

Speaking Acquisition Processes for Second Language Learning

LEUNG, Ricky Chi Yan

Introduction

The development and acquisition of second language (L2) speaking skills significantly enhances a learner's capacity to communicate with others in the target language. The ability to speak fluently in L2 is the goal of many language learners as being able to express one's thoughts, feelings, and ideas in a smooth, quick, and concise manner can result in various personal, educational, social, economic, and professional benefits. When individuals begin L2 learning, they quickly recognize the enormous gap between their L1 and L2 speaking abilities. In fact, Kormos (2006) stated that a L2 learners' knowledge of the target language, even for highly proficient L2 learners, is rarely complete in comparison to their L1. Regarding L2 speaking, learners lack competence in areas of linguistic knowledge such as lexis, syntax, and phonology. In addition to these L2 linguistic challenges, learners have cognitive limitations on neurological processing resources, time pressure during speaking, and decoding L2 input. The considerable effects of L1 speech on L2 speech, known as language transfer, also must be considered as L2 speech production research has shown that knowledge stores such as conceptual memory, lexis, and phonemes are shared in both L1 and L2 and they frequently compete for selection (Kormos, 2006). Language transfer also affects L2 phonological acquisition and encoding where learners might substitute and use L1 phonemes for similar but nonidentical L2 sounds. In their L2 phonemic output, some learners might incorrectly apply L1 phonological rules hardwired into their long-term memory and articulatory system when encoding new L2 lexis or while producing L2 output (Kormos, 2006). Fluency in L2 speech processing is another area that requires the learner's attention in lexical, grammatical, and phonological encoding phases, in contrast to the highly automatized nature of L1 speaking.

Several key processes work together in L2 speaking acquisition. Cognitive processes, linguistic acquisition involving lexical, syntactical, and phonological processing and encoding, and interaction are processes that contribute and explain L2 speech acquisition and processing.

Cognitive Models of Speech Acquisition and Production

The language faculty contains several functional modules to conceptualize, plan, and

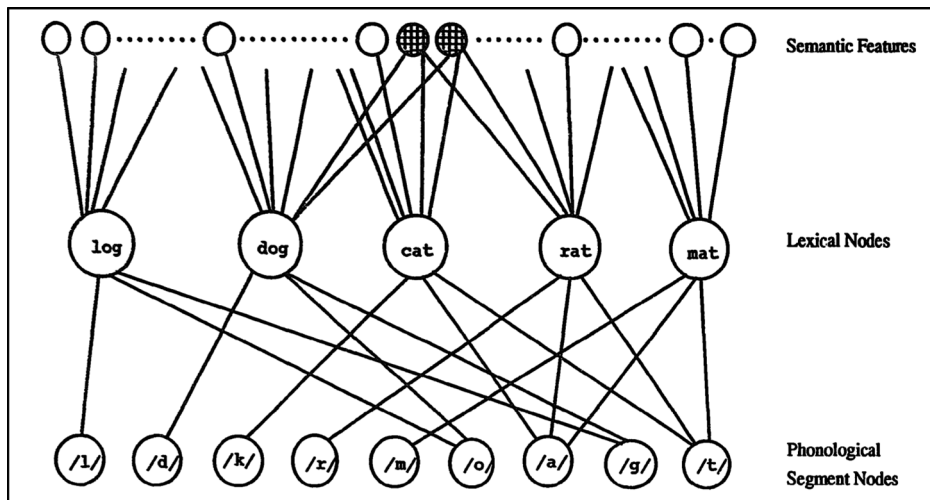
encode speech production. At this stage speakers use declarative memory to formulate the content and form of their message, according to the communicative situation and nature of discourse (Kormos, 2006). Two cognitive models that have attempted to explain the framework for speech processing are Dell's (1986) spread activation model and Levelt's (1989) model of speech production. These models were originally created for modelling L1 speech production processes and in the domain of second language acquisition (SLA), they have been adapted for L2 learning and research purposes.

Spread Activation Model

Dell's (1986) spread activation model proposed that language processing is accomplished through simultaneous and collaborative activation over several distinct processing levels. The spread activation model consists of three primary levels of processing: syntactic, morphological, and phonological. At the syntactic level, words are selected and arranged according to a learner's existing conception of grammatical rules. At the morphological level, words are specified in relation to their respective morphemes. At the phonological level, words are syllabized according to their sound. Semantic, lexical, and phonological nodes are linked by bidirectional inter-level connections (Dell & O'Seaghdha, 1992). An example of a lexical network structure in the spread activation model is shown in Figure 1:

Figure 1

Lexical Network Structure in Dell's Spread Activation Model



Note. Adapted from "Stage of lexical access in language production," by G. S. Dell and P. G. O'Seaghdha, 1992, *Cognition*, 42(1-3), p. 294.

Processing at any level provides activation or feedback to and from other levels, resulting in stable, self-reinforcing activation patterns across the neural network (Dell, 1986). Interaction between levels helps to maintain activation for the processing duration and through the formation of mutually supportive combinations of semantic, lexical, or phonological units, this helps to ensure accuracy (Dell, 1986). Each level consists of connected nodes that represent linguistic units and follow connectionist principles. This results in what Dell and O'Seaghdha (1992) describe as a globally modular and locally interactive model of speech production, in which two subsequent levels present interactions top-down and bottom-up (Dell & O'Seaghdha, 1992)

In the spread activation model, the node of the required category that has the highest level of activation is accessed by the learner (Kormos, 2006). Once this node is accessed, the processing or “spreading activation” proceeds further to lower-level nodes or across levels. At the lexical level, semantically and phonologically related items in the lexicon also receive activation. According to this process of spread activation, a Japanese learner of English, learning the word “ball” would first activate its translation equivalent, 球 (kyuu) due to their semantic relationship. The activation of “kyuu” would then spread to 野球 (yakyuu) or 地球 (chikyuu), due to their phonological relationship. 球 (kyuu) would then activate its translation equivalent, “ball”, due to their semantic relationship. Activation spreads between related representations might share notable similarities in semantics, morphology, phonology, or orthographies as in the case of 球 (kyuu) and 野球 (yakyuu). However, distant or seemingly unrelated links might also be activated, such as in the case of 球 (kyuu) and グローブ (globe), in which they are different concepts but share some thread of linguistic connection through spread activation. Dell (1986) also stated that when a node is activated, activation spreads to all the nodes connected to it. Items with the highest activation that possess combinational rules are selected and according to this model, speech errors are produced because an incorrect item might have a higher activation than a correct item (Kormos, 2006). The different levels might also activate with each other across multiple levels resulting in speech errors. A learner of Japanese might say the lexical form, 寒い水 (samuimizu), when 冷たい水 (tsumetaimizu) is the correct form when referring to cold water. In this case, one can recognize the semantic relation between the intended meaning and its correct form, as well as the phonological similarity, which leads to a possible speech error for the L2 learner. In an examination of oral data on Chinese learners' use of past tense accuracy, Yu and Lowie (2020) found that the learners' L1 might also influence their item selection preferences. As there is no past tense in Chinese, this makes it difficult for learners to learn past tense verbs in English and occasionally, the students confused past tense with the past participle.

Substitution errors involving simultaneous semantic and phonological similarity are an example of the spread activation process (Dell and O'Seaghdha, 1992).

Speech Production Model

Levelt's (1989) speech production model proposed that the processes of speech production are modular, incremental, and unidirectional in one's neurological faculties. Nonlanguage communicative intentions are first generated by the conceptualizer and determine the semantic content of the spoken utterance to be produced. The conceptualized message then goes through the two operations of macroplanning and microplanning. Macroplanning is the elaboration of a communicative goal into a series of subgoals and the retrieval of information to be expressed to realize these subgoals (Levelt, 1993). Microplanning is the process of giving a propositional shape to each conceptual element, and assigning the informational perspective (topic and focus), that will guide the interlocutor's attention (Levelt, 1993). As a learner's L2 proficiency increases, these conceptual elements or "chunks" will become larger, leading to the formation of formulaic sequences. This chunking operation provides proficient L2 speakers with a processing advantage that alleviates the cognitive pressure of speech production and facilitates a more fluent flow of conversation between interlocutors (Yan, 2019). The product of macroplanning and microplanning is a preverbal message which is then assigned to the formulator where the lemmas or lexical items that most accurately represent the semantic content of the preverbal message's elements are selected (Levelt, 1989). Grammatical encoding also takes place in the formulator which comprises of the information needed for formulating syntax and generating the surface structure of an utterance. Another task of the formulator is to select the phonological representations for the selected lemmas or lexical items (de Bot, 1996). The result of the combined processes of the conceptualizer and formulator is the articulatory plan, or one's "internal speech" when conceptualizing an utterance. The next key module in Levelt's speech production model is the articulator. It specializes in the physical execution of the articulatory plan through one's speech organs. Moreover, the articulator considers the possible differences between when the formulator delivers the phonetic plan and its physical execution by the articulator, creating an articulatory buffer which can temporarily store the articulatory and phonetic plans. The articulator then retrieves lexical chunks from the articulatory buffer and preps them for execution, with speech as the final product. Monitoring might also occur, allowing speakers to monitor their internal speech and detect problems before and during articulation by assessing the meaning and form of their output (Levelt, 1989). The autonomous functioning and incremental aspect of each module and consecutive progression is what makes parallel

processing possible within Levelt's model (Kormos, 2006). While one word is being spoken, the selection of lemma and lexeme for other words occur concurrently, while the speaker is also conceptualizing their next message during speech production.

Linguistic Processes of L2 Speech Acquisition and Production

Linguistic processes of L2 speech acquisition and production consists of lexical, syntactic, and phonological acquisition. Without high levels of competence in L2 lexical, syntactic, or phonological knowledge, an individual can produce L2 speech and might still be generally understood (depending on the interlocutor's listening competence), but without the use of appropriate lexis, syntax, or pronunciation, L2 speech and interactions cannot be accomplished above an elementary communicative proficiency. The influence of lexical, syntactical, and phonological L2 acquisition processes are described.

Lexical Acquisition for Production

It is widely accepted among language teachers and SLA researchers that vocabulary knowledge is essential for being able to communicate in L2. Tavakoli and Uchihara (2020) identified that more fluent L2 speakers produced a greater portion of n-grams (a sequence of discourse markers or language chunks) than lower proficiency learners, demonstrating that higher levels of L2 interaction demonstrate significant differences at the learner's lexical proficiency. In early stages and possibly later stages of L2 lexical learning, the primary task is to remember the L2 word's form and meaning. For L2 learners, words are primarily learned as formal items because the meaning is given with a L1 translation or definition (in L1 or L2) rather than being learned within the context itself (Jiang, 2000). At this initial phase, the use of L2 words involves activation of the links between L2 words and their L1 translations or forms in the conceptualizer (de Bot, 1996). The preverbal message first activates the L1 words that contain the intended semantic meanings to match the message chunks. As one's experience and proficiency in L2 increases, simultaneous activation of L2 word forms and lemma information including semantic and syntactic conceptions, might result in strengthened connections between L2 word forms and lemma of its L1 translation (Jiang, 2000). When learners reach higher levels of L2 proficiency, the semantic, syntactic, morphological, and formal conditions about L2 words are established within the learner's lexical store and are processed at much higher speeds than when first learned (Jiang, 2000).

Meara (1997) suggested that L2 lexical encoding consists of a series of events that build connections between newly encountered words and words that already existed in the learner's lexicon. Meara stated that either the L1 translation of a L2 word or a L2 word with associative

links such as synonyms or antonyms can be formulated. One way of enabling deeper processing at early stages of L2 lexical acquisition is to teach L2 learners how to engage in complex strategies of lexical processing such as rote memorization, keyword mapping, and semantic mapping (Sagarra and Alba, 2006).

Syntactical Acquisition for Production

Selinker's (1972) interlanguage hypothesis proposed that syntactical encoding in L2 learning begins with the application of L1 syntactical rules to construct L2 utterances. This process is incremental in which L2 grammatical forms are gradually incorporated and corrected for accuracy, leading to more proficient L2 output. Dulay and Burt (1974) challenged Selinker's hypothesis by suggesting that a learner's L1 grammar system is not always transferred to their L2 system. In Dulay and Burt's study of young children's (ages 5 – 8) SLA, they found that native language interference accounted for a small amount of variance (4.7%) as compared to developmental cognitive strategies (87.1%), implying that a creative construction process independent of both L1 and L2 occur in the L2 syntactical encoding of learners. Two models that describe some of the key elements of L2 syntactical acquisition are Bates and MacWhinney's (1982) competition model and Pienemann's (1998) processability theory.

In Bates and MacWhinney's (1982) competition model, language learning is influenced by the frequency and complexity of the relationships of grammatical forms and communicative functions. L2 learners engage in form-function mapping, a mechanism for organizing surface forms and communicative intentions based on the available input. Learners process and acquire "cues" and the strength or salience of cues is determined by the frequency and availability of the particular form-function mapping in the available input (Kormos, 2006). In both L1 and L2 acquisition, cues compete with and might override each other and the strength of each cue influences the order in which various cues are learned (MacWhinney, 1997). The competition model assumes that L2 learners begin by attempting to transfer the form-function mappings of their L1 to L2. If the L1 transfer does not produce the correct output in L2, the learners will focus on cues in the input and encode the new L2 grammatical structure in a systematic procedure for later recall and production (MacWhinney, 1997). If a syntactical structure has the same function in both L1 and L2, encoding of the structure will be facilitated easier than syntactic structures that are formally similar but have different communicative purposes or value; these structures are deemed to be more difficult to encode (Kormos, 2006).

Pienemann's (1998) processability theory seeks to explain learners' developmental

schedules and is concerned with the constraints of acquisition rather than the representation of grammatical knowledge. Processability theory ascribes that learners can produce only what they can process; the syntactic acquisition operations establish a hierarchy, in which each subordinate phase needs to be at least partially acquired before the next phase can properly function (Pienemann, 1998). L2 learners need to acquire lower order grammatical structures before they can encode following or higher order structures. L2 learners develop along this hierarchical order of processability in acquiring any L2 syntax. When this hierarchy of acquisition is applied to the syntactical structure of any target language, it results in a set of predictions for the order in which learners will develop syntactical and morphological structures in that specified L2. Pienemann (1998) theorized that syntactic encoding takes place in the following sequence: lemma access, category procedure, phrasal procedure, the S-procedure (establishes the positioning of the phrase within the sentence), and subordinate clause procedure. If an L2 learner had not acquired the necessary syntactic encoding procedures at a specific stage, the learner would have to map concepts to surface form from that level onwards (Pienemann, 1998).

Phonological Acquisition for Production

At the beginning stages of L2 phonological acquisition, learners tend to apply L1 rules of encoding to L2 phonology, and find it difficult to modify existing articulatory gestures automatized for L1 production (Kormos, 2006). To describe the L2 phonological acquisition process, Eckman's (1977) markedness differential hypothesis defined "markedness" as the frequency of a particular linguistic structure in all language. When a specific linguistic feature frequently occurs in languages, it is considered unmarked, and when a specific linguistic feature rarely occurs, it is considered marked (Eckman, 1977). If L2 phonemic structures are considered more marked than a learner's corresponding L1 phonemic structure, then the L2 structures are more difficult to acquire. Regarding L2 syntactical encoding, Flege's (1995) speech learning model proposed that when L2 beginners encounter an L2 sound that is not part of their L1 phonological store, they substitute the nearest L1 sound for the target language phoneme. The acquisition of L2 phonemes is constrained by phonetic differences between L2 sounds and corresponding most similar L1 sounds and features of the L1 phonemic inventory. L2 learners gradually establish a new phonological inventory that contains the mental representations for the L2 sounds. Flege (1995) also proposed that the age when learners started learning L2 and the frequency of use also affects the success of the L2 phonemic acquisition process.

Interaction

Interaction is key to the development and acquisition of L2 speaking proficiency because it allows learners to actively produce L2 through speech. Learners who want to be able to communicate orally in L2 need adequate opportunities for spoken output and practice in meaningful contexts. Long's (1996) interaction hypothesis stated that the development of L2 proficiency is enhanced by face-to-face interaction and communication. L2 speaking acquisition is facilitated by the negotiation of meaning with interlocutors, particularly in areas of linguistic or communication difficulty where learners will notice gaps in their L2 linguistic knowledge and output. Real-time interaction facilitates L2 acquisition because it connects input, internal cognitive processes, monitoring, meaning negotiation, and output in productive ways. Long (1996) also highlighted the effect of both positive and negative evidence on L2 production. Positive influence during interaction occurs when accurate L2 utterances are provided by an interlocutor and noticed by the L2 learner. Negative evidence provides valuable information regarding inaccurate or non-permissible L2 linguistic forms. When learners receive negative evidence, this promotes attention to form and allows for the opportunity to produce modified output that is closer to more target-like forms (Long, 1996). Swain's (1995) comprehensible output hypothesis adds onto the conceptualization of learners' "gaps" in their L2 linguistic knowledge, through the process of noticing. The process of noticing can trigger cognitive processes which might generate new linguistic knowledge or consolidate their existing knowledge when modify their output (Swain, 1995). The concept of hypothesis testing also allows learners to test hypotheses about the L2 to confirm what is or is not possible within L2 linguistic features. Swain (1995) noted that a L2 learner's metalinguistic function serves to control and analyze output, and internalizes relevant linguistic knowledge. This relates to Levelt's (1983) monitoring and self-repair model, which is a self-initiated, corrective process utilized to repair erroneous or inaccurate output. Levelt's monitoring and self-repair model consists of three mechanisms. The initial monitoring mechanism involves the speaker's realization that the preverbal plan needs to be changed and reformulated. To avoid errors, speakers might decide to encode new and different information from the original message or they might modify the content of their existing message. The second monitoring mechanism is the correction of linguistic errors that arise during error repairs. The preverbal plan is appropriate, but an erroneously activated word, syntactical structure, or phoneme is selected during the error repair. The third monitoring mechanism, called rephrasing repair involves the modification of the preverbal plan but leaves the content of the message unaltered. Rephrasing repairs are employed when L2 speakers are uncertain about the correctness of their utterance, which makes this type of repair similar to a

communication strategy (Kormos, 2006).

Conclusion

The multitude of cognitive processes, linguistic factors, or other moderating variables involved in the acquisition and production of L2 speaking have not been fully identified nor accounted for in existing SLA research. This research outcome is likely to remain unchanged in the long-term future; L2 speaking development is unstable as it is subject to individual differences among learners that cannot always be quantitatively or qualitatively measured in a precise or generalizable manner. Variations among languages, learning environments, and other sociolinguistic influences are just some of the other significant factors that add onto already existing variations in L2 speaking acquisition and learner proficiency outcomes. From what can be gathered from available research, key L2 acquisition features include cognitive models of speech acquisition and production, linguistic processes such as L2 lexical, syntactical, and phonological acquisition and encoding, and the role of interaction, output, and monitoring.

In contrast to other primary L2 skills such as reading, writing, or listening, speaking acquisition is unique because other interlocutors can play a significant role in aiding a learner's development and proficiency. In today's ever-expanding and interconnected global world, in social, educational, professional, political, and economic arenas, being able to speak and proficiently communicate in L2 has become more important than any other time in history. For many L2 learners, a substantial amount of practice in producing L2 output and interacting with others represents the modern approach to L2 speaking acquisition and development compared to traditionalist views which tend to primarily feature explicit learning. It is important to bridge what is known about cognitive processing and L2 linguistic acquisition with pedagogical methodology. To survive and communicate in the hyper-paced, technology-driven 21st century, L2 learners who desire to participate on a global stage need to be able to exchange their opinions, ideas, and information in L2. Thus, it is imperative to use existing and future research findings to practically scaffold and aid L2 learners' speaking acquisition in a pragmatic and interactionally communicative approach.

References

- Bates, E., & MacWhinney, B. (1982). *Functional approaches to grammar* (E. Wanner, L. R. Gleitman, Eds.). Cambridge University Press.
- de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46(3), 529-555. <https://doi.org/10.1111/j.1467-1770.1996.tb01246.x>
- Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321. <https://doi.org/10.1037/0033-295X.93.3.283>
- Dell, G.S., & O'Seaghdha, P.G. (1992). Stages of lexical access in language production. *Cognition*, 42(1-3), 287-314. [https://doi.org/10.1016/0010-0277\(92\)90046-K](https://doi.org/10.1016/0010-0277(92)90046-K)
- Dulay, H. C. & Burt, M. K. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly*, 8(2), 129-136. <https://doi.org/10.2307/3585536>
- Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27(2), 315-330. <http://dx.doi.org/10.1111/j.1467-1770.1977.tb00124.x>
- Flege, J. E. (1995). *Second language speech learning: Theory, findings and problems* (W. Strange, Ed.). York Press.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1) 47-77. <https://doi.org/10.1093/applin/21.1.47>
- Kormos, J. (2006). *Speech production and second language acquisition*. Routledge.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation, volume 1*. MIT Press.
- Levelt W. J. M. (1993). *Lexical access in speech production* (E. Reuland, W. Abraham, Eds.). Springer. https://doi.org/10.1007/978-94-011-1840-8_11
- Long, M. H. (1996). *The role of the linguistic environment in second language acquisition* (W. C. Ritchie, T. K. Bhatia, Eds.). Academic Press.
- MacWhinney, B. (1997). *Second language acquisition and the competition model* (A. M. B. de Groot & J. F. Kroll Eds.). Lawrence Erlbaum Associates Publishers.
- Meara, P. (1997). *Towards a new approach to modelling vocabulary acquisition* (N. Schmitt, M. McCarthy, Eds.). Cambridge University Press.
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. John Benjamins Publishing Company.
- Sagarra, N. & Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *Modern Language Journal*, 90(2), 228-243. [10.1111/j.1540-4781.2006.00394.x](https://doi.org/10.1111/j.1540-4781.2006.00394.x)

- Selinker, L. (1972) Interlanguage. *International Review of Applied Linguistics*, 10(3), 209-231. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Swain, M. (1995). *Three functions of output in second language learning* (G. Cook, B. Seidlhofer, Eds.). Oxford University Press.
- Tavakoli, P., & Uchiyara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*, 70(2), 506-547. <https://doi.org/10.1111/lang.12384>
- Yan, X. (2019). Unpacking the relationship between formulaic sequences and speech fluency on elicited imitation tasks: Proficiency level, sentence length, and fluency dimensions. *TESOL Quarterly*, 54(2), 460-487. <https://doi.org/10.1002/tesq.556>
- Yu, H., & Lowie, W. (2020). Dynamic paths of complexity and accuracy in second language speech: A longitudinal case study of Chinese learners. *Applied Linguistics*, 41(6), 855-877. <https://doi.org/10.1093/applin/amz040>