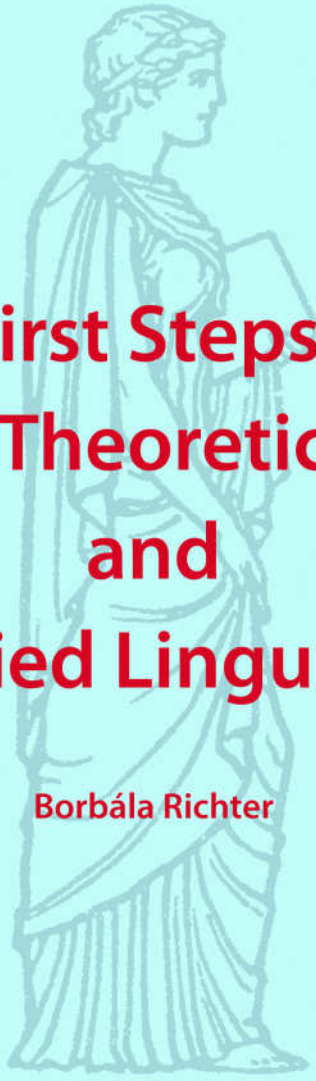


First Steps in Theoretical and Applied Linguistics

Borbála Richter

BÖLCSÉSZ KONZORCIUM

Értékünk az ember
Magyarország célba ér
NEMZETI FEJLESZTÉSI TERV
KODOLÁNYI JÁNOS FŐISKOLA
2006

First Steps in Theoretical and Applied Linguistics

Borbála Richter

Bölcsész Konzorcium
HEFOP

First Steps in Theoretical and Applied Linguistics

Borbála Richter

Bölcsész Konzorcium • Budapest, 2006

SZERZŐK:

Górász Judit
Horváth József
Kenesei Andrea
Kiszely Zoltán
Lázár A. Péter
Richter Borbála
Szemere Pál
Szitó Judit

Szerkesztette: *Richter Borbála*
Nyelvi lektor: *Keith Hardwick*
Szakmai lektor: *Hollóssy Béla*
Szakmai konzulens: *Nikolov Marianne*
Rajzok: *Szitó Judit*
Műszaki szerkesztő: *Nagy Károly*

Table of contents

Foreword

The idea for this book goes back about ten years, to my first impression of higher education in English in Hungary, when I felt that the discrepancy between the standards of the excellent textbooks available and the needs of the local target audience were subtly at odds. Both the academic English used and the assumptions about prior subject knowledge seemed to me conducive to dampening rather than stimulating the interest of a significant proportion of the readers. So when it became possible to apply for a grant to write a book for the new Bachelor of Arts programme in English, it occurred to me that here was the opportunity to fill the perceived gap with a genuinely introductory linguistics textbook written in accessible English, presuming no prior knowledge of linguistics and aimed at 'passing on the flame', the passionate interest in their area of expertise of the writers. It filled me with gratification to find that the consortium distributing the HEFOP grants and the colleagues I asked to collaborate agreed.

The authors of the various chapters all specialise in their field and teach their topic at institutions of higher education here in Hungary. They are therefore particularly suited to the task set them: to introduce their field in a way that makes it accessible to the readers, through the carefully calibrated use of English, while keeping to high academic standards in the content. Experienced lecturers using the book as a textbook will find many a point whence they can move on to deepen their students' understanding. Without limiting the way in which the book may be used, I would suggest it as 'pre-reading' to be set before the student goes to class, as extra reading about fields that may not form part of the curriculum, as a textbook providing the necessary minimum yet wide-ranging knowledge of the field of linguistics as a whole, and as supplementary reading in English for students of linguistics in Hungarian with a working knowledge of English.

As anyone who has ever written a textbook knows, the list of people who should be thanked is very extensive. I would, however, like to mention a few on that list. First of all, the authors, for contributing their knowledge and enthusiasm, and Marianne Nikolov, without whose expert participation I would not have dared to embark on the project. Secondly, Keith Hardwick and Károly Nagy, my infinitely patient and helpful partners in the actual realisation of the book. Thirdly, Béla Hollósy, whose constructive comments and encouraging words of praise gave me the much needed second wind to complete the project. Finally, thanks are due to the Consortium who awarded us the HEFOP grant; Csilla Sárdi of Kodolányi János University College who nudged me into submitting the application; and Éva Stephanides of the Dept of English at KJUC who did everything she could to help her members of staff involved in the project take part despite their already heavy workload, and last, but quite emphatically not least, all the family members of the participants, whose support and endurance were essential. To all the others who contributed in many ways, but whom I have not the space to mention, I trust that the book itself will serve as an expression of my gratitude.

Borbála Richter, editor

Borbála Richter
Kodolányi János University College
Department of English Language and Literature

1. The First Step

Introducing Language and Linguistics

Language is such a central and natural part of our lives that we tend to take it for granted until we come across the wonder of a tiny child becoming able to speak, or until we turn our attention to it as an object of scholarly inquiry. Language is intimately and intricately linked to our ability to think, to be aware of our own existence and of the perspectives of others, to share information and feelings – and to investigate its own nature. This, then, is what linguistics is: the scientific investigation into human language, in all its many forms and aspects.

There are many of these aspects, as we shall see, and there is a branch of linguistics for each approach to the examination of language. Before we mention some of those branches, which will form the major part of this book, let us discuss the nature of language, and the nature of scientific inquiry.

1.1. Language and its features

Language is first of all a communication system. This means that it is a system specialised for communication, unlike, for example, the purring of a cat. As we shall see, it is a very complex phenomenon and very difficult to define precisely. What follows is a list of descriptive characteristics, called design features.

1.1.1. Arbitrariness

In the case of human language, words mean something; they refer, for example, to events or objects in the world. Words are ARBITRARY. In other words, the form of the sign is not related to its meaning. Why is a dog called *dog*? It never told us its name, nor does the word sound like a dog or represent it in any way. Indeed, in Hungarian, *a dog* is *kutya*. Some words do resemble sounds, but even words that refer to sounds are not the same in all languages: in English, dogs say *woof*, *woof* or *bow wow*, while in Hungarian they say *vau vau*. (To find what dogs say in other languages, go to http://www.georgetown.edu/faculty/ballc/animals/dog.html). Even the speech sounds of a language are randomly picked – although all humans have the vocal organs necessary to make the sounds of all languages, individual languages do not use all of these. English, for example, uses about 40 significantly different speech sounds, or PHONEMES These are, in turn, arbitrarily represented by letters. In fact, sometimes one sound [k] is represented by more than one letter (*k* in *kite*, or *c* in *cut*), or different sounds [i] and [ai]are represented by the same letter (*i* in *dig* or *dine*). The way in which sounds are combined into words is also arbitrary, but not haphazard. There are principles

that systematically restrict how this happens. Technically we would say that PHONEMES are constrained in the way they may be combined into MORPHEMES. We can all recognise the word *taxi* and we can all predict that **txia* is not likely to be a word in English or Hungarian. Languages are patterned and their users follow principles – even if they cannot define these 'rules'. Would you have been able to specify a rule that states that Hungarian and English do not use *tx* at the beginning of a word?

1.1.2. Infinite combinations

So arbitrariness does not, mean randomness. If you think about it, this is obvious – you would never dream of saying something like:

** A see tree I big can beautiful and.*

Although as a student of English you might have said:

**Can see I a tree big and beautiful?*

(In this book, when you see an asterisk (*), you know that what follows is ill-formed or ungrammatical, or a reconstructed form of a word.)

Languages organise and combine sounds and arrange words according to principles to create infinite possibilities. This very important characteristic of human language is called duality of patterning – also known as double articulation – and means that a conveniently small number of meaningless elements can be combined to form a large number of meaningful elements. To see how duality of patterning works, let us look at the word *dog*. Individually /d/, /ɒ/ and /g/ are meaningless but together they are meaningful. If we replace one of these elements with another meaningless element, e.g. /d/ with /f/ we get another meaningful word: *fog*. We can do this to coin new words, which may or may not become commonly understood. An example which was not a word until recently but which you probably all recognise – although your grandmother may not – is *blog*, which replaces [d] or [f] with [b] plus [l].

Linked to duality of patterning is the productivity of language. Here, the focus is on the generation of meanings, for example, by recombining existing MORPHEMES, that is, the smallest unit of meaning. From *jog* + *ing*, we get *jogging*. Productivity is usually syntactic. By combining words, we could make the sentence: *I jog*. If we then replace *I* with *they*, we get a similar sentence, *They jog*, with a change of subject (or agent in this case). Clearly, there are limitations on the kinds of sequences we can produce and these form the GRAMMAR of the language. By applying these rules, humans can generate a limitless number of combinations. So at every level of the analysis of language we find a limited number of elements which we can combine with a limited number of rules to generate an infinite number of possibilities.

You, too, are able to say something novel, something that has never before been said or heard, but that a native speaker of the same language would be able to understand. This is one way in which humans differ from animals, which may have to use certain forms of communica-

tion in response to certain stimuli. Human beings can say anything they like in any context. This includes the 'saying the wrong thing'; speaking about something that has already happened, inventing something, or telling lies.

1.1.2.1. Constituency and recursiveness

These off-putting technical terms are linked to the fact that there is no theoretical limit to the length of words, or the number of words, or the number or length of sentences. All linguistic structures are made up of constituents (words or groups of words that function as a unit). More complex units can enter structures where simpler ones are possible. For example,

He entered the room.	He = subject
The man entered the room.	The man = subject
The tall, dark man with the red tie entered the empty room of the big house.	The tall etc man with the red tie = subject

If you like, you could continue by, for example, specifying in more detail what kind of room he entered.

Another feature that can lead to endless sentences is called recursiveness. This property of language lets you apply processes over and over again.

You know what linguistics is.
I know that you know what linguistics is.
He knows that I know that you know what linguistics is.
We all know that he knows that I know that you know what linguistics is or is not or could become.

Or you could do this in a more varied way:

We are all aware that he hopes that I think that you know what linguistics is.

1.1.3. Born to speak

From this it should be clear that no one can learn a language by memorising all the words of a language, or all the sentences of that language. Even if someone *did* learn all the words, which is theoretically possible, it would still not be enough. They would still not be able to speak since there cannot be a list of all possible sentences. We need to look for some other explanation of how humans are able to learn to speak. This includes investigating the brain to see if we are born with some special mechanism for learning language. One argument for this is the commonplace but astounding fact that children acquire language easily and quickly.

The major stages in the development of language also take place in a similar order irrespective of the child's language environment. Another possibility is that language learning is essentially just one form of general learning. The technological revolution caused by the development of computers has led some researchers to suggest that we can use computer models to see if by feeding in language input, for example, a certain number of English verbs and their past tenses, the model can produce output for a form that it was not fed, for example, a regular past tense, like *talked*, or even an irregular past tense, like *sang*. Even if these attempts have not led to clear-cut conclusions, they have led the supporters of the "born to speak" hypothesis to revisit the issue. Whatever the final outcome of this debate may be, it is clear that languages themselves are not inborn. Human beings have to acquire their native language from other speakers: languages span generations and their speakers learn them in the course of their socialisation. They also learn to rely on context to help them understand language. The context of a sentence can be the words or sentences directly before or after it or it could be the broader context of who is saying what to whom, where and when and why.

1.2. Linguistics describes, it does not prescribe

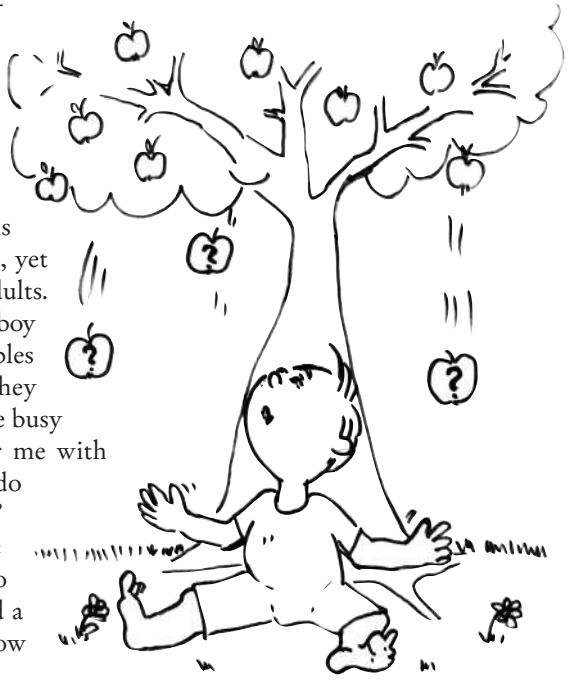
When students start studying linguistics they often think they are going to learn about what is correct or incorrect in a particular language; they expect linguists to be authorities on these questions. By now you will know that this is not the case. Linguistics is the science of language and linguists are scholars who study language – and linguistic behaviour – as phenomena to be studied. Unlike the physicists, linguistics studies human behaviour. Consequently, we have to be aware of our own attitudes and preconceptions. It is important to realise that these may hinder our progress towards understanding how language works. Our ideas about speech that differs from our own may influence the way in which we judge people. Consider the sentence

I ain't never smoked no cigar, man?

As students trying to master Standard English for an examination, you may know that a teacher would say that this sentence is 'wrong'. You could jump to negative conclusions about people who speak like this. On the other hand, you might already have negative ideas about a certain group and so you would disapprove of the language used because such people speak like this. Can you think of a similar example in Hungarian? At this point it is just important to realise that for linguists, all kinds of language use and all kinds of language form the subject of objective study. Although languages share some features, they also vary considerably. Each person has a repertoire of ways of speaking, the choice depending on the situation and the reason for speaking. The way we speak also reveals a great deal about ourselves, or social and geographical origins, for example. One of the functions of 'small talk' is to give the people we are speaking with the opportunity to draw some tentative conclusions about us. We all tell the world who we are when we speak. (You may have noticed that I usually mention speaking, and not writing. This is because writing, important though it is, is second in line after speaking.)

1.3. The scientific method

Even in this brief introduction, various ways of investigating language have been mentioned. From what you have already read, you now have a better idea of what language is. You know that linguistics is the scientific study of it. How does one study something scientifically? What is meant by scientific method? What will studying linguistics mean?



When we go about the everyday business of studying, it is easy to see this as a process of learning what 'those who know' teach us. Yet we should always remember the crucial importance of questions, since they are what move science, in its broadest sense, forward. Asking questions is a skill well-developed in every child, yet often lost by the time we become adults. It is so easy to imagine the little farm boy asking, "Granny, Granny, why do apples fall down from trees? Why don't they fall up?" and the irritated reply of the busy care-giver who says, "Don't bother me with your silly questions. Go out and do something useful, like chop wood!" Fortunately, the little boy – Isaac Newton – grew up to be a man who posed that question again, and found a way to answer it. Now we all know about the law of gravitation.

From this simple example, we can learn a great deal about the scientific method, for little Isaac's question shows that he had noticed that something happened with regularity. He then described this regularity, and looked for a reason to explain it. These three elements come up again and again in any academic field of study: observation, description, and explanation. Before he published his theory of gravitation, Newton also read about and studied what others had done in the past. Scientific knowledge is cumulative, building always on the work of others. Not that this path is simple and orderly: hypotheses are proposed, studied, argued about, verified and/or rejected. Even Newton's theory, which explained all the observed facts and made predictions that were tested by others and found to be correct, so that for many people over many centuries it was considered "the Truth", eventually reached a point where new, more accurate instruments started to allow new facts to be observed that could no longer comfortably fit the theory. Albert Einstein then proposed his Theory of Relativity, explaining the newly observed facts and making predictions, which have since been tested and found to be accurate. Of course, the universe itself has not changed. The

important point to remember is that the best of theories have been found to be incorrect. Mankind's store of knowledge increases with replicating and testing and sometimes repudiating the findings of others. This is true as much for Linguistics as for Physics and indeed even for something as everyday as learning a language or finding out what the funny noise is that your car makes when it is cold.

1.4. Questions which linguists pose

What does all this have to do with Linguistics, in particular? Well, we find all the elements of scientific method here, too. Also, Isaac used language to pose his question to his grandmother, and linguistics is always concerned with language and often with language used in interaction between people. Let's look at Isaac's words again and see what kind of questions we can ask. I have composed 15 groups of questions in all, one for each chapter that follows:

- What Isaac asks is a question, and this can be seen in the way he puts his words. So, what is a question? Can we ask a question in any language? How can we form questions and other sentences? See the chapter on *Syntax*.
- His grandmother hears an uninterrupted stream of sounds yet she understands this as words. How is this possible? See the chapter on *Phonetics and Phonology*.
- The word *fall* includes the idea that something goes in the direction dictated by gravity. How do words 'mean' anything? See the chapter on *Lexical Semantics.*
- The word *fall* often occurs together with the word *down*. What other words does it occur together with? How can we find out what real users of language say? See the chapter on *Corpus Linguistics*.
- If Isaac had been a little Hungarian boy learning English, he might have said, "**Apples are under the tree.*" Why would this be a typical mistake for a Hungarian language learner? See the chapter on *Second Language Acquisition*.
- In our century, if you want proof of your proficiency in a foreign language, you can take a language exam. Would it be fair to set a question testing whether the candidate knows that *fall* occurs together with *down*? See the chapter on *Testing*.
- In our example the boy calls his grandmother *Granny*, but he could also call her *Nana*, as many children in Great Britain would. Why are different words used in different families? See the chapter on *Sociolinguistics*.
- When Isaac was a little boy, English sounded rather different and his question may actually have been put differently. How and why do languages change? See the chapter on *Historical Linguistics.*
- While formulating his question, Isaac might have repeated the word *Granny*, to gain time to make up the question. What is going on in our brains when we perceive or produce speech? See the chapter on *Psycholinguistics*.
- Isaac expected a reply to his question. Why do questions and answers come in pairs? See the chapter on *Discourse Analysis and Text Linguistics*.

- When Isaac grew up and answered his question for the benefit of all of us, he wrote his book in Latin, which was the language of science at that time. Isaac had to know at least two languages and he had to formulate his ideas in a language he did not use for everyday communication. Is proficiency in two languages normal? See the chapter on *Bilingualism*.
- For us to be able to read what Newton said, we would have to read a translation. Would that translation be the equivalent of the original? See the chapter on *Translation*.
- Isaac knows that he has to add an '*s*' to the word *apple*. Does the '*s*' have meaning on its own? See the chapter on *Morphology*.
- If we don't know what *fall* means, and we reach for a dictionary, how will we find it and what information can the dictionary give us? See the chapter on *Lexicography*.
- Instead of giving Isaac an explanation, his grandmother told him to go and chop wood. What do we really do through talk? See the chapter on *Pragmatics*.

1.5. Taking the next step

You, dear Reader, may not find all of these questions interesting, but you are probably curious about some of them. The writers of this book are all curious about aspects of language and we all find our own area the most exciting. In this book we want to tell you a little about various fields and sub-fields of linguistics that we are fascinated by, and elements we find so remarkable that we want to spend our time thinking and learning more about them. We are not going to go into detail, nor are we going to try and cover all the areas that linguistics studies.

This first chapter is an invitation to you to take a tour of linguistics and see what you find, and think about what you like. I have been your guide on this first part of the tour, where I have told you a little about what language is, and what linguistics is. I will stay with you and – I hope – intrigue you with stories and questions that will lead us to the individual chapters, where other guides will take over. It is a journey I know I will enjoy and I hope very much that you will, too, *since these will be only the* ***first steps***. After this, if you have caught the infection, a whole lifetime of questions and possible answers stretches before you. That life-long journey may require you to do hard things, to struggle with difficult concepts and theories, and to battle with mountains of data. Yet it is an exciting journey and perhaps you will set out on that journey today, with the help of this little book.

In the course of your studies you will come across a multitude of terms used by linguists as they try to identify how language(s) work, how they can be learned, and how they can be described and investigated. Sometimes these terms are used in different and even confusing ways. For this reason we have tried in this book to use only the most important or common terms. To help you, we have marked these technical words by using SMALL CAPITALS, and

have added a glossary near the back of the book. Each entry lists a page number as well. In the digital version, there are hyperlinks from the text to the GLOSSARY.

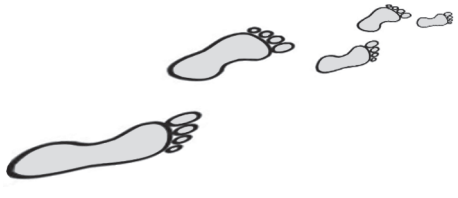
Each chapter introduces you to some of the 'tools of the trade' for that particular area of linguistics, so that you can see the various means used to analyse language in a structured and systematic way. Wherever possible, we have used realistic linguistic examples, mostly from English and Hungarian, to illustrate what we have to say. These examples, as you may already have noticed above, are printed in *italics* so you can easily recognise them. When the pronunciation is significant, we have used phonetic symbols like the ones in the learner dictionaries that you probably use.

Suggested Reading

Aitchison, J. (1992): *Teach yourself linguistics.* Hodder: London
A pleasant, easy-to-read introduction.
Fasold, R. - Connor-Linton, J. (2006): *An Introduction to Language and Linguistics.* CUP: Cambridge:
A brand new book that is similar to this one in many ways. The contents require a better knowledge of English and provide much more in-depth coverage of the topics. Like this book, each chapter is written by someone who teaches courses on that subject. Basically, it is this book's big brother (or sister).
Fromkin,V. - Rodman, R. (19986): *An Introduction to Linguistics.* Holt, Rinehart and Winston: New York
The book that your teachers may have studied from. Still a good general introduction and I, personally, liked the cartoons.
Pinker, S. (1994): *The language instinct.* Harper Collins: New York
A book written for the general public by a serious academic. It is also available in Hungarian.

Web-sites

http://en.wikipedia.org
Lots of useful information; written in clear language. Start at any point, jump to interesting links, and stop when you run out of time. Warning: this can be addictive.
http://www.linguistlist.org
This is a mailing list dedicated to linguistics. It contains postings sent by linguists from all over the world different on topics related to linguistics. Sign up as linguistis-to-be.



*Little Isaac asked, "**Why does the apple fall down?**" The same sentence in the phonetic script: [ˌwaɪdəzðɪ ˌæplfɔːl ˈdaʊn]*

At first glance, this looks incomprehensible. What are the strange symbols?

*You can find pronunciation symbols, or the phonetic script, at the beginning of every English-Hungarian dictionary. Why do we need them? Because the pronunciation of English is fairly unpredictable. The "a" in **apple** sounds /æ/ and in **fall** it sounds /ɔː/. **The** is pronounced /ðɪ/ in the above sentence but it is pronounced /ðə/ before a consonant. As the old joke goes, the pronunciation of English words is so difficult that it would be no surprise if Manchester was pronounced as /ˈlɪvəpuːl/. Liverpool, that is!*

*At second glance, even knowing about the phonetic script, the sentence still looks incomprehensible. Why are the words not separated with spaces? Because the sounds do not come to us neatly separated into words. The fact is that what we hear is generally an unbroken sound stream. Yet we hear 'words' and understand the soundstreams. Indeed, most of us would say that we hear the difference between **bin** and **pin** even if a machine recording does not show a clear cut-off point between the /b/ and /p/*

If you wish to know more about English speech, read on. This chapter is about the sounds of language and about the writing down of those sounds.

Pál Szemere
Kodolányi János University College
Department of English Language and Literature

2. Sounds Good

Phonetics & Phonology

2.1. Sounds and letters

We all speak differently. There are slow speakers and fast speakers, people who speak loudly and ones who speak softly and, like in the opera, there are sopranos and baritones.

We even say the sounds differently, all of us. If you call home, you can tell who answered the phone as soon as you hear the first word. You can tell if it is your mother, sister, grandmother or someone else. Another example: incoming calls are routinely recorded at the fire station. If you call in to say 'There is a bomb in the school building, it can explode any minute' and this call is recorded, police investigators will be able to identify you. It's not only your voice, it is also the way you pronounce the sounds. Do not experiment with it, it was just an example.

There are personal differences in speech, just like in writing, We can recognize our friends' handwriting. Having marked a few tests, teachers can tell who wrote which test paper, just by looking at the handwriting. How? We all write differently.

Even the same person says the same sound differently every now and then. For instance, we can pronounce the sound "sh" in a great many ways. Try saying "Sssh!" for a long time. Smile (like in *she*) and pout your lips (like in *shoe*) while saying "sh" nonstop, and you will hear the difference.

2.2. Sounds and phonemes

Phonology deals with sounds, but of course it does not deal with all the small differences we can observe between two speakers. Phonologists are not interested in the fact that my 'Ssh!' sounds different from your 'Ssh!', the way I normally say it, or that I sound different when I smile to when I do not.

All sounds are different, like letters in people's handwriting. Never mind – we will ignore these small differences. Phonology is not interested in them. All that matters is the difference between "sh" (like in *shoe*) and "s" (like in *Sue*). Similarly, when we read we do not notice the subtle differences between two A's (**a** or **a**). The only thing we want to know is whether it is an A, a D or some other letter (**a** or **d**). Phonology is interested in sound types, which we call PHONEMES. A phoneme is a type of sound, like /d/, regardless of personal differences between people's speech. Phonemes are customarily indicated like this: /k/. Don't say "kay",

just /k/. If you say "kay" you mean the letter "K", and in this chapter we are not going to talk about letters, we are only interested in speech. Here in Table 1 is a phoneme chart for your convenience. A small technicality: phonemes are abstract and cannot be pronounced. We pronounce sounds like [k] representing phonemes like /k/. The notation given in this chapter is phonological.

Consonants:		**Vowels:**	
/p/	**p**ut	/iː/	s**ee**n
/t/	**t**ake	/ɪ/	s**i**n
/k/	**c**ool	/e/	m**e**t
/b/	**b**ee	/æ/	c**a**t
/d/	**d**o	/ʊ/	b**u**sh
/g/	**g**et	/uː/	c**oo**l
/f/	**f**ar	/ɔː/	l**aw**
/v/	**v**et	/ɒ/	p**o**t
/s/	**s**ay	/eɪ/	h**a**te
/z/	**z**oo	/aɪ/	r**i**de
/ʃ/	**sh**ut	/ə/	**a**bout
/ʒ/	mea**s**ure	/eə/	c**a**re
/tʃ/	**ch**ild	/ɪə/	h**e**re
/dʒ/	**J**oe	/ʊə/	s**u**re
/θ/	**th**ink	/aɪə/	f**i**re
/ð/	**th**is	/aʊə/	fl**ow**er
/m/	**m**ean	/ɑː/	c**a**rd
/n/	**n**o	/ʌ/	c**u**t
/ŋ/	si**ng**	/ɜː/	b**ir**d
/l/	**l**ook		
/r/	**r**ight		
/j/	**y**ou		
/w/	**w**hat		
/h/	**h**i		

The phonetic script – symbols and examples

But why would we want to have a set of symbols? Why not use the well-known letters of the Roman alphabet? The reason is that one letter can stand for a number of different phonemes. Pronounce the "A" part of these examples:

cat Kate card care around

Each of them sounds different from all the others. 1 letter → 4 different phonemes. Besides, sometimes the opposite is true: 3 different letters, or in this case letter combinations, stand for 1 phoneme: *graph*, *laugh* and *half* all end in /f/.

spelling (written form)			Pronunciation (spoken form)
*c**a**t*	“A”	↗	/ae/
*K**a**te*		→	/ei/
*c**a**rd*		↘	/a:/
*c**a**re*			/eeö/
***a**round*			/e/

Letters and sounds 1: One-to-many

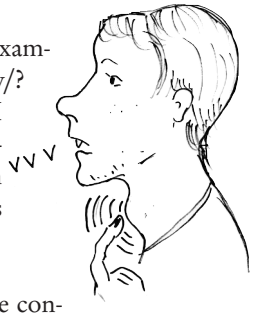
spelling (written form)			Pronunciation (spoken form)
*gra**ph***	“ph”	↘	/f/
*lau**gh***	“gh”	→	
half	“f”	↗	

Letters and sounds 2: Many-to-one

2.3. Feel your voice

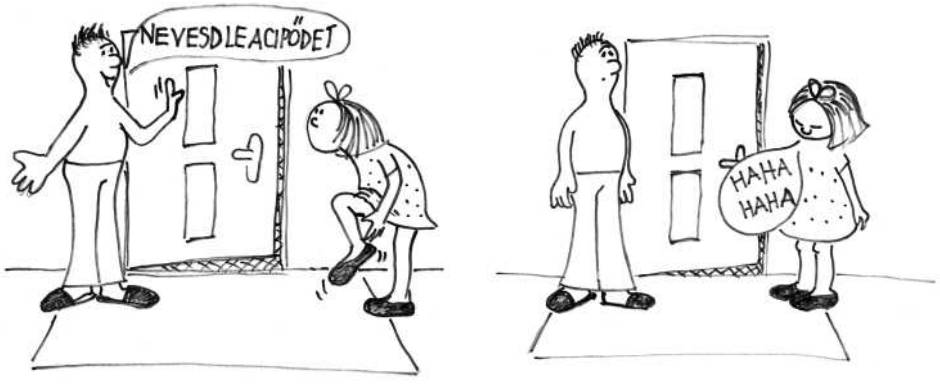
Do you remember when you started primary school? It was easy to learn Hungarian spelling: one letter – one sound. (I’m sorry for English-speaking children, they have a much harder job when learning how to spell.) There are a few exceptions in Hungarian, though. Take the Hungarian words ending in *–sd*, for instance, like *mosd, lásd, nevesd, üsd,* etc. What happens is that “sh” will become “zh” in pronunciation, so in fact we say *mozsd, lázsd, nevezsd, üzsd.* This is called *voicing assimilation.*

What is *voicing*? We are going to look at some examples. What is the difference between /f/ and /v/? Try saying ffffffffffff (for a long time, I mean) and put your finger on your throat – can you feel anything? Try the same with vvvvvvvvvv – can you feel the good vibrations now?



There is a name for these vibrations: voicing. Some consonants are voiced, some are voiceless. Can you guess which is which in our previous example of /f/ and /v/? That’s right, /f/ is voiceless (no vibrations) and /v/ is voiced (vibrations in the throat).

Let us now see a more detailed analysis of what is going on in *mozsd*. When we take the word *mos*, the final sound is "sh", that is voiceless. When we say *mos+d*, we are trying to say a voiceless and a voiced consonant at the end of a word – this is impossible. So what will happen? There is a choice: either the voiceless one (/ʃ/) will have to turn voiced, or the other one, that is the voiced /d/ will have to become voiceless. As it happens, in Hungarian the first scenario is realized and /ʃ/ will turn into /ʒ/.



Does the same thing happen in English, as well? Yes and no. Certainly, there are voiced and voiceless consonants. English words may end in /sh/ as well, like *wash*. We can add a /d/, like in the past form of this verb: *washed*. There is the same conflict: /sh/ and /d/ at the end of a word, the first being voiceless, the second voiced. But the solution, the way out of the conflict is different this time. No native English speaker will say *wazhed*, they will say *washt* instead. Different language – different rules apply.

2.4. R's are us

English R's do not sound the same as Hungarian R's. That is no surprise, neither do German or French R's. The problem is, though, that some R's in some dialects of English don't sound at all. They are silent.

If you ever buy hip-hop music or just browse song titles, you'll surely see words like *brotha*, *sista*, *ganxta* – these are more than just alternative spellings to *brother*, *sister* and *gangster*. They are telling signs that the singers (or shall we say *singaz*?) do not pronounce R's at the end of a word. Are all dialects of English like this?

Definitely not. If we just look at 2 *dialects*, we'll see the main differences. Let us experiment. Look at these examples.

Four. Four people. Four animals.

If you know someone from the South of England and ask this person to read out the three sentences above, you will only hear /r/ in the last sentence. R's are not pronounced at the end of a sentence, neither before a consonant, but they are pronounced before a vowel like /æ/. (As a consequence of this, this linking "r" appears even where the written form has no trace of it, as in *Béla is here* /beilərɪz hɪə/.)

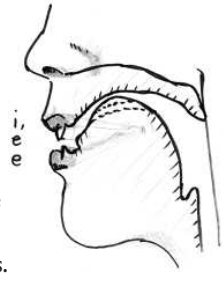
This, however, is not true for the most part of the United States. If you ask an American speaker to read out the same three sentences above, you will hear /r/ three times. In the dialect called General American every letter "r" is pronounced.

So what about *brotha* and *sista*? After all, hip-hop music comes from the US. If you want to find out the answer, you will have to read the chapter on *sociolinguistics* where you can also find more information on language varieties.

2.5. Front and back vowels

What is *in* in Hungarian? It depends – in some words it is *-ban*, in others *-ben*. *HázBAN* but *kézBEN*; *hóBAN* but *vízBEN* but *kínBAN* and *hídBAN*. (Sorry, that was rather cruel to you – *kín* and *híd* are exceptions.) So what is the rule? After "é", "e" and "i" we normally say *-ben*, after "u", "o" and "a" we say *-ban*. Interestingly, this seems to coincide with the place of articulation of vowel sounds.

Why is it like that? Let us take a closer look, now. The vowel in *ház* is a back vowel. It means that the highest point of your tongue is in the back of your mouth as you say it ("mély hangrendű") so it goes with a suffix containing the back vowel *-ban* (also "mély hangrendű"). The vowel in *kéz* contains a front vowel. It means that the highest point of your tongue is in the front of your mouth as you say it ("magas hangrendű"). It is no surprise that it goes with the suffix containing the front vowel *-ben* ("magas hangrendű"). So the basic idea is that front vowels go with front vowel suffixes, back vowels with back vowel suffixes.



The place of articulation of Hungarian i, é & e

In other words, if it is a front vowel (produced in the front part of your mouth) we say *–ben* after it, and if it is a back vowel (produced in the back of your mouth) we say *–ban*. Actually, the whole truth is somewhat more complex than this, but do you think you will be able to tell front vowels from back vowels?

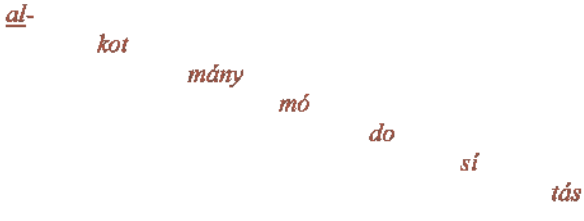
2.6. Are you stressed?

In everyday life we use the word *stress* in sentences like this *"I'm under a lot of stress, the phonology exam is on Monday."* In phonology *stress* is used in a different sense. First we shall talk about stress in words, then stress in sentences.

In Hungarian words the beginning of the word is always louder than the rest, like in *banán*, *automatikus*, *titkosszolgálat* or *alkotmánymódosítás*. We can represent this in many ways, for example, by underlining the first : *banán*, *automatikus*, *titkosszolgálat* or *alkotmánymódosítás*. In fact, it is not only loudness – the height of your voice changes, too. We can represent it like this:

Or like this:

and



A Hungarian word is full of optimism at the beginning, but then the height of voice starts slipping down the slope and there is no way back.

The stress in English words is not always at the beginning (*banana*) and sometimes there are two or more stresses (*automatic*, *Secret Service*, *amendment of the constitution*). English words are less predictable. The height of voice – or pitch – goes up and down like the price of bonds at the stock exchange.

ba- na- na

au- to- ma- tic

sec- ret ser- vice

and

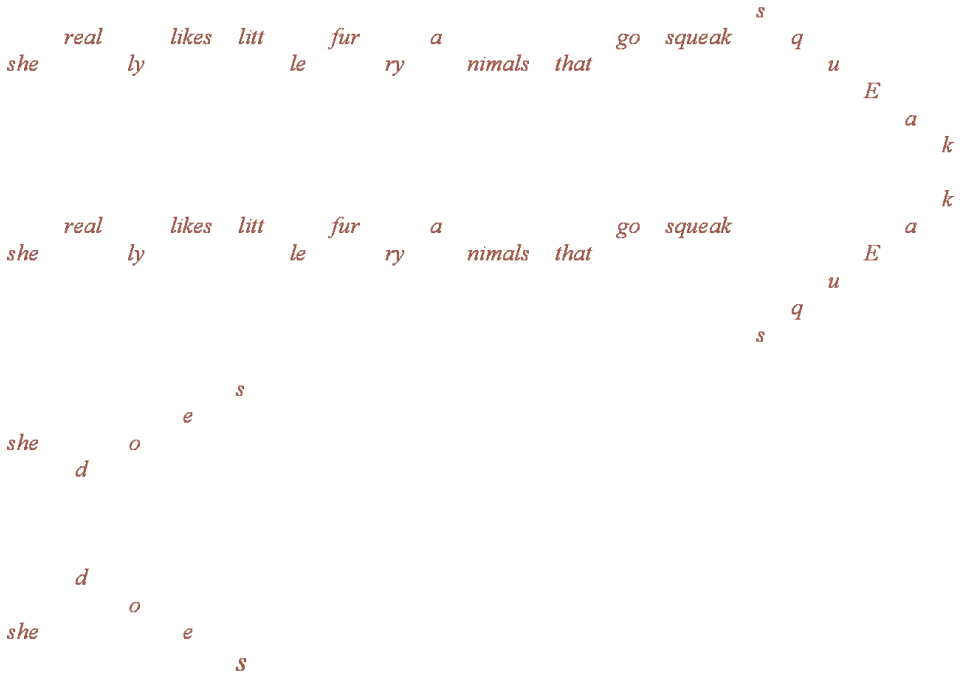
a- mend- ment of the con- sti- tu- tion

Some pairs of words only differ in the place of stress, like *produce* (= "termék") vs *produce* (="termel"), *record* (="rekord") vs *record* (="feljegyez, felvesz"), *subject* (="alattvaló") or *subject* (="alávet"). You can read more about it in the chapter on *Morphology*. As for sentence stress, see the section below.

2.7. Intonation

There is stress in sentences, too. Stressed parts are louder and, to put it simply, higher up on the voice scale than unstressed ones. This will be marked in the examples below.

Intonation is the melody of the sentence. Like in words, the height of our voice changes in sentences. In fact, it is the end of the sentence that matters. Look at the examples.



If it goes up, it is likely to be a question. (You can find more information about questions in the chapter on syntax.) In addition to that, intonation is an effective means to express your emotions and attitude towards other people. Read more about this in our chapters on psycholinguistics and pragmatics.

2.8. Summary

In this chapter you have read about phonology. It might be useful to organize your notes around these headings:

Sounds and letters – we all speak differently but these differences do not matter more than the differences between handwritings.

Sounds and phonemes – try to remember the phonetic script, *i.e.*, the symbols for vowels and consonants.

Feel your voice – some consonants are voiced, others are voiceless.

R's are us – depending on where you come from, you either pronounce all "r"s or you just pronounce some of them.

Front and back vowels – some vowels are produced in the front of your mouth, some others in the back.

Are you stressed? – stress is not everywhere in an English sentence; it is on some words/syllables.

Intonation – just like songs, sentences have recognizable melodies.

Points to Ponder

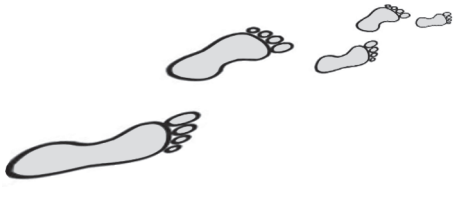
- Why do we need the phonetic script?
- What is voicing assimilation?
- What is the rule in Hungarian vowel harmony?
- Which r's are pronounced by native speakers of English?
- What is a syllable?
- What is stress?
- What is the melody of an English yes-or-no question?

Suggested Reading

Hancock, M. (2003): *English Pronunciation in Use.* Cambridge University Press.
We recommend this book for a first look at English pronunciation.

Nádasdy, Á. (1998): *Background to English pronunciation: (phonetics, phonology, spelling): for students of English at Hungarian Teacher Training Colleges.*
A somewhat more demanding task is to cope with this book.

Nádasdy, Á. (2003): *Practice book in English phonetics and phonology.* Nemzeti Tankönyvkiadó
This is actually course material used by English majors at a number of universities in Hungary.



*Remember little Isaac and the apple-tree in Chapter One? (If not, you are missing out on a story!) Well, from the perspective of this chapter – and history – that particular apple fell down a long, long time ago. But surely that should have read, "…the apple *falled down."? Were you uncertain for a moment? Have you considered why* ***fall*** *is not a well-behaved English verb? Nobody really knows why* ***fall*** *does not take* ***-ed*** *for its past form, like most other English verbs. As you will see in the chapter to follow,* ***fall*** *has irregular morphology, and* ****falled*** *is morphologically ill-formed.*

As to the real answer to the real question of why apples fall down, that is a question for a physicist. A physicist is someone who knows a lot about physics, just like a chemist knows a lot about chemistry. Linguists, as we have seen, study language, and historians study history. Yet these words, although showing some similarity, do not follow a regular pattern. There are no words like ****chemistrist****,* ****historicist****, or* ****languageist*** *- nor* ****chemistrian*** *or* ****languagean****. Of course, there is a word* ***physician*** *– but then that means something else altogether!*

Once you start seeing these anomalies, you can keep yourself amused for a life-time. Read on to find out more about words and how pieces of words can be put together to make new words.

Pál Szemere
Kodolányi János University College
Department of English Language and Literature

3. Words = word + s

Morphology

3.1. Morphemes

In this section I will try to demonstrate that words are made up of meaningful parts. Let us start with apples.

How do you count apples? One apple, two apples, three apples, four ...
It is a simple rule: *1 apple →2 apple + s.*

How do you count mice? One mouse, two mice, three mice, four mice.
The rule is somewhat different: *1 mouse → 2 *mouses mice* (See 10.1)

What do we call this? In the course book you probably used as a beginner learner of English, they called these words IRREGULAR plural nouns. The idea was that nouns have a singular and a plural form in English, just like in any other language. The plural form is often *–s* (or *–es*), to be quite precise). These plurals are regular, they follow a pattern, they observe the rule, they are nice, well-behaved nouns.

Some nouns, though, are different. They have a plural form, but it is not what we would expect. The plural of *child* is not **childs* (but *children*), that of *criterion* is not **criterions* (but *criteria*) and there is no word like **foots* (it is *feet*). What is the rule then? There is no rule. The plural of these nouns is unpredictable.

In other words:

regular plural = singular + (e)s *book → books*
irregular plural = ??? [unpredictable] *mouse → mice*

English and American children often make mistakes about these nouns – they would say things like *mouses* and *gooses* and *fishes* where adult speakers use the forms *mice*, *geese* and *fish*. (This is the way to count fish: one fish, two fish, three fish, four fish.)

This is the sort of thing morphology deals with: the form of words. More precisely, it focuses on the meaningful parts of words. A word like *apples* consists of two meaningful parts: *apple + s*. We all know the meaning of *apple* but why do we say that *-s* has a meaning as well? The meaning here is that there are more of them. (It has GRAMMATICAL MEANING rather than being the name of an object/person/idea.)

3.1.1. Irregular forms

In the above example we have had a look at irregular plural nouns. There is nothing logical about them, you just have to remember that the plural of *foot* is *feet*. There are irregular verbs and adjectives, too. Look at this chart.

base form	past form
smoke	*→ smoked*
arrive	*→ arrived*
use	*→ used*
go	*(→ *goed) → went*
forbid	*(→ *forbidded) → forbade*

Table 1 Regular and irregular verbs

base form	Comparative form
small	*→ smaller*
high	*→ higher*
intelligent	*→ more intelligent*
beautiful	*→ more beautiful*
good	*(→ *gooder) → better*
bad	*(→ *badder) → worse*

Table 2 Regular and irregular adjectives

The top of both tables show regular examples, the bottom part irregular ones. In the irregular examples, when the form is totally unpredictable and unexpected, in fact a completely different word, we speak about SUPPLETIVISM. The form *went* does not follow from *go*, neither does *better* from *good*.

3.1.2. Free and bound morphemes

Meaningful elements like *apple* and the plural *–s* are called MORPHEMES. In *unemployment* we can find three morphemes: *un* + *employ* + *ment*. Out of these three morphemes *employ* can also be a word used on its own – this is what we call a FREE MORPHEME. *Un-* and *–ment*, on the other hand, cannot be used as words, they can only be part of a word. In order to appear in a word they must be attached to another, free, morpheme. These are BOUND MORPHEMES – they are bound to a free one. The traditional name of *un-* is PREFIX (as it is before another morpheme) and that of *–ment* is SUFFIX (it is after another morpheme). *Employ* is the ROOT of the word.

There are some more words that begin with *un-* : *un+happy*, *un+real*, *un+able*, *un+conscious*. I am sure you could list many more words like this – it is a productive process. A process is like a machine: in this case an adjective is the input and *un* + adjective is the output. There are a number of examples where this works. Notice that un- in all the above examples is attached to adjectives. *Happy* is an adjective and *unhappy* is still an adjective. *Un-* is a meaningful element, a morpheme. *Un-* means the opposite of something. *Unhappy* = not happy.

Let us see more examples for a suffix: words that end in *–ment*: *govern+ment*, *retire+ment*, *argu+ment*. Notice that *-ment* is attached to verbs and the resulting word is a noun. We could list a great many words ending in *-ment*, which means this is also a productive process. The meaning of *–ment* = the word ending in *-ment* is a noun.

Now let us take a close look at *quick+er* and *read+er*. Can we say that *-er* is a morpheme in these words? Can we say it is the same morpheme?

We can surely say they are morphemes, as we can list many words following the same pattern as *quicker*: *slower, nicer, uglier, smaller,* etc. In these words *-er* is attached to an adjective and the resulting word is still an adjective, but in comparative form, like in this sentence:

(1) *Ed is quick, but Liz is quicker.*

Can we say the same about *reader*? Similar words: *teacher, learner, writer, walker*. This *–er* is added to verbs and the result is nouns (*teach* v → *teacher* n). So the answer is: yes, *-er* is a morpheme in both *quicker* and *reader*, but not the same morpheme. In fact, it is sheer coincidence that they both have the form *–er*, as their function is completely different.

3.1.3. Morphological coincidences

In the previous section we concluded that *–er* is really two different suffixes: -er[1] and –er[2]. We could say the same about *–s* in *decides* and *umbrellas*, *-ly* in *quickly* and *friendly*, or *-al* in *global* and *arrival*. See Table 3 below.

form	examples	function (= meaning)
-s	*decides, understands, knows, eats*	3rd person singular verb
-s	*umbrellas, computers, cats, bodies*	plural noun

form	examples	function (= meaning)
-ly	*quickly, suddenly, shortly, happily*	adjective → adverb
-ly	*friendly, brotherly, heavenly, bodily*	noun → adjective

form	examples	function (= meaning)
-al	*global, postal, doctoral, autumnal*	noun → adjective
-al	*arrival, betrayal, dismissal, refusal*	verb → noun

Table 3 Morphological coincidences

3.2. Lexemes

What is a verb? Morphologically speaking, a verb is a word that takes endings like 3rd sing –s.

(2) *She arrive|**s** next Friday.*

A word of caution: there are many irregular verbs – not all of them take -ed.

What is a noun? A word that takes endings like plural –s is definitely a noun.

(3) *So far I've lost four umbrella|**s** this year.*

A word of caution: not all nouns do, see *mouse* and *fish* above – not to mention uncountable nouns, which have no plural form at all.

What is an adjective? A word that takes adjectival suffixes like comparative –er and superlative –est.

(4) *Ed is slow but Liz is even slow|**er**, and Gyula is the slow|**est** of them all.*

A word of caution: many adjectives follow a different pattern like
tired→more tired→the most tired.

The above examples show that morphological clues on words do not work in a hundred per cent of the cases. (For more on WORD CLASSes see Ch. 4 *Syntax*) But let us turn to an even more intriguing question. What is a word?

This is a difficult question. For our current purposes and in order to keep things simple at this point let us say that a word is a string of letters between two spaces on a printed page. Nouns, verbs and adjectives are words. Words are listed in the dictionary and, in the case of irregular nouns, verbs or adjectives some other forms are given, too. If you look up the noun *child* in a dictionary it will provide the information that the plural form is *children*, at the verb *understand* you will be warned that the past form is *understood*, at the adjective *good* you will find the irregular comparative and superlative forms *better* and *best*.

Child and *children* (or *book* and *books*) are not two unrelated words, rather two forms of the same word. They are one LEXEME. A lexeme is the name for the various forms of the same word. Book and books are one lexeme. Similarly, *understand* and *understood* (or *arrive* and *arrived*) are one lexeme. *Good*, *better* and *best* are one lexeme, and the same is true for the three forms of regular adjectives. In the dictionary you will not find *children* – search for it at the head word *child*.

3.2.1. Stress as noun marker

Oftentimes one word belongs to two or more word classes. Take these sentences:

(5) *Ronaldo was booked in the first half.*
(6) *I've been reading an excellent book.*

It seems that *book* is a verb in (5) and a noun in (6). Where is this information coded in the dictionary? You may find it at the head word *book* – *book*[1] or *book*[2] or just *book*, depending on the kind of dictionary you use. The difference is that if you see *book*[1] and *book*[2] the editors mean they are two different lexemes, not just two senses of the same lexeme.

In some cases the printed form is the same, but the spoken form is not (see *homographs* in Ch. 2 *Phonology*). As a matter of fact, sometimes the place of stress is the only indicator whether it is a noun or a verb (underlined vowels are stressed vowels)

(7) *I just want to set the r<u>e</u>cord straight.*
(8) *This song was rec<u>o</u>rded back in 1962.*
(9) *Objection, your honour. I obj<u>e</u>ct to this proposal.*
(10) *The Smiths caught sight of a huge circular <u>o</u>bject floating above their home.*

3.2.2. The origin of words

Where do words come from? The word *saxophone* was named after the inventor, Antoine J. Sax, who made musical instruments in Belgium in the 19th century. According to Webster's dictionary *saxophone* first appeared in print in 1851. Other words, in fact most words, have a much less documented origin. See Ch. 7 *Language History.*

The origin of a word is not the meaning of a word. Many people believe that the proper meaning of a word is its original meaning. This is not true. Think of the word screen. An English – Hungarian dictionary lists, among others, the following Hungarian equivalents: *kálya-haellenző, redőny, vívómaszk, szélvédő, szentélyrekesztő fal, vetítővászon, képernyő, fényképészeti rács, rosta, szűrés*. The original meaning must have been either 'rosta' or 'ellenző'. Today, however, the most frequently occurring senses are probably 'képernyő', 'mozivászon' and 'szúnyogháló'.

3.3. Word classes

So far we have been discussing the morphological behaviour of nouns, verbs and adjectives. There are a lot more word classes, though. Look at this short passage.

> (11) *Frong Thampid jurgled the stonkers at a dongle dridge so gorleantly that no rampid flugsoniac could witherate the chorticle under the blistic plinkage before shemming away to the goffle.*

What can *Frong Thampid* be? Probably somebody's name. A noun then. *Jurgled* must be a verb in past form, *the* is 'the', *stonkers* is a noun in plural, it seems, *at* and *a* are actual English words. We cannot be sure about *dongle* (noun or adjective?) but *dridge* must definitely be a noun. And so on.

In this nonsensical text I made up nonexistent nouns (*chorticle*), verbs (*witherate*), adjectives (*blistic*) and adverbs (*gorleantly*) but I could not possibly come up with new articles, prepositions or auxiliaries. Find out more about it in the next section.

3.3.1. Open and closed classes

Let us examine word classes. First, a few examples.

- Nouns (*umbrella*)
- Verbs (lexical) (*arrive*)
- Adjectives (*big*)
- Adverbs (*soon*)

- Prepositions (*in, at, with, under*)
- Pronouns (*her, all, what*)
- Auxiliary verbs (*can, has*)
- Determiners (*an, every*)
- Conjunctions (*and, because*)
- Interjections (*Oops, Ouch*)

In traditional grammar word classes are defined with reference to their meaning. I recall having to memorise definitions like 'Verbs refer to actions, events or existence' back in primary school. This is not the direction we are going to pursue in the current chapter. For morphology, form is very important (in Hungarian it is actually called 'Alaktan'). That is why we will define word classes on a formal basis rather than on the meaning of words.

The easiest way to find out the word class of a word is by looking it up in a dictionary. *Arrive* is marked **v** (short for verb), *should* **aux** (auxiliary) and *after* **prep** (preposition).

About word classes we can ask four questions.

- One: Can you make a new word from this word by adding a suffix/prefix?
- Two: How many words are there in this word class?
- Three: Are there new words in this word class?
- Four: Does it normally carry stress in the sentence?

And the answers are as follows.
One: nouns, verbs, adjectives and adverbs take suffixes, (e.g. *umbrellas, arrived, biggest, sooner*), the other word classes do not, except for the plural –s .

(12) *There were two and's and three in's in her last sentence.*

Two: there are a lot of nouns, verbs, adjectives and adverbs in English, tens of thousands, that is, and there are only a few dozen of the other word classes at most.

Three: we see a great many new words brought about by technological invention or slang usage (take the word *bluetooth* or the expression *pimp my ride*) but these can only be nouns, verbs, adjectives and adverbs. There are no new prepositions.

Four: if you give an English sentence to a newsreader who is a native speaker of English, she will stress nouns, verbs, adjectives and adverbs and typically not stress other word classes.

The above differences suggest that nouns, verbs, adjectives and adverbs are different from other word classes. Grammars call them OPEN CLASSES, the other word classes are CLOSED CLASSES.

3.4. Word-formation

3.4.1. Compounding and other ways of word formation

There are many ways to create new words in every language, but let us look at a frequent technique. You can make a new word if you combine two words. Examples: *police officer*, *word-formation* and *handbook*. (Notice that *word-formation* is hyphenated and *police officer* is actually two words if we look at the spelling, but one lexeme – it is a head word in dictionaries.) During the process of COMPOUNDING, as it is called, the meaning sometimes drifts away from the original meaning, like a *shortlist* is not any list that is short and a *blackbird* is not any kind of bird that is black.

There are a great many other ways to form new words, among others BLENDING (*brunch* from *breakfast* and *lunch*, or *infotainment* from *information* and *entertainment*), ACRONYMS (*ICT* from *information and communication technology*, or *CV* from *curriculum vitae*), CLIPPING (*info* from *information*, or *prep* from *preparatory*), CONVERSION (to *message somebody* – the verb comes from the noun *message*).

3.4.2. Derivation vs. inflection

A suffix is inflectional if the result is just another form of the same lexeme (e.g. *arrive* →*arrive**d***/*arrive**s*** or *umbrella* →*umbrella**s***) and derivational if it creates a new lexeme (*arrive* → *arriv**al*** or *book* → *book**ish***). This may or may not be linked to a change in word class: *-let* in *piglet* is a derivational suffix, and so is *-ship* in *friendship*, although both the original words and the derived words are nouns. Derivation sometimes happens through prefixes in English: *large* → ***en****large*, *polite* → ***im****polite*.

3.5. Language types

In the previous sections the English language was in focus; in this section we are going to look at the morphology of various languages and come up with a linguistic typology. There are several ways to put languages in groups: it can be based on word order, origin/history and other criteria. In this section we will only look at morphological differences. Now look at the examples in the table on the next page.

You may notice two things: first, that the order of elements is reverse if we compare the two languages in examples 13a-13b; 14a-14b. Second, what is one word in Hungarian is likely to be several words English. English tends to isolate the morphemes – that is why it is often called an isolating language.

	1	2	3
	I	*can*	*go*

(13b)

	3	2	1
	Me	*het*	*ek*

(14a)

	1	2
	at	*midnight*

(14b)

	2	1
	éjfél	*kor*

What is one morphologically complex word in Hungarian is often many separate words in English. For example,

(15) *elállatiasodottságában* = [in his condition of being rather similar to an animal].

Hungarian is said to be an AGGLUTINATING language, where you can have a long word made up of several morphemes, which are clearly separated and all of them have identifiable meaning. In the above example: *el+állat+i+as+od+ott+ság+á+ban* and English is an ISOLATING language, where every morpheme is likely to be a new word. There are no clear cases, though, so we know many English words that contain a number of morphemes.

There is a third type of language and again we will see a Hungarian example. What is the root of Hungarian *eszik*? You might want to say it is *esz-*, because there are forms like *eszel*, *eszünk*, etc. But then consider *együnk*, *evett*, *enni* – are we going to say the root is nothing but *e-* ? Perhaps we could say that the root of the word changes. It is like the branch of a tree bending this way and that way, as the wind goes, under the burden of fruits or in the direction of the sunlight. The root of the word is flexible. This is common in INFLECTIONAL languages. In inflectional languages it is not clear where one morpheme ends and the next one begins.

Another name for this type of language is FUSIONAL. This name suggests that a lot of information is compressed into one or very few morphemes. Look at (16a and 16b) below.

(16a) *Látlak*
(16b) *I can see you*

Látlak is just one word but the part *-lak* expresses first person singular subject, second person singular object, and present tense.

3.6. Summary

In this chapter you have read about morphology. It might be useful to organize your notes around these headings:

Morphemes – they are the smallest meaningful elements in a language; they can be free and bound.

Lexemes – various forms of the same word (i.e. small, smaller and smallest) are really the same lexeme.

Word classes – some word classes are lexical, others are grammatical – they are called open and closed classes, respectively.

Word formation – there are ways to make new words out of existing ones; two such ways are derivation and compounding.

Language types – languages can be put into groups according to their morphology. One such categorization is to differentiate between synthetic and analytic languages.

We suggest that you try to do the Tasks immediately after reading the passage they are related to.

Points to Ponder

1. Identify free morphemes and bound morphemes in these words.

Example: *un*|*friend*|*ly* : *un-* & *-ly* are bound morphemes; *friend* is a free morpheme.
work|*er, additional, independent, irrationality, suddenly, brotherly, motherhood, hardship, friendship, battleship, personalities, enlightenment, unthinkable*

Now do the same with these Hungarian words:
kutyákat, érthetetlen, barátainknak, édesen, darabolják

2. Write two sentences per word to illustrate the ambiguity.

Example: a) *Nem akarok* ***vért*** *látni*
b) *Fénylő ércből kovácsolt* ***vért*** *csillogott a lovagokon.*

vállal, szánt, bánt, ért, írt, árt

Now do the same with these English examples:

tinker, claims, suspect, building, bank, mine, lead, can, bear

3. One morpheme or two morphemes? If the word has two morphemes, specify the function of the suffix.

Example: *sing*|*er*, *-er* is a verb →noun sufix

singer, finger, blister, sister, thinker, wonder, deeper, beeper, over, mover, actor, factor, manor, denture, venture

Now do the same with these Hungarian words:

arát, barát, hánt, ment, dönt, szőrt, kért, mart, gondol, gombol, horgas, ordas, vakondok, vakondokok

4. What is the word class of each word in these sentences? Check your answers in a dictionary.

Example: *Dogs chase cats.* Noun, verb, noun

a) *The dog chased a cat.*
b) *The fierce dog chased the black cat.*
c) *The cat is under the bed.*
d) *The dog is barking loudly.*
e) *The cat might be scared to death.*
f) *As we entered the house an incredibly fierce large dog was chasing a pitch black cat round and round in our living room. The cat eventually found refuge under the bed while the dog barked like there was no tomorrow.*

Suggested Reading

Aitchison, J. (1992): *Teach yourself linguistics.* Hodder: London

Akmajian, A. - Demers, R. - Farmer, A. - Harnish, R. (2001): *Linguistics: An Introduction to Language and Communication.* MIT Press

Crystal, D. (ed.) (1997): *The Cambridge Encyclopedia of Language.* 2nd ed. Cambridge University Press: Cambridge.

If you liked this chapter you may want to look at the morphology chapter of these introductory books to find out more. They do not require any previous knowledge.

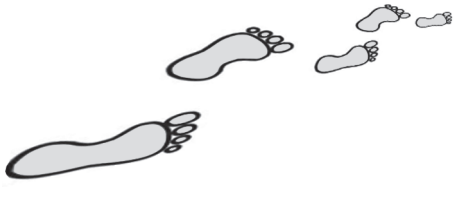
Carstairs-McCarthy, A. (2002):: *An Introduction to English Morphology.* Edinburgh University Press

Bauer, L. (1990): *Introducing Linguistic Morphology.* Edinburgh University Press

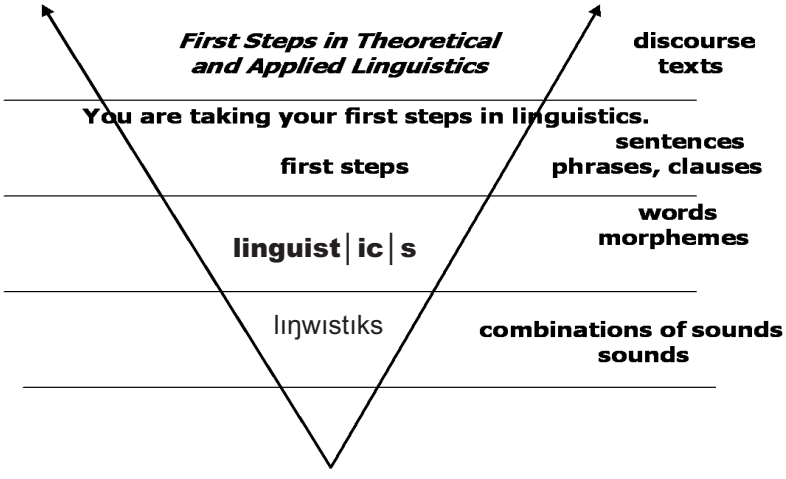
The books above are recommended to the interested reader. They are quite accessible.

Recommended websites:

www.wikipedia.com
www.linguistlist.com



This chapter is going to be about the syntactic level. You will find out things like why words or strings of words ordered randomly are of little use to people who want to communicate. If you set off to visit a foreign country where they speak a language you do not, then it is not enough to simply take a dictionary and hope that you will be able to communicate. You will find out that to make sentences, words have to be combined according to rules and you will find out about how these rules work. You already know a lot about language from the chapters that have gone before. To help you put it all together, the diagram below shows how language can be seen as an upside-down pyramid. The levels are separate but are related to one another. There are rules for combining units at all the different levels, and together they constitute the structure of language.



Judit Górász
Kodolányi János University College
Department of English Language and Literature

4. Up the Linguistic Pyramid

Syntax

4.1. What is syntax?

4.1.1. Introduction

What would you say if somebody asked you what the building blocks of language were? A probable answer is: words. We know that every language has a set of words (called LEXICON);

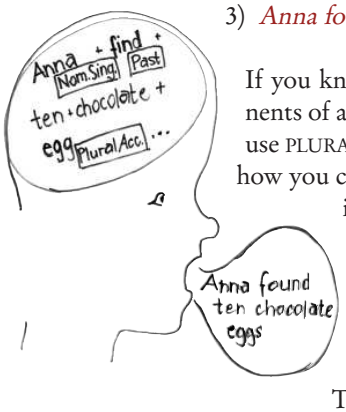
still if you have ever tried to translate even a very short text from Hungarian into English, you must have noticed that simply knowing the English equivalent of each word is not enough. What do you have to know if you want to translate the sentence below?

1) *Anna tíz csokoládétojást talált a kertben húsvétkor.*

Of course, you must know the English translations of all the Hungarian words above (you can find them in a dictionary), but you must also know, for example, that the equivalent of *csokoládétojás* (unlike the Hungarian original) has to appear in the plural form in the English sentence (because of the number) and you must know what this plural form is *chocolate eggs*. This means you have to know "building blocks" smaller than words, i.e., morphemes, e.g., the plural morpheme *-s* in English. However, morphological knowledge is not enough if you want to translate *kertben* or *húsvétkor*, since now you have to know that instead of suffixes (such as *–ben* or *–kor* in Hungarian) English uses prepositions (words placed before nouns) here (in, at). Finally, when you have the components listed below:

2) *[Anna], [ten chocolate eggs], [found], [in the garden], [at Easter];*

you will have to order them in a different way to get a grammatical (i.e., correct) English sentence:



3) *Anna found ten chocolate eggs in the garden at Easter.*

If you know when to use prepositions or how to order the components of a sentence or whether (after numbers, for example) you must use PLURAL or SINGULAR forms of nouns, you know rules that tell you how you can combine the building blocks of language into grammatical units larger than words. That these rules are indispensable is easy to see by looking at the words of our example sentence arranged in a random order:

4) *The found Easter chocolate in ten at garden Anna eggs.*

The group of words in 4) cannot be regarded as an English sentence, while the same words in 3) can. The difference is that 4) ignores the rules which combine the constituents (or parts) of a sentence, while 3) follows them. These rules and the "building blocks larger than words", which are formed with the help of the rules, make up the area of linguistics called syntax. In this chapter you will read about some of the basic concepts of this area.

4.1.2. Syntactic components

What are the "building blocks" or grammatical units of language which are larger than words? We have already mentioned one of them by name: SENTENCE. A sentence minimally

contains at least two basic parts: a subject (in our example, *Anna*) and a predicate (i.e., something that is said about the subject). A sentence, however, may contain more than one subject and more than one predicate: in this case the sentence consists of several "sub-sentences" called CLAUSES, e.g.

5) *Anna was playing the piano, while Sebastian was drawing a picture.*

Sentence 5) is still one sentence but, while 3) consists of only one clause, this sentence consists of two clauses (the word *while* starts the second clause). So, sentences and clauses are syntactic units, but what are the immediate constituents of clauses? Certainly not words, as example 4) has shown us; but rather groups of words like the ones in brackets in 2) above. Such syntactic units are called PHRASES.

To summarize the above: syntax (which constitutes a part of grammar) shows how words are combined into phrases, how phrases form clauses, and how clauses are joined to make sentences. Certain categories of syntax are universal: for, example, every language uses subjects. Other categories are language specific: some languages (e.g., English, Hungarian) have articles (*the*, *a*), while some others (e.g., Russian, Finnish) do not. Even the borderline between syntax and morphology (another area of grammar) may vary. In English the genitive relationship between two nouns may be expressed by changing the form of one of the nouns by adding a suffix (i.e., a morpheme) to it: *the country's history*. The same meaning may also be expressed by means of a preposition: *the history of the country*. In this phrase, both nouns are in their base forms and a separate word is used to convey the genitive ("possessive") meaning. The two structure types can even be combined in a certain meaning: *a friend of Sebastian's*. Notice that in Hungarian, for example, only the morphologically marked genitive is possible; and even that differs from the equivalent English structure, since it is the "possessed", rather than the "possessor" noun whose form is changed by the addition of a suffix: *az ország történelme*. In other languages, such as Spanish, it is always a preposition, not a suffix, that expresses the same meaning: *la historia del país*. Both in Hungarian and Spanish, however, we may find the opposite strategies as well if we examine other structures. Every language uses both structure types, though the proportions vary.

4.2. Generative grammar

4.2.1. Universal grammar

When the various grammatical differences between any two languages are closely examined, linguists often find that differences can be described using simple BINARY parameters: e.g., a language either uses articles or not; the article and/or the adjective either AGREES in its grammatical features with the noun it modifies (as in Spanish) or not (as in English), as the examples below demonstrate:

6) *la niña pequeña* — *the little girl*
las niñas pequeñas — *the little girls*
el niño pequeño — *the little boy*
los niños pequeños — *the little boys*

Moreover, as we saw above, there are certain universal categories which equally characterise every language (known as language universals): nouns, verbs, past tense or the possibility to form questions. Language itself is universal among humans (at the same time it is exclusively human) and human beings very early in their lives (i.e., under the age of three) are universally able to fully learn any language they are exposed to without being formally taught. This is a remarkable ability, especially if we consider how many human beings later in their lives struggle painfully for many years while trying to learn a foreign language.

These and other observations led a group of linguists in the second half of the 20th century to think that the ability to learn a first language is innate in all humans and the speed with which the task is performed by young children must be supported by an innate, so-called UNIVERSAL GRAMMAR. The notion of universal grammar is an important part of the wide-spread and influential theory called generative grammar. (Though numerous linguists have worked and are currently working in this field, the single most important name connected to generative linguistics is that of the American linguist and philosopher Noam Chomsky.) According to the framework known as 'principles and parameters', this innate grammar contains both the universal characteristics of human language (principles) and the binary parameters (such as the ones mentioned above) that the child must set on the basis of the language input (s)he receives; e.g., the child will use articles if and only if the people around him/her use articles as well (which, of course, depends on the syntactic properties of the language they speak). Therefore, parameter setting is language specific. If the language does have the category of articles, the actual phonological form(s) must be learned as well; but this is a LEXICAL (vocabulary-related) question.

4.2.2. Competence

Generative linguists were the first to pay attention to the fact that the number of sentences produced in a language is potentially infinite. This makes syntax unique in comparison with other areas of grammar: every language has a finite set of phonemes, a finite (though large and open) set of words and morphemes. The sentences of a language will never be listed in a "dictionary", like the words, since this would be an impossible task: every one of us regularly produces sentences we have never heard before and even sentences that have never been uttered by anyone before. Unlike words, sentences (and clauses and phrases) are not learned but generated on the basis of a finite set of words and a finite set of rules. The native speaker knows these rules and is able to generate well-formed sentences even without being aware that these rules exist at all. This knowledge of the native speaker is called COMPETENCE and it is often imperfectly reflected in actual speech, where various PERFORMANCE errors ("slips of the tongue", etc.) are possible. The task of a generative linguist is to mirror the native speaker's competence by constructing a formal grammar.

4.2.3. Related sentences

Generative linguists also observed that certain sentences are related to each other. Look at the following examples:

7) *The children have laid the table.*
8) *Have the children laid the table?*
9) *The table has been laid by the children.*
10) *Has the table been laid by the children?*

These sentences are related because they have the same underlying or DEEP STRUCTURE, from which each one can be generated by using various transformational rules (such processes are called TRANSFORMATIONS). Rules tell us, for example, how to generate an active or a passive sentence (e.g., sentences 7) and 9)) or how to form questions (e.g., sentences 8) and 10)). Since different rules are applied to the same deep structure, the result will be different SURFACE STRUCTURES (cf. the different ordering of *the children* and *the table* in sentences 7) and 9) or the different positions of *have* in 7) and 8) – these differences are the results of the MOVEMENTS of certain constituents out of their original, i.e., underlying positions). These surface structures appear in the pronounced or written forms of the sentences.

So, sentences with different surface structures may have the same deep structure. But can we find examples for the opposite? Are there sentences with different deep structures but identical surface structures? Look at the following example:

11) *He was interested in hunting lions.*

The sentence is ambiguous: without a context we cannot tell whether it is the lions that hunt or whether it is the lions that are hunted. This means that here we have two different sentences with identical surface structures. The deep structures, however, are different and the difference lies in the relationship of *hunting* and *lions*.

4.3. Words in syntax

4.3.1. About words...

Now let us return to words, which are the components of phrases, and as such are syntactic units themselves. You know that words are the units of the lexicon. In fact, they are also morphological units (since they consist of morphemes) and even phonological units (since they have stress patterns, etc.). So, which of their characteristics are important from a syntactic point of view?

4.3.2. Categories of words

Take a look at the following lists of words:

A) *girl, mountain, grandfather, cat, table, happiness*
B) *take, arrive, send, give, understand, memorise*

All the words in A) can be preceded by *the*, *a(n)* or by *this* or *that*. For example:

12) *The grandfather thought of the happiness the girl had felt after finding a cat under that table.*

The words in B) cannot occur in these combinations: **the take*, **this understand*. (The * indicates ungrammatical expressions.) They, however, can be preceded by words like *can, must, will*, etc. For example:

13) *It will give me a headache if I must memorise something that I can't understand.*

But **must mountain*, **can cat*, **will girl* are ungrammatical. Besides these, there are other markers which distinguish the two types of words. Phrases like *the girl*, *my grandfather*, *this cat* can be replaced by words like *he*, *she*, *it*, e.g.,

14) *My grandfather was a teacher. He lived in a small village.*

Can anything replace the words in B)? Look at this dialogue:

15) - *They never arrive on time.*
- *I think this time they will.*

By this time you may have recalled from your previous studies that the words in A) are nouns, while the words in B) are verbs. Nouns, verbs and the other word classes (auxiliaries, pronouns, articles, etc.) are characterised by the way they form combinations with each other. Articles occur before nouns but not before pronouns; auxiliaries occur before verbs but not before nouns; combinations formed by an article and a noun (i.e., noun phrases) may be substituted by pronouns; noun phrases, as well as pronouns, may be preceded by prepositions (e.g., *under the table*, *for him*), to mention just a few possibilities.

4.3.3. Subcategories

So far, we have seen that members of the same word class share certain syntactic properties. Still, there is variation within word classes as well, i.e., there are subcategories. For example, certain verbs must be followed by a noun phrase or a pronoun or a clause, which functions as the object of the verb. These verbs are called transitive verbs:

16) *Linda bought a toothbrush.*

But:

17) **Linda bought.*

There are also verbs which cannot have objects. These are intransitive verbs:

18) *Linda was going.*

Nouns give us another example for subcategorisation. There are proper nouns (such as *Anna*, *London* or *The Metropolitan Opera House*), each of which denotes one particular person or object and common nouns (*teacher* or *stone*), which may denote any member of a class. Common nouns are further divided by their ability to take the indefinite article *a(n)* or to occur in plural forms, e.g., *a tree*, *five trees*. Nouns like *tree* are called countable (or count) nouns. Other common nouns are uncountable (or non-count): e.g., *information*, *furniture*. These cannot be preceded by *a(n)* and do not have plural forms, so **informations* and **an information* are ungrammatical.

4.4. From words to phrases

4.4.1. An example

What we have seen so far indicates that words are organized into larger units in accordance with certain rules. Since combinations of words are called phrases, such rules are called phrase structure rules. This implies that phrases have structures; for example, they have obligatory and may have optional constituents; but each phrase has a central, most important constituent, which determines the syntactic properties of the phrase. The term *noun phrase* has been mentioned above; this is the phrase type we will see in some details below. Let us examine a noun phrase e.g.:

the young kangaroo

4.4.2. Testing our phrase

How do we know that it is a phrase, i.e., a structural unit at all? To answer such questions we may use syntactic tests. Look at the following sentences:

19) *The tourist saw the young kangaroo yesterday.*
20) *The young kangaroo was seen by the tourist yesterday.*

These two related sentences (an active and a passive one) consist of mostly the same words, though some of these words appear in different places in the two sentences. Still, the constituents of our phrase remain together in both sentences (while e.g., the word *yesterday*,

which does not belong to the phrase, follows the word *kangaroo* directly only in 19) but not in 20)). Furthermore, if we wanted to use a pronoun instead of *kangaroo*, we would have to replace the whole phrase:

21) *The tourist saw it.*
22) *It was seen by the tourist.*
23) **The tourist saw the young it.*
24) **The it was seen by the tourist.*

These facts indicate that the string of words *the young kangaroo* really forms a syntactic unit, which is itself a constituent in the sentences above; therefore, we may conclude that it is a phrase. (The same method could be applied to prove, for example, that the string of words *tourist saw the* is not a phrase.) Now we may examine the structure of our noun phrase.

4.4.3. Phrase structure

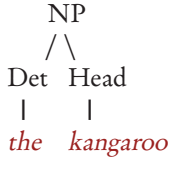
The phrase consists of an article, an adjective, and a noun. Which of these three words will be the most important constituent of the phrase? Could any of these constituents be left out? It seems that only the adjective (*young*) is optional: without it the phrase is still grammatical:

25) *The kangaroo was seen by the tourist.*
26) **The young was seen by the tourist.*
27) **Kangaroo was seen by the tourist.*
28) **Young was seen by the tourist.*
29) **Young kangaroo was seen by the tourist.*

It is also important that the verb which follows our phrase in 25) must agree in number with the noun phrase (which is the subject of the clause) and the number of the phrase is determined by the number of the noun. Cf. 25) and the following sentence:

30) *The kangaroos were seen by the tourist.*

The importance of the noun in this phrase justifies the fact that the phrase is called a noun phrase and the noun (i.e., the most important constituent) is the head of the phrase. The other obligatory constituent in a noun phrase (this time represented by an article) is called a determiner. Optional constituents (e.g., adjectives) are modifiers. A noun phrase (NP) containing only obligatory constituents has the following structure: Det + Head

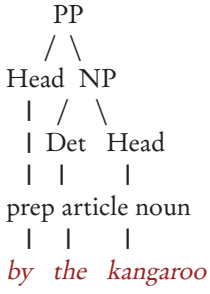


4.4.4. A phrase in another phrase

Of course, much more complicated structures are possible, for example, a phrase can be a part of another phrase:

31) *The tourist was seen by the kangaroo.*

Now our NP is preceded by a preposition with the result that the NP and the preposition together form another phrase: a prepositional phrase (PP), which has the following structure:



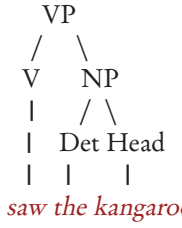
As the tree diagram shows, the head of a prepositional phrase is a preposition. Similarly, the heads of verb phrases are verbs, the heads of adjective phrases are adjectives, and the heads of adverb phrases are adverbs.

4.4.5. On verb phrases

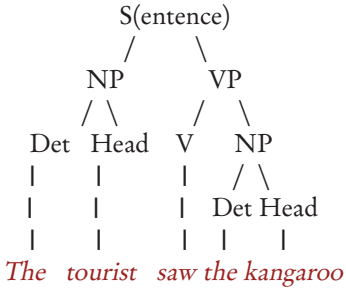
Textbooks on syntax discuss the various phrase types in detail. Do not be surprised, however, if you find that different textbooks use different terminology for the same structures, or, in other cases, if the same term is used in different meanings in different textbooks. A fairly recent theory, for example, considers what we have discussed as nouns phrases to be determiner phrases (DP), claiming that the most important constituents of these phrases are determiners. The arguments which form the basis of this theory cannot be discussed here; interested students of syntax will find the relevant books. This is simply to remind you that linguistics, as any branch of science, develops through the introduction of new approaches and theories, which often cast new light on the studied subject matter, or introduce new points of view to research together with new terms.

The term *verb phrase* (VP) is a term which is used in different meanings by different theories. Generative linguists use it when they refer to the verb and the constituent(s) following it (such as an object NP). An example of a VP and its structure in this meaning is the following:

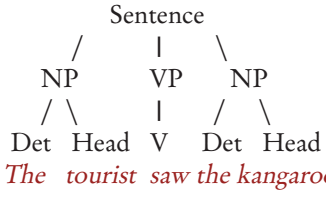
32) *The tourist saw the kangaroo.*



Consequently, the clause in 32) consists of the following phrases:



In other textbooks, however, you may find that verb phrase means the combination of the verb and its auxiliaries e.g., *may have gone* or *will be seen*, and the object does **not** belong to the VP. In consequence, our sentence may be represented as follows:



This, again, exemplifies that a variety of approaches is used in linguistic research.

4.5. From phrases to clauses

4.5.1. Functional clause constituents

As words form phrases, several phrases together form clauses. Phrases have various functions in clauses. NPs may have a variety of functions: they can be subjects (33), objects (34), complements (35) and even adverbials (36):

33) *The kangaroo saw the tourist.*
34) *The tourist saw the kangaroo.*
35) *The tourist was a middle-aged man.*
36) *The kangaroo saw him this morning.*

Saw and *was* are, of course, verbs. Subjects, verbs, objects, complements and adverbials are clause constituents. Subjects usually come before verbs and they typically (though not always!) represent the agent, i.e., something or somebody that does the action expressed by the verb. Objects typically follow the verb and they typically refer to something or somebody that undergoes the action expressed by the verb. 33) and 34) consist of exactly the same words; still they are different sentences, because *the tourist* and *the kangaroo* have reverse functions in them. Adverbials express various circumstances of the action, such as time or place (and many others). Though, as 36) has shown us, an adverbial can take the form of a noun phrase, the typical phrases functioning as adverbials are either adverb phrases, such as *usually* or *very politely*, or prepositional phrases, e.g., *in the park*, *at Christmas*. Complements have two types: subject complements (e.g., in 35) above), which occur after verbs like *be*, *become*, *seem*, etc. and object complements, obligatory constituents (other than adverbs) after verb + object. Both types of complements can be noun phrases or adjective phrases, e.g.:

37) *Charles has become old.* (subject complement)
38) *The song made Julia a star.* (object complement)
39) *I found the play a little boring.* (object complement)

4.5.2. Phrases in clauses

At this point, it may be useful to emphasise that clause constituents are realised by phrases, not by single words (nouns, adjectives, etc.). Therefore, if we analyse the sentence:

40) *The unbelievably impatient young actress offended the elderly secretary.*

We will find the subject is *the unbelievably impatient young actress*, (rather than just *actress*!) while the object is *the elderly secretary*. To prove that these constituents are realised by whole phrases and not only by the heads of phrases, we may try the substitution test:

41) *She offended her.*

If we accept that in 41) the pronoun *she* is a subject, while *her* is an object, we must suppose that what they replace is nothing more or less. (Interested readers may try the passive transformation test as well, which will lead to similar results provided they know that the passive transformation involves a movement by the object of the active sentence to the subject position of the corresponding passive sentence.)

If clause constituents are always phrases, what can we say about the pronominal subjects and objects in 41) above? The answer is that phrases consisting of only one word are perfectly possible; after all if there are one-word-units which are able to function as phrases, they must be phrases. (Another example of a one-word phrase is a proper noun like *Kevin*.) In fact, even a clause may consist of only one phrase, which may, of course, itself consist of one word. In that case, what kind of phrase is *she*? Since pronouns substitute noun phrases (i.e., they perform exactly the same functions as noun phrases do) in sentences, they must be regarded as noun phrases. Of course, these phrases are not typical noun phrases in the sense that they do not even contain nouns (let alone determiners); but, to mention a comparable example from zoology, ostriches are not typical birds, since they cannot fly; nevertheless, they **are** birds, as you would probably agree.

4.6. From clauses to sentences

4.6.1. Coordination

In the previous section, we were discussing clauses, though our example clauses could have been called simple sentences as well. However, not every clause can be called a sentence and, consequently, sentences are not always simple: they may consist of several clauses. The following sentence is an example of a sentence with two clauses:

42) *Bob wants to go out but Jane would rather stay at home.*

This sentence comprises two clauses of equal rank: neither is subordinated to the other. The relationship between the two clauses is called COORDINATION. We may also say that the two clauses are coordinated, or that our sentence is a compound one. The conjunctions that link coordinated clauses (*but, and, or*) are coordinators.

4.6.2. Subordination

In other cases, the clauses of a sentence are not equal: one (or more) may be subordinated to the other. SUBORDINATION means that the subordinate clause functions as one of the constituents (subject, object, complement or adverbial) of another clause. The clause which is not subordinated to any other clauses is the main clause. Since sentences can, in theory, be infinitely complex, it is not surprising that a sentence may contain several subordinate and/or several main clauses. Here are a few examples:

43) *Sebastian has promised that he will help Anna.*
44) *Sebastian has promised that he will help Anna and Anna has promised that she will help Sebastian.*
45) *When Sebastian arrived, he said that he had brought something for Anna.*
46) *Anna says that she wanted to buy the CD as soon as she saw it in the shop window.*

43) consists of a main clause and a subordinate clause functioning as an object. (Cf. the answer to the question: *What did Sebastian promise?*) In 44), we find two (coordinated) main clauses, each of which has a subordinate (object) clause. In 45) one main clause (*he said*) has two subordinate clauses: *When Sebastian arrived* is an adverbial clause (indicating time), while *that he had brought something for Anna* is an object clause. Finally, 46) is a sentence in which the clauses to *buy the CD* (object) and *as soon as she saw it in the shop window* (adverbial) are subordinated to a clause (*that she wanted*) which is itself subordinated to the main clause *Anna says*. The word *that*, which introduces subordinate clauses, is called a subordinator.

4.6.3. Finite and non-finite clauses

In 46) you may have noticed certain structural differences between the last two (subordinate) clauses. Take a look at the following sentences:

47) *Anna promises [that she will help Sebastian].*
48) *Anna promised [that she would help Sebastian].*
49) *Anna promises [to help Sebastian].*
50) *Anna promised [to help Sebastian].*

If we examine the sentences, we will find that the change from present to past in the main clauses in 47) and 48) resulted in a change in the verbs of the subordinate clauses as well; while no similar change happened to the subordinate clauses in 49) and 50). Furthermore, we may notice that the subordinate clauses in 47) and 48) are introduced by *that*; the other two subordinate clauses are not. Still another difference is the presence of the subject *she* in 47) and 48) and its apparent absence in 49) and 50). Clauses of the type represented by the subordinate clauses in 47) and 48) are called finite clauses and the corresponding verb forms (such as *will help* or *promised*) are finite verb forms; while clauses of the type found in 49) and 50), as well as the corresponding verb forms (such as *to help*), are called non-finite.

4.6.4. Finally...

If you have successfully reached the end of this chapter, you may now take the opportunity to test your general understanding of the concepts outlined here, or to try to generate some thoughts of your own in connection with syntax by reading the points below.

Points to Ponder

1. The underlined phrases in the following sentences are ambiguous. Explain the differences in structure that can account for the ambiguity.

1) *This is a toy shop.*
2) *Old English literature societies usually have their own websites.*
3) *Mr Smith is too old to visit.*

2. Compare the following sentence pairs. In the second one you will always find that the place of a constituent is empty. Do you think empty categories may have functions in syntax? If so, explain their functions in these examples.

1) A) *Sue wanted him to leave.*
   B) *Sue wanted Ø to leave.*

2) A) *The lions are hungry.*
   B) *Ø Lions are dangerous.*

3) A) *They have seen her.*
   B) *They Ø saw her.*

3. In English finite clauses the presence of a subject is obligatory. In some other languages, such as Hungarian, subjects are often dropped. Can this difference be explained by other structural characteristics in the two languages? What do you think?

Suggested Reading

Fromkin,V. - Rodman, R. (1998[6]): *An Introduction to Linguistics.* Holt, Rinehart and Winston: New York

A reader-friendly and fascinating introduction to all aspects of linguistics.

Graver, B.D. (1986[3]): *Advanced English Practice.* OUP: Oxford

This book is a collection of really advanced practical exercises for those who want to understand syntactic structures, as well as to use them.

Radford, A. (1997): *Syntactic Theory and the Structure of English: a Minimalist Approach.* CUP: Cambridge

The book introduces the concepts of syntax on the basis of a branch of generative linguistics with easy-to-follow explanations, guiding the reader step by step from the basic terminology to the more complicated ideas.

Since we use language to communicate with each other, it is clear that we will want to understand how language can pass on 'what we mean'. Where can we find meaning in language? How is it that we understand what others mean? Sometimes what the other person says is ambiguous, that is, it can be understood in two different ways. There are two types of ambiguity, lexical and structural. A simple example of the former is "I saw her run to the bank". From this sentence, we don't know if the bank is a financial institution or the edge of a river. Of course, in a real conversation we would probably know. Structural ambiguity is different. If, for example, I were to say, "I ate the biscuits on the bed", you wonder whether I was on the bed while I was eating, or if the biscuits I was eating had been on the bed. A lot of humour works because we are good at interpreting words and sentences in several ways. Of course, taking a joke to pieces to try and find out how and why it works does rather spoil the fun. Studying semantics makes us aware of just how difficult it is to know what meaning is. This is something that every child who has read Lewis Carroll's book Through the Looking Glass knows. Alice is having a frustrating conversation with Humpty Dumpty who is not being clear about what he means and won't accept Alice's objection:

"When I use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean – neither more nor less."
"The question is," said Alice, "whether you can make words mean so many different things."

And that indeed is an important question. After reading the next chapter you will be able to think about words and lexical semantics in a new way.

Péter A. Lázár
Eötvös Loránd University SEAS
English Linguistics Department

5. Words, Meanings, and Their Relationships

Lexical Semantics

5.1. Linguistics — Semantics — Lexical semantics

SEMANTICS is the study of (the various kinds of) meaning in language. Because it has links to logic, mathematics, and philosophy, not all aspects of it are suitable for even a simple presentation in this book. There are highly mathematicized branches within semantics, such as formal semantics, the semantics of logic, etc. (these mainly deal with sentences). Semantics may also deal with DISCOURSE ANALYSIS and PRAGMATICS. Pragmatics is often treated as part of semantics, but when it is looked on as a self-contained branch of linguistics, it is semantics' "neighbour".

Due to its complexity and the centuries-old (linguists', philosophers', logicians') interest in it, meaning has been studied in diverse ways. One narrow, but surely rich enough, area within this field is lexical semantics, which studies word meaning. We, however, include within lexical semantics the meaning of both smaller-than-word units (the morphemes) and larger-than-word units (phrases and sentences).

This chapter looks at the meanings of words, while Chapter 6 surveys how insights from these studies can be used in the making – and, what is more relevant to you, the use – of dictionaries.

5.2. Grammars and dictionaries: mental vs. written

As "grammar" means two things, so does "dictionary". GRAMMAR in one sense is the (knowledge of the) system of language in the native speaker's head. In another, a grammar is a "man-made" description, an account, a model of this system. There are many types of grammars for various purposes, modern and conservative, for natives and foreigners, theoretical and practical, etc. Grammar in the second sense, then, is a model of grammar in the first.

A dictionary or LEXICON is a huge, complex collection or, rather, network of words (and many shorter and longer expressions) in speakers' heads: the MENTAL DICTIONARY or MENTAL LEXICON. It is their knowledge, their individual variety of the collective LEXICON, or VOCABULARY of a language. The secondary sense of "dictionary" should be clear now: the written dictionary, a man-made product. It gives no description or explanation of the mental dictionary – it does not even try to model it. It is also a collection of words. The mind's lexicon is non-linear, an ever-changing (certainly not alphabetised) grid, a network with millions of different types of information for each item listed. There is no such complexity in dictionaries: they are linear, static, and mostly just alphabetical lists. What they contain is rather the WORD STOCK of a language. In the chapter on dictionaries, we will see some things, however, that written dictionaries are better at than mental ones.

5.3. Description vs. prescription

Some grammars tell you what you should (not) say or write, instead of telling you what people do produce in language. This is PRESCRIPTION, instead of a neutral DESCRIPTION of what actually is used. The fate of grammar teaching in Hungarian schools is partly to do with that. You often hear about what is (not) "good Hungarian". Because of this, people may feel that for a linguistic expression to be "good", it must be approved by some authority. That anything not in the dictionary may not even exist. Think about it: could it really be that *vapol* 'use WAP', for instance, is not a Hungarian word because it is not documented (yet)? And did you check all of them? In a game of Scrabble® it is OK to agree not to accept any word that is not in the dictionary (of your choice). But language is nothing like a game of Scrabble®: it is mostly spoken, dynamic, ever-changing, open-ended, and impossible to "fix" or register once and for all.

5.4. What a word is: type vs. token, word form vs. lexeme

So far, we have used the word "word" so you knew what we meant. Or did you? If you do, you will surely be able to count them here:

Some bakers have always known that a baker's dozen is thirteen, while some others don't know that

Most people will – initially – say that the count is 17. So what is the problem?

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Some bakers have always known that a baker's dozen is thirteen, while some others don't know that

Well, some people might insist that No1 & No13 as well as No6 & No17 should be counted as one: after all they are the same word. Then we have just 15 words, like this:

Some bakers have always known that a baker's dozen is thirteen, while some others don't know that

When you count each occurrence of a word in a text, what you are counting are (word) TOKENs. When only different words count, you are counting (word) TYPEs. In the baker sentence, the word (type) *some* is represented by two tokens.

The more careful among you, however, will point out that what is fine for *some* does not work for *that*. The two occurrences of the written *that* do not signal identical copies; they do not have the same meaning or pronunciation, and have a different function. The *that* under No6 has no opposite, while No17 often opposes *this*. No6 often sounds /ðət/, while No17 never has this unstressed form, only the full one: /ðæt/. They just happen to coincide, but are different words. This is a typical pair of HOMONYMS. So how many words are there in the baker sentence? Two identical *some's* but two different *that's*. Back to No16. We have 16 different word types, but since there are two tokens of *some*, the token word count is 17.

Encouraged by the identity of the *some's* but not discouraged by the difference between the *that's*, daring readers now might say that *bakers* and *baker's* are also two "copies", so at an abstract level, No2 and No8 are really the same word, whose two forms are spelt (slightly) differently but sound alike: /beɪkəz/. That would again leave us with 15 words.

Some bakers have always known that a baker's dozen is thirteen, while some others don't know that

When you do not want that, and are considering all forms that differ grammatically, you are counting word forms: *bakers* and *baker's* are two words – word forms – in this sense. If you "collapse" them into one, then these two forms of the noun are one word: what is called the LEXEME BAKER.

An even more radical idea along these lines is that *known* and *know* are the same word. These two forms of the lexeme KNOW are also "the same", although they have different

forms. So how many words? Down to just 14: the two "copies" of KNOW have only been counted once.

Some bakers have always known that a baker's dozen is thirteen, while some others don't know that

When you are not counting (different) word forms but consider the same abstract word only once, you are counting lexemes. Thus *known* and *know* (together with *knows*, *knew*, and *knowing*) are realisations of the same (verbal) lexeme. (The noun *know* – e.g. *be in the know* belongs to another, nominal lexeme KNOW). (The forms *bakers* and *baker's* also make up the same lexeme).

Could we now change points of view and argue that some word in the *baker* sentence is not one but really two? Yes: the word(s) don't. This is traditionally termed a CONTRACTED form. In an abstract sense, we are looking at two words, *do* and *not*. Once, however, the two of them are glued together, they have a pronunciation that is different from those of the components – phonologically this is one "new", independent word. Both get "distorted" in the process: do /duː/ + *not* /nɒt/ → *don't* /dount/. We actually store and retrieve from storage this as one item. Also, it behaves as one word in the syntax. Back to 15, then – because *do* and *not* are sometimes two separate lexemes? The correct answer obviously depends on which sense of "word" we use. You get different answers depending on what you exactly mean.

5.5. What is (not) a word

What does it mean that *don't* – or any LEXICAL ITEM – behaves as one word in, and for, the syntax? How do we know, that is, what is one syntactic word? Here is how: syntactic words are indivisible. They may have "original" components, but there is no way those parts can be separated. Once they are "fed into" the syntax machine, it cannot separate them. Contractions and compounds illustrate this well.

When *do* and *not* are separated, they behave in yes/no questions like this:

People do not use a coffee grinder like that. **Do people ___ not use a coffee grinder like that?**

That is, *do* moves to the front from its position marked with ___, but it leaves not behind: they are different words.

When they are joined, this is how they behave:

You don't use a coffee grinder like that. **Don't you ___ use a coffee grinder like that?**

That is, *don't* (i.e. *do* and *not* together) moves to the front: it acts as one word.

The noun COMPOUND *coffee grinder* contains two nominal components, but after it has been formed, a pair of virtual brackets appears around it and keeps them together. Nothing can come between them; neither member can "wander off"; a syntactic rule cannot "see into" the brackets, that is, cannot refer to just **one** of the components. Phrases do not behave like that.

The expression *grind coffee*, for example, is a phrase. The syntactic rule that produces questions "sees into" it: it can take the question's target – *what* – out of it, and move it to the front:

Phrase: *this gadget* [$_{\text{VerbPhrase}}$ *grinds coffee*] *What does this gadget* [*grind* ____] ?

The rules of syntax, however, stop short at the boundaries of words. You cannot ask the asterisked question in anticipation of an answer, like this:

Compound: *this gadget is a* [$_{\text{Noun}}$ *coffee grinder*] * *What is this gadget a* [___ *grinder*] ?

All this has serious consequences for "wordhood" and may change the way you think about words. For instance, you won't be surprised if you apply what we have just said to *leül* 's/he sits down' and *ülhet* 's/he can sit'.

Which of these is one word – *leül* or *ülhet*? If you have taken the preceding section seriously, then you know that *leül* is two words. Things do come between the two parts; the *le* bit sometimes leaves the *ül* bit; the parts even find themselves in separate clauses: ***Le*** *akartam volna* [*hogy* ***ül****jön*].

Compare that to *ülhet*. Nothing ever comes between the parts; neither part can leave the rest; neither can wander off.

leül –– NOT A WORD	*ülhet* –– A WORD
separation is possible:	separation is impossible:
le is ül	**ül is het*
le se ül	**het se ül*
le Saci ült, én csak...	**ül Saci het, de én...*

5.6. Lexical units larger than words: multi-word lexemes

Native speakers keep in their mental dictionary objects that are shorter than words: these are the morphemes. Mental dictionaries also list objects that are longer than words: *coffee grinder* and *leül* are cases in point. In grammars MULTI-WORD VERBS are often mentioned as examples of such longer-than-word items: sit down, e.g., consists of a verb and a particle, and the two are a PHRASAL VERB. The phrasal verbs of English (along with the prepositional and

phrasal-prepositional verbs); the *leül*-type of "PREFIXED" verb in Hungarian; most compounds in any language; and the IDIOMS of any language – these are lexical units larger than the word, stored in, and retrieved from, memory as units. An idiom is an expression whose meaning is not the total of the meanings of its parts, i.e. not COMPOSITIONAL. Recall that *coffee grinder* was a compound, while *baker's dozen* is an idiom.

Typical examples of multi-word lexemes:

Multi-word verbs: (with idiomatic or non-idiomatic meanings)

phrasal verbs: *go out; put sg on, bring sg up*
prepositional verbs: *look into sg, dwell on sg*
phrasal-prepositional verbs: *look in on sy, look down at sy, zoom in on sg*

Compounds (only nominal type):
red brick, apple core, computer mouse

Idioms: *see the light, not see the wood for the trees, put two and two together*

5.7. Meaning and arbitrariness

One property of language is ARBITRARINESS: this means that there is no connection between form and meaning. The objects of language do not resemble objects "out there": they are ARBITRARY. /maʊs/ is not like the animal itself, and indeed mice have other "names" in other languages. There are many (near) identical-sounding forms (remember: it's the sound not the spelling that counts!) that are associated with different meanings across languages: English *hot* /hɒt/ is 'six', while *heart* is 'back' in Hungarian (if pronounced without the /r/, i.e. /haːt/). *Bär* /beːr/ is (the furry) 'bear' in German, but 'wage' in Hungarian, and even more complicated in English (more about that soon). When two words, each with an arbitrary meaning, are joined together, the joint meaning may be no longer arbitrary, but often the sum of the parts, i.e. compositional – predictable from the parts. Given *coffee* and *grinder*, *coffee grinder* is guaranteed to be 'grinder of coffee'. Not only all idioms, but a good many compound types, as well as many multi-word verbs are idiomatic.

So there is no knowing why *mouse* (or anything else) means what it means? No. This is not a meaningful question if you accept that language is arbitrary. Neither do we really know what meaning as such is. We can, however, ask meaningful questions about the relations of meanings of words (i) among themselves, and (ii) the relations of these to the sentences that contain them.

5.8. Some sense relations

If we learn that

(A) *Andy has managed to find his favourite mouse*

then we will be able to report to some third person that
(B) *Andy has found his favourite mouse*

5.8.1. Entailment

The (A) sentence will guarantee the truth of the (B) sentence. Technically, (A) ENTAILS (B), and the relation is ENTAILMENT. Note that we do know something about the meanings of (A) and (B) and also their relationships without knowing whether Andy's mouse is the hot-blooded or the electronic kind.

If, however, we are told that
(A') *Andy has failed to find his favourite mouse*

then we will also know that
(B') *Andy has not found his favourite mouse*

and again, we need not be clear about the sense of the word *mouse*. Here (A') entails (B').

If we learn that
(C) *Andy has found a dog*

then we will also know that
(D) *Andy has found an animal* Here (C) entails (D)

(if we know the meanings of the words contained)

And if
(E) *Andy has found a Rottweiler*

then we will be also sure that
(F) *Andy has found a dog* Here (E) entails (F)

(provided, again, that we know the meanings of the words contained)

If, in sentences such as (C)–(D) and (E)–(F), where entailment holds in this particular direction – i.e., the (C) sentence entails the (D) sentence, while (E) entails (F) – there is one word of difference between the sentences, then the word in the earlier sentence is a type of the word in the later one: W_C is a type of W_D, while W_E is a type of W_F. Another way of putting this: W_C is a HYPONYM of W_D, while W_E is a hyponym of W_F. In the opposite direction, this looks as follows: W_D is a SUPERORDINATE of W_C, while W_F is a superordinate of W_E.

The relation between such sentences, then, is entailment, while the relation between such words is HYPONYMY, or sense inclusion.

5.8.2. Synonymy

When two structurally identical sentences contain the same words except one (in the same "slot"), and the relation between Sentence 1 and Sentence 2 is mutual entailment – that is, when the truth of the one guarantees the truth of the other and the other way round – then the sense relationship between W1 and W2 is SYNONYMY. If it is true that *Sue has bought a book*, then it *must* be true that *Sue has purchased a book* – and other way round. *Buy* and *purchase*, then, are SYNONYMS. The two sentences that contain them have the same TRUTH VALUE.

If by synonymy, then, we mean that two words have the same COGNITIVE MEANING (also called descriptive meaning or referential meaning), then synonymy does indeed exist – if only to a limited extent.

We simplified above in two ways. The first: synonyms (or just *one* of them) need not be words; they can be any linguistic expression: *look* is synonymous with *take a look*. The second: even when words are involved, it is not the word as such, just one meaning of it that is synonymous with another: *buy into a company* means 'purchase the shares of a company', but there is no **purchase into*; *buy* also means 'believe' – for instance, *I'm not going to buy that* – but here, *purchase* is not its synonym. The idiom *have bought it* is slang for 'be killed', but *have purchased it* just never means that. In sum: it is *not* a requirement for synonyms to meet that they should be identical in all meanings; they do *not* have to display identical behaviour in larger units, including collocations or idioms. Nor is stylistic identity required: even if *buy* "equals" *purchase*, no sane English-speaking person arrives home saying *I've purchased the beer, darling!*

Synonymy is much less frequent than is usually supposed, and is usually over-taught. There's certainly more to it than the dubious Hungarian *kutya–eb*, *kukorica–tengeri* routine. (Have you *ever* used two of these? You know which.) It would not make sense, anyway, for two bits of language to have *exactly* the same meaning!

5.8.3. Two types of opposites: antonyms and complementaries

Another sense relation has already been exemplified: oppositeness. *Manage* is the opposite of *fail*. There, A' entailed the B' sentence, the one containing not. Then *manage* must be the opposite of *fail*.

ANTONYMS are gradable: *big–small*, *hot–cold* are examples. Things can be bigg**er** and small**er**; something *less big* is automatically more small. Antonyms are not an either/or affair; something can be between two extremes: something can be neither. This is when we say that something is "relative": a big mouse is smaller than a small elephant!

The other kind of oppositeness is COMPLEMENTARITY: open–closed and dead–alive are examples. Here, there is no state of being neither the one nor the other. These describe an either/or situation. Nothing can be *more closed* or *less dead*. "Relative" does not come into play here.

5.9. Homonymy vs. polysemy

While words do not typically share the same meaning, it is common for words to mean different things. The technical term for this is AMBIGUITY, and we now look at what this means.

Now, /tʃɜːtʃ/ and /maʊs/; /noʊ/ and /wiːk/; /bæŋk/ are all AMBIGUOUS – in different ways. The string /tʃɜːtʃ/ means 'ecclesiastical building' or 'people of the same religion'; a /maʊs/ is either an animal or a computer accessory. The sound shape /noʊ/ is either *no* or *know*; /wiːk/ is *week* or *weak*, with different spellings (but not before age six for an English-speaking child!). /bæŋk/ means two unrelated things: the financial institution /bæŋk/$_1$ and the river /bæŋk/$_2$.

Traditionally, /tʃɜːtʃ/ and /maʊs/ are POLYSEMY; /noʊ/ and /wiːk/; /bæŋk/ are HOMONYMY. In polysemy, the same polysemous word has several meanings. This is frequent. By contrast, in homonymy, several different homonymous words (often in different word classes) share the same shape. They may be spelt the same way – *bank* – such words are HOMOGRAPHs. Homonymy is less frequent than polysemy – although it is often funny: cf. H. *szerelem* (n) – *szerelem* (v); *követ* (n) – *követ* (n) – *követ* (v).

5.10. Meanings revisited

Our mental dictionaries store huge amounts of (morphological, syntactic, and semantic) information for lexical items. The semantic information includes the various sense relations.

The most elementary of these is hyponymy. That is what the traditional Aristotelian definition is based on. We can define *mouse* thus: 'a rodent that has a pointed snout and relatively large ears and eyes' [modified from the New Shorter Oxford English Dictionary, 1997]. Then, a mouse is a rodent: *mouse* is a hyponym of *rodent*; the meaning of the former includes that of the latter. The definition first places *mouse* in a larger category of rodents, and then describes it within that category.

Be careful: not the objects – animals – themselves are being meant. In their case, inclusion is the other way round: the class of rodents includes the class of mice. Recall that in terms of entailment, this means: The sentence *X saw a* ***mouse*** entails X saw a ***rodent***.

The definition and its relation to meaning, however, raise a lot of questions. When you know this definition, do you know the meaning of *mouse*? Perhaps. And when you know what *mouse* means, will you know this definition? No. (You might not know rodent at all). Do

those who know what *mouse* means have some kind of mental image of mice? Probably. Do they share a similar picture? Likely yes. Exactly the same? Hardly so. What is your mental image of *lecture*? And your lecturer's?

Now pose these same questions to these words: *differ*, *mammal*, and *but*. If you see the difficulties involved, you begin to appreciate the problems facing the study of semantics. It is impossible to have any mental picture of *differ*. You can memorise the definition of *mammal* but can hardly visualise one. You can hardly have either an image or a definition of *but*. So, while meanings in mental dictionaries are very complex and widely differing things, dictionaries simply have to compromise, and provide standardised definitions for all.

Points to Ponder

1. One sometimes comes across estimates as to the number of words in different languages. Think of a few reasons why you should be careful about such claims.

2. What are the Hungarian equivalents for the English *lexicon, dictionary, vocabulary* and *word stock*?

3. Draw Venn diagrams – remember, circles in other circles – for the relationship between animals—dogs—Rottweilers. Now write a definition for all three, and draw Venn diagrams for these. What includes what, in which case?

4. List five words usually considered as synonyms in a Hungarian grammar class (hint: *kutya*). Could you easily add another five? And another five? Do they really have the same meaning in all respects?

5. List five words that have several (un)related meanings (hint: *kulcs*; *kuka*). Can you easily add another (and another, and yet another) five – once you get the hang of it? Try to do the same in English.

6. In Point 7. the example of /beːr/ was mentioned. (In this book, the (r-less British pronunciation of this) word is transcribed as /beə/). Why is this sound shape even trickier than the other examples in the passage? (Hint: it can be associated with two spellings).

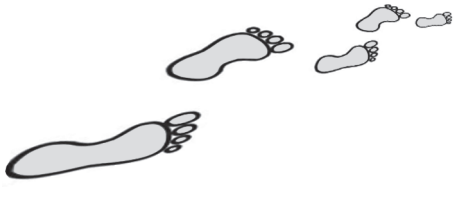
Suggested Reading

Atkinson, M. – Britain, D. et al.(1999): *Linguistics: an introduction.* Part 2, Ch. 12. CUP
An intro to linguistics book, explains basic concepts and recent ideas as well as the applications of these. Is divided into three parts (sounds, words & sentences), has exercises & suggestions for further reading.

Crystal, D. (ed.) (1997): *The Cambridge encyclopedia of language.* 2nd ed. Ch. III–17. CUP
A wonderful volume, offers a wealth of – textual & pictorial – information on all aspects of language & English. Everything you should, and will ever have to, know about language.

Jackson, H. (1988): *Words and their meaning.* Longman
An accessible introductory book devoted to lexicology and lexicography, i.e. meaning in language and dictionaries.

O'Grady, W. – Dobrovolsky, M. – Katamba, F. (1996): *Contemporary linguistics.* 3rd ed. Ch 7. Longman
Another comprehensive – more traditionally organized – introduction to linguistics, looks at not only how language is structured but how it is used functionally & socially.



We may not think about it, but in most of our conversations we tend to use language in predetermined ways. In the following chapter, the knowledge underlying our conversation is examined. If we know what the goal of the person communicating with us is, we should be able to limit the likely interpretations. For example, unless we actually want to deceive the other person, we usually tell the truth. Not that we always keep to the rules! Breaking the rules is one of the sources of humour – and also of irritation and frustration. I recently went to a conference in another town. I had to take a taxi from the railway station but I was a little worried that I would be ripped off. I had heard a lot about taxi-drivers charging huge amounts of money for short trips. So I asked the taxi driver to give me an idea of how much it would cost to get to my destination, which he did. Then I asked if he had a meter and if he could give me a receipt. To both these questions he replied, "Yes." After we had been driving for about five minutes, I noticed that the taxi meter was not running. I asked the driver why this was so, mentioning that I had asked him specifically about the meter. His reply? "You asked me if I had a meter, not if I would use it!" I was furious, feeling that he had tricked me. What do you think? Taking his words literally, he had been honest, but I think he had intentionally deceived me. This chapter will look at how linguists can analyse situations like this and come to general conclusions about how conversation works.

Andrea Kenesei
Pannon University, Veszprém
Institute of English and American Studies

6. What Did You Mean by What You Said?

Pragmatics

6.1. What's pragmatics for?

(1) *Hi, Geoff! Oh, sorry, I forgot to bring you the book you asked for. Will bring it tomorrow, OK?*

(2) *Professor McKenzie, may I ask you for an extension of the deadline for the essay? May I hand it in next week? I've been preparing for an exam in linguistics and haven't been able to complete it. Thank you.*

Examples (1) and (2) tell us something interesting about language and especially the ways we use it. We talk to people differently – we use different words and formulate our sentences in different fashions. It is clear from the above examples that we address our friend/school fellow in a casual way, whereas the addressing of our teacher/educator is much more formal as we must keep the required distance. The relationships with friends and teachers are not the

same, therefore, we must choose the appropriate language to express this. Besides friends and teachers we communicate with family members, shop assistants, doctors, acquaintances and complete strangers – using language differently each time. What factors determine the choices we make in our use of language? They are as follows:

a. Place – classroom, after school, home, shop, doctor's surgery, etc.
b. Time – morning, noon, etc.
c. Our relationships with people – teacher, friend, mother, doctor, assistant, etc.
d. Media – oral conversation, writing a (formal or informal) letter, talking on the phone, email, etc.
e. Situation – class at school, chat with friends, having breakfast, shopping, etc.

6.2. Register – different language choice in different situations

Communicative utterances (complete sentences, fragments, interjections like *Wow*!) are produced by participants whose language use is determined by their social relationships, profession, age, the subject matter and the setting (place and time) of the situation. This particular language is called REGISTER. (N.B.: Hence the expression, *Mind your register!*) Broadly speaking, register is either informal as illustrated in (1) or formal (2). In Hungarian we express this by the address forms – *tegezés*, (*tu* in French) and *magázás*, (*vous* in French). In English, the difference lies in the use of the first name or family name. Following the addressing of the person, we use different modal auxiliaries, lexical items and sentence structures in informal and formal registers. If they are used inappropriately, we either ridicule ourselves or we are impolite; both are to be avoided.

(3) *Can you tell me the time?* (to a friend)
(4) *Would you tell me the time?* (to a person superior to us)

6.3. Interpretation – Is this what you mean? Do you know what I mean?

(5) *"I met the man and his son, coming out of the toilet."*
"I wouldn't have thought there was room for the two of them."
"No, silly, I mean I was coming out of the toilet. They were waiting."

The joke is grounded on something that is so frequent in humour – the introductory sentence can be interpreted in two ways, in other words, it is AMBIGUOUS. However, this ambiguity is theoretical only, our background knowledge makes it possible to arrive at the correct interpretation of the sentence. The reason why we can easily understand such kind of sentences is the particular context, that is, the linguistic and social situation in which it occurs.

The focus of pragmatics is how language (complete communicative stretches) is used in social situations. Therefore, the aim is not that of syntax, for example, which analyses the internal

structure of linguistic units, but approaches language from the external environment. The same utterance gains different interpretations under different circumstances.

(6) *"The party's great! Are we staying a little more?"*
"It's late." [It's time to go home.]
(7) *"They should have arrived by now."*
"It's late." [The train is late.]
(8) *"Let's go to the cinema."*
"It's late." [refusal of an invitation]
(9) *"May I hand in my essay?"*
"It's late." [abandoning a project]

Without knowledge of who makes the above statement, where, when and to whom, or the reference of the sentence-initial pronoun, the statement is meaningless.

6.4. Pragmatic principles: Cooperation makes communication

(10) *A six-year-old child is playing with her ball, which gets behind a row of books and she can't find it. To her question the owner of the books replies: "Why don't you look behind Volume 6 of Dostoyevsky's Collected Works?"*

When people talk, they do this with the intention of communicating something to somebody. This intention involves two things – 1. the speaker wants to put forward ideas / pass on information and 2. the speaker wants the hearer to understand the message. The hearer's intentions are similar – (s)he wants to understand the speaker's information and wants to see why it is sent to him/her. This means that they work together, that is, they are in cooperation in the communicative situation.

Just think about the conversations you have with people – you listen carefully, trying to understand the person who is talking to you, that is, you are cooperative. You also respond to what is being said to you, this also means working together in the conversation, you are still cooperative. Your partner does the same; (s)he says things to you that (s)he thinks you understand; this is cooperation from her/his side. This is what we call the PRINCIPLE OF COOPERATION – without it there is no successful communication. Does it ever happen that you or your partner asks "Pardon?" or "What do you mean?" Of course, it happens every now and then, which means that you haven't been able to express yourselves clearly – you may not have violated the Principle of Cooperation on purpose, however, it has been violated accidentally.

Therefore, the Principle of Communication sometimes seems to be an idealistic picture of communication: people are not always cooperative even if it is not intentional or conscious. Example (10) well illustrates this issue. What is the problem with the answer? Why do we say that the adult is not cooperative? Firstly, the information is not clear – Dostoyevsky does not mean anything to a small child. Therefore, the answer breaches the MAXIM OF MANNER,

which says that we should avoid obscurity (being unknown or difficult to understand) and ambiguity (doubtfulness or uncertainty); we should be brief (short) and orderly (systematically arranged). Secondly, the information is too much (volume, name, title) and too little (not enough to find the toy) at the same time. In this way the answer violates the MAXIM OF QUANTITY, which requires us to be as informative as necessary, in other words, we should not say too much because it distracts the hearer's attention and we should not say too little because the hearer will not understand what we mean.

(11) *Parent: "Someone's eaten the icing off the cake."*
Child: "It wasn't me."

The above example is interesting, as the parent's statement violates one maxim but complies with another. The parent actually attempts to accuse the child of pinching the icing but does it in an indirect way. Thus, although she breaches the Maxim of Quantity, she obeys, however, the Principle of Politeness by avoiding direct accusation. The Principle of Politeness is very often present in our communication - we do not want to hurt the other party so we express ourselves in an indirect way to lessen the weight of what we say.

(12) *"I've lost a diamond ring."*
"Well, Julie was wearing a diamond ring this morning."

This example illustrates the violation of the MAXIM OF RELATION, as the reply is not (does not seem to be) relevant to the topic. This is only a superficial failure because we do understand that the answer is actually an implied insinuation. The MAXIM OF QUALITY is breached when someone makes a statement which is not true or which the speaker is not able to support with adequate evidence. The Maxim of Quality expects us to avoid lies.

6.5. Describing facts vs. changing the state of affairs

A communicative act is produced in three dimensions – intention, utterance and effect. The sender is deliberate in passing on information; then he puts forward the message, which exerts some kind of influence on the receiver. Utterances either describe things or through utterances we make changes in the world. This serves as a ground to differentiate CONSTATIVES (e.g., *I live in Budapest, She likes coffee, or I have two dogs.*) and PERFORMATIVES (see 13). The former are statements describing states of affairs and the latter are utterances which perform actions such as promises, orders, apologies, and so on - actions that bring about changes in the state of affairs. The claim is that constatives can be observed on grounds of truth-falsity, whereas performatives cannot: they are acts themselves.

(13) *I apologise for being rude.* [apology]
I congratulate you on completing your Ph.D. [congratulation]
I approve of forwarding the document. [approval]

6.5.1. Speech acts: Minimal units of communication

Performatives, as stated above, are acts which do things in the world, therefore, they are SPEECH ACTS. Speech acts are truly situation-specific utterances; they are the basic units of communication. The reason is, as described above, that they combine two functions – besides conveying messages they bring about changes in the state of affairs. When saying, "I baptize thee", the utterance, accompanied by the conventional rites, becomes the act of baptising. As in (13), when someone is apologising, it is also an act of apology. Apparently neutral statements also bear communicative forces.

(14) *It's cold in here.*

There is no utterance made without the speaker's intention (Someone close the window!; or, I'm not staying here.; or, Turn up the heating!): this is what we call the illocutionary force of the utterance. The intention manifests itself in the particular utterance – this is the locutionary force of it. The effect exerted is the third force, the perlocutionary (They close the window, or they beg him to stay or they turn up the heating). This tripartite unity is observable not only in speech acts but in every utterance, too. Communication is successful if the perlocutionary force (the effect) and the illocutionary (the intention) coincide, in other words, if the receiver arrives at the same interpretation as intended by the sender. The fact that these three dimensions exist in everything we communicate leads us to the conclusion that the original premises about constatives cannot really be held – eventually, every utterance is a performative.

6.5.2. Direct and indirect speech acts

(15) *I promise I'll be there.*
(16) *I'll be there.*

Is there any difference between the two sentences? On the one hand, there is, but, on the other hand, not really. Structurally, (15) contains a main clause and a subclause, and the verb "promise" constitutes the predicate of the main clause. However, the question that concerns us now is whether (15) and (16) function dissimilarly or not. The pragmatic function of (15) is to convey a promise which is directly expressed by the verb in the main clause. What is the message of (16)? We observe that it is the same as that of (15), since the modal verb "will" expresses promise. Therefore, from the point of view of pragmatic meaning and social aspect, (15) and (16), although they are different in structure, function in the same fashion. Both are promises made to the listener. What does this mean? Do we have to use a performative verb, such as "promise" to convey our intention? As (16) shows, it is not necessary because there are other elements in language to express the same content, and, more importantly, the context ensures the listener's understanding. What else can we say about (15) and (16)? The first sentence conveys the message directly, whereas the second indirectly. This makes us conclude that in language we use direct and indirect speech acts. Also, we realise that most of the time people communicate using expressions that do not express their thoughts in a straightforward

fashion: they imply rather than say things directly. This is the point where pragmatics becomes important, because it examines the intricate network of human communication.

6.5.3. More on indirect communication

(17) *"Let's go to the movies tonight."*
"I have to study for an exam."

How do we (and Speaker$_1$) know that Speaker$_2$ refuses the invitation to the cinema when the answer seems to be about something completely different? The reasons are numerous. Let us see the most important ones. As said above, we use more indirect utterances than direct ones, therefore, we are used to understanding information which is usually hidden or, as pragmaticians say, implied. Let us see the steps that help us understand the process of interpretation.

i. Speaker$_1$ and Speaker$_2$ have made two statements.
ii. Speaker$_1$ is sure that Speaker$_2$ has said something that meets the requirement of the Maxim of Relevance, i.e., he is cooperative.
iii. Speaker$_1$ has made a suggestion. The expected answers are: "Fine!" [acceptance]; Sorry, I can't." [rejection], "Why don't we go to the zoo? [counter-suggestion].
iv. What Speaker$_2$ says is none of the above.
v. Does Speaker$_2$ mean more or something different?
vi. It is common knowledge that preparation for an exam takes time and going to the cinema is waste of time.
vii. It is logical to assume that Speaker$_2$ cannot afford both activities.
viii. Speaker1 makes a proposal because he is able and willing to do it.
ix. Speaker1 knows that what Speaker$_2$ says expresses his inability to accept the pro posal.
x. The conclusion is that the intention of Speaker$_2$ is to reject the proposal of Speaker$_1$.

Although it takes time for you to *read* the steps of understanding how a short dialogue like this is interpreted, the whole process actually flashes through our mind in seconds or even less time. This shows how miraculously the human mind works and that it is pragmatics that tries to make use of linguistics, psychology and sociology to trace the flow of thoughts expressed in and by language.

6.6. Cross-cultural pragmatics – Do we understand each other?

(18) *(a French person's email) Can you tell me how to get from the airport to Rabat?*
(a Moroccan colleague's answer) You can take the train at the airport, with a change at Ain sbaa station and you'll arrive at the Rabat downtown station.

At first sight there is no problem with this conversation: there is an answer to the question. However, the Frenchman sent the email not to request information but to ask for help – he actually wanted colleagues to pick him up at the airport. From the colleague's answer it turns out that she misunderstood the request. Why? Because the Frenchman expressed his wish in an indirect way, which was not perceived by the Moroccan woman. If he had asked her to meet him at the airport explicitly (directly), she would have been ready to do so. This kind of misunderstanding is not rooted in language problems but in cultural differences. Language learners have two types of problems – linguistic and cultural - which result in non-understandings and misunderstandings. The farther the two cultures (the learner's culture and the culture of the learnt language) are from each other, the more problems will occur. These kinds of problems are in the focus of CROSS-CULTURAL PRAGMATICS.

6.7. Summary

As we have seen from the examples and the accompanying explanations, the field of pragmatics is a practical way of looking at how language works and how it is used by people. Our communication is dependent on the setting and the participants, the topic and the media we are using to send and receive information. Speakers send messages in the hope that listeners can follow what they say and listeners make an effort to do so. They need to be cooperative to maintain successful communication, in which we make statements or carry out activities by speaking. Communication is either direct or indirect; however, people are good at interpreting indirectly communicated messages; we apply certain strategies in doing so. Communication is not only language-related; it also contains cultural elements, which are to be learnt by the learner of a language.

Thus, pragmatics is the study of language from the point of view of the users, of the choices they make, the forces they encounter in using language in social interaction, and the effects their use of language has on the other participants in an act of communication. Pragmatics involves semantics, sociolinguistics and non-linguistic context.

Points to Ponder

1. Consider the following excerpt:

"What are we going to do about Baba?" she asked.
"What do you mean?"
"She can't remember anything."
"Did she ask you whether she was taking medicine?"
"No."
"No she's not or no she didn't ask?"
"She didn't ask."
"She was supposed to," I said.
"Well, she didn't."
(Delillo, D.(1986) p.61: *White Noise.* Viking/Penguin: New York)

   a. What is ambiguous in this dialogue?
   b. Where does the ambiguity originate, from a linguistic point of view?
   c. How much context is minimally needed to clear up the ambiguity?
   d. How do the participants resolve the ambiguity?
   e. What is "she was supposed to" referring to?
   f. Do you think the last reply is ambiguous?
   g. How much of it is syntactically and how much pragmatically based?

2. "What's your name?"
   "Betty Skymitch."
   "Spell it, please."
   B - E - T - T - Y

   a. What presupposition is violated?
   b. What makes this conversation funny?

3. I promise to set fire to your house.
   I warn you that you will be awarded the Nobel Prize in literature.
   WARNING: Your lawn will turn brown in November.

   a. What is the problem with these speech acts?
   b. Can you think of any conditions that make any of these speech acts acceptable?

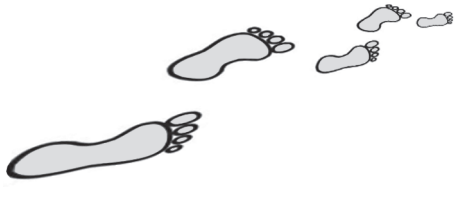
Suggested Reading

Austin, J.L. (1962): *How to Do Things with Words.* Oxford University Press: Oxford
Starting from an exhaustive examination of his well-known distinction of performative utterances from statements, the writer abandons that distinction, replacing it by a more general theory of illocutionary forces of utterances which has important bearings on a wide variety of philosophical problems.

Leech, G. (1983): *Principles of Pragmatics.* Longman: London
This book presents a model of pragmatics: that is, a model which studies linguistic communication in terms of communicative goals and principles of good communicative behaviour. The writer maintains that the language system in the abstract must be studied in relation to a fully developed theory of language use. It builds on speech act theory and the theory of conversational implicature and enlarges pragmatics to include politeness, irony and other social principles of linguistic behaviour.

Levinson, S.C. (1983): Pragmatics. Cambridge University Press: Cambridge
Those aspects of language use that are crucial to an understanding of language as a system, and especially to an understanding of meaning, are the concern of pragmatics. In this book the writer provides an integrative analysis of the central topics in pragmatics – deixis, implicature, presupposition, speech acts, and conversational structure. An introduction and conclusion relate pragmatics to other fields in linguistics and other disciplines concerned with language use – psychology, philosophy, anthropology and literature. Many students in these disciplines, as well as students of linguistics, will find this a valuable textbook.

Mey, J.L. (1993): Pragmatics: An Introduction. Blackwell: Oxford
This is a concise introduction to the field of pragmatics – the study of language from the point of view of its users, of the choices they make, the constraints they encounter in using language in social interaction and the effects their use of language has on other participants in an act of communication. The book includes chapters on what we say when we do not say explicitly what we mean and on the regularities and irregularities of everyday conversation. It aims to introduce the reader to the complexities of language use, and the use of language to social effect.



When we want to use a word and are not sure of its meaning, it is easy to make the mistake of thinking that we can "find out what it really means" by looking at its origins. For example, in the 13th century the English word ***nice*** *meant "foolish, stupid, senseless". Nevertheless, looking at the history of words is a fascinating exercise and sometimes leads us to new inisghts. Although the meaning of words is arbitrary, words of the same language are related to each other and to others in other languages, and each word has its own story or history. The Old English word* ***tunge****, like the word* ***tongue*** *in English today, refers to the muscular organ in the mouth, but also to speech and language. The visually different word* ***language****, meaning the same as it does today, comes from 12th century French. Yet these words have the same ancestor: the Old Latin* ***dingua****, which in turn comes from the supposed Proto-Indo-European word ****dnghwa****. Proto-Indo-European pre-dates writing and is the hypothetical reconstructed ancestral language of the Indo-European family, in others words, of English and German and French and Latin and the Slavic languages and Sanskrit and Greek and … other languages. However, Hungarian is not among them, as every Hungarian schoolchild knows. To find out how we know that* ***tongue*** *and* ***language*** *are related even though they are so dissimilar, read the following chapter.*

Judit Górász
Kodolányi János University College
Department of English Language and Literature

7. The Way from *Tunge* to Language

Language Change and Language History

7.1. Living languages change constantly

7.1.1. Levels of change

Take a look at the following Old English sentence (Old English is the "ancestor" of today's English) and its Modern English equivalent:

Her Cynewulf benam Sigebryht his rices 7 West Seaxna wiotan for unryhtum dćdum buton Hamtunscire.

'Here Cynewulf and the West Saxon wise men deprived Sigebryht of his kingdom for evil deeds except for Hamptonshire.'

Would you have recognised the OE sentence as English? The difference is rather obvious even though some words may seem familiar. If we examine the Old English sentence, we can find evidence of change at all levels of language. Here are a few examples:

PHONOLOGICAL CHANGE (sound change): the y represents a vowel like Hungarian *ü*, which existed in Old English but was lost later.

MORPHOLOGICAL CHANGE (change in the forms or components of words): OE *dæd* 'deed' is present in the sentence in the (plural dative) form *dædum*, while in an equivalent phrase in Modern English there is no case marking (*-s* indicates plural number), what is more, the dative case does not exist at all.

SYNTACTIC CHANGE (change in the way words are combined): the word *unryhtum* ('un' + 'right'), an adjective in the plural dative form, has the same number and case as the noun *dædum*; in other words, the adjective agrees with the noun it modifies. This type of agreement does not occur in Modern English.

SEMANTIC CHANGE (change in meaning): the noun *rices* may perhaps remind you of Modern English *riches*. In fact, the two words are related, but the meanings are somewhat different. The Modern English word means 'money and valuables', while the Old English meaning was 'kingdom'.

LEXICAL CHANGE (change in the "lexicon" or word-stock): as you may have guessed, the words *benam* and *wiotan* have disappeared from the language since the Old English period.

7.1.2. Evidence for change

The branch of linguistics that deals with LANGUAGE CHANGE is called DIACHRONIC OR HISTORICAL linguistics. However, we do not need to be linguists to find evidence for language change. We often hear people say that young people today use words they never heard when *they* were young (for example in Hungarian *szuper* or *ciki*) — as we already know, lexical change is one aspect of language change. When you notice the pronunciation and other differences between British English and American English, you are observing the results of language change. Such a difference is, for example, the fact that in American English words like *short* or *heart* are pronounced with an /r/ sound, while in British English without. Another (syntactic) difference is that in American English the Simple Past is used in many sentences in which the British would use the Present Perfect. Changes like these can add up to major changes: the fact that Latin was a living language two thousand years ago, but is a dead language today; while at the same time we now have Spanish, Italian, French, is a result of language change.

Change is a normal, indeed inevitable feature of every living language: languages changed in ancient times and they change today. Since this is so, language change must *not* be regarded as something bad (as "decline" or "corruption") – it is rather like a natural phenomenon, neither good, nor bad, and it must be accepted as a fact. Nobody would today consider Modern English a "corruption" of Old English (in the same way as a frog is not considered to be a corrupted form of a tadpole). Still, people often resent changes that are in progress in their

lifetime. The only real difference is that the results of earlier changes are familiar to the new generations, while ongoing changes introduce something new to them. This, of course, is not a linguistic difference.

7.2. Dialects and related languages

7.2.1. Dialects

While it may be easy to see that differences between an earlier form of a language and a more modern form of the same language reflect changes in that language, it may not seem to be so self-evident how dialectal differences such as the ones between British English and American English are connected with language change. Every language exists in variation (see Ch. 8 *Sociolinguistics*) and one type of that is geographical (or dialectal) variation. Speakers of the same language in different geographical areas speak different forms (DIALECTS) of the same language. Language change may decrease or increase the differences between dialects. While every language changes constantly, a particular change does not have to take place all over the geographical area(s) where the language is spoken. In fact, not only every language but every dialect has its own history of changes. The more isolated the dialects are from each other, the greater differences may develop between them. The differences may reach an extent where dialects of the same language are not mutually intelligible any more. This has happened, for example, in Chinese.

7.2.2. Language families

This may even raise the question whether the varieties are still dialects of one language or they have become different languages. The borderline which divides dialectal differences from differences between separate languages is by no means easy to define (see 8.3.4.). We do know, however, that dialectal differences have often led to the total separation of the varieties, producing new, independent languages. These languages, in spite of the differences between them, will share a number of similar characteristics as well. Since they originate from the same "ancestor" or PROTO-LANGUAGE, they are said to be GENETICALLY RELATED LANGUAGES, as if they were sisters, i.e., the daughters of the same mother (the proto-language). Such "sisters" are, for example, the Scandinavian languages, including, for example, Swedish, Norwegian and Danish, which separated only during the Middle Ages and still show remarkable similarity today. The Scandinavian languages (the above-mentioned languages as well as Icelandic and Faeroese) form the northern branch of the larger Germanic group of languages. The other Germanic languages are English, Frisian, Dutch, Flemish, Afrikaans, German, Yiddish and several dead languages, for example, Gothic. The Germanic languages are further related to a large number of other languages, which together form the Indo-European group or "family" of languages. The Indo-European family comprises the following main branches:

Albanian
Hellenic (Greek)
Anatolian (e.g., Hittite)
Tocharian
Indo-Iranian
    Iranian (e.g., Persian)
    Indo-Aryan (e.g., Sanskrit, Hindi)
Armenian
Balto-Slavic
    Baltic (Lithuanian, Latvian, Prussian)
    Slavic (e.g., Russian, Polish, Czech)
Germanic
    North (Icelandic, Faeroese, Norwegian, Danish, Swedish)
    East (Gothic)
    West (English, Frisian, Dutch, Flemish, Afrikaans, German, Yiddish)
Celtic (e.g., Welsh, Scottish Gaelic, Irish Gaelic, Breton)
Italic (e.g., Latin)
Romance languages (e.g., French, Spanish, Italian)

7.2.3. Regular differences

Naturally, the more distant the relationship is between any two languages, the more difficult it is to detect what they have in common. With linguistic methods, it is possible to discover the common origin of words even in languages that separated several thousand years ago. (Such studies belong to the area of COMPARATIVE LINGUISTICS.) Words deriving from a common ancestor, such as English *father*, German *Vater*, Dutch *vader* and even Latin *pater*, are called COGNATE WORDS. Since language change is inevitable in all living languages and different languages may undergo different changes, cognate forms will often be rather different from each other. (Compare, for example, Latin *dentis* with English *tooth*.) Furthermore, in the case of words which are similar, mere similarity may be due to various factors, as we will see below. Therefore, linguists look for regular differences rather than any apparent resemblance between cognate words. One example of regular differences is the /f/ ~ /p/ correspondence found between Germanic languages and other Indo-European languages as illustrated below.

Latin	English	German
pater	*father*	*Vater*
piscis	*fish*	*Fisch*
pellis	*fell* ('hide, skin')	*Fell*

7.3. Language contact

7.3.1. Similarities between languages

Regular differences found in large numbers of words can be safely said to originate in the same kind of change in the languages in question (in our present example this change is a regular change from Indo-European */p/ to Germanic /f/) rather than mere coincidence. It is all the more important, since similarities of various kinds can often be found between languages, but similarities are not always the result of genetic relationships. On the contrary, words with similar meaning can have similar phonological forms quite accidentally in any two languages (e.g. Hungarian *ki* – Italian *chi*). In this case there in no regular difference and the number of similar words is not likely to be large enough to result from a relationship between the languages. There are other cases, however, when languages apparently share a considerable amount of related vocabulary (perhaps even structural features) without being related. In such cases the key to the similarities is LANGUAGE CONTACT.

7.3.2. Bilingualism

Languages do not usually exist in isolation from all other languages but are in contact with the languages surrounding them. This contact is realised through the speakers of the languages, who often feel the necessity to communicate with members of other communities, whose native language is different from their own. This communication is made possible by the bilingualism (or even multilingualism) of a large percentage of the world's population. (See Ch. 16 *Bilingualism*) This means that people often use more than one language. The ability to use two (or more) languages means possessing two (or more) grammars and two (or more) sets of vocabulary. In such cases words, sometimes even grammatical structures may enter one language from another. For example a large percentage of English vocabulary is of French origin (*orange, navy, court*, etc.) and even the suffix *–able* entered the language from French, as a result of intensive language contact. In short, languages in contact influence each other. The intensity of this influence depends on the intensity of the contact.

7.3.3. Lexical borrowing

The process when words from one language begin to be used in another language is called LEXICAL BORROWING. Such words in the receiving language are called loanwords. For example, *szoftver* is a loanword in Hungarian from English. Lexical borrowing is an extremely common phenomenon, though certain types of words are borrowed more easily than others. The so-called "content words", i.e., words with lexical meaning, for example, verbs or nouns, constitute an open class in the lexicon of a language, which means that new members (i.e., words) frequently enter the class, while old members may relatively easily disappear. Numerous examples can be found in English, which has borrowed words among others from Latin (*school, data*), French (*letter, veal*), Scandinavian (*skirt, sky*) Italian (*soprano, spaghetti*), Algonquian

(*moose, skunk*), Japanese (*kimono, samurai*) and Hungarian (*hussar*). These are the words which constitute the majority of lexical borrowing. "Function words", however, (the closed class of words with grammatical functions rather than lexical meaning) like pronouns or articles, are less likely to be borrowed: the borrowing of a pronoun indicates closer contact between the languages. An example of the latter is the English pronoun *they* borrowed from Scandinavian in Old English times: Old English and Scandinavian were genetically closely related languages and probably mutually intelligible. The necessity to borrow words from other languages may arise from various sources. New objects or new concepts require new words and the solution is often the adoption of a "ready-made" word already existing in another language. For example, computer language items in Hungarian have been borrowed from English (*e-mail, file, Internet*) or sometimes translated (e.g., *honlap* from *homepage*). (Such loan translations are called CALQUES.) On other occasions the use of a foreign word (regarded as interesting or original) may be due to stylistic motivations (e.g., when *trendy* is used in Hungarian instead of *divatos*); and still another (frequent) situation is when speakers of language A borrow words from language B because language B is spoken by a socially prestigious or dominant community. Numerous examples of the latter can be found in the languages spoken in present-day Russia, which have borrowed words extensively from the Russian language. For example, the word-stock of Erza Mordvin (a Finno-Ugric language) includes loanwords of Russian origin such as, *utka* 'duck' and *volna* 'wave'.

When a foreign word enters a language, it will usually undergo certain changes: if the phoneme inventories of the two languages are not exactly the same (which is typical), the borrowed word will have to be pronounced in accordance with the phonology of the receiving language – unless its use remains rather restricted. Thus "foreign" phonemes are substituted by similar "native" phonemes. For example, the English "*th*" sound is replaced by *sz* in Hungarian (e.g. in the name *Thatcher*). At the same time, the loanword may acquire new morphological forms, i.e., it will be inflected according to the morphological rules of the receiving language. For example, the above-mentioned *szoftver* has received the forms *szoftvert*, *szoftverrel*, *szoftvereim*, etc. in Hungarian.

Moreover, even the original meaning of the loanword may change in the receiving language, as in English the word *china* received the new meaning 'porcelain' (cf. *chinaware*: 'ware from China').

7.3.4. Structural borrowing

However, not only the borrowed word may undergo changes in the receiving language but heavy borrowing from another language may also have effects on the structure of the receiving language. Relatively close contact and a high degree of bilingualism in the receiving community may sometimes result in the adoption of a new phoneme (together with a significant number of words containing it) or at least a change in the distribution of an already existing phoneme. For example, the present-day *dzs* phoneme in Hungarian is the result of 16th century Turkish influence. In Old English *v* was simply a variant of *f*, but the introduction of

numerous French loanwords established it as a separate phoneme. New derivational suffixes (e.g., *-able* as mentioned above) may also be borrowed: if a considerable number of loanwords contain the same suffix, the use of the suffix may be extended to other words as well. It may also happen that loanwords keep their original inflections (e.g., plural forms, cf. English *thesis – theses*, of Greek origin), introducing in this way new inflectional affixes, which is certainly an alteration of the morphological system of the receiving language. The above-described processes are instances of STRUCTURAL BORROWING, which may even extend to the adoption of syntactic characteristics. It must be stressed, however, that structural borrowing requires much closer and more intensive contact than lexical borrowing.

7.3.5. Convergence

Another case when languages in contact influence each other is the situation where several languages in a geographical area develop a series of structural similarities without any strikingly heavy lexical borrowing between them. This development is called CONVERGENCE and the geographical area is called a convergence area (or Sprachbund). An example of a convergence area is the Balkan Peninsula, with languages like Bulgarian (Slavic), Rumanian (Romance), Albanian, and Greek. These languages happen to be Indo-European ones but representing different subgroups within the family and they show certain similarities which otherwise do not characterise their respective families. One such similarity (among many others) is the definite article following (rather than preceding) the noun in several Balkan languages, including Rumanian and Bulgarian, a phenomenon unknown in other (non-Balkan) Romance or Slavic languages. Though the phonological forms of these articles are different in the various Balkan languages, the structural similarity is obvious. The large number of structural similarities among these languages point to the existence of wide-spread multilingualism and extensive interethnic communication in the area. In such a situation, the convergent development of originally rather different languages ultimately facilitates the communication between the language communities: The speakers of one language will easily learn the languages of the others, since they may often use the same structures from language to language, simply substituting the vocabulary of one language for that of another.

7.3.6. The sources of historical linguistics

It should be clear that contact with other languages can result in lexical and even structural changes in a language regardless whether the languages in contact are genetically related or not. This means, on the one hand, that genetically non-related languages may show remarkable similarities if they are or *were* at an earlier point in their history in contact situation and, on the other hand, that genetically related languages may stay or, in time, get into contact with each other after their separation and, through various degrees of borrowing, may develop further similarities in structure and/or vocabulary. What is more, structural similarities may manifest themselves in totally unrelated and geographically distant languages simply because the choices of a language with regard to a structural item are often very few. For example, the

ordering of subject, verb and object will offer a very limited choice in comparison with the large number of the world's languages, which means that many languages will have the same order quite independent of each other. (Typologists have classified languages from various viewpoints and found surprisingly few types.) As a result, a language historian must be rather cautious when trying to reconstruct the history of a language. The task may be supported by written documents in the earlier language forms, the comparison of related languages or dialects, the study of borrowed elements and the examination of present-day alternations in a language as well as a more general knowledge of what is likely and what is *not* likely to happen in a language. In the next section we will see some of the approaches in the history of diachronic research which have had important contributions to the exploration of language history.

7.4. Approaches to language change

7.4.1. The Neogrammarian approach

The 19th century group or "school" of linguists called THE NEOGRAMMARIANS were the first to discover regularity in language change. They noticed that certain types of SOUND CHANGE are not sporadic but tend to happen in large numbers of words (for example, the */p/ > /f/ change mentioned above). They also noticed that present-day irregular forms (such as the irregular plural formation in *foot – feet*) can be derived from earlier regular forms by means of sound changes operating in a certain period in the history of the language. They also described the concept of ANALOGY: a process which helps to regularize language and to eliminate some of the irregularities – a very important device if we know that the regular sound changes themselves are likely to cause irregularity somewhere else in the language (notably in the morphology). The Neogrammarians compared sound changes to the laws of nature and tried to find regular explanations for any apparent exceptions.

7.4.2. The Structuralist approach

Later some linguists found other ways of explaining language change. THE STRUCTURALIST SCHOOL of linguists (dating from the early 20th century) began to see language as a system in which every element has its own place. Consequently, the change in some of the elements will affect in some way the whole system. The Structuralists were especially interested in changes which introduced significant alterations in a linguistic system. One example of these changes is the merger of two phonemes into one, when one phoneme changes in such a way that it becomes identical with another, already existing phoneme of that language. For example, the letter combination *wh* (*which*) in English used to represent a sound /hw/ different from the one spelt as *w* /w/ (*witch*) but later the two sounds merged, i.e., the former became identical with the second. A serious change like this may result in communication problems (e.g., the number of homophonous words may increase considerably) and this might start other changes in other phonemes until a whole chain of changes occurs. The Structuralists

also noticed a preference for symmetry in linguistic systems, which again might be enough to start a chain of changes, once an initial change has taken place – and left a gap – somewhere in the system. One example of a series or chain of changes in English is the vowel change known as the GREAT VOWEL SHIFT. This involved English long vowels (see below 7.5.4.1).

7.4.3. The Generative approach

THE GENERATIVE THEORY (see 4.2.) of language (developed in the second half of the 20th century) regards language changes as changes in the grammar of the language in question. Since grammar is a set of rules, what happens when a particular change occurs in a language is a change in the rules: existing rules may change in certain ways; new rules may be added to the grammar, while others may be lost in a process that involves the acquisition of the language again and again by subsequent generations.

7.5. A glance at the history of English

7.5.1. English arrives in Britain

As we have seen, English belongs to the Germanic group of languages, originally spoken in continental Europe. The ancestor of English spoken on the continent is called Pre-Old English. The separation of this language from the rest of the Germanic family took place in the 5th century, when the tribes known as Saxons, Angles and Jutes left the continent and invaded Britain, which was then inhabited by various Celtic peoples, who – according to medieval chronicles – were currently at war with each other. The name *English*, as the invaders' descendants later started to call themselves, derives from the name of the Angles. Different tribes settled in different regions, establishing a basis for dialectal differences. OLD ENGLISH (or ANGLO-SAXON), as the language spoken before the Norman Conquest is known, is preserved in numerous written documents dating from the tenth century or even earlier, whose largest part was written in the West Saxon (Wessex) dialect.

7.5.2. Old English

7.5.2.1. The structure of Old English

Old English, as we saw at the beginning of this chapter, differed from Modern English at all linguistic levels. The phonological system included certain phonemes that are not present in Modern English (for example, the vowel *y* as mentioned above), while others which now belong to the English phoneme system were not used then or were used only as variants of other sounds pronounced in specific environments. Morphologically, a very important characteristic of Old English was a verbal and nominal inflection system rather rich in comparison with the Modern English one. Nouns were classified for grammatical gender (Masculine,

Feminine or Neuter), inflected for several grammatical cases (Nominative, Accusative, Genitive and Dative) both in the PLURAL and the SINGULAR. Adjectives were likewise inflected and they had to agree with the nouns they modified in number, case and gender. Verbal morphology also showed a greater variety of endings than today. Syntactic patterns were in many ways different as well. For example, unlike in Modern English, questions and negative sentences were formed without the do auxiliary. Clauses in which the object preceded the verb were common, though the today familiar subject–verb–object order was also possible.

The word stock (or lexicon) of English has undergone considerable changes since the Old English period: some words have disappeared; others have entered the language either through internal word formation or from external sources; and most of the words that have been preserved from the vocabulary of Old English have changed their phonological forms and some have changed their meaning. Who would think, for instance, that Modern English *queen* goes back to a word meaning simply 'woman'? Yet, this is the case.

7.5.2.2. Old English in contact situations

Old English, of course, did not exist in isolation from other languages, and the languages it had contacts with left their imprints on English. Latin, for example, was an important source of loanwords: Old English inherited a significant number of Latin loanwords from continental Germanic and the influence of Latin over English continued in Britain as the Anglo-Saxons were converted to Christianity. Another important contact situation was the result of the invasion of Britain by the Vikings (or "Danes"): after a long history of fighting, King Alfred managed to consolidate his kingdom and achieve some kind of peace with the Danes, who settled in the north of Britain and, doubtless, found the means of communication with the English, a process which must have been facilitated by the considerable similarity of the two languages. During this time of relative stability, a huge corpus of Old English texts (chronicles, legal documents, literature, religious writing, etc.) was produced – representing a rich source for linguistic research in modern times.

7.5.3. Middle English

7.5.3.1. Historical background

The next period, MIDDLE ENGLISH, which started at the end of the 11th century, was ushered in by the political event known as the Norman Conquest in 1066. The Normans were French-

speaking people, who, after defeating the English at the battle of Hastings, took over all political power in Britain. As a result, the English-speaking aristocracy was replaced by a French-speaking one. This created a totally new sociolinguistic situation for English, which continued to be spoken as "the language of the people" but lost its earlier status as the language of the royal court, the law, etc. Still, the necessity of communication between the ruling Normans and the oppressed Saxons led to the gradual emergence of a new form of language, neither Anglo-Saxon, nor French, but a new English language, which in time became the language of a nation comprising the descendants of both Saxons and Normans.

7.5.3.2. Vocabulary

It is easy to suspect that the vocabulary of Middle English contained a huge number of words of French origin – and many of these words are still part of the English language. Have you ever wondered why the living animals are called *cow*, *pig* and *sheep* in English while their meat is *beef*, *pork* and *mutton*, respectively? The answer is that the living animals have names of Anglo-Saxon origin, while the names of the meat from them are French loanwords – a curious division, which points out the social differences: those who looked after the animals were Anglo-Saxons, while those who ate them were mainly Normans. Still, the changes that took place in Middle English are much greater than simply the replacement of a part of the Anglo-Saxon vocabulary by French loanwords: very significant changes happened in the grammar as well.

7.5.3.3. Structure

The inflectional suffixes of verbs, nouns and adjectives were greatly reduced in number. This morphological change was the result of a phonological change: namely, that the vowels of unstressed syllables all became /e/, which later changed into /ə/. The change levelled many of the formerly different morphological forms of verbs, nouns and adjectives. One result was the reinterpretation of the remaining s ending found in the plural forms of certain nouns as a general plural morpheme. Another development affected the syntax of the language: clauses with subject-verb-object order gradually gained ground until SVO became the dominant word order pattern. These and other changes, however, did not take place in a uniform way all over the area where English was spoken. Dialectal differences were significant throughout the whole Middle English period.

7.5.4. Modern English

7.5.4.1. The Great Vowel Shift

English entered its next period called MODERN ENGLISH in about 1500. At the beginning of this period (known as Early Modern English) a series of important changes happened in the

vowel system of English. These changes together are called the Great Vowel Shift. It included the raising of long vowels, which means that they started to be pronounced with a tongue position one step higher than before. This explains why English spelling differs so strikingly from other European spelling systems in the representation of vowels: why, for example, the vowel of *feed* (/i:/) is spelt with the letter e in English. The reason is that it goes back to the one step lower /e:/ sound and the spelling still preserves the pre-Vowel Shift situation. A similar change explains the spelling of present-day /u:/ in *boot* as well. The originally "highest" long vowels, original /i:/ and /u:/, instead of being raised, became diphthongs (see Ch. 2 *Phonetics & Phonology*). So, the /aɪ/ in *divine* goes back to long /i:/, as the spelling still shows. Similarly, the vowel in *house* or *town* originates in /u:/; although this time the origin is not indicated in spelling.

7.5.4.2. Other changes

The centuries of the Modern English period have seen the grammar of the language develop into the grammar we know today. To mention just a few examples: the class of modal auxiliaries has emerged with grammatical properties significantly different from those of their earlier forms; the verb-subject order (i.e., inversion) has become restricted to some very specific sentence types; *do* has come to be used as the obligatory auxiliary in questions and negation when no other auxiliary is present. Together with these and other grammatical changes, the sociolinguistic situation of English has also changed. Before, English had been simply the language of Britain but in the Early Modern English period the British Empire started to emerge, which meant the spreading of the English language over continents. In time, new local varieties developed (American English, Australian English, etc. with their own regional and social dialects, of course) besides the British varieties. As a result, English has become a "world language" or a lingua franca used all over the world.

Points to Ponder

1. Try to match the following Modern English words with their earlier meanings. (You may use an etymological dictionary.) What is the relationship between the old meaning and the new one?

Modern English words	**Earlier meanings**
1. bead	a) 'autumn'
2. knight	b) 'dragon'
3. fowl	c) 'boy'
4. worm	d) 'chance'
5. starve	e) 'happy, blessed'
6. harvest	f) 'prayer'
7. silly	g) 'die'
8. fortune	h) 'bird'

(Solution: 1f, 2c, 3h, 4b, 5g, 6a, 7e, 8d.)

2. Among the loanwords borrowed by one language from other languages, we find a significant number of place names. For example, the name *London* and *Avon* (the name of a river in England) are of Celtic origin. In Hungarian, the word *Balaton* originates in Slavic. What do you think is similar in the history of these place names?

3. Language historians who study languages spoken centuries ago may read old manuscripts, which supply information about the words, morphemes or the syntax, but they do not have any sounding material to provide phonological information. At the same time, spelling conventions may vary considerably from time to time and from place to place. In that case, how can linguists find out information about the probable pronunciation of written documents? To what extent is spelling a reliable source of phonological information and what other resources may language historians use?

Suggested Reading

Burrow, J.A. - Turville-Petre, T. (1996[2]): *A Book of Middle English.* Blackwell: Oxford
If you are interested in the Middle English period, you will find this book a useful student guide. Sample texts are included.

McCrum, R. - Cran, W. - MacNeil, R. (2002[3]) (3rd edition): *The Story of English.* Penguin: London, New York
Not a university textbook but a real bestseller on the history of the English language; ideal for those who want to read something entertaining in this field. Accompanied by a BBC documentary series available on VHS and DVD.

McMahon, A.M.S. (1994): *Understanding Language Change.* CUP: Cambridge
An excellent textbook on the various aspects of language change with clear explanations and detailed, easy-to-remember examples. It also introduces many of the important theories on language change, from a critical point of view. Readers are not expected to have much prior knowledge of historical linguistics, though a certain degree of familiarity with the basic concepts of linguistics is recommended.

Pyles, T. - Algeo, J. 1982[3] (1964): *The Origins and Development of the English Language.* Harcourt Brace Jovanovich: New York
If someone with little linguistic knowledge wants to learn about the history of English, this book may be a good choice. It introduces the necessary basic concepts and gives a concise description of the most important changes that happened to English.

Hogg, R. M. (ed.) (1992): *The Cambridge History of the English Language. Vol. I.* (The beginnings to 1066). CUP: Cambridge

Blake, N. (ed.) (1992): *The Cambridge History of the English Language. Vol. II.* (1066-1476). CUP: Cambridge

Lass, R. (ed.) (1999): *The Cambridge History of the English Language. Vol. III.* (1476-1776). CUP: Cambridge
The first three volumes of the series provide information on practically everything you may want to know about the history of English up to the end of the Early Modern Period. The other volumes are worth reading as well.

Knowing what to say in particular situations, and how to phrase what you say, can make a big difference in life. If, for example, a secretary wanted to leave work early to meet her mother at the airport, the dialogue might go something like this:

PR Manager:	***Hello, Cathy. You wanted to see me?***
Secretary:	***Hello, Margaret. Uh, yes, well, I have got a request to make actually.***
PR Manager:	***Mmmm?***
Secretary:	***It's, uh, rather important to me. I wonder if you could possibly let me leave after lunch. My mother is arriving from the US, and I would like to meet her flight.***

*In this case, Cathy's chances are improved by her strategy for making her request: first signalling her intention to make a request (by **I have got a request to make**), then indicating the high status of her request (by the style of phrases such as **I wonder, could possibly**). In this way she creates certain expectations in her boss's mind and also signals her boss's relative power position. She uses a tentative and very polite way of formulating her request. She also adds a piece of personal information. This is a strategy common in many Western cultures. For learners of foreign languages it is important to learn these strategies.*

Other cultures may structure such a conversation differently, for example by starting with the request and adding reasons and expressing politeness after that. But what if Cathy formulated her request in this 'foreign' way?

Secretary:	***Hello, Margaret. I would like to leave work after lunch. My mother is arriving from the US and I want to meet her flight.***

Although this is not wrong or even inappropriate, it is less likely to succeed, and may even be seen as a little cheeky or rude.

This example shows that using language successfully is not just a question of vocabulary and grammar. Using language in society means knowing about the norms adopted in that society and the way they surface in actual language use. In this chapter you will find out more about the relationship between language and society.

Judit Szitó
Kodolányi János University College
Department of English Language and Literature

8. Language IDs

Sociolinguistics

8.1. Introduction

Think of the last time you heard a stranger talk: perhaps you were listening to the radio or received a phone call. Would you be able to form an idea of the person? I suspect you would. You can typically tell the sex and rough age of the person, and whether the language is her mother tongue. Perhaps you could even tell her ethnicity and which part of the country s/he is from. What exactly would you base your opinion on? Accent? Word use? Voice? Anything else?

Ways of talking make a strong impression on all of us. However, the average person is not fully aware of this. He may judge speech as "educated", "elegant" or just the opposite, but he usually qualifies the person who speaks. The result? In a very short time the speaker is categorized. You might want to know why people are like this. This chapter invites you to look behind the curtain of judgments. Whenever one starts to observe what there is outside in the world, he takes a step toward science. Do not be afraid of it, just let your curiosity lead you instead of your instant opinions. Ornithologists study birds' behaviour: this means they observe them and do not think of judging them or forcing them to behave differently; sociologists study society, anthropologists humans, and again do not tell people how they should behave. Similarly, you start "doing sociolinguistics" when you examine how people use language and try to find the things that motivate them. Sociolinguists usually select some features of speech and examine them with relation to society, that is, groups of people and their social behaviour.

In this chapter, you will be invited to find out

- what it means to speak a language / a dialect or "with an accent",
- whether or not the widely held opinion that language is getting more and more corrupt is true,
- what social overtones there are to pronouncing certain sounds such as -r in New York or the *-ng* in *–ing* as [ŋ] or [n]; and
- if (and why) women and men, or the old and the young speak differently.

8.2. What is sociolinguistics?

A student wrote on the Internet: "A language becomes alive ... when spoken by people. It transmits the hearts and minds of a society..." The writer, Sabine, has lived in many countries and has learned several languages, so she writes from experience. Her words express an opinion that many sociolinguists would agree with: language 'comes alive' when used by people.

Sociolinguistics is a relatively new discipline: it was born in the 1960s as a field of study for those who were interested in real language use and function. From the very beginning, sociolinguistics stood in opposition to generative linguistics, also nicknamed "armchair linguistics" after its method of studying made-up sentences. Generative linguists (see Ch. 4 *Syntax*) think that the main question in linguistics is to describe knowledge of language, which is an exclusively human capacity. Therefore, it is enough to describe a single person's idealized knowledge: the person should know the language perfectly, should be placed in a homogeneous speech community, and should never be affected by circumstances such as errors, memory problems or distractions. Obviously, neither language use nor language variation is at issue when one wants to describe a person's knowledge of language under such idealized, homogeneous (and thus nonexistent) circumstances. It is real language use and language variation that sociolinguistics focuses on. Sociolinguists say that language does not exist separately from the community, the users of that language. Therefore, language must be studied as it is naturally used. Also, since speakers of a language form various communities, language will be VARIABLE. This means it will appear in different forms from community to community.

Let us take a popular example to illustrate the manner in which sociolinguistic research is carried out. In the most often quoted sociolinguistic study from the early 60's, William Labov went to three department stores in New York City. These were Saks Fifth Avenue, serving upper classes, Macy's, serving middle classes, and S. Klein, for lower income shoppers. Labov asked the clerks questions that required answers containing two words: "fourth floor". From the way the clerks pronounced the *r*'s he drew conclusions on class differences. The clerks serving the more affluent shoppers at Saks Fifth Avenue said [fɔːθ flɔː], that is, dropped their *r's* far less frequently than the ones at the discount store, S. Klein. Labov found that the upper classes generally use more *r*'s than the middle or lower classes. Another result was that people in general use more *r*'s when they pay attention to their speech. In this research, the frequency of *r*-pronunciation in English indicated a social class. We can say that speakers of English in New York City socially distinguished themselves from others by their *r*-pronunciation.

Similarly to this study, most research carried out in sociolinguistics examines how groups of people actually speak. Researchers often choose communities (such as speakers of New York City), observe in their language use a chosen set of language features called variables (such as the *r* phoneme), and try to relate those variables to various (social or regional) factors in the community (such as lower, middle and upper classes). There may be a lack of correlation between a language form and social categories, but that is also a finding and requires an explanation. For instance in Hungary, Miklós Kontra (1992) found no correlation between the use of *halasztja* / *halassza* and educational level and explains this with the forced restructuring of Hungarian society after World War II. Sociolinguists employ a carefully worked out methodology because they usually want their statements to be valid to many people. For example, they want to be sure that it is true for all upper class people in New York City that they keep their *r*'s in words like *floor* at a frequency between 40% to 64%", and not just the few people they asked. Therefore, they use statistical procedures to carefully select these few from a large group of people, so that the small group, the SAMPLE, is representative of the

large group, the POPULATION. Some sociolinguistic research employs an ETHNOGRAPHIC APPROACH: this means it obtains data from extended (i.e. long-term) contact with only a few people or a small community, and wants to thoroughly describe how the community members communicate. This second kind of research is often used by teachers in their classrooms, folklorists and discourse analysts, or anyone who wishes to collect authentic language data. Studying language used among Gypsy children in Hungary, Zita Réger gave the children tape recorders and asked them to teach the researchers their games. Over 12 years in 13 communities, together with her colleagues, Réger managed to collect about 60 hours of data, the Gypsy Child Language Corpus (Cigány Gyermeknyelvi Szövegkorpusz) (Réger 1987).

In this section, we have seen that sociolinguists are interested in actual language use and language variation, and they study them by either using statistical methods or ethnography. We also learned that people socially distinguish themselves by the way they speak. The variety of language that is related to social groups is called *social dialect* or SOCIOLECT. In the following sections we shall go through major types of variation: regional, social and individual.

8.3. Regional variation

8.3.1. 'Englishes' (+levels of analysis)

The following dialogue is borrowed from a show on a British television channel. A man enters a petrol station shop.

Shop assistant: It's self-serve.
Man: I don't need any petrol, thanks.
Shop assistant: Good. 'Cause we don't have any. Only gas.

Can you see why the short conversation above is funny? In it, the Englishman uses the word *petrol*, and the American shop assistant gas, to name the fuel in cars. The humour comes from the shop assistant's ignorance of the other variety of English. Can you continue listing similar examples? You probably know that British people put on *trousers*, live in *flats* and go to *shops* to buy *sweets*, whereas Americans wear *pants*, live in *apartments* and go to *stores* to buy *candies*. According to the countries or regions where English is spoken as a mother tongue (see Ch. 16 *Bilingualism*), we distinguish at least British (English, Irish, Scottish), American, Canadian and Australian English, but there are a large number of further varieties within those countries as well. In a number of countries which used to be English colonies, such as India or Hong Kong (in Asia) and Nigeria or Kenya (in Africa), English is spoken as an official second language in addition to the dozens of locally spoken vernacular languages. These varieties differ from each other on many linguistic levels, from phonetics and phonology (see Ch. 2 *Phonetics & Phonology* – sounds) through the lexicon (see Ch. 10 *Lexicography* – words) to syntax (see Ch. 4 *Syntax* – sentence grammar). In practical terms there are even more differences, for example, in *orthography* (spelling).

Can you name the most famous English food? Fish and what? Well, it depends on where you are. In Britain, it is fish and chips. But be sure not to ask for *chips* in an American restaurant, for you will not get the kind of fried potato you want. In the USA, the right term is *French fries*. To make things more confusing, however, Americans do use *chips*, but for the popular potato snack which looks very thin and round, and is packaged in a bag. These, by the way, are called *crisps* in Britain. *Chips*, *French fries* and *crisps* are examples of differences on the lexical level. Grammatical differences include what is considered plural (American (AmE) *the police* ***is*** vs. British (BrE) *the police* ***are***) or whether the sick are treated *in hospital* (BrE) or *in the hospital* (AmE). A phonological example is the pronunciation of the *r* sound: in most of England, *r* is not pronounced after a vowel at the end of words (*fathe**r***, *fai**r***) or before a consonant (*he**r**b*, *Ba**r**bie*), but in most of the USA, it is. *Can't* is pronounced with the vowel sound of *far* in Britain but with the vowel sound of *pan* in the States. Orthographic (spelling) differences include *neighbor* (AmE) and *neighbo**u**r* (BrE). Such differences make British, Irish, American, etc. English different *regional varieties* of the same language.

8.3.2. We all speak a dialect

Lexical, grammatical and phonological differences continue to exist on a smaller scale within a country or a region. Regional variants or DIALECTs are the most well-known type of variation and are traditionally studied by dialectologists.

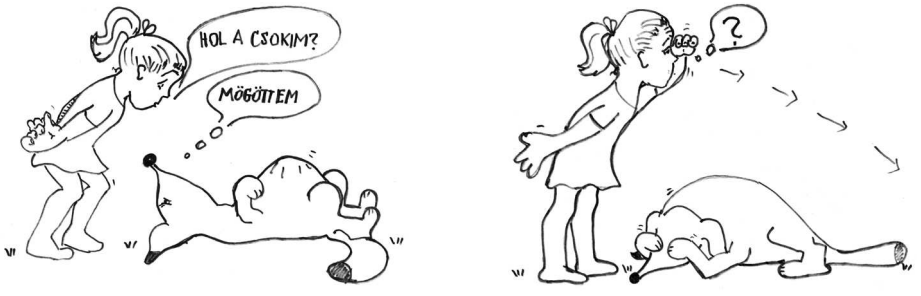
Let us take a phonological example first, the pronunciation of r. Would you please pronounce *father* and *farther*? Now, did you say them differently? Most learners of English do: they pronounce the *r* in *farther*. However, most English people do not: they say [fa:ðə] for both words, because they speak an r-less variety. On the contrary, most Americans keep their r's in words such as *farther*; so they say the two words differently, similarly to you. Nevertheless, this is only a tendency, which means there is regional variation. The *r*-less variety is characteristic of the south of England; but the northern regions, as well as Ireland and Scotland, keep their *r*'s. Likewise, while most Americans in the States do pronounce their *r*'s, some Americans use an *r*-less variety, for instance in Boston or New York (or as some write, "Noo Yawk") City. The variation of the *r* sound has historic origins in both Britain and the United States. (Note: The previously mentioned study of *r*-pronunciation in NYC tells us about sociolects and not dialects.)

The best-known examples of dialectal differences are words and expressions. For *heavy rain*, some say *downpour*, some say *shower* or *toad frog strangler* in the middle and southern states of the American east coast. Some of these lexical varieties (such as the last) occur mostly in the south. You can find out more about lexical variation in the Middle and South Atlantic States of the USA if you click on the LAMSAS link in the Bibliography.

There may be interesting differences in grammar, which appear strange to the outsider. Did you know that an Irishman may say *yous* when he means "you" in the plural (that is, for more than one people)? An American from the South (for example, from Georgia) may use

y'all (you+all) for the same. Standard English does not differentiate between singular and plural *you*.

You probably know very well that your mother tongue, Hungarian, also varies as you travel some distance from where you live. In or around Szeged for example, there are rules for the ö sound to be frequently (but by no means always) used in place of e. You can promise something by saying *möglösz* (instead of *meglesz* 'It will be (done)'). It is common to hear *öszök* (for *eszek*, *eszem*, 'I eat'), or *öttem* (for *ettem*, 'I ate'). So if you have eaten something up you may say *mögöttem* – which sounds exactly like 'behind me' with standard Hungarian pronunciation. People may have different feelings about dialects and accents, yet these varieties continue existing.



If there are a large number of language phenomena (words, grammatical features, sound varieties) that appear only in one region, they distinguish a regional variety called DIALECT. So a dialect is neither "country language" nor some language variety your grandparents speak; it is simply a geographical variety of language. We all speak a dialect of our mother tongue.

Does it surprise you that you speak a dialect? People usually think of others and not themselves as dialect speakers. This is because of the everyday meaning of "dialect": because it is usually considered different from the standard, perhaps incorrect or ill-educated. (You will be even more surprised soon to see that the standard is also a dialect.)

For now it is best to accept that a *dialect is simply a regional variety* in sociolinguistics, and because we usually grow up in one region, we speak that regional variety. Speakers of Hungarian in and around Szeged, for instance, belong to the South Plain Region dialect.

8.3.3. Accent

In the example from Szeged above, it was the ACCENT, that is, the pronunciation pattern that we focused on. Accents represent varieties in pronunciation (phonetics and phonology). An accent tells us which country or part of the country the speaker comes from; accordingly, people are said to have an Irish accent, an American accent, an Edinburgh accent, or a 'Southern drawl' (as they say in the USA). A foreign accent warns that the speaker comes 'from abroad' and is not a native speaker of a language. Since accents are phonetic or phonological

variants, and do not involve lexical and syntactic features, they are not equivalent to dialects. However, dialect regions may be recognized by the speakers' characteristic accent. Received pronunciation (RP) is a highly prestigious accent in England.

So far we have distinguished sociolects (or social dialects, which are social varieties) and dialects (or regional varieties). We also clarified the difference between dialect and accent.Note that we used the word 'language' for larger regional varieties like 'Englishes' (English English, American English, Australian English, etc.), but called smaller regional varieties 'dialects'. We shall see next if there is any difference between them.

8.3.4. Language or dialect?

Is Irish English a language or "just a dialect"? Is it a separate language, like Hungarian, Slovenian or Estonian, or is it just a dialect of English, like London Cockney or Boston English? If the 'thing' spoken in a country is a language, then Irish English should be a language; but one can argue that speakers of other varieties of English can very well understand Irish English, and then, according to this criterion called *intelligibility* (='understandability'), it is a dialect. We cannot really decide if Irish English is a dialect or a language.

It is not always easy to determine if a certain variety should be considered a language or a dialect. Everyday thinking suggests that if speakers from two regions can understand one another (that is, if the two varieties are *mutually intelligible*), they speak dialects; and if they do not understand each other (that is, if the two varieties are mutually unintelligible), the speakers speak two different languages. This criterion often works but sometimes it does not: there

are mutually unintelligible dialects and mutually intelligible languages! In China, for example, Cantonese and Mandarin are mutually unintelligible varieties, that is, the speakers cannot understand one another, yet Cantonese and Mandarin are considered dialects because they are spoken in one country and have a common writing system. On the other hand, Serbians and Croatians, who now live in different countries, understand each other perfectly. In case you think this is perhaps an exception, there will be more examples here. Danish, Swedish, and Norwegian speakers can more or less understand one another; Ukrainian and Russian (and many other Slavic languages) are mutually intelligible; and Hindi (in Northern India) and Urdu (in neighbouring Pakistan) are also mutually intelligible varieties. All these are considered languages, and not dialects, because they are used in different countries or have different historical and cultural backgrounds. For instance, Croats are largely Catholic and use Latin script, whereas Serbs are mostly Orthodox and use Cyrillic letters in their writing.

To make things a bit more complicated, dialects are not really affected or separated by country borders: for instance, German gradually changes as we go from north to more southern regions, even when crossing a country border. Therefore, we can speak about a *dialect continuum*. Nearby languages also become similar, as the Italian spoken near the French border becomes more "French-like" and the neighbouring French variety more "Italian-like". The same can be said about minority languages spoken in a country: ethnic Slovakian or Schwabish in Hungary sound more "Hungarian-like" than 'mainland' Slovakian or German.

As you can see, linguists have tried their best to distinguish dialect and language – but they simply cannot do so by using linguistic means. We cannot really define in linguistic terms where dialects end and languages start. For this reason, and also because of social and cultural factors such as national pride, religion and writing systems, sociolinguists usually let the speakers of that variety decide for themselves if they speak a dialect or a language.

8.4. Individual variation

Is there anyone whose voice you would immediately recognize if you heard it on the phone or on the radio? You can probably think of some talk show host on TV or a politician, or perhaps a member of your family who has a characteristic way of speaking. Perhaps s/he has a slight lisp or speaks fast or uses certain expressions more frequently than others. This type of language variety is called IDIOLECT, and marks individual variation. Although sociolinguistics typically deals with groups, the individual's role should not be underestimated. Researchers usually get data from individuals; this means they meet, ask or contact individuals, who mark their identity by their individual language use.

When children or comedians imitate someone else's speech, they usually choose some of the most characteristic features. You perhaps feel that the way you speak in your mother tongue somehow expresses who you are, that is, your identity. This becomes obvious if you imagine how badly you can be hurt if someone mocks your way of speaking. If you change your way of speaking in any manner (e.g. adopt a new accent or use new words), that means

you adopt a new identity. This often happens to people who move to live in a new region of their country and then slowly accommodate their speech to the local variety. It also happens to anyone who learns a foreign language, especially when trying to imitate its pronunciation and intonation.

For some Hungarians, the rising intonation at the end of "Did you have a good time?" is problematic. Adopting rising intonation in yes / no questions is a strange new behaviour for a Hungarian and some may feel they behave "like a monkey". Changing the ways of talking, and thus speaking a foreign language means creating a new identity.

CHANGING IDENTITIES

8.5. Social variation

Through describing one of William Labov's studies, we showed in section 8.2 that certain ways of speaking could indicate social class, or *socio-economic status*. Linguistic markers of socio-economic status can occur on any level of description, such as phonological (e.g., the pronunciation or dropping of *r*, or the pronunciation of the verb suffix *–ing* as [ɪŋ] or [ɪn]). A linguistic marker that changes its value (depending on the socio-economic status) is called a LINGUISTIC VARIABLE. In the given examples, *r* marks socio-economic status by having a tendency to take on two values: pronounced or dropped. Likewise, the pronunciation of *–ng* (or *–ing*) is also a linguistic variable and has two values, [ŋ] and [n] (or [ɪŋ] and [ɪn]).

Have you ever tried to make out the words of a rapper or a hip-hop song? If you have, you know that African-American people speak a distinct variety of English. The ways African-Americans or Hispanic-Americans or even Hungarians living in a colony in the States speak are called *ethnic varieties.* "We about to get it on" and "But you ain't got nothin' on me" are examples from hip-hop lyrics. Black English Vernacular (BEV) (or as some call it, African-American Vernacular English) is a much-studied variety in sociolinguistics. It was for a long time considered an "ungrammatical", "simple" or "corrupt" version of English, but sociolinguistic research later showed that BEV is able to express complex meanings and has a perfectly systematic grammar, just like any other language. The hip-hop examples illustrate two typical grammatical characteristics: the deleted (missing) "be" and the double (or multiple) negative. Both of them occur regularly and systematically in BEV. This means both the deletion of "be" and multiple negation are perfectly grammatical in that variety of English. Both grammatical features occur in other languages of the world, too (we need not go further than Hungarian), and they do not make any language inferior to others.

Language also varies according to age. If you see films made in the 1950s, you probably notice in them how different the language of conversations is. Also, observe younger or older people talk, and you will quickly notice that they differ from the way you talk with your friends. Sociolinguists think that *age-related variation* is connected to language change, and have shown that teenagers tend to lead innovations.

Finally, sex is also a cause for variation. There are two major topics around *sex and language*. One concerns linguistic inequality and the prejudice against women, in other words, *sexism* in language. Language itself is not sexist, but is able to transmit the views of society. The fact that "he" and "she" are differentiated in English is not sexist – however, to use "he" to refer to people of unknown sex counts as sexism in the English language. There are a great number of words that name the male and only "include the female" as well, for instance, "mankind" or "postman". For "postman" we do not have a female pair, and the existing pair for "mankind" is "womankind", which is used pejoratively. To draw a parallel in Hungarian, present-day Hungarian has had little success in trying to create a female pair for *úr* (sir or Mr.). For example, we can say *Fekete úr*, but no independent female pair (only *Feketéné asszony*); for the female pair of *tanár úr* we can say *tanárnő* (?), which is not quite the same. Back to English: in other cases such as *prince*, *actor* or *hero*, the male form is unmarked and the female form is derived from the male form, with the help of an extra morpheme (*princess*, *actress* and *heroine*). Also, there are more pejorative words for women (usually as sex objects) than for men. This also appears in unequal word pairs, such as the mentioned *mankind* – *womankind*: for example, both *bachelors* and *spinsters* are unmarried, but only the female form is insulting. Researchers working in this field try to educate people so they are more aware of their prejudices.

The other research topic concerns how women and men speak. These are called *varieties related to sex* or *gender varieties*. Perhaps it does not surprise you that men and women talk differently; after all, they distinguish themselves in so many other ways such as clothing, hair length or hobbies. You probably think, you know that women talk much more than men. Well, it depends! When women are with men they usually let the men talk. But the ways they differ are numerous in the world's languages. In Kṻṛux, a small-group Dravidian language in India, verbs have three forms according to the speaker's sex as well as the listener's sex. Men (or women) talk to men using certain verb forms, women talk to women using another set of verb forms, and men talk to women using a third set of verb forms. For instance, the three forms of *you come* is *barday*, *bardin* and *bardi*, respectively. In other languages, it is only the speaker's sex that counts. In French and Russian for instance, the adjective receives a masculine or feminine ending (*Je suis heureux* and *Je suis heureuse* in French or *Ya shastliv* and *Ya shastliva* in Russian for 'I am happy'). Japanese and Thai have very different female and male speech styles, even the first person pronouns differ for men and women. According to one book, Japanese guide dogs are trained in English, because the trainers do not know what the sex of the future owner will be. In English, there are no grammatical, only subtle (=fine) differences according to the sex of the speaker. For instance, when women talk in a formal situation, they use more standard forms. Men, on the other hand, are known to use more local variants, that is, non-standard forms. Even young girls and boys were observed to behave like this: girls used more and boys fewer standard forms of the verb ending *–ing* (standard [ɪŋ] or nonstandard [ɪn]) in words such as *playing*.

This section was about social variation, which included socio-economic status variants, ethnic variants and sex-related (or gender) variants. It also mentioned sexism in language.

8.6. Following norms

8.6.1. Standard

In the previous section we saw that the grammar of BEV is different from the grammar of Standard English. We also learned that women tend to be more status-conscious and employ more standard forms in their speech than men. But what is the standard?

In spite of its added prestige, the STANDARD is just a variety. It is a prestigious dialect that has become distinguished in the course of history, usually by becoming associated with literature, printing and education. For example, southern British English became the standard through the influence of prestigious universities in southern England (e.g. Oxford), where the upper classes sent their children. The standard is also an idealization; therefore, no one really speaks it. A good example of this is Arabic, spoken in various countries in the Near East and Africa (e.g. Syria, Lebanon, Jordan, Saudi Arabia and Egypt). For this language, the norm or standard is set by the Koran (Q'uran), which is a sacred book all Muslims must read. Koranic Arabic is, however, several hundred years old, which means that no one speaks its language anymore: its daughter languages have changed throughout the centuries in about the same fashion as present-day Italian, Spanish, Portuguese and French have developed from Latin. Although no one speaks it anymore, Koranic Arabic remains the norm, a prestigious variety, which all speakers of Arabic hope to follow in educated language use (at school, in writing, on TV).

In sum, a standard variety is a prestigious one usually taught at school and to non-native speakers, spoken by educated people, and used in print and news broadcasts. The standard is not any better than other varieties, because, as sociolinguist Trudgill wrote, standard speakers (if there are any) "swear as much as others".

8.6.2. Descriptive and prescriptive

Speaking about norms: have you ever participated in arguments about what is grammatical or correct to say? In Hungarian, would you say *eszek*, *iszok*, *alszok* or *eszem*, *iszom*, *alszom* to the question *Mit csinálsz?* ('What are you doing?')? In English, some use *fewer people* and some *less people*, some say *didn't used to* and some *didn't use to*. Which is correct? You may have consulted grammar books to find out the correct form, and possibly concluded that people are losing their ability to construct a grammatical sentence. At other times you may have found that even grammar books differed on the same topic. Is it really true that language is getting worse and worse? That would mean the generations following one another carry on an increasingly corrupt language. If you look into this question, you will find that people have been lamenting the fate of their mother tongue for hundreds of years. All those languages are alive and well now, only a little different from what they used to be hundreds of years ago. The fact is that languages vary and change, and for some reason people find it difficult to tolerate these things (see 7.1.2.).

People usually feel strongly about correctness. One of the many reasons for this is that our mother tongue is so much part of our everyday reality that we feel entitled to form an opinion about how it works. Also, rules and correctness are what are taught at school: how to spell words such as Hungarian *gólya*, *játsszuk* or English *though* or *enough*, how to and how not to do certain things such as sentence construction, sentence analysis or verb conjugation (*írnék* or *írnák*). These are PRESCRIPTIVE rules, and the grammar prescriptive grammar, since they prescribe what ought to be done. In Hungarian books for example, *eszem*, *iszom*, *alszom* is accepted and *eszek*, *iszok*, *alszok* is refused, yet a large number of people use the latter forms in everyday conversations.

When we speak, however, we usually do not follow what we learned at school. Five-year-olds can speak their mother tongue very well, although they have not gone to school yet. In the first few years of their lives, children acquire their mother tongue; yet they will not be able to formulate any rules, because they do it unconsciously. In fact, trying to describe this very complex grammar is what many linguists do for a lifetime. The grammars they come up with are DESCRIPTIVE grammars. So when people say *He ain't right* or *Eszek valamit*, they ignore (or are unaware of) the prescriptive rules and instead rely on the ones in their heads. When people observe and describe how others use language, they are said to write descriptive grammars. Everyday thinking about language tends to be prescriptive and judgmental; sociolinguistic thinking tends to be descriptive and factual.

We have seen that certain language varieties may be treated with contempt, whereas others such as the standard are highly valued. Arabic speakers go so far in admiring their standard, Koranic Arabic, as to call their own mother tongues (e.g. Egyptian Arabic or Syrian Arabic) "street languages". They are not alone with this attitude. The idea that certain language varieties or languages are better than others is perhaps as old as humankind and has lead to *language purism*. Language purists want to purify the language from what they think to be corruptions. This has even led to banning languages. Until 1982 Parisian French was established as the standard in France, which meant that other dialects, even languages such as Breton, were banned; all official communication, including schooling, was in French, and if a child was caught speaking Breton at school, he was punished by having to wear a wooden shoe around his neck. Language purism has been highly influential in many European countries up to now. From a purely linguistic (thus descriptive) point of view, however, no variety of language is better or worse: it is only different. The standard is 'just' a prestigious variety, and yes, knowing it is necessary for functioning in a modern society. However, prescribing the use of the standard also gives way to social inequality. The very minimum that professionals, especially ones working with people must do is to tolerate varieties and spread this tolerant attitude.

8.7. Summary and study guide

In this chapter, you have read a few things about sociolinguistics. The present section is trying to help you remember better by providing a short summary and giving some key words for each section.

You learned that sociolinguistics was concerned with natural language use and language variation (and not hypothethical sentences). Because observing language use is only possible when we have data, sociolinguistics relies on specific methods of data collecting: sample, population for data to be processed statistically, and lots of interaction with members of small communities for ethnography. Labov's department store study is an example for the first. If the methodology is good in a study, then the results will be valid for the large population, and not only the small sample.

The next three sections dealt with three types of variation: regional, individual and social. These varieties can all be described on all levels of linguistic description (e.g. phonetics, phonology, morphology, syntax) because variation can occur on any level.

Regional variation includes languages and dialects. Because these two concepts cannot be differentiated by only linguistic means (mutual intelligibility), they are usually distinguished by social and cultural factors (e.g. writing system, religion, history). Accents are not dialects: they involve only pronunciation, people's characteristic ways of talking. An accent tells you where a person is from. A foreign accent warns us that the speaker is not a native speaker of the language concerned.

The individual variety is called idiolect. It marks specific ways of expressing the speaker's identity.

Social variation includes varieties related to socio-economic status, ethnic varieties and gender (or sex-related) varieties. Labov and others showed that *r*-pronunciation or *r*-dropping (a linguistic variable) was related to socio-economic status (higher, middle or lower classes). Black English Vernacular (BEV) is an ethnic variety of Standard American English and its grammar is systematically different from the grammar of the standard. The last social variety discussed was men's and women's speech. In some languages, females and males speak as if they had two different dialects (different verb forms related to speaker's sex, and sometimes to listener's sex as well). In English, women tend to use more standard forms than men.

The last section was about following norms. The first part discussed the standard, which turned out to be a prestigious and idealized variety. The standard sets the norm for people in educated language use. The second part stated that people do not always use their language according to the standard: instead, they follow the grammar in their heads. When we try to describe the rules of a language as it is used, we create a descriptive grammar. This is different from prescriptive grammar, which instructs people how to use language. Language purism is related to prescriptivism. Independently of whether we accept prescriptivism or descriptivism, we need tolerance and acceptance of language variation when we meet people. This is probably the most important thing you can learn from this chapter.

Points to Ponder

1. A tall building has a *lift* in Britain and an *elevator* in the USA; yet *giving someone a lift* in British English (BrE) is *giving someone a ride* (and not *an elevator*) in the USA (AmE). Can you find similar confusing pairs of words / expressions? List them. (See 5.8.2.)
2. West Midlands dialect (England). Translate the following joke about bay windows into standard English, using the glossary provided below:
   "What sort of windas am them?" "They'm bay windas." "Well if they bay windas wot bin them?"
   Glossary: *am* = *are*, *ay* = *is not* (related to *ain't*), *bay* = *are not*, *bin* = *am* or, emphatically, for *are*.
3. Prepare a list of the varieties you may speak, based on the information in the chapter. Can you associate at least one characteristic with each?
4. Click on the Linguistic Atlas of the Middle and South Atlantic States (LAMSAS) website (given under "Kretzschmar" in the Bibliography below). Click on and read "Introduction" (on the left).
   a) "Browse" among the expressions. See what happens if you click on F (or any other category). Get acquainted with the site.
   b) Click on "Analyses" (on the left) and then on "Density estimation maps". You will see a list of expressions that have been put on a map. Click on the first and see some other alternatives for that expression. Possibly also using a map of the USA, can you tell in which state "a week ago Sunday" was most frequently used?
   c) Click on DARE (on the left) and play with it.
5. If you had the opportunity to organize Labov's department store research in your mother tongue, what would be the linguistic variable(s) that could possibly be related to various socio-economic classes? (Think about sounds or words or grammatical forms that are often considered 'incorrect' or 'strange'. They may be indicators of something in society.)
6. Do you agree with the common belief that a linguist always knows what is correct? (Hint: think of the varieties anyone speaks.) Explain why you think what you think.
7. Listen to female speech, especially the voice, on a TV channel in your mother tongue. Then switch to an American (not a British!) channel and listen to the female voice there. Switch to and fro between the channels. Are there any differences between women's voices as represented in the two languages?
8. Have you ever received any comment on your own way of talking? In other words, has anyone ever commented on the variety of language that you use? If yes, what was it and in which variety type would you place it?

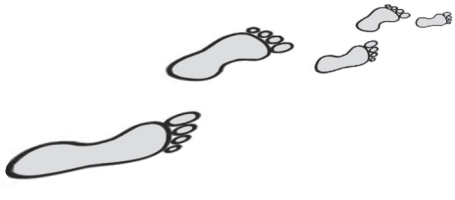
Suggested Reading

Fromkin, V. - Rodman, R. (1998): *An Introduction to Language.* Harcourt Brace: Fort Worth
One of the most popular introductory textbooks on linguistics. Easy to read, great examples and funny cartoons.

Kontra M. (1990): *Fejezetek a South Bend-i magyar nyelvhasználatból.*
A Magyar Tudományos Akadémia Nyelvtudományi Intézete: Budapest
This book is a good example for a sociolinguistic study on the use of Hungarian among émigrés in the USA. Very interesting. The examples will make you smile.

Labov, W. (1966 / 1982): *The Social Stratification of English in New York City.*
Center for Applied Linguistics: Washington
Start reading Labov with this book.

Wardhaugh, R. (2002) (4th ed.): *An Introduction to Sociolinguistics.*
Blackwell Publishers: Oxford / Cambridge
A popular textbook. Very exhaustive. Recommended for those who are interested in sociolinguistics.



If one morning you got up and decided to stop talking, would you lose your knowledge of language? Of course not! Your knowledge of language is internalised, and this is what linguists call competence; what you actually say when you speak is called performance. In the tradition of generative grammar pioneered by Noam Chomsky, it is competence that is the most important and linguists can gain insight into the workings of language by looking inside themselves, by introspection. Another way of studying language is to look at performance, at what people actually say and write. Collections of such data have a long history – think of all the hand-written slips of paper that used to be necessary to compile a dictionary. That was in the years b.c. (before computers). Since then it has become possible to collect and store huge amounts of language data in corpora. This means that it is now possible for linguists to study language with new tools and in new ways. For example, researchers can make use of the same data source repeatedly to study different kinds of language features, or to compare studies of these. This is true for non-native speaker researchers as well. Although it sometimes seems that corpus data and native speaker judgment data are competing ways of investigating language, they can also be seen as complementing each other. You can read about what corpora are, how they are compiled and what they can be used for in the next chapter.

József Horváth
University of Pécs
Department of English Applied Linguistics

9. Keywords in Context

Corpus Linguistics

9.1. Introduction

1. I could not cope with the problem of expressing my ideas in an exact way, consequently I
2. I could not get rid of my second person singular personal pronouns. I continuously gave
3. I could so as to fulfill the requirements of a good essay which is subjective now I know.
4. I tried to be more careful and accurate as a whole. I managed to eliminate most of those
5. I tried to translate expressions word- by-word in lacking an up-to-date dictionary such as
6. I tried to use the language as creatively as I could so as to fulfill the requirements of a
7. I used a lot of abbreviations ("can't" or "isn't") and noteforms (underlining important
8. I wanted a quick result, therefore the presentation of my work was simply awful
9. I wanted to be more wise than I really was. It is best represented by the fact that I wrote a
10. I wanted to have my own special style even if it was ridiculous sometimes to read such

The lines 1 to 10 above are taken from students' essays. They wrote their texts for various courses at the University of Pécs. I asked them to give the scripts to me on computer disks, too. The reason? I wanted to put together several dozen such texts because I was building a

collection – a CORPUS. I wanted to collect their essays to be able to analyze them. The lines 1 to 10 are from that corpus, a collection of English essays written by Hungarian students. (Note that grammar and vocabulary mistakes appear in such student collections.)

As you can see, each line begins with "I", followed by "could" "tried," "used" and "wanted." We do not know who wrote which line, but we can see how the first person singular pronoun is used in each example. In corpus linguistics, such chunks of language use are called CONCORDANCES. They help the learner, the teacher and the researcher focus on particular elements of language. In this chapter on corpus linguistics, you will learn about why this field of linguistics is interesting and how you, too, can benefit from applying its methods and techniques.

9.2. What is corpus linguistics?

Why would anyone bother to collect written and sometimes spoken language DATA? The main reason is that this is the only way we can study language as it is used day in, day out. Just think of the millions and millions of people who are reading and writing in English at this moment, for example – not just you reading this line, but students across the globe, teachers, journalists, all manner of people. They produce and reproduce language. And what the corpus linguist does is record and analyze a tiny part of this mass of NATURALLY OCCURRING LANGUAGE – by compiling a corpus.

According to Leech (1997: 1), "a corpus is a body of language material which exists in electronic form, and which may be processed by computer for various purposes such as linguistic research." On the basis of the billions and billions of words spoken and written down, corpus linguists do the five Ss: they select, structure, store, sort and scrutinize language. In this chapter, we will look at each of these five Ss so that you can do the sixth: study them.

9.2.1. The first S: Selecting

Working with a corpus has to follow a plan so that the resulting collection may be useful for language study. Obviously, of the five stages of corpus work, the first one is the most basic: what should we include? As in every research project, planning is crucial. Of the billions of words one could capture, only a select few can be included, that is, incorporated. The resulting corpus has to represent some type of language use that it aims to describe. Corpus linguistics is an EMPIRICAL field and for us to be able to analyze language that occurs in natural contexts, we have to take account of and select from those contexts. This feature of corpus linguistics deals with the issue of REPRESENTATIVENESS. A corpus has to provide evidence for the language performance of a particular language. We will see two examples of this – the first computer corpus and the largest English corpus today.

The first computer corpus project was carried out by Francis and Kučera in the 1960s in the USA. The result was the BROWN CORPUS, a collection of one million words. It incorporates

written English texts. For this corpus, only parts of texts were selected – that is, no component is a full script. In terms of type, there are informative (non-fiction) and imaginative (fiction) texts in it. That is, for the purposes of this project, one selection criterion was to have two major types of text published in 1960 in the USA. It is a STABLE CORPUS: it has not changed since its development. As such, it is still a useful source of information for language in the last century.

The other corpus is called the BANK OF ENGLISH. If you have used the Collins COBUILD dictionaries, you may already be familiar with it. The Bank of English is the basis of that dictionary. In 1995, this corpus had 200 million words in it – the largest English corpus. As opposed to the Brown Corpus, the Bank of English contains not only written, but spoken language data as well as texts from Britain and other English speaking countries. The team developing it has repeatedly made "the bigger the better" claim. This means that for truly reliable accounts of lexis and grammar, large collections are necessary. In 2005, it had 525 million words in it – more than five hundred times more than the Brown Corpus. One of the most interesting results of a corpus project is that it provides so much objective and reliable data on the frequency of words and phrases. The larger the corpus, the more reliable that data is. Another difference between the Brown and the Bank of English is that the latter is a MONITOR CORPUS. New texts are included in it, and older ones are excluded, so that it always contains the most current texts.

9.2.2. The second S: Structuring

The Brown Corpus has two major divisions: fiction and non-fiction texts are incorporated in it. The Bank of English also has two large groups: written and spoken. But there are other ways in which a corpus can be further divided. It is not only a subdivision, but a feature of the selection process: the structure of the corpus will determine the fine-tuned analyses it will allow.

What other ways are available for structuring a corpus? For this to be clear, it may be best to study Table 1.

By language	monolingual	parallel	
	L1	learner	
By representation	synchronic	diachronic	
	general	specialized	
By text type	written	spoken	combined
By storage	static	dynamic	
By notation	un-annotated	annotated	
By generation	first	second	
By status	set	developing	
By use	linguistic	applied linguistic	

Table 1: A way to classify texts in a corpus

As you can see, there are at least eight ways in which corpora can be defined and further structured. From the previous section on the first S, you already know that both the Brown and the Bank of English are MONOLINGUAL corpora. They represent GENERAL ENGLISH, rather than just a small segment of language users. The Brown, however, is a FIRST-GENERATION corpus, whereas the Bank of English is a SECOND-GENERATION one. All recent corpora belong to the latter category – as they have been facilitated by the increase in computing power. The status of the Brown is set – it is a stable corpus. The Bank of English, by contrast, is developing – it is a monitor corpus.

9.2.3. The third S: Storing

For both first-generation and second-generation corpora, a crucial aspect is where the texts are stored. Obviously, the larger the CAPACITY of a computer, the more data it can hold. Every computer corpus is stored as bits and bytes on a hard disk or other storage medium. This is a technical aspect of the work with corpora. You, too, can see the need for this medium to be reliable. No one who has worked hard on selecting and structuring a corpus would like to lose it if there should be a power-out or other mishap. If you have ever lost a file because, for example, a virus has attacked the system, you will appreciate the importance of this aspect. Corpus linguists or the information technology team that assist their work have to ensure that the data is protected and available for further use.

Besides this level of storage, there are two other criteria to point out. One concerns the law, the other the audience of the corpus. In terms of the legal matters, only texts that have been cleared by the owner of them can be stored in a corpus. For example, when the Bank of English is updated, the team has to procure COPYRIGHT permission for the texts. Even when someone aims to build, say, a corpus of English writing as it appears in blogs, they have to ask for permission from the writers of those blogs. Information on how permission was sought has to be included with the release of the corpus.

The first two aspects of storing corpora were the technological and the legal. The third is concerned with the audience. Yes, corpora certainly have an audience: students, other corpus linguists, people who have an interest in languages. There are several real and virtual forums for discussions about corpus studies, as well as collections of corpora. In recent years, we have witnessed a growth in publicly available corpora. They can be found in that massive virtual reality: the Internet. Even segments of the Bank of English are available for free – and so you, too, can see how these corpora stored and made available for a wide audience can boost your understanding of them.

9.2.4. The fourth S: Sorting

Remember the lines in the first page of this chapter? They are from the student corpus of essays. The lines appear in a CONCORDANCE. In the example, the KEY WORD was the first

person singular pronoun, "I." A concordance program is used for generating the concordance lines from a corpus. All the text is loaded into the program, and then the user can decide how lines should be sorted. You can do the same even without a concordancer. If you have Word, open an English-language file, maybe an essay you have recently written and typed. In this simple activity, your file will be the corpus. Now, choose the "Find" option on the "Edit" menu, and type in the word "the." If you continue to hit the button for the next time the word "the" appears, it will be highlighted and you will see the context of the definite article.

This is similar to how a concordancer highlights keywords. The exception is that each line where the word, in our example now, "the," appeared, would be on the same page, so that the researcher can see the context of each occurrence. In corpus linguistics, the word CO-TEXT is used, rather than context.

The key word can appear at the beginning of the line of a concordance – as you have seen on the first page of this chapter. Wherever they are and whatever method is used for sorting, what matters is that we see the co-text of the keywords. We can see the patterns in which these keywords appear, and the COLLOCATES these words have. Collocates are words that often appear with other words. For example, two frequent collocates of the word "I" in the student corpus were "tried" and "wanted."

9.2.5. The fifth S: Scrutinizing

If you feel that you have become to understand what corpus linguistics is all about, good. Now, it will be even better – for you will learn about what happens after all the other four Ss have been carried out – after the corpus has been selected, structured, stored, sorted. Yes, it will now be analyzed, or, as the title of this section says, scrutinized.

There are as many ways of such scrutiny as there are researchers, but in each such project the following three jobs are done. Word FREQUENCY information is collected, concordances are analyzed, and theories are put forth. To show the first of these ways, I will use the student corpus.

It is called the JPU Corpus, my own collection of student writing from the University of Pécs (which used to be called Janus Pannonius University). It contains over 400 thousand words – a large enough corpus collected by a single person. Table 2 presents a part of the word frequency list.

Rank	Word	Frequency
1	the	32231
2	of	14757
3	to	11602
4	and	10835
5	in	9102
6	a	8526
7	is	6409
8	it	4149
9	that	4123
10	I	3695

Rank	Word	Frequency
11	are	3265
12	they	3195
13	not	3041
14	for	2981
15	be	2916
16	this	2759
17	with	2755
18	as	2732
19	was	2566
20	on	2521

Table 2: The 20 most frequent words in the JPU Corpus

Rank	Word	Frequency
1	students	2164
2	writing	1552
3	essay	945
4	language	898
5	people	773
6	English	747
7	different	746
8	time	729
9	use	680
10	words	660

Rank	Word	Frequency
11	like	651
12	paper	606
13	introduction	587
14	make	554
15	write	553
16	work	549
17	way	539
18	used	531
19	text	524
20	reading	506

Table 3: The 20 most frequent content words in the JPU Corpus

It is clear from the table that the corpus has several essays about language study, particularly about writing.

After working with the frequency information, we can go deeper in the concordances by analyzing them and using them for putting forth theories. An especially useful approach of doing that was developed by a leading expert in the field, Tim Johns. He has worked with students to help them revise their essays and dissertations. A student would come, for example, to get help from Johns on the difference between "reason for" and "reason to." The teacher would show examples, concordances, from a corpus, and help the student see the patterns in which these expressions are used. Having these lines in front of him, the student then can form a theory on the differences and put it into practice in their own writing. Johns has called this the KIBBITZER approach – and has made available 76 such pages of cooperation between student and teacher. Figure 1 presents the concordance lines for "reason for" and "reason to" from Johns.

1. ing to use Mr Yeltsin's poor health as a reason for overthrowing him. The president spen
2. us.`That, he said heavily, `is the best reason for getting the Prince of Wales married
3. er action against EU fraud is not a good reason for holding up legislation which must ge
4. matic venture, he reminded them that his reason for being there was not to act as a role
5. de this year. But that slide is the main reason for fearing higher inflation just when t
6. it was rumoured yesterday that the MoD/s reason for blocking the book was its concern th
7. es. "But the Citizens' is different. My reason for working is to try and enjoy myself.
8. mperfect comprehension. This is just one reason for welcoming the increasing availabilit
9. Gregor Mendel. But that is not the real reason for cutting such people out of your life
10. eeing another soul. This was part of the reason for moving her manufacturing base away f

11. harmonising dialogue. We all have ample reason to be grateful to Spender for his 85 yea
12. ue of agents provocateurs. If we had any reason to suspect that an informant was acting
13. for three days to read it. We have every reason to be grateful to Andrew Davies for serv
14. am Hussein. He is now a democrat and has reason to regret it. >The other is Ali Salim al
15. o suburbia, there has been less and less reason to use the middle of many cities. >The h
16. ts so many times now that we have little reason to believe that this will stop the firin
17. little as possible "because there is no reason to believe that they do it better than b
18. el when he was taken ill, but we have no reason to suppose it was inadequate.' The reali
19. Let us try the second option. Is there reason to believe our intuitions are not genuin
20. egun on the tunnel and there was no real reason to doubt the project's viability - but s

Figure 1: Concordance for "reason for" and "reason to" from Johns's (2000) Kibbitzer page.

There are, of course, many more ways to scrutinize a corpus. If you would like to know more about this field, study the Bibliography and feel what you always are: free. Ask, search – study. In other words: do the sixth S after you have learned about the five S's of corpus linguistics.

9.3. Summary

In this chapter you have read about corpus linguistics – the basics of this empirical field. When preparing for an examination on it, it may be helpful for you to organize your notes around these headings:

The first S: Selecting – representativeness and naturally occurring language; Brown; Bank of English
The second S: Structuring – examples of corpus divisions
The third S: Storing – technically, legally, publicly
The fourth S: Sorting – the concordance, keyword, context, co-text, collocates
The fifth S: Scrutinizing – frequency, concordance analysis, theory
The chapter is not over yet – see the study questions and the bibliography.

Points to Ponder

1. What are the five Ss of corpus linguistics? Describe one of them in detail.
2. Using Google, find a publicly available English corpus. Describe its structure.
3. Look up five verbs with the 4-diamond frequency band in the second edition of the Collins COBUILD dictionary. Write down all the examples of the verbs. Next, open the Bank of English concordance sampler. Type in each of the five verbs and print the concordance lines. On the basis of the concordances, can you say that the examples chosen for the dictionary represent the use of the verbs? If so, how? If not, why not? Write a report of your findings. Length: 400 words.
4. Plan a corpus for English-speaking students who study Hungarian at an intermediate level. They would like to improve their oral skills. What kind of corpus would be best for them? Using terms and processes you have learned about in this chapter, explain your corpus plan. Is there a problem you see that a corpus cannot solve?
5. Do an Internet search to find a free concordance program. Download it from a secure site. Read the manual and load one of your texts into the software. Generate and print the frequency list. Study the list. Give a report of your study and put forth a theory about the way the text can be described on the basis of the word frequency list. Length: 500 words

Suggested Reading

Granger, S. (ed.) (1998): *Learner English on Computer.* Longman: London
A collection of studies that examine European students' written English, based on the International Corpus of Learner English. On http://cecl.fltr.ucl.ac.be you can find the webpage of the developers of the corpus,

Horváth J. (2001): *Advanced Writing in English as a Foreign Language: A Corpus-Based Study of Processes and Products.* Lingua Franca Csoport: Pécs.
Available online at the www.geocities.com/writing_site/thesis site. A study about Hungarian university students' written English, based on the JPU Corpus.

Kennedy, G. (1998): *An Introduction to Corpus Linguistics.* Longman: London
An introductory book on corpus linguistics, including chapters on the design of a corpus, its analysis, and the applications of such analysis. You can read about the author at the http://www.vuw.ac.nz/lals/staff/graeme-kennedy/kennedy.aspx webpage.

Sinclair, J. (1991): *Corpus, Concordance, Collocation.* Oxford University Press: Oxford
A classic book of corpus linguistics by the creator of the Bank of English. You can use a simple online concordance and collocation tool at the http://www.collins.co.uk/books.aspx?group=153 address.

Wichmann, A. – Fligelstone, A. – McEnery, T. – Knowles G. (eds.) (1997): *Teaching and Language Corpora.* Longman: London
A collection of studies, including chapters on the reason for developing a corpus and on the use of corpus learning and teaching. You can send an email to one of the editors, Anne Wichmann, at awichmann@uclan.ac.uk and read about her work on the http://www.uclan.ac.uk/facs/class/humanities/staff/wichmann.htm page.

As a language learner you probably use a bilingual dictionary regularly. Translators depend upon dictionaries to help them understand the texts they have to translate, and also to write their translations. This is especially the case when the person called upon to translate is an amateur or unsure of their own knowledge of one or more of the languages. Yet sometimes the best of dictionaries lets us down, and sometimes using a dictionary incorrectly can lead to embarrassing mistakes. Take for example, the seemingly simple title of a CD-ROM designed to help tourists find out about a town in Hungary: ***XY- város kalauz****. The town has this translated into German, and the title now reads,* ***XY-stadt Schaffner****. Now the German word* ***Schaffner*** *refers to the man who checks your tickets on a bus or train! This is clearly quite the wrong word. Whatever happened here? Well, from your knowledge of Hungarian you can guess that the translator took the wrong meaning of* ***kalauz****. In the next step, the German title is to be translated into English. This process is called relay translation and is sometimes used at conferences where the interpreters can only interpret between certain languages, for example from Danish to English and then someone else interprets from English to Hungarian. So our imaginary translator checks Schaffner in the German-English dictionary, and may come up with* ***conductor****. In the final stage, a Hungarian back-translates the strange and incomprehensible English title to get:* ***XY-város karmestere****!*

Péter A. Lázár
Eötvös Loránd University SEAS
English Linguistics Department

10. What Is in a Dictionary?

Lexicography

10.1. Taking a brief but close look

The purpose of this chapter is to give the reader help in getting about the world – and indeed the vast and not always friendly market – of dictionaries. Lexicography itself is a branch of APPLIED LINGUISTICS: it uses the insights of linguistics to a practical end. This chapter, then, is doubly practical: it shows you how lexicography can be put to use.

On opening a dictionary at random, the layout of pages is the first thing you notice. What a page looks like in layout terms depends, first, on what is within the entries: this is the MICROSTRUCTURE of the dictionary. The structure above the level of entries is the MACROSTRUCTURE. Information can be arranged in several ways both within and across entries.

We have seen in Chapter 5 that both /maʊs/ and /wi:k/ are ambiguous. *Mouse* is polysemous, while *week* and *weak* are homonyms. Alphabetically based dictionaries have just one entry for a word with more than one sense, and two entries for homonymous words. The financial

institution bank$_1$ and the river bank$_2$ are separately entered. When the two homonymous entries are also spelt differently (i.e. are not homographs), obviously they can only go to separate entries. Crucially, homonymy is for the macrostructure, and polysemy for the microstructure.

Homonyms are often themselves polysemous; this is to be expected since polysemy is pervasive. Thus, *weak* itself has 17 meanings, while *week* has 4 senses in the *Random House Webster's Unabridged Dictionary* (CD-ROM), 1999.

To illustrate all this, let us examine the entry for *mouse* in (a) the *Oxford Advanced Learner's Dictionary* on CD-ROM (2005), modified, and (b) *Collins English Dictionary*, 1995. (Layouts have also been slightly modified).

Study the entries below and based on your experience with dictionaries, try to imagine what these dictionaries may be like (in paper format). Also, find different points of view that you can use to talk about dictionaries in general. When you work your way though this chapter, you will have many such criteria at your disposal. At the end of the chapter, you find a "key": two checklists in terms of our classification.

(a) *mouse* /maʊs/ *noun* (pl. mice maɪs/)

1 a small animal that is covered in fur and has a long thin tail. Mice live in fields, in people's houses or where food is stored: a field *mouse* • *a house mouse* • *The stores were overrun with rats and mice.* • *She crept upstairs, quiet as a mouse.* • *He was a weak little mouse of a man.* —see also DORMOUSE.

2 (pl. also mouses) (computing) a small device that is moved by hand across a surface to control the movement of the CURSOR on a computer screen: *Click the left mouse button twice to highlight the program.* • *Use the mouse to drag the icon to a new position.* IDIOMS see CAT

WORD ORIGIN Old English [...] of Germanic origin; related to Dutch *muis* and German *Maus*, from an Indo-European root shared by Latin and Greek *mus*.

(b) *mouse* n. (maʊs), pl. *mice* (maɪs)

1. any of numerous small long-tailed rodents of the families Muridae and Cricetidae that are similar to but smaller than rats. See also fieldmouse, harvest mouse, house mouse. 2. any of various related rodents, such as the jumping mouse. 3. a quiet, timid, or cowardly person. 4. Computer technol. a hand-held device used to control the cursor movement and select computing functions without keying. 5. Slang. a black eye. 6. Nautical. another word for mousing. *vb.* (maʊz) 7. to stalk and catch (mice). 8. (intr.) to go about stealthily. 9. (tr.) Nautical. to secure (a hook) with mousing.

10.2. Reference, encyclopaedia, dictionary

When you do what a LEXICOGRAPHER does – either produce or critically evaluate a dictionary – you are doing LEXICOGRAPHY. (Note that to the average English-speaking user, "dictionary" usually means monolingual dictionary). There are so many different types of dictionaries, however, that there really is no such thing as simply "dictionary". One statement can still be risked: in the most general sense, man-made dictionaries resemble the mental dictionary in a person's head: they list lexical items – both words and objects smaller *and* bigger than words – along with various types of information. Dictionaries are simply tools. They are definitely not meant to be models of mental dictionaries. This is not surprising: we have produced them for centuries but have only recently begun to develop ideas about the human mind which contains the mental lexicon.

Some of what we said in Chapter 5 is only relevant to certain kinds of dictionary. Definition, e.g., is a notion for the monolingual, not the BILINGUAL DICTIONARY. There are, on the other hand, reference works other than dictionaries that also use definitions: ENCYCLOPAEDIAS. With gross oversimplification, dictionaries are reference works about *words* while encyclopaedias are reference works about *things*. Because, however, dictionaries tell you about words-for-*things*, while encyclopaedias obviously cannot help using *words* to describe things, the difference is not so marked. It makes more sense to speak about dictionaries that are more or less ENCYCLOPAEDIC on a scale between the two extremes.

It is best to have some kind of taxonomy of dictionaries so that any claim about them might be easily located. Many typologies using several criteria have been offered; we follow a practical classification (Landau 2001).

10.3. Dictionary types

10.3.1. Age of users

While most dictionaries are obviously for grown-ups, children's dictionaries have an important place in English lexicography. They use a simplified CONTROLLED or GRADED VOCABULARY and go back a long tradition. (The Hungarian "Ablak–Zsiráf", which gives you practice in the early use of reference, is a children's encyclopaedia, while the English likes of it are mostly called "my first" dictionary). These were the precursors of today's learner's dictionaries.

10.3.2. Number of languages

According to the number of languages involved, MONOLINGUAL DICTIONARIES contrast with TRANSLATION DICTIONARIES; some of these latter are BILINGUAL, others MULTILINGUAL. Monolingual dictionaries – called "explanatory" in Hungarian – are chiefly for native speakers.

10.3.3. How much they contain: size

Size refers to how fully the word stock is covered. Size statements are never clearly defined, and notoriously unreliable: some blurbs will mention the number of words (and expressions), some refer to entries or articles, some to headwords, some to headings; yet others mention the number of records, while some refer to the number of lexical units. There is also no general agreement on "entry": just the first word at the beginning of an article, or anything in bold face anywhere. Physical size obviously is of no more help, since page layouts and letter sizes are different, which makes comparison difficult.

By and large, five sizes may be distinguished in current English monolingual dictionaries: dictionaries with upwards of 400,000 entries (e.g. Webster's Third) – these are termed "UNABRIDGED"; dictionaries with 200,000-plus entries such as the *Random House Dictionary* – these are the semi-unabridged ones; college dictionaries with fewer than 200,000 entries; desk dictionaries with 60–80,000 entries; and pocket dictionaries, which are even smaller. (*Abridge* means 'shorten by omissions').

Importantly, the smaller the size, the simpler the structure, the fewer the meanings, and the briefer and simpler the definitions, which may then be replaced by a synonym or two. Also, smaller size guarantees less encyclopaedic character.

10.3.4 . What they contain: coverage

10.3.4.1. Coverage by subject

According to the subject covered, general dictionaries contrast with special(ised), SPECIAL-FIELD DICTIONARIES. These include dictionaries of law, linguistics, economy, computers, medical dictionaries, etc. They tend to be encyclopaedic, and (understandably) normative. The bigger the dictionary, the smaller the difference between general and special-field: this is because the proportion of technical entries in large dictionaries tends to be very high – about 40% in unabridged ones. Indeed, most large volumes are collections of special-field dictionaries merged with a general dictionary. Though the special-field market is narrow, their influence is great exactly because general dictionaries heavily rely on them.

10.3.4.2. Coverage by type of language

Also meant by coverage is another, linguistically more important notion: the type of language covered. Here, general dictionaries contrast with SPECIAL-PURPOSE DICTIONARIES. Their speciality is not subject-related but genuinely linguistic: dialect dictionaries; etymological, pronunciation and spelling dictionaries; usage manuals; synonym finders and antonym finders; dictionaries of taboo and slang; dictionaries of neologisms (i.e. new words); collocational (or combinatory) dictionaries; dictionaries of false friends (English *actual* =/= Hungarian *aktuális*

is the stock example); dictionaries of confusable words – these are all special-purpose works. THESAURUSES and other non-alphabetically arranged works – thematic dictionaries – also belong here.

Coverage can also be considered in terms of the size of the units catalogued: whether a dictionary lists (primarily) words or larger units such as multi-word verbs, idioms, or even longer items such as catchwords, clichés, quotations and proverbs. Dictionaries of foreign words and phrases (mostly French and Latin), e.g. tend to provide longer-than-word units. Dictionaries may also list shorter-than-word lexical units (in use mainly in the written medium), such as abbreviations and/or ACRONYMS.

10.3.4.3. Period of time featured

Along this dimension, DIACHRONIC or historical dictionaries oppose synchronic ones. Etymology may be a feature of any dictionary, but of the etymological dictionary it is an integral part. The largest dictionary of English – the Oxford English Dictionary, OED – provides hundreds of thousands of illustrations of the various senses of words down the centuries, right back to their first occurrence.

10.4. Translation dictionaries

All bilingual dictionaries have a direction, and contain a SOURCE LANGUAGE and a TARGET LANGUAGE: in a Hungarian–English dictionary (direction: H → E; the source language is Hungarian, the target language English. In an English–Hungarian dictionary, SL is English, TL is Hungarian. Most bilingual works are used by two kinds of user: the target-language user and source-language user. If someone is an English-speaking user of an E–H dictionary, s/he is a SL user; an English–Hungarian dictionary has Hungarians as TL users.

10.4.1. Direction, user, function

Bilingual dictionaries can be uni-directional vs. bi-directional: in reality, a two-way dictionary – although sometimes sold (even bound) as one book – is two uni-directional dictionaries combined.

Linked to this is the function of a dictionary (that varies with the occasion). It is either used for comprehension, when messages in a source language are being decoded, e.g. when a Hungarian uses an English–Hungarian dictionary to read English poetry. The E–H dictionary is being used here as a COMPREHENSION DICTIONARY or DECODING DICTIONARY. The other, more active, function is called for when you express yourself, i.e. produce messages in a TL. When, as a Hungarian SL user you open a H–E dictionary, it is being used as a PRODUCTION DICTIONARY or ENCODING DICTIONARY.

Given this, it is a natural expectation that dictionaries differ according to type of user and function. Most, however, ignore this: few E–H dictionaries are specifically for source or target users. Since there are very few English users of both H–E and E–H dictionaries for either function, all of these should target Hungarians in terms of their every feature.

10.5. Monolingual dictionaries

Monolingual dictionaries also target two kinds of user: native speakers on the one hand, and ESL – English as a second language – users on the other. Accordingly, there are native speaker dictionaries and ESL or LEARNER'S DICTIONARIES (these users being most often learners). ESL dictionaries always offer less in terms of quantity but more in terms of sophistication, combining as they do features of language instruction (pronunciation, COLLOCATIONS, style, usage, idioms, and grammar: countability of nouns, transitivity of verbs, comparison, etc.) with those of children's dictionaries (illustration, and a simplified vocabulary). They are for production rather than comprehension. As a consequence of this, many people think that for comprehension native speaker dictionaries, while for production, learner's dictionaries should be used.

Five learner's dictionaries are available at this time.

(i) The oldest, Oxford range, the OALD, or the "Hornby"; (ii) the Longman series, from 1978; and the three relatively new ones: (iii) the Collins–Cobuild dictionaries, (iv) the Cambridge International Dictionary of English (CIDE), and (v) the youngest Macmillan English Dictionary (MED). It is safe to say that any of these is well worth having – especially the CD-ROM versions, which are also more and more affordable – while having two or three of them means possessing fail-proof tools for language learning. This takes us to the different means of access to data.

10.6. Means of access

Access can be understood in physical terms – paper or other media – or arrangement of information in any of those media. Most paper dictionaries are alphabetical, some use a mixture of alphabetical and non-alphabetical methods of presentation (but have an alphabetical Index). The commonest non-alphabetical arrangement is the thematic or SEMANTIC: a THESAURUS is such a dictionary – without definitions. (Roget's Thesaurus, for example, is a household name in most English-speaking homes).

Also in terms of access, "textually based" opposes pictorial: the Oxford Duden, for instance, is a thematically arranged pictorial dictionary without definitions: it exists in both mono- and bilingual versions. Illustrations may be featured in many a general (native speaker and ESL) dictionary.

Computers have created enormous possibilities – both for the lexicographer and the user. These involve both (i) the whole production process: data selection, preparation, CORPUS BUILDING and management, typography, design, production; (ii) and the actual user interface on the CD-ROM or on the Web. While the traditional paper dictionary is now being complemented (and for many, replaced) by the CD-ROM, also more and more people turn to the Web. This is another lexical source that may be put to intelligent use – note the *may*. Here, on-line dictionaries (offering searches while you are connected) must be distinguished from the – usually older *and* less reliable – downloadable ones. Do not use any of them without checking them out first. "Free" has never been a byword for "quality" – except *some* software. Any CD-ROM by a well-known publisher, by contrast, is safe to use and, what is more, has search options and features – examples, at-a-click cross-references, etc. – unimaginable in a paper dictionary.

You should also be aware of the large number of out-of-date – practically worthless – reprinted dictionaries on the market. Given the speed of lexical change on the one hand, and the amount of work needed for any decent dictionary on the other, it is no wonder that they are outdated the day they are published!

10.7. What (not) to look for in a dictionary

Most lay people, who do not appreciate the differences between types of dictionaries, think that they are simply lists of words, so the biggest is best. First, do not judge them by the hype on the cover. (This mainly concerns size). When you buy a dictionary, make sure you have a list of items to be checked. These may be new words you do not want any dictionary to be without, or important words in your special field – hobby, job, whatever – that you can check. But coverage of the lexicon is by no means everything – consider other aspects: friendly layout, speed of lookup, and whether you get the kinds of information you need. Experiment! You cannot have one custom-built, but this way you can hope to have your own "personalised" dictionary that really fits your needs.

10.8. The two dictionaries above can be characterised as follows:

(a) non-encyclopaedic – grownups'– monolingual – desk – ESL (thus mainly production) – general – synchronic (but has etymology, surprisingly) – alphabetical (but has pictures) – CD-ROM dictionary which

- has spoken pronunciations both in British and American English;
- offers a lot of grammatical information: computer mice may be mouses;
- has just two meanings;
- provides (also sentential) examples;
- refers you to *dormouse* (a seemingly similar word that has nothing to do with mice;
- advises that under cat you can find two *mouse* idioms;
- highlights *cursor*, a word that is not in the defining vocabulary

- does not use *rodent* but *animal* as superordinate
- offers a fairly everyday "definition" (with just one potentially unknown word: *cursor*) involving "where mice live"

(b) encyclopaedic – grownups' – monolingual – college – native speaker – general – synchronic (but has etymology, not surprisingly) – alphabetical (has no pictures) – electronic dictionary, which

- has no spoken pronunciations, just transcription, only British English;
- offers some grammatical information;
- has six nominal meanings (!);
- has (three!) verbal meanings in addition to the noun;
- includes slang, and technical/specialised use;
- provides no examples;
- offers no idioms;
- uses *rodent* as a superordinate for *mouse*, and offers a "scientific" – encyclopaedic – definition including Latin terms.

Points to Ponder

1. What kind of a dictionary do Hungarians typically have in mind when they say "dictionary"? And an English-speaking person?
2. Which language do you think has a lot more word forms that happen to be identical, English or Hungarian? (Hint: *bear–bear, still–still; követ–követ, török–török*)
   For *bear*, please consider Item 6 in the Points to ponder section of Chapter 5.
3. Probably – and hopefully, and expectedly by now – you have more than one (English, or perhaps other) dictionary. Try and classify them using the criteria in the chapter.
4. Field work (a): Gather 10 words that you think have recently appeared in the Hungarian lexicon. Provide a "professional" definition for them. Which of them will be around in 5 years, do you think? Those that will deserve an entry in a monolingual "értelmező" dictionary.
5. Field work (b): Gather 10 English words that you think have recently appeared. Find a Hungarian equi¬valent, or translation, for them, and make their dictionary entry.
6. Field work (c): (You should only tackle this after you have covered both this chapter and Ch 5).
   Meanings change fast. One meaning of the English word *semantics* is (almost) pejorative. That mean¬ing is employed, e.g., when someone says, "That's just *semantics*". Find out what this means. Trans¬late this into Hungarian. Is it fair that the word should have this meaning? Why (not)?

   This use of *semantics* is mentioned, e.g. on http://en.wikipedia.org/wiki/ Semantics. Note, however, that (on the Internet a lot more than elsewhere, I'm afraid) all is not gold that glitters.

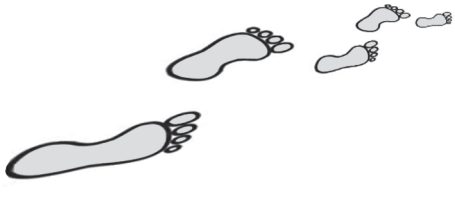
Suggested Reading

Crystal, D. (ed.) (1997): *The Cambridge encyclopedia of language*. 2nd ed. Ch. III-18. CUP
A wonderful volume, offers a wealth of – textual & pictorial – information on all aspects of language & English. Everything you should, and will ever have to, know about language.

Jackson, H. (1988): *Words and their meaning*. Longman
A down-to-earth book specifically devoted to questions of (basically: English) lexicology and lexicography, i.e. meaning in language and dictionaries.

Landau, S. L. (2001): *Dictionaries. The art and craft of lexicography*. 2nd ed. CUP
A book on lexicography (more American than British): on how dictionaries are researched and written, with emphasis on computer technology. Examines and explains all features of dictionaries; illustrations from up-to-date dictionaries.

Jackson, H. (2002): *Lexicography*. Routledge
Reader-friendly and accessible book devoted to (basically: English) lexicography.



Let us return to Isaac and his apple. ***Why does the apple fall down?*** *Translate it into Hungarian:* ***Miért esik az alma lefelé?*** *Which word is which? We can list the English-Hungarian word pairs:*
why *=* ***miért****;* ***apple*** *=* ***alma****;* ***down*** *=* ***lefelé****;* ***fall*** *=* ***?esik****;* ***does*** *=* ***???***
These are called equivalences. But there seem to be a few problems.

Is the verb fall really esik? That is what you would find in most dictionaries: ***fall (v) (fell, fallen): 1. esik.*** *Maybe the equivalent should include* ***esem, esel, esünk, estek*** *and* ***esnek****, all together. Although* ***esik*** *is correct in the sentence above, if you back-translate it, it would be* ***falls****. What is even more difficult to explain is the Hungarian equivalent of* ***does****. There is no single Hungarian word that does the same thing as does does. So what is the Hungarian equivalent? It has none. But both languages ask a question, albeit differently.* ***Why*** *seems to be a simple case, any beginner could translate it. Or could it mean something other than* ***miért****? Why certainly it could!* ***Hát hogyne!*** *So the context is crucial, and explains why* ***fall*** *in the meaning of* ***autumn*** *would be a bad choice.* ***Apple*** *refers to the fruit, but it also appears in the idiomatic cliché* ***the apple of their eye****, which is* ***szemük fénye*** *and not **szemük almája****.*** *In this latter case the unit of translation is not the word.*

Indeed, is equivalence to be sought in words or phrases or clauses, or perhaps entire texts? Are there equivalences at all? Read on to find out more.

Borbála Richter
Kodolányi János University College
Department of English Language and Literature

11. In and Out of Languages

Translation

Have you read a play by Shakespeare? Or any part of the Bible? Or the instructions on how to use a DVD- player? The chances are good that your answer to these questions was "Yes". The next question is whether you read any of these in the original and now I expect the answer to be "No". A question at the core of translation studies is whether the translation is the 'same' work as the original. If it is not, in what ways does it differ? If it is, how is this possible? In fact, how can we provide an answer to this question? These are some of the questions we will look at in this chapter.

The field of translation studies is a new one, although the phenomenon of translation is probably as old as language itself. Translators, thinking about translation in previous cen-

turies before the twentieth century, focussed on many of the issues that still concern scholars today, although the methods used to study translation have developed and been refined. The tension between literal or word-for-word translation and free or meaning-for-meaning translation is fundamental to translating, and thinking about translating. We can trace the principles and justifications that have characterised the debate about translation: from using a word-for-word method in Roman times to make a Latin gloss for a Greek original, through Cicero's arguing for the primacy of keeping the sense when translating speeches, to the translation of the Bible (a task considered special because the Bible was holy and contained the Word of God), up to the present time when knowledge is expanding at an unprecedented rate and the transfer of information is often dependent on translation. Yet it is only relatively recently that the study of translation has moved beyond the subjective and normative evaluation of texts ("What makes a good translation?") to a systematic use of the insights and techniques of linguistics and related disciplines, including *text and discourse analysis*, *sociolinguistics* and *psycholinguistics*.

11.1. Kinds of translation – as observed by a linguist

A good place to begin is with the linguist Roman Jakobson. He starts his discussion of translation with the different linguistic aspects which have to be considered when one wants to translate from one language into another. Jakobson distinguishes between the sign (which can be a word) and what it means, the signified. For example,

Sign:	*The written word 'tree'*
Signifier:	*The letters 't-r-e-e'*
Signified concept:	*The notion 'tree'*

Translation is the replacing of one sign by another. This can happen in three different ways, which Jakobson calls:

1. **Intralingual translation.** This is based on the existence of *synonyms* and happens within a language: what is said in the language can be said in another way with a similar meaning. (For example, paraphrasing, or 'translating' for your grandmother what the bank manager said about the loan repayments).
2. **Interlingual translation.** This is what we usually call translation and the verbal signs of one language are interpreted into the verbal signs of another language. (For example, a guidebook in many languages).
3. **Intersemiotic translation.** In this case interpretation of the signs of a sign system with the signs of another sign system. (For example, traffic lights, where the red means stop).

Although from the above it is clear that all of us spend a lot of time translating in one way or another, translation studies focuses on researching and exploring translation between languages, whether *spoken or written*. (The unrehearsed oral transmitting of a spoken – or a

signed message – from one language to another is referred to as "interpreting", although in everyday use the difference is usually less distinct, at least in English. Interpreting is included in translation studies.)

The term "translation studies" was coined by Holmes in 1972 since, by then, the outlines of a new discipline had become apparent. He identified three main areas of research:

1. theoretical translation studies which seeks to establish general principles, theories and models to explain and predict the processes, functions and products of translation; this uses research from:
2. descriptive translation studies which describes existing translations, their functions in the receiving culture, and the act of translation itself; and
3. applied translation studies which includes translator training, translation aids, translation policy, translation planning and translation criticism.

If you think about the links between these points and the other chapters of this book, you get an idea of the interconnectedness of the various branches of linguistics and applied linguistics. Underlying all the points listed above are some "great questions" of translation, including the relationship between language and reality, and the relationship between the original text and its translation – both globally and in its details.

11.2. Languages and Reality

For our purposes, we will concentrate on the second kind of translation, that is, between languages. The first point to notice is that languages are different but each system is complete and everything can be expressed in every language. Despite this, problems arise because not every word in one language has an equivalent in the other language that means the same. For example:

Hungarian	English	German
fa	*tree*	*Baum*
fa	*wood*	*Holz*
erdő	*wood, forest*	*Wald*

Here we can see that even such a simple group of words is referred to in slightly different ways in the three languages. This does not mean that any of these words is untranslatable, but it does mean that the translator is confronted with a choice, albeit a limited choice. It is this element of choice that makes translation fundamentally a creative activity. In fact, the question of "untranslatability" continues to occupy translators and translation theorists.

Later in this chapter we will come back to this topic. Here it is enough to say that Jakobson considers that there is no untranslatability as such, as any sign can be conveyed in any language. In problematic cases, words can be loaned, words can be created, word components can be translated, or paraphrasing can be used to find a solution.

Another important point made by Jakobson is that languages differ less in what they *may* convey than in what they *must* convey. An example you are all familiar with is the simple problem of gender traditions. In Hungarian we can speak about a person without it becoming clear whether this person is male or female, for example, *Ő is megfürdött mielőtt lefeküdt.* In English we can do this for only a very short time: the moment a pronoun is called for, we have to decide whether to use *he* or *she:*

Before going to bed, ***she****, too, took a bath* or
Before going to bed, ***he****, too, took a bath*

11.2.1. What comes first? Language or reality?

It is an interesting question whether reality is the same for all of us, in other words, whether it is only the linguistic expressions that we use to refer to the different segments of reality that differ. There is a school of thought, often referred to as the Sapir-Whorf Hypothesis after the scholars who wrote about this in the 1950s in America, that language also affects reality in some way. This is a much debated point that has yet to be proved or disproved. In fact, it is wrong to call it a *hypothesis* since it is virtually impossible to test. Nevertheless, it is an intriguing exercise to think about this issue: do you think that language shapes thought or that thought shapes language? As you learn more about linguistics in general and translation in particular, you should remember to revisit this question.

If languages really segment reality differently, then every language community will have its own world view. Some languages will make it easy for their speakers to be very specific about things. Take the Italians who have many, very precise terms for different kinds of pasta. For example, macaroni, spaghetti, orzo, penne. In fact, many languages have simply adopted some of these words. Does this mean that the Italians think differently about pasta? Does it also mean that we cannot find a way to talk about these kinds of pasta in English or Hungarian without 'borrowing' the words themselves? Now that some of us buy penne and use the word, do we have a different world view than before?

11.2.1.1. The process of translation

Various attempts have been made to model what happens during translation. In one view, the process of translation is basically transcoding, with a direct path, one-to-one, from the source language to the target language. If both languages had the same number of elements, similarly distributed and operating in the same way, this model would cover all cases. And machine-translation would work just fine. In another, psycholinguistic, model, the translator reads and

understands the text in one language and forms a universal, non-language specific semantic representation of it, which is then synthesised into the second, language specific, text. A third possibility is that the translator moves from the source language signs back into the real world, the situational reality behind the text. This helps the translator create solutions for translating objects that are typical of a particular culture and for which there are no words in the target language: It is also useful in cases where there is a fixed expression for a certain situation, like *Keep off the grass* for *Fűre lépni tilos!* The next time you translate something, consider your own subjective experience of what you are doing: is any of these models intuitively more accurate for you? Or do you perhaps use all three in different places, for different purposes? Thinking aloud and recording your thoughts are ways researchers have tried to get insight into what happens "inside your head" when you translate.

11.3. Translated text = original text?

If we accept, then, that translation is possible, we come to the question of equivalence. If we understand equivalence to mean having the same meaning and function, then we can see that there is a difference in emphasis. Call to mind all the places in this book where you have already read about meaning. Meaning is not simply contained in individual words, although these are important, but also in how words are related to each other, and how they are used. Translation, to be ADEQUATE, must preserve the meaning of the original even if it cannot always preserve the forms of the original. One way of understanding this is to think of the goal as being thought-for-thought translation. We can think of function as referring to the purpose for which an utterance, whether spoken or written, is made. Commanding, or questioning, or emphasising are all functions that we expect to express through language. Furthermore, the translated text should serve the same function, that is, should play the same role, for its audience as the original did for its audience. (See 5.7.-5.10.)

11.3.1. Equivalence at different levels

Various theorists have discussed equivalence in widely differing ways. Mona Baker, whose book *In Other Words* was written to help translators translate, differentiates between different levels of equivalence. Starting from the bottom, equivalence at word level is often the first element that the translator looks at and struggles with. BILINGUAL DICTIONARIES are often considered to give us equivalents, but they need to be used with caution. (See Ch. 10 on *Lexicography*.) For some terms and between particular pairs of languages there are terms with readily available parallels, like *tree* = *fa* in the example above. Then there are terms that refer to objects that are different culturally but have similar functions, for example, a table to eat off in Japan and in Hungary. Finally, there are the challenging terms which specify cultural specialities. Just think about life in Hungary and you will realise how quickly these problems come to mind: *Túró Rudi* is not a chocolate bar, *disznóvágás* brings to mind a quite different event from a *pig-killing*, and *úttörő* is not quite a *boy scout.*

Grammatical equivalence refers to the fact that grammatical categories and rules may vary across languages. Such differences can lead to changes in the information content of the message. The translator may be forced to put in or take out information depending on the grammatical possibilities offered by the language into which translation is taking place, that is, the TARGET LANGUAGE (TL). If information that is required in the TL is not specified in the SOURCE LANGUAGE (SL), then the translator is faced with a dilemma. A common problem for translators from English into Hungarian is the lack of a grammatical form for expressing respect and/or non-familiarity. The Hungarian translator has to make conscious decisions about the relationships between people. For example, in a love story, at which point do the lovers move from *Ön* or *maga* to *te?* Or if a professor tells a student "Call me William", does this always mean that in what follows *te* (rather than *Ön*) should be used?

Textual equivalence refers to the equivalence between a SL text and a TL text as regards information and cohesion. The focus here is on the ways in which texts achieve coherence and the cohesive devices. Coherence and cohesion can be realised in many ways and the choice of the translator depends on the receptor audience, the purpose of the translation and the text type. (See 10.4.) The languages involved also play an important role: Hungarian, as you may have noticed, codes cohesion very differently from English. For example, *Mari ette meg a kiflit* is different from *Mari a kiflit ette meg,* and they are both opposed to the neutral *Mari megette a kiflit.*

Pragmatic equivalence is the highest level and refers to the ways in which what is not said explicitly is conveyed in the source text and in the target text. This is where cultural differences play a particularly important role. The translator has to try and find a way of recreating what the author intended so that the TL reader understands it. This means working out the implied meanings in the source text and incorporating them in the target text so that they are accessible to the reader.

11.3.2. Approaching equivalence from a different angle

Equivalence is a central category in the theory of translation because it can be approached from so many different angles. You, as a reader, probably assume that equivalence exists. Think back to the question posed at the beginning of this chapter. When you read a translated work, do you think about the fact that it is translated? Do you think about the quality of the translation? How many times have you read a book in both the original and in translation and compared the two? For the reader, equivalence is something that is assumed. Not so for the translator, who has to create the equivalence, nor even for the researcher who researches equivalence. The approach to it can, furthermore, vary greatly. At the one extreme are those who consider that equivalence is a precondition for translation. In other words, equivalence forms part of the definition of translation!

The other approach assumes that equivalence is a question of degree. The aspects in which a text and its translation are equivalent have to be revealed. A classic approach is associated

with Eugene Nida, who distinguished between FORMAL EQUIVALENCE, or formal correspondence, and DYNAMIC EQUIVALENCE. The former focuses on the message itself, on its form and content, while the latter moves the emphasis to equivalent effect. Formal correspondence can distort the grammatical and stylistic patterns of the receptor language and through this also distort the message. Dynamic equivalence aims to elicit the same response from the target audience as the original had on the source audience.

For a text to function in the same way in the target culture, it has to assume the status of an original in the target language, and the reader may not even know that it is a translation. If you read a magazine article about how the human heart works, for example, this is unlikely to be culture specific. Some texts, however, are clearly translations and are recognised as such by the reader. For example, texts like the Gettysburg Address are linked to particular historical events, when a precisely defined audience was being addressed. In this case, the translation is intended to be read as a translation and is not functionally equivalent. (This is not to be confused with *translationese,* which is that strange, unnatural language that comes from source language interference and has aptly been termed a "third language".)

11.4. Theory and practice

Having had a brief look at the very complex issue of equivalence, let us take a look at an example. The example I have chosen is quite banal, since it is not only intricate sentence formulations and complicated ideas that present the translator with challenges. Even in the case of a text taken out of context, like the one below, it is possible to go through some of the various issues that come up in the process of translating.

The first issue is whether we are going to produce a text that reads like an original or one that reads like a translation.

The second issue is whether we can reproduce both the forms (syntax and lexis) and the ideas of the original, or whether we have to give preference to the one or the other.

The third issue is how we can retain the style of the original. This means taking into account, firstly, elements such as the characteristics of the user of the code (i.e. the kind of dialect used which would indicate age, region and social class) and secondly, elements related to the use of the code such as the relationship between the addresser and the ADDRESSEE (how formal, how polite, how personal, how accessible), the mode (spoken or written, how spontaneous, part of a monologue or a dialogue, and how private), and the function of the message. Here we also need to decide whether we can reproduce the same kind of text, and whether the act of communication has taken place in a time close to us (if not, then that will present us with further decisions to take).

The final issue is whether we can translate everything, and whether we will have to add anything.

In this brief chapter, many aspects have had to be left out, but that is not because they are unimportant. In every case, it is relevant to know who the translation is intended for, why it is being done, under what circumstances, and between which languages and cultures.

Imagine a kitchen, where a Hungarian family is eating breakfast and talking before leaving for school or work. The mother says to a child:

Az asztal alá esett a kiflid.

Over to you – translate this into English.

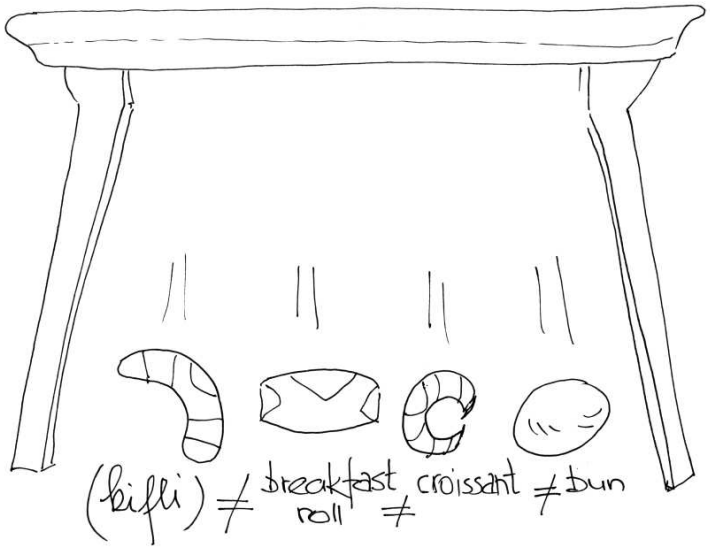
Now let us look at the details together.

Example:	*Az*	*asztal*	*alá*	*esett*	*a*	*kiflid.*
Gloss:	*the*	*table*	*to-below*	*fall-PAST*	*the*	*kifli-your*

1 Some elements are straightforward:
Az/a = *the* (at least in this case)
asztal = *table* (the setting, the participants, the script, and the object *table* are similar in an ordinary Hungarian and an English family)

2. *alá* is not quite as easy to find an equivalent for: *under* translates it, but how equivalent is it? Not really, since *alá* implies a motion in a direction. Does the dictionary help us? The entry for *alá* is *beneath*, *underneath*. These are static as well, telling us about location. It seems we cannot specify the motion in our choice of preposition, this becomes clear in English only together with the verb. The choice between *under* and *underneath* depends on whether we want to give the location as completely and emphatically under. Since this is not indicated, the more general *under* is more natural.
3. Surely the verb *esett* will not present a problem? The verb does not, indeed. However, the tense does. The Hungarian is in the past tense but the simple past tense is only one option in English; a more appropriate solution would be the present perfect. So what is the closest equivalent of *esett*?
4. Many Hungarians eat a *kifli* for breakfast and do not consider this unusual. Who would have thought that such a simple object has no equivalent in English? We could call it a *breakfast roll* or a *bread roll*, but a different object comes to mind for the English-speaker. How about *croissant*, vaguely reflecting the shape? Pity that croissants are made of quite different dough and are associated with France and not Hungary. Also, if you buy a *kroaszan* in Hungary this is clearly not a *kifli*. If enough people in the English-speaking world take to eating "kiflis", the word could come into fashion. An English-speaking family living in Hungary may actually use *kifli*, but we have no information that our example is about such a relatively rare family. Perhaps *a sort of Hungarian breakfast roll* is the closest to an equivalent, but that would certainly make the object seem unusual, which may not be appropriate.

All our choices involve some sort of loss, either of meaning or of style. Our actual decision would probably be made on the basis of what the audience needs to know in the translation context.



5. Next we come to the question of whose *kifli* is under the table. As a general rule, Hungarian uses a suffix to indicate the possessive and English uses a separate function word, so it is not difficult to translate. Still, it does mean that we seemingly omit the second article (*a*) in the English translation, since the English *your* serves to replace both the *a* and the *~d*.
6. What is the relationship between the addresser (the mother) and the ADDRESSEE (the child)? In Hungarian, the use of *kifli**d***, as opposed to *kifli**je***, reflects the real world relationship. In the English language we need not reproduce this, but we also cannot reproduce the implication.
7. Then there is the question of word order. Hungarian allows a lot of choice in word order – although it is probably not fair to call it a free word order language. English allows less, and unusual word order is more marked. In a situation where it is describing what has happened, *Az asztal alá esett a kiflid* is neutral. Another neutral formulation is possible: *A kiflid az asztal alá esett*. No such neutral alternative, telling us where the kifli has fallen, is possible in English. From the English version we could only tell if the emphasis is on telling us **where** or telling us **what** if this were stressed in an oral exchange.
8. Finally, let us give a summarized statement about the issues we outlined before the example. Our translation should read like an original, of a similar, informative, text type; should translate both forms and ideas; retaining the style; and allowing the omission of minor elements in the interests of sounding like an original.

Having read this, take another look at *your* translation, at your target text. Did you take everything into account? Did I perhaps miss something? Is the translation below that tries to specifically resolve all the issues the 'best' one? Which translation is the closest to being an equivalent of the original text?

1. *Your roll has fallen under the table*
2. *It is under the table that your own Hungarian crescent-shaped breakfast bread roll has fallen, my child*
3. *It is your bread roll that has fallen under the table*
4. **The table under fell your kifli*

Now look at the exercises below and re-consider the issues raised above.

Points to Ponder

1. Take a look at *Trilla*, written by Örkény István (in *Egyperces novellák*). The line: *Kicsavarja a papírt az írógépből* is repeated four times before the poem-like story reveals that it is about a woman. Play around with the question of how this short work could be translated into English.
2. Imagine you have to translate an entry in a Hungarian encyclopaedia into English. Look up a few such entries and you will find that they include elements that form chains of reference. Choose an entry about a person in Hungarian history that you admire and translate it, paying particular attention to the ways in which the different people mentioned are referred to.
3. Think about situational expressions. What is the equivalent of the sign which says, *No smoking*, or the cautionary sign outside a house, *Beware of the dog*. If you begin a letter in English with the words *Dear Professor Smith*, would you begin the letter in Hungarian **Drága Kovács Professzor*? Why?
4. Ponder the difference between written and spoken language, and translating written texts and interpreting. Can a clear line be drawn between these two activities? What sort of information about the way in which our minds process language could we get by transcribing simultaneous interpreting?

Suggested Reading

Baker, M. (1992): *In Other Words: a Course-book on Translation.* Routledge: London
A book that helps the reader learn to translate while taking a guided tour of linguistics and considering a wide variety of actual translation examples.

Bell, R.T. (1992): *Translation and Translating: Theory and Practice.* Longman: London
This book presents a model of the process of translation, together with a discussion of many aspects of linguistics that have been integrated into the model. Argues for seeing the clause as the unit of translation.

Klaudy, K. (2003): *Languages in Translation.* Scholastica: Budapest
An excellent introduction to the theoretical aspects of translation studies; it also offers a structured way of looking at actual translational transfer operations. It is a translation from the original Hungarian book.

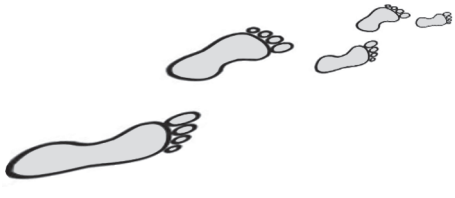
Nida, E.A. – Taber, R. (1969): *The Theory and Practice of Translation.* Brill: Leiden.
An important work that is interesting to read. The focus is on Bible translation.

Venuti, L. (2004): *Translation Studies Reader (2nd Ed).* Routledge: London and New York
An excellent collection of 'classic' writings about translation, with a good introduction to each section of the book. It includes the article by Holmes referred to in this chapter.

Web-sites

http://www.anukriti.net
An interesting site. Check out the "About Translation" section which tells you more about some of the issues, and the researchers, discussed in this chapter, without going into great detail.

http://www.proz.com
This is an Internet forum where translators, interpreters and their clients can exchange views, opinions, and experiences. It provides a way for translators to access the advice of other translators about any more or less specialist question they may have. If you do any actual translating, try it out.



From your own experience of school, would you understand the following situation? It takes place in a primary school in Canada between a student and the teacher of French.

Student: Can I use the stapler, Monsieur?
Teacher: Parle français! [Speak French!]
Student: But it's not nine o'clock yet.

Why is the teacher annoyed? Why does the student think that the time is relevant to the choice of language? Although the teacher and the student know each other and are in the same place at the same time, they quite clearly think the situation is different. All of us use something that resembles a script for our everyday interactions – we know what to expect. The teacher expects the school to be a 'French-speaking zone', while the student expects to speak French only once school has actually started. As we speak, we also express our own identity. In this case the teacher is asserting his authority – which the student subtly rejects. After all, the reply could have been in French! If you start observing what happens in any conversation in your own life, you will find evidence of these kinds of scripts, and you will find examples of seemingly illogical statements that the people speaking understand without effort. You will find out more about how this works in this chapter.

Andrea Kenesei
Pannon University, Veszprém
Institute of English and American Studies

12. Spoken and Written Worlds

Discourse Analysis and Text Linguistics

12.1. Speaking and writing – similar or different?

(1) *Rita [hairdresser, recently admitted to university] (coming back to the desk): D' y' get a lot like me?*
Frank [teacher at the university]: Pardon?
Rita: Do you get a lot of students like me?
Frank: Not exactly, no ...
Rita: I was dead surprised when they took me. I don't suppose they would have done if it'd been a proper university. (from Educating Rita by Willy Russell)

(2) *Educating Rita illustrates how a comedy can also raise serious issues. One of its greatest attractions is its humour, which springs from Rita's and Frank's relationship and the attraction of two very different cultures and backgrounds. (from Introduction to the novel)*

The two examples illustrate two different media of communication – (1) is a dialogue / conversation (in written form here) and (2) is a written text. Are speaking and writing such different ways of communication as they seem at first sight? On the one hand there are dissimilarities; on the other hand there is much that connects them. Both are communicative events. They require the participants – speaker (writer) and hearer (reader) to have an understanding of what is being talked (written) about. This understanding involves the realisation (a) in what situation the participants are, (b) what the relationship between them is, (c) what the circumstances of the communication are, (d) what connects the sentences that are produced, (e) whether the participants have got the same background knowledge of what is being communicated and (f) whether the hearer (reader) interprets the speaker's (writer's) message in the same way as is intended by the latter. Before moving on to discuss the channels of communication, let us see these aspects one by one, through examples (1) and (2).

12.1.1. Situation

(1) Rita and Frank are talking with each other in a room of the university about various things. Rita's question is sort of unexpected – previously they were talking about a picture. As a result, Frank is disturbed by the sudden change in topic and at first he does not understand it. In other words, the CONTEXT (background) of Rita's question is not appropriate for what happens next. This context is built up properly when Rita repeats her question and rewords it. (2) It is easier to understand the second example – the reader knows that it is part of the *Introduction* to the novel. The sentences are well-formed and the information in brackets enables the reader to learn about the novel. The context of the dialogue and the context of the written piece make it possible for the participants and the reader to have a full understanding of the meaning of the spoken utterances and the written sentences. Without context no meaning can be attributed to communicative stretches – words, phrases, sentences or texts.

12.2. Social relationships

(1) The relationship of the participants is important as it determines the language they use; we talk differently to friends, family members and people inferior or superior to us. Three things determine the language that Rita uses: (i) she feels close to Frank, (ii) she's his student and (iii) she comes from the working class. Also, the non-verbal elements – gestures, bodily and facial expressions - all contribute to the context of situation. (2) The relationship between the writer and the reader of a written text is less clear-cut as they are not in immediate contact. This requires the writer to produce a well-formed stretch – grammatically correct sentences which are arranged in a logical sequence. Reader and writer get in touch through the text; therefore it is a must for the writer to provide the necessary information so that the reader can arrive at the right interpretation. The other aspect of the context is the reader's background knowledge such as (i) what makes a novel, (ii) what a university is and (iii) what the roles of teachers and students are.

12.3. Circumstances

(1) Another important factor is what we call SETTING, which involves the place and time of the communicative situation. Rita and Frank are having a conversation at the university which would require a more distanced kind of talk. However, their relationship is closer than the usual teacher-student connection and Rita's social background prevents her from being able to find the desired register. (2) The setting which provides the space and time background of writing and reading is more difficult to describe or define; they can take place irrespective of each other. The temporal setting becomes interesting when we realise that in the case of re-reading a text, the reader can have various interpretations of the text; we can even say that there are as many readings as many times one gets involved in the text. Obviously, the more scientific or referential the text is (like example (2), the fewer possible readings can be imagined, and, the more fictional the written piece is, the more varied the receptions are – depending mostly on the reader's mood. We read a book and it arouses emotions in us. If we reread it again a few years later, the same book can have a different effect.

12.4. Connections

There are certain ties that hold spoken discourses and written texts together. They are grammatical (formal) connections and semantic (meaning) links. In example (1) the grammatical links are as follows: Since Frank does not understand Rita's question, she says it again though with a little bit of difference, in other words, she paraphrases what she has said. This is called REPETITION, which is used to help understanding, or to emphasise the message. The pronoun "they" in Rita's third TURN (her third contribution in the dialogue) refers to the people at the university, who have been mentioned earlier – this is what we call BACKWARD REFERENCE. This is very often used both in conversations as well as in writing as we prefer to avoid saying the very same things several times. The pronoun "it" in the next sentence refers to "university", which is mentioned later, therefore "it" has FORWARD REFERENCE. The grammatical links in (2) cover the following instances: "its", used twice in the second sentence, has backward reference to "Educating Rita", that is, the novel. The word "humour" ends the second sentence and is used again in the third – this is lexical repetition. The first sentence contains "comedy" and "serious issues", which are opposites, almost antonyms. All these grammatical connections are called devices of COHESION.

The semantic connections are more challenging to find and define as they do not appear in the structure of the discourse but on the level of meaning. In example (1) Rita's second turn can be understood only if we know that Frank is a teacher and has got students to teach. For the same reason her first turn is not clear to Frank because it does not relate to the previous turns and the lexical reference "a lot [of students]" is missing. Rita's last turn becomes meaningful only if we know that she has been admitted to university and this fact makes her surprised - she is a hairdresser who wants education. In (2) the third sentence elaborates on, that is, gives more precise instances of "humour", mentioned in the previous sentence. Meaning relationships are collected under the cover term COHERENCE. Coherence also involves

relations which are not found within text and discourse but are said to be outside the text – this is discussed in section 11.5.

12.5. Background knowledge

A spoken dialogue or a written text cannot be understood if the participants or the reader do not possess the necessary knowledge about what is being communicated. This knowledge is not directly expressed in the discourse but implied, that is, indirectly present and referred to. It is vital for the participants to have the same background knowledge otherwise communication is not successful; there is no mutual understanding. In (1) Frank does not understand Rita's question because he at first does not realise what she is talking about; they do not share the necessary common information. In Rita's second turn "students" is understood by both of them because Frank's profession involves dealing with students and Rita has become one. In Rita's third turn "took" is understood as "being accepted" as they both know that she has been admitted to university. It sounds obvious but it is worth mentioning that the concept of "university" is shared by Rita and Frank, however, from two different angles – from the point of view of a teacher and a student. Example (2) is clear for the reader in knowledge of "comedy", "humour" or "culture", just to mention the most important instances. Without possessing the information about these the reader cannot understand the text. This is shared background knowledge between writer and reader.

12.6. Understanding

The communicated message is understood by the receiver (hearer / reader) if (s)he interprets the message in the same way or at least very similarly to what the sender (speaker / writer) intends to say or write. In other words, there is the sender's intention and the receiver's interpretation and communication is successful if the two coincide. The more similar the former are, the more successful communication is. Rita's intention in (1) to get information at first fails because Frank is not able to interpret the question due to a lack of sufficient shared knowledge. It is more challenging to detect how successfully text (2) is interpreted; the reader of the text should be requested to comment on it. If the reader fulfils the criteria of the ideal reader, that is, (s)he possesses the relevant background knowledge discussed in 11.5, then we can expect the right interpretation of the text.

12.6.1. What is communicated and how is it communicated?

In communication we use declarative statements, questions and imperatives (orders) – these are the forms in which information is passed on. However, the same piece of information can appear in different forms; the same content can take the form of a statement or a question or an order. In (1) Rita puts a question, "Do you get a lot of students like me?" We can imagine the same content in the form of a statement, "You get a lot of students like me" or in the

form of an imperative, "Get a lot of students like me!" The content is the same in all the three sentences but the forms and, as a result, the functions of the sentences are different. This content is called PROPOSITIONAL CONTENT.

12.7. Channels of communication – speaking and writing

12.7.1. The power of speech over writing

Example (1) represents spoken communication. The additional information in brackets – coming back to the desk – enables the reader of the play to imagine the setting – the two people are talking in Frank's office and Frank is sitting at a desk. Rita and Frank are talking in person; they can see each other, their bodily movements and facial expressions, they perceive the distance between each other – all these non-linguistic factors help them to understand each other. In spoken interaction the paralinguistic devices (body movements and facial expressions) help the speaker to add very important clues to what is being said and the hearer to arrive at a better interpretation of the message. It does make a big difference what gestures the same propositional content is accompanied by – something that is next to impossible to express in writing. Also, the hearer is able to simultaneously accompany their interpretation while listening to the speaker, thus providing feedback, which may result in the speaker's monitoring and incidental modification of their intention if faced with refusal or disagreement. Therefore, spoken communication has its advantages for both speaker and hearer. The disadvantages are overridden in written texts.

12.7.1.1. Conversation analysis

In example (1) Rita and Frank are talking to each other, that is, they are having a conversation. It takes at least two people to have a conversation, which means that the participants take turns. TURN-TAKING is a complex activity – there are various responses to certain turns. These turns alternate; the speaker's turn is followed by the listener's turn, who, at the same time, becomes the speaker. Turns following each other form pairs. These pairs contain expected or unexpected responses. Rita's first turn is a question, to which she receives an unexpected turn as Frank does not understand her. When Rita repeats her question in her second turn, she receives an answer, which is an expected turn. Frank's first turn reveals that he does not understand her, which is followed by an expected turn – Rita repeats what she has said. Other turns form the following pairs:

	expected turn	unexpected turn
invitation / offer	acceptance	refusal
request	Compliance (action in accordance with a request)	refusal
assessment (evaluation)	agreement	disagreement
blame	denial	admission
question	expected answer	unexpected or no answer

12.7.2. The power of writing over speech

The advantages of written communication are just the reverse of the drawbacks of speech. There is no or only delayed feedback from the reader. However, this is what makes writing more collected. The time allotted for the writer enables him/her to produce a well-formed composition. Of course, the obligation to form a grammatically and structurally adequate stretch of language is a challenge; to maintain coherence is not an easy task. Besides, non-verbal communication is not at hand for the writer. There are other requirements that the writer must bear in mind – punctuation conventions (comma, colon, semi-colon, dash, hyphen), logical organisation of thought, linearisation (proceeding from the beginning to the end) and paragraph development. To comply with the rules of writing creates a challenge for the writer. It does, however, enable the reader to arrive at an interpretation more easily.

12.7.3. Topic boundary in speech and writing

Both oral discourse and written text are about a subject matter, that is, a topic. This topic, however, is divided into further sub-topics, which must be separated from one another. In other words, there is a topic shift between the homogenous parts of communication. This topic shift is indicated by paratones in speech and paragraphs in writing. The beginning of a paratone is an expression that indicates what the speaker intends to talk about; this expression is uttered with a raised pitch. The end of the paratone is indicated with a phrase, which is followed by a pause. The written counterpart of paratone is the paragraph, which must contain a thesis sentence (the main point) and the rest includes sentences supporting the thesis sentence. The beginning of a paragraph is genre-specific, i.e., it depends on the form of the written piece – a guidebook, a poem, a letter, an article, an advertisement or a novel.

12.7.4. Utterances and sentences

Communication takes place in stretches of language, which are differentiated in size (from one word to long texts), structure (syntactically complete or elliptical) and channel (spoken or written). One word is regarded as text in a particular context. For example,

(3) *Danger!*

becomes a written text only if it is displayed as a warning sign by the road or elsewhere. The context that provides meaning for the word and thus creates a text from it is our background knowledge of the particular place where it appears. Elliptical utterances like

(4) *"To the shop."*

gain meaning in the context of a conversation; there is an understanding that it replies to the

question, "Where are you going?" If someone treads on your foot and you shout out

(5) *Ouch!*

it is again the situational context that makes the interjection comprehensible. Can we regard (5) a sentence? Conventionally, we call this type of communicative pieces UTTERANCES. Therefore, every sentence is an utterance but not the other way round – not every utterance is a sentence: utterance is a cover term for each type of communicative stretch.

12.8. Summary

We attempt to describe the most important issues of discourse and text analysis by pinpointing the common features as well as the differences of spoken and written communication. It is revealed that the similarities exceed the dissimilarities as communication is basically the conveying of messages, that is, the passing on information, regardless of its channel. Two examples – a short dialogue from a play and an extract from a reference to the same play – illustrate the structures of communicative stretches, the meanings of messages, the ways with which we reveal these meanings, the communicative situations and circumstances, the grammatical and knowledge-related relations between the parts of conversations and sentences of texts, the necessary background knowledge of the participants, the interpretation of the conveyed messages and the differences between speech and writing.

Points to Ponder

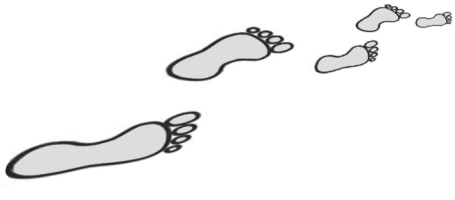
1. What is the function of the following instances of communication?
   a. Oh, no!; Ugh!; Fantastic!
   b. Please help me!
   c. Are you still there?; Can you hear me?
   d. O Rose, thou art sick. / The invisible worm / That flies in the night / In the howling storm
   e. Readings 6, 7, 8 can be downloaded from that site.
   f. What does this mean?
2. How do we know that the second turn in the following conversation is a refusal of the invitation?
   "Let's go to the movies tonight."
   "I have to study for an exam."
3. Do texts represent unilateral (involving one participant) or bilateral (involving two participants) communication?
4. What information does body language convey?
5. Which type of discourse abounds in verbs and which in nouns?
6. Which type uses more repetitions, pronouns, active or passive constructions and conjunctions?

Suggested Reading

Beaugrande, R. de – Dressler, W. (1994): *Introduction to Text Linguistics.* Longman: London
The book provides the first comprehensive survey of a rapidly expanding area of linguistic research. Instead of being limited to the classification of minimal units or to syntactic rules for single sentences, text linguistics is the study of the properties of whole texts and their uses in communicative interaction. The central notion of the book is textuality: what makes the text a unified meaningful whole rather than a string of unrelated words and sentences? The answer lies not in the text as an independent object, but rather in the human activities we perform with it.

Brown, G.–Yule, G. (1989): *Discourse Analysis.* Cambridge University Press: Cambridge
In this textbook, the authors provide an extensive overview of the many and diverse approaches to the study of discourse, but base their own approach on the discipline which is common to them all – linguistics. The main concern is to examine how any language produced by man, whether spoken or written, is used to communicate for a purpose in a context. The discussion is carefully illustrated throughout by a wide variety of discourse types (conversations recorded in different social situations, extracts from newspapers, notices, fiction, graffiti, etc.). The techniques of analysis are described and exemplified in detail for the students to be able to apply them to any language in context that they encounter.

Stubbs, M. (1995): *Discourse Analysis. The Sociolinguistic Analysis of Natural Language.* Blackwell: Oxford
The author believes that sociolinguistics should be about language as it is used in everyday life, in conversational situations, by real people. Much of the data are obtained in genuine, real-life situations and the theoretical issues are accompanied by discussion of their practical implications. The book sheds light on social process and, through its informed interactional perspective, also advances our understanding of the use of structure of language.



Let's take a walk down memory lane… Do you remember learning to ride a bike? Learning to tie a pretty bow – or your shoelaces? Your first picture book? Learning to read? Learning to speak?

You may have answered yes to the first few questions, but it would be most surprising if you said you can remember learning to speak! Of course, your parents may have told you stories of cute things you said, but that is not quite the same thing. Learning all the complexities of language is something all children in normal circumstance learn in just a few years. Only years later, when you are struggling to learn a foreign language, do you realise what a marvellous achievement acquiring the first language was. Studying how children acquire language is one of the things psycholinguists do. They do it without that essential of laboratory experiments, a control group. It would be most unethical to deprive a child of the conditions needed to it to learn to speak just so that we understand the process better. Sadly, it does happen that some children grow up "without language". In 1970, a girl of about 13 was found in California. She had been kept alone in a small room and not spoken to by her parents since she was a baby. She was called Genie, to protect her, and was taken into the care of specialists. Although they tried to teach her to speak, Genie could only learn to produce telegraphic speech, that is, a string of words with only an elementary syntax. She never managed to get control of features of language like inflection and function words. Of course, it would be dangerous to generalise from such a sad case. Most instances of 'wild children' are sad and the children have been traumatised and badly treated. This means that it is hard to know if their development could have been normal if their lives had been normal. Still, there does seem to be a link between age and learning to speak. If you don't learn to speak as a child, you may never be able to do so. Remember Genie's story as you read the following chapter.

Judit Szitó
Kodolányi János University College
Department of English Language and Literature

13. Speak Your Mind

Psycholinguistics

13.1. Introduction: What is psycholinguistics?

We see and hear people talking every day and everywhere: at home, at work, at school, in the street, on the bus, even in their sleep; and we take this for granted. In fact we treat it as unusual if a person is quiet for a long time or has no language. We do not usually stop to think why people talk or what makes them able to have language because we take it as natural. We let psycholinguistics do this for us.

Psycholinguistics was born in 1953 (out of a seminar at Indiana University) to study the mental processes during language use. You may be interested in these questions, too, especially if

your profession will (or does) involve psychology, languages, or meeting adults or children. Perhaps you want to understand or help someone with speaking difficulties or other language-related difficulties (such as dyslexia). Generative linguist Noam Chomsky, who has deeply influenced psycholinguistics, proposed that a central issue of linguistics should be to find out what exactly it is that humans, and only humans, know when they talk; how they acquire that knowledge; and how that knowledge is put to use. Psycholinguistics basically leaves it to linguists to find out what the knowledge of language consists of, and concentrates on the (reformulated versions of the) other two issues. Shortage of space does not permit us to cover all areas of psycholinguistics: we shall try to give very basic answers to the following questions:

1. Do people *learn* their knowledge of language or is it *innate* (=born with them)?
2. *How* do people *acquire language,* especially their first language?
3. What happens when people *produce talk*, and what happens when they try to *comprehend* it?

Let us imagine that two people are holding a conversation (in their mother tongue) and a dog is listening to them. Both people, but not the dog, have the knowledge of language about which we want to know more (1st question); both of them acquired this knowledge at a very young age, and we want to know more about it (2nd question); and, finally, we want to know what is happening when one is using this knowledge to talk, while the other is listening and comprehending (3rd question).

Although these questions look simple, they are not easy to answer. The nature of knowledge is abstract and complex, and so is its acquisition or the ways it is put to use. We cannot open a person's mind to find the knowledge of language neatly packed in a box, neither can we find there a language-producing machine and a language-comprehension device and see how they work; we need to create hypotheses and models. We cannot ask children how they acquire language since they will not be able to answer, neither can we experiment with depriving them of the opportunity to learn their mother tongue, for that would be unethical. Psycholinguists suggest treating the mind as a black box and language production as a means to look into the black box. It is possible to observe language production, measure some of its aspects, and connect the findings to language knowledge in the mind. Although language production or language use (see Ch. 8 *Sociolinguistics*) gives us only indirect information about knowledge of language, it is also the only window offering us a sight.

13.2. Some psycholinguistic aspects of knowing a language

13.2.1. Talking animals?

"Please stop barking, I want to sleep", you tell the neighbour's dog and it says, "All right, I am sorry" and becomes silent. Wouldn't this be nice? Yes, but alas, it will not happen since we know that dogs cannot talk. But let us stop for a second: don't we all know an intelligent

dog somewhere that "nearly speaks"? A dog which perfectly understands what we say and stops barking? And then, is it actually true that only humans have language?

If you think about it, the achievement of some animals is truly astonishing: there are counting horses and parrots, talking apes, or dolphins communicating through a wall. However, if you examine how non-human species communicate naturally, you will find that their communication is limited both in form and content. Bees can perform a dance which signals the direction and distance of the food source, but they cannot 'dance out' if the direction of the food is "up" from where they are, let alone say things such as, "What a beautiful Sunday we had yesterday!", "Let's go on strike!" or, "Do you like our new hive?". Vervet monkeys have over thirty vocalizations and, for example, different call sets for "snake" and other types of danger, but that is all. But what if the problem is just a missing language? Perhaps we should teach the most intelligent of them? So apes, which share over 80% of their genes with humans, have also been taught human languages. When scientists realized that apes could not talk only because apes were not 'built for speech', they started to use sign language with them. That turned out to be a success. But this still leaves us with the question: what makes sign language (or Latin or Swahili or anything that counts as a language) a language? What is it that differentiates a language from gestures, mimics, vocalizations and dances? A few examples will show it to us.

In the 1960s Washoe, a female chimpanzee, learned American Sign Language (ASL) fairly naturally. She grew up in the Gardner family's home. Just like humans, she accepted to use arbitrary (see 1.1.) signs to replace real-world things. Like people, she realized the signs were 'names' for things (SEMANTICITY), even for those that were not present (DISPLACEMENT). So Washoe could talk about things not present, something that animals usually cannot do. She was also able to make up new combinations of signs such as "Roger Washoe Tickle", so she used language CREATIVEly – just like people. What she did not produce was word order, something that human babies seem to know without any teaching. This is called STRUCTURE DEPENDENCE, or sensitivity for structure. This was a problem, too, for another chimp called Nim Chimpsky (named after Noam Chomsky). Nim produced sentences such as "eat drink eat drink", "grape eat Nim eat", "eat Nim eat Nim" and "Nim eat Nim eat", where words are basically thrown together randomly. An English-speaking child, on the contrary, says structured sentences as early as in its two-word stage. In utterances such as "see ball", "get ball" or "want baby" (Scovel, 1998:15-16), the human child spontaneously creates two categories of words and places the two categories in a certain order (*see*, *get* or *want* first and *ball* or *baby* second). Likewise, in "Mummy come" or "Adam put" (Aitchison, 1976:46), two categories are created and ordered (Mummy or Adam first, come or put second). True: an ape is able to learn word order (at least, one chimp named Sarah was), but only when it is trained rigorously (that is, constantly rewarded for good forms). Some scientists say there is continuity between the ways animals and humans communicate, but others think that is true only to yells, groans, cries and other non-linguistic vocalizations that we kept from our past in *addition to* language. It seems that there is discontinuity: that animal communication is different in many respects from human language. All in all, it seems that no animal can be led to use language as humans do, especially if we consider structure essential to language.

Linguists and psycholinguists do consider structure essential to language, and study child language development with regard to it. 'Structure' often means grammaticality of sentences, that is, that the syntax is 'right'. This syntax-centred approach is characteristic of the 1970s and 80s and comes from the syntactic focus in generative grammar. In the child language example above, the first set ("see ball") illustrates verb-object ordering, and the second set ("Mummy come") subject-verb ordering. Structure is also addressed in semantic models. These operate with concepts such as agent (the participant which carries out the action) or action word, instead of subject or verb. According to a semantic model, the child language examples show an action word-affected thing set and an agent-action word set. It is not decided if knowing and acquiring a language is semantics-based or syntax-based, but it is obviously structured.

13.2.2. Innate or learned?

The question, then, is: what makes humans different from animals in their sensitivity to structure? Is it a gift of nature, is it born with them, in other words, is it INNATE (see 4.2.) – or perhaps is it learned? Again, scientists do not agree on the answer. There was a time when people thought babies were born with a "clean slate" of mind (*tabula rasa*) and everything had to be learned through imitation, including language. This is clearly not true: for example, children cannot learn *goed* from adults, yet they often say this for the past tense of *go*. We saw that animals will not talk in the same way as people do, even when they are trained, whereas all normal children grow up talking without special attention paid to it (even in cultures which discourage child – adult communication). But does this prove that we have a 'language gene'? It does not. We all walk and yet we do not want proof for a 'walking gene'. So we need to look for other arguments. Those who are against the innateness of language think that language develops in parallel with other cognitive abilities and is not separate from them. This approach cannot really explain how mentally retarded children can have perfectly good language. Do you think, for example, that it is easier to learn a language than to open a door with a key, or to find the matching shoe for your right foot? For Rebecca, a woman with an IQ index of 60, it was. Yet Rebecca was a complete person when she listened to stories, and was so fluent and composed she would not be spotted as mentally defective when she acted in a theatre group (Sacks 1970/ 1998). Rebecca's linguistic abilities were obviously much better than her other cognitive abilities, which also means her language abilities must be independent from her cognitive abilities. From the many opposing hypotheses, we shall now see a few arguments for the innateness of language capacities.

Nativists think that innate capacities are present right at birth. To demonstrate this, they gave several-hour-old babies a dummy connected to a computer. In this kind of research the rate at which infants suck at the dummy generally shows how interested they are in what they hear or see. The babies' sucking has told us interesting things. For example, that they can recognize human voice, their parents' voices, their mother tongue, and can even tell apart a list of content words (such as *dog*, *walk*, *nice*) from a list of function words (such as *to*, *the*, *him*)! Adjusting the rate of sucking, babies will select from three tape-recorded tales the one

they heard every day for weeks before they were born. Infants have a natural orientation toward language. Within two or three years after being born, babies produce structured utterances, and all go through more or less the same stages of development with respect to the structures they use. For example, children learning English start forming *why*-questions without auxiliaries (*Where Daddy go?*), and then they add the auxiliary to the main verb (*Where you will go?*), and finally they change the order of subject and auxiliary (*Where will you go?*). Another proof is that children ignore structures before they are 'ready' for them. No matter how many times an adult repeats the correct past tense form of *hold* for the child (*Did you say the teacher held the baby rabbits?* or *Did you say she held them tightly?*), she keeps using *holded* in her answers (*She holded the baby rabbits and we patted them.* or *No, she holded them loosely.*) (Cazden in Aitchison, 1976:75). Moreover, the structures children use get more and more complicated although children usually receive feedback only on the content and not on the form of what they say. This means adults tend to correct young children only if the thing they say is untrue, but not when it is ungrammatical. So, although language itself is clearly learned, and not exclusively inherited, nativists say that its foundations are innate.

An important argument for the innateness of language seems to be the fact that language must be learned early in life; that is, there are biological constraints on language acquisition. This is because language is nested in the brain, and as the brain matures it seems to lose its flexibility for language. The hypothesis will be introduced in the next section; now we shall look at where language is located.

13.2.3. The nest in the brain

Perhaps you think it is funny that language 'sits' in an area of the brain? Because, for the most important parts, it does. You have a very good chance that it is around your left ear.

The brain looks like half a walnut: it has two parts. The right and left parts are called hemispheres, and one of them, usually the left, is dominant. This functional separation is called LATERALIZATION. Left cerebral dominance means the left hemisphere takes the lead in controlling brain processes. About 80-95% of people have a left cerebral dominance, which means that language is perceived and produced in the left hemisphere. People with left cerebral dominance usually (though not always) prefer using their right hand, foot and eye.

For most of us, language is located in the left hemisphere around the left ear. Two areas related to speech in the left hemisphere are named after the neurologists who studied them: one is called BROCA'S AREA (above and in front of the left ear) and the other is called WERNICKE'S AREA (around and under the left ear). We learned these things from studying APHASIA, that is, speech disturbance due to injury or other kind of damage in the brain. Broca's aphasiacs basically have fluency problems because Broca's area is involved in organizing the movements of the vocal organs during speech. The other, Wernicke's area, in involved in speech perception; aphasiacs of this kind talk a lot but have comprehension problems and their speech does not make much sense.

Biology also plays a major role in language acquisition, which is the topic of the next section. Babies' brains are extremely flexible and can recover at an amazing speed in the case of damage. As humans grow up, their brains mature and slowly lose this flexibility.

In this section, we have looked at language as an exclusively human capability. First we compared animal communication to human language and identified several features which separated the two. We found that people have a unique sensitivity to structure, which is born with them (innate). Young children's early speech development also shows uniform features. Finally, we looked at where language is located in the brain (lateralization, Broca's and Wernicke's aphasias) and learned that a young baby's brain is flexible enough to take over functions from damaged areas.

13.3. Language acquisition

In this section, we are looking into the second question: How do people acquire language, especially their mother tongue?

13.3.1. Children without a language

In *The Jungle Book*, when grown-up Mowgli leaves the wolf pack and joins people, he learns a human language in a few weeks. He asks a village elder about a tiger skin, "So thou wilt take the hide to Khanhiwara...?", using literary forms that now a poet would not (*thou for you* and *wilt* for *will*). In literature, everything is possible. But what would happen to Mowgli in real life? How long would it take for him to learn language? And would he learn it at all? A 'wild child' learning or not learning language late in life can show something about the biological foundations of language.

Fortunately for them, there are not many Mowglis in real life. A few recent cases received scientific attention. We shall mention three. Isabelle, the daughter of a deaf mute mother, was found in the 1930s in a dark loft around the age of six and a half years. When she started to learn language, she quickly caught up and in two years was not easily distinguishable from other children of her age. Another girl, Genie, was not found until a much later age, which was estimated to be between 11 and 14 in 1970. Up to that time, she had spent her life either harnessed to an infant's potty-chair or in a sleeping bag in a dark room. Her mother was not allowed to talk to her, and her father punished her if she made any noises. When she finally learned language, Genie acquired a fairly large vocabulary but she had problems with grammar. Also, the stages of acquisition proved to be longer for her than for young children. We shall also mention an adult, Chelsea, who was not diagnosed as deaf and, consequently, had no language until the age of 31. This case was not very well studied or publicized and we only know that, when the study finished, she had a good vocabulary but still no grammar (e.g. *Breakfast eat me*).

13.3.2. An explanation

If someone asked *you* to explain why people cannot completely learn any language after a certain age, what would you say (after "I don't know)? You would surely suspect that there is something special about learning and being young. Perhaps the young brain?

To explain child – adult differences in language acquisition, a hypothesis was proposed by Lenneberg in 1967. It was called the CRITICAL PERIOD HYPOTHESIS (see 14.4.3.). This hypothesis suggests that there is a biologically determined period of time within which successful language acquisition can take place. After the critical period closes down, the acquisition of language will be problematic. This concept is similar to that of *imprinting* among animals, for instance among geese, which will consider their mother anything they see in the first twenty minutes of their life and will follow it (Lorenz, 1970). According to the Critical Period Hypothesis, Isabelle was able to acquire language because she was young enough, but Genie and Chelsea were too old for it. The suggested period is related to lateralization of the brain (discussed in subsection 2.3) and closes around the age of ten. Other scientists doubt if these cases offer enough evidence, arguing that Genie's childhood was so horrible that it led to mental retardation and Chelsea's language acquisition was simply not studied long enough. We have much more evidence from second language acquisition (see Ch. 14 *Second Language Acquisition*) research related to age. Later versions of this hypothesis belong to this area of study and suggest *sensitive periods* for various levels of language (for instance, a period up to ages 6-12 for a native-level pronunciation).

If he is like language-deprived Genie or Chelsea, grown-up Mowgli would have to spend years acquiring his first human language and would probably never be able to learn it completely. The reason? One alternative is that his brain would have lost its plasticity for language by his mid-teens, according to the Critical Period (or Sensitive Periods) Hypothesis. Another explanation is that language would be damaged by that time as part of his other damaged cognitive capacities.

13.3.3. Acquiring language

Another, much vaster and much more pleasurable, source of data is studying how normal babies learn their mother tongue. (A corpus called CHILDES contains child language texts from various languages.) Whether or not you have any experience being with pre-school-age children, you can probably name some general stages of their language development. For instance, you know that all babies can cry after birth, at some later stage they babble, and that, when they start talking, they do not immediately start producing whole grammatical sentences but rather words.

Human babies go through similar stages regardless of the language they are learning. *Crying* and *cooing* are part of babies' universal pre-verbal development, and can be regarded as vocal exercises for speech. *Babbling* (from about 6 months on) is seemingly universal, but shows

a tendency toward the mother tongue's phonetic buildup and intonation (in an experiment some adults were able to pick out the babies babbling in their mother tongue). A large part of language development is language-specific, but there are *one-word* and *two-word* stages in most languages. The one-word stage usually starts anywhere between the baby's first and second birthday, although for some boys it may start even later. The brief one-word stage is quickly followed by the equally short two-word stage. In about six months at the most, English-speaking children arrive at their *three-word stage*, and so on. The uniformity of the developmental stages across languages is due to the fact that human biological maturation is independent from specific languages.

For anyone interested in the innateness of language it is more exciting that there is an order in which structures are acquired. There are examples in 13.2 for stages of question formation in child language acquisition. Hungarian child language studies show that accusative *–t* (*babá**t***) and possessive *–é* (*maci**é***) are among the first acquired morphemes (this is because Hungarian is more complex morphologically than English). Again, there are many explanations for the acquisition of this order. For example, the order in which certain functional categories appear may have to do with cognitive maturation, the child's natural ego-centrism, or the cognitive complexity, the simplicity and the frequency of the morpheme or grammatical form. These are non-nativist views. Others think that there is a *Universal Grammar* underlying all languages and the child 'resets the parameters' of this as she learns the language. According to these nativist views, there are complex things that simply cannot be learned from the talk a child hears. This is called the POVERTY OF STIMULUS argument.

This chapter has shown that there are competing theories behind all the psycholinguistic research in first language acquisition. Evidence from late acquisition (e.g. Genie) or normal language development can be explained in different ways, according to the different theories of language. It is clear there is something innate, a 'readiness for language' in human children, and even after the first 6-10 years, a critical or sensitive period for language learning, humans can learn languages to some extent. Explanations differ with regard to how much of language is innate and learned. Nativists think there is a Universal Grammar in the human mind, a grammar so abstract and complex that it cannot be learned from natural language input (poverty of stimulus). Language learning for them means resetting the parameters. Non-nativists offer other reasons, and believe in the role of learning.

13.4. Language production and language comprehension

Suppose at lunch your friend asks what you did at the weekend and you tell him about it. A common situation – and the process you and your friend carry out in your minds is one of the most important topics in psycholinguistics. What happens in people's minds when they spontaneously talk and understand talk? The answer is given in production and comprehension models. Since we are not aware of the processes that take place in the mind, the models are created on the basis of results from studying and measuring normal speech (especially hesitations and slips of the tongue) and abnormal (aphasic) speech. Since there is no way for us

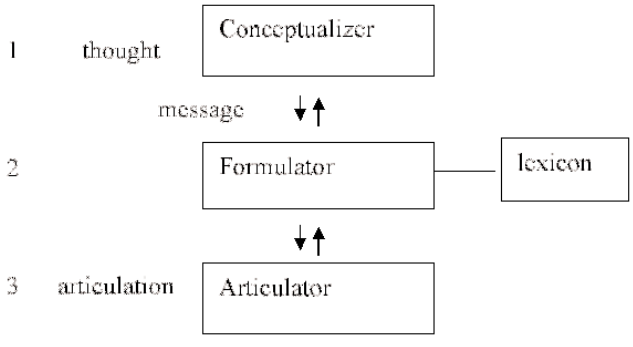
to have direct proof for the reality of the models, they are considered good if they are successful in predicting speech phenomena.

13.4.1. Speech production

If you record speech and then write it down (i.e. transcribe it), you will be surprised at how difficult it is to follow the transcription (i.e. the written version): changes of topic, changes of subject occur vary rapidly, some references are unclear and there is no intonation to help. You will be even more surprised when you randomly look at single sentences and see how 'ungrammatical' and 'unfinished' some of them are. This is basically because spontaneous speech is created 'on the spot' right there and then, and the speaker must plan what to say next while she is speaking.

Speech production models try to capture the process whereby intention turns into speech. The most accepted model of language production is that of Dutch psycholinguist Willem Levelt.

Speech goes through three levels in this model. For the first level, thoughts are formulated in a unit called Conceptualizer (1). This means that the message is generated here but has no linguistic form yet – it may appear in pictures. When it is ready, the message is sent to the next unit, the Formulator (2), where it will be grammatically and phonetically shaped. The Formulator is connected to the Lexicon, from where it takes the words and other units of expression. In the third step, the message goes to the Articulator for articulation (3), that is, for speech. This whole process takes place very fast.



A simplified model of speech production, based on Levelt (1989)

At each level it is possible for the speaker to revise the original message or the form of the message. In the figure above, this is indicated by the arrows pointing upward, meaning that the 'thing' is sent back to a previous level. Revising the form of a message, rethinking a word

choice or deciding otherwise when already starting articulating: all are important for us because it is these revisions that somehow interrupt the flow of speech and can be measured or studied (for instance as hesitations or false starts).

13.4.2. Speech comprehension

Let us get back to your friend listening to you while you are telling her about your weekend and see what she does to understand you. Speech comprehension models differ about allowing or disallowing interaction between levels of perception and comprehension. For instance, if a listener is unsure whether she heard [*f*] or [*h*] in ?*old*, she can use sentence meaning in an interactive model, but not in a serial model, to identify the phoneme. This is because serial models usually start from the lowest levels (such as sounds, like [*f*] or [*h*]) and go only upward, whereas interactive models allow higher-level processing (semantic or syntactic) to affect the processing of lower-level components, like in our example.

Discourse level: To comprehend your weekend story, the listener (your friend) uses her knowledge of the world to understand things you left out. Let us assume you talk about having to stir some sauce over the heat, and for dramatic effect, you skip the sauce burning and get to its smell right away. In this case, the listener must supply the missing information, and she probably will, since people know that a burning smell in the kitchen means something has been burnt! Also, the listener evokes typical mental patterns or scripts that make it easier for her to process. For example, having guests over for lunch or going to a restaurant for lunch involve different scripts: there is a waiter in a restaurant, but not at home, you order food at a restaurant, but not at home, and you pay in a restaurant, but not at home. All in all, the listener must share some background knowledge with the speaker to understand him.

Syntax level: Sentences are processed on the level of syntax. It seems the unit at this level is the clause (see Ch. 4 *Syntax*). The comprehension of a sentence involves breaking it up into constituents and assigning them a semantic interpretation. Therefore syntactic processing is language-specific. In the comprehension of Hungarian speech, for instance, the listener can rely on nominal endings (such as accusative *–t*).

Lexical level: Scientists assume that there is a lexicon that is responsible for storing and recognizing lexical items. It is very different from the alphabetically organized lexicons you know from libraries. In the mental lexicon, the information is grouped according to phonological, syntactic and semantic information. This means the listener can associate a word with others on the basis of its phonological buildup (e.g. similar stress or syllable structure, as in *better* and *butter*), morphology, syntactic position and meaning. The exact way the items in the lexicon are organized is unclear.

This section has been about the processes of speech production and speech comprehension. In both cases three main levels of processing have been named and the processes briefly described.

13.5. Summary

This chapter started with three questions: the first concerned the innate or learned quality of the knowledge of language, the second the acquisition of this knowledge, and the third how the knowledge is put to use. Each question was briefly addressed in one section. According to the first, we can be sure that at least some of the knowledge of language is innate, but it is not clear precisely how much. We came to this conclusion through examining animals' ability for language and children's self-directed language development. Then the biological foundations of language were described. Nativist and non-nativist approaches were used to explain several features of normal language acquisition in children and of late acquisition in two children and an adult. In the last section, we presented a simplified model for speech production and levels of processing for speech comprehension.

Points to Ponder

1. When a chimpanzee learns a human language, it learns a few features together with it. These are mentioned in this chapter (in 13.2.1.) and in the first chapter (in 1.1.). What are they? Make a list and prepare to explain what the features mean.
2. Out of the features that characterize human language (see 13.2.1 and 1.1.), which is the one that seems to be innate in human babies AND the most difficult for an animal to learn?
3. Why do many scientists believe that language is innate in humans? Make a list of reasons or proofs.
4. In the mid-1990s a three-year-old boy was found in a Hungarian village. The boy was raised with dogs, and consequently, could only bark and walk on his hands and knees. What were the chances for him to fully acquire Hungarian?
5. Someone says, "This picture... um on this picture we can see a woman".

a. Something shows that he revised what he had wanted to say. What is it?
b. At what level of speech production did he revise what he had wanted to say?

Suggested Reading

Aitchison, J. (1988[4]): *The Articulate Mammal: An Introduction to Psycholinguistics.* Routledge: London

One of the most popular introductory textbooks in early psycholinguistics. Focuses on syntax and Chomskyan ideas.

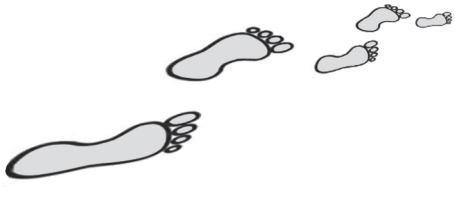
Curtiss, S. (1977): *Genie: a Psycholinguistic Study of a Modern-day "Wild Child".* Academic Press: London / New York

All the interesting details about Genie and her belated language development. Very readable.

Garman, M. (1990): *Psycholinguistics: Central Topics.* Methuen: London

A good overview in English for those who are really interested in psycholinguistics.

Levelt, W.J.M. (1989): *Speaking: From Intention to Articulation.*
MIT Press: Cambridge, Mass.
Recommended for those who want to know more about speech production.
Pinker, S. (1994): *The language instinct.* Harper Collins: New York
(A nyelvi ösztön. Typotex.)
Untraditional (non-textbook) format. Highly recommended for anyone. Witty and easy to read. Defends nativism.



As an experienced language learner, I am sure you have thought about why it is that some people learn a language more successfully than others. Not only that, but that everyone has an individual learning profile – some students love communication games and learn through wanting to take part, others hate role-plays but actually seem to like grammar. Perhaps it is stating the obvious, but motivated learners do better. However, what motivates one learner, like getting good grades, may not move another – who might start making real progress after helping a visitor to Hungary buy a train ticket. You may be one of those learners who have developed useful strategies for learning and for communicating. Or you may just be one of those lucky ones who seem to have been born with a gift for languages. Research into all of these areas improves the way in which learners learn and teachers teach. This chapter takes a look at what is happening in this respect.

Zoltán Kiszely
Kodolányi János University College
Department of English Language and Literature

14. How Do We Learn Languages?

Second Language Acquisition

14.1. Introduction

Imagine a language school which starts a new English course for people who already have some basic knowledge in English. Let us suppose that these students know the English language more or less on the same level, but they have very different backgrounds. Let us take a look at three of them. First, there is a 15-year-old boy who has decided to study English because he wants to understand the lyrics of his favourite music groups. Second, there is a 24-year-old businessman who needs the English language because he wants a better position in his company. Third, there is a 16-year-old girl who speaks German very well, but has decided to study English too, because she realized that speaking only one foreign language will not be enough for her in the future. Having attended some lessons these three students realize that one of them is much more willing to speak than the others, another has more talent in understanding the grammar rules, and the third one understands the reading tasks the best. What can be the reason for these differences? Is it their age? Perhaps their motivation? Or is it something else?

The branch of APPLIED LINGUISTICS that tries to answer questions like these is called second language acquisition (SLA) research. In this chapter, first, you will read about what most experts mean by the term "second language acquisition." Second, you will have the chance to learn about two main issues: (1) the characteristic features of the TARGET LANGUAGE people acquire (learner language); and (2) the explanations of how people acquire a second language.

14.2. Some basic concepts

If you want to learn about second language acquisition research, you have to be aware of the meaning of some terms. Let us see the first one: "SECOND LANGUAGE" is often meant as a language that is learnt in an environment where this language is the main means of communication for the inhabitants of the country. People learn a "FOREIGN LANGUAGE" in an environment where this language is not the main means of communication for the people living there. For example, if Hungarian students attend an English school in England, they learn it as a second language. If they learn English in Hungary, it is a foreign language. In this chapter "second language" and "foreign language" are used interchangeably, or they will be referred to as L2.

The words "ACQUISITION" and "LEARNING" also require some explanation. Acquisition is a subconscious process of picking up language through exposure, like children learn their first language, while learning is a conscious way of studying a language. Again, in this chapter no difference is made between these terms.

14.3. Characteristics of learner language

Remember the times when you were a beginner? Most probably you used to say things like *he go* and *he going*. These examples reflect some common features of learner language and typical examples of low proficiency levels. There are different ways of analysing learner language. These analyses have been organised around four areas: ERROR ANALYSIS; DEVELOPMENTAL PATTERNS; VARIABILITY of learner language; and pragmatic features of learner language.

14.3.1. Error analysis

Error analysis (EA) concentrates on determining whether the ERRORS students make are systematic or not. Systematicity means that students do not make errors just accidentally but there is some regularity in their errors. Error analysis has four stages: identification, description, explanation and evaluation.

The identification of errors is difficult, mainly because there are differences between errors and MISTAKES. Errors occur because there is a gap in a learner's knowledge. A mistake, however, is a lapse in performance, which means that the learner has already learnt that particular structure but for some reason s/he does not use that. (In this chapter, from now on, "errors" and "mistakes" will be used interchangeably.)

Errors can be classified in different ways. A well-known classification contains four error types based upon how differently students use the second language from the way native speakers do:

1. OMISSION: An item is missing from the utterance e.g. – *. . . but I know it impossible.
2. MISINFORMATION: A grammatical form is used incorrectly e.g. *And at home we could jogging or singing.
3. MISORDERING: The word order is wrong e.g. *At home they can more easily learn.
4. ADDITION: There is an extra item in the utterance which should not be there e.g. *. . . and going to home they haven't enough time.

Having identified and described the errors the next step is to give an explanation for them. Errors originate from different sources. They can be the result of students' attempts to use their mother tongue knowledge. These are called TRANSFER ERRORS. For example a typical error of Hungarian learners is *"I didn't do it yet" instead of the correct "I haven't done it yet." The reason for this error is that the Present Perfect Tense cannot be found in Hungarian, so it is difficult for Hungarian students to learn this grammatical structure. Another mistake frequently made by Hungarians when writing in English is the use of a comma in sentences like *"He said, that he would give that book back" instead of the correct "He said that he would give that book back." The explanation for this is that "that" means "hogy" in Hungarian, before which there is always a comma in Hungarian, while in English no comma is used before "that".

Other errors are made regardless of the learners' native language background. The interesting thing is that most errors are systematic, which means that students create rules about the second language they are learning, but these rules are different from the "real" rules of that language. Researchers also found that several errors are universal, meaning that similar errors can be detected in most students' utterances.

The last step in error analysis is to evaluate the errors. There are several ways to do so. An important one is to make a distinction between overt and covert errors. OVERT ERRORS are ones which can be easily detected outside the context in which they occur e.g. *he gone*, from which expression a "has" is missing. Or a common problem for Hungarian learners is that instead of "I suggest that he should go to the cinema" they say "*I suggest to go to the cinema." COVERT ERRORS can only be detected in the context e.g. using the past simple tense instead of the present perfect when talking about an activity which has not been finished yet. For instance, "I didn't see that film" is correct in itself grammatically, but if you talk about a film that you can still watch, it is not because "I haven't seen the film yet" is correct.

Because error analysis mainly concentrated on the errors students make, later researchers started to analyse learner language from a different perspective: they analysed its characteristic features in general not just the errors. They called learner language INTERLANGUAGE, which implies that learner language has its own system of rules, which lies somewhere between the speaker's mother tongue and the target language.

14.3.2. Developmental patterns

Having acquired some basic L2 expressions in the target language students start learning grammar. One of the most interesting features of people's language learning is that regardless of students' age or mother tongue there are some developmental patterns which they follow. These patterns refer to the order the different grammatical elements are acquired and also the stages learners go through while acquiring elements.

The definite order in which the different grammatical structures are acquired is called the ORDER OF ACQUISITION. In all languages young children go from one-word utterances to two-word sentences and to longer stretches of speech. In the case of learning English the order in which some important grammatical elements are acquired is the following:

1. continuous -ing; auxiliary be; plural -s
2. articles; irregular past
3. regular past; 3rd person -s

While the order of acquisition refers to the order in which several language features (grammatical elements) are acquired, the SEQUENCE OF ACQUISITION refers to what phases learners go through while acquiring a language element. Ellis (1997) illustrates this process with the example of the past form of "eat".

1. First, students do not know the irregular past tense; therefore, they say "*eat*".
2. After that they are able to say the correct form "*ate*".
3. Later they overgeneralise the regular past form and say "*eated*".
4. It may occur that mixed forms are created like "*ated*".
5. In the final phase the correct form is produced: "*ate*".

Naturally, the past tense form of every verb is not acquired in the same way as there may be verbs that students find easier or more difficult to memorize. Nevertheless, the existence of developmental patterns is one of the most significant findings of SLA research as it indicates that acquiring a second language is systematic and universal regardless of the learners' background and the context in which they are learning the language.

14.3.3. Variability in learner language

Have you ever noticed that one day you speak English quite well, but on other days you feel you are totally hopeless? It is a totally natural phenomenon. This is called the variability of learner language, which is also systematic.

Variability depends on at least three factors: (1) linguistic, (2) situational and (3) psycholinguistic. An example for how the linguistic context influences learner language is taken from Ellis (1997): Instead of the correct form of the simple past form (*Joe played football*) learners

tend to use wrongly the continuous form (**George playing football*) when there is no adverb of frequency (e.g. always, usually, often, etc.) in the sentence, and when there is (**In Peru, George usually play football every day*), they use the simple present structure.

The situational context also has an impact on the way we use a foreign language. In formal contexts everybody tries to use the correct forms, but in informal context, e.g. when talking to friends, we do not concentrate too much on using the language correctly.

The psycholinguistic context is the third crucial determining factor in the variability of learner language. For instance, if you have enough time to plan what you want to say, you are more likely to produce the correct form. If you have time to prepare for a role-play, you will most probably produce a better performance than when you have to perform the same situation just off-hand.

14.3.4. Pragmatic features

As you could read in Chapter 6, *pragmatics* deals with how language is used in communication. This is the study of appropriateness, which means that even if we use the correct grammatical structures and vocabulary in a certain situation, an utterance can be totally unacceptable because it is not appropriate to use it in a particular context. For example, you cannot tell jokes while listening to the national anthem or when you are at a funeral.

Learner language can also be full of pragmatic problems. For instance, in a formal essay you cannot use expressions like "*Boy, I was happy or what*" as this sentence is far from being acceptable in a formal text.

The interesting issue here is the same as in the case of errors, developmental patterns and variability: are pragmatic errors systematic or not? Of course, most experts answer this question with a clear "yes."

14.4. Possible explanations for the characteristics of learner language

In the sections above you read some information on the characteristic features of learner language. This section will introduce you to some possible explanations of why learner language is as it is. First, external (social), then internal (mental) factors are examined. Third, the individual variations are taken into account.

14.4.1. Learner external factors

An important factor that influences how people acquire a second language is the circumstances in which they learn: the social context. This social context can be of two types: people

can learn languages in (1) natural and in (2) classroom settings. Acquiring a second language in natural contexts means, for instance, immigrants' learning English in the US, while learning a foreign language in classroom settings means studying the language as a subject at school like learning English in Hungary.

Besides the conditions of learning, several other factors make a huge impact on people's SLA. Students' social class and ethnic identity (sociolinguistic variables – see Ch. 8 *Sociolinguistics*) play an important part in their SLA. These variables interact with each other; therefore, it is difficult to tell to what extent they separately affect learners' SLA. Also, there is always an exception to a seemingly obvious trend. For example, they say that middle class learners tend to become more proficient than working class learners. However, several exceptions to these variables can be found easily.

Apart from these, two other external features have a serious effect on SLA: INPUT and INTERACTION. Input is the language that learners are exposed to, which can be quite different from the "real" target language. This may be a simplified version of the target language, which lacks complicated structures and is spoken more slowly than the "real" target language. One special type of input is called FOREIGNER TALK.

Interaction takes place when the target language is used between speakers, for example, in conversations. What seems to be certain is that it is not necessarily the case that the more input a learner receives, the more successful language learner s/he will be.

14.4.2. Learner internal factors

It must have happened to you that you have made so-called Hunglish mistakes, because you "translated" a Hungarian expression into English, but that expression does not exist in Standard English. For example, **a good advice* is not correct in English, but its Hungarian equivalent *egy jó tanács* does exist. This is NEGATIVE TRANSFER, as the learner's L1 is the cause of L2 errors. If students' L1 helps them acquire a SL, we talk about POSITIVE TRANSFER.

Today, experts say that learners' L1, as one source, helps them create their interlanguage rules (see section 14.3.1.). Learners form ideas (hypotheses) about what the rules of the L2 are, then they test them, and if these rules work, they accept them, if not they reject them.

14.4.3. Language learner

It is common knowledge that there are several individual differences in the way people acquire a second language. Besides the previously discussed internal and external factors, there are several other factors or variables that determine individual differences. The most common of these variables are age, aptitude, motivation, and learner strategies.

As you could read earlier there are controversial opinions on the effects of the age factor to SLA. The most debated issue is the so-called CRITICAL PERIOD HYPOTHESIS (CPH – see 13.3.2.). According to this only people starting learning a second language before a certain age period (before the age of 12) can acquire a native-like accent, but those who start learning a second language after this age cannot. Naturally, there are several exceptions to this.

Some issues, however, seem to be clear today. One is that adults are quicker at learning grammar, but younger learners can overtake them in the long run. Second, not only children in a naturalistic setting can acquire native-like accent, but some adults too. Third, the acquisition of grammar may not be affected by age, which means that anybody, regardless of their age, can acquire grammatical rules properly.

APTITUDE is an innate talent for SLA, which indicates the rate of progress, but it does not tell whether somebody can or cannot learn a foreign language. It is a factor which never changes, similar to IQ but language specific.

MOTIVATION includes factors which cause people to learn a second language; it can change all the time. People can learn languages for several reasons. Some of them are interested in the learning task itself (intrinsic motivation). Others are afraid of being punished if they do not learn, or some may be promised to earn more money (extrinsic motivation). Students can also be motivated in an integrative way, which means that they want to be accepted as a member of the L2 community. Learners are instrumentally motivated if a second language is a device for them to get a good job, for example. Last but not least, resultative motivation is the result of successful language learning experience, which further motivates learners. Perhaps, it is not the type of motivation but its intensity that really matters when we wish to explain why some people acquire a second language successfully.

It is also common knowledge that other personality traits also play an important part in successful SLA. Self-esteem, extroversion, anxiety, sensitivity to rejection, empathy, inhibition, and tolerance of ambiguity have different impacts on the way we learn a second language.

LEARNER STRATEGIES are also important determining factors to explain individual differences among language learners. Learner strategies are steps learners take to learn an L2. There are several types: memory strategies (e.g. grouping), cognitive strategies (e.g. guessing meaning of unfamiliar words from context), metacognitive strategies (e.g. planning learning), compensation strategies (e.g. using gestures), affective strategies (e.g. wise risk-taking), and social strategies (e.g. asking questions for clarification). Successful learners tend to use more strategies, pay attention to form and meaning, and are aware of their own learning styles.

14.5. Conclusion

Now that you have read this chapter, you may realize that we are not very much closer to answering the questions posed in the Introduction. It is no wonder. One reason is that this

area of applied linguistics is relatively young; therefore, research results cannot be generalized to a large extent. The other reason is that SLA is a human activity, and as such, there are many factors to take into consideration, which is why it is very difficult to come up with any clear-cut answer.

Having examined the complicated nature of acquiring a second language you can see several things: (1) there are many ways to learn languages; (2) it is not natural at all that learners will acquire a grammatical item when the item is taught; (3) it is possible that learners will not acquire language items in the order the items are taught; (4) learners and their learning strategies may change over time (5) learners rely on the knowledge and experience they have: L1, other L2s, hypotheses about L2 (6) for most adult learners, a complete mastery of the L2 may be impossible (7) and last but not least, there are plenty of individual variations among language learners, which statement leads us back to the situation described at the beginning of this chapter.

Points to Ponder

1. Could you identify any errors in the language of Hungarian learners of English that are typical of Hungarians?
2. What are your motivations for learning foreign languages?
3. Carry out a small-scale research among your classmates about their motivation types for learning English.
4. Carry out a small-scale research among your classmates about how they think the age they started learning English influenced their language learning success.
5. Some people say that Hungary is the home of false beginners in language learning. Why do you think this is the case?

Suggested Reading

Cook, V. (1993): *Linguistics and Second Language Acquisition.* Macmillan: London
If you are interested in how second language acquisition research, which belongs to applied linguistics, can be discussed from the point of view of theoretical linguistics, you will find this book very useful.

Ellis, R. (1994): *The Study of Second Language Acquisition.* Oxford University Press: Oxford
Perhaps the most comprehensive overview of all the issues related to second language acquisition research. Just to illustrate this, among the 800 pages there is a 40-page glossary of terms and a 60-page bibliography.

Krashen, S.D. (1985): The Input Hypothesis: Issues and Implications. Longman: London
One of the best overviews of Stephen Krashen's hypotheses. Krashen is very easy to read!

Larsen-Freeman, D.—Long, M. H. (1991): *An Introduction to Second Language Acquisition Research.* Longman: Harlow

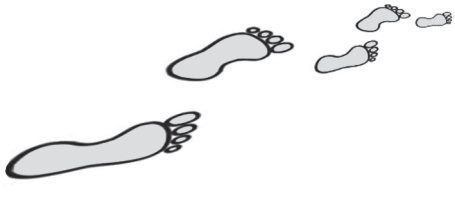
Besides the Ellis book, this is a one of the most comprehensive introductions to the main issues of second language acquisition research. The chapter on SLA theories is extremely interesting.

McLaughlin, B. (1987): *Theories of Second Language Learning.* Edward Arnold: London

A book especially concentrating on clearly describing important second language acquisition theories.

And a website for you to study:

http://homepage.ntlworld.com/vivian.c/SLA/



Do you know some-one who has an intermediate language exam in English, but couldn't buy a sandwich in London? What did he – or she – do in the exam to get a pass mark? How can proficiency exams be designed to make sure that people who pass are really proficient? And what does it mean to be proficient on an intermediate level? In what ways are language proficiency, language teaching and language testing related to each other?

I came across an interesting case recently: a woman, who had spent many years in an English-speaking country, holding down a job there and living with three native speakers, went off to take a proficiency exam. She needed the certificate to get a better job here in Hungary and she was confident of passing easily. She did in fact pass, but only just! If we assume that her knowledge of English was not the problem, what is it that she should have learned to do before attempting the exam? How objective can the markers of letters or compositions be? Do different markers give similar marks? All these elements have to be taken into account in setting a paper to test the language proficiency of the candidates. This chapter goes into some of these questions in more detail.

Zoltán Kiszely
Kodolányi János University College
Department of English Language and Literature

15. How Do We Know What You Know?

Foreign Language Proficiency, Teaching and Testing

15.1. Introduction

Have you ever thought of what it means to know a foreign language? If you speak a foreign language well, does it necessarily mean that you also write well in that language? Have you ever considered the extent to which you know a foreign language? If you have passed an intermediate-level language examination, does that mean your knowledge is on the same level as that of your friend's who has passed the same exam? Moreover, do you happen to know what it really means to know a foreign language on the intermediate level?

These are some of the questions that might occur to you if you do some thinking into the issues of foreign language knowledge, foreign language teaching and foreign language testing. These three issues are strongly related to each other. It means that you can only teach a foreign language or test somebody's foreign language knowledge (proficiency) if you are aware of what it means to know a foreign language. If you do not, you do not seem to know what you should teach and test; and you do not know how to teach and test that. The interesting

thing is that there are several definitions as to what language knowledge is, and this makes it hard even for linguists to give a proper definition. However, there are several examination centres which state that they know what they test since they should not test anybody's knowledge if they do not know what language knowledge covers. But how can they be so sure about what language knowledge means if it is so difficult to define?

Still, it is common knowledge that to know a foreign language you have to be able to comprehend what you hear and read, speak and write in that language. These are the four BASIC SKILLS: listening, speaking, reading, and writing. Some people, however, say that there is a fifth skill as well, which is to be able to MEDIATE between a foreign language and your mother tongue, a usual form of which is written or oral translation. To be able to use these skills, you have to be familiar with the grammatical rules of the language, its pronunciation, and, last but not least, you have to know an ample amount of words and expressions (vocabulary) as well.

So that you can make an attempt to answer the questions asked above, in this chapter, first the connections between knowledge, teaching and testing will be discussed in a historical context. Second, you will familiarize yourself with the basic terminology of modern language testing. Throughout the chapter *language knowledge* and *language proficiency* will be used interchangeably.

15.2. Historical overview: Concepts of language proficiency, language teaching and testing methods

In this section you will read about how the concept of knowing a foreign language has changed in the past one hundred years, and, related to this, what language teaching and testing methods have been dominant throughout this period. The discussed methods and the related concepts of language knowledge do not represent a complete picture; they are the most significant ones. As Communicative Language Teaching is the most widespread approach today, this will be discussed in greater detail.

15.2.1. The Grammar-Translation Method

The oldest and maybe the most well-known language teaching method is the Grammar-Translation Method. It was based upon how Latin and Greek were taught in schools for centuries. The main focus was on teaching grammatical rules, getting students to read and translate literary texts, and to teach them a long list of words that can be found in the texts to be read. The mother tongue was used in the classrooms, and speaking and pronunciation were not practised at all. This method was dominant up to the 1960s, but some of its techniques are still used in language classrooms.

It is not easy to detect what kind of concept of language knowledge there is behind this lan-

guage learning method. It is no wonder that several researchers claim that there is no theory behind this method at all. Nevertheless, it can be seen clearly that out of the five skills only reading and translation were taught, grammatical rules and a long list of vocabulary had to be learnt by heart.

Taking language testing into account in the Grammar-Translation era, we cannot talk about a well-established testing theory. The most significant testing techniques were writing compositions, translations, dictations and transformation tasks. This testing period was mainly characterised by the examiners' intuitive judgements on students' language knowledge.

15.2.2. The Direct Method

The Direct Method, as a reaction to the dominant Grammar-Translation period, started at the end of the 19th century. Its main principles were just the opposite of those of the Grammar-Translation Method. One example of this is that in this method phonetics (see Ch. 2) played an important element. The target language dominated the classroom, mother tongue use was excluded; speaking and listening were the most important skills to be taught; grammar was taught indirectly; everyday vocabulary and sentences were taught; and pronunciation was of high importance.

On the basis of the previously mentioned characteristics, we can see that language knowledge was mainly considered as the ability to speak a foreign language with a proper pronunciation or accent, and proper listening ability was also a significant part of language knowledge.

As far as language testing is concerned in the Direct Method, we cannot talk about a well-established testing theory. The most significant testing techniques were writing descriptions and narratives and also conversations. However, this testing period, similarly to the Grammar-Translation era, was also mainly characterised by the examiners' intuitive judgements on students' language knowledge.

15.2.3. The Audiolingual Method

In the 1950s a new language teaching method was introduced, which, contrary to the Grammar-Translation Method, was based on a solid theoretical background, which was provided by behaviourist psychology and structuralist linguistics. Behaviourists stated that learning is habit formation, which means that the more we practise a language, the better users we will be. Structural linguists made an attempt to describe the phonology (see Ch. 2), morphology (see Ch. 3) and syntax (see Ch. 4) of different languages in as much detail as possible. They also stated that these small elements have to be taught so that students can build sentences on their own later. Another significant viewpoint of theirs was that language is mainly speaking, writing was considered only of secondary importance.

On the basis of behaviourist psychology and structuralist linguistics the most important characteristics of the Audiolingual Method are the following: new language items are presented in dialogues; repetition of learnt language structures is of major importance; pronunciation is important; grammatical explanation is minimal; language laboratories are widely used; and forming correct sentences is significant.

During this period language knowledge was seen as being able to speak a language grammatically correctly with proper pronunciation. Proponents of this method also assumed that if we learn the elements of a language (sounds, words, grammar, sentence structures), we will be able to build grammatically correct sentences and texts in speaking. Writing only helps students reinforce the learnt elements.

In language testing the Audiolingual Method meant the starting point of a period when testing was based upon scientific principles and not on intuitive decisions. The main technique for language testing was the so-called discrete-point technique. Discrete means separate, which meant that if a language can be divided into very small pieces so that it could be taught to students, in the same way, the knowledge of these small elements should be tested in a way that they are separated from each other. The well-known multiple choice grammar tests are a good example for a discrete-point test, in which separate elements of a language are tested on sentence-level, one at a time.

15.2.4. Communicative Language Teaching

Hopefully, you are familiar with Communicative Language Teaching as it is the most popular trend today. It became popular in the 1970s and its supporters claim that language is a means of communication, so the major aim of teaching is to get students to communicate in the target language. In other words, grammatical correctness is not the only important thing that students have to be aware of, but they also have to know how to use the language in very different situations when they meet different people. Pragmatics (see Ch. 6) became a key concept in language teaching. Authentic texts which have not been made for teaching purposes (e.g., newspaper articles) are used in the classroom, providing context for classroom tasks. In the communicative classroom real-life situations are simulated: role plays have to be acted out in such a way that its participants do not know what their partner wants to achieve, they are only aware of what they want to achieve. So, life-like situations have to be solved by communication just as it happens in real life as well.

The concept of language knowledge on which Communicative Language Teaching is based started to develop when a difference was made between language COMPETENCE (see 16.4.1.) and language PERFORMANCE (see 16.3.1. and 9.2.1.). Competence means what we know about a language, about its grammar, how it works. Our foreign language competence can change, hopefully, for the better; the more we learn, the higher our competence level becomes. Language performance is how we actually use our competence in real life situations when we have to listen, speak, read or write in a language. Our performance keeps changing all the

time. You must have experienced it several times that there are days when you speak the foreign language quite well, but on another day you feel you cannot say a word. The aim of language learning for anybody is to become a competent user of that language.

Later the distinction between competence and performance was not enough to describe people's language knowledge; therefore, a new term "COMMUNICATIVE COMPETENCE" was introduced, which included both competence and performance. It covered not only grammatical competence but also the appropriate use of language structures.

Later, researchers felt that this definition was just the starting point and it has to be further developed so that we can have a closer picture of what it means to be a competent language user. They claimed that a communicatively competent user of a language has to have four different competences: grammatical competence, discourse competence, sociolinguistic competence and strategic competence. Grammatical competence means the ability to understand the word-for-word meaning of expressions and sentences, and to form grammatically correct sentences by using the rules of the language. Discourse competence is concerned with the ability to create texts in which sentences are logically connected to each other. In other words, this is the ability to create coherent and cohesive texts; the previous one means linking ideas in a text that make a text logical and understandable, while the latter refers to the grammatical and lexical elements that link the sentences of a text together (COHESION and COHERENCE). Sociolinguistic competence covers the appropriate application of sociocultural rules; for example, you have to know that there are differences between the way you talk to your teachers and to your friends. Strategic competence mainly refers to learners' verbal and non-verbal strategies to compensate for their lack of knowledge in the target language.

Based upon this notion of language knowledge, communicative language testing has several features which were not characteristic of earlier testing periods. First of all, quite obviously, communicative tests are intended to measure the elements of communicative competence. They are integrative in nature, which means that two or more language skills are tested in the same task. Authentic texts are used for testing. Grammar is not tested in separate sentences but with the help of texts containing several sentences. Common techniques in communicative language testing are, for example, compositions, GAP-FILLING TASKS and CLOZE-TESTS. The evaluation of these tests is carried out by at least two examiners (RATERS). Naturally, these techniques are not always present in every communicative test, for instance, it is not the case that every task is integrative in a communicative test; however, these are still the basic characteristics.

15.3. Basic notions of language testing

So far you have come to terms with the historical changes of the concept of language proficiency, and some important teaching and testing methods. In this section you will learn about some of the major issues in modern language testing. You will read about the

characteristics of good tests, test types, the washback effect of tests, and the main steps that should be followed to develop a good test.

15.3.1. Characteristics of good tests

As far as testing is concerned several issues must be taken into account. One of these is that in a language classroom students' language knowledge can be assessed from a lot of aspects: their listening, speaking, reading and writing skills, their vocabulary and grammar can all be tested. Naturally, all these skills have to be assessed if we are interested in somebody's general language proficiency. This is the issue of VALIDITY, which is of key importance in language testing. To put it differently, a valid test should measure what it intends to measure. For example, imagine the following: in the grammar school you learnt English and you got a grade after each semester for your general English proficiency. However, you had to write only grammar and vocabulary tests and never had to make any oral presentations. Your final grade was based upon the test scores in grammar and vocabulary but knowing the language does not only mean being good at grammar and knowing a lot of words, but also being able to read, write, understand and speak in that language. So, the final grade does not reflect a complete picture of the students' real language knowledge; in other words, it does not test what it should.

Another example of validity problems is the case of assessing somebody's speaking ability only with one task: reciting a text which s/he learnt by heart. Even an extraordinary performance in this task does not necessarily mean that spontaneous conversation would be just as easy for the student. Therefore, we can state that this test did not test what it wanted to: the student's speaking skills in general.

The third example is from the realm of writing: if you have to write a composition entitled "Compare the advantages and disadvantages of nuclear and water power stations, and state clearly which you prefer.", this task is not valid if it intends to measure students' composition skills as the ones who have some knowledge of the topic will probably write a more well-organized text than those who do not know much about the topic. The reason for this is that the topic is too specific.

Apart from validity, RELIABILITY is another key factor in testing, and it covers two things. One is the extent to which test scores are consistent: if candidates took the test again tomorrow after taking it today, would they get the same result? If yes, the scores are reliable, if not, there must be some problems with that test. The other meaning of reliability refers to the examiners' work: if two examiners, separately from each other, give the same score to the same student essay, the INTER-RATER RELIABILITY of that test is high. If their scores are very different, the inter-rater reliability is low. INTRA-RATER RELIABILITY is high if one rater gives a certain score to an essay on a certain day, and some days later the same rater gives the same score to the same essay.

It is common knowledge among language testers that a test can be reliable without being valid, but it cannot be valid without being reliable (see "Points to Ponder" at the end of this chapter).

15.3.2. Test types

Based on what is tested, we can talk about ACHIEVEMENT TESTS, which are based on a syllabus or a textbook and intend to test how successfully students managed to learn the material covered in that textbook over a week, term, or year. PROGRESS TESTS are very similar to achievement tests as they intend to measure progress during a course. For instance, in a 15-unit textbook there can be three progress tests altogether, one after every fifth unit, and one achievement test at the end of the book. We can also talk about PROFICIENCY TESTS, which are not related to any SYLLABUS as they are intended to test students' level of proficiency. Any language examination system accredited in Hungary ("érettségi", *társal*KODÓ, ECL, etc.) belongs to this category.

Based on why we test language knowledge, there are DIAGNOSTIC TESTS intended for diagnosing candidates' strengths and weaknesses. The aim of PLACEMENT TESTS is to measure proficiency as related to levels or groups, for example, to find out in what language level group somebody should start their studies in a language school. A FILTER TEST aims to filter out candidates whose level of proficiency is below a certain level. A university entrance test, which in today's Hungary is the advanced-level final examination ("érettségi"), is a typical example for this test type.

As discussed earlier, based on how we test we can distinguish discrete point and integrative tests. DISCRETE-POINT TESTS measure how well students know separate elements of a language always on sentence-level one at a time. Multiple choice grammar tests belong to this category. For example:

On the first day, we came across a young couple and their (1) _____ son.
(A. seven-year-old B. seven years old C. seven-years-old D. seven year old)

INTEGRATIVE TESTS test two or more (or a number of) skills together: For instance, you have to read a letter to which you have to reply. In this task your reading and writing skills are tested at the same time because you have to understand the letter so that you can reply to that.

Of course, a test can have several purposes at the same time, for example, a proficiency test can be a filter test at the same time.

15.3.3. Washback effect of tests

A test can have a positive or a negative WASHBACK EFFECT on teaching and learning. It is positive if the aims of the course and the testing are the same: similar tasks can be found both in the exam and during the language course. It is negative, however, if it is testing that determines the content of the course. It happens if tasks different from the ones that are in the exam are not dealt with during the course at all. In this case, students do not learn the language but rather they prepare for the exam. If you learn a foreign language according to the communicative approach, you should not change your learning methods just because you have to take an examination.

15.3.4. A standard model of developing a good test

Think of the "érettségi" test you had to complete not a long time ago. Both the content and the layout seemed quite professional, did it not? No wonder, as it must have caused a hard time for the testers to make the test look and be so masterly. A test, mainly if it is a high-stakes test like the "érettségi", has to be compiled in a rigorous way, which, in a nutshell, goes like the following:

First the test questions (TEST ITEMS) have to be written by somebody who knows how to write good items (TESTERS). After that this item-writer has to show other item-writers what s/he wrote and they have to discuss how acceptable those items are. If there are some wrong items, they must be changed. In the next phase the test must be tested (PILOT PHASE) – some students should complete the proposed test. These students cannot be the ones we want to test. The results have to be statistically analysed. If there are some items which do not test students' competence well enough, they have to be changed. After that the test has to be formatted and students can complete that. Then the tests have to be evaluated. Also, the passmark has to be set, which means that you have to state what the minimum score is students have to achieve so as not to fail the test.

15.4. Consolidation

In this chapter you could see how differently the concept of language knowledge has changed over a century. As a result, not only the teaching but also the testing methods have changed. Nowadays, by language knowledge we understand communicative competence, the teaching of which is the aim of today's most popular language teaching approach, Communicative Language Teaching. As far as testing is concerned, from an intuitive approach it has developed to be an independent subfield in applied linguistics with its own scientific principles. Perhaps the most important issue in language testing is that any language test can be of high quality if it measures well what it intends to, in other words, if it tests exactly the concept of language knowledge its developers have.

Points to Ponder

1. Think about your language learning experience. What teaching methods did your previous teachers apply? Which did you prefer? Why?
2. Can you state that a test containing only grammar items does not test what it should? Why?
3. Why do they say that a test can be reliable without being valid, but it cannot be valid without being reliable?
4. Ask some of your group-mates to complete the reading tasks in two different intermediate language examinations. Compare their results. If there are big differences between their results, why do you think this could happen? If their results are very similar, what does it tell you about the tests?
5. Compare the advanced-level final examination and an accredited examination with each other. Do you think there is a difference between what the developers of these tests think of language knowledge?

Suggested Reading

Alderson, J. C., Clapham, C. and Wall, D. (1995). *Language Test Construction and Evaluation.* Cambridge University Press: Cambridge.

Perhaps the best and most readable introduction to how a language test should be created, tried out, administered and evaluated. Highly recommended for anybody wanting to have easy access to the tricks of language testing.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing.* Oxford University Press: Oxford.

In this book you can learn about a communicative competence model, which is called communicative language ability, and also about the basic notions of language testing.

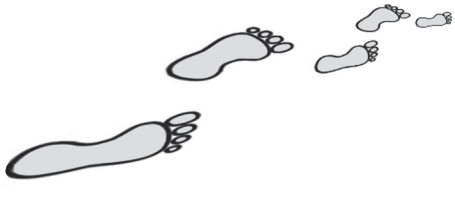
Bachman, L. F. and Cohen, A. D. (eds.) (1998). *Interfaces Between Second Language Acquisition and Language Testing Research.* Cambridge University Press: Cambridge.

This book is for readers interested in both second language acquisition research and the latest issues in language testing and also how these two seemingly distant areas are connected to each other.

Brown, H. D. (1987). *Teaching by Principles: An Interactive Approach to Language Pedagogy.* Prentice Hall: New Jersey.

One of the best introductory books to language teaching for would-be language teachers. It discusses the theoretical background to language teaching, gives ideas how the four skills should be taught effectively and also talks about planning and managing language classes.

Stern, H. H. (1983). *Fundamental Concepts of Language Teaching.*
Oxford University Press: Oxford.
A classic book on the theoretical and historical background to language teaching. The chapters of the book are the following: Clearing the Ground, Historical Perspectives, Concepts of Language, Concepts of Society, Concepts of Language Learning, and Concepts of Language Teaching.



It happens frequently that when I meet people for the first time, the conversation comes round to what languages I speak, where I learnt them, and what my mother tongue is. Sometimes this happens because my first name is strange to English ears, sometimes because I speak German with an accent, and sometimes because when I speak Hungarian, it sounds a little unusual in some way. Although I don't mind answering these questions, the answers are never easy. Perhaps the best answer to what my native language is, is that I am bilingual, since I learnt Hungarian and English very nearly at the same time. Yet I speak, and certainly read and write, English better. As to whether I am bilingual in German, a language I learnt in my thirties? The answer to that can only be "It depends". What it depends on, forms part of the content of this chapter.

Borbála Richter
Kodolányi János University College
Department of English Language and Literature

16. Double Trouble?

Bilingualism

16.1. What about you?

As you start to read this chapter, think about what you are doing. You are probably a Hungarian who has been learning English as a foreign language for some years. Perhaps you have been able to spend some time in an English-speaking country. Certainly you like English enough to want to continue studying it. But are you a bilingual? Do you yourself call yourself a bilingual? How well does someone have to be able to speak more than one language in order to be termed a bilingual? Is this term reserved for people who speak two languages equally well? Or do you think being bilingual has more to do with growing up speaking two languages?

16.2. Mother tongue plus one (or more)

What can be said is that any person who is bilingual by one of these definitions is proficient to some degree, and uses in some way, a language other than his *mother tongue*, or *native language*, or *first language*, or *dominant language* – as you can see, even this basic term mother tongue is full of pitfalls when we move out of everyday language use and try to define it for purposes of study. What is your mother tongue? The language of your mother, almost certainly. What, however, if the language spoken by your mother is not the language of your father? Or of the community you live in? What if you only speak your mother's language to her, and speak another language at school, where you learn to read and write that language on a much higher level? What is then your mother tongue?

We can clarify the issue with the help of four criteria: origin, identification, competence and function (Skutnabb-Kangas, 1984)

Origin:	the language/s learnt first, or in which the first long-lasting verbal contacts happened
Identification:	
Internal:	the language one identifies with (this is decided by the speaker)
External:	the language one is identified with (this is decided by other people)
Competence:	the best-known language
Function:	the most-used language

16.2.1. The simplest case

For people who speak only one language (monolinguals), living in a community that speaks that same language, the four criteria are all met by the one language. This is probably true for most Hungarians, and it is traditionally considered true for Americans and even the British. However, about half of the world grows up speaking more than one language. This number is a rough estimate, since you can imagine how difficult it would be to conduct a survey on this topic. It is clear that these naturally bilingual people can have more than one mother tongue, and that their mother tongue can change during their lifetime (except as regards origin).

16.2.2. The question of language rights

If we look at the fourth criterion, namely function, we can also see that the issue of mother tongue can be very political: many people in powerless situations are not able to choose which language they want to use most. An example of this is the Deaf person who may claim Sign language as a mother tongue in a society where deaf people are not thought to have a language at all, or where the hearing outsiders think that the deaf child's mother tongue should be, or should become, the oral language the parents speak (or a coded form of this language).

This is linked to the issue of proficiency as well. Very often being bilingual is not a matter of choice. It is generally necessity or opportunity that decides which language will be spoken best. Education is expensive and decisions about the language of instruction are usually made on government level: the language of the home and the language of school may not be the same.

Allowing people to use their own language in official situations is even more expensive. Think about the position of Hungarian in the European Union. Would it not be easier just to decide that all European Union citizens should speak just one language? If your instinctive response was to feel uncomfortable with that idea, then you have recognised that language and identity are closely linked. Despite our personal uniqueness, our personal characteristics develop in a social context, speaking a particular language means belonging in some way to a speech community. Membership in more than one speech community complicates an al-

ready complex pattern. Researching this dynamic interrelationship is an important area in studies of bilingualism.

Language choice reflects social patterning and shows how a bi- or multi-lingual society is structured and the functions of the languages used there. Fishman suggests that one way of investigating this is to answer the question 'Who speaks which language to whom and when?

16.2.3. The question of labelling

Likewise, there can be conflict between the way people see themselves and others see them. A particular person could claim a mother tongue even without being able to speak that language! For example, some people feel they belong to an ethnic group, but they may not learn their mother tongue in infancy. In cases where the political system is sympathetic towards their wishes, they, or their children, may be able to learn their mother tongue in school. There are programmes designed to bring a language back to life, such as Welsh in Wales, or Gaelic in Ireland. Other programmes help groups maintain or reclaim their language while they live in a larger language community speaking another language. In Hungary this is true for the German minority, where, for many members of this group, German is a 'grandmother tongue'. The younger generation is being helped to relearn the language of their forefathers by making a place for this in the schooling system. The middle generation, on the other hand, was generally discouraged from learning German in their home or at school.

16.2.4. The consequences of choice

Taken to an extreme, official policies forbidding the use of certain languages can lead to a language being forgotten or not used. In the worst case, we can even speak of LANGUAGE DEATH or LINGUICIDE.

Even when a language does not die, it may change, and the situations in which it is used may change. Bilinguals can often choose which language to use. Their choice may depend on many things, including language proficiency, the prestige of the language or of its users, the relationship between certain people in certain roles, and so on. Many minor decisions taken on a daily basis by many individuals can add up to a major trend in time. In this way the Hungarians of Felsőőr or Oberwart (these names refer to the same place, in Austria) have moved towards becoming speakers of German, a process called language shift. The two languages, Hungarian and German, existed side by side and affected each other and each others' speakers. In Europe, if you look at the areas alongside the national borders, you can frequently find areas where two, or even more, languages co-exist.

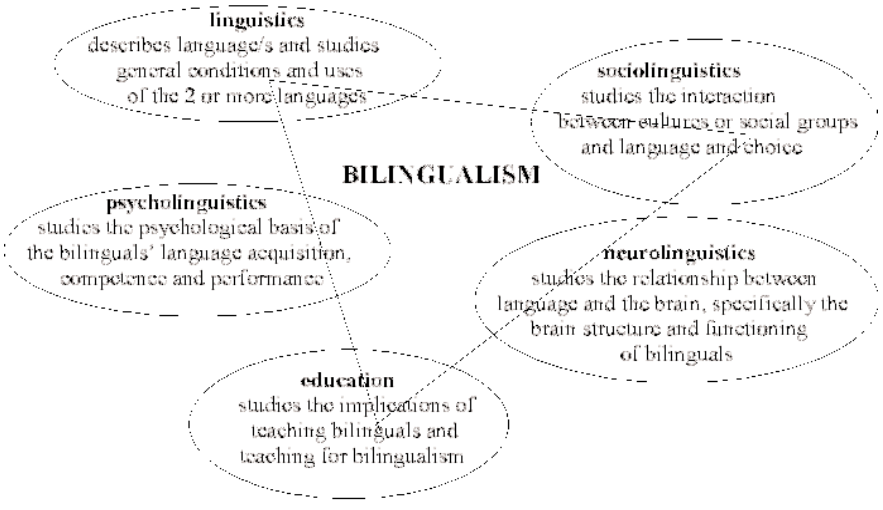
Language planning theory can be used to help save languages or reverse language shift. When a minority language is threatened, the situation can be analysed by placing it on a scale, where the one end urges the recording of the language for possible later reconstruction and moves through stages to the relatively unthreatened status of a minority language that is available in higher education, central government and the national media.

16.3. The scholarly focus

From the above, you can already have some idea of the scope and range of studies of bilingualism. After the chapters on Sociolinguistics and Psycholinguistics, you could have expected this chapter to be sub-titled "Bilinguistics", rather than Bilingualism. The fact that the name does not follow the pattern of other areas of applied linguistics shows that bilingualism, at least at the moment, is less a field of linguistics than a focus. This chapter is called "Double Trouble" because the study of bilingualism means that specialists in various fields deal with it, bringing with them their own perspectives and models, and the methods that have helped them in the past – but now their focus has to be doubled. In fact, every chapter that you have read so far is relevant in some way to the study of bilingualism and bilinguals.

16.3.1. Bilingualism within other fields

Bilingualism has become an emerging focus of interest, with the interest and attention coming mainly from 'pure' linguistics, sociolinguistics, psycholinguistics, neurolinguistics and education, although other perspectives are possible.



16.4. Misconceptions about bilingualism

From what has been said so far, it is clear that bilingualism is a characteristic of individuals, groups and societies, and studies of it can focus on various elements and implications. Let us look now at some commonly held ideas about bilinguals and see what researchers have confirmed or refuted.

- **Misconception 1.** True bilinguals never mix their languages; people who mix their languages can be considered only 'semi-lingual' since they can speak neither language like a native speaker.
- **Misconception 2.** A child needs to be fluent in one language before it should be taught a second language. But only very young children can become maximally bilingual (i.e. close to native speaker proficiency in both languages), after a certain age this is no longer possible.
- **Misconception 3.** Two languages confuse a child and growing up bilingual leads to lower intelligence: bilingualism always implies loss, a splitting of finite cognitive potential.
- **Misconception 4.** There is no need to teach immigrant children their "new" language, they will pick it up easily on the playground.

Before we go on, think about these statements and decide what you think. Most of them could have been said differently, for example, "At school older children learn languages better and faster than younger schoolchildren", or "Even adults can learn a second language so well that no-one can tell that they are not native speakers". There are many beliefs about bilingualism and bilinguals and the interesting thing is that studies of bilingualism often show contradictory results. Let us look at these misconceptions in turn.

16.4.1. Misconception 1. Code-switching

If you have ever observed a bilingual family in action, you may have been surprised by the number of times the members mixed their languages. This is called CODE-SWITCHING and it used to be assumed that it reflected some deficiency in the language competence of the speaker. You may recognise it as a technique used in language learning to overcome a communicative stumbling block caused by limited lexical resources. In this case it is often accompanied by hesitation. Research has, however, found that code-switching can be used to achieve multiple communicative purposes. In many bilingual communities it is seen as normal. The speakers communicate fluently, without hesitation, pauses or changes in rhythm, pitch or intonation. Such a bilingual mode of communication is rule-governed, and can serve many purposes, including the demonstration of solidarity. Investigating the kinds of rules that govern code-switching can help us to learn about the nature of language in a general sense, providing evidence of both its flexibility and inflexibility.

Bilinguals learn to separate their languages but may choose not to do so. A bilingual has the choice of activating or deactivating a language, depending on the context. However, there is never total deactivation of one language: the bilingual chooses a base or MATRIX LANGUAGE but keeps the other language 'active', allowing words, phrases, clauses and sentences to be included from that language as well. From a psycholinguistic perspective, the interaction procedure is still unclear and the challenge is to explain how all this can happen so fast and so efficiently. It is clear that bilingualism in the individual is not simply a form of monolingualism doubled, nor can it be studied in isolation, without its social and cognitive correlates.

16.4.2. Misconception 2. Critical period and order of acquisition.

This is an important question for the study of language acquisition in general, as current theories based on monolingual children will at some point have to include the phenomenon of bilingual language acquisition. SUCCESSIVE, or CONSECUTIVE, BILINGUALISM is common, but so too is becoming bilingual simultaneously. Recognising this can be important for families where bilingualism is not a necessity but a choice. In today's globalising world, knowing several languages can be an advantage and in many educated families it is assumed that their children will become fluent in at least one foreign language in the course of their education, a phenomenon known as ELITE BILINGUALISM.

Weinreich (1953) made a distinction between compound and co-ordinate bilinguals, and the testing of this distinction is still continuing. COMPOUND BILINGUALS supposedly have one semantic system and two language codes, while CO-ORDINATE BILINGUALS have two semantic systems and two codes. Compound bilinguals acquire their languages in the same context, at the same time, while co-ordinate bilinguals acquire their languages in different contexts, at different times. For example, a child may grow up learning two languages from infancy, while another may acquire one language as an infant and the second language later – with puberty possibly marking a decisive point for difference in brain activity pattern.

Modern techniques in neuroscience make it possible to map brain activity using neuroimagery, with implications for understanding the representation of language/s in the cortex. One of the questions that comes up again and again in research on bilingualism is whether bilinguals store language in the same way as monolinguals do. Do children who grow up bilingual from the beginning have a different mental lexicon from children who acquire a second language a few years later? Neurolinguistics can help us test the various ideas about this subject. Findings so far have been contradictory, indicating both shared and divergent representations of languages in the bilingual brain.

One question of great interest is whether such research can help us gain clarity on whether or not there is a critical period for language learning. Recent findings suggest that early bilinguals show similar activity in certain areas in the brain while late bilinguals show a different pattern of activation. Such research can be supported or refuted by looking at aphasiacs – people who have suffered damage to portions of their brain that are responsible for language. While patterns of damage vary, individuals with Broca's aphasia are non-fluent, speaking in short, meaningful phrases that are produced with great effort, while individuals with Wernicke's aphasia are fluent, but speak in long, meaningless sentences. Permanence of damage or patterns of recovery can help us test hypotheses about the anatomical representation of language. Bilingual aphasiacs need to be assessed in all their languages. The recent development of a Bilingual Aphasia Test consists of an evaluation of the patient's multilingual history, the assessment of language disorders in the various languages, and an assessment of translation abilities and interference detection in each language pair. It is available in many languages, and its use will have the added advantage of making future studies comparable.

16.4.3. Misconception 3. Cognitive disadvantage

Perhaps you disagree, thinking that bilingualism is a good thing, that children growing up bilingual are fortunate and even that being bilingual may make a person more intelligent. Yet for many years studies published in America supported the connection between bilingualism and lower intelligence. The researchers who undertook these studies ignored biographies published at the time, which reported no negative consequences of bilingualism. The turning point came with a publication of a study by Peal and Lambert (1962) that questioned the previous findings. By looking closely at the studies, they were able to identify problems in methodology and research design that explained their results. The children examined generally came from immigrant families, in difficult socio-economic circumstances; they were assumed to be 'bilingual' without that being clearly defined or tested; and the tests were often administered in their weaker language. To test the hypothesis that bilingualism was the cause of cognitive deficit the 'control' group should have been as similar as possible in all ways, with the single exception of being monolingual. Peal and Lambert tried to do this in their study and came to the opposite conclusion: bilingualism was linked to some cognitive advantage. Specifically, the bilinguals scored significantly higher on most measures of verbal and non-verbal intelligence.

Research now seems to show positive effects of bilingualism on the linguistic and educational development of children. Studies suggest that bilingual children who have developed both their languages have a better understanding of language, increased language awareness, and are able to use language more effectively. Bialystok, for example, investigated preschool bilinguals and found evidence for a cognitive advantage due to their ability to selectively attend to relevant information.

Peal and Lambert's conclusions have therefore been cautiously borne out. I say 'cautiously' because these newer studies are careful to define exactly what they are studying and to make only justified claims. It is clear, however, that the number of studies showing positive effects for bilingualism far outweigh the ones that point to negative effects. These positive effects are related to mental flexibility and concept formation, memory, language awareness, and creativity. It seems that bilinguals transfer skills and knowledge across languages.

16.4.4. Misconception 4. Natural acquisition needs no support

Education is an area in which most attention is paid to 'normal' children. It is only recently the educational systems in our western civilization have recognised bilingualism as a normal condition. Yet it seems that bilingual children still often struggle to achieve their full potential in school. Despite the potentially positive effects of bilingualism, there are still studies that show children scoring lower on various measures. One explanation for this is that children seem to 'pick up' basic conversational skills with ease and this can be mistakenly interpreted by teachers and parents as native-like language skills appropriate for their age group. Numerous studies have confirmed that learning academic language skills takes significantly longer. This is an area where applying insights from research can have immediate benefits. These issues need to be addressed in curriculum design and teaching practice.

While there seems to be little proof that children need to acquire one language before starting to acquire a second, it does seem to be the case that school children who achieve not only a high level of language proficiency in their mother tongue but also develop cognitive academic skills in that language, succeed better in school. What they have learnt in one language can be transferred to the other, thereby counteracting the negative effects of having to study in a second language. From this we can hypothesize that there is a common cognitive ability underlying behaviour in both languages.

16.5. Summary

In this chapter we have looked at what bilingualism is, both as regards individuals and as regards society. From this it has been possible to give you a taste of the broad range of research topics and issues that bilingualism offers. The interdisciplinarity of this focus of applied linguistics has become clear. Perhaps you have also had to reconsider some ideas you may have had about bilingualism. As the last chapter of this book, the intention has been to pull to-

gether some of the recurring themes of the book as a whole, demonstrating that while the various branches of theoretical and applied linguistics focus on differing aspects of language and language use, all the parts are interconnected and form part of a greater whole.

Points to Ponder

1. Consider what you know of the situation of the 13 official minorities in Hungary. In what way is their position different from that of dual medium schools teaching languages of wider communication, such as English, French and German? Would you call Hungary a multilingual state?

2. WITZ
Situation: an English lesson. The teacher of English is from Germany and learnt English originally as a foreign language; she does not speak Hungarian. The students at a secondary school for the German ethnic minority in Hungary

> Teacher: (showing students a cartoon) *What are these pictures about?*
> Student: *It's funny, a story. It is a ...*
> (asking another student: "Mi a *vicc* angolul?")
> Teacher: (recognising the word from the German word "Witz")
> *Joke, the word you want is joke.*

Think about the elements of this situation. What language is being spoken? When? To whom? Why? Who is bilingual here? Who is accessing more than one language system? Who is accessing more than one mental lexicon?

3. Imagine you have children of your own. Would you like them to grow up bilingual? Why? Think about the conditions which would make this decision easier/harder to carry out.

Suggested Reading

Grosjean, F. (1989). Neurolinguistis, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3-15.
An interesting article that distinguishes between monolinguals and bilinguals.

Hamers, J. F. ,Blanc, M.H.A. (2000) Bilinguality and Bilingualism, 2nd edition, Cambridge: CUP
A completely revised edition of a classic, comprehensive study of bilingualism. The authors call the use of two languages on societal level bilingualism and by the individual , bilinguality. This book is a very up-to-date survey of recent research, focussing on empirical studies. It gives a balanced account of recent research.

Romaine, Suzanne (1995). *Bilingualism.* Oxford: Blackwell.
Another revised edition of a well-known book, Romaine provides an authoritative introduction to the sociolinguistics and psycholinguistics of bilingualism.

Skutnabb-Kangas, T. (1984). *Bilingualism or Not?* Clevedon: Multi-lingual Matters
The author is a vocal and dedicated supporter of language rights. In this work she investigates, among other matters, the issue of defining a mother tongue.

Wei, Li (ed.) (2000) *The Bilingualism reader* London ; New York : Routledge.
This is a collection of `classic' articles by acknowledged scholars and functions as a basic source book for students interested in bilingualism, focussing on code-switching and language acquisition in bilingual children. The book does not, however, cover work on bilingual education, language maintenance, language shift and language attitude.

Further Steps

Bibliography

Aitchison, J. (1988[4]): *The Articulate Mammal: An Introduction to Psycholinguistics.* Routledge: London.

Aitchison, J. (2003[3]): *Words in the Mind: An Introduction to the Mental Lexicon.* Blackwell Publishers: Oxford.

Akmajian, A.– Demers, R. A.– Farmer, A. K.– Harnish, R. M. (1990): *Linguistics. An Introduction to Language and Communication.* The Massachusetts Institution of Technology: Cambridge.

Alderson, J. C. – Clapham, C. – Wall, D. (1995): *Language Test Construction and Evaluation.* Cambridge University Press: Cambridge.

Aronoff, M. (2004): *What is Morphology?* Blackwell Publishers: London.

Atkinson, M. – Britain, D. et al. (1999): *Linguistics: An Introduction.* Part 2, Ch. 12. Cambridge University Press: Cambridge.

Austin, J. L. (1962): *How to Do Things with Words.* Oxford University Press: Oxford.

Bachman, L. F. (1990): *Fundamental Considerations in Language Testing.* Oxford University Press: Oxford.

Bachman, L. F. – Cohen, A. D. (eds.) (1998): *Interfaces Between Second Language Acquisition and Language Testing Research.* Cambridge University Press: Cambridge.

Bachman, L. F. – Palmer, A. S. (1996): *Language Testing in Practice.* Oxford University Press: Oxford.

Baker, C. (2000): *A Parents' and Teachers' Guide to Bilingualism.* 2nd Edition. Multilingual Matters: Clevedon, England.

Baker, M. (1992): *In Other Words: a Course-book on Translation.* Routledge: London

Bank of English (2006): http://www.titania.bham.ac.uk/

Barnhart, S. (no date): *The Tower of Babel and Its Lessons.* www.lewrockwell.com/barnhart/barnhart7.html

Bauer, L. (1990): *Introducing Linguistic Morphology*. Edinburgh University Press: Edinburgh.

Baugh, A. C. – Cable, T. (2002): *A History of the English Language.* London: Routledge.

Beardsmore, H. B. (1982): *Bilingualism: Basic Principles.* Avon: Tieto.

Beaugrande, R. de – Dressler, W. (1994): *Introduction to Text Linguistics.* Longman: London.

Bell, R.T. (1992): *Translation and Translating: Theory and Practice.* Longman: London

Bialystok, E. (2001): *Bilingualism in Development: Language, Literacy & Cognition.* Cambridge University Press: Cambridge.

Biber, D. S. – Conrad, S. – Leech, G. (2002): *Longman Student Grammar of Spoken and Written English.* Pearson Education Limited: Harlow.

Blake, N. (ed.) (1992): *The Cambridge History of the English Language. Vol. II. (1066-1476).* Cambridge University Press: Cambridge.

Bolinger, D. – Sears, D. A. (1981): *Aspects of Language*. Harcourt Brace Jovanovich, Inc: New York.
Brown Corpus. (1961).
http://clwww.essex.ac.uk/w3c/corpus_ling/content/corpora/list/private/brown/brown.html
Brown, G. – Yule, G. (1989): *Discourse Analysis*. Cambridge University Press: Cambridge.
Brown, H. D. (1987): *Teaching by Principles: An Interactive Approach to Language.* NJ: Prentice Hall Regents: Englewood Cliffs.
Burrow, J. A. – Turville-Petre, T. (1996): *A Book of Middle English*. Blackwell Publishers: Oxford.
Bynon, T. (1977): *Historical Linguistics*. Cambridge University Press: Cambridge.
Canale, M. (1983): From Communicative Competence to Communicative Language Pedagogy. In Richards, J. C. – Schmidt R. W. (eds.) *Language and Communication* (pp. 2-27). Longman: London.
Carstairs-McCarthy, A. (2002): *An Introduction to English Morphology.* Edinburgh University Press: Edinburgh.
Chomsky, N. (1957): *Syntactic Structures*. Mouton: The Hague.
Chomsky, N. (1975): *Reflections on Language*. Pantheon Books: New York.
Chomsky, N. (1965): *Aspects of the Theory of Syntax*. The M.I.T. Press: Cambridge.
Clark, H. H. (1996): *Using Language*. Cambridge University Press: Cambridge.
Collins COBUILD English Dictionary. (1995): 2nd ed. HarperCollins: London.
Cook, G. (1994): *Discourse and Literature*: *The Interplay of Form and Mind.* Oxford University Press: Oxford.
Cook, V. (1993): *Linguistics and Second Language Acquisition.* Macmillan: London.
Crystal, D. (ed.) (1997): *The Cambridge Encyclopedia of Language*. 2nd ed. Cambridge University Press: Cambridge.
Crystal, D. (1985): *Linguistics*. Penguin: London.
Cummins, J. (2000): *Language, Power, and Pedagogy. Bilingual Children in the Crossfire.* Multilingual Matters: Clevedon, England.
Curtiss, S. (1977): *Genie: A Psycholinguistic Study of a Modern-day "Wild Child"*. Academic Press: London/New York.
De Villiers, P. A. – de Villiers, J. G. (1986): *Early Language*. Fontana Press.
É. Kiss, K. (2004): *Anyanyelvünk állapotáról*. Budapest: Osiris Kiadó.
Ellis, R. (1994): *The Study of Second Language Acquisition*. Oxford University Press: Oxford.
Ellis, R. (1997): *Second Language Acquisition*. Oxford University Press: Oxford.
Fasold, R. (1990): *Sociolinguistics of Language*. Oxford: Blackwell Publishers.
Fasold, R. – Connor-Linton, J. (2006): *An Introduction to Language and Linguistics.* Cambridge University Press: Cambridge:
Finegan, E. (1989) (2nd ed.): *Language, Its Structure and Use*. Fort Worth: Harcourt Brace College Publishers.
Fischer, O. – van Kemenade, A. – Koopman, W. – van der Wurff, W. (2000): *The Syntax of Early English*. Cambridge University Press: Cambridge.
Fishman, J.A. (1965): Who speaks which language to whom and when? *Linguistique* 2, 67-88.

Fowler, R. (1986): *Linguistic Criticism*. Oxford University Press: Oxford.
Fromkin, V. (1980): *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. Academic Press: London / New York.
Fromkin, V. – Rodman, R. (1998): *An Introduction to Linguistics.* Holt, Rinehart and Winston: New York.
Gardner, R. – Lambert, W. (1972): *Attitudes and Motivation in Second-Language Learning.* Newbury House: Rowley, MA.
Garman, M. (1990): *Psycholinguistics: Central Topics*. Methuen: London.
Granger, S. (ed.) (1998): *Learner English on Computer*. Longman: London.
Graver, B. D. (1986): *Advanced English Practice.* Oxford University Press: Oxford
Grosjean, F. (1989): Neurolinguistis, beware! The bilingual is not two monolinguals in one person. *Brain and Language,* 36, 3-15.
Haegeman, L. M. V. (1994[2]): *Introduction to Government and Binding Theory*. Blackwell Publishers: Oxford.
Hamers, J. F. – Blanc, M. H. A. (2000): *Bilinguality and Bilingualism.* Cambridge University Press: Cambridge.
Hogg, R. M. (ed.) (1992): *The Cambridge History of the English Language. Vol. I. (The beginnings to 1066).* Cambridge University Press: Cambridge.
Holmes, J.S. (1972): The Name and Nature of Translation Studies. In: Venuti, L. (2004): *Translation Studies Reader (2nd Ed)*. Routledge: London and New York
Horváth J. (2001): *Advanced Writing in English as a Foreign Language: A Corpus-Based Study of Processes and Products*. Lingua Franca Csoport: Pécs.
Hudson, R.A. (1980): *Sociolinguistics*. Cambridge University Press: Cambridge.
Hughes, A. (1989): *Testing for Language Teachers.* Cambridge University Press: Cambridge.
Jackson, H. (1988): *Words and their Meaning.* Longman: London and New York.
Jakobson, R. (1959: *On Linguistic Aspects of Translation.* In: Brower. R.A. (ed.), *On Translation*, (pp. 232-239). Harvard University Press: Cambridge, Mass.
Johns, T. (2000): *Tim Johns EAP Page*. http://www.eisu.bham.ac.uk/johnstf/timeap3.htm
Katamba, F. (1993): *Morphology*. MacMillan Press Limited: London.
Kennedy, G. (1998): *An Introduction to Corpus Linguistics*. Longman: London.
Klaudy, K. (2003): *Languages in Translation.* Scholastica: Budapest
Kontra, M. – Váradi, T. (1992): Sociolinguistics in Budapest. In: *Studies in Spoken Languages: English, German and Finno-Ugric.* Linguistics Institute, Hungarian Academy of Sciences: Budapest.
Krashen, S. – Terrell T. D. (1983): *The Natural Approach*. Pergamon Press: Oxford.
Krashen, S. (1981): *Second Language Acquisition and Second Language Learning*: Pergamon Press: Oxford.
Krashen, S. D. (1985): *The Input Hypothesis: Issues and Implications*. Longman: London.
Kretzschmar, W. Jr. (director) (no date): *Linguistic Atlas of the Middle and South Atlantic States* (LAMSAS) website (http://us.english.uga.edu/lamsas/)
Kretzschmar, W. A. Jr. – Schneider, E.W. (1996): *Atlas by the Numbers. An Introduction to Quantitative Analysis of Linguistic Survey Data*. Manuscript.

Labov, W. (1966/1982): *The Social Stratification of English in New York City*. Center for Applied Linguistics: Washington.
Labov, W. (1972): *Sociolinguistic Patterns*. University of Pennsylvania Press: Philadelphia.
Landau, S. L. (2001): *Dictionaries. The Art and Craft of Lexicography.* 2nd ed. Cambridge University Press: Cambridge.
Language Learning and Technology. (2001): Special Issue on Using Corpora in Language Teaching and Learning. Volume 5, Number 3. http://llt.msu.edu/vol5num3/default.html
Larsen-Freeman, D. – Long, M. H. (1991): *An Introduction to Second Language Acquisition Research.* Longman: Harlow.
Lass, Roger (ed.) (1999): *The Cambridge History of the English Language. Vol. III. (1476-1776).* Cambridge University Press: Cambridge.
Leech, G. (1983): *Principles of Pragmatics.* Longman: London.
Leech, G. (1997): Teaching and Language Corpora: A Convergence. In: Wichmann, A. – Fligelstone, A. –McEnery, T. – Knowles, G. (eds.) *Teaching and Language Corpora* (pp. 1-23). Applied Linguistics and Language Study. Longman: London.
Levelt, W. J. M. (1989): *Speaking: From Intention to Articulation.* MIT Press: Cambridge, Mass.
Levinson, S. C. (1983): *Pragmatics.* Cambridge University Press: Cambridge.
Lightbown, P. – Spada, N. (1993): *How Languages are Learned.* Oxford University Press: Oxford.
McCrum, R. – Cran, W. – MacNeil, R. 2002 (3rd edition): *The Story of English.* London, New York: Penguin.
McDonough, S. (1981): *Psychology in Foreign Language Teaching.* Allen & Unwin: London.
McLaughlin, B. (1987): *Theories of Second Language Learning.* Edward Arnold: London.
McMahon, A. M. S. (1994): *Understanding Language Change.* Cambridge University Press: Cambridge.
McNamara, T. F. (2000): *Language Testing.* Oxford University Press: Oxford.
Mey, J. L. (1993): *Pragmatics. An Introduction.* Blackwell Publishers: Oxford.
Mey, J. L. (1999): *When Voices Clash. A Study in Literary Pragmatics.* Mouton de Gruyter: Berlin.
Meyer, C. F. (2002): *English Corpus Linguistics: An Introduction.* Cambridge University Press: Cambridge.
Mitchell, B. (1971[2]): *A Guide to Old English.* Blackwell Publishers: Oxford.
Myers-Scotton, C. and Jake, J. (2001): Explaining aspects of code switching and their implications. In Nicol (ed.) *One Mind, Two Languages: Bilingual Language Processing.* Blackwell: Oxford.
Nida, E.A. – Taber, R. (1969): *The Theory and Practice of Translation.* Brill: Leiden.
O'Grady, W. – Dobrovolsky, M. – Katamba, F. (1996): *Contemporary Linguistics.* 3rd ed. Ch 7. Longman: London & New York.
O'Halloran, K. (2003): *Critical Discourse Analysis and Language Cognition.* Edinburgh University Press: Edinburgh.
Paradis, M. and Libben, G. (1987): *The Assessment of Bilingual Aphasia.* EMS: Bologna.

Peal, E. – Lambert, W.E. (1962): The relationship of bilingualism to intelligence. *Psychological Monographs* 76, 27, 1-23.
Pinker, S. (1994): *The Language Instinct.* Harper Collins: New York.
Plag, I. (2003): *Word-Formation in English.* Cambridge University Press: Cambridge.
Pyles, T. – Algeo, J. (1964): *The Origins and Development of the English Language.* Harcourt Brace Jovanovich: New York.
Quirk, R. – Greenbaum, S. (1991): *A Student's Grammar of the English Language.* Longman: New York
Radford, A. (1997): *Syntactic Theory and the Structure of English: a Minimalist Approach.* Cambridge University Press: Cambridge.
Radford, A. – Atkinson, M. – Britain, D. – Clahsen, H. – Spencer, A. (1999): *Linguistics.* Cambridge University Press: Cambridge.
Richards, J. C. – Rodgers, T. S. (2001): *Approaches and Methods in Language Teaching.* Cambridge University Press: Cambridge.
Romaine, S. (2000): *Bilingualism.* Blackwell Publishers: Oxford.
Saville-Troike, M. (1989) (2nd ed.): *The Ethnography of Communication.* Blackwell Publishers: Oxford.
Sax, O. (1970/1985): *The Man Who Mistook his Wife for a Hat: And Other Clinical Tales.* Touchstone: New York.
Scovel, T. (1998): *Psycholinguistics.* Oxford University Press: Oxford.
Second Language Acquisition Topics. http://homepage.ntlworld.com/vivian.c/SLA/
Sinclair, J. (1991): *Corpus, Concordance, Collocation.* Oxford University Press: Oxford.
Skinner, B. F. (1957): *Verbal Behavior.* Appleton-Century-Crofts, Inc: New York
Skutnabb-Kangas, T. (1984): *Bilingualism or Not?* Multilingual Matters: Clevedon, England.
Skutnabb-Kangas, T. (2000): *Linguistic Genocide in Education or Worldwide Diversity and Human Rights?* Mahwah, NJ: Lawrence Erlbaum Associates.
Spencer, A. – Zwicky, A. M. (2002): *The Handbook of Morphology.* Blackwell Publishers: London.
Stern, H. H. (1983): *Fundamental Concepts of Language Teaching.* Oxford University Press: Oxford.
Strang, B. M. H. (1970): *A History of English.* Methuen: London.
Stubbs, M. (1995): *Discourse Analysis. The Sociolinguistic Analysis of Natural Language.* Blackwell Publishers: Oxford.
Stubbs, M. (1996): *Text and Corpus Analysis. Computer-assisted Studies of Language and Culture.* Blackwell Publishers: Oxford.
Szirmai M. (2001): *The Theory and Practice of Corpus Linguistics.* Kossuth Egyetemi Kiadó: Debrecen.
Trudgill, P. (1995): *Sociolinguistics.* Penguin: London.
van Dijk, T. A. (1977): *Text and Context. Explorations in the Semantics and Pragmatics of Discourse.* Longman: London.
Venuti, L. (2004): *Translation Studies Reader (2nd Ed).* Routledge: London and New York
Wardhaugh, R. (2002) (4th ed.): *An Introduction to Sociolinguistics.* Blackwell Publishers: Oxford/Cambridge.

Wei, Li (ed.) (2000): *The Bilingualism Reader*. Routledge: London; New York.
Weinreich, U. (1953): *Languages in Contact*. Mouton: The Hague.
Weir, C. J. (1990): *Communicative Language Testing*. Prentice Hall: London.
Wichmann, A. – Fligelstone, A. – McEnery, T. – Knowles G. (eds.) (1997): *Teaching and Language Corpora*. Longman: London.
Wikipedia: en.wikipedia.org/wiki/English_English
Yamada, J. (1988): The Independence of Language: Evidence from a Retarded Hyperlinguistic Individual. In: Hyman, L. M. – Li, C. N. (eds.) *Language, Speech and Mind. Studies in Honour of Victoria A. Fromkin*. Routledge: New York.

Afterword

So we come to the end of the book ***but not the last step****. I hope you enjoyed the book and found parts of it fascinating. Most important of all, it should be clear that these were really just the first steps and that there are many paths you can choose to walk down. We tried to give you an idea of what linguistics is and what linguists do.*

This is perhaps the right place to mention that other linguists reading this book probably want to add, "But what about… ?" They are quite justified, since there is so much that we could not say. In fact, there are whole sub-disciplines of linguistics that we could only mention or had to leave out completely. As you may have noticed, the previous section is entitled ***"Further steps"****. It is a bibliography, full of books that you can consult to learn more. But even this bibliography is only a selection. The more you know, the more you will realise how much there is still to learn. I trust that you think this is a cause for optimism! Happy reading!* ***Enjoy the journey!***

Borbála Richter, editor

Glossary and Index

Headword: explanation (page number/s)

accent: a typical pattern of pronunciation, which informs the listener where the speaker comes from. A dialect may be fastest recognized from its accent. (101)

achievement test: based on a syllabus or a textbook and intends to test how successfully students managed to learn the material covered in that textbook. (177)

acquisition: a subconscious process of picking up language through exposure (Stephen Krashen's interpretation) (cf. learning). (162)

acronym: a way of word formation where the first letters (or initials) of the words are combined (like *UK* for *United Kingdom*) (42, 123)

addition: an error type in learners' language; there is an extra item in language learners' utterance which should not be there e.g. **We didn't went there*. (163)

addressee: the person or persons to whom something is spoken or written. (137)

adequacy: an adequate translation realizes in the target language the textual relationships of a source text without breaching its own linguistic system. Closely linked to the concept of acceptability which is determined by the norms of the target culture. (133)

agglutinating: a language where one word can be made up of many morphemes and each morpheme has a clearly identifiable function eg *un* + *friend* + *ly*. See also fusional language. (43)

agreement (also: concord): a situation where a certain word form requires a corresponding form of another word (e.g., subject-verb ~). (49)

ambiguity: multiple meaning. (69, 74)

analogy: a change in which regular forms influence less regular forms, increasing regularity in grammar (e.g. *book* in OE had the forms *bōc - bēc*, by analogy it became *book – books*); in child language: young children often replace irregular forms with forms conforming to the regular pattern (*mans*). (90)

Anglo-Saxon: another name for Old English and its age, culture, etc. (based on the names of two of the tribes that settled in Britain). (91)

antonymy: gradable oppositeness e.g. *big–small* (= antonyms). (68)

aphasia: speech disturbance due to injury or other kind of damage in the brain. (153)

applied linguistics: a discipline which attempts to solve real-world language-based problems. It utilizes knowledge from a wide variety of fields: linguistics, anthropology, language pedagogy, sociology, psychology, education, etc. (161)

aptitude: an innate talent for second language acquisition, which indicates the rate of progress, but it does not tell whether somebody can or cannot learn a foreign language. (167)

arbitrariness: the feature of language that normally there is no connection between form and meaning. (17, 66, 151)

backward reference: an instance of an expression referring to another. In general, an anaphoric expression is represented by a pro-form or some kind of deictic. It occurs when an expression co-refers with a prior expression. (143)

Bank of English: the largest corpus of authentic English, containing spoken and written texts. (113)

basic skills: reading, writing, listening and speaking. (172)
bilingual dictionary: a dictionary involving two languages. (121, 133)
binary (~ system, ~ features, etc.): characterising a system, etc. with two components. In linguistics, ~ features are marked with + or – (e.g., the consonant /b/ is +voiced, while /p/ is –voiced). (49)
blending: a way of word formation where parts of two words make up the new word (like *Bollywood* from *Bombay* and *Hollywood*). (42)
bound morpheme: a morpheme that cannot be used as a word on its own. (37, 44)
Broca's area: the area of the brain above and in front of the left ear. It is involved in organizing the movements of the vocal organs during speech. (153)
Brown Corpus: the first computer corpus of written English. (112)
calque: loan translation; a type of borrowing: a word is translated morpheme by morpheme (*home/page – hon/lap*). (88)
capacity, of computers: the technical ability of the hard disk and memory that allows corpus linguists to save information reliably. (114)
clause: a syntactic unit between the level of phrases and the level of sentences. It typically contains a subject and a verb and may contain other constituents. A sentence may comprise one or more ~s. ~s may be classified according to their forms and functions. (49)
clipping: a way of word formation where a part of the word is used (like *exam* for *examination*). (42)
closed classes: word classes which do not typically take suffixes and there are no new members in them. (41, 44)
cloze-test: there are several types of cloze-tests: in its well-known form every n^{th} word is deleted from a text and students have to find out what words are missing from the text. (175)
code-switching: is using two languages simultaneously or interchangeably during communication, as part of a single interactional whole / speech exchange. It can be a single word, a phrase, or one or more sentences, and can be intrasentential (within a sentence) or intersentential (between sentences). (186)
cognate words: words which have a common origin historically (e.g., English *father* and German *vater*). (86)
cognitive meaning (also: descriptive meaning, referential meaning): the basic, factual meaning of a sentence that determines its truth value. (68)
coherence: refers to the main principle of organisation accounting for the underlying functional connectedness or identity of a piece of spoken or written language. It involves such factors as the language users' knowledge of the world, the inferences they make and the assumptions they hold. It refers to the ideas in a text that make a text logical and understandable. (143, 175)
cohesion: refers to the grammatical and lexical elements that link the sentences of a text together. (143, 175)
collocate: a word that appears near another word. (115)
communicative competence: knowing the rules of a language and being able to use that language appropriately. Mainly consists of grammatical competence, discourse competence, sociolinguistic competence and strategic competence. (175)

comparative linguistics: the branch of linguistics which focuses on the characteristics of different languages with a common historical origin. (86)
competence: the native speaker's unconscious knowledge of the system of rules in his or her language (cf. performance). (50, 174)
complementarity: a kind of oppositeness seen, e.g. in the pair *open–closed* (= complementaries). (69)
compositional: the meaning of an expression is compositional if it cannot be reduced to those of the parts. (66)
compound bilinguals: have one semantic system and two language codes. (186)
compounding: a way of word formation where two words are combined to make one. (42, 65)
concordance: a list of examples from a corpus. (112)
constative: an utterance which is a descriptive statement and can be analysed in terms of truth values. (76)
context: refers to specific parts of an utterance or text near a unit which is the focus of attention. The meaning of a unit is determined by its context, which is specified in terms of the unit's relations. (115, 142)
contraction: two words contracted into one, so one loses its vowel, i.e. its pronounceability. (64)
controlled vocabulary: a limited/simple vocabulary used, e.g. in the definitions of learner's dictionaries; same as graded vocabulary. (121)
convergence: language change in which dialects or languages become more similar to each other (the opposite process is called divergence). (89)
conversion: a way of word formation where a word is used. (42)
co-ordinate bilinguals: have two semantic systems and two language codes. (186)
coordination: the relationship of syntactic units of equal status (e.g., two phrases or two clauses) linked within a larger syntactic unit (cf. subordination). Coordination may be marked by coordinators, such as *and*. (58)
copyright: protection of an article, photograph, or other creative work. (114)
corpus building: designing and collecting the corpus for lexicographic, or any other, purposes. (125)
corpus: a collection of texts for linguistic analysis. (112)
  first-generation ~: a corpus collected before the 90s. (114)
  monitor ~: a corpus that keeps changing with new texts added to it reguarly. (113, 114)
  second-generation ~: a large corpus collected since the 90s. (114)
  stable ~: a corpus that never changes as texts are not added to it. (113)
co-text: in a concordance, the words before and after the keyword. (115)
covert error: an error type that can only be detected in the context e.g. using the past simple tense instead of the present perfect when talking about an activity which has not been finished yet. (163)
creative: to use language creatively is to make up infinite, possibly novel forms or meanings out of a finite set of means. *Colorless green ideas sleep furiously* (Chomsky). (151)
Critical Period Hypothesis (CPH) (Lenneberg, 1967): this hypothesis suggests that there is a biologically determined period of time within which successful language acquisition can take place. After the critical period closes down (around 10 years of age), the acquisition of language will be problematic. (155, 167)

cross-cultural pragmatics: aims at understanding the extent to which non-shared (cultural) knowledge affects and modifies the retrieval of intended meaning. (79)
data: any information collected for linguistic analysis. (112)
decoding dictionary (same as a comprehension dictionary): opposed to an encoding dictionary, used e.g. for reading. (123)
deep structure: (a notion used in generative grammar:) the underlying (or abstract) syntactic representation of a sentence, containing the constituents of the sentence and the relationship between the constituents (cf. surface structure). (51)
description: describing what actually is used in language; it is opposed to prescription. (62)
descriptive: a rule or a grammar is descriptive when it describes language facts. cf. prescriptive. (107)
developmental patterns: refer to the order the different grammatical elements are acquired and also the stages learners go through while acquiring one element. (162)
diachronic (historical) linguistics: the branch of linguistics focusing on the development of languages over time. (84)
diagnostic test: intends to diagnose candidates' strengths and weaknesses. (177)
dialect: a regional / geographical variety of language. (85, 100)
discourse: refers to a continuous stretch of language larger than a sentence. It is a set of utterances which constitute a speech event. (130)
discourse analysis: the study of discourse, its genres, structures; the importance of discourse for understanding the meanings of its parts – including words in context. (61, 128)
discrete-point test: measures how well students know separate elements of a language always on sentence-level, one at a time. (177)
displacement: remoteness of the sign (in the language) from its referent (in the real world). displacement in language means humans can talk about things not present at the time and place of talking. (151)
dynamic equivalence: is receptor oriented and the focus is on 'equivalent effect'. It can include adaptations in grammar, lexicon and cultural information, in order to achieve natural language use. Correspondence in meaning is considered more important than correspondence in style. (135)
elite bilingualism: a conscious decision in favour of bilingualism by parents from stable, middle-class backgrounds who are in a position to support the education process with back-up involvement. (186)
empirical: the kind of research that focuses on observation and experiments. (112)
encoding dictionary (same as a production dictionary): opposed to a decoding dictionary, and used e.g. for essay writing. (123)
encyclop(a)edia: a reference work about *things* – it is opposed to a dictionary, a reference work about *words*. (121)
entailment: a sense relation whereby the truth of a sentence guarantees the truth of another sentence. (67)
error analysis: the analysis of students' errors. (162)
error: occurs because there is a gap in learners' knowledge, which means they cannot express something in the foreign language as they have not learnt that before. (162)

ethnographic approach: a methodology used when we are interested in how small communities of people behave (talk) in various everyday, natural situations, such as at home, in the pub, while shopping or in the classroom. The strengths of ethnography include reliable results. The results of ethnographic research will not be statistically processed. (99)

filter test: intends to filter out candidates whose level of proficiency is below a certain level. (177)

foreign language: learnt in an environment where the native speakers do not speak that language as their mother tongue. (162)

foreigner talk: a simplified language that native speakers use when they talk to non-native speakers. (166)

formal equivalence: refers to equivalence at the level of form, aiming at matching the message in the receptor language as closely as possible with the different elements in the source language. (135)

forward reference: occurs when an expression co-refers with a latter expression in the discourse. (143)

free morpheme: a morpheme that can also be a word. (37, 44)

frequency: the number of times a word or phrase appears in a corpus. (115)

fusional: a language that compresses a lot of information into a morphological form. Eg in *lát* + *lak* the function of *–lak* is rather complex. See also agglutinating language. (43)

gap-filling task: any language task in which missing information has to be written into a text. (175)

general English: English used for regular purposes, rather than only in a special job or science. (114)

generative theory (also: generative linguistics, generative grammar): a theoretical approach to language associated primarily with the name of Noam Chomsky (some of its basic ideas are *Universal Grammar, innateness hypothesis, transformation, deep structure* and *surface structure*). (91)

genetically related languages: languages deriving from a common source; from this viewpoint languages may be classified as belonging to language families (often represented by the family tree model), e.g. the Indo-European family. (85)

grammar: (1) the (knowledge of the) system of language in the native speaker's head; also, (2) a "man-made" description/account/model of this system. (18, 41, 49-50, 62, 65, 87, 91, 104-107)

grammatical meaning: the meaning of suffixes and prefixes. (36)

Great Vowel Shift: a series of vowel changes in early modern English (the systematic raising of long vowels). (93)

homograph: words that not only have the same shape as another but also the same spelling: *bear* 'support; hold' – *bear* 'big animal with claws'. (69)

homonymy: two words that have an identical shape, e.g. past and passed are homonymous. (69)

hyponym: when the sense of word D is included in that of word A, then word A is a hyponym of word D. thus *Alsatian* is a hyponym of *dog*, because the meaning of *dog* is included in that of *Alsatian*: an Alsatian is a large wolf-like *dog*. (63, 67)

idiolect: an individual variety of language. (103, 108)

idiom: an expression whose meaning is not the total of the meanings of the parts, i.e. not compositional. (66)
inflectional (= fusional): a language is inflectional if the root changes when a suffix is added. (42, 89)
innate: if a quality or ability is innate, it is present at birth. (50, 152)
input: the language that learners are exposed to, which can be different from the "real" target language. Stephen Krashen states that *comprehensible input*, which is just above students' current abilities, is a necessary requirement for successful language acquisition. (166)
integrative test: tests two or more (or a number of) skills at the same time, for instance reading and writing in the same task. (177)
interaction: takes place when the target language is used between speakers, for example, in conversations. (166)
interlanguage: a language that has been developed by the learners of a second language who have not fully acquired it, but only approximated it, preserving some features of their first language in speaking or writing the target language, and creating innovations. (163)
inter-rater reliability: at least two raters evaluating the same test. If they give similar scores the inter-rater reliability index is high; if they give different scores, it is low. (176)
intra-rater reliability: one rater evaluating a test at least two times. It is high if the rater gives a certain score to an essay on a certain day, and some days later s/he gives the same score to the same essay. (176)
irregular: a form that does not follow the rule. it is unpredictable. (35-39, 90)
isolating (= analytic): a language that separates morphemes into words. (42)
key word: the central word selected for concordance analysis. (114)
Kibbitzer: Tim Johns's special service to help students use concordances for language study. (116)
L1: people's mother tongue or first language. (166)
L2: an umbrella term for "foreign language", "second language" and "target language". (162)
language change: change in a language over a period of time; it involves all aspects of language (sounds, morphemes, words, meanings, etc.); it is a universal process, since all living languages are subject to change. (84)
language contact: in this situation, members of different linguistic communities regularly interact with each other, which leads to a certain degree of bilingualism, with the result that the languages in the contact situation begin to influence each other (i.e. they borrow words or structures from each other). (87)
language death or **linguicide**: is the intentional or unintentional causing of the death of a language. (183)
lateralization: the process whereby the dominance of one brain hemisphere develops. (153)
learner strategies: approaches that learners use to try to learn an L2. There are memory strategies, cognitive strategies, metacognitive strategies, compensation strategies, affective strategies, and social strategies. (167)
learner's dictionary: a kind of monolingual dictionary specially designed for learners, with many features of language instruction. (120)
learning: a conscious way of studying a language (Stephen Krashen's interpretation) (cf. acquisition). (162)

lexeme: all the forms of a word (nouns: singular and plural, adjectives: base, comparative and superlative, etc.). (39, 42, 44, 63-66)
lexical: relating to the vocabulary. (50)
lexical borrowing: the introduction of a word from one language or dialect into another. (87)
lexical change: change in the vocabulary of a language or dialect, e.g., when certain words fall out of use or when new words are formed or borrowed. (84)
lexical item: a word or any item in the lexicon that has to be listed separately; a listeme. (64)
lexicography: editing/writing of dictionaries; a lexicographer is an editor/writer of dictionaries. (99)
lexicon: a complex network of lexical items in the speaker's head; same as mental lexicon/dictionary. (47, 62, 84, 92, 121, 157-158)
linguistic variable: a ~ is a linguistic feature that can take on different values. e.g. in English, [r] can be pronounced (in various ways) or dropped. (104)
macrostructure: the structure of a dictionary above the level of entries; it is opposed to the *microstructure*. (119)
matrix language: is one of the two languages involved in code-switching. It has the more dominant role and provides the syntactic frames for the phrases. The "embedded Language" provides the inserts to the matrix language. (186)
maxim of manner: states that the contribution should be perspicuous (expressed clearly) – it should be orderly and brief, avoiding obscurity and ambiguity. (75)
maxim of quality: states that speakers' contributions ought to be true, that is, they should not say what they believe to be false, nor should they say anything for which they lack adequate evidence. (76)
maxim of quantity: states that the contribution should be as informative as required and should not be unnecessarily informative. (76)
maxim of relation: states that contributions should be relevant to the purpose of the exchange. (76)
mediation: translating or summarizing the content of a text from one language into another. Written translation is a typical example for this. (172)
microstructure: the structure of a dictionary within the entries, i.e. below entry level; it is opposed to *macrostructure*. (119)
Middle English: the form of English spoken in the period between 1100 and 1500. (92)
misinformation: an error type in learners' language; a grammatical form is used incorrectly e.g. **The dog ated the chicken.* (163)
misordering: an error type in learners' language; the word order is wrong e.g. **What daddy is doing?* (163)
mistake: a lapse in performance, which means that the learner has already learnt that particular structure but s/he does not use that. (162)
Modern English: English as spoken after 1500. The period until about 1800 is called early modern English. (93)
monolingual dictionary: a dictionary involving just one language, with no different source and target language; it is opposed to a translation dictionary. (121)
monolingual: a dictionary or corpus that contains examples form one language. (114)
morpheme: the smallest meaningful part of language. (18, 37, 42, 43, 44, 48, 62, 65)

morphological change: change in the morphemes, i.e. stems and affixes, of a language or dialect. (84)

motivation: includes factors which make people learn a second language; it is a dynamic factor, which can change all the time. (167)

movement: the moving of sentence constituents from one syntactic position to another during the derivation of sentence structures, e.g., the sentence - *what is he doing?* - has undergone the so-called *wh*-movement: the question word ("*wh*-word"), being an object, originates in post-verbal position (*he is doing what?*), but, as a question word, it must move to the beginning of the sentence (in accordance with the rules of grammar); therefore, the *wh*-~ is one step in the production of a well-formed English *wh*-question. (51)

multilingual dictionary: a dictionary involving several languages; same as polyglot dictionary. (121)

multi-word verb: a longer-than-word item consisting of a verb plus one/two other words; they can be phrasal verbs, prepositional verbs, or phrasal-prepositional verbs. (65)

naturally occurring language: spoken or written language people use for communication. (112)

negative transfer: occurs when learners' L1 is the cause of their L2 errors. (166)

Neogrammarians: a group of 19th century linguists who focused on language change; their most important hypothesis is that sound laws (regular sound changes) operate blindly and without inexplicable exceptions; they also introduced the concept of analogy. (90)

Old English: the form of English spoken until about the end of the 11th century, also known as Anglo-Saxon. (91)

omission: an error type in learners' language; an item is missing from the utterance e.g. **My name Alberto.* (163)

open classes: word classes that take suffixes and have new members. (41)

order of acquisition: a definite order in which the different grammatical structures are acquired. (164)

overt error: one which can be easily detected outside the context in which it occurs e.g. **He sad*, from which expression an "is" is missing. (163)

performance: the way we actually use our competence in real life situations when we have to speak, listen, read or write in a language; the actual use of competence in the comprehension and production of language (cf. competence). (50, 112, 174)

performative: refers to a type of sentence where an action is performed by virtue of the sentence having been uttered. (76)

phoneme: idealized speech sound (consonant or vowel), ignoring small differences between individual speakers. (17, 18, 26, 27, 33)

phonological change: change in the phonemes or stress patterns of languages or dialects (e.g., the pronunciation of individual phonemes may be altered, or the whole system may change by the loss and/or introduction of phonemes). (50, 84, 88, 89, 158)

phrasal verb: a multi-word verb consisting of a verb and a particle. (65)

phrase: a syntactic unit, which typically consists of more than one word, below the clause level. ~s are the immediate constituents of clauses. (49)

pilot phase: the trialling period of a test, when a group of students complete a test so that testers can see which items are good enough to remain part of that test, and which are the ones that have to be changed or deleted. (178)

placement test: measures proficiency as related to given levels/groups. (177)
plural: the specification for grammatical number meaning 'more than one', usually signalled by a plural marker (e.g., the plural morpheme attached to nouns, *-s* in English). (cf. singular). (35-41, 48, 89, 90, 92)
polysemy: a kind of ambiguity where a word has more than one meaning, i.e. is polysemous, e.g. *pass* means 1. 'move/go by' 2. 'gain a required mark' 3. 'give a required mark' (etc). (69)
population: the conclusions of a study are often intended to be valid for a large number of people (items), that is, the population. (99)
positive transfer: when students' L1 helps them acquire a second language. (166)
poverty of stimulus: the logical problem of language acquisition that there is no sufficient language data for a child to learn his/her native language but they still do (cf. Universal Grammar). (156)
pragmatics: the study of how language is used in social situations. (61)
prefix: the bound morpheme before the stem of the word. (37, 42, 66)
prescription: stating with authority what correct usage is; it is opposed to description. (62)
prescriptive: a rule or a grammar is prescriptive when it prescribes what to do, that is, when it tells speakers how (not) to speak (cf. descriptive). (107)
principle of cooperation: states that speakers try to cooperate with each other when communicating: they attempt to be informative, truthful, relevant and clear. Listeners assume that a speaker is following these criteria. Speakers may break this principle, but conversation proceeds on the assumption that they do not. (75)
proficiency test: not related to any syllabus as it intends to test students' level of proficiency. (177)
progress test: intends to measure progress during a course. (177)
propositional content: refers to the unit of meaning which constitutes the subject-matter of a statement, question or order. (145)
proto-language: the common ancestor of the languages of a language family, e.g., Proto-Indo-European; a ~ is usually known as a result of linguistic reconstruction (in the absence of surviving contemporary documents), i.e. the attested forms of the daughter languages are compared and their probable common origins are established. (85)
rater: examiner who evaluates language learners' performance. (175)
register: refers to a variety of language defined according to its use in social situations. (74)
reliability: the extent to which test scores are consistent. (176)
repetition: refers to the recurrence of lexical items to indicate the repetition, extent or confirmation of a phenomenon. (143)
representativeness: an important feature of a corpus – corpus linguists should have enough examples to be able to describe a language with it. (112)
root: what is left of the word if all suffixes and prefixes are removed. (37)
sample: a sample is used in research where the data will be processed using statistical methods. A sample is carefully selected so that it represents the population. It consists of a small number of people (items), ideally not fewer than 30. (98)
second language: learnt in an environment where the native speakers' mother tongue (L1) is this language. (162)
semantic change: change in the meanings of linguistic units. (84)

semanticity: if a sign is arbitrarily assigned to something, the sign will "mean" that thing. That is semanticity. (151)

semantics: the study of (various kinds of) linguistic meaning. (61)

sentence: the largest unit of grammar, containing one or more clauses. A ~ comprising one clause is simple, other ~s are compound (with coordinated clauses) or complex (when one or more of its clauses are subordinated to a main clause). (19, 32, 48, 56, 58-60, 147)

sequence of acquisition: refers to what phases learners go through while acquiring one language element. (164)

setting: refers to the place and time where and when the communicative event is situated. (143)

singular: the specification for grammatical number meaning 'one' (vs. 'more than one'). It is usually the (morphologically) unmarked number specification (cf. English *book* – singular – and *book**s*** – plural; though in the case of the third person forms of verbs the singular is marked in English, while the plural is unmarked: *writes* - *write*). (48, 92)

sociolect: a social variety of language. (99)

sound change: a change affecting the phonological system of a language. (90)

source language: (1) the first language in a bilingual dictionary, from which translations are provided into the target language. (2) The language from which something is translated; sometimes called the donor language. (134)

special-field dictionary: a dictionary that covers a specific subject; opposes special-purpose dictionary. (122)

special-purpose dictionary: a dictionary that covers a specific type of linguistic unit; it is opposed to special-field dictionary. (122)

speech act theory: refers to a theory which analyses the role of utterances in relation to the behaviour of speaker and hearer in interpersonal communication. A speech act is a communicative activity, defined with reference to the intentions of speakers while speaking and the effects they achieve on listeners. (77)

standard: a standard is a language variety. It is a socially distinguished, prestigious one usually taught at school and to non-native speakers, spoken by educated people, and used in print and news broadcasts. Historically, it developed from a dialect, which became associated with high culture and education. Linguistically, the standard is equal to other varieties. (106)

structural borrowing: the introduction of structural units (e.g., phonemes or morphemes) or, more generally, structure types (e.g., the use of prepositions) from one language or dialect into another. (89)

Structuralist school: (also: structuralism, structuralist linguistics) a theoretical approach to language dating from the early 20th century and based on the ideas of Ferdinand de Saussurre. The ~ views language as a system of signs. (90)

structure dependence: sensitivity for structure in language. (151)

subordination: the relationship in which syntactic units of different status are linked (cf. coordination). A subordinate clause is a constituent of the main clause. (The markers of ~ are subordinators, e.g., *that* or *when*). (58)

successive, or **consecutive, bilingualism**: learning one language after already knowing another. It differs from simultaneous bilingualism where two languages are learnt as "first languages" or mother tongues. (186)

suffix: the bound morpheme after the stem of the word. (37, 42)

superordinate: when the sense of word A includes the sense of word D, then word D is a superordinate of word A. *Dog* is a superordinate of *Alsatian*, because the sense of *Alsatian* includes that of *dog*: an *Alsatian* is a large wolf-like dog. (67)

suppletivism: when the derived form of a word is not root+suffix, but a one-morpheme substitute (*go* → *went*). (37)

surface structure: (in generative grammar) the final stage in the syntactic representation of a sentence, i.e. the one that is manifested in actual speech or writing (cf. deep structure). (51)

syllabus: an outline or summary of topics covered in a language course. (177)

synonymy: an infrequent relation between words (= synonyms) that have similar meanings and so can mutually replace each other. (68)

syntactic change: change in the syntactic patterns of language, i.e. in the way words and phrases form combinations with other words and phrases; or change in the linear ordering of syntactic constituents. (84)

target language: (1) the second language in a bilingual dictionary, into which translations are given from the source language; (2) the second/foreign language learners learn/acquire; L2; (3) The language into which something is translated; sometimes called the receptor language. (134, 161)

test item: a question in a test. It may be multiple choice, matching, true or false, etc. (178)

tester: a language testing expert who develops tests. (178)

thesaurus: a non-alphabetical, notionally/semantically/thematically arranged reference book, usually without definitions. (123)

token: in counting words in a text, each copy of a word is a token of that word (e.g., there may be 18 or just 8 tokens of *you* on any page) (63)

transfer error: error that is the result of language learners' attempt to use their mother tongue knowledge. (163)

transformation: a linguistic operation which links the different levels of syntactic representation. (51)

translation dictionary: a dictionary involving more than one language; opposed to monolingual dictionary. (121)

truth value: certain types of sentences can be either true or false; this is their truth value. (68)

turn: conversation is seen as a sequence of conversational turns, in which the contribution of each participant is seen as part of a coordinated and rule-governed behavioural interaction. (143)

turn-taking: refers to the alternation of turns (speech contributions) in a conversation. It is governed by the main rule that only one person should talk at a time. How children learn the conventions governing turn-taking is an issue which has attracted interest in language acquisition. (143)

type: if you count each word only once in a text, that will be the type (i.e., however long the page, there can only be one type for *you*) (63)

unabridged: not shortened/abridged from a larger dictionary; the largest kind available. (122)

Universal Grammar: (1) a formal grammar whose aim is to define the universal characteristics of human language; (2) the grammar innate in every human being, which enables young children to construct the specific grammars of their native languages on the basis of the often imperfect language input they receive from their environment (cf. poverty of stimulus). (50)

utterance: a natural unit of speech bounded by breaths or pauses. It can be one word, some words, one sentence or several sentences. (147)
validity: the extent to which a test measures what it is intended to measure. (176)
variability: refers to the variable nature of learner language. a learner may use grammatically perfect language in one context but may also use ungrammatical forms in another. (162)
variable: (1) adj. in general, anything that is ~ appears in different forms depending on the circumstances. A language is variable when it has varieties or variants; (2) n. see *linguistic variable*. (98)
vocabulary: (1) a person's individual lexicon; (2) same as *word stock*, the collective lexicon of a language. (62)
washback effect of tests: the effect of tests on teaching and learning; positive if the aims of the course and the testing are the same, but negative if they are different. (178)
Wernicke's area: the area of the brain around and under the left ear. It is involved in speech perception. (153)
word class: lexical categories like nouns, verbs, auxiliaries, prepositions, etc. (39-41, 44, 52)
word stock (same as *vocabulary*): the collective lexicon of a language. (62)

Kiadta a Bölcsész Konzorcium

A Konzorcium tagjai:
Eötvös Lóránd Tudományegyetem
Pécsi Tudományegyetem
Szegedi Tudományegyetem
Debreceni Egyetem
Pázmány Péter Katolikus Egyetem
Berzsenyi Dániel Főiskola
Eszterházy Károly Főiskola
Károly Gáspár Református Egyetem
Miskolci Egyetem
Nyíregyházi Főiskola
Kodolányi János Főiskola
Szent István Egyetem

Szerkesztette: Richter Borbála
Nyelvi lektor: Keith Hardwick
Szakmai lektor: Hollóssy Béla
Szakmai konzulens: Nikolov Marianne
Rajzok: Szitó Judit
Műszaki szerkesztő: Nagy Károly

A kötet megjelenése az Európai Unió támogatásával,
a Nemzeti Fejlesztési Terv keretében valósult meg:
A felsőoktatás szerkezeti és tartalmi fejlesztése – HEFOP -3.3.1-P.-2004-09-0134/1.0

ISBN 963 9704 91 1

Nyomás: ALTO Nyomda, Székesfehérvár
Felelős vezetők: Molnár Tamás, Vass Tibor

Bölcsész Konzorcium HEFOP Iroda
H-1088 Budapest, Múzeum krt. 4/A.
tel.: (+36 1) 485-5200/5772 – dekanbtk@ludens.elte.hu

A projekt az Európai Unió támogatásával,
a Nemzeti Fejlesztési Terv keretében valósult meg:
A felsőoktatás szerkezeti és tartalmi fejlesztése
HEFOP -3.3.1-P.-2004-09-0134/1.0