

To appear 2024 in: Yuichi Suzuki (ed.): *Practice and automatization in second language research: Perspectives from Skill Acquisition Theory and Cognitive Psychology*, Second Language Acquisition Series, Routledge.

Chapter 6. Individualization of practice through Intelligent CALL

Simón Ruiz, Patrick Rebuschat, Detmar Meurers

Abstract

Intelligent computer-assisted language learning (ICALL) uses methods from Artificial Intelligence to explicitly model learners and the target language, to analyze the language that learners produce and learning materials. In this chapter, we explore and exemplify the potential of ICALL for second language (L2) practice by spelling out how these tools can facilitate individualization of practice. They can support *input practice* through the selection and enhancement of materials that are rich in the pedagogically targeted, developmentally proximal language for a given learner, and it can adaptively provide *output practice* activities and meta-linguistic feedback to promote L2 declarative knowledge and its proceduralization. Complementing the discussion of how ICALL can promote L2 acquisition, we highlight the opportunity for large-scale collection of learning process and product data in individualized interventions delivered digitally using ICALL systems, a unique option for investigating instructed L2 acquisition capable of zooming in to aptitude-treatment interactions.

1. Introduction

In second language (L2) learning and teaching, practice can be understood as an array of activities that involve the use of the L2 in a deliberate, systematic, and repeated manner, with the aim of developing knowledge and skill in learners (DeKeyser, 2007; Suzuki, this volume). One way to facilitate L2 practice is through the use of technology (Brett & González-Lloret, 2011). The field of computer-assisted language learning (CALL) studies how digital technology can facilitate the language learning process (Chen et al., 2021). Intelligent CALL (ICALL) adds methods from Artificial Intelligence (AI) to explicitly model learners and the target language, to analyze the language that learners produce as well as learning materials (Meurers, 2021). This not only makes it possible to systematically support adaptive L2 practice in digitally enhanced learning environments, but also provides L2 researchers with an experimental platform to undertake research on language learning at scale (i.e., involving many learners) in authentic contexts (Meurers & Dickinson, 2017; Meurers et al., 2019).

The aim of this chapter is to contribute to a better understanding of how ICALL can promote L2 acquisition, and how it can also serve as an effective research tool to investigate instructed L2 acquisition (e.g., Petersen, 2010; Chen, 2018; Meurers et al.,

2019; Ruiz et al., 2021; Ziegler et al., 2017). Given the theme of this volume, we explore and exemplify the potential of ICALL systems for L2 practice by spelling out how these tools can support individualization of practice, i.e., sequencing of practice activities and feedback for a given learner (DeKeyser, 2010) to promote the acquisition of L2 declarative knowledge (lexical and syntactic information) and then its subsequent proceduralization and automatization (DeKeyser, 2017).

We illustrate how ICALL systems enable second language acquisition (SLA) researchers to obtain ecologically-valid insights into the interaction of a substantial number of parameters that research on aptitude-treatment interactions (ATIs) has identified as important for learning (Cronbach & Snow, 1977; DeKeyser, 2007, 2021). We argue that ICALL systems provide a unique opportunity for SLA research given that they support large-scale studies of learning in authentic education contexts and the use of learning analytics to explore and conduct fine-grained analyses of learning processes and outcomes. However, we also emphasize that this connection between ICALL and SLA has yet to unfold its full potential and call for a closer collaboration between computational linguistics and applied linguistics, between software developers interested in educational technology, and L2 researchers.

The chapter is organized as follows. In the next section, we provide an overview of the affordances and potential uses of ICALL, and we illustrate the spectrum with concrete systems. We then focus on the nexus between ICALL and L2 practice by zooming in on how ICALL can implement the principles of effective L2 practice (Suzuki, this volume). Finally, we consider some of the limitations and challenges associated with the use of ICALL systems to support SLA research and outline steps needed to enable more effective collaboration to realize the full potential.

2. Linking ICALL to SLA and the principles of effective L2 practice

2.1. Components of ICALL

ICALL integrates AI methods into CALL, making it possible to explicitly model the language to be learned and the learner engaged in the learning process (Meurers, 2021). Most prominently, ICALL integrates the use of natural language processing (NLP) methods. NLP is a subarea of AI that allows ICALL applications to automatically process the language used by learners and the learning materials. Listing all the different ways that language can be used to express a given meaning and all the possible ways that language learners may attempt to do so results in thousands of alternatives that would need to be considered even for single sentences (cf. Nagata, 2009, for an example). The systematic well-formed and ill-formed variability of learner language therefore requires

dedicated mechanisms for analysis (Meurers & Dickinson, 2017).

NLP provides methods to analyze the potentially infinite set of sentences that could be uttered. Traditional NLP techniques include tokenizing (i.e., splitting text into meaningful units), part-of-speech tagging (i.e., identifying the parts of speech of words), and syntactic parsing (i.e., determining the grammatical structure of sentences). While the linguistic characterization of the forms and structures used by learners helps identify relevant learner language characteristics (Meurers, 2015), to also identify properties of ill-formed parts, additional mechanisms such as *mal*-rules or limitation relaxation can be added (Heift & Schulze, 2007; Meurers, 2021). In sum, NLP can be seen as providing *linguistic intelligence* for ICALL, which is important for a range of uses, from characterizing the linguistic complexity of learner input to analyzing learner output to providing immediate feedback (Petersen & Sachs, 2015; Meurers, 2021).

For ICALL applications that are designed to play the role of an intelligent teaching system, further AI components must be added to model the learner and adapt the instruction. Five decades ago, Hartley and Sleeman (1973) spelled out the components required for Intelligent Teaching/Tutoring Systems (ITS): (i) a model of the domain to be taught (expert/domain model), (ii) a model of who the learner is and their learning progress (learner/student model), (iii) a set of "teaching operations" (including explicit activity models), and (iv) a method for selecting the best next teaching operation for a given learner and goal (didactic/pedagogy model). For language learning it is difficult to comprehensively spell out the domain to be taught in a formal way. Correspondingly, language learning is considered an "ill-defined domain" (Fournier-Viger et al., 2010) and has received very little attention in ITS research, with no systems for foreign language learning being included in the extensive meta-analysis of ITS effectiveness by Kulik and Fletcher (2016). Yet, as motivated in Meurers et al. (2019), it is increasingly feasible to develop ITS for foreign language teaching, and indeed several systems have successfully been used in real-life contexts, such as the E-Tutor (Heift, 2010), Robo-Sensei (Nagata, 2009), TAGARELA (Amaral & Meurers, 2011), and the i-tutor (Choi, 2016) in university contexts, and the FeedBook (Meurers et al., 2019) in secondary schools.

While the specification of the domain and activity models still requires substantial manual effort by foreign language teaching experts, endeavors such as the English Grammar Profile project (EGP, <https://englishprofile.org/english-grammar-profile>), which spell out the functional can-do statements of the Common European Framework of Reference (Council of Europe, 2020) and link them to the detailed inventory of lexical and grammatical language means (i.e., words and structures) used to realize those functions, provide valuable guidelines for the development of curricula and

ICALL systems. Such explicit specifications of the language means being acquired can be combined with NLP methods analyzing learning materials and activities to automatically generate the activity models needed for ICALL systems (Quixal et al., 2021). There is also substantial research on the learner model component for L2 learning (Schulze, 2012), which can track the individual interaction with the learning activities and system. This makes it possible to draw inferences about the learner's knowledge and beliefs about the language means being learned (Heift & Schulze, 2007). To support valid inferences going beyond linguistic domain knowledge as such, learner models have also been extended to record information on potential learning strategies and analogies used (Bull et al., 1995) and on task appropriateness, task strategies, and L1 transfer observed in the learner interactions with the system (Amaral & Meurers, 2008). By interpreting learner models in relation to both the domain and activity models, the didactic model enables the ITS to adaptively select activities for a given learner (macro-adaptivity) and to provide immediate scaffolding feedback (micro-adaptivity) with the goal of optimally supporting an individual learning trajectory.

2.2. Relating ICALL to SLA concepts

When considering the components of ICALL systems and how they facilitate the analysis of learner input and output, support the (macro-adaptive) selection of the next learning activity and provide (micro-adaptive) feedback, we can establish direct links to core SLA concepts (see Leow, 2019, for a recent review).

The automatic analysis of learner input transparently links to the widely agreed assumption that exposure to the L2 is essential for learning to take place. As pointed out by Ellis (2017), supporting **input practice** is important, especially for beginner-level learners. In this sense, Ellis highlights that input practice also plays an important role in frameworks such as Task-Based Language Teaching (Jackson, 2022), which is more typically associated with production tasks. In section 3.1, we discuss several ICALL systems that support **input enrichment for comprehension practice**. Search engines such as FLAIR and KANSAS allow teachers to easily identify texts that contain language structures that are relevant to their classes, thus supporting real-life teaching and learning. Further, the more experimental SyB system introduced in section 3.2 aims at establishing a direct adaptive link to identify developmentally proximal input by relating the linguistic complexity of the learner's own writing and the complexity of the input. This makes it possible to empirically explore different ways to operationalize what constitutes developmentally proximal input that optimally fosters acquisition, a challenge that has eluded SLA research since Krashen's (1985) proposal that learners require

comprehensible input at a level beyond their current ability (“i+1”) (see Lichtman & VanPatten, 2021, for recent discussion). Apart from lexical acquisition research (Schmitt et al., 2011), what constitutes developmentally proximal only seems to be explicitly established for a few language means considered in research on developmental sequences (Pienemann & Johnston, 1987) and teachability (Pienemann, 1989), though for English the work on criterial features of different proficiency levels provides some breadth (Hawkins & Filipović, 2012).

Even input that is rich in developmentally proximal language generally is not sufficient for L2 acquisition to go beyond Basic Varieties (Klein & Perdue, 1997). Learners must *notice* in the input the aspect of language that is to become intake for learning (Schmidt, 1990). To provide the necessary Focus-on-Form (Long, 1991), one option is to increase the salience through input enhancement (Sharwood Smith, 1993). The WERTi and the VIEW systems discussed in section 3.3 automate the provision of **input enhancement**, from coloring for input practice in a reading task to interactive exercises practicing forms in the reading text context. Given the automated nature of the enhancement, both what the learner is reading as well as the nature of the language aspect being enhanced, can be individually selected given the learner's interests and needs in terms of language development.

Complementary to input and noticing, it is important for L2 learners to produce language (e.g., DeKeyser, 2007; Suzuki, 2022; Swain, 2005). On the one hand, output requires the learner to commit to the language forms needed to realize the intended function, whereas for processing input one can rely on good-enough processing, i.e., it is sufficient to pick up enough of the functionally relevant information that is typically encoded redundantly in the input. Indeed, the production of language also plays a central role for practice in Skill Acquisition Theory (DeKeyser, 2017), given the need to turn declarative knowledge about language into automated, fluent behavior. This, in turn, requires paying attention in order to focus on what is to be said rather than on how to use language to do it. The complexity of the activities selected for practice and the feedback provided during practice should be determined based on the individual learner characteristics (DeKeyser, 2017). The inclusion of feedback, as one of the most effective factors in real-life education (Hattie & Timperley, 2007) requires immediate analysis of the learner responses (cf. Fu & Li, 2022), which correspondingly has been a major focus of ICALL research developing ITS (Heift & Schulze, 2007). When ITS take on the role of selecting activities of the appropriate complexity for a given learner (macro adaptivity) and providing feedback to scaffold the learning process (micro adaptivity), the social dimension of such tutor-supported practice also becomes readily apparent: adaptive

tutoring systems support scaffolded practice (i.e., instructional guidance during practice) of developmentally proximal language means in activities in the Zone of Proximal Development (ZPD; Vygotsky, 1978) of a given, individual learner. ZPD is understood here as the distance between what learners can do on their own in the L2 and what they can do with guidance from an ICALL system. In section 3.4, we make this concrete by discussing the FeedBook, an ITS for learners of English offering a range of practice activities that provide immediate scaffolded feedback on meaning and form (Meurers et al., 2019). By *immediate scaffolded feedback* we here mean that the FeedBook provides feedback on learner answers in the form of *explanations* that are given *during* activity completion (see Figure 5, for an instance).

In the following, we explore each of the mentioned ICALL-SLA links and the different practice opportunities they provide. We characterize each system's focus on comprehension, production, or both, the nature of the language domain that is targeted, and aspects of the empirical validation of the realized practice approach. We then zoom in on the characteristics of the practice offered by these systems and how they relate to the principles of effective L2 practice. In section 4, we zoom back out to address automatization and ICALL by discussing limitations related to the use of ICALL systems for SLA research and by formulating an agenda for reaching the true potential of ICALL systems for SLA research.

3. Principles of effective L2 practice

Before delving into the details of how ICALL systems can support the effectiveness of L2 practice, a brief reminder of the qualities of effective L2 practice (Suzuki, this volume) is necessary. For practice to be effective, it needs to be *deliberate*, meaning that practice should provide opportunities for repetition where learners are asked to consciously commit language information to memory in order to later proceduralize and automatize it. To that end, learners need to have an internally or externally motivated goal and need to focus their attention on the language they are learning as well. Practice should also be *systematic*, that is, learners should engage in planned (not random) repetitions of language use. This, in turn, should enhance the retrieval of linguistic information that will become proceduralized. Effective practice should also provide error feedback. Feedback is a pillar of learning, as it modulates how quickly (and successfully) we learn (Dehaene, 2020). In L2 learning the positive effect of feedback has also been widely recognized (cf. Nassaji & Kartchava, 2021). Moreover, effective practice should also promote the *transfer* of learning. This suggests that practice should be carried out in instructional contexts that would allow for learners to *transform* what they have been taught (i.e., actively adjusting

and adapting prior knowledge to new situations or goals) into effective real-life L2 use (Larsen-Freeman, 2013). Finally, effective practice should also be challenging. More specifically, practice should present learners with instructional content at different levels of learning difficulty to optimize long-term retention. In this sense, learning difficulty needs to be *desirable* in that it needs to be given in the right amount: if practice is too easy for students, it will demotivate them; if it is too difficult (*undesirable* difficulty), it will hinder further processing of information in procedural memory and thus reduce the opportunities for learning (Suzuki et al., 2019). The degree of difficulty will depend upon practice condition (e.g., the type of feedback learners receive), linguistic factors (e.g., language complexity of learning materials), and the learners' individual differences (e.g., cognitive abilities such as working memory) (DeKeyser, 2017). In sum, effective L2 practice should be *deliberate* (intentional, goal-directed), *systematic* (planned, not random), sufficiently *challenging* (not too easy and not too difficult), and should support *transfer-appropriate* processing (i.e., practice should ensure a close match between learning conditions and future use of the L2).

3.1. Input enrichment: Facilitating encounters with pedagogically targeted language means

Supporting learners to select texts that are of interest and rich in the pedagogically targeted, teacher-selected language means is a goal that ICALL applications can help realize in practice. For instance, the FLAIR (Form-Focused Linguistically Aware Information Retrieval; <http://flair.schule>) system is a search engine that can help language teachers find texts on the web that contain a high number of instances of user-selected linguistic forms (Chinkina & Meurers 2016). FLAIR combines a standard web search for a particular content with a reranking step prioritizing those search results that are rich in the specified linguistic forms and at a selected readability level (i.e., quantitative linguistic complexity; Carrell & Wise, 1998, p. 290; Chen & Meurers, 2019). It supports such input enrichment for all language means mentioned in the German secondary school curriculum for English. FLAIR also integrates result previews that are visually enhanced to support quick identification of the occurrences of the targeted language forms. For example, Figure 1 shows the interface after a teacher selected phrasal verbs as targeted language forms and searched for "climate change" on the bbc.com site, asking for 50 results.

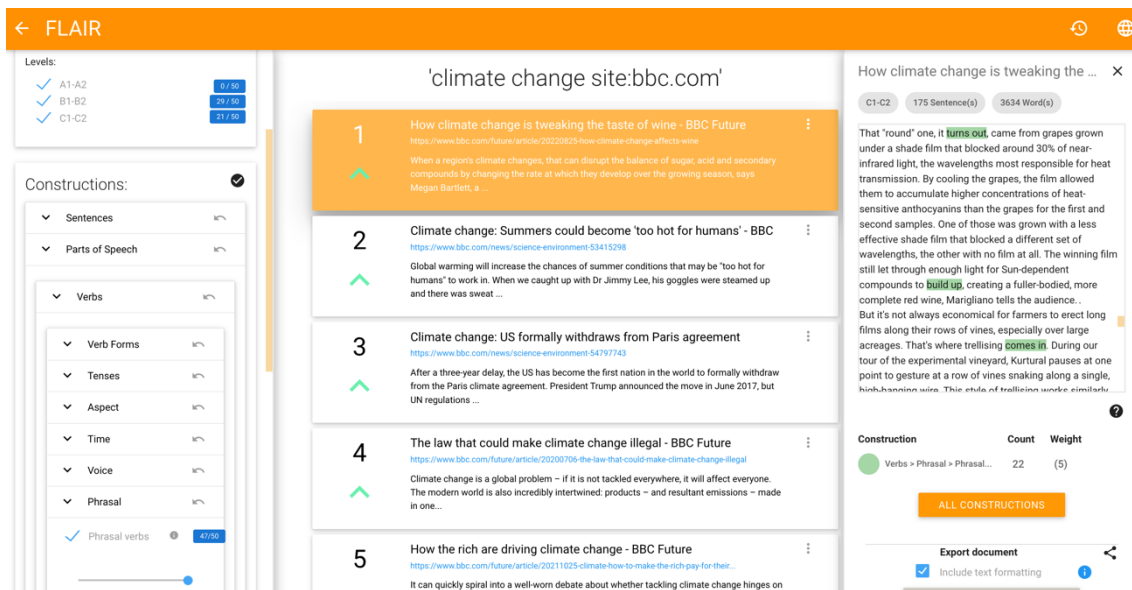


Figure 1. The FLAIR interface: settings panel, list of search results, and result preview.

On the left, we see that 29 of the web search results were at the B1-B2 level and 21 at the C1-C2 level. The top result was a long text (175 sentences) containing 22 phrasal verb occurrences, for which the text preview on the right shows three visually highlighted ones. The specification of the desired readability level and the targeted language means typically is taken care of by a teacher, who then provide a link to FLAIR containing their settings. The students can use the system autonomously to search for any contents they are interested in, with the system ensuring that the teacher specifications of the language level and forms are taken into account in the ranking and presentation of the results.

To explore whether reranking of search results can successfully ensure richer representations of pedagogically relevant language means, Chinkina & Meurers (2016) inspected the distribution of potential pedagogical targets in the search results. For example, Figure 2 shows the top 55 web search results FLAIR obtained for the query “2016 US presidential elections”.

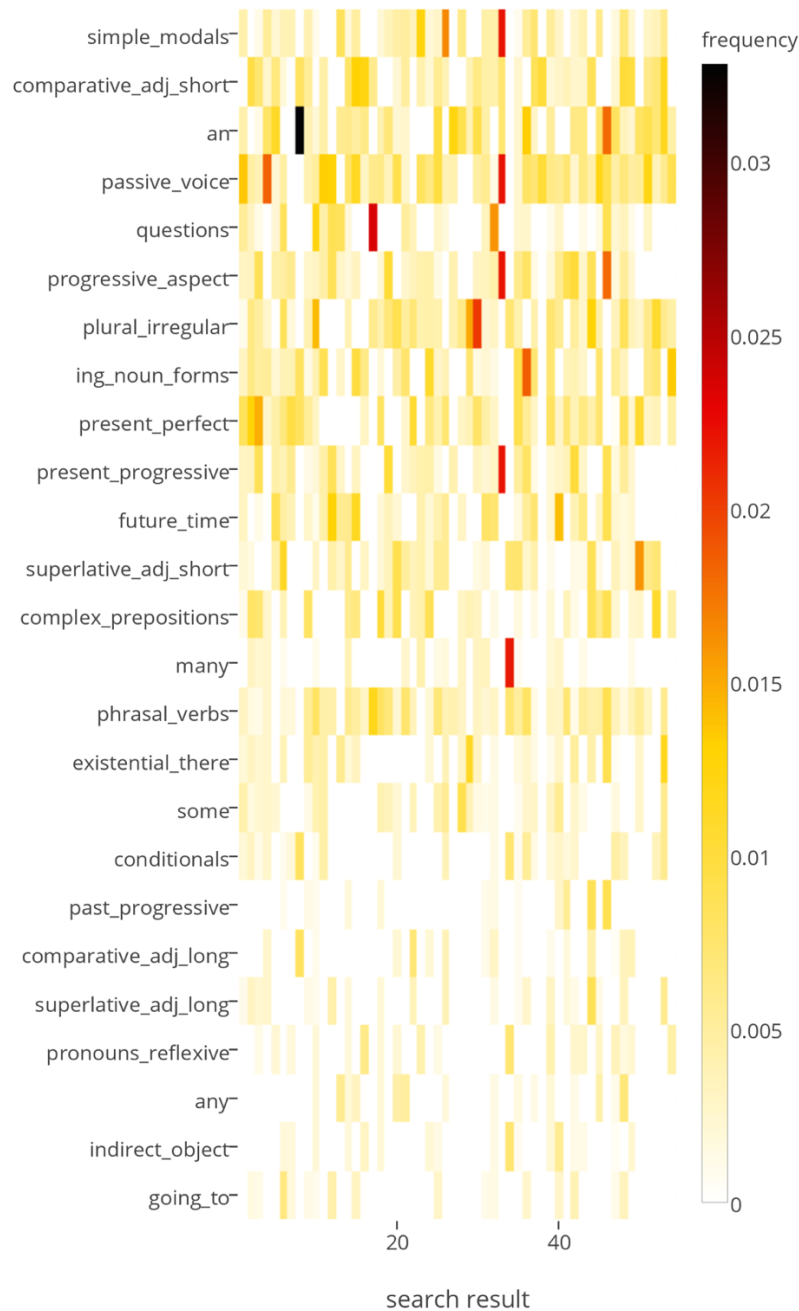


Figure 2. Heatmap showing the distribution of grammatical constructions (y-axis) across the 55 top FLAIR search results (x-axis) for the query “2016 US presidential elections”.

Figure 2 shows that the 55 top search results shown on the x-axis are highly variable regarding how many instances of the language means on the y-axis are found in the given search result. Reranking the results to prioritize specific language means thus is a meaningful approach to ensure the learner is exposed to texts that are rich in the targeted language means.

Another input enrichment ICALL system, KANSAS (<https://kansas-suche.de>) supports the input enriched retrieval of texts for low literacy learners (Weiss et al., 2018). Since texts at low levels of complexity are relatively rare on the web, the system offers both a general web search and a web search focused on a collection of web sites that are dedicated to providing reading material for low literacy readers. Conceptually more interesting is that KANSAS also supports searches in a curated set of reading materials compiled from a range of sources, including crawled web material. Supporting searches in such a curated text base that can be linguistically analyzed offline, i.e., before the user enters a query, makes it possible to systematically and comprehensively rank all texts that contain the pedagogically relevant language target instead of only reranking the small set of top results returned by a standard search engine for the user-specified content query. This also facilitates identifying texts with rare (combinations of) language characteristics, to visualize the frequency distribution of each language target, and to offer filtering as well as (de)prioritizing of language means across the full corpus.

Relating the approach to the principles of effective L2 practice

Both the FLAIR system and the KANSAS system can be used to assist language teachers in carrying out extensive reading as an input practice activity. In extensive reading, learners read large amounts of reading materials at their proficiency level (cf. Grabe & Stoller, 2011). By implementing extensive reading through these ICALL systems, learners can not only self-select reading materials containing the language characteristics pre-selected by the teacher, but they also reap the benefits of doing extensive reading, such as developing themselves as independent readers and reading for meaning, information, and pleasure (Suk, 2017). The skills and knowledge developed through extensive reading supported by these ICALL systems thus should transfer to real-world reading contexts, thereby instantiating transfer-appropriate practice (Suzuki, this volume). To substantiate whether the self-selected, authentic nature of the input enriched materials indeed improves learning outcomes in a way that transfers, systematic field studies will need to be conducted with these freely usable ICALL systems.

3.2 Input enrichment: Providing developmentally proximal input

The SyB application (Syntactic Benchmark, <http://complexity.schule>) is an ICALL system that links input and output-based L2 instruction by means of automated linguistic complexity analysis of both learner input and learner production (Chen & Meurers, 2017; Chen, 2018). More specifically, the SyB system uses a pedagogic target language corpus of leveled reading material as a benchmark corpus and NLP to automatically analyze a set of syntactic complexity features (for technical details, see Chen & Meurers, 2017; 2019). When a language learner using the system enters a short text they wrote into the interface, the SyB system analyzes the learner text in terms of the same set of complexity features. The system then provides the user with a selection of reading materials that for a user-selected complexity feature (or all features combined) are just above the complexity of their own writing. The user can manually select a “challenge level” to select how “just above” is spelled out. This is illustrated in Figure 3.

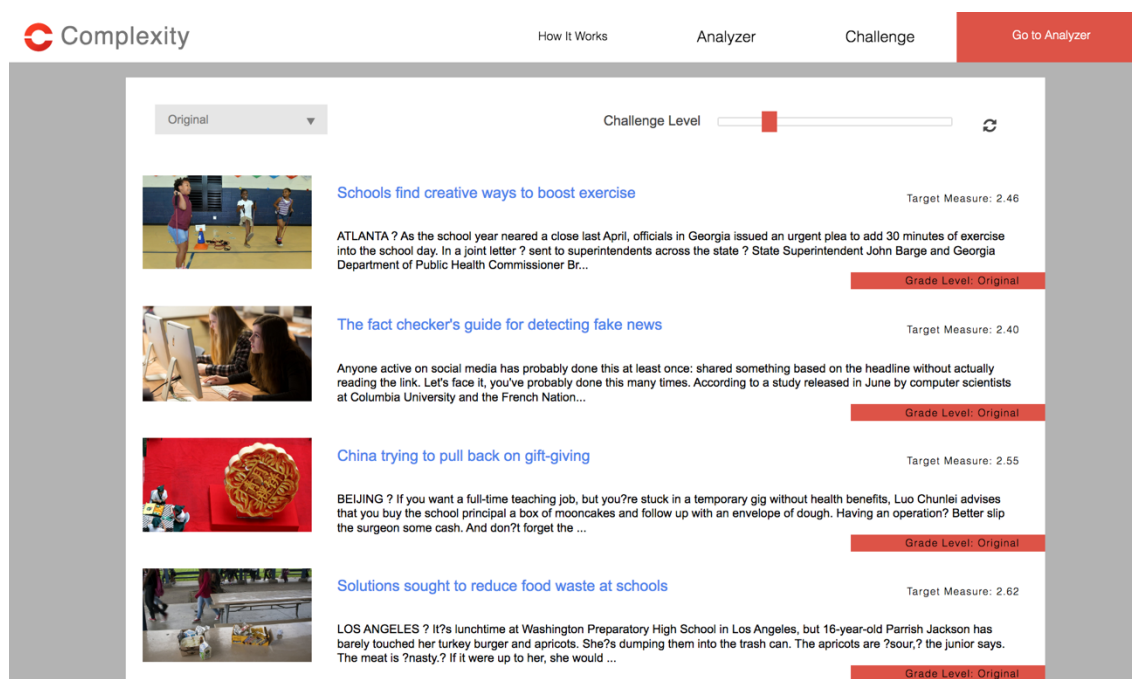


Figure 3. The Challenge Window in the SyB system. Learners are presented with L2 texts based on the syntactic complexity of their language production.

In sum, the ICALL system assigns individually-adaptive reading input based on learners' language production, attempting to match where the learners are in terms of their individual development to texts that are developmentally proximal. SyB thus aims to offer adaptive, individualized L2 reading practice based on the learner's interlanguage characteristics.

Chen et al. (2022) evaluated the effectiveness of individualized $i + 1$ input, where i is operationalized in terms of the syntactic complexity of their language production and the $+ 1$ derived from the selected challenge level. The study examined whether texts selected based on the complexity of the learner's L2 production promote proficiency development, including experimental conditions differing in the challenge level. To directly relate the reading and the writing components, a continuation writing task was chosen, in which the learner continues writing the partial text given as reading input. The study showed that learners align the linguistic complexity of their writing with that of the reading materials, which confirms the potential value of this type of ICALL-supported input complexity challenging practice for L2 learning. However, longer studies are needed to establish whether the alignment in complexity amounts to a lasting learning effect or only a local alignment that does not actually incrementally develop the baseline of the learner.

Relating the approach to principles of effective L2 practice

Concerning the principles of effective L2 practice, an ICALL system like SyB provides the opportunity to experimentally explore the principle of desirable difficulty. Learners can systematically be challenged to process target language that is more complex than what they can produce, and it was shown that this supports complexity alignment – though more evidence is needed to determine whether this translates to language development. The SyB approach also instantiates individualization of L2 practice as learners are pushed to understand and produce L2 language relative to their individual learner proficiency. The SyB system therefore aims to optimize learning by providing appropriate instruction that dynamically adapts to learners' individual differences (DeKeyser, 2021).

3.3. Input enhancement: Fostering noticing of pedagogical target forms

As an ICALL tool focused on input instruction, WERTi (Working with English Real Texts interactively; <http://purl.org/icall/werti>) provides automatic input enhancement of authentic texts (Meurers et al., 2010). The system employs NLP techniques such as syntactic parsing to identify and visually enhance the pedagogically targeted language means in webpages. The WERTi prototype was reimplemented as the web extension Visual Input Enhancement of the Web (VIEW; <http://purl.org/view>), which supports different web browsers, languages, and target patterns (Ziegler et al., 2017). WERTi offers different types of L2 instruction in an authentic, real-life web-based setting. It supports the automatic coloring of targeted language forms in online texts (receptive practice) as well as learner interaction with the text containing these forms through automatically generated clicking, multiple-choice, or fill-in-the-blank activities (dynamic

practice). The WERTi system can also provide a log file for each learner, keeping a detailed record of learner interactions with the system, such as when and how an activity was completed and the number of activities tackled. On this basis, the approach has been empirically validated in a pilot and a more extensive study. The pilot study examined whether automatic input enhancement of news texts supports the acquisition of implicit and explicit knowledge of English articles (Ziegler et al., 2017). The full study (Ruiz Hernández, 2019) employed an extended version of WERTi that not only housed a pedagogical intervention targeting English phrasal verbs, but also integrated newly-developed web-based individual difference tests (Ruiz et al., 2019; 2021).

Relating the approach to principles of effective L2 practice

The WERTi system instantiates three principles of effective L2 practice. Firstly, the WERTi system fosters deliberate practice, as the clicking, multiple-choice, or fill-in-the-blank activities supported by the system requires learners to make a conscious, *deliberate* effort to complete them, which may contribute to the proceduralization of the targeted language. For instance, in the multiple-choice gap-filling activity, learners are required to consciously interact with the online text by clicking on gaps in the text (see Figure 4).



Figure 4. Automatic enhancement in the extended version of WERTi (Ruiz et al., 2021). Learners are asked to fill out the gaps in the text by clicking on the dropdown menus.

In this activity, different options are displayed once learners click on the gap. Learners then have to select the option that correctly completes the gap – for example, in Ruiz et

al. (2021), the correct phrasal verb option in the context of the sentence. The activity thus requires learners to process and (correctly) understand the targeted language form during comprehension in order to complete it (task-essentialness, Loschky & Bley-Vroman, 1993). Ruiz et al., (2021) found that the group who read texts with automatically generated gaps showed a stronger effect (i.e., higher gains) for both comprehension and production of L2 vocabulary than those students who read texts without gaps. Secondly, the WERTi system also provides automatic, *immediate feedback* to learners, which has been indicated as more effective in supporting learning than delayed feedback (Fu & Li, 2022). In the case of the multiple-choice activity, for instance, learners receive color-coded feedback: if the answer is correct, it turns green; if the answer is incorrect, it is highlighted in red. As additional support, a smiley face is displayed next to each blank space that inserts the correct answer into the gap when it is clicked on. Learners are instructed to use the smiley face only as a last-resort strategy. Finally, the WERTi system can also promote *transfer* of learning (Larsen-Freeman, 2013) given that the system offers ecologically-valid practice activities to learners reading news articles on the web, outside of the school context. More research is necessary to empirically substantiate how effective ICALL systems such as WERTi are in supporting the transfer of skills and knowledge into real-life L2 use.

Another way to enhance input to support L2 practice with ICALL applications that would warrant more exploration and an effectiveness study is the use of automatically-generated questions to create a functional need to process targeted linguistic features, i.e., **functionally-driven input enhancement** (Chinkina & Meurers, 2017). NLP methods can automatically generate questions for any user-selected reading material with the primary goal of requiring the learner to cognitively process the targeted language means. When combined with an input enrichment system like FLAIR, one can ensure that the user-selected texts for which questions are to be generated are sufficiently rich in the targeted forms. The automatic generation of questions then is a way to provide a minimal activity structure that ensures input is read and practice can be monitored. In terms of the quality that can be achieved with current NLP approaches, such automatically-generated questions have been found to be indistinguishable from human-created questions (Chinkina et al., 2020).

3.4. Feedback: Scaffolding learners completing a range of closed to open activities

The FeedBook system (Rudzewitz et al., 2017; <http://feedbook.website/en>) is a web-based ICALL system offering input and output-based instruction. This intelligent tutoring system is a digital workbook that was developed to support English language learning

and teaching in secondary schools in Germany. While the original FeedBook was based on an established, printed workbook, the NLP also supports the generation of exercises (Heck & Meurers, 2022). Crucially, the NLP also allows the FeedBook to automatically provide immediate scaffolded linguistic feedback. This feedback is aimed at supporting individual learners when completing a range of form and meaning activities (e.g., short answer, fill-in-the-blank, reading/listening activities), covering all language means on the seventh-grade English curriculum of German secondary schools. To that end, the system compares student answers to the space of well-formed and ill-formed variants automatically determined for the target answers (Rudzewitz et al., 2018) to parametrize and generate one of 188 meta-linguistic feedback templates. Figure 5 shows the feedback provided in an exercise targeting comparative forms, while Figure 6 illustrates the provision of focus on meaning feedback in a reading comprehension exercises, followed by incidental focus on form on the tense of the verb in the student's revised response. Students can also inspect their open learner model and teachers can use the FeedBook to monitor student progress. The effectiveness of the feedback provided by the FeedBook was empirically evaluated in a randomized controlled field trial (Meurers et al., 2019), the gold-standard empirical study design in medical and education research. The study was carried out during a full school year in 14 seventh-grade classes in four German high schools, with the FeedBook being used as digital replacement for the printed workbook, without training or changes to the usual teaching. The digitally administered intervention allowed researchers to collect substantial, ecologically-valid longitudinal data. Every student used the system, but within-class randomization was used to distinguish which student received the scaffolded feedback for which of the language means on the curriculum. The learners who received the scaffolded feedback outperformed those who did not, with a medium size effect ($d = 0.56$). The findings add to the evidence of the effectiveness of practice with incremental, scaffolded feedback that is provided while learners are working on exercises at the appropriate level for them.

B 1 Off to Greece again
 Mr Lambraki is checking flights to Greece. Read the information he has found on the two airlines and use the adjectives below to compare them.
 • LIFBR: Comparison of adjectives

expensive (ticket) · early (departure) · attractive (shopping on board) · good (choice of food offered on board) · healthy (food and drinks) · suitable (airport) · cheap (tickets for shuttle bus) · friendly (service on board) · easy (online booking)

Midair	Air-Con
<ul style="list-style-type: none"> London – Athens from 39 pounds departure 7.00 am non-stop small choice of duty free articles for shopping on board low-calorie and vegetarian food available¹ from Gatwick only 28 miles from London tickets for shuttle bus are 10 euros 	<ul style="list-style-type: none"> London – Athens from 57 pounds departure 12.15 pm via Berlin all international brands for shopping on board snacks: crisps and chocolate bars from Stansted only 40 miles from London tickets for shuttle bus are 10 euros
<p>customers' comments:</p> <ul style="list-style-type: none"> service mostly friendly but a little bit stressed² online booking quick and well-organized 	<p>customers' comments:</p> <ul style="list-style-type: none"> the service staff³ on board were extremely friendly and very helpful online booking not very clear and consequently takes ages

¹ available – erhältlich ² stressed – gestresst

1. The tickets at Air-Con *expensiver* than at Midair.

2.

Feedback für "The tickets at Air-Con *expensiver* L..."

When an adjective has three or more syllables, we form the comparative with 'more' and the superlative with 'most'.

Hilfreich?
 Ja Nein

Figure 5. Feedback to a learner response in a focus-on-forms exercise in the FeedBook.

B 1 Gillian's diary
 Read Gillian's diary entry and complete these sentences.

Friday 23rd September

We'll drive to the north coast tomorrow and have a look at two boarding schools with Gwynn. His sister went to Wildings and he says it would be great for me. If they had a football team, it wouldn't be so bad, but it's all so girly-girly with horses and ballet dancing and everything, yuk! I don't want to go to the school in Llandysul either. It still feels like Gwynn and Mum just want me to go to boarding school because of the new baby. Miss my pals like crazy, miss London and my old school. If my friends were here, boarding school would actually be fun. Being the new girl at school without any friends will be horrible. ↯ And it's all Gwynn's fault! The village where we live now is the worst. 20 minutes to the nearest supermarket. If there was a shopping

centre, I could at least go shopping. But there's nothing, no shops, no cinema, no nothing ... only sheep!

My room is really nice and big though and we have a fab garden which is great for playing football. If I made some friends in the village, we could have a great time there. I met Gruffudd, the boy from next door, this afternoon. He seems nice. He started talking to me in Welsh and I couldn't understand ANYTHING. He then spoke English and told me he plays rugby. Well, it's not football but I might have to

1. Gwynn thinks Wildings School would be great for Gillian because *it is great*

Feedback für "it is great"

There seems to be important information missing in your answer. Please have a look at the highlighted passage in the text.

Hilfreich?
 Ja Nein

15 horrible. ↯ And it's all Gwynn's fault! The village where we live now is the worst. 20 minutes to the nearest supermarket. If there was a shopping

2. Gwynn thinks Wildings School would be great for Gillian because *his sister goes there*

Feedback für "his sister goes there"

We are talking about something that happened in the past. Please use the simple past, not the simple present.

Hilfreich?
 Ja Nein

Figure 6. Feedback on meaning (above left) followed by incidental focus on form (bottom right) incrementally provided to learner responses in a reading comprehension activity.

Relating the approach to principles of effective L2 practice

The FeedBook system transparently illustrates the critical principle of feedback for effective L2 practice. As an intelligent web-based workbook, the FeedBook system provides instantaneous, interactive formative feedback to learners, supporting the individual L2 practice with immediate feedback. It can provide automatic systematic feedback not only on form-oriented exercises, but also on meaning-based tasks, raising the learner's awareness of linguistic accuracy as well as promoting effective L2 practice (Suzuki, this volume). By modeling both the activity, the learner, and the language (Meurers et al. 2018), the FeedBook system can offer individual feedback that is appropriate given the learner's current L2 knowledge and skills in an authentic education environment (Meurers et al., 2019). The FeedBook also promotes deliberate practice, as the activities learners complete in the system are intended not only to complement the declarative knowledge provided in the classroom, but to assist proceduralization and further automatization of this knowledge through deliberate study, completing a variety of closed and more open activity types.

4. Insights into automatization: ICALL's limitations and a look ahead

We have seen how ICALL systems can promote L2 development by providing learners with opportunities for effective individualized input and output practice. This individualization of practice is possible in ICALL environments as these can model the core components of the language learning process – a rich representation of the learner, the target language, and the learning activity (Meurers & Dickinson, 2017). Given the ready availability and use of digital devices in authentic learning contexts, it becomes possible to monitor learning processes and products at scale (Alexopoulou et al., 2022). Where ICALL tools provide substantial benefits to learners in real life so that one can design and longitudinally study different interventions and the range of parameters at stake in ecologically valid contexts. This clearly holds great potential in the use of ICALL systems to advance SLA research, especially research on practice and automatization (Suzuki, 2022), but potentially also for the sociocognitive approaches to SLA (e.g., Ellis, 2010). Detailed system logs make it possible to track increased fluency in language use or to investigate the effect of different exposure conditions (from implicit to explicit, Norris & Ortega, 2000) on subsequent acquisition on a fine-grained level of analysis (e.g., Hui et al., in press; Ruiz et al., 2021; Ziegler et al., 2017). ICALL is also critical to achieve the important, yet elusive goal of adapting L2 instruction to cater to individual characteristics of learners. In most L2 instructed contexts, individualization of instruction may not be feasible due to logistical, financial, or social reasons (DeKeyser, 2021) – but

ICALL applications such as the ones discussed above can in principle support such adaptive individualized L2 practice considering individual learner differences.

However, despite an increasing interest in ICALL and Big Data in SLA (Meurers & Dickinson, 2017; Alexopoulou et al., 2022), the potential of research at the intersection of ICALL and SLA is only slowly starting to be realized. What are some of the challenges currently limiting the use of ICALL systems to systematically research L2 acquisition, practice effects and effective automatization?

A first limitation is that many ICALL systems are not evaluated to test their educational effectiveness. Indeed, traditionally ICALL research is more focused on specific issues and conceptual advances, so that most systems are not even developed to the point where they could be systematically used and evaluated with real-life learners (Amaral & Meurers, 2011). The commercial ICALL systems becoming available, such as Duolingo, clearly are tested before their official release – yet this testing will primarily concentrate on usability, functionality, and reliability of the software. Serious evaluations of learning outcomes are only starting to become available (e.g., Jiang et al., 2022). This arguably reflects the challenge of bringing experts from multiple disciplines together. Developers of ICALL systems based in industry or computer science departments understandably prioritize software development and usability and typically will not have the background or interest to contribute to our understanding of how languages are optimally acquired. They generally also lack the methodological expertise to run randomized controlled experiments to determine how different instructional manipulations in their system affect learning trajectory and outcomes. Conversely, while some SLA researchers are experts in empirically studying the effects of instruction on language development, they generally will not have the software development skills needed to develop or set up a sophisticated ICALL system.

A second, related limitation concerns the malleability of the different ICALL systems. Are we evaluating a commercial system like Duolingo (e.g., Jiang et al., 2022), which is closed in the sense that we as researchers cannot make modifications in our evaluation studies? Or are we focusing on more open but more prototypical systems like WERTi (Meurers et al., 2010), where the developers support collaboration with SLA researchers and make changes to the system to study its educational effectiveness (Ziegler et al., 2017; Ruiz Hernández, 2019)? The advantage of the former strategy is that no programming skills are required as the system is ready for use. However, we are limited by the pedagogical choices, the functionality and the contents supported by the developers. This will generally make it difficult or impossible to test specific hypotheses (e.g., the utility of different feedback types) or even to include appropriate control groups.

When pursuing the second strategy, working with an open ICALL system generally requires very time consuming collaboration in multidisciplinary teams to develop or modify systems, especially when the systems are supposed to be usable by learners in real-life learning settings. However, as part of a sustained, mutually beneficial collaboration, this option clearly holds greater potential in terms of contributing to theories of SLA.

A third limitation is related to the functionality provided by an ICALL system, which is focused on helping users to learn a new language. As such it will not necessarily facilitate the collection of data needed to conduct SLA research. From consent forms via questionnaire data to cognitive tests, a platform for conducting SLA experiments needs to provide functionality that goes beyond what is needed for an ICALL system to work. Such tasks thus may have to be administered outside the ICALL system, as was done by Ruiz et al. (2021), who tested working memory and declarative memory on a separate platform and sequenced the use of these different online tools in a web-based experiment. With modular web service architectures becoming standard, one can hope that researchers will increasingly provide reusable modules (e.g., standard cognitive tests, pre-/posttest and questionnaire templates) that can be used in an SLA online experiment toolkit, which could substantially reduce the overhead of running studies with an ICALL system.

A final limitation relates to the translation of SLA research into ICALL contexts. Large-scale data collection and learner analytics are only useful if the ICALL system is pedagogically sound to begin with. There seems little point in investigating the effectiveness of an ICALL system that clearly fails to incorporate decades' worth of insights from instructed SLA research or educational psychology. The development of ICALL systems thus requires, in addition to technical expertise, a solid understanding of the research literature on how languages are learned and taught. At the same time, research tends to focus on specific aspects and often isolates individual parameters in an experimental context. The research thus does not necessarily provide the empirical breadth and consider all the factors that are needed to use the insights in a real-life learning context. When it comes to applying these SLA research insights to the development and evaluation of an ICALL system, we thus face challenges on both the empirical and the conceptual side. For example, while for the acquisition of questions, there is substantial SLA research, e.g., on developmental sequences, for most of the other language means on the foreign language school curriculum this is not the case – yet, any ICALL system will need to provide enough breadth for learners in an authentic learning context to be willing to engage with it. On the conceptual side, we need to reflect on how to operationalize key SLA constructs such as noticing, awareness, and automatized

knowledge in an ICALL context. How do we measure noticing or awareness in the absence of verbal reports or eye-movement data? How do we define fluency and automatized knowledge in an ICALL context? And how should we distinguish between implicit and explicit knowledge when we cannot use standard experimental tasks? While some of the questions are specific to the digital domain, others arise for any research attempting to transfer laboratory insights to ecologically valid learning contexts so that there is hope that the rise in interest in instructed SLA and real-life language learning will help us make progress in this regard.

In this chapter, we have characterized different ICALL systems and argued that they can be used to deepen our understanding of how different domains of language are acquired and how they can be optimally taught. However, the full potential of ICALL systems for SLA research is yet to be realized. An essential step for this to happen is to establish an ongoing dialogue between computational and applied linguists, ICALL and SLA researchers, software developers interested in educational technology and researchers interested in language learning and teaching. Establishing this type of dialogue and leading it to fruitful collaboration is not trivial, as we know from personal experience working across disciplinary boundaries (see Rebuschat et al., 2017, for example). It requires a willingness to engage with a range of different research literatures and methodologies and a willingness to publish in journals across different fields, which can be a challenge also in terms of journal and tenure reviewers with a more narrow focus. Naturally, it also requires financial resources that take the multidisciplinary demands into account. To facilitate sustained progress, an important step would be the formation of a larger community of researchers who are willing to engage in cross-disciplinary research at the intersection of CALL, SLA, and NLP. Such a group is needed to take the lead in clarifying key constructs, discussing joint objectives, and aligning research agendas. We hope the present volume can lay the foundation for more collaboration along these lines.

5. Conclusion

We discussed how ICALL can support individualization of L2 practice, as well as how it can be a valuable tool to investigate the role of practice on L2 learning. Concerning individualization of practice, we showed how ICALL applications can offer practice on a large scale that can be tailored to individual learners (e.g., Meurers et al., 2019), addressing what DeKeyser calls “true individualization” (2021, p. 2). In such a setup, ICALL can help explore ATIs in instructed language learning, which can lead to the design and implementation of more effective pedagogical interventions. We also discussed that ICALL applications can serve as a research tool that allows L2 researchers

to collect empirical evidence at scale in ecologically-valid contexts not only on the product but also the process of L2 learning (Hui et al., in press; Ruiz et al., 2021). Finally, we identified several issues currently limiting the ability to use ICALL systems to study L2 acquisition, practice effects, and automatization. We discussed the lack of empirical assessment of the educational effectiveness of existing tools, the challenge of adapting system to SLA researchers' experimental designs, and the need to empirically broaden SLA insights and establish how one can operationalize and measure important SLA concepts in ICALL systems in a real-life learning context. Overall, the discussion presented here underscores the importance of more cross-disciplinary work between ICALL and SLA and realize its potential to advance our understanding of second language acquisition and improve language learning and teaching.

References

- Alexopoulou, T., Meurers, D., & Murakami, A. (2022). Big data in SLA: Advances in methodology and analysis. In N. Ziegler & M. Gonzalez-Lloret (Eds.), *The Routledge handbook of second language acquisition and technology* (pp. 92-106). Routledge.
- Amaral, L., & Meurers, D. (2008). From recording linguistic competence to supporting inferences about language acquisition in context. *Computer Assisted Language Learning*, 21(4), 323-338. <https://doi.org/10.1080/09588220802343454>
- Amaral, L., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, 23, 4-24.
- Brett, D., & González-Lloret, M. (2011). Technology-enhanced materials. In M. H. Long, Doughty, & C. J. (Eds.), *The handbook of language teaching* (pp. 351-369). Wiley.
- Bull, S., Brna, P., & Pain, H. (1995). Extending the scope of the student model. *User Modeling and User-Adapted Interaction*, 5, 45-65.
- Carrell, P., & Wise, T. (1998). The relationship between prior knowledge and topic interest in second language reading. *Studies in Second Language Acquisition*, 20(3), 285-309. <https://doi.org/10.1017/S0272263198003015>
- Chen, X. (2018). *Automatic analysis of linguistic complexity and its application in language learning research*. Doctoral dissertation. Universität Tübingen, Germany. <http://dx.doi.org/10.15496/publikation-27277>
- Chen, X., & Meurers, D. (2017). Challenging learners in their individual zone of proximal development using pedagogic developmental benchmarks of syntactic complexity. In *Proceedings of the joint 6th workshop on NLP for computer assisted language learning and 2nd workshop on NLP for research on language acquisition at NoDaLiDa 2017* (pp. 8-17). ACL.
- Chen, X., & Meurers, D. (2019). Linking text readability and learner proficiency using linguistic complexity feature vector distance. *Computer Assisted Language Learning*, 32(4), 418-447. <https://doi.org/10.1080/09588221.2018.1527358>
- Chen, X., Meurers, D., & Rebuschat, P. (2022). ICALL offering individually adaptive input: Effects of complex input on L2 development. *Language Learning & Technology*, 26(1), 1-21. <https://hdl.handle.net/10125/73496>
- Chen, X. L., Zou, D., Xie, H. R., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning & Technology*, 25(3), 151-185.

- Chinkina, M. & D. Meurers (2016). Linguistically-Aware Information Retrieval: Providing Input Enrichment for Second Language Learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. San Diego, CA: ACL (pp. 188-198).
- Chinkina, M. & Meurers D. (2017). Question generation for language learning: From ensuring texts are read to supporting learning. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. Copenhagen, Denmark (pp. 334-344).
- Chinkina, M., Ruiz, S., & Meurers, D. (2020). Crowdsourcing evaluation of the quality of automatically generated questions for supporting computer-assisted language teaching. *ReCALL*, 32(2), 145-161.
- Choi, I. (2016). Efficacy of an ICALL tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, 29(2), 334-364. <https://doi.org/10.1080/09588221.2014.960941>
- Council of Europe (2020), *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*, Council of Europe Publishing, Strasbourg. Available at <http://www.coe.int/lang-cefr>
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.
- Dehaene, S. (2020). *How we learn: The new science of education and the brain*. Penguin.
- DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.
- DeKeyser, R. (2010). Practice for second language learning: Don't throw out the baby with the bathwater. *International Journal of English Studies*, 10(1), 155–165. <https://doi.org/10.6018/ijes/2010/1/114021>
- DeKeyser, R. (2017). Knowledge and skill in SLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 15-32). Routledge.
- DeKeyser, R. (2021). *Aptitude-treatment interaction in second language learning*. John Benjamins.
- Ellis, R. (2010). Theoretical pluralism in SLA: Is there a way forward? In Seedhouse, P., Walsh, S., Jenks, C. (Eds.), *Conceptualising 'learning' in applied linguistics*. Palgrave Macmillan. https://doi.org/10.1057/9780230289772_3
- Ellis, R. (2017). Task-based language teaching. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 108-125). Routledge.

- Fournier-Viger, P., Nkambou, R., & Nguifo, E. M. (2010). Building intelligent tutoring systems for ill-defined domains. In Nkambou, R., Mizoguchi, R. & Bourdeau, J. (Eds.). *Advances in intelligent tutoring systems* (pp. 81-101). Springer.
- Fu, M., & Li, S. (2022). The effects of immediate and delayed corrective feedback on L2 development. *Studies in Second Language Acquisition*, 44(1), 2-34. <https://doi.org/10.1017/S0272263120000388>
- Grabe, W., & Stoller, F.L. (2011). *Teaching and researching reading* (2nd ed.). Routledge.
- Hartley, J. R., & Sleeman, D. H. (1973). Towards more intelligent teaching systems. *International Journal of Man-Machine Studies*, 5(2), 215-236. [https://doi.org/10.1016/S0020-7373\(73\)80033-1](https://doi.org/10.1016/S0020-7373(73)80033-1)
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hawkins, J. A., & Filipović, L. (2012). *Criterion features in L2 English: Specifying the reference levels of the Common European Framework*. Cambridge University Press.
- Heck, T., & Meurers, D. (2022). Parametrizable exercise generation from authentic texts: Effectively targeting the language means on the curriculum. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)* (pp. 154-166). <http://dx.doi.org/10.18653/v1/2022.bea-1.20>
- Heift, T. (2010). Developing an intelligent language tutor. *CALICO Journal*, 27(3), 443-459.
- Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. Routledge.
- Hui, B., Rudzewitz, B., & Meurers, D. (in press). Learning processes in interactive CALL systems: Linking automatic feedback, system logs, and learning outcomes. *Language Learning & Technology*.
- Jackson, D. (2022). *Task-Based Language Teaching (Elements in Language Teaching)*. Cambridge: Cambridge University Press.
- Jiang, X., Rollinson, J., Plonsky, L., Gustafson, E., & Pajak, B. (2022). Evaluating the reading and listening outcomes of beginning-level Duolingo courses. *Foreign Language Annals*, 54(4), 974-1002. <https://doi.org/10.1111/flan.12600>
- Nassaji, H., & Kartchava, E. (Eds.). (2021). *The Cambridge handbook of corrective feedback in second language learning and teaching*. Cambridge University Press.
- Klein, W. & C. Perdue (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, 13(4), 301-347.

- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- Kulik, J. A., & Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42-78. <https://doi.org/10.3102/0034654315581420>
- Larsen-Freeman, D. (2013). Transfer of learning transformed. *Language Learning*, 63, 107-129.
- Leow, R. P. (2019). Theoretical underpinnings and cognitive processes in instructed SLA. In Leow, R.P. (Ed.). (2019). *The Routledge handbook of second language research in classroom learning*. Routledge.
- Lichtman, K., & VanPatten, B. (2021). Was Krashen right? Forty years later. *Foreign Language Annals*, 54, 283-305.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, C. Kramsch & R. Ginsberg (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). John Benjamins.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based learning. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123-167). Multilingual Matters.
- Meurers, D. (2015). Learner corpora and natural language processing. *The Cambridge handbook of learner corpus research*, 537-566. <https://doi.org/10.1017/CBO9781139649414.024>
- Meurers, D. (2021). Natural language processing and language learning. In C. A. Chapelle (Ed.), *The concise encyclopedia of applied linguistics*, 817-831. Wiley. <https://doi.org/10.1002/9781405198431.wbeal0858.pub2>
- Meurers, D., & Dickinson, M. (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, 67(S1), 66-95. <https://doi.org/10.1111/lang.12233>
- Meurers, D., De Kuthy, K., Möller, V., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2018). Digitale Differenzierung benötigt Informationen zu Sprache, Aufgabe und Lerner. Zur Generierung von individuellem Feedback in einem interaktiven Arbeitsheft [Digital differentiation requires information on language, task, and learner. On the generation of individual feedback in an interactive workbook]. *FLuL–Fremdsprachen Lehren und Lernen*, 47(2), 64-82. <http://purl.org/dm/papers/Meurers.DeKuthy.ea-18.pdf>
- Meurers, D., De Kuthy, K., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school.

- Annual Review of Applied Linguistics*, 39, 161-188.
<https://doi.org/10.1017/S0267190519000126>
- Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., & Ott, N. (2010). Enhancing Authentic Web Pages for Language Learners. In *Proceedings of the 5th workshop on innovative use of NLP for Building Educational Applications (BEA) at NAACL-HLT 2010* (pp. 10-18). Los Angeles, California: Association for Computational Linguistics.
- Nagata, N. (2009). Robo-Sensei's NLP-based error detection and feedback generation. *CALICO Journal*, 26(3), 562-579.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
<https://doi.org/10.1111/0023-8333.00136>
- Petersen, K. A. (2010). *Implicit corrective feedback in computer-guided interaction: Does mode matter?* Doctoral Dissertation. Georgetown University.
<http://hdl.handle.net/10822/553155>
- Petersen, K., & Sachs, R. (2015). The language classroom in the age of networked learning. In Leow, R. P., Cerezo, L., & Baralt, M. (Eds.), *Technology and L2 learning: A psycholinguistic approach* (pp. 3-22). De Gruyter Mouton.
- Pienemann, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics* 10(1), 52-79.
- Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45-142). National Curriculum Resource Centre.
- Quixal, M., Rudzewitz, B., Bear, E., & Meurers, D. (2021). Automatic annotation of curricular language targets to enrich activity models and support both pedagogy and adaptive systems. *Proceedings of the 10th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL)*. 38-54.
- Rebuschat, P., Meurers, D., & McEnery, T. (2017). Language learning research at the intersection of experimental, computational and corpus-based approaches: *Language Learning*, 67(S1), 6-13.
- Rudzewitz, B., Ziai, R., De Kuthy, K., & Meurers, D. (2017). Developing a web-based workbook for English supporting the interaction of students and teachers. In *Proceedings of the Joint 6th Workshop on NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa 2017* (pp. 36-46).

- Rudzewitz, B., Ziai, R., De Kuthy, K., Möller, V., Nuxoll, F., & Meurers, D. (2018). Generating feedback for English foreign language exercises. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications* (pp. 127-136). <http://dx.doi.org/10.18653/v1/W18-0513>
- Ruiz Hernández, S. (2019). *Individual differences and instructed second language acquisition: Insights from intelligent computer assisted language learning*. Doctoral dissertation. Universität Tübingen, Germany. <http://dx.doi.org/10.15496/publikation-27963>
- Ruiz, S., Rebuschat, P., & Meurers, D. (2019). Measuring individual differences in cognitive abilities in the lab and on the web. *PLoS ONE* 14(12): e0226217. <https://doi.org/10.1371/journal.pone.0226217>
- Ruiz, S., Rebuschat, P., & Meurers, D. (2021). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research*, 25(4), 510-539. <https://doi.org/10.1177/1362168819872859>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 206-226.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43.
- Schulze, M. (2012). Learner modeling in intelligent computer-assisted language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley. <https://doi.org/https://doi.org/10.1002/9781405198431.wbeal0671>
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15, 165-179. <https://doi.org/10.1017/s0272263100011943>.
- Suk, N. (2017). The effects of extensive reading on reading comprehension, reading rate, and vocabulary acquisition. *Reading Research Quarterly*, 52(1), 73-89. <https://doi.org/10.1002/rrq.152>
- Suzuki, Y. (2022) Automatization and practice. In Godfroid, A., & Hopp, H. (Eds.). *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 308-321). Routledge.
- Suzuki, Y., Nakata, T., & DeKeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *Modern Language Journal*, 103, 713–720.

- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Lawrence Erlbaum, 471-484.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Weiss, Z., S. Dittrich & D. Meurers (2018). A linguistically-informed search engine to identify reading material for functional illiteracy classes. In *Proceedings of the 7th Workshop on Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL)*. Association for Computational Linguistics.
- Ziegler, N., Meurers, D., Rebuschat, P., Ruiz, S., Moreno Vega, J. L., Chinkina, M., Li, W., & Grey, S. (2017). Interdisciplinary research at the intersection of CALL, NLP, and SLA: Methodological implications from an input enhancement project. *Language Learning*, 67, 209-231. <https://doi.org/10.1111/lang.12227>