LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

CONCEPTUAL REVIEW ARTICLE

A Shared Platform for Studying Second Language Acquisition

Brian MacWhinney
Carnegie Mellon University

The study of second language acquisition (SLA) can benefit from the same process of datasharing that has proven effective in areas such as first language acquisition and aphasiology. Researchers can work together to construct a shared platform that combines data from spoken and written corpora, online tutors, and Web-based experimentation. Many of the methods and tools for building this platform are already available as a result of earlier work on corpus sharing in first language acquisition. By working together on a shared platform in a coordinated manner, researchers will be able to construct a rich new empirical basis for the study of SLA.

Keywords data-sharing; corpus linguistics; language learning; spoken language; multimedia; tutorials

Introduction

The study of second language acquisition (SLA) is an inherently interdisciplinary endeavor, deriving inputs from education, linguistics, neuroscience, psychology, and sociology. This is because second language (L2) learning itself is a multidimensional process, operating across a long timespan for individuals with very different sets of abilities interacting in diverse social configurations (MacWhinney, 2015a). Experimentation can elucidate parts of this process, but a full understanding of the course of SLA requires a combination of (1) experimental data; (2) measures of individual learner preferences, motivations, experiences, and aptitudes; and (3) corpus data documenting the course of SLA. This article will outline a vision for a shared infrastructure that can harvest and store data of this type for the advancement of SLA theory and practice.

Correspondence concerning this article should be addressed to Brian MacWhinney, Carnegie Mellon University, Department of Psychology, 5000 Forbes, Pittsburgh, PA 15213-3815. E-mail: macw@cmu.edu

DOI: 10.1111/lang.12220

This article examines, in sequence, the following nine issues regarding the construction of this shared infrastructure:

(1) a review of the current status of open spoken learner corpora from SLA-Bank, BilingBank, and elsewhere on the Web;
(2) the importance of a common SLABank protocol for the collection of SLA learner data (both spoken and written) across languages, sites, and projects;
(3) the eight data types that should be included in SLABank;
(4) methods for data collection and analysis for SLABank;
(5) the eCALL platform for collection of online data for inclusion in SLA-Bank;
(6) the construction of Web-based experiments that can supplement SLABank corpus data;
(7) the construction of online measures of individual differences in cognitive and perceptual skills predictive of success in L2 learning;
(8) additional components of the eCALL platform such as captioned video and support for language learning "in the wild"; and
(9) prospects and problems facing the implementation of this shared platform.

SLABank Corpora

Corpora play a major role in the study of SLA. In the closely related field of first language acquisition, the creation of a shared database of both cross-sectional and longitudinal data has led to major empirical advances. The Child Language Data Exchange System (CHILDES; MacWhinney, 2000) has relied on community datasharing to construct a 60-million-word database (with 2 terabytes of media and an additional 90 million words of annotation) containing over 180 corpora documenting language learning in 30 languages. Since its inception in 1984, the CHILDES database and its related computational tools have generated over 6,500 published articles. Over the last 2 years, CHILDES has received 2.5 million visits through the Web.

SLA research could benefit from corpus sharing in the same way that child language has (Fletcher, 2014). We have already made an initial beginning in that direction through the creation of SLABank, BilingBank, and the CHILDES Biling corpora. SLABank (talkbank.org/SLABank) includes 25 corpora with 4.2 million words from learners of Czech, English, French, German, Hungarian, Mandarin, and Spanish. BilingBank (talkbank.org/BilingBank) includes 11 corpora with 2.4 million words from adult bilinguals, often involved in codeswitching. The CHILDES database includes 31 corpora in the Biling section (childes.talkbank.org/data/Biling) with 4.9 million words. All of these corpora involve spoken language interactions, mostly in naturalistic settings. The majority of these corpora include audio that has been linked to or is being linked to the transcriptions on the utterance level to allow for careful analysis of errors, interactional features, and phonological patterns.

SLABank, Interoperability, and Federated Data

SLABank is one of 20 topic-specific collections of spoken language corpora available from the homepage at http://talkbank.org. To optimize interoperability, all of the transcripts in TalkBank use a common transcription format called Codes for the Human Analysis of Transcripts (CHAT) that supports transcription conventions from conversation analysis (CA; Atkinson & Heritage, 1984), child language, International Phonetic Alphabet coding, UNICODE (unicode.org), speech act analysis (Ninio & Wheeler, 1986), and sociolinguistic analysis. Transcripts in CHAT can be automatically converted to extensible markup language (XML) in accord with a fully documented XML Schema (talkbank.org/software/xsddoc/) within which each coding convention is linked to its full description in the downloadable CHAT manual. Transcripts in CHAT can be automatically validated for format accuracy and converted to XML using the CHATTER Java utility. The CHAT XML format is the same format as that used by the Phon program (www.phon.ca) for detailed phonological analysis and acoustic analysis of learner pronunciations based on complete integration with Praat (praat.org) within Phon. Using a module in the Pepper conversion system (Zipser & Romary, 2010) that was designed for CHAT, learner corpora in TalkBank can be imported to the ANNIS system (corpus-tools.org) for analysis on phonological, morphological, grammatical, referential, and discourse levels (Lüdeling, Walter, Kroymann, & Adolphs, 2005; Reznicek, Lüdeling, & Hirschmann, 2013). Using the CLAN program, videos of learner and classroom interactions can be exported to video annotation software such as ELAN (https://tla.mpi.nl/tools/tla-tools/elan/) and ANVIL (www.anvil-software.org) or other annotation systems, such as EXMARaLDA (exmaralda.org). Files in CHAT can be processed for automatic speech recognition (ASR) using methods built into the SpeechKitchen framework (speechkitchen.org). This method can be useful for analyzing script-based productions and word repetition data. CLAN also includes routines for automatic insertion of captions or subtitles from CHAT files into videos.

TalkBank sets a high priority on the importance of coordinating efforts across major international projects. Toward this end, TalkBank is the first center outside of Europe within the CLARIN federation (clarin.eu). To connect with researchers internationally, TalkBank provides metadata to the Virtual Language Observatory linguistic metadata harvesting system, as well as the Open Language Archives Community harvesting system (language-archives.org). TalkBank has also received the Data Seal of Approval (datasealofapproval.org) for its data, metadata, curation, permanent ID, and sustainability methods.

SLABank Tools

The TalkBank system has developed an extensive series of tools for analyzing spoken language transcript data. Based on over 30 years of software development and transcript analysis, there are systems for automatic computerized analysis of a wide variety of features in spoken language protocols. To run these analyses smoothly, corpora must be transcribed accurately following all the conventions of CHAT. Their accuracy is then checked by an XML validator called CHATTER. Once this is done, CLAN's MOR program can automatically produce a complete and disambiguated part of speech tagging, and CLAN's MEGRASP can produce a complete dependency grammar analysis (Sagae, Davis, Lavie, MacWhinney, & Wintner, 2010). Using these tags, CLAN can compute measures such as type-token ratio (TTR), vocabulary diversity (Malvern, Richards, Chipere, & Purán, 2004), moving average type–token ratio (Covington & McFall, 2010), Computerized Propositional Density Idea Rater (Brown, Snodgrass, Kemper, Herman, & Covington, 2008), Index of Productive Syntax (Scarborough, 1990), and Developmental Sentence Score (Lee, 1974). The EVAL program combines all of these various analyses and profiles into a single push-button computational analysis. Within seconds, EVAL can compare a given transcript against a given reference database across a cluster of 50 or more measures, ranging from syntactic profiles to error analysis and fluency. For each measure, the program gives the standard deviation value of the current transcript in comparison with the overall distribution in the database. EVAL has been used in recent studies in child language (Bernstein Ratner & MacWhinney, 2016; Brundage, Corcoran, Wu, & Sturgil, 2016) and aphasiology (MacWhinney, Fromm, Forbes, & Holland, 2015). For L2 studies, this method will allow a researcher to compare participants in terms of learner level or other grouping characteristics, once adequate comparison data become available.

Another important TalkBank feature is the linkage between transcripts and media. These links are created during transcription and can then be used for later playback either over the Web using the TalkBank browser or on the desktop using media and transcripts that are downloaded from the Website. Linkage to media is crucial for analyses of conversation, gesture, and the details of phonology. Detailed phonological analysis can be achieved by exporting from CHAT to the Phon format for further analysis in the Phon program, which was also developed by the TalkBank project. There are also complete linkages between Phon and Praat for further acoustic analysis. This is particularly important for the study of the growth of L2 fluency.

TalkBank has also developed methods for exporting the results of CLAN analyses to Excel, SPSS, and R (www.r-project.org). In particular, we have written R scripts for subjecting SLABank data to Dynamic Systems Theory growth curve analysis (Verspoor, de Bot, & Lowie, 2011), sociolinguistic R-Brul analysis, and various forms of time series analysis.

Other Corpora

Although there are many corpus resources for the study of SLA, none of these other resources provide the full range of curation, analysis, metadata sharing, open access, format systematicity, transcription precision, and interoperability provided by SLABank. Many other corpora also come with difficult restrictions on data access. For example, the TLA (tla.mpi.nl) has archived a series of 20 corpora relevant to SLA, but none of them can be freely downloaded.There is a comprehensive list of learner corpora from the Catholic University of Louvain at https://www.uclouvain.be/en-cecl-lcworld.html. Most of these are written language corpora, except for six corpora from SLABank that are also included in this list. As of August 2016, there are 16 corpora with active Websites that are listed as available to other researchers. Of these, 12 only permit keyword searches in the corpus, but not complete downloading of the data. There then remain four corpora of written narratives that are freely downloadable over the Web. We hope to work with the contributors of these four corpora to reformat them into CHAT for addition to SLABank. The Centre for English Corpus Linguistics (CELC) and the University of Louvain has itself generated interesting and useful corpora using a variety of consistent elicitation and coding methods (https://www.uclouvain.be/en-426993.html). Some of these corpora are available through CD-ROM, but without audio and with tight usage restrictions, limiting further circulation or reformatting. There is also a corpus of 60 million words from essays produced by learners of English called EFCamDat (https://corpus.mml.cam.ac.uk/efcamdat/; Geertzen, Alexopoulou, & Korhonen, 2013), which is becoming fully available (personal communication) and which would be an excellent potential addition to SLABank.

SLABank Protocols

Ideally, many of the corpora we have discussed can be reformatted and curated for inclusion in SLABank. In addition to making current corpora more available, we want to consider ways of collecting new corpora that can maximize the possibility of making systematic comparisons across learner levels, educational approaches, elicitation methods, and languages, as is now being done for data in CHILDES and AphasiaBank using the EVAL program.

To maximize comparability across new data collections, we need to construct a collection of shared protocols, much like the shared protocol for TalkBank's AphasiaBank database (talkbank.org/AphasiaBank/protocol). This collection of protocols could include the following components:

1. *Spoken language protocol data*. Ideally, these data would be data gathered through use of a standard spoken language protocol of the type used for AphasiaBank, as described in detail at http://aphasia.talkbank.org/protocol. The protocols used by the LINDSEI and NESSI corpora collected by the CELC could form a part of this protocol.
2. *Written language protocol data*. The database should also include written corpora. Currently, there are dozens of written L2 corpora, although most are not openly available. If the SLA community and language educators could implement a common protocol that uses a Web-based method for collecting written language samples, it would be easy to configure a new database of written samples. This collection method would include a core set of writing tasks. It would also include Web-based methods for collecting anonymized demographic information and systems for entering data into CHAT transcripts for automatic lexical, morphosyntactic, and discourse analysis (Amaral & Meurers, 2011; Meurers, 2012). The major challenge involved in the analysis of written corpora is the requirement that incorrect spellings be normalized to standard orthographic versions to support automatic morphological and syntactic analysis. Fortunately, there has been a great deal of work on this topic (Flor, Futagi, Lopez, & Mulholland, 2015), and we can use these methods for the written component of SLABank.
3. *Naturalistic recordings*. Although naturalistic recordings seldom allow for experimental controls, they are important for studying interactional practices (Gardner & Wagner, 2005). For example, the Vienna-Oxford International Corpus of English at http://voice.univie.ac.at/pos/ and the related Asian Corpus of English document the usages of English in widely varying natural contexts by speakers from many different L1 backgrounds. Another interesting format for the collection of naturalistic data involves the

creation of "language villages" in which L2 learners are encouraged to interact with native speakers of the target language without resorting to English. An example of this format is the Icelandic Village organized for learners of Icelandic in Rejkjavik (Clark, Wagner, Lindemalm, & Bendt, 2011). Language villages are very different from virtual reality systems such as Second Life. In language villages, real people are selling goods in real shops and engaging in face-to-face interactions with learners.

4. *Longitudinal corpora*. The biggest gap in current data on SLA is that we have no openly available densely collected (Maslen, Theakston, Lieven, & Tomasello, 2004) longitudinal SLA data. The FLLOC and SPLLOC corpora in SLABank are important longitudinal corpora. However, they are not as extensive and as dense as we would wish. Initially, it would be most efficient to collect data from younger learners who are working intensively to acquire L2 proficiency. For example, the learner who figures in the Ice-Base corpus in SLABank acquired significant control of Icelandic within a year. If we could collect dense longitudinal data of this type, we could better evaluate proposals regarding critical periods (DeKeyser & Larson-Hall, 2005), fossilization (Birdsong, 1999), input-driven learning (Long, 1996), transfer (MacWhinney, in press), and dynamic systems effects (van Geert & Verspoor, 2015; Verspoor et al., 2011). Naturalistic longitudinal corpora could also be supplemented by online tests for growth incomprehension (Vandergrift, 2007), elicited production (Erlam, 2006), vocabulary (Horst, Cobb, & Nicolae, 2005; Wesche & Paribakht, 1996), and sentence processing (Schmitt & Underwood, 2004). For the recording and analysis of naturalistic data, we could use ASR technology of the type originally developed by the LENA Foundation (lenafoundation.org) for the study of child language development. This system generates and processes day-long recordings in the home. Although these recordings are not transcribed, they are marked for utterance length and speaker identity using a variety of ASR methods. This new data type is now being stored, curated, and analyzed by TalkBank in the context of the new National Science Foundation (NSF) HomeBank project at homebank.talkbank.org. Data of this type, collected in environments involving a significant usage of L2, would be ideal for the study of the actual process of L2 learning.
5. *Classroom recordings*. In the field of the Learning Sciences, video recordings of classroom interactions are a major method for developing theory and practice (Goldman, Pea, Barron, & Derry, 2014). TalkBank has collected materials in this format within its ClassBank section (http://talkbank.org/ClassBank). A great deal of SLA research also focuses

on the process and results of classroom teaching. Given this emphasis, it is remarkable that video recordings of successful classroom practices are not made systematically available. Ideally, this segment of the database could be used to provide real-life illustrations of methods such as task-based language teaching (Van den Branden, 2010), processing instruction (VanPatten, 2011), content and language integrated learning (Dalton-Puffer, 2011), concept-based instruction (Erickson, 2007; Negueruela & Lantolf, 2006), peer interaction and collaborative learning (Swain, 2000), or even total physical response (Asher, 1969). For each of these instructional types, video materials could be collected across several sessions of instruction to demonstrate ways in which learners advance through instruction and how they deal with feedback and language challenges. There are currently hundreds of videos at youtube.com that could be used as the basis for a systematic inventory and comparison of instructional methods. These materials can be made available both as videos at http://Iris-database.org and in the form of transcripts linked to the video from http://sla.talkbank.org.

Browser-Based Experiments

The elaboration of SLA theory requires data from both corpora and experimentation. Corpora are important because they track the development of L2 abilities. Experiments are important, because they allow the tight control of stimulus variables that is not possible in real-life situations. SLA theory must be grounded on data from both of these sources. Fortunately, recent developments in HTML5 (http://www.w3.org/TR/html5/) make it possible to develop experimental materials with audio and video that can be deployed through Web browsers on desktops, laptops, tablets, and mobile devices. We will consider three types of data collection that are using this new format: online cognitive tutors, online individual difference measurements, and online experiments.

Online Cognitive Tutors

With support from NSF, the Pittsburgh Science of Learning Center has applied learning theory from Adaptive Control of Thought–Rational (Koedinger, Aleven, Roll, & Baker, 2009), the Knowledge-Learning-Instruction framework (Koedinger, Corbett, & Perfetti, 2012) and the Competition Model (MacWhinney, 2015b) to develop a series of online cognitive tutors for mathematics, science, and L2 learning. All of the data produced by students (typing, keystrokes, timing, etc.) interacting with these tutors are transmitted online to local MongoDB databases at Carnegie Mellon University (CMU). These data are then later included in a larger database for the study of cognitive processes in learning called DataShop (http://learnlab.org), which provides free access to many data sets and supports a variety of statistical and data-mining analyses.

Within this framework, we have developed a series of Web-based tutors, which we refer to as eCALL or experimentalized CALL tutors (Presson, Davy, & MacWhinney, 2013). These tutors are designed to help learners in acquiring difficult aspects of Cantonese, English, French, German, Latin, Mandarin, and Spanish. These tutors are currently being used by thousands of students at universities and public schools around the world. We have implemented nine cognitive tutors that are not only useful in themselves, but also demonstrate the potential of developing tutors for L2 research. Each of these is available for testing and use from sla.talkbank.org.

1. *The Pinyin Tutor*. This tutor is designed to help learners accurately decode the phonological shape of Mandarin words. Learners listen to the sound of a target word and then use a keypad to enter the syllables in terms of initials, finals, and tones. Lessons are graduated, beginning with simple monosyllables and progressing to longer words and phrases. This system is currently in use at 60 universities or high schools, reaching about 2,000 learners each year. Using data collected online from these sites, Kowalski, Gordon, and MacWhinney (2014) created machine-learning models that isolated the most difficult aspects of Mandarin pronunciation for these students. The data also showed that beginners learn better with restricted rather than unrestricted vocabulary, but that intermediates learn better with unrestricted vocabulary. There is also a version of this tutor for the learning of Cantonese.
2. *French dictation*. For intermediate learners of French, accurate spelling is often a challenge. Our French dictation tutor helps learners with this process by providing complete diagnosis of each correct and incorrect letter in the learner's output.
3. *Virtual reality*. To explore the use of virtual reality for language learning, we have created a series of rooms populated with common objects such as tables and lamps. Learners listen to sentences in Spanish instructing them to place objects in particular locations. This activity focuses on developing rapid comprehension of spatial relations in complex sentence constructions. After 45 minutes of practice with this tutor, learners have shown a significant improvement in processing of these structures (Presson, MacWhinney, & Heilman, 2010).
4. *English articles*. This tutor focuses on teaching the correct usage of English articles (the, a, an, and zero)—a major challenge for learners whose languages do not have articles. The instruction focuses on a set of

25 cues for article selection. Experimentally, it compares teaching through rules with teaching through examples (Zhao, 2012; Zhao, Koedinger, & Kowalski, 2014).
5. *English prepositions*. This tutor uses diagrams based on Cognitive Grammar (Langacker, 1987) to help learners understand the meanings of phrases that extend the spatial meanings of prepositions figuratively, as in *watch over* as an extension of the spatial meaning of *over* (Lakoff, 1987).
6. *Wikipedia article tutors*. This tutor takes pages from Wikipedia and creates multiple-choice pull downs for specific word types. Currently, it has been implemented to allow learners to select articles in English and German. However, the method is quite general and can be easily extended to other closed class choices.
7. *Spanish conjugation*. This tutor was used to conduct an experimental investigation of the learning of regular and irregular forms in the present, imperfect, and perfective tenses (Presson, Sagarra, MacWhinney, & Kowalski, 2013). The results supported the hybrid morphology model (Gor & Cook, 2010; MacWhinney, 1978).
8. *French gender*. This tutor was used to investigate the learning by novices of the cues to French nominal gender (Presson, MacWhinney, & Tokowicz, 2014). With only 90 minutes of practice, learners' ability to judge gender accurately rose from 62% to 78%. Moreover, this ability was retained after 2 months with no further exposure to French.
9. *German case*. This tutor compares concept-based instruction with more conventional instruction for case marking in German transitive sentences.

Each of these tutors includes some form of experimental comparison to test for the operation of basic learning principles, such as the effects of graduated interval recall (Pavlik et al., 2007; Pimsleur, 1967) in the Pinyin Tutor, prototypes for categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) in the French gender tutor, cue simplicity (MacWhinney, 1997) in the German case tutor, embodied representation (Barsalou, 1999) in the virtual reality tutor, concept-based instruction (Negueruela & Lantolf, 2006) in the German case tutor, lexical binding to phonological structure (Vihman, 2015) in the Pinyin tutor, and input token vs. type frequency (Ellis, O'Donnell, & Römer, 2015) in the French gender tutor. In most cases, these treatment comparisons are built into the design of these tutors in a way that allows for within-subject statistical analysis. To achieve this, students are randomly assigned to one of two conditions in a Latin Square design. Within that design, they learn one set of items under one of the conditions and the other set under the other condition.

Because condition assignment is balanced across subjects, these results can be analyzed in terms of a within-subjects design, thereby significantly increasing statistical power.

There are many commercial resources for Web-based L2 learning. Companies such as DuoLingo, Rosetta Stone, Fluenz, Babbel, Busuu, Livemocha, Pimsleur, Udemy, and Pearson offer a wide array of courses and activities. The Apple Store and Google Play offer a dazzling array of apps for learning vocabulary, phrases, and parts of grammar. However, from the viewpoint of SLA researchers, these commercial resources suffer from two problems. The first is that they provide no integration with classroom activities. In fact, some are presented as complete replacements for the classroom. Although these online replacement courses can succeed in some ways (Chenoweth, Ushida, & Murday, 2006; Murday, Ushida, & Chenoweth, 2008), they often provide no opportunities for spoken language production; and when language production is required, feedback is often incomplete or inappropriate. The second problem with these materials is that they provide no data for the study of L2 learning. The commercial enterprises that have developed these materials often view datasharing as a threat to their business model. There are some exceptions to this rule. For example, we are working with Constant Therapy (constanttherapy.com) to acquire data for evaluating the efficacy of their online therapy for aphasia. However, such openness to datasharing is certainly the exception. This means that, if SLA researchers want to be able to use the Web to study the process of L2 learning, they will need to build their own resources for this study.

Online Measures of Individual Differences

SLABank has also begun to construct a set of Web-based language measures for assessing learner aptitudes at http://sla.talkbank.org/tasks. The measures currently implemented include Digit Span (Miller, 1956), Flanker (Eriksen & Eriksen, 1974), and Letter-Number Sequencing (Wechsler, 1997) tasks. These are provided with instructions and stimuli for English, Mandarin, Cantonese, German, Dutch, French, Italian, and Korean. We are planning to add additional psychometric tests measuring perceptual accuracy, sentence repetition, auditory discrimination, auditory memory, statistical learning, synonym generation, continuous performance task, Stroop, visual memory, attentional network task, go/no-go, n-back, and number series. Alongside these psychometric tasks, we will make available a series of questionnaires to assess learner preferences, motivations (Dörnyei, 2009), language history (Li, Sepanski, & Zhao, 2006), experiences, and attitudes. We will also select suitable material from Irisdatabase.org (Marsden, Mackey, & Plonsky, 2016), which currently hosts a large quantity of results from digit span, vocabulary, and questionnaire data, although none of these are currently implemented online.

Online Experiments

The focus of our experimental work has been on the evaluation of treatment effects within cognitive tutors. However, we have also configured some nontutorial Web-based experiments that are hosted on our servers, but administered through signup using Amazon Mechanical Turk (AMT). These are prefaced with consent forms to satisfy Institutional Review Board (IRB) requirements. The examples we have posted at sla.talkbank.org include a study of usage of German gender and a cloze test of the ability of native speakers and learners to guess the identity of German separable prefixes occurring at the end of sentences. Depending on the goals of the study and the nature of the participant group, an experimenter may decide to use either experiments within Web-based tutors or direct running of experiments using recruitment and payment methods such as AMT. The newly developed Experiment Factory system (https://expfactory.github.io/) provides methods that will allow researchers to configure new experiments directly for administration over the Web. Experiments built using the Experiment Factory framework can be hosted directly from the sla.talkbank.org Website.

The Language Partner

The systems we have described so far focus on the collection of new SLABank corpora and systems for Web-based tutors (eCALL), individual difference measures, and online experiments. The full shared platform that we are envisioning includes additional methods that will encourage the student to engage more actively with the L2 and its culture. We can refer to this complete system as the Language Partner. Figure 1 illustrates the shape of this larger system.

The six shaded components represent the major activity types in the system: experiments, measures, interactive media, basic skills tutors, corpora, and situated learning activities. We have described the role of experiments, measures, corpora, and basic skills tutors in previous sections.We will describe the role of Interactive Media and Situated Learning Activities below.

Each of these components includes additional nonshaded modules (not all listed in Figure 1) that transmit detailed learning data to a central MongoDB database from which data-mining methods (Kowalski et al., 2014) can construct individual and group learner models. The contents of individual modules can be tailored to the requirements of individual instructors using

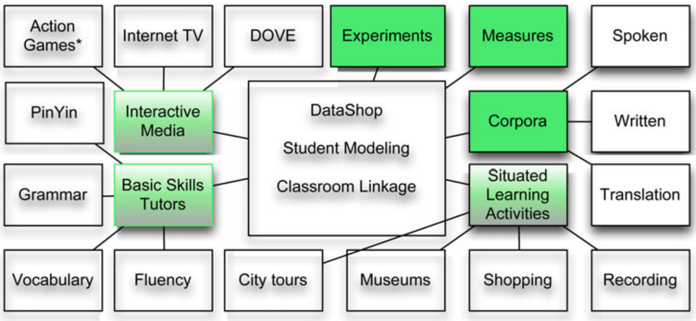

Figure 1 The Language Partner.

different textbooks or methods. Each instructor can view the students' results in a custom Webpage for that class.

Interactive Media

The Web abounds with opportunities for students to listen and watch programs in a L2. It is important for instructors to point learners in the direction of materials that can be useful for them. Often this requires matching learner reading level and interests with available materials (Heilman, Zhao, Pino, & Eskenazi, 2008). However, even when such a match has been achieved, it may be difficult to control the playback of L2 materials and well-linked captions are not always available. To address this problem, we have developed the DOVE (Deploying Online Video for Education) system that provides captioning with playback and repetition control for L2 video. There is evidence that learners working with captioned materials can achieve marked increases in language learning (Debuse, Hede, & Lawley, 2009; Mitterer & McQueen, 2009; Winke, Gass, & Sydorenko, 2010).

Situated Learning Activities

Finally, we can use the computer to promote and support real-world interactions in the community. The goal here is for the student to participate in real-life interactions of the type that Wagner (2015) refers to as Language Learning in the Wild (LLW). For example, a learner of Icelandic recorded her interactions 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 iOS or Android applications such as Recorder for audio or the built-in camera, learners can record interactions with native speakers in sites such as restaurants, museum tours, excursions, or homes. This can be done in the context of prestructured City Tours of the type illustrated at sla.talkbank.org using a tour of the CMU campus and the Frick Museum. Like the Icelandic corpus, these records can then be analyzed either for pedagogical or research purposes. Back 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 for automatic lexical and morphosyntactic analysis or Praat (Boersma & Weenink, 1996) and Phon for phonological analysis (Rose, 2010).

LLW can share methods with task-based language teaching (Skehan, 2003; Van den Branden, 2010). 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. Learners will then record their interactions with people in the coffee shop to see how well they were able to communicate and where there are still gaps. As such, LLW can be embedded in a pretask/posttask focus on form cycle. Another method for making information available would be to create 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. The major task would be to convince instructors of the value of integrating the LLW approach with classroom work and to adapt the technology to their specific needs.

The Ideal World

In this article, we have reviewed the outlines of an integrated platform for the study of L2 learning. This platform would present learners with a wide variety of language learning activities. Some may be required as parts of the classroom curriculum; others are made available as optional learning exercises. In both cases, we can evaluate not only how much these activities further language learning, but also whether or not learners enjoy particular activities more than others. The motivational aspects of many of these activities can be enhanced through gamification (Aleven, Myers, Easterday, & Ogan, 2008; Thorne, Black, & Sykes, 2009) and score posting. Some activities, such as captioned video, city tours, and action games are already high in entertainment value. Others, such as translation and dictation, will require the student to devote more attention and discipline to learning the material at hand.

All of these activities are being programmed using the HTML5 framework that allows for the creation of a single body of code that can be delivered through any modern browser and that can also be reconfigured for use on a tablet as an app. Whether the activity is running in a browser or as an app, the data are transmitted back to a MongoDB database and the anonymized results are available to all researchers either from sla.talkbank.org or through DataShop at pslcdatashop.org. Because all corpora are formatted in CHAT, they can be analyzed using the CLAN programs. In addition, CHAT files can be output in XML for analysis through other programs and data are easily sent to R and other statistical programs.

Many of these exercises are designed in ways that allow them to be readily adapted for other languages. For example, the Spanish conjugation tutor can be adapted for Latin, Arabic, or other languages. Similarly, the French dictation tool can be used for other languages. We are also designing a vocabulary flashcard-type application that is linked to a core database for translations and images that can be redeployed for new languages. All of these require many hours, even years, of programming effort. However, the design of the system is such that multiple institutions and researchers could contribute to the effort, when resources are available.

Open access is a crucial feature of the proposed system. Access must be open for learners, instructors, and researchers alike. We will make all materials and data available, although some more personally sensitive materials will be protected through passwords. For learners, we are assuming that, by default, access will be controlled through classroom assignments, but independent use of the materials is also encouraged. For instructors, we are developing methods that allow them to configure materials as needed for their particular curriculum. For researchers, there is the expectation that users will also be contributors.

The Real World

We have described an ideal world in which researchers, instructors, and learners work together for the common good. However, the real world places certain limits and requirements on this cooperation. One requirement is that there must be advance planning and proper configuration of IRB releases and consent forms to allow for datasharing. A second limitation involves the need for monetary resources for what is clearly a major programming effort. Without a major source of funding, it will be difficult to move forward quickly. However, we can still make progress through collaboration based on a diverse set of funding sources. The third barrier is the fact that researchers will not always receive academic credit for their contributions to this system. TalkBank corpora receive DOI and ISBN numbers, but many institutions will only give credit for journal publications. This means that some researchers will need to focus on those aspects of the system that permit experimentation, rather than those that simply maximize learning or corpus development. Even here, resources are needed to configure segments of the system for particular experiments.

The greatest barrier to rapid development of this system stems from the fact that we need to devise ways in which the eCALL system can fit in smoothly with classroom practice (Gleason, 2013). It is important that the eCALL approach be configured not as a replacement for the language teacher, but as a way of maximizing the value of the classroom by allowing it to focus on exactly those activities that cannot properly be supported by the computer or the wider community.

How to Proceed

Researchers can contribute to this effort in six ways. First, they can contribute learner corpora to SLABank, following the procedures at http://talkbank.org/share/contrib.html. Second, they can propose and test elements of a core protocol for standardized collection of both spoken and written corpora. Third, they can formulate and program measures of learner abilities to be hosted through the SLABank server. Fourth, they can design Web-based language tutors and materials using HTML5 methods and/or Experiment Factory. Fifth, they can use current SLABank corpus and experiment data to test empirical hypotheses. Sixth, and perhaps most importantly, they can organize and seek funding for new data collection projects whose results will interface with SLABank. The TalkBank staff (reachable at macw@cmu.edu) is prepared to encourage and support all of these contributions.

Final revised version accepted 6 September 2016

References

Aleven, V., Myers, E., Easterday, M., & Ogan, A. (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, CA: IEEE Computer Society.

Amaral, L. A., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, *23*(01), 4–24. doi:10.1017/S0958344010000261

Asher, J. (1969). The total physical response approach to second language learning. *The Modern Language Journal*, *53*, 3–17.

Atkinson, J., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis*. New York: Cambridge University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Bernstein Ratner, N., & MacWhinney, B. (2016). Your laptop to the rescue: Using the Child Language Data Exchange System archive and CLAN utilities to improve child language sample analysis. *Seminars in Speech and Language*, *37*, 74–84.

Birdsong, D. (1999). Introduction: Whys and why nots of the Critical Period Hypothesis. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 1–22). Mahwah, NJ: Erlbaum.

Boersma, P., & Weenink, D. (1996). *Praat, a system for doing phonetics by computer*. Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam. Retrieved from http://www.fon.hum.uva.nl/praat/

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, Instruments, and Computers*, *40*, 540–545. doi:10.3758/BRM.40.2.540

Brundage, S., Corcoran, T., Wu, C., & Sturgil, C. (2016). Developing and using big data archives to quantify disluency and stuttering in bililngual children. *Seminars in Speech and Language*, *37*, 117–127.

Chenoweth, A., Ushida, E., & Murday, K. (2006). Student learning in hybrid French and Spanish courses: An overview of language online. *CALICO Journal*, *24*, 115–145.

Clark, B., Wagner, J., Lindemalm, K., & Bendt, O. (2011). *Språkskap: Supporting second language learning "in the wild."* Paper presented at the INCLUDE, London.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*, 94–100. doi:10.1080/02687038.2012.693584

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204.

Debuse, J., Hede, A., & Lawley, M. (2009). Learning efficacy of simultaneous audio and on-screen text in online lectures. *Australasian Journal of Educational Technology*, *25*, 748–762. doi:10.14742/ajet.1119

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88–108). Oxford, UK: Oxford University Press.

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

Ellis, N., O'Donnell, J. J., & Römer, U. (2015). Usage-based language learning. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 163–180). New York: Wiley.

Erickson, H. L. (2007). *Concept-based curriculum and instruction for the thinking classroom*. Thousand Oaks, CA: Corwin Press.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Perception and Psychophysics*, *16*, 143–149. doi:10.3758/bf03203267

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464–491. doi:10.1093/applin/aml001

Fletcher, P. (2014). Data and beyond. *Journal of Child Language*, *41*, 18–25. doi:10.1017/S0305000914000191

Flor, M., Futagi, Y., Lopez, M., & Mulholland, M. (2015). Patterns of misspellings in L2 and L1 English: A view from the ETS Spelling Corpus. *Bergen Language and Linguistics Studies*, *6*, 107–132. doi:10.15845/bells.v6i0.811

Gardner, R., & Wagner, J. (2005). *Second language conversations*. New York: Continuum.

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013). *Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT)*. Paper presented at the Proceedings of the 31st Second Language Research Forum. Somerville, MA: Cascadilla Proceedings Project.

Gleason, J. (2013). Dilemmas of blended language learning: Learner and teacher experiences. *CALICO Journal*, *30*, 323–341. doi:10.11139/cj.30.3.323-341

Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2014). *Video research in the learning sciences*. Mahwah, NJ: Erlbaum.

Gor, K., & Cook, S. (2010). Nonnative processing of verbal morphology: In search of regularity. *Language Learning*, *60*, 88–126.

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.

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning & Technology*, *9*, 90–110.

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 897–964). New York: Routledge.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*, 757–798.

Kowalski, J., Gordon, G., & MacWhinney, B. (2014). Statistical modeling of student performance to improve Chinese dictation skills with an intelligent tutor. *Journal of Educational Data Mining*, *6*, 3–27.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar: Vol. 1: Theory*. Stanford, CA: Stanford University Press.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaires: A web-based interface for bilingual research. *Behavior Research Methods*, *38*, 202–210. doi:10.3758/BF03192770

Long, M. (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 Press.

Lüdeling, A., Walter, M., Kroymann, E., & Adolphs, P. (2005). Multi-level error annotation in learner corpora. *Proceedings of Corpus Linguistics*, *2005*, 15–17.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, *43*, 1–123.

MacWhinney, B. (1997). Implicit and explicit processes. *Studies in Second Language Acquisition*, *19*, 277–281.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.) Mahwah, NJ: Erlbaum.

MacWhinney, B. (2015a). Introduction: Language Emergence. In B. MacWhinney & W. O'Grady (Eds.), *Handbook of Language Emergence* (pp. 1–32). New York: Wiley.

MacWhinney, B. (2015b). Multidimensional SLA. In S. Eskilde & T. Cadierno (Eds.), *Usage-based perspectives on second language learning* (pp. 22–45). New York: Oxford University Press.

MacWhinney, B. (in press). Entrenchment in second language learning. In H.-J. Schmid (Ed.), *Entrenchment*. New York: American Psychological Association.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2015). *Research using AphasiaBank*. Paper presented at the Academy of Aphasia, Tucson, AZ.

Malvern, D., Richards, B., Chipere, N., & Purán, P. (2004). *Lexical diversity and language development*. New York: Palgrave Macmillan.

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1–21). New York: Routledge.

Maslen, R. J., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of*

Speech, Language, and Hearing Research, *47*, 1319–1333. doi:10.1044/1092-4388(2004/099)
Meurers, D. (2012). Natural language processing and language learning. *The Encyclopedia of Applied Linguistics*. doi:10.1002/9781405198431.wbeal0858
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, *63*, 81–97.
Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, *4*(11), e7785. doi:10.1371/journal.pone.0007785
Murday, K., Ushida, E., & Chenoweth, N. A. (2008). Learners and teachers perspectives on languageonline. *Computer Assisted Language Learning*, *21*, 125–142. doi:10.1080/09588220801943718
Negueruela, E., & Lantolf, J. P. (2006). Concept-based instruction and the acquisition of L2 Spanish. In R. Salaberry, A. Barbara, & A. Lafford (Eds.),*The art of teaching Spanish: Second language acquisition from research to praxis* (pp. 79–102). Washington, DC: Georgetown University Press.
Ninio, A., & Wheeler, P. (1986). A manual for classifying verbal communicative acts in mother-infant interaction. *Transcript Analysis*, *3*, 1–83.
Pavlik, P., Presson, N., Dozzi, G., Wu, S., MacWhinney, B., & 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.
Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, *51*, 73–75.
Presson, N., Davy, C., & MacWhinney, B. (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). Amsterdam: John Benjamins.
Presson, N., MacWhinney, B., & Heilman, M. (2010). *A simulated environment for the embodied learning of Spanish prepositions*. Paper presented at the Second Language Research Forum, University of Maryland.
Presson, N., MacWhinney, B., & Tokowicz, N. (2014). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*, *35*, 709–737. doi:10.1017/S0142716412000550
Presson, N., Sagarra, N., MacWhinney, B., & Kowalski, J. (2013). Compositional production in Spanish second language conjugation. *Bilingualism: Language and Cognition*, *16*, 808–828. doi:10.1017/S136672891200065X
Reznicek, M., Lüdeling, A., & Hirschmann, H. (2013). Competing target hypotheses in the Falko corpus. In. A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 101–123). Amsterdam: John Benjamins.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rose, Y. (2010). The PhonBank initiative and second language phonological development: Innovative tools for research and data sharing. In A. Henderson (Ed.), *English pronunciation: Issues and practices (EPIP): Proceedings of the First International Conference* (pp. 223–241). Chambéry, France: Université de Savoie.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language*, *37*, 705–729.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics*, *11*, 1–22. doi:10.1017/S0142716400008262

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 173–190). Amsterdam: John Benjamins.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*, 1–14.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. *Sociocultural Theory and Second Language Learning*, *97*, 114.

Thorne, S. L., Black, R. W., & Sykes, J. M. (2009). Second language use, socialization, and learning in Internet interest communities and online gaming. *The Modern Language Journal*, *93*(S1), 802–821.

Van den Branden, K. (2010). *Task-based language education*: Cambridge, UK: Cambridge University Press.

van Geert, P., & Verspoor, M. (2015). Dynamic systems and language development. In B. MacWhinney & W. O'Grady (Eds.), *Handbook of language emergence* (pp. 537–556). New York: Wiley.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*, 191–210.

VanPatten, B. (2011). Input processing. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 268–281). New York: Routledge.

Verspoor, M., de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language development*. New York: John Benjamins.

Vihman, M. (2015). Perception and production in phonological development. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 437–457). New York: Wiley.

Wagner, J. (2015). Language learning in the wild. In S. Eskilde & T. Cadierno (Eds.), *Usage-based perspectives on second language learning* (pp. 46–67). New York: Oxford University Press.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—third edition*. San Antonio, TX: Psychology Corporation.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*, 13–40.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*, 65–86.

Zhao, H. (2012). Explicitness, cue competition, and knowledge tracing: A tutorial system for second language learning of English article usage. Doctoral disseration, Carnegie Mellon University, Pittsburgh, PA.

Zhao, H., Koedinger, K. R., & Kowalski, J. (2014). *Knowledge tracing and cue contrast: Second language English grammar instruction*. Paper presented at the Cognitive Science, Berlin, Germany.

Zipser, F., & Romary, L. (2010). *A model oriented approach to the mapping of annotation formats using standards*. Paper presented at the Workshop on Language Resource and Language Technology Standards, LREC 2010.