Recognizing Visual Words 6C

1. Intro – ask students how they think they recognize words…
   1. The magic moment – the time between word recognition and accessing it’s meaning (balota)
   2. How quick do you think the magic moment is?
2. Basic Methods and Findings
   1. Basic methods – brain imaging, eye movements, measuring naming, lexical decision, categorization times, and t-scope identification
   2. Eye movements in reading
      1. Techniques
         1. Limbus tracking – infra red beam is bounced off the white of the eye (limbus) – this procedure is good at horizontal movements, but not vertical
         2. Purkinje system – uses the reflection of the cornea and lens to calculate what you are looking at – good at vertical and horizontal (picture?)
      2. Saccades – fast movement of the eye, jumps forward and backward
         1. 20 to 60 msec
      3. Fixation is about 200 / 250 msec – the still period between jumps
         1. Information taken in a fixation is about 15 characters to the right, 3-4 to the left
            1. Span is based on reading skill
         2. Fovea – periscope of sharp vision (show perception eye balls here)
            1. 2 degrees of visual angle, about 7 characters of average size text.
         3. Periphery – visual acuity is worse about 5 degrees of visual angle
      4. Regressions – when we need to move backwards in a text, when we make a mistake reading or the information is ambiguous.
      5. Measures
         1. First fixation duration – amount of time the eye spends looking at the first fixation
         2. Total gaze time – which might include time you went back to the first fixation
      6. Models
         1. EZ Reader Model – attention, visual processing, oculomotor control all together determine when and where eyes move to
            1. We look at one point, then visual attention moves until acuity is bad, then the whole thing shifts forward to where attention is to make acuity better, (by saccade)
            2. Attention is serial, one word at a time depending on identification (familiarity check 🡪 word meaning pulled up 🡪 integration into linguistic structure)
            3. Saccades are programmed after familiarity check
            4. If our interpretation turns out to be incorrect, a higher level process can intervene and makes us go backwards to check.
   3. Reaction time (naming, ldt, t-scope)
      1. Naming task – shown a word, they say it out loud (name it), and the time it takes them to start to pronounce the word is measured.
         1. Average naming latency is around 500msec (ELP)
      2. Lexical decision task – person must decide if a string of letters is a word or nonword (show examples of both)
         1. Measure reaction times and error rates
         2. Problem: speed accuracy trade off – the faster you go the more errors you make, which is why you analyze both
         3. Reaction time averages are from 500 msec to 1 sec – depends on many factors (guess)
      3. Tachistoscope – a machine used to present words for extremely short durations (used before computers were around or could be programmed to do this accurately)
         1. T-scope identification – show words for a short time and find the critical point in which people can no longer identify words (threshold)
         2. Subliminal perception – your perception of words is affected even though you don’t know you saw anything (priming)
      4. Semantic categorization task – asks a participant to do something with the meaning of the word
         1. Is an apple a fruit or a vegetable; is an apple smaller or larger than a chair?
      5. Priming – presenting material before the word to which a response has to be made
         1. For example, show DOCTOR, then NURSE and people respond to NURSE in naming or lexical decision task
         2. SOA – stimulus onset asynchrony – the time between the prime and the target; changing the SOA can help you understand visual recognition, you get stronger priming for long SOAs and less for short SOAs.
3. Word Recognition – what makes it easier/harder?
   1. Interfering with identification
      1. Stimulus degradation – making them harder to read, harder to recognize, breaking up/masking the letters
      2. Backwards masking – presenting another stimulus immediately after the target
         1. Energy masking – putting a mask after the stimulus that is unstructured (noise masking), like random TV fuzz
            1. Interferes with feature identification because there are new features which overwrite the old ones
         2. Pattern masking – structured mask, like other letters
            1. Interferes at the letter level because there are new letters to process
      3. Perception without awareness – when you mask/degrade information to a level where participants claim they can’t see the information, but you can cause priming or lexical access
         1. Highly controversial, the tasks to do this are tricky
      4. Partial word masking – when you present a word and delay showing people some of the letters
         1. Hurts you more when you delay consonants than vowels, (probably due to the visual to sound mapping – vowels change depending on the consonant, but consonants are more constant, which helps you figure out the word).
   2. Frequency, familiarity, and age of acquisition
      1. Frequency – extremely important, commonly used words are easier to recognize and are responded to more quickly
         1. Frequency norms – google (new new new!), Kucera and Francis (brown corpus), CELEX, HAL, SUBLEX, etc. (show websites, facebook frequency data from lab)
         2. However, these things are not the greatest – compare with the old norms and new norms…word use changes
         3. People rate some words as more frequent than others, even though they are the same (giggle versus cohere)
         4. Obviously expertise interacts with this as well
      2. Age of acquisition – age at which you first learned a word
         1. Example, giant is learned early, but it’s a low frequency word
         2. Words that you learn early are worked with more quickly than late learned words
         3. Another problem is cumulative frequency – how often the words occurs in a person’s lifetime
         4. Brain/network model wise – as we grow older, our brains are less plastic and models of word learning have shown that early learned words have stronger representations (activations) and when you learn words later, even with the same frequency they have lower representations
            1. (example: statistics)
      3. Word length
         1. First – how do you measure this? Number of letters, number of syllables, how long it takes you to say it
         2. Number of letters doesn’t seem to have an effect as long as they are in the buffer of visual recognition (5 to 12 characters so you don’t have to make a saccade)
         3. It does take you longer to name words that have more syllables, probably because of the prep time to pronounce those different sounds (moving your mouth, remembering how to form the sound)
      4. Neighborhood effects
         1. Orthographic neighbors – words that look the same as the target word. Sometimes defined as the number of words you can create from changing one letter in the word
         2. N is the neighborhood size or density
         3. Words with more neighbors are easier to recognize
      5. Word-nonword differences
         1. Words are faster than nonwords (a string of letters that does not form a word)
         2. Pseudo words – nonwords that are pronounceable
         3. We are much faster at rejecting non plausible words (TNVK) than we are pseudowords (TINK)
      6. Repetition priming – priming by repeating a stimulus
         1. We are much better at recognizing words when they have been repeated (although strangely sometimes we ignore words we’ve already seen – repetition blindness…why the opposite effects?)
         2. Frequency attenuation – repetition priming helps more for low frequency words than high frequency words
      7. Form based priming – also called orthographic priming – facilitation at recognizing words when they look the same
         1. Only really helps when you mask the words, at a conscious level it can actually hurt you.
      8. Semantic priming – facilitation when words are related in meaning
      9. Other factors:
         1. Grammatical category (mad libs)
         2. Imagine ability
         3. Meaningfulness
         4. Concreteness
         5. At lot of these factors are task dependent – their usefulness depends on what they are having you do
   3. Attentional processes in visual word recognition
      1. Reading is an automatic process – see the stroop task
      2. Automatic processes – processing that is unconscious, fast, obligatory, facilitatory, does not involve working memory space, does not have dual task problems
      3. Attentional/controlled processes – processing that requires central resources, requires working memory space, accessible to consciousness
      4. Priming studies (neely 1977 classic) show that at short SOAs you get fast automatic priming, long SOAs you get controlled priming, both which help.
      5. Predictive validity and proportion effect – the more related cue-target pairs you use, the more priming you get and vice versa
      6. We basically have early automatic processes and slower later controlled processes that make up priming
4. Different tasks different results
   1. Lexical decision tasks
      1. Problem is post access effects – measures participants decision making and strategies
         1. Expectancy generation – we are unconsciously thinking of what word would appear next
         2. Post lexical checking – checking if target matches or is related to the first cue word.
   2. Naming
      1. Does not seem to show inhibition processes – so maybe all automatic processing
      2. Mediated priming – priming through a third part (lion – stripes – mediated by tiger)
         1. You see this more in naming, but not in LDT because the controlled processes eliminate it.
      3. Problem – we hesitate when we do not know how to pronounce the word, so there are different types of strategies.
   3. Frequency effects – where do they occur?
      1. Experiment: participants were trained on a frequency distribution of fake words.
         1. They then were shown blurred symbols and produced the same frequency distributions of guesses at fake words for those symbols.
      2. Frequency seems to create a *response* *bias* or *guessing*
      3. Is frequency linked to recognition or generation?
         1. Frequent words need less evidence to be identified
         2. Repeated exposure lowers the word recognition threshold
         3. Seems to be both
      4. Is frequency practice?
         1. All that seems to matter is relative frequency, overall practice does not help/hurt
      5. LDT versus Naming and Frequency
         1. LDT is more sensitive to frequency because it uses more controlled process factors, which frequency is helpful for
         2. You can see strong frequency effects in naming if you use only regular words (not irregular pronunciations, which eliminate frequency effects because you are having to look up the sound to spelling rules)
         3. The problem with both of these is the frequency of the orthographic neighbor (blur and blue), hurts you in LDT, helps you in naming
5. Is there a specific visual word recognition system
   1. We’ve had much less time evolutionarily to read, so it’s probably evolved out of another perception and cognitive system
   2. Is word processing separate?
      1. Give people a picture naming task, some with words and some with out
      2. Pictures help related words in a LDT (but not as much as words prime words pictures prime pictures)
      3. Brains: fusiform gyrus (left visual cortex) seem to response to words and pseudo words, even when you change the features of the words (which have different cells in another visual area like cAsE).
   3. Two visual recognition properties
      1. Holistic processing – like for faces
      2. Processing of complex multiple parts – like for words
      3. Agnosia – disorder of object recognition – should always have one of these two parts
         1. Prosopagnosia – can’t recognize faces
         2. Dyslexia – can’t recognize words
      4. Not 100% supported
6. Meaning based facilitation of word recognition
   1. Types of semantic priming
      1. Associative priming – association is when two word are related by context and word use
         1. Talk about the nelson norms and how they are created
         2. Give examples from everyday culture
      2. Nonassociative semantic priming – words that are only related by meaning or feature overlap
         1. Give examples
         2. Talk about the McRae / Vinson norms
      3. Most research has mixed the two together, just saying that they are studying semantic priming
         1. When you separate the two out, you get strong associative priming
         2. Research is very mixed on the automatic-ness of these two. Most people argue for low level lexical access priming (association – picture would help) that is automatic, then higher level controlled pure semantic processing
         3. Many studies show no priming, some studies show priming for naming but not LDT, others show both. (sheesh!)
   2. Sentence context and visual word recognition
      1. Priming from context is the extra bump over the association of the words
      2. You should brush your teeth every single \_\_\_\_ day is faster than year or hour.
      3. You seem to get facilitation for a highly predictable word (probably more frequent word), over a word that would work but is not as predictable, especially over a word that doesn’t make sense – tends to be inhibitory
         1. The janitor cleaned the floor with the \_\_\_\_ (mop, solvent, lawnmower)
      4. This mechanism is probably a controlled process – you are generating the expected word
      5. ERP results suggest that priming and context worked at the same time – found N400 attenuation that varied based on the amount of association and sentence context
      6. RSVP task show that people will misperceive a word based on context
         1. She ran her best time yet in the RICE last weekend
         2. People see and report this as the word race.
   3. Meaning based priming studies summary
      1. We can distinguish between associative and semantic priming
      2. There are automatic and attentional control mechanisms that create priming facilitation
         1. Attentional processes – checking items to make sure they are right ones, expectancies, integrating word into larger context
      3. Sentence context seems to provide a constraint on the words activated in the system
      4. Spread of activation for priming to occur (brain explanation, connectionist modeling)
         1. Show an example of spreading activation, items that are close in a network will get activated together
      5. Compound cue theory – search memory for information that is related to a compound cue (both the cue and the target)
7. Morphology – processing complex words
   1. Storage
      1. Full listing hypothesis – all entries for all versions of words (kiss, kisses, kissing)
      2. Obligatory decomposition hypothesis – listing the stem of the word with the rules on how to change them to different versions
         1. Also store a set of exceptions
         2. Affix stripping – to recognize a complex word, we must strip it down
      3. Dual pathway hypothesis – half way between the first two, we wouldn’t list all the different versions just a lot of the common ones for a short cut
   2. Support studies
      1. In LDT tasks words that look like they have affixes take longer to recognize as if we were trying to break them down (results)
      2. Homophones – words that sound the same
         1. When people are asked to judge if a word has a homophone, they have a hard time with words that have an affix
         2. Fined – find, they don’t remember fine because it’s stored as fine and –ed.
      3. Words are recognized faster if the stem (without affixes) is more frequent (seeming, mending – seem is more frequent recognized faster)
      4. Some support for dual pathway because of weird compound words – like buttercup
         1. It’s meaning is different than it’s separate parts, so it’s not primed by milk or spoon
   3. Transparency
      1. Phonologically transparent – if shared part of words sound the same
         1. Friend and friendly, sign and signal
      2. Semantically transparent – meaning is obvious from it’s parts
         1. Happiness is predictable with unhappiness
         2. Depart and department – semantically opaque
      3. Results: we seem to decompose semantically transparent words and leave opaque words together (supports dual pathway)
8. Models of visual word recognition
   1. Four questions to answer:
      1. Automatic or interactive processing – what effect of top down processing
      2. Serial or parallel process
      3. Activation – can it spread to another level before processing at that level is done or must it wait?
      4. How do we find items? Search or storage bin number?
   2. Basic types
      1. Lexical instance models – perceptual access to a lexicon
         1. Lexicon stores the individual words, no rule based component that converts letters to sounds
         2. Either they have a serial search or activation based direct access models
   3. Forster’s autonomous serial search model (page 194)
      1. Model is analogous to searching through a dictionary. Words are ordered in some fashion (alphabetical), then when you get to the word – meaning, phonology, etc. is stored.
      2. Stages of model
         1. Perceptual processes first
         2. Identify words by using a serial search through the lexicon (called an access file)
            1. Files are modality specific (orthographic, phonological, syntactic-semantic)
            2. Access files point to the master file
         3. Master file includes the meaning of the word
      3. *Bins* are used to speed up system – access files are stored in separate bins based on the first syllable or first few letters of word
         1. Within the bin items are stored by frequency with high frequency items first
      4. Literature support
         1. Frequency causes priming
         2. Search is autonomous – not affected by context so that’s a controlled post access issue
         3. Repetition priming occurs by changing bin frequency
         4. Illegal words are easy to reject because they don’t match any bins, but pseudo words take forever because you have to search each bin
         5. Rank hypothesis – relative frequency of word is much more important when it comes to how fast you will find the word
            1. This model accounts for that by bin separation – one bin has frequency of 100,000 and 10, while the second bin is 20 and 10. Both the 100,000 and 20 have the same relative frequency and should be primed the same (and are).
      5. Evaluation
         1. Word recognition is very fast, how could it possibly be serial?
         2. How do we learn to pronounce nonwords?
         3. Forster added that the processing is parallel, so that all bins are searched at the same time – makes the whole system faster
   4. Logogen model (Morton)
      1. Logogen – simple feature counter for every word, as the features accumulate it adds up until the threshold is reached and the word is recognized
      2. Lexical access is direct – some simultaneously and some parallel
      3. Features and demons
         1. Demons monitor perceptual input for particular features
         2. The more evidence there is for a feature, the louder the demon shouts
         3. This is actually very close to the idea of how perception works
         4. Nick research picture of feature // perception
      4. Each logogen has a resting activation and as evidence is presented it raises activation – so if you are shown a letter t, all the words with t will activate
         1. Similar to action potentials, once the threshold is reached, the logogen will fire and a word will be recognized
      5. Model parts (page 195 picture)
         1. This model does not distinguish between context and perceptual information, so they both help in recognition
         2. More frequent items have lower thresholds
         3. Nonwords will be rejected if no logogen has fired after a certain time period has passed
      6. Problems with the original model
         1. Before a single logogen did all modalities of the word – visual or auditory
         2. The problem with that is that the model predicts that all modalities would give you equal priming
            1. Many experiments show that modalities are best if matched (visual-visual) with priming
         3. The new model included separate logogens for input and output
            1. Seems that the output of the priming doesn’t matter as much as the input of the priming
         4. Four logogen systems are necessary – reading, writing, listening, speaking
      7. This model really paved the way for the most dominant view of word recognition
   5. Interactive activation models of word recognition (McClelland and Rumelhart) – page 197 picture
      1. Model was originally designed to investigate the *word superiority effect* (letter perception changes based on the word the letter is in)
      2. Parts
         1. Input level of visual feature units
         2. Letter level
         3. Output level of word units
         4. Connections are excitatory (positive make it go yay!) or inhibitory (negative make it go no!).
         5. Each unit is connected to each other at the same level – creating competition for activation (basically if you decide the letter T is there, you can’t activate CAKE and then when you decide the word is TRAP, CART gets downplayed)
      3. Process
         1. When a unit is activated -> sends activation in parallel to everything it’s connected to.
         2. This will increase activation in some, decrease others
         3. Units that get an increase will start to decrease connections on the same level
         4. Activation also flows back down…making words easier to see
      4. The model is highly interactive – so any evidence is used in the system
   6. Hybrid models (becker)
      1. Combine both the connectionist/logogen models and serial search models
      2. Beckers
         1. Sensory set or perceptual processes that generate stimuli
         2. Semantic set or context top down processes that generate stimuli (ordered by frequency)
         3. These two are checked and verified against the data
         4. Verification is serial
         5. If no match is found, then you go to the sensory set
9. Lexical Ambiguity
   1. Types of lexical ambiguity
      1. Homophones – words that sound the same (bank, pen)
      2. Heterographic homophones – words with different spellings that sound the same (knight, night, soul, sole)
      3. Homographs – words that look the same but are pronounced differently (lead metal, lead to the store)
      4. Polysemous words – words with multiple meanings (bank, straw, ball, letter)
         1. Multiple meanings – words have completely unrelated meanings
         2. Multiple senses – word have related meanings – twist (coil, turning, sprain an ankle, distort the meaning)
      5. Ambiguity advantage – we are faster to recognize ambiguous words with multiple senses
         1. Maybe because they are in the lexicon several times?
   2. Work on ambiguity
      1. Early models
         1. Context guided single reading lexical access model – context restricts the access process so that only relevant meanings can be processed
         2. Ordered access – all the senses of a word are activated in frequency order (pen is a writing utensil versus animal holder)
            1. These are then checked against context
         3. Multiple access model – all senses are activated and appropriate one is chosen when context permits
      2. Early experiments
         1. Exp 1: Participants were shown:
            1. After taking the right turn at the intersection, I…
            2. After taking the left turn at the intersection, I…
            3. Which do you think takes longer to fill in?
            4. The first sentence because there are multiple meanings for the word right…(correct right side) and seems to indicate we activate all the meanings at once.
         2. Exp 2: participants heard:
            1. The farmer put his straw beside the machine.
            2. The farmer put his hay beside the machine.
            3. Which took longer to recognize? - The first one because straw has multiple meanings.
         3. Exp 3: Dichotic listening task (explain two ears, two messages)
            1. Attended: the spy put out the torch as a signal to attack.
            2. Unattended: the spy extinguished the torch in the window.
            3. Unattended: the spy displayed the torch in the window.
            4. They were asked to paraphrase the attended ear message. It was biased by which version of the unattended ear they heard.
         4. Exp 4: People are slow to find ambiguous words when they match the most frequent context.
            1. The accountant filled his pen with ink
            2. The farmer put the sheep in the pen.
            3. Indicates that frequency interacts with activation of all meanings.
         5. Exp 5: Priming shows that selective access is occurring
            1. Save bank money
            2. River bank money
            3. Day bank money
            4. Money is fastest with the first one – which should all be the same if all forms of bank were activated.
      3. Swinney’s experiment
         1. Used a cross modal priming technique – you have to respond to a visual LDT while listening to auditory material
            1. Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several (spiders, roaches, and other) bugs in the corner of the room
         2. BUGS – the () phrase biases you toward insects
         3. People got a LDT after the word bugs or corner
            1. ANT, SPY, SEW
            2. At the word bugs, we got priming for both ANT and SPY (indicating that quickly both versions were available)
            3. At the word corner, only got activation for the word ANT (or relevant meaning)
      4. Summary:
         1. When we hit an ambiguous word, we activate all meanings
         2. Context quickly rejects the non related meanings (200msec or less)
         3. Less frequent meanings take longer to access because they need more activation to cross threshold
         4. Appears that sematic representation of sentence is essentially word by word
   3. Recent research
      1. Effects of meaning frequency and prior context
         1. Question is whether or not all meanings are accessed in the same way in that initial moment (are more frequent meanings activated stronger or slightly first or what?)
         2. Balanced ambiguous words – alternative meanings of words are relatively equal in frequency versus unbalanced words
         3. Models for when the disambiguating information comes *before* ambiguous word
            1. Selective access model – prior information constrains access so only appropriate meaning is accessed
            2. Reordered access model – availability of the appropriate meaning is increased (more activation)
            3. Autonomous access model – context has no effect, but successful integration (called integration model) of one meaning with ambiguous word eliminates activation and search for other meanings.
      2. Exp for meaning frequency and context
         1. People were shown:
            1. Having been examined by the king, the *page* was soon marched off to bed (positive)
            2. Having been hurt by the bee-sting, the *page* was soon marched off to bed (negative)
            3. Just as Henrietta had feared, the *page* was soon marched off to bed (neutral)
         2. Predictions of models
            1. Integration model predicts that context will not help so all meanings will be activated

Therefore: *page* will be longer in the negative condition than positive and neutral

Later context will take longer in the positive and neutral conditions

You can’t successfully integrate in the negative condition so the search takes longer

* + - * 1. Reordered access model predicts context will have an effect

Processing the positive condition will be quick because the non-dominant meaning of the word page will be bumped up.

The negative and neutral conditions don’t help, so they will take longer

* + - 1. Reordered access model wins!
         1. Context sensitive model – meaning frequency and biasing context interact and work together.
  1. Also evidence that the characteristic of the ambiguous word affects interpretation of meaning.
  2. Summary:
     1. Autonomous view – automatically access all meanings, use context to select the appropriate meaning
     2. Interactive view – context enables selective access of appropriate senses
     3. Pattern of access depends on frequencies of meaning, extent of disambiguating context
     4. Both the characteristics of the ambiguous word and the type of disambiguating context matter