Artificial Intelligence in Foreign Language Learning and Teaching

Schmidt, Torben; Strasser, Thomas

Published in:
Anglistik - International Journal of English Studies

DOI:
10.33675/ANGL/2022/1/14

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

TORBEN SCHMIDT AND THOMAS STRASSER

Artificial Intelligence in Foreign Language Learning and Teaching: A CALL for Intelligent Practice

1. Introduction

Practice and focus on form play a crucial and decisive role in foreign language learning. But what would an intelligent, adaptive foreign language learning environment look like if all students could individually practice their language skills with exercises tailored to their individual skill levels, interests, and motivation? How could all learners be supported and challenged according to their abilities, so that they all have the opportunity to achieve specific learning goals in a self-directed manner? And how could digital media contribute to the kind of learning that adapts to the individual student's needs in heterogeneous foreign language classrooms?

In the past years, digital technologies have become scientific and practical focal points in the English language teaching (ELT) world. Whether digital media [are] "friend or foe" (Grimm et al. 2015), technology-enhanced language learning (TELL) has been part of an international discourse, varying between "euphoric proposals," "pessimistic stances," and "opinions which stress that the risks of digital media need to be addressed" (2015, 210). Regardless of general TELL, research studies have shown that "technology can influence the processes and outcomes of education, and many countries are investing in technological support for teaching and learning" (Paiva and Bittencourt 2020, 448). The dynamic development of new technologies and the concomitant digital transformations result in significant challenges both for society as a whole and at all levels of the education system.

One of the latest technological developments, which raises more and more interest in connection with these questions, is artificial intelligence (AI). There are many different definitions of what exactly AI is. However, there seems to be a certain semantic lowest common denominator: "Artificial intelligence (AI) is a broad term used to describe a collection of technologies that can solve problems and perform tasks to achieve defined objectives without explicit human guidance" (Healey 2020, 3). The development of so-called narrow AIs (less complex algorithms that are good at doing one thing in particular, e.g. Siri, AI-powered vacuum cleaners) explicitly shows that data-driven, multi-layered technologies based on algorithms have transformed from a niche discipline into a highly relevant technology for educational, including language learning, purposes. The potential of these programs to analyze unprecedented amounts of data, collected in real time, combined with novel methods from the field of AI, are bringing the optimization of teaching and learning processes into particular focus. Thus, Big Data, combined with sophisticated analytical processes (learning analytics) give hope for a new era of personalized learning, formative assessment, and activating, student-centered, and collaborative forms of learning (Cope and Kalantzis 2016), since, by means of educational data mining, learning processes could theoretically become

Anglistik: International Journal of English Studies **33.1 (Spring 2022): 165-184.**

highly individualized. In this context, new technologies affect the educational system much more substantially than merely on the level of didactic surface structures. This is because they blur the boundaries between formally organized learning environments in schools and informal learning opportunities in leisure time and at home (Seufert et al. 2018). They challenge learners and teachers to adopt an entirely new approach to these new digitally enhanced educational spaces. However, due to the complexity of such systems, the development of AI-based foreign language learning environments that adapt to learner heterogeneity, thus enabling high-level subject learning and practice, is still in its infancy. Although foreign language learning programs are among the most widely used applications on the Internet, and Computer-Assisted Language Learning (CALL) software of various kinds has widely been used for several decades, until today, only very few products have more than rudimentary features of intelligent adaptive systems. In some aspects, only a small number of applications provide a learning experience beyond linear paths, simple right-wrong feedback routines, or a 'one-size-fits-all'-approach (Blume et al. 2017). A great desideratum is to intensify interdisciplinary cooperation between (computer-)linguists, experts in teaching English as a foreign language (EFL), educational scientists, psychologists, computer scientists and interface designers in order to develop smarter foreign language learning systems.

This paper aims to critically examine the potentials and challenges of developing AI-enhanced CALL-software to enhance adaptive, individualized, and, in certain respects, intelligent practice in the foreign language classroom. The following section provides an introductory, practical discussion of terms, methods, and common application types of (narrow) AIs in foreign language learning processes. We will then, as the main focus of this paper, address features of intelligent practice in the foreign language classroom, the possibilities of using AI to create adaptive learning environments, and typical architectural elements of intelligent learning systems. Subsequently, this paper provides insights into two current research and development projects. The article concludes with a utopian outlook into the foreign language classroom of the year 2040.

2. (Narrow) AIs: General Typologies and Practical Examples

By 2023, more than 40% of global capital in the education market will be spent on AI and education (e.g. also language learning apps, intelligent lexical visualizing tools (MARKETSANDMARKETS 2018). As a result, more and more companies and educational stakeholders are investing in AI-powered technologies. This section outlines how three (narrow) AI-powered tools can be used and methodologically exploited in the EFL classroom. As a contrast, their current technical and methodological limitations are discussed as well. Starting with these rather generic tools, the paper will then discuss the use of AI-tools that support digital practice, so-called Intelligent Tutoring Systems (ITS) in the subsequent sections.

Before these AI-powered applications and their methodological potentials and limitations are discussed, it seems to be of relevance to specify the following general AI key concepts in terms of AI-powered language learning:

1. **Natural Language Processing (NLP)** is an area that combines AI and linguistics in general and is concerned with the automated processing of human language. It addresses the generation and analysis of written and spoken language, though speech processing is often regarded as a separate subfield. NLP can be seen as the applied side of computational linguistics, the interdisciplinary field of research concerned with formal analysis and modeling of language and its applications at the intersection of linguistics, computer science, and psychology (Meurers 2020, 817). Intelligent Language Tutoring Systems (ILTS) deal with lexical, morphological and syntactic aspects of language as well as aspects of meaning, discourse, and the relation of these to the extra-linguistic context (2020, 817).
2. **Machine Learning (ML)** is a part of AI. It refers to systems that derive information from or learn based on experience. Machine learning "helps us find solutions to many problems in vision, speech, recognition, and robotics" (Alpaydin 2014, 3). Furthermore, it can be claimed that ML refers to "programming computers [and their corresponding applications & software] to optimize a performance criterion using example data or past experience" (2014, 3). Hence, example data must be categorized in order to use it for NLP purposes.
3. **Deep Learning (DL)** is a subfield of AI that uses Artificial Neural Networks (computing systems resembling specific neural networks of a human's brain) to learn from extensive data sets (Schmidhuber 2015, 86). Deep Learning mainly focuses on vision-based categories (e.g. distinction of images), but can also be used for NLP purposes.

Based on this general overview of AI-technologies, this section identifies Artificial Intelligences embedded in an EFL-context and provides general scenarios of how these specific tools can be methodologically implemented for the language learner's benefit. Several experts suggest that AI-powered tools "offer the possibility of learning that is more personalized, flexible, inclusive, and engaging" (Luckin et al. 2016, 11), and there is evidence that technology-enhanced language learning per se has become more dynamic, assistive and adaptive (Schmidt and Strasser 2018). Especially for assistive and adaptive language learning scenarios (the algorithm supports the teacher and learner and creates an individual learning path), so-called narrow AIs come into play. They could be categorized in the following way (Baker and Smith 2019, 5):

1. **Learner-facing AI tools**: supportive AI-tools that help students improve in a particular subject matter through specific practice patterns, reflective feedback

mechanisms or behavioural drills, e.g., applications like *Babbel* that provide immediate feedback based on the learner's input (mixed tenses, verb forms).

2. **Teacher-facing systems**: teacher-centred tools that aim to minimize the teacher's workload, mainly in automated processes (like grading, feedback mechanisms, classroom management, administrative issues), e.g., *GradeScanner*, an application that automatically grades multiple-choice tests.
3. **System-facing AI tools**: algorithms that provide processed data mainly for institutional administrators or stakeholders, e.g., software that processes a student's transcript and calculates his/her future performance.

The following examples seek to examine AI potentials and challenges for language learning purposes (Figure 2). Even if there is a contentious scientific discussion as to whether an individual application or software is AI-powered or not (Banerjee 2020; Berendt et al. 2020; Kannan and Munday 2018; Paiva and Bittencourt 2020), the following emerging AI-labelled (narrow AIs) technologies are shaping the educational market.

2.1 Machine Translation

Machine translation is the process of (automatically) translating a particular text (spoken or written) from one language to another. Machine translation applications use an extensive linguistic corpus and a myriad of complex algorithms to translate written texts automatically. One of the most popular AI-powered translation websites at the moment is *DeepL* (www.deepl.com). It uses deep learning technology based on neural network technology. These networks are not recurrent but convolutional, which means that *DeepL* processes words, phrases, and utterances in synchronous threads similar to image processing procedures (Merkert 2017).

There are several evident potentials concerning the use of *DeepL* or the like in the EFL-classroom. Machine translation could be helpful for 'understanding-the-gist' or for global comprehension scenarios where language learners need a quick translation in order to proceed with a specific activity. Another use might occur when learners want to get a more precise understanding of specific passages: They can merely click on the word to get the meaning and/or a synonym of the word/phrase. The integrated dictionary/thesaurus helps the learners find semantic approximations for their translation. This interactive corpus generally offers pragmatically coherent translations (the system often recognizes co-text and context); learners can at least passively/receptively extend their receptive lexicon.

In general, machine translation software like *DeepL* performs relatively well in simple syntax and phrases and/or utterances, as well as in specific discursive contextual/intercultural situations (e.g., slang, dialects, regional idioms). However, in many cases, there is still a need for the language learner's linguistic expertise.

Especially in writing texts, learners might come to rely on the translations provided by *DeepL* without developing autonomous skills in paraphrasing or writing. Furthermore, learners have to be careful about unreflectively applying the provided lexical algorithms. Quite frequently, the program lacks expertise in identifying particular contextualized usages of semantic items or attending to textual coherence. Especially in the field of technical and medical terminology, the machine does not always provide an entirely correct translation, since the complicated jargon often requires substantial technical and medical expertise, which the algorithm does not (yet) have.

Therefore, it has to be emphasized that translations might not be the most effective way to learn vocabulary. Moreover, translation software like *DeepL* and *Google Translate* often do not include minority languages, which are an essential part of authentic language learning processes.

2.2 AI Writing Tools

Built-in spell-checkers in text processing software have dramatically improved over the last few years. In the beginning, they merely analyzed a user's orthographic infelicities. The built-in software apps subsequently extended their functions by integrating grammar checking tools. However, most of the early programs were limited in their ability to understand words in context and more nuanced grammar functions.

Since we communicate more and more digitally, our texts and discursive artefacts are being disseminated more efficiently through various channels and networks, and there has been a particular demand to scan our written discourse more reliably. Several AI-powered writing tools like *Grammarly* (www.grammarly.com) have entered the market, offering spelling & grammar checks and functions that analyze the clarity, coherence, engagement, and delivery of a text, and can thus scan written text more reliably. Based on an extensive corpus and algorithms, *Grammarly* also offers linguistic recommendations and adaptions for a text based on given parameters regarding register and delivery. The system's context-dependent suggestions are based on knowledge and predictions about the target group (reader), the level of formality (informal, neutral, formal), the domain (academic, business, general, email, casual, creative), the tone (neutral, confident, joyful, optimistic, friendly, urgent, analytical, respectful) and the intent (inform, describe, convince, tell a story).

AI-powered writing tools like *Language Tool* or *Grammarly* might help the learner develop a specific awareness of a language's versatility. The basic version offers a relatively basic approach of a language checker, whereas the premium version is rather costly. The learner might realize that language is not a mere process of spelling and grammar checking, but also strongly depends on context and coherence. Since the learner has to initially write a text before it is analyzed, the classic productive writing skill is not being 'threatened' or replaced by the machine. *Grammarly* not only indicates whether an

expression or passage within the written text is incorrect, but it also offers linguistic feedback.

By offering colour-coded tagging, learners immediately see the nature of their linguistic deficiencies. The red category indicates mistakes or errors in the field of grammar and spelling. The blue category focuses on clarity, which means that the program analyzes whether sentences are too long or syntactically too complex. Turquoise markings provide feedback regarding delivery, i.e., if the used words and phrases convey the writer's arguments properly. The pink category refers to delivery (register), which means if the used words and phrases are target-group adequate (Figure 1). All the linguistic infelicities or redundancies are marked with a context-dependent feedback window, which means that the system also provides alternatives and explanations as to why, for example, the use of the passive voice is not ideal in a given passage. For learners, this can be of great help in improving their writing skills. Not only do the learners get an overall score for their initial productive performance, but *Grammarly* also offers them coursebook-like explanations concerning various lexical and grammatical phenomena. Therefore, these narrow AIs are paradoxically holistic language learning (writing) tools that do not exclusively focus on deficiencies, but also on improvement potential in various ELT-relevant domains such as vocabulary and grammar acquisition as well as productive skills.

Similar to the limitations of AI-powered translation websites, such writing tools cannot guarantee 100% linguistic accuracy, since pragmatically coherent texts are strongly context-dependent. The context consists of many cultural implications so that one system alone cannot guarantee complete appropriateness. Furthermore, systems like *Grammarly* frequently show redundant feedback recurrence patterns, which means that they often use the same feedback suggestions for different types of mistake.

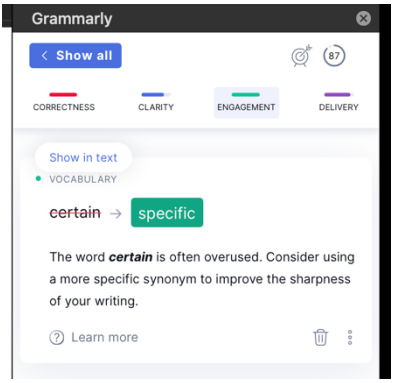

Figure 1: AI-powered spell and style check

2.3 Chatbots/Virtual Chat Agents

In general, chatbots intend to imitate the discursive behaviour of humans. Since chatbots have access to massive linguistic corpora, they are becoming more and more intelligent. As a result, companies in the field of educational technology have started developing their own virtual bots or agents because "chatbots provide customized answers in response to your messages and can even grade your performance or give tips on what you need to improve" (Lisovets 2017). Especially for beginners, chatbots might be helpful, since basic words/phrases (e.g., within the A1/A2 range) often have clear, formulaic structures without substantial semantic, syntactic or lexical complexity. Therefore, the machine is not likely to fail pragmatically. In general, a chatbot is intended to be user-friendly, offering a straightforward interface and the ability to be used ubiquitously (mobile and desktop version, temporally and spatially delimited use). Chatbots like the *Virtual Talk App* (App Store), *Mondly* (www.mondly.com) or *Memrise*, *Babbel* and *Duolingo* (App Store) use a multimedia interface, so that learners can practice the target language with audio, video and images while talking (in the case of integrated speech recognition) or writing with the chatbot. Especially anxious students, eager to avoid face-to-face conversation in the real-life classroom, might benefit from these conversational scenarios with a relatively anonymous chatbot, since mistakes or personal details will not be made public and paralinguistic indications of boredom or judgment are absent.

Despite these potential advantages, the issue of data privacy needs to be explicitly addressed. Language learning is a very personal process that should not be shared with chatbots where the server location is unclear or when not in accordance with the EU General Data Protection Regulation (GDPR). Companies like *Memrise*, *Babbel* and *Duolingo* guarantee safe storage of the learner's data in their terms of use. Whether this is really the case must be decided by the individual learner or their parents or teachers and/or even regional legal experts (government authorities). Another challenge of automated, algorithm-based chatbots is that often (especially in more sophisticated conversations) the bot displays "semantic problems with [not always] complex issues and cultural specifics, i.e. they provide […] answers or turns that do not make sense (similar to the Siri phenomenon)" (Strasser 2021, 100; 101).

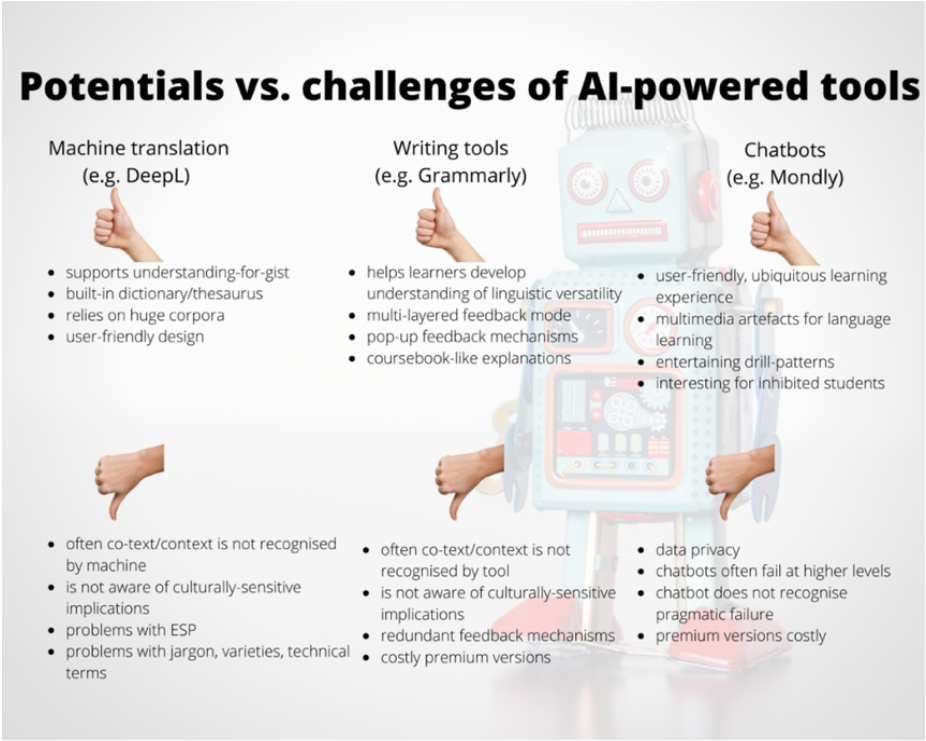

Figure 2: Language learning potentials and challenges of three main AI application types (created with www.canva.com)

3. The Importance of Intelligent Practice in Communicative TBLL

Complex, communicative learning tasks and Task-Based Language Learning (TBLL) as a methodological approach (Nunan 1989; Willis 1996) are at the heart of modern foreign language teaching. TBLL enables the integrated development of functional communicative competencies as required by the *Common European Framework of Reference for Languages* (Council of Europe 2001) and corresponding national European educational standards. The overriding goal is to enable "Teilhabe an den eigenen lebensweltlichen Diskursen [der Lernenden] und zugleich an analogen oder damit verbundenen fremdsprachigen Diskursen in anderen kulturellen Kontexten" (Hallet 2011, 56). Ideal foreign language learning tasks should therefore 1) promote the use of the target language as a communicative activity, 2) focus on the content and meaning of a message, 3) enable a high degree of authenticity in language action, 4) lead to a clearly defined linguistic product or outcome and, 5) on a higher level, allow

learners to communicate and act authentically (Mertens 2010). According to the Output Hypothesis (Swain 2005), this kind of language production is essential for language learning.

Practice, understood as "repetitive, purposeful, and goal-oriented use of language-related activities" (Siebold 2007, 64) that are functional, transferable, contextualized, receptive and productive, plays a crucial and decisive role in such communicative, competence- and task-oriented lessons. Foreign language learners need to focus on form in the context of practice to develop the skills and competencies (Lightbown 1998) necessary for communicative settings. Doughty (2001, 211) adds that "(t)he factor that consistently distinguishes focus on form from other pedagogical approaches is the requirement that focus on form involves learners briefly and perhaps simultaneously attending to form, meaning, and use during one cognitive event." This kind of joint processing facilitates the cognitive mapping of form, meaning, and use that is fundamental to language learning. Therefore, internalizing linguistic structures based on explicitly conveyed or incidentally acquired declarative rule knowledge plays an important role in communicative, task-based language learning (DeKeyser 2005). Müller-Hartmann and Schocker-von Ditfurth (2010, 205; 206), relating to Willis' (1996) ideas regarding the role of focus on form in TBLT, also emphasize the necessity of form-focused activities in different phases of a task cycle. Funk specifies that, in any case, exercises must be "directly and recognizably useful in mastering the communicative target tasks" (Funk 2016, 69, our translation). However, this does not imply that practice needs to consist of mechanical drills in the image of the audiolingual method. According to Bär (2016), Blell (2016), Heymann (2005), Kieweg (2014), Klippel (1998) and Schmidt (2016), intelligent practice in EFL learning should instead

- focus on the needs of the learner in order to communicate (fluently and freely) in the target language (competence orientation);
- be authentic in terms of language and content;
- cultivate autonomy;
- be scaffolded;
- provide feedback (linguistic correctness, communicative adequacy, strategic approach);
- be individualized and differentiated;
- adapt to the learner's affective and cognitive needs;
- be goal-oriented and transparent;
- and develop meta-cognitive (language) learning skills.

In this way, the role of practice in a communicative foreign language classroom differs from earlier historical approaches focusing on the rote acquisition of identified structures without a communicative purpose. The goal is to individualize the exercise processes in order to adaptively support learners in developing their language skills

within the framework of competence-oriented tasks. It is critical that these exercise phases are highly differentiated and individualized, so that the learners work on appropriately challenging exercises in their *Zone of Proximal Development* (ZPD), which are tailored to their individual learning requirements, especially in social interaction with other learners (Vygotsky 1986). In this process, they also need to receive scaffolding and high-quality feedback (Aljaafreh and Lantolf 1994; Hattie and Timperley 2007; Kleppin 2008). The goal is not mastery of the language system as such in these activities. Instead, the practice, with its targeted focus on form(s) should derive from an analysis of the target task in anticipation of the linguistic challenges it poses. Such practice may also be warranted after a task, after learners' linguistic deficits that become apparent during task processing or performance have been identified. Whether such practice occurs prior to the culmination in a task product or subsequently, it is necessary in order to consolidate relevant language and speech resources (Bär 2016, 12).

The fundamental question arises as to whether and, if so, to what extent and how digital learning programs and learning systems can be used to achieve these goals, thereby providing the framework for highly effective, intelligent practice. AI-driven Intelligent Tutoring Systems (ITS), as digital learning and support systems, theoretically have the potential to address some of the key challenges of creating individualized, adaptive practice environments to better support foreign language learners. But what do current language learning programs offer in these areas?

A first hint could be gleaned from the fact that, despite the long history of almost 30 years of ITS in education, programs for foreign language learning are entirely absent from Kulik and Fletcher's (2016) meta-analysis of 50 controlled evaluations on the effectiveness of these systems. Their absence in this meta-analysis is not an oversight on the part of the authors. Instead, it is fundamentally related to the fact that language is an ill-defined domain, in contrast to mathematics, computer science, natural sciences, and other disciplines for which various ITS and adaptive learning systems have been successfully developed. The main challenge in the development of Intelligent Language Tutoring Systems (ILTS) is establishing a valid automatic analysis of the learner language (interlanguage), using Natural Language Processing (NLP) based on students' written and oral answers, contributions, products etc., since the range of possible variations in human language in general and learner language more specifically is difficult to anticipate and model precisely. Moreover, this (computational) linguistic modeling must then go hand-in-hand with task, exercise and learner modeling, including the analysis of task and exercise effects using learner corpora. Thus, apart from some quite promising development projects at universities (e.g. Heift 2010; Nagata 2009; Amaral and Meurers 2011; Meurers et al. 2019a; 2019b), there are hardly any ILTS programs that are being widely used in school practice for foreign language learning, especially in Germany. Of course, there is a very diverse range of language learning programs available. But there is a significant difference between traditional

Computer-Assisted Instruction (CAI) or *Computer Assisted Language Learning* (CALL) and ILTSs, the latter having the capability of interactive problem-solving, intelligent error analysis, adaptive feedback and scaffolding, and curriculum sequencing. These types of programs remain rare. An analysis of 50 current foreign language learning programs in Blume et al. (2017) shows that many currently available, widely used programs do not offer these elements. Available applications are mostly of very poor quality in terms of feedback; the exercises offered are typically only restrictedly oriented towards communicative purposes; there is frequently a decided focus on isolated grammar and vocabulary practice; and the exercise selection and sequencing are usually inflexibly predetermined by the program. Individualization and adaptivity require AI methods which most programs currently available do not (yet) offer. In short, the market is flooded with a great deal of CALL, but ICALL is only very limited.

Teachers need information and knowledge about the exercises' linguistic requirements, the individual learners, their errors and mistakes, strengths and weaknesses, affective attributes, etc. In order to provide an adaptive learning environment and individualized, differentiated practice, an adaptive intelligent computer-based learning support system for foreign language learning also requires a specific architectural structure with three main pillars, as closely interrelated areas, to realize intelligent and adaptive practice:

1. A **domain model** that contains information about the language or certain parts of language addressed in the exercises (e.g., an ITS to practice verb tenses needs a complete grammatical model of the English tense-aspect-mood system), an expansive pool of exercises addressing all kinds of phenomena and different degrees of difficulty, combined with feedback, scaffolding, and tutorials, all based on an empirically validated model of difficulty for the relevant grammatical concepts.
2. An **evaluation model** that constantly assesses the individual and all learners' performance in all activities, identifies types of errors, monitors learning time, number of attempts, use of feedback, scaffolding, hints and tutorials.
3. A **learner model** that gathers and updates data on the learner's mastery of the domain model features as the learner progresses through content.

All three pillars form the foundation for the **adaptation model** of the **adaptation engine,** which steers the overall learning process and determines and coordinates response types (feedback, new content, revision, etc.), depending on data from all three models mentioned above (Meurers et al. 2018). And this is by no means a trivial task, even for a computer program: Developing a software that uses learning analytics, NLP, and error analyses to conduct individual modeling of learners, in order to ascertain their respective level of competence, to assign and sequence exercises and implement

intelligent feedback and scaffolding that help learners develop their skills, requires interdisciplinary teams of experts to be involved in the development process.

In the following section, we will briefly describe two current research and development projects that deal with questions regarding the construction and development of AI-supported focus on form-learning environments for digitally enhanced intelligent EFL practice.

4. Project 1: Dynamic Difficulty Adaptation in EFL Focus on Form-Practice

Advances in computer technology and artificial intelligence create opportunities for developing adaptive foreign language learning technologies which are sensitive to individual learner characteristics and competencies. An intelligent adaptive learning system would theoretically maintain learners in their Zone of Proximal Development and in a permanent state of flow (Csíkszentmihályi 1990) so that a perfect balance between skills and challenge is achieved. In order to realize this goal, it is essential to define, measure and analyze difficulty levels. If a system can predict the difficulty of an exercise (degree of difficulty, difficulty-influencing features), it can manipulate these factors and present the learner with exercise items that perfectly meet his/her skills, thus facilitating learning progress by dynamically adapting to the individual's evolving language ability. In a two-year interdisciplinary research project at Leuphana University in Lüneburg, researchers from the fields of teaching and learning EFL, English linguistics, AI, and educational sciences conducted a study that aimed to contribute to models of (semi-)automatic difficulty scoring of grammar exercise items (focus on English tenses), which were to be used in dynamic difficulty adaptation in an intelligent language tutoring system (Pandarova et al. 2019). In it, methods from item response theory and machine learning were combined with linguistic item analysis in order to calibrate the difficulty of an initial exercise pool of cued gap-filling items (CGFIs) and to isolate CGFI features predictive of item difficulty. The term CGFI refers to limited-production exercise items, where learners read a short text and fill in the gaps using cues usually consisting of one or two words that must be transformed to fit the context:

1. The Taj Mahal _____ (build) around 1640.
2. Laura: Where's Julie? Isn't she here? Mark: She isn't, I _____ (not, see) her all day.
3. That man looks familiar. I _____ (definitely, see) him somewhere before.

Multiple item features at the gap, context and CGFI levels were tested, and relevant predictors of item difficulty were identified at all three levels. Though the primary focus of these items is on the form and meaning of English tenses, the examples show that a number of epiphenomenal features, including voice, polarity, person/number inflection,

word order and irregular morphology, are also targeted. A 40-minute test was administered to 678 ninth- and tenth-graders in two preparatory high schools (*Gymnasium*) and two integrated comprehensive schools (*Integrierte Gesamtschule*) in Lower Saxony. 292 exercise items were completed by a mean of 91.68 students (SD = 3.86), with each student working on a mean of 38.86 items (SD = 2.42). Therefore, a matrix design consisting of 90 unique booklets with a subset of 44 items in random order was used. Overall, 26 gap-level features, 4 context features, and 6 item-level features that had been identified in the preceding analyses of the exercise items were tested in experiments to predict difficulty – with the test data producing 26,772 data points. Difficulty calibration was performed based on a Rasch model. To model CGFI difficulty, several ridge regression models with different CGFI feature sets were built, using the scikit learn library for Python.

The research project presented here and the results obtained indicate the steps being taken towards the development of the domain and adaptation model of an intelligent, adaptive learning system, targeting English tenses and using CGFI as the main exercise format. Based on the difficulty model developed in this study, an algorithm can then be written that anticipates the degree of difficulty and weighs the difficulty-influencing features of the exercise items, relates them to the learner's individual skills (e.g. based on the analysis of mistakes made), estimates the probability that a learner will solve an exercise item correctly, and selects and sequences exercise items adapted to this individual learner's ability. The system could then also anticipate where additional scaffolding is necessary (e.g. more or less detailed explanations and hints) and select the kind of feedback or feedforward that perfectly suits the learner's needs. A future, desirable development step would then be to train the system to identify relevant features in new exercise items and rate and predict difficulty automatically. Consequently, the system would also be able to generate new exercise items or recommend exercise items with feature constellations specifically tailored to the learner's current needs. However, these steps and the complete technical realization of the project are still pending for this research and development project. This study illustrates how a plethora of data must be generated and carefully analyzed, even for such a small-scale project on specifically targeted grammatical forms of a particular level, to construct parts of the domain model of an ILTS.

5. Project 2: Interact4School

Interact4School, funded by the German Federal Ministry of Education and Research (BMBF) from 2020-2023, is an interdisciplinary research and development project between the English department at the Leuphana University in Lüneburg, the computer linguistics and educational departments at the University of Tübingen, the learning and teaching with digital media research group at the Technical University in Dortmund

and multiple cooperating teachers in various schools. In comparison to the project described in the preceding section, this project goes one step further and investigates the conditions, potentials and challenges of integrating focus on form practice (with an ILTS) in meaning-based, task-based, communicative foreign language teaching and learning. One main focus of this project is the further technical and functional development (feedback and scaffolding, user experience, motivational elements, gamification) of *FeedBook*, a fully equipped ILTS (including learner and exercise modelling) developed at the University of Tübingen (Meurers et al. 2018; 2019a; 2019b). Here, an entire printed workbook, accompanying a German EFL coursebook with all kinds of exercise types (from closed to more open formats, from grammar and vocabulary to reading and listening comprehension), was turned into an interactive digital learning environment, using methods and principles of AI. While most EFL websites and learning apps mainly consist of simple true/false feedback mechanisms, which do not adequately reflect the realization potential of natural language processing potential, *FeedBook* uses NLP to offer versatile and individual feedback (188 different feedback types) and scaffolding based on a large data-set, sophisticated input analysis and anticipation of learner responses, and substantial linguistic and curricular expertise (Figure 3). First analyses clearly show that students who received specific, formative feedback from the *FeedBook* system benefitted significantly more from using the program than students who did not receive specific feedback (Meurers at al. 2019a).

Interact4School is now primarily concerned with the synergetic connection of this digitally supported, differentiated focus on form (FoF) practice with task-oriented, communicative foreign language teaching, with the teacher and the entire learning group. For this purpose, four communicative, task-oriented EFL teaching units, including all digital exercises and pre-communicative pre-activities leading to the target task, were developed. While digital technology plays a primary role in the FoF practice phase, it is subordinate to communicative activities in face-to-face phases. Using a mixed-methods approach, the research team uses language testing, classroom videography, questionnaires, interviews with learners and teachers, as well as usage data provided by the software to explore, in particular, the conditions and effects of adaptive, individualized practice with intelligent feedback and scaffolding on classroom-based communicative task performance and language learning processes in general. The study combines a quantitative study of approximately 30 classrooms in Baden-Württemberg and North Rhine-Westphalia with a qualitative analysis of two classrooms in Hamburg. First results are expected for spring 2022.

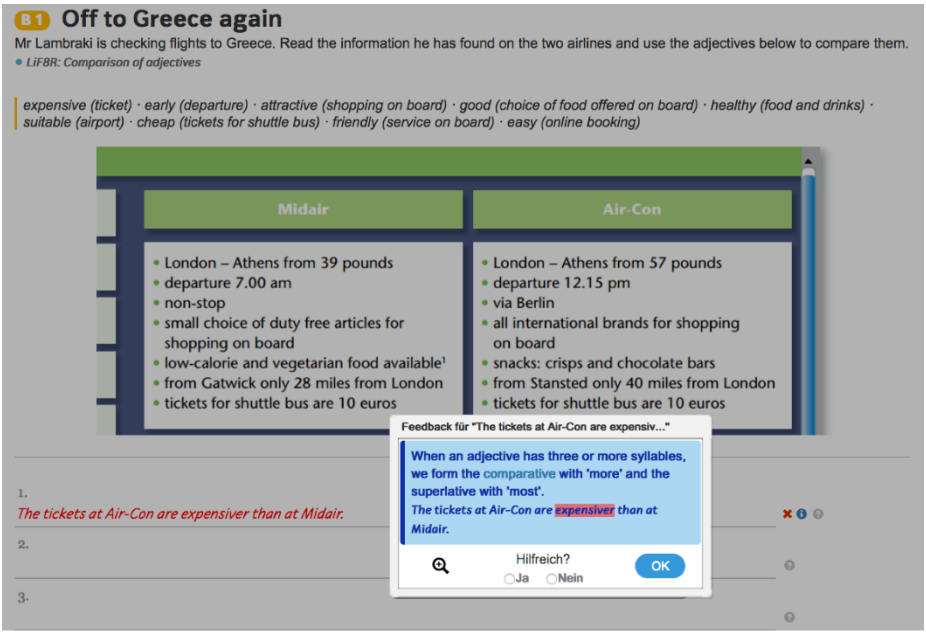

Figure 3: Example of feedback in *FeedBook* system

6. A Utopia: The AI-enhanced Foreign Language Classroom in 2040

In 2040, the foreign language classroom will be an evidence-based blended learning environment with bring-your-own device solutions in all classrooms and subjects and high-speed Internet connections, as well as standardized content management systems and hybrid course books and learning materials that form the infrastructural backbone. Learners from different backgrounds will have equal opportunities to participate. In 2040, the printed textbook will have become an interactive, multimedia, adaptive learning and practice environment that is perfectly adapted to the needs of both face-to-face teaching and the phases of individual practice and self-directed learning. Experts in the fields of foreign language pedagogy, computational linguistics, learning psychology, machine learning and Big Data, as well as multimedia designers, have been working closely together to create this digital learning environment. The digitally enhanced classroom of the year 2040 will help to diagnose needs and learning progress, providing direct access to differentiated and needs-based support services. The learning platform deployed in 2040 is thus an AI-based digital learning support system, a resource, a tool for students and teachers, which will be used to add value to the learning process. The classroom of the future will cleverly combine the advantages of digital

learning with proven, computer-free methods, content and tasks for face-to-face teaching, which will remain indispensable and highly significant for successful learning. And even in 2040, we will still have and need teachers: well-trained, data-literate teachers who are competent, critical, and reflective in their use of media and technology support, and who use empirically established digital scenarios that are based on valuable learning content and are linked to meaningfully constructed, challenging learning tasks and exercise opportunities designed to support individual learning. A sustainable approach to make teachers become acquainted with AI technology within a rather non-technocratic narrative would probably "involve a hybrid transition period until the point is reached where people felt confident in the replacement of a truer assessment of an individual's learning and strengths" (Haddad 2021, 26).

Works Cited

Aljaafreh, Ali, and James P. Lantolf. "Negative feedback as regulation and second language learning in the zone of proximal development." *The Modern Language Journal* 78.4 (1994): 465-483.

Alpaydin, Ethem. *Introduction to Machine Learning*. Cambridge, MA: The MIT Press, 2014.

Amaral, Luiz, and Detmar Meurers. "On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning." *ReCALL* 23.1 (2011): 4-24. DOI: 10.1017/S0958344010000261.

Baker, Toby, and Laurie Smith. *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Nesta, 2019. <https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf> [accessed 25 March 2021].

Banerjee, Dibyendu. "Natural Language Processing (NLP) Simplified: A Step-by-step Guide." *Data Science Foundation* (2020). <https://datascience.foundation/sciencewhitepaper/natural-language-processing-nlp-simplified-a-step-by-step-guide> [accessed 18 March 2021].

Bär, Marcus. "Vom Üben als notwendigem Übel zum funktionalen und intelligenten Üben." *Üben und Übungen beim Fremdsprachenlernen. Perspektiven und Konzepte für Unterricht und Forschung.* Eds. Eva Burwitz-Melzer, Frank Königs, Claudia Riemer, and Lars Schmelter. Tübingen: Narr, 2016. 9-18.

Berendt, Bettina, Allison Littlejohn, and Mike Blakemore. "AI in education: Learner choice and fundamental rights." *Learning, Media and Technology* 45.3 (2020): 312-324. DOI: 10.1080/17439884.2020.1786399.

Blell, Gabriele. "Üben im Fremdsprachenunterricht: Wenn das Üben unmerklich in den Unterricht integriert werden kann, so könnte es weiter bestehen…" *Üben und Übungen beim Fremdsprachenlernen. Perspektiven und Konzepte für Unterricht*

und Forschung. Eds. Eva Burwitz-Melzer, Frank Königs, Claudia Riemer, and Lars Schmelter. Tübingen: Narr, 2016.19-29.

Blume, Carolyn, Torben Schmidt, and Inke Schmidt. "An Imperfect Union? Enacting a Heuristic for Digital Games for Language Learning." *Zeitschrift für Fremdsprachenforschung* 28.2 (2017): 209-231.

Council of Europe. *Common European Framework of Reference for Languages. Learning, Teaching, Assessment*. Strasbourg: Council of Europe, 2001.

Cope, Bill, and Mary Kalantzis. "Big Data Comes to School: Implications for Learning, Assessment, and Research." *AERA Open* 2.2 (2016): 1-19.

Csíkszentmihályi, Mihaly. *Flow: The Psychology of Optimal Experience*. New York: Perennial Modern Classics, Harper & Row, 1990.

DeKeyser, Robert. "What Makes Learning Second-Language Grammar Difficult? A Review of Issues." *Language Learning* 55.1 (2005): 1-25. DOI: 10.1111/j.0023-8333.2005.00294.x.

Doughty, Catherine. "Cognitive underpinnings of focus on form." *Cognition and Second Language Instruction.* Ed. Peter Robinson. Cambridge: Cambridge University Press, 2001. 206-255.

Europarat/ Rat für kulturelle Zusammenarbeit. *Gemeinsamer Europäischer Referenzrahmen für Sprachen. Lernen, lehren, beurteilen*. München: Langenscheidt, 2001.

Funk, Hermann. "Was heißt Time on task? Oder: Warum übt man ausgerechnet im Fremdsprachenunterricht oft nicht das, was man am Ende können soll?" *Üben und Übungen beim Fremdsprachenlernen. Perspektiven und Konzepte für Unterricht und Forschung.* Eds. Eva Burwitz-Melzer, Frank Königs, Claudia Riemer, and Lars Schmelter. Tübingen: Narr, 2016. 62-76.

Grimm, Nancy, Michael Meyer, and Laurenz Volkmann. *Teaching English*. Tübingen: Narr Francke Attempto, 2015.

Haddad, Sam. "The complex role of AI in exam marking." *Digital Learning* 0723 (2021): 26.

Hallet, Wolfgang. *Lernen fördern: Englisch. Kompetenzorientierter Unterricht in der Sekundarstufe 1.* Seelze: Kallmeyer, 2011.

Hattie, John, and Helen Timperley. "The Power of Feedback." *Review of Educational Research* 77.1 (2007): 81-112.

Healey, Justin. *Artificial Intelligence (Volume 450)*. Thirroul: The Spinney Press, 2020. <https://library.camhigh.vic.edu.au/ais/downloadfile/Qj0xOTU1NDk1NTgmVT02Mjk3OQ==/Artificial%20Intelligence.pdf> [accessed 18 March 2021].

Heift, Trude. "Developing an Intelligent Language Tutor." *CALICO Journal* 27.3 (2010): 443-459.

Heymann, Hans Werner. "Was macht Üben intelligent?" *Pädagogik* 57.11 (2005): 6-11.

Kannan, Jayar, and Pilar Munday. "New Trends in Second Language Learning and Teaching through the lens of ICT, Networked Learning, and Artificial Intelligence." *Círculo de Lingüística Aplicada a La Comunicación* 76.0 (2018): 13-30. DOI: 10.5209/CLAC.62495.

Kieweg, Werner. "Das Üben üben. Übung, Wiederholung und Konsolidierung von Wissen und Können im Fremdsprachenunterricht." *Der fremdsprachliche Unterricht. Englisch* 48.131 (2014): 2-8.

Kleppin, Karin. *Fehler und Fehlerkorrektur*. Berlin: Langenscheidt, 2008.

Klippel, Friederike. "Systematisches Üben." *Englisch lernen und lehren. Didaktik des Englischunterrichts*. Ed. Johannes-Peter Timm. Berlin: Cornelsen, 1998. 328-341.

Kulik, James A., and J.D. Fletcher. "Effectiveness of Intelligent Tutoring Systems." *Review of Educational Research* 86.1 (2016): 42-78.

Lightbown, Patsy M. "The importance of timing in focus on form." *Focus on Form in Classroom Second Language Acquisition*. Eds. Catherine Doughty and Jessica Williams. Cambridge: Cambridge University Press, 1998. 177-196.

Lisovets, Volodymyr. "Essentials of Artificial Intelligence for Language Learning." *Intellias – Intelligent Software Engineering*, 2017. <https://www.intellias.com/how-ai-helps-crack-a-new-language/> [accessed 18 March 2021].

Luckin, Rose, Wayne Holmes, Mark Griffiths, and Laurie B. Forcier. *Intelligence unleashed: An argument for AI in education*. London: Pearson, 2016. <https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf> [accessed 25 March 2021].

MARKETSANDMARKETS. "AI in Education Market by Technology." *Markets and Markets*, 2018. <https://www.marketsandmarkets.com/Market-Reports/ai-in-education-market200371366.html?gclid=Cj0KCQiArvX_BRCyARIsAKsnTxMJisM_vhF9ZxJhIy-JP-GNvR-P5niupBb2l0tUrvHFK1M7LbX6rbIaAowdEALw_wcB> [accessed 18 March 2021].

Merkert, Johannes. "Maschinelle Übersetzer: DeepL macht Google Translate Konkurrenz." *heise online*, 2017. <https://www.heise.de/newsticker/meldung/Maschinelle-Uebersetzer-DeepL-macht-Google-Translate-Konkurrenz-3813882.html> [accessed 22 March 2021].

Mertens, Jürgen. "Aufgabenorientiertes Lernen." *Metzler Lexikon Fremdsprachendidaktik*. Ed. Carola Surkamp. Stuttgart: Metzler, 2010. 7-9.

Meurers, Detmar. "Natural Language Processing and Language Learning." *The Concise Encyclopedia of Applied Linguistics*. Ed. Carol A. Chapell. Chichester: Wiley, 2020. 817-831. <http://www.sfs.uni-tuebingen.de/~dm/papers/Meurers-19.pdf> [accessed 18 March 2021].

Meurers, Detmar, Kordula De Kuthy, Verena Möller, Florian Nuxoll, Björn Rudzewitz, and Ramon Ziai. "Digitale Differenzierung benötigt Informationen zu Sprache, Aufgabe und Lerner zur Generierung von individuellem Feedback in einem

interaktiven Arbeitsheft." *Fremdsprachen Lehren und Lernen* 47.2. (2018): 64-81. <http://www.sfs.uni-tuebingen.de/~dm/papers/Meurers.DeKuthy.ea-18.pdf> [accessed 18 March 2021].

Meurers, Detmar, Kordula De Kuthy, Florian Nuxoll, Björn Rudzewitz, and Ramon Ziai. "Scaling up intervention studies to investigate real-life foreign language learning in school." *Annual Review of Applied Linguistics* 39 (2019a): 161-188.

Meurers, Detmar, Kordula De Kuthy, Florian Nuxoll, Björn Rudzewitz, and Ramon Ziai. "KI zur Lösung realer Schulherausforderungen: Interaktive und adaptive Materialien im Fach Englisch." *Schulmanagement-Handbuch*. Ed. Thomas Rieke-Baulcke. Berlin: Cornelsen, 2019b. 65-84.

Müller-Hartmann, Andreas, and Marita Schocker-von Ditfurth. "Task-based Language Teaching und Task-supported Language Teaching." *Handbuch Fremdsprachendidaktik*. Eds. Wolfgang Hallet and Frank G. Königs. Seelze-Velber: Kallmeyer, Klett, 2010. 203-206.

Nagata, Noriko. "Robo-Sensei's NLP-Based Error Detection and Feedback Generation." *CALICO Journal* 26.3 (2009): 562-579. <https://journal.equinoxpub.com/Calico /article/view/17297/19753> [accessed 18 March 2021].

Nunan, David. *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press, 1989.

Paiva, Ranilson, and Ig Ibert Bittencourt. "Helping Teachers Help Their Students: A Human-AI Hybrid Approach." *Artificial Intelligence in Education*. Eds. Ig Ibert Bittencourt, Mutlu Cukurova, Kasia Muldner, Rose Luckin, and Eva Millán. Cham: Springer Switzerland International Publishing, 2020. 448-459. DOI: 10.1007/978-3-030-52237-7_36.

Pandarova, Irina, Torben Schmidt, Johannes Hartig, Ahcène Boubekki, Roger Dale Jones, and Ulf Brefeld. "Predicting the Difficulty of Exercise Items for Dynamic Difficulty Adaptation in Adaptive Language Tutoring." *International Journal of Artificial Intelligence in Education* 29 (2019): 342-367. DOI: 10.1007/s40593-019-00180-4.

Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." *Neural Networks* 61.117 (2015): 85-117. DOI: 10.1016/j.neunet.2014.09.003.

Schmidt, Torben. "Chocolate-covered Drill and Practice? Möglichkeiten und Grenzen des 'gamifizierten', adaptiven Übens in Fremdsprachenlern-Apps." Eds. Eva Burwitz-Melzer, Frank G. Königs, Claudia Riemer, and Lars Schmelter. Tübingen: Narr Francke Attempto Verlag, 2016. 200-210.

Schmidt, Torben, and Thomas Strasser. "Media-Assisted Foreign Language Learning – Concepts and Functions." *Teaching English as a Foreign Language: An Introduction*. Eds. Carola Surkamp and Britta Viebrock. Stuttgart: Metzler, 2018. 211-231.

Seufert, Sabine, Josef Guggemos, and Eric Tarantini "Digitale Transformation in Schulen. Kompetenzanforderungen an Lehrpersonen." *Beiträge zur Lehrerinnen- und Lehrerbildung* 36.2 (2018): 175-193.

Siebold, Jörg. "Aufgabe/Task und Übung/Exercise." *Praxis Fremdsprachenunterricht* 4.4 (2007): 63-64.

Strasser, Thomas. "AI in the EFL-classroom. Clarifications, potentials and limitations." *Digital Teaching and Learning: Perspectives for English Language Education.* Eds. Christiane Lütge and Thorsten Merse. Tübingen: Narr Francke Attempto, 2021. 85-102.

Swain, Merrill. "The output hypothesis: Theory and research." *Handbook on Research in Second Language Teaching and Learning*. Ed. Eli Hinkel. Mahwah, NJ: Lawrence Erlbaum Associates, 2005. 471-484.

Vygotsky, Lev S. *Thought and Language.* Cambridge, MA: MIT Press, 1986.

Willis, Jane. *A Framework for Task-Based Learning*. London: Longman, 1996.