

# From sound to meaning: hearing, speech and language



## About this free course

This free course is an adapted extract from the Open University course SDK228 *The science of the mind: investigating mental health*: [www3.open.ac.uk/study/undergraduate/course/a332.htm](http://www3.open.ac.uk/study/undergraduate/course/a332.htm).

This version of the content may include video, images and interactive content that may not be optimised for your device.

You can experience this free course as it was originally designed on OpenLearn, the home of free learning from The Open University -

[www.open.edu/openlearn/health-sports-psychology/health/health-studies/sound-meaning-hearing-speech-and-language/content-section-0](http://www.open.edu/openlearn/health-sports-psychology/health/health-studies/sound-meaning-hearing-speech-and-language/content-section-0)

There you'll also be able to track your progress via your activity record, which you can use to demonstrate your learning.

The Open University, Walton Hall, Milton Keynes, MK7 6AA

Copyright © 2016 The Open University

## Intellectual property

Unless otherwise stated, this resource is released under the terms of the Creative Commons Licence v4.0 [http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en\\_GB](http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en_GB). Within that The Open University interprets this licence in the following way:

[www.open.edu/openlearn/about-openlearn/frequently-asked-questions-on-openlearn](http://www.open.edu/openlearn/about-openlearn/frequently-asked-questions-on-openlearn). Copyright and rights falling outside the terms of the Creative Commons Licence are retained or controlled by The Open University. Please read the full text before using any of the content.

We believe the primary barrier to accessing high-quality educational experiences is cost, which is why we aim to publish as much free content as possible under an open licence. If it proves difficult to release content under our preferred Creative Commons licence (e.g. because we can't afford or gain the clearances or find suitable alternatives), we will still release the materials for free under a personal end-user licence.

This is because the learning experience will always be the same high quality offering and that should always be seen as positive – even if at times the licensing is different to Creative Commons.

When using the content you must attribute us (The Open University) (the OU) and any identified author in accordance with the terms of the Creative Commons Licence.

The Acknowledgements section is used to list, amongst other things, third party (Proprietary), licensed content which is not subject to Creative Commons licensing. Proprietary content must be used (retained) intact and in context to the content at all times.

The Acknowledgements section is also used to bring to your attention any other Special Restrictions which may apply to the content. For example there may be times when the Creative Commons Non-Commercial Sharealike licence does not apply to any of the content even if owned by us (The Open University). In these instances, unless stated otherwise, the content may be used for personal and non-commercial use.

We have also identified as Proprietary other material included in the content which is not subject to Creative Commons Licence. These are OU logos, trading names and may extend to certain photographic and video images and sound recordings and any other material as may be brought to your attention.

Unauthorised use of any of the content may constitute a breach of the terms and conditions and/or intellectual property laws.

We reserve the right to alter, amend or bring to an end any terms and conditions provided here without notice.

All rights falling outside the terms of the Creative Commons licence are retained or controlled by The Open University.

Head of Intellectual Property, The Open University

Designed and edited by The Open University

# Contents

Introduction	4
Learning Outcomes	5
1 Overview	6
2 The brain's task: the structure of language	7
2.1 Preliminaries	7
2.2 Generativity and duality of patterning	10
2.3 From ear to phoneme: the phonological problem	12
2.4 From phoneme to meaning: the semantic problem	21
2.5 From phoneme to sentence structure: the syntactic problem	22
2.6 Summary of Section 2	24
3 The brain's solution: the machinery of language	24
3.1 Speech perception	24
3.2 The anatomy of the language system	29
3.3 Specialisation within language areas: aphasia	31
3.4 Specialisation within language areas: brain scanning	37
3.5 Electrophysiological studies of language processing	38
3.6 Summary of Section 3	44
4 Conclusions	44
5 Questions and answers	46
Keep on learning	50
Acknowledgements	51

# Introduction

---

This course looks at how language is understood, which includes hearing and how sounds and words are interpreted by the brain. It takes an interdisciplinary approach and should be of wide general interest.

This OpenLearn course is an adapted extract from the Open University course :  
[SDK228 \*The science of the mind: investigating mental health\*](#).

# Learning Outcomes

---

After studying this course, you should be able to:

- recognise definitions and applications of key terms relating to mental health
- understand and apply basic grammatical terminology
- describe briefly the different types of sounds used in speech in both acoustic and articulatory terms
- outline the key features of human language as compared to the vocalisations of other species
- describe the complex psychological processes involved in decoding even simple sentences of spoken language.

# 1 Overview

---

As you walk down the street one day, you hear a voice from somewhere behind you that seems to be discussing this course. It says:

‘My dad’s tutor’s no joker, and he told me the TMA’s going to hit home with a bang.’

You turn to find the face behind the voice, which is a gravelly Glaswegian baritone, but the man has gone, leaving you to ponder what he has said. Let us call his sentence example (1). We will come back to it throughout the course.

At the same moment, in Amboseli National Park in Kenya, a group of vervet monkeys (Figure 1) is foraging on the ground near a large baobab tree. A young male on the periphery of the group suddenly stands on his hind legs and gives a loud triple barking sound. The other monkeys have no doubt what this means: a snake is in the vicinity. The monkeys group together, scouring the grass for the location of the predator.



Figure 1 Vervet monkeys (*Cercopithecus aethiops*)

Copyright © Tony Camacho/Science Photo Library

In both these scenarios, a primate brain performed one of its most remarkable tricks. It took a pattern of vibration off the air and turned it into a very specific set of meanings. How the brain achieves this trick is the subject of this course. The monkey example is



interesting because it has been seen as a model of the early stages of human language. Vervet monkeys have several distinct alarm calls – *snake*, *leopard* and *eagle* are the best studied – each one of which can be said very definitely to have a meaning. We know this because on hearing an eagle call played back by a researcher's tape recorder, the monkeys scan the skies, whilst when a leopard call is played, they climb a tree (Seyfarth et al., 1980). Thus quite different associations are being evoked in the vervet brain.

But in this course, the vervet monkey example will mainly be used as an illustration of how different human language is from the communication system of any other primate. The computational task for the human brain in understanding a single sentence is vastly more complex than the vervet case, at every level. It is so complex that a whole field of research – linguistics – is devoted to investigating what goes on, and it requires a whole set of brain machinery that we are only beginning to identify. Hearing and understanding a sentence, or the reverse, where a thought in the brain is turned into a string of buzzes, clicks and notes we call speech, is the crowning achievement of human evolution and the defining feature of human mental life. No other species that we know about comes even close, as the examples below will illustrate.

### Activity 1

Write down the main ways in which you think human language and the call system of the vervet monkey might differ. Keep this list with you and compare it with the differences mentioned in the text as you go through the first half of this course.

This course is in two parts. In the first half (Section 2), we will investigate the nature and structure of human language, keeping our vervet monkey example on hand at all times as a comparison. The aim of this first half is to come to a precise understanding of just what the task is that the brain has to perform in processing language. This paves the way for the second half of the course (Section 3), where we consider different lines of evidence on how and where in the brain this feat is achieved.

## 2 The brain's task: the structure of language

### 2.1 Preliminaries

To talk about how human language works, we need to establish the meaning of some key terms. The study of language and languages is called linguistics, and linguistics relates closely to biological psychology, as we shall see. Linguists talk about the **grammar** of a language. By this they don't mean a set of rules about how people should speak. They mean the set of subconscious rules we actually use in formulating phrases and sentences of speech. In this sense, there is no such thing as bad grammar as far as linguistics is concerned. There are just many different languages and dialects whose grammars differ from each other to a greater or lesser extent. When linguists discuss grammars, they consider them to be psychologically real. That is, when they say, for example, that there is

a rule for forming the plural in English which is 'Add an s', this is actually a hypothesis that somewhere in the brain of English speakers there is a neural network which carries out this procedure. So a grammar is really a set of hypotheses about the brain.

Grammars have several parts. **Phonology** is the set of rules about how sounds can and cannot be put together in a language. For example, *brick* is a perfectly good English word, whereas *btick* is not possible. This is because the rules about which sounds can go together in English do not allow two heavy consonants like *b* and *t* to go together at the beginning of a syllable. These phonological rules can differ somewhat from language to language. The word *pterodactyl* is pronounced as written in Greek, where sequences of consonants are allowed. In English, though, such sequences are not allowed, so when we adopted the Greek word, it was pronounced *terodactyl*. Sequences that are allowed by the grammar are called **grammatical**; those that are not allowed are called **ungrammatical**. The convention in linguistics is to mark ungrammatical sequences with an asterisk, as in *\*btick*.

**Morphology** is the part of the grammar that deals with the structure of words. Rules like 'Add s to make the plural' are morphological rules. **Syntax** deals with the rules that govern the way individual words are put together into sentences. For example, in English, sentences (2a) and (2b) below are grammatical and (2c) is not, because of the syntactic rules of the language.

(2a) The cat sat on the mat.

(2b) On the mat sat the cat.

(2c) \*The cat sat the mat on.

You can often guess the meaning of sentences that violate syntactic rules from the context and the words used, but you have a strong intuition that they are not quite right nonetheless. **Semantics** is the part of linguistics that deals with meaning. Meaning and grammaticality are two rather separate things. For example, in (3) below, one of the sentences is syntactically fine but meaningless, at least if taken literally, whilst the other is ungrammatical but has a clear meaning.

(3a) Colourless green ideas sleep furiously.

(3b) The man wanted going to the cinema.

### SAQ 1

In example (3), which sentence is ungrammatical and which is meaningless?

#### Answer

Sentence (3a) is meaningless (how can something be colourless and green? how can an idea sleep, especially in a furious manner?), but nonetheless is put together just like a grammatical English sentence. Sentence (3b) seems to mean something like 'The man wanted to go to the cinema' or 'The man wanted taking to the cinema', but we know that it is not a good sentence.

Grammatical rules often involve classes of words. The key classes are **nouns** and **verbs**. Nouns typically denote objects. *Book*, *chair*, *man*, *cloud* and *water* are all nouns. Verbs typically denote actions or processes. *Go*, *come*, *ask*, *eat*, *harass* are all verbs. Verbs



have a **subject**, the person or thing doing the action, and they may have an **object**, the person or thing to which the action is done. A verb with no object is called intransitive. *To sleep* is an intransitive verb, since you can say *I slept* but not *\*I slept the man*. *To kick* is a transitive verb, since you don't just kick in general, you kick someone or something.

### SAQ 2

Can you think of any exceptions to the generalisation that nouns denote objects and verbs denote actions or processes?

#### Answer

Nouns like *arrival*, *explosion* and *assault* seem to denote actions or processes rather than objects.

Note that nouns such as *arrival*, *explosion* and *assault* are derived from verbs, in this case *to arrive*, *to explode*, *to assault*. They are called verbal nouns. The verb *to be* doesn't really describe a process; in fact, it doesn't really have any meaning of its own, but instead is used to join words together in a neutral way, as in *I am tall*.

As well as nouns and verbs there are other, less important categories of words, such as adjectives, adverbs and prepositions.

- Adjectives modify nouns to give us more information about their properties. Colour words are adjectives, as in 'the *green* man' or 'the *black* cat'.
- Adverbs modify verbs, giving us more information about how the action was done, as in 'he arrived *stealthily*' or 'she shouted *loudly*'.
- Prepositions and conjunctions are the little linking words that connect everything together, like *to*, *with*, *on*, *for*, *under*, *but*, *however* and *and*.

### SAQ 3

For each of the following sentences, identify the nouns and verbs. Say whether the verb is transitive or intransitive, and identify the subject and object.

(4a) The man kicked the ball.

(4b) The boy sulked.

(4c) The dog ran after the cat with the attitude problem.

#### Answer

**(4a)** *Man* and *ball* are nouns, *kicked* is the verb. The verb is transitive. The subject is *the man* and the object is *the ball*.

**(4b)** *Boy* is a noun and the verb is *sulked*. The verb is intransitive, with *the boy* as its subject.

**(4c)** *Dog*, *cat* and *attitude problem* are nouns. The verb is *ran after*, which is transitive. *The dog* is the subject and *the cat with the attitude problem* is the object.

## 2.2 Generativity and duality of patterning

Let us now reconsider the sentence you heard in the imaginary scenario at the beginning of this course. Here it is again.

(1) My dad's tutor's no joker, and he told me the TMA's going to hit home with a bang.

### Activity 2

Before reading this section, try writing down what stages you think the brain might go through in turning the sound of sentence (1) above into its meaning. Think about what it is that makes the tasks difficult.

The vervet monkey call system, as we saw, involves a mapping between sounds and meanings, just as human language does. However, it differs from human language in two crucial respects. The first of these is called **generativity**. The three major vervet calls – *snake*, *leopard* and *eagle* – are meaningful units in their own right; you don't need to say anything else, you just call. The calls cannot be combined into higher-order complexes of meaning. A snake call followed by a leopard call could, as far as we understand it, express only the presence of a snake and a leopard. It could not express the proposition that a snake was at that moment being hunted by a leopard, or vice versa, or the idea that leopards are really much more of a nuisance than snakes. This means that the number of meanings expressible in the vervet system is closed, or finite. There are only as many meanings as there are calls. Human languages, by contrast, allow the recombination of their words into infinitely many arrangements, which have systematically different meanings by virtue of the way they are arranged.

In the vervet system, the monkeys can simply store in their memory the meaning associated with each call. Human language could not work this way. Consider example sentence (1) above. You have almost certainly never heard or read this exact sentence before. In fact, it is highly unlikely that anyone, in the entire history of humanity, has ever uttered this exact sentence before me today. Yet we all understand what it means. We must therefore all possess some machinery for making up new meanings out of smaller parts in real time. This is what is known as the **generative** capacity of language; the ability to make new meanings by recombining units. The vervet system is not generative, whereas human language is.

Vervet calls are indivisible wholes; they cannot be analysed as being made up of smaller units. Words, by contrast can be broken down into smaller sound units. Thus language exhibits what is known as **duality of patterning** (Figure 2). At the lowest level, there is a finite number of significant sounds, or **phonemes**. The exact number varies from language to language, but is generally in the range of a few dozen. The phonemes can be combined into words fairly freely; however there are restrictions, known as phonological rules, about how phonemes can go together. Words in their turn combine into sentences. However, as we have just said, not all combinations of words are grammatical. Which combinations are allowed depends on syntactic rules.

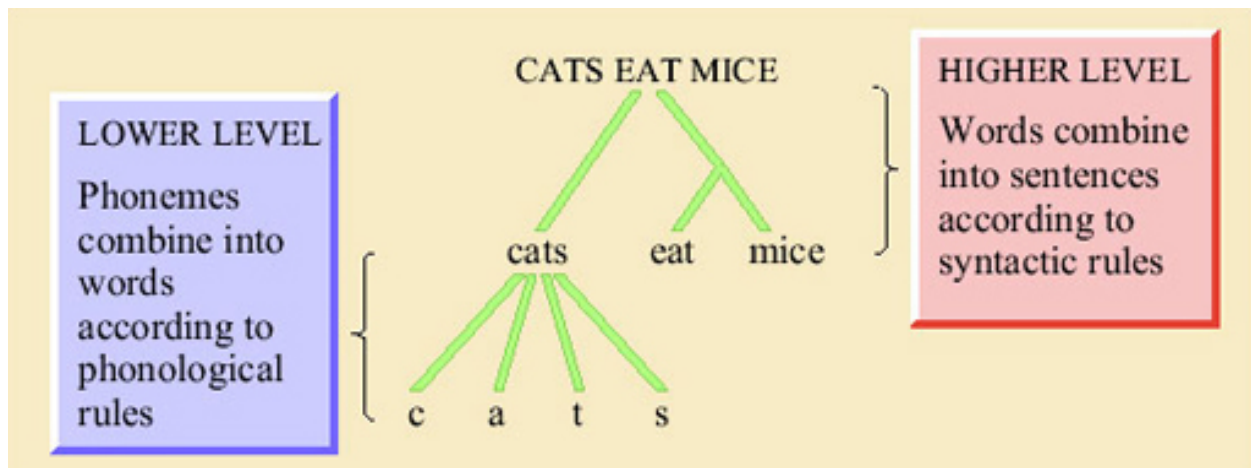


Figure 2 The two levels of patterning in human language

There are some differences between the higher and lower levels of patterning in language. Phonemes, the basic unit of the lower level, have no meaning at all, whereas words, the basic unit of the higher level, typically carry meaning. The meaning of the word *bed* has nothing at all to do with the fact that the phonemes making it up are /b/, /e/ and /d/. (See Box 1 for an explanation of the / / notation.) If you change one phoneme, for example the /d/ to a /t/, then you have a word that is not just different but completely unrelated in meaning – *bet*. You could imagine a hypothetical linguistic system in which particular phonemes had special relationships to meanings; for example, in which words for furniture all began with /b/, or words for body parts all contained an /i/. No human language is like that, however. You cannot predict the meaning of a word, even in the vaguest terms, from the phonemes that make it up.

The higher level of patterning is quite different. The meaning of a sentence *is* largely a product of the meanings of the individual words that it contains. Syntactic rules serve to identify which word in the sentence plays which role, and also to 'glue together' the relationships between the words. Consider these examples.

- (5) The cat bit the dog.
- (6) The cat which was bitten by the dog was thirsty.

In (5), we know that the cat was the biter and the dog the bitten because of a pattern in English syntax which says that the first noun is generally the subject of the sentence. In (6), there are two possible participants to which the state 'thirsty' could be attached – the dog could be thirsty or the cat could be thirsty. The syntax tells us that it must be the cat. Without syntax, no-one would be able to tell who was biting, who was bitten, and who was thirsty in (5) and (6), however much they knew about the behaviour of cats and dogs.

You might say that there is nothing that remarkable about understanding a sentence of spoken English. You listen out for the phonemes; as they come in, you store them in short-term memory until you have enough to make a word. Then the word is passed on to the meaning centres of the brain, where its meaning is activated, and the phonemes making up the next word start coming through. You continue this process until the whole sentence is in. You use your knowledge of syntax to clear up any uncertainties about who did what to whom, and there you are: the meaning. Simple really.

This account underestimates the crucial complexity of linguistic processing in several ways. We will explore this complexity by considering in detail how a sentence (sentence (1) from the opening of this course) could be understood. We will see that there are three

areas of really difficult problems that the brain's linguistic system solves effortlessly. These are the phonological problem, the semantic problem, and the syntactic problem.

## 2.3 From ear to phoneme: the phonological problem

The phonological problem is the problem of knowing which units (words, calls) are being uttered. The speech signal is a pattern of sound, and sound consists of patterns of minute vibrations in the air. Sounds vary in their frequency distribution. The sound of a flute playing is relatively **harmonic**. This means that the energy of the sound is concentrated at certain frequencies of vibration. A plot of the energy of a sound against the frequency at which that energy occurs is called a spectrogram. A spectrogram for a flute's note is shown in Figure 3a. As you can see there are slim coloured bands, and black spaces in between. The coloured bands are the regions of the frequency spectrum where the acoustic energy is concentrated, whereas in the black areas there is little or no acoustic energy. The lowest coloured band corresponds to the **fundamental frequency** of the sound. This is where the most energy is concentrated, and it is the fundamental frequency which gives the sensation of the pitch of the sound. The higher bands are called the **formant frequencies**. In a 'pure' tone, their frequencies are mathematical multiples of the fundamental (in acoustics in general, they are also called overtones or harmonics, but in relation to speech, they are always called formants). The relative strengths of the different formants determine the timbre or texture of the sound.

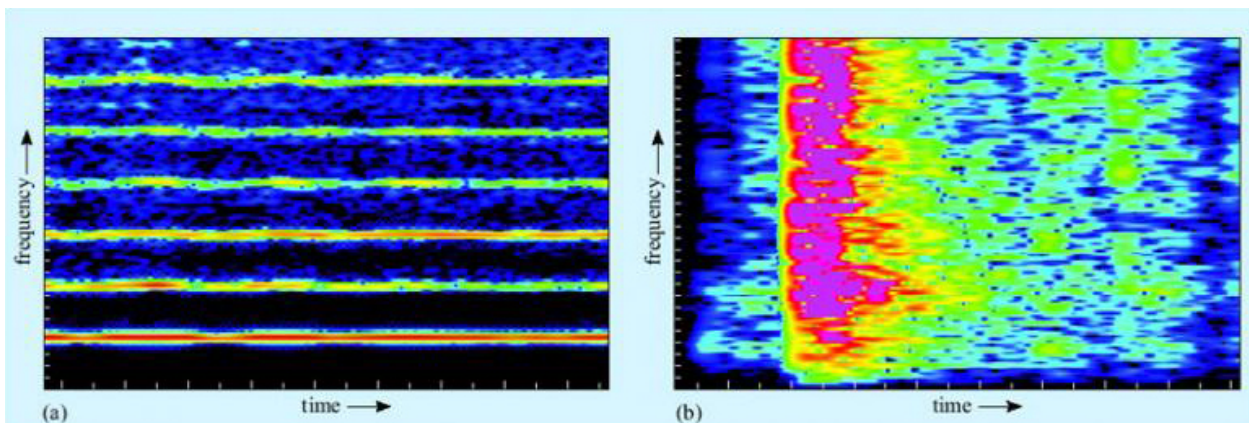


Figure 3 Spectrogram of (a) a single note from a flute; and (b) a box being dropped on the floor. Black represents no acoustic energy at a given frequency. The 'hotter' the colour (i.e. towards the red or yellow end of the spectrum), the greater the concentration of energy at that frequency. Note in (a) that the greatest concentration of energy is in the lowest band (the fundamental frequency), with regular and increasingly faint harmonics at higher frequencies. The harmonics have clear gaps between them, which creates the feeling of a pure and tuneful note. In (b), by contrast, the acoustic energy is smeared across the whole frequency range, making it sound like a noise rather than a note.

Courtesy of Daniel Nettle

Courtesy of Daniel Nettle

**SAQ 4**

In terms of fundamental and formant frequencies, why might a violin, a flute, an oboe and a human voice producing the same note sound so different?

**Answer**

The fundamental frequency is necessarily the same in all cases, since the pitch of the note is the same. The relative strength of the different formants is the main source of the different qualities – thin, reedy, soft, full or whatever – of the notes.

In contrast to harmonic sounds are sounds in which the acoustic energy is dispersed across the frequency spectrum, like the box being dropped in Figure 3b above. These are experienced as noises rather than tones (it is impossible to hum them), and they appear on the spectrogram as a smear of colour.

For the vervet monkeys, the phonological problem is not too difficult. There are only three major alarm calls, and their acoustic shapes are radically different and non-overlapping (Figure 4). They are also different from the other vocalisations vervets produce during social encounters. Thus the incoming signal has to be analysed and matched to a stored representation of one of the three calls.

The human case is more complex. The first stage is to extract which phonemes are being uttered. Phonemes come in two major classes, consonants and vowels. Vowels are harmonic sounds. They are produced by periodic vibration of the vocal folds which in turn causes the vibration of air molecules. The frequency of this vibration determines the pitch (fundamental frequency) of the sound. Different vowels differ in quality, or timbre, and the different qualities are made by changing the shape of the resonating space in front of the vocal folds by moving the position of the lips and tongue relative to the teeth and palate. This produces different spectrogram shapes, as shown in Figure 5. (The conventions used in this chapter to represent spoken language are given in Box 1 below.)

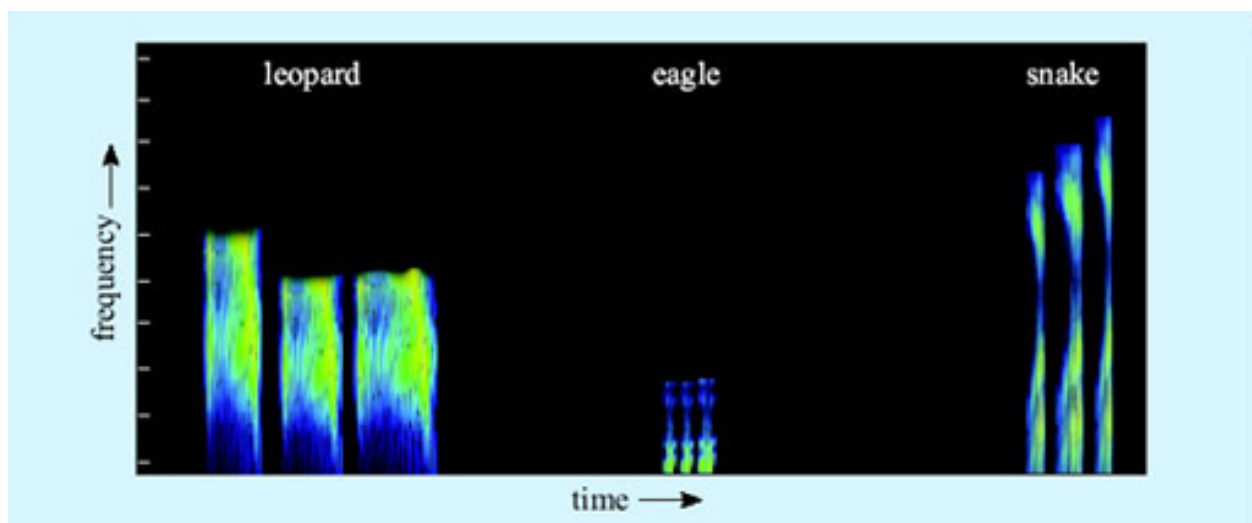


Figure 4 Spectrogram of the three vervet alarm calls



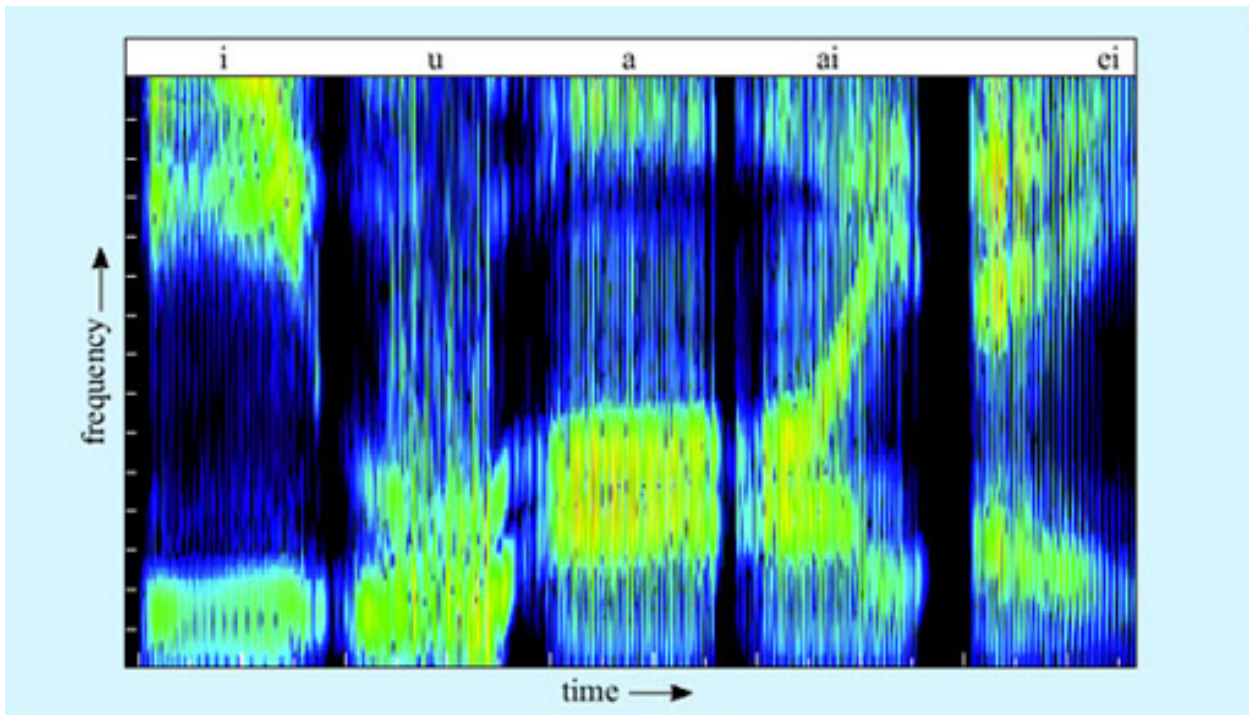


Figure 5 The spectrogram shapes of five vowels, spoken by the author. The vowels correspond to the vowel sounds in *beat*, *boot*, *bet*, *bite* and *bait*. What distinguishes the different vowels is not the absolute frequency of the formants but their position relative to each other

Courtesy of Daniel Nettle

### Box 1: Representing spoken language

The spelling we usually use to represent English in text does not relate very systematically to the sounds we actually make. Consider, for example, the words *farm* and *pharmacy*. The beginnings of the words are identical to the ear, and yet they are written using different letters. The reasons for this are usually historical, in this case due to *pharmacy* coming into English from Greek. The letter *r* is also there as a historical remnant – the *r* in *farm* is now silent, but it used to be pronounced, and still is in some varieties of English, for example in South-West England and in Scotland.

For linguists it matters what sounds people actually produce, so they often represent spoken language using a system called the International Phonetic Alphabet (IPA). Sequences of speech transcribed in IPA are enclosed in slash brackets, / /, or square brackets, [ ], to distinguish them from ordinary text. Many of the letters of the IPA have more or less the same value as conventional English spelling. For example, /s/, /t/, /d/ and /l/ represent the sounds that you would expect. The IPA is always consistent about the representation of sounds, so *cat* and *kitten* are both transcribed as beginning with a /k/.

The full IPA makes use of quite a lot of specialised symbols and distinctions that are beyond our purposes here, so we have used ordinary letters but tried to make the representation of words closer to what is actually said wherever this is relevant to the argument. We have used the convention of slash brackets to indicate wherever we have done this, so for example, the word *cat* would be /kat/ and *sugar* would be /shuga/.



## SAQ 5

Is the vowel of *bit* higher or lower in pitch than the vowel of *bat*?

## Answer

There is no inherent difference in pitch (fundamental frequency) between *bit* and *bat*. You can demonstrate this by saying either of them in either a deep or a high pitched voice. The difference between them is in the relative position of the higher formant frequencies.

The relative position of the first two formant frequencies is crucial for vowel recognition. Artificial speech using just two formants is comprehensible, though in real speech there are higher formants too. These higher formants reflect idiosyncracies of the vocal tract, and thus are very useful in recognising the identity of the speaker and the emotional colouring of the speech. Some of the vowels of English have a simple flat formant shape, like the /a/ and /i/ in Figure 5 above. Others, like the vowels of *bite* and *bait*, involve a pattern of formant movement. Vowels where the formants move relative to each other are called diphthongs.

**Consonants**, in contrast to vowels, are not generally harmonic sounds. Vowels are made by the vibration of the vocal folds resonated through the throat and mouth with the mouth at least partly open. Consonants, by contrast, are the various scrapes, clicks and bangs made by closing some part of the throat, tongue or lips for a moment.

## SAQ 6

Make a series of different consonants sandwiched between two vowels – *apa, ata, aka, ava, ama, afa, ada, aga, a'a* (like the Cockney way of saying *butter*). Where is the point of closure in each case and what is brought to closure against what?

## Answer

*apa, ama* – The two lips together

*ata, ada* – The tip of the tongue against the ridge behind the back of the top teeth

*aka, aga* – The back of the tongue against the roof of the mouth at the back

*ava, afa* – The bottom lip against the top teeth

*a'a* – The glottis (opening at the back of the throat) closing

The acoustics of the consonants are rather varied. Some consonants produce a burst of broad spectrum noise – look for instance at the /sh/ in Figure 6. Others have relatively little acoustic energy of their own, and are most detectable by the way they affect the onset of the formants of the following vowel.

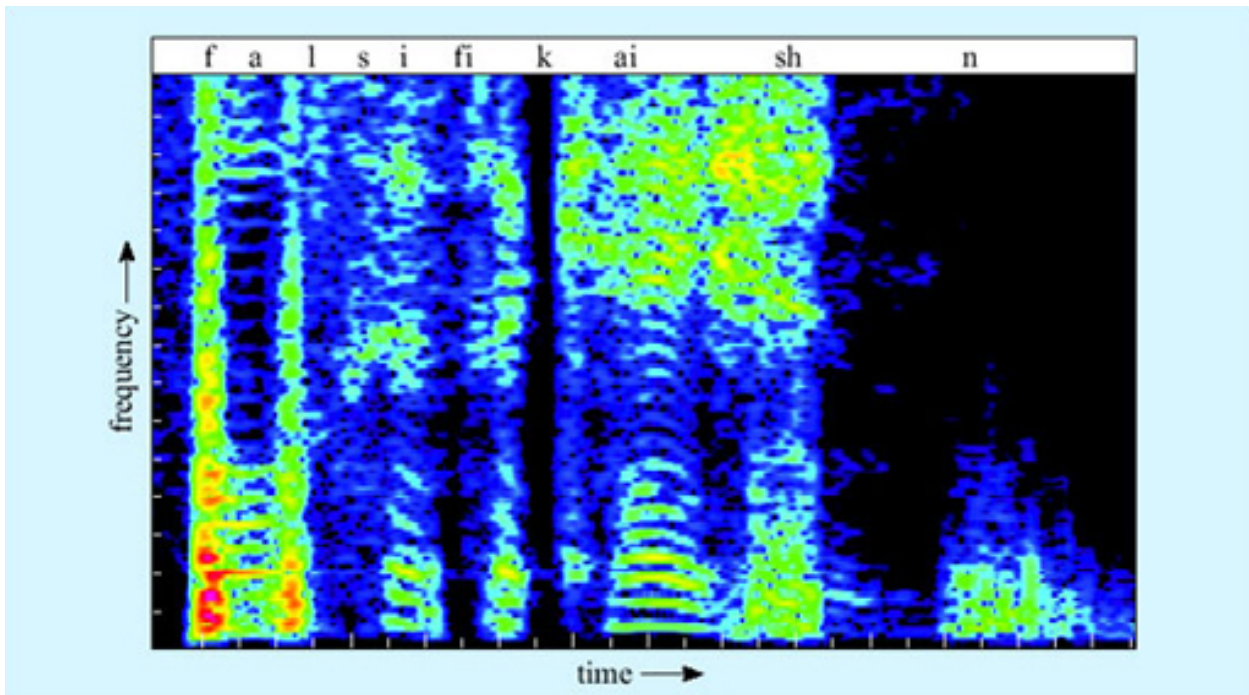


Figure 6 Spectrogram of the word *falsification* spoken by the author. Note the period of silence (black areas) in the word. This corresponds to the closure during the formation of the /k/ sound. The /sh/ corresponds to a burst of noise over most of the spectrum

Courtesy of Daniel Nettle

Courtesy of Daniel Nettle

Phonemes make up small groups called syllables. Typically a syllable will be one consonant followed by one vowel, like *me*, *you* or *we*. Sometimes, though, the syllable will contain more consonants, as in *them* or *string*. Different languages allow different syllable shapes, from Hawaiian which only tolerates alternating consonants and vowels (which we can represent as CVCVCV), to languages like Polish which seem to us to have heavy clusters of consonant sounds. English is somewhere in the middle, allowing things like *bra* and *sta* but not allowing other combinations like *\*tsa* or *\*pta*.

### SAQ 7

Pause for thought: Why do you think language might employ a basic structure of CVCVCV?

#### Answer

Long strings of consonants are impossible for the hearer to discriminate, since identification of some consonants depend upon the deflection they cause to the formants of the following vowel. Long strings of vowels merge into each other. Consonants help break up the string of vowels into discrete chunks. So an alternation of the two kinds of sound is an optimal arrangement – after all, the babbling of a baby uses it.

The task of identifying phonemes in real speech is made difficult by two factors. The first is the problem of *variation*. Phonemes might seem categorically different to us, but that is the product of our brain's activity, not the actual acoustic situation. Vowels differ from each

other only by degree, and this is also true for many consonants. A continuum can be set up between a clear /ba/ and a clear /da/ (Figure 7). Listening to computer-generated sounds along this continuum, the hearer hears absolutely /ba/ up to a certain point, then absolutely /da/, with only a small zone of uncertainty in between. In that zone of uncertainty (and to some extent outside it), the context will tend to determine what is heard. What the listener does not experience is a sound with some /ba/ properties and some /da/ properties. It is heard as either one or the other, an effect known as **categorical perception**.

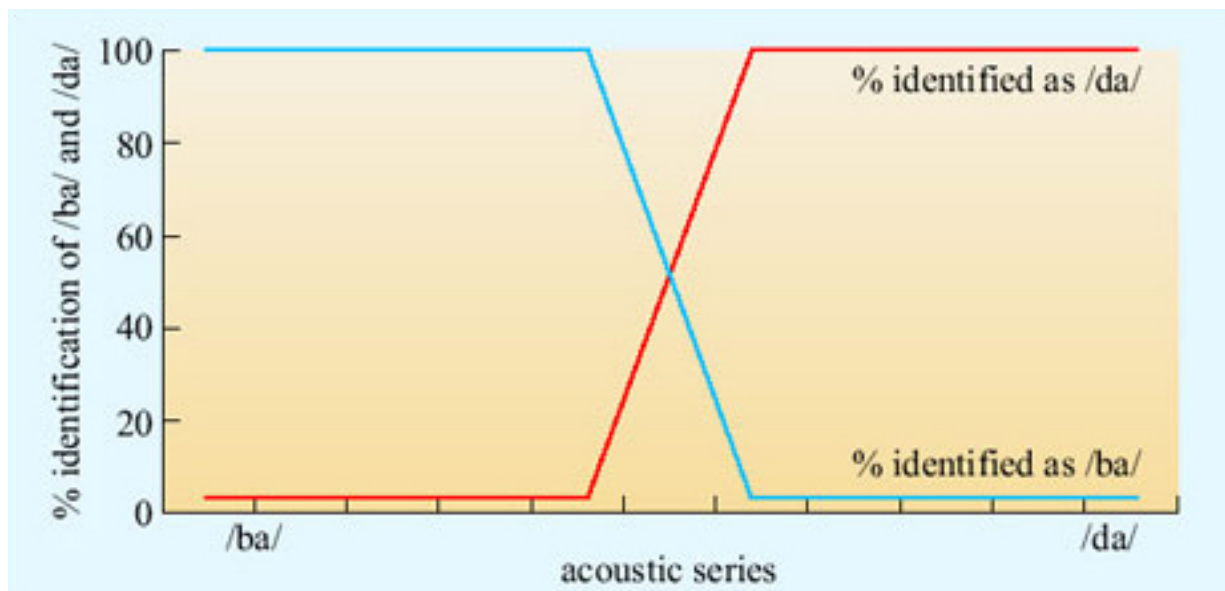


Figure 7 Categorical perception of /ba/ versus /da/ as a function of change in a computer-generated speech signal. The signal is varied continuously in a linear fashion, but what the hearer hears is a non-linear change; /ba/ until a certain point, where after the sound quite suddenly becomes /da/

Holly, R. et al. (1997) 'Neurobiology of speech perception', *Annual Review of Neuroscience*, 20. Copyright © 1997 by Annual Reviews Inc. All rights reserved

Holly, R. et al. (1997) 'Neurobiology of speech perception', *Annual Review of Neuroscience*, 20. Copyright © 1997 by Annual Reviews Inc. All rights reserved

The hearer cannot rely on the absolute frequency of formants to identify vowels, since, as we have seen, you can pronounce any vowel at any pitch, and different speakers have different depths of voices. So a transformation must be performed to extract the position of (at least) the first two formants relative to the fundamental frequency. But that is not all. Different dialects have slightly different typical positions of formants of each vowel; within dialects, different speakers have slightly different typical positions; and worst of all, within speakers, the relative positions of the formants change a little from utterance to utterance of the same underlying vowel (Figure 8). The hearer thus needs to make a best guess from very messy data.

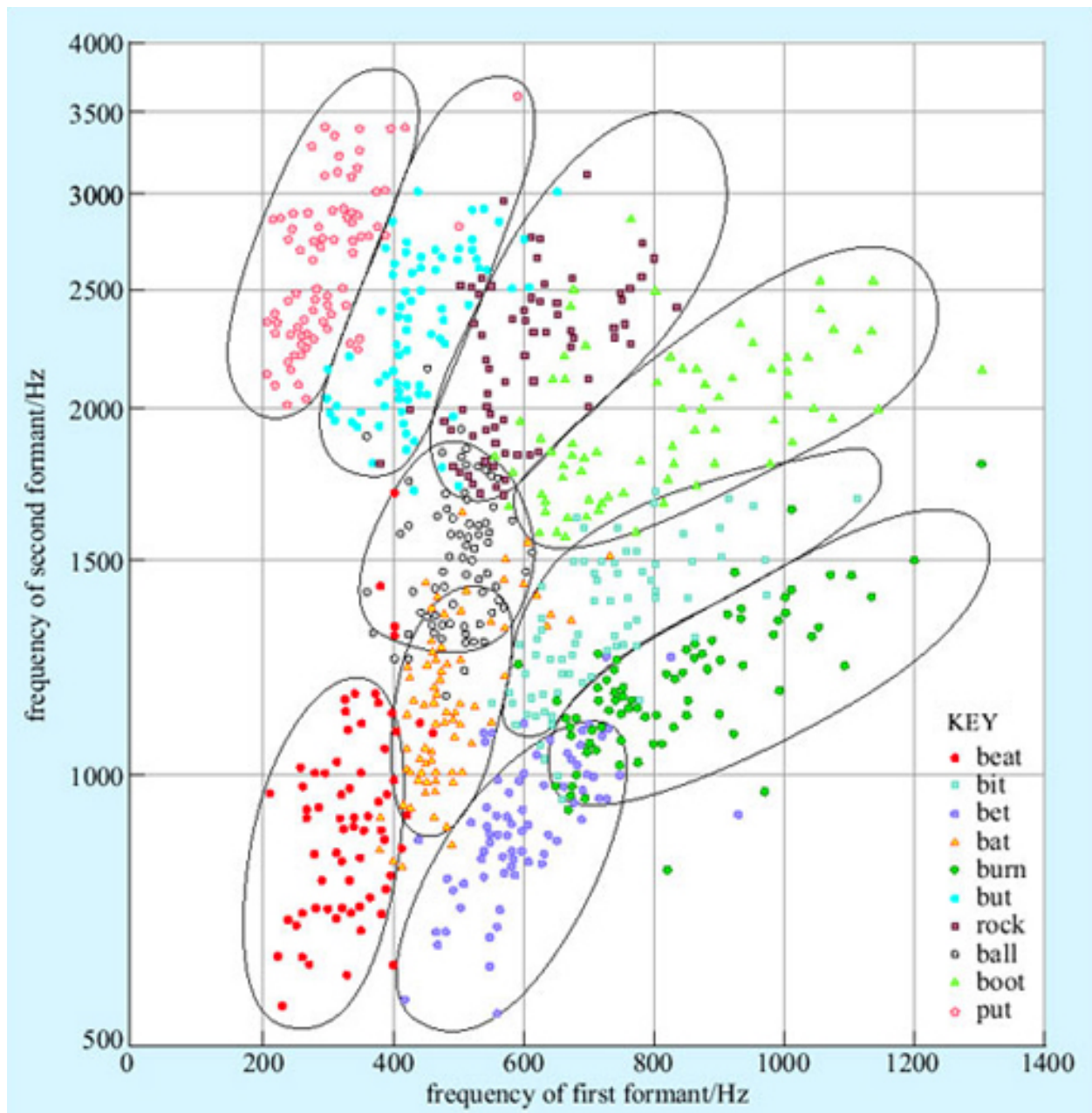


Figure 8 Acoustic realisation of different English vowels in samples of normal speech from several speakers. The graph shows the first and second formants of the vowel sound from the different words shown in the key. Note that the vertical scale is not linear

Peterson, G. E. and Barney, H. L. (1952) 'Control method used in a study of vowels', *The Journal of the Acoustical Society of America*, Vol. 24, No. 2, March 1952. The Acoustical Society of America. Used by permission of the American Institute of Physics

Peterson, G. E. and Barney, H. L. (1952) 'Control method used in a study of vowels', *The Journal of the Acoustical Society of America*, Vol. 24, No. 2, March 1952. The Acoustical Society of America. Used by permission of the American Institute of Physics

This is made more difficult by the second factor, which is called **co-articulation**. The realisation of a phoneme depends on the phonemes next to it. The /b/ of *bat* is not quite the same, acoustically, as the /b/ of *bit*. We are so good at hearing phonemes as phonemes that it is difficult to consciously perceive that this is so, except by taking an extreme example, as in the exercise below.

## SAQ 8

Listen closely to the phoneme /n/ in your own pronunciation of the word *ten*, in the following three contexts – *ten newts*, *ten kings*, *ten men*. Say the words repeatedly but naturally to identify the precise qualities of the /n/ in each case. Are they the same? If not, what has happened to them?

## Answer

You will probably find that the articulation of the /n/ is 'dragged around' by the following consonant – towards the /ng/ of *long* in *ten kings*, and towards /m/ in *ten men*. If this is not clear, try saying *ten men tem men* over and over again (or alternatively *ten kings teng kings*). You soon realise that there is no acoustic difference whatever between the two phrases. This is an example of assimilation, a closely related phenomenon to co-articulation.

Co-articulation makes the task of the hearer even harder, because they have to undo the co-articulation that the speaker has put in (unavoidably, since co-articulation is unstoppable in fast connected speech). A sound which is identical to an /m/ which the listener has previously heard might actually be an /m/, but it might equally be an /n/ with co-articulation induced by the context. An upward curving onset to a vowel might signal a /d/ if the vowel is /e/, but a /b/ if the vowel is /a/. Yet the listener does not yet know what the vowel is. They are having to identify the signal under multiple and simultaneous uncertainties. Speakers use their expectations about what is to follow to resolve these uncertainties.

Other problems arise in the extraction of words from sound. Consider Figure 9, which is part of the spectrogram of the author producing sentence (1). Our brains so effortlessly segment speech into words that we are tempted to assume that the breaks between the words are 'out there' in the signal. But as you can see, there are no such gaps in speech. There are moments of low acoustic intensity (black areas), but they do not necessarily coincide with word boundaries. Co-articulation of phonemes can cross word boundaries.



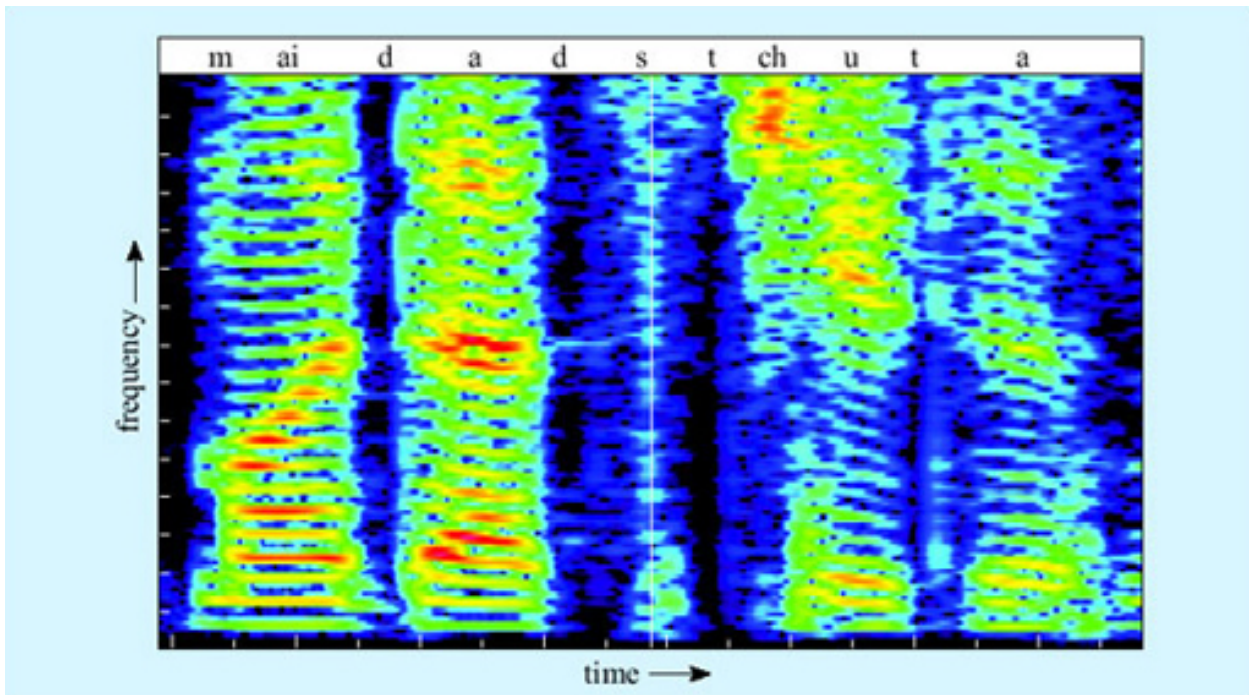


Figure 9 Spectrogram of the phrase *my dad's tutor*, spoken by the author

Courtesy of Daniel Nettle

Courtesy of Daniel Nettle

Strings of phonemes have multiple possible segmentations. *My dad's tutor* could be segmented, among many other possibilities, as:

- (7a) [my] [dads] [tu] [tor]
- (7b) [mide] [ad] [stewt] [er]
- (7c) [mida] [dstu] [tor]

The signal rarely contains the key explicitly. The hearer can exploit knowledge of English phonological rules, for example to exclude (7c) on the grounds that it contains an impossible English syllable. Beyond that, knowledge of words must come into play. Segmentation (7b) is phonotactically fine, but doesn't mean anything in English. Our segmentations always align on those solutions that furnish a string of real words, so (7a) would be chosen. If your name was (7b), you would have extreme trouble introducing yourself, however clearly you spoke, whereas any segmentation containing whole words is much easier to understand.

### SAQ 9

Can you see any problems that might arise with achieving either phoneme identification before segmentation into words, or segmentation into words before phoneme identification?

#### Answer

As we saw above, identification of phonemes often depends upon the word context. On the other hand, segmentation into words depends upon identifying the phonemes. In short it seems likely that neither can always be achieved prior to the other.



We will consider models of how the brain actually does it in Section 3.

## 2.4 From phoneme to meaning: the semantic problem

For a vervet monkey, once an alarm call has been assigned to the correct phonological class – *leopard*, *snake* or *eagle* – then the task is straightforward indeed. Each of these sound patterns is connected in long-term memory to some 'prototype' of the predator in question, including what it looks like, what it does and so on. The *eagle* call always means eagle, whatever the context. The activation of the brain trace for the *eagle* call invariably activates the brain trace for the eagle meaning. That is how the monkeys know to look up after an eagle call and down after a snake call.

The classic way of describing language processing in psychology textbooks makes it seem that human language is not much different. Each word, these textbooks say, is linked to a non-linguistic 'prototype' of what it means. For *eagle*, this would be pretty similar to the vervet's representation of the *eagle* call's meaning. The trouble with this account is that much of the meaning in human language is not really like that.

Many words in everyday language do not carry any meaning, at least not in anything like the *eagle* sense. Consider sentence (8).

(8) It is raining.

There is only one element in (8) with any independent meaning. The *it* certainly does not refer to anything. (You can't seek clarification of (8) by asking *What is raining?*) The *is* has no meaning. Indeed, the verb *to be* in general is completely devoid of intrinsic meaning. So why are there three words in this sentence, which really consists of a single meaningful element – *raining*?

Many words in language are there only because the syntax of a language demands that sentences conform to certain templates; in English, for example, that they have both a verb and a subject. The subject is usually the doer in a sentence. But there is no real subject in the fact that it is raining today – it is just a state of affairs. Thus English syntax demands we fill in the sentence with what is largely meaningless material – *it* – to make the sentence well-formed. Italian, which is more relaxed about the overt expression of a subject, allows example (8) to be captured with a single word (9), which is also a well-formed sentence.

(9) Piove.

Other languages, like Hebrew and Arabic, relax the requirement for every sentence to have a verb, allowing sentences like (10).

(10) Aniiayaf.

I tired (masculine).

I am tired.

However they differ, languages all require a wealth of material whose meaning is either nearly zero, or is filled in entirely from context. Thus for some words, what you look up in long-term memory is not a prototype of the thing to which the word refers, but a further instruction to look for something else in the sentence. Inferring which function a word is carrying in a sentence is subtle. Consider (11).

(11a) Paula and Mary ran panting in from the field, where the dog was barking. It was running now.

(11b) Paula and Mary ran panting in from the field, where the dog was barking. It was ploughed now.

(11c) Paula and Mary ran panting in from the field, where the dog was barking. It was raining now.

These sentence pairs differ only by a single word. In (11a), *it* is generally taken to mean the dog. In (11b), the *it* cannot be read as identifying the dog; it must mean the field. In (11c), the *it* stands for neither dog nor field, but is a purely syntactic *it* filled in by the need for the verb to have a subject. Thus, having identified *it* as the word spoken in this position in the sentence, the brain has to try out several possibilities for its meaning in the context, until it comes up with one that makes some kind of sense. Note that you cannot begin to work out what the *it* means in (11) until you have had the key word *running*, *ploughed* or *raining*. This is a very important point. In processing language in general, you cannot process the meaning of all the elements in the order in which they come. Consider (12).

(12a) As they crashed into the bank, the robbers leapt from their car.

(12b) As they crashed into the bank, the men leapt from their canoe.

The meaning of *bank* could be to do with money or to do with rivers. This cannot be resolved until later information arrives. *Bank* must thus be kept in a pending tray somewhere until more data are available. Linguists thus believe the brain must have a storage pad in which sentences are assembled and disassembled, with the words represented in a way which is still neutral as to which meaning is the right one.

Thus, identifying the meaning of a word depends not just on the phonological form of the word, but on two other things. These are the rest of the sentence, and also the wider context of the discourse. This latter is important, since your reading of (12b) would be quite different in a story about three guys who decide to rob Barclays disguised as the British kayaking team on a promotional tour.

The resolution of meaning depends on a dynamic interplay of these three elements. Often the meaning of a set of words departs almost entirely from the meaning you would expect from them individually. Just consider example (1). It has nothing to do with hitting, nothing to do with homes, and only a very metaphorical bang. Sentences at least as complex as this are everyday occurrences, yet you decode their meanings in a completely automatic way without even being aware of the work your brain is having to do.

## 2.5 From phoneme to sentence structure: the syntactic problem

In the vervet monkey system, calls stand by themselves. Thus there is no syntax. Syntax can be thought of as working like road traffic rules do. It doesn't much matter which side of the road you drive on, as long as there is some clear convention. Similarly in (13), it is necessary to understand the difference between (13a) and (13b) without ambiguity, by having some rule or other about which noun phrase comes first. England may differ from most of the rest of the world in terms of the side of the road that it drives on, but English agrees with 99 per cent of all languages in the syntactic convention that the first noun phrase encountered is generally taken to be the subject.

(13a) The cat bit the dog.

(13b) The dog bit the cat.

Syntax is concerned with such problems as the assignment of roles (subject, object, etc.) to the different participants in a sentence, and the binding together of different meanings. The syntactic task involved in even a straightforward sentence is of fearsome complexity. Consider for example (14).

(14a) The dog bit the green cat.

(14b) The dog which bit the cat was green.

(14c) The dog which bit the cat your uncle Marvin warned you would betray you on your fourteenth birthday was green.

In all these sentences, there are at least three important word meanings activated – *dog*, *cat* and *green*. We can assume that these meanings are different bundles of sensory features in long-term memory somewhere, which get activated by the recognition of their respective word forms. The question is how *green* gets hooked to the correct one of the other two traces, *dog* and *cat*. This is known as the **binding problem**. How the brain solves the binding problem is an active area of research.

You might assume from (14a) that knowing which bindings to make in language was quite straightforward. The brain, encountering an adjective, just has to look for the nearest noun phrase and bind the meanings together. Thus in (14a), *green* must be the cat not the dog because *green* is next to *cat*. However, in (14b), *green* is closest to *cat* in the sentence, but its meaning applies to the dog. In (14c), *green* is seventeen words away from *dog*. After *dog* come several other things which could be green: the cat, you and uncle Marvin. However, the only possible sense of this sentence for any native speaker of English is that the dog is green.

This tells us some very significant things about syntactic processing. The first is that the binding of different words together depends not upon their linear position in the sentence, but on their underlying syntactic relationships. The processing of syntactic relations cannot be completed until the whole sentence is finished.

The brain's analysis of sentences depends upon identifying key words, which then offer 'slots' into which to bind the meanings of other elements. Thus in (14a), the verb *bit* offers two slots for noun phrases to bind to, one for the subject and one for the object. Any noun phrase which precedes the verb binds into the subject position. The first noun in a noun phrase is the head noun; the head noun offers slots to which other words like adjectives and articles (*green* and *the*) can bind. A whole clause can bind onto a head noun – this is what happens in (14c), where the clause *which bit the cat ... birthday* binds to *the dog*. The whole thing then becomes a super-noun phrase which binds as a whole into the subject position of *bit* – that is, the subject of *bit* is not strictly *the dog* but *the dog which bit the cat ... birthday*. Another way of looking at the non-binding of *green* to *the cat* is to say that *the cat* is already bound (into a noun phrase) and so is not free to go hooking up with verbs that might be floating around.

There are many other subtle syntactic processes, most of which we cannot consider here. Consider (1) again.

(1) My dad's tutor's no joker, and he told me the TMA's going to hit home with a bang.

The brain has to assign a value to the *he*. It could be *my dad* that told me, or it could be *my dad's tutor*. The phrase *with a bang* could modify the way the TMA is going to hit home, or

it could modify the way I was told about it. The former interpretation is preferred, but if you changed it to *with a smile*, then your interpretation flips to the latter. This is a case where the semantics of a word informs the syntactic analysis of the sentence.

Overall, then, the brain attempts to bind the words that the phonology has identified into a syntactic package that is well-formed, and liaises with the semantics to see if a sensible meaning is thereby produced. This is a dynamic process with interaction in real-time between the different components, which finally crystallises on a unique meaning for at least most sentences. It is many orders of magnitude more complex than the vervet system, or indeed any other cognitive system that has yet been described. We will consider in Section 3 how the brain does it.

## 2.6 Summary of Section 2

Human language is a complex communication system that allows the generation of infinitely many different messages by combining the basic sounds (phonemes) into words, and combining the words into larger units called sentences. The way the sounds combine is governed by phonological rules, and the way the words combine is governed by syntactic rules.

Phonemes can be divided into the vowels, which are made by vibration of the vocal folds, and consonants, which are abrupt sounds made by bringing two surfaces in the vocal tract together. Different phonemes have different acoustic shapes, but in connected speech these are variable because of the influence of the sounds before and after (co-articulation), so the hearer has to make a best guess using information from the context.

The meaning of a sentence is more complex than just the sum of the meanings of the words in the sentence. The hearer must also perform a syntactic analysis of the sentence to establish which meanings relate to which others, and to fill in elements that are ambiguous. The syntactic relationships within a sentence are not simply reflected by the order in which the words arrive, so the hearer has to keep the sentence in a working memory buffer whilst different interpretations are tried out in order to come up with one that is both grammatical and meaningful.

# 3 The brain's solution: the machinery of language

## 3.1 Speech perception

Now that we have examined the processes involved in understanding a sentence in some detail, we will turn to the issue of how the brain achieves the task. We will begin with the initial capture and analysis of the speech signal.

Vibrations in the air are channelled by the structure of the external ear into the ear canal (Figure 10). In the middle ear, they encounter a taut membrane or eardrum stretched across the ear canal. The vibrations in the air set up sympathetic vibrations in the eardrum. Inside the middle ear is a set of three tiny bones, the **auditory ossicles**. The

auditory ossicles in their turn are caused to vibrate by the vibrations of the eardrum. The inner end of the auditory ossicles abuts a fluid-filled coiled structure called the **cochlea**. Vibration of the ossicles is transferred into the fluid of the cochlea and particularly into a thin membrane that runs along its length called the **basilar membrane**. Adjacent to the basilar membrane is a layer of small receptor cells, each with tiny cilia or hairs on it. Indeed, these receptors are known as **hair cells** (Figure 11). Movement of the basilar membrane causes movement of the hairs, which is converted into changes in electrical activity within the cell. Hair cells form synaptic connections with adjacent neurons, and thus electrical changes within them trigger neuronal action potentials. These neurons join into the **auditory nerve**, which relays the action potentials from the ear to the brain.

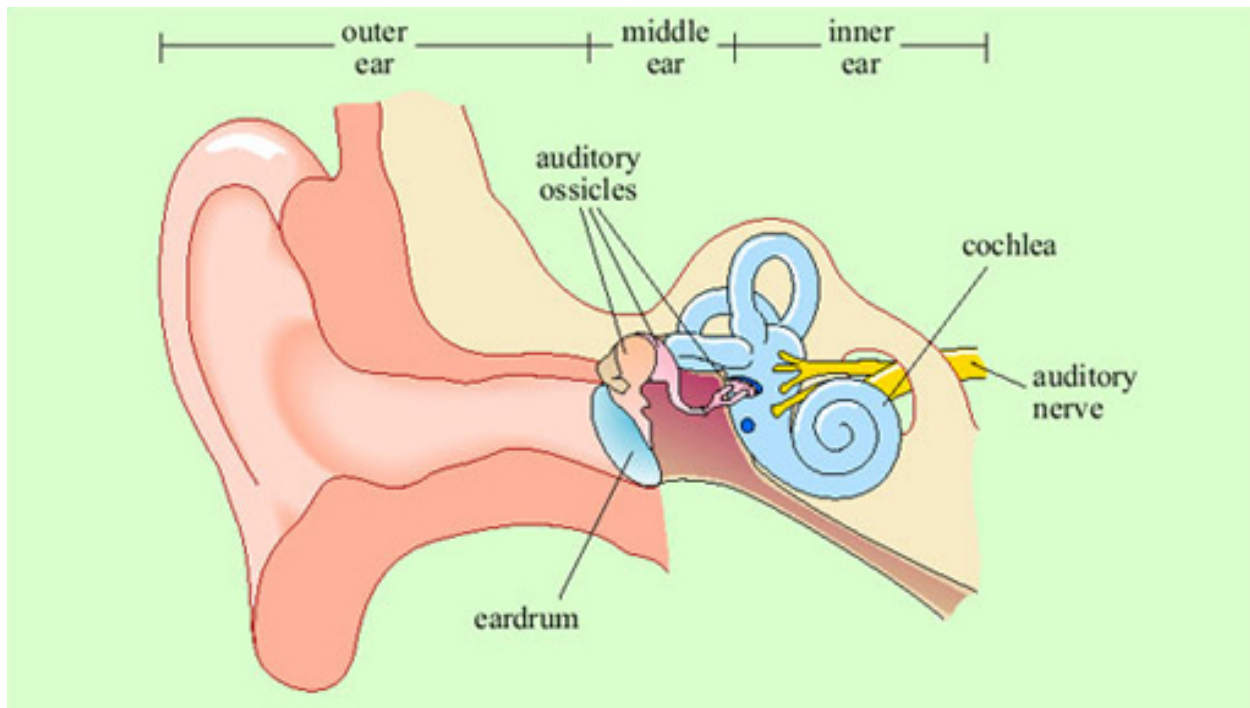


Figure 10 A cutaway diagram of the human ear





Figure 11 Scanning electron micrograph of auditory hair cells

Copyright © VVG/Science Photo Library

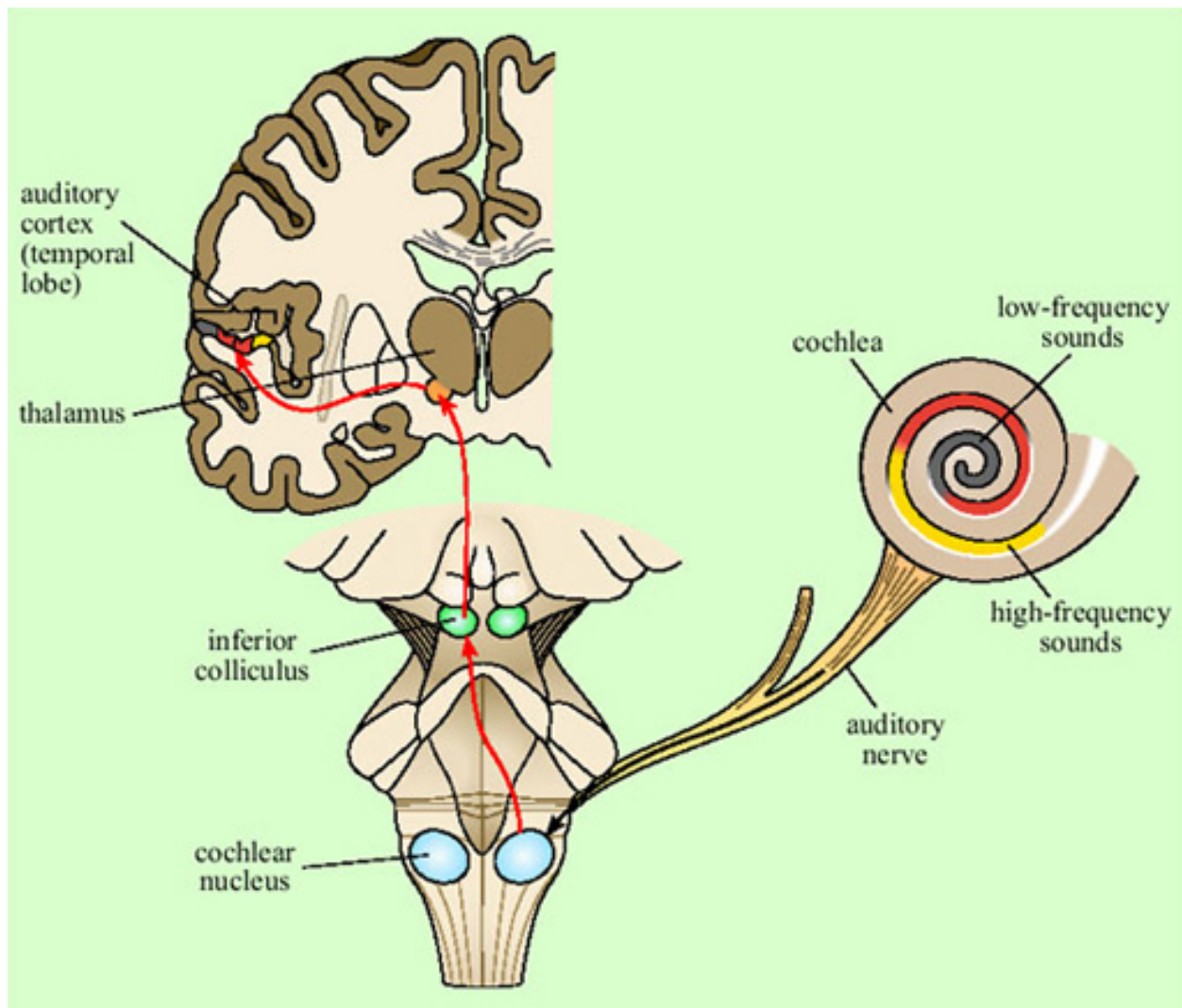
Copyright © VVG/Science Photo Library

The pattern of vibration of the eardrum and ossicles is simply a reflection of the pattern of vibration in the air. In the cochlea, transformation of the signal begins. Because of the structure of the cochlea, high-frequency vibrations cause displacement of the basilar membrane at the outer end, and low-frequency vibrations cause displacement further along. Thus different hair cells at different positions will respond to different frequencies of sound, by virtue of being adjacent to different sections of the basilar membrane.

Thus, two crucial things have happened at the cochlear stage. First, a mapping of sounds of different frequencies onto different places on the basilar membrane has been set up. This is called **tonotopic** organisation. Second, any sound which consists of patterns of acoustic energy at several different frequencies will have been broken down into its component frequencies. This is because each formant within the complex sound will cause vibration at a different position along the basilar membrane and hence cause different subsets of hair cells to respond. Action potentials generated in the neurons that connect to the hair cells are transmitted to the brain via the auditory nerve. What will be transmitted to the brain, then, already contains information about pitch (coded by which cells are firing), and a preliminary breakdown into formants.

The auditory nerve feeds into the brainstem (at the cochlear nucleus), from where the auditory pathway ascends into the medial geniculate nucleus of the thalamus, the relay station in the middle of the forebrain (Figure 12). The main route for auditory information from here is to the auditory cortex of the superior temporal lobe on both sides of the brain (though there is also another pathway to the amygdala, not shown on the figure). Neurons in the auditory cortex generally respond to information from the ear on the opposite side of the body, though some integration of information from the two ears occurs at the brainstem level.





**Figure 12** Connections from the ear to the primary auditory cortex. The arrow indicates the route of the auditory information

Martini et al. (2000, Figures 18–17, p. 483) in Toates, F. (2001) 'The other sensory systems', in *Biological Psychology*, Pearson Education

Martini et al. (2000, Figures 18–17, p. 483) in Toates, F. (2001) 'The other sensory systems', in *Biological Psychology*, Pearson Education

Representation of the signal in the primary auditory cortex is tonotopic. That is, cells at different locations respond to sounds at different frequencies, resulting in a 'map' of the frequency spectrum of the sound laid out across the surface of the brain (Figure 13). Recognition of sounds depends not on the absolute pitch of the formants but on their relationship to each other. We assume that this is processed in deeper layers of the auditory cortex, though exactly where or how is not yet understood.

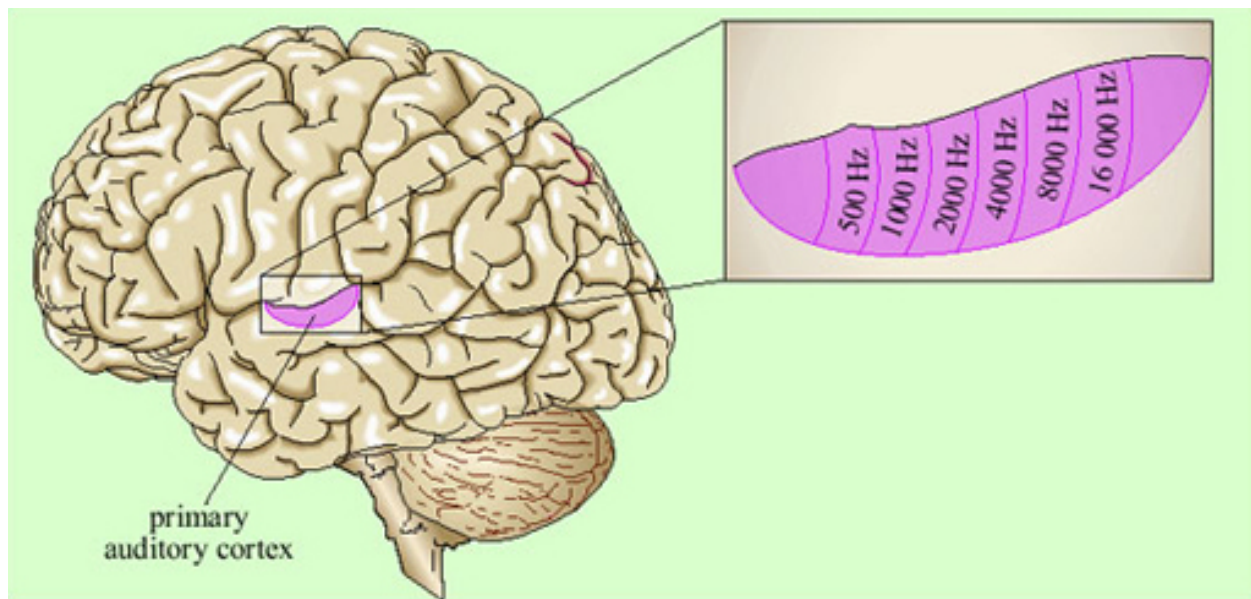


Figure 13 Tonotopic organisation within the primary auditory cortex

There is some evidence that within the primary auditory cortex, there are populations of neurons specialised for speech. This was shown by brain imaging experiments that compared patterns of activation in response to speech, scrambled cocktails of speech sounds, and non-speech sounds which were matched to the speech sounds on basic acoustic features (Moore, 2000). Areas in the superior temporal sulcus, on both sides of the brain, responded preferentially to real and scrambled speech (Figure 14), rather than to the other sounds.

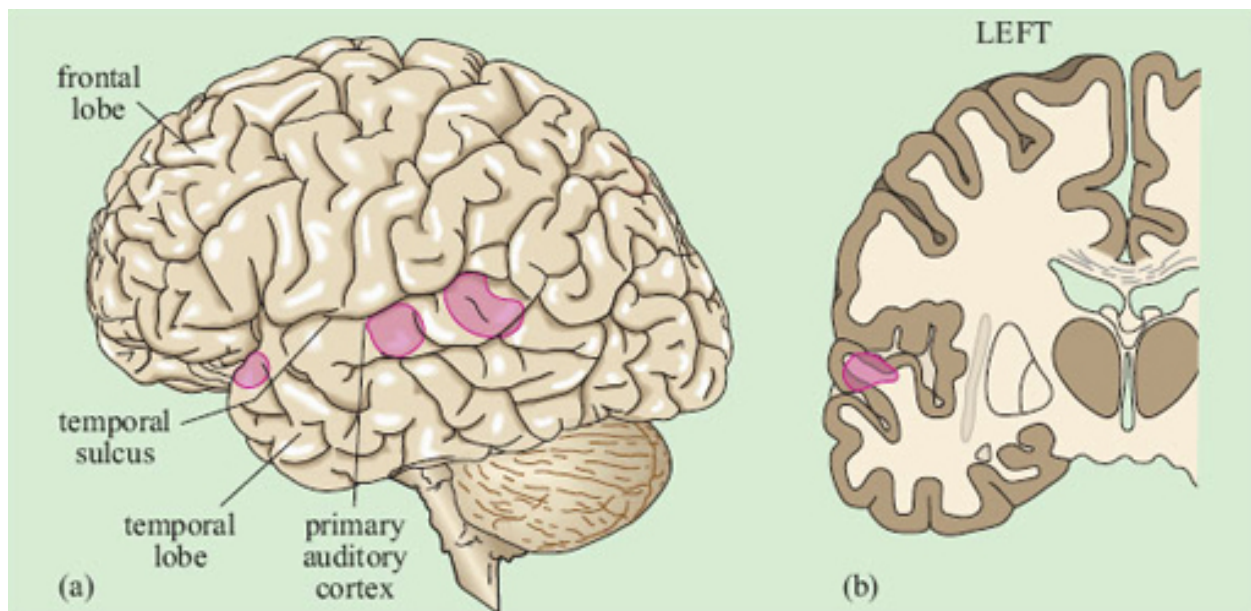


Figure 14 Cortical areas that respond preferentially to speech or scrambled speech sounds, as opposed to non-speech sounds. (a) Lateral view; (b) coronal section

Listening to speech produces activation on both sides of the brain in the auditory cortex. Damage to the superior temporal lobe on either side causes difficulties with speech recognition, though the pattern of the difficulties may be somewhat different on the two sides.

Thus it seems that the initial perception of speech is processed by some of the neurons in the auditory areas of the superior temporal lobe, on both sides of the brain. As we have seen, though, there is a great deal more to language than identifying phonemes and formants; similarly, there are many more brain areas which, when damaged, cause problems with speech and language. We must thus consider the architecture of the whole language processing system.

## 3.2 The anatomy of the language system

Perhaps the best-known generalisation about the language system is that it is represented on one side of the brain – usually the left – more than the other. Many lines of evidence support this view. Specific impairments to linguistic abilities are known as **aphasia**, and aphasia results much more often from damage to the left hemisphere of the brain than from damage to the right. It is also possible to temporarily deactivate one or other hemisphere. This is usually done as an investigative prelude to brain surgery, particularly in cases of epilepsy. It can be done by injecting a fast-acting drug into the carotid artery on one side of the body, or alternatively by weak electrical stimulation of part of the cortex. Deactivation of the left hemisphere causes a disruption of speech and language much more often than does deactivation of the right hemisphere. These techniques have also led to estimates that language is processed predominantly in the left hemisphere in about 97 per cent of right-handed people but in about only 60 per cent of left-handers.

We argued in the previous section that the initial perception of speech, like that of other sounds, was carried out bilaterally, whereas in this section we have argued that language is usually lateralised to the left hemisphere. There must thus be a point in the analysis of the speech signal where mechanisms on just one side ‘take over’ from the bilateral auditory areas. Brain scanning studies show that this is indeed the case: the additional activation in response to speech which is understood, compared to speech which is in an unknown language, is mainly on the left.

Further insight into such lateralisation comes from ‘split-brain’ patients. These are people who, because of severe epilepsy, have had the corpus callosum, the bundle of fibres connecting the two hemispheres, cut. This means that there is no direct neural connection between the left and right hemispheres. In these individuals, stimuli presented to the right hand, right ear, and right side of the visual field will mainly be processed by the left hemisphere, and stimuli presented to the left-hand side are processed by the right hemisphere. Objects presented to the left hemisphere can be named and talked about, whereas those presented to the right hemisphere generally cannot.

The right hemisphere does have some linguistic abilities. In split-brain patients the left hand can reach out and choose an object whose name has been presented to the right hemisphere, just as the right hand can choose an object whose name has been presented to the left hemisphere. However, the right hand can also choose the drawing that illustrates a sentence like *The girl kisses the boy*, whereas the left hand cannot correctly discriminate this situation from one in which the boy kisses the girl. Similarly, the right hemisphere cannot identify the scenes represented by *The dog jumps over the fence* versus *The dogs jump over the fence*, and so on. Thus, right-hemisphere language ability is limited to the phonological analysis of individual words, and access to their concrete meanings; sentence processing is the pure province of the left.

The areas within the left hemisphere which are most important for speech and language have been identified by several lines of evidence. Post-mortem examination of the brains



of people affected by aphasia identified two main areas which were most often damaged – one in the frontal lobe and one further back, in the temporal lobe, close to the junction of temporal and parietal lobes. These areas are known respectively as Broca's area and Wernicke's area after the scientists who first identified their associations with language. Because of their respective positions, they are sometimes also called the anterior and posterior language areas. Evidence from non-aphasic individuals generally supports the importance of these areas. In one procedure, the cerebral cortex can be weakly electrically stimulated as a prelude to brain surgery for epilepsy. The patient can be awake during this procedure – since the brain contains no nociceptive neurons, only local anaesthesia is required. Thus, the patient can perform a verbal task whilst the electrical stimulation is moved around different areas of the brain.

Electrical stimulation disrupts speech and language activities mainly when it is on the left of the brain, and only in specific areas (Figure 15). These areas are in the general regions of Broca's and Wernicke's areas, as predicted. There is also an additional area in the left superior frontal lobe. This is in part of the cortex generally associated with motor functions, and may be important for the execution of speech production more than for language processing per se. There is also evidence for the involvement of tissues further down into the temporal lobe than the extension of the classical Wernicke's area/posterior language area. (Subcortical structures appear to be important, too, but they are beyond the scope of this course.)

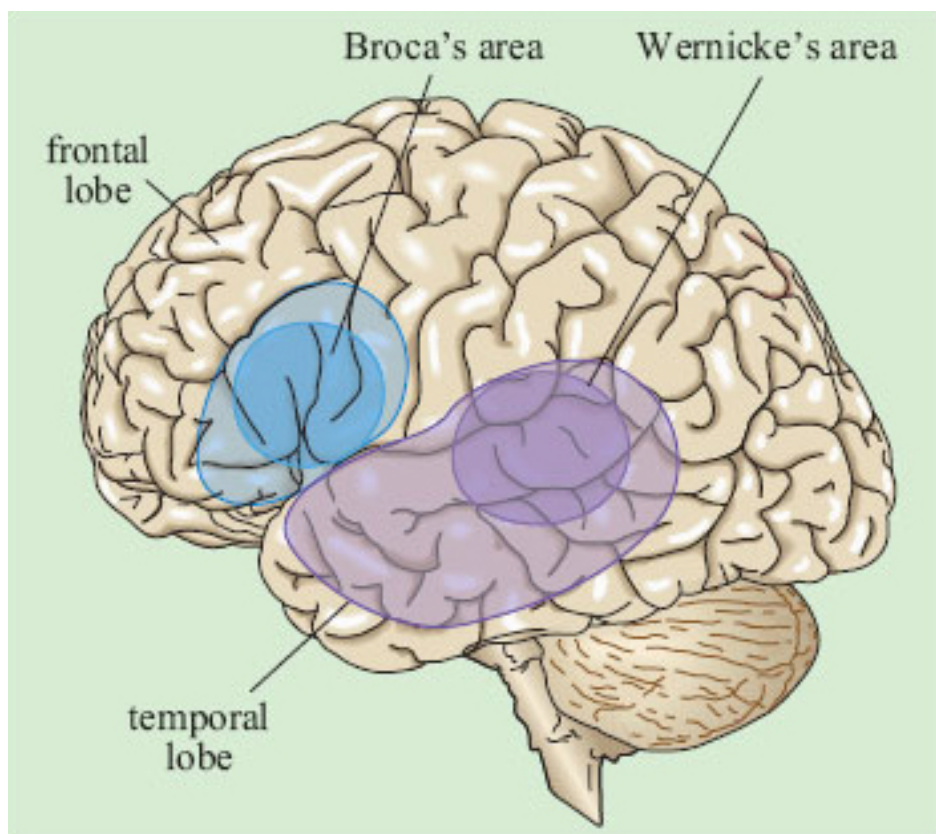


Figure 15 The lighter toned areas are the areas of the left hemisphere in which electrical stimulation disrupts speech and language. The two darker toned areas are Broca's and Wernicke's areas

Brain imaging studies support the map shown above in Figure 15. Tasks to do with speech and language lead to increased regional blood flow – which is suggestive of greater neuronal 'work' – across a broad range of structures in the left hemisphere,

generally within the triangle demarcated by Broca's and Wernicke's areas and the horn of the temporal lobe. There is strong evidence, which we will review in the next three sections, that this large area does not operate as an undifferentiated whole in the processing of language. Nor should we expect it to; as we saw in Section 2, decoding a sentence involves several distinct subtasks. The key questions to occupy researchers have thus been which subpart of the language cortex is associated with which subtask of overall language processing, and how does the processing of a sentence proceed in time through the various subareas and subtasks? The main lines of evidence we have to answer these questions come from three sources: aphasic patients, brain scanning and electrophysiological studies. We will look at each of these in turn in the next three sections.

### 3.3 Specialisation within language areas: aphasia

Aphasia is caused by localised brain damage, for example due to a stroke or an automobile accident. General intellectual functioning is not necessarily impaired, as the person can still perform non-linguistic tasks. Nor is the understanding and production of language necessarily completely abolished. Instead, there are highly specific patterns of impairment in the way language is processed.

Aphasia is divided into two main types, fluent and non-fluent. For reasons which will become apparent, they are also known as Wernicke's and Broca's aphasia. In non-fluent, or Broca's, aphasia the person has a marked problem with speech production. Speech rate is slow, and the articulation of speech is laboured and distorted. Speech lacks its normal intonation contours, being instead pronounced in rather a monotonic way. The typical feel of non-fluent aphasic speech is given by a sample from a patient in (15). The patient was attempting to explain that he had come to the hospital for dental surgery. The dots indicate long pauses.

(15) Yes ... ah ... Monday ... er ... Dad and Peter H ... [his own name] and Dad ... er ... hospital ... and ah ... Wednesday ... Wednesday nine o'clock ... and oh ... Thursday ... ten o'clock, ah doctors ... two ... an doctors ... and er ... teeth ... yah

#### SAQ 10

Describe as specifically as possible the ways in which this sample of non-fluent aphasic language is similar to and differs from normal language use.

#### Answer

In the sample, all the key referring words (the people involved, places and times, doctors, hospital and teeth) have been appropriately selected. Thus there is no apparent deficit in selecting the correct referring words on the basis of their meaning. These are all nouns, however; there are no verbs. More generally, there is no evidence of words being glued together into higher structures like sentences or phrases.

It was once thought that the comprehension abilities of people with non-fluent aphasia were normal. Thus, non-fluent aphasia was conceptualised as a disorder of language production that did not affect language comprehension. It is indeed the case that people with non-fluent aphasia can sometimes understand even quite complex sentences, in sharp contrast to what they produce. However, some subtle investigations over the last

few decades have shown that language comprehension is also affected, and thus that the disorder is one of language in general, rather than just production (Berndt et al., 1997). Consider (16).

- (16a) The dog chased the cat.
- (16b) The cat chased the dog.
- (16c) The girl watered the flowers.
- (16d) The flowers were watered by the girl.
- (16e) The dog was chased by the cat.

People with non-fluent aphasia can generally pick out the picture corresponding to (16a) or (16b) correctly. It might seem that the difference between (16a) and (16b) is purely syntactic; the words are after all the same in each case. Thus people with non-fluent aphasia seem able to use syntax in understanding if not in production. They can also correctly identify the picture associated with (16c) and (16d). Sentence (16d) is passive; this means that the usual generalisation that the first thing you come across is the subject and the second one the object does not hold in this case. Again, the fact that people with non-fluent aphasia often get the right meaning suggests that they are using syntax. However, in (16a) to (16c), you can get the right meaning simply by assuming the first thing you come across is the agent or doer. In (16d), you can guess the correct meaning simply from the individual meanings of the words present; flowers can't water a girl, so the correct meaning has to be the other way around. Thus, you could get (16a) to (16d) correct without really using any sophisticated syntax, but just by general reasoning from the words present.

Sentence (16e) is different. Using the principle of subject first is a red herring, since it would lead you to assume that the dog was the chaser not the chased; and, unlike (16d), either assignment of roles is plausible. The only way you can tell which meaning is intended in (16e) is by applying a syntactic analysis. People with non-fluent aphasia are very bad at sentences like (16e). They tend to get the roles the wrong way around. In a similar way, non-fluent aphasics are fine with (17a), but bad with (17b).

- (17a) The man who pushed the woman is old.
- (17b) The man who the woman pushed is old.

Once again, the explanation must be to do with syntax. In (17a), the subject of the verb *to push* is *the man*, which comes before the verb, so even though the sentence is moderately complex, the agent is coming first. In (17b), *the man* is the object of the verb *to push*; at some more logical level, the sentence means something like *The woman pushed the man and that man was old*. This logical meaning in (17b) can only be extracted by a syntactic analysis, because the linear order of the words is not the same as their logical order in the meaning.

### SAQ 11

What does the pattern of non-fluent aphasic comprehension deficit, coupled with the speech production data, suggest is going on in non-fluent aphasia?

#### Answer

It seems that people with non-fluent aphasia do not have access to the syntactic mechanisms. These are the processes which deduce the logical relations between



different words in a sentence from the form of a sentence. People with non-fluent aphasia are probably guessing at the meaning of sentences just from the individual words involved, and not from access to the syntactic structure of the sentence. For this reason, non-fluent aphasia has also been called *agrammatic aphasia*.

The brain damage in non-fluent aphasia was always thought to be in Broca's area in the left frontal lobe (see Figure 15). This generalisation is probably too simplistic; studies of many aphasic and other neurological patients show that damage to Broca's area does not always produce these symptoms, and that damage to other areas does sometimes produce them. Nonetheless, the most frequent source of these symptoms is damage to Broca's area and the most frequent result of damage to Broca's area is these symptoms.

### SAQ 12

From the aphasia evidence, what is the function of Broca's area?

#### Answer

Broca's area seems especially involved in the syntactic analysis of a sentence. (It may also have other functions to do with the production of speech, which is why articulation in Broca's patients tends to be distorted.)

In stark contrast to non-fluent aphasia is fluent aphasia. In fluent aphasia, the patient has very obvious comprehension problems. He makes semantic errors with words, such as pointing to his ankle when asked to point to his knee. By contrast, the articulation, intonation and fluency of speech sound normal. For this reason, fluent aphasia was initially thought of as a disorder of comprehension rather than of production. However, linguistic output is far from normal, as illustrated by (18), a fluent aphasic's description of a picture.

(18) Well this is ... mother is away here working her work o'here to get her better, but when she's looking, the two boys looking in the other part. One their small tile into her time here. She's working another time because she's getting, too.

### SAQ 13

Compare and contrast the language of (18) to the language of the non-fluent patient in (15).

#### Answer

Unlike (15), the fluent patient's output is in full sentences, with verbs as well as nouns. There is evidence of grammatical relations; for example, 'is working' is inflected to agree with 'mother'. The non-fluent sample was economically worded but informative, since all the key elements were specified. Here the speech is fluent but the information content is very low. This is because the words selected tend to lack concreteness and specificity of meaning ('working her work' rather than what she is actually doing; 'the other part' rather than the exact place). At other times, it seems that the wrong content words have been selected, and as a consequence the sentence is uninterpretable ('One their small tile into her time here').

## SAQ 14

From the above, what aspects of language processing would you guess are impaired in fluent aphasia?

## Answer

Syntax is generally unaffected. There seems to be a specific deficit in semantics. That is, there is trouble hooking the concrete meanings of words, especially nouns, to the objects to which they refer.

This view is reinforced by the fact that fluent aphasia can often be linked to another condition known as **anomia**. Anomia is a deficit in accessing the names of objects. It can take several forms, including the inability to give the name of an object when confronted with a picture of it, or the inability to find a name from a description or definition. One of the most striking discoveries about anomia is that the impairment can be specific to particular categories of things.

There are many patients who have trouble retrieving abstract names, like *supplication*, *pact* or *culture*, whilst still being able to find relatively concrete ones like *cheese* or *thimble*. This is perhaps not surprising; the abstract meanings tend to be rarer, more complex, and learned later in life, so they might be expected to be the first ones abolished by brain damage. What is striking is that there are also patients in whom the pattern of preservation is the other way around. Some examples of the definitions that this second group of patients volunteer are given below in Table 1.

**Table 1 Definitions of abstract and concrete words produced by patients with a superiority for abstract words**

Word	Meaning given
<i>Patient A.B.</i>	
Supplication	Making a serious request for help
Pact	Friendly agreement
Cabbage	Eat it
Geese	An animal but I've forgotten precisely
<i>Patient S. B.Y.</i>	
Malice	To show bad will against somebody
Caution	To be careful how you do something
Ink	Food – you put on top of food you are eating – a liquid
Cabbage	Use for eating, material – it's usually made from an animal
<i>Patient F.B.</i>	
Society	A large group of people who live in the same manner and share the same principles
Culture	A way to learn life's customs; it varies from country to country
Duck	A small animal with four legs

Thimble	We often say sewing thimble
---------	-----------------------------

(Adapted from E.M. Saffran and A. Sholl (1999))

These patients clearly have real trouble with the meanings of things that are based on concrete sensory (visual, touch or taste) properties, whereas they have no trouble with meanings that are expressible entirely in terms of other words. There are patients in whom the impairment is even more specific. There are individuals who have trouble finding the names of living things, but are fine with manufactured objects such as tools, and there are other patients with the opposite pattern.

How are we to interpret the loss of concrete word meaning in fluent aphasia and anomia? The best explanation is that there are neuronal circuits somewhere that carry the meanings of individual words by connecting the cells that respond to the phonological shape of that word to other cells which represent its non-linguistic attributes such as smell, colour, taste and so on. These circuits appear to be disrupted in fluent aphasia. In that disorder, the machinery of language is all working but it is not anchored to the concrete, non-linguistic meaning of nouns. These word-meaning circuits clearly have some functional specialisation within them, so that the area that represents living things is different from that which represents manufactured objects, which is different again from that which represents abstract concepts.

Damage in fluent aphasia is generally located in the vicinity of Wernicke's area, near the junction of the temporal and parietal lobes. Anomias for various types of nouns are generally a consequence of bilateral temporal lobe damage (Figure 16). From this we conclude that the posterior and temporal parts of the language areas are specialised towards the linking of particular words to their concrete meanings. The distinction between living and non-living things probably arises as a consequence of the different modalities we use to identify them. Living things are generally distinguished by their perceptual, particularly visual, properties. A 'thing' being an elephant is related to the way it looks. The circuits associated with anomias for living things are down in the temporal lobe, adjacent to circuits from the visual system that we know are involved in identifying objects. Non-living things, like tools, are more defined by what we do with them (a saw is for cutting wood). It is likely that the circuits for the meanings of these words are hooked up to motor circuits for performing the related actions.

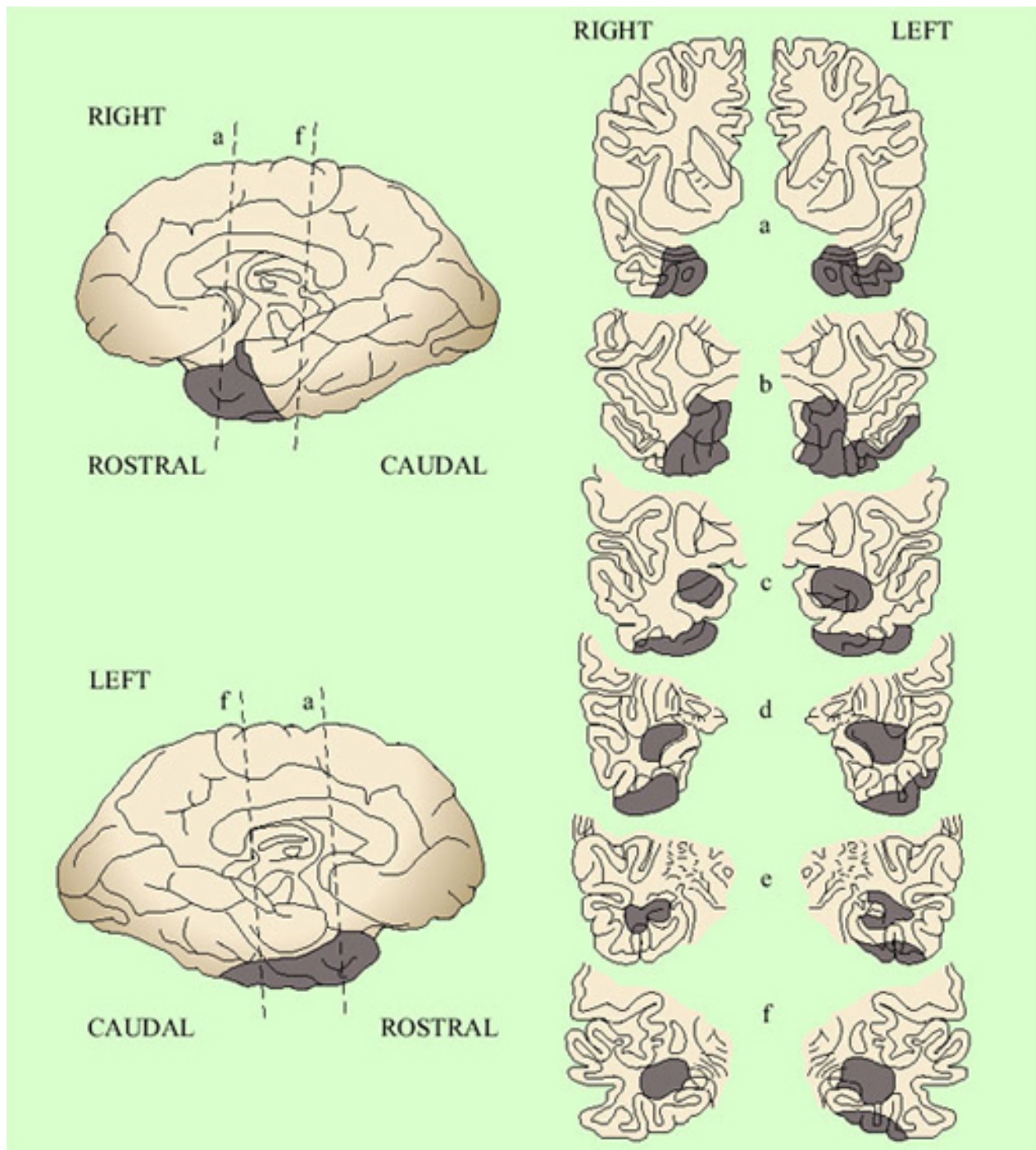


Figure 16 The areas of brain damage in an anomic patient (F.B.) who was more accurate with abstract than with concrete word meanings. The dark areas are the areas with cortical damage. The series of slices a–f on the right of the figure are coronal sections moving from anterior to posterior

Sirigu, A. et al. (1991) 'The role of sensorimotor experience in object recognition', *Brain*, Part VI, p. 2558, by permission of Oxford University Press

Sirigu, A. et al. (1991) 'The role of sensorimotor experience in object recognition', *Brain*, Part VI, p. 2558, by permission of Oxford University Press

Anomia, as described here, is a problem with nouns in particular. There are word-finding difficulties associated particularly with verbs, too. They are more particularly associated with frontal lobe damage and non-fluent aphasia. This is a satisfying finding. Verbs are much more 'syntactic' words than nouns. They do not have meaning outside of sentences

in quite the way a concrete noun can have, and they are intimately tied up with the creation of sentences, the assigning of subject and object roles and so on. Thus this finding generally supports the view outlined here that sentence processing has a more anterior basis, and individual word meaning has a more posterior one.

## 3.4 Specialisation within language areas: brain scanning

Is there any evidence from the undamaged brain that the view derived from aphasia is indeed correct? The most useful methodologies here use either PET or functional MRI (fMRI) scanning to establish which parts of the brain are active in particular tasks. The difficulty is that a standard linguistic task, such as understanding a sentence's meaning, involves phonology and syntax and semantics, and thus is not helpful when trying to tease out which of these subtasks happens in which areas.

Many studies have looked at the pattern of activation produced in the brain by single words. The areas especially active are widespread and somewhat variable, but generally include the auditory cortex on both sides, other parts of the left temporal lobe, and Wernicke's area. A study by Karin Stromswold from the Massachusetts Institute of Technology aimed to identify the areas specialised for the processing of syntax (Stromswold et al., 1996). Her team set up two different conditions of sentence processing. In one condition the participants heard sentences like *The child spilled the juice that stained the rug*, whereas in the other they heard sentences like *The juice that the child spilled stained the rug*. Both of these contain the same words. The first is syntactically quite simple because the order of the nouns in the sentence mirrors their logical relations (child spilled juice, juice stained rug). The second is more complex as the order of its elements does not reflect the logical relations. The areas of the brain specialised for syntax should be more active in the second condition than the first.

The most significant difference between the first and second conditions was indeed that Broca's area was much more active in the second (Figure 17). This finding confirms that of several other studies. Thus the view from aphasia seems confirmed; the anterior language areas are specialised for syntax (and verbs and sentence construction), whereas the posterior and temporal ones are more specialised for individual word meanings (and nouns and concreteness). This is doubtless a simplification. There is evidence of significant variability between individuals, and the distinction between areas and subtasks is not watertight. Moreover, we usually do all the subparts of linguistic processing interactively and simultaneously, so something that affects any one part will probably affect them all to a greater or lesser extent. Nonetheless, research in this area is allowing us to understand the anatomy of the language faculty in greater and greater detail.



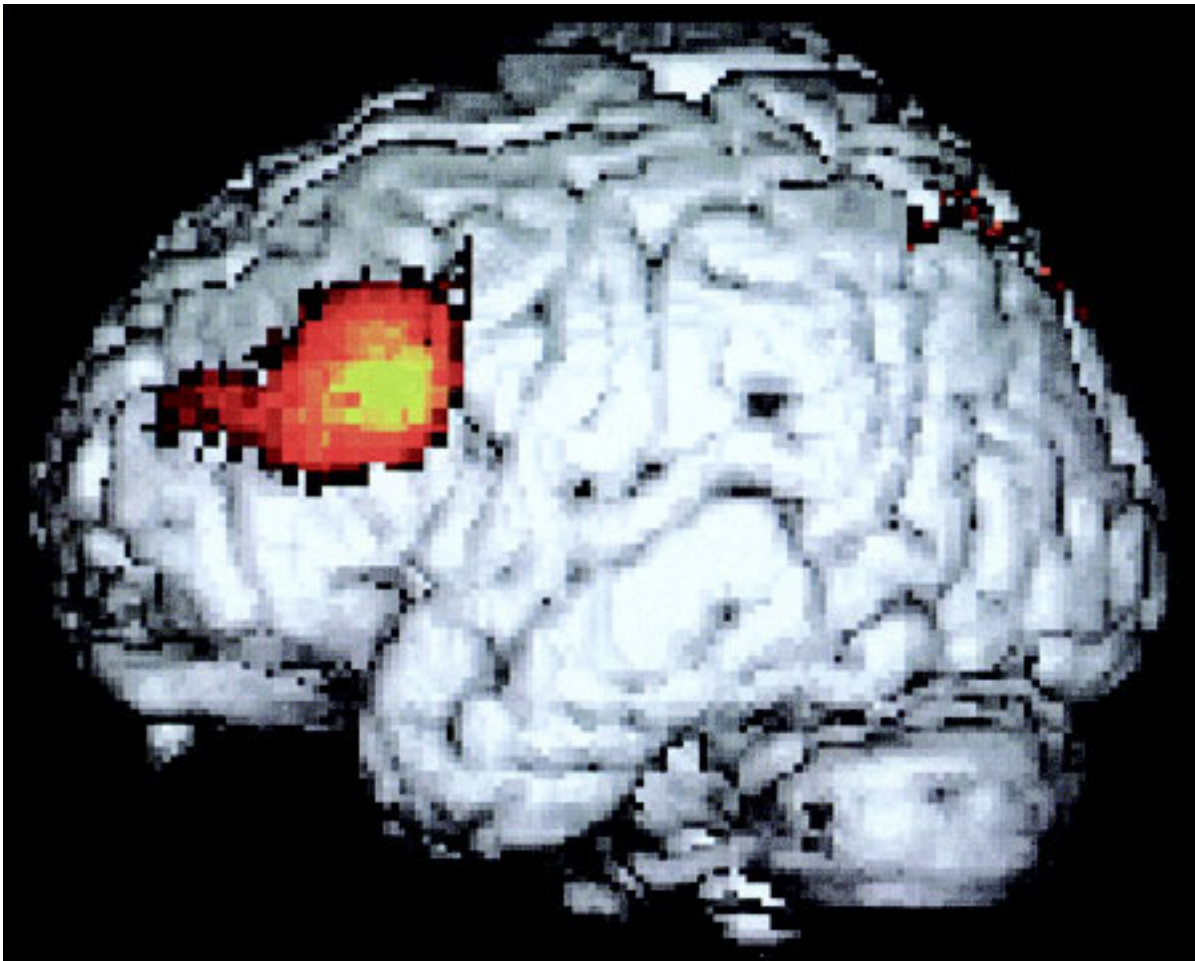


Figure 17 The area of greatest additional brain activity in a series of syntactic as opposed to non-syntactic language tasks imaged using PET scanning. Note that this area corresponds to Broca's area

Indefrey, P. et al. (1998) 'Specific response of the left inferior frontal sulcus to syntactic processing' (manuscript), by permission of Dr Peter Indefrey

Indefrey, P. et al. (1998) 'Specific response of the left inferior frontal sulcus to syntactic processing' (manuscript), by permission of Dr Peter Indefrey

## 3.5 Electrophysiological studies of language processing

Brain imaging and aphasic studies helped us localise the subparts of language processing within the brain. However, they have shed little light on how processing unfolds in real time. This is because contemporary brain imaging is quite poor at showing changes in activity through time in fine detail, so it is hard to pick up something that may be happening slightly before something else.

In Section 2, we identified several tasks that the language processor has to perform – phonological analysis, syntactic analysis, retrieval of word meaning and so on. We stressed that you cannot always complete one without reference to the others. For example, a signal which is phonologically ambiguous might be resolved by the context, or the rest of the sentence might determine which of several possible meanings should be

given to a particular word. But which happens first? Do they all take place in parallel? Do they each work independently, or is there cross-talk between them?

There is a general controversy within linguistics about just how much the different processing components interact. In one school of thought, for example, the identification of phonemes and word boundaries goes on autonomously, without access to the likely meanings of words or the context. This is a *modular* model, with processing going on in separate watertight subsystems. The alternative would be an *interactive* approach. In an interactive model, phoneme recognition and segmentation processes would already be influenced by information about what meanings are likely to be conveyed in the context. Even the modular model admits there must be some interaction between different processes. For example, in (7) (in Section 2.3), we rejected the second segmentation (7b) on the basis not of its phonological implausibility but because it doesn't mean anything. The difference between the two models lies in where they see the interaction happening. In an interactive account, considerations of meaning enter into the very process of identifying and segmenting the words, blocking (7b) and causing the processor to output (7a).

(7a) [my] [dads] [tu] [tor]

(7b) [mide] [ad] [stewt] [er]

In a modular account, phonological processing uses acoustic information only. It then outputs to the semantic and syntactic processors 'here are two possible segmentations of this signal' (i.e. 7a and 7b). The semantic and syntactic processors then accept one and suppress the other.

Rather similar issues arise with ambiguous words, as in (19):

(19a) The robber walked into the bank.

(19b) The canoe crashed into the bank.

It has long been known that if the word *bank* is flashed onto a screen, then for quite a long time afterwards, all the meanings associated with banks are primed in a person's brain. The classic way of testing this is what is called a **lexical decision task**. The participant is shown a string of letters and has to decide if they make up a word. The time it takes him to do this is recorded. People who have earlier seen the word *bank* on a screen are subsequently much quicker to decide that *money* is a word. They are also quicker to decide that *river* is a word than those who have not seen the word *bank* earlier. However, they are no quicker to decide that *giraffe* is a word. Thus, the word *bank* partially activates the whole realm of meanings related to it, such that you are then quicker to recognise other elements of that realm when you encounter them. The question, then, is whether in (19a), only the *money house* meaning is ever accessed for *bank*, or whether the *river* meaning is also accessed. If the former is the case, it would be evidence for an interactive account; elements of the context affect the processing of *bank* to the extent that the *river* meaning is never accessed. If the latter is the case, it would be evidence for a modular account; *bank* primes *river*, automatically, and without regard for the context.

Interestingly, what happens in this case is that *river* is primed for just a few seconds after presentation of (19a), but the priming then fades away rapidly. By contrast, *money* is primed immediately after presentation, and then stays primed. Sentence (19b) produces the opposite effect. This can be taken as evidence for a modular account; all the possible meanings of *bank* are primed, automatically, by the processing of *bank*. It is only subsequently that the integration of the word *bank* into the sentence context suppresses the other meanings.

This stage-like model is supported by direct recording of brain activity whilst linguistic processing is going on. This recording is done by placing electrodes over the scalp, and they pick up small changes in electrical activity caused by changes in neuronal activity in the brain. The participant is then given an experimental and a control task to do.

Differences between the patterns of electrical activity in the control and the experimental task can then be investigated. This technique is known as event-related potential (ERP) recording. It lacks the precision of localisation that is provided by brain imaging but it has the advantage of very fine temporal resolution, and thus is informative about the time course of the brain's response to a stimulus.

ERPs are hard to interpret without well-designed experiments. The classic ERP paradigm for studying language is to construct sentences that are for some reason hard to process, and compare the ERP trace of these sentences to that of ordinary ones (Friederici, 2002). For example, sentences where one word is semantically out of context and thus for which it is hard to find a meaning, produce a deviation in the activity trace across the middle of the brain about 400 milliseconds after the deviant word (Figure 18). This deviation is known as the N400. It is fairly consistent in two respects: it begins at around 400 milliseconds after the key word, and it is only provoked by *semantic* anomalies, that is sentences which are grammatically correct but weird in meaning.

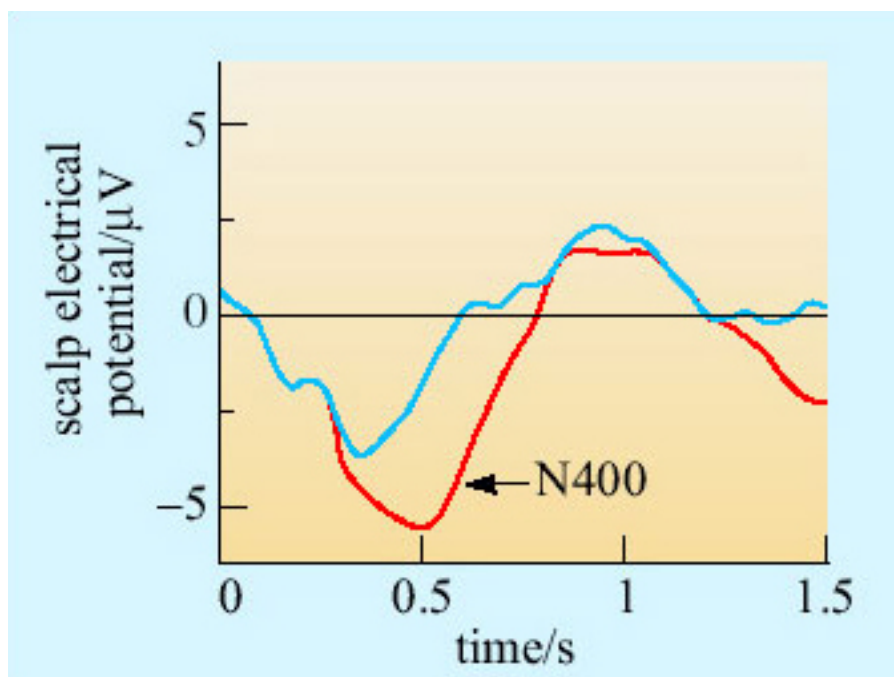


Figure 18 Event-related potential trace for *The shirt was ironed* (blue line) versus *The thunderstorm was ironed* (red line). The key word is *ironed*, and the deviation is called the N400

Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier

Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier

## SAQ 15

In Figure 18, which sentence is semantically normal, and which is semantically anomalous?

## Answer

*The shirt was ironed* is semantically normal. *The thunderstorm was ironed* is semantically anomalous. We do not expect thunderstorms to be ironed – indeed it is hard to see what this could mean. (*The thunderstorm was ironed* produces an N400 deviation.)

Sentences where the difficulty is syntactic provoke two kinds of deviation. An ungrammatical sentence, like *The shirt was on ironed* produces a deviation in activity over the left frontal lobe, often within a couple of hundred milliseconds of the word *ironed* (Figure 19). This deviation is called the LAN (left anterior negativity).

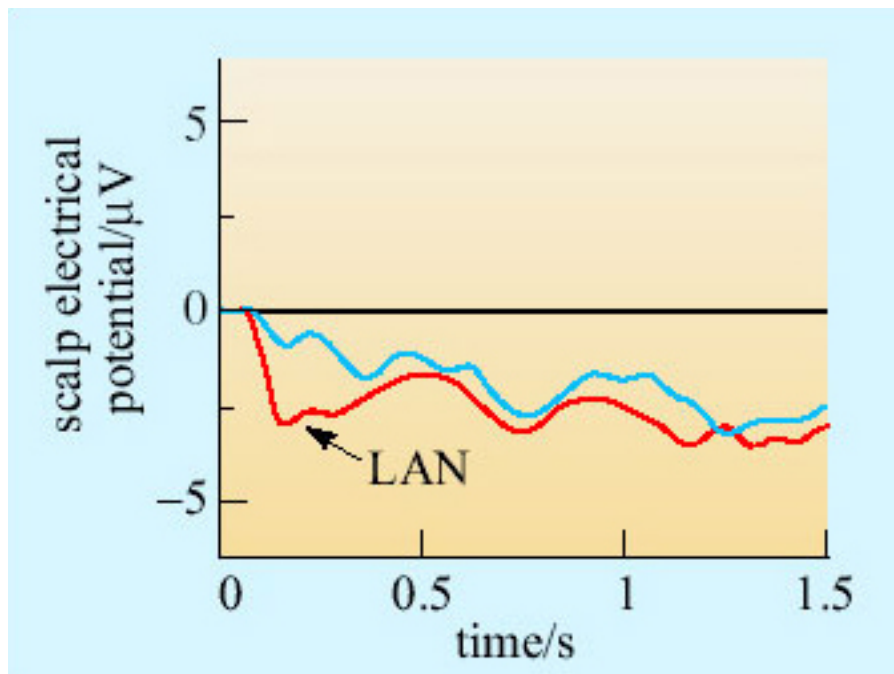


Figure 19 Anterior event-related potential trace for *The shirt was ironed* (blue line) versus *The shirt was on ironed* (red line). The key word is *ironed*, and the deviation is called the LAN

Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier

Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier

## SAQ 16

Why do you think the LAN is observed specifically over the left frontal lobe?

## Answer

This is the site of Broca's area, which we believe to have important syntactic functions.

A second deviation is also produced by syntactic complexities, somewhat later and more posterior in the brain. This deviation is called the P600, and typically peaks around 600 milliseconds after the key word (Figure 20). The P600 is produced by grammatical violations, and also by sentences that are syntactically tricky or ambiguous. Sentences like *The shirt was on ironed* produce a LAN and a P600. Other sentences, like *The horse raced past the barn fell* produce a P600 but no LAN.

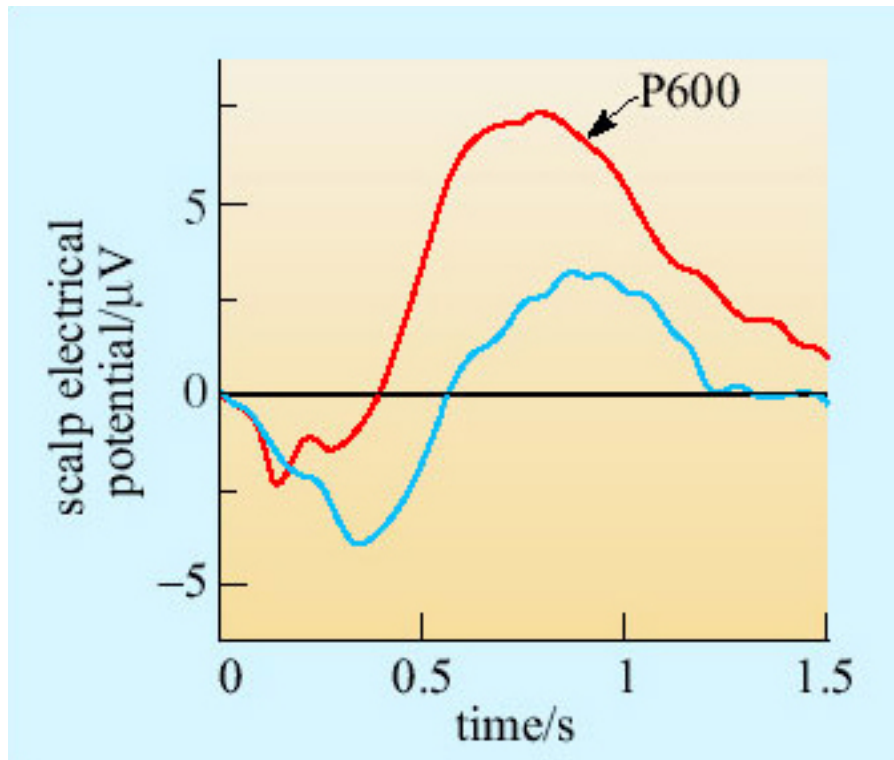


Figure 20 Posterior event-related potential trace for *The shirt was ironed* (blue line) versus *The shirt was on ironed* (red line). The key word is *ironed*, and the deviation is called the P600

Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier

Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier

### SAQ 17

What do you think the cause of the P600 might be, and how does this differ from the LAN?

#### Answer

Both the P600 and the LAN follow syntactic anomalies. The LAN is triggered by gross violations of syntax which are immediately obvious. In *The shirt was on ironed* there is no way that the string of words can ever be grammatical, since *on* must be followed with a noun phrase, and *ironed* is a verb. *The horse raced past the barn fell* is not ungrammatical. There is no violation and therefore the LAN is not produced. However, it is a difficult sentence to integrate conceptually; one tends to start off down one track



and then have to switch interpretations. The P600 reflects this extra cognitive work: the search for an appropriate conceptual representation of the sentence.

Note that the LAN happens more quickly than the response to the semantics of the words (the N400), and that the P600 occurs after the initial syntactic and semantic responses have occurred. This has been taken as evidence for a three-stage account of sentence processing. First, there is an initial syntactic analysis of the sentence, mainly in or around Broca's area. This happens within a couple of hundred milliseconds, and produces the LAN. It probably identifies the key lexical elements, whose individual meanings must then be processed.

Semantic processing of the key components of the sentence then follows. This happens further back in the brain, and produces the N400. Finally, the structure and meaning of the sentence is established by integrating semantic and syntactic information. This is straightforward for simple sentences, but where the words or structure are ambiguous, it may mean revising the initial syntactic analysis or suppressing alternative meanings. This is probably the stage at which the unwanted senses of *bank* are suppressed in (19). It is the stage at which you decide whether sentences like *They are hunting dogs* is a description of dogs or a description of what the hunters are doing. The extra work that may be required at this stage is reflected in the P600 which is produced by complex and ambiguous sentences.

Neat evidence for this three-stage model comes from the response to doubly anomalous sentences, like (20).

(20) The thunderstorm was in ironed.

### SAQ 18

What ERP responses might you expect that (20) would produce?

#### Answer

You would expect a LAN, because the sentence is grossly ungrammatical. You would also expect an N400, because of the semantic anomaly of *thunderstorm ... ironed*.

What is observed in response to (20) is in fact a LAN, but no N400. This means that the sentence gets blocked at the initial syntactic analysis stage, and is not 'passed on' to the semantic system for meaning analysis.

Such evidence supports a stage-like and somewhat modular account of sentence processing. First the phonology makes available a representation of the sentence. Then an initial syntactic analysis (in the left frontal lobe) assigns a basic structure to the sentence, and then passes on the key words for processing the meaning. The semantic system then makes available the possible meanings. This all happens within half a second. Finally, there is a conceptual integration of syntax and semantics, starting at about 600 milliseconds. Note that the sentence itself must be kept available in a buffer or short-term memory all this time, in case the parsing proves wrong and a re-analysis is necessary.

## 3.6 Summary of Section 3

Sound waves received by the ear are turned into neural activity by a complex mechanism involving the eardrum, the bones in the middle ear, and the hair cells within the cochlea. The auditory nerve carries the signal from the ears to the brainstem, from where it passes via the thalamus to the auditory areas of the cerebral cortex. In the cortex, speech sounds are extracted from the incoming signal. There are neural circuits in the auditory cortex that are specialised for speech and language as opposed to other types of sound.

Production and perception of speech takes place predominantly in one hemisphere of the brain, usually the left. Several areas within the left hemisphere are involved. Broca's area, in the frontal lobe, seems to be crucial for syntactic operations in both production and perception of speech. Wernicke's area, in the temporal lobe, seems to be crucial for accessing the concrete meanings of words. The evidence for the distinction in function between posterior and anterior language areas comes from the study of aphasia, that is, problems with language resulting from brain injury, as well as brain scanning.

Evidence from electrophysiological recording suggests that decoding a sentence involves several stages. First, there is an initial syntactic analysis of the structure of the sentence. Then the meanings of the individual words are accessed. Finally, the meanings of the individual words and the structure of the sentence are integrated to produce a coherent overall meaning. This all happens within one second of the final word being uttered.

## 4 Conclusions

### Activity 3

Read back over Section 3. Make two columns on a piece of paper, one headed 'finding' and one headed 'evidence'. Make a list of key findings we have established about the processing of language in the brain. For example, your first finding might be 'Language localised to the left', and the evidence might be 'aphasia, brain scanning', and so on. The findings might be of two sorts: the evidence of locations (i.e. where things happen), and the evidence of processes (i.e. the sequence in which they happen).

The first half of this course showed just how complex human language is; in the second half, we have established some preliminary findings about how the brain masters it. We clearly have a lot of finely structured special neural machinery to pull off this complex task. We have hardly touched on where this machinery comes from (though see Box 2). Scientists have only just scratched the surface of language. It has been hard to identify the relevant processes in the brain because the precise identification of neural activity in space and time is something we have only just started being able to do. Moreover, just to achieve a correct characterisation of the behaviour whose neural bases we are seeking took a long time.

## Box 2: The language instinct?

We have seen in this course that there is a great deal of apparently specialised neural machinery for language, from language-specific areas in the auditory cortex to Broca's area and its role in syntax. Where does this specialised machinery come from? Is it the effect of frequent practice from an early age that makes these structures become dedicated to language?

An influential school of thought in linguistics argues that the capacity for language is innate. Of course, this does not mean that English, German or Swahili is encoded genetically. Rather, the acquisition of a language is a genetically initiated and genetically guided process. It certainly seems like there is something in this view. The acquisition of language unfolds in a similar pattern across different cultures. It does not seem to matter whether parents attempt to explicitly teach their children or not; as long as there is some linguistic input around, children will acquire a language. At peak rate, children acquire something like one new word per hour. The linguist Noam Chomsky has argued for decades that this process is much more like growth in stature than learning to play the piano; as long as there is adequate nutrition, it will just happen because it is programmed to do so.

The main argument for Chomsky's position has always been logical rather than empirical. The child has to infer a grammar, quickly, on the basis of the sentences flying around above his head. These sentences contain positive examples of what a grammatical sentence is, but no negative evidence. Parents do not say *\*the cat sitting is mat the on* followed by a small electric shock to condition children away from ungrammatical sentences. Rather, the child has to figure out the grammar with no examples of what sort of things it does *not* allow. Chomsky has argued that it would be impossible to do this without fairly strong innate principles to guide the process (for example, the innate knowledge that the language will contain nouns and verbs, subjects and objects, put constraints on what sequences of sounds are possible, etc.). The data alone are not enough to allow the acquisition of language; you need some guiding principles already laid down.

The exact content of this 'start-up pack' is very hard to determine. So too is the adequacy of Chomsky's argument, which is called the **argument from the poverty of the stimulus**. We simply do not know nearly enough about the overall architecture of cognition, or the mental representation of language, to know how much needs to be specified beforehand and how much will just emerge in interaction with the environment and other cognitive systems. However, it does seem fairly clear that language is a species-specific behaviour, and thus part of our biology and evolution rather than our cultural history.

Interesting light has been shed recently on the innateness debate by the study of a genetic disorder called **specific language impairment (SLI)**. Affected individuals have rather specific problems with grammar, like getting the inflection right in *The man wants to go* versus *\*The man want to go*. In some families, the disorder is controlled by a single gene, called FOXP2 (Lai et al., 2001), and all the affected individuals in these families carry a particular allele that is not present in other families. Intriguingly, FOXP2 is present in other primates but the gene there differs from that in humans. Is FOXP2 one of the genes encoding Chomsky's innate linguistic principles?

The story seems unlikely to be a simple one. It has recently been shown that people affected by SLI have a wide range of other difficulties, from IQ scores to complex facial movements. If the gene involved is a gene for grammar, it is also a gene involved in other things too.

There are still many unanswered questions. For example, the binding problem is the question of how the meaning *green* gets stuck to the meaning *dog*, and not something else, in (14c). Our best guess about how this happens neurally is as follows. There is a localised circuit which encodes the meaning of *green*, and this starts firing in response to a sentence with the word *green* in it. There is another circuit for *dog*, which fires when the word *dog* is present. If the syntax says that these two meanings are bound together in the particular sentence, then their firing becomes synchronous. If they are in the same sentence but not bound together, they both fire, but asynchronously.

Now this account may very well be correct, but it poses an obvious problem. How then do we process the meaning of (21)?

(21) The green cat is beside the blue cat, not the green dog.

The *green* circuit would have to be synchronous with both *cat* and *dog*. The *cat* circuit would have to be synchronous with both *green* and *blue*. Yet the meaning of the sentence is not a cat that is both green and blue. The bindings are kept separate. We have no real idea how this is done. This is the wonderful thing about human language. Like the pattern of an ice crystal, the closer you probe it, the more complexities of structure there are. The adventure is only just beginning ...

## 5 Questions and answers

### Question 1

Define each of the following: grammar, phonology, syntax, semantics, noun, verb, subject, object.

### Answer

Grammar: The set of unconscious rules or principles that the speaker of a human language brings to bear when producing or understanding speech.

Phonology: The part of grammar dealing with permissible combinations of sounds into words.

Syntax: The part of grammar concerned with the rules for legitimately combining words into sentences.

Semantics: The part of grammar dealing with the meanings of words and sentences.

Noun: A class of word that can take the role of the subject of a sentence. A noun usually denotes an object or person, though this is not always the case.

Verb: A class of word that usually requires the presence of at least one noun to form a grammatical sentence. Verbs usually denote actions or processes.

Subject: The primary grammatical role a noun may play in a sentence. The role of the subject may be as the doer or initiator of an action, or bearer of a quality.

Object: Any noun in a sentence other than the subject. It often denotes the target or recipient of an action.

## Question 2

For each of the following words, identify the vowels and the consonants which are produced when the word is spoken. Bear in mind that English spelling doesn't bear a close relationship to the way the language is spoken today. The first example is done for you. Note: The precise spelling you use is unimportant for present purposes – just try to spell it like it actually sounds. The point of the question is to identify phonological units. (These will be slightly different depending on which dialect of English you have.)

nightmare	nai	t	m	ee
	V	C	C	V
psychology				
subtle				
meaning				
monkey				
yearning				

## Answer

These answers are for the author's own Southern England dialect of English.

psychology	s	ai	k	o	l	o	dg	ee
	C	V	C	V	C	V	C	V
subtle	s	u	t	u <sup>1</sup>	l			
	C	V	C	V	C			
meaning	m	ee	n	i	ng <sup>2</sup>			
	C	V	C	V	C			
monkey	m	u	n	k	ee			
or	m	u	ng	k	ee			
	C	V	C	V	C			
yearning	y	e <sup>3</sup>	n	i	ng <sup>2</sup>			
	C	V	C	V	C			

<sup>1</sup> This is a sound, very common in English, called schwa, which has no consistent representation in our spelling.

<sup>2</sup> This is pronounced as one sound.

<sup>3</sup> This is a difficult vowel to capture using the ordinary alphabet, but it is a single sound.

Did you find yourself tempted to put in letters because they are there in the spelling, only to find on careful thought that they are not in the spoken form at all?



### Question 3

Use the information you noted down as part of Activity 1 to fill in the following table comparing the vocal communication systems of vervet monkeys with human language, using the options on the right:

	Vervet calls	Human language	
Number of meanings			large, small
Generativity			present, absent
Elements can be combined			yes, no
Meaning depends on context			yes, no
Syntax			present, absent
Noun/verb distinction			present, absent
Acoustic distinctiveness of different signals			great, small

### Answer

	Vervet calls	Human language
Number of meanings	small	large
Generativity	absent	present
Elements can be combined	no	yes
Meaning depends on context	no	yes
Syntax	absent	present
Noun/verb distinction	absent	present
Acoustic distinctiveness of different signals	great	small

### Question 4

Explain in your own terms why each of the following sentences might be difficult for people to process when heard aloud:

- 1 The dog it was that the cat chased.
- 2 The boy the man the woman knew liked went away.
- 3 The horse raced past the barn fell.

### Answer

- 1 The first noun you come to is not actually the subject of the sentence. The subject is the cat, and the object is the dog, but they appear in the opposite order.

- 2 This is what is known as a centre-embedded sentence. This means that the 'went away' actually binds with 'the boy', but there is other material embedded in the middle which makes this difficult to see.
- 3 This is what is called a 'garden-path' sentence. As you read it, you at first think it means that the horse raced past the barn. It actually can only mean that 'the horse [which was] raced past the barn fell' but this only becomes clear when you reach the 'fell'. This means you have to revise your sketch of what the structure of the sentence is. It is normal to have to read it a few times before seeing that it was not the barn that fell, and the horse did not race, but was raced by someone else.

### Question 5

How do hair cells detect sound, and how is this signal converted into neural activity?

#### Answer

Hair cells have tiny hairs (cilia) on them. Vibrations from the air are transmitted into the cochlea, where they cause the hairs to move. Within the cell, movements of the cilia are converted into electrical impulses. Hair cells have synapses with neurons of the auditory nerve, and so their electrical changes become action potentials which are transmitted to the brain.

### Question 6

What are the symptoms of fluent aphasia? How do these differ from those of non-fluent aphasia?

#### Answer

In fluent aphasia, the sentences are complete and the grammatical relations are often normal. However, the information content is very low. The words chosen are either generalised or inappropriate in meaning. By contrast, in non-fluent aphasia, the right content words are selected but they are not strung together into full sentences.

### Question 7

How do we know that the capacity for language is localised predominantly in the left hemisphere?

#### Answer

There are several lines of evidence:

- 1 Brain damage to the left hemisphere causes aphasia more often than does damage to the right hemisphere.
- 2 Injection of a rapid-acting anaesthetic into the left carotid artery (which suppresses the left cerebral hemisphere) tends to cause loss of speech, whereas injection into the right carotid does so less often. A similar procedure can be done with electrical stimulation of the cortex.
- 3 'Split-brain' patients whose corpus callosum has been severed tend to be able to talk about the stimuli on the right side of their visual field or in their right hand. The right side is connected to the left hemisphere of the brain.

- 4 Brain imaging techniques such as PET scanning.

### Question 8

What are the three stages of sentence processing that have been identified by studies using ERPs? What is the name of the ERP component associated with each stage?

#### Answer

Phonology makes available a basic representation of the sentence.

- 1 An initial syntactic analysis assigns a basic structure to the sentence (starting within 200 ms). This stage is called LAN.
- 2 The meaning of key words is processed (normally within 400 ms). This stage is known as N400.
- 3 Syntax and semantics are integrated (normally within 600 ms). If they do not fit together, the sentence, which is still stored in short-term memory, is re-analysed. This stage is known as P600.

## Keep on learning



## Study another free course

There are more than **800 courses on OpenLearn** for you to choose from on a range of subjects.

Find out more about all our [free courses](#).

## Take your studies further

Find out more about studying with The Open University by [visiting our online prospectus](#).

If you are new to university study, you may be interested in our [Access Courses](#) or [Certificates](#).

## What's new from OpenLearn?

[Sign up to our newsletter](#) or view a sample.

For reference, full URLs to pages listed above:

OpenLearn – [www.open.edu/openlearn/free-courses](http://www.open.edu/openlearn/free-courses)

Visiting our online prospectus – [www.open.ac.uk/courses](http://www.open.ac.uk/courses)

Access Courses – [www.open.ac.uk/courses/do-it/access](http://www.open.ac.uk/courses/do-it/access)

Certificates – [www.open.ac.uk/courses/certificates-he](http://www.open.ac.uk/courses/certificates-he)

Newsletter –

[www.open.edu/openlearn/about-openlearn/subscribe-the-openlearn-newsletter](http://www.open.edu/openlearn/about-openlearn/subscribe-the-openlearn-newsletter)

## Acknowledgements

The content acknowledged below is Proprietary (see [terms and conditions](#)). This content is made available under a

[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Licence](#)

Grateful acknowledgement is made to the following sources for permission to reproduce material in this course:

Course image: [Thomas Rousing](#) in Flickr made available under [Creative Commons Attribution 2.0 Licence](#).

The content acknowledged below is Proprietary and is used under licence.

Taken from book 5 pages 49-81 and pages 83-85 of *Sound to meaning: Hearing, speech and Language*

Figure 1 Copyright © Tony Camacho/Science Photo Library;

Figure 3, 5, 6, 9 Courtesy of Daniel Nettle;

Figure 7 Holly, R. et al. (1997) 'Neurobiology of speech perception', *Annual Review of Neuroscience*, 20. Copyright © 1997 by Annual Reviews Inc. All rights reserved;

Figure 8 Peterson, G. E. and Barney, H. L. (1952) 'Control method used in a study of vowels', *The Journal of the Acoustical Society of America*, Vol. 24, No. 2, March 1952.

The Acoustical Society of America. Used by permission of the American Institute of Physics;

Figure 11 Copyright © VVG/Science Photo Library;

Figure 12 Martini et al. (2000, Figures 18–17, p. 483) in Toates, F. (2001) 'The other sensory systems', in *Biological Psychology*, Pearson Education;

Figure 16 Sirigu, A. et al. (1991) 'The role of sensorimotor experience in object recognition', *Brain*, Part VI, p. 2558, by permission of Oxford University Press;

Figure 17 Indefrey, P. et al. (1998) 'Specific response of the left inferior frontal sulcus to syntactic processing' (manuscript), by permission of Dr Peter Indefrey;

Figure 18, 19, 20 Reprinted from *Trends in Cognitive Sciences*, Vol. 6, No. 2, Friederici, A. D., 'Towards a neural basis of auditory sentence processing', p. 81, Copyright © 2002, Elsevier.

**Don't miss out:**

If reading this text has inspired you to learn more, you may be interested in joining the millions of people who discover our free learning resources and qualifications by visiting The Open University - [www.open.edu/openlearn/free-courses](http://www.open.edu/openlearn/free-courses)