concepts of 'man', 'woman', 'boy' and 'girl' all share the feature HUMAN, 'man' and 'woman' share the feature ADULT, 'boy' and 'girl' share the feature YOUNG, 'man' and 'boy' share the feature MALE, and 'woman' and 'girl' share the feature FEMALE:

		MALE	FEMALE
HUMAN	ADULT	'man'	'woman'
	YOUNG	'boy'	'girl'

The features MALE and FEMALE clearly belong within the same axis, the **semantic category** of gender, while the features ADULT and YOUNG belong within another axis, the semantic category of age.

Such analysis into semantic categories and features is useful in bringing out both parallels and differences between sets of concepts. For instance, there are both clear parallels and one clear difference between the concepts for human beings above and the terms for ovine, bovine and equine beings below:

		MALE	FEMALE
OVINE	ADULT	'ram'	'ewe'
	YOUNG	'lamb'	

		MALE	FEMALE
BOVINE	ADULT	'bull'	'cow'
	YOUNG	'calf'	

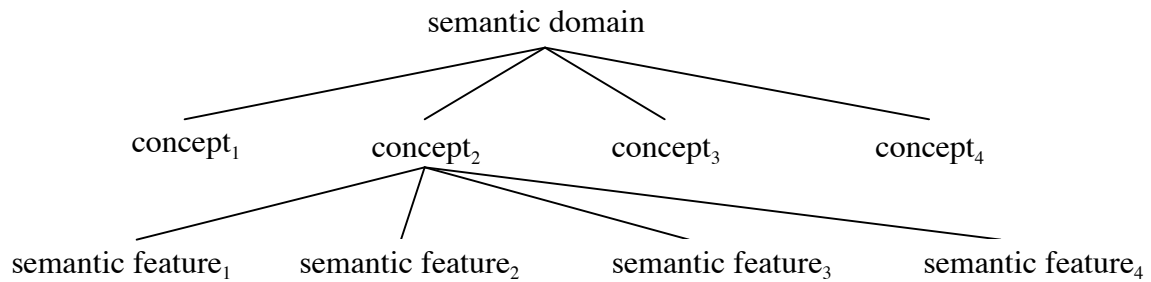|		MALE	FEMALE
EQUINE	ADULT	'stallion'	'mare'
	YOUNG	'foal'	

All four sets of concepts are analyzable in terms of the same semantic categories and features: gender (MALE vs. FEMALE) and age (ADULT vs. YOUNG). The difference is that young humans are subdivided into male and female, wheres the young non-human animals are not. In other words, the concepts for young human beings are **marked** (or specified) for gender, whereas the concepts young ovine, bovine and equine beings are **unmarked** (or neutral or unspecified) for gender.[1]

Strictly speaking, we should distinguish more clearly between concepts and the actual words of a living language. Although the English language does not

---

[1] Note some other differences between the four sets of terms analyzed here: In humans, the terms marked for gender and age are all commonly used in everyday language, while the unmarked term *person* is slightly more technical. In ovines and equines, the unmarked terms *sheep* and *horse* as well as the young terms *lamb* and *foal* are common, while *ram* and *ewe* as well as *stallion* and *mare* are more specialized. In bovines, there is no unmarked term at all, the species usually being referred to by the adult female term *cow*, though *ox* and *calf* are also common terms. Cf. the discussion of basic level terms later in this chapter.

distinguish between terms for 'male lamb' and 'female lamb', 'male calf' and 'female calf', 'male foal' and 'female foal', this does not necessarily mean that speakers of English do not have these mental concepts. Most speakers of English are certainly able to form such concepts. It is only that their vocabulary does not prompt them to do so.

From what has been said above, one might get the impression that each semantic domain contains a number of concepts, each of which may in turn be analyzed into semantic features:

semantic domain

$concept_1$ $concept_2$ $concept_3$ $concept_4$

$semantic\ feature_1$ $semantic\ feature_2$ $semantic\ feature_3$ $semantic\ feature_4$

In fact, however, semantic domains (like *color*, *anatomy*, *zoology*, *alimentation*, and *dairying*) are also concepts, and so are semantic features (like MALE, FEMALE, ADULT, and YOUNG). To a large extent, semantic analysis implies analyzing concepts in terms of other concepts.

Furthermore, the difference between semantic domains and semantic features should not be exaggerated. In fact, features like HUMAN, OVINE, BOVINE, EQUINE, MALE, FEMALE, ADULT, and YOUNG may also be used to form classes of concepts that share the same feature, so-called **semantic classes**. These are very similar to what we have called semantic domains.

2.1.1 Lexicalization

In addition to being shaped by our social experience and our genetic endowment, our concepts are also partly shaped and reinforced by our language. Different languages **conceptualize** the same reality in different ways. English, for instance, treats blue and green as different colours, while other languages, like the Mexican language Tarahumara, treat them as variants of the same colour. English has one term for 'cousin', while many European languages distinguish between male and female cousins (French *cousin* vs. *cousine,* Spanish *primo* vs. *prima,* German Vetter vs. Base), and Chinese has eight different terms for 'cousin' based on gender (male vs. female), age (older vs. younger) and whether or not they have the same family name (paternal vs. maternal).

Learning a new language, therefore, often implies learning new conceptual distinctions, like the one between 'blue' and 'green', 'male cousin' and 'female cousin', 'paternal cousin' vs. 'maternal cousin', etc. In other cases, we learn to look away from conceptual distinctions routinely made by our own language. Speakers of French, Spanish and German who learn the English word *cousin,* also learn a new concept, viz. one that is unmarked with regard to gender. Chinese-speaking learners of English will have to learn to look away from not only the distinction between male and female, but also that between older and younger, as well as paternal and maternal cousins, since Chinese also lacks an unmarked term corresponding to English *cousin.* And while English makes a conceptual distinction between 'blue' and 'green', it lacks

an unmarked term corresponding to Tarahumara *siyóname* (which covers both blue and green).

Basically, language contributes to conceptualization in two ways, through lexicon and grammar. A concept is given **lexical expression** (or it is being **lexicalized**) when a word (or some other lexical item) is used to represent it, as when German *Vetter* and *Base* lexicalizes the concepts of 'male cousin' and 'female cousin'. A concept is given **grammatical expression** (or it is being **grammaticalized** or **grammaticized**)[2] when a morphological or syntactic construction is used to represent it, as when English marks the conceptual distinction between 'substance' and 'object' by distinguishing between mass nouns (*water, clay, love*) and count nouns (*house, sculpture, idea*). Morphologically, mass nouns have no plural form, while count nouns do; syntactically, mass nouns cannot be directly preceded by numerals or words like *many,* while count nouns can. A large number of languages across the world do not make an equally clear distinction between the two types, because their nouns do not have plural forms and may not be directly preceded by numerals, and because the same word is used for *much* and *many*. We will come back to the distinction between mass nouns and count nouns later in this chapter.

Whether or not a concept is lexicalized in a given language is not always an either-or question. It makes sense to talk of **degrees of lexicalization**. For instance, it has been proposed that all languages lexicalize the concepts of 'man' and 'woman'. In most languages this is done by means of simple words like English *man* and *woman*.[3] The corresponding Chinese words, however, are complex terms consisting of the word *nán* 'masculine' or *nǚ* 'feminine' plus the word *rén* 'person':

> *nán-rén* 'man'
> *nǚ-rén* 'woman'

Japanese goes one step further and adds the grammatical particle *no* between *otoko* 'masculine' or *onna* 'feminine' and *hito* 'person', marking clearly that each expression consists of two separate words (in addition to the particle), the first of which modifies the second:

> *otoko no hito* 'man'
> *onna no hito* 'woman'

Even in the Japanese case, however, one may still argue that *otoko no hito* and *onna no hito* are fixed expressions, and that the concepts of 'man' and 'woman' are lexicalized, although they are represented by fixed, idiomatic phrases rather than single words. The degree of lexicalization, however, is much weaker than in English. We get the following correspondence between linguistic form and degree of lexicalization:

---

[2] The terms 'grammaticalization' and 'grammaticization' are, however, more often used in another meaning, viz., the historical development of a word into a grammatical marker, as when the English numeral corresponding to modern *one* developed into the article *an* or *a*.

[3] Historically, even *woman* is a complex term, going back to Old English *wīfman,* from *wīf* 'woman' and *man* 'person'.

Degrees of lexicalization	STRONG	MEDIUM	WEAK
Linguistic form	simple word	complex word	fixed phrase
Examples	*man* *woman*	*nán-rén* *nǔ-rén*	*otoko no hito* *onna no hito*

The difference in degree of lexicalization is reflected in actual language use. While the English terms *man* and *woman* are often used even if the gender of the person involved is irrelevant, the Chinese terms *nán-rén* / *nǔ-rén* and the Japanese terms *otoko no hito* / *onna no hito* are normally only used when the gender of the person is highlighted. In contrast to its Chinese and Japanese equivalents, the English word *person* sounds slightly formal and technical. Thus, although it may be true that the concepts of 'man' and 'woman' are lexicalized in all living languages, they are so to different degrees and in different ways.

2.1.2 Prototypes

In the **semasiological situation**, our perspective goes from language to the world. Let us imagine that you are talking to Li, a native speaker of Chinese, in his home. You ask him to show you an appropriate instance of the kind of object that is designated by the Chinese word *bēizi* 'drinking vessel'. Hopefully, Li will show you a drinking vessel of some kind, but there are many kinds of drinking vessels. They vary in material, size, and shape, they come with and without a handle or two, with and without a lid etc. Some of the variation is shown in PICT. 1.

PICTURE 1. Drinking vessels

Probably, Li will be of the opinion that some drinking vessels are better representatives of the category that *bēizi* is used about than others, and, since he wants to help you in your study of the Chinese language, he will show you a good representative of the category, an instance that we may talk about as a **prototypical** instance. Most concepts have prototypical instances, or **prototypes**, entities to which a word is typically applied. We may guess that a *bēizi* made of porcelain (also known as china!) will be closer to Li's idea of the prototypical *bēizi* than those made of glass, paper, silver or some other material. He will probably point to something resembling a cup, like one of those in the upper row in PICTURE 1, especially the two in the middle. In fact, *bēizi* is translated into English as 'cup' in most Chinese–English dictionaries, but this translation is based on the prototypical meaning, and in fact all the drinking vessels in PICTURE 1 can be referred to as *bēizi*. It is tempting to translate *bēizi* into English as 'drinking vessel, and in particular a porcelain cup' as a reflection of its usage.

Just as concepts may differ from language to language (*bēizi* covering both English *cup* and *glass*), so may prototypical representatives of the same concept. In one study, for instance, it turned out that native English speakers in Britain tended to regard the potato as a protypical vegetable, while many immigrants in the country did not regard it as a proper vegetable at all.

2.1.3 Hyponymy

In the **onomasiological situation** our perspective goes from the world to language. You show Li a cup, and ask him to tell you what Chinese word(s) he would regard as an appropriate name or designation for it. We expect him to mention at least the word *bēizi*. But there are many other possible designations. Li might have called it *chábēi* 'teacup', *qìmǐn* 'vessel', or even *dōngxi* 'thing, object'! The four words *chábēi*, *bēizi*, *qìmǐn*, and *dōngxi* may all be used correctly about the cup. These words belong to different *levels* in a system of classification, with an increasing degree of generality: teacups, cups, and vessels are kinds of things; teacups and cups are kinds of vessels, and teacups are kinds of cups. In semantics, a less general term is a **hyponym** of the more general term, and this phenomenon is referred to as **hyponymy**. *Jiǔbēi* 'winecup' and *chábēi* 'teacup' are hyponyms of *bēizi* 'drinking vessel (and in particular a china cup)', which in turn is a hyponyms of *qìmǐn* 'vessel'. The more general term is referred to as a **hyperonym** or simply a **superordinate** term, so that *qìmǐn* 'is superordinate to *bēizi*, and *bēizi* is superordinate to *jiǔbēi* and *chábēi*.

Li did not refer to the cup as *bēizi* just by coincidence, but because cups are habitually called *bēizi* in Chinese. In the hierarchy of terms that can be applied to cups, *bēizi* is the **basic level term**, the term used unless there are good reasons to do otherwise. The basic level is the highest level at which we can form a mental image of a concept. We can easily form a mental image of a teacup or a cup, but vessels vary too much in shape, and we have to go back down to the basic level. Terms above the basic level tend to have a technical or scientific flavor, which is the case with Chinese *qìmǐn* 'vessel'.

Let us imagine that you are asking Aamadu, a native speaker of Fula, what he sees on PICT. 2. He will probably tell you that it is a *nagge* 'cow'. But he could also have called it *jamale* 'cow with big patches of black and white' or *wumale* 'cow without horns'. Alternatively, he could have stated that it is a picture of *ndabbawa wuro* 'domestic animal' (literally, 'animal of the village'), simply *ndabbawa* 'animal', or *huunde*, a noun meaning 'animal' or 'thing' more generally.[4]

But *nagge* is clearly the basic level term applied to the «entity» in PICT. 2 (just as *cow* is the basic level term in English). It is the term a native speaker of Fula comes up with when seeing the picture. The terms *jamale* and *wumale* are technical terms for cow experts (who, admittedly, there are lots of among the Fula, traditionally a people of cattle nomads), while simply telling that it is *ndabbawa wuro* 'domestic animal' or *ndabbawa* 'animal' sounds like a conscious attempt to be less informative than expected. We can easily form a mental image—or make a drawing—of a cow, while the same cannot be said of a domestic animal, which is a generalization across cows, horses, sheep and goats. These latter animals, on the other hand, are also easy to conceptualize.

PICTURE 2
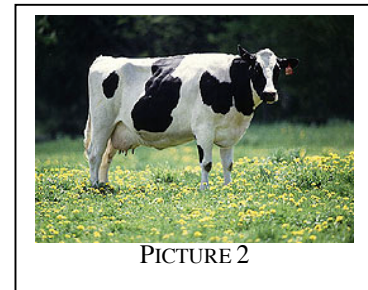
Basic level terms vary from language to language. For instance, both English *cow* and Fula *nagge* are gender-specific terms, while the corresponding basic level term in Chinese, *niú,* is not specified with regard to gender, and its hyponyms *gōng-niú* 'ox' and *mǔ-niú* 'cow' are clearly less basic. As it happens, it is a general tendency for Chinese basic level terms to be gender-neutral where English and many other languages have gender-specific terms. As we saw above, *man* and *woman* are basic level terms in English, the hyperonym *person* being slightly more technical in flavour, while the Chinese basic level term is the gender-neutral *rén* 'person', the hyponyms *nán-rén* 'man' and *nǔ-rén* 'woman' being less basic and therefore much less commonly used.

2.1.4 Polysemy

Meaning is not only about relations between language and the outside world, but also between elements within the language itself. We have already seen this in the case of hyponymy discussed above. In the present subsection, we shall look at three other types of semantic relations.

The first of these is **polysemy**, the phenomenon of a single word having several related meanings. Take, for instance, the English word *date,* which may refer

---

[4] *Huunde* means 'animal' explicitly in expressions like *huunde ladde* 'wild animal' (literally, 'animal/thing of the bush').

to, among other things, a point in time or an appointment (especially an appointment with romantic undertones). The relationship between 'point in time' and 'appointment [at a certain point in time]' is obvious; these meanings are clearly related. In these senses, therefore, *date* is only one polysemous word:

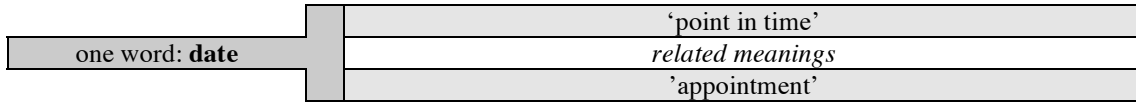|	'point in time'
one word: **date**	*related meanings*
	'appointment'

FIGURE 2. The polysemy of the English noun *date*: one pronunciation, two (or more) related meanings

The same word may even be used to refer to the person with whom one has an appointment, especially if the appointment is a romantic one. The relationship between 'appointment' and 'person with whom one has an appointment' is also clear, and again we have a case of polysemy.

Polysemy may be fruitfully contrasted with **homonymy**, the phenomenon of two words with unrelated meanings having, by accident, the same pronunciation. As it happens, the English spelling *date* (and its pronunciation [dɛit]) may also represent another word, which refers to the fruit that grows on a date palm. The meaning of this word is clearly unrelated to the meanings of the word *date* 'point in time; appointment; person with whom one has an appointment'. The two words *date* are homonyms:

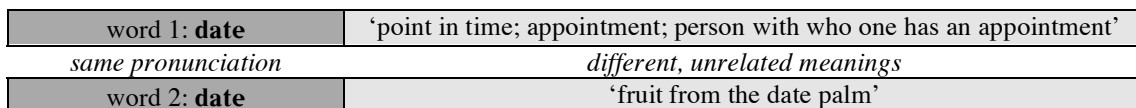| word 1: **date**	'point in time; appointment; person with who one has an appointment'
same pronunciation	*different, unrelated meanings*
word 2: **date**	'fruit from the date palm'

FIG. 3. The homonymy of the two English words *date*: one pronunciation, two unrelated meanings

Both words are in fact polysemous, the fruit term also sometimes referring to the whole fruit tree (the date palm). Homonymy, therefore, holds between independent words each of which may be polysemous:

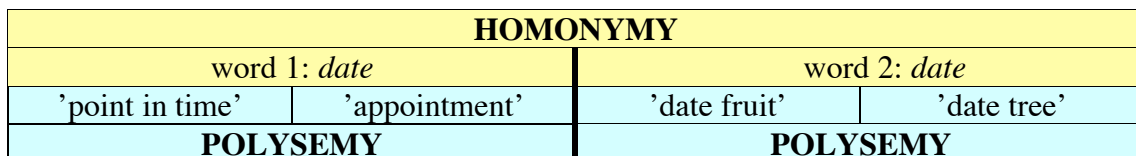| HOMONYMY			
word 1: *date*		word 2: *date*	
'point in time'	'appointment'	'date fruit'	'date tree'
POLYSEMY		POLYSEMY	

FIG. 4. The homonymy and polysemy of the two English words *date*

Actually, homonymy is not a semantic relation at all. The meanings in question are unrelated, though they happen to be expressed through the same form.

When different words (with different pronunciations) have identical meanings, we have **synonymy**. The words *date* (in one of its meanings) and *rendezvous* are synonyms:

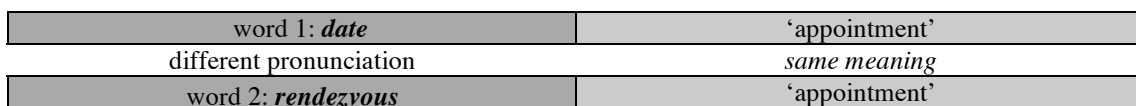| word 1: ***date***	'appointment'
different pronunciation	*same meaning*
word 2: ***rendezvous***	'appointment'

FIGURE 5. The synonymy between English *date* and *rendezvous*: two pronunciations, one meaning

However, *date* and *rendezvous* are not totally synonymous, only partially so. Not only do they have different connotations, a *date* being much more casual than a

rendezvous, but the two words do not have the same set of related meanings. The word *rendezvous* does not have the meaning 'point in time', but it does have other meanings that *date* lacks, such as 'an appointed place of meeting' and 'a base for naval ships or for military units'. Both words are polysemous, but in different ways:

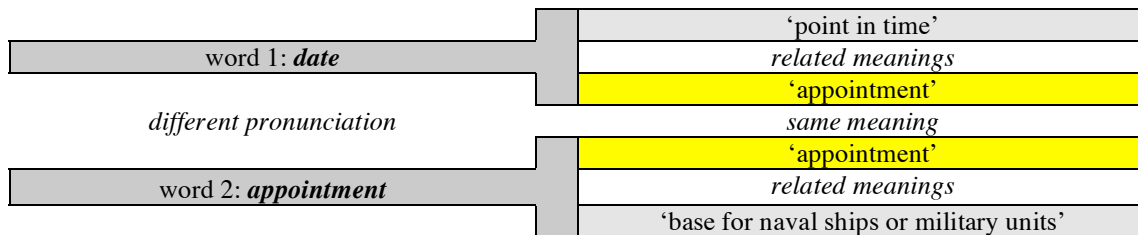|	
	'point in time'
word 1: ***date***	*related meanings*
	'appointment'
different pronunciation	*same meaning*
	'appointment'
word 2: ***appointment***	*related meanings*
	'base for naval ships or military units'

FIG. 6. The partial synonymy between English *date* and *appointment*

In natural languages, total synonymy is very rare, if it exists at all. Synonymy nearly always means **partial synonymy**.

The English noun *foot* is highly polysemous, and has at least the following five meanings:

---
*The meanings of **foot***

I. 'the part of the leg beneath the ankle'
II. 'a length unit; 12 inches'
III. 'a rhythm unit in poetic meter; a stretch of syllables of which one has primary stress'
IV. 'lower part of an object (such as a mountain)'
V. 'the part of a stocking that covers the foot'

On the basis of *foot*, we shall take a closer look at how **semantic relatedness** can be described in a more precise way.

The meaning 'the part of the leg beneath the ankle' is placed as *number I* on the list. This is the **primary** (or *basic* or *central*) **meaning**, from which the other **extended meanings** are most naturally derived. There is an associative relationship between the primary meaning and meaning number II, since the length unit *foot* is originally based on the length of the body part. There is also an associative relationship between the primary meaning and meaning number III, since the rhythm unit *foot* reflects the common habit of using the foot to beat the rhythm while reciting poetry. Both of these relationships are instances of **metonymy**:

Metonymy: an expression that basically designates one entity comes to be used of another entity in the same domain. The two entities are closely associated in space, time or cause, or by one being a part of the other.

Cases where one entity is part of the other are common. In now obsolete usage among anatomists, *foot* could refer to the whole leg. This is referred to as **pars-pro-toto**.
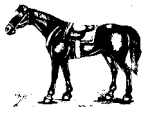
The relationship between the primary meaning and meaning IV is different. The foot of a person and the foot of a mountain are not associated in space, time or cause, or by one being a part of the other. Instead, this usage of *foot* reflects the fact that we may, to some extent, structure a mountain in terms of our own bodies. Just as human feet are (usually) the lowest parts of the human body, so the foot of a mountain is the lowest part of a mountain. This is called **metaphor**:

Metaphor: one domain is thought of or understood in terms of another domain.

The human body is one domain and the topography of mountains is another. When mountains are thought of in terms of the human body, it follows that the lowest part of a mountain may be called its *foot*. The foot of a human being and the foot of a mountain are analogous entities in two different domains. Metaphors are *mappings between different domains*, where the structure of one domain, the **source domain** (like the human body) is imposed on another domain, the **target domain** (like the topography of mountains). As a general rule, the source domain is a more familiar knowledge structure than the target domain.

An interesting example of structure imposition from a familiar source domain to a less familiar target domain is found in the Fula dialect of Maasina in Mali, where the bicycle is called *puccel njamndi*, literally 'little iron horse' (from *puccel* 'little horse' and *njamndi* 'iron'). This name indicates that the speakers have imposed a «horse structure» on the bicycle. On the basis of this creative and fruitful construal of a new technical device, the Fula speakers started using equine terminology for the different parts of the bicycle, as illustrated in TABLE 1.

TABLE 1. *Maasina Fula horse and bicycle terms*

TERMS	MEANING IN THE SOURCE DOMAIN	MEANING IN THE TARGET DOMAIN
koyngal	'foot, leg'	'wheel'
yitere	'eye'	'lamp'
kirke	'saddle'	'seat'
	HORSE	BICYCLE

Let us finally return to the English word *foot* and compare its primary meaning to meaning V. On the one hand, the part of a stocking that covers the human foot is spatially associated with the human foot; thus we may see this as an instance of metonymy. On the other hand, as in the foot of a mountain, this part of the stocking is also its lowest part, and the two entities belong to different domains (body part vs. clothing); thus we may also see this as an instance of metaphor.

2.2   Word classes

Words belong to different **word classes** (or parts of speech), each class being characterized by a mixture of semantic and grammatical properties. Different languages have different word classes, though some word classes seem to be more or less universal.

2.2.1 Nouns

The two most basic word classes are **nouns** (like *man, dog, sunflower, stone, water, clay* etc.) and **verbs** (like *kill, fall, grow, like* etc.). The distinction between nouns and verbs is found in most languages. A few languages have been reported to lack this distinction, though this is controversial. In the following word pairs from the native

American language Nootka, the same stem is used for nounlike and verblike meanings:

[qoˑʔas] 'a man' vs. [qoˑʔas-ma] '[he] is a man'
[wɑ̥lɑːk] 'a going; a trip' vs. [wɑ̥lɑːk-mɑ] '[he] goes'.

It remains true, however, that the vast majority of languages in the world do make a distinction between nouns and verbs.

The basic semantic properties of nouns and verbs are more or less the same everywhere. Nouns typically refer to **entities** (objects or substances), while verbs typically refer to **processes** (events or states). Both entities and processes may be **concrete**, such as the entity *ball* and the process *walk,* or **abstract**, such as the entity *democracy* and the process *become*. Furthermore, processes may be **entified** (seen as entities) and referred to by nouns, e.g. *singing* (in *his singing never ended*).

The grammatical characteristics of nouns and verbs vary a lot from language to language. In many languages, including English, nouns typically function as subjects and objects. English verbs may also be used as subjects or objects, but only in the infinitive (*to sing*) or the gerund (*singing*) forms. In English, nouns, but not verbs, may occur directly after the articles *the* and *a(n)*. And while English verbs are inflected for tense, mood, person and number, nouns are only inflected for number, if at all.

COMMON NOUNS VS. PROPER NOUNS

One group of words that is usually included among nouns is names like *Tom, Brazil, January* and *Adidas*. While ordinary nouns are referred to as **common nouns**, such names are referred to as **proper nouns**. Like common nouns, proper nouns refer to entities, though they do not have a general meaning, only a specific reference. Like common nouns, proper nouns typically function as subjects and objects. But unlike common nouns, they do not occur directly after the articles *the* and *a(n),* and they are never inflected for number. The term 'noun' is primarily used to refer to common nouns.

COUNT NOUNS VS. MASS NOUNS

In many languages, there is a clear distinction between **count nouns** and **mass nouns**. English count nouns, such as *horse*, share the following grammatical characteristics:

- they have a singular and a plural form (*horse* vs. *horses*)
- they may occur with words such as *many* or *few* (*many/few horses*)
- they may occur with numerals (*two horses*)
- they may occur with the indefinite article (*a horse*)

In contrast, English mass nouns, such as *smoke*, share the following characteristics:

- they do not have a plural form (**smokes*)
- they occur with *much* or *little* rather than *many* or *few* (*much/little smoke*)
- they do not occur with numerals (**two smokes*)
- they do not occur with the indefinite article (**a smoke*)

Count nouns typically refer to **objects**, while mass nouns typically refer to **substances**.

> • An **object** is a sharply delineated entity, and if it is divided into smaller pieces, these pieces are no longer instances of the same object.
> • A **substance** is not sharply delineated, and if it is divided into smaller pieces, these pieces are instances of the same substance.

While a horse cut in two does not give two horses, smoke remains smoke even if divided in two.

The distinction between count nouns and mass nouns plays a central role in the grammar of English, as it does in Swahili and many other languages of the world.[5] There are also many languages, however, where it plays at best a peripheral role, since their nouns make no distinction between singular and plural forms, they make no distinction between *many/few* and *much/little*, numerals never precede nouns directly, and they have no indefinite article. This is true of a large number of languages along the Pacific Rim, in East and Southeast Asia, Polynesia and native America.

CLASSIFIERS

In languages where the distinction between count noun and mass noun plays a peripheral role, numerals do not directly modify any nouns. In these languages, numerals are typically followed by a **classifier** (sometimes called a measure word). Classifiers resemble English nouns like *cup, piece* and *spray* in *a cup of tea, a piece of meat* and *a spray of flowers*, but there are also many differences:

1. Classifiers are used in connection with all nouns, not just mass nouns. It is as if one is saying *one piece [of] man* instead of *one man*.

2. The meaning of many classifiers is weakened. In a way, one does not really say *one piece [of] man*, since the word corresponding to *piece* often has no clear meaning at all.

3. Unlike English *cup, piece* and *spray*, classifiers clearly constitute a separate word class, and most of them are never used as ordinary nouns. In Chinese, for instance, the noun for 'cup' is *bēizi,* while the classifier for 'cup' is *bēi.*

PRONOUNS

Words that typically "stand for" a noun or a noun phrase are called **pronouns**. The most common pronouns are **personal pronouns**, like the following English forms:

---

[5] The correspondence between the count/mass noun distinction and the distinction between objects and substances is only approximate. In English, for instance, *cloud* is a count noun, although dividing a cloud in two does produce two clouds, and *furniture* is a mass noun, although cutting a piece of furniture in two does not produce two pieces of furniture. Note that many words have both count noun and mass noun usages, such as the mass noun *oil* (denoting the substance) vs. the count noun *oil* (pl. *oils*, denoting a type of oil).

		1st person	2nd person	3rd person		
				masculine	feminine	neuter
Singular	Subject form	*I*	*you*	*he*	*she*	*it*
	Object form	*me*		*him*	*her*	
Plural	Subject form	*we*		*they*		
	Object form	*us*		*them*		

As far as we know, all languages have personal pronouns. Personal pronouns are always distinguished according to person (1st, 2nd and 3rd), though quite a few languages have two variants of the 1st person plural, an inclusive form (the 'we' that includes 'you') and an exclusive form (the 'we' that excludes 'you'). Personal pronouns are also usually distinguished according to number (singular vs. plural; in some languages the dual and even the trial are given special forms), even in languages that do not distinguish singular and plural forms of nouns. The distinctions of case (subject form vs. object form) are much less common, though there are many languages with a much more fine-meshed case system than English.

A related set of pronouns are **reflexive pronouns**, which are typically used to refer back to the subject of the clause of which they are a part. Compare the following two sentences:

A. *She killed herself.*
B. *She killed her.*

In A, the reflexive pronoun *herself* must refer to the same person as the subject pronoun *she*, while in B, the personal pronoun *her* must refer to some other person than the subject pronoun *she*.

English reflexive pronouns are differentiated on the basis of person, number and gender: *myself, yourself, himself, herself, itself, ourselves, yourselves* and *themselves*. Many other languages have only one reflexive pronoun (corresponding to English *self*), but often use it in conjunction with personal pronouns.

The languages of the world differ from each other in the types of pronouns they possess. Consider the following two examples of such differences:

1. English has a number of possessive pronouns (*my/mine, your/yours, his, her(s), its, our(s), their(s)*), while other languages use personal pronouns plus possessive markers instead (cf. Japanese *watashi no* 'my/mine' from *watashi* 'I' and the possessive marker *no*).

2. English has four types of pronouns for which other languages, like Chinese, have only one. Depending on the context, many Chinese indefinite pronouns (*shéi* 'who(ever)', *shénme* 'what(ever)/which(ever)', *nǎ*/*něi* 'which(ever)') may be used where English uses interrogative pronouns (*who(m), what, which*), negative pronouns (*no, no one, none, nobody, nothing, neither*), universal pronouns (*all, every, each, everyone, everybody, everything*) and indefinite pronouns (*some, someone, somebody, something, either, any, anyone anybody, anything*).

It is not uncommon to call Japanese *watashi no* a possessive pronoun, it may be translated by English *my/mine*. It is evern more common to call Chinese *shéi*, *shénme*, *nǎ/něi* interrogative pronouns, since they most commonly occur in contexts where English would use *who(m), what* or *which*. By doing so, however, we are in fact imposing an English terminology on languages for which they are not suitable.

2.2.2 Verbs

DYNAMIC VS. STATIVE VERBS

Verbs may be **dynamic** or **stative**, depending on the kind of process they refer to:

• A **stative process** is a sequence of uniform states.
• A **dynamic process** is a sequence of heterogeneous states.

If you make a movie of a dynamic process, you get a sequence of pictures where you can observe a gradual change from one picture to the next. If you make a movie of a stative process, you get a sequence of identical pictures. In FIG. 7, the three pictures in each row represent three states following each other in time, with the first state being the one on the left and the last state the one on the right. The earth going round is a typical dynamic process, while a house standing on the ground at the end of a small road is a typical stative process.

A dynamic process	🌍	🌍	🌍
A stative process	🏠	🏠	🏠
	1st state	*2nd state*	*3rd state*

FIGURE 7. *A dynamic and a stative process*

Verbs like *run, die* and *roll* are dynamic, while verbs like *have* (in the sense of 'possess'), *love* and *know* are stative.

The difference between stative and dynamic processes plays an important role in the grammar of many languages. A number of grammatical constructions are fully acceptable with dynamic verbs like *run*, but less so with stative verbs like *know*:

1) Dynamic verbs have a progressive form, while stative verbs do not:[6]
    a) *John is running*.
    b) **John is knowing the answer*.
2) Verbs like *force* can be used with dynamic verbs, but not with stative verbs:
    a) *I forced John to run*.
    b) **I forced John to know the answer*.
3) Dynamic verbs have imperative forms, while stative verbs do not.
    a) *Run!*
    b) **Know the answer!*

---

[6] Note how the McDonald's chain makes the stative verb *love* into a dynamic verb by saying *I'm lovin' it*.

4) Dynamic verbs may occur in the "pseudo-cleft construction"; stative verbs may not:

    a) *What John did was run.*
    b) **What John did was know the answer.*

English has relatively few stative verbs, most of them denoting mental processes. Non-mental (as well as some mental) stative processes are usually expressed by adjectives (see below). In many languages across the world, however, there is no sharp distinction between verbs and adjectives.

LEXICAL VS. AUXILIARY VERBS

In English, a small subgroup of verbs called **auxiliary verbs** (or just auxiliaries) are clearly distinguished from other verbs (known as **lexical verbs**) in a number of ways:

1. Auxiliaries typically precede other verbs without the use of *to* (as in *I hate <u>to</u> say this*):
a) *You <u>may</u> go home.*
b) *I <u>have</u> seen you.*
c) *He <u>is</u> dancing.*

2. They form questions by inversion rather than by means of *to do* (as in *<u>Do</u> you study linguistics?*):
a) *<u>May</u> I go home?*
b) *<u>Have</u> you seen me?*
c) *<u>Is</u> he dancing?*

3. They are negated by the simple negator *not,* without the help of *to do* (as in *He <u>does</u> not study linguistics*):
a) *You <u>may</u> not go home.*
b) *I <u>have</u> not seen you.*
c) *He <u>is</u> not dancing.*

Many auxiliaries also have deviant patterns of inflection. While most languages have words resembling English auxiliaries in meaning and usage, the distinction between lexical verbs and auxiliary verbs is often much less clearly marked than in English.

THE COPULA

One verb with a very special position is the **copula**. A copula is used to link a subject and predicate, and it may imply a number of different semantic relations between the two, such as:

1. Identity: *When the area behind the dam fills, it will <u>be</u> a lake.*
2. Class membership: *He <u>is</u> a bachelor.*
3. Property: *The house <u>is</u> blue.*
4. Location: *I <u>am</u> here.*

Some languages allow other semantic relations to be expressed by the copula, as when the Chinese copula *shì* is used to express a possessive relation:

Zhāng xiānsheng shì ge   Rìběn tàitai.
'Mr. Zhang is [one who has] a Japanese wife.'

A few languages have no copula at all.

2.2.3 Adjectives

In English, **adjectives** also constitute a large and important word class.

Semantically, adjectives typically refer to **properties**: *red, hard, cold* etc. Like entities and processes, properties may also be abstract, such as *serious* and *brilliant*. And like processes, properties are sometimes entified and referred to by nouns, e.g. *happiness* and *brilliance*.

Syntactically, adjectives typically have two different functions, **attributive** and **predicative**. In English, attributive adjectives mostly precede the noun whose property they describe:

> <u>red</u> *dust*
> <u>hard</u> *times*
> <u>cold</u> *water*

The opposite is the case in many other languages, such as French:

> *jupe* <u>*noire*</u> 'black skirt'
> *film* <u>*interessant*</u> 'interesting film'

In English, predicative adjectives occur after verbs like *be, become, turn* and a few others:

> *The dust became* <u>*red*</u>
> *Times are* <u>*hard*</u>
> *The water has turned* <u>*cold*</u>

Other languages, such as Hungarian, do not require a verb before a predicative adjective:

> *Róbert öreg* 'Robert [is] old'

As in many other languages, English adjectives have a comparative form (either *happier* or *more happy*) and a superlative form (either *happiest* or *most happy*).

While nouns and verbs are more or less universal, there are many languages that do not have a significant word class corresponding to English adjectives. In Southeast Asian languages, for instance, most words corresponding to English adjectives are verbs. *Red, hard, cold* etc. are not only properties, but also states, and in this they resemble stative verbs, such as *like* or *fear*.

2.2.4 Adverbs

The English words called **adverbs** constitute a very heterogeneous group. Semantically, they cover a number of different meanings, such as:

manner: *slowly, badly*
degree: *very, somewhat*
time: *then, now*
place: *here, everywhere*
cause: *therefore, thus*
concession: *nevertheless*
condition: *otherwise*

Syntactically, they typically modify verbs (*walk <u>slowly</u>*), adjectives (<u>*really*</u> *noisy*), other adverbs (<u>*very*</u> *badly*), or whole clauses or sentences (*He <u>then</u> left the country*), but some words usually classified as adverbs are not modifiers at all (*go <u>home</u>*).

Some types of adverb are closely related to other word classes. For instance, English manner adverbs are usually derived from adjectives, most often by adding the suffix *-ly* (as in *slowly* and *badly*), and they resemble adjectives in having comparative (*more slowly*) and superlative forms (*most slowly*). They are also semantically related, since adjectives may be seen as describing the properties of entities, while manner adverbs may be seen as describing the properties of processes.

In English, there is also a strong connection between time and place adverbs and prepositions, the same word often being used as both adverb and preposition:

Adverb: *He went <u>up</u> and <u>down</u>.*
Preposition: *He went <u>up</u> and <u>down</u> the road.*

In such cases, one and the same word simultaneously belongs to two different word classes.

The class of adverbs, if it exists at all, is equally heterogeneous in other languages.

2.2.5 Adpositions

In their concrete meaning, English **prepositions** (like *on, at, in, to, until, from, of, off, over, above, under, below, with, by, about, among, before, after, behind, through* etc.) typically locate an entity in space or time:

<u>*under*</u> *the tree*
<u>*after*</u> *midnight*

But prepositions are also often used in more abstract senses:

<u>*under*</u> *surveillance*
he is named <u>*after*</u> *his grandfather*

The majority of prepositions may also be used as adverbs (see above). Prepositions are never inflected. When followed by pronouns, they are always followed by the oblique form (*me, him, her, us, them*) rather than the subject form (*I, he, she, we, they*).

In other languages, such as Japanese, words corresponding to English prepositions are typically placed after the noun:

Tookyoo <u>kara</u> 'from Tokyo'
Kyooto <u>made</u> 'to Kyoto'

Such words are called **postpositions**. Even English arguably has three postpositions:

ten years <u>ago</u>
ten years <u>hence</u>
ten miles <u>away</u>

Both prepositions and postpositions may be referred to by the common term **adposition**.

In some languages, adpositions do not constitute a separate word class. For instance, Chinese prepositions (like *cóng* 'from' and *zài* 'at, on, in') are a subclass of verbs, while Chinese postpositions (like *yǐqián* 'before' and *lǐmiàn* 'inside' are a subclass of nouns.

2.2.6 Determiners

In English, the class of **determiners** basically consists of the **demonstratives** *this/these* and *that/those* and the **articles** *a(n)* (indefinite) and *the* (definite). They typically stand at the very beginning or (in some languages) the very end of a noun phrase, making it more or less definite.

Many languages have no articles, choosing either to leave the noun unmarked for definiteness or to use a numeral meaning 'one' instead of the indefinite article and a demonstrative instead of the definite article.

Quite a few languages have a larger number of demonstratives than English. For instance, both Japanese and Thai make the following three-fold distinction:

	'this'	'that'	'that over there'
Japanese	*kore*	*sore*	*are*
Thai	*nîi*	*nân*	*nóon*

2.2.7 Numerals

Virtually all languages have **numerals**. Some languages are reported to have very few of them, such as the Australian aborigine language Pitjantjatjara, which apparently only has the numerals 'one', 'two' and 'three'. Only the Amazonian language Pirahã is reported to have no numerals whatsoever.

Like many languages, English makes a basic distinction between **cardinal numbers** (*one, two, three* etc.) and **ordinal numbers** (*first, second, third* etc.).

2.2.8 Conjunctions

Conjunctions are words that bind elements together within or beyond the sentence.

Coordinative conjunctions like *and, or, but* are used to bind together words, phrases, clauses, sentences or stretches of speech larger than the sentence. The elements on each side of a coordinative conjunction are conceived as being of more or less equal weight, and often (though far from always) it is possible to change the

order of the elements without changing the meaning of the whole expression: *you and me* vs. *me and you*.

   Subordinative conjunctions (sometimes called complementizers) include one subordinate clause within a superordinate clause:

> *If you touch me, I'll shout.*
> *He told me that you would come.*
> *A girl that I knew appeared in the newspaper.*

In the first example, the subordinate if-clause has the same syntactic function within the superordinate clause as adverbs like *tomorrow* and *otherwise*, and such clauses are called **adverbial clauses**. In the second example, the subordinate that-clause has the same syntactic function within the superordinate clause as noun phrases like *your story* or *many lies*, and such clauses are called **nominal clauses**. In the third example, the subordinate that-clause has the same syntactic function as an attributive, and such clauses are called **relative clauses**.

   In English, many subordinative conjunctions are also used as prepositions and/or adverbs, such as *before* and *after*:

> *He gave it to me before he left.* (subordinative conjunction)
> *He gave it to me before last summer.* (preposition)
> *He gave it to me before.* (time adverb)

But this is not necessarily the case in other languages.

2.2.9 Particles

The term **particle** is used about any small, unstressed grammatical word that does not readily fall into any other word class. In Japanese grammar, for instance, the term is often used about phrase-final case markers (like *wa* for topic, *ga* for subject, *wo* for object and *ni* for indirect object) and sentence-final mood markers (like *ka* for question).

2.2.10 Interjections

All languages appear to have a separate class of **interjections**, resembling English words like *wow, psst, mhm* and *wham*. The phonological make-up of interjections is often different from that of other words, and in any other word class, words like *psst* and *mhm* would have been impossible.

   Interjections may be divided into four basic subclasses: expressive (*ouch, oh, wow, aha*), directive (*hush, psst, hey*), phatic (*mhm, yes, no, huh*) and descriptive ideophones (*wham, thud, bang*).

2.2.11 Open vs. closed word classes

Of all the word classes examined above, some are **open** classes that readily admit new members, while others are **closed** and seldom admit new members. Nouns, verbs, adjectives, (some types of) adverbs and (in their own special way) interjections are

the clearest examples of open word classes, while determiners, pronouns, classifiers, adpositions, conjunctions and auxiliaries are all closed sets that do not easily change.

2.2.12 Content words vs. function words

To a large extent, the distinction between open and closed word classes coincides with the distinction between **content words** and **function words**. Content words are full words with a lexical meaning, denoting concrete or abstract entities, processes, properties etc. The meaning of function words, on the other hand, is more akin to the meaning of grammatical constructions. Usually,

Often, the meaning expressed by function words in one language is expressed by grammatical constructions in another language. For instance, the meaning expressed by English prepositions (like *in, at, from* etc.) is often expressed by special forms of the noun called case forms in Finnish and Hungarian.

Sometimes, even one and the same language has two alternative ways of expressing the same meaning, either with a function word or with a grammatical construction. For instance, *more* in *more happy* is equal in meaning to the comparative ending *-er*. The pronoun *he* is also very close in meaning to the verb ending *-s* (as in *[he] runs*), except that the latter contains no information about gender.

2.3 Word structure: inflection

Morphology is the study of word structure, of how smaller meaningful units are combined into words, as in English *sing-er-s*. Morphology may be divided into two branches, **inflection** and **word formation**. The present subsection will be concerned with inflection.

2.3.1 Lexeme and word form

We use the word *word* when talking about the vocabulary of the English language, as when we say *Some people believe that the English language has more **words** than most other languages, probably because English dictionaries have so many pages.* When talking about the word from the perspective of vocabularies and dictionaries, we shall introduce the term **lexeme**, reflecting the fact that linguists usually refer to the vocabulary of a language as its *lexicon* and that professional dictionary-makers are called *lexicologists*. We shall write lexemes with SMALL CAPITALS, usually followed by one or more subscript letters telling about the word class of the lexeme. For example, we have *nouns* like GIRL$_N$ and BOOK$_N$, *verbs* like SING$_V$ and WRITE$_V$, and *adjectives* like GOOD$_{Adj}$ and BAD$_{Adj}$.

In many languages, lexemes show up in different inflectional forms, which we shall refer to as **word forms**. For example, *sing, sings, sang, sung, singing* are the word forms of the lexeme SING$_V$ and *girl* and *girls* are the word forms of the lexeme GIRL$_N$. We may define **lexeme** and **word form** in the following way:

A **word form** is a word with meaning and form.
A **lexeme** is a family of word forms that are inflectionally related.

We shall come back to the meaning of *inflectionally related*.

For several reasons, it is important to distinguish between *lexemes* and *word forms*. It would be very confusing to count *word forms* when you discuss the size of the vocabulary of a language. In Chinese, there is a lexeme FÁNGZI$_N$ 'house', which has only one word form, *fángzi*, since Chinese nouns do not have any inflection. In Turkish, there is a lexeme EV$_N$ 'house', which has 84 word forms. EV$_N$ has singular forms and plural forms, and among both the singular forms and the plural forms there are six different case forms, as shown in TABLE 7.

Expression	Meaning	Expression	Meaning
ev	'house; singular, absolutive'	*evler*	'house; plural, absolutive,
evi	'house; singular, accusative'	*evleri*	'house; plural, accusative'
evin	'house; singular, genitive'	*evlerin*	'house; plural, genitive'
eve	'house; singular, dative'	*evlere*	'house; plural, dative'
evde	'house; singular, locative'	*evlerde*	'house; plural, locative'
evden	'house; singular, ablative'	*evlerden*	'house; plural, ablative'

TABLE 7. *Some word forms of Turkish EV$_N$*

For example, *evden* means 'from (the) house' and *evlere* 'to (the) houses'.

In addition, Turkish nouns may have six different *possessive suffixes*, telling who is the owner, and these suffixes can be combined with singular, plural, and all the case forms. A few illustrations are given in TABLE 8, where all the possessive suffixes are added to the absolutive forms (notice that possessive suffixes follows the plural suffix *–ler*), and in TABLE 9, where forms with the 1sg possessive suffix is combined with different case forms (notice that possessive suffixes precede the case suffixes).

Expression	Meaning	Expression	Meaning
evim	'house; 1sg, singular, absolutive'	*evlerim*	'house; 1sg, plural, absolutive,
evin	'house; 2sg, singular, absolutive'	*evlerin*	'house; 2sg, plural, absolutive'
evi	'house; 3sg, singular, absolutive'	*evleri*	'house; 3sg, plural, absolutive'
evimiz	'house; 1pl, singular, absolutive'	*evlerimiz*	'house; 1pl, plural, absolutive'
eviniz	'house; 2pl, singular, absolutive'	*evleriniz*	'house; 2pl, plural, absolutive'
evleri	'house; 3pl, singular, absolutive'	*evleri*	'house; 3pl, plural, absolutive'

TABLE 8. *More word forms of Turkish EV$_N$*

Expression	Meaning	Expression	Meaning
evim	'house; 1sg, singular, absolutive'	*evlerim*	'house; 1sg, plural, absolutive,
evimi	'house; 1sg, singular, accusative	*evlerimi*	'house; 1sg, plural, accusative
evimin	'house; 1sg, singular, genitive	*evlerimin*	'house; 1sg, plural, genitive
evime	'house; 1sg, singular, dative	*evlerime*	'house; 1sg, plural, dative
evimde	'house; 1sg, singular, locative	*evlerimde*	'house; 1sg, plural, locative
evimden	'house; 1sg, singular, ablative	*evlerimden*	'house; 1sg, plural, ablative

TABLE 9. *Even more word forms of Turkish EV$_N$ 'house'*

For example, *evim* means 'my house' and *evlerimiz* 'our houses', while *evlerimde* means 'in my houses'.

It would be meaningless to claim that for every Chinese noun there are 84 Turkish nouns, and that therefore Turkish has a much bigger vocabulary than Chinese. Vocabulary comparisons should be based upon *lexemes*, and the number of word forms in a lexeme does not influence the size of the vocabulary.

2.3.2 Morphemes

It is not difficult to discover that the Turkish word forms in TABLES 7–9 can be divided into smaller parts, each of which has its own meaning. These parts are called **morphemes**:

Morphemes are the *smallest meaningful parts that words can be divided into*.

Here are the morphemes of the Turkish word form *evlerimde* 'in my houses':

	morpheme 1	morpheme 2	morpheme 3	morpheme 4
Expression	*ev*	*ler*	*im*	*de*
Meaning	'house'	'plural'	'1sg possessive'	'locative'

TABLE 10. *The morphemes of the Turkish word form **evlerimde** 'in my houses'*

We find the morphemes of a word by comparing it to other words. We notice that *ev* occurs in all the words in the three tables, and the only constant meaning is 'house'. We may also compare *evlerimde* 'in my houses' with *ellerimde* 'in my hands', where the morpheme *el* 'hand' can be identified. The element *ler* occurs in all and only the words having the meaning 'plural', *im* occur in all and only those having the meaning 'my' or '1sg possessive', and *de* occurs in all and only the words with the meaning 'locative'.

Now, take a look at some word forms of KÖY$_N$ 'village' in TABLE 11. We shall refer to a list of the word forms of a lexeme as a **paradigm**. This is therefore a *partial paradigm* for KÖY$_N$.

Expression	Meaning	Expression	Meaning
köy	'village; singular, absolutive'	*köyler*	'village; plural, absolutive,
köyü	'village; singular, accusative'	*köyleri*	'village; plural, accusative'
köyün	'village; singular, genitive'	*köylerin*	'village; plural, genitive'
köye	'village; singular, dative'	*köylere*	'village; plural, dative'
köyde	'village; singular, locative'	*köylerde*	'village; plural, locative'
köyden	'village; singular, ablative'	*köylerden*	'village; plural, ablative'

TABLE 11. *Some word forms of Turkish KÖY$_N$ 'village'*

Evi 'house; singular, accusative' contains the element *i* 'accusative', while *köyü* 'village; singular, accusative' contains the element *ü* 'accusative'. In the same way, *in* 'genitive' is found in *evin* 'house; singular, genitive' and and *ün* 'genitive' is found in *köyün* 'village; singular, genitive'. There is a system to it: we find *i* and *in* when the preceding vowel is *i* or *e*, and we find *ü* and *ün* when the preceding vowel is *ü* or *ö*. The plural forms *köyleri* 'village; plural, accusative' and *köylerin* 'village; plural, genitive' confirm this generalization. Cf. also the word forms *gün* 'day; singular, absolutive', *günü* 'day; singular, accusative', *günün* 'day; singular, genitive', *diş* 'tooth; singular, absolutive', *dişi* 'tooth; singular, accusative' and *dişin* 'tooth; singular, genitive' (the letter *ş* is pronounced [ʃ], like in English *dish* [dɪʃ]).

Few linguists would say that *i* 'accusative' and *ü* 'accusative' are two different morphemes, but rather two **variants** of the same morpheme. It is also very common to use the term **allomorph** instead of **variant**.

The fact that morphemes have variants (or allomorphs) is rather unfortunate for our morpheme definition. One may claim that we find *allomorphs* when words are divided into meaningful parts, and that morphemes are not the smallest meaningful parts that words can be divided into, but rather *families of allomorphs with the same meaning*.

2.3.3 Base and operation

Let us continue our analysis of Turkish word forms, starting with *evlerimde* 'house; 1sg, plural, locative'. We often think of word forms like this as having been **formed** by adding one morpheme to another, as illustrated in TABLE 12.

	Base	*ev*
1st step:	Add ler to the right of the base (= *ev*)	*ev+ler* ⇒ *evler*
2nd step:	Add im to the right of the base (= *evler*)	*evler + im* ⇒ *evlerim*
3rd step:	Add de to the right of the base (= *evlerim*)	*evlerim + de* ⇒ *evlerimde*

TABLE 12. *Bases and suffixes*

Each step consists of adding a morpheme to the right of a **base**; we shall come back to a definition of **base** immediately.

Adding a morpheme to the base is an example of applying a **morphological operation** to the base. Morphological operations can change the base in almost any logically possible way, by **adding**, **modifying**, and **subtracting** elements. A fourth type may also be established, which implies **converting** the base without any formal changes. We shall give examples below.

From our discussion it follows that a **base** is simply an element that a morphological operation can apply to. A base consisting of one morpheme only is referred to as a **root**. Therefore, the base *ev* in TABLE 12 is a root, while the bases *evler* and *evlerim* are not.

ADDING ELEMENTS TO THE BASE
A morpheme that is added to the base is called an **affix**, and there are affixes of three kinds:

a **suffix** is added *to the right of* the base
a **prefix** is added *to the left of* the base
an **infix** is added *into* the base

Suffixes are the only affixes found in Turkish. A language with lots of prefixes is Swahili, as in the word forms *m-toto* 'child' (singular) and *wa-toto* 'children' (plural).

Infixes are a much rarer phenomenon, but it is common *inter alia* in the Malayo-Polynesian language Tagalog of the northern Philippines, where, as shown in TABLE 14, the past tense morpheme is an infix *in*, which is inserted after the leftmost consonant of the base.

present		past	
ibigay	'give; present'	*ibinigay*	'give; past'
ipaglaba	'wash (for); present'	*ipinaglaba*	'wash (for); past'
ipambili	'buy (with); present'	*ipinambili*	'buy (with); past'

TABLE 14. *Tagalog verbal word forms*

Infixes split morphemes into two parts, so that one part comes before and the other after the infix.

MODIFYING THE BASE

In cases of **modification**, no special element is added to the base; instead, the operation changes one or more of the sounds in the base, as in English plurals like *men* (from *man*) and *geese* (from *goose*) and English past forms like *spat* (from *spit*) and *wrote* (from *write*). While such forms are exceptional in English, they are regular in Arabic. Arabic nouns are divided into two types, those that form the plural by adding an affix and those that form the plural by modification, which in this case implies changing the vowels, as shown in TABLE 15.

SINGULAR		PLURAL	
qalb	'heart; singular'	*qulūb*	'heart; plural'
kalb	'dog; singular'	*kilāb*	'dog; plural'
kitāb	'book; singular'	*kutub*	'book; plural'

TABLE 15. *Some Arabic nouns*

SUBTRACTING ELEMENTS FROM THE BASE

In the Nilo-Saharan language Murle, which is spoken in southern Sudan, the plural of nouns is formed by **subtracting** the last consonant from the singular form; cf. TABLE 16.

SINGULAR		PLURAL	
nyoon	'lamb; singular'	*nyoo*	'lamb; plural'
wawoc	'white heron; singular'	*wawo*	'white heron; plural'
onyiit	'rib; singular'	*onyi*	'rib; plural'

TABLE 16. *Some Murle nouns*

CONVERTING THE BASE

Conversion means leaving the base formally unchanged (despite the meaning change), as in the English plural *sheep* (from *sheep*). Conversion creates **synchretism**, which is *homonymy among the word forms of a paradigm.*

REDUPLICATING THE BASE

Reduplication means repeating the base or part of it, as in Malay *anak-anak* 'children' (from the stem *anak* 'child').

MORE COMMENTS ON OPERATIONS

The plural forms of the Arabic noun *kutub* 'books' cannot at all be divided into a morpheme meaning 'book' and a morpheme meaning 'plural', the way we can divide English *books* into *book+s*. Word forms formed by non-affixing operations are not easy to analyze within the morpheme model.

2.4   Word structure: derivation and compounding

We will now turn our attention to the formation of new words. Consider the words *player* and *football*. Both consist of two morphemes, *play-er* and *foot-ball*. But there is a difference. In *play-er,* only the first morpheme may act as a word in its own right, whereas in *foot-ball*, both morphemes may act as words of their own. Both *player* and *football* are examples of word formation, but while *player* is an example of **derivation**, *football* is an example of **compounding**.

Both *player* and *football* constitute **inflectional bases**, that is, the bases that inflectional operations (such as the adding of the plural suffix *-s*) are applied to: *players, footballs*. Inflectional bases are also known as **stems**. In the study of **word formation** we analyze the morphological structure of stems.

If the stem cannot be divided into smaller meaningful parts (or if the stem cannot be analyzed as a base plus one or more operations), the stem is a **root**. Both *play, foot* and *ball* are roots, while *player* and *football* are the results of derivation and compounding, respectively.

2.4.1 Derivation

The stem *player* can be divided into two morphemes, *play* and *er*. While *play* is both a stem and a root, however, *er* is neither, because it does not occur as the stem (inflection base) of any English lexeme. Instead *er* must be identified as a suffix added to verb stems to form stems of nouns designating actors. From the stem *sing* we get *singer*, and from the stem *ride* we get *rider*.

The suffix *er* is a **derivational affix**, that is, an affix added to a base to form a new stem. The stem *player* is derived from the stem *play* by suffixing *er*. In general, we may talk about **derivational operations**, which include *affixing, modification, subtraction*, and *conversion*.

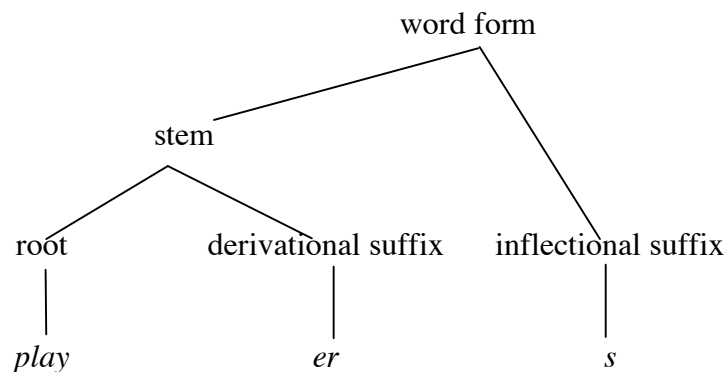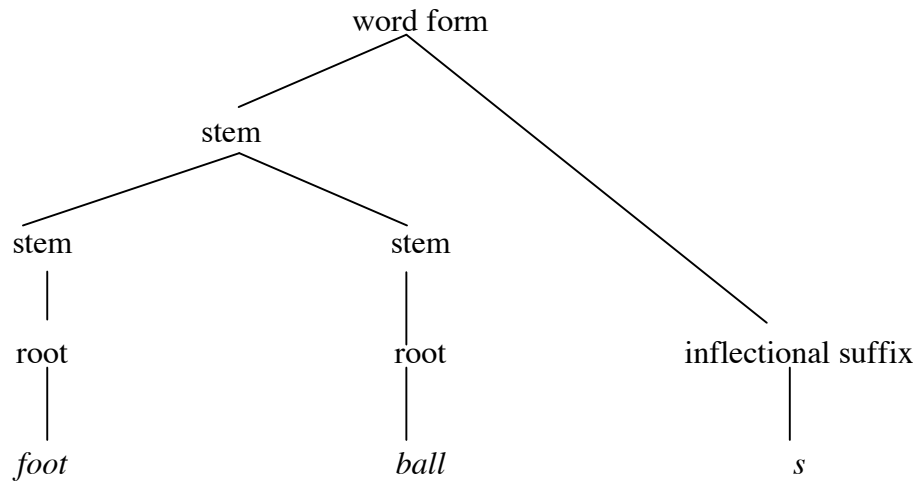The word form *players* (plural of *player*) has the morphological structure in FIGURE 14.

FIGURE 14. *The morphological structure of the word form **players***

2.4.2 Compounding

The stem *football* can be divided into two morphemes *foot* and *ball*. Both of them are both stems and roots. Consequently, the stem *football* is a **compound**, that is, a stem formed from two or more other stems.

The structure of the plural word form *footballs* is presented in FIGURE 15.

```
                          word form
                         /         \
                   stem             \
                  /    \             \
            stem        stem          \
             |           |             \
           root        root         inflectional suffix
             |           |                  |
           foot        ball                 s
```

FIGURE 15. *The morphological structure of Turkish **önsözde***

The two parts of the compound stem are also stems. In this case, they are also roots, though that does not have to be the case.

2.5   Language and thought

Most people see language as a vehicle for the expression of thoughts that are already there independently of the words and grammatical structures that express them, as seems to be implied in Saussure's model of the speech situation. But it has also been claimed that language contributes to the shaping of thought, and that different languages do so in different ways. This idea is often referred to as the **Sapir-Whorf hypothesis** (or just Whorfianism), after the linguists Edward Sapir and Benjamin Lee Whorf.

There are two versions of this hypothesis. The stronger version, called **linguistic determinism**, asserts that thought is completely determined by language, while the weaker version, called **linguistic relativity**, asserts instead that ways of thinking tend to be partly shaped by language. Linguistic determinism would leave little room for creative thinking that transcends the limitations of language, and this view has been largely discredited. Linguistic relativity is a much more plausible idea.

In a later chapter, we shall discuss such perspectives in relation to the typology of motion verbs (see 6.5). In the following, we shall look at a number of studies that seem to support the idea of linguistic relativity.

2.5.1 Memory and problem-solving

A number of psychological studies indicate that linguistic labels do make a difference for thinking, especially for memorizing and problem-solving.

In one experiment, two groups of persons were shown several meaningless shapes (see examples below). In one group, the shapes were given random names, in the other group, the shapes were not given names at all. It turned out that the group for whom the shapes had been given names were more prone to remember the names afterwards.

CAT   SCREEN   RABBIT   DOG
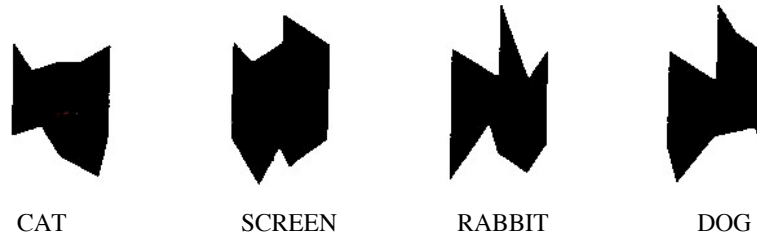
In another experiment, two groups of persons were shown the same figures, but with different names. For instance, the figure O—O was presented as *eyeglasses* in one group and *barbells* in another group. When asked to reproduce the figures afterwards, the eyeglass group tended to distort the figure towards O⌢O, while the barbell group tended to distort it towards O—O.

In yet another experiment, two groups of persons were placed in front of a table with a candle, a box of tacks, and matches. They were asked to fasten the candle to the wall and make it burn without dripping. In order to solve the problem, they needed to realize that they could use the box not just as a container for the tacks, but also as a container for the candle, to be fixed to the wall. Both groups were shown a figure that illustrated the problem, but in one case each object was given a name (*box, tacks, candle* and *matches*, as illustrated below), while in the other case no names were given. The name group solved the problem after an average of 0.61 minute, while the no-name group needed an average of 8.82 minutes.
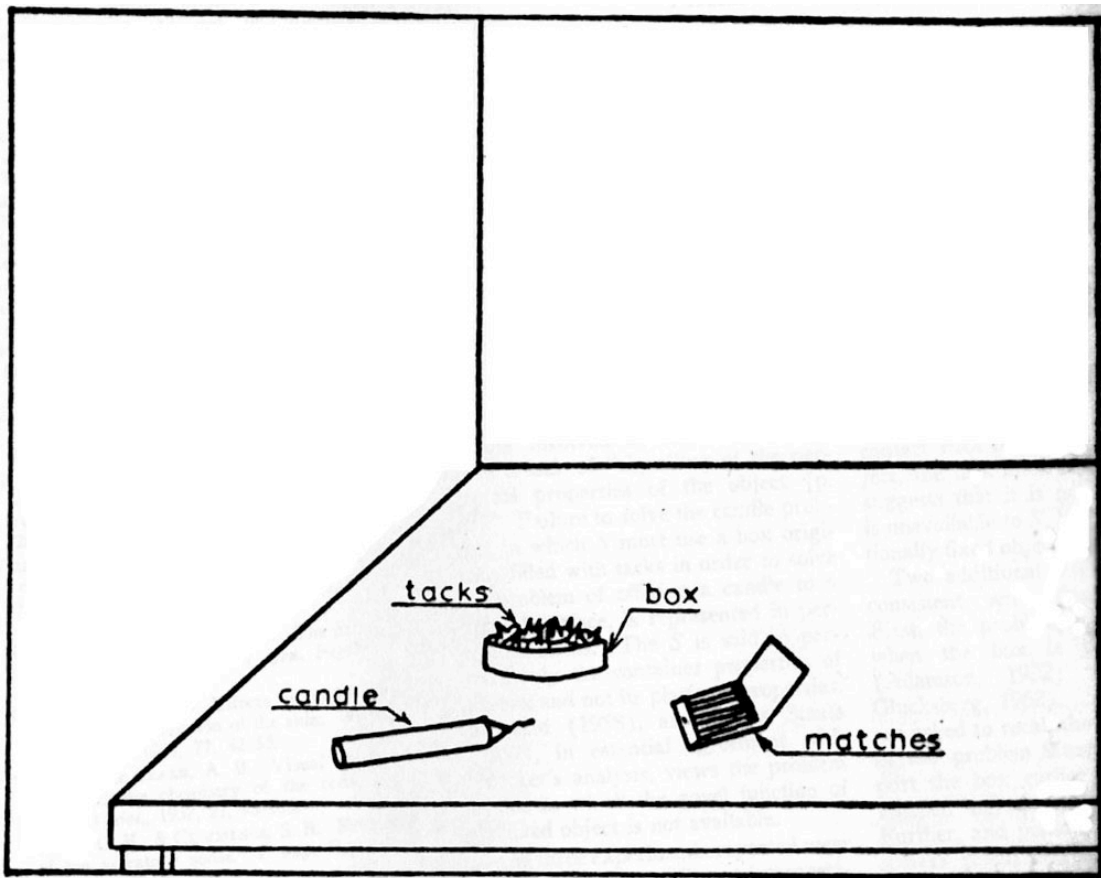

These and other experiments indicate that providing a name to an object does influence our ability to remember the object and to solve problems involving the object, as well as our tendency to distort our memory of it. The problem is to

determine the consequences of this for natural languages, which are infinitely more complex than the experimental situations just described.

2.5.2 Colour perception

In contrast to English, the Mexican Indian language Tarahumara does not have one term for 'green' and one for 'blue', but instead has a single term, *siyóname*, that covers both. In other words, where English has two concepts, 'green' and 'blue', Tarahumara has only one. How does this difference affect the perception of colours among speakers of the two languages?

In one experiment, speakers of both languages were presented with three colour chips at a time. Each time, all three chips, say, A, B and C, were of different shades of colour on the scale from green to blue, with chip B being somewhere in between chips A and C:
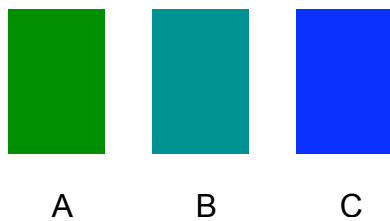
The subjects were asked to determine whether the distance in colour between chips A and B was greater than the distance between chips B and C or the other way around. It turned out that when the borderline between English *green* and *blue* went between chips B and C, the English speakers tended strongly to feel that the distance between chips B and C was greater than that between chips A and B, even when the actual distance between chips A and B, as measured independently, was greater. Speakers of Tarahumara did not make a similar systematic distortion.

The proposed explanation for this is that English speakers solve a difficult problem (that of determining distances between colours) by resorting to a "name strategy". If chip A and B are both called *green,* while chip C is called *blue,* the name strategy prompts the English speaker to decide that chip C is more different from chip B than chip A is, even when the opposite is in fact the case. This strategy is not available to Tarahumara speakers, since their vocabulary does not distinguish between 'green' and 'blue'.

In a second experiment, English speakers were presented with the same triads of colour chips, but in a way that only enabled them to see two chips at a time, either A and B or B and C. When they were shown chips A and B, the experimenter said: "You can see that this chip (points to A) is greener than this chip (points to B)." Everybody agreed. And when they were shown chips B and C, the experimenter said: "You can see that this chip (points to C) is bluer than this chip (points to B)." Again everybody agreed. Thus, all subjects were prompted to use both the terms *green* and *blue* to refer to chip B. When they were subsequently asked to judge the relative distance between A and B as opposed to B and C, the systematic distortion found in the first experiment had disappeared. The proposed explanation is that the name strategy was no longer available, since they had already referred to chip B by both terms. This suggests that the use of the name strategy was indeed the correct explanation for the systematic distortion in the first experiment.

To sum up, whether or not a language distinguishes between 'green' and 'blue' does seem to influence the perception of these colours. To some extent, language influences the way we perceive the world.

2.5.3 Why are East Asians so good at maths?

Children from East Asian countries have been generally shown to perform better at mathematics tests than children from Western countries. Many explanations have been proposed, including parental emphasis, pedagogical techniques and cultural differences. One suggested explanation is the different way numerals are constructed in the languages of the two areas.

In general, numerals in Chinese, Japanese and Korean are more systematically transparent than numerals in English and other European languages. This is especially true of numbers between 11 and 99, as shown in the following table, where East Asian numerals are directly translated into English for ease of comparison, and Japanese pronunciations are given as one East Asian example:[7]

	English	**East Asian**	**Japanese**
11	eleven	ten-one	jū-ichi
12	twelve	ten-two	jū-ni
13	thirteen	ten-three	jū-san
20	twenty	two-ten	ni-jū
21	twenty-one	two-ten-one	ni-jū-ichi
22	twenty-two	two-ten-two	ni-jū-ni
30	thirty	three-ten	san-jū
99	ninety-nine	nine-ten-nine	kyū-jū-kyū

For numbers between 1 and 10, there is no such difference, and the same is basically true of numbers from 100 upwards.

Is the systematic transparency of East Asian numerals one of the reasons why they perform so well in mathematics tests? Maybe. In one experiment, Japanese and American first-grade children were given a set of blocks to represent numbers. White blocks represented units, while purple blocks represented tens, so that one purple block was equivalent to ten white blocks. The children were asked to read a number on a card and then to use the blocks to show that number. On first trial, American children tended to use only the white blocks (representing units), while Japanese children tended to use a combination of white and purple blocks, indicating that they had a better grasp of the decimal number system. Only on second trial did the American children begin to use the purple blocks as well. In other experiments, Chinese and Korean first-graders have been shown to perform more or less like their Japanese peers.

---

[7] In some respects, the Japanese system is more complex than both the English and the Chinese ones, since virtually every numeral has at least two different pronunciations, one indigenous and one borrowed from Chinese. With the exception of 4 and 7, the borrowed pronunciation is the preferred one.

## Two ways of saying twelve

The American way

The Japanese way
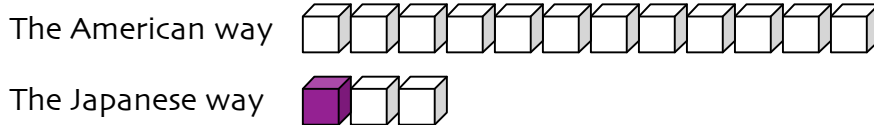
Significantly, the range of numbers between 11 and 99 showed the greatest differences between East Asian and American kids. In other tests, Chinese children have been shown to be better at counting between 11 and 99 than English-speaking children, but no better at counting between 1 and 10 or beyond 99. This suggests that the systematic transparency of East Asian numerals between 11 and 99 is indeed one of the factors influencing mathematics performance. East Asian numerals clearly stimulate children to think in tens, while the less transparent European system contains less of a stimulus to go beyond the amassment of single units.

2.5.4 Form perception

In the American Indian language Navaho, the form of a verb for handling an object varies with the form or shape of the object. For instance, if one asks somebody to hand over an object, the form of the verb will vary according to whether the object is long and flexible like a piece of string, long and rigid like a stick, or flat and flexible like paper or cloth. In general, Navaho-speaking children use these verb forms correctly as early as the age of three. Thus, the grammar of Navaho classifies objects according to form or shape in a way that English grammar does not.

In an experiment, Navaho children from the same reservation and living under similar circumstances were divided into two groups according to whether English or Navaho was their dominant language. Both groups were given pairs of objects where each object differed from the other in two respects, such as colour and size, colour and shape/form, or size and shape/form. For instance, one of the pairs consisted of a yellow stick and a piece of blue rope of comparable size, differing from each other in both colour and shape/form:
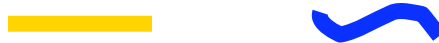
Then they were given a third object, which differed from each of the two others in only one respect, such as a piece of yellow rope, which differed from the yellow stick in shape/form and from the piece of blue rope in colour:

The children were asked which of the original objects went best with the new one. Children that grouped the piece of yellow rope with the yellow stick were noted as classifying objects on the basis of colour, while children that grouped the piece of yellow rope with the piece of blue rope were noted as classifying objects on the basis of shape/form.

As predicted, Navaho-speaking children tended more strongly to classify objects on the basis of shape/form than English-speaking children. In both groups, classification based on shape/form increased with age, but later and less marked in

children speaking English than in children speaking Navaho. These results seem to confirm the linguistic relativity hypothesis.

When, however, the same experiment was performed on English-speaking white middle-class children in Boston, their performance was more similar to the Navaho-speaking children than to the English-speaking children from the reservation. Their preference for shape/form-based classification was slightly weaker in the earliest age group, but later was actually stronger than that of the Navaho-speaking children. The proposed explanation is that other non-linguistic factors associated with social class overrule the effect of language. This explanation seems to be confirmed by a later experiment with English-speaking schoolchildren in lower-class Harlem, whose performance was very close to the English-speaking children in the Navaho reservation.

2.5.5 Objects vs. substances

In English, there is a clear distinction between **count nouns** and **mass nouns**. Count nouns, such as *horse,* share the following characteristics:

- they have a singular and a plural form (*horse* vs. *horses*)
- they may occur with words such as *many* or *few* (*many/few horses*)
- they may occur with numerals (*two horses*)
- they may occur with the indefinite article (*a horse*)

In contrast, mass nouns, such as *smoke*, share the following characteristics:

- they do not have a plural form (*smokes)
- they occur with *much* or *little* rather than *many* or *few* (*much/little smoke*)
- they do not occur with numbers (*two smokes)
- they do not occur with the indefinite article (*a smoke)

Count nouns refer to **objects**, while mass nouns refer to **substances**. An object is a sharply delineated entity, and if it is divided into smaller pieces, these pieces are no longer instances of the same object (a horse cut in two does not give two horses). A substance is not sharply delineated, and if it is divided into smaller pieces, these pieces are instances of the same substance (smoke remains smoke even if divided in two). This distinction plays a central role in the grammar of English, as it does in Swahili and many other languages of the world.[8]

In Japanese, the distinction between count nouns and mass nouns is grammatically unimportant, for the following reasons:

- there is no distinction between singular and plural forms of nouns
- there is no distinction between *many/few* and *much/little*
- no full nouns may occur with numerals alone
- there is no indefinite article

---

[8] The correspondence between the count/mass noun distinction and the distinction between objects and substances is only approximate. In English, for instance, *cloud* is a count noun, although dividing a cloud in two does produce two clouds, and *furniture* is a mass noun, although cutting a piece of furniture in two does not produce two pieces of furniture. Note that many words have both count noun and mass noun usages, such as the mass noun *oil* (denoting the substance) vs. the count noun *oil* (pl. *oils*, denoting a type of oil).

All Japanese nouns resemble English mass nouns in requiring a **classifier** (also known as a measure word) to intervene between a numeral and the noun itself. Just as English has *two clouds of smoke,* so Japanese has what may be literally rendered as *two pieces of horse*:

> *uma-ga[9]   ni-hiki*
> horse-SUBJ two-CLASS
> 'two horses'

Thus, the distinction between count nouns and mass nouns is not a central part of Japanese grammar. The same is true of the Mexican Indian language Yucatec and a number of other languages across the world.

How does this grammatical difference between English and Japanese affect the way speakers of these languages perceive the world around them? In one experiment, speakers of Japanese and English were given entities of three different kinds: complex objects, simple objects and substances. *Complex objects* had complex shapes and distinct functions, such as a lemon juicer:

Simple objects were made of a solid substance such as clay and were formed into very simple shapes such as the shape of a kidney:

Substances were non-solid stuff such as sand, but were arranged into characteristic shapes, such as an elongated S-shape:
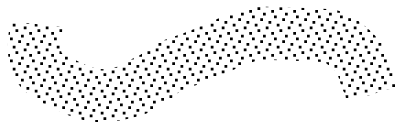
Complex objects were the most prototypical objects, since they clearly could not serve their function if broken up into pieces. Simple objects were an in-between case, with seemingly accidental shapes which did not *prima facie* indicate particular functions that might be lost if the objects were broken into pieces. In some respects,

---

[9] The particle *-ga* marks nominative case and must be replaced by *-wo* if the phrase occurs as object, *-wa* if the phrase occurs as topic etc.

the simple objects were quite close to substances, especially since the substances used were also formed into characteristic shapes.

The participants were presented with an entity (such as a lemon juicer, a kidney-shaped clay object or an S-shaped line of sand), which was given a novel name, such as *dax*. Then they were asked to compare this entity with two other entities, one that had the same shape, but different material, and one that had the same material, but different shape: "Look at this dax! Can you find the tray that also has the dax on it?" Thus, they were forced to classify the object on the basis of either shape (indicating that they saw the dax as an object) or material (indicating that they saw the dax as a substance).

As might be expected, both groups tended to classify a complex object on the basis of shape, while a substance was much more often classified on the basis of material (especially by the Japanese). The main difference lay in the way they treated the in-between group of simple objects. For English speakers, the tendency to classify simple objects on the basis of shape was almost (though not quite) as strong as for complex objects. For Japanese speakers, simple objects were just as often classified on the basis of material as shape, indicating that they were in fact perceived as an in-between case.

The sharp distinction between count nouns and mass nouns in English seems to have prompted English speakers to make a more unequivocal distinction between objects (including simple objects) and substances. The Japanese speakers, on the other hand, seemed to treat the distinction as a gradient one, where complex objects like a lemon juicer were the most typical objects and substances like sand the most typical substances, while simple objects like the kidney-shaped object made of clay were treated like in-between cases.

Similar results have been found when contrasting English with other languages without a clear-cut distinction between count nouns and mass nouns, such as the Mexican Indian language Yucatec.

2.5.6 How do words influence thought?

Experimental situations are far removed from the complexities of real life, and it is difficult to know how to interpret the results of the studies above. In some cases, it is uncertain whether it is language that influences thought, thought that influences language, or some other factor that influences both language and thought. On the whole, however, the studies do provide support for linguistic relativity. Language does seem to have a certain influence on thought.

How does this influence take place? In other words, which aspects of thought are influenced by language, and which aspects of language may influence thought?

The psychological studies referred to in 2.5.1 indicate that providing a linguistic label to an object reinforces (and sometimes distorts) our *memory* of that object. They also suggest that linguistic labels help us in *problem-solving* involving the labelled objects.

The linguistic studies referred to in 2.5.2 to 2.5.5 indicate that linguistic categories influence at least three different levels of thought:

1. *Perception,* as when the "name strategy" is used in forming judgements about the relative distance between colours.

2. *Classification,* as when the shape/form categories of the Navaho verb system prompt Navaho-speaking children to classify objects on the basis of shape/form, or when the distinction between count nouns and mass nouns in English prompt English speakers of all ages to distinguish sharply between objects and substances.

3. *Abstract thinking*, as in the case of mathematics and counterfactual reasoning.

In chapter 6 of this book, we will return to another type of linguistic relativism, where the properties of motion verbs in different languages influences the way speakers of these languages describe (and, presumably, perceive) the same sequence of events.

In many of the studies discussed, the crucial factor is how languages **conceptualize** the world differently, either through their grammar (as when the distinction between open and counterfactual conditionals is clearly expressed in the grammar of English, but not Chinese) or through their vocabulary (as when blue and green shades of colour are referred to by two words in English and one word in Tarahumara). Such differences are not insurmountable. If needed, speakers of Chinese may express the counterfactuality distinction (for instance, by adding a sentence like *The doctor hasn't arrived* in front of the conditional sentence), and speakers of Tarahumara may express more or less the same colour distinctions as English (for instance, by referring to the blue-green colour of the sky, as opposed to the blue-green colour of grass). But they can hardly do so without **foregrounding** distinctions that an English speaker may take for granted, as part of the **backgrounded** information. The way from backgrounded information to habitual thinking is not so far. Speakers of English are prompted by their language to think habitually in terms of the distinction between open and counterfactual conditionals, and in terms of the distinction between blue and green, in a way that speakers of Chinese and Tarahumara, respectively, are not.

Conceptualization is not always an advantage. English numerals like *twelve* and *twenty* more clearly represent individual concepts than Japanese numerals like *ten-two (jū-ni)* and *two-ten (ni-jū)*. But this is exactly the reason why the Japanese numerals are so transparent. The decimal number system is more strongly built into the form of Japanese numerals, and this may be one of the reasons why Japanese children perform so well in mathematics tests.

Language is, of course, not the only factor influencing habitual thinking. We saw that while Navaho-speaking children were prompted by their language to classify objects on the basis of shape/form, English-speaking middle-class children in Boston did the same, although they were not prompted by their language to do so. Language does seem to have a certain effect on the way we think, but the effect may often be overruled by other factors.