Introduction to English Linguistics

A Companion to the Seminar

(Revised and Abridged)

László Varga

Eötvös Loránd University
Budapest
2010

Preface

The seminar called *Introduction to English Linguistics* is offered in English to first year students in weekly sessions. Since for most students this seminar is the only place where the topics of the course are discussed in English, teachers of this seminar often have to explain the material to their students before (or instead of!) doing exercises or discussing problems, and so a considerable part of precious seminar time is wasted on lecturing.

The present book offers a solution. It contains ten units: each discusses an important topic in English and each is followed by exercises. Thus it provides ample seminar material for about 12 weeks. Some of the units may be curtailed or even omitted at the teacher's discretion.

The units should be assigned for home study *before* the actual seminar session at which they will be discussed. Thus the students will have an opportunity to get acquainted with the main ideas and the special English vocabulary of each topic before coming to the seminar, and this frees their tutors from having to lecture during the seminar, enabling them to concentrate on discussion and practical problem-solving activities.

I hope both students and teachers of the *Introduction to English Linguistics* seminar will find this book a useful companion to the course. If for any topic further reading is required, I recommend the relevant chapters of Fromkin, V. & R. Rodman (1998) *An Introduction to Language*. 6th ed. (Fort Worth, etc.: Harcourt Brace College Publishers), or Radford, A., M. Atkinson, D. Britain, H. Clahsen & A. Spencer (1999) *Linguistics, An Introduction*. (Cambridge: CUP).

August 2010

László Varga
Eötvös Loránd University
Budapest

Acknowledgements

The contents of this book have been derived from a number of sources. The sources include standard introductory textbooks, such as Aitchison, J. (1978) *Linguistics*. 3rd ed. (Teach Yourself Books. Hodder & Stoughton), Akmajian, A., R. A. Demers & R. M. Harnish (1979) *Linguistics: An Introduction to Language and Communication*. (Cambridge, Mass., London, England: The MIT Press), Fromkin, V. & R. Rodman (1998) *An Introduction to Language*. 6th ed. (Fort Worth, etc.: Harcourt Brace College Publishers), or Radford, A., M. Atkinson, D. Britain, H. Clahsen & A. Spencer (1999) *Linguistics, An Introduction*. (Cambridge: CUP), but also textbooks of more specific kinds, such as Lyons, J. (1977) *Semantics*. Cambridge: CUP, or Cook, V. J. & M. Newson (1996) *Chomsky's Universal Grammar*. 2nd ed. (Oxford: Blackwell), and many more. I would like to express my indebtedness to the authors of all of them.

My thanks are also due to my colleagues Ádám Nádasdy and Péter A. Lázár, both of Eötvös Loránd University, for their expert opinions on the manuscript of this book. Needless to say, I alone am responsible for any weaknesses that may have remained.

László Varga

Table of Contents

Unit 1
Language

1.1 Communication and signs

This book is an introduction to language and linguistics. Since language can be described as the most effective means of human communication, we will first briefly examine what we mean by communication.

Communication in general can be defined as the transmission of information (= transfer of a message) between a source and a receiver by means of signs. A **sign** is something physical, which represents something other than itself.[1] Signs have an exponent, a meaning and a set of referents. The **exponent** of a sign is its physical manifestation, something which can be perceived (heard, seen, touched, etc.) by the receiver of the message; e.g. a gesture, a facial expression, a picture, a roadsign, the sounding of a horn, a word, the smell of burning, etc. The individual things, qualities, actions, states in the world to which a sign refers are the referents (= denotata) of the sign, and these together constitute the **reference** (= extension) of the sign. In addition to having an exponent and reference, a sign is also associated with meaning. The **meaning** of a sign is the concept which it evokes in its users and which can be identified with a set of semantic features; this set can be called the sign's intension. For instance, the English word *girl* evokes a concept that includes the features 'young, female, human'. The signs used in a communication system constitute a code.

Signs can be divided into three basic kinds: they can be symbolic, iconic and symptomatic.

- When the exponent of a sign bears an arbitrary relationship to the sign's referents, the sign is a **symbol**. For instance, the colours used in traffic lights are symbolic: it is a matter of arbitrary convention that the red light means 'stop' and the green light means 'go', in principle it could be the other way around. The vast majority of the words of human languages are symbolic signs: their physical form (pronunciation) and their referents are arbitrarily associated, consider e.g. English *table*, Russian *stol*, German *Tisch*, Spanish *mesa*, etc., which all refer to the same kind of thing but all sound different.

[1] According to a nice old definition by St. Augustine (died in 430), a sign is something that gives itself to the senses but something beyond itself to the mind.

- However, when there is a natural resemblance between the exponent and the referents of the sign, the sign is an **icon**. For instance, the stylised silhouette of a man or a woman on a public lavatory door is an iconic sign. A small minority of the words in languages, viz. **onomatopoeic** words, e.g. English *buzz*, *dingdong*, *miaow*, *cuckoo*, etc., are signs that are partly iconic. (But even these are partly arbitrary, i.e. partly symbolic, because their counterparts in other languages are never quite the same.)
- And finally, when the exponent of a sign is mechanically linked to its source in such a way that it is a spontaneous reflection of the state of the source, the sign is a **symptom** (= indicium). Blushing is the symptom of embarrassment, a trembling voice is the symptom of excitement, smoke is the symptom of fire, etc. Symptoms do not need interpretation in a language, they are interpreted simply by there being a direct association between two states of affairs. They are not chosen and sent deliberately but follow automatically from certain states of affairs. (Human beings, however, are capable of producing some symptoms deliberately. This happens in playacting or deceving, e.g. when somebody deliberately assumes a tired or a drunken way of walking, even when they are not tired or drunk at all.)

1.2 Linguistic communication: the use of language

Human beings communicate in lots of ways, but the most effective way of human communication is **linguistic communication**, i.e. the use of language. The basic signs used in linguistic communication are words.[2] Words (and morphemes) are predominantly symbolic signs, though a small subset of them (onomatopoeia) are partly iconic. The linguistic signs and the rules for their combinations used by a community constitute a **linguistic code** (a language).

Linguistic communication takes place in the following way. Speaker A, in his/her mind, selects words from the language and combines them according to the rules of the language, i.e. encodes the message. Then his/her articulatory organs – or, in the case of writing, his/her hands – realise the signs: transmit the string of words into a physical signal, a stretch of sound or writing. This signal is perceived by the ears – or, in the case of writing, by the eyes – of Speaker B, who then decodes the message, i.e. reconstructs the message in his/her mind. These steps are shown in (1).

[2] In Unit 5 we shall see that many words can be decomposed into even smaller meaningful parts called morphemes, e.g. the English third person singular present tense verb form *enjoys* contains the morphemes *en-*, *joy* and *-s*, but since the meanings of such composite words often do not obviously follow from the meanings of the constituent morphemes, we claim that the basic signs are words rather than morphemes.

(1) Linguistic communication

nervous system	articulatory organs or hands	sound waves or writing	auditory organs or eyes	nervous sytem
Speaker A		→ signal →	Speaker B	

If Speaker A and Speaker B do not share the linguistic code (do not speak the same language), linguistic communication cannot take place between them.

While animal communication happens automatically and instinctively, conditioned by the situation, relying mostly on symptomatic (and iconic) signs, linguistic communication between human beings is predominantly intentional, relying mostly on symbolic signs, and can be totally independent of the situation.

A comparison of human languages and animal communication systems can be made in terms of the so called **design features of language** (a set of features characterising human languages).[3]

- We have already mentioned one of these, viz. **arbitrariness**, i.e. the absence of a natural bond between sign-exponents and their referents (cf. 1.1). Arbitrariness is present in the words (and morphemes) of human languages (not counting the onomatopoeic ones, which are partly natural). Interestingly, a certain degree of arbitrariness is present in bee-dancing, too, which is a term for the special movements that bees perform with their wings and bodies in order to communicate to their fellow-bees about the direction and distance of a source of nectar. This can be regarded as arbitrary because there is no obvious connection between the form of the dance and the distance from the hive, but this arbitrariness is of a very limited kind, manifesting itself only in connection with localising food.

- Probably the most important design feature of language is **duality** (= double articulation). By this we mean that every human language is organised into two layers. The first is a layer of basic sounds called phonemes, such as /æ/, /k/, /t/, which are meaningless in isolation and take on meaning only when combined in certain ways.[4] The second is a layer of meaningful units (morphemes, words, sentences), which result from combining the basic sounds,

[3] First set up by the American linguist Charles Hockett.

[4] *Sounds in general* can be transcribed (represented on paper) by symbols put between square brackets, as in [æ], [k], [t], but the *phonemes* of a particular language are represented by symbols put between slants, as in /æ/, /k/, /t/ (see Unit 4 below).

as in e.g. /kæt/ *cat*, /tæk/ *tack*, /ækt/ *act*, or from combining meaningful units, as in e.g. *un-friend-li-ness*, or in *The man saw the lion*. There is no evidence that duality is present in any animal communication system.

- Another important design feature is **patterning**. This means that every language has certain permitted ways of combination, i.e. ways in which phonemes can be combined into words and words into sentences. For instance, in English we can have /kæt/, /tæk/, /ækt/, but not */ktæ/ or */tkæ/ or */ætk/.[5] Similarly, we can have *He is happy*., *Is he happy?*, and even *Happy he is*., but not **Is happy he?* This feature, too, is missing from animal communication systems. Other design features are creativity and displacement.
- **Creativity** (= productivity) means that human beings can produce and understand an infinite number of new messages that they have never heard before. This feature is present in bee-dancing, because the exact location of nectar can always be different and so new messages are always possible, but this is a very limited kind of creativity, existing in connection with localising food again.
- Finally, **displacement** can be defined as the ability to use language in connection with things and events remote in space and time. For instance we can talk about past, future and distant events, hypothetical objects, and we can even lie. Displacement is certainly present in bee-dancing, but only in connection with food, and it is only spatial displacement, not temporal. (Bees can inform their fellow-bees about the whereabouts of a source of nectar at the time of the communication, but they cannot pass on any information about the availability of a source of nectar in the future or in the past.)

The following table sums up those design features that we have discussed, and the corresponding values in bee-dancing.

(2) Design features of human language and an animal communication system

	human language	bee-dancing
duality	+	–
patterning	+	–
creativity	+	limited
arbitrariness	+	limited
displacement	+	limited

[5] The star (asterisk) before an item means that the item is ill-formed (ungrammatical).

1.3 The discrete nature of language

Linguistic communication, i.e. the use of language, is characteristically vocal and verbal behaviour, involving the use of discrete language elements.

- It is **vocal** because it is crucially associated with the articulatory (vocal) organs.
- It is **verbal** because words play a central part in it.
- Thirdly, it involves the use of **discrete** language elements, which differ from one another discretely (on an either-or basis) rather than gradually (on a more-or-less basis).

Words as lexical items are discrete because they differ from one another on an either-or basis. This means that two word-realisations either represent the same word or two different words. For instance, whenever English speakers utter the word *pen*, there will always be some physical difference between the realisations, but these variations are not only gradual and hardly noticeable but also insignificant, and so each rendering will be taken as realising the same word *pen*. However, when the English words *pen*, *pan*, *car*, *picture* and *camera* are uttered, the words are discretely different, regardless of the fact that in physical terms *pen* and *pan*, both starting with /p/ and ending with /n/ are more similar to each other than to the others. In the relevant sense, they are all discretely different because if you change one of them for another in a particular sentence you may get a completely different sentence, with a completely different meaning, as in (3).

(3) This { pen / pan / car / picture / camera } is very expensive.

Words are composed of basic sounds called **phonemes**. The latter are discrete, too, because two phoneme-realisations either represent the same phoneme or two different phonemes. The phonemes of a language are those sounds that are capable of distinguishing otherwise identical words. If you replace one phoneme with another in a particular word, you may get a different word which no longer means the same, cf. (4).

(4)

/ p	e	n/	*pen*
	æ		*pan*
	ɔː		*pawn*
	ɪ		*pin*
	ʌ		*pun*

When the only difference between two words is that one has one phoneme where the other has another phoneme, the two words constitute a **minimal pair**. Thus, e.g. /pen/ and /pæn/, or /pen/ and /pɔːn/, or /pæn/ and /pɔːn/, etc., are minimal pairs in English.

In English the phonemes /e/, /æ/ and /ɔː/ are discretely different, even though from a strictly physical point of view /e/ and /æ/, both being "front" vowels, are more similar to each other than to /ɔː/, which is a "back" vowel. But this similarity is irrelevant, because they are capable of distinguishing words that are otherwise identical.

Phonemes can be looked upon as **segmental element**s, because they are in a sense the smallest building blocks (= segments) of words and sentences. But words and their sequences in sentences also contain **suprasegmental elements**, which are called so because they are "superimposed" upon units that are or can be larger than segments, such as e.g. syllables.[6]

The most significant suprasegmental elements are stress patterns and pitch patterns, and we shall now briefly look at these to prove that they are discrete, too. **Stress** is a degree of the prominence of a syllable**. Stress patterns** are patterns of syllabic prominence. Syllabic prominence can be achieved by various means. The innumerable degrees of syllabic prominence that are physically possible and may actually occur in real speech can be grouped into a few, discrete degrees in English: non-stress, tertiary stress, secondary stress, and primary stress (in order of increasing strength). In a somewhat simplified account we can define these in the following way:

- A **tertiary-stressed** syllable is louder than the **unstressed** ones, i.e. it is extra-loud.
- A **secondary-stressed** syllable is extra-loud and pitch prominent, i.e. it is associated with some pitch change, but this pitch change is not the initiation of a nuclear pitch pattern, only a step-up or step-down in pitch.
- By contrast, a **primary-stressed** syllable is extra-loud and pitch prominent in the sense that it initiates a nuclear pitch pattern. A nuclear pitch pattern is the characteristic final melody in an intonational phrase, e.g. a falling

[6] Suprasegmental elements are also known as **prosodic elements**.

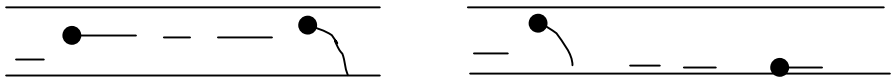
contour, a rising contour, a falling-rising contour, etc.[7] The stress degrees so defined are discrete: a syllable is either extra-loud or not, and when it is extra-loud, it is either pitch prominent or not, and when it is pitch prominent, it either has a nuclear contour or not.

When we want to show which syllables are stressed in words, we may put the symbol ˈ before a primary-stressed syllable, the symbol ˌ before a secondary-stressed syllable, and the symbol ₒ before a tertiary-stressed one, as in the examples in (5): *Monongahela* (name of a river in the USA) and *individualistic*. The stress symbols used in the lines of text of (5) indicate stress degrees only, they do not show intonation. Intonation here is indicated by schematic drawings
representing the pitch heights of the syllables of the examples.

(5)a. /məˌnɒŋgəˈhiːlə/ b. /ˌɪndɪₒvɪdjʊəˈlɪstɪk/

In English, stress patterns are able to distinguish words that are otherwise identical. For example, the noun *insult* has the primary stress on its first syllable, whereas the verb *insult* has it on the second syllable, cf. /ˈɪnsʌlt/ vs. /ɪnˈsʌlt/. But they can also distinguish utterances, i.e. spoken sentences. For instance, in (6) the interpretation of the utterances depends on where the primary stress falls in them. In (6a) the primary stress falls on *come* and the utterance implies 'but he didn't come', whereas in (6b) it falls on *thought* and the utterance implies 'and he did come'.

(6)a. We ˌthought he would ˈcome. b. We ˈthought he would ₒcome.



Pitch patterns (= tones) are permanent pitch configurations that are carried by syllables or syllable sequences. The commonest pitch patterns are the falling (i.e. high-low), rising (i.e. low-high), falling-rising (i.e. high-low-high), rising-falling (i.e. low-high-low), high level (i.e. high), and low level (i.e. low) tones.

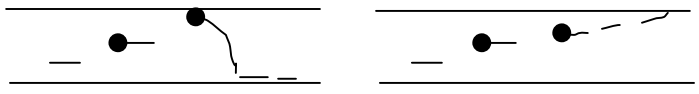
[7] In English, in addition to the loudness and pitch features that characterise them, stressed syllables also have a full (unreduced) vowel quality.

• In some languages, such as e.g. Chinese or Thai, pitch patterns are used as **lexical tones** or "word melodies", because they can distinguish the meanings of words that are segmentally identical. Such languages are called **tone-languages**. For example, in Thai there are several different words which are all composed of the sound string *kha*, but have different pitch patterns and so different meanings. With a falling pitch pattern *kha* means 'kill' (7a), with a rising one it means 'leg' (7b), with a low level one it means 'spice' (7c), with a high level one it means 'trade' (7d). (The special symbols used in the examples indicate the various pitch patterns.)

(7)a. /᷄kha/ 'kill' b. /᷅kha/ 'leg' c. /_kha/ 'spice' d. /¯kha/ 'trade'

• In other languages, such as English or Hungarian, there are no word-melodies, but pitch patterns are used in **intonation**, i.e. as parts of "utterance melodies", because they distinguish the meanings of utterances that are in other respects identical. Such languages are called **intonational languages**. Compare for example the English utterances (8a) and (8b), differing only in the final parts of their intonations, i.e. in the nuclear contours they have. Intonation can be transcribed (i.e. represented on paper) by means of **tonetic stress marks**, i.e. graphic symbols which simultaneously indicate stress and intonation. Thus in both (8a) and (8b), the first syllable (*they*) is unstressed (no symbol), the second syllable (*came*) is secondary-stressed and has a high level tone (symbol: ˈ). The third syllable (*yes*-) in (8a) is primary-stressed and initiates a *falling* nuclear contour (symbol: ˧˩), while in (8b) it is primary-stressed and initiates a *rising* nuclear contour (symbol: ˧˥). This difference in intonation shows that (8a) is a statement, and (8b) is a yes-or-no question. The vertical lines (|) indicate the boundaries of the intonational phrases.

(8)a. | They ˈcame ˧˩yesterday. | b. | They ˈcame ˧˥yesterday? |



By means of the tonetic stress marks we can transcribe the stressing and the intonation of utterances within the line of text and no separate drawings are necessary. (The drawings in (8a) and (8b) are only there to help you visualise what the tonetic stress marks stand for.)

In (6a) and (6b) above, the suprasegmental transcription symbols in the lines of text above the intonational drawings showed the stressing of the utterances but not their intonation. The intonation was shown in separate drawings below the lines of text. If we apply tonetic stress marks to them, as in (9a) and (9b), we do not need intionational drawings:

(9)a. (=6a) |We ˈthought he would ˧˥come. |
b. (=6b) |We ˧˥thought he would ₒcome. |

Of course there is room for variation within the pitch patterns. For example, a fall can start at any pitch from very high to mid-low pitch. Nevertheless, a fall is a discrete pitch pattern, because it is not a rise, not a fall-rise, not a rise-fall, and not a level tone, either. Two tone-realisations realise either the same tone, or two different ones.

1.4 Paralanguage

In the preceding section we have seen that the use of language is characteristically vocal and verbal behaviour, involving the use of discrete elements. However, *accompanying* and occasionally even *replacing* language, we also find behaviour which is not vocal or, if vocal, not verbal and not discrete. Variations in this kind of behaviour, used during and instead of linguistic communication, are called **paralinguistic features** or **paralanguage**.[8]

Among the non-vocal features of paralanguage we have to mention gestures, bodily movements, facial expressions which we make while we are speaking or instead of speaking. These include e.g. bowing, waving, winking, raising our eyebrows, putting our finger across our lips, shaking our head, nodding, etc. The vocal paralinguistic features include various meaningful noises, such as *hm*, *pff*, the wolf whistle (a not very polite way of young men's whistling to a pretty girl), throat clearing, etc. Some of these are intentional, some are unintentional. Throat clearing, for example, can be either. When it is unintentional, it is merely symptomatic, and shows that you have a cold. But it can be used deliberately (symbolically), too, e.g. as a warning.

The suprasegmental part of human communication, which we sketched in 1.3, is vocal, non-verbal behaviour, and it is partly linguistic, partly paralinguistic. The stress degrees are linguistic because they are discrete, but

[8] The prefix *para-* comes from Greek and means 'beside, accompanying'.

the general loudness-level of a certain part of an utterance is gradable, and so paralinguistic. The pitch patterns used in word-melodies and in intonation are linguistic because they are discrete, but their vertical extent (range) and the general pitch height of certain parts of utterances (key) are gradable, i.e. paralinguistic variations. Similarly, tempo, pause-length, voice-quality variations are gradable, and however important and informative they may be, they belong to paralinguistic rather than linguistic communication.

Exercises, problems, and other tasks

1. What is communication?
2. How do we define a sign and what are the three parts of a sign?
3. What is the exponent, the meaning and the reference of the English word *tree*?
4. What are symbols, icons and symptoms?
5. What are onomatopoeic words?
6. What are the English counterparts of the following Hungarian words: *kukurikú*, *bimbam*, *tiktak*. What is your conclusion?
7. When a dog opens its mouth and shows its teeth in a threat to bite, the sign it uses is partly iconic, partly symptomatic. Explain.
8. How does linguistic communication take place?
9. What do we mean by arbitrariness as a design feature of human language? Illustrate it.
10. What is duality?
11. What is patterning?
12. Find out how English speakers pronounce *Tbilisi*, *Gdansk*, *Xerxes* (by looking them up in e.g. J. C. Wells' *Longman Pronunciation Dictionary*.) What happens to the initial consonant cluster? Why?
13. Combine the foll. phonemes in as many ways as you can to form existing English words: /s/, /p/, /t/, /ɒ/. Point out some impossible combinations, too. What are your conclusions?
14. How do the English pronounce the letter combinations *kn*, *ps*, *mb* in *knee*, *knowledge*, *psychology*, *psalm*, *comb*, *dumb*, *acknowledge*, *rhapsody*, *cucumber*? Try to find an explanation.
15. What do we mean by creativity and by displacement as design features of human language?
16. Linguistic communication is characteristically vocal and verbal behaviour, involving the use of discrete language elements. Explain the key words *vocal*, *verbal*, *discrete*.

17. What are types and tokens?
18. How do we define phonemes? Why are they segmental elements?
19. What is a minimal pair? Do *collar* and *colour* constitute a minimal pair? And *monkey* and *donkey*?
20. What are suprasegmental elements, and why are they called so?
21. What is stress, and how many degrees have we distinguished in English? In what sense are these degrees discrete?
22. Find three English noun–verb pairs like ˈ*insult* vs. *in*ˈ*sult*. What do they differ in? Why can we not find similar pairs in Hungarian?
23. What are pitch patterns and what are the two basic ways in which they are used in human languages? In what sense are pitch patterns discrete?
24. Chinese is said to be a tone-language. Why?
25. What is intonation?
26. What is paralanguage? Mention non-vocal and vocal paralinguistic features.
27. In certain cultures shaking one's head means 'yes' and nodding means 'no' What does this show?
28. Do you know what these paralinguistic vocal noises mean: [tsk], [mmm], [pɸ]? Can you do the wolf whistle?
29. Classify the following signs:
    a. involuntary cough.
    b. cough for getting attention
    c. nod of head ('Yes')
    d. *Uh-huh* ('Yes')
    e. *Yes.*
    f. *Bzzz* (sound of a bee)
    g. hand indicating height from ground ('So high.')
    h. waving the hand ('Good-bye.')
    i. blushing
    j. sweating.

Unit 2
The Study of Language (i)

2.1 Language: Externalised and internalised

A language is a linguistic code, which its speakers know and use, and which manifests itself in its speakers' linguistic knowledge and in the actual utterances that its speakers make in linguistic communication. Consequently, language can be regarded as existing in essentially two modes. On the one hand it can be looked upon as a body of objective facts (strings of sounds or letters) produced and perceived by its users in linguistic communication. On the other hand it can be regarded as the language users' knowledge which makes linguistic communication possible, an internal property of the human mind. One of the greatest figures in modern linguistics, Noam Chomsky, has called these two modes of language **Externalised Language (E-language)** and **Internalised Language (I-language)**, respectively.

The dominant kind of language study in the first half of the 20th century, viz. Structuralist Linguistics (see Unit 3), concentrated on E-language. It aimed at collecting samples of E-language, i.e. samples of the actual products of linguistic communication, as objects independent of the mind, and then describing the regularities (patterns, structures) found in those samples. Since then, however, the interest and emphasis of language study has shifted to I-language, i.e. to the knowledge that native speakers of a language possess and use when they communicate linguistically. Generative Linguistics (see Unit 3) aims at modelling the I-language of the native speaker, i.e. his/her **linguistic knowledge** or **internal grammar**.

2.2 Components of language

A natural language (whether we look upon it as E-language or I-language) has several components. The central ones are phonology, morphology, syntax, and semantics.

- **Phonology** includes the phonemes (basic sounds) and the discrete suprasegmental elements (stress patterns, tones, intonation) in the language. The phonological component also contains rules that regulate how phonemes can be combined in morphemes and words. For example, the sequences /kæt/

and /tæk/ are phonologically well-formed in English, but */ktæ/ or */tkæ/ are phonologically ill-formed.

- Another component is **morphology**. This includes the morphemes and the rules for combining them to derive and inflect words in a particular language. (For the time being we define morphemes as the smallest meaningful units of a language. We will make this definition more precise in Unit 5.) In English, for instance, the morpheme *-ion* can be added to the verb *elect* (which is a vocabulary item) and the result is the noun *election* (which is a new vocabulary item **derived** from the former one). In a similar way, the plural morpheme *-s* can be added to the noun *election* to obtain the plural form of the same noun: *elections* (which is not a new vocabulary item but the **inflected** variant of an already existing one). The morphological rules of English tell us that the sequence *un-friend-li-ness* is a morphologically well-formed word, while **friend-li-un-ness* is not.
- **Syntax** is the component of language that contains the rules for putting together words in phrases and phrases in sentences. For example, the English sentence *He went to London*. is syntactically well-formed, whereas **To he London went*. is syntactically ill-formed.
- Finally, languages also contain a system of meanings: this component is known as **semantics**. The semantic rules specify which sentences are semantically normal and which are semantically anomalous. For instance, *This woman is the mother of three girls*. is semantically normal but [!]*This woman is the father of three oil-wells*. is anomalous.[1]

In addition, we can also separate a special component in which all the central components may play a role, viz. a **lexicon**. This is a list of the vocabulary items of a language and it contains all idiosyncratic information about those vocabulary items (such as the unpredictable aspects of their phonology, morphology, syntactic behaviour, and meaning). Words, once formed and established as vocabulary items, are stored in the lexicon, from where they can be retrieved as wholes and do not have to be put together again from their constituent morphemes every time they are used by a speaker.

Native speakers of a language have linguistic knowledge: they *know* their language. They possess I-language, they have an **internal grammar**.[2] They know the elements and the rules in the various components of their language, after all they use those elements and obey those rules all the time and, on the basis of this knowledge, they can tell whether a string of words in their

[1] In this book, the raised exclamation mark [!] before a sentence indicates that the sentence is semantically anomalous.

[2] The word *grammar* is used here in a broad sense to include phonology, morphology, syntax, and semantics. It can also be used in a narrow sense to include only morphology and syntax.

language is grammatical or not. But most speakers are unable to explain to their children or to their foreign friends *why* one string of words is grammatical in their language and another is not. This is because their linguistic knowledge (internal grammar) is **intuitive** (subconscious), and they cannot express it explicitly (i.e. clearly and definitely).

2.3 Linguistics and its branches

If we want to obtain explicit knowledge about language, we must study language systematically and objectively, i.e. we must deal with **linguistics**. Linguistics seeks explicit knowledge *about* language, by submitting it to systematic and objective study. A study that is systematic, objective, and seeks explicit knowledge is scientific. Linguistics is the scientific study of language (i.e. E-language and/or I-language).

The product of linguistics is an objective, systematic, and explicit account of (some aspect of) language, i.e. an **explicit grammar**.

A **linguist** is a person who is professionally engaged in the scientific study of some aspect of language (i.e. of one particular language or of several languages or of human language in general). From this definition it follows that someone who knows a number of languages (i.e. a polyglot) is not necessarily a linguist, and a linguist is not necessarily someone who knows a number of languages.

Linguistics, or its product, a grammar, has branches corresponding to the central components of language. **Phonology** is the study of the phonemes and their combinations in words and morphemes, and also of the discrete suprasegmental elements in words and sentences. **Morphology** is the study of word derivation and word inflection in terms of constituent morphemes. **Syntax** is the study of sentence formation. **Semantics** is the study of the meaning of words and sentences. **Lexicology** is the study of the lexicon, i.e. the phonological, morphological, syntactic, and semantic properties of vocabulary items. All these are summed up in (1).

(1) The central branches of linguistics/grammar

phonology	morphology	syntax	semantics
lexicology			

Moreover, all these can be studied from a **synchronic** point of view (how they constitute a particular state of language at a particular point of time),

or from a **diachronic** (historical) point of view (how they change through time).[3]

In a somewhat broader concept of linguistics there are phonetic and pragmatic components, too. **Phonetics** is closely related to phonology, it is the study of the production, physical properties and perception of the actual sounds realising the phonemes and of the suprasegmental elements of speech. **Pragmatics** is close to semantics and the difference is not always quite clear. We can say that while semantics examines what sentences and words mean in themselves, pragmatics studies the ways in which they obtain different interpretations when uttered in different situations. For instance, if I put the question *Can you play the piano?* to a person I am interviewing in a room where there is no piano, my utterance will count as a real yes-or-no question. But if I say the same utterance to a person who is known to be a good pianist, and I point towards a piano at the same time, my utterance will count as a request to play.

The scope of linguistics can be extended further. It can include **sociolinguistics**. This is an interdisciplinary branch of study (relevant to both linguistics and sociology), studying the different varieties of a language used by different geographical and socio-cultural subsections of a community, or varieties used by the same group of speakers in different social situations. **Psycholinguistics**, another interdisciplinary subject, deals with areas such as the mental processes that take place when we produce and receive linguistic messages, or the processes of native language acquisition. And finally, linguistics can be put in the service of a large number of other fields, some more practical, some more theoretical, such as e.g. foreign-language teaching, speech therapy, successful advertising, literary criticism, stylistics, etc. These involve various kinds of **applied linguistics**. For example, when a doctor wants to cure a patient who suffers from aphasia (i.e. who has lost – partly or completely – the ability to use language), the doctor will have to know about the language system. In such cases linguistics helps the doctor in his/her work.

The present book offers an introduction to the basic notions in the main branches of linguistics. Unit 4 introduces Phonetics and Phonology, Unit 5 Morphology, Unit 6 Syntax, Unit 7 Semantics, Unit 8 Pragmatics, Unit 9 Sociolinguistics, and Unit 10 historical linguistics.

In the rest of this unit we take a brief look at the history of language study before and at the time of the appearance of modern linguistics in the 20th century.

---

[3] These terms come from Greek, where *khronos* means 'time', *syn-* means 'together' and 'alike', and *dia-* means 'through'. They were introduced by Ferdinand de Saussure (see below).

2.4 Traditional Grammar

Languages began to be studied a very long time ago: in the 5th century BC or earlier, but it is only since the 19th century that we can speak about linguistics. It was in the 19th century that historical language study began to meet the criteria of scientificness and only in the 20th century that the study of contemporary languages became scientific in today's sense of the word.

Earlier language study can be called **Traditional Grammar**. In principle, this kind of language study dealt with the contemporary state of languages but it often mixed its synchronic statements with diachronic ones.

Traditional Grammar was not sufficiently scientific. (a) It was not explicit enough: it was often too vague in its statements and its definitions were often too loose. For example, the noun was defined as "the name of a person, place or thing", although there are lots of words that we intuitively feel to be nouns even though they are not the names of persons, places or things, e.g. *reflection*. (b) It was not systematic enough: it ignored spoken language and was preoccupied with written language, especially with the written language of older literarary works. (c) It was not objective enough: it was often **prescriptive** and puristic rather than descriptive, i.e. instead of recording what the language examined was like, traditional grammarians often tried to prescribe what it should be like. In these attempts they relied on their subjective wishes and speculations and on historical, logical and aesthetic arguments, and on analogy with Latin. For example, they argued that the split infinitive, which is quite common in English, was incorrect: "You shouldn't say *to humbly apologize*, you should say: *to apologize humbly*". The idea that the split infinitive was wrong was based on Latin. It was believed that, since a Latin infinitive was only one word, its English equivalent should also be as near to one word as possible. Traditional grammarians thought that language change was harmful and they fought against it.

With all its weaknesses, however, Traditional Grammar accumulated a great number of facts about individual languages and elaborated linguistic terminology. Modern linguistics would not have been born if there had been no Traditional Grammar to prepare the way for it.

2.5 Comparative Philology

Comparative Philology was the dominant kind of language study in the 19th century. It was scientific in several respects. However, it narrowed down the concept of language study to a study of the history and genetical relationships of languages and of the written records that were available.

This kind of linguistics emerged after the discovery that Sanskrit was related to Latin and Greek. The discovery was made in 1786, by a British government official working in India, Sir William Jones. Throughout the 19th century, language scholars tried to establish genetical relationships between languages. That was the time when the various language families and branches were discovered, for example the Germanic branch (of which English is a member) and a Proto-Indo-European parent language was reconstructed. In Comparative Philology the study of language was beginning to develop towards an autonomous, independent branch of study. Language began to be studied for its own sake. Besides, this kind of language study had an objective method: it was based on textual evidence, i.e. E-language facts, found in earlier written records of language, and it also tried to show language change in a systematic way, as a process determined by rules. (In the last quarter of the 19th century, a group of scholars in and around Leipzig, nicknamed the **Neogrammarians**, claimed that language changes were not just accidental events or optional tendencies, but "laws".)

Meanwhile, the study of the contemporary state of languages went on in the non-scientific (or not sufficiently scientific) framework of Traditional Grammar.

2.6 The Beginnings of Modern Linguistics in Europe, Saussure

Modern linguistics emerged almost simultaneously in Europe and the USA in the early decades of the 20th century.

In Europe the study of language at the beginning of the 20th century was characterised by two features: the inheritance of a long period of Traditional Grammar, and the predominantly historical interest of 19th century Comparative Philology. Modern linguistics appeared as a kind of revolt against this background. The first great figure of modern linguistics in Europe, **Ferdinand de Saussure**, a Swiss scholar, was a comparative philologist himself (a professor of Sanskrit at the University of Geneva), but his ideas about language and language study went far beyond the limitations of Comparative Philology.[4]

- He was the first to emphasise the difference between (a) language as an abstract system, residing in the collective consciousness of the community (which he called **la langue**) and (b) language as the realisation of that system (which he called **la parole**).

[4] We know his revolutionary ideas from a posthumous book, *Cours de Linguitique Générale,* which was published by his students in 1916.

- He separated the **synchronic** and **diachronic** aspects of lsnguage study, and argued for the primacy of the former by saying that the synchronic aspect deals with language as a collection of simultaneous facts, existing as a state at a particular point of time, whereas the diachronic regards language as a succession of states, so it is the states that have to described first.
- According to Saussure, linguistic signs enter into two kinds of relationship: syntagmatic and paradigmatic. The **syntagmatic relationship** is a linear (horizontal, chain) relationship, which exists between the signs that follow one another in a complex unit. For example, the four words in *This coffee is strong*. are in a syntagmatic relationship: they are placed one after the other along the syntagmatic axis, and each of the words has a particular environment or CONTEXT which consists of the other words on its left and right.

The **paradigmatic** relationship is a vertical (choice) relationship, which exists between a sign present in a particular environment and all the other signs that could replace it while still yielding a well-formed complex unit. For instance, *coffee* in the above sentence is in a paradigmatic relationship with *tea*, *student*, *girl*, *wall*, *light*, *whisky*, *cigar*, etc., see (2).

(2)

This { tea / coffee / student } is strong

Since Saussure's time the notion of these two relationships has been extended to phonemes as well, see (3).

(3)

/b { e / ʌ / iː / ɪ / ɔː } t/

bet
but
beat
bit
bought

2.7 The Beginnings of Modern Linguistics in America, the Sapir—Whorf Hypothesis

Linguistic research in the USA also began in the early decades of the 20th century, but with a different motivation. Here it was found that the languages of the American Indian population (the Amerindian languages) were threatened with extinction and so the main aim was to describe these languages as quickly and

accurately as possible. Modern American linguistics in the first half of the 20th century was usually called **structural(ist) or descriptive linguistics**.

The Amerindian languages did not make a traditional approach possible. They existed only in a spoken form, they had no earlier written records, they were very different from most of the languages studied until then, and the linguists who wanted to describe them did not speak them, so no prescriptive and puristic statements could be made about them. Briefly: these languages forced language-scholars to adopt a non-traditional approach to language, based on objectivity, systematicness and explicitness.

American descriptivists tried to describe each language in its own terms and they emphasised (even exaggerated) the differences between languages. One of them, Martin Joos, said: "Languages differ from one another without limit and in unpredictable ways." This is the essence of **linguistic relativism**. Linguistic relativism is the assumption that any natural language can be totally different from other natural languages.

But some linguists went even further. Sapir and especially Whorf thought that languages not only differed from one another without limit but also that the language of a community determined the way in which that community saw the world. This latter view is called **linguistic determinism**.[5] The combination of linguistic relativism and linguistic determinism became known as **the Sapir—Whorf hypothesis**. According to the strong version of the hypothesis the individual is not free in his experience of the world, because the vocabulary and grammatical categories of his native language determine the ways in which he can interpret his experience. For instance, the American linguist Boas discovered that in Eskimo there are several different words for different kinds of snow, whereas in English there is only one generic term: *snow*. Other linguists collected similar facts from other languages. (For instance, the Navajo language has no separate words for blue and green but has two separate words for different shades of black; the Hopi language does not distinguish present, past and future tenses; in Kwakiutl the distinction between singular and plural number is not obligatory, etc.) On the basis of such examples the conclusion was drawn that people belonging to different cultural-linguistic groups not only *spoke* differently but also *thought* differently: i.e. each cultural-linguistic community lived in the "prison" of its language. This conclusion, however, cannot be accepted. It is true that different languages cut up reality in different ways, but this is because different communities find different things important in their life. The fact that the English have no separate words for different kinds of snow does not mean that they cannot see these differences, only that they are not significant to them. When these

[5] This idea was not quite new: Wilhelm von Humboldt made a similar claim at the beginning of the 19th century.

differences do become important, the English can paraphrase and say "falling snow", "hard packed snow", "powdery snow", etc. The main counter-argument against the strong form of linguistic determinism is the possibility of translation. Translation is possible for most of the time and although we cannot always translate everything with the same ease, we are nevertheless usually able at least to paraphrase or explain what we mean in any language.

However, the weak form of the Sapir—Whorf hypothesis, according to which language influences thought, seems to be correct. Certain things are less **codable** (i.e. less expressible) in some languages than in others. The **codability** of an aspect of reality in a particular language means having a word for it, or at least the possibility of a simple paraphrase. People tend to notice and remember the things that are codable in their language better than things that are not codable. But differences in codability between languages are of secondary importance: it is only the less basic, culture-specific concepts that may present codability problems. The essential things are equally codable because they are equally relevant to all cultures.

Exercises, problems, and other tasks

1. What do we mean by E-language and I-language?
2. Define morphology, syntax, phonology, and semantics.
3. What do we mean by the lexicon?
4. Why do we say that most native speakers' knowledge of their language is intuitive?
5. What makes a study scientific?
6. How do we define linguistics?
7. What is a linguist?
8. Discuss phonetics.
9. Discuss pragmatics.
10. What is the difference between synchronic linguistics and diachronic linguistics? Who distinguished them first?
11. What are interdisciplinary subjects? Explain sociolinguistics. Explain psycholinguistics. Mention fields of applied linguistics.
12. What are phonemes?
13. What are morphemes? Can you identify the morphemes in *unexplainability?*
14. Discuss the difference between a. ¡*Colourless green ideas sleep furiously*. and b. **Late got he morning up this.*
15. Characterise Traditional Grammar.

16. What is the difference betwen the descriptive and prescriptive approaches to the investigation of language?
17. Comment on the following quotations:
    a. "Double negation is illogical."
    b. "'It is me' should be replaced by 'It is I' because Latin required the nominative case after the verb corresponding to 'to be'."
    c. "The regular plural of English nouns is formed by adding the letter -*s*."
    d. "The noun is the name of a person, place or thing."
    e. "A preposition is a word that is put before another word. Therefore it should not be used to end a sentence."
18. Characterise Comparative Philology. How did it start? In what sense was it scientific?
19. Study the divisions of the Indo-European parent language, provided in Appendix B.
20. Does synchronic linguistics necessarily mean the study of the present day state of language?
21. What did Saussure mean by *langue* and *parole*? Match the following adjectives, nouns and sentences with *langue* and *parole*, respectively:
    a. *potential*
    b. *He speaks English.*
    c. *individual*
    d. *actual*
    e. *social*
    f. *language system*
    g. *He is speaking English.*
    h. *behaviour*
22. Discuss the syntagmatic and paradigmatic aspects of the following:
    Your friend may come.
    1 2 3 4
    Compare the number of words that can replace 1, 2, 3 and 4. Which sets of words are open? Which ones are closed?
23. What do we call the kind of linguistics prevalent in the USA in the first half of the 20th century and why?
24. Discuss linguistic relativism, linguistic determinism, and the Sapir—Whorf hypothesis (its strong and weak forms).
25. What is meant by codability and by lexical gaps?
26. Here are a few examples of lexical gaps. Can you translate *busójárás*, *kopjafa*, *másfélszobás* into English? Or *privacy*, *brackish*, *brass-rubbing* into Hungarian?

27. Make sure you know the following terms: phonology, phonetics, morphology, syntax, semantics, pragmatics, lexicology, sociolinguistics, psycholinguistics, applied linguistics, Traditional Grammar, prescriptive attitude, linguistics, linguist, explicit grammar, synchronic point of view, diachronic point of view, langue, parole, syntagmatic relationship, paradigmatic relationship, linguistic relativism, linguistic determinism, codability.

Unit 3
The Study of Language (ii)

3.1 The Great Synthesis of American Structuralist Linguistics

Leonard Bloomfield, and his followers, the Bloomfieldians, thought that a linguist should collect observable data, i.e. real utterances, and analyse these data, i.e. segment and classify the physical features of the utterances collected. A body of such data (a set of observed and collected utterances) is a **corpus**. Using a corpus for linguistic investigation is called the "corpus-based" or inductive procedure. In Chomsky's terminology this means that American structuralism was preoccupied with discovering and describing the E-language aspect of natural languages.

The Bloomfieldians dealt with phonetics, phonology, morphology, and syntax, but rejected semantics, thinking that the study of meaning would only be possible when human knowledge had become far more advanced. The only aspect of meaning that they paid attention to was whether two forms (signs or sign-combinations) had the same meaning or different meanings. They used a strictly **formal analysis**. This was an analysis without reference to meaning, and it was based on an examination of distribution and constituency.

- The **distribution** of a language element (i.e. of a phoneme or morpheme or word) is the sum of all the environments in which it occurs. If two language elements always occur in different environments, i.e. they occur in mutually exclusive environments, then there is not even one environment in which one could replace the other. To put it differently: they never enter into a paradigmatic relationship with each other. In this case we say that the two language elements have totally different distributions: they are in **complementary distribution**. This means that where one of them can occur, the other cannot occur, and vice versa. For instance, the English phoneme /l/ has two variants, and they are in complementary distribution.[1] The "clear" variant [l] (which is like Hungarian /l/) occurs before vowels, e.g. [ˈhelɪn] *Helen*, and the "dark" variant [ɬ] (which is pronounced with a cupped tongue, i.e. with a raising of the back part of the tongue) occurs elsewhere, i.e. before consonants, e.g. [heɬp] *help* and in word-

[1] We shall call such variants of a phoneme *allophones* (see Unit 3 below).

final position, e.g. [heɬ] *hell*. By contrast, if the distributions of two language elements are not entirely different, i.e. there is at least one common environment in which one could replace the other, the two elements are **not in complementary distribution**. In this case they are either in contrast or in free variation.

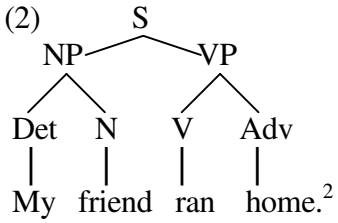
Two language-elements are in **contrast** in a particular unit if replacing one by the other changes the meaning of the unit. For example, English /e/ and /i:/ in the environment /m—t/ are in contrast because /met/ does not mean the same as /mi:t/.

If however replacing one language element by another in a particular unit does not change the meaning of the unit, they are in **free variation** in that unit. For instance, the same two phonemes /e/ and /i:/ are in free variation in the environment /—kəˈnɒmɪks/, because /i:kəˈnɒmɪks/ means the same as /ekəˈnɒmɪks/.

• The other important method of formal analysis which the Bloomfieldians introduced was **constituent analysis**. (The Bloomfieldians themselves called it "immediate constituent analysis" or "IC analysis".) This means cutting syntactic units (or words) into their constituents, then the constituents into their constituents, and so on until we reach the individual words (or morphemes). Cutting a unit into its constituents is based on the test of **substitution** (replacement). For instance, the sentence *My friend ran home.* can be divided into two: [*My friend*] and [*ran home*] because *My friend* can be replaced by a simpler constituent, e.g. *Peter*, as in *Peter ran home*; and because *ran home* can also be replaced by a simpler constituent, e.g. *slept*, as in *My friend slept*. So we divide the sentence into [*My friend*] and [*ran home*], and then, through further applications of the substitution test, these parts can be divided into even smaller constituents. Constituent analysis can be visualised in essentially two ways, viz. by **bracketings**, as in (1), or by **tree diagrams**, as in (2).

(1) $[_S[_{NP}[_{Det}$ My$][_N$ friend$]]$ $[_{VP}[_V$ ran$]$ $[_{Adv}$ home$]]]$.

(2)



[2] The syntactic representations in (1) and (2) and in other sections of this book are strictly preliminary and will be substantially modified in your later syntax studies.

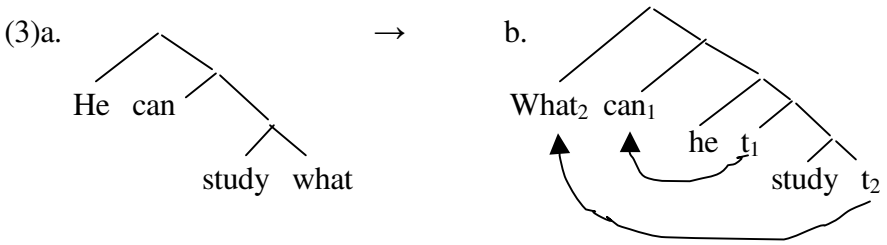
The constituents in the representations in (1) and (2) are labelled, S stands for Sentence, NP for Noun Phrase, VP for Verb Phrase, Det for Determiner, N for Noun, V for Verb, and Adv for Adverb. Trees and bracketings do not have to be labelled but the labelled ones are more informative than the unlabelled ones.
For further information on syntactic representations, see Unit 6.

Constituent analysis was suitable for resolving certain ambiguities, by showing different constituent structures, e.g.: (*old* (*men and women*)) vs. ((*old men*) *and women*).

3.2 Generative Linguistics

However, there were lots of ambiguities which constituent analysis could not resolve. For instance, *The lamb is ready to eat*. has two distinct meanings (is ambiguous), but the American structuralists could give it only one analysis: ((*The lamb*)(*is* (*ready* (*to eat*)))). Their analysis remained on the surface and could not disambiguate structures which were different in the deep.

The growing dissatisfaction with the limitations of structuralist linguistics led to the emergence of a radically new type of linguistic analysis towards the end of the 1950s. This has become known as transformational-generative linguistics, or just **generative linguistics** (= generative grammar), for short. This kind of analysis distinguishes two levels of syntactic analysis: a **surface structure** or **S-structure** (which was recognised by the structuralists, too) and an underlying abstract **deep structure** or **D-structure** (which was not recognised by the structuralists). Transformational-Generative grammar is **transformational** because it explains surface structure as being derived from deep structure by a series of changes: **transformations**. For instance, the S-structure in (3b) is derived from the D-structure in (3a).



The S-structure in (3b) is an **ordinary wh-question**, whose more abstract, underlying representation is the D-structure in (3a). But the latter can come to the surface unchanged, too, as *He can study what?*, and then it is an **echo wh-**

question, which can be used e.g. as a surprised response to *He can study chiromancy*.

In this framework, ambiguous sentences have identical surface structures but different deep structures, according to the different meanings. For instance, the ambiguous sentence *The lamb is ready to eat* (whose two meanings can be paraphrased as 'The lamb can eat' and 'Somebody can eat the lamb') is derived from two different deep structures. Synonymous sentences like *It rained yesterday*. and *Yesterday it rained*., however, derive from one common deep structure and differ only on the surface.

Transformational-generative grammar is **generative,** because it can generate (i.e. produce, define and explicitly characterise) all and only the grammatical sentences of a language. This means that (a) by applying the rules of the grammar, we always get a syntactically well-formed sentence, (b) this kind of grammar generates all the well-formed sentences of a language, i.e. not only those that have been uttered but also those that have not been uttered but could be uttered, and are, thus, potential sentences of the language. The number of possible grammatical sentences in any language at any one time is infinite, but the rules which make this infinite variation possible are finite (otherwise the native speaker would not be able to learn them).

The founder and most influential representative to this day of generative linguistics has been the American linguist **Noam Chomsky**, whose works have found a great many followers all over the world. Since its appearance the theory has been modified and remodified several times and several new proposals have been made and are still being made by Chomsky himself and by others.[3]

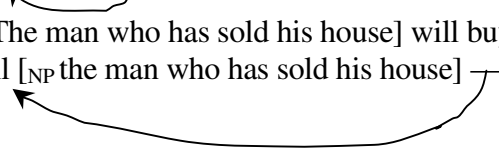
As we saw above, the Bloomfieldians were uninterested in general theoretical questions, emphasised the differences between individual languages, and thought that the main purpose of linguistics was to describe individual languages. In contrast, Chomsky holds that linguistics should be primarily concerned with **Universal Grammar**, i.e. with the principles that are the properties of all human languages. One of these **principles** is **structure-dependence**, which means that operations in a sentence apply to phrases and not just words, i.e. these operations require a knowledge of the structural relationships of words rather than just their linear sequence. For instance, when English speakers transform a declarative sentence into a yes-or-no interrogative, the auxiliary they move is not simply "the second word" of the declarative sentence,

[3] Chomsky's generative works include *Syntactic Structures* (1957*), Aspects of the Theory of Syntax* (1965), *The Sound Pattern of English* (with M. Halle, 1968), *Language and Mind* (1972), *Lectures on Government and Binding* (1981), *Barriers* (1986), *Principles and Parameters Theory* (with L. Lasnik 1993), *The Minimalist Program* (1995).

as a superficial observer might think on the basis of (4a), but rather the word after the entire Noun Phrase that occupies the subject-slot of the declarative sentence, as is shown in (4b).

(4)a. [$_{NP}$ John] will buy a car. →
Will [$_{NP}$ John] ___ buy a car?

b. [$_{NP}$ The man who has sold his house] will buy a car. →
Will [$_{NP}$ the man who has sold his house] ___ buy a car?



In Hungarian sentences, too, the syntactic constituents are not individual words but structural units composed of words (occasionally consisting of single words), i.e. phrases, such as e.g. *meghívtam* 'I invited', *a szomszéd gyerekeit* 'my neighbour's children', *egy belvárosi fagylaltozóba* 'to an inner-city ice-cream bar', cf. (5a), (5b), (5c).

(5)a. [Meghívtam] [a szomszéd gyerekeit] [egy belvárosi fagylaltozóba].
b. [Meghívtam] [egy belvárosi fagylaltozóba] [a szomszéd gyerekeit].
c. [A szomszéd gyerekeit] [meghívtam] [egy belvárosi fagylaltozóba].
d. [Egy belvárosi fagylaltozóba] [meghívtam] [a szomszéd gyerekeit].

According to Chomsky, a generative grammar is a model for the native speaker's intuitive knowledge of the language (i.e. his internal grammar), a decisive part of which is Universal Grammar and is genetically inherited. Chomsky calls the native speaker's language-knowledge **competence** (or – to use his more recent term – I-language). But the knowledge of language, competence, has to be distinguished from the actual use of that knowledge in real-life situations, i.e. from **performance**. Performance is the actual use of competence and it involves individual and situational features, imperfections, errors, memory limitations, time limitations on the length of sentences, life-span limitations on the number of sentences actually produced by the individual, etc. Chomsky's distinction between competence and performance reminds us of Saussure's distinction between langue and parole. But while Chomsky uses the term *performance* in very much the same sense as Saussure used the term *parole*, there is considerable difference between *competence* and *langue*. Saussure's *langue* was static: it was the system of linguistic signs. Chomsky's competence is dynamic: it puts the generation of sentences in the centre of attention. Another difference is that Saussure thought of *langue* as being in the collective consciousness of a community. Chomsky thinks of competence as knowledge whose basis is given to every normal human being

by birth, in the sense that its structure is related to the structure of the human mind and so the basis of competence is a universal characteristic of the human species.

On the basis of their competence, native speakers can do several things:

- They can produce and understand an infinite number of new grammatical sentences in their language.
- They can distinguish between grammatical and ungrammatical formations (*He went to London* vs. **Went London he to*).
- They can interpret elliptical sentences, i.e. sentences with missing elements (*Peter is happy but John isn't*).
- They can perceive ambiguity (*The lamb is ready to eat*).
- They can perceive synonymy (*The duck crossed the road* vs. *The road was crossed by the duck*).
- They can idealise utterances, i.e. they can disregard the imperfections and idiosyncratic features of performance and reconstruct the grammatical sentences which the utterances realise (*? *The thought of those poor children were really … WAS really ... bothering me*.).[4]

The last point has a very important consequence: generative linguistics has an **I-language approach** to the study of language. Earlier, both Saussure and the American structuralists in the first half of the 20th century were convinced that the way to *la langue* led through the observation of *la parole*. In other words, linguistic analysis had to be based on a corpus of data taken from the linguistic behaviour (actual language-use) of people, i.e. from *parole* or performance. This can be called the **E-language approach**. By contrast, generative linguists think that linguistics is concerned with far more than what can be found in a corpus. Thus, even if we do use a corpus for linguistic work, we shall have to "idealise" the data, i.e. free them from the imperfections and idiosyncrasies of performance. This is what native speakers automatically do when they understand other native speakers' utterances. They do so intuitively, on the basis of their competence (or I-language). But then the real task of linguistics should be the study of the native speakers' competence (and especially the part of it which can be regarded as Universal Grammar). This is more important than the actual utterances found in a corpus. Competence can be examined by asking questions about intuitions. Consequently, the linguist has the right to use his own and other people's intuitions in linguistic analysis. And if the linguist is a native speaker of the language he examines, he can ask and answer questions about his own intuitions. Examining one's own intuitions concerning language is a kind of introspection. In other words, generative linguists can base their theories not (only) on empirical

[4] In this book, the sign combination *? indicates performance mistakes.

facts but on **introspection** and on native speakers' **intuitions**. However, this does not mean that they give up objectivity because their theories can be submitted to subsequent empirical verification. (Only their method is different from the **inductive** method of the preceding decades: their method is **deductive**, proceeding from theories to empirical facts.) But the focus of their attention is undoubtedly on I-language: they are interested not so much in the empirical facts themselves as rather in the knowledge that enables speakers to produce those empirical facts.

Since competence resides in the individual language-user's mind and is a device of the reasoning activity of human beings, it is a mental, psychological phenomenon. Consequently, by studying what linguistic competence is and how it works, we are actually studying what the mind is and how the mind works. If language-competence is part of the human mind, then the study of this competence, i.e. linguistics, is part of the study of the mind, i.e. psychology. In other words: Chomsky's conclusion is that **linguistics is a branch of cognitive psychology**. Generative linguistics, then, has extended the status of psycho-linguistics from being a mere branch of linguistics, to being the dominant branch of modern linguistics.

Exercises, problems, and other tasks

1. What do we mean by the formal analysis of linguistic elements and units?
2. Here is a nonsensical sentence invented by Charles Fries (American descriptivist) to show that it is possible to analyse sentence structure without reference to meaning: *Woggles ugged diggles*. Analyse it, and transform it into a yes-no-question, negative, passive, singular.
3. When are two language elements in complementary distribution?
4. What is contrast and what is free variation? Comment on the Hungarian vowels between /f/ and /l/ in the pairs *fel – föl* and *felém – fölém*.
5. If two language elements never occur in the same environment, are they in a. contrast, b. free variation, or c. complementary distribution?
6. Can clear [l] and dark [ɬ] distinguish minimal pairs in English? Explain.
7. Reveal the constituent structure of the following sentences by drawing unlabelled tree diagrams for them:
   a. *The child found a puppy*.
   b. *The mouse ran up the clock*.
   c. *The hungry mouse ate up the old cheese*.
   d. *I met a foreign language teacher*.
   Comment on the last example.

8. How are the D-structure and S-structure connected in Transformational-Generative grammar?
9. Which of the following sentences are synonymous?
    a. *Pigs prefer large turnips*.
    b. *Pigs don't prefer large turnips*.
    c. *Large turnips are preferred by pigs*.
    d. *What pigs prefer is large turnips.*
10. When is a grammar generative?
11. What is Universal Grammar?
12. What do we mean by structure dependency?
13. How does Chomsky define competence and performance?
14. What can native speakers do on the basis of their competence?
15. Comment on these utterances in relation to native speakers' competence:
    a. *He went to London*. vs. **Went London he to*.
    b. *Peter is happy but John isn't*.
    c. *The lamb is ready to eat*.
    d. *The duck crossed the road*. vs. *The road was crossed by the duck*.
16. If you know Hungarian, comment on the following Hungarian utterances, taken from native speakers' performance:
    a. *? *Azokba a ... hm ... ódon falak között azér mosis … most is hűvös lehet*. = *Azok között az ódon falak között azért most is hűvös lehet*.
    b. *? *Nem tudod, mi volt te abban…*
    c. *? *Akkor mint betűszedőként működött*.
    d. *? *Csak hát hogy itt van ez a …izé, örrsov… orrsövényferdülés … probléma*.
17. In what sense can we say that Generative Linguistics has an I-language approach to the study of language?
18. Make sure you know the following terms: corpus, formal analysis, distribution, complementary distribution, contrast, free variation, constituent analysis, S-structure, D-structure, transformation, generative, Universal Grammar, competence, performance

Unit 4
Phonetics and Phonology, the Study of Sounds and Phonemes

4.1 Phonetics

Phonetics is the science of human speech sounds. It has three subfields or branches.

- The oldest branch, and also the one which is the most relevant in foreign language teaching, is **articulatory** phonetics. This examines the articulatory (vocal) organs and their role in the production of speech sounds.
- The second branch is **acoustic** phonetics. This deals with the physical properties of speech sounds as they travel through the air in the form of sound waves.
- The third branch is called **auditory** phonetics, which examines the way in which human beings perceive speech sounds through the medium of the ear.

When people speak, they produce physically continuous stretches of sound, which those who know the language in which the utterance was made can analyse into strings of individual speech sounds. For instance, the English word *fish*, when pronounced, is a continuous stretch of sound and not [f]+[ɪ]+[ʃ]. Still, speakers of English know that there are three distinct sounds in that word: [f] and [ɪ] and [ʃ]. The minimal distinct sounds that we distinguish one after the other in the physical continuum of speech are called **speech sounds** (= phones).

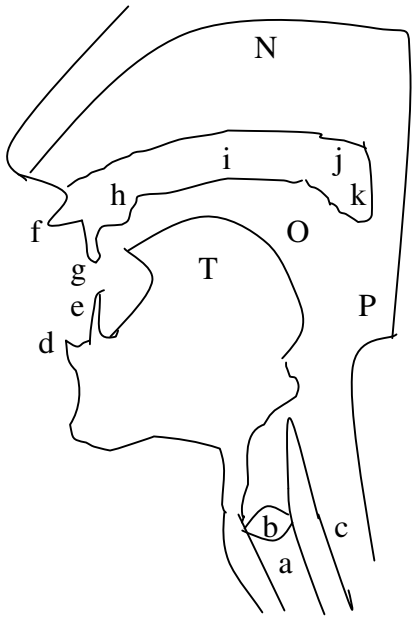
Phonetics attempts to examine all and only the speech sounds used in human languages. For instance, the sound [y] (the initial sound of the Hungarian word *üveg*) is a human speech sound because it occurs in several languages, including Hungarian, French, German, even though it does not occur in many other languages, e.g. English, Spanish, Italian. So [y] has to be dealt with in phonetic terms. By contrast, the sounds we produce when we sneeze or belch are not speech sounds in any language, so they have no place in phonetics. The number of speech sounds that phoneticians distinguish in the world's languages is around one hundred.

4.2 The tasks of phonetics

The main task of phonetic science is twofold, it is to provide a notation and description for each speech sound. By notation we mean a system of transcription symbols whereby we can make an accurate and unambiguous record of what goes on in speech. This is necessary because conventional letters cannot do this job properly. The correspondence between letters and sounds is indirect. In English, for example, there are 26 basic letters but considerably more speech sounds that we can distinguish. The transcription system which contains symbols for the hundred or so speech sounds that can be distinguished in human language is a special kind of alphabet, known as the **International Phonetic Alphabet** (= IPA).[1] In this system each phonetic symbol stands for one and only one speech sound. Sometimes supplementary marks (diacritics) are added to the symbols, e.g. the raised letter *h* indicates aspiration of the initial sound [t] in the word [t^{h}u:ɬ] *tool*. Phonetic transcriptions are enclosed in square brackets: [ ]. They are detailed, and called **narrow transcriptions**. The degree of detail (narrowness) depends on the analyst's purposes.

The other main aim of phonetics is the description (characterisation) of speech sounds. This is done in terms of phonetic features. In order to understand these features, we have to get acquainted with the **articulatory organs**. These are schematically shown in (1) below. The diagram represents the side-view of the front part of a human head and neck.

(1)



a. wind-pipe = trachea
b. larynx, with the glottis
c. food-pipe = oesophagus
d. lower lip
e. lower teeth
f. upper lip
g. upper teeth
h. alveolar ridge
i. palate = hard palate
j. velum = soft palate
k. uvula
P. pharynx
O. oral cavity
N. nasal cavity
T. tongue

[1] This was devised in 1888 by the International Phonetic Association and has been widely used ever since.

One characteristic feature of speech sounds is, for instance, the presence or absence of vocal cord vibration during the production of the sound. The air coming from the lungs by way of the wind-pipe (a) arrives at the larynx (b). This is where the vocal cords are situated, forming an opening between them called the glottis. When the vocal cords are together and the air stream passing through between them makes them vibrate, the sound produced will be **voiced** (or [+voice]), e.g. [b, d, g, v, ð, z, ʒ]. When the vocal cords are apart and so the air stream passes through freely, without causing vibration of the vocal cords, the sound produced will be **voiceless** (or [–voice]), e.g. [p, t, k, f, θ, s, ʃ].

Another feature of speech sounds is, for example, the presence or absence of nasality. The air, leaving the glottis, arrives at a cavity called the pharynx (P), from which it can go on to two further cavities: the nose and the mouth, i.e. the nasal cavity (N) and the oral cavity (O), respectively. These two are separated from each other by the roof of the mouth. The roof has several parts. Just behind the upper front teeth (g) is the alveolar ridge (h), then comes the hard palate or palate (i), followed by the soft palate or velum (j). When the back of the velum, i.e. the uvula (k) is raised, the passage through the nose is cut off and the air can only escape through the mouth. Sounds produced in this way are **oral** [–nasal], e.g. [b, d, g]. If, however, the back of the velum is lowered, the air can escape through the nose and the mouth. Sounds produced this way are **nasal** [+nasal], e.g. [m, n, ŋ].

The **consonants** which occur in the world's languages can be described in terms of place and manner of articulation. Here we shall concentrate on the most important English consonants only. We distinguish eight classes of these consonants according to **place of articulation**. *Bilabials* ([p, b, m, w]) are produced between the two lips, *labiodentals* ([f, v]) between the upper front teeth and the lower lip, *dentals* ([θ, ð]) between the upper front teeth and the tip of the tongue, *alveolars* ([t, d, s, z, n, l, r]) between the alveolar ridge and the front of the tongue, *palatoalveolars* ([ʃ, ʒ, tʃ, dʒ]) in the postalveolar region, *palatals* in the area of the hard palate ([j]), *velars* ([k, g, ŋ]) in the area of the soft palate or velum, and finally *glottals* ([ʔ], called the 'glottal stop', and [h]) are produced in the glottis.[2] According to **manner of articulation** we distinguish six classes. *Plosives* (also known as *oral stops*) [p, b, t, d, k, g, ʔ]: a complete closure is made between two articulatory organs, behind which the

[2] The glottal stop [ʔ] is produced by the vocal cords when their tightly closed position is suddenly burst open; it may replace [t] in certain contexts, as in [ˈskɔʔlənd], or indicate a syllable boundary between two vowels, as in [kəuˈʔəpəreɪt].

air-pressure builds up and is then released explosively. *Fricatives* (also known as *spirants*) [f, v, θ, ð, s, z, ʃ, ʒ, h]: two articulatory organs form a narrowing so that the air stream passing through causes friction. *Affricates* [tʃ, dʒ]: complete closure is made but is released slowly, so that friction can be heard. *Nasals* (also known as *nasal stops*) [m, n, ŋ]: complete closure is made somewhere in the mouth but the air escapes continuously through the nose. *Liquids*: these are sounds of the types [l, r]. *Glides* [w, j]: there is a narrowing but it is not narrow enough to cause friction.

Plosives, fricatives and affricates are produced with a stricture impeding the flow of air, and therefore they can be called *obstruents*; while nasals, liquids and glides are produced with a relatively free airflow, and can be called *sonorants*.

The most important English consonants that we have dealt with are shown in (2). The columns represent the place of articulation, the rows the manner of articulation.

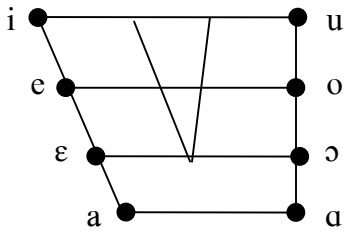
(2) English Consonants

		Bi-labials	Labio-dentals	Den-tals	Alve-olars	Pala-toal-ve-olars	Pal-atals	Vela rs	Glot-tals
Obstruents	Plo-sives	p b			t d			k g	ʔ
	Frica-tives		f v	θ ð	s z	ʃ ʒ			h
	Affri-cates					tʃ dʒ			
Sonorants	Nasals	m			n			ŋ	
	Li-quids				l r				
	Glides	w					j		

Vowels can be represented with regard to the horizontal and vertical tongue position within the oral cavity. If you raise the front of your tongue as close to the hard palate as you can without actually reaching it, you produce a close (high) front vowel: [i]. If you lower the front of your tongue as far from the hard palate as possible, you get an open (low) front vowel: [a]. Now if you

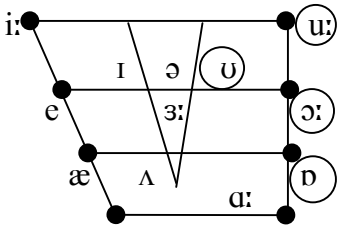
divide the distance between the tongue positions for [i] and [a] into three equal parts, you get the half-close front [e], and the half-open front [ɛ]. If you do the same movements with the back of your tongue, you will get the close back vowel [u], the half-close back [o], the half-open back [ɔ], and the open back [ɑ]. The 8 vowels so obtained are called **cardinal vowels**. They do not necessarily occur in every language, they should rather be regarded as theoretical vowels or orientation points which indicate the limits within which the tongue can move in the human mouth to produce vowels, and with reference to which all vowels of all languages can be accommodated. The trapezium formed by the cardinal vowels is called the **Cardinal Vowel Chart**, see (3).

(3) Cardinal Vowel Chart



The most important **simple vowels** of English are shown in (4). They are called simple because the particular tongue position characterising the vowel in each case is steady throughout producing the vowel. The vowels in the triangle of the chart are *central* vowels, those on the left of the triangle are *front*, those on the right of the triangle are *back* vowels. The encircled vowels are produced with *lip-rounding*: they are round vowels. The vowels whose symbols have a colon (:) attached to them are *long* vowels.

(4) English Simple Vowels



In English there are **diphthongs** as well. A diphthong is a complex vowel during the production of which one tongue position is changed into another but

no new syllable is formed. For instance, the vowels in the words *height*, *hate*, *house*, *hose*, i.e. [aɪ, eɪ, ɑʊ, əʊ], are diphthongs.[3]

It needs to be emphasised that there are far more distinguishable speech sounds (both consonants and vowels) in English than the ones we have presented in (2) and (4), but we have only concentrated on the most important ones.

Consonants and vowels together can be called segments.[4] Since phonetics primarily deals with these, the major part of phonetics is **segmental phonetics**. But phonetics has to deal with other aspects of human speech as well, viz. aspects characterising larger units than segments. This kind of phonetics is called **suprasegmental phonetics**. The suprasegmental aspects of speech include **intonation** (the meaningful melody of utterances) and **stress** (the extra prominence of a syllable over the other syllables in a word or phrase). We have already mentioned these notions in Unit 1, and in this introductory course we are not going to say more about them.

4.3 Phonology

While phonetics deals with the articulatory, acoustic and auditory aspects of actual speech sounds, **phonology** ignores all non-distinctive detail and limits its attention strictly to the really distinctive speech sounds, i.e. the basic sounds or phonemes, which form systems in a particular language. The key notion of phonology is that of contrast.

A **phoneme** is an abstract minimal sound unit of a particular language, which, when realised, is capable of distinguishing different words in that language. Phonemes can be discovered by the **minimal pair technique**. If replacing one sound by another results in a different word, the two sounds represent different phonemes and the two words form a minimal pair. For instance, the English consonants [k] and [s] represent two different phonemes because they distinguish e.g. [li:k] *leak* and [li:s] *lease*, and since the two words are otherwise identical, they form a minimal pair. The minimal pair technique is based on the notion of **paradigmatic relationship**, which we first introduced in Unit 2. (As you will remember, this is a "vertical" or "choice" relationship, which exists between a language element present in a particular environment

[3] In fact we may even distinguish **triphthongs**, in which the vowel has three tongue positions one after the other, as in e.g. *fire* and *power*, containing the triphthongs [aɪə] and [ɑʊə], respectively.

[4] This is because once they were looked upon as the smallest building blocks, i.e. segments, of speech. Today segments are regarded to be decomposable into features, cf. 4.3 below.

and all the other language elements which could replace it while yielding a well-formed complex unit.)

By means of the minimal pair technique we can distinguish 44 phonemes in Standard British English. (The pronunciation of Standard British English is sometimes referred to as **Received Pronunciation**, or just **RP**; this is the kind of pronunciation which has the highest social prestige, see Unit 9 below.) Actually, the so-called "important sounds" of English which we saw in (2) and (4), except for the glottal stop [ʔ], are all phonemes of Standard British English. In addition to them, however, there are a large number of other consonants and vowels in Standard British English which are distinct (i.e. we can distinguish them) but which are not distinctive (i.e. they do not distinguish English words) and so they are not separate phonemes, only variants of existing phonemes.

When we transcribe speech sounds from the point of view of the phonemes that they represent, we ignore all non-phonemic (i.e. non-distinctive) detail, and use a **phonemic transcription**. This is normally put between slashes: / /. In phonemic transcription we use as many symbols as there are phonemes. Consequently there are fewer symbols in phonemic transcription than in narrow phonetic transcription. For example, the phonemic transcription of the word *tool*, /tuːl/, omits non-phonemic details such as the aspiration of the initial [t] or the darkness of the final [ɬ]. These would be included in a narrow phonetic transcription: [t^{h}uː ɬ].

The myriads of actual speech sounds or phones that realise a phoneme in a language can be grouped into a small number of allophones. **Allophones** are the positional alternants of a phoneme: they are phonetically similar and are in complementary distribution. (In 3.1.2 we explained that if two language elements occur in mutually exclusive environments, then they are said to be in **complementary distribution**.) For instance, in Standard British English the phoneme /l/ has two allophones: a clear [l], which occurs before vowels, and a dark [ɬ], which occurs elsewhere, cf. *lip* [lɪp] and Helen [ˈhelən] vs. *film* [fɪɬm] and *hill* [hɪɬ]. Other examples include the English phoneme /p/, which also has two allophones: an aspirated [p^{h}] at the beginning of a stressed syllable and an unaspirated [p] elsewhere, as in *port* [p^{h}ɔːt] and *sport* [spɔːt]. Another example: any English vowel gets a nasal allophone when it is adjacent to a nasal consonant but an oral allophone elsewhere, cf. *pen* [p^{h}ẽn] vs. *pet* [p^{h}et]. The allophones of a phoneme are conditioned by the environment and so their properties are predictable or redundant. They never occur in a paradigmatic relationship with one another and so they cannot be in contrast. That is why they do not perform a distinctive function, and cannot be separate phonemes in

the same language. (However, what is non-distinctive in one language may be distinctive in another.) Since allophonic variations are not reflected in phonemic transcription, the examples given in this paragraph are phonemically transcribed as /lɪp/, /ˈhelən/, /fɪlm/, /hɪl/, /pɔ:t/, /spɔ:t/, /pen/, /pet/.

From the adjectives *phonetic* and *phonemic* the terms *etic* /ˈetɪk/ and *emic* /ˈi:mɪk/ have been abstracted, referring to two kinds of approach which can be distinguished in various types of linguistic studies. The **etic approach** deals with all data of a given kind, while the **emic approach** studies the structuring of data into systemic abstract entities on the basis of their distinctive power in a given language. In the field of sounds the emic approach is concerned with phonemes, the etic approach with allophones and phones.

Structuralist phonology (i.e. the phonology of Structuralist Linguistics, especially in the 1950s and 1960s) looked upon phonemes as the ultimate building blocks of language. Generative phonology (i.e. the phonology in Generative Linguistics) has claimed that phonemes should be decomposed into bundles (sets) of binary **distinctive features**. (*Binary* here means 'having two values'.) Take, for example, the English phonemes /p/, /b/, /m/. They all share the properties of being consonantal [+consonantal] and being pronounced with the lips [+labial], but only /b/ and /m/ are voiced [+voice], and only /m/ is [+nasal], and so on. We can draw a chart which shows the properties of each phoneme, with a + sign if the property is present with a positive value, and with a − sign if the property is present with a negative value, as is illustrated in the the partial specification of /p, b, m/ in (5).

(5) Feature Specifications (partial)

	p	b	m
consonantal	+	+	+
labial	+	+	+
voiced	−	+	+
nasal	−	−	+

Any feature which distinguishes one phoneme from another is a distinctive feature (DF). For instance, /p/ and /b/ differ in voice, /b/ and /m/ differ in nasality. Each phoneme, then, can be characterised as a bundle of DF specifications, i.e. a column of + and − marks representing the values of the features.

Phonemic and allophonic changes can be described by rules. In generative linguistics these rules typically have the following form:

(6) Rule Format

$$A \rightarrow B \,/\, X \_\_\_\_ Y$$

This means that "A becomes B between X and Y", where A, B, X, Y indicate the variables (X or Y may be missing, too), the arrow indicates the change, the slash indicates the environment, and the horizontal dash indicates the exact place of the change. For instance, nasal assimilation, which is an allophonic change turning [bɪn] into [bĩn], can be described as (7). (V stands for vowel, C for consonant.) The rule expresses that a vowel becomes nasal before a nasal consonant.

(7) Nasal Assimilation

$$V \rightarrow [+\text{nasal}] \,/\, \underline{\quad\quad} \begin{matrix} C \\ [+\text{nasal}] \end{matrix}$$

Exercises, problems, and other tasks

1. Compare the ways in which the letter *a* is pronounced in *map, many, ago, village* and the ways in which the sound [ɪ] is written in *sit, busy, village, women*.
2. Which English words are pronounced as transcribed here? /maɪs, ʃɪp, ðə, huː, eɪtθ, bæk, bɔːt, bəʊt, ɑʊt, dʒɔɪ, ˌeθnəˈgræfɪk, ɪŋkənˈsiːvəbəl/.
3. Which of the following words end with voiceless consonants and which end with voiced ones? *touch, pig, maze, lip, lathe, sit, use (!)*
4. Why do English speakers say [gʊd ˈbɔːdɪg] instead of [gʊd ˈmɔːnɪŋ] when they have a cold?
5. Pronounce the initial sounds of the following words and then determine the place and manner of articulation of each: *foot, tooth, box, chips, think, cup.*
6. Characterise English simple vowels (and also a few English diphthongs) in terms of horizontal and vertical tongue position.
7. Which English simple vowels are produced with lip-rounding?
8. How many syllables are there in the word *rain*? Why?
9. Transcribe the pronunciation of the English word *pill*, *lip*, *help*, *ten* phonetically and phonemically.
10. Transcribe phonemically: *sun, son, dam, damn, colour, collar, monkey, donkey, heat, hot, not, knot, gone, gun, ram, lamb, very, bury, birth,*

worth, sword, board, head, though, rough. Which of them are minimal pairs?

11. Which is more similar to /m/: /p/ or /b/? Why?
12. What is the general rule format in generative linguistics?
13. Try to write a rule for the elision (omission) of /t/ in words and phrases like *postman* and *must be*.
16. What are allophones?
17. Comment on the emic and etic approach.
18. Is nasality a distinctive feature of English vowels? And of English consonants?
19. Learn the IPA phonemic transcription symbols provided in Appendix A.
20. Make sure you know the following terms: phonetics, articulatory phonetics, acoustic phonetics, auditory phonetics, speech sound (phone), transcription symbol, narrow transcription, wind-pipe (trachea), larynx, glottis, vocal cords, alveolar ridge, palate, velum, roof of the mouth, pharynx, oral cavity, nasal cavity, tongue, voiced, voiceless, oral, nasal, consonant, vowel, bilabial, dental, alveolar, palatoalveolar, palatal, velar, glottal, plosive (stop), fricative (spirant), affricate, liquid, glide, cardinal vowels, front vowels, central vowels, back vowels, rounded, diphthong, stress, intonation, syllable, phoneme, allophone, complementary distribution, aspiration, emic approach, etic approach, distinctive features, assimilation, elision.

Unit 5
Morphology, the Study of Morphemes and Words

5.1 Words

The term **word** can be used in different senses. On the one hand, vocabulary items, i.e. entries in the dictionary (e.g. *take*), are called words, but on the other hand the different inflected forms of a word (e.g. *take, takes, taking, took, taken*) are also called words. Moreover, some words, e.g. (*life insurance*), are said to be written in two "words". To avoid confusion, we shall use the following terms: lexeme, syntactic word, and orthographic word.

- A **lexeme** is a unit of the lexicon (an entry in the dictionary, a vocabulary item), which is an uninflected abstract form that underlies all its inflected variants. To distinguish lexemes from their inflected variants it is customary to use capital letters to indicate lexemes. For instance, the lexeme TAKE underlies the inflected variants *take, takes, taking, took, taken.*
- A **syntactic word** is an inflected variant of a lexeme (including the zero-inflection), so *take, takes, taking, took, taken* are syntactic words. **Inflection** means varying the shape of a lexeme in such a way that its grammatical relation to other lexemes within the phrase or sentence becomes clear. Consider e.g. the sentence in (1):

(1) He takes them.

In this sentence the verb *takes* is a predicate in the 3rd person singular present tense, preceded by *he* (the subject pronoun in the nominative case), and followed by *them* (the object pronoun in the accusative case). All three words are inflected, even the apparently uninflected *he* can be regarded as zero-inflected. (Compare Hungarian *kap-ok, kap-sz, kap-0*, where the last form is not uninflected but zero-inflected.) The whole set of inflected variants of a lexeme is called a **paradigm**. The forms *take, takes, taking, took, taken* constitute the paradigm of TAKE. The members of such a paradigm are syntactic words.

- The third sense in which the word *word* is popularly used is a unit of writing: it is a stretch of graphic symbols with a space on either side and no space within. This will be called **orthographic word**. For example, the lexeme

LIFE INSURANCE is two orthographic words, but the two lexemes in *I'm* are just one orthographic word.

The three senses of the word *word* are not equally important. In our linguistic studies it is only the lexemes and syntactic words that have to be taken into consideration, and orthographic words are irrelevant. Lexemes can be likened to **types**, syntactic words to **tokens**, i.e. particular instances of the abstract types. Lexemes (and their inflected variants, the syntactic words) belong to different **syntactic categories** (= word classes, parts of speech). Nouns, verbs, adjectives, adverbs and prepositions are **content words**, others, e.g. conjunctions, pronouns, auxiliaries are **function words**. Traditional grammarians (see Unit 2) tried to define the syntactic categories of lexemes on the basis of meaning. (Just to remind you, e.g. a noun was said to be "the name of a person, place or thing", which, however, is blatantly false in the case of many words that native speakers use as nouns.) American structuralists (see Unit 3) defined the syntactic categories of lexemes not on the basis of meaning but on the basis of form. This involved an examination of word endings (e.g. any word ending in *-ness* is a noun), and of the ways in which the words enter into larger constructions (e.g. any word that fits the dash in the frame *The — is there*. is a noun).

5.2 Morphemes

To start with a working definition, which we will make more precise later, we can say that **morphemes** are the smallest meaningful units of language, which cannot be subdivided without losing their meaning. They are abstract units, indicated between braces: { }. Lexemes and syntactic words are composed of one or more than one morpheme. For instance, the lexeme TEACHER consists of two morphemes: {teach}{-er}, the lexeme ALBATROSS consists of one: {albatross}.[1] When we realise morphemes, we produce morphs. **Morphs** are the physical realisations of morphemes. The billions of actual morphs realising an abstract morpheme by actual speakers in actual situations can be grouped into a few phonologically different shapes, so called allomorphs. **Allomorphs** are the positional alternants of a morpheme: they have the same meaning and are in complementary distribution. (The latter means that they occur in mutually exclusive environments, cf. Unit 3.)

There is a perfect parallel between the morph – allomorph – morpheme series on the one hand, and the phone – allophone – phoneme series on the

[1] To express this more technically, we can say that *teacher* is bimorphemic (= having 2 morphemes), *albatrosss* is monomorphemic (having 1 morpheme).

other. Just like the phone (speech sound) and the allophone are the concern of the **etic approach** and the phoneme of the **emic approach**, the morph and the allomorph are the concern of the etic approach and the morpheme of the emic approach, cf. Unit 4.

The phonological differences between the allomorphs of a morpheme are often due to the phonological environment, i.e. the phonological differences are often **phonologically conditioned**. For instance, {-s}, the abstract plural morpheme in English has three regular allomorphs. When the last sound of the noun is a sibilant (i.e. /s, z, ʃ, ʒ, tʃ, dʒ/), the allomorph will be /ɪz/, as in e.g. *boxes*, *bushes.* When the last sound of the noun is a voiceless non-sibilant, the allomorph will be /s/, as in *books*, *plates*. And elsewhere, i.e. where the last sound of the noun is a voiced non-sibilant, the plural morpheme will be realised as /z/, as in *bags*, *apples*, *potatoes*. In other cases the phonological differences of the allomorphs can be due to **lexical conditioning**. For instance, the plural morpheme is realised as /ən/ when it is attached to the noun *ox*. Here it is not the last sound of the noun that is responsible for the alternation but the entire lexeme OX. The phonological difference of the allomorphs can also be caused by **morphological conditioning**. This happens e.g. in the plural noun *houses*, i.e.{house}{-s}, where the first morpheme is realised as /haʊz/ before the plural morpheme, although it is realised as /haʊs/ when it stands alone as a singular noun. Here one morpheme affects the realisation of another.

Morphemes can be grouped into two types on the basis of whether or not they can form monomorphemic words.

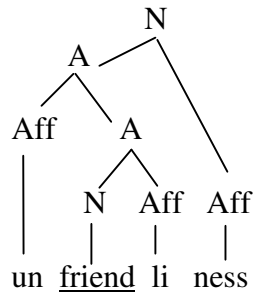
- If they can occur by themselves as whole words, (i.e. if they can form monomorphemic words), then we call them **free morphemes**. For instance, {house}, {albatross}, {kangaroo}, {lullaby}, {table}, etc. are free morphemes.

- But there are also morphemes which must be attached to other morphemes within words, these are called **bound morphemes**. For example, the plural morpheme {-s}, or the adverb-forming morpheme {-ly} are bound morphemes. Most bound morphemes are **affixes**. In English, these are either suffixes (following stems) or prefixes (preceding stems). **Suffixes** in English are either inflectional or derivational. If you add an inflectional suffix to a stem, you do not create a new lexeme, you only produce another inflected variant (i.e. another syntactic word) of the same lexeme. For example, {-s} is an inflectional suffix, because by adding it to the stem {boy}, we get *boys*, which is just another syntactic word belonging to the paradigm of BOY. However, if you add a derivational suffix to a stem, you create another lexeme. For example, {-hood} is a derivational suffix, because by adding it to the stem {boy}, you

produce a new lexeme BOYHOOD, which is the starting point of a new paradigm. **Prefixes** in Present-Day English are always derivational, e.g. {en-}, added to the stem {joy} gives rise to a new lexeme ENJOY.[2]

A **stem** is that part of a word which remains if we remove the suffix or prefix that has entered the word last. The stem is not necessarily a single morpheme, e.g. the stem of *unfriendliness* is *unfriendly*, the stem of *unfriendly* is *friendly*, and the stem of *friendly* is *friend*, cf. (2). If we remove all affixes, we arrive at the absolute stem, called **root** (also known as base), which is always a single morpheme. Thus, the root of *unfriendliness* is {friend}, underlined in (2).

(2)



The root is usually a free morpheme (as in e.g. *unfriendliness*), but there are also roots which are bound. For example, in words like *include, conclude, preclude, exclude*, etc. the prefix {in-}, {con-}, {pre-}, {ex-}, etc. is followed by the root {-clude}, which is not a free form since it never occurs alone as a monomorphemic word. Moreover, the meaning of {-clude} is unclear, in fact it is dubious whether it has any meaning at all. (If you know Latin you may think that {-clude} means 'to close', but this can hardly be felt in e.g. *conclude*. Besides, native speakers of English do not think of Latin when they use such words.) Although we do not know if {-clude} has a meaning or what that meaning is, we still regard it as a morpheme, because its pronunciation /klu:d/ systematically varies with /klu:s/ when it is followed by the suffix {-ive}, as in *inclusive* or *conclusive*, and with /klu:ʒ/ when followed by the suffix {-ion}, as in *inclusion*, *conclusion*. To put it in another way, {-clude} has allomorphic variants: /klu:d/, /klu:s/ and /klu:ʒ/, which shows that it is a morpheme.

Since, as we have just seen, the criterion of meaning cannot always be used, we shall revise our original definition, which we gave at the beginning of this section, in the following way: **Morphemes** are the smallest meaningful

[2] While inflectional affixes correspond to what are called *ragok* and *jelek* in Hungarian linguistics, derivational affixes correspond to *képzők*.

units of language or the units of allomorphic variation, which cannot be subdivided without losing their meaning or losing their allomorphic variability. To put it more informally, morphemes are recurring word-parts which have some constant variants, and which are typically but not necessarily meaningful. This definition will cover all morphemes that we have considered so far.

5.3 Segmentability of words into morphemes

There are words which are easy to segment into morphemes, e.g {un-}{friend}{-li}{-ness}, {girl}{-s}, {smoke}{-ed}, etc. Languages in which most words are of this kind (i.e. in which most words are sequences of separable morphemes) are called **agglutinating languages**. For instance, Hungarian is a typical agglutinating language, cf. e.g. {pénz}{-telen}{-ség}{-em}{-től} ('money-less-ness-my-from' i.e. 'from my not having money').

There are also many words which are monomorphemic, i.e. which are composed of single morphemes. In these, morphemes coincide with words, e.g. *go, coffee, elephant*. Languages in which most words consist of single morphemes are called **isolating languages**. Classical Chinese is one of them.

Finally, there are words in which the constituent abstract morphemes are fused together in an inseparable way, e.g. the English words *took* and *mice* consist of {take}{-ed} and {mouse}{-s}, respectively. Also, there are words in which some morphs represent inseparable fusions of abstract morphemes, e.g. the Russian nominal inflectional suffix *–u*, as in e.g. *lampu* ('lamp-Acc.'), simultaneously realises {Feminine}, {Singular}, {Accusative}. Languages in which the fusion of morphemes is typical are called **fusional** (= inflecting) **languages**. Latin is a typical fusional language.

Of course, these language types, established on the basis of the morphological make-up of the majority of their words, are not pure types. English, for instance, is a mixture of all three, but it is predominantly (statistically) isolating because a large part of its words are monomorphemic. By contrast, Hungarian is predominantly agglutinating.

5.4 Word formation

In addition to the borrowing of **loanwords** from other languges, e.g. ALCOHOL from Arabic) or the introduction of **coinages** (lexemes artificially invented, e.g. XEROX), there are also ways in which we can produce new lexemes, making use of old ones. These ways are called **word formation processes**. We shall now briefly discuss the most common word formation processes.

- One of the major word-forming processes is **derivation** (= affixation), i.e. creating a new lexeme by means of adding a derivational prefix or suffix to an old lexeme. For instance, the lexeme KINGDOM is derived from the stem {king}, to which the derivational suffix {-dom} has been added, or the lexeme IMPOLITE is derived from the stem {polite}, with the derivational prefix {in-}, or the lexeme UNHAPPINESS is derived from the stem UNHAPPY (itself derived from HAPPY), by adding the derivational suffix {-ness}. The lexemes produced by affixation can be called **derivative words**, or simply just **derivatives**.

- It can happen that a lexeme is assigned to another word class (part of speech) without changing its form. This is called **conversion** (also known as zero affixation), which is extremely common in English, see e.g. $\text{BOTTLE}_N \rightarrow \text{BOTTLE}_V$, $\text{DAILY}_A \rightarrow \text{DAILY}_N$, $\text{MILK}_N \rightarrow \text{MILK}_V$. A special subtype of conversion is called **approximate conversion**, in which lexemes undergo a small but systematic change in pronunciation and are thereby assigned to a different word class. Sometimes this "small change" is a stress shift (with some concomitant changes in vowel quality), as in e.g. $\text{SUS}'\text{PECT}_V$ and $'\text{SUSPECT}_N$, $\text{PER}'\text{MIT}_V$ and $'\text{PERMIT}_N$, $\text{CON}'\text{VICT}_V$ and $'\text{CONVICT}_N$, $'\text{ENVELOPE}_N$ and $\text{EN}'\text{VELOP}_V$, etc. Another kind of approximate conversion is changing the voice value of the final fricative in some lexemes, having a voiceless final fricative in nouns and a voiced one in verbs, e.g. HALF_N /hɑːf/ → HALVE_V /hɑːv/, USE_N /juːs/→ USE_V /juːz/, WREATH_N /riːθ/→ WREATHE_V /riːð/. Changing the voice value of the final fricative is sometimes accompanied by vowel change, too, e.g. GLASS_N /glɑːs/ → GLAZE_V /gleɪz/.

- The next major word-forming process is **compounding**. This means bringing together two roots or two lexemes to produce a new lexeme, called a **compound**, as in e.g. 'BLACKMAIL, 'GOLDFISH, 'WHITE HOUSE (where the president of the US lives), 'HAY FEVER, 'CHRISTMAS-TREE, ˌCHRISTMAS 'PUDDING, etc. Members of a compound may be compounds themselves, cf. e.g. 'RAILWAY-ˌSTATION ATˌTENDANT, etc. Typically (but not always) compounds bear the main stress on their initial member. They are usually (but not always) written in one orthographic word.

Other word-formation processes include clipping, blending, backformation and the formation of acronyms.

- **Clipping** means shortening a lexeme and thus producing a more informal variant, e.g. PHOTOGRAPH → PHOTO, INFLUENZA → FLU, EXAMINATION → EXAM, etc.

- **Blending** is putting together lexemes but at least one of these lexemes is present only in a fragmentary form, as in e.g. FOG + SMOKE → SMOG, BREAKFAST + LUNCH → BRUNCH, etc. The lexemes so produced are **blends**.
- A kind of reverse affixation takes place in **backformation**, which means establishing, on the basis of analogy with derivatives, the apparently existing stem of a lexeme which looks like a derivative, although it is not a real derivative. For instance, pairs like SUPERVISION_N and SUPERVISE_V suggest that if there is a word TELEVISION_N, there should also be a word TELEVISE_V, although the former was not derived from the latter. We say that TELEVISE has been backformed from TELEVISION. Similarly, DONATE has been backformed from DONATION (by analogy with pairs such as CREATION and CREATE).
- Finally, **acronym formation** means forming a lexeme from the initial letters or larger parts of words; the lexemes so created are **acronyms.** Many of these are pronounced as words, e.g. RADAR for 'radio detecting and ranging', NATO for 'North Atlantic Treaty Organization'. In many cases, however, they are pronounced as sequences of letters, e.g. BBC, YMCA, etc.

Exercises, problems, and other tasks

1. Enumerate the syntactic words belonging to the lexemes BE and HAPPY.
2. Identify the syntactic category of each of the lexemes in *The camel-driver smoked the cheapest cigars*.
3. Identify the morphemes in: *The camel-driver smoked the cheapest cigars*.
4. When two syntactic words are pronounced differently and spelt identically, they are **homographs**, e.g. *read* /riːd/ - *read* /red/; when they are spelt differently and pronounced identically, they are **homophones,** e.g. *raise* /reɪz/ - *rays* /reɪz/ ; when they are both pronounced and spelt identically, they are **homonyms**, e.g. *bear* /beə/ 'medve' - *bear* /beə/ 'hord'. Pick out the homographs, homophones and homonyms from: *bow*, *scent*, *sent*, *rain*, *reign*, *rein*, *fair*, *fare, lead* (V), *lead* (N), *sun*, *son, one*, *won*.
5. Transcribe the words in (4) phonemically.
6. What are the regular allomorphs of the English past tense suffix {-ed} (as in *ended*, *laughed*, *begged*)? Explain this dependence phonologically.
7. What can condition the shape of an allomorph? Think of the past tense form of *skip*, *shut*, *keep*.
8. Enumerate the inflectional suffixes of Present-Day English.

9. What is {ge-}in Old English *ic habbe geleorned* ('I have learnt')? Are there any inflectional prefixes in Present-Day English?
10. Why do we say that *unfriendly*, rather than *friendliness*, is the stem of *unfriendliness*?
11. Give a labelled bracketing representation of the morphemic structure of the word *unfriendliness*.
12. Draw labelled tree diagrams to show the morphemic structure of these words: *hospitalisation*, *organisation*, *desirability*, *ungentlemanliness*; *greenhouse, oil well, red-hot, breastfeed, dog food box*.
13. List the bound morphemes to be found in these words: *misleads*, *submit, previewer*, *shortened*, *unhappier*, *fearlessly*, *permitted*. Classify the bound morphemes you have found into roots and derivational or inflectional affixes.
14. Discuss each process of word formation and collect examples of your own.
15. Make sure you know the following terms: lexeme, syntactic word, orthographic word, inflection, paradigm, parts of speech, content words, function words, morpheme, allomorph, morph, free morpheme, bound morpheme, affix, suffix, prefix, inflectional and derivational affixes, stem, root (base), agglutinating language, isolating language, fusional or inflecting language, loanword, derivation, derivative word, conversion, compound, clipping, blend, backformation, acronym, homograph, homophone, homonym

Unit 6
Syntax, the Study of the Structure of Phrases and Sentences

6.1 Sentences and phrases

Syntax is the study of sentence structure. Sentences are composed not directly out of words but of constituents which may consist of more than one word, called phrases. A **phrase** is an expression which is a constituent in a sentence and is the expansion of a **head** (i.e. key word). For instance, the constituent *the king* in (1), or the constituents *my brother* and *an expensive car* in (2) are Noun Phrases, abbreviated as NPs, because their key elements are the nouns (Ns) *king*, *brother* and *car*, respectively.[1] It can happen that a phrase is realised by a single word, for example the NPs *John*, *Mary* and *apples* in (3) consist of the Ns *John, Mary* and *apples,* and nothing else. In (4) *he* is a special NP because its head is a pronoun rather than a noun.

(1) The king laughed.
(2) My brother bought an expensive car.
(3) John gave Mary apples.
(4) He went home.

(1)-(4) are sentences. The terms **sentence** and **clause** can be used synonymously. A sentence or clause is an expression which minimally contains a subject and a predicate, and which may also contain other types of elements, viz. complements and adjuncts. For instance, (1) consists of just a subject and a predicate. The NP *the king* is the **subject**, and the Verb Phrase (VP), composed of a single verb (V) *laughed*, is the **predicate**.

A **complement** is a constituent whose presence is structurally "dictated" (required or licensed) by a particular word. The presence of the complement "follows" from the presence of the word which it is a complement of. For instance, in (2) above the NP *my brother* is the subject, the V *bought* is the predicate, and the NP *an expensive car* is a complement, more particularly a

[1] What we call *Noun Phrase* (*NP*) here has been recently reanalysed in many syntactic works as *Determiner Phrase* (*DP*). However, in this introductory course we shall continue to use the old term *Noun Phrase* (*NP*) instead.

direct object, of the verb *bought*. An object is a particular kind of complement. In (3) above the subject is the NP *John*, the predicate is the V *gave*, and there are two complements, the NP *Mary*, functioning as an indirect object, and the NP *apples* functioning as a direct object. In (4) the complement of the V *went* is the Adverb Phrase (AdvP) *home*, consisting of the single adverb (Adv) *home*.

The subject and the complement(s) together are said to be the arguments of the predicate. **Arguments** are the participants (entities) that are necessarily involved in the situation identified by the predicate. For example, in (2) the predicate *bought* has two arguments: the subject (somebody did the buying), and the object (something was bought). In English, subjects typically occur in the **nominative case** (*I*, *he*, etc.), whereas objects occur in the **accusative case** (*me*, *him*, etc.), but observable case-marking is restricted to pronouns. Another difference between subjects and complements is that, in English, verbs agree with their subjects in person and number but do not agree with their complements. Also, subjects in English typically precede verbs, while complements follow them.

In addition to the subject, verb and complement(s), the sentence or clause may also contain constituents which are not structurally required by the verb but add optional information about place, time, manner, purpose, etc. Such constituents are called **adjuncts**. Some of these function as adverbials, e.g. the Prepositional Phrase (PP) *on Tuesday* in (5) is a time adverbial, the Adverb Phrase (AdvP) *very quickly* in (6) is a manner adverbial. Some of the adjuncts function as attributes within noun phrases, e.g. the Adjective Phrase (AP), realised by a single Adjective (A) *expensive* in (5), is an attribute of *car*.

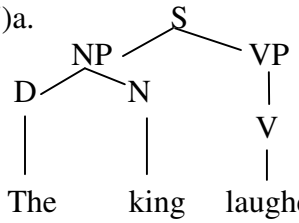
(5) My brother bought an expensive car on Tuesday.
(6) He went home very quickly.

The terms *subject*, *predicate*, *object* (direct and indirect), *adverbial*, *attribute*; *complement* and *adjunct* refer to **grammatical functions** which constituents may perform in the sentence, whereas terms such as *NP*, *VP*, *AP*, *AdvP*, *PP*, *N*, *V*, *A*, *Adv*, *P*, etc. refer to **syntactic categories**, i.e. they name the grammatical category to which the constituent belongs. These two sets of terms are fairly independent of each other, e.g. an NP can function as subject, or as object, or as the complement of a preposition, or even as adverbial (e.g. the NP *last year*). Similarly, the function of **adverbial** can be performed by an AdvP (*very quickly*), a PP (*on Tuesday*), an NP (*last year*) or even by an embedded clause (e.g. *when I was writing a letter*).

6.2 Representation

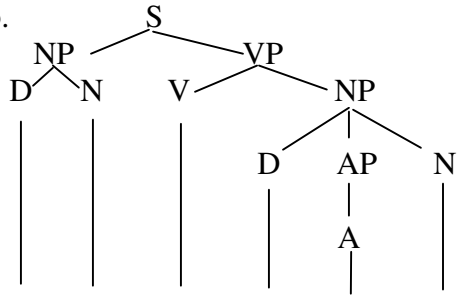
The constituent structure of sentences can be represented in essentially two ways: by means of labelled **tree diagrams**, and by means of labelled **bracketings** (see Unit 2). Although the two ways of representation are logically equivalent, we prefer tree diagrams because they help visualise structure better than bracketings do. Tree diagrams are like uprooted trees, with branches and nodes. The **nodes** in a tree diagram are the topmost point, the bottom points, and all those intermediate points at which the tree branches. The **labels** are the abbreviated names of the categories to which the constituents belong. The new labels in (7) are S, D, Pron, Aux, and DegP; these stand for Sentence, Determiner, Pronoun, Auxiliary, and Degree Phrase, respectively. Please note that the complements in (7b), (7c), (7d), viz. *an expensive car*, *Mary*, *apples*, *home*, are sisters of the verb, while the adjuncts, viz. *on Tuesday* and *very quickly* in (7e) and (7f), are adjoined to the VP, with which they form a higher VP.[2]

(7)a.



[$_S$[$_{NP}$[$_D$ The][$_N$ king]] [$_{VP}$[$_V$ laughed]]].

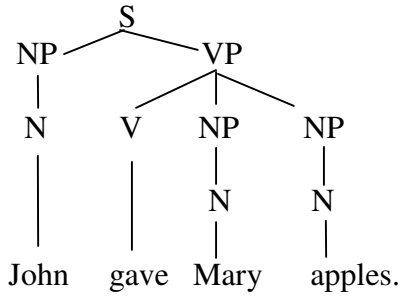
b.



[$_S$[$_{NP}$[$_D$ My][$_N$ brother]] [$_{VP}$[$_V$ bought][$_{NP}$[$_D$ an][$_{AP}$[$_A$ expensive]][$_N$ car]]]].

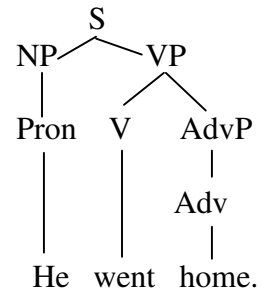
[2] The syntactic analyses embodied in the tree diagrams of this book are strictly preliminary and are likely to be substantially modified in your later studies.

c.



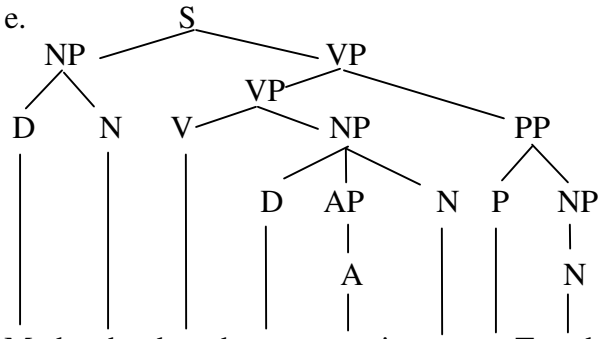
[S[NP[N John]][VP[V gave][NP [N Mary]] [NP [N apples]]]].

d.



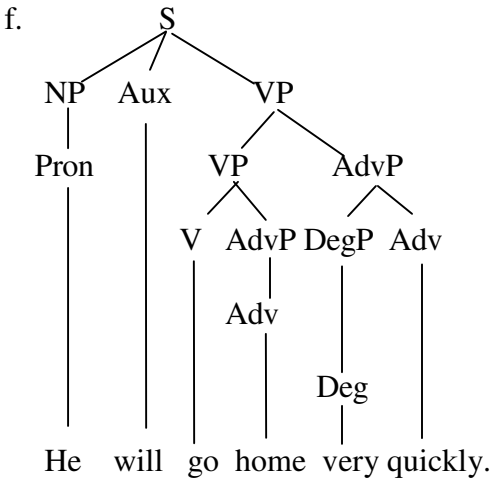
[S[NP[Pron He]][VP[V went][AdvP[Adv home]]]].

e.



[S[NP[D My][N brother]] [VP [VP[V bought][NP[D an][AP[A expensive]] [N car]]]
[PP[P on][NP[N Tuesday]]]]].

f.



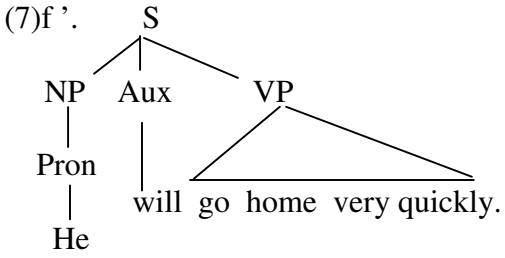
[$_{S}$[$_{NP}$[$_{Pron}$ He]] [$_{Aux}$ will] [$_{VP}$ [$_{VP}$ [$_{V}$ go][$_{AdvP}$[$_{Adv}$ home]]]
[$_{AdvP}$[$_{DegP}$[$_{Deg}$very]][$_{Adv}$ quickly]]]].

In the last example, (7f), the auxiliary *will* stands as a separate constituent outside the VP, although we may intuitively think the auxiliary should be part of the VP. One of the reasons why we analyse it as being outside the VP is that the VP may be deleted independently of the Aux, see e.g. (8).

(8) Speaker A: Will he go home very quickly?
Speaker B: Yes, he will ~~go home very quickly~~.

When we do not want to specify the internal structure of a particular constituent, we may replace the part of the tree diagram corresponding to it by a triangle. For instance, if, for any reason, we wish to ignore the internal structure of the VP *go home very quickly* in (7f), we may use a triangle for this part of the sentence, see (7f', to be read as "seven-ef-prime").

(7)f '.



[$_{S}$[$_{NP}$ [$_{Pron}$ He]] [$_{Aux}$ will] [$_{VP}$ go home very quickly]].

A node in a tree is said to **dominate** (i.e. contain) all the nodes below it that are linked to it by a line. A string of words (which consists of minimally one word) is a **constituent** in a tree if there is a node which exclusively dominates it, i.e. dominates all and only the words in that string. For instance, in (7f) each word is a separate constituent because each one is exclusively dominated by a node (*he* by the node Pron, *will* by the node Aux, *go* by the node V, *home* by the node Adv, *very* by the node Deg, and *quickly* by the node Adv), but the strings *go home* and *very quickly* are also constituents because they are exclusively dominated by the lower VP and the AdvP, respectively, and the string *go home very quickly* is a constituent, too, because it is exclusively dominated by the higher VP. However, the words *home very* do not form a constituent in (7f) because there is no node in this tree which would dominate these two words and only these two words.

When a node dominates lower nodes without the intervention of intermediate nodes, we speak about **immediate domination.** A string of words is called an **immediate constituent** (IC) in a tree when there is a node which immediately dominates all and only the words in that string. Thus, the immediate constituents of the sentence in (7f) are *He*, *will*, and *go home very quickly*, because these are the NP, Aux and VP which are immediately dominated by the sentence. The sentence is "mother" to its immediate constituents, the immediate constituents are "daughters" to the sentence, and "sisters" to each other. The immediate constituents (i.e. daughters) of the VP *go home very quickly* are the lower VP *go home* and the AdvP *very quickly*. The immediate constituents of the lower VP *go home* are the V *go* and the AdvP *home*, and those of the AdvP *very quickly* are the DegP *very* and the Adv *quickly*.

6.3 Simple and complex sentences

Until now, all the constituents (apart from the topmost ones) within our example sentences have been phrases and lexical items of various kinds: NPs and Ns, VPs and Vs, APs and As, AdvPs and Advs, PPs and Ps, DegPs and Degs, Auxes and Ds. None of the constituents was a sentence (S). Therefore we can say that all our examples so far have been simple sentences. A **simple sentence** is a sentence which contains no lower sentence (clause) embedded in it; to put it in another way, it is a sentence which has no S-node other than the topmost S-node in it.

However, it can happen that a non-topmost constituent within a sentence is itself a sentence. This is the case in (9), where the complement (more precisely the object) of the verb *believes* is not an NP but an S. This lower

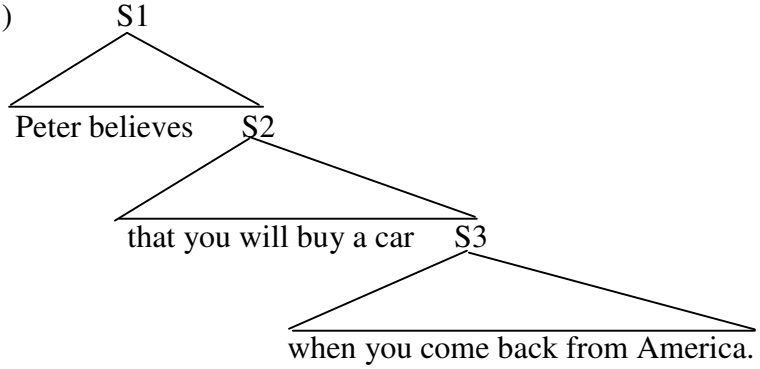
sentence (S2) functions as a complement clause within the higher sentence (S1).

(9) [$_{S1}$ Peter believes [$_{S2}$ that you will buy a car]].

The phenomenon in which a constituent contains constituents of the same category as itself is known as **recursion**. For instance, in our previous examples (7e) and (7f) we saw that a VP contained a lower VP. In (9), however, recursion applies to the category S, so here we can speak about sentential or clausal recursion. A sentence containing a lower sentence embedded in it is called a **complex sentence**. (9) is a complex sentence, because it contains two sentences: a higher one, called **matrix clause**: *Peter believes (that) you will buy a car*, and a lower one, called **embedded clause** or **subordinate clause** (or just **subclause**, for short): *(that) you will buy a car*. It can happen that a subclause has its own subclause and so the upper sublause is the matrix clause of the lower, as in (10), whose simplified tree representation is given in (11).

(10) [$_{S1}$ Peter believes [$_{S2}$ that you will buy a car [$_{S3}$ when you come back from America]]].

(11)



The topmost matrix clause minus the subclause it contains is also known as the **main clause**. So in both (9) and (10) the main clause is *Peter believes …* Let us now consider two different kinds of subclause in (12a) and (12b).

(12)a. I didn't know [George/he collects stamps].
b. I've never known [George/him collect stamps].

In (12a) the verb *collects* carries the inflectional suffix *-s*, which shows that the verb is inflected for agreement with its subject (third person singular) and simultaneously for present tense. We regard **tense** as an inflection on the first

auxiliary or, if there is no auxiliary, on the verb in the sentence, consequently we distinguish only two tenses in English: present tense, e.g. *collect-s* and *collect-0* or *will-0*, and past tense, e.g. *collect-ed*, or *will-ed* = *would*.[3] By contrast, in (12b) the verb *collect* does not agree with *George* and it is tenseless. The verb in (12b) is uninflected for agreement and tense. A further difference is that in (12a) the subject of the embedded clause, *George*, can be replaced by a pronoun in the nominative case: *he*, but in (12b) *George* can only be replaced by a pronoun in the accusative case: *him*.

In (13a) the auxiliary *will* is inflected for tense (it is in the present tense: *will-0*, its past tense form would be: *would*, i.e. *will-ed*). This is in contrast with the untensed particle *to* of the infinitive in (13b). And just like in (12a) and (12b), the subject of the bracketed clause, *George*, can be replaced by the nominative case pronoun *he* in (13a) and the accusative case pronoun *him* in (13b).

(13)a. I expect [George will win]. / I expect [he will win].
b. I expect [George to win]. / I expect [him to win].

From (12) and (13) we can conclude that sentences or clauses can be finite and nonfinite. A **finite clause** has a subject in the nominative case and contains a verb or an auxiliary inflected for tense / agreement. A **non-finite clause** does not have a nominative subject and does not contain a verb or auxiliary inflected for tense / agreement. The subclauses in (12a) and (13a) are finite, whereas those in (12b) and (13b) are non-finite.

The subject of an English non-finite subclause can be an invisible pronoun called **PRO** (pronounced: 'big pro'), too, as in (14a) and (14b).

(14)a. [PRO to swim here] is dangerous.
b. We want [PRO to buy a new printer].

In (14a) the PRO has a general interpretation ('anyone'), whereas in (14b) it inherits the features of its **antecedent**, the main clause subject *we*, and so PRO, like *we*, is also first person plural.[4] PRO satisfies the requirement that we have

[3] Although traditional school grammars speak about 12 "tenses", viz. the simple, continuous, perfect, and perfect continuous "tenses" in present, past, and future, we do not follow this tradition because it mixes up tense with aspect and modality, and words with phrases.

[4] An antecedent can be defined as a constituent from which another constituent (usually later in the sentence or in the discourse) derives its interpretation. For instance, *the boy* is the antecedent of *he* in: *The boy was eating rapidly because he was very hungry*.

set up for the subjects of non-finite clauses: it is not a subject in the nominative case.

Non-finite verb forms are the bare infinitive and *to*-infinitive forms (e.g. (*to*) *write*), the *-ing* form (e.g. *writing*), and the *-en* form (e.g. *written*) of verbs. (Though the latter two are inflected, they are not inflected for tense and agreement!) To sum up: the bracketed subclauses in (12b), (13b), (14a) and (14b) are non-finite, those in (9), (10), (12a) and (13a) are finite. All the main clauses are finite.

Finally, consider the bracketed subclause in (15), *which we saw last week*. This is part of the NP *the shoes which we saw last week.*

(15) [$_{S1}$ I've bought [$_{NP}$ the shoes [$_{S2}$ which we saw last week]]].

Here the NP itself is the object (complement) of the matrix verb *bought*. The subclause modifies (is an adjunct to) the noun *shoes*. Since *which* relates to (refers back to) *shoes*, it is called a Relative Pronoun and the subclause which contains it is called a **Relative Clause**. More precisely, we can say it is a Defining (or Restrictive) Relative Clause because it helps identify the referent of the word *shoes*, i.e. tells us which particular shoes the speaker is actually talking about.

6.4 Compound (coordinated) sentences

A compound or coordinated sentence contains two or more, equally important, lower sentences (clauses) placed side by side in coordination. There are essentially three kinds of relationship between the coordinated clauses: additive, adversative, and disjunctive. The label *Conj* stands for conjunction.

(16) Additive: [$_{S}$ [$_{S}$ Her daughter was a teacher] [$_{Conj}$ and] [$_{S}$ her son was studying arts]].
(17) Adversative: [$_{S}$ [$_{S}$ I asked him] [$_{Conj}$ but] [$_{S}$ he refused]].
(18) Disjunctive: [$_{S}$ [$_{S}$ I can go to my friends] [$_{Conj}$ or] [$_{S}$ my friends can come to me]].

6.5 Sentence and utterance

Sentences have to be distinguished from utterances. A **sentence** (mondat) is any string of words produced by the sentence-forming rules of a language, these rules are stored in native speakers' competence. (By **competence** we mean the native speaker's intuitive knowledge of language, see Unit 2 above.) So sentences are

constructs of competence, they are ideal, abstract entities. For instance, *Peter smokes cheap cigars* is an English sentence because it has the structure of an English sentence.

By contrast, an **utterance** (megnyilatkozás) is typically the physical realisation of a sentence in a real situation of language use, i.e. in performance. (**Performance** is the actual use of competence and it involves individual and situational factors, see Unit 2.) Since utterances belong to performance, in spontaneous speech they often contain imperfections, such as hesitations, false starts, lack of concord, etc., especially if the speaker is tired or excited or embarrassed. For instance, (19) is an attested utterance made by a Hungarian artist in a TV interview, and it realises a hybrid of two different but synonymous Hungarian sentences shown in (20).

(19) *?Akkor mint betűszedőként működött.
then as typesetter-as worked-3sg
(20)a. Akkor mint betűszedő működött.
b. Akkor betűszedőként működött.
'At that time he was working as a typesetter.'

"Incorrect" utterances like (19) are often made by native speakers but they do not seem to matter because, on the basis of their competence, Hungarian speaker-hearers automatically interpret (19) as either (20a) or (20b). That is why such mistakes are usually not corrected and often not even noticed.

The utterances we make are not necessarily the realisations of complete sentences. It may happen that sentences are left unfinished because the speakers change their minds in the middle of the sentence and begin a new one, or because they are interrupted by someone, or because they are shot dead before they have finished, etc. So it may happen that an utterance consists of a fragment of a sentence or that parts of it belong to different languages. **Ellipsis** (= omission) of predictable constituents is quite common.

Exercises, problems, and other tasks

1. What is the difference between an adverbial and an adverb?
2. Identify the subject, predicate (verb), complement(s), adjunct(s) in examples (7a)-(7f).
3. What kind of complement(s) and adjunct(s) do we find in (7a)-(7f)?
4. Distinguish grammatical functions and categories. What functions can an NP perform? What categories can perfom the function of Adverbial?

5. Identify the constituents and immediate constituents in (7b) and (7e). Describe their relationships using the metaphorical mother-daughter-sister-terms.
6. In (7b), why are *brother bought*, *bought an*, *expensive car* not constituents?
7. Try to reveal the constituent structure of the following sentences and phrases by drawing trees for them: a. *The big dog followed the small boy*. b. *My friend ran home*. c. *The president went to the wedding last Sunday*. d. *The reporter realised that the minister had lied*.
8. The following expressions are ambiguous. Try to disambiguate them by means of unlabelled tree diagrams:
   a. *old men and women* b. *I met a foreign language teacher on the train*.
9. What is the difference between a simple sentence and a complex sentence?
10. Find examples of recursion among the tree diagrams (1)-(10).
11. Identify the function of the bracketed subclause in the following:
    a. *I don't know* [*where the plane crashed*]. b. [*Where the plane crashed*] *is still uncertain*. c. *Let's meet* [*where the plane crashed*].
12. How do we define tense and how many tenses do we distinguish in English? What is the difference between the forms *may* and *might*, *shall* and *should*, *write* and *wrote*? How do we analyse the string *will go*?
13. What controls PRO in a. *I promised Mary* [PRO *to buy it*]. b. *I persuaded Mary* [*PRO to buy it*]. c. [*PRO to buy it now*] *would be immoral*.
14. Identify and comment on the subclause in [*the dentist* [*who*(*m*) *you've consulted*]].
15. Compare and establish the difference between the subclauses in [*the dentist* [*who*(*m*) *you've consulted*]] and [*my father*, [*who*(*m*) *you've consulted*]]. Note that only in the latter do we have a comma before the relative pronoun.
16. Draw simplified trees on the basis of the bracketings for the compound sentences (16)-(18).
17. If you are a native speaker of Hungarian, comment on what the following utterances show:
    a. *? *Pedig tudom, hogy az egyik legnagyobb emberek egyike*.
    b. *? *Elviszem ezt a könyveket a Péternek*.
    c. *? *Azokba a…hm…ódon falak között azér mosis…most is hűvös lehet*.
    d.*? *Mi lenne a …öööö… mivel tudnának kiegyezni?*
    e. *? *Az ördög megint a részletekben zajlik*.
18. Comment on this utterance: *? *It's uh … it's hu not … I mean … actually well I have just sort of … er … sort of thought of going to … bed*.

19. In which of the following utterances does ellipsis depend on the linguistic context? a. *Anybody need a lift?* b. *No, tomorrow.* c. *Looking for me, Terry?* d. *Sorry*. e. *In Bristol*. f. *Yes, I have*.
20. Make sure you know the following terms: phrase, head, sentence (clause), subject, object, adverbial, predicate, complement, argument, nominative case, accusative case, genitive case, adjunct, node, domination, constituent, immediate constituent, simple sentence, recursion, complex sentence, matrix clause, embedded clause (subordinate clause), main clause, tense, finite clause, non-finite clause, PRO, antecedent, compound sentence, utterance, ellipsis; noun, pronoun, verb, auxiliary, adverb, adjective, numeral, article, determiner, preposition, interjection, conjunction; parts of speech; masculine gender, feminine gender, neuter gender; singular number, plural number, declarative, interrogative, imperative, exclamatory; relative clause, adverbial clause.

Unit 7
Semantics, the Study of Meaning

7.1 Kinds of meaning

Semantics is the study of the meaning of meaningful units.[1] It is, however, notoriously difficult to define what **meaning** is, and linguistics – though it has offered several (partial) solutions –, is still searching for a satisfactory definition.

Meaning is not homogeneous. The most important, central kind of meaning can be called **cognitive meaning**. In the case of declarative sentences, this is a state of affairs described by the sentence, which can be true or false. The cognitive meaning of a sentence is sometimes called **propositional meaning** or **proposition**. In the case of words, cognitive meaning is the contribution that the word (lexeme) systematically makes to the cognitive meaning of sentences. The cognitive meaning of lexemes is sometimes called **sense**.

The sentence in (1) describes a state of affairs, and its cognitive meaning is the set of conditions which have to be fulfilled to make the sentence true.

(1) The girl went to the garden.

There are other kinds of meaning, too. For instance, we can say (2), which is different from (1) in terms of **stylistic meaning**, although cognitively identical with it. Or, we can say (3a) and (3b), which, being a question and an imperative, respectively, cannot be treated as being either true or false, but they can be claimed to have a questioning and a commanding **speech act meaning**, respectively. Or, we can say (4), where, in addition to the cognitive meaning that the sentence has, we have considerable **emotive (affective) meaning**, too (expressed not only by the word *wow* but also by its special intonation).

(2) The damsel made her way to the garden. (formal, archaic style)
(3)a. Did the girl go to the garden? (question)
b. Let the girl go to the garden. (command)
(4) Wow! The girl went to the garden! (emotional)

[1] The term *semantics* comes from Greek *sēmantikos*, which means 'meaningful, significant'.

By a somewhat arbitrary decision, we claim that semantics is concerned with cognitive meaning, while the other kinds of meaning are the concern of pragmatics (see Unit 7). In the present unit we shall only deal with cognitive meaning.

7.2 Approaches to word meaning

It has often been thought that word meaning is primary and sentence meaning secondary. Let us examine this assumption.

We distinguish two approaches based on the primacy of word (lexeme) meaning: the referential theory and the conceptual theory.

- The **referential theory** of word meaning assumes that lexemes mean what they refer to (i.e. what they "name"). This view concentrates on the **referents** (= extensions, denotata) of lexemes. This seems correct in the case of proper names, e.g. the name *Buckingham Palace* refers to the object Buckingham Palace in London. The theory can be extended to non-names as well: common nouns (e.g. *boy*) can be regarded as referring to sets of individual objects, verbs (e.g. *eat*) as referring to actions, adjectives (e.g. *big*) as referring to properties of individuals, and adverbs (e.g. *happily*) as referring to properties of actions. There are, however, serious problems with this theory. There are lexemes that do not refer to anything in the extralinguistic word, e.g. *fairy*, or lexemes that refer to something that used to exist in the past but no longer exists today, e.g. *dinosaur*, but we cannot deny that they have meaning. And, last but not least, there are lexemes which perform grammatical functions in sentences (so called **function words**, such as *if*, *very*, *why*, *and*, etc.), and can in no way be thought of as referring to anything in the world. But they do have meaning.
- The other approach is the **conceptual theory** of word meaning, which is based on the **concepts** with which lexemes are associated. Under this theory what a lexeme means is the sum of the most essential features of the concept associated with the lexeme, i.e. a set of **semantic features**, (= intension**)** which native speakers have to know and agree upon. These features are pieces of information by which the meaning of a lexeme can be – at least partially – specified. Breaking up the meaning of a lexeme (i.e. the concept associated with it) into semantic features is called **componential analysis** (= lexical decomposition, intensional definition**)**. For instance, native speakers of English agree that the meaning of the noun *assassin* contains the following semantic features: 'person' who 'murders' 'important people'. Or, the semantic

features of the verb *die* are: 'animate being' 'becomes' 'not alive'.[2] Or, the semantic features of the noun *man* are: 'male', 'adult', 'human'. The semantic features need not be scientifically correct. Consider, for example, the lexeme *whale*, whose popular conceptualisation does not necessarily contain the feature 'mammal' (many speakers are not aware of this), although the feature 'mammal' is undoubtedly part of the scientific definition of what a whale is.

The conceptual theory also has its problems. The most obvious one is that a large number of lexemes are not associated with concepts at all. Again we can think of many function words that are meaningful because they affect the meaning of sentences (e.g. *if*, *very*, *why*, *and*, etc.) but are not definable in terms of concepts or essential features of concepts. And proper names like *George Bush* or *The Louvre* also contribute to the meaning of sentences, but they are not associated with concepts.

From this it follows that we cannot propose a definition of word meaning suitable for all words, without taking sentence meaning into consideration. Since word meaning cannot be given an independent characterisation, our original assumption, viz. that word meaning is primary, has to be replaced by the view that sentence meaning is primary. If we give sentence meaning an independent characterisation, then the meaning of any word can be defined as the contribution it systematically makes to the meaning of the sentences in which it occurs. This will cover not only the meaning of content words but the meaning of function words and the meaning of proper names, too. It is important to emphasise that we do not deny word meaning, we only identify it with the contribution the word makes to the meaning of the sentence. We shall return to the question of word meaning after we have discussed sentence meaning.

7.3 Sentence meaning

We accept a truth-based account of sentence meaning. According to this, what a declarative sentence means is the set of the conditions that are necessary and sufficient for the sentence to be true. These are called the **truth conditions** of the sentence. Consider (5).

(5) A boy saw a mouse.

[2] A sentence such as *The engine died*. is metaphorical: we treat *the engine* as if it was a living being, which can *die,* i.e. 'stop operating'.

This sentence is true if and only if an individual that has the features which we attribute to boys (i.e. 'human', 'male', 'non-adult') perceived through his eyes another individual that has the features we attribute to mice, (i.e. 'small' 'rodent'). This set of conditions, which minimally guarantees that (5) is true, is the meaning of (5).

If, however, we replace the verb *saw* by the verb *killed*, as in (6), the sentence will have a different set of truth conditions, i.e. a different meaning: a boy ('human', 'male','non-adult') caused another individual, having the features that we attribute to mice (i.e. 'small', 'rodent'), to die.

(6) A boy killed a mouse.

Both (5) and (6) refer to different states of affairs, which can be true or false.

7.4 Sense relations between words

As we saw in (5) and (6), the contribution that a lexeme makes to the cognitive meaning of a sentence, i.e. the cognitive meaning (sense) of the lexeme, can be revealed if we replace a lexeme with another in a sentence and see whether the cognitive meaning of the sentence changes or not, and if it does change, how it changes. This activity involves a comparison of lexemes in terms of their contributions to the cognitive meaning of the sentence, i.e. in terms of their senses. The relationships between lexemes established on the basis of their senses are called **sense relations**.

- One of these is **synonymy**, which means that two or more lexemes have the same cognitive meaning (even though they may differ stylistically), e.g. *damsel* (formal, archaic), *girl* (neutral), *bird* (informal). Since *girl* and *damsel* are cognitively synonymous, the sentences under (1) and (2), repeated here for your convenience as (7a) and (7b), are also synonymous, even though stylistically different. (7a) and (7b) must both be true or both be false.

(7)a. The girl went to the garden.
b. The damsel made her way to the garden.

- Another sense relation is **ambiguity**, of which we distinguish two kinds, homonymy and polysemy. **Homonymy** (or perfect ambiguity) means that two or more phonologically and orthographically identical lexemes have completely different, unrelated meanings, e.g. $ball_1$ ('round object that you can throw or kick') vs. $ball_2$ ('social event at which you can dance'). **Polysemy** (or imperfect ambiguity) means that the meaning of one lexeme is metaphorically

extended on the basis of some similarity, cf. *leg* (of a man) vs. *leg* (of a table). It often happens that the metaphorical connection that once used to exist between such lexemes fades or is lost altogether and so what started out as a set of polysemous items becomes a set of homonymous items, cf. e.g. *horn*$_1$ ('the hard pointed part that grows on the head of cattle') vs. *horn*$_2$ ('kind of musical instrument played by blowing') vs. *horn*$_3$ ('apparatus in a car which makes a loud warning sound'). Originally, a *horn*$_2$ was made of a *horn*$_1$, and a *horn*$_3$ was a kind of *horn*$_2$, but many native speakers are no longer aware of this connection. The presence of an ambiguous lexeme in a sentence makes the sentence ambiguous, too. This is shown in (8a) and (8b).

(8)a. We waited by the bank. ('by the building of the financial institution').
b. We waited by the bank. ('by the riverside').

- Another sense relation is **oppositeness** or **antonymy**, with subtypes called complementary, gradable and relational opposites.[2] **Complementary opposites** are lexemes in such a relationship that the negation of the meaning of one lexeme gives us the meaning of the other, e.g. *dead* vs. *alive* (because 'not dead' means 'alive' and 'not-alive' means 'dead'). **Gradable opposites** are gradable lexemes, relative to some norm, e.g. *large* vs. *small*. (A small elephant is not a small animal, it is only small for an elephant, a large mouse is not a large animal, it is only large for a mouse.) More of one is less of the other, e.g. *smaller* means 'less large', *larger* means 'less small'. One member of gradable opposites is normally unmarked, the other is marked. It is the unmarked member that is used in questions of degree unless we have some good reason to use the other one; cf. *How old are you?* is unmarked, *How young are you?* is marked. **Relational opposites** are lexemes referring to symmetrically opposite aspects of the same situation, cf. e.g. *employer* vs. *employee*. (If Peter employs you, you are his employee and he is your employer.) Replacing a lexeme by its opposite in a sentence causes the original sentence and the new sentence to have opposite or **incompatible meanings**. This means that they cannot both be true at the same time.

(9)a. John is dead.
b. John is alive.

- The last sense relation we mention is **hyponymy** or **logical inclusion**. This is the relation between a cognitively superordinate, i.e. more

[2] The term *antonymy* can be used broadly, for all kinds of lexical oppositeness, but some scholars use it narrowly: only in connection with gradable opposites.

general, lexeme and the more specific lexemes that are cognitively subordinated to it. For instance, *tulip*, *rose*, *daisy*, *carnation*, *lily*, *daffodil*, etc. are all hyponyms in relation to *flower* (and co-hyponyms in relation to one another), because the sets of semantic features that they all have include the feature 'flower'.

There are semantic features which are present in the sense of a number of lexemes, e.g. 'female' or [–male] is present in the nouns, proper names, verbs and adjectives enumerated in (10).

(10)a. tigress, doe, ewe, hen, mare, vixen, cow, actress,
queen, girl, maiden, widow, nun, woman, sister,
mother
b. Agnes, Sue, Eve
c. to give birth, to breastfeed
d. pregnant, buxom, etc.

The noun phrase in (11a) and the sentence in (11b) are semantically anomalous because they attempt to reconcile incompatible semantic properties:

(11)a. !my brother called Sue
b. !My brother is pregnant.

If, in a sentence, we replace a hyponym lexeme with its superordinate lexeme, the original, first sentence is said to **entail** the new one. One sentence entails another sentence if the truth of the first guarantees the truth of the second, and the falsity of the second guarantees the falsity of the first. For instance, (12a) entails (12b). (Check this for yourself: if (12a) is true, is (12b) true, too? If (12b) is false, is (12a) false, too? If the answer to both questions is 'yes', then (12a) entails (12b).)

(12)a. Mary picked daisies.
b. Mary picked flowers.

7.5 The cognitive meaning of sentences

The cognitive meaning (or proposition) of a sentence depends on three factors.

- First it depends on the cognitive meanings of the sentential constituents, whether these are content words such as *man*, *dog* and *kick*, or function words such as *the*, or proper names such as *Mr. Brown*, or affix-morphemes such as the past tense suffix {-ed}, cf. (13a) vs.(13b).

(13)a. Mr. Brown kicked the man. ≠
b. Mr. Brown kicked the dog.

- Secondly, it depends on the functional labels that the constituents have, which is often, though not always, mirrored by the order of the constituents, cf. (14a) and (14b).

(14)a. [$_{subject}$ The sheriff] kicked [$_{object}$ the man]. ≠
b. [$_{subject}$ The man] kicked [$_{object}$ the sheriff].

- Thirdly, the cognitive meaning of a sentence depends on its structure. In (15a) the adverb *fast* modifies both verbs, in (15b) it modifies only the second verb.

(15)a. They ((run and swim) fast).
b. They (run and (swim fast)).

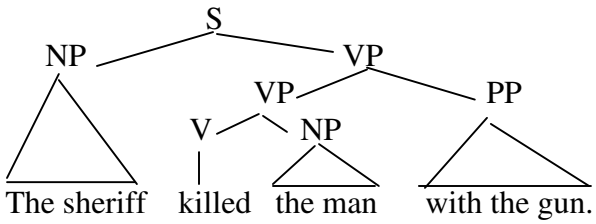
As we talked about sense relations between lexemes, so we can talk about **proposition relations** between sentences. These are relations between sentences on the basis of their cognitive meanings.

- Sentences can be **synonymous**, in this case they are each other's **paraphrases**. The synonymy of sentences may result from lexical synonymy, as we saw in (7a) and (7b) above. But sentential synonymy can also be achieved structurally, as in the active–passive pair in (16a) and (16b).

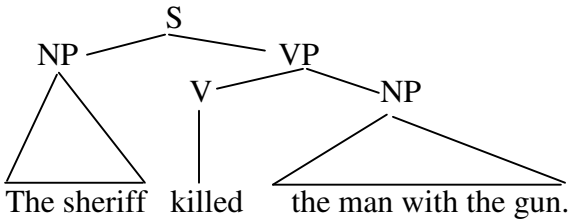
(16)a. The dog crossed the road.
b. The road was crossed by the dog.

- Sentences can also be **ambiguous**. This means that two sentences are composed of the same constituents in the same order but they have different meanings. This again can have lexical reasons (one of the words being ambiguous), as in (8a) and (8b) above. Or, ambiguity can be caused by structural differences, as in (17a) and (17b).

(17)a. The sheriff killed the man with the gun. ('The sheriff fired the gun at the man')



b. The sheriff killed the man with the gun. ('The sheriff fired at the man who had the gun.')



Furthermore, ambiguity may also result from the different functions a particular constituent can perform, see (18a) and (18b).

(18)a. [$_{\text{subject}}$ The lamb] is ready to eat. ('The lamb will eat.')
b. [$_{\text{object}}$ The lamb] is ready to eat. ('Somebody will eat the lamb.')

- The remaining two proposition relations, viz. **incompatibility** and **entailment** are not brought about by structural means or by different functions, they are solely due to the contribution of lexemes, as we saw in the example pairs of (9) and (12), respectively.

Exercises, problems, and other tasks

1. Compare the senses of *kill*, *murder*, *assassinate*.
2. For each group of words, state what semantic features they share and how the A words and the B words are related. (Use the + and − values of such features as 'animate', 'male', 'human', 'concrete', 'solid'.)
   (i) A. book, house, mountain, road, car
   B. peace, love, sincerity, fear, bravery
   (ii) A. lamp, pen, desk, house, ship, bike
   B. wine, alcohol, rice, jam, mud
   (iii) A. husband, man, son, monk, chief

B. bull, stallion, drake, ram, rooster.

3. Synonymous lexemes may differ in their non-cognitive (e.g. stylistic) meaning, cf. *toilet* and *loo*, or *shinbone* and *tibia*. Comment and give at least two similar examples.
4. Look up the meaning of the lexeme *bottleneck* and comment on it.
5. What is the difference between homonyms, homophones and homographs? Illustrate them. (You may consider Exercise 4 in Unit 5.)
6. Comment on the Hungarian lexeme *toll* ('pen').
7. Identify the following opposites: *good – bad*, *expensive – cheap, false – true*, *parent – offspring*, *beautiful – ugly*, *lessor – lessee*, *pass – fail, hot – cold*, *legal – illegal*, *poor – rich*, *fast – slow*, *asleep – awake, husband – wife*, *before – after*
8. What are *suffix* and *prefix* the hyponyms of?
9. Comment on the proposition relation between *Yesterday it rained*. and *It rained yesterday*.
10. The following sentence is both structurally and lexically ambiguous: *I saw him walking by the bank.* Paraphrase all its meanings.
11. Turn the labelled tree diagrams in (17a) and (17b) into labelled bracketings.
12. What relation is there between the A sentences and the B sentences:
    i) A. *The police wounded the burglar*. B. *The burglar is injured*.
    ii) A. *The house is red*. B. *The house is not white*.
13. Describe the oddness of the following sentences, using semantic features: [!]*The television drank my water*. [!]*Colourless green ideas sleep furiously*. (N. Chomsky)
14. Make sure you know the following terms: cognitive meaning, proposition(al meaning), stylistic meaning, speech act meaning, emotive meaning (affective meaning), referential theory of word meaning, conceptual theory of word meaning, semantic features, componential analysis, truth conditions, sense relations, synonymy, ambiguity, homonymy, polysemy, oppositeness, complementary opposites, gradable opposites, relational opposites, hyponymy, incompatibility, entailment, paraphrase

Unit 8
Pragmatics, the Study of Language Use in Particular Situations

8.1 Pragmatics

Pragmatics is the study of various aspects of language use; it deals with the ways in which language-users use and interpret words and utterances in particular situations. By words and utterances we mean lexemes and sentences used in particular situations, and by situations we mean linguistic and physical contexts.

Pragmatics is not easy to separate from semantics and it is to some extent an arbitrary decision where we draw the line between them. The central concerns of the two, however, stand out fairly clearly. While semantics primarily examines the cognitive meaning of lexemes and sentences, pragmatics primarily examines what the speaker means by the lexemes (words) and sentences (utterances) used in particular situations; i.e. it is a study of intended "speaker meaning".

We shall deal with the following aspects of language use: (a) the role of context and presuppositions, (b) language functions and speech acts, (c) conversational implicatures.

8.2 The role of context and presuppositions

Ambiguous words and utterances are usually disambiguated by means of the **linguistic context**. In a narrow sense, the linguistic context is provided by the environment of the ambiguous word within the utterance, i.e. by the other words around the ambiguous word. For instance, although the lexemes $BANK_1$ ('riverside') and $BANK_2$ ('financial institution') are homonyms (see Unit 6), they are not normally confused when they occur in particular linguistic contexts, as in (1a) and (1b):

(1)a. The right bank of the River Danube in Budapest is nice and hilly.
b. The bank has announced an increase in interest rates.

In a broader sense, the linguistic context comprises the other utterances around the ambiguous utterance within a discourse. By **discourse** we mean the physical product of language use in a particular situation; it consists of all the utterances made in the same situation. For example, although the sentence *The sheriff killed the man with the gun.* is ambiguous in itself (see 7.5), it gets disambiguated when uttered in different discourses, see (2a) and (2b).

(2)a. There were two people waiting for him round the corner: a man with a gun and a woman with a knife. The sheriff killed the man with the gun.
b. John gave the sheriff the gun the man had dropped. The sheriff killed the man with the gun.

Another disambiguating factor is the **physical context**. For instance, when you see the word BANK written on the front of an elegant building in a city, you will know that what you see is not the edge of a river but an institution dealing with money matters. There are utterances containing pronouns and pronoun-like place- and time-adverbials which refer to the personal, locational and temporal characteristics of the situation and whose meaning is relative to the situation, because they can be interpreted only if the speaker's immediate physical context is known. Such pronouns and pronoun-like elements are called **deictic expressions**. For instance, the underlined elements in (3) are deictic. The phenomenon of using deictic elements is known as **deixis**.

(3) I'll have to do that next week because they're not here now.

It can happen that a positive sentence and its negative counterpart both presuppose that a particular state of affairs (proposition) is true and known not only by the speaker but also by the hearer. This proposition is called a **presupposition.** A presupposition is a proposition that follows from both a positive sentence and its negative counterpart, and which both the speaker and the hearer assume to be true. For example, the presupposition of (4a) and (4b) is (5).

(4)a. Your brother wants to see you.
b. Your brother doesn't want to see you.

(5) You have a brother.

8.3 Language functions and speech acts

One way of dealing with language use is in terms of **language functions**. It is customary to distinguish six types of language function.

- The **cognitive function** (= propositional or descriptive function): this is the communication of a state of affairs, e.g. *Today is Monday*. or *The table is in the middle of the room*.
- The **expressive function** (= affective function): this is the expression of the speaker's attitudes, feelings, emotions, e.g. *Damn!* or *Shit!* or *Oh!*
- The **directive function.** this is influencing the hearer's behaviour or attitude, e.g. *Come here!* or *Could you lend me two thousand dollars until Friday?*
- The **phatic function**: this is establishing and maintaining contact with the hearer, e.g. *Hi there*., *Nice to see y*ou. or *Can you follow me?*
- The **metalinguistic function**: talking about language in order to clarify certain aspects of it, sometimes to ensure that communication can take place undisturbed, e.g. *The word "violin" is of Italian origin*.
- The **poetic function** (= aesthetic function): this is the use of language primarily for its own sake, i.e. for the pleasure it gives speaker and hearer through its sound and rhythm, rather than for performing any of the other functions, e.g. *Pat a cake, pat a cake, baker's man*.

A more recent classification of various types of language use has been provided by **speech act theory**, first propagated by the language philosophers Austin and Searle. The central notion in this theory is **illocutionary act**, which is the act the speaker performs in and while saying an utterance.[1] An illocutionary act realises the speaker's communicative intention, which can be of hundreds of different kinds, e.g. asserting, stating, reporting, complaining, promising, inquiring, warning, suggesting, ordering, requesting, thanking, greeting, etc. Illocutionary acts can be made fully explicit if we use **performative verbs**. A performative verb is so called because it explicitly performs an illocutionary act; i.e. it explicitly expresses the speaker's communicative intention, and as such it stands in the first person singular and can be prefixed by the adverb *hereby*. In some situations the use of the performative verb is obligatory. Consider the underlined verbs in (6), (7) and (8), where the verbs *name*, *declare* and Hungarian *üdvözölni* ('to greet'), all in first person singular, present tense, indicative, carry out the very acts of naming, declaring and greeting.

(6) I <u>name</u> this ship the Rainbow.

[1] The word *illocution* derives from "in+locution", i.e. 'in speaking'.

(7) I declare the meeting open.
(8) Üdvözlöm.

Of course, the performative verb need not be there in most cases, because the speaker's communicative intention can be obvious in the situation without making it explicit. Consider the sentences in (9), (10) and (11), in which the use of the performative verb is not obligatory. The verbs underlined in the (a) sentences are performative verbs and they explicitly carry out the illocutionary acts of asserting, suggesting and promising. But instead of the (a) sentences we may actually utter the (b) sentences; with these we can perform the same illocutionary acts as with their (a) counterparts.

(9)a. I (hereby) assert that he speaks excellent English.
b. He speaks excellent English.

(10)a. I (hereby) suggest that you should leave.
b. You should leave.

(11)a. I (hereby) promise that I'll be there.
b. I'll be there.

The identification of language functions or illocutionary acts depends on various factors and is not always easy, for several reasons.

- First, functions or speech acts may overlap in an utterance. Take, for instance, the sentence *The door's too low*. This can simultaneously be a report, an assertion, a warning, and a complaint when you say it after you've hit your head against the door-bar and want to warn other people to mind their heads; i.e. it can simultaneously have a cognitive, expressive, directive and phatic function.
- Secondly, functions and illocutionary acts are not consistently matched by sentence forms. The same grammatical form can be used in a wide variety of different functions or speech acts. The imperative, for example, can be used in giving advice, suggestions, commands, demands, prayers, requests, etc. Conversely, the same functional category or speech act can be realised by a wide variety of different forms. For instance, the sentences in (12) can all be interpreted as expressions of the same directive function or persuasive act.

(12)a. I'd leave if I were you.
b.You ought to leave.
c. You'd better leave.
d. I hope you leave.

e. I want you to leave.
f. Why don't you leave?
g. When are you leaving?
h. It's time you left.
i. Please, leave.

Out of these (12i), *Please, leave*., can be called a **direct speech act** because it is realised by the most obvious linguistic means, the imperative. The others are **indirect speech acts** because they use syntactic structures that are more usually associated with other acts.

- Thirdly, the interpretation of the function or illocutionary act represented by an utterance requires knowledge of the situation (physical and linguistic context) in which the utterance is made. Consider, for example, the sentence *Can you play the piano?* This counts as an ordinary question if you say it to someone in a room where there is no piano and the conversation is about who can play what musical instrument. But it will count as a request to play if it is said to someone who is known to be a good pianist, in a room where there is a piano, cf. Unit 2. In other words, we simply cannot say what the function or illocution of a sentence is if we take it in isolation from the context or situation in which it is uttered.

8.4 Conversational implicatures, Grice's maxims

When we establish the "speaker meaning" of a sentence uttered in a particular situation, we rely, among other things, on **conversational implicatures**. These are implications following from the utterance on the basis of **Grice's maxims** (named so after Grice, the philosopher who invented them). Two of Grice's maxims are (13a) and (13b):

(13)a. Grice's maxim 1: Make your contribution as informative as is required but not more informative than is required.
b. Grice's maxim 2: Be relevant.

Consider, for instance, the response by Speaker B in (14).

(14) Speaker A: Have you cleaned your room and done the shopping?
Speaker B: I have cleaned my room.

On the basis of maxim (13a), we may rightfully assume that the implicature is: *I have not done the shopping*, i.e. this is how we can interpret Speaker B's

utterance in the given context. This interpretation, however, is not necessarily correct and can be cancelled, as is shown in (15).

(15) Speaker A: Have you cleaned your room and done the shopping?
Speaker B: I have cleaned my room.
Speaker A: So you haven't done the shopping.
Speaker B: Oh, yes, I've done the shopping, too.

The last example is the exchange in (16).

(16) Speaker A: The doorbell rang.
Speaker B: I'm in the bath.

In the given situation the first utterance, by Speaker A, can only be understood as a request towards Speaker B to go and answer the door. However, Speaker B is in the bath, and his response, on the basis of maxim (13b), can only be interpreted as *I can't go, so please go yourself.*

Exercises, problems, and other tasks

1. What are the deictic expressions in the following statement? *I am busy now, so you can't do that here.*
2. What is one obvious presupposition of a speaker who says:
   a. *Your watch is broken*. b. *The King of France is bald*. c. *John regrets that he broke your window*. d. *The pregnant teacher went on holiday*.
3. In what functions can you use the sentence *It is cold this morning*.?
4. What illocutionary acts can you perform in saying the sentence in 3?
5. Comment on the function of the Hungarian nursery rhyme: *Egyedem, begyedem, tengertánc.*
6. Which utterance does not contain a performative verb?
   a. *I deny your charge*. b. *I bequeath you my gold watch*. c. *I sentence you to three years in prison*. d. *I know you*. e. *I resign* f. *I apologise*.
7. Why can we not interpret the following utterances as promises?
   a. *I've brought you a box of chocolates*. b. *I'll kill you*.
8. Mention two of Grice's maxims.
9. Discuss possible interpretations of
   a. *You've left the door open*. b. *I can hear someone laughing*. (uttered by an invigilating teacher in an exam room where students are supposed to be working silently). c. *My mouth is parched*. (said to a barman) d. *My mouth is parched*. (said to a doctor).

10. Explain the abnormality of the following exchange:
    Speaker A: *Excuse me, I'm looking for the English Department*.
    Speaker B: *That's very nice of you, go ahead*.
11. Make sure you know the following terms: linguistic context, physical context, discourse, deixis, deictic expressions, presupposition, language functions, cognitive function (propositional function, descriptive function), expressive function (affective function), directive function, phatic function, metalinguistic function, poetic function (aesthetic function), illocutionary act, performative verb, direct speech act, indirect speech act, conversational implicatures.

Unit 9
Language Variation

9.1 The identity and variability of language

A natural language is not just one homogeneous code. Any natural language in the world exists in several varieties at the same time. All these varieties have their own sets of rules: all of them are codes for those sections of the community that use them. The existence of language varieties side by side is called **language variation** (= synchronic variability). Moreover, the coexisting varieties are in a constant change along the dimension of time, too, this phenomenon is called **language change** (= diachronic variability). (For the terms *synchronic* and *diachronic* see Units 2 and 3.)

The first question we have to discuss briefly is the problem of **language identity**, viz. what makes us decide whether two linguistic codes are two separate languages or just varieties of one language? One may say the criterion is mutual understandability, but this often breaks down between codes that are regarded as belonging to the same language (e.g. northern Chinese speakers and southern Chinese speakers do not necessarily understand each other's speech), moreover, it can bring together codes that are regarded as separate languages (e.g. Swedes and Danes often understand each other's speech fairly well). Therefore we have to admit that mutual understandability is not a safe criterion. Language identity is a socio-psychological concept, one language is the sum of all the varieties that their users are culturally and politically conditioned to regard as one and the same language. (There is, thus, a certain amount of truth in the humorous saying: "a language is a dialect with an army and a navy".) So English, like any other natural language, is an abstraction, it is a cover term for all the linguistic codes that are, or have been, or will be, regarded as English. It refers to a bundle of partly similar, partly different codes: "Englishes".

In this unit we deal with language variation, i.e. the simultaneously existing varieties of English. Since these varieties constitute a particular aspect of the relations between language and society, you can regard this chapter as a preliminary introduction to **sociolinguistics**, as well. Language variation can be discussed in terms of user-related and use-related variation.

9.2 User-related variation: dialect, sociolect, pidgin, creole, child language, gender differences

The most obvious user-related language varieties involve the user's geographical and social position.

- That variety of a language which is used in a certain geographical area is called **regional dialect** or just **dialect**, for short. Dialects may differ in vocabulary, pronunciation and even morphology and syntax. The boundaries between dialects are not as clearcut as political boundaries or topographical features. They can be established by collecting linguistic features characteristic of the area. The line marking the limit of the distribution of a linguistic feature on a map is called an isogloss. For instance, in a particular area within the state of Pennsylvania (USA), the local word for 'drought' is *drooth*. The line drawn around this area on the map is an isogloss. Other language features observed in this area may have slightly different geographical distributions, so the isoglosses based on these other features will not necessarily perfectly coincide with the isogloss for *drooth* but there will be considerable overlap between them. A dialect is a more or less congruent bundle of isoglosses.
- It often happens that one of the regional varieties gains social-political priority over the others and becomes the **standard variety** (or prestige variety), which is used for education, scholarship and state administration all over the country. The standard variety is no longer restricted to the geographical area where it was originally used but is associated with people who are educated, who are at the top of the socio-cultural scale, no matter where they live. The standard is no longer a regional dialect, it is rather a **social dialect**, or **sociolect**. A sociolect is a variety of language used by people in the same socio-cultural position.

It is important to emphasise that the standard variety has a higher social prestige, but is not linguistically better than the other varieties. For instance, **Standard English** was originally a regional dialect used in the South-East of England and its emergence as the standard was accidental from a linguistic point of view. The fact that it was this particular variety rather than a northern variety that became the standard is due to historical, political, cultural, economic reasons (think of the significance, in this respect, of the capital city and the great universities in the region). Standard English has two major national subvarieties, Standard British and Standard American, neither of which is linguistically superior to the other. The two display remarkable uniformity, the greatest difference between them is probably in pronunciation. The ideal type of pronunciation of Standard British English is called **Received Pronunciation**, or RP (so called because by the 19th century this had become

the only socially acceptable pronunciation in polite society in England, notably the pronunciation of those people who were received at court). The pronunciation associated with Standard American English is called **General American**, or GA.[1]

Standard British English, with its RP, is the language of the educated people at the top of the socio-cultural scale in Britain. The lower you go along this scale, the more you find that people mix the standard with dialectal (regional, local) features on the one hand, and with sociolectal (non-regional) features that generally characterise the language of less educated people on the other hand. Those near the bottom of the socio-cultural scale nearly always use **non-standard** varieties, which may coincide with regional dialects but may also cut across dialect boundaries. Here are a few examples: *He want it*., *I wants it*., *That was the man has done it*., *He don't know nothing*., *I ain't got no car*., etc. (The last two examples illustrate **double or multiple negation**, a non-standard sociolectal feature used by uneducated English speakers in very different geographical areas.) One must not think, however, that examples of this sort are incorrect. They simply belong to other codes than the standard. They are perfectly well-formed within the varieties to which they belong and obey the rules of those varieties. (This is why the derogatory label *substandard* is unjustified and should be avoided. We recommend the use of *non-standard*, instead.)

- A third type of user-related language variation is **pidgin**. A pidgin is usually the simplified version of a European language, containing features of one or more local languages, used for occasional communication between people with no common language, in West Africa or in the Far East. For example, Melanesian Pidgin English (called *Tok Pisin*) is used in Australian New Guinea and the nearby islands. While a pidgin is not a native language, it can become the native language of a community (e.g. through intermarriage between people who have been brought together on a plantation from different linguistic backgrounds, and who have the pidgin as the only common language they can use for communication with one another). When a pidgin becomes the native language of a community, it is called a **creole**. For instance, in Jamaica, in addition to Standard English, there exist several kinds of Creole English.

Finally, one could add to the list of user-related varieties the linguistic features that are attributable to the age and sex of the language user. Apart from the features of **child language**, however, such features are not sufficiently systematic to form clearly identifiable varieties. For instance, although one can

[1] Since the way in which a language is pronounced is called *accent*, RP and GA are also accents. Needless to say, there are dialectal and foreign accents as well, so one can speak English with a Yorkshire accent, or with a German accent, or with an RP accent, etc.

spot a few features that tend to occur more often in the language of female speakers than in the language of male speakers (and vice versa), it would be unjustified to separate feminine and masculine varieties of English.

9.3 Use-related variation: spoken and written varieties, styles, registers

There are different types of use-related language variation.

- The first type of use-related variation is conditioned by the **medium** of language use, i.e. by speech and writing. The language we speak is generally different from the language we write. When we write, we are often more careful and use longer sentences because the addressee is not present and so cannot rely on the situation (physical context), but can always go back to the beginning of the sentence and read it again if necessary. But it seems that a finer distinction of media is required because there are different subtypes of speech and different subtypes of writing, and these differences trigger corresponding linguistic differences. For instance, the language we use in face-to-face talks tends to differ from the language of public lectures, which in turn is very different from the language of telephone conversations. Or, the language of text-messages on your mobile phone is clearly different from the language of your personal letters, though both are written varieties.
- The second type of use-related variation is **style**. This is conditioned by the language users' relative social status and attitude towards their interlocutors (e.g. they can talk to equals, to people in higher or lower social positions, to older or younger people, to children, they may talk to someone who they have never seen before or to someone who is an old friend of theirs, etc.) We recognise a **neutral** or unmarked style, which does not show any obvious colouring brought about by relative social status and attitude. On either side of this we can distinguish sentences which are markedly formal or informal. Compare the sentences in (1).

(1)a. Formal: I wonder if you'd mind switching off the light.
b. Neutral: Would you please switch off the light?
c. Informal: Switch off the light, will you?

Formal style is usually impersonal and polite, used in public speeches, serious polite talk, serious writing (official reports, regulations, legal and scientific texts, business letters, etc.). A very formal style can be called **rigid**, it is nearly always written and standard. **Informal** (= colloquial) style characterises private conversations, personal letters between intimates and popular newspapers. A

very informal style can be called **familiar**, this may involve the use of non-standard features, four-letter words, and slang expressions. **Slang** can be defined as very informal language, with a vocabulary composed typically of coinages and arbitrarily changed words, such as the ones often created by young speakers. Some slang expressions are associated with particular groups of people, so we can distinguish e.g. army slang, school slang, etc., to this extent slang is partly user-related. After a time, some slang expressions die out or become old-fashioned, e.g. *to take a shufti at something* ('to take a look at something'), but some may pass into ordinary colloquialism (i.e. informal standard), e.g. *to slag someone off* ('to criticise someone') is a British slang expression half-way towards becoming a standard vocabulary item.

When we use language, we must use sentences that are not only grammatical and meaningful but also **stylistically appropriate**, i.e. matching the stylistic requirements of the situation. For instance, the sentence *Be seated.* is perfectly grammatical and meaningful, but would be ridiculously inappropriate if we said it to a friend of ours in our home (unless we wanted to sound humorous).

- The third type of use-related language variation is **register**, which is conditioned by the subject matter in connection with which the language is being used. Each field of interest, activity, occupation is associated with a special vocabulary, and it is mainly these vocabulary differences that underlie the different registers. Thus we can talk about the registers of sports, religion, medicine, computer engineering, cookery, weather forecasts, etc. Think, for example, of the word *shotputting*, which is hardly ever used outside the sports register, or the word *blackboard*, which is only used in the register of school teaching. When the register of a field is full of technical terms which those who have received no training in that field cannot understand, it is referred to as **jargon** (think e.g. of the jargon of computer engineers or the jargon of linguists). Criminals' jargon can be called **argot** or **cant**. Since the most frequent and most favourite topics of one's speech or writing are related to one's occupation, registers are partly user-related, too.

9.4 Idiolect, code switching, diglossia

The total of all the varieties of a language that a person knows is the person's idiolect. An **idiolect**, then, is the amount of a language that an individual possesses. The ability to change from one variant to another is **code switching**. For instance, a doctor switches codes when he speaks of a bone as *tibia* to his colleagues in the hospital and as *shinbone* to his family at home.

It can happen that two distinct varieties of a language co-occur in a speech community, one with a high social prestige (such as e.g. Standard English, learnt at school, used in church, on radio programmes, in serious literature, and generally on formal occasions), and one with a low social prestige (e.g. a local dialect, used in family conversations and other informal situations). The sociolinguistic term for this situation is **diglossia**, and an individual having diglossia is a **diglossic**. (These terms are not to be confused with **bilingualism** and **bilingual**, which mean 'knowledge of two languages' and 'person knowing two languages', respectively.)

Exercises, problems, and other tasks

1. Are the boundaries between regional dialects as clear and sharp as political boundaries? What is an isogloss? How can you define a dialect in terms of isoglosses?
2. Discuss the standard variety. How does it emerge? Is it more correct than the other (non-standard) varieties? Is it more beautiful? Is it purer? Is it more versatile and flexible than the other varieties? Why?
3. The form *ain't* is extremely common in non-standard varieties. Find out how it is used, on the basis of the following data: *It ain't coming*. *We ain't going*. *It ain't there*. *I ain't done it*. *He ain't got one.* BUT: *I haven't a clue*. and not **I ain't a clue*. What does this show?
4. In American Black English *Where ya been lately?* replaces the standard *Where've you been lately?* but **Where've been lately?*, **Where been lately?* and **Where ya lately?* are impossible. What does this show? Are there no rules in non-standard varieties?
5. Find differences between British and American English lexis, phonology, morphology, syntax, and orthography.
6. Why do we teach British English in most schools in Hungary?
7. Explain RP, GA, and the term *accent*. Does the Queen speak with an accent?
8. In Tok Pisin, the phrase *ka bilong yu* means 'your car'. What do you think *ka bilong mi* means?
9. Try to find differences between male speech and female speech in English or in your native language.
10. Identify the styles of these:
    (i) a. *Peter went home early because he felt tired*. b. *Feeling tired, Peter went home early*. c. *Pete felt tired, so he went home early*.

(ii) a. *When his dad died, Eddie had to get another job*. b. *On the decease of his father, Mr Grey was obliged to seek alternative employment*. c. *After his father's death, Edward had to change his job*.

11. On the basis of the data in Exercise 10, try to identify some characteristic features of informal and formal style in English.
12. What is "wrong" with *Hello, old chap!*, said to the dean by a student?
13. Identify the varieties in each of the following:

    a. *To be taken three times a day, after meals*. b. *Dear Sir,…* c. *Three Die in Hotel Blaze* d. *Ladies and Gentlemen!* e. *He don't know nothing*. f. *Aphasia results from cortical lesion*. g. *No daddy come*. h. *The price of gas is terrible, isn't it Eddie?* i. *Be seated*. j. *They chucked a stone at the cops, and then did a bunk with the loot*.
14. Scientific reports, legal documents, printed essays in English do not usually contain contracted forms (e.g. *I'll*, *you're*, *isn't*, etc.). Why not?
15. Make sure you know the following terms: dialect (regional dialect), socioloect (social dialect), standard variety, Standard English, Received Pronunciation (RP), General American (GA), non-standard varieties, double or multiple negation, pidgin, creole, medium-conditioned varieties, style, neutral style, formal and rigid styles, informal and familiar styles, slang, stylistic appropriateness, register, jargon, argot (cant), idiolect, code switching, diglossia, bilingualism.

Unit 10
Language Change

10.1 Periods in the history of English

All languages are in a constant process of change along the dimension of time. This phenomenon is called **language change** or the **diachronic variability** of language. English has undergone considerable changes in the three main periods of its history. These periods are Old English (OE), roughly from 450 to 1100, Middle English (ME), from about 1100 to about 1500, and Modern English (ModE), from roughly 1500 to the present.

The **Old English** period started when three Germanic tribes coming from the Continent, viz. the Angles, the Saxons and the Jutes, settled down in what is known today as England, in the 5th century. They spoke Germanic dialects, from which the various OE dialects developed directly.[1] Today OE texts are largely unintelligible even to the English. Consider the example in (1), which is taken from a 10th century document (Aelfric's homily on St. Gregory):

(1) *þā sǣde him man þæt hī of Engla-lande wǣron and þæt*
then said him someone that they of England were and that

ðǣre ðēode mennisc swā wlitig wǣre[2]
that country's people so comely were

'And then someone told him that they were from England and that the people of that country were so comely.'

The OE example differs from its ModE counterpart in many respects. For instance, it contains a short [a] and a long [ā], as well as diphthongs that later disappeared, e.g. [ēo]. Words like *ðēod* ('country') and *wlitig* ('comely, handsome, beautiful') have disappeared from use. The word *hī* means 'they'. There are considerable syntactic differences, too: for example in the first clause the verb precedes the indirect object and the indirect object precedes the

[1] The main OE dialects were Northumbrian, Mercian, West Saxon and Kentish.
[2] The letter *þ* (called 'thorn') and the letter *ð* (called 'barred d') were both pronounced as the /ð/ in intervocalic position and as /θ/ elsewhere.

subject. Morphology is also quite different. The suffix *-on* in *wǣron* indicates that the verb is in the past tense and plural. In OE there was an elaborate inflection system for both verbs and nominal phrases.

In 1066, with the Norman Conquest, a new era began, which is referred to as **Middle English**. In this period the English language changed more radically than in any other period of its history. Vast numbers of French words entered English and by the end of the period English had lost most of its inflections, and the quality of many of its original sounds had changed considerably. From the main dialects of ME[3] eventually a Mid-South-Eastern dialect (around London) emerged as the dominant one and this provided the basis for what later became Standard English. The following extract, (2), is not too difficult to understand for present-day readers. It is from Chaucer's *The Parson's Tale* (end of the 14th c.). Chaucer wrote in the London dialect.

(2) *Of the roote of contricion spryngeth a stalke that bereth braunches and*
of the root of contrition springs a stalk that bears branches and

leves of confessioun, and fruyt of satisfaccioun.
leaves of confession and fruit of satisfaction

'From the root of contrition springs a stalk, leaves of confession and fruits of satisfaction.'

By about 1500, English had essentially become a language which, though looking archaic to us, can be recognised as English and understood without much difficulty even today. This is the beginning of the **Modern English** period. The following illustration is from Shakespeare's *Hamlet, Act V, Scene I* (early 17th c.)

(3) – *… why was he sent into England?*
– *Why, because he was mad: he shall recover his wits there; or, if he do not, 'tis no great matter there.*
– *Why?*
–*'Twill not be seen in him there; there the men are as mad as he.*
– *How came he mad?*
– *Very strangely, they say.*

[3] The main dialects of ME were Northern, West-Midland, East-Midland, Southern and Kentish.

10.2 Examples of changes

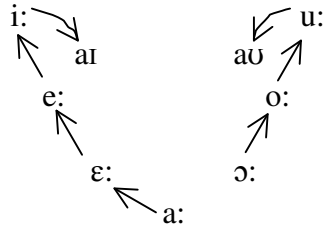
An examination of the changes that have occurred in English during the past 1500 years shows that they have affected all parts of the language. We will illustrate this claim with a few examples.

- Let us start with some **phonological changes**. These are sound changes that directly affect a language's phonological system. For instance, the sound [ŋ] was originally an allophone of /n/, which appeared before /k, g/ in English, just as it does in Hungarian words like *munka* and *inga* today. During Middle English, /g/ was lost in word-final position after a nasal consonant, leaving [ŋ] as the final consonant in words like *sing*. The loss of the final /g/ in words created minimal pairs such as *sin* and *sing*, i.e. the phoneme /n/ split into two phonemes: /n/ and /ŋ/.

Sound change was also a common type of phonological change. By the end of the 14th century the old /ɑː/ sound was replaced by /ɔː/, and words such as e.g. *stān*, *bān* and *gāst* became *stōne* ('stone'), *bōne* ('bone') and *gōst* ('ghost'), respectively.

A major phonological change in the history of English took place approximately between 1400 and 1600. It is known as the **Great Vowel Shift**. The seven long vowels of Middle English underwent the following change. The highest vowels, /iː/ and /uː/, became the diphthongs /aɪ/ and /aʊ/, respectively. In addition to this, each of these vowels was replaced by the next higher vowel. All this is schematically shown in (4), with some examples in (5).

(4) The Great Vowel Shift (GVS) (1400-1600)



(5)

ME	GVS	Later diphthongisation	ModE
/tiːd/	/iː/ → /aɪ/		/taɪd/
/huːs/	/uː/ → /aʊ/		/haʊs/
/geːs/	/eː/ → /iː/		/giːs/
/brɛːkən/	/ɛː/→ /eː/	→ /eɪ/	/breɪk/
/goːs/	/oː/ → /uː/		/guːs/
/brɔːkən/	/ɔː/→ /oː/	→ /əʊ/	/brəʊkən/
/naːme/	/aː/ → /ɛː/	→ /eɪ/	/neɪm/

- There have been considerable **morphological changes**, too. In Old English, there was a rich conjugation system for verbs: the verbs had different endings depending on person, number and tense. Nouns were divided into three gender classes: masculine, feminine and neuter, and each gender class was associated with a different set of case endings in both singular and plural, and there was concord between nouns and their adjectives, too. This is shown in (6), illustrating the declension of *sē gōda wind* ('the good wind').

(6)

	Singular	Plural
Nom.	sē gōda wind	þā gōdan windas
Acc.	þone gōdan wind	þā gōdan windas
Gen.	þæs gōdan windes	þāra gōdra winda
Dat.	þǣm gōdan winde	þǣm gōdum windum

This elaborate system of declension had disappeared by the end of the Middle English period and with the overwhelming majority of nouns it was only the non-genitive singular form and the form with the suffix *-s* (standing for non-genitive plural, genitive singular and genitive plural) that remained distinguished.

- As for **syntactic changes**, there are statistics which show that in 1200 the direct object (O) was put before the verb (V) in 53 % of all cases and after the verb in 47 %. By about 1500 this had changed completely: the direct object was put before the verb in only 2 % of the cases and after the verb in 98 %. The verb-object word order had become dominant.

(7)

	1200	1500
OV	53 %	2 %
VO	47 %	98 %

While in Old and Middle English the inversion involved in the formation of questions could apply to all verbs, in the Modern English period the inversion rule was gradually changed to apply solely to auxiliary verbs. In the *King James Version* of the Bible (1611) we still find both main verbs and auxiliary verbs inverted with the subject, as shown in (8). Later this possibility ceased to exist for main verbs and *do*-insertion became the established way of asking questions when no auxiliary was present.

(8)a. Answerest thou nothing? (Mark 14)
b. How much owest thou onto my lord? (Luke 16)
c. See ye not all these things? (Matthew 24)
d. What think ye? (Matthew 26)
e. Tell us, when shall these things be? (Matthew 24)
f. What will ye give me…? (Matthew 26)

In addition to phonological, morphological and syntactic changes, **lexical changes** have also taken place in English over the past 1500 years. English has borrowed a large number of lexemes from other languages, especially from French, which was in large part the consequence of the Norman Conquest. (Lexical items borrowed from other languages are called **loanwords**.) The French-speaking Normans who conquered England in 1066 and their descendants gradually learnt English over the next decades, but they continued to use French words to refer to political, judicial and cultural notions. These words were in turn borrowed by native English speakers. The list under (9) shows you some of the French loanwords in English and the fields they belong to.

(9) Government: *tax*, *revenue*, *government*, *royal*, *state*, *parliament*, *authority*, *prince*, *duke*, *slave*, *peasant*, *nation*, *crown*, *society*
Religion: *prayer*, *sermon*, *religion*, *chaplain*, *friar*, *saint*, *charity*
Law: *judge*, *defendant*, *jury*, *evidence*, *jail*, *verdict*, *crime*, *attorney*, *court*
Medicine: *medicine*, *physician*
Culture: *art*, *sculpture*, *fashion*, *satin*, *fur*, *ruby*
Warfare: *army*, *navy*, *battle*, *soldier*, *enemy*, *captain*

Today, if you examine the 5000 most frequent words in English, you will find that 40 % of them are of English (Germanic) origin, 39 % of French origin, 12 % of Latin origin and 9 % of other. But if you concentrate on the first 1000 most frequent words, you will find that the ratio of words of English origin is considerably higher:

(10)

most frequent words in English	source language			
	E	Fr	Lat	Other
5000	40%	39%	12%	9%
1000	83%	11%	2%	4%

Words have often changed their meaning, too. These changes involve the following processes: **semantic broadening** (the meaning of a word becomes more general than its earlier meaning), **semantic narrowing** (the meaning of a word becomes less general than its earlier meaning), and **semantic shift** (the word loses its earlier meaning and acquires a new one). These processes are illustrated in (11).

(11)

Process	Word	Old meaning	New meaning
Semantic Broadening	BIRD	'small, domesticated winged creature'	'any winged creature'
	DOG	'a hunting canine'	'any canine'
Semantic Narrowing	FOWL	'any winged creature'	'a small, domesticated winged creature'
	HOUND	'any canine'	'a hunting canine'
Semantic Shift	SILLY	'happy, blessed'	'foolish'
	IMMORAL	'unusual'	'unethical'

Exercises, problems, and other tasks

1. Which historical dates mark the beginning and the end of the Old English and Middle English period?
2. Analyse the word order in the following line (from an 8th c. translation of the Venerable Bede's Latin history of England): A*nd Seaxan þā sige geslōgan* (and Saxons the victory won) = 'And Saxons won the victory.'
3. What were the main linguistic tendencies in the period of Middle English?
4. Give an outline of what happened in the course of The Great Vowel Shift.
5. Questions like *What think ye?* were possible until the end of the Early Modern English period. What do these show?
6. What words of French origin are used to refer to the meat of these animals: *pig*, *cow*, *calf*, *sheep*.
7. Try to identify the source of the following lexical borrowings into English: *comrade*, *motto*, *poodle*, *tornado, czar*, *smuggle*, *gin*, *toboggan*, *bagel*, *kindergarten*, *balcony*, *banana*, *coach*, *wigwam*, *mosquito*, *casino*, *sauerkraut*. Use a dictionary that contains etymological information (e.g. The Concise Oxford Dictionary or Webster's Ninth New Collegiate Dictionary).
8. The words *aunt* and *mete* (*meat*) used to mean 'father's sister' and 'food', respectively. What kinds of semantic changes do they exemplify?

9. Make sure you know the following terms: Old English, Middle English, Modern English, semantic broadening, semantic narrowing, semantic shift.

Appendix A: IPA symbols for the phonemes of Standard British English (RP)

Consonants	Vowels
/p/ *p*en	/æ/ b*a*d
/b/ *b*ack	/ɑː/ c*al*m
/t/ *t*ea	/e/ b*e*d
/d/ *d*ay	/ɪ/ sh*i*p
/k/ *k*ey	/iː/ sh*ee*p
/g/ *g*et	/ɒ/ p*o*t
/m/ *m*ouse	/ɔː/ c*augh*t
/n/ *n*ice	/ʊ/ p*u*t
/ŋ/ si*ng*	/uː/ b*oo*t
/l/ *l*ed	/ʌ/ c*u*t
/r/ *r*ed	/ɜː/ b*ir*d
/f/ *f*at	/ə/ *a*go
/v/ *v*iew	/eɪ/ m*a*ke
/θ/ *th*ing	/aɪ/ b*i*te
/ð/ *th*en	/ɔɪ/ b*oy*
/s/ *s*un	/ɑʊ/ = /aʊ/ n*ow*
/z/ *z*ero	/əʊ/ = /oʊ/ n*o*te
/ʃ/ *sh*ip	/ɪə/ h*ere*
/ʒ/ rou*ge*	/ɛə/ = /eə/ th*ere*
/h/ *h*ot	/ʊə/ p*oor*
/tʃ/ *ch*air	
/dʒ/ *j*ump	
/w/ *w*et	
/j/ *y*es	

Appendix B: The Indo-European family of languages