COSC421: Neural Models of Language

Lecture 6: Language networks in the brain

Alistair Knott

Dept. of Computer Science, University of Otago

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

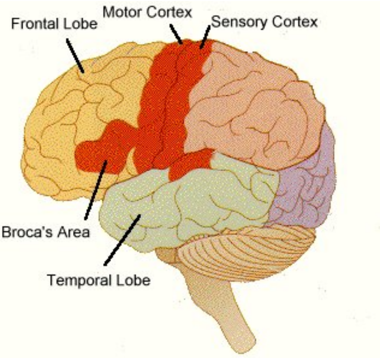
Outline of today's lecture

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.

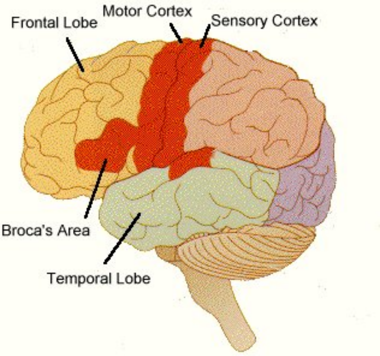


Neural representations of phonological structure

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.

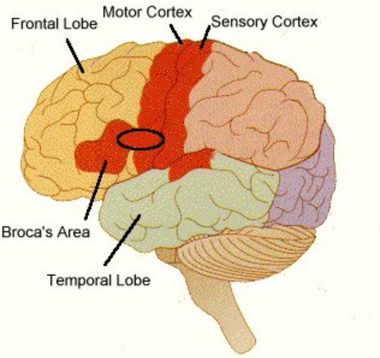


Neural representations of phonological structure

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**.

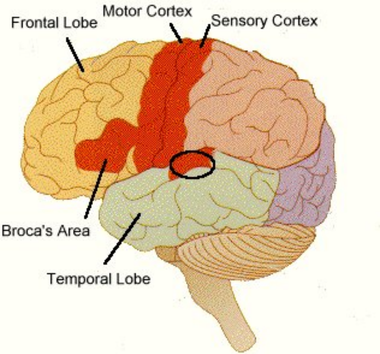


Neural representations of phonological structure

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).

  h*appy vs* **h***ope*
  l*ift vs goa***l***kick*
  t*ake vs abou***t** *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

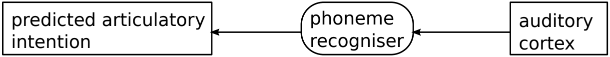
1. Evidence from the syntactic discipline of **phonology.**
(Study of the sound system of language.)

- Westermann and Miranda (2004): infants learn the relationship between intended articulatory gestures and acoustic signals by **babbling**.
    - They activate a random articulatory intention, and execute it.
    - 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')

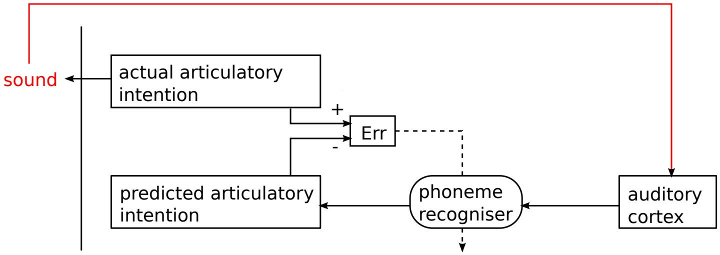
A circuit mapping auditory signals onto phonemes

Here's the circuit, very roughly.



A circuit mapping auditory signals onto phonemes

Here's the circuit, very roughly.
Here's how the circuit could be trained during 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*.

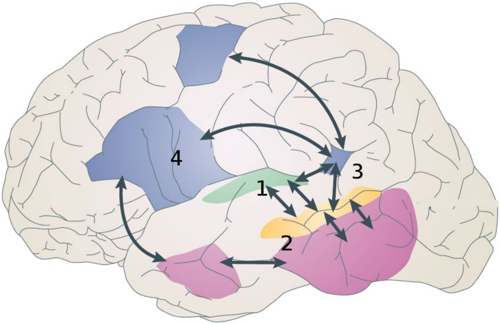
Evidence for a mirror system for phonemes

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 *h**ea**d* or *h**a**d*.
  - 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

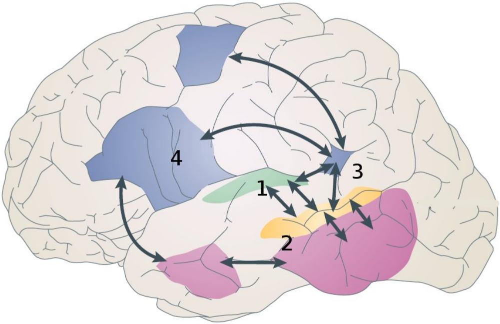
An influential model by Hickok and Poeppel (2007):



1. Auditory cortex: acoustic representations
2. Middle/posterior STS: phonological representations
3. Left temporoparietal junction / inferior parietal cortex: waystation
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):

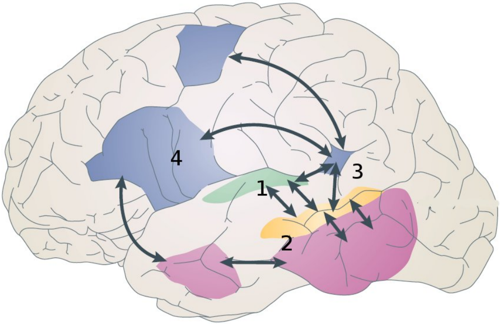


Note that this pathway is similar to the pathway for action recognition:

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.)

The units of the phonological input-output mapping

What are the units in which the direct mapping between phonological inputs and outputs is expressed?

- We know they're larger than individual phonemes. (Because of coarticulation.)

One proposal is that the units are **syllables**.

- A speech signal is a sequence of phonemes—but these phonemes are also grouped into units called **syllables**. (*pho-neme-sare-group-tin-to-sy-lla-bles*)
- Wheeldon and Levelt (1995) gave subjects a subvocal self-monitoring task: 'translate this word into your own language, and tell us whether the translation contains the phoneme *p*'. They found that it took longer to answer this question depending on which syllable the word was in (but position within syllables didn't matter).

Syllables

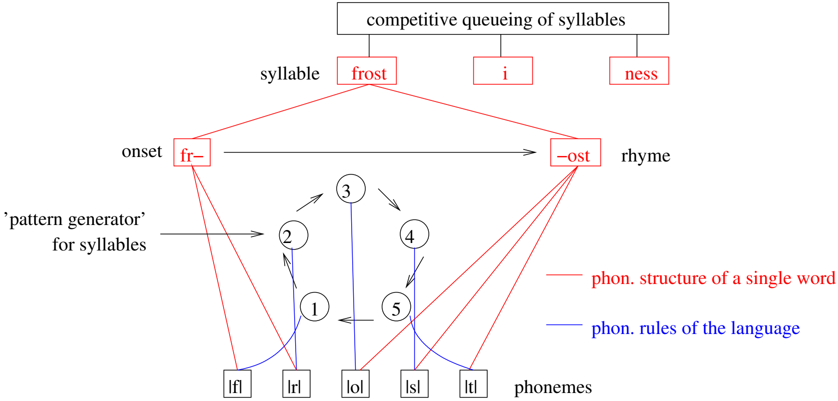
A syllable is a *rhythmic* unit of speech, organised around a 'high-sonority' sound (a vowel).
A syllable has hierarchical structure of its own:

- It 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!

Syllable representations: Hartley & Houghton's model

Hartley and Houghton's (1996) model uses localist representations to encode each syllable in the language.



Evidence for a sublexical route from hearing to speech

The main evidence for a sublexical route from hearing to speech 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 phonological
  (*dog*→*dod*)
- Or semantic
  (*dog*→*puppy*)
- Conduction aphasia tends to result from damage to temporoparietal cortex.

The phonological loop

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

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

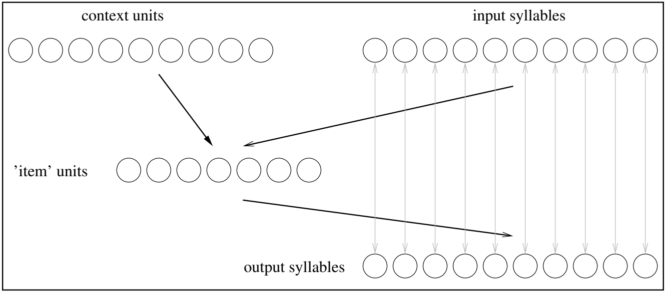
Immediate recall experiments for syllable / digit sequences:

- **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.

Burgess and Hitch's full model of phonological WM

In B&H's complete model, there are input and output phonological units, which are connected bidirectionally, one-to-one.

- These represent the 'direct' pathway between inputs and outputs.
- The connections from output to input units create a 'loop'.



Phonological input and output buffers

Burgess and Hitch's model is a model of how to *echo input* phonological sequences.
There's also evidence for a separate form of phonological working memory, involved in *preparing output* phonological sequences.

- E.g. **Spoonerisms** (*barn door* $\rightarrow$ *darn bore*).
  (The 'd' comes earlier than it should do: speaker must be planning it in advance.)

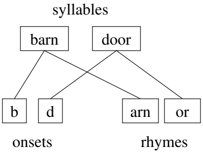
There must be a separate **phonological output buffer** where we prepare multiple syllables in parallel when planing speech.

- In this buffer, the activation of a syllable builds up over time, until it reaches a threshold and is pronounced.

Phonological input and output buffers

Hartley and Houghton's model of syllables produces Spoonerisms quite nicely.

- Whole syllables are active in parallel, and activate their associated onset and rhyme units in parallel.
- Onset and rhyme units compete separately.



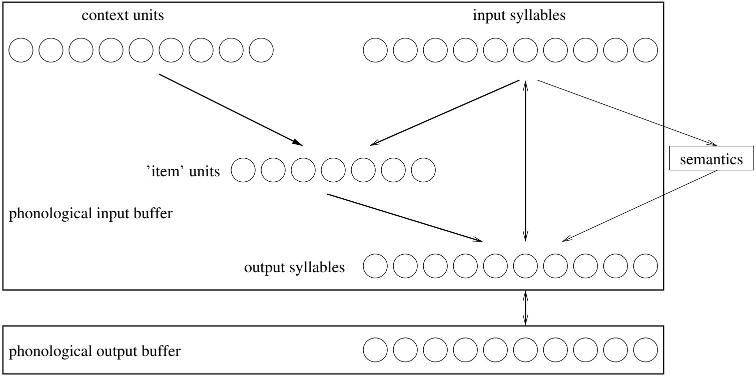
Phonological input and output buffers

Evidence for a distinction between input and output buffers:

- **Reproduction conduction aphasia**: patient with intact digit span, but makes phonological errors on *complex words* 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)

Phonological input and output buffers

Here's a more elaborate model of phonological working memory.



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*

Morphology

Actually the mapping between phonology and meaning happens at a level below whole words.

A single word can contain several meaningful units (**morphemes**).

prefix	stem	suffix
un-	*happy*	
	walk	*-ed*
	dog	*-s*

Morphology

Actually the mapping between phonology and meaning happens at a level below whole words.

A single word can contain several meaningful units (**morphemes**).

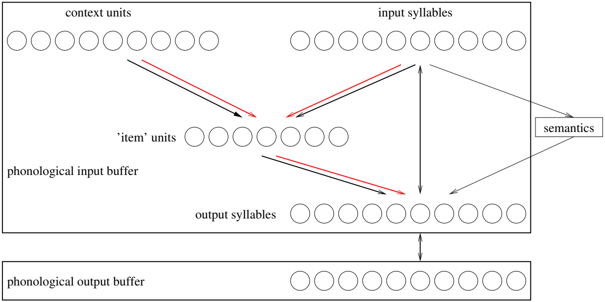
prefix	stem	suffix
un-	*happy*	
	walk	*-ed*
	dog	*-s*

I'll focus on **word stems** for now. (I'll discuss **inflections** later.)

Phonological representations of words

Where are phonological word representations?

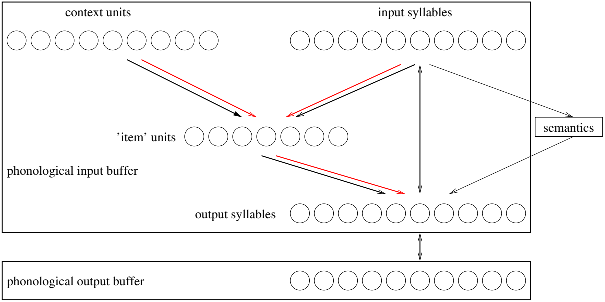
Burgess and Hitch: they are *long-term connections* in the phonological WM system.



Phonological representations of words

Where are phonological word representations?

Burgess and Hitch: they are *long-term connections* in the phonological WM system.



But we eventually need word representations to link to semantics..

Neural representations of ‘word meanings’

A simple proposal, for concrete nouns and action verbs:

- Concrete nouns denote object representations in IT.
- Action verbs denote motor programmes in premotor cortex.

Neural representations of ‘word meanings’

A simple proposal, for concrete nouns and action verbs:

- Concrete nouns denote object representations in IT.
- Action verbs denote motor programmes in premotor cortex.

Evidence for this proposal?

- Tranel *et al.* (2001): people with damage to IT had difficulties naming pictures of objects.
- Hauk *et al.* (2004): reading words denoting mouth, arm or leg actions activates the same premotor areas active during actual movements.

Neural representations of 'word meanings'

A simple proposal, for concrete nouns and action verbs:

- Concrete nouns denote object representations in IT.
- Action verbs denote motor programmes in premotor cortex.

Evidence for this proposal?

- Tranel *et al.* (2001): people with damage to IT had difficulties naming pictures of objects.
- Hauk *et al.* (2004): reading words denoting mouth, arm or leg actions activates the same premotor areas active during actual movements.

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

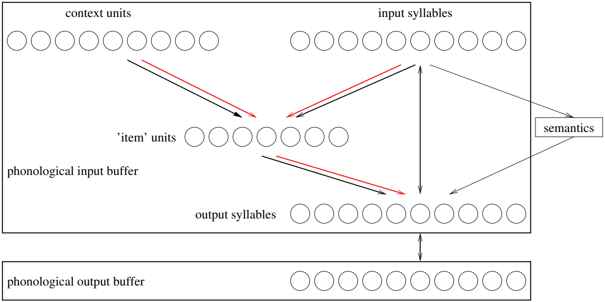
Evidence for a distinct phonological input lexicon: **pure word deafness**.

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

Evidence for a distinct phonological output lexicon: **anomia**.

- Patients who have difficulty producing nouns, but whose speech is otherwise fluent. (No problems *interpreting* nouns, or *repeating* nouns.)

Phonological input and output lexicons



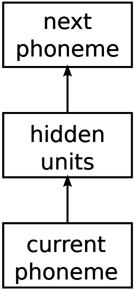
Circuits linking word forms and word meanings

A good model of the circuits mapping between word forms and word meanings are **simple recurrent networks** (**SRN**s: Elman 1990).

- SRNs are very widely used in models of language.
- They have a neat way of representing temporal sequences as declarative patterns of activation.

SRNs for representing words

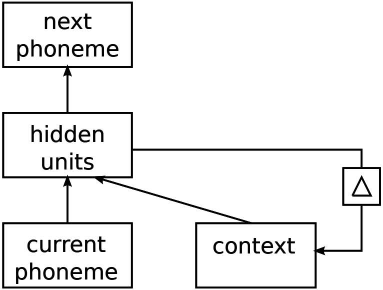
A SRN learns to map one item in a sequence onto the *next* item, using a layer of hidden units.



SRNs for representing words

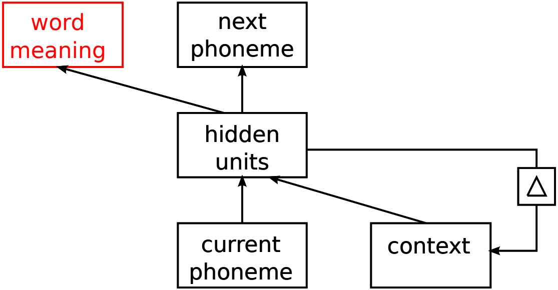
The key feature of a SRN is its use of a set of **context units** as an additional input.

- The context units at time t are a *copy* of the hidden layer at $t-1$.



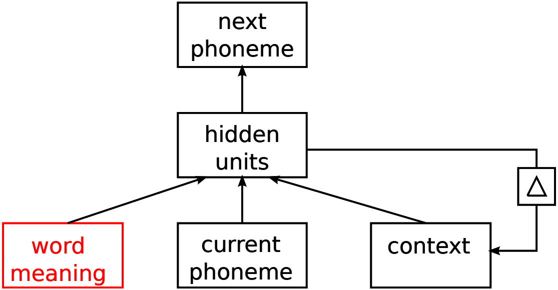
SRNs for representing words

When the SRN is configured for **word interpretation**, we can add an extra *output* layer, representing the meaning of a word.



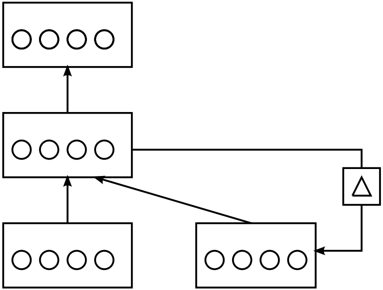
SRNs for representing words

When the SRN is configured for **word production**, we can add an extra *input* layer, representing the meaning of a word.



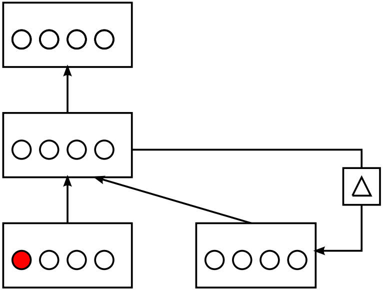
SRNs for representing words

Here's a SRN with four input/output units: each is a localist encoding of an item in a sequence.



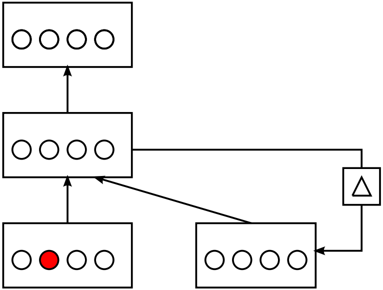
SRNs for representing words

Here's a sequence of inputs. We want the network to learn this sequence.



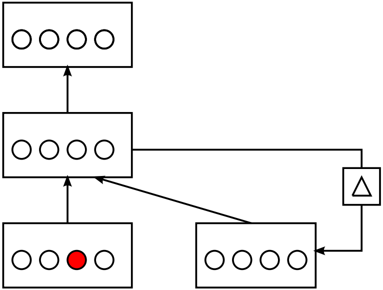
SRNs for representing words

Here's a sequence of inputs. We want the network to learn this sequence.



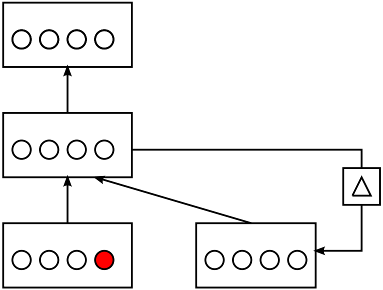
SRNs for representing words

Here's a sequence of inputs. We want the network to learn this sequence.



SRNs for representing words

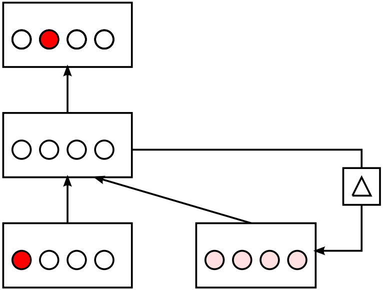
Here's a sequence of inputs. We want the network to learn this sequence.



SRNs for representing words

Training happens when each input is presented.

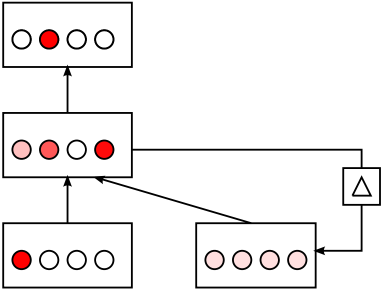
- We begin by setting a standard *initial context* representation, and we train the network to map the first input to the second in this context.



SRNs for representing words

Training happens when each input is presented.

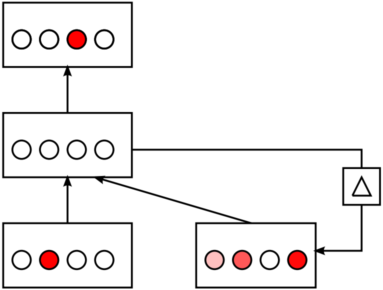
- The network learns to represent this mapping with a particular pattern of hidden unit activation.



SRNs for representing words

Training happens when each input is presented.

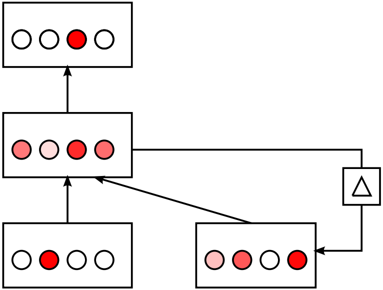
- At the next time point, we copy the hidden unit pattern into the context units, and present the next training pattern.



SRNs for representing words

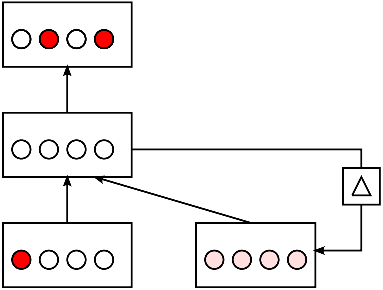
Training happens when each input is presented.

- The network learns to represent *this* pattern in a new pattern of hidden units. And so on.



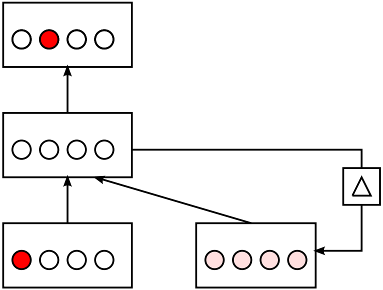
SRNs for representing words

After training, when you present a given input and context, the network will output a *distributed* representation of 'what it saw next' during training.



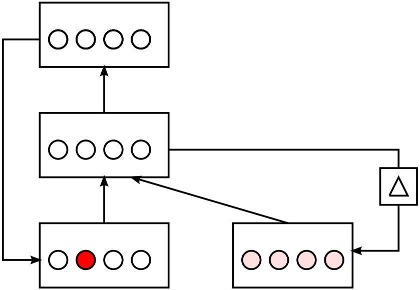
SRNs for representing words

From that input you can pick a single winner…



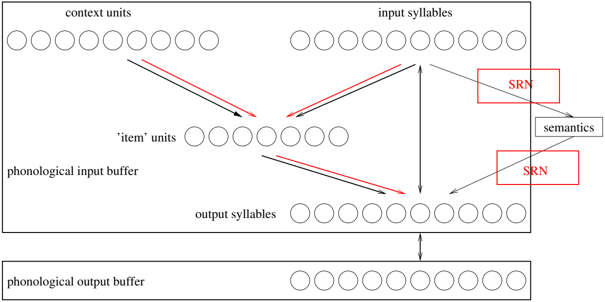
SRNs for representing words

And copy that winner into the input layer to predict the next item in the sequence. (And so on iteratively.)



SRNs for input and output lexicons

We can model the functions which map between phoneme sequences and semantics as SRNs.

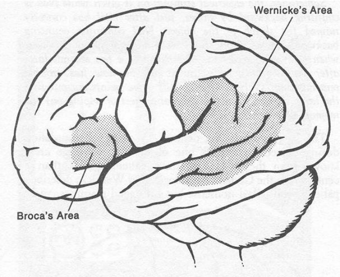


Neural areas involved in 'the lexical route'

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 rough picture:



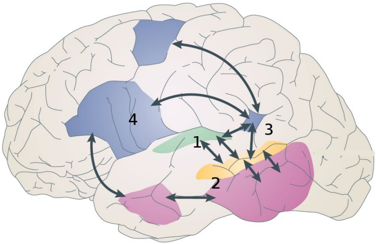
Traditional cortical language areas, inferred from the results of strokes.

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.

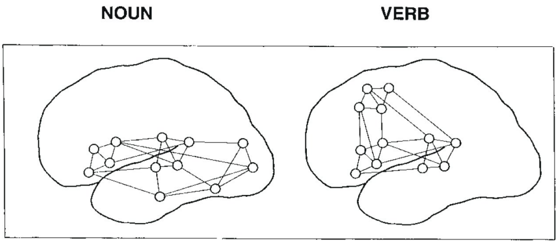
- **Wernicke’s aphasics** talk in fluent, well-formed sentences—but their utterances are meaningless. They can’t name objects, or understand single words.

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



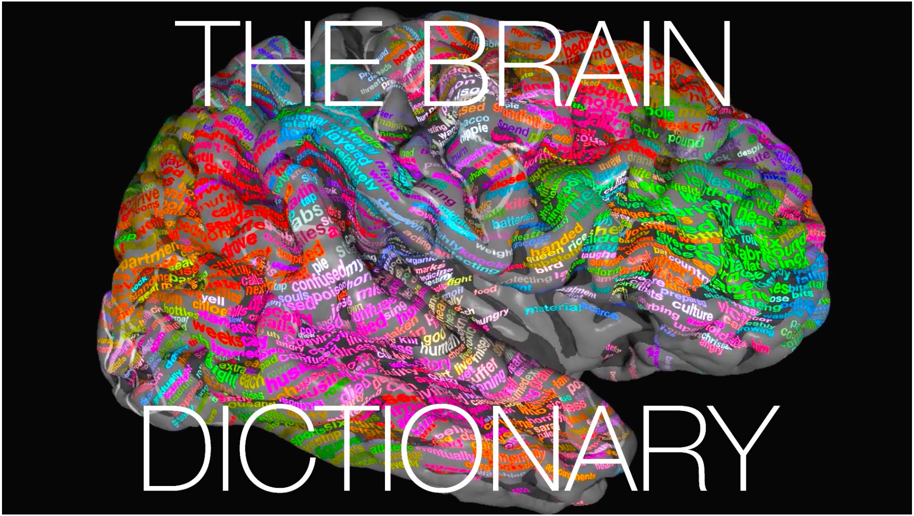
Word representations: Summary

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.

Huth *et al.* (2016) might change all that



Huth *et al.* (2016)

A recent study from Jack Gallant's lab (Huth *et al.*, Nature 2016) has found evidence for much more localised representations of words, using multivoxel pattern analysis.

- Subjects heard stories, and fMRI activity was recorded. The researchers looked for fMRI patterns associated with the presentation of each word.
- Each word W was encoded as a high-dimensional feature vector. (Dimension d of the vector represented the number of times some keyword K_d co-occurred close to W in a big corpus of text.)

https://www.youtube.com/watch?feature=player_embedded&v=k61nJkx5aDQ

http://gallantlab.org/huth2016

To me, these localised representations feel like 'convergence zones': sparse assemblies chosen to hold *associations* between large distributed representations.

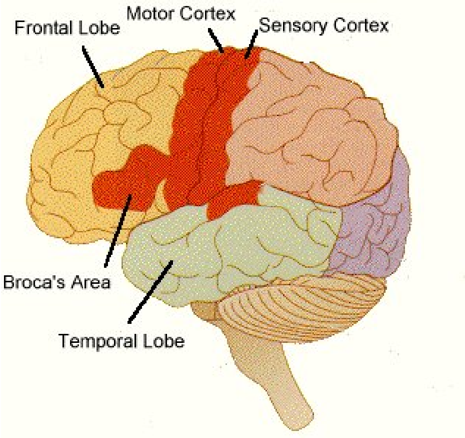
Outline of today's lecture

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

The neural locus of syntactic processing

There are two key areas involved in ‘syntactic processing’:

- **Broca’s area** (left inferior prefrontal/premotor/motor cortex)
- Anterior temporal cortex. (Friederici *et al.*, 2000)



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.

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

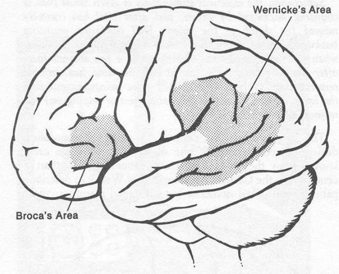
Broca's aphasics probably 'understand' sentences by understanding individual words (and using world knowledge / canonical word order).

- Anterior temporal cortex is perhaps involved in learning canonical word orderings. (Knoblauch & Pulvermueller, 2005)

Broca's area and syntax/semantics distinctions

A rough model:

- Broca's area is where syntactic processing is done (both for input and output)
- Anterior temporal cortex does 'surface' syntactic processing (during input)
- Wernicke's area (and associated temporal/premotor 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. (But they certainly don't *always* co-occur!)

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).