Challenging Learners in Their Individual Zone of Proximal Development Using Pedagogic Developmental Benchmarks of Syntactic Complexity

Xiaobin Chen and **Detmar Meurers**
LEAD Graduate School and Research Network
Department of Linguistics
Eberhard Karls Universität Tübingen, Germany
{xiaobin.chen,detmar.meurers}@uni-tuebingen.de

Abstract

This paper introduces an Intelligent Computer Assisted Language Learning system designed to provide reading input for language learners based on the syntactic complexity of their language production. The system analyzes the linguistic complexity of texts produced by the user and of texts in a pedagogic target language corpus to identify texts that are well-suited to foster acquisition. These texts provide developmental benchmarks offering an individually tailored language challenge, making ideas such as Krashen's i+1 or Vygotsky's Zone of Proximal Development concrete and empirically explorable in terms of a broad range of complexity measures in all dimensions of linguistic modeling.

1 Introduction

The analysis of linguistic complexity is a prominent endeavor in Second Language Acquisition (SLA) where Natural Language Processing (NLP) technologies are increasingly applied in a way broadening the empirical foundation. Automatic complexity analysis tools such as CohMetrix (McNamara et al., 2014), the L2 Syntactic Complexity Analyzer (Lu, 2010), and the Common Text Analysis Platform (Chen and Meurers, 2016) support studies analyzing interlanguage development (Lu, 2011; Lu and Ai, 2015; Mazgutova and Kormos, 2015), performance evaluation (Yang et al., 2015; Taguchi et al., 2013), and readability assessment (Vajjala and Meurers, 2012; Nelson et al., 2012).

In this paper, we introduce a new system called *Syntactic Benchmark (SyB)* that utilizes NLP to create syntactic complexity benchmarks and identify reading material individually challenging learners, essentially instantiating the next stage of acquisition as captured by Krashen's concept of i+1 (Krashen, 1981) or relatedly, but emphasizing the social perspective, Vygotsky's Zone of Proximal Development (ZPD; Vygotsky, 1976).

In terms of structure of the paper, we first locate our approach in terms of the Complexity, Accuracy, and Fluency (CAF) framework in SLA research. Then we review approaches adopted by earlier studies in developmental complexity research, including problems they pose for a pedagogical approach aimed at offering developmental benchmarks. We propose and justify a solution, before presenting the architecture and functionality of the SyB system.

2 Development of Syntactic Complexity

The three-part model of development distinguishing Complexity, Accuracy, and Fluency has gained significant popularity among SLA researchers (Wolfe-Quintero et al., 1998; Skehan, 2009; Housen et al., 2009; Bulté and Housen, 2012) since it was first delineated by Skehan (1989). It provides SLA researchers with a systematic and quantitative approach to development. Among the CAF triplet, complexity arguably is the most researched and most "complex" due to its polysemous and multidimensional nature (Bulté and Housen, 2012; Vyatkina et al., 2015). Complexity in the SLA literature has been used to refer to task, cognitive, or linguistic complexity (Housen et al., 2009). In the present paper, we investigate complexity from a linguistic perspective, where it is concisely characterized by Ellis (2003) as "the extent to which language produced in performing a task is elaborate and varied". While the linguistic complexity construct consists of a range of sub-constructs at all levels of linguistic modeling, such as lexical, morphological, syntactic, semantic, pragmatic and discourse (Lu, 2010; Lu, 2011;

Xiaobin Chen and Detmar Meurers 2017. Challenging learners in their individual zone of proximal development using pedagogic developmental benchmarks of syntactic complexity. *Proceedings of the Joint 6th Workshop on NLP for Computer Assisted Language Learning and 2nd Workshop on NLP for Research on Language Acquisition at NoDaLiDa 2017.* Linköping Electronic Conference Proceedings 134: 8–17.

Lu and Ai, 2015; Ortega, 2015; Mazgutova and Kormos, 2015; Jarvis, 2013; Kyle and Crossley, 2015), the focus in this paper is on syntactic complexity.

In line with Ellis's (2003) definition of linguistic complexity, Ortega (2003) characterized syntactic complexity as the range of syntactic structures and the elaborateness or degree of sophistication of those structures in the language production, which we adopt as the operational definition in this paper. The uses of syntactic complexity analysis in SLA research include (i) gauging proficiency, (ii) assessing production quality, and (iii) benchmarking development (Ortega, 2012; Lu and Ai, 2015).

The development of syntactic complexity in language produced by learners is closely related to the learner's proficiency development. While the goal of language acquisition is not as such to produce complex language, advanced learners usually demonstrate the ability to understand and produce more complex language. With increasing proficiency, the learners are expanding their syntactic repertoire and capacity to use a wider range of linguistic resources offered by the given grammar (Ortega, 2015), thus producing "progressively more elaborate language" and "greater variety of syntactic patterning", constituting development in syntactic complexity (Foster and Skehan, 1996). As a result, syntactic complexity is often used to determine proficiency or assess performance in the target language (Larsen-Freeman, 1978; Ortega, 2003; Ortega, 2012; Vyatkina et al., 2015; Wolfe-Quintero et al., 1998; Lu, 2011; Taguchi et al., 2013; Yang et al., 2015; Sotillo, 2000).

Besides the practical side of performance assessment and placement, in SLA research the developmental perspective is considered to be "at the core of the phenomenon of L2 syntactic complexity" (Ortega, 2015). However, it is also the least addressed and understood phenomenon of syntactic complexity in SLA research (Vyatkina et al., 2015; Ortega, 2012). Understanding the development of syntactic complexity would enable SLA researchers to determine trajectories of the learners' development and set benchmarks for certain time points or across a given time span. On the practical side, such work could help language teachers select or design appropriate learning materials, and it can provide a reference frame for testing the effectiveness of instructional interventions. Hence researching syntactic complexity from a developmental perspective is of far-reaching relevance and applicability.

2.1 Development of Syntactic Complexity in Learner Corpora

A number of longitudinal and cross-sectional studies have been conducted to investigate the relationship between syntactic complexity and learner proficiency, aimed at finding (i) the most informative complexity measures across proficiency levels (Lu, 2011; Ferris, 1994; Ishikawa, 1995), (ii) the patterns of development for different syntactic measures (Bardovi-Harlig and Bofman, 1989; Henry, 1996; Larsen-Freeman, 1978; Lu, 2011), or (iii) discovering a developmental trajectory of syntactic complexity from the learner production (Ortega, 2000; Ortega, 2003; Vyatkina, 2013; Vyatkina et al., 2015).

With a few exceptions (Vyatkina, 2013; Tono, 2004), one thing these studies have in common is that they analyze the syntactic complexity development of learners based on their production. This seems natural since it investigates complexity development by analyzing the production of the developing entity, i.e., the learners. In principle, a longitudinal learner corpus with a continuous record of productions from individual learners over time would seem to enable us to determine the developmental trajectory and linguistic complexity benchmarks. However, this approach encounters some challenges that make it suboptimal for determining developmental benchmarks in practice.

First, the approach is dependent on learner corpora varying significantly on a number of parameters such as the learners' background, the tasks eliciting the production, and the instructional settings, etc. Significant effects of such factors on the syntactic complexity of learner writing have been identified in a number of studies (Ellis and Yuan, 2004; Lu, 2011; Ortega, 2003; Sotillo, 2000; Way et al., 2000; Yang et al., 2015; Alexopoulou et al., 2017). Consequently, the developmental patterns or benchmarks constructed from different learner corpora, elicited using different tasks, etc. are likely to vary or even contradict each other. For example, the correlation between subordination frequency and proficiency level have been found to be positive (Aarts and Granger, 1998; Granger and Rayson, 1998; Grant and Ginther, 2000), negative (Lu, 2011; Reid, 1992), or uncorrelated (Ferris, 1994; Kormos, 2011). It is difficult to build on such conflicting findings in practice.

Second, the NLP tools used for the automatic complexity analysis do not work equally well when applied to the language produced by learners at varied proficiency levels. Complexity analysis is currently performed using tools developed for different analysis needs (McNamara et al., 2014; Lu, 2010; Kyle and Crossley, 2015; Chen and Meurers, 2016). They enable fast and robust analysis of large corpora, in principle making the conclusions drawn from these analyses more powerful. However, analyzing learner data can pose significant challenges to the NLP components, which were usually developed for and tested on edited native language, as found in newspapers. While some NLP tools were shown to be quite reliable for analyzing the writing of learners at upper intermediate proficiency or higher (Lu, 2010; Lu, 2011), their robustness for lower-level writing or for some types of task (e.g., not providing reliable sentence delimiting punctuation) is questionable, requiring dedicated normalization steps and conceptual considerations (Meurers and Dickinson, 2017). This may well be why developmental profiling has rarely been done for learner language below upper-intermediate proficiency levels, as Ortega and Sinicrope (2008) observed. This currently limits the possibility of determining developmental benchmarks or trajectories across the full range of proficiency levels.

Last but not least, second language proficiency development is systematically affected by individual differences, making complexity research findings from learner data chaotic and hard to generalize. For example, Vyatkina et al. (2015) observed a "non-linear waxing and waning" (p. 28) for different modifier categories in a longitudinal learner corpus. Norrby and Håkansson (2007) identified four different types of morphosyntactic complexity development in a corpus of Swedish adult learner language, referred to as "the Careful", "the Thorough", "the Risk-taker", and "the Recycler". The analysis of morphological development in English L2 acquisition presented by Murakami (2013; 2016) also highlights the importance of accounting for individual variation in modeling L2 development. As a result, given the current state of affairs and without complex models integrating a range of factors, developmental benchmarks based on learner corpora are of limited practical use for proficiency placement or performance assessment. Naturally this does not mean that research into developmental patterns based on learner corpora is not important or relevant for SLA. On the contrary, the dynamic and adaptive nature of language acquisition means that it is challenging and interesting to approach language development in a way accounting for individual differences (Larsen-Freeman, 2006; Verspoor et al., 2008; Verspoor et al., 2012), task effects (Alexopoulou et al., 2017), and other factors. For benchmarking and developmental tool development it is useful to look for a more stable data source though.

2.2 Developmental Benchmarks of Complexity in a Pedagogic Corpus

Considering the challenges just discussed, we explore the analysis of syntactic complexity in pedagogic language corpora compiled from well-edited target language (TL). A pedagogic TL corpus is a corpus "consisting of all the language a learner has been exposed to" (Hunston, 2002), or more realistically "a large enough and representative sample of the language, spoken and written, a learner has been or is likely to be exposed to via teaching material, either in the classroom or during self-study activities" (Meunier and Gouverneur, 2009). An optimal TL corpus for benchmarking syntactic complexity development would be one that includes texts targeting learners at any proficiency level, i.e., covering the full spectrum.

The advantages of a pedagogic corpus for developmental benchmarking are two-fold: First, pedagogic corpora can be constructed to exhibit a linear development of complexity measures, as shown by Vyatkina (2013) and confirmed here later. While the developmental trajectory in learner productions is "bumpy" and influenced by individual differences, task, and other factors discussed earlier, the pedagogic corpus can be written in a way targeting increased linguistic complexity. This is desirable if one wants the class to follow an instructional progression enriching grammatical forms in line with the pedagogic input they receive (Vyatkina, 2013). Pedagogically, it should be easier for language teachers to select instructional materials based on a linear benchmark of linguistic complexity, especially if one has evidence of the students' proficiency using that same scale.

Second, the problem of the NLP tools being challenged by learner language, especially that of the low-proficiency learners, is avoided since pedagogic corpora contain texts with grammatically well-formed and edited articles. Considering the high accuracy of current NLP for such text material, the developmental benchmark constructed from a pedagogic corpus using automatic complexity analysis tools should be highly reliable. It should be acknowledged that no benchmarking system can avoid analyzing learner language if the system is used for proficiency placement purposes (unless additional, external language tests are used). However, complexity benchmarks constructed based on a TL corpus are more reliable than a comparison with a benchmark computed based on learner corpora. If the NLP tools fail to process the learner production to be compared to the benchmark because of grammar errors, resulting in placing the student on a lower level of the TL benchmark, the placement in a sense still is indicative of the aspect of the learner language that needs to be improved.

In sum, the above review suggests that a developmental perspective to syntactic complexity aimed at teaching practice can be meaningfully approached with the assistance of a pedagogic corpus consisting of texts targeting learners in a wide spectrum of language proficiency. In the following section, we will introduce an NLP-based system based on this idea.

3 The Syntactic Benchmark System

Syntactic Benchmark (SyB) is an Intelligent Computer Assisted Language Learning (ICALL) system that analyzes the syntactic complexity of a text produced by a learner and places the text onto a developmental scale constructed from a comprehensive pedagogic corpus. The system aims at helping learners place the syntactic complexity level of their writings with regard to the pedagogic benchmark and identify the syntactic areas where further improvement is needed. The system is able to visualize the developmental benchmark for different syntactic complexity measures and the learner's position on the benchmark for the selected complexity index. Based on the complexity level of the user's language output, SyB then proposes appropriately challenging texts from the pedagogic corpus. Reading these texts providing "i+1" input should help the user advance in language proficiency. The size of the "+1", i.e., the degree of the challenge and the overall proficiency level that the learner assumes being at currently are manually specified by the user.

Figure 1 shows the Data Window, into which the learner enters a text they wrote to identify its level in terms of syntactic complexity in relation to the TL benchmark corpus. In Figure 2, we see the Visualization Window providing the result of the analysis for the selected complexity feature (here, the Mean Length of Clause measure). The boxplots show the results for each text in each level in the TL benchmark corpus, and a red line indicates the measure's value for the learner text. Selecting the "Challenge" button leads to the Search Result Window shown in Figure 3. It provides a search result list with links to TL articles intended as i+1 input material for the learner. The texts are slightly above the level of the learner text in terms of the selected complexity measure, with the degree of the challenge being determined by the user setting. The learner also specifies the overall proficiency level they assume to be in so that the text challenging them in terms of the selected complexity measure is selected from the pool of texts intended for that overall proficiency level.

In the following, we take a closer look at the SyB components.

3.1 The Pedagogic Corpus

The pedagogic TL corpus used for constructing the syntactic complexity benchmark consists of 14,581 news articles from the educational website Newsela[1], which is a website that provides news articles on a wide range of topics. Each article on the website is adapted into five reading levels (including an "original" level, which is the article in its unadapted form) by human editors. Newsela uses the Lexile Framework (Lexile, 2007) for text leveling and provides a grade to Lexile mapping for converting from Lexile scores to US grade levels. Since the grade level is easier to understand for most users, the SyB system uses grade levels as benchmarking levels. For copyright reasons, the SyB system does not store the original articles from Newsela. It only keeps records of the complexity statistics of the articles and the Search Result Window provides the results in terms of links to the text on the Newsela web site.

[1] `https://newsela.com`

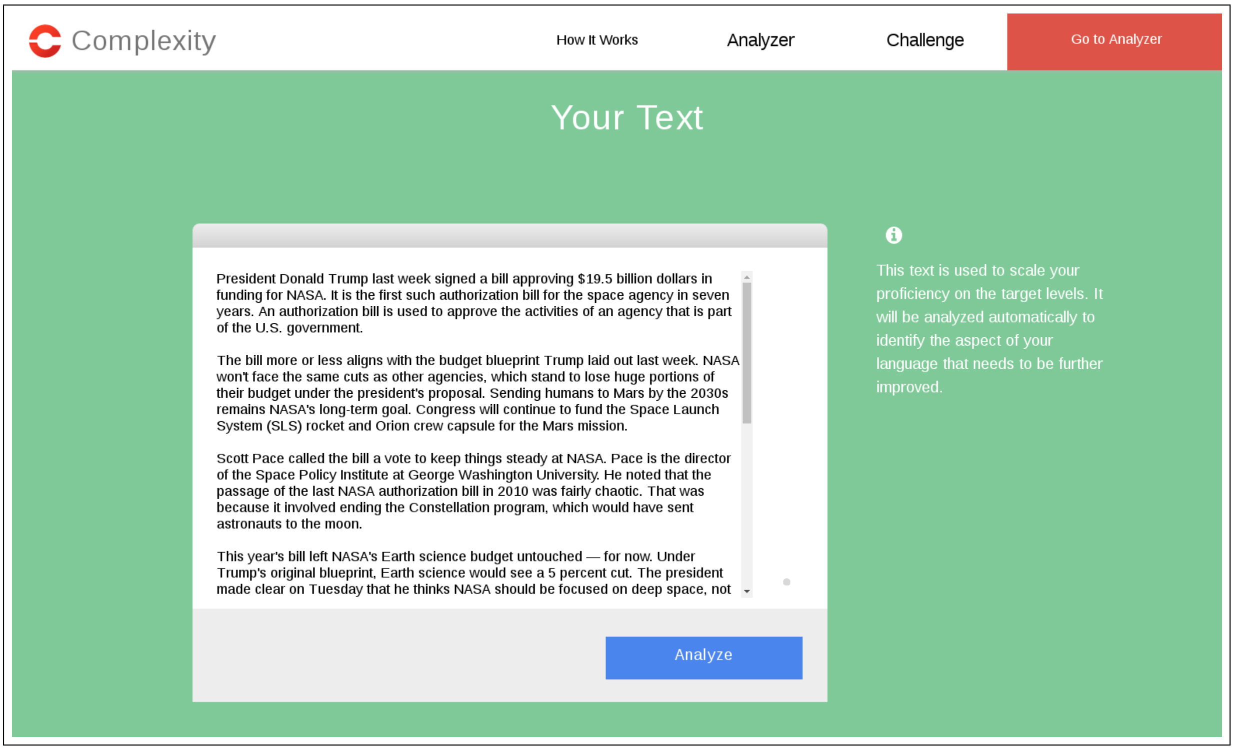

Figure 1: The Data Window of the Syntactic Benchmark Analyzer, where users can paste a composition to identify their level in relation to the TL benchmark corpus

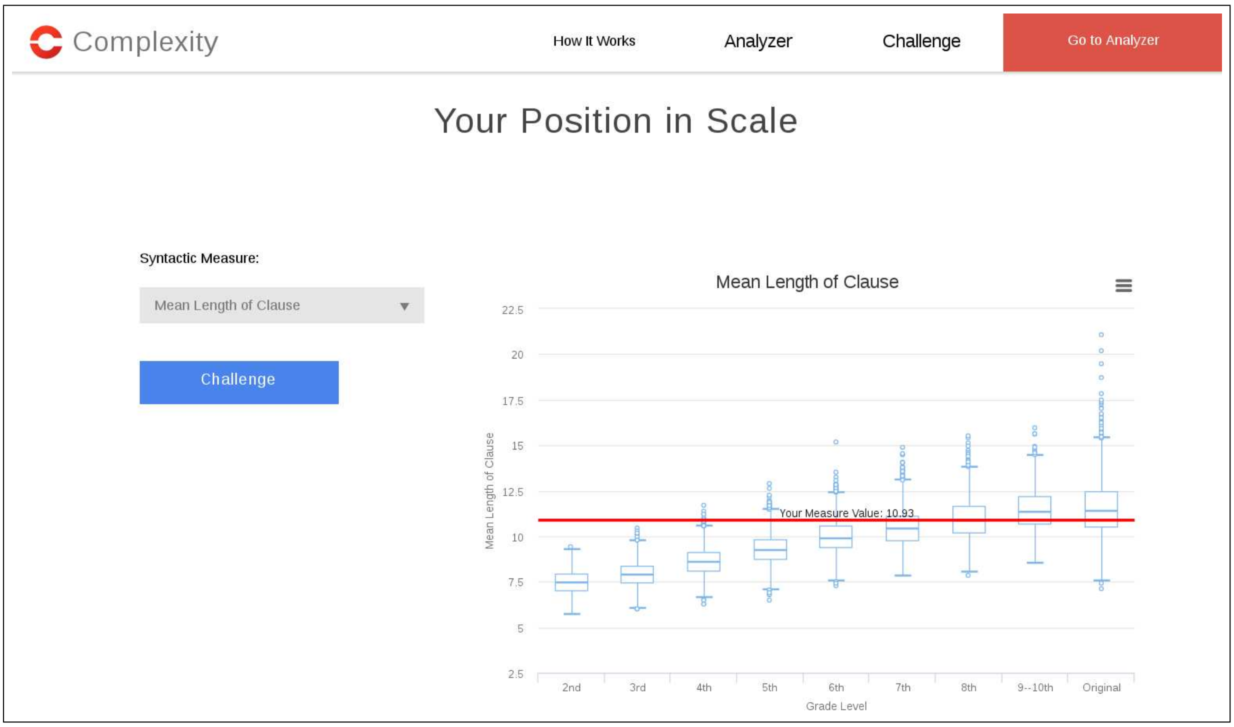

Figure 2: The Visualization Window showing the users' level (red line) for the selected syntactic complexity measure (here: Mean Length of Clause) in relation to the TL benchmark corpus

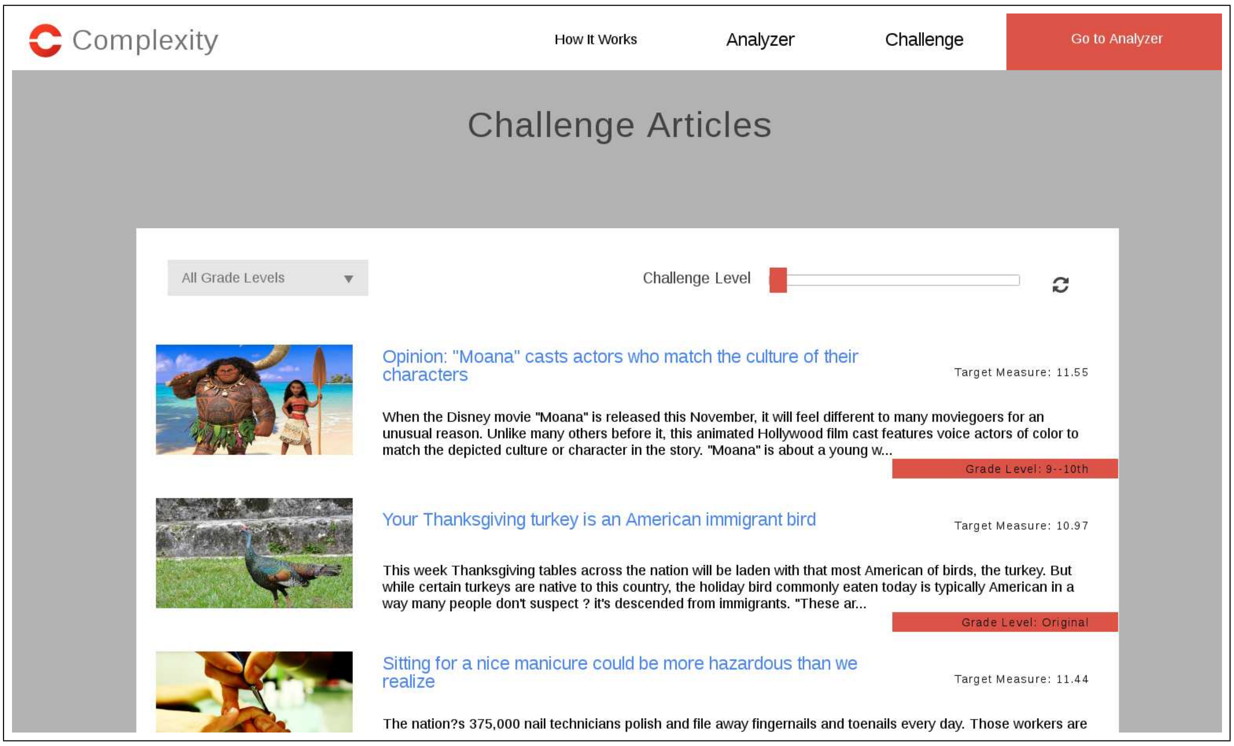

Figure 3: The Search Result Window supporting selection of TL articles based on the learner production's syntactic complexity level (and user-specified degree of challenge and overall target grade level)

3.2 NLP Processing

Each article in the Newsela TL reading corpus was processed with an NLP pipeline consisting of a sentence segmenter, a tokenizer and a parser from the Stanford CoreNLP Toolkit library (Manning et al., 2014). Tregex (Levy and Andrew, 2006), a utility for tree pattern matching, was used to extract syntactic units such as coordinate phrases, clauses, and T-units from the parse tree of a sentence.

We used the Tregex patterns of Lu's (2010) L2 Syntactic Complexity Analyzer and calculated the same set of 14 syntactic indices suggested in his study (p. 479, Table 1). This set of syntactic features have also been used in developmental syntactic complexity studies and proved to be valid and reliable (Larsen-Freeman, 1978; Ortega, 2003; Wolfe-Quintero et al., 1998). The SyB system currently uses a replication of Lu's processing pipeline, which was shown to have achieved a very high level of reliability in a number of studies (Lu, 2010; Lu and Ai, 2015; Yang et al., 2015; Ai and Lu, 2013; Lu, 2011).

In future work, we plan to integrate the broad range of linguistic complexity measures offered by our Common Text Analysis Platform (Chen and Meurers, 2016).

3.3 Benchmarking and Challenging

For each of the 14 syntactic measures, a benchmark box plot of the measure values by grade level was created. Whenever the user pastes or enters a representative production and chooses the measure they are interested in, the SyB system calculates the chosen measure value from the user text and draws a horizontal red line across the benchmark box plot to signify the relative position of the user text's complexity level on the TL corpus benchmark. Figure 2 shows an example of a benchmark plot and the learner text as measured by the same complexity index, Mean Length of Clause.

The system then selects from the TL corpus those articles that challenge the user in terms of specific syntactic complexity as measured by the user's choice of complexity indicator. The user is also given choices of the overall target grade levels of the texts and the level of challenge they want to receive (Figure 3). The range of challenge levels matches the range of the syntactic measure calculated from the TL corpus. The complete challenge range is divided into ten sections and controlled by a range slider with those steps, shown as the red slider in the top-right corner of Figure 3.

Each article in the Newsela TL reading corpus comes with the overall evaluation of reading level by the editors. Since there is significant overlap in the range of complexity measure values across target reading levels, it is useful to let the user determine the overall pool of texts that they want the system to select from using the selected complexity measure. In SyB, the overall reading level of the challenge texts is selected using the drop-down listbox in the top-left corner of Figure 3. The current system then only evaluates a single complexity feature of the learner's production (in the case of Figure 2, Mean Length of Clauses) and proposes texts at an appropriately challenging levels based on this single aspect, selected from the pool of texts at the user-selected overall level.

This is not optimal because whether a text poses challenges to specific readers also depend on other factors, such as the lexical complexity, the learners' language competence including aspects such as strategic competence, their world and domain knowledge, and so forth. An alternative method we intend to explore in the future is to compute a broad range of complexity measures using the NLP from our Common Text Analysis Platform (Chen and Meurers, 2016) so that each text is represented by a vector encoding the results for each complexity measure for that text (which could also include dimensions for other factors to be considered, such as measures of the user's domain knowledge for different topics or subject domains). The overall i+1 challenge can then be computed using a vector distance metric (Manhattan, Euclidean, etc.). Perhaps most attractively, one could combine the two approaches, with the vector-based overall comparison replacing the current manual setting of the global level determining the set of texts to be considered, and the challenge being determined by the user-selected single complexity measure as in the current approach.

The hypothesis behind the overall setup is that by reading the challenging texts, the users will "align" (Wang and Wang, 2015) to the target levels of syntactic complexity, hence promoting their TL proficiency. Whether this hypothesis is correct and which approach works best for determining input material appropriately challenging learners is an empirical question. Answering it should also provide important insights into the question how Krashen's notion of an i+1 (or Vygotsky's ZPD) can be operationalized in terms of measurable features such as linguistic complexity.

4 Summary and Outlook

This paper introduced the ICALL system SyB for benchmarking syntactic complexity development based on a TL corpus. A TL corpus can provide a consistent, linear, and complete instantiation of incremental complexification for different aspects of linguistic complexity. Current NLP technologies are more robust for analyzing such TL corpora than for analyzing learner corpora. As a result, syntactic complexity benchmarks in TL corpora may be more applicable and relevant for instructional use than models of linguistic complexification based on learner corpora, which are harder to analyze automatically, exhibit significant individual variation, task effects, and other uncontrolled factors. However, this hypothesis remains to be validated empirically in actual teaching practice. Future research also needs to investigate which level of challenge for which of the complexity measures at which domain of linguistic modeling is most effective at fostering learning, i.e., what constitutes the best +1 for which aspect of linguistic complexity (for learners with which individual characteristics). Last but not least, while the SyB system provides users with options to control the syntactic complexity and overall reading challenge levels, the system does not take into account the gap between the active ability exhibited in production and the passive ability used for comprehension. The receptive and productive knowledge were found to differ within learners in a number of studies (Zhong, 2016; Schmitt and Redwood, 2011).

We plan to empirically evaluate the system's effectiveness in providing input individually tailored to the i+1 in terms of linguistic complexity as a means to foster learning. It will also be interesting to compare this kind of individual adaptation of the complexity of the input based on the complexity analysis of the learner's production with the input enrichment supported by a teacher-based selection of the constructions targeted to be learned as supported by the FLAIR system (Chinkina and Meurers, 2016).

Finally, it will be interesting to enhance the system by making the texts it suggests for reading adaptive not only to what the learner is capable of producing, but also to how well the learner understands the articles suggested by the system. We are currently developing a production task module where the learner is asked to produce output after reading the complexity challenge texts. This will make it possible to analyze (i) whether there is uptake of the increasingly complex language being read and (ii) how the complexification impacts the user's comprehension of the challenging texts. In principle, the system could then be extended to adapt the subsequent text challenges based on a combination of these form and meaning factors.

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Xiaobin Chen is a doctoral student at the LEAD Graduate School & Research Network.

References

Jan Aarts and Sylviane Granger. 1998. Tag sequences in learner corpora: a key to interlanguage grammar and discourse. In Sylviane Granger, editor, *Learner English on Computer*, pages 132–141. Longman, London; New York.

Haiyang Ai and Xiaofei Lu. 2013. A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In Ana Daz-Negrillo, Nicolas Ballier, and Paul Thompson, editors, *Automatic Treatment and Analysis of Learner Corpus Data*, pages 249–264. John Benjamins.

Theodora Alexopoulou, Marije Michel, Akira Murakami, and Detmar Meurers. 2017. Analyzing learner language in task contexts: A study case of task-based performance in EFCAMDAT. *Language Learning*. Special Issue on "Language learning research at the intersection of experimental, corpus-based and computational methods: Evidence and interpretation".

Kathleen Bardovi-Harlig and Theodora Bofman. 1989. Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11(1):17–34.

Bram Bulté and Alex Housen. 2012. Defining and operationalising l2 complexity. In Alex Housen, Folkert Kuiken, and Ineke Vedder, editors, *Dimensions of L2 Performance and Proficiency*, pages 21–46. John Benjamins.

Xiaobin Chen and Detmar Meurers. 2016. CTAP: A web-based tool supporting automatic complexity analysis. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, Osaka, Japan, December. The International Committee on Computational Linguistics.

Maria Chinkina and Detmar Meurers. 2016. Linguistically-aware information retrieval: Providing input enrichment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 188–198, San Diego, CA.

Rod Ellis and Fangyuan Yuan. 2004. The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26(1):59–84.

Rod Ellis. 2003. *Task-based Language Learning and Teaching*. Oxford University Press, Oxford, UK.

Dana R. Ferris. 1994. Lexical and syntactic features of esl writing by students at different levels of l2 proficiency. *TESOL Quarterly*, 28(2):414–420.

Pauline Foster and Peter Skehan. 1996. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3):299–323.

Sylviane Granger and Paul Rayson. 1998. Automatic profiling of learner texts. In Sylviane Granger, editor, *Learner English on Computer*, pages 119–131. Longman, New York.

Leslie Grant and April Ginther. 2000. Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2):123–145.

Kathryn Henry. 1996. Early l2 writing development: A study of autobiographical essays by university-level students of russian. *The Modern Language Journal*, 80(3):309–326.

Alex Housen, Folkert Kuiken, Jane Zuengler, and Ken Hyland. 2009. Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30(4):461–473.

Susan Hunston. 2002. *Corpora in Applied Linguistics*. Cambridge University Press.

Sandra Ishikawa. 1995. Objective measurement of low-proficiency efl narrative writing. *Journal of Second Language Writing*, 4(1):51 – 69.

Scott Jarvis. 2013. Capturing the diversity in lexical diversity. *Language Learning*, 63:87–106.

Judit Kormos. 2011. Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2):148 – 161.

Stephen D. Krashen. 1981. The fundamental pedagogical principle in second language teaching. *Studia Linguistica*, 35(1–2):50–70, December.

Kristopher Kyle and Scott A Crossley. 2015. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4):757–786.

Diane Larsen-Freeman. 1978. An ESL index of development. *TESOL Quarterly*, 12(4):439–448.

Diane Larsen-Freeman. 2006. The emergence of complexity, fluency, and accuracy in the oral and written production of five chinese learners of english. *Applied Linguistics*, 27(4):590–619.

Roger Levy and Galen Andrew. 2006. Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *5th International Conference on Language Resources and Evaluation*, Genoa, Italy.

Lexile. 2007. The Lexile Framework® for reading: Theoretical framework and development. Technical report, MetaMetrics, Inc., Durham, NC.

Xiaofei Lu and Haiyang Ai. 2015. Syntactic complexity in college-level English writing: Differences among writers with diverse l1 backgrounds. *Journal of Second Language Writing*, -:in press.

Xiaofei Lu. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474–496.

Xiaofei Lu. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level esl writers' language development. *TESOL Quarterly*, 45(1):36–62, March.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Diana Mazgutova and Judit Kormos. 2015. Syntactic and lexical development in an intensive english for academic purposes programme. *Journal of Second Language Writing*, 29:3–15.

Danielle A. McNamara, Arthur C. Graesser, Philip M. McCarthy, and Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press, Cambridge, M.A.

Fanny Meunier and Céline Gouverneur. 2009. New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In Karin Aijmer, editor, *Corpora and Language Teaching*, pages 179–201. John Benjamins, Amsterdam.

Detmar Meurers and Markus Dickinson. 2017. Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, 67(2). `http://dx.doi.org/10.1111/lang.12233`.

Akira Murakami. 2013. *Individual Variation and the Role of L1 in the L2 Development of English Grammatical Morphemes: Insights From Learner Corpora*. Ph.D. thesis, University of Cambridge.

Akira Murakami. 2016. Modeling systematicity and individuality in nonlinear second language development: The case of english grammatical morphemes. *Language Learning*, 6(4):834–871.

Jessica Nelson, Charles Perfetti, David Liben, and Meredith Liben. 2012. Measures of text difficulty: Testing their predictive value for grade levels and student performance. Technical report, The Council of Chief State School Officers.

Catrin Norrby and Gisela Håkansson. 2007. The interaction of complexity and grammatical processability: The case of swedish as a foreign language. *International Review of Applied Linguistics in Language Teaching*, 45:45–68.

Lourdes Ortega and C. Sinicrope. 2008. Novice proficiency in a foreign language: A study of task-based performance profiling on the STAMP test. Technical report, Center for Applied Second Language Studies, University of Oregon.

Lourdes Ortega. 2000. *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners*. Unpublished doctoral dissertation, University of Hawaii, Manoa, HI.

Lourdes Ortega. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4):492–518.

Lourdes Ortega. 2012. A construct in search of theoretical renewal. In B. Szmrecsanyi and B. Kortmann, editors, *Linguistic complexity: Second language acquisition, indigenization, contact*, pages 127–155. de Gruyter, Berlin.

Lourdes Ortega. 2015. Syntactic complexity in l2 writing: Progress and expansion. *Journal of Second Language Writing*, 29:82–94.

Joy Reid. 1992. A computer text analysis of four cohesion devices in english discourse by native and nonnative writers. *Journal of Second Language Writing*, 1(2):79–107.

Norbert Schmitt and Stephen Redwood. 2011. Learner knowledge of phrasal verbs: A corpus-informed study. In Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin, and Magali Paquot, editors, *A Taste for Corpora. In Honour of Sylviane Granger*, pages 173–207. John Benjamins Publishing Company, Amsterdam.

Peter Skehan. 1989. *Individual Differences in Second Language Learning*. Edward Arnold.

Peter Skehan. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4):510–532.

Susana M. Sotillo. 2000. Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4(1):82–119.

Naoko Taguchi, William Crawford, and Danielle Zawodny Wetzel. 2013. What linguistic features are indicative of writing quality? a case of argumentative essays in a college composition program. *TESOL Quarterly*, 47(2):420–430.

Yukio Tono. 2004. Multiple comparisons of IL, L1 and TL corpora: the case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In Guy Aston, Silvia Bernardini, and Dominic Stewart, editors, *Corpora and Language Learners*, pages 45–66. John Benjamins.

Sowmya Vajjala and Detmar Meurers. 2012. On improving the accuracy of readability classification using insights from second language acquisition. In Joel Tetreault, Jill Burstein, and Claudial Leacock, editors, *In Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–173, Montral, Canada, June. Association for Computational Linguistics.

Marjolijn Verspoor, Wander Lowie, and Marijn Van Dijk. 2008. Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2):214–231.

Marjolijn Verspoor, Monika S. Schmid, and Xiaoyan Xu. 2012. A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3):239–263.

Nina Vyatkina, Hagen Hirschmann, and Felix Golcher. 2015. Syntactic modification at early stages of L2 german writing development: A longitudinal learner corpus study. *Journal of Second Language Writing*, 29:28–50.

Nina Vyatkina. 2013. Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, 97(S1):11–30.

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

Chuming Wang and Min Wang. 2015. Effect of alignment on l2 written production. *Applied Linguistics*, 36(5).

Denise Paige Way, Elizabeth G. Joiner, and Michael A. Seaman. 2000. Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of french. *The Modern Language Journal*, 84(2):171–184.

Kate Wolfe-Quintero, Shunji Inagaki, and Hae-Young Kim. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy & Complexity*. Second Language Teaching & Curriculum Center, University of Hawaii at Manoa, Honolulu.

Weiwei Yang, Xiaofei Lu, and Sara Cushing Weigle. 2015. Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28:53–67.

Hua Flora Zhong. 2016. The relationship between receptive and productive vocabulary knowledge: a perspective from vocabulary use in sentence writing. *The Language Learning Journal*, Advanced Access.