Age & Language Acquisition
• **Age & L2 morphosyntax: Questions of Ultimate Attainment**

• **How: Correlational studies + Work on age differences and L2 morphosyntactic**

• **Ultimate attainment is usually measured by comparing L2 learners’ responses on grammaticality judgement tasks**

• “Learners who began acquiring the L2 before a certain age, which these studies locate to be around puberty, will tend to exhibit intuitions that are very close to those of native speakers of that language. The late learners’ intuitions, by way of contrast, are not likely to be in the native speaker range, and this holds true regardless of the number of years, since they arrived in the L2 environment past puberty.”
L2 morphosyntactic knowledge along the age of onset continuum

Studies comparing knowledge of morphosyntax associated with varying ages of onset for L2 acquisition.

- Main research question: Are age and morphosyntactic attainment systematically related?
- Researchers’ interpretation of results: Johnson and Newport interpreted their data in support of a critical/sensitive period. Birdsong and Molis’s replication did not support all of the original findings.

| Johnson and Newport (1989) | 46 L1 Chinese and Korean adult speakers of L2 English who were college educated and had been living in the US for at least five years took a 276-item grammaticality judgement task. There was a statistically significant negative correlation of $r = -0.77$ between age of arrival (which ranged from 3 to 39) and grammaticality judgement score. The correlation was larger when only the 3 to 15 age group was examined ($r = -0.87$) and disappeared for the 17 to 39 group. The youngest group (3 to 7 years old when they arrived in the US) scored within the range of the NS control group, the adolescent group (who had arrived between 8 and 16 years of age) showed scores linearly declining with age and the group of adults (who had arrived at between 17 and 39 years of age) scored variably, without age holding any systematic relationship with their grammaticality intuitions. One late arrival (at age 23 upon arrival) scored 92% accurate, as high as native speakers |
• In Johnson and Newport (1989) the relationship between age and grammatical intuitions abruptly disappears after around puberty.
Birdsong and Molis (2001) Exact replication of Johnson and Newport: 61 college-educated L1 Spanish speakers of L2 English took Newport and Johnson’s grammaticality judgement task. The 29 early acquirers had arrived in the US between age 3 and 16 and had a mean length of residence of 12.2 years in the L2 environment. The 32 late acquirers had arrived at age 17 or older and had a mean length of residence of 10.5 years. There was a statistically significant negative correlation of $r = -0.77$ for the full sample between age of arrival and grammaticality judgement score. The early arrivals exhibited no variation, as they obtained near-perfect scores. There was a statistically significant negative correlation of $r = -0.69$ between age of arrival for the late arrival speakers and grammaticality judgement score. 13 late arrivals obtained 92% or higher scores, 3 above 95%. (The study also examined reported amount of L2 use. Amount of current L2 use was strongly related to judgements)
• In Birdsong and Molis (2001) grammaticality scores keep gradually declining across all ages beyond puberty.
• **Supporting studies** for both findings:

• DeKeyser (2000) produced findings that resonate with those of Johnson and Newport (1989) with 57 Hungarian US immigrants.

• Flege et al. (1999) obtained a pattern of results similar to that of Birdsong and Molis (2001) with 240 Korean permanent residents in the United States.
• Another way to view morphosyntax and age: Studies that are specifically designed to investigate the upper limits of successful late L2 learning.

• The focus is, like in Ioup et al. (1994), on exceptional learners who seem to have reached native-like ultimate attainment, often with L2s other than English. Even though these cases have traditionally been considered purely exceptional, Birdsong speculates that they may actually account for as much as 5 per cent to 25 per cent of cases of learners who are given ‘a fair chance of success’ (Birdsong, 1999b, pp. 14–15). The method of choice has been to compare their performance on grammaticality judgement tasks to that of native-speaking controls, sometimes using retrospective interviews to probe learners’ explanations of their choices. The question at stake in this kind of study is: How native is really ‘near-native’?
Differences between near-native and native morphosyntactic knowledge

Studies scrutinizing knowledge of morphosyntax in L2 speakers who achieve exceptionally high levels of ultimate attainment and are identified as near-native outside the laboratory (typically, cases of L2 acquisition after puberty)

- Main research question: Can some exceptionally successful L2 acquirers be indistinguishable from native speakers in their morphosyntactic knowledge?
- Researchers’ interpretation of results: Coppieters: No, near-native speakers’ L2 knowledge is different from true monolingual native speakers. Birdsong: Yes, some rare, exceptional near-native speakers cannot be distinguished from native speakers even under tight laboratory scrutiny

**Coppieters (1987)**

21 L2 French speakers, all of whom were highly successful and educated French users who had begun learning the L2 after puberty. They did a grammaticality judgement task and were interviewed. Their average grammatical intuitions on the task were three standard deviations away from the average of native speaker controls. Their rationalizations for their judgements during the interview were different from those of native speakers. Subtle syntactic-semantic and morphosemantic differences of knowledge distinguished nativeness from near-nativeness
Birdsong (1992)

Partial replication of Coppieters: 20 L2 French speakers all of whom were highly successful and educated French users who had begun learning the L2 after puberty. Their age of arrival in France was 19 to 48 and their L1 was English. Of these, 15 participants performed on a grammaticality judgement task within the native speaker range. There was a negative correlation of $r = -0.51$ between age of arrival and scores in this late-starter-only sample.
Evidence on L2 morphosyntax from cognitive neuroscience:

- FMRI, ERP

- Some researchers have shown that localization of language functions in the brain is less lateralized in late bilinguals (more right hemisphere activation is observed) than in early bilinguals and monolinguals.

- Helen Neville and her lab in the United States have produced evidence that, when engaged in certain kinds of L2 syntactic processing, the bilingual brains of people who began learning their L2 later in life (eight years or older in most of these studies) show clear different activation patterns from those of monolingual and early bilingual brains.
• Such age-related differences disappear when brain activation is inspected during the processing of L2 semantic stimuli.

• For example, Weber-Fox and Neville (2001) investigated Chinese–English bilinguals who were first immersed in the L2 environment anywhere between age one and past age 15. Those bilinguals who were exposed first to the L2 after year seven processed closed-class words (i.e. function words like with, the, some) differently from the early bilinguals and the monolingual controls, whereas open-class words (i.e. content words like nose, stored, glad) yielded no major differences in brain activation patterns across groups.
• Germany-based researcher Anja Hahne (2001) also found that her Russian–German bilingual participants, all of whom had learned German as their L2 after the age of 10, processed syntactically anomalous sentences of the kind Das Geschäft wurde am geschlossen (‘The shop was being on closed’) statistically significantly different from monolinguals, whereas no differences were found between the two groups with semantically anomalous sentences.
• Ullman (2001) have suggested that the learning of syntactic functions (in the L1 or the L2) is fundamentally different from the learning of semantic features. Specifically, they propose that syntax involves computational learning mechanisms and is constrained by a biological schedule, and that semantics draws on associative learning mechanisms and is free from critical period constraints.
• **Against:** Lee Osterhout and colleagues (Osterhout et al., 2002) demonstrated that the different neural activation patterns uncovered for function versus content words could be also explained by the differential word length in both kinds of stimuli (content words are typically longer, and this naturally *can affect processing*). Italian researchers Daniela Perani and Jubin Abutalebi (2005) suggest that *it is not the age of onset but the degree of active use of the L2 that matters when explaining degrees of brain activation*. They argue that the neural systems serving L2 and L1 grammatical processing are the same, and that higher attained proficiency and higher daily exposure to the L2 are independently correlated to lower activation patterns. Furthermore, they report on studies that show that, **even when attained proficiency is kept constant, the brains of L2 speakers who have less daily exposure to the L2 exhibit higher degrees of activation in the left prefrontal cortex.** They claim this parallels the general neurocognitive finding that increased practice leads to lower levels of neural activation, because with more practice the same processing task will consume less resources.
• (Osterhout et al., 2008) have initiated a research program that involves measuring the brain activity of zero-level beginning learners while they process L2 stimuli, longitudinally as they progress through their regular college-level foreign language courses. They have found that brain activation patterns can change in degree and location just after experiencing about four months or 80 hours of college instruction. At least for certain L2 forms, the brain’s activation patterns become similar to the patterns observed in fluent L1 users.
• Marinova-Todd et al. (2000) pointed out, given what we know about the *plasticity* of the brain, any age-related differences in brain location and neural activity patterns may be as much a result of the brain’s architecture shaping how subsequent linguistic experience is processed and used for L2 learning, as it could be the result of the brain having been shaped by previous experience.
• L2 Phonology and Age:

• Scovel (1988, 2000): *speech has a special status when it comes to critical periods because ‘pronunciation is the only part of language which is directly “physical” and which demands neuromuscular programming’* (1988, p. 62).

• He concluded that, in study after study, *non-native speaking samples were consistently and accurately detected by native-speaker judges*.

• The evidence accumulated since 1988 overwhelmingly shows that *foreign-sounding accents are likely to develop when the L2 is first learned later in life*. 
• Acquisition of phonology may be more impervious to non-biological influences such as L2 use and education, and therefore more strictly tied to biological schedules, than other areas of the L2.
• Opposite position to Scovel is:

• Flege found that phonetic categories or mental representations of speech sounds in the L1 are stabilized by age five to seven. After that point, new phonetic contrasts will be processed through such an L1 filter, and hence it is more difficult, although not biologically impossible, to detect and produce L2 categories that are not salient. Ironically, then, foreign accents may arise ‘not because one has lost the ability to learn to pronounce, but because one has learned to pronounce the L1 so well’ (Flege, 1999, p. 125).

• In other words, instead of viewing neurophysiological maturational constraints as the main explanatory factor for the development of L2 phonology, as Scovel does, or as a result of neurofunctional reorganization during development, as cognitive neuroscientists do, Flege puts the explanatory emphasis on psychoperceptual and phonetic causes related to previous massive experience with the mother tongue.
• exceptional learners shared two features.

• They had all received considerable amounts of high-quality L2 instruction and

• they all self-reported high levels of motivation and concern to sound native-like.
WHAT CAUSES THE AGE EFFECTS? BIOLOGICAL AND OTHER EXPLANATIONS

‘age is an index of the state of development of the L1 system. The more fully developed the L1 system is when L2 learning commences, the more strongly the L1 will influence the L2’

Language attainment differences can be expected, but they are probably a consequence of experience and socialization, and not biological or insurmountable in nature.

Fundamental Difference Hypothesis, which posits that the acquisition process undergone by children and adults is fundamentally different because children possess the innate ability to intuit the L1 grammar, whereas adults have lost this ability and thus need to resort to problem solving and conscious attention to handle L2 learning.

Plasticity, myelination, pubertal increases in estrogen or testosterone

Late and adult L2 acquisition generally results in lower levels of ultimate attainment and more individual variability than is observed for L1 and very early L2 acquisition.
Bilingualism and Age:

The realization that age effects may be present in additional language acquisition much earlier than previously thought, perhaps by age four.

(Sebastián-Gallés et al., 2005) investigated lexical representations of early versus simultaneous bilinguals by asking them to tell apart Catalan L2 words and non-words on a lexical decision task. They found that participants who had started to be exposed to Catalan at age 4 or earlier, but not from birth, did less well on this lexical decision task than participants for whom both Catalan and Spanish were available from birth.

Moving the onset of age effects into the very first years of life also blurs the traditional distinction between L2 and bilingual learners.
• The actual relative amounts of L2 and L1 use at the time of study may be central to the task of gauging age effects. This is the so-called issue of language activation (also called language dominance in bilingual studies)
• Cook (2008) explains the dangers of comparing L2 users with monolinguals eloquently:

• “There is no reason why one thing cannot be compared to another; it may be useful to discover the similarities and differences between apples and pears. SLA research can use comparison with the native speaker as a tool, partly because so much is already known about monolingual speakers. The danger is regarding it as failure not to meet the standards of natives: apples do not make very good pears. Comparing L2 users with monolingual native speakers can yield a useful list of similarities and differences, but never establish the unique aspects of second language knowledge that are not present in the monolingual.”
What could, then, be taken as a fair point of comparison to gauge attainment in age studies? Citing the work on multicompetence and bilingualism by Cook (1991) and Grosjean (1989), Singleton (2003) suggests that:

The appropriate comparison in the investigation of age effects in L2 acquisition is not between post-pubertal L2 beginners and monoglot native speakers but between post-pubertal L2 beginners and those who begin to acquire an L2 in childhood.
Cognitive-Interactionist Perspective on L2 learning
• Cognitive-interactionism is associated with the work in developmental psychology by Jean Piaget (e.g. 1974) and refers to the position that multiple internal (cognitive) and external (environmental) factors reciprocally interact (hence the word ‘interactionist’) and together affect the observed processes and outcomes of a phenomenon.

• Internal cognition is assumed to be the locus of learning (hence the word ‘cognitive’ in the term) and that a clear separation between cognitive-internal and social-external worlds is presupposed, since how the two interact is the object of inquiry.
• The Case of **WES**

• He stands for someone who became particularly adept at ‘initiating, maintaining, and regulating relationships and carrying on the business of living’ in his additional language (Schmidt, 1983, p. 168) but remained unable to master the L2 grammar despite what seemed to be sufficient time and ideal environmental conditions.
• In his early thirties, Wes emigrated from Tokyo to Honolulu by choice, in a financially and socially comfortable position, in pursuit of expanded international recognition in his already well-established career. Perhaps two features can be singled out as most defining of Wes’s personality. One is his strong professional identity as an artist, captured in excerpt (1) from an oral letter recorded into the third year of the study (Schmidt, 1983, p. 158):

(1) you know I’m so lucky / because ah my business is painting / also my hobby is painting / ... this is my life / cannot stop and paint / you know nobody push / but myself I’m always push /

• The second defining feature is Wes’s predisposition towards communication. His was the kind of social personality that avidly seeks people and engages in skilfully designed reciprocal interaction. This is illustrated in excerpt (2), also recorded around the same time as the previous excerpt:

(2) well / I like talk to people you know / um / I’m always listen then start talk / then listen / always thinking my head / then talk / some people you know only just talk, talk, talk, talk /
• In Schmidt’s estimation, by the third year of study, he was using English in his daily interactions between 75 and 90 per cent of the time.
His linguistic ability: For all three years, Wes’s verbs were characterized by overuse of –ing attached consistently to certain verbs denoting activities (e.g. joking, planning, training, touching), the use of past in only high-frequency irregular forms that can be memorized as items (e.g. went, sent, told, saw, said, met, bought), a complete absence of –ed and an overwhelming preference to make interlocutors understand the intended tense and aspect of his messages via lexical means such as adverbs (e.g. all day, always, right now, yesterday, tomorrow).

In other words, over three years of rich exposure to and meaningful use of English, Wes’s temporal L2 system remained rudimentary, stuck in the transition between the lexical marking stage and the next stage of development, where tense and aspect morphology begins to deploy.

Likewise, his articles and plurals improved minimally, from practically no occurrence in the beginning of the study to accuracy in up to a meagre third of the relevant cases, but even then in great part because of repeated occurrence of these forms in chunked phrases like n years old or n years ago (for plural –s), and a little (bit) X (for use of the indefinite article a).
• Schmidt proposed that ‘sensitivity to form’ or the drive to pay attention to the language code (p. 172) seems to be the single ingredient missing in Wes’s efforts to learn the L2.

• Wes was driven as a learner by an overriding investment in ‘message content over message form’ (p. 169). As he himself puts it, ‘I know I’m speaking funny English / because I’m never learning / I’m only just listen / then talk’ (p. 168).

• Schmidt concluded that positive attitudes and an optimal environment will afford the linguistic data needed for learning, but that the learning will not happen unless the learner engages in active processing of those data. In other words, grammar acquisition cannot be successful without applying ‘interest’, ‘attention’ and ‘hard work’ (p. 173) to the task of cracking the language code.

• Noticing Hypothesis
• **ACCULTURATION AS A PREDICTIVE EXPLANATION FOR L2 LEARNING SUCCESS?**

• The L2 environment engenders in learners certain attitudes that have affective and social–psychological bases and that must be considered if we want to understand L2 learning.

• **Pidginization Hypothesis, also known as the Acculturation Model:** It is predicted that great social distance between the L1 and L2 groups (as is the case of circumstantial immigrants, who speak a subordinate minority language and are surrounded by a powerful language of the majority), and an individual’s affective negative predispositions towards the target language and its members (e.g. culture shock, low motivation) may conspire to create what he characterized as a bad learning situation that causes learners to stagnate into a pidgin-like state in their grammar, without inflections or mature syntax. Conversely, he predicted that the more acculturated a learner can become (that is, the closer to the target society and its members, socially and psychologically), the more successful his or her eventual learning outcomes will be.
• INPUT FOR COMPREHENSION AND FOR LEARNING:

• The environment affords learners input, or linguistic data produced by other competent users of the L2.

• Also in the late 1970s, Stephen Krashen at the University of Southern California formally proposed a central role in L2 learning for input in his Comprehensible Input Hypothesis.
• According to Krashen, the single most important source of L2 learning is comprehensible input, or **language which learners process for meaning and which contains something to be learned, that is, linguistic data slightly above their current level. This is what Krashen termed i+1.**

• Learners obtain comprehensible input mostly through listening to oral messages that interlocutors direct to them and via reading written texts that surround them, such as street signs, personal letters, books and so on. When L2 learners process these messages for meaning (which they will most likely do if the content is personally relevant, and provided they can reasonably understand them), grammar learning will naturally occur.

• Krashen proposed this role for input on the assumption that **the mechanisms of L2 learning are essentially similar to the mechanisms of L1 learning:** in order to build an L1 grammar, children only need to be exposed to the language that parents or caretakers direct to them for the purpose of meaning making.
• **Against:** The strong claim that comprehensible input is both necessary and sufficient for L2 learning proved to be untenable in light of findings gleaned by Schmidt (1983) and by many others, who documented minimal grammatical development despite ample meaningful opportunities to use the language, even with young L2 learners – for example, children attending French immersion (Swain, 1985) and regular English-speaking schools (Sato, 1990). Input is undoubtedly necessary, but it cannot be sufficient.

• **Comprehension vs. Acquisition**


• **Interaction and Negotiation for Meaning:**

• Michael Long proposed the **Interaction Hypothesis**. The hypothesis grew out of work conducted for his dissertation at the University of California Los Angeles, in which college-level ESL learners were paired to interact with English native-speaking pre-service and in-service teachers of ESL.

• It extended Krashen’s proposal by connecting it in novel ways with studies in discourse analysis that had entered the field via the work of SLA founder Evelyn Hatch (1978) and work done on caretaker speech and foreigner talk in neighbouring disciplines. At the time, **Long agreed with Krashen that learning happens through comprehension, and that the more one comprehends, the more one learns.**

• However, he departed from the strong input orientation of the times by focusing on interaction and proposing that **the best kind of comprehensible input learners can hope to obtain is input that has been interactionally modified**, in other words, adjusted after receiving some signal that the interlocutor needs some help in order to fully understand the message.

• **Interactional modifications are initiated by moves undertaken by either interlocutor in reaction to (real or perceived) comprehension problems, as they strive to make meaning more comprehensible for each other, that is, to negotiate for meaning.** Typically, negotiation episodes begin with clarification requests if non-understanding is serious (e.g. whaddya mean? uh? pardon me?), confirmation checks when the interlocutor is somewhat unsure she has understood the message correctly (e.g. you mean X? X and Y, right?) and comprehension checks if one interlocutor suspects the other speaker may not have understood what she said (e.g. you know what I mean? do you want me to repeat?).
• **Output and Syntactic Processing During Production:**

• Where there is interaction, learners engage by necessity not only in comprehending and negotiating messages but also in making meaning and producing messages, that is, in **output**.

• Merrill Swain (1985) at the University of Toronto formulated her **Pushed Output Hypothesis (you will also see the terms Comprehensible Output Hypothesis and Output Hypothesis used interchangeably)**. She did so drawing on results of large-scale assessment of the linguistic outcomes of French immersion schools in Ontario, an English-speaking province of Canada. Specifically, she compared the oral and written performances of children who had studied in immersion schools against the performances on the same tasks by same-age L1 French peers. She found patterns that remarkably resonate with Schmidt’s (1983) findings for Wes.
• Comprehension does not usually demand the full processing of forms. During comprehension (e.g. when children read textbooks and listen to teacher explanations in school), it is possible to get the gist of messages by relying on key content words aided by knowledge of the world, contextual clues, and guessing.

• For example, in yesterday I walked three miles, we may hear ‘yesterday’ and not even need to hear the morpheme –ed in order to know our interlocutor is telling us about something that happened in the past.

• By the same token, reliance on this kind of lexical processing is less possible during production, because the psycholinguistic demands of composing messages force speakers to use syntactic processing to a much greater extent. Thus, Swain proposed that ‘producing the target language may be the trigger that forces the learner to pay attention to the means of expression needed in order to successfully convey his or her own intended meaning’ (p. 249).
NOTICING AND ATTENTION AS MODERATORS OF AFFORDANCES IN THE ENVIRONMENT:

Can we then conclude that acculturated attitudes, comprehensible input, negotiated interaction and pushed output are the four ingredients we need to explain optimal L2 learning?

Attention to formal detail in the input seemed to be missing and perhaps needed. The insights Schmidt gained from studying Wes and from a later case study of himself learning Portuguese during a five-month stay in Rio de Janeiro (Schmidt and Frota, 1986) led him to formally propose the Noticing Hypothesis in the early 1990s (best explained in Schmidt, 1995). He claimed that, in order to learn any aspect of the L2 (from sounds, to words, to grammar, to pragmatics), learners need to notice the relevant material in the linguistic data afforded by the environment. Noticing refers to the brain registering the new material, with some fleeting awareness at the point of encounter that there is something new, even if there is no understanding of how the new element works, and possibly even if there is no reportable memory of the encounter at a later time.

He concluded: The more L2 learners notice, the more they learn, and that learning without noticing (that is, subliminal learning), even if it exists in other domains of human learning, plays a minimal role in the challenging business of learning a new language.
• Thus, attention and noticing act as filters that moderate the contributions of the environment.
Negative Feedback During Meaning and Form Negotiation:

A final benefit of the environment, not discussed so far, is that it may provide learners with information about the ungrammaticality of their utterances. When the interlocutor has the actual intention to provide such negative information, then we may want to speak of error correction. However, more often than not, it is impossible (for the researcher as much as for the parties involved in the interaction!) to decide whether the intention to correct was at work. Therefore, we will prefer the term negative feedback over error correction or the near-synonymous corrective feedback (both of which imply a clear pedagogical intention to correct) and also over negative evidence (which is used in formal linguistic discussions about what linguistic abstract information would be needed to reset certain values within the limits available in Universal Grammar; Beck et al., 1995).

From the perspective of cognitive-interactionist researchers, negative feedback may come about as part of negotiating meaning or form. For example, a clarification request (e.g. sorry?) is offered when intelligibility is low and meaning itself needs to be negotiated. Nevertheless, it may convey to the learner an indication, albeit a most implicit and indirect one, that some ungrammaticality is present:

(Learner): what happen for the boat?
(Interlocutor): what?
(Learner): what’s wrong with the boat?

(McDonough, 2005, p. 86)
• At the other extreme, **explicit corrections overtly focus on the form at fault** and occur when a teacher clearly indicates to a student that some choice is non-target.

• Somewhere in the middle are **recasts and elicitations**. **Recasts occur when an interlocutor repeats the learner utterance, maintaining its meaning but offering a more conventional or mature rendition of the form.**

• **Elicitations include moves such as asking how do we say X? or directly asking the interlocutor to try again.** When they occur in classrooms, the teacher may initiate an other-repetition and pause in the middle of the utterance at fault to let the student complete it correctly.
• **Summary:**

- The five environmental ingredients that together contribute to (but do not guarantee) optimal L2 learning are: acculturated attitudes, comprehensible input, negotiated interaction, pushed output, and a capacity, natural or cultivated, to attend to the language code, not just the message. These five ingredients were likely present in a case like Julie (see Chapter 2, section 2.2), the first of several exceptionally successful learners discovered since the mid-1990s. The last ingredient, attention to the language code, was fundamentally missing from a case like Wes (see Chapter 4, section 4.1), who epitomizes the frequently attested phenomenon of mixed learning success.

- Neither positive attitudes towards the target language and its speakers nor abundant and meaningful comprehension of L2 messages are in and of themselves sufficient for second language learning to be successful, although both are certainly important ingredients in a highly complex environmental equation.

- For successful grammar acquisition, attention to form is probably necessary. This attentional focus on form can be externally achieved by instruction or internally sought by self-study and self-directed analysis of the linguistic material available in the environment.
• Negotiation for meaning, other- and self-initiated output modification, negotiation of form during collaboration, and negative feedback of varying degrees of explicitness all carry potential for learning, provided they occur under optimal conditions that recruit attention to the language code. They facilitate psycholinguistic and metalinguistic processes of segmenting the input, noticing gaps and holes, parsing messages syntactically, monitoring and hypothesis testing; these are in turn processes that help L2 learners crack the language code.

• Cognitive-interactionist researchers agree that negative feedback (or the implicit or explicit indication that some part of an utterance is ungrammatical) is better overall than entirely ignoring errors. Much less agreement has been reached as to when, how and why negative feedback works, when it does.

• The value of comprehension versus production for acquisition is an ill-understood conundrum that causes disagreement among SLA researchers. Some view learning as driven by comprehension exclusively and assign production a role for fluency-building. Others claim that productive, meaningful language use is in itself a catalyst for learning.

• Grammatical competence appears to evolve in ways that are less amenable to incidental benefits from the environment than other aspects of the language to be learned, such as vocabulary, discourse competence, and so on. It also seems to hold a special status in language acquisition. Specifically, grammar (a) requires more interest, attention and hard work than other aspects of the language to be learned; (b) may even require more time to simmer and deploy than the learning of other aspects of an L2; and (c) can act as a gatekeeper to development in other areas of the L2 beyond formulaic repertoires, particularly sociolinguistic competence.

• What matters in the linguistic environment is not simply ‘what’s out there’ physically or even socially surrounding learners, but rather what learners make of it, how they process (or not) the linguistic data and how they live and experience that environment.
Cognition and L2 Learning
• **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’)

• Information Processing Theories

• Emergentism

• Memory & Attention
• Information Processing (IP)

• 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, 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 Assumptions:

• The human cognitive architecture is made of **representation** and **access**.

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

• Cognitive resources such as **attention and memory are limited**.
Information processing theories distinguish between representation (or knowledge) and access (or processing).

Linguistic representation is comprised of three kinds of knowledge: 1) grammatical, 2) lexical and 3) 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 entails the activation or use of relevant knowledge via two different mechanisms known as automatic and controlled processing.

Human cognition is supported by both automatic and controlled processing. Information processing psychologists believe that all human perception and action, as well as all thoughts and feelings, result from the interaction of these two kinds of processing.
• **Automatic processes** require **small effort** and **take up few cognitive resources**.

• **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).

• By contrast, **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** 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**.

• If several demands are competing for controlled processing, they will be prioritized and certain processes will wait in line, so to speak, while only one is being executed. This is what we call **a limited capacity model of information processing**. The model predicts that performance that draws on controlled processing is more variable and more vulnerable to stressors than performance that draws on automatic processing. Therefore, a widely employed method in the study of automaticity is the dual-task condition.
• **The power of practice:** *proceduralization, automaticity*

• **Skills Acquisition Theory:** Skill acquisition theory *defines learning as the gradual transformation of performance from controlled to automatic.*

• This transformation *happens through relevant practice* over many trials, which enables controlled processes gradually to be withdrawn during performance and automatic processes to take over the same performance. The process has been called *proceduralization or automatization* and *entails the conversion of declarative or explicit knowledge (or ‘knowledge that’) into procedural or implicit knowledge (or ‘knowledge how’).*

• It is important to realize that *the learning of skills is assumed to start with the explicit provision of relevant declarative knowledge.* Thus, L2 learners (particularly instructed learners) *begin with explanations explicitly presented* by their teachers or in textbooks and, through practice, this knowledge can hopefully convert into ability for use, or implicit-procedural knowledge made up of automatic routines.
• **How does practice work?**

• It helps *proceduralization* of new knowledge by allowing 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.

• However, the power of practice is not constant over time. There is a well-known *power law of learning*, by which practice will at some point yield no large returns in terms of improvement, because optimal performance has been reached (Ellis and Schmidt, 1998).

• In addition, *proceduralization is skill-specific*. Therefore, practice that focuses on L2 production should help automatize production and practice that focuses on L2 comprehension should help automatize comprehension (DeKeyser, 1997).

• 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).
• Misconceptions:

• Two misinterpretations of skill acquisition tenets are common:

  A. that automaticity is simply accelerated or speedy behaviour; and

  B. that L2 learners simply accumulate rules that they practise until they can use them automatically.

• Read the study by DeKyser (PP. 85-87)
• Long Term Memory (LTM):

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. Much of the knowledge encoded in long-term memory is explicit-declarative, that is, verbalizable and consciously recalled. Explicit-declarative memory supports recollection of facts or events, and it is served by the hippocampus in the human brain. As much knowledge, or probably more, is encoded in implicit-procedural memory. These are 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.

• Tooling (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’. Episodic memory corresponds to a more recent type of memory in evolution, believed to have evolved from semantic memory. It ‘allows people to consciously re-experience past experiences’ and also to think of their future (Tulving, 2002, p. 6).
• **LTM and Vocabulary Knowledge:**

• **A word** is established in long-term memory when the **link between a form and its meaning is made.** However, knowing a word means a lot more: it includes the **strength, size and depth of the knowledge represented in memory.**

• Vocabulary knowledge strength concerns the **relative ability to use a given known word productively or to recognize it passively.** Thus, **strength is a matter of degree of proceduralization in implicit memory.**

• It is typically found that **learners know more words receptively than productively,** particularly if they are infrequent or difficult words, and that this gap becomes smaller as proficiency develops.

• The size of the **mental lexicon,** which **refers to 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.**

• **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. As Meara (2007) has argued for many years now, the notion of depth of vocabulary knowledge assumes the existence in implicit long-term memory of networks of meaning-based and form-based associations across the entire mental lexicon.
• **Working Memory:**

• By contrast to long-term memory, which is about representation and is unlimited, *working memory is about access and is limited.*

• Nick Ellis (2005): ‘If I ask you what $397 \times 27$ 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). In other words, we need 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 help define working memory.

• First, unlike long-term memory, working memory is of limited capacity.

• A second characteristic is temporary activation. Activation is so central to working memory.
Short Term Memory

Table 5.1
Memory tasks and benchmarks in the study of storage memory capacity

**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, keponen, 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.
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.”
• **Attention and L2 Learning:**

  • One main characteristic of attention is that its capacity is limited.

  • Secondly, because attention is limited, it is also thought to be selective. Only one attention-demanding processing task can be handled at the same time.

  • A third definitional feature is that attention can be voluntary, in the sense that it can be subject to cognitive, top-down control that is driven by goals and intentions of the individual.

  • A fourth characteristic is that attention controls access to consciousness.
• The focus has been on processes and outcomes of learning under three attentional conditions, which can be summarized as:

• **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).

• In a nutshell, SLA researchers have asked themselves whether L2 learning is possible without intention, without attention, without awareness and without rules.
Is it possible to learn about the L2 incidentally, as a consequence of doing something else in the L2, or does all L2 learning have to be intentional?

Incidental L2 learning is possible indeed. The learning of L2 vocabulary during pleasure reading is an incidental type of learning that has been found to be possible in the L2 as well as in the L1.

BUT:

We cannot ignore the fact that intentions wax and wane and fluctuate during online processing and that, in the end, it is online attention that is at stake in a cognitive understanding of L2 learning.

While learning without intention is possible, people learn faster, more and better when they deliberately apply themselves to learning.
• Can new L2 material be learned without attention?

• Is detection only sufficient for L2 learning or is noticing necessary?

• **Detection** is defined as registration outside focal or selective attention (Tomlin and Villa, 1994), whereas **noticing** is defined as detection plus controlled activation into the focus of conscious attention (Schmidt, 1995).

• Nelson Cowan’s (1988, 2001, 2005) **unified model of memory and attention** offers a framework for envisioning this problem as one that depends on a continuum in the quality of attention (from *low-level, automatic attention* to high-level, controlled attention), rather than on an all-or-nothing dichotomy between unattended and attended processing.
Which of the two extreme qualities of attention (low-level automatic detection or high-level, controlled activation) leads to learning? Or can both result in learning?

Detection at the periphery of focal attention is all that is necessary for L2 learning, whereas detection plus controlled activation into focal attention is facilitative of learning, but not necessary. Gass (1997) also agrees that noticing facilitates L2 learning but cannot be considered necessary. By contrast, Schmidt (1994, 2001) has maintained that detection involving peripheral attention is not enough for L2 learning, on the grounds that novel material that is attended peripherally could never be encoded in long-term memory. Instead, detection plus controlled activation into the focus of attention is needed for L2 learning: ‘what learners notice in input is what becomes intake for learning’ (1995, p. 20).
• Nick Ellis concedes that the Noticing Hypothesis may be right, but only if accompanied by an Implicit Tallying Hypothesis (2002a, p. 174), which imposes two provisos:

  • (a) noticing is necessary only for new elements with certain properties that make low-attentional learning unlikely, but not for all aspects of language to be learned, and

  • (b) noticing may be necessary only for the initial registration of such ‘difficult’ elements so as to make an initial representation in long-term memory possible, but not for subsequent encounters. This is because ‘once a stimulus representation is firmly in existence, that stimulus ... need never be noticed again; yet as long as it is attended to for use in the processing of future input for meaning, its strength will be incremented and its associations will be tallied and implicitly cataloged’ (Ellis, 2002a, p. 174).

• Schmidt did shy away from his initial claim that noticing is the ‘necessary and sufficient condition for converting input to intake’ (1990, p. 129), and since then his position has been that ‘more noticing leads to more learning’ (1994, p. 18). That is, noticing is facilitative of L2 learning
Learning without Rules:

- Can grammar generalizations result from experiencing L2 data without explicit knowledge being provided at the outset of the learning process? Or even without the learner actively and consciously searching to discover generalizations behind the language data she experiences?

- United States psychologist Arthur Reber was the first to expend sustained effort into the study of implicit learning, which he defined as learning without rules.

- Is learning without rules is about symbolic or associative learning?

- Those with Reber, believe they can learn without rules or awareness of rules, the proposal is that implicit (unconscious) processing leads to the abstraction of rules that are symbolically represented in the mind, only that they happen to be inaccessible to consciousness. That is, their theories of implicit learning are abstractionist and symbolic.

- Learning of underlying statistical structure, rather than learning of underlying rules (Shanks, 2005).
In the end, then, is L2 learning possible without rules?

Robinson concludes that, 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.
Emergentism:

Emergentism refers to a contemporary family of theories in cognitive science that have coalesced out of increasingly critical examinations of the tenets of information processing theories.

It is sometimes characterized as a post-cognitivist paradigm.

"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. (2006, p. 577)"

Three important tenets on which emergentist approaches build are 1) associative learning, 2) probabilistic learning and 3) rational contingency (Ellis, 2006a). From these three principles derive the ‘simple learning mechanisms’ to which Ellis and Larsen-Freeman (2006) refer in the quote above.
• **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. Ellis (2006a) explains that the human architecture of the brain is neurobiologically programmed to be sensitive to the statistical properties of the input and to learn from them.

• **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. However, this kind of probabilistic calculation is not a slave of whatever is experienced by the human brain as a contiguous temporal or spatial surface pattern. Instead, 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. Specifically, the processor makes best-evidence predictions about outcomes based upon (a) the overall statistics extracted by accumulated experience, (b) the most recent relevant evidence, (c) attention to cues detected to be present and (d) the clues provided by the context.
• **Usage-based learning**, or the position that language use and language knowledge are inseparable, because we come to know language from using it. Hence the specification in the earlier quote by Ellis and Larsen-Freeman (2006) that *learning from exposure comes about ‘as part of a communicatively-rich human social environment’ and is experienced ‘by an organism eager to exploit the functionality of language’.*

• Among others, US cognitive scientist [Michael Tomasello](https://www.mpig-berlin.mpg.de/en/staff/michael-tomasello), now at the Max Planck Institute in Germany, has been instrumental in advancing a view of language acquisition that is usage-based, where grammar concepts emerge out of communicative and social needs: ‘people construct relational and semantic categories in order to make sense of the world and in order to communicate with one another’ (Abbot-Smith and Tomasello, 2006, p. 282). Importantly, this commitment to usage-based learning means that two traditional distinctions in linguistics and information processing, respectively, are transcended: competence and performance, and representation and access. Furthermore, meaning (rather than rules) is held to be of primary importance in understanding the language faculty. For this reason, the linguistic schools that best suit the emergentist project are cognitive linguistics (Langacker, 2008) and corpus linguistics (Gries, 2008).
Another broad-scope tenet of emergentism is that **cognition is grounded, and therefore language is too.** By this, it is meant that our species’ experience in the world and the knowledge that we abstract from such experience is always structured by human bodies and neurological functions.

The final tenet that is worth highlighting in this synoptic examination of emergentism is that **language acquisition, like the acquisition of other forms of cognition, is a self-organizing dynamical system.** This entails viewing the phenomenon to be explained (e.g. language learning) as a system (or ecology) composed of many inter-connected parts that self-organize on the basis of multiple influences outside the system; these influences provide constraints that afford self-organization, but no single cause has priority over others.