

Lecture notes 3.2 cognition and SLA

Points of highlight

- Information processing theories assume that: (a) the human cognitive architecture is made of representation and access; (b) mental processing is comprised of automatic or fluent (unconscious) operations and voluntary or controlled (conscious) operations; (c) cognitive resources such as attention and memory are limited; and (d) performance is variable and vulnerable to stressors, as shown during dual-task performance.
- Memory is composed of two types, long-term memory and working memory, which interact.
- Long-term memory is about representations and it is unlimited; it can be explicit-declarative (the facts and events we know and can talk about) or implicit-procedural (knowledge we do not know we hold, but which affects our behaviour and supports our skills, habits and performance). Another distinction that has received less attention in SLA is between semantic memory (decontextualized knowledge) and episodic memory (knowledge encoded with information about the lived experiences in which we acquired it).
- How L2 vocabulary is encoded in long-term memory has resulted in research about the strength, size and depth of word knowledge, as well as various theories about the content of the entries in the mental bilingual lexicon and how the L1 and L2 lexicons interact during lexical access.
- Working memory is about access and it is limited; it controls what information can be stored momentarily (the issue of storage) and how well and how long it can be activated and integrated with already known information in long-term memory (the issue of processing); it is the site for controlled processing and consciousness.
- Attention is thought to be central to understanding L2 learning. Attention is limited and selective, and it can be voluntary and accessible to consciousness. Several questions have been investigated in SLA with regard to attention.
- Is L2 learning possible without rules? In the absence of rules, low-level associative learning is certainly possible. This kind of learning draws on data-driven processes supported by memory. With rules, learning proceeds by drawing on controlled operations and conceptually driven processes supported by conscious attention. Both types of processing of new L2 material can lead to learning, and both can interact. A pending question for future research is whether all aspects of an L2 are equally learnable by implicit means or whether some particularly complex aspects of the L2 may require conceptually driven processing in order for associations and representations to be formed.
- An imminent emergentist turn has made inroads into several research programmes about cognition in SLA. Emergentism refers to a contemporary family of theories in cognitive science that reconceptualizes information processing as an associative, probabilistic, rational, usage-based, grounded, dynamic and, in sum, emergent adaptation of the agent to the environment.

- It is difficult to predict how long it will take for emergentist SLA to really come to fruition and so far the publications offer discussions and expositions more than they offer L2-specific empirical evidence. Emergentism will likely flourish in cognitive SLA in future years, however, judging from the pervasiveness of emergentist thought in many other areas of the cognitive sciences.

Introduction

Cognition refers to how information is processed and learned by the human mind (the term comes from the Latin verb *cognoscere*, 'to get to know').

Key question:

what it takes to 'get to know' an additional language well enough to use it fluently in comprehension and production?

Limited knowledge in this field:

our capacities to investigate the relevant questions are shaped by the pace at which new theories and methods to inspect the workings of human minds and brains become available (typically in neighboring disciplines) and the rate at which SLA researchers become conversant in them.

Evidence often used

Behavioural and neurobiological evidence

Key theories and issues:

Information processing; emergentism

Memory and attention in L2 learning

1. INFORMATION PROCESSING IN PSYCHOLOGY AND SLA

Beginning in 1970s

- What is a IP?

In a nutshell, the human mind is viewed as a symbolic processor that constantly engages in mental processes. These mental processes operate on mental representations and intervene between input (whatever data get into the symbolic processor, the mind) and output (whatever the results of performance are).

Performance vs behavior

Performance, rather than behaviour, is a key word in information processing theories. This is because inferences about mental processes can only be made by inspecting what is observable during processing while performing tasks, rather than by inspecting external behaviour in response to stimuli, as behaviourists used to do.

- Key assumption in IP used in current SLA research about cognition

First, the human cognitive architecture is made of representation and access.

Second, mental processing is dual, comprised of two different kinds of computation: automatic or fluent (unconscious) and voluntary or controlled (conscious).

Third, cognitive resources such as attention and memory are limited.

- Representation (knowledge) vs access (processing)

'knowing what is in the library, plus how the contents are classified and related to one another, must be distinguished from retrieving desired information from the books at a given time'

Representation (knowledge)

Linguistic representation is comprised of three kinds of knowledge: grammatical, lexical

and schematic or world-related.

New L2 knowledge is stored in the mind and has to be accessed and retrieved every time it is needed for use in comprehension or production.

Access: automatic vs controlled processing

Automatic processes require small effort and take up few cognitive resources, and therefore many automatic processing routines can run in parallel. During automatic processing, cognitive activation is triggered bottom up by exogenous sources in the environment (something outside the processor, that is, some aspect of the data in the input or environment).

controlled processing is activated by top-down, endogenous sources (by something inside the processor, that is, by voluntary, goal-directed motivation in the individual's mind), and it is handled by what we call the central executive.

Controlled processes therefore allow us self-regulation, but they require a lot more effort and cognitive resources than automatic processes, and thus cannot operate in parallel; they are serial.

subject to a bottleneck effect

limited capacity model of information processing

predictions: performance that draws on controlled processing is more variable and more vulnerable to stressors than performance that draws on automatic processing. (dual task to investigate automaticity)

2. THE POWER OF PRACTICE: PROCEDURALIZATION AND AUTOMATICITY

Skill acquisition theory (a particular kind of IP)

Adaptive Control of Thought theory (Anderson, 1983)

Key argument:

learning as the gradual transformation of performance from controlled to automatic relevant practice over many trials enables controlled processes gradually to be withdrawn during performance and automatic processes to take over the same performance—proceduralization or automatization (the conversion of declarative or explicit knowledge (or 'knowledge that') into procedural or implicit knowledge (or 'knowledge how'))

how does practice work?

(1) Practice allows the establishment and strengthening of corresponding links in long-term memory. The more this knowledge is accessed via practice, the easier it will become to access it without effort and without the involvement of the central executive at a future time.

(2) the power of practice is not constant over time.

power law of learning: practice will at some point yield no large returns in terms of improvement, because optimal performance has been reached (Ellis and Schmidt, 1998)

(3) proceduralization is skill-specific.

(4) The final outcome of the gradual process of proceduralization or automatization is

automaticity, which is defined as automatic performance that draws on implicit-procedural knowledge and is reflected in fluent comprehension and production and in lower neural activation patterns (Segalowitz, 2003).

Two misinterpretations of skill acquisition tenets

- (a) automaticity is simply accelerated or speedy behavior
response: a qualitative change is reached once performance is automatized
- (b) that L2 learners simply accumulate rules that they practise until they can use them automatically
response: repeated practice changes the knowledge representation itself by making the stored knowledge become more elaborated and well specified, or more analysed. by the time they become automatized, rules may be just different from the declarative rules that were initially committed to memory.

An exemplary study of skill acquisition theory in SLA: DeKeyser (1997).

3. Long term memory

Long-term memory is about representation. It is virtually unlimited in its capacity and it is made of two kinds: explicit-declarative memory and implicit-procedural memory.

explicit-declarative memory

verbalizable and consciously recalled. Explicit-declarative memory supports recollection of facts or events, and it is served by the hippocampus in the human brain

implicit-procedural memory

things that we know without knowing that we know them. Implicit-procedural memory supports skills and habit learning, and it is served by the neocortex in the human brain

another important difference

semantic and episodic memory (see Tulving, 2002)

Semantic memory pertains to relatively decontextualized knowledge of facts that 'everyone knows'. Episodic memory involves knowledge of the events in which people are personally involved or 'the events we've lived through'.

4. LONG-TERM MEMORY AND L2 VOCABULARY KNOWLEDGE

Key question: What does it mean to remember a word?

established in long-term memory when the link between a form and its meaning is made

the strength, size and depth of the knowledge represented in memory

- Vocabulary knowledge strength
- degree of proceduralization in implicit memory

findings:

learners know more words receptively than productively, particularly if they are infrequent or difficult words, and that this gap becomes smaller as proficiency develops

- Size of the mental lexicon

the total number of words known and represented in long-term memory

Size is often related to the relative frequency with which words are encountered in the input that surrounds learners, since high-frequency words usually make it into long-term memory earlier in the learning process than low-frequency words.

interesting findings about vocabulary size in L2 learning (e.g. Nation and Waring, 1997; Nation, 2006)

in L1 a five-year-old child begins school with an established vocabulary of about 5,000 word families, and a typical 30-year-old college-educated adult ends up knowing about 20,000 word families (Nation and Waring, 1997).

For L2 users, new vocabulary presents a formidable challenge. They need to learn about 3,000 new words in order to minimally follow conversations in the L2, and about 9,000 new word families if they want to be able to read novels or newspapers in the L2 (Nation, 2006).

- Vocabulary depth

resides in the realm of both explicit and implicit memory and refers to how well the known words are really known, that is, how elaborated, well specified and structured (or how analysed, in Bialystok's 2001 sense) the lexical representations are.

Depth of knowledge includes whether L2 learners know how a word sounds (/di-'zərt/ in a meal and /'de-zərt/ in a landscape), how it is spelled (exude, not exhude), how many other word parts it can appear with (pre-, -ment, -er, -s, -ing), what is likely to precede or follow a word (make a decision, do exercise; mental state, state of affairs/mind), how many meanings the word may have (demonstrate = to show and to protest), in what registers different synonyms may be preferred (weather, climate), or how frequently and in what contexts the word will occur (many oaths in court but not in hospitals, many incisions in hospitals but not in court).

- Other long-term memory-related questions in L2 vocabulary content of the lexical representations encoded in memory for bilinguals?

5. Working memory

working memory is about access and is limited

- definitions:

Nick Ellis (2005): 'If I ask you what $397 * 27$ is, you do not look up the answer from long-term memory, you work it out' (p. 338). Peter Robinson (1995) describes it as 'the workspace where skill development begins...and where knowledge is encoded into (and retrieved from) long-term memory' (p. 304)

- functions of working memory

to hold information (a storage function) as well as to integrate new information with known information already encoded in long-term memory (a processing function).

Working memory handles automatic and controlled processing. Importantly, thus, it is the site for the executive control, which supports controlled processing (Baddeley and Hitch, 1974), and also the site of consciousness (Baars and Franklin, 2003).

As Nick Ellis (2005) explains, working memory 'is the home of explicit induction, hypothesis formation, analogical reasoning, prioritization, control, and decision-making.

It is where we develop, apply, and hone our metalinguistic insights into an L2. Working memory is the system that concentrates across time, controlling attention in the face of distraction' (p. 337).

- Two characteristics of working memory
limited capacity: under normal conditions information can be remembered in working memory for about two seconds only

temporary activation

working memory is just the part of memory that becomes activated during a processing event (Cowan, 2005)

- work memory related studies in SLA
individual differences
hypothesis: since memory is involved in information processing in pervasive ways, people who have better working memory capacities can learn an L2 more efficiently. Thus, working memory capacity is posited to help predict learning rate and ultimate levels of attainment in the L2.

Two facts about L2 memory capacity

working memory capacity is smaller in the L2, when compared to the L1.

as L2 proficiency develops, this lag in working memory capacity between the L2 and L1 should become smaller

6. HOW WORKING MEMORY IS MEASUREED?

Two approaches

- Memory as storage: passive working memory tasks

Digit span recall tasks are one of the oldest methods used in psychological research to measure storage capacity. Put simply, participants are asked to repeat increasingly longer sequences of numbers, sometimes in backward order. Great care is taken to minimize opportunities for subvocal rehearsal, since this is a good strategy to 'stretch' memory capacity. As Miller (1956) concluded in a seminal paper, average-memory adults have a working memory span in their L1 of about seven digits, which means that they can remember sequences of about seven digits accurately at least 50% of the time

Word span tasks are also frequently used. Most people can repeat sequences of five to six unrelated words, but after that they experience increasing memory difficulties. Rather than thinking in terms of digits or words, however, it is more accurate to think of memory capacity in terms of chunks, or pieces of information that are already linked and stored together in long-term memory. For example, you will probably be able to remember equally well the string '*Nicole, Gary, Tom, Katherine, Penélope, Sean*' and the double-length string '*Nicole Kidman, Cary Grant, Tom Cruise, Katharine Hepburn, Penélope Cruz, Sean Connery*'... that is, if you know these actors' names well and have 'chunked' their first and family names together for each of them

Non-word repetition span tasks are preferred to word span tasks by some researchers, precisely to eliminate recall-enhancing strategies that draw on long-term memory, like grouping and chunking. Examples of non-words are *johmbe, zabide, wakime, migene, shosane, tisseke, chakume* and *nawase* (taken from an L2 study by Williams, 2005) and *lus, vip, kug, taysum, kepponen, woogalamic* and *reutterpation* (taken from an L1 study by Gathercole et al., 1999)

Sentence repetition tasks (also called elicited imitation tasks) are another way to measure working memory storage capacity. We know that the typical human memory span for sentences is about 16 words, a sequence much longer than the typical word span of five or six isolated words. This is because in sentences we perceive words as 'chunked' into phrases. Our knowledge of grammar (which is stored in long-term memory) helps us group words into phrases and remember sentences better than isolated sequences of digits or words

- Memory as dynamic processing: active working memory tasks

Reading span task:

In these tasks, people are asked to read sentences presented on cue cards and to comprehend or evaluate what they are reading. Immediately after reading all sentences, they are asked to recall the last word appearing on each sentence or the few selected words that were underlined on each card. This kind of reading span task is assumed to reflect how well people can maintain information in short-term memory (the targeted words they are asked to recall out of the sentence stimuli) while another processing task is executed simultaneously (reading the sentences for comprehension).

- Nature of working memory

'[working memory] capacity is not about individual differences in how many items can be stored per se but about differences in the ability to control attention to maintain information in an active, quickly retrievable state. [It] is not directly about memory – it is about using attention to maintain or suppress information. Greater [working memory] capacity does mean that more items can be maintained as active, but this is a result of greater ability to control attention, not a larger memory store'

7. ATTENTION AND L2 LEARNING

- What is attention?

attention is another essential component of cognition. Remember that under normal conditions simple activation of a stimulus in working memory will last for a few seconds and then fade away. Here is where attention comes in; it heightens the activation level of input in working memory, allowing it to remain there for longer through rehearsal and thus making it available for further processing and for entering long-term memory.

- Features of attention
 - 1) its capacity is limited
 - 2) selective
 - 3) voluntary (it can be subject to cognitive, top-down control that is driven by goals and intentions of the individual)
 - 4) controls access to consciousness (Under normal conditions, participants can tell researchers about their conscious perceptions, thoughts or feelings while they attend to some aspect of a task, through the method of thinking out loud (Ericsson and Simon, 1993) or via some other retrospective method, as discussed in detail by Gass and Mackey (2000) in relation to SLA.)

- key question in relation to SLA:
what quality of attention is necessary for L2 learning?

Learning under three attentional conditions: incidental (i.e. learning without intention, while doing something else), implicit (i.e. learning with no intervention of controlled attention, usually without providing rules and without asking to search for rules) and explicit (i.e. learning with the intervention of controlled attention, usually summoned by the provision of rules or by the requirement to search for rules)

8. SYMBOLIC VS ASSOCIATIVE LEARNING: Robinson (2009)

what would happen if no explicit information were made available to L2 learners at the outset of training?

Conclusion: in the absence of rules, low-level associative learning that draws on data-driven processes supported by memory is certainly possible. Learning without rules leads to the formation of memories of instances that can be accessed more easily, allowing for faster performance, but without knowledge that can be generalized to new instances. That is, without the initial provision of rules (without an explicit learning condition), learning is bottom-up (i.e. data and memory driven), and it does not lead to knowledge of a systematic rule of some kind. With rules, learning proceeds by drawing on high-level attention and conceptually driven processes supported by conscious attention, resulting in generalization with awareness.

9. AN EMERGENT TURN IN SLA?

- A manifesto of emergentism is SLA
Emergentists believe that simple learning mechanisms, operating in and across the human systems for perception, motor-action and cognition as they are exposed to language data as part of a communicatively-rich human social environment by an organism eager to exploit the functionality of language, suffice to drive the emergence of complex language representations.

Ellis and Larsen-Freeman, 2006: 577

- Three important tenets of emergentist approach
 - 1) associative learning, probabilistic learning and rational contingency (Ellis, 2006)
 - 2) Associative learning means that learning happens as we form memories of instances or exemplars we experience in the input, in a process of automatic extraction of statistical information about the frequency and sequential properties of such instances.
 - 3) Probabilistic learning posits that learning is not categorical but graded and stochastic, that is, it proceeds by (subconscious) guesswork and inferences in response to experience that always involves ambiguity and uncertainty (Chater and Manning, 2006)
 - 4) The probability calculations of the human mind are guided by principles of rational contingency, or automatically computed expectations of outcomes on the basis of best possible evidence (Chater and Manning, 2006; Ellis, 2006b).

- Additional tenets

Usage-based learning: the position that language use and language knowledge are inseparable, because we come to know language from using it.

- Sum up: Second language learning under an emergentist perspective has the potential to look less like development that proceeds teleologically towards the ultimate attainment of a so-called native grammar, and more like a complex deployment of human multi-language capacities as a function of experience in the world. The new approach, by redefining cognition as emergent, helps envision additional language acquisition less as a formal, deterministic and symbolic feat and more as an ecological phenomenon, ‘a dynamic process in which regularities and system emerge from the interaction of people, their conscious selves, and their brains, using language in their societies, cultures, and world’ (Ellis, 2007, p. 85).