Knowledge beyond usage

Potential and limitations of cognitive corpus linguistics for pedagogy

David Stringer
Indiana University

Much interesting research has been conducted in cognitive and corpus linguistics that has potential implications both for our understanding of how language works and for how it is put to use. In her keynote paper, Gilquin (2022) presents a thought-provoking argument for combining insights from these two major subfields in a cognitive corpus linguistic approach to pedagogy. While this synthesis is appealing and the argumentation is clearly presented, there are several reasons to rein in this unbridled enthusiasm for exposing learners to ideas whose validity, efficacy, and generalizability remain largely untested. There is ample scope for cognitive corpus linguistics to expand and significantly inform materials development and syllabus design, but first there is a need to engage more meaningfully with established findings and concepts in mainstream linguistics, second language acquisition (SLA), and language pedagogy.

This proposal is engaging in several ways: the argumentation is well-structured, and Gilquin (2022) advocates convincingly for both cognitive and corpus linguistic applications in the classroom, while being frank about certain limitations of both approaches. Cognitive linguistic research on the teaching of prepositions, phrasal verbs, and idioms (e.g., Boers and Lindstromberg, 2006; Tyler and Huang, 2018) has indeed introduced some very exciting new possibilities for language teaching, especially for vocabulary and fixed expressions. Several arguments for the usefulness of corpora in the language classroom are similarly persuasive. Researchers and teachers alike have generally warmed to the idea that corpora can provide authentic language materials in a variety of genres, and that such materials are a more naturalistic kind of input than artificially constructed sentences (though it is unlikely that they can replace all types of input: natural input from contexts of conversational interaction; input from formal, written texts, etc.).

In terms of vocabulary acquisition, the use of concordances (Gilquin, 2021) seems very promising as an intermediary type of input between traditional word

https://doi.org/10.1075/pl.22008.str | Published online: 13 July 2022
Pedagogical Linguistics 3:2 (2022), pp. 166–173. ISSN 2665-9581 | E-ISSN 2665-959X

lists and the kind of multiple exposure that all teachers know is necessary for lexical acquisition, given that all words have multiple senses, contextual nuances, and very often more than one kind of syntactic framing (Stringer, 2019). While not a substitute for natural language input in the course of meaningful activities, "condensed exposure" (Gabrielatos, 2005: 10) through concordances is likely to be effective. It is well understood that successful vocabulary acquisition requires multiple exposure and contextual support (Nash & Snowling, 2006; Stahl, 2005). Concordances could be yet another element in a teacher's toolkit that already includes, for example, structured read-alouds, discussion sessions, and independent reading experiences (Cunningham, 2005).

Gilquin (2022) is refreshingly honest about several limitations on the applications of both cognitive and corpus linguistic research to pedagogy. As for the former, she accurately observes that studies have been mostly restricted to "a few pet subjects", and teaching efficacy has been assessed only on the basis of "small-scale and short-term experiments" (p. 111). Arguably, these lines of research seem relevant primarily to aspects of linguistic knowledge that are represented in the mental lexicon. Gilquin also concedes that the advantages of corpora are not all-encompassing. Corpora are fundamental in the calculation of frequency, and frequency clearly plays a major role in acquisition, but it is widely understood that "frequency is not everything" (p. 115). Indeed, functional items such as determiners, quantifiers, auxiliaries, and inflectional morphemes are among the most frequent items in any corpus, yet have been described as constituting the "bottleneck" for second language (L2) acquisition (Slabakova, 2008, 2016). Gilquin also recognizes that data-driven learning of the type advocated here is time-consuming, certain concepts such as embodiment have restricted application, and that L1-L2 comparisons are not always practicable in all classrooms (p. 185).

Nevertheless, despite an admirable candor, this passionately optimistic keynote paper is somewhat undermined by a notable lack of reference to linguistics and applied linguistics more generally. With regard to linguistic assumptions, it is broadly accepted that there is a huge gap between knowledge and usage. This is not just the somewhat controversial claim of a "poverty of the stimulus" (Hornstein & Lightfoot, 1981; Schwartz & Sprouse, 2013). It is also the more widely accepted observation that language production is prone to error due to time pressure, competing lexical activations, memory lapses, and the limitations of working memory, leading to a gap between what we know and what we say. Beyond the reassurance that the corpus is not intended to supplant the teacher "as a repository of knowledge" (p. 112), it is important to recognize that a corpus is not a repository of knowledge at all, but a repository of usage. Similarly, it cannot be true that "quantitative distributions [...] are part of the grammar" (Bybee,

2010: 122), as is claimed here; distributions are measurements of the output of the grammar, not the grammar itself.

The fact that the concept of grammatical knowledge (as distinct from usage) has gone up in smoke in this analysis is unsurprising given the background literature that informs the proposal. The idea that language is not an automous mental faculty (p. 118) is a core tenet of usage-based approaches to language that deny the very existence of grammar. As argued by Ellis (2003: 63–64), "the knowledge of a speaker/hearer cannot be understood as a grammar, but rather as a statistical ensemble of language experiences that changes slightly every time a new utterance is processed". The acquisition-theoretic correlates of this position are that learning basically involves pattern-finding and then memorization of chunks in the input. However, Bardovi-Harlig & Stringer (2017) have shown that even conventional expressions such as *Sorry I'm late* are not stored and retrieved as unanalyzed chunks by learners, but rather exist as memorized processing routines, often with open slots, and have to be recreated in language production. They analyzed over 9000 responses from three different elicited production tasks, and discovered ample evidence of creative construction, both in terms of interlanguage grammar (e.g., *I'm sorry for lating, I'm sorry for be late, I'm sorry about my late*) and unconventional but grammatical forms (e.g., *Sorry for being late, I am sorry to come late, Sorry for coming late*). They concluded that the production of even high-frequency formulaic expressions reflects – rather than drives – interlanguage grammatical knowledge, and that eventual well-formedness of such expressions is dependent on autonomous syntactic development (p. 84).

It is telling that the idea that "all language units are meaningful" (p. 109) can only be maintained by ignoring mainstream linguistic analyses of syntax and phonology. In corpus linguistics, several "lexical bundles" that are identified as meaningful units, such as *in the middle of the* or *in the front of the* (Tremblay et al., 2011), in fact contain meaningless elements with purely syntactic functions, such as the functional proposition *of*, and are not well-formed products of mental grammar, as they are not built from legitimate phrasal units. *Of* is necessary only because case cannot be assigned by nouns such as *middle* and *front*. (Similarly, we can say *He bakes bread* but *of*-insertion is required in *He is a baker of bread*). Case-assigning *of* heads a prepositional phrase; *the* heads a determiner phrase. The existential reality of syntactic phrase structure has been proven time and again in linguistics through distributional evidence involving substitution or movement, as demonstrated in every introductory syntax textbook. The idea that chunk-based learning can also apply to "fully abstract combinations such as the ditransitive structure [Subj V Obj1 Obj2]" (Gilquin, 2022, p. 114) can only be posited by ignoring decades-worth of research on the dative alternation and reducing the relation between the relevant objects to Thing One and Thing Two (Seuss, 1957).

The idea that we can acquire language simply by paying attention to meaningful units crashes most obviously on the rocks of phonology. The various elements and organizational units of phonology, both segmental and suprasegmental, are never acquired through the processing of meaning. Babies converge on a language-specific phonemic inventory before understanding their first word (Jusczyk, 1997), and eventually elaborate prosodic levels of representation such as phonological and intonational phrases whose boundaries are, importantly, different from syntactic structure. As Carroll (2010, p. 231) has observed, the notion of simply extracting phrases from input is ill-defined in all forms of usage-based linguistics precisely because no distinction is made between syntactic and phonological structures. If word strings acting as input are prosodically bounded, such boundaries cannot be used to infer knowledge of syntax. These comments on both syntax and phonology are *not* for the purpose of denying the importance or relevance of memorizing constructions, from idioms to frequent syntactic patterns, but to stress some limitations when the target for acquisition is the entirety of the grammar.

As stated at the outset, this proposal would also benefit from a more serious engagement with previous and current research in SLA and pedagogy. For example, almost all activities mentioned or exemplified in Gilquin (2022) operate on the assumption that language acquisition involves explicit instruction and conscious learning: the idea of "guided self-learning" (Rudzka-Ostyn, 2003), the noticing and remembering activities of Lindstromberg and Boers (2008), the metalinguistic explanations for prepositions or motion event framing (Langacker, 1987; De Knop & Meunier, 2015), and the idea that learning of periphrastic causatives might be facilitated by showing color-coded word clouds of commonly appearing verbs (blue for descriptive verbs, purple for mental processes, fuchsia for verbs of involuntary movement; Gilquin, 2022, p. 128). Such strategies may assist in the conscious learning of lexical items or more abstract metalinguistic knowledge, but their scope is limited by certain agreed-upon observations in SLA: much of the acquisition of syntax and phonology is unconscious and results in implicit knowledge; learners come to acquire some aspects of linguistic knowledge that cannot be explained in terms of the input; and there are limits on what can acquired through explicit instruction (see, e.g., VanPatten, Keating, & Wulff, 2020).

Another claim that merits re-examination in terms of the SLA literature is the overemphasis on individual differences that typifies usage-based approaches. It is argued here that language acquisition is strictly an individualized process in which there are no common paths of development or endstates (p. 128). Of course, it is true that individual differences exist, in terms of age, motivation, working memory, or personality (for a critical review of the individual differences liter-

ature, see Dornyei & Ryan, 2015). However, to imply that we cannot generalize across thousands of individuals would be absurd. There are clear universal sequences in the acquisition of particular languages that all learners go through, with variations that hold in cases of particular L1s. For summaries of such research on the L2 English acquisition of negation, question formation, possessive pronouns, relative clauses, and past tense marking, see Lightbown & Spada (2013, pp. 51–59).

It is claimed that "[w]hen it comes to teaching applications, corpora offer undeniable advantages over experimental data" (p. 113), such as greater authenticity and contextualization. However, it is not clear that this comparison bears scrutiny. Most pedagogy-related experimentation in SLA tests the validity of learning hypotheses or the efficacy of instructional strategies, and as such, it is not in opposition to either cognitive proposals or corpora. While there is no citation of experiments over which corpora supposedly have an advantage, it is possible to imagine work at which this criticism could be legitimately aimed. For example, the many studies on the effects of working memory capabilities on language learning have arguably no real-world application, as public schools, private institutes, and universities are never going to organize student groups based on this variable, which means that teachers will always have students with variation in this particular capacity. However, there have been so many valuable experimental studies on L2 instructional effects that overview chapters are plentiful (Long, 2015, on interaction; Bardovi-Harlig, 2015, on pragmatics; Granena & Yilmaz, 2021, on Task-Based Language Teaching and corrective feedback), and meta-analyses are often cited (Norris & Ortega, 2000, on implicit vs. explicit instruction; Saito & Plonsky, 2019, on teaching pronunciation; Shintani et al., 2013, on input vs. output focused activities).

In fact, there is a need for experimental evaluation of the suggestions in this paper. For example, it is argued that learning will be facilitated if a teacher "[reads] out a text loudly and emphatically, with long pauses between chunks, so that learners can notice the chunks more easily" (p. 118), or if learners stare at color-coded verbs in a word cloud (p. 128), or if printed corpora include boldface or boxed items (p. 124), of if "cognitively inspired explanations" of prepositional use are followed by error detection exercises (p. 126). There is no convincing evidence of the efficacy of any of these activities, and experimental data should be brought to bear on such proposals before any suggestion of widespread application in classrooms.

Cognitive corpus linguistic approaches could also potentially find support in the SLA literature. The data-driven approach to corpus learning materials advocated here recalls similar arguments for "*a posteriori* error analysis" in the 1970s. Gradman (1971) maintained that predicting patterns of interlanguage on the basis

of comparative language study was a waste of time and resources, and teaching materials should be based on existing evidence of the errors that learners actually do make (see Schachter, 1974 for critical discussion). More recently, in a special issue devoted to corpora in language learning and teaching, Bardovi-Harlig et al. (2017) tested the relative efficacy of teacher-developed corpus-based materials and teacher-guided corpus searches by students in the learning of pragmatic routines. They found that both were useful: the former for improving the clarity of the illocutionary force of turns in conversation, and the latter to focus learners' attention and to motivate subsequent independent searches.

In conclusion, while there is a convincing case to be made for the exciting potential of cognitive corpus linguistics to facilitate language teaching and learning, this would undoubtedly be made stronger by considering the endeavor from a broader perspective. Beyond the useful and inspiring references from the subfields of cognitive and corpus linguistics in this paper, a more open engagement with general linguistics, SLA, and pedagogy would help provide a more solid foundation based on previous findings and on the experimental methodology needed to vet teaching materials and practices.

References

Bardovi-Harlig, K. (2015). Designing instructional effect studies for L2 pragmatics: A guide for teachers and researchers. In S. Gesuato, F. Bianchi, & W. Cheng (Eds.) *Teaching, learning and investigating pragmatics: Principles, methods and practices* (pp. 135–164). Cambridge Scholars Publishing.

Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines. *Language Learning & Technology* 21(3): 76–103.

Bardovi-Harlig, K. & Stringer, D. (2017). Unconventional expressions: Productive syntax in the L2 acquisition of formulaic language. *Second Language Research*, 33(1): 61–90. https://doi.org/10.1177/0267658316641725

Boers, F., & Lindstromberg, S. (2006). Cognitive linguistic applications in second or foreign language instruction: Rationale, proposals, and evaluation. In G. Kristiansen, M. Achard, R. Dirven, & F. J. Ruiz de Mendoza Ibáñez (Eds.), *Cognitive linguistics: Current applications and future perspectives* (pp. 305–355). Berlin: de Gruyter.

Bybee, J. (2010). *Language, usage, and cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Carroll, S. E. (2010). Explaining how learners extract 'formulae' from L2 input. *Language, Interaction and Acquisition* 1: 229–50. https://doi.org/10.1075/lia.1.2.05car

Cunningham, A. E. (2005). Vocabulary growth through independent reading and reading aloud to children. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 45–68). Mahwah, NJ: Erlbaum.

De Knop, S. & Meunier, F. (2015). The 'learner corpus research, cognitive linguistics and second language acquisition' nexus: A SWOT analysis. *Corpus Linguistics and Linguistic Theory*, 11(1): 1–18. https://doi.org/10.1515/cllt-2014-0004

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge. https://doi.org/10.4324/9781315779553

Ellis, N.C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 33–68). Oxford: Blackwell. https://doi.org/10.1002/9780470756492.ch4

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESLEJ*, 8(4), 1–35.

Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247. https://doi.org/10.1111/ijal.12317

Gilquin, G. (2022). Cognitive corpus linguistics and pedagogy: From rationale to applications. *Pedagogical Linguistics*, VOL: pages. https://doi.org/10.1075/pl.22014.gil

Gradman, H. (1971). The limitations of contrastive analysis predictions. *Proceedings of the Pacific Conference on Contrastive Linguistics and Language Universals (PCCLLU): University of Hawai'i Working Papers in Linguistics* 3/4: 11–16. University of Hawai'i.radman.

Granena, G., & Yilmaz, Y. (2021). A psycholinguistically motivated methodology for task-based language teaching. In M. Ahmadian & M. Long (Eds.), *The Cambridge handbook of task-based language teaching* (pp. 305–325). Cambridge University Press. https://doi.org/10.1017/9781108868327.018

Hornstein, N. & Lightfoot, D. (1981). Introduction. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: The logical problem of language acquisition* (pp. 9–31). New York: Longman.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Langacker, R. W. (1987) *Foundations of cognitive grammar, Volume 1: Theoretical prerequisites*. Stanford: Stanford University Press.

Lightbown, P. & Spada, N. (2013, 4th ed.). *How languages are learned*. Oxford: Oxford University Press.

Lindstromberg, S. & Boers, F. (2008). *Teaching chunks of language: From noticing to remembering*. Innsbruck: Helbling Languages.

Long, M. (2015). Experimental perspectives on classroom interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 60–73). Malden, MA: John Wiley & Sons. https://doi.org/10.1002/9781118531242.ch4

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders*, 41(3), 335–354. https://doi.org/10.1080/13682820600602295

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417–528. https://doi.org/10.1111/0023-8333.00136

Rudzka-Ostyn, B. (2003). *Word power: Phrasal verbs and compounds. A cognitive approach*. Berlin: de Gruyter. https://doi.org/10.1515/9783110197235

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69, 652–708. https://doi.org/10.1111/lang.12345

Schachter, J. (1974). An error in error analysis. *Language Learning* 24(2): 205–214. https://doi.org/10.1111/j.1467-1770.1974.tb00502.x

Schwartz, B.D. & Sprouse, R. (2013). Generative approaches and the poverty of the stimulus. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 137–158). New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139051729.010

Seuss, Dr. (1957). *The cat in the hat*. New York: Random House.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning*, 63(2), 296–329. https://doi.org/10.1111/lang.12001

Slabakova, R. (2008). *Meaning in the second language*. New York/Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110211511

Slabakova, R. (2016). *Second language acquisition*. Oxford: Oxford University Press.

Stahl, S. (2005). Four problems with teaching word meanings (and what to do to make vocabulary an integral part of instruction). In E.H. Hiebert & M.L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 95–114). Mahwah, NJ: Lawrence Erlbaum.

Stringer, D. (2019). Lexical semantics: Relativity and transfer. In Erdogan, N. & Wei, M. (Eds.), *Applied linguistics for teachers of culturally and linguistically diverse learners* (pp. 180–203). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-5225-8467-4.ch007

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence from Self-Paced Reading and Sentence Recall Tasks. *Language Learning* 61(2): 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Tyler, A., Huang, L., & Jan, H. (2018) (Eds.). *What is applied cognitive linguistics? Answers from current SLA research*. Berlin: de Gruyter. https://doi.org/10.1515/9783110572186

VanPatten, B., Keating, G.D., & Wulff, S. (Eds.). (2020). *Theories in second language acquisition: An introduction* (3rd ed.). New York: Routledge. https://doi.org/10.4324/9780429503986

Address for correspondence

David Stringer
Dept. of Second Language Studies
Indiana University
Ballantine Hall 717
Bloomington, IN 47405
USA

ds6@iu.edu
https://orcid.org/0000-0003-0006-0823

Publication history

Date received: 25 May 2022
Date accepted: 2 June 2022
Published online: 13 July 2022