N

Natural Language Processing and Language Learning

DETMAR MEURERS

As a relatively young field of research and development started by work on cryptanalysis and machine translation around 50 years ago, natural language processing (NLP) is concerned with the automated processing of human language. It addresses the analysis and generation of written and spoken language, although speech processing is often regarded as a separate subfield. NLP emphasizes processing and applications and as such 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. In terms of the language aspects dealt with in NLP, traditionally lexical, morphological, and syntactic aspects of language were the center of attention, but aspects of meaning, discourse, and the relation to the extralinguistic context have become increasingly prominent in the last decade. A good introduction and overview of the field is provided in Jurafsky and Martin (2009).

This entry explores the relevance and uses of NLP for language learning, focusing on written language. (For a recent overview of technology targeting pronunciation, see Pennington & Rogerson-Revell, 2019, chap. 5.) The entry focuses on showing the relevance, characterizing the techniques, and delineating the uses of NLP for language learning. More historical background and discussion can be found in Nerbonne (2003), Heift and Schulze (2007, 2015), and Heift (2017).

Two broad uses of NLP related to language learning are described. On the one hand, NLP can be used for the *analysis of learner language*, that is, words, sentences, or texts produced by language learners. This use includes the development of NLP techniques for the analysis of learner language by tutoring systems in intelligent computer-assisted language learning (ICALL; see Heift, 2017), automated scoring in language testing, as well as the analysis and annotation of learner corpora. On the other hand, NLP for the *analysis of native language* can also play an important role in the language learning context. Applications in this second domain support the search for and the enhanced presentation of native language reading material for language learners as well as the generation of exercises and tests based on authentic materials.

NLP and the Analysis of Learner Language

Intelligent Language Tutoring Systems

Intelligent language tutoring systems (ILTSs) use NLP to provide individualized feedback to learners working on activities, usually in the form of workbook-style exercises as in the E-Tutor (Heift, 2010), Robo-Sensei (Nagata, 2009), TAGARELA (Amaral & Meurers, 2011), the i-tutor (Choi, 2016), and the FeedBook (Meurers, Kuthy, Nuxoll, Rudzewitz, & Ziai, in press). The NLP analysis may also be used to individually adjust the sequencing of the material and to update the learner model (see Schulze, 2011). Typically, the focus of the analysis is on errors of linguistic form made by the learner, even though in principle feedback can also highlight correctly used forms, or it can target aspects of meaning or the appropriateness of a learner response given the input provided by an exercise.

What motivates the use of NLP in a tutoring system? To be able to provide feedback and keep track of the students' abilities in a learner model, an ILTS must obtain information about their abilities. How this can be done depends directly on the nature of the activities and the learner responses they support, that is, the ways in which they require the learner to produce language or interact with the system. The interaction between the activity or task given to a learner and the learner response is an important topic in language assessment (see Bachman & Palmer, 1996) and task-based language teaching and learning (Ellis, 2009), and it arguably is crucial for determining the system analysis requirements of different activity types (Quixal & Meurers, 2016).

For exercises that explicitly or implicitly require the learner to provide responses using forms from a small, predefined set, it often is possible to anticipate all potential well-formed and ill-formed learner responses, or at least the most common ones given a particular learner population. The intended system feedback for each case can then be explicitly specified for each potential response. In such a setup, knowledge about language and learners is exclusively expressed in this extensional mapping provided offline when the exercise is created. Sometimes targets are allowed to include regular expressions (http://en.wikipedia.org/wiki/Regular_expression) to support a more compact specification. The online process of comparing an actual learner response to the anticipated targets is a simple string comparison requiring no linguistic knowledge and thus no NLP. Correspondingly, the quiz options of general course management systems (Moodle, ILIAS, etc.) can support such language exercises just as they support quizzes for math, geography, or other subjects. The same is true of the general Web tools underlying the many quiz pages on the Internet (e.g., http://www.eslcafe.com/quiz/). Tools such as Hot Potatoes (http://hotpot.uvic.ca/) make it easier to specify some exercise forms commonly used in language teaching but also use the same general processing setup without NLP.

For many types of language learning activities, however, extensionally specifying a direct and complete mapping between potential learner input and intended feedback is not feasible. Nagata (2009, pp. 563–4) provides a clear illustration of this with an exercise taken from her Japanese tutor ROBO-SENSEI in which the learner reads a short communicative context and is asked to produce a sentence in Japanese that is provided in English by the system. The learner response in this exercise is directly dependent on the input provided by the exercise (a *direct response* in the terminology of Bachman & Palmer, 1996), so that a short, seven-word sentence can be defined as target answer. Yet after considering possible well-formed lexical, orthographic, and word order variants, one already obtains 6,048 correct sentences which could be entered by the learner. Considering incorrect options, even if one restricts ill-formed patterns to wrong particle and conjugation choices, one obtains almost a million sentences. Explicitly specifying a mapping between a million anticipated responses and their corresponding feedback clearly is infeasible. Note that the explosion of

possible learner answers illustrated by Nagata already is a problem for the direct response in a constrained activity, where the meaning to be expressed was fixed and only form variation was anticipated. A further dimension of potential variation in responses arises when going beyond the analysis of language as a system (parallel to *system-referenced* tests in language assessment; see Baker, 1989) to an analysis of the ability to use language to appropriately complete a given task (*performance-referenced*).

In conclusion, for the wide range of language activities supporting significant well-formed or ill-formed variation of form, meaning, or task-appropriateness of learner responses, it is necessary to abstract away from the specific string entered by the learner to more general classes of properties by automatically analyzing the learner input using NLP algorithms and resources. Generation of feedback, learner model updating, and instructional sequencing can then be based on the small number of language properties and categories derived through NLP analysis instead of on the large number of string instances they denote.

What is the nature of the learner language properties to be identified? Research on intelligent language tutors has traditionally focused on *learner errors*, identifying and providing feedback to learners on their errors. Automatic analysis can also be used in an ILTS to identify well-formed language properties to be able to provide positive feedback or record in a learner model that a given response provided evidence for the correct realization of a particular construction, lexical usage, or syntactic relation. All approaches to detecting and diagnosing learner errors must explicitly or implicitly model the space of well-formed and ill-formed variation that is possible given a particular activity and a given learner. Insights into activity design and the language development of learners thus are crucial for effective NLP analysis of learner errors.

Errors are also present in native language texts, and the need to develop *robust NLP* which works in suboptimal conditions (due to unknown or unexpected forms and patterns, noise) has been a driving force behind the shift from theory-driven, rule-based NLP in the 1980s and 1990s to the now dominant data-driven, statistical, and machine-learning approaches. However, there is an important difference in the goal of the NLP use in an ILTS compared to that in other NLP domains. NLP is made robust to gloss over errors and unexpected aspects of the system input with the goal of producing a result, such as a syntactic analysis returned by a parser, or a translation provided by a machine translation system. The traditional goal of the NLP in an ILTS, on the other hand, is to identify the characteristics of learner language and the features of learner responses that diverge from the expected targets in order to provide feedback to the learner. In this case, error characterization is the goal of the abstraction performed by the NLP; errors are not something to be glossed over by robustness of processing.

Writers' aids such as the standard spell and grammar checkers (Dickinson, 2006) share the ILTS focus on identifying errors, but they rely on assumptions about typical errors made by native speakers which do not carry over to language learners. For example, Rimrott and Heift (2008) observe that, "in contrast to most misspellings by native writers, many L2 misspellings are multiple-edit errors and are thus not corrected by a spell checker designed for native writers." Tschichold (1999) also points out that traditional writers' aids are not necessarily helpful for language learners since learners need more scaffolding than a list of alternatives from which to choose. Writers' aids tools targeting language learners, such as the ESL Assistant (Gamon et al., 2009), therefore provide more feedback and, for example, concordance views of alternatives to support the language learner in understanding the alternatives and choosing the right one. The goal of writers' aids is to support the second language user in writing a functional, well-formed text, not to support them in acquiring the language, as is the goal of an ILTS. Where writing well-formed and well-structured texts is the goal, advanced learners can also benefit from the quickly developing market for

automatic writing evaluation tools, such as the following: https://writeandimprove.com, http://noredink.com, http://grammarly.com, or http://criterion.ets.org.

NLP methods for the diagnosis of learner errors fall into two general classes. On the one hand, most of the traditional development has gone into language licensing approaches, which analyze the entire learner response. On the other hand, there is a growing number of pattern-matching approaches which target specific error patterns and types (e.g., preposition or determiner errors), ignoring any learner response or part thereof that does not fit the pattern.

Language licensing approaches are based on formal grammars of the language to be licensed, which can be expressed in one of two general ways (see Johnson, 1994). In a *validity-based* setup, a grammar is a set of rules, and recognizing a string amounts to finding valid derivations. Simply put, the more rules that are added to the grammar, the more types of strings can be licensed; if there are no rules, nothing can be licensed. In a *satisfiability-based* setup, a grammar is a set of constraints and a string is licensed if its model satisfies all constraints in the grammar. Thus, the more constraints that are added, the fewer types of strings are licensed; if there are no constraints in the grammar, any string is licensed.

Corresponding to these two types of formal grammars, there are essentially two types of approaches to analyzing a string with the goal of diagnosing learner errors. The *mal-rule approach* follows the validity-based perspective and uses standard parsing algorithms. Starting with a standard native language grammar, rules are added to license strings which are used by language learners, but not in the native language, that is, so-called *mal*-rules used to license learner errors (see, e.g., Sleeman, 1982; Matthews, 1992, and references therein). Given that a specific error type can manifest itself in a large number of rules—for example, an error in subject–verb agreement can appear in all rules realizing subjects together with a finite verb—*meta-rules* can be used to capture generalizations over rules (Weischedel & Sondheimer, 1983). The *mal*-rule approach can work well when errors correspond to the local tree licensed by a single grammar rule. Otherwise, the interaction of multiple rules must be taken into account, which makes it significantly more difficult to identify an error and to control the interaction of *mal*-rules with regular rules. To reduce the search space resulting from rule interaction, the use of *mal*-rules can be limited. In the simple case, the *mal*-rules are only added after an analysis using the regular grammar fails. Yet this only reduces the search space for parsing well-formed strings; if parsing fails, the question remains which *mal*-rules need to be added.

The ICICLE system (Michaud & McCoy, 2004) presents an interesting solution by selecting the groups of rules to be used based on *learner modeling*. It parses using the native rule set for all structures which the learner has shown mastery of. For structures assumed to currently be acquired by the learner, both the native and the *mal*-rules are used. And for structures beyond the developmental level of the learner, neither regular nor *mal*-rules are included. When moving from traditional parsing to parsing with probabilistic grammars, one obtains a further option for distinguishing native from learner structures by inspecting the probabilities associated with the rules (see Wagner & Foster, 2009, and references therein).

Different from such statistical approaches based on rules and the insights they capture, the most recent research on grammatical error correction (GEC) mostly focuses on detecting and correcting errors in written text as a type of translation problem. The idea is to map ill-formed to well-formed text directly using statistical machine translation methods (Junczys-Dowmunt & Grundkiewicz, 2016) or current neural network approaches (Chollampatt & Ng, 2018), which is the quantitative state of the art for GEC, but very limited in relevance for research and applications for which linguistic rules and (mis)conceptualization play a role.

The second group of language licensing approaches is typically referred to as *constraint relaxation* (Kwasny & Sondheimer, 1981), which is an option in a satisfiability-based

grammar setup or when using rule-based grammars with complex categories related through unification or other constraints which can be relaxed. When parsing is treated as a general constraint satisfaction problem, a general purpose conflict detection algorithm can be used to diagnose learner errors (Boyd, 2012). The idea of constraint relaxation is to eliminate certain constraints from the grammar: for example, specifications ensuring agreement, thereby allowing the grammar to license more strings than before. This assumes that an error can be mapped to a particular constraint to be relaxed—that is, the domain of the learner error and that of the constraint in the grammar must correspond closely. Instead of eliminating constraints outright, constraints can also be associated with weights controlling the likelihood of an analysis (Foth, Menzel, & Schröder, 2005), which raises the interesting issue of how such flexible control can be informed by the ranking of errors likely to occur for a particular learner given a particular task. Other proposals combine constraint relaxation with aspects of *mal*-rules. Reuer (2003) combines a constraint relaxation technique with a standard parsing algorithm modified to license strings in which words have been inserted or omitted, an idea which essentially moves generalizations over rules in the spirit of meta-rules into the parsing algorithm.

For *pattern matching*, the most common approach is to match a typical error pattern, such as the pattern looking for *of cause* in place of *of course*. By performing the pattern matching on the part-of-speech tagged, chunked, and sentence delimited learner string, one can also specify error patterns such as that of a singular noun immediately preceding a plural finite verb (e.g., in *The baseball team are established*.). This approach is commonly used in standard grammar checkers and, for example, realized in the open source LanguageTool (Naber, 2003; http://www.languagetool.org), which was not developed with language learners in mind, but readily supports the specification of error patterns that are typical for particular learner populations.

Alternatively, pattern matching can also be used to identify *context patterns* that are likely to include errors. For example, determiners are a well-known problem area for certain learners of English. By using a pattern that identifies all nouns (or all noun chunks) in the learner response, one can then make a prediction about the correct determiner to use for this noun in its context and compare this prediction to the determiner use (or lack thereof) in the learner response. Given that determiners and preposition errors are among the most common English learner errors found and that the task lends itself well to the current machine learning approaches in computational linguistics (and raises the interesting general question of how much context and which linguistic generalizations are needed for predicting such functional elements), these error types have received particular attention (see, e.g., De Felice, 2008, and references therein).

Complementing the analysis of form, for an ILTS to offer meaning-based, contextualized activities it is important to provide an *automatic analysis of meaning* aspect: for example, to determine whether the answer given by the learner for a reading comprehension question makes sense given the reading. Although most NLP work in ILTSs has addressed form issues, some work has addressed the analysis of meaning (Delmonte, 2003; Ramsay & Mirzaiean, 2005; Bailey & Meurers, 2008) and the issue is directly related to work in computer-assisted assessment systems outside of language learning: for example, for evaluating the answers of short answer questions (see Pérez Marin, 2007; Ziai, 2018, and references therein) or in essay scoring (Shermis & Burstein, 2013). In terms of NLP methods, it also directly connects to the growing body of research on recognizing textual entailment and paraphrase recognition (Androutsopoulos & Malakasiotis, 2010; Dzikovska et al., 2013).

Shifting the focus from analysis techniques to the interpretation of the analysis, researchers need to optimize the results in view of the specific needs of the context. Like the activity type and the learner that play an important role in defining the space of variation to

be dealt with by the NLP, the interpretation and feedback provided to the learner needs to be informed by activity and learner modeling (Amaral & Meurers, 2008). While feedback in human–computer interaction cannot simply be equated with that in human–human interaction, the results presented by Petersen (2010) for a dialogue-based ILTS indicate that results from the significant body of research on the effectiveness of different types of feedback in instructed SLA (second language acquisition) can transfer across modes, providing fresh momentum for research on feedback in the CALL (computer-assisted language learning) domain (e.g., Pujolà, 2001).

A final important issue arising from the use of NLP in ILTSs concerns the resulting lack of teacher autonomy in materials development. Quixal, Preuß, García-Narbona, and Boullosa (2010) explore putting the teacher back in charge of designing their activities with the help of an *ICALL authoring system*—a complex undertaking since in contrast to regular CALL authoring software, NLP analysis of learner language needs to be integrated without presupposing teachers' understanding of the capabilities and limits of the NLP. Nevertheless, authoring systems are an important vehicle for moving these technologies from laboratories to the learning environments where they can be used by language learners.

Learner Corpora

Little interaction between ILTSs and learner corpus research is visible in the research in this area. Their independence is perhaps because ILTSs traditionally have focused on exercises requiring limited language production from students whereas most learner corpora consist of essays of much greater length. Nevertheless, the analysis of learner language in the annotation of learner corpora can be seen as an offline version of the online analysis performed by an ILTS (Meurers, 2015).

What motivates the use of NLP for learner corpora? In contrast to the automatic analysis of learner language in an ILTS providing feedback to the learner, the annotation of learner corpora essentially provides an index to learner language properties in support of the goal of advancing our understanding of acquisition in SLA research and to develop instructional methods and materials for foreign language teaching. Corpus annotation can support a more direct mapping from theoretical research questions to corpus instances (Meurers, 2005). For a reliable mapping between theory and data, it is crucial for corpus annotation to provide only reliable distinctions which are replicably based on the evidence found in the corpus and its meta-information, for which clear measures are available (Artstein & Poesio, 2009).

What is the nature of the learner language properties to be identified? Just as for ILTSs, much of the work on annotating learner corpora has traditionally focused on learner errors, for which a number of *error annotation schemes* have been developed (see Díaz Negrillo & Fernández Domínguez, 2006, and references therein). So far, there is no consensus on the external and internal criteria—that is, which error distinctions are needed for which purpose and which distinctions can reliably be annotated based on the evidence in the corpus and any meta-information available about the learner and the activity for which the language was produced. An explicit and reliable error annotation scheme and a gold standard reference corpus exemplifying it is an important next step for the development of automatic error annotation approaches, which need an agreed upon gold standard for development as well as for testing and comparison of approaches. Automating the currently manual error annotation process using NLP would support the annotation of significantly larger learner corpora and thus increase their usefulness for SLA research and teaching.

Since error annotation results from the annotator's comparison of a learner response to hypotheses about what the learner was trying to say, Lüdeling, Walter, Kroymann, and Adolphs (2005) argue for making the *target hypotheses* explicit in the annotation. Such target

specification would also make it possible to specify alternative error annotations for the same sentence based on different target hypotheses in multilevel corpus annotation. However, Fitzpatrick and Seegmiller (2004) report unsatisfactory levels of inter-annotator agreement in determining target hypotheses. It is an open research issue to determine for which type of learner responses written by which type of learners for which type of tasks such target hypotheses can reliably be determined. Target hypotheses might have to be limited to encoding only the minimal commitments necessary for error identification; and in place of target hypothesis strings, they might have to be formulated at more abstract levels: for example, lemmas in topological fields, or sets of concepts the learner was trying to express. In either case, if the target hypotheses are made explicit, the second step from target hypothesis to error identification can be studied separately and can be realized more reliably (Rosen, Hana, Štindlová, & Feldman, 2014).

Returning to the general question about the nature of the learner language properties which are relevant, SLA research essentially observes correlations among linguistic properties, whether erroneous or not. And even research focusing on learner errors needs to identify correlations with linguistic properties—to identify overuse/underuse of certain patterns, or measures of language development, for example. While the use of NLP tools trained on the native language corpora is a useful starting point for providing a range of linguistic annotations, an important next step is to explore the creation of annotation schemes and methods capturing the *linguistic properties of learner language* (see Meurers & Dickinson, 2017, and references therein).

In terms of using NLP for providing *general measures of language development*, as captured by the triad complexity, accuracy, and fluency (CAF; Housen & Kuiken, 2009). The automatic analysis of linguistic complexity in learner language has been a particularly active area of research (see Lu, 2014; Kyle, 2016; Chen, 2018, and references therein). The elaborateness and variedness of language use can be identified at all levels of linguistic modeling, including morphology, lexicon, syntax, and discourse.

To identify specific patterns that are characteristic of language development, it often is necessary to linguistically annotate the data (Meurers & Dickinson, 2017), which for large corpus resources requires automatic analysis. For example, Hawkins and Buttery (2009) identify so-called criterial features distinguishing different proficiency levels on the basis of part-of-speech tagged and parsed portions of the Cambridge Learner Corpus, and Alexopoulou, Geertzen, Korhonen, and Meurers (2015) illustrate this with a study of relative clause development in the very large EFCamDat learner corpus (https://corpus.mml.cam.ac.uk/efcamdat2). Using the second release of that corpus, containing 1.2 million texts written by 175,000 learners, Alexopoulou, Michel, Murakami, and Meurers (2017) highlight that the valid interpretation of linguistic complexity analysis also requires taking into account the properties of the task for which the writing was produced.

Learner corpora are also systematically analyzed with NLP methods to identify crosslinguistic effects, such as the transfer of characteristics of one's native language to text written in a second language. NLP research in this area was popularized by a series of shared tasks on native language identification (Malmasi et al., 2017), with approaches exploring both shallow, surface-based, and deep linguistic features (Jarvis & Crossley, 2012; Meurers, Krivanek, & Bykh, 2014; Bich, 2017). While most work focused on non-native English writing, Malmasi and Dras (2017) provide a genuinely multilingual approach.

NLP and the Analysis of Native Language for Learners

The second domain in which NLP connects to language learning derives from the need to expose learners to authentic, native language and its properties and to give them

opportunities to interact with it. This area of research includes work on searching for and enhancing authentic texts to be read by learners as well as the automatic generation of activities and tests from such texts. In contrast to the ILTS side of ICALL research covered in the first part of this entry, the NLP in the applications under discussion here is used to process native language in authentic texts, hence referred to as ATICALL.

Most NLP research is developed and optimized for native language material and it is easier to obtain enough annotated language material to train the statistical models and machine learning approaches used in current research and development. Therefore, in principle a wide range of NLP tools with high-quality analysis is available—even though this does not preempt the question which language properties are relevant for ATICALL applications and whether those properties can be derived from the ones targeted by standard NLP.

What motivates the use of NLP in ATICALL? Compared to using prefabricated materials such as those presented in textbooks, NLP-enhanced searching for materials in resources such as large corpora or the Web makes it possible to provide on-demand, individualized access to up-to-date materials. It supports selecting and enhancing the presentation of texts depending on the background of a given learner, the specific contents of interest to them, and the language properties and forms of particular relevance given the sequencing of language materials appropriate for the learner's stage (see, e.g., Pienemann, 1998).

What is the nature of the learner language properties to be identified? Traditionally, the most prominent property used for selecting texts has been a general notion of readability, for which a number of *readability formulas* were developed (DuBay, 2004). The traditional measures are based on shallow, easy-to-count features, typically average sentence and word lengths, but current machine learning methods informed by a broader range of linguistic characteristics are substantially more accurate (see, e.g., Xia, Kochmar, & Briscoe, 2016; Crossley, Skalicky, Dascalu, McNamara, & Kyle, 2017; Weiss & Meurers, 2018), including some commercial systems (Nelson, Perfetti, Liben, & Liben, 2012). Interestingly, readability analysis substantially benefits from the integration of complexity features originally designed to measure second language development (Vajjala & Meurers, 2012). Work in the Coh-Metrix project emphasizes the importance of analyzing text cohesion and coherence and of taking a reader's cognitive aptitudes into account for making predictions about reading comprehension (McNamara, Graesser, McCarthy, & Cai, 2014). For languages other than English, morphological features become more prominent (Dell'Orletta, Montemagni, & Venturi, 2011; François & Fairon, 2012; Hancke, Vajjala, & Meurers, 2012), for which finite-state NLP approaches provide particularly rich information (Reynolds, 2016), also readily supporting exercise generation on that basis (Antonsen & Argese, 2018). For reading practice with a focus on vocabulary acquisition (Cobb, 2008), several projects have emphasized the relevance and impact of individual learner models (Heilman et al., 2010; Walmsley, 2015).

Going beyond readability, based on the insight from SLA research that awareness of language categories and forms is an important ingredient for successful second language acquisition (see Lightbown & Spada, 1999), a wide range of linguistic properties have been identified as relevant for language awareness, including morphological, syntactic, semantic, and pragmatic information (see Schmidt, 1995, p. 30). In response to this need, the FLAIR system (Chinkina & Meurers, 2016) supports linguistically aware Web searches, which makes it possible to systematically enrich the input of language learners with the kind of language patterns to be acquired next. By integrating automated linguistic complexity analysis, it also becomes possible to retrieve reading material in the zone of proximal development of a learner by matching the complexity of the material to that of text written by this learner (Chen & Meurers, 2019).

Complementing the question of how to obtain material for language learners, there are several strands of ATICALL applications which focus on the enhanced presentation of and

learner interaction with such materials. One group of NLP-based tools, such as COMPASS (Breidt & Feldweg, 1997), Glosser (Nerbonne, Dokter, & Smit, 1998) and Grim (Knutsson, Pargman, Eklundh, & Westlund, 2007), provides a *reading environment* in which texts in a foreign language can be read with quick access to dictionaries, morphological information, and concordances. The Alpheios project (http://alpheios.net) focuses on literature in Latin and ancient Greek, providing links between words and translations, access to online grammar reference, and a quiz mode asking the learner to identify which word corresponds to which translation.

Another strand of ATICALL research focuses on supporting language awareness by *automating input enhancement* (Sharwood Smith, 1993), that is, by realizing strategies which highlight the salience of particular language categories and forms. For instance, WERTi (Meurers et al., 2010) visually enhances Web pages and automatically generates activities for language patterns which are known to be difficult for learners of English, such as determiners and prepositions, phrasal verbs, the distinction between gerunds and *to*-infinitives, and *wh*-question formation. Complementing the visual input enhancement of forms, Chinkina and Meurers (2017) propose to automatically generate questions as a form of functionally driven input enhancement. One can view such automatic input enhancement as an enrichment of data-driven learning (DDL). Where DDL has been characterized as an "attempt to cut out the middleman [the teacher] as far as possible and to give the learner direct access to the data" (Boulton 2009, p. 82, citing Johns, 1994), in visual input enhancement the learner stays in control, but the NLP uses "teacher knowledge" about relevant and difficult language properties to make those more prominent and noticeable for the learner.

A final, prominent strand of NLP research in this domain addresses the *generation of exercises and tests*. Most of the work has targeted the automatic generation of multiple choice cloze tests for language assessment and vocabulary drills (see, e.g., Liu, Wang, Gao, & Huang, 2005; Sumita, Sugaya, & Yamamoto, 2005, and references therein). Issues involving NLP in this domain include the selection of seed sentences, the determination of appropriate blank positions, and the generation of good distractors. The VISL project (Bick, 2005) also includes a tool supporting the generation of automatic cloze exercises, which is part of an extensive NLP-based environment of games and corpus tools aimed at fostering linguistic awareness for dozens of languages (http://beta.visl.sdu.dk). Finally, the Task Generator for ESL (Toole & Heift, 2001) supports the creation of gap-filling and build-a-sentence exercises such as the ones found in an ILTS. The instructor provides a text, chooses from a list of learning objectives (e.g., plural, passive), and selects the exercise type to be generated. The Task Generator supports complex language patterns and provides formative feedback based on NLP analysis of learner responses, bringing us full circle to the research on ILTS we started with. Currently, the most advanced approach in this line of research is the Language Muse Activity Palette (Burstein & Sabatini, 2016).

In conclusion, the use of NLP in the context of learning language offers rich opportunities, both in terms of developing applications in support of language teaching and learning and in terms of supporting SLA research—even though NLP so far has only had limited impact on real-life language teaching and SLA. More interdisciplinary collaboration between SLA and NLP will be crucial for developing reliable annotation schemes and analysis techniques which identify the properties that are relevant and important for analyzing learner language and analyzing language for learners.

SEE ALSO: Automatic Speech Recognition; Computer-Assisted Language Learning Effectiveness Research; Computer-Assisted Pronunciation Teaching; Corpus Linguistics in Language Teaching; Innovation in Language Teaching and Learning; Learner Corpora; Mobile-Assisted Language Learning

References

Alexopoulou, T., Geertzen, J., Korhonen, A., & Meurers, D. (2015). Exploring big educational learner corpora for SLA research: Perspectives on relative clauses. *International Journal of Learner Corpus Research*, *1*(1), 96–129.

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning*, *67*, 181–209.

Amaral, L., & Meurers, D. (2008). From recording linguistic competence to supporting inferences about language acquisition in context: Extending the conceptualization of student models for intelligent computer-assisted language learning. *Computer-Assisted Language Learning*, *21*(4), 323–38.

Amaral, L., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, *23*(1), 4–24.

Androutsopoulos, I., & Malakasiotis, P. (2010). A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, *38*, 135–87.

Antonsen, L., & Argese, C. (2018). Using authentic texts for grammar exercises for a minority language. In *Proceedings of the 7th workshop on NLP for Computer Assisted Language Learning* (pp. 1–9). Stockholm, Sweden: LiU Electronic Press. Retrieved April 4, 2019 from https://aclweb.org/anthology/papers/W/W18/W18-7100/

Artstein, R., & Poesio, M. (2009). Survey article: Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 1–42.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, England: Oxford University Press.

Bailey, S., & Meurers, D. (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications (BEA)* (pp. 107–15). Columbus, Ohio: Association for Computational Linguistics. Retrieved April 4, 2019 from https://aclweb.org/anthology/W08-0913

Baker, D. (1989). *Language testing: A critical survey and practical guide*. London, England: Edward Arnold.

Bich, S. (2017). *"Is there choice in non-native voice?" Linguistic feature engineering and a variationist perspective in automatic native language identification* (PhD thesis). Eberhard-Karls Universität Tübingen, Germany. Retrieved April 4, 2019 from http://hdl.handle.net/10900/77443

Bick, E. (2005). Grammar for fun: IT-based grammar learning with VISL. In P. Juel (Ed.), *CALL for the Nordic languages* (*Copenhagen studies in language*, pp. 49–64). Copenhagen, Denmark: Samfundslitteratur. Retrieved April 4, 2019 from http://beta.visl.sdu.dk/pdf/CALL2004.pdf

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 81–106.

Boyd, A. (2012). *Detecting and diagnosing grammatical errors for beginning learners of German: From learner corpus annotation to constraint satisfaction problems* (PhD thesis). The Ohio State University. Retrieved April 4, 2019 from https://www.semanticscholar.org/paper/Detecting-and-Diagnosing-Grammatical-Errors-for-of-Boyd/f35eedf18536b697853d157108eabc393e43f469

Breidt, E., & Feldweg, H. (1997). Accessing foreign languages with COMPASS. *Machine Translation*, *12*(1–2), 153–74.

Burstein, J., & Sabatini, J. (2016). The Language Muse Activity Palette. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 275–80). New York, NY: Routledge.

Chen, X. (2018). *Automatic analysis of linguistic complexity and its application in language learning research* (PhD thesis). Eberhard Karls Universität Tübingen, Germany. Retrieved April 4, 2019 from http://hdl.handle.net/10900/85888

Chen, X., & Meurers, D. (2019). Linking text readability and learner proficiency using linguistic complexity feature vector distance. *Computer-Assisted Language Learning.* Advance online publication. doi: 10.1080/09588221.2018.1527358

Chinkina, M., & Meurers, D. (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)* (pp. 188–98). San Diego, CA: Association for Computational Linguistics. Retrieved April 4, 2019 from https://www.semanticscholar.org/paper/Linguistically-Aware-Information-Retrieval%3A-Input-Chinkina-Meurers/d1e6e66b181b5912c101a19d345e56a0c5c28bba

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)* (pp. 334–44). Copenhagen, Denmark: Association for Computational Linguistics. Retrieved April 4, 2019 from https://aclweb.org/anthology/W17-5038

Choi, I.-C. (2016). Efficacy of an ICALL tutoring system and process-oriented corrective feedback. *Computer Assisted Language Learning*, *29*(2), 334–64.

Chollampatt, S., & Ng, H. T. (2018). A multilayer convolutional encoder–decoder neural network for grammatical error correction. In *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)* (pp. 5755–62). New Orleans, LA: AAAI.

Cobb, T. (2008). Necessary or nice? The role of computers in L2 reading. In Z. Han & N. J. Anderson (Eds.), *L2 reading research and instruction* (pp. 144–72). Ann Arbor: University of Michigan Press.

Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, *54*(5–6), 340–59.

De Felice, R. (2008). *Automatic error detection in non-native English* (Ph.D. thesis). St Catherine's College, University of Oxford. Retrieved from https://ora.ox.ac.uk/objects/uuid:5192f0cb-6e4d-4730-bb54-a97a73d603ed/download_file?file_format=pdf&safe_filename=DeFelice_DPhilThesis.pdf&type_of_work=Thesis.

Dell'Orletta, F., Montemagni, S., & Venturi, G. (2011). READ-IT: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies* (pp. 73–83). Edinburgh, Scotland: Association for Computational Linguistics.

Delmonte, R. (2003). Linguistic knowledge and reasoning for error diagnosis and feedback generation. *CALICO Journal*, *20*(3), 513–32.

Díaz Negrillo, A., & Fernández Domínguez, J. (2006). Error tagging systems for learner corpora. *Revista Española de Lingüística Aplicada (RESLA)*, *19*, 83–102.

Dickinson, M. (2006). Writer's aids. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed.). Oxford, England: Elsevier.

DuBay, W. H. (2004). *The principles of readability*. Costa Mesa, CA: Impact Information. Retrieved April 4, 2019 from http://www.impact-information.com/impactinfo/readability02.pdf

Dzikovska, M., Nielsen, R., Brew, C., Leacock, C., Giampiccolo, D., Bentivogli, L., … & Dang, H. T. (2013). SemEval-2013 task 7: The joint student response analysis and 8th Recognizing Textual Entailment Challenge. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 263–74). Atlanta, GA: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/S13-2045

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–46.

Fitzpatrick, E., & Seegmiller, M. S. (2004). The Montclair electronic language database project. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective*. Amsterdam, Netherlands: Rodopi.

Foth, K., Menzel, W., & Schröder, I. (2005). Robust parsing with weighted constraints. *Natural Language Engineering*, *11*(1), 1–25.

François, T., & Fairon, C. (2012). An "AI readability" formula for French as a foreign language. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 466–77). Jeju Island, Korea: Association for Computational Linguistics. Retrieved April 4, 2019 from https://www.aclweb.org/anthology/D12-1043

Gamon, M., Leacock, C., Brockett, C., Dolan, W. B., Gao, J., Belenko, D., & Klementiev, A. (2009). Using statistical techniques and Web search to correct ESL errors. *CALICO Journal*, *26*(3), 491–511. Retrieved April 4, 2019 from http://klementiev.org/publications/calico09.pdf

Hancke, J., Vajjala, S., & Meurers, D. (2012). Readability classification for German using lexical, syntactic, and morphological features. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)* (pp. 1063–80). Mumbay, India: COLING. Retrieved April 4, 2019 from http://aclweb.org/anthology/C12-1065

Hawkins, J. A., & Buttery, P. (2009). Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. In *Studies in language testing: The social and educational impact of language assessment*. Cambridge, England: Cambridge University Press.

Heift, T. (2010). Developing an intelligent language tutor. *CALICO Journal*, *27*(3), 443–59.

Heift, T. (2017). History and key developments in intelligent computer-assisted language learning (ICALL). In S. L. Thorne & S. May (Eds.), *Language, education and technology* (3rd ed., pp. 289–312). Cham, Switzerland: Springer.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. New York, NY: Routledge.

Heift, T., & Schulze, M. (2015). Tutorial computer-assisted language learning. *Language Teaching*, *48*(4), 471–90.

Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi, M., Juffs, A., & Wilson, L. (2010). Personalization of reading passages improves vocabulary acquisition. *International Journal of Artificial Intelligence in Education*, *20*, 73–98.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461–73.

Jarvis, S., & Crossley, S. A. (Eds.). (2012). *Approaching language transfer through text classification: Explorations in the detection-based approach. Second language acquisition*. Bristol, England: Multilingual Matters.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293–313). Cambridge, England: Cambridge University Press.

Johnson, M. (1994). Two ways of formalizing grammars. *Linguistics and Philosophy*, *17*(3), 221–48.

Junczys-Dowmunt, M., & Grundkiewicz, R. (2016). Phrase-based machine translation is state-of-the-art for automatic grammatical error correction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1546–56). Austin, TX: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/D16-1161

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Knutsson, O., Pargman, T. C., Eklundh, K. S., & Westlund, S. (2007). Designing and developing a language environment for second language writers. *Computers & Education*, *49*(4), 1122–46.

Kwasny, S. C., & Sondheimer, N. K. (1981). Relaxation techniques for parsing grammatically ill-formed input in natural language understanding systems. *American Journal of Computational Linguistics*, *7*(2), 99–108.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (PhD thesis). Georgia State University. Retrieved April 4, 2019 from http://scholarworks.gsu.edu/alesl_diss/35

Lightbown, P. M., & Spada, N. (1999). *How languages are learned*. Oxford, England: Oxford University Press.

Liu, C.-L., Wang, C.-H., Gao, Z.-M., & Huang, S.-M. (2005). Applications of lexical information for algorithmically composing multiple-choice cloze items. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP* (pp. 1–8). Ann Arbor, MI: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/W05-0201

Lu, X. (2014). *Computational methods for corpus annotation and analysis*. New York, NY: Springer.

Lüdeling, A., Walter, M., Kroymann, E., & Adolphs, P. (2005). Multi-level error annotation in learner corpora. In *Proceedings of Corpus Linguistics*. Birmingham. Retrieved April 4, 2019 from https://pdfs.semanticscholar.org/c7e8/a9530d77ae1a85388f337cbaaba8b14a3e89.pdf

Malmasi, S., & Dras, M. (2017). Multilingual native language identification. *Natural Language Engineering*, *23*(2), 163–215.

Malmasi, S., Evanini, K., Cahill, A., Tetreault, J., Pugh, R., Hamill, C., … & Qian, Y. (2017). A report on the 2017 Native Language Identification Shared Task. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 62–75). Copenhagen, Denmark: Association for Computational Linguistics.

Matthews, C. (1992). Going AI: Foundations of ICALL. *Computer Assisted Language Learning*, *5*(1), 13–31.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, England: Cambridge University Press.

Meurers, D. (2005). On the use of electronic corpora for theoretical linguistics: Case studies from the syntax of German. *Lingua*, *115*(11), 1619–39.

Meurers, D. (2015). Learner corpora and natural language processing. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 537–66). Cambridge, England: Cambridge University Press.

Meurers, D., & Dickinson, M. (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, 67(2), 66–95.

Meurers, D., Krivanek, J., & Bykh, S. (2014). On the automatic analysis of learner corpora: Native language identification as experimental testbed of language modeling between surface features and linguistic abstraction. In A. Alcaraz Sintes & S. Valera (Eds.), *Diachrony and synchrony in English corpus studies* (pp. 285–314). Frankfurt am Main, Germany: Peter Lang.

Meurers, D., Kuthy, K. D., Nuxoll, F., Rudzewitz, B., & Ziai, R. (in press). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics 39*.

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)* (pp. 10–18). Los Angeles: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/W10-1002.pdf

Michaud, L. N., & McCoy, K. F. (2004). Empirical derivation of a sequence of user stereotypes for language learning. *User Modeling and User-Adapted Interaction*, *14*(4), 317–50.

Naber, D. (2003). *A rule-based style and grammar checker* (Master's thesis). Universität Bielefeld, Germany.

Nagata, N. (2009). Robo-Sensei's NLP-based error detection and feedback generation. *CALICO Journal*, *26*(3), 562–79.

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Technical Report, The Council of Chief State School Officers. Retrieved April 4, 2019 from https://achievethecore.org/content/upload/nelson_perfetti_liben_measures_of_text_difficulty_research_ela.pdf

Nerbonne, J. (2003). Natural language processing in computer-assisted language learning. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 670–698). Oxford, England: Oxford University Press.

Nerbonne, J., Dokter, D., & Smit, P. (1998). Morphological processing and computer-assisted language learning. *Computer Assisted Language Learning*, *11*(5), 543–59.

Pennington, M. C., & Rogerson-Revell, P. (2019). Using technology for pronunciation teaching, learning, and assessment. In M. C. Pennington & P. Rogerson-Revell (Eds.), *English pronunciation teaching and research* (pp. 235–86). London, England: Springer.

Pérez Marin, D. R. (2007). *Adaptive computer assisted assessment of free-text students' answers: An approach to automatically generate students' conceptual models* (PhD thesis). Universidad Autonoma de Madrid, Spain.

Petersen, K. (2010). *Implicit corrective feedback in computer-guided interaction: Does mode matter?* (PhD thesis). Georgetown University, Washington, DC.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam, Netherlands: John Benjamins.

Pujolà, J.-T. (2001). Did CALL feedback feed back? Researching learners' use of feedback. *ReCALL*, *13*(1), 79–98.

Quixal, M., & Meurers, D. (2016). How can writing tasks be characterized in a way serving pedagogical goals and automatic analysis needs? *CALICO Journal*, *33*, 19–48.

Quixal, M., Preuß, S., García-Narbona, D., & Boullosa, J. R. (2010). AutoLearn's authoring tool: A piece of cake for teachers. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 19–27). Los Angeles, CA: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/W10-1003.pdf

Ramsay, A., & Mirzaiean, V. (2005). Content-based support for Persian learners of English. *ReCALL*, *17*, 139–54.

Reuer, V. (2003). Error recognition and feedback with lexical functional grammar. *CALICO Journal*, *20*(3), 497–512.

Reynolds, R. (2016). *Russian natural language processing for computer-assisted language learning: Capturing the benefits of deep morphological analysis in real-life applications* (PhD thesis). UiT—The Arctic University of Norway. Retrieved April 4, 2019 from https://munin.uit.no/bitstream/handle/10037/9685/thesis.pdf

Rimrott, A., & Heift, T. (2008). Evaluating automatic detection of misspellings in German. *Language Learning and Technology*, *12*(3), 73–92.

Rosen, A., Hana, J., Štindlová, B., & Feldman, A. (2014). Evaluating and automating the annotation of a learner corpus. *Language Resources and Evaluation*, *48*(1), 65–92.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawai'i Press.

Schulze, M. (2011). Learner modeling in intelligent computer-assisted language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics: Technology and language*. Oxford, England: Wiley-Blackwell.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*, 165–79.

Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook on automated essay evaluation: Current, NY*. London, England: Routledge.

Sleeman, D. (1982). Inferring (mal) rules from pupil's protocols. In *Proceedings of the 5th European Conference on Artificial Intelligence (ECAI)* (pp. 160–4). Orsay, France.

Sumita, E., Sugaya, F., & Yamamoto, S. (2005). Measuring non-native speakers' proficiency of English by using a test with automatically-generated fill-in-the-blank questions. In *Proceedings of the Second Workshop on Building Educational Applications Using NLP* (pp. 61–8). Ann Arbor, MI: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/W05-0210

Toole, J., & Heift, T. (2001). Generating learning content for an intelligent language tutoring system. In *Proceedings of NLP-CALL Workshop at the 10th Int. Conf. on Artificial Intelligence in Education (AI-ED)* (pp. 1–8). San Antonio, TX.

Tschichold, C. (1999). Grammar checking for CALL: Strategies for improving foreign language grammar checkers. In K. Cameron (Ed.), *CALL: Media, design & applications*. Exton, PA: Swets & Zeitlinger.

Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)* (pp. 163–73). Montreal, Canada: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/W12-2019.pdf

Wagner, J., & Foster, J. (2009). The effect of correcting grammatical errors on parse probabilities. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)* (pp. 176–9). Paris, France: Association for Computational Linguistics. Retrieved April 4, 2019 from http://aclweb.org/anthology/W09-3827

Walmsley, M. (2015). *Learner modelling for individualised reading in a second language* (PhD thesis). The University of Waikato, New Zealand. Retrieved April 4, 2019 from http://hdl.handle.net/10289/10559

Weischedel, R. M., & Sondheimer, N. K. (1983). Meta-rules as a basis for processing ill-formed input. *Computational Linguistics*, *9*(3–4), 161–77.

Weiss, Z., & Meurers, D. (2018). Modeling the readability of German targeting adults and children: An empirically broad analysis and its cross-corpus validation. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*. Santa Fe, NM: International Committee on Computational Linguistics.

Xia, M., Kochmar, E., & Briscoe, T. (2016). Text readability assessment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 12–22). San Diego, CA: Association for Computational Linguistics. Retrieved April 4, 2019 from https://www.aclweb.org/anthology/W16-0502

Ziai, R. (2018). *Short answer assessment in context: The role of information structure* (PhD thesis). Eberhard-Karls Universität Tübingen, Germany.

Suggested Readings

Rebuschat, P., Detmar, M., & McEnery, T. (Eds.). (2017). Language learning research at the intersection of experimental, computational and corpus-based approaches. *Language Learning*, *67*(S1). doi: 10.1111/lang.12243

The Proceedings of the Workshop Series on *Innovative Use of NLP for Building Educational Applications (BEA)*, organized by the Special Interest Group (SIG) for Building Educational Applications of the Association for Computational Linguistics. Accessible from https://sig-edu.org/bea/past

The Proceedings of the Workshop Series *Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL)*, organized by the ICALL SIG of the North European Association of Language Technology. Accessible from https://spraakbanken.gu.se/eng/research/icall/nlp4call