COGNITIVE APPROACHES TO SLL

1. INTRODUCTION

Universal Grammar-based researchers,
are interested in the development of L2 grammars from a purely linguistic point of view:
- description of the linguistic systems
- emphasis on the *language* dimension of SLL
- interested in property theories

see language as a separate innate module distinct from other aspects of cognition

Cognitivists,
investigate hypotheses which come from the field of cognitive psychology and neurology
- emphasis on the *learning* component of SLL
- interested in transition theories

view SLA as one instantiation (or working example) of learning
believe we can understand the SLA process by understanding how the human brain processes info → cognitive psychology and neurology

> **Note:** UG theorists are interested in competence; cognitive theorists are concerned with how learners access this linguistic knowledge in real time, or in the strategies they might employ when their incomplete linguistic system lets them down, etc.

2. PROCESSING APPROACHES

Scholars
- may or may not believe language is a separate module
- concerned to develop transition or processing theories to complement property theories
- focus primarily on the computational dimension of language learning, that is:
  - the way in which the brain's processing mechanisms deal with the L2
  - how their ability to process the L2 develops over time

2.1. Information Processing Models of L2 Learning

Models in this approach investigate
- how different memory stores (sensory memory; short-term memory; long-term memory) deal with L2 information
- how this information is automatized and restructured through *repeated activation*

Basic assumptions:
- Humans are viewed as autonomous and active.
- The mind is a general-purpose.
- Complex behavior is composed of simpler processes.
- Component processes can be isolated and studied independently of other processes.
- Processes take time.

- The mind is a limited-capacity processor.

2.1.1. Attention-processing model

Learners first resort to **controlled processing** in the L2 which

- involves the temporary activation of a selection of information nodes in the memory
- requires a lot of attentional control
- is constrained by the limitations of the short-term memory
- is typical of anyone learning a brand new skill

Through repeated activation processing become **automatic**

Automatized sequences

- stored as units in the long-term memory → available very rapidly, unconsciously, and effortlessly whenever the situation requires it
- require minimal attentional control on the part of the subject → automatic processes in parallel

Note: The distinction between controlled and automatic processing is one of routinization NOT one of conscious awareness

Attention to formal properties of language	*Information processing*	
	Controlled	*Automatic*
Focal	(Cell A) Performance based on formal rule learning	(Cell B) Performance in a test situation
Peripheral	(Cell C) Performance based on implicit learning or analogic learning	(Cell D) Performance in communication situations

Learning: movement from controlled to automatic processing via practice (repeated activation) → through this shift, controlled processes are freed to deal with higher levels of processing → incremental nature of learning

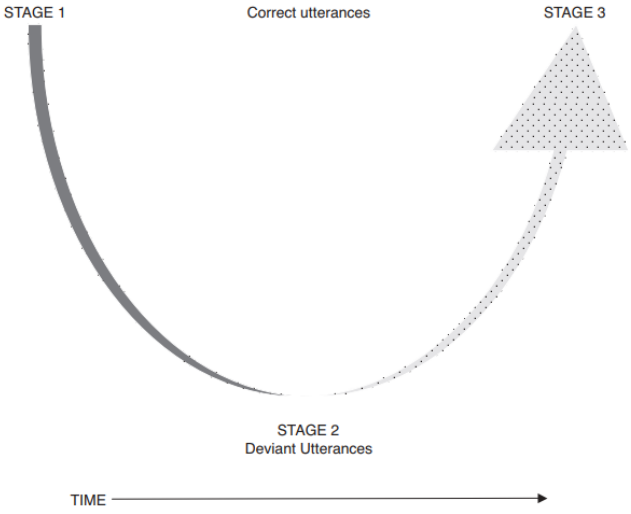
Learning & **restructuring**:

Learning is the inclusion of additional information → make changes to parts of the existing system or reorganizing the current system, i.e., restructuring

changes are discontinuous or qualitatively different from a previous stage

Time 1	Time 2	Time 3	Time 4
I am no go.	I am no go.	I am no go.	I am no go.
No look.	No look.	Don't look.	Don't go.
I am no run.	I am don't run.	I am don't run.	I am no run.
No run.	Don't run.	Don't run.	Don't run.

Restructuring account for some of the variability characteristic of learner language: restructuring destabilizes some structures in the interlanguage:



2.1.2. Active Control of Thought (ACT) model

Practice: **declarative** knowledge (i.e., knowledge *that* something is the case) → **procedural** knowledge (i.e., knowledge *how* to do something).
essential differences between them:

- Declarative knowledge in an all-or-none manner VS. procedural knowledge can be partially possessed
- Declarative knowledge acquired suddenly, by being told VS. procedural knowledge acquired gradually, by performing the skill
- Declarative knowledge communicated verbally VS. procedural knowledge cannot

Move from declarative to procedural knowledge:

- **Cognitive stage:**
  a description of the procedure is learnt, that is, explicit knowledge is stored as facts: e.g. information from a teacher + concrete exemplars
  a lot of attentional control required
- **Associative stage:**
  a method for performing the skill is worked out
  proceduralization
    is achieved after a few trials
    reduces demands on working memory
  errors are likely during the associative stage
  knowledge at this stage is prone to restructuring
- **Autonomous stage:**
  the skill becomes more rapid and automatic

extensive practice needed → decrease time required to perform the skill, error rate, amount of attention required
automatized knowledge outside attentional control → difficult to change or delete

2.1.3. Skill acquisition theory (SAT)

Learning shows a similarity in development:
initial representation of knowledge → initial changes in behavior → eventual fluent, spontaneous, largely effortless, and highly skilled behavior

- **Cognitive, declarative or presentation stage:** learners establish some new explicit knowledge
- **Associative, procedural or practice stage:** information about how to do something is put into action and the skill is performed.
- **Autonomous, automatic, or production stage:** before proceduralized knowledge can be used quickly, a great deal of automatization has to take place through extensive practice

Power law of practice

In SLL: communicative practice serves as a device for proceduralizing knowledge of linguistic structures that have been first presented declaratively

Practice needs to be skill-related → procedural knowledge is context- and skill-specific and cannot easily be transferred to other tasks → uni-directionality or specificity of the practice effect
Declarative knowledge is transferable to other contexts, skills, or tasks → bi-directionality (i.e., can be utilized in the development of different skills).

Problems of SAT:

- no explanation for the orders and sequences of acquisition
- insisting that all knowledge starts out in declarative form

Skill acquisition theory provides accounts for at least six key SLA phenomena:

(1) *Why some structures never seem to enter the interlanguage at all?*
(2) *Why native-like forms are used in some contexts but not others?*
(3) *Why learning is incremental?*
(4) *Why there are differences between individual learners?*
(5) *Why there is fossilization?*
(6) *Why some structures are more likely to fossilize than others?*

2.2 Theories of Second Language Processing

Explore the factors controlling the way in which L2 learners process the linguistic input

2.2.1 Input processing (IP)

working memory is limited in capacity → it is difficult for learners to attend concurrently to different stimuli in the input → main concern: how learners allocate attentional resources during online processing = what causes learners to detect certain stimuli in the input and not others

This area of research deals with how learners

parse sentences → assign form-meaning relationships → comprehend utterances → convert L2 input into intake

linguistic data processed from the input and held in working memory for further processing

Input processing does not offer

- a complete model of normal processing of input
- any explanation of how intake becomes integrated into the developing IL system

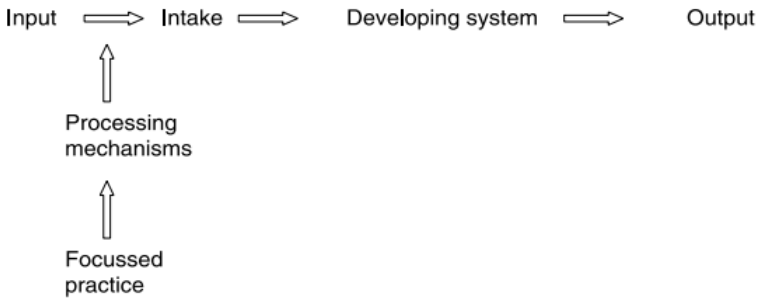
IP theory offers a set of principles that explain the apparent failure of L2 learners to process the linguistic forms encountered in L2 input:

1) *The Primacy of Meaning Principle:*
   a) *The Primacy of Content Words Principle:* e.g., The **cat** is **sleep**ing.
   b) *The Lexical Preference Principle:* e.g., I studi**ed** well **yesterday**.
   c) *The Preference for Non-redundancy Principle:* e.g., The cat is sleep**ing** vs. The **cat** sleep**s** ten hours **every day**.
   d) *The Meaning Before Non-meaning Principle:* e.g., Mary thinks **that** he is smart vs. He loves **that** girl.
   e) *The Availability of Resources Principle*
   f) *The Sentence Location Principle*

2) *The First Noun Principle*
   a) *The Lexical Semantics Principle:* e.g., The fence **kicked** the horse
   b) *The Event Probabilities Principle:* e.g., The child **scolded** the mother
   c) *The Contextual Constraint Principle:* e.g. John is in the hospital because **Mary attacked him** (lo ataco Maria)

Processing instruction – explicit and input-based – takes as its basis how learners process input and make form-meaning relationships: → If we know what learners are doing wrong at the level of input processing, can we create pedagogical intervention to push them away from non-optimal processing?

Traditional approach to teaching grammar



IP approach to teaching grammar

➔ IP attempts to deal with not just a linguistic difficulty, but with a problematic processing strategy

2.2.2. Autonomous induction theory (AIT)

Suzanne Carroll reminds us that the understanding of SLA processes requires:

- A theory of the representation of language in the mind (i.e., property theory)
- A theory of how language is processed, both receptively and productively
- A theory of changes in linguistic competence (i.e., transition theory)

Jackendoff:

Integrative processors integrate primitives of some level of representation into complex structures → Carroll claims our mental representations of language involve a number of distinct modules (prosodic structures, morphosyntactic structures, conceptual structures)

Correspondence processors link the distinct autonomous representations → Carroll states these processors have as their input one level of representation (say phonology) and their task is to map it onto or correspond it to units of a level higher up (say morphosyntax).

Carroll proposes a version of inductive learning (i-learning) → initiated when we fail to parse incoming language stimuli adequately using our existing mental representations = language acquisition device is triggered when the parsing system fails, e.g., *That's the cat whom the dog bit*

- **Note:** Adjustments lead to learning
- **Note:** Parsing presupposes linguistic competence; acquisition presupposes the absence of it

as children learn their L1, they develop processing procedures that are tuned to the specific grammatical properties of the L1 → language users reveal preferences for parsing sentences in particular ways even when other analyses are grammatically possible

Filter hypothesis

preferences transfer

fossilization

UG is necessary to explain how a learner comes to have a representational system capable of encoding phonological and morphosyntactic information → though children are not empiricist learners, adults learn L2 empirically

2.2.3. Nativization model

According to Andersen's model, L2 acquisition consists of two general processes:

Nativization:

learners make the input conform to their own internalized view of what constitutes the L2 system ←→ they simplify the learning task by forming hypotheses based on knowledge that they already possess ➔ pidginization

Denativization:

learners accommodate to an external norm ←→ they adjust their interlanguage systems to make them fit with the input, making use of inferencing strategies ➔ depidginization

2.2.4. Multidimensional model - Meisel, Clashen and Pienemann

L2 acquisition has two sides:

developmental sequence → unaffected by aspects of the learner or of the environment ; dependent on general factors of language processing

variational sequence → responds to differences in the learner or the situation ; based on learner variables, e.g. the extent to which the learners are integrated into the target culture

The developmental sequences reflect the systematic way in which learners overcome processing constraints (of a general cognitive nature) → explains why learners pass through stages

Canonical order strategy: no permutation of constituents in a structure, e.g. You have a cat?

Initialization/Finalization strategy: no permutation involving the movement of an element from within a structure occurs, but movement of an initial element in a structure to final position and vice versa, e.g. Where the cats are?

Subordinate clause strategy: no permutation of any element in a subordinate clause but movement of an element from within a main clause to another position in the clause, e.g. Can you say where the cat is?

Note: The three strategies are hierarchical

The variational sequences reflect the overall orientation to the learning task which is the product of social-psychological factors:

Segregative orientation → lack of interest in contact with native speakers, discrimination on the part of native speakers, and a general lack of motivation

Integrative orientation → learners with either a strong desire to assimilate into the L2 culture or in those with a strong instrumental need to learn the L2

2.2.5 Processability theory

Requirements:

processing component (based on Levelt's (1989) model of language generation)

theory of grammar (Lexical Functional Grammar, Halliday, 1985)

Pienemann believes:

language acquisition is the gradual acquisition of computational mechanisms → how learners acquire the computational mechanisms that operate on the linguistic knowledge

production and comprehension of L2 forms can only take place to the extent that they can be handled by the linguistic processor

The processing challenge is that learners must learn to exchange grammatical information across elements of a sentence → **feature unification**

This psychological process takes place in the following phases:

- the identification of grammatical information in the lexical entry
- the temporary storage of that information
- its utilization at another point in the constituent structure

The view on language production is based on Levelt (1989). Premises:

- Processing components are generally not consciously controlled (i.e., speakers do not need to be aware of the grammatical structures they produce)
- Processing is incremental (i.e., the speaker can start producing an utterance without having planned all of it)
- The output of the processor is linear, although it may not be mapped onto the underlying meaning in a linear way (for instance, the idea produced first does not need to occur first in natural events, e.g., *Before I drove off, I started the engine*).
- Grammatical processing has access to a temporary memory store that can hold grammatical information

The ability to match features across elements in a sentence develops gradually = Learners are claimed to have a **Hypothesis Space**, which develops over time:

- Level 1: lemma access, e.g., producing a simple word such as *play / Where is my book?*
- Level 2: category procedure; lexical morphemes
- Level 3: phrasal procedure, e.g., matching gender between *Det* and *N*
- Level 4: simplified S-procedure; exchange of information from internal to salient constituent
- Level 5: S-procedure; inter-phrasal morphemes; exchange of information between internal constituents, e.g., subject-verb agreement
- Level 6: subordinate clause procedure

→ learners will be able to share information across elements in a sentence in gradually less local domain:

Teachability hypothesis:

- L2 learners follow a fairly rigid route in their acquisition of certain grammatical structures → *Stages of acquisition cannot be skipped through formal instruction*
- L2 learners can only operate within their Hypothesis Space → *Instruction will be most beneficial if it focuses on structures from the next stage*

2.2.6. The efficiency-driven processor

Provide explanations for language acquisition by reference to more basic non-linguistic factors, e.g. physiology, perception, working memory, etc. → no innate linguistic constraints on the language processor

William O'Grady (2005) presents an account in which language structure arises from an *efficiency-driven online linear* computational processor.

Two different cognitive systems:

Lexicon → as a repository of information about a language's words and morphemes, including information about:

category membership (N, V, etc.)
combinatorial propensities (co-dependencies)

Co-dependencies could be thought of as the glue that holds a sentence together to create a meaningful proposition, e.g. *drink:* V, <N N> → *Mary drank tea*.

How co-dependencies are to be resolved?

Computational system → operates on these words and morphemes, combining them in particular ways to construct phrases and sentences

These computational operations are carried out by working memory → *Minimize the burden on working memory* → the computational system should operate in the most efficient manner → *Co-dependencies must be resolved at the first opportunity*:

It strives to interpret incoming language and produce outgoing language so that as little information as possible needs to be stored. Production:
[Mary drank] → [Mary [drank tea]]

(In comprehension) processing costs are caused by

having to revise an interpretation and so reactivate representations within WM
not knowing which elements of a sentence resolve a co-dependency
having an item left unresolved in the sentence

EDP and Language acquisition. Language acquisition is about the creation of mappings between form and meaning. *Mary drank tea*

1. Interpret the first nominal (Mary)
   MARY
2. Access the meaning of the transitive verb drink; find its agent argument (MARY) to the left
   DRINK <agent: MARY>
3. Interpret the nominal to the right (TEA); treat it as the verb's patient argument
   DRINK <agent: MARY; patient: TEA>

Mary drank tea. ← form (sentence)
↓↑
DEANK <agent: MARY; patient: TEA> ← meaning

The same sequence of three operations is repeatedly activated, ultimately forming a **computational routine** → routines improve the speed and efficiency of the processor as they are gradually strengthened

Processing amelioration: improvements to the speed and efficiency of the processor, via the creation and strengthening of processing routines:

Development: as computational routines are formed
Acquisition: as computational routines become entrenched

Overgeneralization → the computational routine of producing *–ed* becomes automatized, making it less costly to overgeneralize than to inhibit the automatic routine
Interference → entrenched computational routines block or inhibit other routines developing: SVO to SOV

3. EMERGENTISM

L2 learning is **bottom up** → language learning taps into the same, general, cognitive mechanisms that drive basic human learning in order to extract structures and patterns from the language input

Usage-based view of development → knowledge of language is created and strengthened in response to opportunities to interpret and/or form utterances in the course of communication

Learning → extraction of meaningful patterns from the language input they are exposed to → formal aspects of language **emerge** from language use and experience, rather than being either innate, or learned as abstract structures

Note: There is no real way to explain how people come to know more than what they are exposed to

3.1. Input-Related Factors

One characteristics of the input affecting input processing is **cue** — word order, vocabulary, morphology, and intonation

Learner's task is to discover the particular form-function mappings → the forms of natural languages are governed in the service of communicative functions → It is functionalist

Competition Model → the 'competition' that arises among the different cues that signal functions like 'agent' (there is language-specific strength assigned to cues)

e.g., *John kicks the ball* → cues are word order, knowledge of the lexical items, animacy criterion, subject–verb agreement
e.g., *That teacher we like a lot* → cues are word order, animacy criterion, case, subject–verb agreement

The usefulness of a cue:

Cue availability → how often the cue is present when a particular pattern is being interpreted
Cue reliability → how often a cue points to a particular interpretation
Conflict validity → whether a cue wins or loses when it appears in competitive environments
In English, word order is a highly available & highly reliable cue for identifying subject
In English, agreement is highly reliable & often unavailable for identifying subject
➔ there is language-specific strength assigned to cues

Interference?

The resolution of these conflicts is such that learners

a) resort to their L1 interpretation strategies
b) resort to a universal selection of meaning-based cues
c) gradually adopt the appropriate L2 biases as their L2 proficiency increases

Learning → readjustment of which cues are relevant to interpretation and a determination of the relative strengths of those cues

3.2. Learner-Related Factors: Associative Learning CREED

Construction-based → learning involves learning and recycling 'constructions':
concrete lexical items such as *book*;
formulae such as *once upon a time*;
slot-and-frame constructions such as *give* [*someone*] [*something*];
more open abstract schemata such as [*subject V Obj1 Obj2*].

Rational → language representations in the mind are tuned to predict the linguistic constructions that are most likely to be relevant in the ongoing discourse context

Exemplar-driven → language learning is formulaic (*Good+daytime* or *NounStem-PL*) ; acquisition of these productive patterns is based on exemplars → every time the language learner encounters an exemplar of a construction, the language system compares this exemplar with memories of previous encounters

Emergent → regularities emerge as learners determine structure from language usage ; learning responds to and emerges out of the learner's experiences of the language rather than being the result of innately constrained rules

Dialectic → interaction with others or instructional events such as conscious learning help to mitigate the aspects of associative learning that cause L2 learning problems

3.2.1. Learners' use of frequency in the input

Role of frequency:

the more times a stimulus is encountered, the faster and more accurately it is processed
a critical component used in the calculation of probabilities

Humans have innate abilities to pick out cues and to calculate statistical probabilities implicitly without awareness → learning language is a statistical process: it requires the learner to acquire a set of associations between constructions and their function/semantic interpretations

3.2.2. Overshadowing and attention blocking

Overshadowing

Where a feature in the input is redundant → the feature may not be processed in a way that is useful for learning, e.g., *Yesterday, I carried that box.*

This processing failure is thought to happen when:

- other features are more salient,
- world knowledge is likely to predict a communicatively satisfactory meaning
- other features have been activated many more times before, in the L1

Attention blocking

Overshadowing can over time lead to attention blocking:

- if *x* has always expressed a particular meaning/function, it is difficult to associate *x* with another meaning or function, e.g. *-ly*
- if a particular meaning has reliably been expressed using *x*, it is difficult to associate a different or an additional language feature with that same meaning, e.g. *indefiniteness*

3.2.3. Construction learning

Language learning = learning constructions or conventionalized form-meaning mappings

Are learners assisted by factors such as frequency or **prototypicality**?

- *verb locatives → go somewhere*
- *verb object locatives → put something somewhere*
- *ditransitives → give somebody something*

3.2.4. Statistical learning and connectionist accounts

Statistical learning / associative learning → acquisition = statistical regularities are absorbed via implicit learning → **complex adaptive system:** knowledge emerges gradually, driven by the exemplars learners are exposed to in social interaction

Connectionist model = **parallel distributed processing (PDP)** (Rumelhart & McClelland, 1986) → neurally inspired models of human information processing: nodes connected by pathways

Learning → as the network/learner is able to make associations between units:

- Human mind is predisposed to look for associations between elements, (i.e., find regularities in the input = extract probabilistic patterns) and create links between them
- Associations become weakened through non-activation ≠ become strengthened through exposure to repeated patterns

Learners learn languages both through repeated exposure to linguistic patterns in the input and through opportunities to repeat sequences in production:

- Linguistic knowledge takes the form not of rules or items but of an elaborate system of weighted connections → rule-like behavior ≠ rule-governed behavior

Symbolist vs. connectionist theories

- Symbolists
  - adopt an abstract view of linguistic representation: linguistic knowledge consists of a universal set of symbols and rules for combining them
  - make a distinction between property theory and transition theory
- Connectionists
  - view linguistic knowledge as a complex network of associations
  - no distinction is make between representation (product) and learning (process) mechanisms = no belief in the distinction between competence and performance