An NLP tool for the analysis of constructional diversity

Haerim Hwang
Korea University
(haerimhwang@korea.ac.kr, haerimhwang.github.io)

2 October 2021
ALAK 2021

Overview

- Literature review
- Development of an NLP tool: Constructional Diversity Analyzer

  Hwang, Choe & Zenker (2020)
- Validation/Application of the indices included in the Constructional Diversity Analyzer

  Hwang & Kim (under review)
- Discussion

  Hwang & Kim (under review)

Literature Review

Traditional syntactic complexity indices

Category	Index name	Index description
Length of production	Mean length of clause	Number of words per clause
	Mean length of sentence	Number of words per sentence
	Mean length of T-unit	Number of words per T-unit
Sentence complexity	Clauses per sentence	Number of clauses per sentence
Subordination	Clauses per T-unit	Number of clauses per T-unit
	Complex T-unit ratio	Number of complex T-units per T-unit
	Dependent clauses per clause	Number of dependent clauses per clause
	Dependent clauses per T-unit	Number of dependent clauses per T-unit
Coordination	Coordinate phrases per clause	Number of coordinate phrases per clause
	Coordinate phrases per T-unit	Number of coordinate phrases per T-unit
	T-units per sentence	Number of T-units per sentence
Particular structures	Complex nominals per clause	Number of complex nominals per clause
	Complex nominals per T-unit	Number of complex nominals per T-unit
	Verb phrases per T-unit	Number of verb phrases per T-unit

Lu (2010)

Length of production

- *Mean length of T-unit*
  - Significantly explained the variation in the scores of TOEFL essays collected from adult L2 learners with diverse L1 backgrounds

Cumming, Kantor, Baba, Erdosy, Eouanzoui & James (2005)

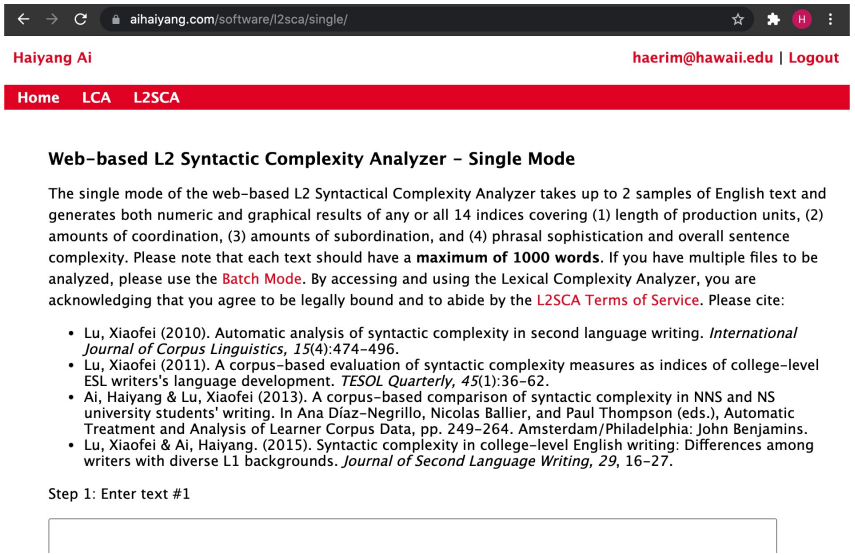
Rate of subordination

- *Mean number of dependent clauses per clause*
  - Significantly accounted for the quality of narratives produced by Spanish-speaking L2 learners of English

Lorenzo & Rodríguez (2014)

Lu (2011)

- L2 Syntactic Complexity Analyzer



https://aihaiyang.com/software/l2sca/

Lu (2011)

- In the analysis of the essays written by college-level Chinese L2 learners of English, Lu found several syntactic complexity measures as strong indicators of L2 proficiency
- *Complex nominals per clause* and *mean length of clause* were identified as the best indices for differentiating between proficiency levels, followed by *complex nominals per T-unit*, *mean length of sentence*, *mean length of T-unit*, *coordinate phrases per clause*, and *coordinate phrases per T-unit*

Research gap

- Theoretical validity:
  e.g. Length-based indices are more associated with writing fluency than with complexity and thus fall short of capturing overall L2 writing proficiency, which subsumes multidimensional constructs including fluency, accuracy, and complexity

  Norris & Ortega (2009)

- The traditional "large-grained" indices obscure the more sophisticated linguistic variation in development

  Kyle & Crossley (2017)

Kyle & Crossley (2017)

- Proposal of an alternative: Syntactic sophistication indices
- Usage-based constructionist approaches
  - Argument structure construction (hereafter, **construction**):
    a form–meaning pairing that denotes a clause's meaning

    Goldberg (1995)
  - Frequent constructions are learned earlier and more easily
  - Constructions are initially used with a prototypical
    path-breaking verb

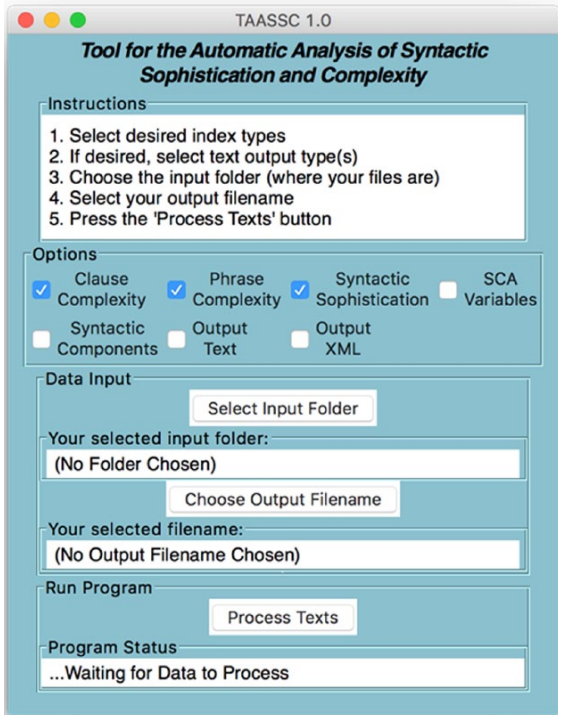
    Ninio (1999)

Kyle & Crossley (2017)

- Finer-grained indices of syntactic sophistication
  - *Frequency of constructions*
  - *Frequency of main verb lemmas*
  - *Frequency of verb-construction combinations*
  - *Association strength of verb-construction combinations*
  - Many others

Kyle & Crossley (2017)

- Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC)



Kyle (2016), https://www.linguisticanalysistools.org/taassc.html

Kyle & Crossley (2017)

- The analysis of TOEFL essays from L2 learners of English with diverse L1 backgrounds revealed that usage-based sophistication indices explained greater proportion of the variance (R^2 = .142) in essay scores than traditional syntactic complexity indices (R^2 = .058)

→ Validity of the syntactic sophistication indices as alternative measures of syntactic complexity

Research gap

- Constructions defined in Kyle and Crossley (2017) are closer to constructional patterns rather than abstract argument structure constructions defined in Goldbergian Construction Grammar:

  e.g. *He gave me the pictures* vs. *He gave me the pictures in the morning*

  [subject–verb–indirect object–direct object] *[subject–verb–indirect object–direct object–prepositional phrase]*

- There are currently no NLP tools that compute the **diversity of different constructions**

Development of an NLP tool: Constructional Diversity Analyzer

Hwang, Choe & Zenker (2020)

Assumptions

- Usage-based approaches to language acquisition
  - Less frequent constructions are learned later
  - **Constructional diversity** increases as learner proficiency goes up

Ellis, O'Donnell & Römer (2013), Tomasello (2003)

→ More proficient learners can produce constructions with lower frequencies, thus showing higher **constructional diversity**

→ **Constructional diversity** can serve as a reliable indicator of language development/proficiency

Goal

- To develop automated indices related to **constructional diversity** as measures of syntactic complexity indicating L2 development

  → To this end, I developed the application computing the **constructional diversity** and **proportion of individual constructions** (as well as verbal diversity) in Python

an updated version of the Python code reported in Hwang, Choe & Zenker (2020)

Constructional Diversity Analyzer (CDA)

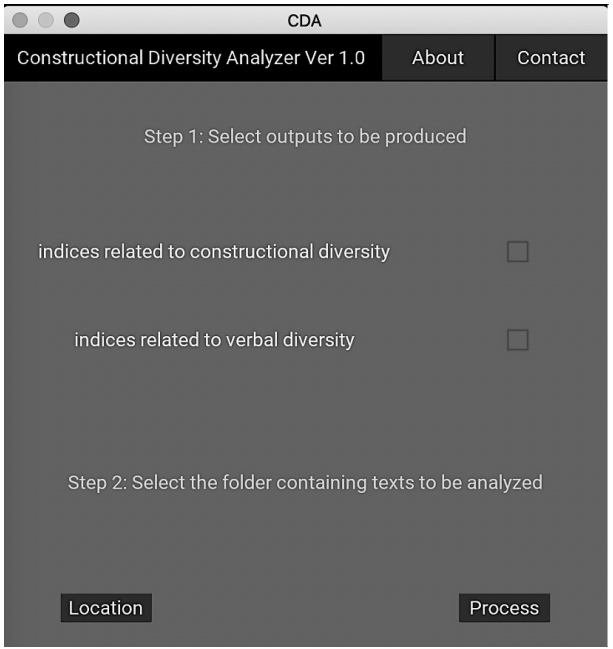


Figure 1. Graphical User Interface of the Constructional Diversity Analyzer

https://haerimhwang.github.io/tools/English-constructional-diversity-analyzer

Calculating constructional diversity in CDA

1. Label each clause in the text as one of the 11 constructions using the spaCy module

Honnibal & Montani (to appear)

Constructions in CDA

- The first six constructions come from Construction Grammar

Goldberg (1995)

  - Caused-motion: *Mary faxed a letter to John.*
  - Ditransitive: *Mary faxed John a letter.*
  - Intransitive motion: *The fly buzzed into the room.*
  - Intransitive resultative: *The pond froze solid.*
  - Phrasal verb: *The girl looked the name up.*
  - Transitive resultative: *The boy painted the barn red.*

Constructions in CDA

- The other five are common English constructions

Hornby (1978), Tomasello (2003)

  - Attributive: *The boy is a student.*
  - Passive: *The sheet was folded.*
  - Simple intransitive: *The man worked.*
  - Simple transitive: *The man kicked the ball.*
  - *There*-expletive: *There is a house.*

Classification algorithms in CDA

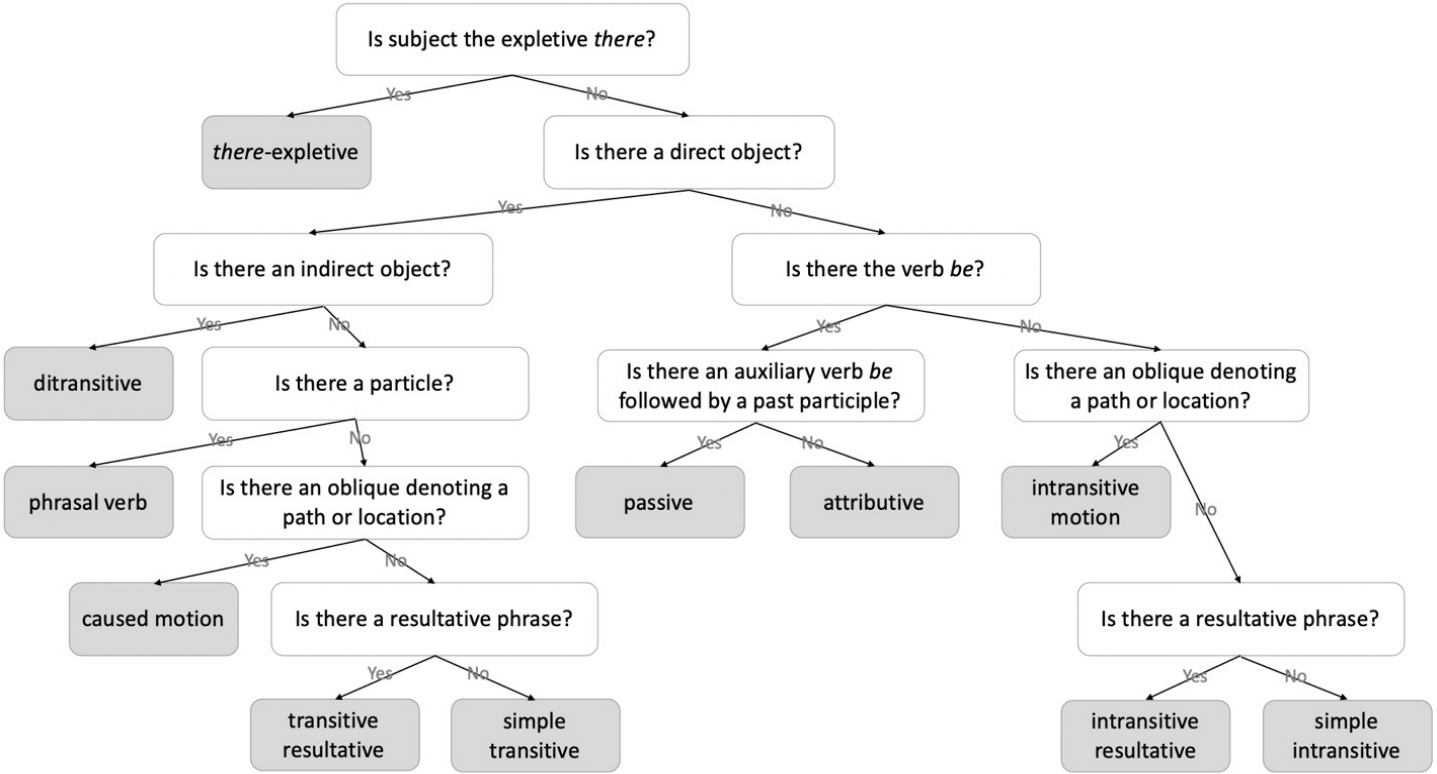


Figure 2. Classification algorithms applied in CDA

Calculating constructional diversity in CDA

1. Label each clause in the text as one of the 11 constructions using the spaCy module

Honnibal & Montani (to appear)

2. Calculate log_{10} *(type frequency + 1)* for each text

Laplace smoothing technique
Manning, Raghavan & Schutze (2008)

data normalization
Compton, Gilbert, Kearns & Olson (2020)

Measuring the use of individual constructions in CDA

1. Compute a proportion of each construction
2. Arcsine-transform the proportion for each construction to approximate normality and stabilize the variance of the data

Studebaker (1985)

Evaluating the classification accuracy in CDA

- Compared the tool's output to judgments made by two human annotators on a test dataset of 1,000 clauses from the American National Corpus

Reppen, Ide, & Sunderman (2005)

  - Inter-annotator reliability: Cohen's Kappa = 1.00, after resolving inconsistencies
  - Classification accuracy: recall = 0.82, precision = 0.86, F1 = 0.82

Validation/Application of the indices in the Constructional Diversity Analyzer

Hwang & Kim (under review)

Research questions

- RQ 1:
  Does the **constructional diversity** predict learner writing proficiency?
- RQ 2:
  Which constructions contribute more to predicting learner writing proficiency?

Data

- 3,286 argumentative essays written by university students in Korea (collected through the Yonsei English Learner Corpus project)

Rhee & Jung (2012)

- Topics
  - allowing physical punishment at schools
  - banning smoking at public places
  - going to military service
  - using animals in medical experiments
  - using cellphones while driving
  - using real names on the internet

Data

- Proficiency levels
  - Each essay was rated by trained English native speakers based on the Common European Framework of Reference for Languages, which categorized the essays into one of the 12 levels: A1, A1+, A2, A2+, B1, B1+, B2, B2+, C1, C1+, C2, C2+

    (Council of Europe, 2011)

  - We excluded A2+, C1+, C2, and C2+, which contained less than three essays

Data

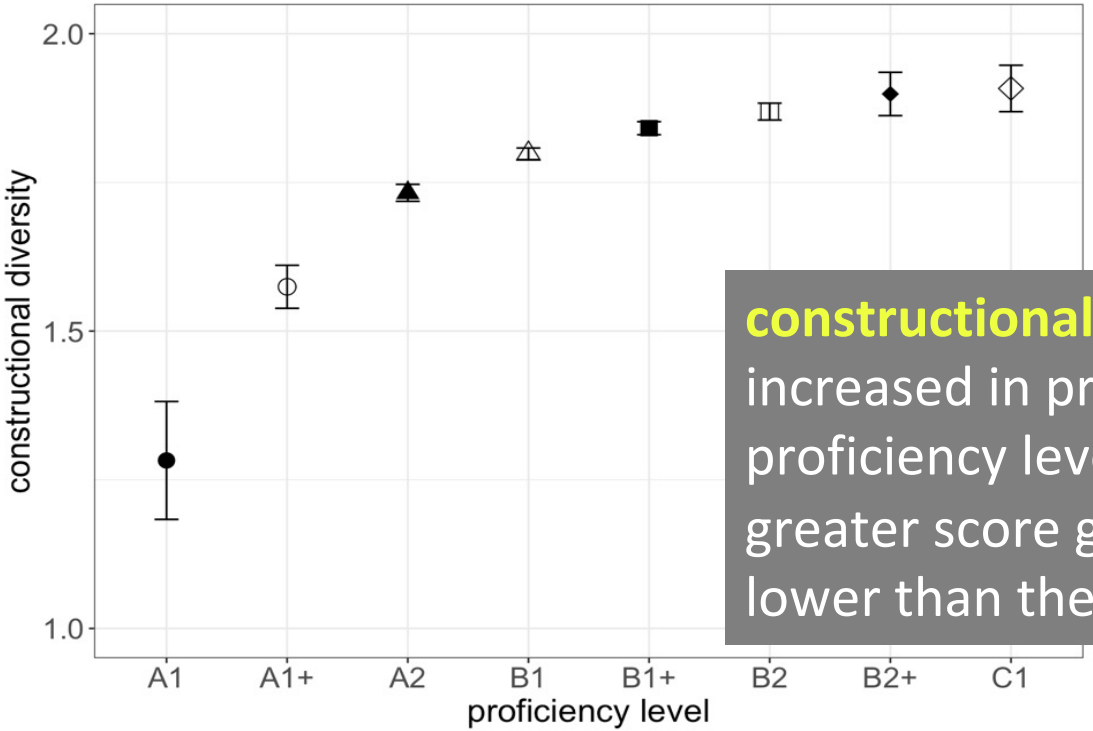
- Information of the sample essays

Proficiency level	Number	Mean number of word tokens (*SD*)	Mean number of clauses (*SD*)
A1	41	38.12 (51.82)	7.17 (9.43)
A1+	185	119.38 (79.36)	19.74 (13.07)
A2	684	194.91 (76.85)	31.58 (13.02)
B1	1173	234.81 (62.94)	37.22 (11.06)
B1+	705	271.27 (40.89)	42.38 (8.55)
B2	378	285.72 (39.82)	43.88 (8.15)
B2+	81	299.62 (44.94)	45.05 (8.04)
C1	37	307.16 (6.36)	45.16 (11.60)

Analysis

- RQ 1:
  Does the **constructional diversity** predict learner writing proficiency?
  → simple linear regression analysis with **constructional diversity** scores as a predictor of proficiency
- RQ 2:
  Which constructions contribute more to predicting learner writing proficiency?
  → stepwise regression analysis that tested the effects of the **proportion of individual constructions** as predictors of the proficiency levels

Results: Constructional diversity



constructional diversity scores increased in proportion to the proficiency levels, with a greater score gap between the lower than the higher levels

Figure 3. Constructional diversity scores by proficiency levels

Results: Constructional diversity

- The simple linear regression analysis revealed a significant effect of **constructional diversity** scores in the variability of the proficiency levels (*Estimate* = 3.307, *SE* = 0.131, *t* = 25.198, *r* = .403, $p < .001$), explaining 16.2% of the variance (R^2 = .162)

→ **Constructional diversity** successfully predicted the proficiency levels

Results: Use of individual constructions

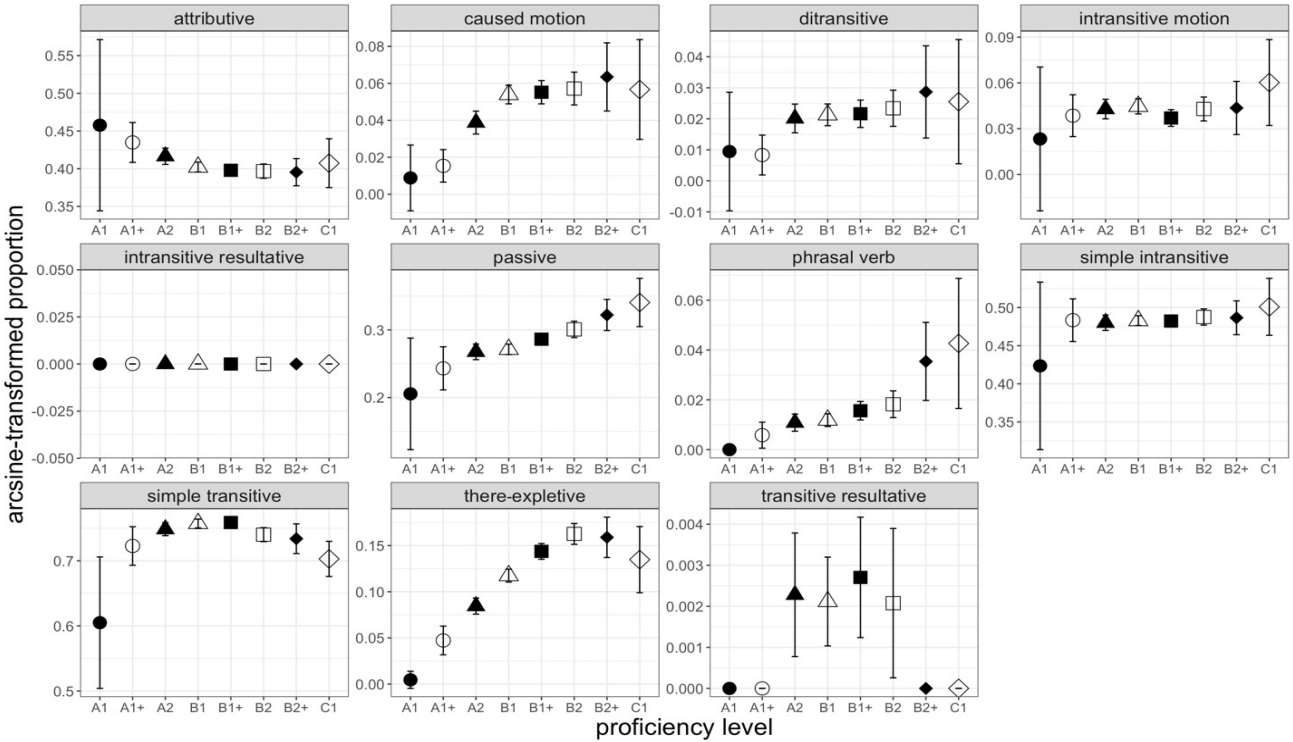
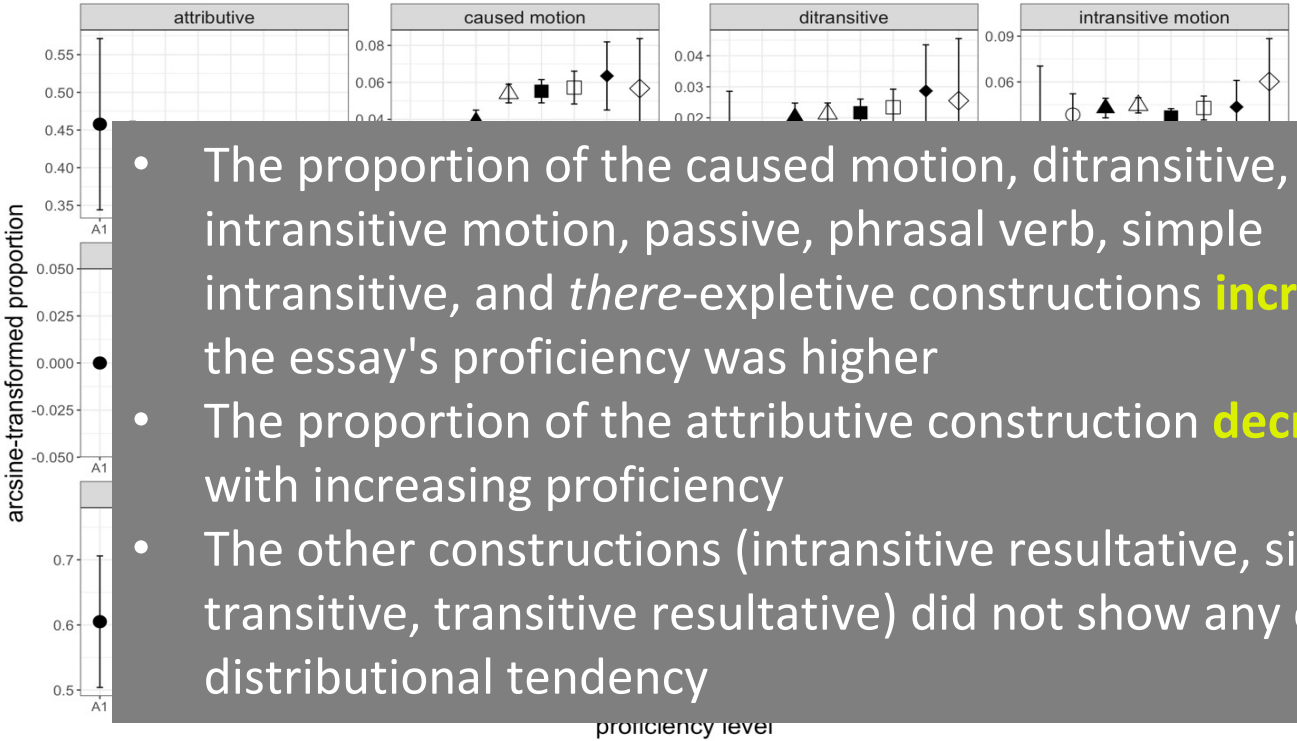


Figure 4. Proportions of individual constructions by proficiency levels

Results: Use of individual constructions



- The proportion of the caused motion, ditransitive, intransitive motion, passive, phrasal verb, simple intransitive, and *there*-expletive constructions **increased** as the essay's proficiency was higher
- The proportion of the attributive construction **decreased** with increasing proficiency
- The other constructions (intransitive resultative, simple transitive, transitive resultative) did not show any clear distributional tendency

Figure 4. Proportions of individual constructions by proficiency levels

Results: Use of individual constructions

- The final model from the stepwise regression analysis included seven constructions, which collectively accounted for 11.5% of the variance in proficiency scores (r = .339, R^2 = .115, p < .001)
- The four most contributive constructions were the *there*-expletive construction, followed by the passive, phrasal verb and caused motion constructions, indicating more instances of these constructions with increasing proficiency levels
- The less contributive predictors were the simple intransitive, ditransitive and attributive constructions

Results: Use of individual constructions

- Except for the ditransitive construction, the relative contributive power of each construction was aligned with the construction's frequency and complexity, with less frequent and more complex constructions contributing more to the model

→ Constructional frequency and complexity reliably predict the proficiency levels

Discussion

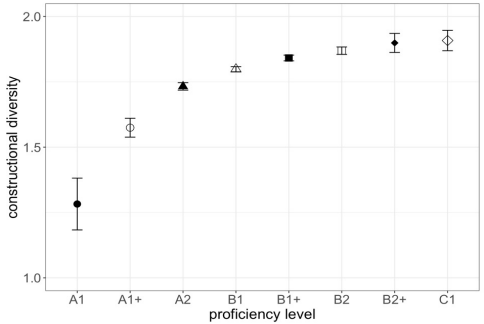
Hwang & Kim (under review)

Validity of constructional diversity

- The predictive value of **constructional diversity** (proposed based on the solid background of the constructionist approach) underscores its validity as an alternative to the traditional syntactic complexity measures, such as length-based and subordination-based indices, which have been criticized for the lack of theoretical motivation for using them to assess L2 writing proficiency

  Bardovi–Harlig (1992), Biber, Gray & Poonpon (2011), Norris & Ortega (2009), Rimmer (2006)

Proficiency effect



- The explanation power was stronger in the essays at lower proficiency levels: Whereas beginning-level learners relied on a few construction types, the **constructional diversity** rapidly increased as they achieved higher proficiency
- This finding is consistent with the constructional development pattern in children

Ellis et al. (2013), Goldberg (2013), Sethuraman (2002), Tomasello (2003)

Complex/infrequent constructions

- The complex constructions exhibited stronger predictive power
  - Passive: a unique configuration with a particular meaning of 'X is affected by Y' and the non-canonical mapping of Patient–Action–Agent to the subject–verb–object word order
  - Caused-motion: In English, the path and manner of motion is encoded separately in a preposition and a verb, whereas the Korean counterpart conflates both path and manner via a combination of verbs
  - Ditransitive: Dispreferred over the prepositional dative construction in Korean

Attributive construction

- In the final model from the stepwise regression anlaysis, only the attributive construction had an inverted relationship with the proficiency levels
- It is conceivable that L2 learners acquire the attributive construction at early stages before acquiring other complicated constructions, given the relatively early mastery of the copula *be* by English-speaking children

Dulay & Burt (1974), Larson–Freeman (1976), Krashen (1982)

[Argumentative Essay from an L2 Learner with the Lowest Proficiency Level (PPT710)]
Title is very hard subject. I don't have the power. I think that is no impact my life. So, I don't think that. It is not important. Very important is our heart. It is the real mind person. We are together!

Next Step

- Improve accuracy of the CDA
- Test the CDA with more data (e.g. spoken data) in diverse contexts (e.g. L1-English context)
- Release the CDA for researchers and language teachers

References

Bardovi–Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*, 390–395.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*, 5–35.

Compton, D. L., Gilbert, J. K., Kearns, D. M., & Olson, R. K. (2020). Using an item-specific predictor to test the dimensionality of the orthographic choice task. *Annals of Dyslexia*, *70*, 243–258.

Council of Europe. (2011). *Manual for language test development and examining for use with the CEFR*. Produced by ALTE on behalf of the Language Policy Division, Council of Europe. Council of Europe.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, *10*, 5–43.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*, 37–53.

Ellis N. C., O'Donnell M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, *63*(s1), 25–51.

References

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

Goldberg, A. E. (2013). *Constructionist approach*. Oxford University Press.

Honnibal, M., &, Montani, I. (2017). spaCy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. https://spacy.io

Hornby, A. S. (1954). *Guide to patterns and usage in English*. Oxford. University Press.

Hwang, H., Choe, A., & Zenker, F. (2020, April). Construction counter: A tool to measure (nonnative) language development. Brown Bag Series, University of Hawai'i at Mānoa, Honolulu, USA.

Hwang, H., & Kim, H. (under review). Automatic analysis of constructional diversity as a predictor of EFL students' writing proficiency. *Applied Linguistics*.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

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

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing, 34*, 513–535.

References

Larsen–Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, *26*, 125–134.

Lorenzo, F., & Rodríguez, L. (2014). Onset and expansion of L2 cognitive academic language proficiency in bilingual settings: CALP in CLIL. *System*, *47*, 64–72.

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

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language*, *26*, 619–654.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in SLA: The case of complexity. *Applied Linguistics, 30*, 555–578.

Reppen, R., Ide, N., & Suderman, K. (2005). *American National Corpus (ANC) second release LDC2005T35* (DVD). Linguistic Data Consortium.

Rhee, S. C., & Jung, C. K. (2012). Yonsei English Learner Corpus (YELC). Paper presented at the First Yonsei Corpus Symposium. Seoul, Korea.

References

Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, *23*, 497–519.

Sethuraman, N. (2002). The acquisition of verbs and argument structure constructions (Unpublished PhD dissertation). University of California, San Diego, CA.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research*, *28*, 455–462.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Special thanks to:

Kristopher Kyle, Yangon Rah, Bonnie D. Schwartz, Hyun-Kwon Yang, and all classmates who took syntax courses at Seoul National University in 2011–2012 and SLS 680R at the University of Hawai'i in Fall, 2018 with me