Introduction

Aim of this chapter is to understand the unconscious (non-conscious?) operations that allow humans to process language.

Focus is on language perception - production is covered in chapter 7.

Covers:

(i) Models of recognition - of spoken and written words. These usually involve assuming we have access to a mental lexicon.

(ii) What the mental lexicon contains, and how it might be organised.

(iii) The process of comprehension beyond the mental lexicon - e.g. how grammar is used to construct sentence meaning.

Word recognition

English speakers - knowledge of between 50,000 and 100,000 words. Common words easy to describe and use; less common words represented more vaguely; unknown words may be capable of being pronounced accurately even without knowing what it means - e.g. tarantella.

Mental lexicon therefore stores information about word meanings and pronunciations. Recognising words means accessing the lexicon as quickly as possible.

Spoken word recognition

Speech is acquired by almost all humans, has been around long enough for some aspects to be considered innate. c.f. reading, which is new in evolutionary terms.

Speech is somewhat similar to written words - boundaries exist between words (though not

necessarily silences - different to 'white space' of written text in this regard.)

(i) Segmenting the speech stream

Two main models:

(i) **pre-lexical models**, which rely on characteristics in the speech stream that may indicate a boundary. The rhythm of speech is also important - the **metrical foot** in English, for example (strong syllable followed by zero or more weak syllables.) Most words that have a meaning in English are like this - grammatical words (of, it) do not - **Cutler and Carter**. Could therefore be used to segment words in English.

Cutler and Norris- support from playing pairs of nonsense syllables and asked participants to report any familiar words embedded in the speech (wordspotting task.) e.g. mintayve and mintesh - easier to spot mint in mintesh as it is a strong followed by a weak syllable - the first example is two strong syllables.

Cutler and Otake - French and Japanese have different rhythmic units but they were able to demonstrate similar results - suggests we 'tune in' early in life to our native language and so segment according to its rhythmic characteristics.

Not the whole story - a word like 'confess', which starts with a weak syllable would be mis-segmented if this was the only method we used. So there is a role for:

(ii) **lexical models** - boundaries are determined by our knowledge of how words sound - their phonological representation.

Such models require us to recognise each word and then predict the word boundary at the end of it. (**Marslen-Wilson and Welsh**.) If you are able to recognise the first word before it finishes, then you can predict where it ends - e.g. last syllable of confess = /s/ - so can predict the start off the next word. OK for long words - but problematic for very short ones lots of backtracking would be required. e.g. compare 'confess tomorrow or die!' with 'own up now or die!'

Cross linguistic differences in speech segmentation suggest the ability is learnt, rather than innate, and develop as we are exposed to language.

Saffran et al - demonstrated using a 'head turning' procedure to see how 8m.o. infants perceived an artificial language speech stream of three syllable nonsense words. The artificial language was played from one speaker; jumbled syllables from another.

Found infants could pick out 'words' from the stream of syllables in around 2 minutes - head turned more often after that time towards the 'novel' jumbled loudspeaker. The explanation is therefore that they are attending to the statistical information about the co-occurrence of syllables in the speech stream (as the artificial language was designed to ensure there were no acoustic of rhythmic cues in the speech stream.)

Ability to learn statistical information from patterns is universal - operates for adults too - and for non-speech stimuli such as tones and shapes - later work by **Saffran et al**.

Speech segmentation may therefore make use of an **implicit learning ability** - also shared by other primates (**Hauser et al**.)

(ii) Parallel activation

Words in speech have to be continuously evaluated against possible identities - a process known as parallel activation. The identity of a word can therefore be determined before the end of the word is heard.

Marslen-Wilson et al - cohort model as an explanation.

Start of a word activates the **word-initial cohort**.

As more of the word is heard, recognition is a process of eliminating potential matches until the **uniqueness point** is reached – a single word is identified.

However, the goal of word recognition is the ability to access our stored knowledge about meaning, which research indicates can occur earlier than this point.

Marslen-Wilson: cross-modal priming experiment.

A spoken prime word is heard, followed by a visual target word. Task is to decide as quickly as possible if the target is a real word. Semantic similarity between prime and target leads to faster recognition – e.g. Confess primes Sin. But also found confe still facilitates the recognition of sin, compared to unrelated primes and targets. Confe also primes wedding – suggesting meanings of confess and confetti are briefly accessed when the word confess is heard.

Also implies word meanings are accessed before we identify a word – ensures we have the relevant meaning to hand by the time identification takes place.

However, **Gaskell and Marslen-Wilson** have also shown that meaning activation is limited – also from cross-modal priming data. Many meanings activated at the same time shows a weak priming effect c.f. when only a few meanings are activated.

However, a key characteristic of speech perception seems to be access to too much, rather than too little information to improve the chances of the correct meaning being found as quickly as possible.

(iii) Lexical competition

Later alternatives to the cohort model use the activation and competition metaphor to describe the efficiency of recognition. [See ch4 – IAC].

Each word in the lexicon is associated with an **activation level** during recognition – reflecting the probability of it being that word. [Different to the cohort model as that is a simple member/not a member paradigm.]

The lexical competition model (e.g. TRACE – **McClelland and Elman**) uses continuously varying levels of activation to reflect the probability of a particular word being the one that needs to be recognised – sensible, as the information in the speech stream will match some words better than others.

TRACE – connectionist model. Three levels of representation:

Phonetic features level (bits of phonemes) -> **Phoneme level** -> **Word level** (one node per known word)

The speech stream is therefore changing patterns of activation at the phonetic features level.

As well as activation, inhibition occurs from the links between nodes at the word level – if a word is activated, it inhibits others – the competitive element of the model.

Activation and competition models are common to other models of language and cognition (e.g. face recognition.)

For speech recognition, it provides a subsidiary method of making sense of the speech stream – e.g. confess and fester overlap (1st syllable = 2nd syllable) – therefore both words are activated bottom-up – but only one will remain active through the lexical competition mechanism. Implies word segmentation can be solved implicitly – if confess wins, the boundary is at the end of the syllable 'fess' – if fester wins, the boundary is at the start of the syllable.

McQueen et al – wordspotting expt. supports this model. Looked at time taken to spot words like 'mess' in two types of sequences.

'duhmess' – first 2 syllables of domestic
'nuhmess' – no longer word possible

Detection rates slower for first example than second. It appears that lexical competition allows two key processes to occur simultaneously – i.e. word identification and segmentation.

Visual word recognition

Raises specific issues – e.g. how does out recognition system determine where our eyes should fixate and for how long.

(i) Models of visual word recognition

For speech recognition, TRACE was a development of the IAC model (**McClelland and Rumelhart**) – shares characteristics with the IAC model for face recognition [ch. 4].

Three levels of representation:

Visual features level -> Letters level -> Word level

Activation flows bottom-up and top-down.

Top-down activation can help explain lexical effects on lower-level processing. e.g word superiority effect – WSE. Letter detection is easier when the letter is part of a word (e.g. 'i' easier to detect in slim than spim.) 'Slim' provides a secondary source of activation for 'i' but as spim is not a word no secondary activation occurs in that case.

However – **Grainger and Jacobs** – demonstrated WSE could be explained using an IAC variant without top-down feedback. Proposed letter detection is based on two different levels of representation – activation of letter nodes **or** word nodes.

Controversy is whether top-down info. Is required. –

not only in the area of word recognition (auditory and visual) but in other areas of perception.

(ii) Mappings between spelling and sound

There are links between the **orthography** (spelling) and the **phonology** (sound) of words.

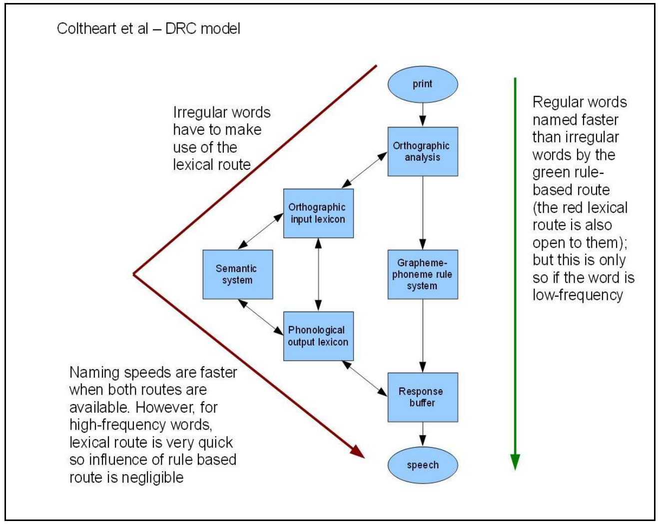
Reading aloud often characterised in terms of two different mechanisms:

Assembled phonology - pronunciation is based on a set of mappings between letters and sounds - e.g. ‘b’ in bell = /b/ phoneme. Works well for regular words.

Addressed phonology - pronunciation relies on stored information on the whole word in the mental lexicon. Works for irregular pronunciations - e.g. pint.

[**Pseudohomophones** - non-words pronounced to sound like real words - e.g. Brane, noyz - experimental / rock lyric use!]

The two different mechanisms are represented in dual-route models of reading - e.g. **Coltheart et al** DRC model.



Glushko - properties of neighbouring words (e.g. similar spellings) impact recognition rates as well as word regularity.

e.g. wade is regular and neighbouring words with the same final letters are also consistently pronounced (made, jade, spade …) => fast recognition.

c.f. wave is regular but neighbouring words have inconsistent pronunciations (have, slave don’t rhyme.) => slower recognition.

Finding implies simple dual route models like DRC are incomplete as the same rules are applied to regular items irrespective of if neighbouring words are regular or irregular.

Seidenberg and McClelland - proposed a single connectionist n/w could be produced to account for both regularity and neighbour properties.

Van Orden - phonological representations are present in silent reading (as well as when reading aloud.) - even if they adversely impact performance.

Visually presented words e.g. rose - asked if they were a category member (e.g. is a rose a flower.) Participants found it difficult to reject homophones - e.g. rows and also pseudo-homophones - e.g. roze.

Implies the pronunciation was being activated - even in silent reading. Similar findings for other languages.

Possibly because as speech has innate characteristics, it makes sense for the visual recognition system to ‘latch onto’ it. **Frost** - does the phonological system get bypassed in later life as reading becomes more skilled and less effortful?

(iii) Eye movements in reading

Speech perception mostly passive; reading is more active - need to direct eyes to do it. Eye tracking therefore a useful technique to aid understanding.

Eye movements when reading consist of **saccades**, followed by **fixations** (typically 200ms), in which information is processed.

Rayner and Duffy - fixation duration dependent on frequency of word usage => fixations can be used as a measure of processing difficulty.

As well as eye tracking as an aid for understanding how we read, **Tanenhaus et al** used it to see how we interpret a visual scene when we hear a spoken sentence. e.g. Candy and candle present - can use eye tracking to correlate at what point someone looks at the candle when asked to pick up the candy.

Movement of eyes across a page is not linear - some words get skipped, others need multiple fixations. Regressive saccades are used. Function words (of, it) more likely to be skipped than content words; regressions tell us if a word has been misinterpreted - e.g. because of ambiguity. (**Starr and Rayner**).

O’Regan and Jacobs - words identified most quickly if fixated at their **optimal viewing point** (OVP). Usually in the middle of a word or slightly to the left for longer words. Near the middle makes biological sense - visual acuity is best in the foveal region of the retina.

Shilcock et al argued the bias to left of centre is due to the relative informativeness of the left hand side of words. Shorter words (e.g. it) have an OVP outside of the word - perhaps explaining why we tend not to fixate on them.

McConkie and Rayner - moving window over text on a computer screen that shifts with participant’s gaze. If window is small, reading is difficult; if the window is larger, reading is largely unaffected. => information is also retrieved from the parafoveal region - wider, but reduced acuity.

Also suggests in English perceptual span is limited - 15 chars to right of fixation; 3 chars to the left. In right to left languages (e.g. Hebrew) asymmetry is swapped.

The mental lexicon

Identification is the first step to understanding a word - therefore need to also understand how word meaning is accessed during recognition.

Definitions:

Semantic content - how word meanings are stored

Semantic organisation - how meanings are related

Morphology

= size of units in the mental lexicon - it may not necessarily be a whole word, as these can be divided into **morphemes** (the smallest meaningful unit in a word.) This could therefore be the real basic unit - e.g. Turkish - long words - so a lexicon based on morphemes would be more efficient and smaller.

Inflectional change (e.g. adding the plural ending to a word) is known as **inflectional morphology**. Major modifications - e.g. adding -ness to an adjective to change it into a noun or -ly to turn and adjective into an adverb = **derivational morphology**.

Irregularities exist in both morphological changes (e.g. irregular plurals such as mouse to mice) and the relationship between morphemes and the whole word may not always be deterministic of meaning (e.g. as in depart and department.)

Two cognitive approaches to recognition:

Full-listing approach - words made of many morphemes recognised in the same way as single morpheme words; word is the basic unit of the mental lexicon.

Decompositional (**Taft and Forster**) - words decomposed to morphemes as perceived; morpheme is the basic unit of the mental lexicon.

Marslen-Wilson et al - priming method used to see if morphemic units exist in the mental lexicon. Rationale - if words represented as morphemes a priming effect should occur between words containing the same morpheme. Finding was that priming only occurred if there was some shared meaning between the words - e.g. cruelty primes cruel but casualty did not prime casual.

Suggests the two extreme positions aren't correct - and that a more pragmatic view of the mental lexicon has greater support.

Accessing word meanings

(i) Semantic representations

Models of language perception become less well defined at the point where a word is recognised and a meaning needs to be retrieved - as meanings vary dependent on context and on the person.

Two theories are popular:

(a) **Spreading activation** models - e.g. **Collins and Loftus** - words represented by nodes (as in TRACE/IAC), but the difference is the links between nodes represent semantic relationships. Their original model used different types of links to express relationships e.g. canary 'is a' bird; canary 'has' wings. (Other models use unlabeled links to connect similar words together.) Once a word is recognised, activation spreads to the semantic network to generate a set of facts about a word.

(b) **Featural** theory - words meanings are represented as a set of **semantic features**. Argues the mental lexicon contains many features; each word representation contains a subset of these. e.g. features for canary include 'has wings', 'can fly' etc.

Can be incorporated into connectionist models - enabling recognition models and semantic representations to be linked. Activation of a written/ spoken representation activates a set of semantic nodes (node = semantic feature.) - e.g. **Masson**

The theories can be investigated using techniques such as semantic priming.

Most robust effects found on pairs of **associated words** - the strength of association is often measured by asking someone to say or write down the first word that comes into mind.

Nelson et al - cheddar -> cheese (90%); swiss (3%)

Problem with this method is variability of results - e.g. near synonyms (portion -> part); antonyms (gain -> lose); context related (law -> break)

Researcher therefore look at weakly associated words that still have some semantic link - e.g. horse and sheep.

Lucas, metastudy, concludes non-associative priming effects are robust.

Kellenbach et al - investigation of words linked by visual appearance - e.g. button and coin. Two measures of priming:

(a) standard reaction time test
(b) Event Related Potential (ERP) technique

No effect found for (a) but a robust effect for (b) => even if semantic link too weak to be detected conventionally, one still exists. Perception of a word therefore activates semantic information.

(ii) Semantic ambiguity

Resolving ambiguity (e.g. the word 'bank'). Two opposing views:

(a) **Autonomous** view - all meanings of an ambiguous word are accessed and then the contextually compatible one is selected.

(b) **Interactive** view - some inappropriate meanings are ruled out before they are fully accessed.

Swinney - cross modal semantic priming expt - evidence for the autonomous view. Participants heard homonyms (like 'bugs') in unbiased and biased contexts and were asked to make a lexical decision between 3 alternatives - 2 meanings of the word and an unrelated control word. Regardless of bias, he found both target meanings were primed.

If delay between sentence and targets was around 1s, only the contextually appropriate meaning appeared to be activated - suggesting the window where ambiguous meanings can be access is less than this.

Lucas - metastudy - more priming is shown for appropriate meanings than inappropriate ones - so would favour the interactive view.

Sentence comprehension

Sentences are nearly always novel - implies perception at this level is not just about recognition but is instead a constructive process. Could be thought about as being a process of building a model of information to be communicated [ch. 12]. Each word has a particular function (grammatical or syntactical) in a sentence - the process of constructing sentences is known as **parsing**.

Syntax

Definition: Mutually agreed conventions for word order.

In a typical sentence, each word has a specific syntactic role - can be thought of as a hierarchy. This hierarchy follows phrase structure rules. Phrase structure grammar is a linguistic analysis of a sentence but it is possible to see how it may be applied to language processing.

Parsing is difficult, as the same word can be used in different contexts. E.g. 'spotted' - can be used as a verb or an adjective; 'yacht' - verb or noun. **Altmann** - more than 50 grammatically allowable interpretations of 'Time flies like an arrow' - they don't have to make sense - e.g. **Chomsky**'s 'Colourless green ideas sleep furiously'.

Different models of parsing cope with such ambiguities in different ways.

Models of parsing

Parsing could perhaps take place only at major syntactic boundaries - **delayed parsing** - (e.g. the end of a sentence), but current models tend to assume it happens **incrementally**, as such a strategy would maximise the availability of information for a response.

e.g. **Tyler and Marslen-Wilson** - use of ambiguous phrases like 'landing planes'. ('landing' can be a verb or an adjective, depending on context).

If parsing is delayed until a syntactic boundary reached, there should be no effect of preceding context about if the word following 'landing planes' should be 'is' or 'are'. Method was to listen to a sentence fragment such as 'If you're trained as a pilot, landing planes …' (prime) and present a visual response ('is' or 'are' in this case.)

Found speed of naming a response depended on the context of the preceding phrase - appropriate continuations named quickly, compared to inappropriate ones - a result incompatible with the delayed parsing hypothesis (no effect for appropriateness is incorporated into it.)

Garden path model - **Frazier** - is an incremental model. The parser makes a decision about the correct alternative to pursue based on syntactic information alone - 'garden path' as you can end up at a dead end if the decision you take is wrong, and so have to backtrack. **Bever** - 'The horse raced past the barn fell' is an example - 'raced' is not often used in this way.

Garden path model assumes parsing is serial and has an autonomous component - i.e. initial evaluation of a word's role in a sentence is based on syntactical factors alone.

A contrasting account is the constraint based model - e.g. **MacDonald et al**, which assumes parsing is parallel and interactive. In other words, parallel as more than one potential parse can be evaluated at the sane time (similar to the cohort model of word recognition). They are interactive accounts, as frequency and semantic plausibility are held to be able to influence parsing immediately.

The mental lexicon is also required in such models - by assuming that how words combine with other words is stored there. Parsing is therefore like putting a jigsaw puzzle together.

Is parsing autonomous?

Support for autonomous parsing (and **Frazier**'s garden path model) comes from **Ferreira and Clifton.** Looked at how people interpret verbs in phrases like 'the defendant examined …' Garden path interpretation suggests examined as a main verb - 'the defendant examined his hands' is more expected than it being used as a reduced relative clause - 'the defendant examined by the lawyer turned out to be unreliable'.

Eye tracking (fixation on region after the verb) showed people have difficulty with the second sentence - even when the meaning of the word preceding the ambiguous verb should have reduced the plausibility of the main verb interpretation - e.g. 'the *evidence* examined by the lawyer turned out to be unreliable'.

However, **Trueswell et al** found some of the contexts used by **Ferreira and Clifton** were not as constraining as the example above (e.g. 'the car towed …'). Similar eye-tracking experiment found more constraining semantic contexts lessened or eliminated the 'garden path' effect.

Constraints on parsing

Ambiguity in spoken sentences can be reduced by the intonations used – e.g. 'Jane hit the man with the hammer' could have the emphasis on 'Jane' or on 'the man with the hammer'. Listeners can and do make use of such information in helping to resolve ambiguity – **Warren**.

Trueswell demonstrated that the lexical frequency of how words are used (e.g. 'The horse raced past the barn fell' – past participle use of 'raced' is uncommon) also influences the way sentences are parsed – prior knowledge of the way words are normally used can influence parsing in cases of ambiguity.

Tanenhaus et al – visual information about the environment can be a source of information that can eliminate the garden path effect. Using eye tracking techniques and various visual contexts, the ambiguity from a sentence like 'put the apple on the towel in the box' is lessened – e.g. If the visual objects consisted of an apple on a towel and an empty towel, then eye tracking shows the empty towel is looked at as a possible destination for the apple. If, however, there is an apple on a towel, an apple on a napkin and an empty towel, then eye tracking shows that the empty towel is not looked at – 'the apple on the towel' is 'put in the box'.

Conclusion

Processes involved in word recognition and understanding sentences can be modelled in terms of competition between different possibilities – which we remain consciously unaware of most of the time.

There seems to be a significant level of linkage between subsystems – e.g. between perceiving a written word and the speech subsystem. Makes sense – as we need to minimise the amount of effort required to comprehend language to operate efficiently.