Week 7a. Cognitive/processing approaches

Cognitive approaches

- So far:
  - The state of knowledge (what they (can) know)
  - Is the form of an L2er’s knowledge like a L1er’s?

- Cognitive approaches:
  - Less concerned with what knowledge is acquired, more with how that knowledge is acquired.
  - Tend to view language learning as being learning of a complex skill, using general learning mechanisms (not specific to language).
  - More plausible for L2A (less so for L1A) but generally contrary to a “UG in L2A” view…

Today’s mixed bag of topics

- Information processing (McLaughlin)
  - Controlled vs. automatic processing
  - Fossilization, intake and input processing.
  - Pedagogical predictions?
- Competition model (MacWhinney & Bates)
- Learning strategies (O’Malley & Chamot)
- Connectionism (Sokolik & Smith)
- Operating Principles (Slobin, Anderson)

Information processing

- Central assumptions:
  - The mind is a general-purpose symbol-processing system.
  - The mind is a limited capacity processor.
- The mind as a slow computer with limited RAM.
- (This version primarily McLaughlin’s)

Information processing

- Refining the model further:
  - Conscious tasks require attention.
  - Attention is limited (by memory, processing power).
  - Well-practiced tasks become “automatic”.
  - Automatic processes no longer require resources, leaving the resources available for other conscious tasks.

Information processing

- Intuitively illustrated by learning to drive a standard transmission car.
- Initially, much conscious attention.
  - Clutch in, ease up on the accelerator, move gearshift to the next position in the reverse-N pattern, clutch out gradually while returning pressure to the accelerator.
- Eventually, an almost unconscious process; attention freed up to focus on things like finding the turnoff or finding a place to merge into traffic.
Information processing

• The information processing approach extends this intuition about what happens when you learn to drive a standard transmission car to what happens as you learn a second language.
  – You learn the rules (explicitly?).
  – You practice using them over and over.
  – Eventually, they become habit, “automatic”.

Controlled processing

• First step: controlled processing of information, conscious attention to processing.

• Limitations on attention: only one process can be controlled at once (otherwise, interference).

• Controlled processes are flexible, adaptable.
  – Can be altered in the face of contradicting evidence
  – Can be applied to novel situations.

• But, slow and resource-intensive.

Automatic processing

• Learning: transfer of controlled processes to long-term memory: automatic processes.

• In long-term memory, automatic processes:
  – Rapid
  – Do not tax working memory, attention (allowing these resources to be used elsewhere)

• But difficult to suppress or alter.

Fossilization

• An explanation for the (otherwise mysterious) phenomenon of fossilization in L2A.

• (Recall: Fossilization is reaching a plateau of grammatical performance where systematic errors persist regardless of further training. Idea: A process which has become automatic before it had been learned properly.)

Restructuring

• Controlled⇒Automatic, requires/prefers storing in a way that fits, leading to restructuring of the existing automatic processes.

• Example: U-shaped learning curve:
  – (automatic) unanalyzed forms (walked, went)
  – (automatic) analyzed forms (walk-ed)
  – Prompts reorganization, generalizing a regular rule for adding past tense ending -ed).
  – Then: over-regularization (go-ed).

Restructuring vs. fossilization?

• Automatic knowledge is frequently restructured during L2A, so why is it that systematic errors fossilize, again?

• Guess: automatic knowledge has developed to a point where restructuring is a) difficult (magnitude) or b) not required by new additions to automatic knowledge.
Restructuring

• To date: The accounts of what is involved in restructuring automatic knowledge are hopelessly vague.
  – Intuition
  – Qualitative shifts observed.
• But a vital area of the model for useful predictions about the end state.
  – Is this the stage where UG exerts its influence over L2A?
  – Only language systems of a certain “shape” can “fit” in the brain?
  Restructuring of automatic knowledge to maintain the shape of the knowledge?

Input ≠ intake

• Krashen: Learner must get comprehensible input (mixture of structures acquired and structures not yet acquired) to advance.
• Intake: Input that is used in grammar-building.
• VanPatten on input ⇒ intake
  – Intake (detection) of grammatical form requires attention to form.
  – Limited resources, competes with attention to content.

Comprehensible input

• We have already seen some evidence that L2ers tend to proceed through similar stages as they gain proficiency.
• Coupled with the idea of comprehensible input being required to advance, suggests there is an order in which structural information can be used.
• If true—can we tailor the input provided to L2ers to maximize its effectiveness by specifically provide input relevant to the next stage?

VanPatten’s Input Processing Model

• Learners process input for meaning before they process it for form
  – Content words first
  – Lexical (vs. grammatical/functional) items first
  – More meaningful morphology (vs. less) first
  – Less meaningful morphology (e.g., agreement) is at the bottom of the totem pole. It gets attention only if there is any left to spare, if the communicative meaning is clear with little or no cost to attention.

VanPatten’s Input Processing Model

• Learners possess a default strategy that assigns the role of agent to the first noun (phrase) they encounter in a sentence (word order).
  – Can be overridden by lexical semantics and event probabilities
  – Learners adopt other processing strategies for grammatical role assignment only after their developing system has incorporated other cues (case, stress)

Input ≠ intake

• Inuktitut—input:
• Qasuirsarvigssarsingitluinarpuq
  ‘Someone did not find a completely suitable resting place.’

Qasu-iir-sar-vig-ssar-si-ngit-luin-ar-nar-puq
Not place for find completely 3sg
Implications so far for instruction

• McLaughlin’s controlled ⇒ automatic knowledge model suggests that frequent use/repetition of grammatical constructions will speed the move to automatic processing and allow for new (controlled) processes to be undertaken. With additional risk of fossilization.
• VanPatten’s input processing model suggests that input can be molded to be more useful—by structuring the input in ways that make attention to meaning less of an issue.

Learnability/Teachability

• Pienemann explored this under the title Learnability/Teachability Theory.
• Looked at the L2A of German (ZISA), which appears to proceed in stages.
• Proposed that certain linguistic features of the target language are only learnable at certain times (after features of prior stages have been learned), and proposed certain theoretically-based methods of teaching (to move learners from one stage to the next).
• Almost nothing seems to affect the observed order of the stages—teaching out of order is not as effective.

Competition Model

• MacWhinney and Bates: essentially an expansion of the idea of a “default strategy” for interpretation (i.e. VanPatten’s first noun (phrase) = agent principle).
  – (Note: They don’t consider it as such, they consider it to be at the heart of how grammar works, but we need not interpret it that way)

• M&B identify several cues that people use to assign noun phrases to their roles in the sentence.
  – Word order (subject first)
  – Case marking (subjects and objects marked as such)
  – Subject-verb agreement
  – Animate agents are generally subjects

• M&B: learners can use all of these cues, but limited resources and differential cue perceptibility require them to prioritize.
• Different cue priorities are appropriate for different languages.
  – Case marking is not a reliable cue for English, is for Japanese
  – Word order is a reliable cue for English but is not for Italian or Japanese
  – Subject-verb agreement is not a reliable/easily perceptible cue for English but is for Italian

• Another point of M&B’s research concerned the transfer of use of cues appropriate in the L1 to the IL.
• Presupposes that L1 speakers have and use these cues in their L1 as well.
  – M&B interpret this as this being a fundamental property of L1 knowledge…
  – …but we can also interpret it as a means of dealing with ambiguity in the L1 when it arises and the grammar doesn’t make the call. May well be a learned skill. Squarely in the domain of L1 processing.
Competition Model

• Whatever the root cause, the results are possibly of interest from a pedagogical point of view, given the apparent transfer of cues from L1 to IL/L2.
  – Chinese: animacy
  – Dutch: case inflection/subject-verb agreement
  – English: word order
  – French: clitic pronoun agreement
  – German: subject-verb agreement/animacy
  – Italian: subject-verb agreement
  – Japanese: animacy
  – Spanish: subject-verb agreement

Competition Model

• One other hint that this might be really a processing strategy distinct from grammatical language competence is that there is evidence that bilingual speakers will sometimes combine the cue priorities of the two languages and use the same cues to interpret ambiguity in both languages.
• One study even found that native speakers of Chinese learning L2 (immersion at 12-16) exhibited reverse transfer, applying English cues to interpret Chinese sentences.

Learning strategies

• If attention is under conscious control, and attention plays a crucial role in converting input to intake, we might also expect that learning to focus attention certain ways (learning strategies) can assist in the L2A process.
• O’Malley and Chamot 1990:
  – Explored what learning strategies L2’ers employed
  – Differentiated successful strategies
  – Experimented with teaching strategies as a skill (intending to make the strategy automatic).
  – Seemed to yield only marginal rate improvements, though.

Learning strategies

• Somewhat related, connectionism is a theory of memory and learning in a computation model that is roughly modeled on the connections between neurons, modeling learning in abstractly the same way it appears to be done in the brain.
• Points of connection: frequency-based associations, an (in some ways) more detailed model of learning.
• Connectionism tends to be looked on (by its practitioners) even more than other approaches as an alternative to innate language structure.

Connectionism

• A connectionist system is actually very simple conceptually.
• It consists of a set of interconnected nodes (“neurons”).
• Each connection has a certain strength and polarity.

Connectionism

• Each node has an activation level and a threshold value.
Connectionism

- Each node has an activation level and a threshold value.
- When a node reaches the threshold level, it fires—and transfers its activation (additively or subtractively) along the connections.

Connectionism

- If this pushes a connected node over its threshold, it fires.

Connectionism

- If this pushes a connected node over its threshold, it fires.
- And so forth...
- Clearly, the interactions can quickly become mind-bogglingly complex.

Connectionism

- Certain nodes are designated as input nodes. These have an activation level driven by the perceptual system. So, maybe the node will be active if the currently perceive word starts with t…

Connectionism

- Other nodes are designated as output nodes. The status of these nodes determine the system’s reaction to the input.
- So, it’s a complex way to compute a function from input (patterns) to output (patterns).
Connectionism

• Finally, the learning aspect. The way neural nets are trained is generally:
  – Provide an input with a known “correct” output.
  – Check the output the system provides.
  – If the system’s output doesn’t match the correct output, adjust the connection weights in the network using a general “back-propagation” algorithm to make it come closer next time.

Connectionism

• After a lot of training, the neural net can produce the appropriate outputs from the given inputs.
  – The neural net has abstracted out the systematicity in the input data—but in ways that are almost always far too mathematically complicated to fathom.

Connectionism

• Then, when presented with novel inputs, the neural net will generalize its training to make decisions. This was previously considered to be a sure signal of following a rule.
  – When trained on rule-governed material, it tends to “follow the rule” even with novel forms. Neural nets are also great pattern recognizers—they latch onto any kind of statistical regularity. Medical diagnosis, image reconstruction, …

Connectionism

• What’s the point?
  – The point is that neural nets can “learn” rule-like behavior from statistical regularity without being taught the rule (and in fact without there even being a rule). There are just neurons and connections (vaguely like the human brain).
  – The learning has the same “practice reinforces” property attributed to controlled⇒automatic processing.

Connectionism and L2A

• Sokolik and Smith (1992) created a neural net to determine the gender of French nouns.
  – Trained from zero (connection weights set to zero), successfully generalized to be able to guess with a high degree of accuracy the gender of novel nouns.
  – Trained from a previous distribution (“L2”), learning this was less successful and slower.
  – So, do we now have the answer? A model of language and language learning in the brain?

Connectionism and L2A

• This French gender case is a fairly celebrated one in the world of connectionism and L2A.
  – But—what is the alternative to a connectionist view of this? Regular memory storage (e.g., memorization) is almost certainly done in some kind of associative network like these neural nets.
Connectionism and L2A

- There is a sense in which what we have got here is not much more interesting than a model of how we remember things that start with \( p \) vs. things that don’t. It’s not that it isn’t a problem, nor that it might not be interesting to look at interference effects from pre-existing (L1) associations, but it doesn’t get us anywhere near a model of actual grammatical rules in the brain.

Connectionism is too hard

- Connectionist research often has this property, looking at a very small problem on the boundary between grammatical knowledge and lexical memorization, where it is not at all clear that we could generalize the results to language as a whole grammatical system (or even come close to understanding what the network is even doing).
- At this point, connectionism is too hard—for a network large enough to do anything interesting and predictive, the generalizations it reaches will be completely inaccessible to us analyzing it from the outside.

Perceptual saliency

- Slobin (1979) proposed and pursued a view under which the universal characteristics of language (in this case, concerning L1A) are primarily due to our perceptual mechanisms and a need to organize the world around us. The idea is that to the extent that languages show universal properties it is because our language learning is molded by our (presumably innately specified in some sense) perception and organization of the linguistic input.

Perceptual saliency

- To this end, Slobin proposed a number of operating principles that children follow, including:
  - Pay attention to the ends of words
  - There are linguistic elements which encode relations between words.
  - Avoid exceptions.
  - Underlying semantic relations should be marked overtly and clearly.
  - The use of grammatical markers should make semantic sense.

A moment of skepticism

- As an aside, I don’t really understand what these are supposed to be buying us except pure description. First: Do children invariably follow these principles? Are they innate? Second: There are linguistic elements which encode relations between words? The use of grammatical markers should make semantic sense? How do kids even interpret these without some prior knowledge of what linguistic elements, words, grammatical markers are and what makes semantic sense?

Perceptual saliency

- Nevertheless, learning a language is learning a language in a certain, real sense—regardless of what’s “built in”. Languages differ and pretty much any kid can learn any language. Everyone needs to learn the words (or maybe roots) in the language and hear the morphological patterns to acquire the morphological system.
- Whatever we believe about the status of these principles in L1A (general learning principles about memorizing and storing/categorizing any kind of information?), however, they may still hold some (descriptive) truth in the acquisition of L2A.
Operating principles in L2A…

• Roger Anderson pursued this avenue for L2A, proposing three primary operating principles that underlie L2A, two of which are the One-to-one principle and the Relevance principle.

• One-to-one principle: An interlanguage system should be constructed in such a way that an intended underlying meaning is expressed with one clear invariant surface form (or construction).
  – Presumably means given the choice, L2er opts for one form per meaning. Initially.

• Relevance principle: If two or more functors apply to a content word, try to place them so that the more relevant the meaning of a functor is to the meaning of the content word, the closer it is placed to the content word. If you find that a notion is marked in several places, at first mark it only in the position closest to the relevant content word.
  – I take “functor” to be a derivational morpheme. How do you judge “relevance” in a non-circular way? Note that if closer to the content word means “lower in the tree” then we have a similar prediction but made on much more stable structural grounds in Vainikka & Young-Scholten’s proposal.

Cognitive approaches in summary

• Views L2A from the perspective of psychology and operational models of learning.

• Practitioners often do—but we need not—presume that the problem of language acquisition (L1A/L2A) can be completely solved by proposing the right general learning theory.

Where we ended up in our mixed bag of topics

• Information processing (McLaughlin)
  – Controlled vs. automatic processing is very successful in modeling novice⇒expert shifts in complex skills. Little puts L2A in this category except intuition and the suggestive evidence from fossilization.
  – Processing input into intake and the idea that progress requires comprehensible input makes interesting pedagogical predictions, which for one reason or another remain to be tested.

• Competition model (MacWhinney & Bates)
  – We can take the influence of cues to be characterizations of how L1 ambiguity is processed, which is applied (transferred) to the L2A task when “ambiguity” is more severe.

• Learning strategies (O’Malley & Chamot)
  – An intriguing idea but with very minimal results in the end.

• Connectionism (Sokolik & Smith)
  – Overly ambitious; the only problems of a manageable size are not of much interest in L2A and/or won’t generalize reliably to broader language phenomena.

Operating principles in L2A…

• Relevance principle: If two or more functors apply to a content word, try to place them so that the more relevant the meaning of a functor is to the meaning of the content word, the closer it is placed to the content word. If you find that a notion is marked in several places, at first mark it only in the position closest to the relevant content word.
  – I take “functor” to be a derivational morpheme. How do you judge “relevance” in a non-circular way? Note that if closer to the content word means “lower in the tree” then we have a similar prediction but made on much more stable structural grounds in Vainikka & Young-Scholten’s proposal.

Cognitive approaches in summary

• L1A pretty clearly cannot be accounted for under a general learning theory, but many aspects of L1A are likely to be susceptible to a cognitive-style explanation. Ambiguity resolution, memorization of words and their properties (irregular forms). Depending on whether “UG is involved in L2A”, much of L2A may also be in the domain of general learning theories as well.

• …But it seems that evidence of Subjacency is still not in the input, even for the fanciest current connectionist model.
Where we ended up in our mixed bag of topics

- Operating Principles (Slobin, Anderson)
  - The status of these “principles” is far from clear, but they appear to be weakly predictive descriptions of things we could probably explain through UG principles of grammar and principles of knowledge organization.