What Is Proficiency: Updating Our Current Notions

Trevor Sitler
Kansai University

Reference Data:

Sitler, T. (2025). What is proficiency: Updating our current notions. In B. Lacy, R. P. Lege, & M. Swanson (Eds.), *Moving JALT Into the Future: Opportunity, Diversity, and Excellence*. JALT. https://doi.org/10.37546/JALTPCP2024-42

This paper attempts to inform Task-Based Language Teaching research (TBLT) and TBLT pedagogy by doing a Register Analysis (RA) with lexical bundles to describe the language used by different proficiency levels on a picture description task in the NICT-JLE corpus. Through a description of the language used by each proficiency level, this methodology was able to identify the syntactical form necessary to complete the task, as well as the structures that increase communicative efficiency. The structure 3rd Person/Noun + VP was found to be an essential structure used by all levels, while higher levels varied this structure by using present participles, partitives, and verbs of perception. The data gained from such an analysis can inform the use of focused tasks in focus-on-form instruction in TBLT, especially in the design of input tasks tailored to specific proficiency levels.

本稿は、タスク中心言語教授法（TBLT）の研究および教育実践に貢献することを目的としている。そのために、NICT-JLEコーパスにおける絵描写タスクで使用される言語を、異なる習熟度レベルごとに語彙連鎖を用いたレジスター分析（RA）を行った。各習熟度レベルの言語の特徴を記述することで、本手法はタスク遂行に必要な統語形式や、コミュニケーション効率を高める構造を特定することができた。「三人称／名詞＋動詞句」の構造はすべてのレベルで使用される重要な構造である一方、高習熟度レベルでは現在分詞、部分表現、知覚動詞を用いてこの構造が多様化されていることが明らかになった。このような分析から得られたデータは、TBLTにおけるフォーカス・オン・フォーム指導、特に特定の習熟度に合わせたインプット・タスクの設計に有益な示唆を与える。

Task-based language teaching (TBLT) is an approach to language teaching characterized by a focus on meaning rather than form or grammar. Bygate et al. (2022) claim that one of the principles in which tasks operate is the "Meaningfulness Principle," and completing a specific task goal is emphasized more than acquiring a particular grammatical form. In this approach, learners are taught to have a "functional command" of a language, not just learn about grammatical features (Long & Ahmadian, 2022, p. xxvi).

The dominant means by which students are evaluated within TBLT research is through the Complexity, Accuracy, Fluency (CAF) framework (Housen & Kuiken, 2009). While this framework has been very good at explaining how different factors, such as task complexity, influence student output, the CAF framework has been criticized for its reliance on holistic measures (Biber et al., 2020) and for its lack of reference to communicative success (Pallotti, 2009). Biber et al. (2020) have claimed that overreliance on holistic measures is unable to provide information on specific syntactical features and how they function in the task. This, joined with Pallotti's (2009) criticism that CAF factors are not immediately related to communicative success means that the dominant framework used to evaluate students in TBLT research is not connected to meaning. Furthermore, a lack of information on specific syntactical structures used in specific tasks and task types deprives teachers and researchers of important information on how different tasks can elicit different lexico-grammatical features (Bygate, 2020).

Crawford and Zhang (2021) have pointed out the similarities between TBLT and Register Analysis research, as both traditions seek to understand how non-linguistic features influence linguistic output. TBLT investigates how the difficulty of a task, the amount of planning time, etc., influences the students' interlanguage. RA similarly investigates how the functional goals of a particular situation influence the linguistic choices the speaker or writer may make (Biber & Conrad, 2019). The particular context of language use is known as a register, and treating tasks themselves as a register in an RA would allow for the investigation of how specific lexico-grammatical features are related to the functional goals of a task. Doing a contrastive interlanguage language analysis (CIA) (Gilquin & Granges, 2015), in which the lexico-grammatical features of different proficiency levels are compared and contrasted with each other, may also show

patterns of development from low to high levels. This method can supply TBLT research with information on specific grammatical structures that are currently lacking in the repertoire of students at different proficiency levels.

TBLT and Focus on Form

This paper will first discuss how specific syntactic structures are taught in a Task-Based classroom. TBLT 's emphasis on meaning has important implications for how specific grammatical structures are taught. Long (1991) drew a distinction between two kinds of language instruction: focus-on-form and focus-on-forms. In the latter, the primary focus of the class is on specific linguistic features that the teacher has selected before class. The focus is on the grammatical form of the language. In contrast, focus-on-form emphasizes meaning. Attention to the form comes from a desire to communicate effectively during a communicative task. This is the type of instruction often employed in TBLT classrooms. Doughty (1999) argues that during focus on form, students simultaneously focus on form, meaning, and use in an attempt to communicate their meaning.

While the focus on form can be incidental, in which forms are meant to "arise naturally out of the performance of the task" (Ellis et al., 2002, p. 421), this study will focus on planned focus on form, in which the teacher has a specific, targeted structure in mind. It is important to note, however, that this is different from the above-mentioned focus-on-forms in that the main focus of the task continues to be on meaning and communication, and the students are not explicitly aware that a specific form is being targeted.

This targeted structure can be incorporated into input-based tasks. In contrast to production tasks, in which learners are expected to produce linguistic forms, input-based tasks are structured around the comprehension input provided by the instructor (Erlam & Ellis, 2018). Students are expected to comprehend the input and, by doing so, pick up the linguistic features used in the task. An example provided in Erlam and Ellis (2018, p. 499) called 'Flyswat' involves showing pictures of different articles of clothing on the board. The teacher calls out a certain piece of clothing, and the students are expected to hit the piece of clothing that corresponds to the teacher's word with their flyswatter. Ellis (2020) notes that beginner students especially react very well to input tasks. Shintani (2016) and Shintani and Ellis (2010) both showed grammar acquisition through repetition of input-based tasks. These studies investigated schoolchildren in Japan in which learners listened to instructions requiring them to distinguish between singular and plural nouns.

Loschky and Bley-Vroman (1993) describe three ways targeted language can be incorporated into a task: essentialness, utility, and naturalness. Task-essentialness refers to how necessary a particular form is to the completion of a task. Task utility refers to how useful the structure is in completing the task, though again, it may not be essential. Naturalness indicates the degree to which the targeted structure can be thought of as arising naturally from the task, though it may not be completely necessary to complete the task.

Focus on form can also take place during conversation, either before or after an error has been made. The latter, known as reactive focus on form, can be done during conversation when there is a breakdown in communication. The interaction hypothesis developed by Long (1996) claims that the ideal setting for language acquisition to take place is this negotiation of meaning that occurs when communication breaks down. In contrast, there can also be a negotiation of form, in which communication takes a brief pause so that a linguistic error can be addressed despite no breakdown in communication.

Literature Review

In an attempt to provide data to better inform teachers who would like to incorporate planned focus on form in their classrooms, this study conducted an RA focusing on lexical bundles used during a picture description task. Lexical bundles are strings of three or more words that appear frequently in each discourse (Biber et al., 1999). Lexical bundles are considered an accessible, though as of yet, underutilized and under-researched means of doing an RA (Crawford & Zhang, 2021). Doing an RA with lexical bundles entails first determining the length of the bundle to be researched, which in most cases is a four-word bundle, as well as the minimum frequency and minimum dispersion of the bundle across texts. After this step, each bundle is categorized into two different taxonomies: a structural taxonomy and a functional taxonomy. For the sake of brevity, this paper will only consider the structural taxonomy. Biber et al. (2004), in their study of lexical bundles across a variety of university registers, found three main subcategories within the structural taxonomy: Verb Phrase (VP) fragments (e.g., *is going to be*), dependent clause (DC) fragments (e.g., *if you want to*), and noun phrase/prepositional phrase (NP/PP) fragments (e.g., *one of the things*). These can be further subdivided into more specific structures. For example, VP fragments can be divided further into (connector +) 1st person/2nd person pronoun + VP fragment (e.g., *you don't have to*). See Appendix A for further details. Biber et al. (2004) found that there was a

significant difference in lexical bundles used across registers, with written registers using more NP/PP bundles and spoken registers using more VP bundles.

There has also been research using lexical bundles on learner corpora using CIA. Chen and Baker (2010) found that L1 Chinese students used more VP fragments within academic English writing, while native and expert writers used more NP/PP fragments. A similar study comparing L1 Korean students' English Language argumentative essays with essays written by L1 English students also found that non-native writers tended to overuse VP and underuse NP/PP-based bundles. Combined with the results from Biber et al. (2004), these studies suggest that non-native and lower proficiency learners use language that is more typical of speech in writing assignments, as the study identified VP fragments as being prominent in conversation, while NP/PP fragments were typical of writing.

A study of a spoken corpus of Korean students in an EFL setting found that VP-based fragments were the most common among both native speakers and learners (Lee & Zipagan, 2018). However, native speakers tended to use a larger variety of bundles, including, for example, more NP/PP-based bundles.

The above studies, however, do not consider the specific task the students are performing. There is also no information about the performance of specific proficiency levels. To fill this gap, this study investigated the specific syntactic structures used in a specific task and then analyzed how they differ between proficiency levels. This information can be used to supply teachers with information about the kind of syntactical structures used in a particular task, which can then be used in the creation of focused tasks in a TBLT classroom.

Methodology

There were two basic research questions in this study:

RQ1. Can relative frequency of lexical bundles indicate the essential syntactic structures for a particular task?

RQ2. Can a developmental progression be found between proficiency levels?

Data

This study used the National Institute of Information and Communications Technology Japanese Learner Corpus (NICT-JLE) (Izumi et al., 2004). This spoken corpus contains 1.2 million words from 1,281 Japanese speakers of various ages and occupations. It contains recordings from the results of which determine the students' proficiency level. The Standard Speaking Test (SST) was developed for Japanese learners based on the Oral Proficiency Interview (OPI), which was originally developed by the American Council for the Teaching of Foreign Language (ACTFL) (Izumi et al., 2004). Detailed descriptions of the levels can be found in Appendix B (ACTFL Proficiency Guidelines, 2012). The original rating groups students into nine different levels. However, due to the large disparity in the number of texts for each group, these were combined into four groups: Beginner, Pre-Intermediate, Intermediate, and Advanced. Information on text and token count can be seen in Table 1.

The SST is a one-on-one interview that lasts for fifteen minutes and consists of three tasks: picture description, negotiation, and narration. This study will focus on the "messy classroom" task in the description task type. In this task, students had to describe a picture of a messy and chaotic classroom. A sub-corpus of student responses for this task was made, and information on texts and task type can be found below. The interviewer's words were deleted, and only the students' words were analyzed.

Table 1

Texts and Token Counts for Each Proficiency Level

Category	Beginner	Pre-Intermediate	Intermediate	Advanced
Participants/Text	8	25	41	54
Token Count	1803	6754	14028	21591

This specific task was chosen because, unlike other tasks, all proficiency levels were included, and the distribution of tokens was relatively equal. No cut-off frequency was used, so even lexical bundles with a frequency of one were used. This was due to the overall small size of the sub-corpora.

Data Analysis Instruments

After the creation of the sub-corpus, lexical bundles were extracted using Antconc (Anthony, 2024). Further analysis and tagging were done using Excel. Lexical bundles that included repetitions (e.g., *the teacher is*) were excluded from this analysis. Non-standardized bundles, that is, bundles containing an error (e.g., *he listen music and*) were also labeled. From this data, graphs were made showing the distribution of the grammatical variation of this bundle type across proficiency levels.

Data Analysis Procedures

This study used the Register Analysis Methodology to categorize bundles. Register Analysis is primarily a quantitative approach, though there is a qualitative element as well, as the researcher must determine how the various lexico-grammatical items are functioning in context (Biber & Conrad, 2019). Lexical bundles were first categorized according to the structural taxonomy used by Biber et al. (2004) and mentioned above. Then, the most common structural bundle was determined. After that, all variations of the bundle (whether or not it included a present participle, for example) were described. The use of these variations was then described for each proficiency level in order to determine if a developmental progression could be found.

Results

The first step consisted of extracting the lexical bundles and then sorting them into the three structural categories of VP, NP/PP, and DC. The VP-based category of bundles contained the largest number of instances, so this category was further sorted into smaller structures. The structure that was used the most by all proficiency levels was the 3rd Person/Noun + VP structure.

Table 2 shows the number and frequency of 3rd Person/Noun + VP structure for each proficiency level. The second column refers to the number of specific, unique types used, while the latter refers to how often the bundles were used. The number of bundles is close to the frequency of bundles, indicating that most levels did not use individual bundles more than once. The advanced group is the exception, with the frequency of bundles exceeding the number of bundles used, indicating that this level reused many of the same bundles.

Table 2

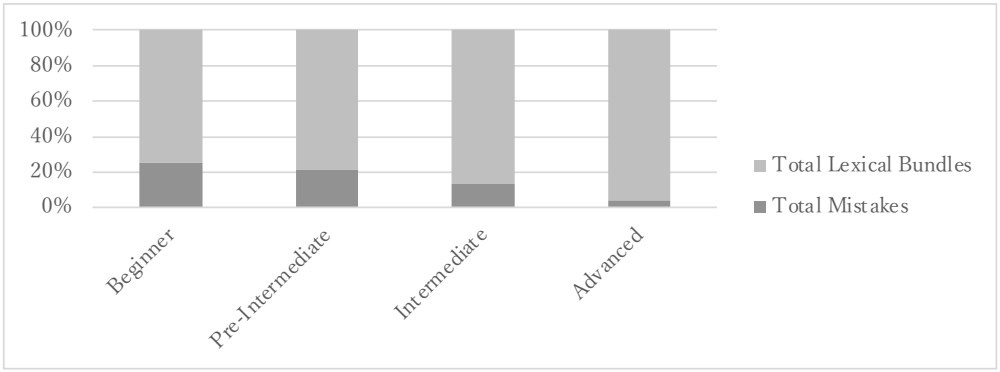
Number and Frequency of Bundles for Each Proficiency Level

Level	Number of Bundles	Frequency of Bundles
Beginner	60	61
Pre-Intermediate	90	99
Intermediate	259	283
Advanced	426	540

Figure 1 shows the ratio of non-standard bundles to total bundles used. For each group, the ratio gets smaller, with beginners starting out with over 20% and advanced being under 10%.

Figure 1

Ratio of Total Bundles Used and Total Amount of Mistakes



For the next step, bundles were further analyzed into different grammatical variations. Table 3 shows the number of bundle tokens for each variation, with the percentage of the total in parentheses. Four main variations of the %rd Person/Noun + VP structure were found: Present participle (e.g., *the teacher is teaching*), conjunctions (e.g., *and one boy is*), partitives (e.g., *some people are talking*), and verbs of perception (e.g., *students seem to be*).

Table 3

3rd Person/ Noun + VP Variations Across Proficiency Levels

Proficiency Level	Present Participle	Conjunction	Partitive	Verbs of Perception
Beginner	24 (40%)	11 (18%)	1 (2%)	1 (2%)
Pre-Intermediate	34 (38%)	18 (20%)	5 (6%)	1 (1%)
Intermediate	111 (42%)	52 (20%)	21 (8%)	9 (3%)
Advanced	205 (48%)	83 (19%)	40 (9%)	21 (5%)

The most common variation among the advanced users is the use of the present participle. Advanced learners use these almost twice as often as the intermediate level, who use them twice as much as the pre-intermediate level. Next, conjunctions (e.g., *but, so, because, and*) nearly doubled with each proficiency level. Partitive verbs (e.g., *some, most*) likewise increased drastically with proficiency level. Lastly, verbs of perception (e.g., *look, seems*) were characteristic of advanced student speech and intermediate student speech to a lesser extent. In contrast, they were rarely used by pre-intermediate and beginner-level students.

Discussion

This section will describe the language used by each level in more detail and give pedagogical recommendations based on that description.

Description of Each Proficiency Level

Beginner

Beginner speech was characterized by a repetition of the basic structure, 3rd Person/Noun + VP. Beginner speech was also characterized by the use of non-standard bundles (e.g., *and these girls is*), which made up 20% of beginner speech, as seen in Figure 2. Many of the non-standard bundles in the beginner group involved subject/verb agreement (e.g., *and these girls is, these girls is chatting, he eat chewing gum*).

There was an overall lack of lexical bundle tokens and lexical bundle frequency, indicating both a lack of overall variety of bundles used as well as the frequency of the individual bundles. Beginners attempted to solve the task's goal, in this description of a picture, through the use of a limited number of variations of the 3rd Person/Noun + VP structure.

However, the beginner level's use of present participles (e.g., *the teacher is teaching*) and conjunctions (e.g., *and the boy is*) suggests that these two syntactic structures may be acquired earlier than partitives and verbs of perception, which only start to appear frequently at later proficiency levels.

Pre-Intermediate

Compared to the beginner level, the pre-intermediate level is characterized by a modest increase in bundle variety and frequency. There was also a decrease in non-standard bundles at 20%. Students who have moved from beginner to pre-intermediate have been able to standardize further the use of 3rd Person/Noun + VP and use this more frequently than beginner students.

However, there is still overreliance on the basal form of 3rd Person/Noun + VP and issues with non-standardization. Table 3 shows only a small amount of variation in this structure compared to other levels, as well as a general lack of bundles.

Intermediate

In addition to more standardization of bundles, speech in the intermediate group is characterized by an increase in syntactic variation, as seen in Table 3. The use of all four syntactic structures measured in this study more than doubled from pre-intermediate to intermediate. The overall number of bundles used, as well as the frequency of bundles, also increased. The jump from pre-intermediate to intermediate entails the ability to vary the 3rd Person/Noun + VP structure with different syntactic features. The most common variation used was the addition of the present participle, which, as can be seen in Table 3, drastically increases from pre-intermediate to intermediate. While not as drastic, there was also a large increase in conjunctions. The intermediate level also began to introduce more partitives as well as verbs of perception.

The jump from pre-intermediate to intermediate entails more than just the smooth use of the base form of the 3rd Person/Noun + VP structure. It also entails a large jump in the amount of language used and modifications in tense through using the present participle. Lee and Zipagan (2018) and Chen and Baker (2016) also note that higher level students tend to use more grammatically diverse bundles than beginner level students.

Advanced

The speech of advanced learners is characterized by an increase in the total number of bundles used and an increase in the frequency of bundles. Advanced students not only had more resources available, but they could also use these resources more frequently to complete the task. The advanced group shows not only a large jump in the number of present participles used but also an increase in syntactic structures rarely used in the prior levels, such as partitives and verbs of perception, which were used by advanced level students a total of 40 and 21 times, respectively. This is in line with Lee and Zipagan (2018) and Chen and Baker (2016), which found more structural variation in advanced students' speech and writing.

Pedagogical Implications

This section will use the above data in order to give pedagogical advice for teachers using picture description tasks in a task-based context. The data above is best suited to provide teachers with a general idea of the type of errors made during a description task and the grammar to be targeted in a pre-emptive focus on form. It should be noted that the majority of the errors made were not errors that would cause a major breakdown in communication. The subject-verb disagreement that often occurred at the beginner and pre-intermediate levels would unlikely cause a breakdown in communication needed for the kind of negotiation of meaning argued for by Long (1996). This is even more true for the higher levels, as the absence of partitives, verbs of perception, and even present participles would be highly unlikely to cause a complete breakdown in communication.

As such, a focus on form would have to be didactic and explicit. However, this data is best served to inform the creation of a pre-emptive focus on form, especially through the design of focused input tasks. First, the data will be considered in light of Loschky and Bley Vroman's (1993) three conditions mentioned above: task essentialness, task utility, and task naturalness. The basal form of 3rd Person/Noun + VP can be considered essential to complete this task. It is hard to imagine how a student could describe a picture without using this form. The students who could not use this form smoothly are considered beginners. In terms of naturalness, as the data comes from a real recording of student language, all the language here can be considered natural.

The variations in the 3rd Person/Noun + VP (e.g., *present participles, partitives, etc.*) seen in the higher levels can be seen as increases in task utility. While they may not be essential to communication, they increase efficiency. Present-participles are the best at describing continuous action, as is depicted in pictures. Conjunctions combine several actions together, just as within the picture, several different types of action can be seen to be happening. Partitives divide and categorize different people, allowing for more specific descriptions. Finally, verbs of perception allow the speaker to show their subjectivity and provide an opportunity for hedging.

Next, issues regarding input task design will be discussed. The descriptions of each proficiency level provide ample data for task design. Beginner and pre-intermediate input tasks should focus on the basal form of 3rd Person/Noun + VP. It has also been shown that beginners do well with input tasks that focus more on comprehension than output (Shintani, 2016; Shintani & Ellis, 2010). For example, beginners can listen to a short audio sample and then choose the picture the audio describes.

Moving up to intermediate and advanced levels entails not just the use of targeted structures but a large increase in fluency as well. It should be noted that overemphasizing form may inhibit fluency (Ellis et al., 2002), so for the intermediate stage, it is important that the teacher allows ample time for speaking in addition to the input task. In terms of structures, intermediate and advanced students should be provided with input tasks with partitives and verbs of perception. Since a lack of these features by themselves is unlikely to cause a breakdown in communication, the teacher needs to supply more explicit corrective feedback so as to direct the students' attention to these useful structural features.

Limitations and Conclusion

This study attempted to discover a developmental progression among proficiency levels from a sub-corpus of a single picture description task in the NICT-JLE corpus. Lexical bundles were first extracted and then categorized by their main structural component according to Biber et al.'s (2004) structural taxonomy. Then, the most commonly used structural bundle was identified to find variation among proficiency levels. The 3rd Person/Noun + VP structure was further analyzed, and then variations of this bundle were labeled, identifying four main variations. These variations were then measured in each proficiency level.

It was found that there was a large amount of variation between proficiency levels within this single structure. Beginner and pre-intermediate students were characterized by reliance on the basal 3rd Person/Noun + VP as well as non-standard versions of this structure. For the higher levels, what separated the advanced students from the pre-intermediate students was not an absence of miscommunication, but rather the ability to add more nuance and meaning to their sentences. This may call for more emphasis on focused input tasks, in order to prepare students to use the language that will complete the communicative goal in the most effective and efficient way possible. This may also call for a more explicit, didactic interpretation of focus on form.

There were some limitations to this study. First, it was quite small, only including 128 students. There were also large discrepancies in the number of texts for each level, which may have affected the results. For the pedagogical advice, further action research can investigate the extent to which the proposed recommendations result in acquisition of the target structures recommended.

Overall, this study was able to provide specific syntactic information regarding task performance in a TBLT context. A Register Analysis with lexical bundles can show

the linguistic parameters and boundaries of a particular task, as well as the syntactical resources needed to complete a task in the most efficient way possible. While students are free to choose any words they feel would communicate their message, the communicative rules and structure of the task itself limit these linguistic possibilities. Future RAs can help to explicate the linguistic possibilities in a variety of different tasks and task types.

Bio Data

Trevor Sitler is currently pursuing a Ph.D. in foreign language education at Kansai University. He holds an MA in TESOL from the University of Birmingham and has been teaching and living in the Kansai region of Japan for a decade. He is currently working as an adjunct lecturer at Kindai University, Ryukoku University, and Ritsumeikan University.

References

ACTFL Proficiency Guidelines. (2012). *ACTFL proficiency guidelines*. Available at: ACTFL | ACTFL Proficiency Guidelines 2012 (Accessed 10 September 2023).

Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntConc

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Longman.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at… Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*, 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D., & Conrad, S. (2019). *Register, genre, and style*. Cambridge University Press.

Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, *46*, 100869. https://doi.org/10.1016/j.jeap.2020.100869

Bygate, M. (2020). Some directions for the possible survival of TBLT as a real-world project. *Language Teaching*. *53*(3), 275–288. https://doi.org/10.1017/S0261444820000014

Bygate, M,. Samuda, V., & den Branden, K. (2022). A pedagogical rationale for task-based language teaching for the acquisition of real-world language use. In M. Ahmadian and M. Long, (Eds.) *The Cambridge handbook of task-based language teaching* (pp.27–52). Cambridge University Press. https://doi.org/10.1017/9781108868327.003

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, *14*(2), 30–49. http://dx.doi.org/10125/44213

Crawford, J., & Zhang, M. (2021). How can register analysis inform task-based language teaching? *Register Studies*, *3*(2),180–206. https://doi.org/10.1075/rs.20021.cra

Chen, Y.-H., & Baker, P. (2016). Investigating criterial discourse features across second language development: lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics, 37*(6), 849–880. https://doi.org/10.1093/applin/amu065

Doughty, C. (1999). Cognitive underpinnings of focus on form. *University of Hawai'i Working Papers in ESL, 18*,1–69. http://hdl.handle.net/10125/40801

Ellis, R. (2020). Task-based language teaching for beginner-level young learners. *Language Teaching for Young Learners*, *2*(1), 4–27. https://doi.org/10.1075/ltyl.19005.ell

Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus on form. *System*, *30*, 419–432. https://doi.org/10.1016/S0346-251X(02)00047-7

Erlam, R., & Ellis, R. (2019). Input-based tasks for beginner-level learners: An approximate replication and extension. *Language Teaching*, *52*(4), 490–511. doi:10.1017/S0261444818000216

Granger, S., & Gilquin, G. (2015) Learner language. In D. Biber and R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press. https://doi.org/10.1017/cbo9781139764377.024

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*, 461–473. https://doi.org/10.1093/applin/amp048

Izumi, E., Uchimoto, K., & Isahara, H. (2004). The NICT JLE Corpus: Exploiting the language learners' speech database for research and education. *International Journal of the Computer, the Internet and Management*,*12*(2), 119–125.

Lee, K., & Zipagan, M. (2018). Korean English Learners' use of lexical bundles in speaking. *Journal of Asia TEFL*. *15*(2), 276–291. http://dx.doi.org/10.18823/asiatefl.2018.15.2.2.276

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In C. N. Candlin and N. Mercer (eds.), *English language teaching in its social context* (pp. 182–190). Routledge.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie and T.K. Bhatia (Eds.) *Handbook of second language acquisition* (pp. 413–468). Academic Press. https://doi.org/10.1016/b978-012589042-7/50015-3

Long, M., & Ahmadian, M. (2022). Preface the origins and growth of task-based language teaching. In Ahmadian, M. and Long, M. (Eds.) *The Cambridge Handbook of Task-Based Language Teaching* (pp.xxv-xxxii). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108868327.001

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes and S. Gass (Eds.) *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Multilingual Matters.

Pallotti, G. (2009). CAF: Defining, refining and differentiation constructs. *Applied Linguistics*, *30*, 590–601. https://doi.org/10.1093/applin/amp045

Shintani, N. (2016). Input-based tasks in foreign language instruction for young learners. Amsterdam: John Benjamins. https://doi.org/10.1075/tblt.9.03the

Shintani, N., & Ellis, R. (2010). The incidental acquisition of plural-s by Japanese children in comprehension-based lessons: A process-product study. *Studies in Second Language Acquisition, 32*(4), 607–637. https://doi.org/10.1017/S0272263110000288

Appendix A

Structural Types of Lexical Bundles

(Source: Biber et al., 2004)

Lexical Bundles that incorporate *verb phrase* fragments
1a. (connector +) 1st/2nd person pronoun + VP fragment
Example: *you don't have to, I'm not going to*
1b. (connector +) 3rd person pronoun + VP fragment
Example: *it's going to be, that's one of the*
1c. Discourse marker + VP fragment
Example: *I mean you know, you know it was, I mean I don't*
1d. Verb Phrase (with non-passive verb)
Example: *is going to be, is one of the*
1e. Verb Phrase (with passive verb)
Example: *is based on the, can be used to*
1f. *yes-no* question fragements
Example: *are you going to, do you want to*
1g. WH-question fragments
Example: *what do you think, how many of you*
Lexical bundles that incorporate *dependent clause* fragments
2a. 1st/2nd person pronoun + dependent clause fragment
Example: *I want you to, I don't know if, I don't know why*
2b. WH-clause fragments
Example: *what I want, what's going to happen*
2c. *If*-clause fragments
Example: *if you want to, if you have a*
2d. (verb/adjective+) *to*-clause fragments
Example: *to be able to, to come up with*
2e. *That*-clause fragments
Example: *that there is a, that I want to*
Lexical bundles that incorporate *noun phrase* and *prepositional phase* fragments
3a. (connector +) Noun phrase with *of*-phrase fragment
Example: *one of the things, the end of the*
3b. Noun phrase with other post-modifier fragment
Example: *a little bit about, those of you who*
3c. Other noun phrase expressions
Example: *a little bit more, or something like that*
3d. Prepositional phrase expressions
Example: *of the things that, at the end of*
3e. Comparative expressions
Example: *as far as, greater than or equal*

Appendix B

Description of the Three Main Levels for the ACTFL

(Source: ACTFL Proficiency Guidelines, 2012)

Advanced
Speakers at the Advanced level engage in conversation in a clearly participatory manner in order to communicate information on autobiographical topics, as well as topics of community, national, or international interest. The topics are handled concretely by means of narration and description in the major time frames of past, present, and future. These speakers can also deal with a social situation with an unexpected complication. The language of Advanced-level speakers is abundant, the oral paragraph being the measure of Advanced-level length and discourse. Advanced-level speakers have sufficient control of basic structures and generic vocabulary to be understood by native speakers of the language, including those unaccustomed to non-native speech.
Intermediate
Speakers at the Intermediate level are distinguished primarily by their ability to create with the language when talking about familiar topics related to their daily life. They are able to recombine learned material in order to express personal meaning. Intermediate-level speakers can ask simple questions and can handle a straightforward survival situation. They produce sentence-level language, ranging from discrete sentences to strings of sentences, typically in present time. Intermediate-level speakers are understood by interlocutors who are accustomed to dealing with non-native learners of the language.
Novice
Novice-level speakers can communicate short messages on highly predictable, everyday topics that affect them directly. They do so primarily through the use of isolated words and phrases that have been encountered, memorized, and recalled. Novice-level speakers may be difficult to understand even by the most sympathetic interlocutors accustomed to non-native speech.