Introduction to English linguistics

A companion to the seminar

Compiled, revised and abridged by
László Varga

Newly revised and abridged in 2017 by
Péter A. Lázár

Eötvös Loránd University, Budapest
2017

Preface to the 2010 version

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).

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
Eötvös Loránd University
Budapest
August 2010

Preface to the 2017 version

Seven years have passed since the last abridged and revised version of this Companion became available in 2010. It has now seemed prudent to streamline the text somewhat, which has meant further abridgment, just a few changes, and even fewer minor additions. The old format, the ten topics and their sequence, and above all, the original concept, and the general drift of the units, however, have all been preserved.

The text, which has always been an introductory one right from the first, then yet "unabridged", version back in 2007, has remained eminently useful but only as an entry-level introduction to linguistics. This has always meant, among other things, that the latest advances in linguistic thinking have never been included, and no school of linguistics or approach to language has ever been singled out. Neither has this deliberately conservative attitude to the subject matter changed now, no matter how fast – indeed radically – most branches of linguistics have been changing.

The key notions in each Unit are in **bold face** *– familiarity with these is obviously desirable for the understanding of a given topic. Even more importantly, students should make sure that they have a reliable knowledge of all the terms printed in* **this colour***, which are of particular importance.*

Should the interested student wish to delve deeper into any of these topics, find about some new developments in the areas (not) treated here, or gain insight into some of the more recent approaches to language, I also recommend the two introductory books listed above, in 2010. Their more recent editions are:

Fromkin, V. & R. Rodman, Hyams, N. *An Introduction to Language*. 9th international ed. Wadsworth Cengage Learning. 2011.

Radford, A., M. Atkinson, D. Britain, H. Clahsen & A. Spencer (1999) *Linguistics, An Introduction*. CUP. 2009.

Also of interest may be other books, recent and not so recent, elementary and less so, introducing (by far not just) first year students to linguistics, including:

Akmajian, A., R. A. Demers, A.K, Farmer & R. M. Harnish *Linguistics: An Introduction to Language and Communication.* The MIT Press. 1979.

Baker, A. E. & K. Hengeveld *Linguistics.* Wiley-Blackwell. 2012.

Dobrovolsky, W. & F. Katamba *Contemporary linguistics*. Longman. 1996.

Finch, G. *How to study linguistics.* 2nd ed. Palgrave Macmillan. 2003.

Hazen, K. *An Introduction to Language.* Wiley-Blackwell. 2015.

Meyer, Ch. F. *Introducing English Linguistics.* CUP. 2009.

O'Grady, W., M. Dobrovolsky & M. Aronoff *Contemporary Linguistics An Introduction.* 3rd U.S. Edition. Bedford/St.Martin's. 1997.

Richter, B. *First Steps in Theoretical and Applied Linguistics.* Bölcsész Konzorcium Budapest. 2006.

Yule, G. *The study of language,* 4th ed. CUP. 2010.

Péter A. Lázár
Eötvös Loránd University
Budapest
August 2017

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 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. (Here is what St. Augustine said about the sign: "[it] is something that gives itself to the senses but something beyond itself to the mind".) Signs have an exponent, a meaning, and a set of referents. The **exponent** of a sign is its physical manifestation, which can be perceived (heard, seen, touched, etc.) by the receiver of the message; it can be e.g. a gesture, a facial expression, a picture, a road sign, the sounding of a horn, a word, the smell of burning. The individual things, qualities, actions, and states in the world which a sign refers to are the **referents** of the sign, and these together constitute the **reference** of the sign. In addition to having an exponent and reference, signs are also associated with meaning. The **meaning** of a sign is the concept which it evokes in its users. The signs used in a communication system constitute a **code**.

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

- When the exponent of a sign has an **arbitrary** relationship to the sign's referents, the sign is a **symbol**. For instance, the colours used in traffic lights are symbols, i.e. they 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 *man*, Russian *čelovek*, German *Mann*, Spanish *hombre*, etc., which all refer to the same kind of thing – have the same reference – but all sound different.

- However, when there is a natural link, a resemblance between the sign's exponent and its referents, 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 – the **onomatopoeic** words, e.g. English *buzz*, *dingdong*, *miaow*, *cuckoo* – 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.

- Finally, when the exponent of a sign is mechanically linked to its source in such a way that the sign is a spontaneous reflection of the state of the source, the sign is a **symptom**. 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 the given state of affairs. Human beings, however, are capable of producing some symptoms deliberately. This happens in acting, or deceiving, 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**. (In Unit 5 we will see, though, that many words can be decomposed into smaller meaningful parts called **morphemes**, e.g. the English third person singular present tense verb form *enjoys* contains the morphemes *en-*, *joy* and *-s*.) Words and morphemes are predominantly symbolic signs, though some of them (the onomatopoeic ones) are partly iconic. The linguistic signs and the rules for their combinations used by a language community constitute a **linguistic code**.

Linguistic communication takes place in the following way. Speaker A, in her/his mind, selects words, and combines them according to the rules of the language, i.e. **encodes** the message. Then her/his articulatory organs – or, in the case of writing, her/his 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 her/his mind. Note that in sign (or signed) languages, which we have ignored here but which are as rich and complex as any spoken (oral) language, the signs are realised by the body, the hands/arms, and the face.

The above steps are shown in (1).

(1) Linguistic communication

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

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

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

A comparison of human languages and animal communication systems (animal "languages") can be made in terms of what are called **design features** of language, features that characterise human languages.

- One of these – **arbitrariness**, i.e. the absence of a natural bond or link between the exponents and referents of signs – has already been mentioned. Arbitrariness is present in the words (and morphemes) of human languages (except the onomatopoeic ones, which are partly natural). Interestingly, a certain degree of arbitrariness is also present in bee dancing (the special movements that bees perform with their wings and bodies to communicate to their fellow bees about the direction and distance of a source of nectar). This is arbitrary because there is no obvious connection between the form of the dance and the distance from the hive. This arbitrariness, however, is of a very limited kind, manifesting itself only in connection with localising food.

- Probably the most important design feature of language is **duality** (also called **double articulation**). By this is meant that every human language is organised into two layers.

The first is a layer of basic sounds called phonemes, such as /æ/, /k/ and /t/, which are meaningless in isolation, and take on meaning only when they combine in certain ways.

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

The second is a layer of meaningful units (morphemes, words, phrases, sentences), which result from combining the basic sounds, e.g. /kæt/ *cat*, /tæk/ *tack*, /ækt/ *act*, or from combining meaningful units, 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.

- **Creativity**, another important design feature means that humans 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, which exists in connection with localising food again.
- Another important design feature is **patterning**. This means that every language has certain possible, i.e. permitted ways of combination in which phonemes can be combined into words and words into sentences. For instance, English has /kæt/, /tæk/, /ækt/, but not */ktæ/ or */tkæ/ or */ætk/. (The star – the **asterisk –** before an item means that it is **ill-formed**, i.e. **ungrammatical**.) Similarly, *He is happy*, *Is he happy?* and even *Happy he is* are possible, but not **Is happy he?* The design feature of patterning, too, is missing from animal communication systems.
- Finally, **displacement** (Hungarian: eltolás; elvonatkoztatás; áthelyezhetőség) is the design feature that can be defined as the ability to use language in connection with things and events that are remote in space and time. For instance, humans can talk about past, future, and distant events, hypothetical objects; we can even lie. Displacement is again present in bee dancing, but (again) only in connection with food; moreover, 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 (that is, "now"), but they cannot pass on information about the availability of a source of nectar in the future or in the past.

The following table sums up the design features discussed so far, and the corresponding values in bee dancing.

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

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

1.3 The discrete nature of language

Linguistic communication, i.e. the use of language, is characteristically (a) vocal and (b) verbal behaviour, and it involves the use of (c) discrete language elements.

(a) It is **vocal** because it is associated with the articulatory (vocal) organs.

(b) It is **verbal** because words play a central part in it.

(c) It involves the use of **discrete** language elements: this means that these elements differ from one another discretely (i.e. 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: 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 these realisations, but these variations are gradual, hardly noticeable, and also insignificant – so each rendering will be taken as realising the same word *pen*. However, when the words *pen*, *pan*, *car*, *picture* and *camera* are uttered, they 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 one of them is changed for another in a particular sentence, it may result in 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**. These 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, 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/ are minimal pairs in English.

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

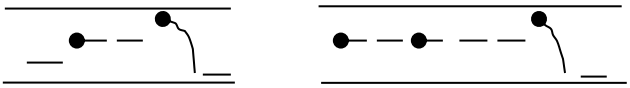
Phonemes are **segmental elements**, because they are the smallest building blocks (i.e. segments) of words and sentences. But words and their sequences in sentences also contain **suprasegmental elements**, which are called so because they are "superimposed" upon (placed on top of) units that are or can be larger than segments, such as e.g. syllables. Suprasegmental elements are also known as **prosodic elements**.

The most significant suprasegmental elements are stress patterns and pitch (i.e. height) patterns; these are discrete, too. **Stress** is a degree of the prominence of a syllable. Stress in English has four discrete degrees: **non-stress**, **tertiary** stress, **secondary** stress, and **primary** stress (in order of increasing strength).

When you would like to show which syllables are stressed in any word, you may put the symbol ˈ before a primary-stressed syllable, and the symbol ˌ before a secondary-stressed syllable, as in the examples in (5): *Monongahela* (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 drawings representing the pitch of the syllables.

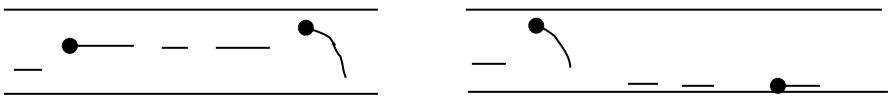
(5)a. /məˌnɒŋgəˈhiːlə/ b. /ˌɪndɪˌvɪdjʊəˈlɪstɪk/



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: /ˈɪnsʌlt/ vs. /ɪnˈsʌlt/.

But they can also distinguish sentences: for instance, in (6) the interpretation depends on where the primary stress falls. In (6a) it falls on *come* and the sentence implies 'but he didn't come', whereas in (6b) it falls on *thought* and the sentence implies 'and he did come'.

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

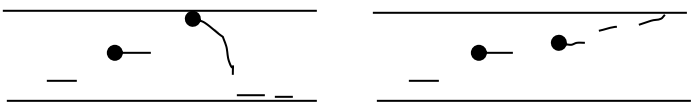


Pitch patterns (= **tones**) are carried by syllables or syllable sequences. The commonest pitch patterns are the **falling** (high-low), **rising** (low-high), **falling-rising** (high-low-high), **rising-falling** (low-high-low), **high level** (high), and **low level** (low) tones.

In languages such as English or Hungarian, pitch patterns are just used in **intonation**, they form the meaningful melodies of spoken sentences rather than of words. Such languages are called **intonational languages**. Compare the English utterances (7a) and (7b), which differ only in the final parts of their intonations. Intonation can be transcribed by means of graphic symbols which simultaneously indicate stress and intonation.

The difference in intonation shows that (7a) is a statement, and (7b) is a yes-no question.

(7)a. They ˈcame ˩yesterday. b. They ˈcame ˧yesterday?



By means of these graphic symbols (the **tonetic stress marks**) the stressing and the intonation of sentences can be transcribed within the line of text, and thus 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.

1.4 Paralanguage

The use of language is characteristically vocal and verbal behaviour, and it involves the use of discrete elements. However, accompanying and occasionally even replacing language, behaviour which is not vocal is also found, or, if it is vocal, is not verbal and is not discrete. Variations in this kind, which are used during and instead of linguistic communication, are called **paralinguistic features** or **paralanguage**.

The **non-vocal** features of paralanguage include gestures, bodily movements, facial expressions which we make while we speak or instead of speaking, e.g. bowing, waving, winking, raising our eyebrows, putting our finger across our lips, shaking our head, nodding.

The **vocal** paralinguistic features include various meaningful noises, such as *hm*, 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, sketched in 1.3, is vocal and non-verbal, and it is partly linguistic, partly paralinguistic. The stress degrees are linguistic because they are discrete, but the general loudness 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; their vertical range and the general pitch height of certain parts of utterances, however, are gradable, i.e. paralinguistic variations. Similarly: tempo, pause length, voice quality variations are gradable, so they belong to paralinguistic rather than linguistic communication.

Exercises, problems, and other tasks

1 What is communication?

2 What is the definition of 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 Classify the following signs: (a) involuntary cough (b) cough for getting attention (c) nod of head ('Yes') (d) *Uh-huh* ('Yes') (e) *Yes* (f) hand indicating height from ground ('So high') (g) waving the hand ('Good-bye') (h) blushing (i) sweating.

6 What are onomatopoeic words?

7 What are the English counterparts of the following Hungarian words: *kukurikú*, *bimbam*, *tiktak*. What is your conclusion?

8 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.

9 How does linguistic communication take place?

10 What is meant by arbitrariness as a design feature of human language? Illustrate it.

11 What are duality and patterning?

12 Combine the following 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?

13 How are the letter combinations <kn>, <ps>, <mb> in *knee*, *knowledge*, *psychology*, *psalm*, *comb*, *dumb*, *acknowledge*, *rhapsody*, *cucumber* pronounced in English? Find an explanation.

14 What do we mean by creativity and by displacement as design features of human language?

15 Linguistic communication is characteristically vocal and verbal behaviour, involving the use of discrete language elements. Explain the key words *vocal*, *verbal*, *discrete*.

16 How are phonemes defined? Why are they segmental elements?

17 What is a minimal pair? Do *collar* and *colour* constitute a minimal pair? And *monkey* and *donkey*?

18 What are suprasegmental elements?

19 What is stress? How many degrees are distinguished? In what sense are they discrete?

20 Why cannot noun–verb pairs like ˈ*insult* vs. *in*ˈ*sult* be found in Hungarian?

21 What is intonation?

22 What is paralanguage? Mention non-vocal and vocal paralinguistic features.

Unit 2

The study of language (i)

2.1 Language: external and internal

A language is a linguistic code, which its speakers know and use, and which manifests itself in its speakers' linguistic **knowledge** on the one hand, and in the actual **texts** (both spoken and written) that those speakers produce in linguistic communication, on the other. Consequently, language can be regarded as existing in essentially two modes: on the one hand, it can be looked upon as the language users' knowledge which makes linguistic communication possible, an internal property of the human mind. On the other hand, it can be regarded as a body of objective facts (texts as products; strings of sounds or letters) produced and perceived by its users.

One of the most well-known 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 – structuralist linguistics – 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 it described the regularities (the patterns, or structures) that were found in those samples. Since then, however, the emphasis of language study has shifted to I-language, i.e. to the linguistic knowledge that native speakers possess and put to use when they communicate. Generative linguistics (see Unit 3) aims at **modelling** the I-language of the native speaker, i.e. their linguistic knowledge or **internal grammar**.

2.2 Components of language

A natural language (whether you 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 and the 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/, /ækt/ and /tæk/ are phonologically well-formed, but */ktæ/, */ætk/ and */tkæ/ are phonologically ill-formed, in English.

- Another component is **morphology**. This includes the morphemes and the rules for combining them to **derive** (i.e. to form) and to **inflect** words in a particular language. (For the time being we define morphemes as the smallest meaningful units of a language. This definition will be made 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 *elect-ion* (which is a new vocabulary item that is **derived** from *elect*). In a similar way, the plural morpheme *-s* can be added to the noun *election* to obtain the plural form of the same noun: *election-s.* This, however, is not a new vocabulary item, but the **inflected** variant of an already existing one. The **morphological** rules of English allow the sequence *un-friend-li-ness* to be a morphologically well-formed word, while **friend-li-un-ness* is ill-formed.

- **Syntax** is the component of language that contains the rules for putting together words in phrases, and phrases in sentences. For example, the 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 **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. (In this book, the raised exclamation mark ! before a sentence indicates that it is semantically anomalous.)

In addition, a component is recognized in which all the central components may play a role: a **lexicon**. This is a list of the vocabulary items of a language, which contains all idiosyncratic (particular, individual) information about those vocabulary items (such as the **unpredictable** aspects of their phonology, morphology, syntactic behaviour, and meaning). Words, once they have been 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 ("on line") every time they are used by a speaker.

Native speakers of a language have linguistic knowledge: they know their language. The language they possess is I-language, which is an **internal grammar**. (Note that while 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.) Native speakers know both the elements and the rules (these are not man-made rules!) in the various components of their language, and on the basis of this knowledge, they can tell whether a string of words is **grammatical** (**well-formed**) or not (**ungrammatical**, or **ill-formed**). But most speakers are unable to explain (either to their children or to their foreign friends) why it is that one string of words is grammatical and another is not. This is because their linguistic knowledge is **intuitive** (**subconscious**, **implicit**), and they cannot express it explicitly.

2.3 Linguistics and its branches

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

The product of linguistics is a **systematic**, **objective**, 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). Someone who knows a number of languages (i.e. a polyglot) is called a linguist in English, but this person is not necessarily a professional linguist in the sense relevant for us now – and a linguist does not necessarily know a number of languages.

Linguistics has branches that correspond to the central components of language. **Phonology** is the study of the phonemes and their combinations in words and morphemes, and also of the suprasegmental elements. **Morphology** is the study of the derivation and inflection of words 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:

phonology	morphology	syntax	semantics
lexicology			

Moreover, all these can be studied from a **synchronic** point of view, i.e. looking at a particular state of language at a particular point of time, or from a **diachronic** (= historical) point of view, i.e., how they change through time. (The Greek word *khronos* means 'time', *syn-* means 'together', and 'alike', and *dia-* means 'through'.)

In a somewhat broader concept of linguistics there are (not just phonological but) phonetic and (not just semantic but) pragmatic components, too.

Phonetics, which is closely related to phonology, is the study of the **production**, **physical properties**, and **perception** of the actual sounds realising the phonemes and of the suprasegmental elements.

Pragmatics is close to (or part of) semantics, and the difference is not even always clear. While semantics examines what sentences and words mean in themselves, pragmatics studies what speakers mean, i.e. the ways in which the sentences and words get different interpretations in different situations. If you put the question *Can you play the piano?* to a person you are interviewing in a room where there is no piano, your sentence will count as a real yes-no question. But if you say the same sentence (which is not a real question this time) to a person who is a good pianist, and you point to a piano at the same time, your sentence will count as a request to play.

Sociolinguistics may further extend the scope of linguistics. This is an interdisciplinary branch of study (relevant to both linguistics and sociology), which studies 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 foreign-language teaching, speech therapy, advertising, literary criticism, stylistics, etc. These involve various kinds of **applied linguistics**. For example, when doctors want to cure an **aphasia** patient, i.e. someone who has lost – partly or completely – the ability to use language, they will have to know about the language system.

This booklet is 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.

2.4 Traditional Grammar

Languages began to be studied a long time ago, in the 5th century BC or earlier; but it is only since the 19th century that you can speak about linguistics. It was in the 19th century that historical language study began to meet the criteria of **scientificness**, and it was 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**. This kind of study dealt with the contemporary state of languages, but it often mixed its synchronic statements with diachronic ones. Traditional Grammar was not sufficiently scientific.

- It was not **explicit** enough: it was often too vague in its statements, and its **definitions** were 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 feel to be nouns even though they are not the names of persons, places or things – e.g. *reflection*.
- It was not **systematic** enough: it ignored the spoken language, and was preoccupied with the **written** language, especially of older literary works.
- It was not **objective** enough: it was often **prescriptive** and **puristic** rather than **descriptive**. Instead of recording what the language being 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 e.g. **Latin**. For example, they argued that the **split infinitive** – which is quite common in English – was incorrect: "One shouldn't say *to humbly apologize*, one ought to say: *to apologize humbly*". The idea that the split infinitive was wrong was based on Latin: it was believed that, since you could not insert anything in the middle of a Latin infinitive, its English equivalent should not be interrupted either. 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 has elaborated linguistic terminology much of which has been in use ever since. Modern linguistics would not have been born if there had been no Traditional Grammar to prepare the way.

2.5 Comparative Philology

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

This kind of linguistics emerged after the discovery that Sanskrit was related to Latin and Greek. The discovery was made by William Jones in 1786. Throughout the 19th century, language scholars tried to establish genetic relationships between languages. Various **language families** and **branches** were discovered, e.g. the Germanic branch, of which English is a member. A Proto-Indo-European **parent language** was reconstructed. The study of language was beginning to develop towards an **autonomous** 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. It also tried to show language change in a **systematic** way, as a process that is determined by rules. A group of German scholars actually claimed that language changes were not accidental events, or optional tendencies, but "**laws**" (Lautgesetze).

Meanwhile, the study of the contemporary state of languages went on in the still 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 20th century.

In Europe the study of language at the beginning of the 20th century was characterised by two features: the inheritance of Traditional Grammar, and the mainly 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 **comparative philologist**, was a professor of Sanskrit himself, but his ideas about language and language study went far beyond Comparative Philology.

- He emphasised the difference between (a) language as an **abstract** system, to be found only in the **collective** consciousness of the community (this abstract system he called **la langue**); and (b) language as the **realisation** of that system (which he called **la parole**).
- He separated the **synchronic** and **diachronic** aspects of language study, and argued for the primacy of the former.
- He taught that linguistic signs enter into two kinds of relationship: **syntagmatic** and **paradigmatic**.

The **syntagmatic relationship** is a horizontal – chain, "and" – 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 which consists of the other words on its left and right.

The **paradigmatic relationship** is a vertical – choice, "or" – relationship, which exists between a sign that is 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.

(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	t	*bet*
	ʌ		*but*
	iː		*beat*
	ɪ		*bit*
	ɔː		*bought*

Exercises, problems, and other tasks

1 What is meant mean by E-language and I-language?

2 Define morphology, syntax, phonology, and semantics.

3 What is meant by the lexicon?

4 Why do we say that most native speakers' knowledge of their language is intuitive (subconscious, intuitive)?

5 What makes a study scientific?

6 How do you define linguistics? Who is a linguist? (2 meanings!)

7 Discuss phonetics and pragmatics.

8 What is the difference between synchronic and diachronic linguistics? Who distinguished them first?

9 What are interdisciplinary subjects? Explain sociolinguistics and psycholinguistics. Mention fields of applied linguistics.

10 What are phonemes?

11 What are morphemes? Identify the morphemes in *unanalysability?*

12 What is the difference between ¡*Colourless green ideas sleep furiously* and **Got he late up.*

13 Characterise Traditional Grammar.

14 What is the difference between descriptive and prescriptive approaches to the (study of) language?

15 Comment on these quotations:
(a) "Double negation is illogical."
(b) "The regular plural of nouns is formed by adding the letter *-s*."
(c) "The noun is the name of a person, place or thing."

16 How did Comparative Philology begin? In what sense was it scientific?

17 Does synchronic linguistics necessarily mean the study of the present day state of language?

18 What did Saussure mean by *langue* and *parole*? Match the following expressions with these terms.

potential
He speaks English
individual
actual
social
language system
He is speaking English
behaviour

19 Discuss the syntagmatic and paradigmatic aspects of the following:

Your friend may come
1 2 3 4

20 Compare the number of words that can replace 1, 2, 3 and 4. Which sets of words are open? Which ones are closed?

Unit 3

The study of language (ii)

3.1 The Beginnings of Modern Linguistics in America

Linguistic research in the USA also began in the early decades of the 20th century, but with a different motivation. The languages of the American Indian (Native American) population, the Amerindian languages, were **threatened with extinction**, and so the main aim of linguistics was to describe these languages as quickly and accurately as possible. Modern American linguistics in the first half of the 20th century was 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 the languages studied until then, and the linguists who wanted to describe them did not speak them – so no prescriptive statements could be made about them. These languages simply forced language scholars to adopt a **non-traditional** approach to language, based on explicitness, systematicness and objectivity.

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

Some linguists went even further: they 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 view is called **linguistic determinism**. The combination of linguistic relativism and linguistic determinism became known as **the Sapir–Whorf hypothesis** after the scholars who propagated them. According to the **strong version** of the hypothesis, the vocabulary and grammatical categories of an individual's native language determine the ways in which (s)he can interpret his/her experience. For instance, it was discovered that in Eskimo there were several different words for different kinds of snow. Other linguists collected similar facts (also grammatical ones) from other languages: for instance, the Hopi language does not distinguish present, past and future tenses. On the basis of such examples the conclusion was drawn that people belonging to different cultural-linguistic groups not only spoke but also thought differently: that each such community lived in the prison of its language.

This conclusion 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. The fact that English has no separate words for different kinds of snow does not mean that its speakers cannot see these differences, only that they are not significant to them. When they do become important, English speakers can paraphrase and say "hard packed snow", "powdery snow", etc. The main counterargument against the strong form of linguistic determinism is the possibility of **translation**. Translation is possible for most of the time, and although not everything can be translated with the same ease, you are usually able to paraphrase or explain what you mean in any language.

However, a **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 some aspect of reality in a 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. But differences in codability between languages are of secondary importance: only the less basic, culture-specific concepts may present codability problems. The essential things are equally codable because they are equally relevant to all cultures.

3.2 The Great Synthesis of American Structuralist Linguistics

Leonard Bloomfield and his followers thought that a linguist should collect observable data, i.e. real "texts", and analyse these, i.e. segment and classify the physical features of the utterances collected. A body of such data (i.e. 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 Noam Chomsky's terminology, this means that American structuralism was preoccupied with discovering and describing E-language, i.e. the E-language aspect of natural languages.

The Bloomfieldians dealt with phonetics, phonology, morphology, and syntax, but had little to say about semantics. The only aspect of meaning they paid attention to was whether two forms (signs or sign combinations) had the same or different meanings.

They used a strictly **formal analysis**: an analysis without reference to meaning. It was based on an examination of **distribution** and **constituency**.

- The **distribution** of a language element (i.e. of a phoneme, 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 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, and vice versa. For example, the English phoneme /l/ has two variants, and they are in complementary distribution. (Such variants of a phoneme are called **allophones**.)

The "clear" [l] – which is like Hungarian /l/ – occurs before vowels, e.g. [ˈhelɪn] *Helen;* the "dark" [ɬ] – which is pronounced with a cupped tongue – occurs **elsewhere**, i.e. before consonants, e.g. [heɬp] *help,* and in word-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 in **non-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 by replacing one by the other, the meaning of the unit changes. For example, English /e/ and /iː/ in the environment /m___t/ are in contrast because /met/ does not mean the same as /miːt/.

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 /iː/ are in free variation in the environment /—kəˈnɒmɪks/, because /iːkəˈnɒmɪks/ means the same as /ekəˈnɒmɪks/.

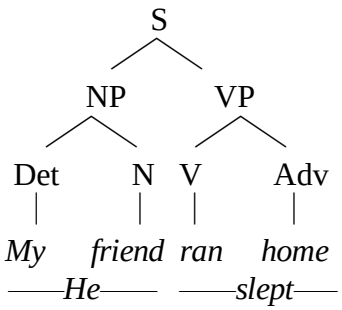
- The other important method of formal analysis that the Bloomfieldians introduced was **constituent analysis** ("immediate constituent analysis", or "IC analysis"). This means cutting syntactic units (or

words) into their constituents, then those constituents into their constituents, and so on, until the individual words (or morphemes) are reached.

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]*. Cf. [*Peter*] [*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** and **represented** in essentially two ways: by **brackets**, as in (1), or by **tree diagrams**, as in (2).

(1)a. [$_{S}$[$_{NP}$[$_{Det}$ *My*][$_{N}$ *friend*]] [$_{VP}$[$_{V}$ *ran*] [$_{Adv}$ *home*]]]

(1)b.



(Note that the syntactic representations in this book are preliminary and will be substantially modified in your later syntax studies.)

The constituents in (1a) and (1b) 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 brackets do not have to be labelled, but the labelled ones are more informative.

Constituent analysis can resolve certain **structural ambiguities** by showing different constituent structures, e.g.:

(2) [*old* [*men and women*]] vs. [[*old men*] *and women*]

or simply, using just two pairs of brackets:

(3) *old* [*men and women*] vs. [*old men*] *and women*

3.3 Generative Linguistics

There are many ambiguities like this t constituent analysis cannot **resolve** (= **remove**). Consider the sentence *The lamb is ready to eat*. It has two distinct meanings – it is ambiguous – and because the words mean the same in both of its meanings ("**readings**"), it can only be the different structures that cause the ambiguity.

Constituent analysis can only give it one analysis: [[*The lamb*] [*is* [*ready* [*to eat*]]]]. This kind of analysis remains on the **surface** and cannot disambiguate structures which are **underlyingly** different, i.e. different "in the deep".

The dissatisfaction with the limitations of structuralist linguistics led to a radically new type of linguistic analysis towards the end of the 1950s. This is known now as **generative linguistics** (= generative grammar). Back then, this kind of analysis distinguished two levels of syntactic analysis: a **surface structure** and an underlying abstract **deep structure** (this latter was not recognised by the structuralists). These two levels are still with us in some approaches in different forms.

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 now eat' and 'Somebody can eat the lamb now') is derived from two different deep structures.

Synonymous sentences like *It rained yesterday* and *Yesterday it rained* derive from one common deep structure and differ only on the surface.

Generative grammar is said to be able to **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, you 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 native speakers would not be able to learn them).

The founder of generative linguistics is the American **Noam Chomsky**. Since its appearance the theory has been modified and remodified several times and several new proposals have been and are still being made by Chomsky himself and now mainly by others.

Bloomfield and his school argued that the purpose of linguistics was to describe individual languages. In contrast, Chomsky holds that linguistics should be primarily concerned with **Universal Grammar (UG)**, i.e. with the **principles** that are the properties of all human languages. One of these principles is **structure dependence**, which means that syntactic operations apply to phrases, not words. If this is so, then these operations require a knowledge of the structural relationships of words rather than just their **linear sequence**.

When you change a **declarative** sentence into a yes–no **interrogative**, the **auxiliary** that you must move to the front of the sentence is not simply the second word of the declarative sentence, as a superficial observer might think on the basis of (4a). Rather, it is 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 it → Will [$_{NP}$ John] — buy it?

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

In Hungarian sentences, too, the syntactic constituents are not individual words but structural **units** composed of words (and occasionally consisting of just a single word), i.e. **phrases**, such as *elvittem* 'I took', *a legkisebb gyerekeket* 'the youngest children', *egy fagylaltozóba* 'to an ice cream parlour': (5a), (5b), (5c).

(5)a. [Elvittem] [a legkisebb gyerekeket] [egy fagylaltozóba]

b. [Elvittem] [egy kis fagylaltozóba] [a legkisebb gyerekeit]

c. [A legkisebb gyerekeit] [elvittem] [egy fagylaltozóba]

When "**word order**" is mentioned", what is really meant is the order of these chunks/units of language, i.e. **constituent order**.

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**, i.e. **innate** in humans. Chomsky calls the native speaker's knowledge of language **competence** (or I-language). But that knowledge, competence, must be distinguished from the **use** of that knowledge, 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 that the individual can produce, etc.

Chomsky's distinction between **competence** and **performance** may remind one of Saussure's distinction between **langue** and **parole**. But while Chomsky uses the term *performance* in very much the same sense as Saussure used parole, there is a difference between competence and langue. Saussure's langue is static: it is the system of linguistic signs. Chomsky's competence is dynamic: it focuses on the **generating** (**generation**) of sentences. Also, 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, so the basis of competence is a universal characteristic of the human species.

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

- produce and understand an infinite number of **new** grammatical sentences in their language
- distinguish between **grammatical** and **ungrammatical** formations in their language (*He went there* vs. **Went there he*)
- interpret **elliptical** sentences, i.e. sentences with missing elements (*Peter is happy but John isn't* ___)
- perceive **ambiguity** (*The lamb is ready to eat*)
- perceive **synonymy** (*The duck crossed the road* vs. *The road was crossed by the duck*)
- **idealise** utterances, i.e. they can disregard e.g. the imperfections of performance, and "reconstruct" the grammatical sentences which the utterances realise.

The last point has an important consequence: recall that generative linguistics has an **I-language approach** to (the study of) language. Earlier, linguists 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 linguistic behaviour (actual language use), i.e. from parole or performance. This was the **E-language approach**. By contrast, generative linguists think that linguistics is concerned with far more than what can be found in a corpus. Even if you do use corpora for linguistic work, you have to idealise the data, i.e. free the corpus from the **imperfections** and **individual** features of performance. **Idealisation** is what native speakers automatically do when they understand other speakers' utterances.

But then the 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, linguists have the right to use their own and other people's intuitions in linguistic analysis. And if the linguist is a native speaker of the language (s)he examines, (s)he can ask and answer questions about her/his own intuitions. Examining one's own intuitions concerning language is a kind of **introspection**: generative linguists can base their theories not (only) on **empirical** facts but also on introspection and on native speakers' **intuitions**. This does not mean that they give up objectivity, because their theories can be submitted to subsequent **empirical verification**. Their method, however, is different from the **inductive** method of the preceding decades: it is **deductive**, proceeding from theories to empirical facts. But the focus of their attention is on I-language: they are interested not so much in the empirical facts themselves as in the knowledge that enables speakers to produce those facts.

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 an American descriptivist) to show that it is possible to analyse sentence structure without reference to meaning: *The woggles ugged a diggle*. Analyse it, and transform it into (a) a yes-no-question, (b) negative, (c) passive, (d) singular.

3 What is the name of the kind of linguistics in the USA in the first half of the 20th century?

4 Discuss linguistic relativism, linguistic determinism, and the Sapir–Whorf hypothesis in both its strong and weak forms.

5 What is meant by codability?

6 When are two language elements in complementary distribution?

7 What is contrast, and what is free variation?

Comment on the vowels /ɪ/ and /aɪ/ in *sit* and *sight*, and in /dɪˈrektə/ and /daɪˈrektə/. Comment on the vowels between /f/ and /l/ in the Hungarian pairs *fel* 'up'– *föl* 'up' and *felém* 'towards me' – *fölém* 'above me'.

8 If two language elements never occur in the same environment, are they in contrast, free variation, or complementary distribution?

9 Can clear [l] and dark [ɬ] distinguish minimal pairs in English? Explain.

10 Why is constituent order a better term than word order?

11 Reveal the constituent structure of these 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. (Find the ambiguity.)

12 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*

13 When is a grammar generative?

14 What is Universal Grammar?

15 What is meant by structure dependence?

16 How are competence and performance defined?

17 What can native speakers do on the basis of their competence?

18 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*

19 What does it mean that generative linguistics has an I-language approach to the study of language?

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 branches.

- **Articulatory** phonetics examines the articulatory (vocal) organs and their role in the production of speech sounds.
- **Acoustic** phonetics deals with the physical properties of speech sounds.
- **Auditory** phonetics examines the way in which human beings perceive speech sounds.

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 examines 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. By contrast, the sounds we produce when we sneeze or cough 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 phonetics is twofold: to provide a **notation** for each speech sound and a **description** for them. By notation we mean a system of **transcription symbols** whereby an accurate and unambiguous record can be made 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 all languages.

In English, for example, there are just 26 letters (this is easy to count) but considerably more speech sounds. 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). In this system each **phonetic symbol** stands for one and only one speech sound. Sometimes **supplementary marks** (called **diacritics**) are added to the symbols, e.g. the raised letter *h* indicates aspiration of the initial sound [t] in the word [tʰuːɬ] *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 aim of phonetics is the description (characterisation) of speech sounds. This is done in terms of **phonetic features**. To understand these features, one has to get acquainted with the **articulatory organs**. (This will be done in your phonetics and phonology courses.)

One characteristic feature of speech sounds is the presence or absence of the **vibration** of the **vocal cords** (**vocal folds**) during the production of a sound. The air coming from the lungs by way of the **wind pipe** arrives at the **larynx**. This is where the vocal cords are, forming an opening between them, which is called the **glottis**.

When the vocal cords are together and the air stream that passes through between them makes them **vibrate**, the sound produced will be **voiced**, e.g. [b, d, g, v, ð, z, ʒ]. When the vocal cords are apart and so the air stream passes through freely, i.e. without causing vibration of the vocal cords, the sound produced will be **voiceless** (or **unvoiced**), e.g. [p, t, k, f, θ, s, ʃ].

Another feature of speech sounds is the presence or absence of **nasality**. The air, leaving the glottis, arrives at a cavity called the **pharynx**, from which it can go on to two further cavities: the nose and the mouth, i.e. the **nasal cavity** and the **oral cavity**. These are separated from each other by the roof of the mouth. The roof has several parts: just behind the upper front teeth is the **alveolar ridge**, then comes the **hard palate**, followed by the **soft palate** or **velum**. When the back of the velum, i.e. the **uvula** 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]. When 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].

Consonants can be described in terms of **place** and **manner** of articulation (in addition to voice).
In English, eight **classes** of consonants are distinguished according to **place of articulation**:

bilabials ([p, b, m, w]), produced between the two lips
labiodentals ([f, v]), produced between the upper front teeth and the lower lip
dentals ([θ, ð]), produced between the upper front teeth and the tip of the tongue
alveolars ([t, d, s, z, n, l, r]), produced between the alveolar ridge and the front of the tongue
palato-alveolars ([ʃ, ʒ, ʧ, ʤ]), produced in the post-alveolar region
palatals ([j]), produced in the area of the hard palate
velars ([k, g, ŋ]), produced in the area of the soft palate or velum
glottals ([ʔ], called the 'glottal stop', and [h]), produced in the glottis.

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 [ˈskɒʔlənd] or [ˈluːʔən], or indicate a syllable boundary between two vowels, as in [kəʊˈʔɒpəreɪt].

According to the **manner of articulation** six classes are distinguished:

plosives [p, b, t, d, k, g, ʔ]
fricatives (also known as spirants) [f, v, θ, ð, s, z, ʃ, ʒ, h]
affricates [ʧ, ʤ]
nasals (also known as nasal stops) [m, n, ŋ]
liquids [l, r]
glides [w, j]

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

In (1), the columns represent the place of articulation, the rows the manner of articulation.

(1) English consonants

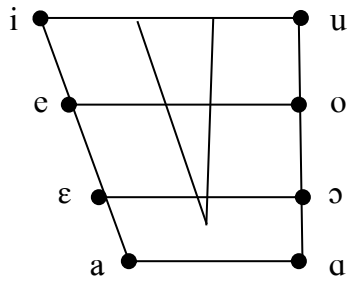
		Bilabials	Labiodentals	Dentals	Alveolars	Palatoalveolars	Palatals	Velars	Glottals
Obstruents	Plosives	p b			t d			k g	ʔ
	Fricatives		f v	θ ð	s z	ʃ ʒ			h
	Affricates					tʃ dʒ			
Sonorants	Nasals	m			n			ŋ	
	Liquids				l, r				
	Glides	w					j		

Eight pairs of consonants show a voiceless–voiced opposition: these pairs are in the boxes that have two symbols, the voiceless coming first. The only exception is the box with /l/ and /r/, which are not opposed this way but are simply the two (voiced) liquids.

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].

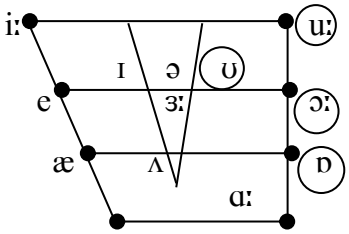
The eight vowels you get this way are called **cardinal vowels**. They do not necessarily occur in every language, but should rather be regarded as orientation points which indicate the limits within which the tongue can move to produce vowels, and with reference to which all vowels of all languages can be accommodated. The **trapezium** formed by the cardinal vowels is called the **Cardinal Vowel Chart**, see (2).

(2) Cardinal Vowel Chart



The most important **simple vowels** of English are shown in (3). They are simple because the tongue position characterising them is steady throughout producing the vowel. The vowels in the triangle are **central**, those on the left of the triangle are **front**, those on the right are **back** vowels. The circled vowels are produced with **lip rounding**: they are **round** vowels. The vowels whose symbols have a colon (:) are **long** vowels.

(3) English Simple Vowels



English has **diphthongs** as well: complex vowels during whose production one tongue position is changed into another, but no new syllable is formed. The vowels in the words *height*, *hate*, *house*, *hose*, i.e. [aɪ, eɪ, aʊ, əʊ] e.g. are diphthongs.

Sometimes even **triphthongs** are distinguished, in which the vowel has three tongue positions one after the other, as in e.g. *fire* and *power*, which contain the triphthongs [aɪə] and [aʊə], respectively.

Consonants and vowels together can be called segments. Since phonetics primarily deals with these, the major part of phonetics is **segmental phonetics**. But phonetics deals with other aspects of human speech as well: 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 sentences) and **stress** (the extra prominence of a syllable).

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 to the really **distinctive** speech sounds, i.e. the 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, can distinguish 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, then the two sounds represent different phonemes, and the two words form a **minimal pair**. For example, the English consonants [k] and [s] represent two different phonemes because they distinguish e.g. [liːk] *leak* and [liːs] *lease*, or [kəʊl] *coal* and [səʊl] *soul.*

The minimal pair technique is based on the notion of **paradigmatic relationship**. By means of the minimal pair technique 44 phonemes can be distinguished 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. The sounds in (1) and (3) above are all phonemes of Standard British English, except for the glottal stop [ʔ]. In addition to them, however, there are a large number of other consonants and vowels in Standard British English which are distinct (i.e. can be distinguished) but which are not **distinctive** (i.e. they do not distinguish words) and so they are not separate phonemes, only **variants** of existing phonemes.

When speech sounds are **transcribed** from the point of view of the phonemes that they represent, all **non-phonemic** (i.e. **non-distinctive**) detail is ignored, and a **broad**, **phonemic transcription** is used. This is 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 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: [tʰuːɬ].

The myriads of actual speech sounds (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 they are 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 [ˈhelɪn] vs. *film* [fɪɬm] and *hill* [hɪɬ]. Other examples include the English phoneme /p/, which also has two allophones: an **aspirated** [pʰ] at the beginning of a stressed syllable, and an **unaspirated** [p] elsewhere, as in *port* [pʰɔːt] (aspirated) and *sport* [spɔːt] (unaspirated). Another example: any English vowel gets a **nasal** allophone, e.g. [ẽ], nasal [e], when it is adjacent to a nasal consonant, but gets an **oral** allophone, e.g. [e] elsewhere, [e.g. *pen* [pʰẽn] vs. *pet* [pʰet]. 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. Since allophonic variations are not reflected in phonemic transcription, the examples given in this paragraph are phonemically transcribed as /lɪp/, /ˈheləɴ/, /fɪlm/, /hɪl/, /pɔːt/, /spɔːt/, /pen/, /pet/.

The terms *etic* and *emic* have been abstracted from the adjectives phonetic and phonemic. They refer 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 (i.e. ability to distinguish meanings) in a given language. In the field of sounds the emic approach is concerned with phonemes, the etic approach with allophones and phones. (We will come back to this issue in 5.2).

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, eɪtθ, bæk, bɔːt, bəʊt, aʊt, dʒɔɪ, ˌeθnəˈgræfɪk, ˌɪŋkənˈsiːvəbəl, tʌb, ʃɒk

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 something like [gʊd ˈbɔːdɪg] instead of [gʊd ˈmɔːnɪŋ] when they have a cold? (Hint: oral vs. nasal).

5 Pronounce the initial sounds of the following words and determine the place and manner of articulation of each: *foot, tooth, box, chips, think, cup.*

6 Which English simple vowels are produced with lip rounding?

7 How many syllables are there in the word *rain*? Why?

8 Transcribe the pronunciation of *pill, lip, help, ten* phonetically and phonemically.

9 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?

10 What are allophones?

11 Comment on the emic and etic approach.

12 Is nasality a distinctive feature of English vowels? And of English consonants?

13 Study the IPA phonemic transcription symbols provided in the Appendix.

Unit 5

Morphology: the study of morphemes and words

5.1 Words

The term **word** can be used in different senses. To avoid confusion, we use the following terms to separate these senses: **lexeme**, **syntactic word**, and **orthographic word**.

- A **lexeme** is a unit of the lexicon (an entry in the lexicon, a vocabulary item of a language), which is an **uninflected** abstract form that underlies all its **inflected** variants (i.e. its different forms). To distinguish lexemes from their inflected variants (small) capitals may be used 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*

Here, the verb *takes* is in the 3rd person singular present tense; *he* is a pronoun in the nominative case; *them* is a 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* 'I get', *kap-sz* 'you get', *kap-Ø* '(s)he gets', where the last form is not uninflected but **zero-inflected**.

The set of inflected variants of a lexeme is called a **paradigm**. The forms *take, takes, taking, took, taken* make up the paradigm of TAKE. The members of such a paradigm are syntactic words.

- An **orthographic word** is a unit of writing: a stretch of graphic symbols with a space on either side (and no space within). For example, the lexeme LIFE INSURANCE is two orthographic words, but the two lexemes in *I'm* are just one orthographic word.

These three senses of "word" are not equally important. In your linguistic studies it is mainly the lexemes and syntactic words that have to be considered – orthographic words are mostly irrelevant.

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 defined the syntactic categories on the basis of meaning. The American structuralists defined the syntactic categories on the basis of form. This involves an examination of 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 will be modified later: **morphemes** are the smallest meaningful units of language, which cannot be subdivided without losing their meaning. They are abstract units, which may be indicated between braces: { }. Lexemes and syntactic words are composed of one or more than one morpheme. For example, the lexeme TEACHER consists of two morphemes: {teach}{-er}; the lexeme ALBATROSS consists of one: {albatross}. (More technically: *teach-er* is **bimorphemic**, i.e. it has 2 morphemes, and *albatross* is **monomorphemic**, i.e. it has 1 morpheme.)

When morphemes are realised, morphs are produced. **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 different shapes, called allomorphs. **Allomorphs** are the **positional alternants** of a morpheme: they have the same meaning, and are in **complementary distribution**.

There is a perfect parallel between the morph—allomorph—morpheme series on the one hand, and the phone—allophone—phoneme series on the other. Just like phones (speech sounds) and allophones are the concern of the **"etic" approach** and phonemes are the concern of the **"emic" approach**, morphs and allomorphs are studied by the etic approach and morphemes by the emic approach.

- 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 (realised as) /ɪz/, as in e.g. *boxes*, *bushes.*
    - When the last sound is a voiceless non-sibilant, the allomorph will be /s/, as in *books*, *plates.*
    - And elsewhere – i.e. where the last sound is a voiced non-sibilant – the plural morpheme will be realised as /z/, as in *bags*, *apples*, *potatoes.*

- In other cases the phonological differences between the allomorphs is due to **lexical conditioning**. For example, 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 – you get *oxen*. If it was the last sound that conditioned, i.e. dictated or decided the variant of the plural, it would be *oxes* – as in the case of *fox–foxes*.

- The phonological difference between the allomorphs can 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. Note that the plural morpheme itself is phonologically conditioned: the last sound of the plural noun is a sibilant, /z/. This means that although the noun *house* is irregular, it is not the plural morpheme that is morphologically conditioned but the stem, which thus exists in two (graphically indistinguishable) alternants: /haʊs/ and /haʊz/.

Morphemes can be grouped into two types on the basis of whether they can form monomorphemic words.

If they can occur by themselves as words, (i.e. if they can form monomorphemic words), they are **free morphemes**. For example {house}, {albatross}, {kangaroo}, {lullaby} are free morphemes.

Morphemes which must be attached to other morphemes within words 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, affixes are either suffixes (which follow 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, 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}, you get *boys*, which is just another syntactic word belonging to the paradigm of the lexeme BOY. However, if you add a derivational suffix to a stem, you create another, new 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.

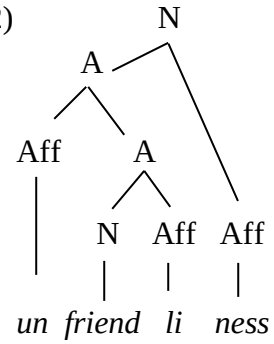
Prefixes in English are always derivational, e.g. {en-}, added to the stem {joy} gives rise to a new lexeme ENJOY.

Inflectional affixes correspond to *ragok* and *jelek* in Hungarian linguistics; derivational affixes correspond to *képzők*.

A **stem** is that part of a word which remains if the suffix or prefix that has entered the word last is removed. The stem is thus 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*, see (2).

If you remove all affixes, you arrive at the **absolute stem**, called **root**, which is obviously always a single morpheme. Thus, the root of *unfriendliness* is {friend}, see (2).

(2)



The root of a word is usually a free morpheme (as in e.g. *unfriendliness*), but there are roots which are **bound**. For example, in words like *in-clude, con-clude* and *ex-clude*, etc. the prefix {in-}, {con-} and {ex-}, etc. is followed by the root {-clude}, which is not a free morpheme, 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 in English. If someone knows Latin they 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 it is not clear whether {-clude} has a meaning, or what that meaning is, {-clude} is still regarded as a morpheme, because its pronunciation, /klu:d/, systematically varies

- with /kluːs/, when it is followed by the suffix {-ive}, as in *inclusive*, *exclusive* or *conclusive*, and
- with /kluːʒ/, when it is followed by {-ion}, as in *inclusion*, *exclusion* or *conclusion*.

So, {-clude} has three allomorphic variants: /kluːd/, /kluːs/ and /kluːʒ/, and it is this that shows that it is a morpheme.

Since, as we have just seen, the criterion of meaning cannot always be used (not all morphemes have meaning, to put it simply), we now revise our definition given at the beginning: **morphemes** are the smallest meaningful 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. Easy **segmentation** means that morphemes, which are (usually) units of both form and meaning, are easy to tell apart.

Languages in which most words are of this kind (i.e. in which most words are sequences of **separable** morphemes) are called **agglutinating languages**. Hungarian is a typical agglutinating language, cf. e.g. {pénz}{-telen}{-ség}{-em}{-ből} (‘money-less-ness-my-from’, i.e. ‘from my not having money’.

In words which are monomorphemic, i.e. composed of single morphemes, 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 morphemes are **fused** together in an inseparable way. For example, the words *took* and *mice* consist of {take}{-ed} and {mouse}{-s}, respectively. Also, there are words in which some morphs represent inseparable **fusions** of morphemes. For example, the Russian nominal inflectional suffix *-u*, as in e.g. *lampu* (‘lamp-Acc.’), simultaneously realises the morphemes {Feminine}, {Singular}, and {Accusative}. Languages in which this fusion of morphemes is typical are called **fusional** (or **inflecting**) **languages**. Latin is a typical fusional language.

These language types, which are 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 languages (e.g. ALCOHOL from Arabic) or the introduction of **coinages** (lexemes artificially invented, e.g. XEROX, KLEENEX), there are also ways in which new lexemes are produced, making use of old ones. These ways are called **word formation processes**. We briefly discuss the commonest of these.

- One of the major word forming processes is **derivation**, i.e. creating a new lexeme by adding a **derivational affix** (prefix or suffix) to an existing lexeme.

For example, the lexeme KINGDOM is derived from the stem {king}, to which the **derivational suffix** {-dom} has been added; the lexeme IMPOLITE is derived from the stem {polite} with **the derivational prefix** {in-}; 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 **derivatives**.

- When a lexeme is assigned to another word class without a change in its form, the process is known as **conversion** (also called zero affixation), which is extremely common in English. For example: $\text{BOTTLE}_N \rightarrow \text{BOTTLE}_V$; $\text{DAILY}_A \rightarrow \text{DAILY}_N$; $\text{MILK}_N \rightarrow \text{MILK}_V$.

A subtype of conversion is called **approximate conversion**, in which lexemes undergo a small but systematic change in pronunciation, and are assigned to a different word class.

Sometimes this change is a **stress shift** (sometimes with some changes in **vowel quality** that go with it). For example, $\text{SUS}'\text{PECT}_V$ 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 $\text{EN}'\text{VELOP}_V$. (ENVELOPE and ENVELOP also differ in the vowel of the last syllable – ˈenvələʊp vs. /ɪnˈvɛləp/ – and this is why they are spelt differently.)

Another kind of approximate conversion is changing the **voice** value of the final **fricative** in some lexemes: these have a voiceless final fricative in nouns and a voiced final fricative in verbs. For example, HALF_N /hɑːf/ → HALVE_V /hɑːv/; USE_N /juːs/→ USE_V /juːz/; WREATH_N /riːθ/→ WREATHE_V /riːð/. The changing of the voice value of the final fricative is sometimes accompanied by vowel change. For example, 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**, e.g. BLACKMAIL, GOLD FISH, HAY FEVER, CHRISTMAS TREE.

Members of a compound may be compounds themselves, e.g. RAILWAY STATION ATTENDANT. Typically compounds bear the main stress on their initial member. They are written in one orthographic word, or separate orthographic words, or with a hyphen.

Minor word formation processes include clipping, blending, backformation and acronym formation.

- **Clipping** means shortening a lexeme and thus producing a more informal variant. For example, PHOTOGRAPH → PHOTO; INFLUENZA → FLU; EXAMINATION → EXAM; UNIVERSITY → UNI; INFORMATION → INFO.

- **Blending** is putting together lexemes but at least one of these lexemes is present only in a fragmentary form. The lexemes so produced are **blends**. For example, FOG + SMOKE → SMOG; BREAKFAST + LUNCH→ BRUNCH.

- **Backformation** is a kind of **reverse affixation**, 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. For example, 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 TELEVISION was not derived from TELEVISE). What happens is that the verb TELEVISE has been backformed from the noun. Similarly, DONATE has been backformed from DONATION (by analogy with pairs such as CREATION and CREATE), and LIAISE has been backformed from LIAISON. The verb *baby-sit* has been backformed from the noun *baby-sitter*. Note that the linguistic term BACKFORM has itself been formed from BACKFORMATION, by backformation.

- **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, FBI, YMCA. In this latter case they are sometimes called **initialisms**.

Exercises, problems, and other tasks

1 Enumerate the syntactic words belonging to the lexemes BE, HAVE and HAPPY.

2 Identify the syntactic category 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 pronounced and also spelt identically, they are **homonyms**, e.g. *bear* /beə/ (the animal) - *bear* /beə/ (the verb *carry*).

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*. (Hint: the case of *keep* is like *house*.)

8 Enumerate the inflectional suffixes of English.

9 Why is it that *unfriendly*, rather than *friendliness*, is the stem of *unfriendliness*?

10 Give a labelled bracketing representation of the morphemic structure of the word *unfriendliness*.

11 Draw labelled tree diagrams to show the morphemic structure of these words: *hospitalisation, organisation, desirability, ungentlemanliness*; *greenhouse, oil well, red-hot, dog food box*.

12 List the bound morphemes in these words: *misleads, submit, previewer, shortened, unhappier, fearlessly, permitted*. Classify the bound morphemes you have found into roots and derivational or inflectional affixes.

13 Discuss each process of word formation, and collect examples of your own.

Unit 6

Syntax: the study of the structure of phrases and sentences

6.1 Sentences and phrases

Syntax is the study of the structure of sentences and phrases. 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. a key word). For example, 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 heads are the nouns (Ns) *king*, *brother* and *car*, respectively.

(What is called Noun Phrase here is nowadays called Determiner Phrase. In this course, however, we continue to use the old term, NP.)

It can happen that a phrase is realised by a **single word**: 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: complements and adjuncts. For example, (1) consists of just a subject and a predicate. The NP *the king* is the **subject**, and the Verb Phrase (VP), which is composed of a single verb (V) *laughed*, is the **predicate**.

A **complement** is a constituent whose presence is structurally licensed (i.e. permitted) or even required 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 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 verb *went* is the Adverb Phrase (AdvP) *home*, which consists of the single adverb (Adv) *home*.

The subject and the complement(s) together are 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 who did the buying), and the object (something that was bought).

In English, subjects typically occur in the **nominative case** (*I*, *he*, etc.), whereas objects occur in the **accusative case** (*me*, *him*, etc.). Observable (visible) **case marking** is restricted to pronouns.

Also, subjects in English typically precede verbs, while complements follow them.

Another difference between subjects and complements is that, in English, verbs **agree** (show **agreement**) with their subjects in **person** and **number** but do not agree with their complements.

In addition to the subject, verb and complement(s), the sentence (clause) may contain constituents which are not structurally licensed (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) *expensive*, which is realised by a single Adjective (A) 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 (both direct and indirect), Adverbial, Attribute, Complement and Adjunct are **grammatical functions** which constituents may perform in the sentence.

Terms such as NP, VP, AP, AdvP, PP, N, V, A, Adv, P, etc. are **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; as the complement of a preposition (e.g. the NP *the tree* in the PP *in the tree*); or 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 by an embedded clause (e.g. *when I was writing a letter*).

6.2 Representation

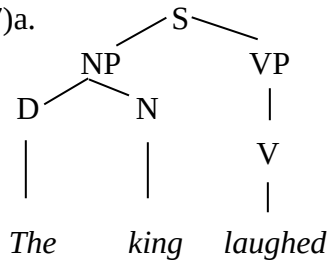
The constituent structure of sentences can be represented in two ways: by means of **labelled tree diagrams**, and by means of **labelled bracketing**. Although the two ways are logically equivalent, tree diagrams are preferred because they help visualise structure better than brackets do. Tree diagrams 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 of the constituents. The new labels in (7) are S, D, Pron, Aux, and DegP; these stand for Sentence, Determiner, Pronoun, Auxiliary, and Degree Phrase. Please note that the complements in (7b), (7c), (7d) – *an expensive car*, *Mary*, *apples*, *home* – are sisters of the verb, while the adjuncts – *on Tuesday* and *very quickly* in (7e) and (7f) – are sisters of the lower VP, with which they form a higher VP.

(The syntactic analyses in these tree diagrams are strictly preliminary and will be substantially modified in your later studies.)

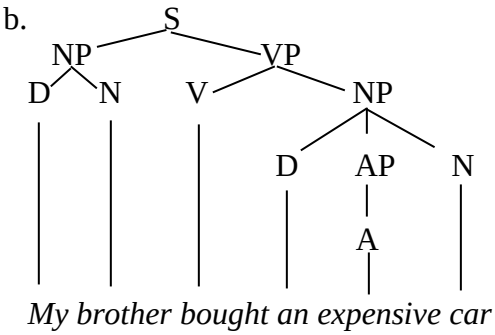
(7)a.



The same structure represented with labelled brackets (in larger print for better visibility):

(7)a'. [$_S$ [$_{NP}$ [$_D$ *The*] [$_N$ *king*]] [$_{VP}$ [$_V$ *laughed*]]]
(7a' is to be read as "seven-ay-prime".)

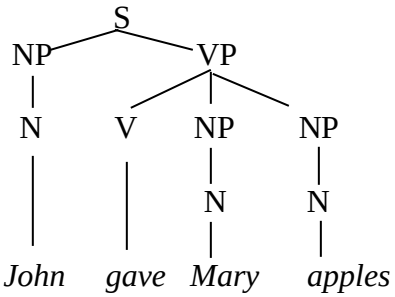
b.



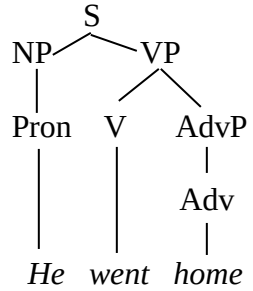
The same structure with labelled brackets (in larger print for better visibility):

(7)b'. [$_S$ [$_{NP}$ [$_D$ *My*] [$_N$ *brother*]] [$_{VP}$ [$_V$ *bought*] [$_{NP}$ [$_D$ *an*] [$_{AP}$ [$_A$ *expensive*]] [$_N$ *car*]]]]

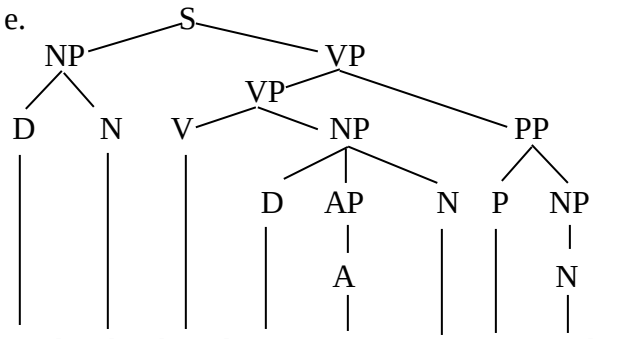
c.



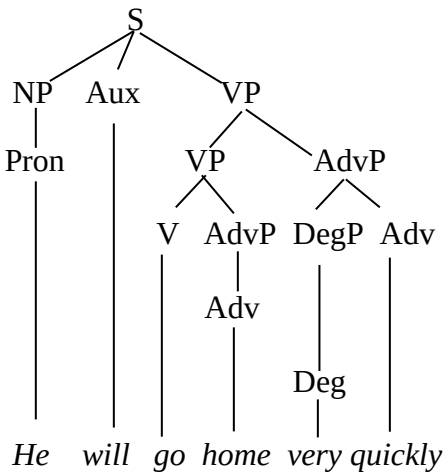
d.



e.



f.

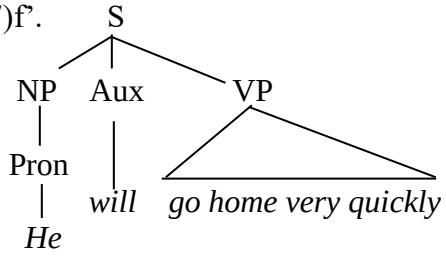


In the last example, (7f), the **auxiliary** *will* stands as a separate constituent outside the VP, although we may intuitively think that the auxiliary should be part of the VP (like this: *will go home*). One of the reasons why it is analysed as being outside the VP is that the VP may be **deleted** independently of the Aux, leaving an **elliptical** sentence, see (8).

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

When you do not want to specify the internal structure of a particular constituent, you may replace the part of the tree diagram corresponding to it by just a triangle. For example, if you wish to ignore the internal structure of the (higher/larger) VP *go home very quickly* in (7f), you may use a triangle for this part of the sentence, see (7f').

(7)f'.



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 word is exclusively dominated by a node: *he* is dominated by the node Pron, *will* by the node Aux, *go* by V, *home* by Adv, *very* by Deg, and *quickly* by the node Adv. The strings *go home* and *very quickly* are also constituents because they are exclusively dominated by the lower VP and the AdvP, respectively. 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 because there is no node 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.

6.3 Simple and complex sentences

Until now, all the constituents (apart from the topmost ones) in our example sentences have been phrases and lexical items of various kinds: NP's and N's, VP's and V's, AP's and A's, AdvP's and Adv's, PP's and P's, DegP's and Deg's, Aux's and D's. None of the constituents was a sentence (S). That is, all the examples so far have been simple sentences. A **simple sentence** is a sentence which contains no lower sentence (i.e. no 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.

However, a non-topmost constituent within a sentence can itself be a sentence. This is the case in (9), where the complement (more precisely: the object) of the verb *believes* is (not an NP but) a S. This lower sentence, the *buy* clause, (S2), functions as a complement clause within the higher sentence, the *believe* clause, (S1).

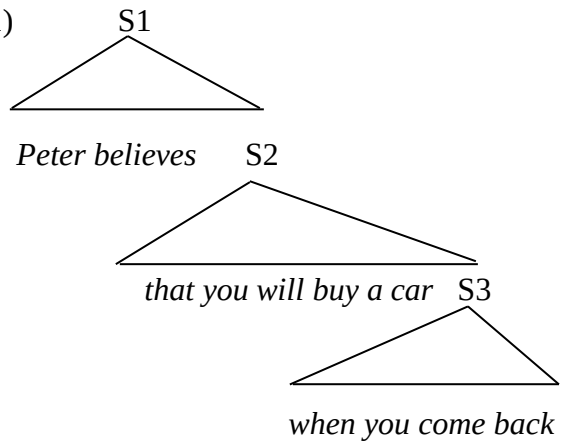
(9) [$_{S1}$ *Peter believes* [$_{S2}$ *that you will buy a car*]]

The phenomenon in which a constituent contains constituents of the same category as itself is known as **recursion**. For example, in examples (7e) and (7f) a VP contains a lower VP. In (9), however, recursion applies to the category S: here **sentential recursion** or **clausal recursion** happens.

A sentence containing a lower sentence (i.e. a clause) embedded in it is a **complex sentence**. (9) is a complex sentence, because it contains two sentences: a higher one, called the **matrix clause**: *Peter believes (that) you will buy a car*, and a lower one, called **embedded clause** or **subordinate clause** (or just **subclause**): *(that) you will buy a car*. It can also happen that a subclause has its own subclause, and so the upper subclause is the matrix clause of the lower one, 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*]]]

(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 …* .

Finally, consider the bracketed subclause in (12), *that we saw last week*. This is part of the NP *the shoes that we saw last week.*

(12) [$_{S1}$ *I've bought* [$_{NP}$ *the shoes* [$_{S2}$ *that we saw last week*]]]

Here the NP itself is the complement (more precisely: the object) of the matrix verb *bought*. The subclause modifies the noun *shoes* (it is an adjunct to it). Since *that* relates to (refers back to) *shoes*, it is traditionally called a **Relative Pronoun** and the subclause that contains it is a **Relative Clause**. More precisely, 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 three kinds of relationship between the coordinated clauses: **additive**, **adversative**, and **disjunctive**. The label "Conj" stands for **conjunction**: *and, but, or*.

(13) **Additive**: [$_{S}$ [$_{S}$ *Her daughter was a teacher*] [$_{Conj}$ *and*] [$_{S}$ *her son was studying arts*]]
(14) **Adversative**: [$_{S}$ [$_{S}$ *I asked him*] [$_{Conj}$ *but*] [$_{S}$ *he refused*]]
(15) **Disjunctive**: [$_{S}$ [$_{S}$ *I can go to them*] [$_{Conj}$ *or*] [$_{S}$ *they can come to me*]]

6.5 Sentence and utterance

Sentences have to be distinguished from utterances. A **sentence** is a string of words produced by the sentence forming rules of a language. So sentences belong to competence; they are ideal, abstract entities. For example, *Peter smokes cheap cigars* is an English sentence because it has the structure of an English sentence.

By contrast, an **utterance** is the physical realisation of a sentence in a situation of language use, i.e. in performance. 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. (See, for example, the spontaneous utterance by a native speaker in Exercise 15 at the end of this Unit.)

Such "incorrect" utterances are often made by native speakers but they do not matter because speaker-hearers automatically interpret them correctly on the basis of their competence. That is why such mistakes are usually not corrected and often not even noticed. (See, e.g. the spontaneous utterance by a native speaker in Exercise 14 at the end of this Unit.)

The utterances that humans 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, and begin a new one, or because they are interrupted by someone, or because they are shot dead before they have finished, etc. **Ellipsis** (= **omission**) of predictable constituents is quite common. The dialogue in (8) above, repeated here for convenience as (16), illustrates the ellipsis of a constituent (of the VP):

(16) *Will he go home quickly?*
Yes, he will ~~go home quickly~~

Exercises, problems, and other tasks

1 What is the difference between an adverbial and an adverb?

2 Identify the subject, predicate, 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 perform the function of Adverbial?

5 Identify the constituents and immediate constituents in (7b) and (7e).

6 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*

7 The following expressions are ambiguous. Disambiguate them by means of unlabelled tree diagrams:
(a) *old men and women*
(b) *a foreign language teacher*

8 What is the difference between a simple sentence and a complex sentence?

9 Find examples of recursion among the tree diagrams (1)–(11).

10 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*]

11 Identify and comment on the subclause in [*the dentist* [*who you've consulted*]].

12 Compare and establish the difference between the subclauses in *the dentist* [*who you've met*] and *my father,* [*who you've met*]. Note that the latter has a comma.

13 Draw simplified trees on the basis of the bracketing for the compound sentences (13)–(15).

14 Comment on this utterance: *It's uh … it's erm… not … I mean … actually well I've just sort of … er … sort of thought of going to … bed*

15 In which of the following utterances does ellipsis depend on the linguistic context?
(For which of them do you need to know the context in order to understand it?)
(a) *Anybody need a lift?*
(b) *No, tomorrow*
(c) *Looking for me, Terry?*
(d) *Sorry*
(e) *In Bristol*
(f) *Yes, I have*

Unit 7

Semantics: the study of meaning

7.1 Kinds of meaning

Semantics is the study of the meaning of linguistic units. 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 also called **propositional meaning** or simply **proposition**.

In the case of words, cognitive meaning is the contribution that the word (lexeme) makes to the cognitive meaning of sentences. The cognitive meaning of lexemes may be called **sense**.

The sentence in (1) describes a state of affairs, and as such it has a certain cognitive meaning.

(1) *The girl went to the garden*

There are other kinds of meaning, too. For example, (2) is different from (1) in terms of **stylistic meaning**, although it is **cognitively** identical with it.

(3a) and (3b), being a **question** and an **imperative**, cannot be treated as being either true or false, but they have a (questioning and a commanding) **speech act meaning**.

In (4), in addition to the cognitive meaning, there is considerable **emotive (affective) meaning**, 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: Hungarian e.g. *leány(zó); vette útját*)
(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)

Simplifying somewhat, it can be said that semantics is concerned with cognitive meaning, while the other kinds of meaning are the concern of pragmatics.

7.2 Approaches to word meaning

Two approaches to word (lexeme) meaning can be distinguished: the referential theory and the conceptual theory.

- The **referential theory** assumes that lexemes mean what they refer to (i.e. what they "name"). This view concentrates on the **referents** of lexemes. This seems correct in the case of proper names: e.g. *Buckingham Palace* refers to the building 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*) can be seen 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 world**, e.g. *fairy*; there are lexemes that refer to something that used to exist but no longer does today, e.g. *dinosaur* – but it cannot be denied 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 certainly do have meaning.

- Another 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 that is associated with the lexeme, i.e. a set of **semantic features**, which native speakers know. 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** (or **lexical decomposition)**. For instance, native speakers of English agree that the meaning of the noun *assassin* contains the following semantic features: 'person' who 'murders' 'important people'. The semantic features of the verb *die* are: 'animate being' 'becomes' 'not alive'. The semantic features of the noun *man* are: 'male', 'adult', 'human'. The semantic features need not be scientifically correct. Consider, e.g. the word *whale*, whose popular conceptualisation does not contain the feature 'mammal' (many speakers are not aware of this), although the feature 'mammal' is 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 you 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.

7.3 Sense relations between words

The contribution that a lexeme makes to the cognitive meaning of a sentence, – i.e. the cognitive meaning (the 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 involves a comparison of lexemes in terms of their senses. The relationships between lexemes established on the basis of their senses are called **sense relations**.

Simplifying again: it may be difficult to know about meaning, but it is relatively easy to study the sense relations.

- One of these sense relations 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 cognitively 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 two kinds are distinguished: homonymy and polysemy.

- **Homonymy** means that two or more **phonologically** and **orthographically identical** lexemes (i.e. lexemes pronounced and spelt identically) have completely different, unrelated meanings, e.g. *ball*$_1$ ('round object you throw') vs. *ball*$_2$ ('social event').

- **Polysemy** means that the meaning of one lexeme is metaphorically extended on the basis of some similarity, e.g. *leg* (of a human) vs. *leg* (of a table); *back* (of a human) vs. *back* (of a chair).

It may happen that the metaphorical connection that used to exist between such lexemes is lost, and so what started out as one **polysemous** item becomes two – or more – **homonymous** items; e.g. *horn*$_1$ (on the head of cattle) vs. *horn*$_2$ ('musical instrument') vs. *horn*$_3$ ('device in a car which makes a warning sound'). Originally, a *horn*$_2$ was made of a *horn*$_1$, and a *horn*$_3$ was a kind of *horn*$_2$, but most native speakers are no longer aware of this connection.

The presence of an ambiguous lexeme in a sentence makes the sentence ambiguous. This is shown in (8a) (homonymy) and (8b) (polysemy):

(8)a. *We waited by the bank* ('the financial institution') vs. *We waited by the bank* ('the riverside')
b. *The foot was visible* ('human foot') vs. *The foot was visible* ('of the hill')

- Another sense relation is **oppositeness** or **antonymy**, with subtypes called complementary, gradable and relational opposites. Please note that the term *antonymy* can also be used broadly, for all kinds of lexical oppositeness. Different types are distinguished:

  - **Complementary opposites** are lexemes in such a relationship that the negation of the meaning of one lexeme gives the meaning of the other, e.g. *dead* vs. *alive* (because 'not dead' means 'alive' and 'not alive' means 'dead').
  - **Gradable opposites** are (i) gradable lexemes, (ii) **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. In these cases, (iii) more of the one is less of the other, e.g. *smaller* means 'less large', *larger* means 'less small'.

    One member of gradable opposites is normally (iv) **unmarked**, the other is **marked**. It is the unmarked member that is used in (v) questions of degree, unless there is some good reason to use the other; e.g. *How old are you?* is unmarked, *How young are you?* is marked.
  - **Relational opposites** are lexemes that refer to symmetrically opposite aspects of the same situation, e.g. *employer* vs. *employee*. If someone employs you, you are their employee and that person 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 Doe is dead*
b. *John Doe is alive*

- The last sense relation is **hyponymy** or **logical inclusion**. This is the relation between a cognitively **superordinate** – more general – lexeme and the more **specific** lexemes that are **cognitively subordinated** to it. For example, *tulip*, *rose*, *daisy*, *carnation*, *lily*, *daffodil* 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'. When you define *tulip* or *rose*, the

word *flower* will be in the definition.

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 in (10).

(10) a. *tigress, hen, mare, cow, vixen, actress, queen, girl, maiden, widow, nun, woman, sister, mother*
b. *give birth, breastfeed*
c. *pregnant, buxom*, etc.

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

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

If a hyponym lexeme is replaced with its superordinate lexeme in a sentence, the original sentence **entails** the new one. One sentence entails another 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 example, (12a) entails (12b). If (12a) is true, (12b) is true, too. If (12b) is false, (12a) is false, too. But this does not work the other way round: just because (12b) is true is no guarantee that (12a) is true – she may have picked any other kind of flowers; and just because (12a) is false does not guarantee that (12b) is false – exactly because she may have picked any other kind of flowers, not daisies.

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

7.4 The cognitive meaning of sentences

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

- First, it depends on the **cognitive meanings** of its 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}, e.g. (13a) vs.(13b).

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

- Second, the cognitive meaning of a sentence depends on the **grammatical function** of the constituents, which is often – though not always – mirrored by the **order** of the sentence constituents, cf. (14a) and (14b).

(14) a. [$_{\text{Subject}}$ *The sheriff*] *kicked* [$_{\text{Object}}$ *the man*] ≠
b. [$_{\text{Subject}}$ *The man*] *kicked* [$_{\text{Object}}$ *the sheriff*]

- Third, 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 you can talk about sense relations between lexemes, so you can talk about **proposition relations** between sentences. These are relations between sentences on the basis of their cognitive meanings.

- Sentences can be **synonymous**, in which case they are (each other's) **paraphrases**. The synonymy of sentences may result from **lexical synonymy**, as in (7a) and (7b) above.

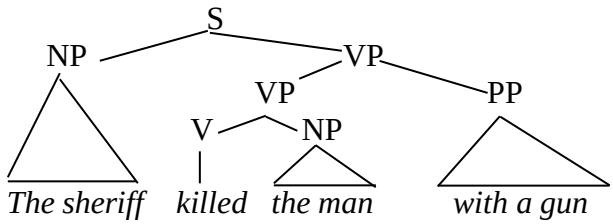
Sentential synonymy can also be achieved **structurally**, as in the active–passive pair in (16a) and (16b).

(16)a. *The dog crossed the road*

b. *The road was crossed by the dog*

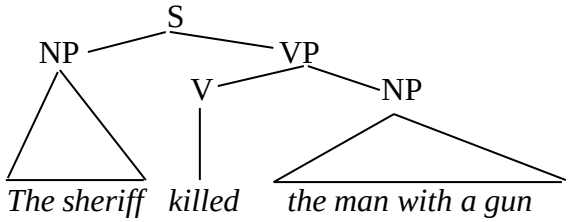
- Sentences can also be **ambiguous**. This means that two sentences are composed of the same words in the same order, but they have different meanings. This again can have lexical reasons (one of the words is ambiguous), as in (8a) and (8b) above. Or, ambiguity can be caused by structural differences,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

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



SVOA pattern

b. *The sheriff killed the man with a gun* '... fired the man who had a gun'



SVO pattern

Furthermore, ambiguity may also result from the different (**semantic**) **roles** that a particular constituent can play, see (18a) and (18b). In (18a), the constituent *the lamb* plays the role of **Agent** (= the doer of an activity), while in (18b), it plays the role of **Patient** (= the person or thing undergoing an activity).

(18) a. [$_{Agent}$ The lamb] is ready to eat (‘The lamb will eat’)

b. [$_{Patient}$ The lamb] is ready to eat (‘Somebody will eat the lamb’)

The remaining two proposition relations, **incompatibility** and **entailment**, are never brought about by structural means or by different functions; they are solely due to the contribution of lexemes, as the example pairs of (9) and (12) show, respectively.

Points 7.3 and 7.4 summarised

(19)

<table>
<tr><td></td><td>words</td><td></td><td>sentences</td><td></td><td colspan="2">also caused by
factors having to do with</td></tr>
<tr><td rowspan="2">①
(cognitive)
synonymy</td><td rowspan="2">synonymous
lexemes

girl ~ damsel</td><td rowspan="2">→</td><td rowspan="2">– sentential synonymy
– (cognitively)
synonymous sentences
(= paraphrases)
The girl went
to the garden
~ The damsel made her
way to the garden</td><td rowspan="2">←</td><td>structure
The dog
crossed the
road
~
The road was
crossed by the
dog</td><td>role</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>②
ambiguity
(2 types)</td><td>ambiguous
lexemes
homonymy: ball,
bank, bat
polysemy: leg,
head, back, point</td><td>→</td><td>ambiguity
The ball was not
expensive
We waited by the bank
That’s not a very good
point</td><td>←</td><td>The sheriff
killed the man
with a gun</td><td>The lamb
is ready
to eat</td></tr>
<tr><td>③
antonymy
(3 types)</td><td>antonymous
lexemes
dead ↔ alive
large ↔ small
employer ↔
employee</td><td>→</td><td>incompatibility
(example with
complementaries)
JD is dead ↔
JD is alive</td><td colspan="3"></td></tr>
<tr><td>④
hyponymy</td><td>cognitively super-
/subordinate
flower: rose,
tulip</td><td>→</td><td>entailment
Mary picked daisies
— Mary picked
flowers</td><td colspan="3"></td></tr>
</table>

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 + 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. Give at least two similar examples.

4 What is the difference between homonyms, homophones and homographs? (cf. p. 38)

5 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*

6 What are *suffix* and *prefix* the hyponyms of?

7 Comment on the proposition relation between *Yesterday it rained* and *It rained yesterday*.

8 The following sentence is both structurally and lexically ambiguous: *I saw him walking by the bank.* Paraphrase all its meanings.

9 Turn the labelled tree diagrams in (17a) and (17b) into labelled bracketings.

10 What relation is there between the A sentences and the B sentences:

(a) A. *The police wounded the burglar* B. *The burglar is injured*
(b) A. *The house is red* B. *The house is not white*

Unit 8

Pragmatics: the study of language use

8.1 Pragmatics

Pragmatics is the study 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 as they are used in particular situations; by situations we mean linguistic and physical contexts.

Pragmatics is not easy to separate from semantics. It is to some extent an arbitrary decision where the line is drawn between them. The central concerns of the two, however, stand out fairly clearly. While **semantics** primarily examines the **cognitive meaning** (the **sense**) 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 deal with the following two aspects of language use: (a) the role of context and presuppositions, (b) language functions and speech acts.

8.2 The role of context

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 example, the lexemes BANK$_1$ ('riverside') and BANK$_2$ ('financial institution') are homonyms, and they may cause ambiguity (the sentence *We waited by the bank* has two meanings), but they are not normally confused when they occur in particular linguistic contexts, as in (1a) and (1b):

(1)a. *The right bank of the Danube 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, it gets disambiguated when uttered in different discourses, e.g. (2a) and (2b).

(2)a. *There were two people waiting for him: 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 example, when you see the word BANK written on the front of a building in a city, you will know that what you see is not the edge of a river.

8.3 Deixis

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 **deictic expressions**. For example, 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*

8.4 Presuppositions

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. *You brother wants to see you*
b. *Your brother doesn't want to see you*

(5) *You have a brother*

8.5 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** (= **propositional,** or **descriptive**) **function**: the communication of a state of affairs, e.g. *Today is Monday* or *The piano is in the middle of the room*.
- The **expressive** (= **affective**) **function**: the expression of the speaker's attitudes, feelings, emotions, e.g. *Damn!* or *Shit!* or *Oh!*
- The **directive function**: influencing the hearer's behaviour or attitude, e.g. *Come here!* or *Could you lend me two thousand dollars?*
- The **phatic function**: 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, e.g. *The word "violin" is of Italian origin* or *"Violin" is of Italian origin* or *Violin is of Italian origin*.
- The **poetic** (= **aesthetic**) **function**: the use of language primarily for its own sake, i.e. for the pleasure that 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 classification of various types of language use has been provided by **speech act theory**. The central notion in this theory is **illocutionary act**, which is the act that the speaker performs in and while saying an utterance. (The word *illocution* derives from "in+locution", i.e. 'in speaking'.) 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 explicit if **performative verbs** are used. A performative verb is so called because it explicitly performs an illocutionary act, i.e. it explicitly expresses the speaker's communicative intention; 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 *üdvözöl* '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*
(7) *I <u>declare</u> the meeting open*
(8) *<u>Üdvözlöm</u>*

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. In (9), (10) and (11) 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. Instead of the (a) sentences you may utter the (b) sentences; with these you perform the same illocutionary acts as with their (a) counterparts.

(9)a. *I (hereby) <u>assert</u> that he speaks excellent English*
b. *He speaks excellent English*
(10)a. *I (hereby) <u>suggest</u> that you should leave*
b. *You should leave*
(11)a. *I (hereby) <u>promise</u> 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 the sentence *The door's too low*. This can 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.
- Second, 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, e.g. 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.

- Third, the interpretation of the function or illocutionary act represented by an utterance requires knowledge of the situation (the physical and linguistic context) in which the utterance is made. Recall, e.g. *Can you play the piano?* in Unit 2. In other words, you simply cannot say what the function or illocution of sentences is if you take them in isolation from the context or situation in which they are uttered.

Exercises, problems, and other tasks

1 What are the deictic expressions in the following? *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*

3 In what functions can you use this 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 Discuss possible interpretations of

(a) *You've left the door open*
(b) *My mouth is parched* (said to a barman)
(c) *My mouth is parched* (said to a doctor)

Unit 9

Language Variation

9.1 The identity and variability of language

A natural language is not one homogeneous code. Any natural language 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 varieties side by side is called **language variation** (i.e. **synchronic variability**).

Moreover, the coexisting varieties are in constant change along the dimension of time: this is called **language change** (i.e. **diachronic variability**).

The first question that has to be discussed is the problem of **language identity**. This is the question of how you decide whether two linguistic codes are two separate languages or just varieties of one. Many think that the criterion is **mutual understandability** – but this often breaks down between codes that are regarded as belonging to the same language (e.g. northern 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 do understand each other's speech fairly well). Therefore, mutual understandability is not a safe criterion.

Language identity is not a linguistic but 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 truth in the humorous saying, "a language is a dialect with an army and a navy".

English, like any other natural language, is an **abstraction**, 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, this chapter is 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, pidgins and creoles

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

- **Regional dialect** or just **dialect** is the variety of a language which is used in a certain geographical area. Dialects may differ in vocabulary, pronunciation, even morphology and syntax. The boundaries between dialects are not as clear-cut as political boundaries. They can be established by identifying the linguistic features characteristic of the area. The line marking the limit of the spread – **distribution** – of a linguistic feature on a map is called an **isogloss**. For example, 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 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 not restricted to the geographical area where it was originally used, but is associated with people who are educated and are at the top of the **socio-cultural scale**, no matter where they live. The standard is no longer a regional dialect: it is 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 although the standard variety has a higher social prestige, it is not linguistically better than the other varieties. For example, **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 has historical, political, cultural, economic reasons: the significance of the capital 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 in pronunciation. The prestige type of pronunciation of Standard British English is **Received Pronunciation**, or **RP** (so called because by the 19th century this had become the socially acceptable pronunciation in "polite society" in England). The pronunciation associated with Standard American English is **General American**, or **GA**. Since the way in which a language is pronounced is an **accent**, RP and GA are also accents. There are dialectal and foreign accents as well, so one can speak English with a Yorkshire accent, or with a German accent, with an RP accent, etc.

Standard British English is the language of 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 (i.e. regional, local) features on the one hand, and with sociolectal 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** ("substandard") 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 guy has done it*, *He don't know nothing*, *I ain't got no problem*. The last two examples illustrate **double** or **multiple negation**, a non-standard **sociolectal** feature used by less educated speakers in very different geographical areas. Examples of this sort, however, are not incorrect linguistically: 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" should perhaps be avoided and "non-standard" used instead.

- A **pidgin** or **pidgin language** is a user-related language variant that is usually the simplified version of a European language (mostly English or French), which contains features of one or more local languages, and is used for occasional communication between people with no common language, in West Africa or in the Far East, e.g. Melanesian Pidgin English (Tok Pisin). A pidgin is not anyone's native language, but it can become the native language of a community, e.g. through

intermarriage between people from different linguistic backgrounds, 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 a **creole**.

- 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 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

- The **medium** of language use, i.e. **speech** vs. **writing**, conditions the first type of use-related variation. The language you speak is generally different from the language you write. When you write, you 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 and read them again if necessary. But a finer distinction of media is required because there are different subtypes of speech and different subtypes of writing, and these differences cause corresponding linguistic differences. For example, the language used in face-to-face talks differs from the language of public lectures, which in turn is very different from the language of telephone conversations. The language of text messages or emails is different from the language of personal letters, though both are written varieties.

- **Style** is the second use-related variation, which is conditioned by the 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 or children; they may talk to someone who they have never seen before or to an old friend of theirs, etc. A **neutral** or **unmarked** style can be recognised, which does not show any obvious features brought about by relative social status and attitude. On either side of this neutral style sentences can be distinguished which are markedly formal or informal. Compare the sentences in (1).

(1)a. Very formal: *I was wondering if it might be at all possible for you to switch off the light*
b. Formal: *I wonder if you'd mind switching off the light*
c. Neutral: *Would you please switch off the light?*
d. Informal: *Switch off the light, will you?*
e. Familiar: *Switch it off, OK?*

Formal style is 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.

Familiar is a very informal style, which may involve the use of **non-standard** features, four-letter words, and slang expressions.

Slang is 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 e.g. army slang, school slang, etc. can be distinguished. 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'), *gold brick* ('swindle'), but some may pass into ordinary colloquialism – i.e. into (informal) standard. For example, *slag smb off* ('criticise') was a British slang expression that is now a standard (informal) item.

When you use language, you 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 it would be ridiculously inappropriate if you said it to a friend in your home (unless you wanted to sound humorous).

- **Register** is the third type of use-related language variation, which is conditioned by the **subject matter** in connection with which the language is being used. Each field of interest, activity or occupation is associated with a **special** (**specialised**) **vocabulary**, and it is mainly these vocabulary differences that underlie the different registers. Thus you can talk about the registers of sports, religion, medicine, computer engineering, cookery, weather forecasts, etc. The word *shot-putting*, e.g. is hardly ever used outside the sports register; the word *blackboard* is only used in the context of school teaching.

When a register of a field is full of technical terms which those with no training in that field cannot understand, it is a **jargon**. Think of the jargon of computer engineers, or the jargon of linguists.

Argot or **cant** is criminals' jargon.

Since the most frequent and most favourite topics of one's speech or writing are obviously related to one's occupation, registers are partly **user-related**, too. (Note that "register" is sometimes used in other senses, mainly to mean "style".)

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**. A doctor switches codes when (s)he speaks of a bone as *tibia* to her/his colleagues in the hospital and as *shinbone* to her/his family.

It can happen that two distinct varieties of a language co-occur in a speech community. In such situations one has **high social prestige** (e.g. Standard English, which is learnt at school, used in church, on radio programmes, in serious literature, and generally on formal occasions), while the other has **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?

2 What is an isogloss? How can you define a dialect in terms of isoglosses?

3 Discuss "standard variety". How does it emerge?

4 Is a standard more correct than the other (non-standard) varieties? More aesthetic? Purer? Is it more versatile and flexible than the other varieties?

5 In a non-standard variety of English *Where ya been lately?* replaces the standard *Where've you been lately?* but **Where've been lately?* **Where been lately?* **Where ya lately?* are impossible. What does this show? Are there no rules in non-standard varieties?

6 Find differences between British and American English lexis, phonology, morphology, syntax, and orthography.

7 Explain RP, GA, and the term "accent". Does the Queen speak with an accent?

8 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*

9 (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*

10 On the basis of the data in the previous exercise, identify some characteristic features of informal and formal style.

11 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 up again*
(i) *Be seated*
(j) *They chucked a stone at the cops, and then did a bunk with the loot*

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 **diachronic variability**.

English has undergone considerable changes in the three main periods of its history: these periods are Old English (OE), from about 450 to 1100; Middle English (ME), from cca. 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 – the **Angles**, **Saxons**, and **Jutes** – settled down in what is today England, in the 5th century. They spoke Germanic dialects, from which the various Old English dialects developed directly. Today OE texts are largely unintelligible even to English speakers. 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 ðǣre ðēode mennisc*
then said him someone that they of England were and that that country's people
'Then someone told him that they were from England and that the people of that country

swā wlitig wǣre
so comely were
were so comely'

(The letter þ (called 'thorn') and the letter *ð* (the 'barred d', called 'eth') were both pronounced as /ð/ between vowels, and as /θ/ elsewhere).

This OE example differs from its Modern English counterpart in many respects.

- 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* ('handsome') have disappeared from use. The word *hī* means 'they'.
- There are considerable syntactic differences: e.g. in the first clause, the indirect object precedes the 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 English 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 the language had lost most of its inflections, also, the quality of many of its original sounds had changed considerably. From the main dialects of ME 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, from Chaucer's *The Parson's Tale* (end of the 14th c.), is not too difficult to understand for present-day readers.

(2) *Of the roote of contricion spryngeth a stalke that bereth braunches and leves of confessioun,*
of the root of contrition springs a stalk that bears branches and leaves of confession
and fruyt of satisfaccioun
and fruit of satisfaction
'From the root of contrition springs a stalk, leaves of confession, and fruits of satisfaction'

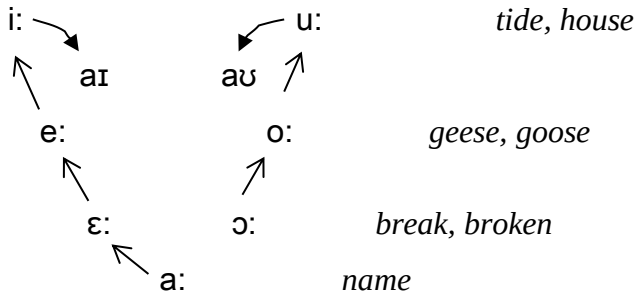
By about 1500, English had become a language which (though it now may look archaic), can be recognised as English, and understood without much difficulty even today. This is the beginning of the **Modern English** period.

10.2 Examples of changes

The changes that have occurred in English during the past 1500 years have affected all parts of the language.

- 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 the long vowels was replaced by the next higher vowel. All this is shown in (3), with examples in (4).

(3) The Great Vowel Shift (1400–1600)



(4)

Middle English	GVS	Later diphthongisation	Modern English	
tiːd	iː → aɪ		taɪd	*tide*
huːs	uː → aʊ		haʊs	*house*
geːs	eː → iː		giːs	*geese*
brɛːkən	ɛː→ eː	→ eɪ	breɪk	*break*
goːs	oː → uː		guːs	*goose*
brɔːkən	ɔː→ oː	→ əʊ	brəʊkən	*broken*
naːme	aː → ɛː	→ eɪ	neɪm	*name*

- There have been considerable **morphological changes**. In Old English, there was a rich **conjugation** system: 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 (this is called **declension**). There was **concord** (= agreement) between nouns and their adjectives. This is shown in (5), illustrating the declension of *sē gōda wind* ('the good wind').

(5)

	Singular	Plural
Nominative	*sē gōda wind*	*þā gōdan windas*
Accusative	*þone gōdan wind*	*þā gōdan windas*
Genitive	*þæs gōdan windes*	*þāra gōdra winda*
Dative	*þæm gōdan winde*	*þæm gōdum windum*

This elaborate system of declension had disappeared by the end of the Middle English period.

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

(6)

	year 1200	year 1500
OV ('*him see*' type)	53 %	2 %
VO ('*see him*' type)	47 %	98 %

While in Old and Middle English **inversion**, which is involved in the formation of questions, could apply to all verbs, in the Modern English period the inversion rule was gradually changed to apply only to **auxiliaries**. In the *King James Version* of the Bible (1611) you still find both main verbs and auxiliary verbs inverted with the subject, as in (7). Later this possibility ceased to exist for main verbs and ***do* insertion** became the established way of forming questions when no auxiliary was present.

(7) 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, huge **lexical changes** have taken place 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, 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 (8) shows you some of the French loanwords in English and the fields they belong to.

(8) Government: *government, royal, parliament, authority, prince, duke, 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*

The core vocabulary of English is still Germanic. Today, of the 5000 most frequent words, 40% 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 find that the ratio of words of English origin is considerably higher:

(9)

most frequent words	source language:			
	English	French	Latin	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
- **semantic shift**: the word loses its earlier meaning and acquires a new one

These processes are illustrated in (10).

(10)

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'

Exercises, problems, and other tasks

1 Analyse the word order in the following line:
A*nd Seaxan þā sige geslōgan* (and Saxons the victory won) = ‘And Saxons won the victory.’

2 What were the main linguistic tendencies in the period of Middle English?

3 Give an outline of what happened in the course of The Great Vowel Shift.

4 Questions like *What think ye?* were possible until the end of the Early Modern English period. What do these show?

5 What words of French origin are used to refer to the meat of these animals: *pig, cow, calf, sheep, deer?*
What explains that there are different lexical items for the animal and its meat?

6 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.

7 The words *aunt* and *mete* (i.e. meat) used to mean ‘father’s sister’ and ‘food’, respectively.
What kinds of semantic changes do they exemplify?

Appendix

IPA symbols for **RP** used in this book

Consonants	Vowels
/p/ pen	/æ/ bad
/b/ back	/ɑː/ calm
/t/ tea	/e/ bed
/d/ day	/ɪ/ ship
/k/ key	/iː/ sheep
/g/ get	/ɒ/ pot
/m/ mouse	/ɔː/ caught
/n/ nice	/ʊ/ put
/ŋ/ si*ng*	/uː/ boot
/l/ led	/ʌ/ cut
/r/ red	/ɜː/ bird
/f/ fat	/ə/ *a*go
/v/ *v*iew	
/θ/ thing	/eɪ/ make
/ð/ then	/aɪ/ bite
/s/ sun	/ɔɪ/ boy
/z/ zero	/aʊ/ now
/ʃ/ ship	/əʊ/ = /oʊ/ n*o*te
/ʒ/ rouge	/ɪə/ here
/h/ hot	/eə/ there
/tʃ/ chair	/ʊə/ poor
/dʒ/ jump	
/w/ wet	
/j/ yes	