COSC451: Artificial Intelligence

Lecture 14: How 'surface language' is represented in the brain

Alistair Knott

Dept. of Computer Science, University of Otago

Recap

Assumptions:

- A syntactic theory must posit two syntactic representations for a sentence: a 'deep' underlying representation (LF) and a 'surface' representation (PF).
- The LF of a sentence describing a concrete event can be interpreted as a description of the process of 'internally replaying' the sequence of SM signals involved in experiencing that event (as stored in working memory).

I haven't said anything about how PF is represented in the brain, or how a child learns a mapping from LF to PF.
This is the topic for the next couple of lectures.

Neural representations of language

When we process language, what areas of the brain are involved?

Neural representations of language

When we process language, what areas of the brain are involved?

'Processing language' can involve interpretation or generation, in three different media.

	Medium		
	Speech	Writing	Sign language
Interpretation	hearing	vision (pattern recog)	vision (action recog)
Generation	articulation	hand/arm movements	

Neural representations of language

When we process language, what areas of the brain are involved?

'Processing language' can involve interpretation or generation, in three different media.

	Medium		
	Speech	Writing	Sign language
Interpretation	hearing	vision (pattern recog)	vision (action recog)
Generation	articulation	hand/arm movements	

I'll focus on speech.

Outline of today's lecture

1. Neural representations of phonological structure
2. Neural representations of words
3. The neural locus of syntactic processing

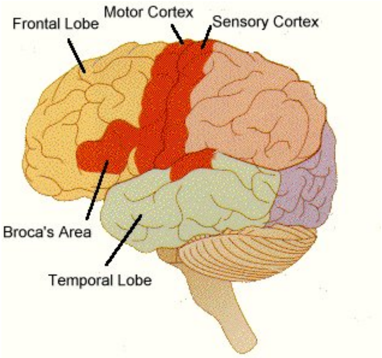
Outline of today's lecture

1. Neural representations of phonological structure
2. Neural representations of words
3. The neural locus of syntactic processing

Neural representations of phonological structure

What is a speech signal?

- If you're generating it, it's a sequence of articulatory gestures.
- If you're listening to it, it's an auditory signal.

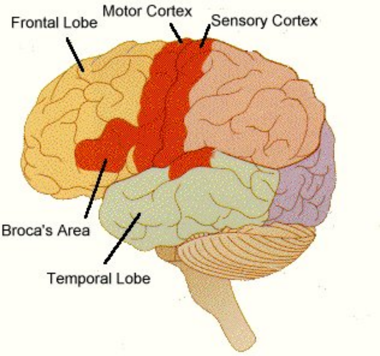


Neural representations of phonological structure

What is a speech signal?

- If you're generating it, it's a sequence of articulatory gestures.
- If you're listening to it, it's an auditory signal.

The endpoints for interpretation and production are auditory and motor cortex.

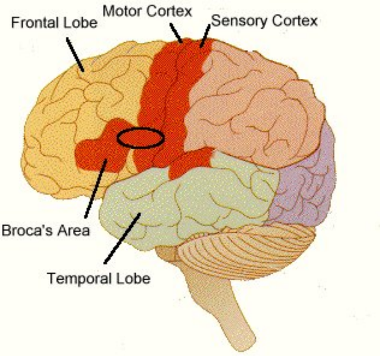


Neural representations of phonological structure

What is a speech signal?

- If you're generating it, it's a sequence of articulatory gestures.
- If you're listening to it, it's an auditory signal.

Articulatory gestures originate in the premotor/motor cortex.

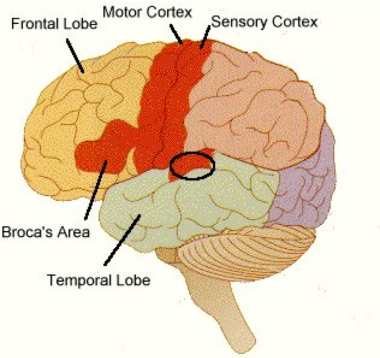


Neural representations of phonological structure

What is a speech signal?

- If you're generating it, it's a sequence of articulatory gestures.
- If you're listening to it, it's an auditory signal.

Auditory signals originate in the auditory cortex.



A mirror system for phonemes

Children need to work out equivalences between acoustic signals and articulatory signals.

- They need to learn to copy the sounds they hear.
- They need to learn to *speak* using *auditory* training signals.

There's evidence that children learn to represent speech sounds as articulatory actions.

Note the analogy with the mirror system for action recognition:

Action recog:	We represent visual signals as motor programmes
Speech recog:	We represent auditory signals as speech actions

Evidence for a mirror system for phonemes

1. Evidence from the syntactic discipline of phonology.
(Study of the sound system of language.)

Evidence for a mirror system for phonemes

1. Evidence from the syntactic discipline of phonology.
(Study of the sound system of language.)

- The 'units' of spoken language are called phonemes.
- It turns out there's no direct mapping from acoustic signals to phonemes. This is because of coarticulation: when we're producing one phoneme, we're also preparing the next one(s).

  happy vs hope
  lift vs goalkick
  take vs about one

- So what's a phoneme?
  What do all the instances of h/l/t have in common?

Evidence for a mirror system for phonemes

1. Evidence from the syntactic discipline of phonology.
(Study of the sound system of language.)

- Proposal: a phoneme is an articulatory gesture.
  - E.g. l is the gesture of touching the tip of the tongue behind the teeth…
  - E.g. t is the gesture of 'sealing' the tongue against the palate…
- We can recognise *intended articulatory gestures*, just as we can recognise intended hand/arm actions.

Evidence for a mirror system for phonemes

1. Evidence from the syntactic discipline of phonology.
(Study of the sound system of language.)

- Westermann and Miranda (2004): infants learn the relationship between intended articulatory gestures and acoustic signals by babbling.
  - They generate a random articulatory intention.
  - This results in a sound, which they hear.
  - They create (Hebbian) associations between the heard sound and the intention.
  - These associations allow them to recognise the articulatory intentions *of other speakers*.

Note the similarity with the model of how the mirror system for hand/arm actions is trained. (Called 'motor babbling')

Evidence for a mirror system for phonemes

2. Evidence from experiments.

Evidence for a mirror system for phonemes

2. Evidence from experiments.
   - Wilson *et al.* (2004) had subjects listen to or pronounce a nonsense syllable.
     - An area of ventral premotor cortex lit up in both cases.

Evidence for a mirror system for phonemes

2. Evidence from experiments.

- Wilson *et al.* (2004) had subjects listen to or pronounce a nonsense syllable.
  - An area of ventral premotor cortex lit up in both cases.
- Fadiga *et al.* (2002) had subjects listen to words whose pronunciation either required the tongue (*birra*) or did not (*baffo*).
  - They recorded activity from their tongue muscles (while TMS was applied to the motor cortex to amplify motor signals).
  - Hearing *birra* produced more tongue activity than hearing *baffo*.

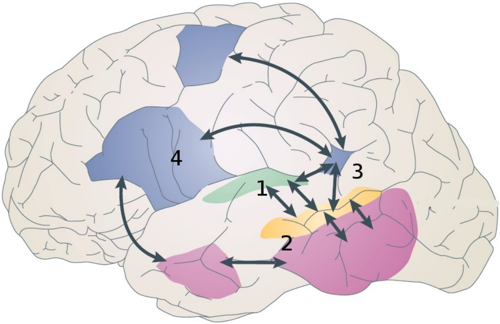
Evidence for a mirror system for phonemes

2. Evidence from experiments.

- Wilson *et al.* (2004) had subjects listen to or pronounce a nonsense syllable.
  - An area of ventral premotor cortex lit up in both cases.
- Fadiga *et al.* (2002) had subjects listen to words whose pronunciation either required the tongue (*birra*) or did not (*baffo*).
  - They recorded activity from their tongue muscles (while TMS was applied to the motor cortex to amplify motor signals).
  - Hearing *birra* produced more tongue activity than hearing *baffo*.
- Ito *et al.* (2009) used robotic devices to stretch the skin around subjects' mouths to mimic the gestures needed to pronounce *head* or *had*.
  - Auditory stimuli were delivered in synch with stretch signals.
  - Subjects' perception of these stimuli was biased in the direction of the imposed motor movements.

Pathways involved in the mirror system for phonemes

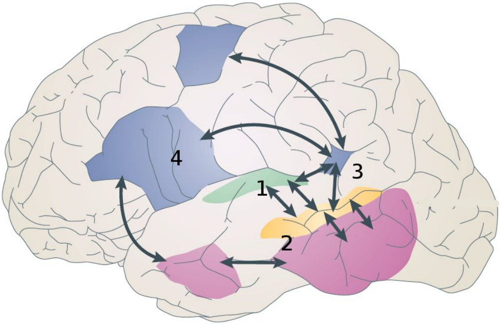
An influential model by Hickok and Poeppel (2007):



1. Auditory cortex: acoustic representations
2. Middle/posterior STS: phonological representations
3. Left temporoparietal junction / inferior parietal cortex
4. Broca's area (left premotor/motor cortex): articulatory plans

Pathways involved in the mirror system for phonemes

An influential model by Hickok and Poeppel (2007):

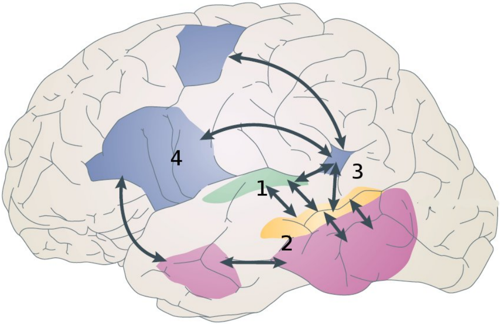


Note that this pathway is similar to the pathway for action recognition:

Action recog.	Visual cortex → STS → IP cortex → PM cortex
Speech recog.	Auditory cortex → STS → IP cortex → PM cortex

Pathways involved in the mirror system for phonemes

An influential model by Hickok and Poeppel (2007):



H&P suggest that this ‘mirror system’ route from hearing to speech is *purely phonological*: doesn’t involve accessing words/meanings.

(A separate ‘meaning-based’ route runs through temporal cortex.)

Evidence for a 'sublexical' route from heard to spoken phonemes

The main evidence for a sublexical route from heard to spoken phonemes is from conduction aphasia.

- Patients can successfully understand words, and successfully use words (e.g. in naming tasks), but are impaired at *repeating* words.
- Repetition errors can be semantic (*dog*→*cat*)
- Or phonological (*dog*→*dod*).
- Conduction aphasia tends to result from damage to temporoparietal cortex.

The phonological loop

It's not just single phonemes which use the sublexical route from hearing to speech: *sequences* of phonemes can use it as well.

Recall: we have a special form of working memory for phonological sequences.

Immediate recall experiments for phoneme / digit sequences:

- Phonological similarity effects show these sequences are stored phonetically.
- Lexical effects show that word representations can be involved.
- Interference by articulatory suppression shows that sequences need to be *rehearsed* to be retained.

What is phonological STM for?

It's involved in *learning words*.

- An infant hears words, and has to associate these with semantic representations of objects/actions in the current scene.
- Cross-situational learning: keep a sliding window of the most recent *n* words you have heard, and associate each semantic representation with every word in that window.

Is it involved in mature sentence processing?

- Some patients have badly impaired phon. STM, but intact abilities to generate speech (Shallice and Butterworth, 1977) or interpret it (Vallar and Shallice, 1990)
- Patients with impaired phon. STM tend to have language deficits (Martin and Saffran, 1997)

Phonological input and output buffers

One way of resolving the conflict is to hypothesise separate phonological input and phonological output buffers.

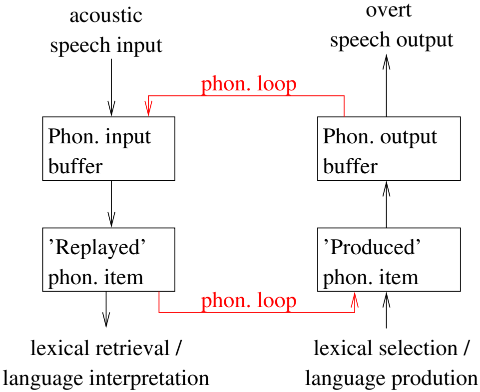
- The phon. input buffer stores an incoming signal as a sequence of phonemes.
  The signal fades over time (unless it's refreshed).
  In Hickok and Poeppel's model, this buffer is held in STS.

- The phon. output buffer is a *planned articulatory sequence*.
  Activity of assemblies in this buffer grows over time (until a firing threshold is reached).
  Prefrontal/premotor areas are likely to be involved. In particular, Broca's area.

Phonological input and output buffers

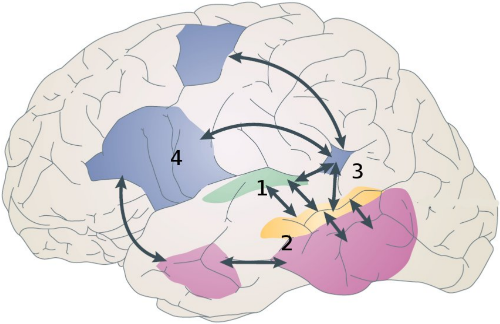
A model of rehearsal in phonological STM:

- Auditory input is sent to the phon. input buffer.
- Items are ‘read out’ of the input buffer directly to the output buffer.
- Items are (covertly) ‘read out’ of the output buffer, and ‘refresh’ the input buffer.



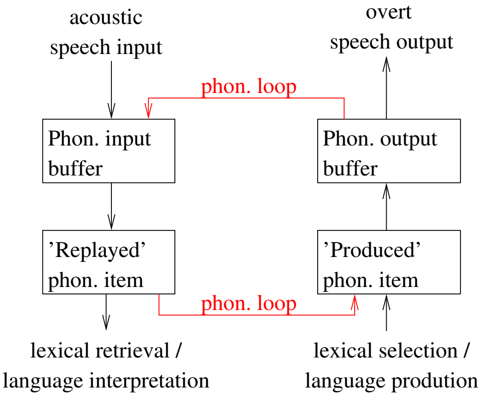
Phonological input and output buffers

The direct links between phon. input and output buffers are likely to overlap with Hickok and Poeppel's direct sublexical route from auditory to articulatory representations.



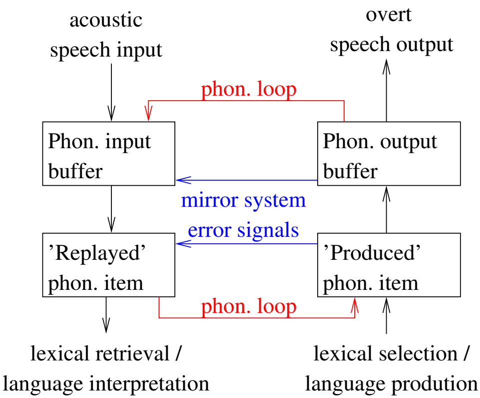
Phonological input and output buffers

The direct links between phon. input and output buffers are likely to overlap with Hickok and Poeppel's direct sublexical route from auditory to articulatory representations.



Phonological input and output buffers

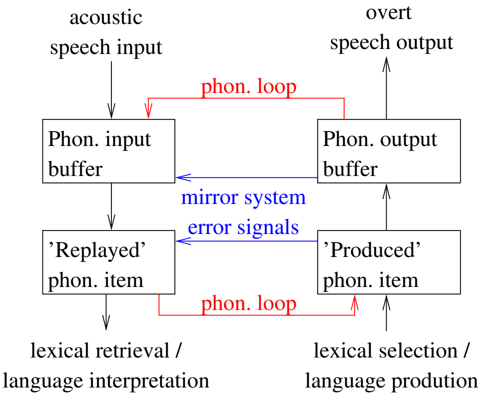
The direct links between phon. input and output buffers are likely to overlap with Hickok and Poeppel's direct sublexical route from auditory to articulatory representations.



Phonological input and output buffers

Q: do we convert input to output reps at the level of individual phonemes, or at the level of phoneme sequences?

- Probably a bit of both.



Phonological input and output buffers

Evidence for the phonological output buffer:

- Spoonerisms in normal speech (*barn door* $\rightarrow$ *darn bore*).
  (A CQ model of prepared phon. sequences is good for modelling these—see esp. Hartley and Houghton, 1996.)

Evidence for a distinction between input and output buffers:

- Reproduction conduction aphasia: patient with intact digit span, but makes phonological errors when generating speech, or repeating word sequences.
  (Intact input buffer; damaged output buffer)
- Patients with bad phon. STM, but intact abilities to generate speech.
  (Damaged input buffer; intact output buffer)

Outline of today's lecture

1. Neural representations of phonological structure
2. Neural representations of words
3. The neural locus of syntactic processing

Neural representations of words

How are words represented in the brain?

Words can be understood in three different ways:

- Phonologically (as regularly occurring phonological sequences)
  supercalafragilisticexpialidocious
- Semantically (as ‘word-sized’ semantic representations)
  DOG, CHASE
- As *mappings* between the above phonological and semantic reps
  DOG $\leftrightarrow$ *dog*

Phonological representations of words

There are three layers of phonological structure in a word.

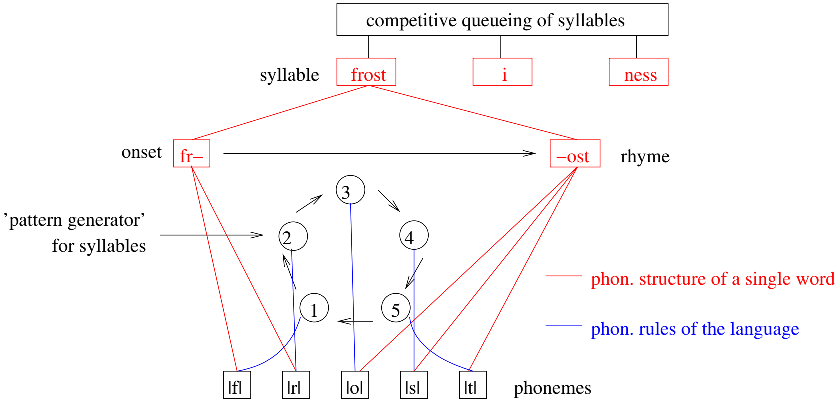
- A word is a sequence of syllables (e.g. *frost-i-ness*)
- A syllable has an onset and a rhyme (e.g. *fr-ost*)
- The onset and rhyme are both sequences of phonemes
- The onset is a sequence of 0, 1 or 2 consonants (*fr*)
- The rhyme is a vowel followed by 0–2 consonants (*o-st*).

Recall that phonological representations are *motor* representations.
All motor representations are hierarchical!

The best network models of phonological structure (e.g. Hartley and Houghton, 1996) use localist assemblies to represent units at different levels of structure.

Phonological representations of words

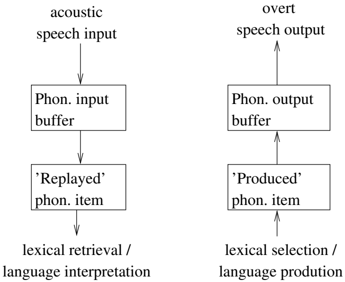
Hartley and Houghton's model of phonological word representations:



Phonological representations of words

Where are phonological word representations?

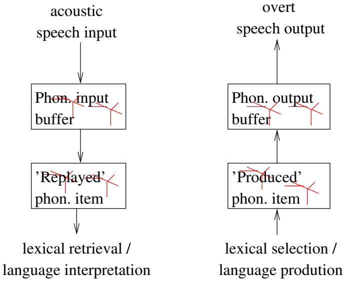
They're probably best thought of as *structure* within the phonological input and output systems.



Phonological representations of words

Where are phonological word representations?

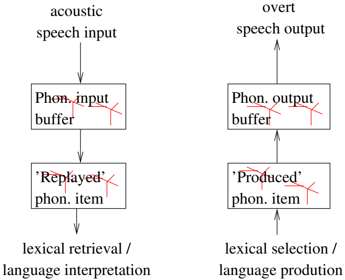
They're probably best thought of as *structure* within the phonological input and output systems.



Phonological representations of words

Where are phonological word representations?

They're probably best thought of as *structure* within the phonological input and output systems.



I'll look at evidence that there are separate phonological input and output lexicons in a moment.

Neural representations of ‘word meanings’

We already have a story here—at least, for concrete nouns & verbs.

- Concrete nouns denote object representations in IT.
- Action verbs denote planned motor/attentional actions in PFC.

There’s obviously more to word meanings than this.
But SM denotations *ground* the interconnected web of word meanings in the world.

Words as associations between forms and meanings

Phonological word representations must obviously map to semantic word representations, so that we can understand and produce words.

Our model of phonological representations predicts that the input lexicon is separate from the output lexicon. And this indeed seems to be the case.

Phonological input and output lexicons

Evidence for a distinct phonological input lexicon: pure word deafness.

- Patients who can read, write and speak words normally, but *cannot understand spoken words*.

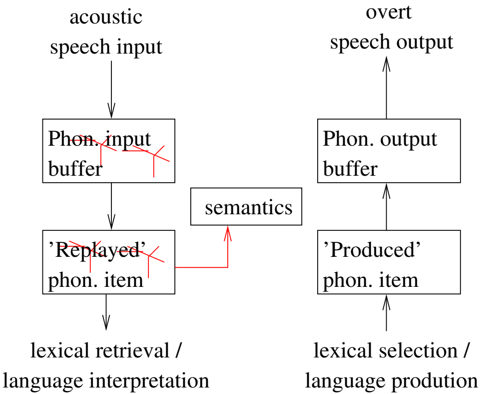
Evidence for a distinct phonological output lexicon: tip-of-the-tongue states.

- You can't find the word for a concept.
  But you remember it as soon as you hear it.

Input and output lexicons: Summary

The phonological input lexicon is:

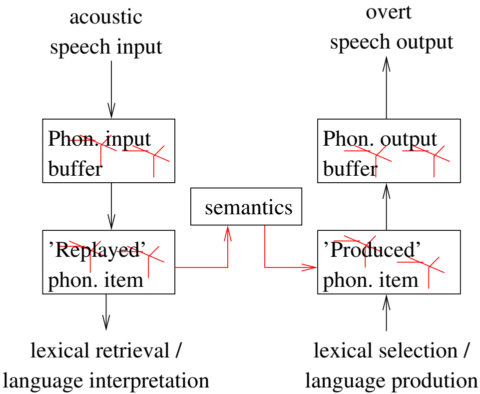
- Structure in the phonological input buffer;
- Links from assemblies in this buffer to semantic assemblies.



Input and output lexicons: Summary

The phonological output lexicon is:

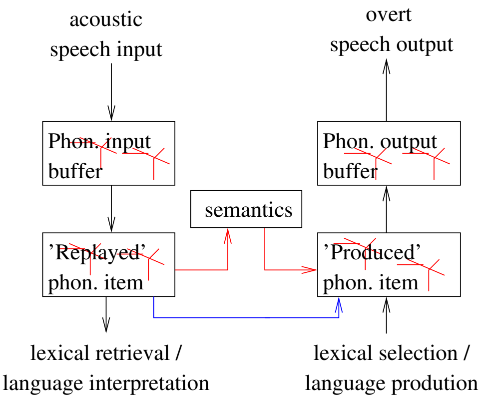
- Structure in the phonological output buffer;
- Links from assemblies in this buffer to semantic assemblies.



Input and output lexicons: Summary

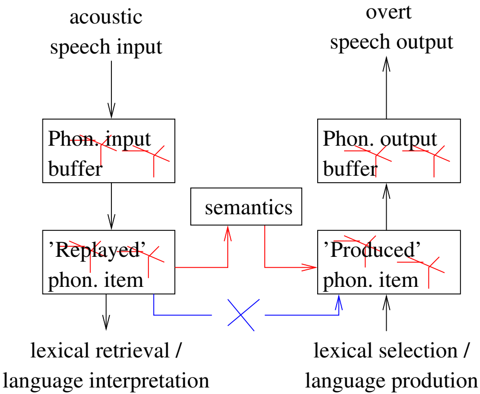
Note: there are now two pathways from hearing to speaking:

- One via a sub-lexical ‘sensory-motor’ route;
- One via lexical semantics.



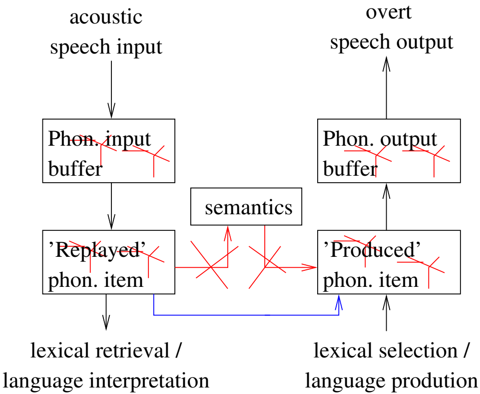
Input and output lexicons: Summary

Damage to the sub-lexical route results in a form of conduction aphasia in which patients make semantic errors when repeating words..
E.g. *mirror* → *reflection*.



Input and output lexicons: Summary

Damage to the lexical route results in an ability to echo speech without understanding it. (Sometimes called echolalia.)

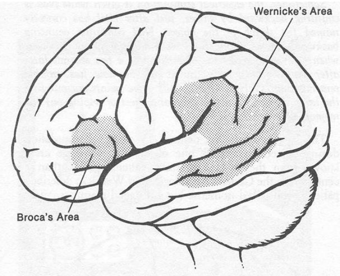


Neural areas involved in ‘the lexical route’

Traditionally, ‘Wernicke’s area’ is seen as an area involved in storing associations between phonological and semantic word reps.

- Wernicke’s aphasics talk in fluent, well-formed sentences—but their utterances are meaningless. They can’t name objects, or understand single words.

W’s area is notoriously ill-defined. Here’s a functional mapping:



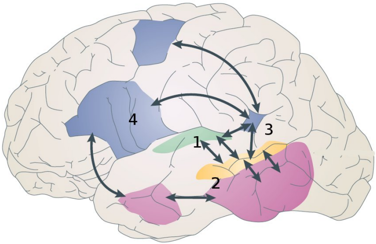
Traditional cortical language areas, inferred from the results of strokes.

Neural areas involved in 'the lexical route'

Traditionally, 'Wernicke's area' is seen as an area involved in storing associations between phonological and semantic word reps.

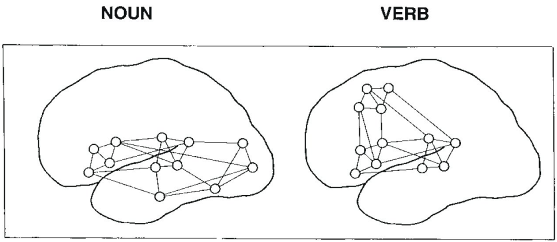
- Wernicke's aphasics talk in fluent, well-formed sentences—but their utterances are meaningless. They can't name objects, or understand single words.

Hickok and Poeppel model the 'lexical route' as involving posterior and anterior temporal areas too:



Word representations: Summary

Pulvermüller *et al.* (1999) generated fMRI images showing the areas activated when subjects are presented with (written) concrete nouns and verbs.



Conclusion: words are represented as *large-scale neuronal assemblies*, comprising a mixture of phonological (input and output) representations, semantic representations and neural pathways linking these representations.

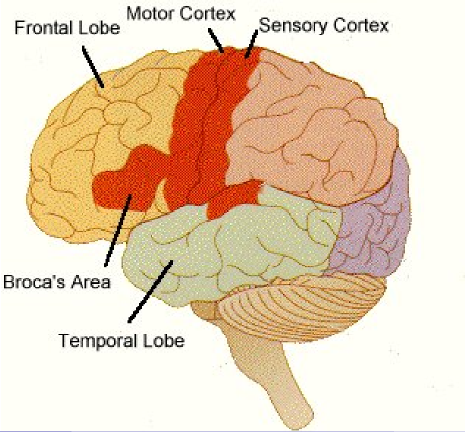
Outline of today's lecture

1. Neural representations of phonological structure
2. Neural representations of words
3. The neural locus of syntactic processing

The neural locus of syntactic processing

Where does 'syntactic processing' happen?

This one is a bit more straightforward: it happens in Broca's area (left inferior prefrontal/premotor/motor cortex).



Broca's area and input/output distinctions

We have already suggested that Broca's area is one of the key sites for the phonological output lexicon—i.e. for articulatory planning of spoken words.

- Damage to Broca's area *seems to* impair language production more than language comprehension. (Speech is very *nonfluent*.)

But in fact:

- Broca's aphasics can generate *single words* quite well. (So the phonological output lexicon doesn't seem damaged!)
- (Hickok and Poeppel: the phon. output lexicon probably involves all of the sensory-motor pathway for phonemes, including auditory areas.)

Broca's area and syntax/semantics distinctions

A better proposal is that Broca's area is where syntactic processing is done, both for input and output.

- Broca's aphasics speak ungrammatically, as well as disfluently.
- They also have difficulties understanding complex sentences.

However, Broca's aphasics can understand and produce individual words.

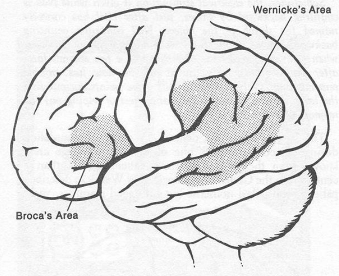
They probably 'understand' sentences by understanding individual words (and using world knowledge / canonical word order).

- *The man grabbed the cup* - easy
- *The boy was chased by the girl* - hard
- *The boy who the girl chased lives next door* - hard

Broca's area and syntax/semantics distinctions

A pretty straightforward idea:

- Broca's area is where syntactic processing is done (both for input and output)
- Wernicke's area (and associated temporal areas) are where word forms are mapped to word meanings (and vice versa).



Traditional cortical language areas, inferred from the results of strokes.

What processing is done in Broca's area?

The big question: what does Broca's area do?

- We know it 'processes syntax'.
- Problem: linguists can't agree about what syntax is. (There are lots of models of grammar, and they all leak.)
- Maybe we can get some ideas by looking at the kind of processing which seems to happen in Broca's area.

Broca's area and general serial cognition

Broca's area includes portions of prefrontal cortex.

- We've already seen that PFC is involved in the planning of action sequences.
- Sentences are sequences of words: perhaps Broca's area specialises in learning how to sequence words?

Some specific ideas:

- Goschke *et al.* (2001): Broca's aphasics *can* learn *simple* sequences.
- Conway and Christiansen (2001): Broca's aphasics are impaired in learning *hierarchical structure* in sequences.
- Dominey *et al.* (2003): Broca's aphasics are impaired in sequencing *abstract patterns* (E.g. XYYX)
- Greenfield (1991): children under 2 use Broca's area for regular sensorimotor sequencing

Broca's area and sensorimotor mechanisms

Rizzolatti and Arbib (1998) suggest that Broca's area is the human homologue of macaque F5 (the mirror neuron area).

- Note: F5 is where object and action representations are *combined* in the macaque.
- One of the important things about language is that it is compositional: the meaning of a sentence is a function of the meanings of its individual words (e.g. *grab cup*).
- Maybe the SM combination mechanism was co-opted by evolution to support compositional linguistic representations.

Note: Broca's aphasia often co-occurs with ideomotor apraxia: a condition where patients are unable to imitate hand/arm movements they observe.

Summary

We've looked at:

- A mirror system for individual phonemes and short phoneme sequences
- A (largely overlapping) system for phonological STM, involving phonological input and output buffers
- Hierarchically structured phonological representations of words
- Separate phonological input and output lexicons
- Distinct 'lexical' and 'sub-lexical' routes from hearing to speech
- A dissociation between knowledge of word meanings (held in Wernicke's area / related temporal areas) and syntactic knowledge (held in Broca's area).