ASC analyzer: A Python package for measuring argument structure construction usage in English texts

Hakyung Sung[1,2] and **Kristopher Kyle**[1]

[1]Department of Linguistics, University of Oregon, Eugene, OR, USA
[2]Department of Psychology, Rochester Institute of Technology, Rochester, NY, USA
`hksgla@rit.edu, kkyle2@uoregon.edu`

Abstract

Argument structure constructions (ASCs) offer a theoretically grounded lens for analyzing second language (L2) proficiency, yet scalable and systematic tools for measuring their usage remain limited. This paper introduces the ASC analyzer, a publicly available Python package designed to address this gap. The analyzer automatically tags ASCs and computes 50 indices that capture diversity, proportion, frequency, and ASC-verb lemma association strength. To demonstrate its utility, we conduct both bivariate and multivariate analyses that examine the relationship between ASC-based indices and L2 writing scores.

1 Introduction

Linguistic complexity has long been recognized as an important construct in second language (L2) production research. It is commonly conceptualized along two complementary dimensions: absolute complexity and relative complexity (Bulté and Housen, 2012; Bulté et al., 2025). Absolute complexity refers to the structural properties of language, where complexity increases with the number and interrelation of constituent units. In contrast, relative complexity pertains to the cognitive effort involved in using particular forms, typically operationalized via their relative frequency and the strength of their statistical contingencies. To date, a wide range of lexicogrammatical units have been proposed to quantify complexity dimensions, including argument structure constructions (ASCs).

ASCs are clausal-level lexicogrammatical patterns, each anchored by a main verb and a specific argument configuration (e.g., Goldberg, 1995, 2013; Diessel, 2004; Ellis and Larsen-Freeman, 2009). In L2 research, two main approaches have examined their linguistic complexity. One builds on Goldberg's (1995) inheritance hierarchy, which organizes ASCs by semantic role complexity and posits that learners acquire them in a developmental sequence—from simpler constructions (e.g., simple transitives) to more complex ones (e.g., transitive resultatives). Empirical studies have operationalized this trajectory by analyzing the diversity or proportion of ASCs in learner texts (e.g., Hwang and Kim, 2023; Kim et al., 2023).

The other line of research focuses on the relationship between verbs and constructions. It posits that language learners initially tend to produce ASCs with semantically prototypical verbs (i.e., those that strongly instantiate verb-specific argument patterns), which gradually generalize into more abstract constructions (Ninio, 1999). For instance, learners may first acquire the ditransitive construction using prototypical verbs like "give" (e.g., "She **gave** him a book"), before extending it to less prototypical verbs such as "offer" or "send". This developmental trajectory has often been assessed using measures such as the relative frequency and statistical contingency between verbs and constructions (e.g., Ellis and Ferreira-Junior, 2009; Kyle and Crossley, 2017). While a growing body of empirical research has supported both developmental patterns (§ 2.1), scalable and systematic tools for extracting and analyzing ASC-based indices remain underdeveloped.

To address this gap, we present ASC analyzer, a Python package that leverages a RoBERTa-based ASC tagger (Sung and Kyle, 2024b) trained on a gold-standard ASC treebank (Sung and Kyle, 2024a). The tool automatically labels ASCs and computes a suite of indices capturing their diversity, proportion, frequency, and ASC-verb lemma association strength. We also demonstrate the application of the tool through a sample analysis of 6,482 English learner essays from the ELLIPSE corpus (Crossley et al., 2023), examining the relationship between ASC-based indices and L2 English writing proficiency.

2 Background

2.1 Empirical findings on ASC usage in L2 production

From a usage-based constructionist perspective, language is a network of form-meaning pairings (i.e., constructions) that emerge through repeated exposure and use (Fillmore, 1988; Goldberg, 1995; Langacker, 1987). The constructions develop from actual language use and are shaped by patterns of frequency, distribution, and co-occurrence in the input and output (Bybee, 2010; Diessel, 2015; Ellis, 2012; Stefanowitsch and Gries, 2003). As learning is usage-driven, linguistic knowledge accumulates incrementally, shaped by each learner's unique language experience. Empirical studies in this framework examine constructions at varying levels of granularity (e.g., words, phrases, clauses, discourse), with particular attention to clausal-level ASCs, which are schematic form–meaning pairings that encode core semantic relations such as motion, causation, and transfer (Goldberg, 1995).

A body of L2 research has illustrated how ASC usage can be investigated across different L2 modalities and proficiency scores. In L2 writing, for example, Hwang and Kim (2023) found that more proficient L2 writers tend to produce a higher proportion of complex ASCs such as resultatives. Kim et al. (2023) further demonstrated that ASC-based indices outperform traditional T-unit measures in predicting writing proficiency. Another line of research highlights the role of verb-construction pairings. Kyle and Crossley (2017) found that L2 essay scores were negatively correlated with the relative frequency of these pairings but positively correlated with their strength of association, suggesting that advanced learners favor less frequent but more strongly associated verb–construction combinations.

Although less studied, L2 speaking shows similar patterns. Choi and Sung (2020) found that ASC use (especially transitive constructions) explained most of the variance in L2 fluency. Kim and Ro (2023) reported that advanced L2 speakers produced a wider range of verb–construction combinations. A recent study by Sung and Kyle (2025) further confirmed these findings, showing that ASC-based indices alone accounted for 44% of the variance in L2 oral proficiency scores.

2.2 ASC tagger

In this context, reliable identification of ASCs is essential for investigating their relationship with L2 proficiency in large-scale learner corpora (Kyle and Sung, 2023). To meet this need, prior studies have explored a range of tagging methodologies, including dependency parsing (e.g., O'Donnell and Ellis, 2010; Römer et al., 2014; Kyle and Crossley, 2017), rule-based approaches built on top of dependency structures (e.g., Hwang and Kim, 2023; Kim et al., 2023), and methods that leverage semantic role labels (e.g., Jeon, 2024; Kyle and Sung, 2023). Of particular relevance, Kyle and Sung (2023) introduced a supervised ASC tagger trained on a treebank that integrates semantic information across key construction types. Their system targeted nine ASC types, each defined by a characteristic mapping between semantic and syntactic frames (Appendix A).

Building on this supervised training approach and the selected ASC types, Sung and Kyle (2024b) evaluated multiple training strategies and found that a RoBERTa-based tagger trained on a combined L1 and L2 gold-standard treebank (Sung and Kyle, 2024a) achieved high F1 scores across L2 writing (0.915) and L2 speaking (0.928) domains. The results suggest that the tagger is reasonably robust across L2 production modes, providing a foundation for downstream tools that compute ASC-based indices for corpus-based L2 proficiency analysis.

3 ASC analyzer architecture

ASC analyzer is designed to compute interpretable indices that quantify the use of ASCs in English texts. Based on ASC annotations generated by the ASC tagger, the analyzer transforms these labels into a set of operational metrics.

As illustrated in Figure 1, the analyzer processes ASC-tagged output from input texts and calculates four families of indices. Diversity and proportion are text-internal measures that reflect the range and distribution of ASC types and ASC-verb lemma pair types within each text. In contrast, frequency and strength of association (SOA) are text-external measures, computed by comparing ASC usage in the input texts to norms from reference corpora, capturing how often input texts include common or strongly associated ASC-verb combinations. Note that based on the F1 scores reported in Sung and Kyle (2024b), the Gold L1+L2 model was used to

Input

For example, a student who made a presentation was said to do very well at the front of the class. But people who said that might be his friends just to make him feel good. Nevertheless, self-esteem comes from achievement.

ASC tagger (L1+L2)

ASC-tagged text

For example, a student who made a presentation was said to do very well at the front of the class.

But people who said that might be his friends just to make him feel good.

Nevertheless, self-esteem comes from achievement.

ASC analyzer

diversity | frequency
proportion | strength of association

Reference corpus

ASC tagger (L1)

Processed databases bundled in the analyzer

Extracted norms

a snippet of ASC-verb norm database

A set of Indices calculated for each construct

Output
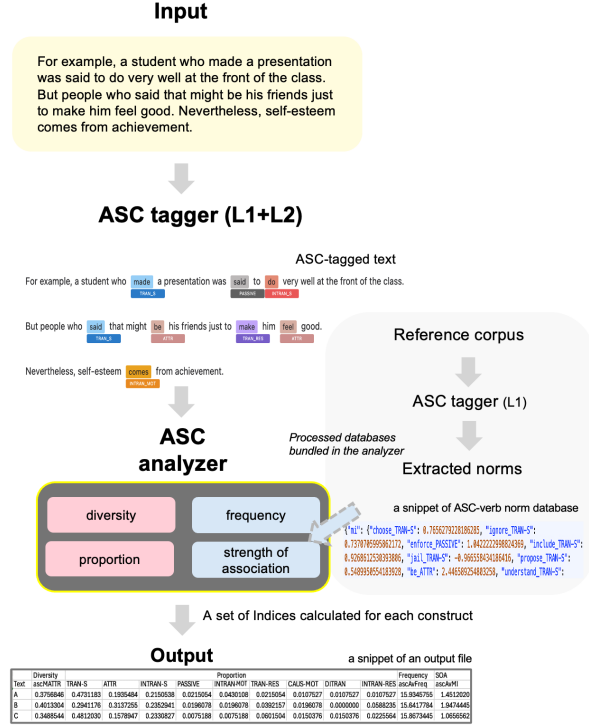
a snippet of an output file

Figure 1: High-level architecture of ASC analyzer

process input texts due to its higher accuracy in L2 contexts, while the Gold L1 model was applied to the reference corpus for its more stable performance in L1 contexts. See Appendix B for detailed F1 scores of each tagger.

3.1 ASC-based Indices

Below we formalize each index family. Implementations in Python follow the equations verbatim.

Diversity. The moving-average type-token ratio (MATTR; Covington and McFall, 2010) with a sliding window w (default: 11) for a token sequence X of length N is defined as:

$$\text{MATTR}_w(X) = \frac{1}{N-w+1} \sum_{i=1}^{N-w+1} \frac{|\text{types}(X_{i:i+w-1})|}{w},$$
$$\text{if } N \geq w+1,$$

We derive three variants: $ascMATTR$ (ASC tokens), $ascLemmaMATTR$ (ASC-verb lemma pairs), and $ascLemmaMATTRNoBe$ (ASC-verb lemma pairs excluding *be*).

Proportion. For each construction type c, we define its proportion (Hwang and Kim, 2023) in the text as

$$\text{Prop}_c(X) = \frac{f_c}{N_{\text{ASC}}},$$

where f_c is the number of tokens of type c in X, and N_{ASC} is the total number of ASC tokens in X. This yields nine variants, one per ASC type (e.g., $ATTR\_Prop$).

Frequency. Let an input text contain M tokens $t_1, \ldots, t_M$, where each t_i is matched up to its raw frequency $f^{\text{ref}}(t_i)$ in a reference corpus (excluding types with $f^{\text{ref}} < 5$). Defining

$$\ell(t_i) = \ln\big(f^{\text{ref}}(t_i)\big),$$

we compute a frequency index as:

$$\text{Freq} = \frac{1}{M} \sum_{i=1}^{M} \ell(t_i).$$

Two variants are derived, depending on the selected token sets: $ascAvFreq$ (ASC tokens) and $ascLemmaAvFreq$ (ASC verb-lemma pairs), following the approach of Kyle and Crossley (2017).

SOA. For each ASC–verb lemma pair (c, v), SOA scores are computed from frequency counts in a reference corpus, where $a = f_{c,v}$, $b = f_{\bar{c},v}$, $c = f_{c,\bar{v}}$, and $d = f_{\bar{c},\bar{v}}$, with total corpus size $N = a + b + c + d$. The expected frequency of the pair is given by:

$$\text{E}(c, v) = \frac{(a+b)(a+c)}{N}$$

Based on these values, we define four pointwise association metrics: mutual information (MI), t-score (T), and two ΔP values:

$$\text{MI}(c, v) = \log_2 \left(\frac{a}{\text{E}(c, v)} \right)$$

$$\text{T}(c, v) = \frac{a - \text{E}(c, v)}{\sqrt{a}}$$

$$\Delta P_{\text{Lemma}}(c, v) = \frac{a}{a+b} - \frac{c}{c+d}$$
$$\Delta P_{\text{Structure}}(c, v) = \frac{a}{a+c} - \frac{b}{b+d}$$

We derive two text-level SOA indices: $ascAv*$, the mean score across all ASC-lemma tokens (e.g., $ascAvMI$), and $t*$, a type-specific mean computed only over tokens labeled with ASC type t (e.g., $ATTR\_AvMI$). This indexing approach follows Gries and Ellis (2015) and Kyle and Crossley (2017).

3.2 Reference corpora for norm extraction

As briefly explained, frequency and SOA are text-external measures that reflect how closely an input text aligns with constructional norms from large external corpora. In its current version, the analyzer draws on two reference corpora:

cow We used a subset of the English Corpus of the Web (EnCOW; Schäfer, 2015; Schäfer and Bildhauer, 2012). It contains 360,783,433 tokens, 15,439,673 sentences, and 39,838,785 automatically tagged ASCs.

subt We used the SUBTLEX-US corpus of American film and television subtitles (Brysbaert et al., 2012; Brysbaert and New, 2009). The version used here comprises 76,965,430 tokens, 164,686 word types, 5,128,462 sentences, and 5,665,251 tagged ASCs across 8,388 subtitle files.

4 Using ASC analyzer: From installation to application

4.1 Installation and quick start

First, install the required dependencies and the ASC analyzer package:

```
pip install spacy
pip install spacy-transformers
python -m spacy download en_core_web_trf
pip install asc-analyzer
```

Next, view the available options:

```
python3 -m asc_analyzer.cli --help
```

To analyze a directory of input texts and save the features to CSV, run:

```
python3 -m asc_analyzer.cli \
  --input-dir "/path/to/texts" \
  --output-csv "/path/to/output.csv" \
  --source "cow"    # or "subt"
```

4.2 Application: ELLIPSE Corpus

To demonstrate the utility of the ASC analyzer in L2 research, we conducted both bivariate and multivariate analyses using a large-scale ESL writing dataset. We used 6,482 essays from the ELLIPSE corpus (Crossley et al., 2023), a reliability-filtered subset of U.S. statewide writing assessments spanning grades 8–12 across 29 prompts. Each essay includes six analytic scores for cohesion, syntax, vocabulary, phraseology, grammar, and conventions. To construct a composite proficiency index, we averaged the four subscores most aligned with constructional usage (syntax, vocabulary, phraseology,

and grammar). The constructional norms were derived from the COW to compute frequency and SOA indices.

4.3 Modeling the relationship between ASC use and L2 writing proficiency

Bivariate correlations: Pearson correlations were computed between each ASC-based index and the composite writing score, retaining only those with $|r| \geq 0.10$ (Cohen, 2013).[1] As shown in Table 1, $ascMATTR$ yielded the strongest positive correlation ($r = 0.26$), while the frequency-based index $ascAvFreq$ showed the strongest negative correlation ($r = -0.22$). Although the correlations were modest overall, the results align with previous findings: more proficient L2 writers tend to use a wider variety of ASC types (Hwang and Kim, 2023) and rely less on highly frequent, but strongly entrenched, verb–construction pairings—except in the case of simple transitives (Kyle and Crossley, 2017).

Construct	Index	r
Diversity	ascMATTR	**.26**
	ascLemmaMATTR	.16
	ascLemmaMATTRNoBe	.11
Proportion	ATTR_Prop	-.11
	CAUS.MOT_Prop	.06
	DITRAN_Prop	.06
	INTRAN.MOT_Prop	.04
	INTRAN.RES_Prop	.13
	INTRAN.S_Prop	.06
	PASSIVE_Prop	.19
	TRAN.RES_Prop	.16
	TRAN.S_Prop	-.13
Frequency	ascAvFreq	**-.22**
	ascLemmaAvFreq	-.15
SOA	asc_AvMI	.12
	CAUS.MOT_AvMI	.09
	DITRAN_AvMI	.11
	INTRAN.MOT_AvMI	.08
	INTRAN.RES_AvMI	.20
	INTRAN.S_AvMI	.11
	PASSIVE_AvMI	.15
	TRAN.RES_AvMI	.14
	TRAN.S_$\Delta P_{Structure}$	-.14

Table 1: Correlations between ASC-based indices and L2 writing scores

Multivariate regression: Indices that passed the bivariate filter were entered into an AIC-based

---

[1] Within each SOA family, we retained only the index most strongly correlated with the scores.

model selection procedure ($\Delta\mathrm{AIC} < 4$; Akaike, 2003; Tan and Biswas, 2012), with multicollinearity controlled beforehand. The final model retained 12 ASC-based predictors and explained a modest proportion of variance in writing scores ($R^2_{\mathrm{adj}} = 0.143$; Table 2), with an overall correlation of $r \approx 0.38$.

Predictor	Estimate	SE	t	p	Rel. Imp. (%)
Intercept	3.001	0.106	28.26	<.001	–
ascMATTR	0.892	0.204	4.37	<.001	16.4
ATTR_Prop	-0.902	0.106	-8.48	<.001	8.1
DITRAN_AvMI	0.013	0.003	4.58	<.001	4.2
INTRAN.RES_AvMI	0.036	0.004	9.95	<.001	15.2
INTRAN.RES_Prop	-1.082	0.502	-2.15	.031	3.4
INTRAN.S_AvMI	0.052	0.007	7.32	<.001	6.1
PASSIVE_AvMI	0.052	0.006	8.10	<.001	9.2
PASSIVE_Prop	2.242	0.294	7.62	<.001	12.2
TRAN.RES_AvMI	0.023	0.005	5.09	<.001	5.7
TRAN.RES_Prop	0.650	0.240	2.71	.007	5.9
TRAN.S_ΔP_Structure	-8.372	1.174	-7.13	<.001	7.8
TRAN.S_Prop	-0.518	0.101	-5.14	<.001	5.7
$R^2 = 0.145$ (adj. 0.143); RSE $= 0.521$; $F(12, 6469) = 91.1$, $p < .001$					

Table 2: Summary of the regression model predicting L2 writing scores

4.4 Comparative evaluation with other models

Comparison with a syntactic complexity-based model: To evaluate the explained variance of ASC-based indices, we compared their predictive power against a multivariate model composed of syntactic complexity measures widely used in L1/L2 acquisition research (Hunt, 1965; Lu, 2011; Ortega, 2003).

Drawing from prior studies (Biber et al., 2016; Kyle, 2016; Kyle and Crossley, 2017), syntactic complexity was operationalized using a set of text-internal indices that capture structural elaboration and grammatical maturity in learner writing. These indices were grouped into three broad categories based on the syntactic units they quantify: (1) Unit length, which includes measures such as the mean length of clause; (2) Clausal complexity, which captures the frequency and depth of embedded clause constructions; and (3) Phrasal complexity, which reflects the internal modification and elaboration of noun phrases. The full list of representative indices is provided in Appendix C.

Following the same procedure used for the ASC-based indices, we built a regression model using the suite of syntactic complexity indices. The final model yielded a lower adjusted $R^2 = 0.077$. This comparison suggests that, for this corpus, ASC-based indices accounted for a greater portion of the variance in L2 writing scores than models based solely on syntactic complexity measures.

Comparison with an alternative lexicogrammatical complexity model: In studies of this kind, it is also important to examine whether newly proposed indices offer unique explanatory power beyond existing measures of lexicogrammatical complexity. To address this, we compared the ASC-based indices against a second baseline model composed of well-established lexicogrammatical indices, which primarily capture complexity at the word and bigram levels (Bulté et al., 2025).

Following prior research (Kyle and Eguchi, 2023; Paquot, 2018), the lexicogrammatical indices in this study fall into three main categories: (1) Syntactic dependency bigrams, which measure the SOA between syntactically linked words (e.g., verb and object pairs); (2) Contiguous lemmatized bigrams, which capture lexical co-occurrence patterns independent of syntactic structure; and (3) Word-level indices, reflecting lexical sophistication (e.g., frequency, concreteness, contextual and associative distinctiveness) and diversity. Full descriptions of these indices are provided in Appendix D.

The baseline model, which included only word- and phrase-level lexicogrammatical indices, achieved an adjusted $R^2 = 0.363$, while the combined model incorporating both lexicogrammatical and ASC-based indices yielded a higher adjusted $R^2 = 0.390$, reflecting an increase of $\Delta R^2 = 0.027$. This result suggests that ASC-based indices may capture an additional variance in L2 writing scores, offering complementary insights into constructional aspects of language use not fully accounted for by existing lexicogrammatical measures.

5 Conclusion

This study introduced the ASC analyzer, an open-source toolkit designed for L2 researchers and applied linguists interested in examining ASC usage in English texts. Through a proof-of-concept analysis, we demonstrated how ASC-based indices can quantify constructional patterns in L2 writing and examined their relationships with writing proficiency scores. We also compared their explanatory power against traditional syntactic complexity indices and assessed how much additional variance they capture beyond existing lexicogrammatical measures. Additional information about the package is available at `https://github.com/hksung/ASC-analyzer`.

Limitations

Several limitations should be acknowledged. First, the outputs of the ASC tagger may be influenced by model-internal biases and training data limitations, which can affect the accuracy and reliability of the extracted indices. As one reviewer noted, certain ASC types (e.g., intransitive resultatives) were underrepresented in the training data, potentially limiting performance for these low-frequency but pedagogically relevant constructions.

Second, the constructional norms used to calculate frequency and SOA scores were derived from a limited set of reference corpora. While these corpora provide useful native-speaker baselines, they may not fully capture the range of registers or genres present in the target texts.

Third, as proof of concept, this work focused on modeling rather than interpretation and did not conduct a detailed linguistic analysis of ASC usage.

Acknowledgments

References

Hirotugu Akaike. 2003. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Douglas Biber, Bethany Gray, and Shelley Staples. 2016. Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5):639–668.

Marc Brysbaert and Boris New. 2009. Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior research methods*, 41(4):977–990.

Marc Brysbaert, Boris New, and Emmanuel Keuleers. 2012. Adding part-of-speech information to the subtlex-us word frequencies. *Behavior research methods*, 44:991–997.

Bram Bulté and Alex Housen. 2012. Defining and operationalising l2 complexity. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 32:21.

Bram Bulté, Alex Housen, and Gabriele Pallotti. 2025. Complexity and difficulty in second language acquisition: A theoretical and methodological overview. *Language Learning*, 75(2):533–574.

Joan Bybee. 2010. *Language, usage and cognition*. Cambridge University Press.

Jungyoun Choi and Min-Chang Sung. 2020. Utterance-based measurement of l2 fluency in speaking interactions: A constructionist approach. *English Teaching*, 75(1):105–126.

Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.

Michael A Covington and Joe D McFall. 2010. Cutting the gordian knot: The moving-average type–token ratio (mattr). *Journal of quantitative linguistics*, 17(2):94–100.

Scott Crossley, Yu Tian, Perpetual Baffour, Alex Franklin, Youngmeen Kim, Wesley Morris, Meg Benner, Aigner Picou, and Ulrich Boser. 2023. The english language learner insight, proficiency and skills evaluation (ellipse) corpus. *International Journal of Learner Corpus Research*, 9(2):248–269.

Holger Diessel. 2004. *The acquisition of complex sentences*, volume 105. Cambridge University Press.

Holger Diessel. 2015. Usage-based construction grammar. In Ewa Dabrowska and Dagmar Divjak, editors, *Handbook of Cognitive Linguistics*, pages 296–322. De Gruyter Mouton.

Nick C Ellis. 2012. Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual review of applied linguistics*, 32:17–44.

Nick C Ellis and Fernando Ferreira-Junior. 2009. Construction learning as a function of frequency, frequency distribution, and function. *The Modern language journal*, 93(3):370–385.

Nick C Ellis and Diane Larsen-Freeman. 2009. Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 59:90–125.

Charles J Fillmore. 1988. The mechanisms of" construction grammar". In *Annual Meeting of the Berkeley Linguistics Society*, pages 35–55.

Adele E Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Adele E Goldberg. 2013. Argument structure constructions versus lexical rules or derivational verb templates. *Mind & Language*, 28(4):435–465.

Stefan Th Gries and Nick C Ellis. 2015. Statistical measures for usage-based linguistics. *Language Learning*, 65(S1):228–255.

Kellogg W Hunt. 1965. *Grammatical structures written at three grade levels*. 8. National Council of Teachers of English.

Haerim Hwang and Hyunwoo Kim. 2023. Automatic analysis of constructional diversity as a predictor of efl students' writing proficiency. *Applied Linguistics*, 44(1):127–147.

Jeeyoung Jeon. 2024. A corpus-based analysis of english argument structure constructions in esl learner writings. Master's thesis, Seoul National University. Unpublished master's thesis.

Hyunwoo Kim and Eunseok Ro. 2023. Assessment of sentence sophistication in l2 spoken production: Expansion of verbs and argument structure constructions. *System*, 119:103175.

Hyunwoo Kim, Gyu-Ho Shin, and Min-Chang Sung. 2023. Constructional complexity as a predictor of korean efl learners' writing proficiency. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 36(2):436–466.

Kristopher Kyle. 2016. *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Ph.D. thesis, Georgia State University, Atlanta, GA.

Kristopher. Kyle and Scott Crossley. 2017. Assessing syntactic sophistication in l2 writing: A usage-based approach. *Language Testing*, 34(4):513–535.

Kristopher Kyle and Masaki Eguchi. 2023. Assessing spoken lexical and lexicogrammatical proficiency using features of word, bigram, and dependency bigram use. *The Modern Language Journal*, 107(2):531–564.

Kristopher Kyle and Hakyung Sung. 2023. An argument structure construction treebank. In *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+ NLP, GURT/SyntaxFest 2023)*, pages 51–62.

Ronald W Langacker. 1987. Nouns and verbs. *Language*, pages 53–94.

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.

Anat Ninio. 1999. Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of child language*, 26(3):619–653.

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.

Matthew O'Donnell and Nick Ellis. 2010. Towards an inventory of english verb argument constructions. In *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, pages 9–16.

Magali Paquot. 2018. Phraseological competence: A missing component in university entrance language tests? insights from a study of efl learners' use of statistical collocations. *Language Assessment Quarterly*, 15(1):29–43.

Ute Römer, Matthew Brook O'Donnell, and Nick C Ellis. 2014. Second language learner knowledge of verb–argument constructions: Effects of language transfer and typology. *The Modern Language Journal*, 98(4):952–975.

Roland Schäfer. 2015. Processing and querying large web corpora with the cow14 architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora*, pages 28–34.

Roland Schäfer and Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 486–493, Istanbul, Turkey. European Language Resources Association (ELRA).

Anatol Stefanowitsch and Stefan Th Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International journal of corpus linguistics*, 8(2):209–243.

Hakyung Sung and Kristopher Kyle. 2024a. Annotation scheme for English argument structure constructions treebank. In *Proceedings of The 18th Linguistic Annotation Workshop (LAW-XVIII)*, pages 12–18, St. Julians, Malta. Association for Computational Linguistics (ACL).

Hakyung Sung and Kristopher Kyle. 2024b. Leveraging pre-trained language models for linguistic analysis: A case of argument structure constructions. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7302–7314, Miami, Florida, USA. Association for Computational Linguistics.

Hakyung Sung and Kristopher Kyle. 2025. Usage-based analysis of l2 oral proficiency: Characteristics of argument structure construction use. *Studies in Second Language Acquisition*, page 1–27.

Y.Ĩ. Tan, M. and Rahul Biswas. 2012. The reliability of the akaike information criterion method in cosmological model selection. *Monthly Notices of the Royal Astronomical Society*, 419(4):3292–3303.

A   Target ASCs and semantic-syntactic representations

This table is reproduced from Table 1 in Sung and Kyle (2024a).

ASC (Tag)	Semantic frame	Syntactic frame
Attributive (ATTR)	theme–VERB–attribute	nsubj–cop–root
Caused-motion (CAUS_MOT)	agent–VERB–theme–destination	nsubj–root–obj–obl
Ditransitive (DITRAN)	agent–VERB–recipient–theme	nsubj–root–iobj–obj
Intransitive motion (INTRAN_MOT)	theme–VERB–goal	nsubj–root–obl
Intransitive simple (INTRAN_S)	agent–VERB	nsubj–root
Intransitive resultative (INTRAN_RES)	theme–VERB–result	nsubj–root–advmod
Passive (PASSIVE)	theme–aux–$V_{passive}$	nsubj:pass–aux:pass–root
Transitive simple (TRAN_S)	agent–VERB–theme	nsubj–root–obj
Transitive resultative (TRAN_RES)	agent–VERB–theme–result	nsubj–root–obj–xcomp

B   F1 scores across ASC types by model and domain

This table, adapted from Table 2 in Sung and Kyle (2024b), reports F1 scores by ASC tag across two taggers: one trained only on the L1 treebank (Gold L1) and another trained on a combined L1+L2 treebank (Gold L1+L2). Each model is evaluated on three test sets (L1, L2 writing, and L2 speaking) to assess cross-domain robustness.

ASC Tag	Gold L1			Gold L1+L2		
	L1	L2-writing	L2-speaking	L1	L2-writing	L2-speaking
ATTR	0.972	0.954	0.986	0.968	0.971	0.988
CAUS_MOT	0.818	0.833	0.710	0.857	0.867	0.710
DITRAN	0.919	0.914	0.842	0.865	0.881	0.947
INTRAN_MOT	0.800	0.770	0.789	0.772	0.807	0.843
INTRAN_RES	0.750	0.788	0.800	0.625	0.813	0.833
INTRAN_S	0.779	0.806	0.817	0.808	0.803	0.865
PASSIVE	0.920	0.775	0.938	0.940	0.865	0.909
TRAN_RES	0.884	0.800	0.625	0.881	0.792	0.625
TRAN_S	0.931	0.929	0.927	0.936	0.943	0.948
Weighted Avg.	0.908	0.900	0.905	0.912	0.915	0.928

C   Syntactic complexity indices

Dimension	Index	Description
Clause	mlc	Average number of words per finite clause
	mltu	Average number of words per T-unit
	dc_c	Number of dependent clauses per clause
	ccomp_c	Frequency of finite complement clauses
	relcl_c	Frequency of relative clauses per clause
	infinitive_prop	Proportion of "to + verb" constructions
	nonfinite_prop	Proportion of nonfinite (gerund/participial) clauses
Phrase	mean_nominal_deps	Average number of nominal dependents per noun
	relcl_nominal	Relative clauses modifying nominals
	amod_nominal	Adjectival modifiers of nominals
	det_nominal	Determiners modifying nominals
	prep_nominal	Prepositional phrases modifying nominals
	poss_nominal	Possessive modifiers of nominals
	cc_nominal	Coordinating conjunctions in noun phrases

D   Word and bigram-level lexicogrammatical indices

Dimension	Index	Description
Bigram	n_amod_{T, MI, MI2, DP*}	SOA scores for noun–adjective dependencies
	v_advmod_{T, MI, MI2, DP*}	SOA scores for verb–adverb dependencies
	v_dobj_{T, MI, MI2, DP*}	SOA scores for verb–object dependencies
	v_nsubj_{T, MI, MI2, DP*}	SOA scores for verb–subject dependencies
	lemma_bg_{T, MI, MI2, DP*}	SOA scores for lemmatized word bigrams
Word	amod_freq_log	Log frequency of adjectives
	advmod_freq_log	Log frequency of adverbs
	adv_manner_freq_log	Log frequency of manner adverbs
	mverb_freq_log	Log frequency of main verbs
	lex_mverb_freq_log	Log frequency of lexical main verbs
	noun_freq_log	Log frequency of nouns
	cw_lemma_freq_log	Log frequency of content word lemmas
	b_concreteness	Word concreteness ratings
	mcd	Contextual distinctiveness (entropy-based)
	usf	Associative distinctiveness (from USF norms)
	MATTR_11	Moving-average type–token ratio (window = 11)

Note. DP* indicates various types of ΔP scores, computed using either the left or right word as the cue (or the head or dependent in the case of dependency bigrams). All scores were calculated, and for the regression model, only the score showing the strongest relationship with scores was included—consistent with the treatment of the SOA scores in the baseline model (see Footnote 1).