Part I: **Perspectives on usage in L2 learning and teaching**

Brian MacWhinney

Multidimensional SLA

Complex natural phenomena, such as human language, are shaped by processes operating on very different scales in both time and space (MacWhinney, in press). Consider the case of timescales in Geology. When geologists study rock outcrops they need to consider the results of general processes such as vulcanism, orogeny, glaciation, continental drift, erosion, sedimentation, and metamorphism. Within each of these larger processes, such as vulcanism, there are many microprocesses operating across smaller timescales. For example, once the pressure in the magma chamber reaches a certain level, there can be a slow outpouring of lava or sudden explosions. Pressure can be released through steam vents with geysers operating at regular intervals. The lava may enter lakes or oceans forming pillows or it may rest in underground chambers forming columnar basalt. The variations in these volcanic processes and their interactions with each other and plate tectonics are extensive. The same is true of human language. Within human populations, the ability to articulate and process sounds has emerged across millennia of ongoing changes in physiology and neurology. Within particular language communities, ongoing change is driven by language contact, dialect shift, and group formation. Within individuals, language learning involves a continual adaptation for both first and second languages. Within individual conversations, all of these forces come together, as people work out their mutual plans, goals, and disagreements, using language. Each of these space-time frames interacts with the others at the actual moment of language use.

To fully understand the process of second language acquisition (SLA), we must place it within this multidimensional context, both theoretically and practically. In this paper, I will attempt to show how looking at SLA through this lens forces a radical restructuring of our understandings about how languages are learned and how learning can be optimized. Most importantly, the view of language learning as emerging from a multidimensional competitive integration allows us to formulate new methods for second language pedagogy.

To explain these linkages between theory and practice, we should first consider what it means to view second language learning as a multidimensional emergent process. This emergentist approach rests on three basic principles often expressed in systems theory (Beckner et al., 2009; von Bertalanffy, 1968). These are the principles of competition, hierarchical structure, and timeframes. Each of these principles plays an important role in understanding second language learning as a multidimensional emergent process, and in helping us understand how we can optimize this process.

1 Competition

Competition is fundamental to biological processes. Darwin (1859) showed how the evolution of the species emerges from the competition between organisms for survival and reproduction. The three basic evolutionary processes Darwin identified were proliferation, competition, and selection. Proliferation generates variation through mutation and sexual recombination. Organisms with different compositions compete for resources or rewards such as food, shelter, and the opportunity to reproduce. The outcome of competition is selection through which more adaptive organisms survive and less adaptive ones disappear.

The combined operation of proliferation, competition, and selection is the major engine driving change in all biological and social systems. Emergentist approaches to language (MacWhinney, 1999) also view linguistic structures as arising from the processes of proliferation and competition. For the organism as a whole, the fundamental functional pressure is to reproduce. For language, the fundamental functional pressure is to communicate efficiently in ways that allow the listener to decipher messages efficiently and accurately. As MacWhinney, Bates, and Kliegl (1984) noted, "the forms of natural languages are created, governed, constrained, acquired and used in the service of communicative functions."

The handmaiden of competition is cooperation. As Bates and MacWhinney (1982) observed, humans have a great many ideas that they would love to express all at once, but language only allows us to say one thing at a time. One way in which language addresses this problem is by allowing motives to form coalitions. Bates and MacWhinney (1982) characterized the possible solutions to competition as: (1) peaceful coexistence, (2) divide-the-spoils, and (3) winner-take-all. Based on this analysis and the linkage of cue strength to cue validity (Brunswik, 1956), MacWhinney, Bates, and colleagues (MacWhinney & Bates, 1989) conducted a series of empirical investigations of cue processing in first and second language learning called the Competition Model, which has now resulted in over 100 published empirical studies (see http://psyling.psy.cmu.edu/papers for a bibliography). This research program has shown how second language learning emerges from competitions between first language cues and the reliability and costs of cues in the new language. Although Competition Model work has given us a good understanding of the core process of competition, it has not yet come to grips with the full multidimensional nature of second language learning.

2 Hierarchical Structure

Language is multidimensional in terms of both structure and process. The traditional levels of linguistic structure include phonology, morphology, lexicon, and syntax. Beyond these, there are language structures for conversational interaction, mental model construction, and sociolinguistic group formation. All of these structures emerge over time in response to external factors from human physiology and society. In neural terms, language use relies on virtually every region of the brain from the hippocampus, basal ganglia, and cerebellum to the anterior temporal and frontal cortex. In physical terms, much of the action is located in our vocal tract, face, tongue, and lungs, but we also use our hands, eyes, and posture to supplement the message with gesture and other signs. To decode these complex messages we rely on audition, vision, and the other senses. Social forces also shape language through processes such as migration, social class, and memesis.

Complexity arises from the hierarchical recombination of small parts into larger structures. For biological evolution, the smallest parts are the genes. For the brain, the smallest parts are the neuronal assemblies that generate competing ideas (D. T. Campbell, 1960; Edelman, 1987). In his seminal article entitled "The Architecture of Complexity", Simon (1962) analyzed higher-level cognitive processes as hierarchically structured combinations of elementary information processes. These elementary pieces are configured in modules whose structure is (only) partially decomposable.

2.1 An Example

These basic architectural principles can be illustrated by the four levels of structure that emerge during protein folding (N. A. Campbell, Reece, & Mitchell, 1999; MacWhinney, 2010). In this process, the primary structure of the protein is determined by the sequence of amino acids in the chain of RNA used by the ribosome as the template for protein synthesis. This sequence conveys a code shaped by evolution; but the physical shape of a specific protein is determined by processes operating after initial RNA transcription. The next structure to emerge is a secondary structure of coils and folds created by hydrogen bonding across the amino acid chain. These forces can only shape the geometry of the protein once the primary structure emerges from the ribosome and begins to contract. After these second structures have formed, a tertiary structure emerges from hydrophobic reactions and disulfide bridges across the folds and coils of the secondary structures. Finally, the quaternary structure derives from the aggregation of polypeptide subunits based on the ternary structures. This final structure allows each protein to serve its unique role, be it oxygen transport for hemoglobin or antigen detection for antibodies. In this partially decomposable emergent system, each level involves a configuration of components from lower levels, but the biochemical constraints operative on each level are unique to that level and only operate once that level has emerged during the process of folding. If a given protein operates successfully, it promotes the adaptation of the whole organism, eventually leading to evolutionary selection for the DNA sequence from which it derives. This can be viewed as a type of backwards or downwards causality between levels (Andersen, Emmeche, Finnemann, & Christiansen, 2000).

As we have seen, protein folding produces emergent structures that are then subject to the mechanisms operating at these new structural levels. The same is true for language. When articulatory gestures combine into words, they produce structures that link to meanings through emergent lexical patterns. Once words are available, they can be joined into combinations that are then subject to new mechanisms such as the tendency to place related items next to each other, as stated in Behaghel's Law (Behaghel, 1923). Different structural levels trigger the operation of different mechanisms of emergence, such as episodic encoding, generalization, topological organization, structure mapping, or common ground. For a discussion of mechanisms of emergence see MacWhinney (in press)

2.2 Interlocking Linguistic Hierarchies

The principles of elementary units, partial decomposability, level-specific constraints, and backwards causality also apply to the study of language learning, where the interactions between levels and timeframes are so intense. The linguistic systems of auditory phonology, articulatory phonology, lexicon, syntax, mental models, and communicative structure are represented in partially distinct neuronal areas (Hagoort, 2013) and each displays hierarchical composition between levels. For example, lexical items are composed of syllables that group into prosodic feet to produce morphemes. Morphemes combine to produce compounds, derivations, and longer formulaic strings (Sidtis, 2014). Articulatory form emerges from motor commands that group hierarchically into gestures that eventually produce syllabic structures. Syntactic patterns can be coded at the most elementary level in terms elementary level in terms of item-based patterns, which then group on the next level of abstraction into constructions, and eventually general syntactic patterns. Mental models are structured in terms of grammatical roles that link to embodied interpretations of events (MacWhinney, 2008c). At the most elementary level, communicative structures involve speech acts that can group into adjacency pairs from which emerge higher-level structures such as topic chains and narrative structures. Each of these hierarchies is tightly linked to others. For example, syntax and lexicon are tightly linked on the level of the item-based pattern and in terms of the local organization of parts of speech in the lexicon (Li, Zhao, & MacWhinney, 2007). Given the interactive nature of these interlocking hierarchies, full decomposition or reductionism is clearly impossible. Instead, the primary task of systems analysis is to study the ways in which the various levels and timeframes mesh. To express these interactions in the terms of the Competition Model, we need to measure the strength of competing forms or patterns and their interactions during both online and offline processing (Labov, 1972).

This view of language as an emergent hierarchy has important consequences for second language acquisition theory and practice. The basic hierarchy emerges during first language learning. However, for each level, the second language learning must acquire new structures and linkages. In part, these new structures can be acquired from the natural process of language interaction (Krashen, 1982). However, as discussed in detail in MacWhinney (2012), adults face a set of risk factors that reduce the effectiveness of mere exposure to L2 input. These are the factors of entrenchment, negative transfer, parasitism, misconnection, and isolation. Entrenchment arises from the fact that linguistic structures become locked into place in cortical maps that are then resistant to restructuring during second language learning. Negative transfer arises from the inappropriate use of a first language structure in the second language. Parasitism arises when learners access second language meanings through translation to their first language, rather than accessing meaning directly. Misconnection arises when the connections of white matter tracts to cortical areas lead to inefficient transfer of information. Social isolation is a general risk for older learners who do not integrate well in the second language community. To combat these risk factors, adults can rely on the protective processes of resonance, positive transfer, internalization, chunking, and participation. Resonance relies on methods such as recoding and interactive activation to encode new linguistic forms in multiple dimensions for more reliable retrieval. Positive transfer works to map semantic and pragmatic strutures that are similar between the first and second languages. Internalization involves learning to think in the second language. It can include exercises such as recital of songs or poems and echoing of newly learned phrases. Chunking involves the creation of larger units for more fluent production, and participation involves the formation of social bonds and support systems to maximize contact with speakers of the second language. To maximize the operation of these protective processes, learners can rely on particular methods of support from instructional sources and the community. Some of this support can be provided by language immersion (Clark, Wagner, Lindemalm, & Bendt, 2011). However, for certain linguistic skills, focused training will be even more effective (Presson, Davy, & MacWhinney, 2013). To understand how we can reach this balance, we need to further consider the impact of timeframes on second language learning.

3 Timeframes

To understand the mechanics of competition and hierarchical pattern combination, we must examine inputs from processes operating across contrasting timeframes (MacWhinney, 2005; MacWhinney, Malchukov, & Moravcsik, in press). Broadly speaking, we can distinguish four major timeframes:

1. **Processing.** Processing occurs during the moment of speaking. For the analysis of this moment, psycholinguists have focused on the neural basis for online processing of words and sentences during production and comprehension, whereas conversation analysts have focused on the interactional basis for the ways in which we take turns and share ideas. During speaking, all of these processes work closely together.
2. **Consolidation.** Online processing leads to the storage of experiential traces in memory. Some traces last for only seconds, others persist across decades. Memory processes can also support the emergence of higher levels of structure through generalization that vary through the course of a human lifespan.
3. **Social Diffusion.** Linguistic forms diffuse through processes of social memesis (Mesoudi, Whiten, & Laland, 2006) across interactional networks. Sociolinguists have shown that the changes triggered by these processes can extend across days or centuries.
4. **Genetic Diffusion.** We can also trace the diffusion and consolidation of genetic support for producing spoken and written language. We do not have space to consider genetic timeframes here, but the reader may wish to consult MacWhinney (2008a) for an emergentist analysis of language evolution.

For convenience, we refer to these inputs as “timeframes”, although it would be more accurate to call them “space-time frames”, because they involve both unique spatial configurations and unique temporal configurations. For example, social memesis operates within both the spatial frame of face-to-face interaction and the spatial frame of communication over the Internet. Differences in the configuration of these two spatial frames can lead to variation in the linguistic devices being used.

The forces operating on these various timeframes must have their effects on language use during speaking and listening. Sometimes, the relevant structures may lie dormant for months or years before achieving activation. For example, the *what's X doing in Y* construction found in *what is this fly doing in my soup* (Kay & Fillmore, 1999) only surfaces rarely. When it occurs, it expresses a unique configuration of shock or pretended shock regarding some untoward condition, and either enough social solidarity to withstand the intended irony or else a power differential that allows for expression of some level of approbation or even accusation. In order to operate effectively, this pattern must become consolidated into long-term memory in a way that permits efficient retrieval when this unique situational configuration arises. The various sociolinguistic and affective assignments needed to activate this pattern depend on the computation of the status of personal relations as they have developed across days, months, and years. These computations must then be linked to more immediate practical judgments regarding the unexpected nature of the condition (i.e., the fly in the soup). If the relevant, but rare, preconditions are not fulfilled, we may select a more neutral statement, such as *oh goodness, there is a fly in my soup*.

Two recent papers (MacWhinney, 2014, in press) present a fuller analysis of the ways in which linguistic processes mesh across timeframes. For our present purposes, what is important is to consider how different aspects of language learning align with inputs across different frames in time and space. Consider the learning of the correct articulation and perception of English /r/ and /l/ by speakers of Japanese. In the natural mode, this learning may require years of exposure, if it can be achieved at all. Attempts to teach this contrast (Ingvalson, Holt, & McClelland, 2012) indicate that it is difficult to direct learner's attention to the crucial role of the third formant (F3) in determining this contrast for this particular area of phonological space (Mann, 1986). What learning does occur may be based on piecemeal acquisition of combinations of cues that only generalize case-by-case (Bradlow, Akahni-Yamada, Pisoni, & Tokhura, 1999) across repeated training periods and through actual language usage. The problems that Japanese learners have with these sounds contrast with the fact that, in general, adults who receive focused phonological training can achieve nativelike articulation. Focused training can achieve similar results for other linguistic systems. For example, advanced English learners who are still making mistakes in their use of articles (a, an, the, and zero) can improve control of this skill based on only three hours of focused training (Zhao, 2012). Although focused training can achieve powerful effects, it is still necessary for learners to make use of this training in the longer timeframes of ongoing reading and conversation to consolidate these gains.

When thinking about ways to support second language learning, it is important to consider the ways in which the brain uses neural mechanisms to support different timescales of processing. MacWhinney (in press) introduces this topic by showing how the neurochemical processes that operate to consolidate memories in bees (Menzel & Giurfa, 2001) are sensitive to the ecology of pollen collection. Short-term consolidation processes work for gathering within patches of a certain type of flower. Longer-term consolidation operates for gathering across patches and flower types. Very long-term consolidation operates across days of activity. Each of these consolidation processes is designed to record patterns within specific timeframes. The same is true for the consolidation of memories in language learning through the operation of the hippocampus upon cortical storage (McClelland, McNaughton, & O'Reilly, 1995). Short-term consolidation is important for the maintenance of topics and styles during single interactions. Across interactions, consolidation works to acquire new words and chunks. Still longer-term consolidation is based on generalization, structure mapping, and proceduralization. To maximize consolidation on these various levels, input to the language learner must be presented consistently across the timeframes appropriate for the specific type of learning involved (Pavlik et al., 2007).

3.1 Support for Language Learning

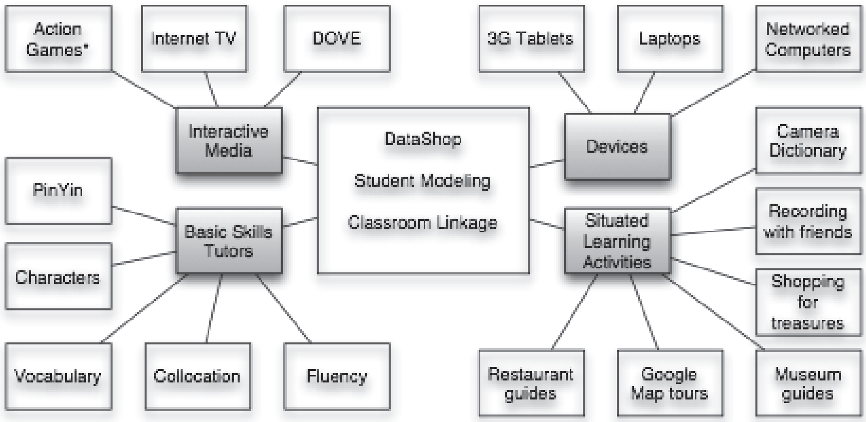
Given the multidimensionality of language learning, how can we develop materials to support meshed emergent learning across diverse structural levels and timeframes? One approach is to provide maximally natural input through work or study abroad or communicative method teaching in classrooms. These methods achieve holistic support, but not the type of targeted support needed to overcome the various risk factors facing the adult learner. Here, we will explore the construction of an integrated system of methods combining both holistic and analytic methods. This system – named eCALL – takes advantage of new methods for online language exposure, while still relying on support from the classroom environment. This system is being developed at Carnegie Mellon University by John Kowalski with support from a National Science Foundation grant to the Pittsburgh Science of Learning Center.

The eCALL system is designed to take advantage of the torrent of recent improvements in computer bandwidth, connectivity, operating systems, processors, programming languages, and high-resolution touch screens. These improvements have led to a rapid expansion in online resources such as games, dictionaries, grammars, translators, multilingual media, and Wiki pages. Researchers can now use these methods to gather rich longitudinal data on the actual process of second language learning, as it occurs in both instructed and naturalistic situations. Moreover, they can create a variety of online experiments to test the effects of alternative instructional treatments. Because of this, we can refer to this new field as experimentalized computer-assisted language learning or eCALL.

At the time of this writing, the eCALL system at http://talkbank.org/SLA includes:

- Individual difference measures to evaluate learner aptitudes.
- Games and tutors for learning vocabulary, Chinese characters, conjugations, and phonology.
- Fluency training using repetition, picture probes, and translation.
- Links to online resources, including dictionaries, TTS, television channels, and podcasts.
- Web pages with maps, instructions, and video dialogs that lead the learner through Pittsburgh, including sites such as museums, gardens, and shopping areas.
- The DOVE system for learning from subtitled video.
- Pictures and audio that help the learner with tasks such as ordering food in a Chinese restaurant.
- Methods for tracking learner progress in detail and reporting progress to instructors and learners.

The overall shape of the system is given in Figure 1 with the goal of illustrating concretely how we are thinking about the construction of the fuller eCALL system. Each of the pieces mentioned in this figure has been implemented, but many of these implementations are still preliminary. Although the system is currently accessed primarily from laptop computers, the software is programmed in HTML5, which allows for uniform deployment through browsers on tablets, laptops, and networked computers. This means that the system can be used anywhere the user has access to the Internet. At the center of the system described in Figure 1, are methods for recording learner usage patterns over the Internet in order to understand student preferences and track the progress of their learning in terms of individualized statistical models.



4 eCALL and topics in SLA

Online systems can track three types of learner data: learner error patterns, learner usage patterns, and experimental results. The eCALL system is designed to track all three types of data for storage in a central repository called DataShop. Error data are important in theory and model construction, particularly if they can be collected systematically across time. Usage data can be tracked both within and across the various modules of the eCALL system. For example, within the Pinyin Tutor, we can know how long students work with which levels of the tutor. If we find that students are only using the tutor for perhaps ten minutes at a time, we may wish to consider adding gamification features to increase the time that users will want to work with the tutor. Systems built using designs like that found in DuoLingo (http://duolingo.com) can track how users advance across levels and when and where they stop interacting with the system. We can also track preference patterns by comparing the selection of activities across eCALL modules. For example, we may find that learners of Chinese spend 15 minutes with the Pinyin Tutor module and then shift to working with subtitled video for 45 minutes. We can track such patterns across days, and use these data to guide further construction of the system.

In addition to gathering data on errors and usage, we can use the eCALL system to evaluate learning processes experimentally. To do this, we can use either between-groups or within-groups designs. For between-group designs, we must construct parallel versions of tutors, such as the Spanish conjugation tutor (Presson, Sagarra, MacWhinney, & Kowalski, 2013). We then assign users randomly to one of these versions. This method has the pedagogical disadvantage that it could lead to weaker learning for one set of students. Moreover, between-groups designs have weaker statistical power to detect the effects of experimental treatments. To correct these problems, we can use Latin Squares designs to create within-groups comparisons in which some target forms (words, sounds, sentences, constructions) are taught using one method (e.g. feedback vs. no feedback) and others systematically selected from the same larger set are taught in another way. The selection of items across treatments is counterbalanced across learners. However, this type of design can only be used if there is no significant transfer between items. Together, these two designs allow us to extend experimental methods to web-based tutors, thereby improving the experimental grounding of CALL systems (Felix, 2005). By refining and improving existing models through ongoing experimentation, we can ratchet up the quality of instruction, even as we continue to collect data that will improve our understanding of the basic cognitive mechanisms underlying second language learning. An iterative process of this type, in which theory informs intervention and interventions inform theory, can provide a solid basis for improving SLA theory.

In this section, we consider work that explores the roles of five instructional factors that have figured prominently in SLA research: the provision of corrective feedback, explicit rule instruction, repeated practice, modeling student knowledge, and gamification. Within each of these topics, we need to understand how language emergence is determined by competition, hierarchical structure, and timescales.

4.1 Corrective feedback

Corrective feedback (CF) can be a key engine for learning in the classroom (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Ellis, 2009). However, experiments and meta-analyses (Lyster, 2004; Lyster, Saito, & Sata, 2013; Sato & Lyster, 2012) have shown that there are wide variations in the effectiveness of CF depending on how it is provided, by whom, and to whom. Successful delivery of optimal levels of CF in the classroom requires a high level of instructor skill and engagement. Instructors must quickly diagnose the nature of the learner error and choose a feedback method, while avoiding interruptions of the normal flow of classroom interaction (Gardner & Wagner, 2005).

Providing CF through the computer can address some of these problems. When the computer provides CF, learners may be frustrated, but they will not suffer from embarrassment. Furthermore, the speed of computer processing makes it possible to provide immediate feedback that is predictably contingent on the nature of the learner's response. Many CALL programs only provide feedback after full blocks of trials, thereby losing one of the important advantages of online instruction. For this reason, our eCALL tutors all provide immediate CF, thereby operating within a timescale that allows the most effective comparison between the learner's response and the CF. Feedback can also be tailored to the individual learner's characteristics, the operative instructional theory, and specific instructional goals. Moreover, during interactions in a classroom, learners are being constantly exposed to errors that they themselves may never make, whereas the errors being corrected by the computer are specific to a particular learner interacting with a particular program.

Computerized CF may come in three basic forms: correction, recasting, and diagnosis. Simple correction involves telling the learner whether their response is right or wrong. Recasting involves also providing information about the correct answer. If the item only involves a binary true-false choice, then there is no difference between correction and recasting. However, when responses are more complicated, recasting can provide much more information. For example, if the learner translates *you went to the store* into Spanish as *tú fui al mercado*, the computer's recasting response would be *tú fuiste al mercado* in which the first person verb *fui* is restated or recast as *fuiste*. Just as an instructor could emphasize the word *fuiste* when providing this recasts, the computer can highlight the word being corrected. In order to provide diagnostic feedback, the computer would have to give the student some explicit rule or category, such as "*fui* is 1^{st} or 3^{rd} person preterite and you need 2^{nd} person preterite *fuiste* in this case." The Pinyin Tutor provides such diagnostic feedback by indicating whether the student's response has the wrong tone or wrong letter and in which position in which syllable. If the learner's response violates some rule of Pinyin spelling, the computer gives that rule and explains what alternative forms can be used. Thus, diagnostic feedback often includes explicit rule instruction.

There are three findings from the classroom CF literature that we should be able to apply to eCALL.

1. First, the evidence for better uptake of CF in the form of prompts by younger learners (Lyster & Saito, 2010) could suggest that eCALL instruction that targets younger learners should emphasize this form of CF.
2. The relation between high working memory capacity (WMC) and the ability to benefit from recasts (Goo, 2012; Sagarra, 2007) may also operate in the eCALL context.
3. The potential advantage of prompts over explicit correction in the classroom might possibly extend to eCALL instruction (Sanz & Morgan-Short, 2004). Recent classroom research has suggested that prompts may be more effective

than recasts in promoting learning (Ammar & Spada, 2006; Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Sheen, 2007; Yang & Lyster, 2010).

Implementing prompts through the computer is relatively easy. The standard classroom method is just to say please try again, but the computer can also deliver a simple repetition with a question mark at the end, clarification requests such as *pardon*? or partial feedback that repeats the correct segment and asks the student for a revision. There are good theoretical reasons to explore the role of prompts in eCALL. Tutorial design theory (Koedinger, Pavlik, McLaren, & Aleven, 2008) has characterized this issue as "the assistance dilemma". The dilemma here is how to decide when instruction should provide information and assistance to students and when it should request students to generate this information. On the one hand, assistance can function as scaffolding, but it can also serve as a crutch. Similarly, asking students to generate their own correct forms can function as effective teaching or it can lead to confusion and imposition of an unnecessary cognitive load. The best approach to this dilemma depends largely on details of the material to be learned, as well as characteristics of the learner.

In practice, the study of contrasting types and effects of CF in eCALL is still in its infancy. The work we will consider here has only been able to examine a few segments of this general problem space. One area of emphasis has been on the effects of diagnostic CF, as opposed to simple correctness feedback. Zhang (2009) examined this contrast in the context of a system called the PinyinTutor (http://talkbank.org/pinyin) that helps learners of Chinese practice the dictation into Pinyin of Chinese words and phrases. In this system, the instructor configures a set of about 20 words for practice, usually based on the words being studied in class and the textbook during that week. Each trial begins with the audio presentation of a Chinese word, such as *qing1wa1* "frog" and the learner's task is to write the word in Pinyin. If the transcription is correct, the learner moves on to the next item. If it is incorrect, the system tells the learner which segment or tone is wrong. Learners can also ask the system to play the sound of the form they have typed. Analyses of the effectiveness of this system showed that the experimental group that received diagnostic CF attained an 18 % improvement in accuracy, as opposed to the control group subjects who received simple correctness feedback and who attained a 10 % improvement across the duration of the study. This advantage for the training group was highly significant. In this task, there are many possible errors: the letters typed could be an illegal sequence in the Pinyin system, the initial or final sound of either syllable could be incorrect, there could be an incorrect number of syllables in the student response, and the tone of one or both syllables could be incorrect. In the diagnostic version of the program, students are given feedback regarding the exact nature of each error type. Moreover, each of these errors leads to different predictions about what the student needs to know, as well as different estimates of his or her current knowledge state (Kowalski, Gordon, & MacWhinney, 2013). In this system, response and latency data collected by the HTML5 program are transferred to servers at Carnegie Mellon University (CMU) for ongoing computation of adaptive feedback. The PinyinTutor data are also sent to the CMU DataShop repository (http://pslcdatashop.web.cmu.edu) for further offline analysis and possible future reanalysis. Providing correctness feedback immediately after the student types an answer makes correcting mistakes easier, and is important for student uptake (Ellis, 2009), but this basic feedback can be supplemented by diagnostic CF targeted at the specific component of Pinyin typing that led to the error.

Speech recognition technology also has the potential to provide immediate feedback on pronunciation, a feature that requires a particular type of feedback that does not rely on the student's own perception (Ehsani & Knodt, 1998). Refinement of pronunciation requires an outside observer to monitor and provide corrective feedback, as speakers are often unable to compare their pronunciation to a model. Cucchiarini, Neri and Strik (2009) further suggest that feedback on pronunciation should occur a) in a stress-free environment, b) in real-time, and c) be individualized for each speaker. Computer software kits like EduSpeak® (Franco et al., 2010) use phone-level mispronunciation detectors that, when they are able to produce reliable transcriptions, are comparable to human raters, and can provide feedback in real time and based on individual performance. These kits can be used to replace native speaker human listeners, which are often in short supply for second language learners. Speech recognition can also be used for the automatic scoring of oral fluency by focusing on temporal dynamics of speech, such as word count, length and rate of speech, and so on. The SpeechRater(TM) system uses this information, as well as rough estimates of number of repetitions and corrections, to provide real-time feedback to speakers on their oral performance (Zechner, Higgins, Xi, & Williamson, 2009). Speech recognition systems can provide feedback on both segmental (phonemes or syllables) and suprasegmental (prosody and intonation) features of the language (see Ehsani & Knodt 1998 for an overview).

4.2 Explicit rule instruction

In addition to varying CF, computers can implement different levels of explicitness in the presentation of linguistic patterns and rules. The value of explicit metalinguistic information for adult learners is a central and ongoing question in SLA theory. Meta-analyses (Norris & Ortega, 2000; Spada & Tomita, 2010) have indicated a general positive effect for explicit rule presentation. In an earlier review of this issue, MacWhinney (1997) concluded that explicit rule instruction was most useful when the rule to be learned was quite simple. In such cases, the rule can be kept active in working memory and used to match to incoming positive exemplars, thereby consolidating learning on both the explicit and implicit levels (MacWhinney, 2012). For example, in a computerized study of the learning of the Spanish counterfactual conditional, Rosa and Leow (2004) found that the most effective instruction involved explicit corrective rule feedback concurrent with each test trial. The rule governing this construction in Spanish is complex, but Rosa and Leow were able to formulate it on computer screens in a way that was concrete and memorable (p. 196). Although it is important to provide rule feedback immediately, this does not mean that the consolidation of learning from this feedback is itself immediate. Instead, we may want to think in terms of longer-term timescales or processing wheels that operate to convert this declarative feedback into specific automatic procedures (Anderson & Fincham, 1994; DeKeyser, 2001).

A major problem with the conclusions reached in meta-analyses of the effects of explicit feedback is that the studies involved have typically involved untimed measures that may not assess the proceduralization of the relevant skills. When evaluated in this way, explicit instruction could be viewed as "teaching to the test." It is clear that research needs to disentangle the effects of explicit rule instruction from the testing of these effects. There are three ways to do this. One way is to administer posttests that require generalization of the newly acquired knowledge to constructions not involved in the training. For example, training of French gender marking on the article should be able to support improvement in gender marking on the adjective. Generalization tests of this type are rare, but they can prove quite useful in assessing this issue. A second way of testing generalization is to examine changes in speed of processing. We report some initial attempts in this direction below. A third way is to examine retention. Many of the studies discussed here have administered repeated posttests across intervals of weeks and months to assess long-term retention. The assumption here is that explicit rule formulations should be more prone to loss than implicit or proceduralized learning.

To illustrate how we can use eCALL to approach this issue, consider two multi-session training experiments for novice learners of French learning to categorize nouns by grammatical gender (Presson, MacWhinney, & Tokowicz, in press). In these experiments, the rules for the orthographic cues to gender (e.g., words ending in *-age* are almost always masculine, or words referring to months and seasons are masculine) were presented in an eCALL tutorial. Learners were given a French noun and their task was to select its gender by choosing either *la* or *le*. They received correctness feedback with either no additional information, explicit orthographic cue statements (e.g., *-age* –> le), or highlighting of the relevant ending. The computerized training interface allowed for immediate feedback, randomization to feedback conditions, automatic data logging, and monitoring of participant progress. By computerizing the task, we were also able to test the prediction that, although explicit feedback might lead to better performance and greater retention after delay, this advantage could come at the cost of less rapid performance under time pressure, due to the additional time required to process explicit information. To that end, the program presented post-tests both with and without a time pressure constraint (a response deadline of 1400ms). This additional testing condition showed clearly that explicit cue feedback led to better learning and retention with no time pressure, but that explicit cue feedback led to greater accuracy even with a time pressure constraint. Furthermore, the addition of time pressure during training did not produce greater improvement, suggesting there may be less of a trade-off between learning explicit cues or rules and rapid online behavioral performance than often assumed. This study also showed that, with only 90 minutes of practice, learners' ability to judge the gender of French words rose from 62 % accuracy to 78 % accuracy (change is 50 %). Moreover, this ability was retained two months later, even though these novice learners were receiving no further exposure to French in the interval. The validity of the cues involved in this training was between 90 % and 97 %, and the training did not include high frequency exception words. Thus, a fuller study of learning of French gender in future experiments will need to include both training on the valid cue patterns and additional training on exception words to see how well the two types of training can be integrated.

4.3 Repeated Practice

A third issue in SLA theory is the role of practice. Critics of systems that require repeated trials have referred to such methods as "drill and kill". However, experimental work points to a strong linkage between practice and learning. At the very beginning of experimental psychology, Ebbinghaus (1885) studied the ways in which memory for new items and associations could be promoted by distributed practice that focused on the generation of remembered items. The ways in which consolidation of memories arise across timeframes has been a frequent subject for both psychological (McClelland et al., 1995) and neuropsychological (Wittenberg, Sullivan, & Tsien, 2002) models.

The role for such graduated interval recall and generation in vocabulary learning has been studied frequently within SLA (Barcroft, 2007; McNamara &

Healy, 1995; Pimsleur, 1967; Royce, 1973). Barcroft (2013) provides ananalysis of the various factors that can support or detract from vocabulary learning. His input-based incremental (IBI) approach emphasizes the importance of processes that forge a link between L1 and L2 items, as opposed to activities that focus on meanings in the languages separately. The formation of these links can be controlled through eCALL systems that emphasize the construction of resonant (MacWhinney, 2012) links between words in terms of paradigmatic associations, sensory associations, derivation, synonyms, and sentential context. Studies in the eCALL framework (Kowalski et al., 2013; Pavlik et al., 2007; Zhang, 2009) have used graduated interval methods to promote better learning and retention. To achieve the fullest effects of this approach, it must be used across long timescales and fully supported by classroom instructors.

In addition to training on grammar and vocabulary, computers can also be used to practice speaking. As described above, advances in speech recognition technology have allowed language learners to practice their pronunciation even without a teacher (for an overview of the use of speech recognition in language learning, see Eskenazi 2009). However, even without the use of speech recognition, speaking practice with a computer has shown to lead to improvements in both fluency and accuracy of speech. Yoshimura and MacWhinney (2007) had English learners of Japanese repeatedly read Japanese sentences and recite them from memory. They heard a sentence spoken by a native Japanese speaker, then read the sentence aloud six times, then repeated it from memory three times. Through these repetitions, they not only began to repeat the sentences with greater fluency but showed improvements in phonological accuracy of their repetitions as well.

Davy and MacWhinney (in preparation) used a task in which adult English learners of Spanish listened to a native Spanish speaker and repeat what they hear multiple times,. They found that this procedure leads to greater accuracy and fluency on later sentence production tasks compared to untrained sentences. Learners benefitted more from training that allowed them to practice complex sentences in separate phrasal groups than through complete sentence repetition. The study also showed how fluency training could be controlled through picture stimuli, thereby eliminating any reliance on pure echoic production of sentences (Erlam, 2006).

4.4 Student Modeling

SLA researchers have devoted considerable attention to understanding the ways in which individual differences in attention, memory, background, and motivation can affect the course of second language learning (Dörnyei, 2009; O'Malley & Chamot, 1990; Skehan, 1998). Despite the acknowledged importance of these factors, it is difficult to use this information to reshape classroom instruction. However, computers excel in their ability to collect systematic data on individual differences and to use those data along with performance data to tailor further instruction. Individual differences in memory, attention, and auditory accuracy can be collected systematically using online tests of the type available now at http://talkbank.org/SLA/tasks. Language background information can be gathered using questionnaires, such as the Language History Questionnaire (LHQ 2.0) at http://blclab.org/language-history-questionnaire (Li, Zhang, Tsai, & Puls, 2014).

In addition to characterizing individual differences that operate across longscale timeframes, eCALL programs can measure processes on shorter timescales across trials and sessions. Kowalski and Gordon (2013) used the 3 million responses to trials in the Pinyin Tutor to construct a general group-based model of learning of tones and segment combinations. They then used new incoming data from individual students to further tune the general model to match the abilities of the current student. They then used that model to optimize the presentation of additional new trials.

Zhao (2012) also created student models within an eCALL tutor for correct use of the English article system. She first created a list of cues for selecting which article (*a*, *an*, *the*, or no article) to use in various contexts. She differentiated between cues that are rule-based (e.g., when introducing new information into the conversation, use the indefinite article "a", as in "I just bought a new car.") and those that are feature-based (e.g., when giving a street name, do not use an article, as in "turn onto Fifth Avenue" instead of "turn onto the Fifth Avenue"). By tracing student performance on practice trials with each type of cue, she found that students could acquire proper use of rule-based cues easily without explicit instruction, but explicit instruction was needed for learning of feature-based cues. After only three hours of using the English Article Tutor, students showed a 23 % improvement in accuracy of article selection, moving from 53 % accuracy to 76 % accuracy (where chance accuracy is 33 %).

4.5 Gamification

Both SLA researchers and classroom teachers understand the importance of providing motivating contexts for language lessons. In the classroom environment, this is done through smooth lesson planning, careful preparation of textbooks, and establishment of entertaining interactive formats. Within eCALL, motivation is mostly controlled through methods that are collectively called gamification. Aleven et al. (2008) analyze computer learning games in terms of the three dimensions of mechanics, dynamics, and aesthetics (MDA). The mechanics are the rules and game movements; the dynamics arise when a player interacts with the game; and the aesthetics are the features that keep the player involved with the game. For SLA, we are particularly interested in the aesthetics, which may fall into these eight motivational categories: sensation, fantasy, narrative, challenge, fellowship, discovery, expression, and submission. Each of these motivations can be implemented through specific design features. For example, challenge can be implemented by keeping scores in which one competes against oneself or others. Sensation is implemented by attractive visual and auditory design. Narrative can be implemented by creating an adventure game within a given historical or fantasy context. We must be cautious in applying all of these features indiscriminately, as it remains to be seen whether such manipulations of the training environment can improve learner compliance to a training regimen, and whether they improve or harm learner outcomes in various language skills.

We are now seeing a proliferation of websites that emphasize the importance of both practice and gamification for second language learning. Sites like Memrise (http://www.memrise.com), Quizlet (http://quizlet.com), or Anki (http://ankisrs.net) are centered around vocabulary drills, allowing users to review vocabulary at any time on their computers or mobile devices, either by using provided “decks” of words or by creating their own custom lists. Because these commerical sites will not provide data for SLA research, it will be necessary for the SLA community to build such sites through community efforts. Once this is done, games can become a huge source of data for SLA theory and the further elaboration of eCALL.

5 Technical Considerations

In addition to their aid in the rigorous control of pedagogically relevant variables, computers can also ease the arduous processes of material creation and data collection and analysis. In this section we consider the use of computers in automatizing tasks that, when done by hand, can be prohibitively time-consuming: the generation of practice and testing materials, and data logging. We then consider how this system can be constructed as a general community resource to which individual projects can add additional modules and data.

5.1 Automating Materials Generation

Technology can also aid in the generation of eCALL materials, either for practice (Brown & Eskenazi, 2004; Heilman, Zhao, Pino, & Eskenazi, 2008) or testing (Feeney & Heilman, 2008; Pino & Eskenazi, 2009). This generation can be a difficult and time-consuming task for teachers and researchers, who must consider elements such as student level, relevant interests, subject matter, and representation of certain target vocabulary items or grammatical elements. Any degree of computerized automation of this task would be a great help to educators as they could then devote the time spent creating tests and training materials to other instructional activities.

For example, the REAP Project (or REAder's Practice) is designed to provide extensive reading practice for ESL learners, with the goal of increasing vocabulary. This project scrapes the Internet for reading passages containing vocabulary from the Academic Word List (Coxhead, 2000). The passages are then analyzed for content information and reading level so that students can read on topics that are both interesting for them and at a level appropriate for them. Whereas some of the passages are screened, either by researchers or by the students, many are simply mined and presented directly to students. The REAP Project also works to semi-automate the process of creating assessments to check whether students were reading the text, and measure vocabulary knowledge. For example, Feeney and Heilman (2008) discovered that simply by generating a list of unique words that appeared in the text and asking students to choose between that list and a similar list that also contains random words, they can see whether the students are actually reading. Student performance on this task correlated significantly with post-reading vocabulary assessments, suggesting that reading the passages leads to vocabulary growth.

Another example of semi-automated generation of practice trials comes from an L2 preposition tutor asking learners to move objects around a virtual room by following instructions presented in the L2 using spatial prepositions (Presson, MacWhinney, & Heilman, 2010) – available at http://talkbank.org/SLA/prepositions). For example, a learner can see (in Spanish) *Pick up the ball to the left of the plant and put it on the sink* and would click the object (ball) to move it to the target location (on the sink). The goal of this training intervention was to use spatial and enactment cues to strengthen the gains from practice comprehending preposition words. The tutor itself was created by manually segmenting target areas around each object (e.g., *above* and *under*) such that many objects could be arranged into multiple configurations. For example, a fixed object such as a chair, or a movable object such as a ball, could have another object to its left or right, above or below it, or more generally near it. Therefore, using those manual specifications, the program was able to automatically generate a much larger number of room scenes, each containing several fixed and movable objects that could be manipulated when following L2 instructions. This trial generation would have been extremely difficult if each individual trial had to be created separately and by hand, and the degree of automation results in a practically unlimited number of trials from a relatively limited practice set (although a greater variety of objects and locations could be highly beneficial to motivation and generalization).

In addition to generating natural language stimuli for real learners, computerization also aids in the generation of artificial and miniature languages that can be used to conduct tightly controlled laboratory studies of language learning processes (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Opitz & Friederici, 2007). These statistically reliable but sometimes complex stimuli can allow for systematic manipulation of language frequency and structure, as well as the processing difficulty for naive learners. For example, de Graaff (1997) used an artificial language in a computerized training paradigm to show that explicit instruction improved L2 grammar learning relative to no explicit instruction, and that this advantage was present for both simple and complex grammatical structures, and for both morphology and syntax. In this case, computerization was used to "optimally control the input and exposure" (p. 253), reflecting the advantage of both artificial grammar and computerized training.

5.2 Data logging

A core feature of eCALL is the capacity to log learner performance, instantaneously and automatically. With Internet connectivity, data from remote access to training materials can be logged to a central repository, and can be easily exported, manipulated, and analyzed. Without automatic logging, an experimenter often must code and log all responses manually, and sometimes must use coarse aggregate data (e.g., time to finish taking a test) instead of fine-grained trial- or student action-level data (e.g., time to finish each problem independently), which provide more detailed information about student behavior. Earlier, we noted how data from the PinyinTutor system (Zhang, 2009) is transmitted continually to servers at CMU. This system is now in use at 60 locations internationally, yielding semester-long learning records from 5844 students. As these data come in, they are collected into web pages that allow each instructor to monitor the usage and progress of each of their students. These scores can be used as a component of the class grade, thereby freeing the instructor from the task of grading pinyin dictation assignments. At the same time, the computer provides students with summary scores that allow them to track their own progress.

The benefit of automatic logging is especially clear in the emerging field of the neuroscience of second language learning. In running a neuroimaging experiment, brain data must be matched to the behavioral events that trigger or reflect the activation being recorded. This correspondence allows researchers to analyze the brain correlates of behavioral and perceptual events, which can provide key evidence into the mechanisms of second language learning (Abutalebi, 2008; Hernandez, Hofmann, & Kotz, 2007; Osterhout et al., 2008; van Hell & Tokowicz, 2010). Without a fine-grained temporal log of behavioral events and stimulus presentation, used to tag events on the resulting brain data, it would be difficult or impossible to associate changes in brain activity to events in the world. For example, Morgan-Short and colleagues (2010) showed different event-related potential (ERP) responses in high and low proficiency L2 learners for noun-article and noun-adjective agreement errors in a computer-generated artificial grammar, which was possible because of the ability to match automatically generated time-stamped stimulus presentation logs (thereby differentiating signals from article and adjective trials) to corresponding ERP data.

6 Language Learning in the Wild

Successful language learning depends crucially on having opportunities for conversational interaction with native speakers (Gass, Mackey, & Pica, 1998; Long, 1983, 1996). Such interactions can support acculturation (Firth & Wagner, 2007; Pavlenko & Lantolf, 2000), the growth of the use of L2 for inner speech or “thinking in the second language”, and improved motivation for learning (Dörnyei, 2009). The Unified Competition Model considers participation and internalization as primary protective factors promoting both L1 and L2 learning.

One approach to promoting conversational interactions relies on computer-mediated communication (CMC) in chatrooms. Chat sites like LiveMocha (http://www.livemocha.com) or systems such as Second Life (http://secondlife.com/destination/chinese-island) promote onine conversations, sometimes with other learners and sometimes with paid native speakers. It has been suggested that CMC may provide opportunities for negotiation of meaning of the type observed in face-to-face interactions (Varonis & Gass, 1985). However, Peterson (2006) found that these systems produce only low levels of meaning negotiation. In terms of the overall effects of chat rooms, there have been no controlled studies, and student reactions to these systems are often mixed (Peterson, 2010; Yao, 2009).

In the current eCALL system, we are exploring an alternative approach to CMC based on “language learning in the wild” or LLW (Wagner, this volume).

This approach seeks to use the computer as a backup method for the promotion of actual interactions in the community. The goal here is for the student to participate in real-life interactions of the type promoted in the Språkskap Project (Clark et al., 2011) for later analysis in the classroom. For example, a learner of Icelandic recorded her interactions across several months in a bakery in Rejkjavik. These data were transcribed and the transcripts were linked to audio records. The resulting corpus, called IceBase, is available to researchers from http://talkbank.org.

By using applications such as Recorder for audio or the built-in camera, learners can use mobile devices to record interactions with native speakers in sites such as restaurants, museum tours, excursions, or homes. Like the Icelandic corpus, these records can then be analyzed either for pedagogical or research purposes. In the classroom, these materials could help students understand conversational practices, pragmatic norms, linguistic forms, and methods for negotiating meaning. For researchers, the corpora can be analyzed by programs such as CLAN (MacWhinney, 2008b) for automatic lexical and morphosyntactic analysis or Praat (Boersma & Weenink, 1996) and Phon for phonological analysis (Rose & MacWhinney, in press). Within Praat, researchers (de Jong & Wempe, 2009) are developing methods for linking transcripts to audio at the word level. Once these methods are available, we will be able to conduct increasingly powerful analyses for fluency and phonological accuracy.

Although the collection of recordings from interactions for analysis in the classroom is an excellent method, there is often too much noise and movement for good recording. In other cases, it may be socially unacceptable to make recordings. This means that we need to consider other methods for linking the classroom to the real world. One method allows the classroom to focus on providing materials that will facilitate the real world interactions. For example, if students are asked to order coffee and pastries from a coffee shop, the classroom will review the names of the various coffee and pastry types and the standard ways of asking for items and responding to questions. Another method for making information available relies on QR (quick response) codes for segments of the relevant information that would be posted at the restaurants, museums, homes, or offices participating in the LLW program. Systems of this type have been implemented by many libraries, and we can view this as an extension of that approach. We can also build online dictionary facilities that respond to SIRI-like voice activation to retrieve relevant information at each site. At this point, technology is no longer the limiting factor in constructing methods for LLW. Our major task is to convince instructors of the value of integrating these methods with classroom work.

7 Conclusion

Both first and second language learning are shaped by competition, emergent hierarchical organization, and the meshing of inputs from divergent timeframes. Although first and second language learners have access to the same basic learning processes, second language learners are more subject to the risk factors of entrenchment, negative transfer, misconnection, parasitism, and isolation. To counteract these risk factors, second language pedagogy can emphasize the protective roles of resonance, positive transfer, proceduralization, internalization, and participation. A particularly promising way of doing this is to build an eCALL system that integrates a wide variety of learning experiences across varying space-time frames. This system can be designed to provide complete support for all levels of second language learning from vocabulary practice up to language learning in the wild. It should be constructed as a community effort that will provide rich, anonymized longitudinal data to the whole SLA community. Construction of this system can allow classroom teachers to focus on what they do best and to link their instruction to interactions in the real world. Once this system is fully constructed, we will be able to use learners themselves as the final arbiters of what works best for what aspects of language learning.

Recent advances have removed all the major technological barriers to construction of this system. What we need now is support for its construction from teachers, SLA researchers, and funding agencies. The best way to progress in this direction is to work together to continue construction of the eCALL system we have described here.

References

Abutalebi, J. 2008. Neural aspects of second language representation and language control. *Acta Psychologica, 128*, 466–478.

Aleven, Vincent, Myers, Eben, Easterday, Matthew, & Ogan, Amy. 2008. Toward a framework for the analysis and design of educational games. In G. Biswas, D. Carr, Y. S. Chee & W. Y. Hwang (Eds.), *Proceedings of the 3rd IEEE Conference on Digital Game and Intelligent Toy-Enhanced Learning* (pp. 69–76). Los Alamitos: IEEE Computer Society.

Ammar, A., & Spada, N. 2006. One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*, 543–574.

Andersen, Peter, Emmeche, Claus, Finnemann, Niels, & Christiansen, Peder (Eds.). 2000. *Downward causation: Minds, bodies, and matter*. Aarhus, Denmark: Aarhus University Press.

Anderson, John, & Fincham, J. M. 1994. Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 1322–1340.

Bangert-Drowns, Robert, Kulik, Chen-Lin, Kulik, James, & Morgan, Mary Teresa. 1991. The instructional effects of feedback in test-like events. *Review of Educational Research, 61*, 213–238.

Barcroft, Joe. 2007. Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*, 35–56.

Barcroft, Joe. 2013. Input-based incremental vocabulary instruction for the L2 classroom *Innovative research and practices in second language acquisition and bilingualism*. New York: John Benjamins.

Bates, Elizabeth, & MacWhinney, Brian. 1982. Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York: Cambridge University Press.

Beckner, Clay, Blythe, Richard, Bybee, Joan, Christiansen, Morten H, Croft, William, Ellis, Nick C, … Schoenemann, Tom. 2009. Language is a complex adaptive system: Position paper. *Language Learning, 59*(s1), 1–26.

Behaghel, O. 1923. *Deutsche Syntax*. Heidelberg: Winter.

Boersma, Paul, & Weenink, David. 1996. Praat, a system for doing phonetics by computer *Tech. Rep. 132*. Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam.

Bradlow, A., Akahni-Yamada, R., Pisoni, D., & Tokhura, Y. 1999. Training Japanese liseners to identify English /r/ and /l/: Long term retention of learning perception and production. *Perception and Psychophysics, 61*, 977–985.

Brown, J., & Eskenazi, M. 2004. *Retrieval of authentic documents for reader-specific lexical practice.* Paper presented at the InSTIL/IOCALL Symposium, Venice.

Brunswik, E. 1956. *Perception and the representative design of psychology experiments*. Berkeley, CA: University of California Press.

Campbell, D. T. 1960. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67*(6), 380–400.

Campbell, N. A., Reece, J. B., & Mitchell, L. G. 1999. *Biology: Fifth Edition*. Menlo Park: Addison Wesley.

Clark, Brendon, Wagner, Johannes, Lindemalm, Karl, & Bendt, Olof. 2011. *Språkskap: Supporting second language learning "in the wild"*. Paper presented at the INCLUDE, London.

Coxhead, A. 2000. A new academic word list. *TESOL Quarterly, 34*, 213–238.

Cucchiarini, Catia, Neri, Ambra, & Strik, Helmer. 2009. Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication, 51*, 853–863.

Darwin, Charles. 1859. *On the Origin of Species*. London: John Murray.

de Graaff, R. 1997. The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249–275.

de Jong, Nivja, & Wempe, T. O. N. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*, 385–390.

DeKeyser, Robert. 2001. Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125–151). Cambridge: Cambridge University Press.

Dörnyei, Zoltán. 2009. *The psychology of second language acquisition*. New York: Oxford University Press.

Ebbinghaus, H. 1885. *Über das Gedächtnis*. Leipzig: Duncker.

Edelman, Gerald. 1987. *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Ehsani, F., & Knodt, E. 1998. Speech technology in computer-aided learning: Strengths and limitations of a new CALL paradigm. *Language Learning and Technology, 2*, 45–60.
Ellis, R. 2007. The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 339–360). Oxford: Oxford University Press.
Ellis, R. 2009. Correct feedback and teacher development *L2 Journal 1*, 3–18.
Ellis, R., Loewen, S., & Erlam, R. 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*, 339–368.
Erlam, Rosemary. 2006. Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*, 464–491.
Feeney, C., & Heilman, M. 2008. *Automatically generating and validating reading-check questions*. Paper presented at the Ninth International Conference on Intelligent Tutoring Systems.
Felix, U. 2005. Analyzing recent CALL effectiveness research: Towards a common agenda. *Computer Assisted Language Learning, 18*, 1–32.
Firth, A., & Wagner, J. 2007. Second/Foreign Language Learning as a Social Accomplishment: Elaborations on a Reconceptualized SLA. *Modern Language Journal, 91*, 800–819.
Franco, H., Bratt, H., Rossier, R., Rao Gadde, V., Shriberg, E., Abrash, V., & Precoda, K. 2010. EduSpeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing, 27*, 401–418.
Gardner, R., & Wagner, J. 2005. *Second language conversations*. New York: Continuum.
Gass, Susan, Mackey, Alison, & Pica, Terri. 1998. The role of input and interaction in second language acquisition: An introduction. *Modern Language Journal, 82*, 299–307.
Goo, Jaemyung. 2012. Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*, 445–474.
Hagoort, Peter. 2013. MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4*, 1–13.
Heilman, M., Zhao, L., Pino, J., & Eskenazi, M. 2008. *Retrieval of Reading Materials for Vocabulary and Reading Practice*. Paper presented at the The 3rd Workshop on Innovative Use of NLP for Building Educational Applications.
Hernandez, A. E., Hofmann, J., & Kotz, S. A. 2007. Age of acquisition modulates neural activity for both regular and irregular syntactic functions. *NeuroImage, 36*, 912–923.
Ingvalson, Erin, Holt, Lori, & McClelland, James. 2012. Can native Japanese listeners learn to differentiate /r-l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*, 255–274.
Kay, Paul, & Fillmore, Charles. 1999. Grammatical constructions and linguistic generalization: The "what's X doing Y?" construction. *Language, 75*, 1–33.
Koedinger, K. R., Pavlik, Phillip, McLaren, B. M., & Aleven, V. 2008. *Is it better to give than to receive? The assistance dilemma as a fundamental unsolved problem in the cognitive science of learning and instruction*. Paper presented at the Proceedings of the 30th Annual Conference of the Cognitive Science Society Austin, TX.
Kowalski, John, Gordon, Geoff, & MacWhinney, Brian. 2013. Refining an assessment for improving dictation skills of Chinese syllables. *Journal of Educational Data Mining*.
Krashen, S. 1982. *Principles and practice in second language acquisition*. New York: Pergamon Press.
Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Li, Ping, Zhang, Fan, Tsai, Erlfang, & Puls, Brendan. 2014. Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(03), 673–680.
Li, Ping, Zhao, X., & MacWhinney, Brian. 2007. Dynamic self-organization and early lexical development in children. *Cognitive Science, 31*, 581–612.
Long, Michael. 1983. Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126–141.
Long, Michael. 1996. The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic
Lyster, Roy. 2004. Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*, 399–432.
Lyster, Roy, Saito, Kazuya, & Sata, Masatoshi. 2013. Oral corrective feedback in second language classrooms. *Language Teaching, 46*, 1–40.
Lyster, Roy, & Saito, Tetsuya. 2010. Oral feedback in classroom SLA. *Studies in Second Language Acquisition, 32*, 265–302.
MacWhinney, Brian. 1997. Implicit and explicit processes. *Studies in Second Language Acquisition, 19*, 277–281.
MacWhinney, Brian. 2005. The emergence of linguistic form in time. *Connection Science, 17*, 191–211.
MacWhinney, Brian. 2008a. Cognitive precursors to language. In K. Oller & U. Griebel (Eds.), *The evolution of communicative flexibility* (pp. 193–214). Cambridge, MA: MIT Press.
MacWhinney, Brian. 2008b. Enriching CHILDES for morphosyntactic analysis. In H. Behrens (Ed.), *Trends in corpus research: Finding structure in data* (pp. 165–198). Amsterdam: John Benjamins.
MacWhinney, Brian. 2008c. How mental models encode embodied linguistic perspectives. In R. Klatzky, B. MacWhinney & M. Behrmann (Eds.), *Embodiment, Ego-Space, and Action* (pp. 369–410). Mahwah: Lawrence Erlbaum.
MacWhinney, Brian. 2010. A tale of two paradigms. In M. Kail & M. Hickmann (Eds.), *Language Acquisition across Linguistic and Cognitive Systems* (pp. 17–32). New York: John Benjamins.
MacWhinney, Brian. 2012. The logic of the Unified Model. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 211–227). New York: Routledge.
MacWhinney, Brian. 2014. Competition across time. In B. MacWhinney, A. Malchukov & E. Moravcsik (Eds.), *Competing motivations in grammar and usage*. New York: Oxford University Press.
MacWhinney, Brian. (in press). Language Emergence. In B. MacWhinney & W. O'Grady (Eds.), *Handbook of Language Emergence*. New York: Wiley.
MacWhinney, Brian (Ed.). 1999. *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.
MacWhinney, Brian, & Bates, Elizabeth (Eds.). 1989. *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
MacWhinney, Brian, Bates, Elizabeth, & Kliegl, Reinhold. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*, 127–150.
MacWhinney, Brian, Malchukov, Andrej, & Moravcsik, Edith (Eds.). (in press). *Competing motivations in grammar and usage*. New York: Oxford University Press.

Mann, Virginia. 1986. Distinguishing universal and language dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition, 24*, 169–196.
McClelland, James L., McNaughton, Bruce, & O'Reilly, Randy. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.
McNamara, D. S., & Healy, A. F. 1995. A generation advantage for multiplication skill training and nonword vocabulary acquisition. In A. F. Healy & L. E. Bourne (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 132–169). Thousand Oaks: Sage.
Menzel, R., & Giurfa, M. 2001. Cognitive architecture of a mini-brain: the honeybee. *Trends in Cognitive Sciences, 5*, 62–71.
Mesoudi, Alex, Whiten, Andrew, & Laland, Kevin. 2006. Towards a unified science of cultural evolution. *Behavioral and Brain Sciences, 29*, 329–383.
Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. 2010. Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60*, 154–193.
Norris, J. M., & Ortega, L. 2000. Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.
O'Malley, M., & Chamot, A. 1990. *Learning strategies in second language acquisition*. Cambrdige: Cambridge University Press.
Opitz, B., & Friederici, Angela. 2007. Neural basis of processing sequential and hierarchical syntactic structures. *Human brain mapping, 28*, 585–592.
Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., … Hirschensohn, J. 2008. Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*, 509–521.
Pavlenko, A., & Lantolf, J. 2000. Second language learning as participation and the (re) construction of selves. In A. Pavlenko & J. Lantolf (Eds.), *Sociocultural theory and second language learning* (pp. 155–178). Oxford: Oxford University Press.
Pavlik, Phillip, Presson, Nora, Dozzi, G., Wu, S., MacWhinney, Brian, & Koedinger, K. R. 2007. The FaCT (Fact and Concept Training) System: A new tool linking Cognitive Science with educators *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1379–1384). Nashville, TN: Cognitive Science Society.
Peterson, Mark. 2006. Learner interaction management in an avatar and chat-based virtual world. *Computer Assisted Language Learning, 19*, 79–103.
Peterson, Mark. 2010. Computerized Games and Simulations in Computer-Assisted Language Learning: A Meta-Analysis of Research. *Simulation and Gaming: An Interdisciplinary Journal, 41*(1), 72–93.
Pimsleur, P. 1967. A memory schedule. *Modern Language Journal, 51*, 73–75.
Pino, J., & Eskenazi, M. 2009. *Semi-automatic generation of cloze question distractors' effect of students' L1*. Paper presented at the SLaTE Workshop on Speech and Language Technology in Education, Pittsburgh.
Presson, Nora, Davy, Colleen, & MacWhinney, Brian. 2013. Experimentalized CALL for adult second language learners. In J. Schwieter (Ed.), *Innovative research and practices in second language acquisition and bilingualism* (pp. 139–164): John Benjamins.

Presson, Nora, MacWhinney, Brian, & Heilman, Michael. 2010. *A simulated environment for the embodied learning of Spanish prepositions*. Paper presented at the Second Language Research Forum, University of Maryland.

Presson, Nora, MacWhinney, Brian, & Tokowicz, Natasha. (in press). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*.

Presson, Nora, Sagarra, Nuria, MacWhinney, Brian, & Kowalski, John. 2013. Compositional production in Spanish second language conjugation. *Bilingualism: Language and Cognition, 16*, 808–828.

Rosa, E. M., & Leow, R. P. 2004. Computerized task-based exposure, explicitness, type of feedback, and Spanish L2 development. *Modern Language Journal, 88*, 192–216.

Rose, Y., & MacWhinney, Brian. (in press). The Phon and PhonBank initiatives.

Royce, J. M. 1973. Memory effects for test-like events during acquisition of foreign language vocabulary. *Psychological Reports, 32*, 195–198.

Sagarra, Nuria. 2007. From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 229–248). Oxford: Oxford University Press.

Sanz, C., & Morgan-Short, K. 2004. Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*, 35–78.

Sato, Masatoshi, & Lyster, Roy. 2012. Peer interaction and corrective fedback for accuracy and fluency development. *Studies in Second Language Acquisition, 34*, 591–626.

Sheen, Y. 2007. The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A colleciton of empirical studies* (pp. 301–322). Oxford: Oxford University Press.

Sidtis, Diana. 2014. Formulaic language: acquisition, use, and neurological foundations. In B. MacWhinney & W. O'Grady (Eds.), *Handbook of Language Emergence*. New York: Wiley.

Simon, Herbert. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society, 106*, 467–482.

Skehan, P. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.

Spada, Nina, & Tomita, Yasuyo. 2010. Interactions between type of instruction and type of language features: A meta-analysis. *Language Learning, 60*, 263–308.

van Hell, Janet, & Tokowicz, Natasha. 2010. Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. Second Language Research, 26, 43–74. *Second Language Research, 26*, 43–74.

Varonis, E., & Gass, S. 1985. Non-native/Non-native conversations: a model for negotiating meaning. *Applied Linguistics, 6*, 71–91.

von Bertalanffy, L. 1968. *General system theory: Foundations, development, applications*. New York: George Braziller.

Wittenberg, Gayle, Sullivan, Megan, & Tsien, Joe. 2002. Synaptic reentry reinforcement based network model for long-term memory consolidation. *Hippocampus, 12*, 637–647.

Yang, Yingli, & Lyster, Roy. 2010. Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition, 32*, 235–263.

Yao, Tao-chung. 2009. The current status of Chinese CALL in the United States. *Journal of Chinese Language Teachers Association, 44*, 1–23.

Yoshimura, Y., & MacWhinney, Brian. 2007. The effect of oral repetition in L2 speech fluency: System for an experimental tool and a language tutor. *SLATE Conference*, 25–28.
Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. 2009. Zechner, K., Higgins, D., Xi, X. & Williamson, D. M. 2009. Automatic scoring of nonnative spontaneous speech in tests of spoken English. Speech Communication 51: 883–895. *Speech Communication, 51*, 883–895.
Zhang, Y. 2009. *A tutor for learning Chinese sounds through pinyin.* (Ph. D.), Carnegie Mellon University, Pittsburgh.
Zhao, Yun. 2012. *Explicitness, cue competition, and knowledge tracing: a tutorial system for second language learning of English article usage.* (Ph.D.), Carnegie Mellon University, PIttsburgh.