# A Theoretical Rationale for the Importance of Patterning in Language Acquisition and the Implications for Data-driven Learning

*Anne O'Keeffe (Mary Immaculate College, University of Limerick)*

*Abstract*

This article makes the case for the importance of exposure to language patterns in the context of second language acquisition. From the theoretical perspective of the Usage-Based (UB) model of acquisition, I argue that in the process of first and second language acquisition, the learner attends to frequently used form-meaning pairings that they experience. These then become entrenched as grammatical knowledge in the mind of the speaker (Ellis and Ferreira-Junior 2009). From a data-driven learning (DDL) perspective, I argue that the UB model provides a strong rationale for this approach because it offers a principled means of curating data from corpora to aid L2 learning. Work that brings the UB model and corpus linguistics (CL) together is emerging. For example, Römer (2019) explores patterns of verb-argument constructions (VACs) acquired by German beginner-level learners of English. Building on Römer's work, this study focuses on certain patterns and explores these using data from the 55-million-word Cambridge Learner Corpus. It illustrates how patterns develop across proficiency levels in learner English and compares this with L1 patterns (using the British National Corpus, Burnard and Aston 1998). Overall, it is argued that, through a UB understanding of how language is acquired, DDL has the potential to bring an acceleration of language experience to the learner. However, the curation of language input needs to be sensitive to and differentiated for stages of learning. It is argued that UB theory coupled with CL analysis can aid this process.

Keywords: data driven learning (DDL); corpus linguistics (CL); usage-based model (UB); second language acquisition (SLA); verb-argument constructions (VACs)

*1. Introduction*

In very simplistic terms, data-driven learning (DDL) is an instructional approach that involves giving learners of a language multiple exposures to patterns of the target language. The curation of these patterns is ideally driven by a student's curiosity which leads them from forms to meaning(s). The ideal user of DDL is motivated to investigate and induce meanings from patterns of language use and ultimately to store these patterns so that they can form part of their repertoire of language, which can be expanded over time. While years of research has gone into showing that this technique carries pedagogical merit in the classroom and is worth the technical effort on the part of the teacher and the student (Boulton 2012; Boulton and Cobb 2017; Chambers 2007; Cobb and Boulton 2015; Lee et al. 2019; Mizumoto and Chujo 2015; Vyatkina and Boulton 2017), very little thought has gone into the rationale for why repeated encounters with patterns of language might be a good idea from a theoretical perspective of second language acquisition (O'Keeffe 2021a, 2021b).

The cognitive process of grappling with patterns and their meanings is seen as a fundamental process in the usage-based model of language acquisition, where, for example, in first language acquisition, forms encountered by a child are paired with meanings and refined over and over through experience (Pérez-Paredes et al. 2020). Patterns become 'entrenched as grammatical knowledge in the speaker's mind' (Ellis and Ferreira-Junior 2009: 188). This article will first explore existing work on the importance of form-meaning mapping of patterns of language and explore what this means in the context of second language acquisition. From a data-driven learning perspective, I will argue that the UB model provides a strong theoretical rationale for the DDL approach and, as a result, more attention needs to be paid to mediating the curation of DDL materials. This argument echoes calls made over the years for the need to engage more with second language acquisition (Cobb 1999, 2005; Mukherjee 2006; Johansson 2009; Flowerdew 2015; Myles 2015) and to the need to make links with the UB model (Römer 2019; Pérez-Paredes et al. 2020; O'Keeffe 2021a, 2021b). By engaging with the UB model, there is also scope to learn how DDL can be enhanced. For example, it may inform us as to how to make DDL more useable with learners who have a lower level of proficiency. Following through with this perspective in this article, empirical data will be examined to investigate the patterns that

learners acquire at different levels and it will be argued that this kind of information is valuable for DDL design.

*2. Usage-Based Model: Patterns and form-meaning mapping*
*2.1 Usage-based model in first and second language contexts*
The connection between language acquisition and human experience is long-established in the field of cognitive linguistics (see Langacker 1987, 1991; Lakoff 1987; Croft and Cruse 2004; Goldberg 1995; Barlow and Kemmer 2000 among others). Constructivists within first language acquisition studies have long advocated for a usage-based model of acquisition (for instance, Slobin 1987; Tomasello 1992, 2003) and laterally it has been applied to second language acquisition contexts also (see Pérez-Paredes et al. 2020 for a summary). Within a first language acquisition perspective, the idea that we acquire our first language by building it up through our experiences with language can be observed over time. For example, a toddler interacting with a parent, guardian or carer might hear words and phrases repeatedly and will start to tally these patterns and link them to contextual meanings (see Ellis 2003). For instance, a child hears a phrase, such as *Mammy's gone* a number of times. Then they may also encounter phrases like *Mammy's gone in the car*; *All gone*; *Where's daddy gone? It's gone!* (see Pérez-Paredes et al. 2020). Based on processing and grappling with the link between these patterns of language and their contextual meaning, the child first might begin to produce or *construct* language output by putting two single words together in holophrases (word combinations). For example, from these patterns (above) that a child experiences, they might construct the two-word phrase *car gone* while pointing to the driveway where the family car is normally parked (see Ellis 2003). The next stage we expect is the abstracting of grammatical patterns where syntactic slots are filled and expanded. Therefore, *car gone* might expand to:

> *The car's gone,*
> *Where's the car gone?*
> *The juice is gone,*
> *Where's the Lego gone?*

Within this usage-based perspective, the child will gradually start to apply this accruing knowledge of form and meaning to make more patterns. In essence, pattern building is key to language acquisition and experience of language is key to pattern building (Pérez-Paredes et al. 2020).

As illustrated above, patterns build up from holophrase formula to an abstraction of a pattern and its meaning which can lead to more 'low-scope' pattern formation (see Ellis 2003). Essentially, we see a transition from learning about what words go together (based on the language that is experienced) to learning about patterns of complementation[1], collocation[2] and colligation[3], and this seems to happen on a verb-by-verb basis as more new language is experienced (see Pérez-Paredes et al. 2020). In this way, the mind acquires *construction patterns* of form and meaning. Constructions vary in terms of their complexity but, as Wulff and Ellis (2018) point out, the more often a speaker encounters a particular construction, or combination of constructions, the more *entrenched* it becomes. To say that a construction is entrenched means that it has become *automatized* as a routine chunk of language that is stored and activated by the language user as a whole, rather than 'creatively' assembled on the spot (De Smet and Cuyckens 2007: 188). This essentially means a unit of meaning has been subconsciously stored in the brain of the language user. As language users, we have, to quote Wulff and Ellis, 'a huge warehouse of constructions that vary in their degree of complexity and abstraction' (2018: 39).

The usage-based model has gained much support in second language acquisition contexts (Bybee 2008; Ellis 2012a, 2012b; Ellis et al. 2016; Römer 2019; Pérez-Paredes et al. 2020). However, it is more complex to create a scenario that parallels the hypothetical toddler described above for the typical second language learner. Second languages can be acquired at different ages; they can be acquired through very different experiences,

---

[1] Complements, along with subject, verb, object and adjunct, are one of the five major elements of clause structure. In clauses with linking verbs (*be*, *seem*, *become*), complements which follow the verb and which add information about the subject are called subject complements (underlined): *He is a teacher; They seem surprised.* Object complements add more information about an object: *This makes her very happy.* (adapted from Carter et al. 2011).

[2] Collocation refers to how words go together or form fixed relationships (e.g., *make a wish* rather than *do a wish*; *scenic view* but not *scenic picture.* (adapted from Carter et al. 2011).

[3] Colligation refers to how words form grammatical patterns with other words. For example, the colligational pattern *in the* is in many multi-word units but not all. We can say *in the beginning* but not *in the start,* for example (McCarthy et al. 2010: 156).

from instructed classroom situations (e.g., a child learning a second language in school) to fully immersive scenarios (an individual immigrating to a country where a new language is spoken and has to be acquired possibly without instruction). In second language acquisition, there is normally no equivalent relationship to the child-carer one we associate with first language acquisition. Cognitively, a second language learner's experience can be mediated by conscious learning and instruction (i.e., they can consciously decide to study irregular verbs whereas a child cannot; they can decide to look up a word or phrase in a dictionary, etc.). The patterns that a learner of a second language encounters may be formally sequenced in a syllabus that is outside of their choosing. A child, while acquiring a first language, is also going through a myriad of social, cognitive, emotional and moral development milestones, whereas most second language learners will have reached most developmental stages or indeed may be adults. The differences are many. Nonetheless, we acknowledge that a second language learner has already gone through the cognitive process of grappling with patterns and meaning-mapping in their first language and thus they are equipped and practiced in this subconscious process of language acquisition already. Therefore, in theory, they should be able to apply a parallel process to a second language. As Tyler and Ortega (2018: 3) note, there is no doubt that 'learning a language is one of the most complex accomplishments humans achieve' and that 'this is true for the first language learner and perhaps even more so for the second language learner'.

Pérez-Paredes et al. (2020) point to UB evidence that suggests that the process of additional language learning involves intentional reading and pattern finding and that it develops along a similar cline as first language acquisition (described above), moving from formula to low scope patterns to fully abstracted constructions (Ellis 2003). As discussed above in relation to first language acquisition, second language learners also move from holophrase (e.g., *I'd like to…*) to low scope slot and frame system (e.g., *I went/walked/jogged to the cinema/shop/restaurant*) to a fully abstracted formulaic chunk (*He came to the conclusion that…*). A core tenet of the UB model is that our knowledge of language comes from experiencing and using it 'as part of a communicatively-rich human social environment' (Ellis and Larsen-Freeman 2006: 577). However, for many learners of a second language, the classroom may be the only environment where language is experienced. If we accept, within the UB model, that

frequent encounters with language patterns are essential for language acquisition, then any means of intensifying a learner's experience with language should benefit the process even in the limited social context of a classroom and this is where we may find important connections for DDL.

*2.2 Constructions*
Within the field of cognitive linguistics, the term *construction* has been used to refer to the forms that have been mapped to meanings and become conventionalised within a speech community and entrenched in the mind of the user (Ellis 2012a, 2012b). Essentially, they are morphological, lexical and syntactic units that have been mapped through use to particular semantic, pragmatic and discourse functions (Ellis 2012a). The verb is usually seen as an important part of a construction because of its role in providing ways of talking about events and their participants (Perek 2015). Verbs, more so than other content word classes, according to Perek (2015: 1) 'are rarely uttered in isolation but are usually accompanied by certain other words called *arguments* of the verb'. Within UB studies, the term verb-argument construction (VAC) is core to empirical corpus-based work, as we shall return to below in relation to the work of Römer (2019).

*3. Usage-Based Model and DDL*
There is great resonance for DDL in the UB view of language acquisition—consider this quote from Ellis:

> Frequency is a key determinant of acquisition because "rules" of language, at all levels of analysis from phonological, through syntax to discourse, are **structural regularities** which emerge from learners' lifetime analysis of the distributional characteristics of language input. (Ellis 2012b: 196) [bold emphasis added]

Within DDL, we strive to help learners become aware of 'structural regularities' and patterns of form and meaning (Johns and King 1991; Boulton and Cobb 2017; Lee et al. 2019; Pérez-Paredes 2019; Pérez-Paredes et al. 2020). A key difference is perhaps that the UB model refers to the acquisition of structural regularities that emerge from learners' *lifetime* analysis of the distributional characteristics of language input. As discussed above, the second language acquisition process can differ greatly in terms of how much language experience a learner has in their

'lifetime' as a learner and much of it may be within the context of formal instruction mediated by an external syllabus.

However, it can be argued that in the context of DDL, an acceleration of language experience can be brought to the learner: a type of 'input flooding' (after Sharwood Smith 1991, 1993). In line with the words of Gabrielatos (2005: 10), DDL can offer learners a type of 'condensed exposure' that can aid lexical and pattern awareness. We can possibly drive a type of intensification of the cognitive process through 'grappling' with patterns. However, this begs the questions: what input will best promote language acquisition and development, and does this depend on the level of the learner (i.e., their stage of development)? In essence, how do we know whether we are curating the right patterns within this process of condensed exposure?

From many studies over the years and, above all, from principled meta-analysis studies of work on DDL, such as Boulton and Cobb (2017) which looked at 64 empirical studies, it has been noted that DDL is successful for university-based students (both undergraduate and postgraduate), and for learners who are at intermediate to advanced levels of language competency (Boulton and Cobb 2017: 386). The fact that grappling with patterns in DDL is not daunting for intermediate to advanced-level learners, O'Keeffe (2021a, 2021b) hypothesizes, is because learners at these levels have already abstracted many patterns and have a critical level of understanding of these in terms of mapping their forms and meanings. Through a UB lens, it can be speculated that learners from intermediate level upwards have already gained from building on low-scope patterns in the L2 and they are thus equipped to build on the cognitive processes that have already been used to acquire their L1. However, in this article, it is argued that we should not write off learners who are at lower levels of proficiency from benefitting from the advantages of DDL. To achieve this, there is a need for careful and principled design in terms of how DDL is used at these lower levels (A1 to B1) so as to structure the process of acquisition based on a UB-based understanding of the process.

If we work with the insights from the UB model (as discussed above from a theoretical perspective), then we may be better able to curate and mediate the learning process for lower-level learners so that they can experience language patterns that are differentiated to their level. Recent empirical work by Römer (2019), which focused on patterns acquired by

beginner-level German learners of English, offers the starting point for our case study in which data from the Cambridge Learner Corpus will be used to gain insight into the importance of patterns even at A1 level and how this kind of knowledge might be of use in DDL.

*4. Case study of pattern learning from A1 level*

Römer (2019) explores patterns of verb-argument constructions (VACs) acquired by German beginner-level learners of English and how this repertoire develops across levels. Using data from the *Education First-Cambridge Open Language Database[4]* (EFCAMDAT) (see Alexopoulou et al. 2015), Römer looks at a sample of German L1 learners of English. Among the top 10 most frequent VACs (see section 2.2) in the beginner-level data, Römer (2019: 275ff) lists structures like:

> *My name is Anna* (*be* copula constructions)
> *I'm happy* (nominal subject + copula *be* + an adjectival complement)
> *They live in Cologne* (verb + prepositional phrase starting with *in*)
> *There are many things near my house* (existential *there,* followed by a form of copula *be,* followed by a nominal subject)

Römer notes that longer, more complex VAC patterns are rare or non-existent at A1. While this is to be expected, quantifying and detailing which VAC patterns are most frequent at this level has great potential to inform DDL design because it allows us draw on the cognitive store of patterns that has been acquired by learners. This can help design DDL tasks that are not just level appropriate but are timely in terms of the stage and process of acquisition within a UB perspective. In other words, as O'Keeffe (2021a, 2021b) notes, this kind of information has the potential to offer us a prototype for calibration of materials or input for DDL that is informed by a theory of second language acquisition.

Römer's (2019) results shows an exponential growth in VAC types acquired from A1 to C1. We see change in the repertoire in terms of what

---

[4] At the time of writing, the *EF-Cambridge Open Language Database* (EFCAMDAT) is a publicly available 83-million-word corpus resource that contains written samples from thousands of adult learners of English as a second language, worldwide. The data collected is based on 1 million assignments written by 174,000 learners, across a wide range of levels (CEFR stages A1–C2). See https://philarion.mml.cam.ac.uk/.

patterns learners most use. There is a clear tendency towards greater complexity of patterns, most notably in relation to both nominal and adjectival complementation. For example, Römer (2019) found that the A1 learners rely on a narrow range of verbs. In addition to the copular *be*, the top 10 patterns were limited to the verb lemmas: *live*; *have*; *meet*; *like* and *see* (within the top ten VAC combination patterns), whereas at B2 level, the top patterns included the verbs: copular *be*, plus *have*, *think*, *let*, *apply*, *find*, *want*, *observe* in more complex patterns of use (e.g., [lemma *observe*] + direct object). Findings of this nature are useful for a teacher wanting to use DDL with beginner level students because they offer a window into the learning process. We see the patterns that are core at A1 level which learners then build on in their acquisition process.

This case study will delve deeper into just one of Römer's (2019) patterns: the existential *there* + copula *be* + a nominal subject pattern. The aim is to look at its typical use among A1-level learners in general, using the A1 data from the 53-million-word Cambridge Learner Corpus (CLC) (2019 version), and observe how this pattern develops at other levels. The version of the CLC that will be used in this case study comprises data from learners from over 150 first language backgrounds[5]. The case study will also draw on typical use of the same pattern in the British National Corpus (BNC).

First of all, using the CLC, via Sketch Engine, we extract 'candidates' for the existential *there* + copula *be* + a nominal subject pattern using the following search string [lemma="there"][lemma="be"][][][]. The results for the top 20 most frequent patterns used by A1 learners are given in Table 1. Note that 'lemmas' rather than word forms are recalled. Lemmas refer to head words that incorporate all of the related forms, for example the lemma *GO* will incorporate all the possible forms such as *go*, *goes*, *going*, *gone*, *went*. This will allow for more condensed results. For example, the result for the most frequent pattern in Appendix A is *There BE a lot of*. This can include any of the following:

> *There/there is a lot of*
> *There's/there's a lot of*
> *There/there are a lot of*

---

[5] Percentage breakdown of first languages in CLC 2019 version: 13.9% Spanish -Latin American; 11.3% Chinese; 8.8% German; 7.6% Portuguese; 7.1% Italian; 5.8% Spanish-European; 5.5% Greek; 4.2% French; 3.4% Polish; 3.4% Swiss German; 29% Other.

> *There/there was a lot of*
> *There/there were a lot of*

Henceforth, where a capitalised and italicised form is used, this refers to a lemma. Also note that the search string that was used to recall the patterns also brought up items with *BE + GO* (e.g., *There is going to be a lot*) which do not fit with the *there* + copula *BE* + a nominal subject pattern that we are investigating, so these were eliminated from the results.

Table 1. Top 20 lemmatised patterns of existential *There* + copula *BE* + a nominal subject pattern from Cambridge Learner Corpus (2019 version) A1 learners (at performance level[6])

|     | **Lemmatised pattern**       | **Freq.** | **PMW[7]** |
| --- | ---------------------------- | --------- | ---------- |
| 1.  | *There BE* a lot of          | 269       | 4.46338    |
| 2.  | *There BE* a bed ,           | 203       | 3.36827    |
| 3.  | *There BE* all my *FRIEND*   | 56        | 0.92918    |
| 4.  | *There BE* a concert .       | 50        | 0.82962    |
| 5.  | *There BE* a concert on      | 48        | 0.79644    |
| 6.  | *There BE* a concert of      | 42        | 0.69688    |
| 7.  | *There BE* a desk ,          | 40        | 0.6637     |
| 8.  | *There BE* a bed and         | 35        | 0.58074    |
| 9.  | *There BE* my house .        | 34        | 0.56414    |
| 10. | *There BE* my bed ,          | 34        | 0.56414    |
| 11. | *There BE* a computer ,      | 29        | 0.48118    |
| 12. | *There BE* two *BED* ,       | 24        | 0.39822    |
| 13. | *There BE* a concert ,       | 23        | 0.38163    |

---

[6] A1 Performance level refers to data that was graded at this level. For example, if a learner failed the A1 exam, their data is not included, etc.

[7] PMW stands for 'per million words', where 'words' refers to the number of individual words, or tokens, in a corpus (including repetitions of words). This is a 'normalisation' process used in corpus linguistics so that descriptive frequency results can be compared between corpora of difference sizes. It is calculated in the following way (where 'raw' refers to the actual number of occurrences): number of raw occurrences / number of words in the corpus x 1,000,000. In cases of high occurrences, it is common to normalise to 100, 1000, 10,000 also.

| 14. | *There BE* my family and | 23 | 0.38163 |
|---|---|---|---|
| 15. | *There BE* a TV , | 23 | 0.38163 |
| 16. | *There BE* a table , | 22 | 0.36503 |
| 17. | *There BE* a sofa , | 21 | 0.34844 |
| 18. | *There BE* a rock concert | 21 | 0.34844 |
| 19. | *There BE* a concert at | 20 | 0.33185 |
| 20. | *There BE* a big bed | 20 | 0.33185 |

Noticeably, Table 1 shows that the following patterns prevail:
1. *There + BE + a lot of*
2. *There + BE + a* [concrete noun relating to household items]
3. *There + BE + a* [event]

Further examination of the most frequent pattern *There + BE + a lot of*[8] shows that the following items most frequently come after *a lot of* (see Table 2).

Table 2. Top 20 most frequent items to follow *There BE a lot of* in A1 CLC data

| | **Lemmatised pattern** | **Freq.** | **PMW** |
|---|---|---|---|
| 1. | *THERE BE* a lot of *PEOPLE*[9] | 36 | 0.59733 |
| 2. | *THERE BE* a lot of *THING* | 19 | 0.31526 |
| 3. | *THERE BE* a lot of *GAME* | 12 | 0.19911 |
| 4. | *THERE BE* a lot of *FURNITURE* | 11 | 0.18252 |
| 5. | *THERE BE* a lot of *BOOK* | 11 | 0.18252 |
| 6. | *THERE BE* a lot of my | 10 | 0.16592 |
| 7. | *THERE BE* a lot of *FRIEND* | 7 | 0.11615 |
| 8. | *THERE BE* a lot of *PRESENT* | 5 | 0.08296 |

---

[8] Methodological note: the following search query was used in Sketch Engine using A1 performance level data in the CLC: [lemma="there"][lemma="be"][word="a"][word="lot"][word="of"][].

[9] In some cases, such as lines 1 and lines 5, lemmas are used where there is only one possible option in standard English. However, because at A1, countability errors are frequent, I wanted to capture attempted uses of the pattern such as *There are a lot of peoples*; *There was a lot of furnitures*.

| 9. | *THERE BE* a lot of fun | 5 | 0.08296 |
|---|---|---|---|
| 10. | *THERE BE* a lot of kind | 5 | 0.08296 |
| 11. | *THERE BE* a lot of beautiful | 5 | 0.08296 |
| 12. | *THERE BE* a lot of *SHOP* | 5 | 0.08296 |
| 13. | *THERE BE* a lot of *ROOM* | 5 | 0.08296 |
| 14. | *THERE BE* a lot of good | 4 | 0.06637 |
| 15. | *THERE BE* a lot of *COLOUR* | 4 | 0.06637 |
| 16. | *THERE BE* a lot of *PICTURE* | 3 | 0.04978 |
| 17. | *THERE BE* a lot of different | 3 | 0.04978 |
| 18. | *THERE BE* a lot of money | 3 | 0.04978 |
| 19. | *THERE BE* a lot of *ACTIVITY* | 3 | 0.04978 |
| 20. | *THERE BE* a lot of think[10] | 3 | 0.04978 |

From Table 2, we see that the phrase *a lot of* mostly quantifies the word *people,* plus concrete everyday objects. We also see the uncountable noun *room* used to refer to an amount of space available. Apart from concrete nouns, we also see the emergence of attributive adjectives after *a lot of* (*fun*, *kind*, *beautiful*, *good*, *different*, *famous*) and the determiner *my*.

This kind of information can prove useful to DDL design. It allows us to say, based on patterns 1, 2 and 3 above, in a principled evidence-based manner that the following patterns are important at A1 level:

1. *There BE + a lot of + nouns*:
   everyday concrete nouns (*people*, *thing*, *game*, *furniture*, *book*, *friend present*, *shop*, *picture*)
   abstract nouns (*money*, *activity*, *colour*, *room)*
2. *There BE + a lot of + adjective + noun*
   Most common adjectives used: *fun*, *kind*, *beautiful*, *good*, *different*, *famous*

Obviously, as these patterns are drawn from A1 level data, they reflect the syllabus and the tasks at this level where typical learner competencies include talking about oneself, one's family and friends and about places and events.

---

[10] These three patterns refer to the misspelling of *thing*.

The evidence from these patterns makes a strong case for DDL-based work at beginner's level on the existential *there* pattern followed by a form of copula *be,* followed by a nominal subject. Specifically, to focus on activities around these patterns, as well as the emerging *There BE + a lot of + determiner* pattern should prove very productive for beginner-level learners. While it might be counter-argued that these patterns do not need to be taught if they are going to develop anyway, we note that these results are based on learners who have reached A1 level. Those taking an A1 level course would therefore benefit from awareness-raising work on these core patterns.

The *There BE +* a nominal subject pattern forms the basis of much development across other levels and concomitant with this, there is evidence that the pattern evolves in how it is used. Appendix A shows existential *There BE +* a nominal subject pattern across all the levels of the Common European Framework of Reference (CEFR) for languages from A1 to C2 using the CLC. Notice (in Appendix A) that the pattern *There BE + a lot of* is consistently the most used. What emerges at A1 level is a pattern that is built upon developmentally across the levels of proficiency (see shaded cells in Appendix A). By A2, we see the use of negation emerging (*There BE not a lot of*) but we can also observe the emergence of many other patterns for quantification using the *There BE +* a nominal subject pattern. If we trace solely the development of patterns which are used in quantification, we see that while A1 learners start out with just one way of doing this through *There BE + a lot of*, the pattern continues to build so that by A2, learners already have many variations of patterns in their repertoire in terms of how they can express quantity using the fundamentals of a pattern that they acquired at A1 (see Table 3).

Table 3. Patterns from top 20 most frequent items (from Appendix A) based on *There BE +* a nominal subject used to express quantity by learners from A1 to C2

| A1 | B1 | C1 |
|---|---|---|
| *There BE* a lot of | *There BE* a lot of<br>*There BE* [number] parking space<br>*There BE* no parking space<br>*There BE* not enough parking | *There BE* a lot of<br>*There BE* a lack of<br>*There BE* not enough MINIBUS<br>*There BE* not enough time<br>*There BE* many people who |

| | | |
|---|---|---|
| | *There BE* only four *OFFICE*<br>*There BE* over [number] people<br>*There BE* not a lot<br>*There BE* only four *OFFICE*<br>*There BE* no parking<br>*There BE* a fall in<br>*There BE* no parking place | *There BE* nothing to do<br>*There BE* no room service<br>*There BE* a wide range<br>*There BE* a number of<br>*There BE* more and more<br>*There BE* no money *LEAVE*<br>*There BE* a variety of<br>*There BE* over [number] people |
| **A2** | **B2** | **C2** |
| *There BE* a lot of<br>*There BE* not a lot<br>*There BE* many *KIND* of<br>*There BE* many *THING* to<br>*There BE* not many people<br>*There BE* not so much<br>*There BE* not so many<br>*There BE* not so much<br>*There BE* not so many<br>*There BE* more thing to<br>*There BE* also a lot<br>*There BE* more and more<br>*There BE* so many people<br>*There BE* all kind of | *There BE* a lot of<br>*There BE* over [number] people<br>*There BE* a lack of<br>*There BE* more than [number]<br>*There BE* a number of<br>*There BE* a general upward<br>*There BE* a wide range<br>*There BE* not a lot<br>*There BE* so many *THING*<br>*There BE* lot of *THING* | *There BE* a lot of<br>*There BE* many people who<br>*There BE* some people who<br>*There BE* a number of<br>*There BE* no money *LEAVE*<br>*There BE* a lack of<br>*There BE* more and more<br>*There BE* a wide range<br>*There BE* too many people<br>*There BE* a great number<br>*There BE* also a lot<br>*There BE* so many people<br>*There BE* many *REASON* for<br>*There BE* a wide variety |

While, on one hand, this kind of evidence makes a strong case for the usefulness and the potential of DDL in accelerating the development of this pattern through careful curation and task design, it could be argued that such curation goes against the free exploratory spirit of DDL. This

more 'domesticated' approach (see Meunier 2019) may also have an important place in DDL design. This is a point we shall return to below.

In the final part of the case study of this one core pattern that A1 learners use, we will compare the learner results with the British National Corpus (BNC) (Burnard and Aston 1998). Table 4 illustrates the results for the *There BE* + a nominal subject pattern in the BNC. Similar to the CLC data, at all levels (see Appendix A), we see that the pattern *There BE* + *a lot of* is the most commonly used pattern.

Table 4. Top 20 lemmatised patterns of *THERE BE* + a nominal subject in the British National Corpus (BNC)

|     | **Lemmatised pattern**      | **Freq.** | **PMW**   |
|-----|-----------------------------|-----------|-----------|
| 1.  | *There BE* a lot of         | 1486      | 13.22703  |
| 2.  | *There BE* a number of      | 1071      | 9.53307   |
| 3.  | *There BE* , however ,      | 789       | 7.02296   |
| 4.  | *There BE* no doubt that    | 726       | 6.4622    |
| 5.  | *There BE* no need to       | 680       | 6.05274   |
| 6.  | *There BE* no point in      | 546       | 4.86      |
| 7.  | *There BE* no reason why    | 483       | 4.29923   |
| 8.  | *There BE* no sign of       | 412       | 3.66725   |
| 9.  | *There BE* no need for      | 406       | 3.61384   |
| 10. | *There BE* , of course      | 325       | 2.89286   |
| 11. | *There BE* , *THERE BE*     | 322       | 2.86615   |
| 12. | *There BE* no evidence that | 296       | 2.63472   |
| 13. | *There BE* no reason to     | 284       | 2.52791   |
| 14. | *There BE* nothing wrong with | 274     | 2.4389    |
| 15. | *There BE* no such thing    | 264       | 2.34989   |
| 16. | *There BE* a great deal     | 257       | 2.28758   |
| 17. | *There BE* a need for       | 254       | 2.26088   |
| 18. | *There BE* no question of   | 246       | 2.18967   |
| 19. | *There BE* one or two       | 219       | 1.94934   |
| 20. | *There BE* no way of        | 192       | 1.70901   |

The BNC results in Table 4 bring to light some other interesting findings when we compare them with the learner language presented earlier (Table 3):

1. On qualitative examination, many of the most frequent patterns in the BNC relate to evaluation and expression of stance: *There BE no doubt that*; *There BE* no need to; *There BE* no point in; *There BE* no reason why; *There BE* no sign of; *There BE* no need for; *There BE* , of course; *There BE* no question of; *There BE* no reason to; *There BE* nothing wrong with; *There BE* no such thing; *There BE* a need for; *There BE* no question of

2. The majority of the Top 20 patterns in Table 4 involve negations (12 out of 20, 60%): *There BE* no doubt that; *There BE* no need to; *There BE* no point in; *There BE* no reason why; *There BE* no sign of; *There BE* no need for; *There BE* no evidence that; *There BE* no reason to; *There BE* nothing wrong with; *There BE* no such thing; *There BE* no question of; *There BE* no way of

3. We notice, in Table 4, a tendency to break the pattern up with discourse markers: *There BE , however*; *There BE , of course.*

4. Unlike the learner results, the top 20 exponents are not dominated by expressions of quantification.

In Table 5, when we look at Appendix A in comparison with the results from the BNC in Table 4, we see that the C1 and C2 levels have some of the BNC features (listed above), including patterns that involve negation and which can be used to mark stance and evaluation; for instance, *There BE no doubt that*; *There BE no need to*; *There BE no point in.* However, it is noted that at C1, these patterns are often literal (in the sense of being specific or concrete such as *There BE no social programme*; *There BE not enough minibus*) while at C2, even though there are fewer items, there are overall more figurative markers of stance such as *There BE no doubt that*; *there BE no need to*; *there BE no point in*; *there BE no need for*[11]. The BNC, by contrast, shows all figurative items in the top 20.

---

[11] Figurative language refers to formulaic and often metaphoric uses that do not draw on the literal meaning of the components of the phrase. For example the phrase *you nailed it* can have a literal meaning when someone has just hammered a nail or it can be used figuratively to praise someone for doing something well.

Table 5. Patterns which involve negation top 20 of CLC and BNC results (bold marks figurative items)

| C1 | C2 | BNC |
|---|---|---|
| *There BE* **no doubt that** | *There BE* **no doubt that** | *There BE* **no doubt that** |
| *There BE* no social programme | *There BE* **no need to** | *There BE* **no need to** |
| *There BE* **no need to** | *There BE* **no point in** | *There BE* **no point** |
| *There BE* **no point in** | *There BE* no money leave | *There BE* **no reason why** |
| *There* BE not enough minibus | *There BE* **no need for** | *There BE* **no sign of** |
| *There BE* not enough time | | *There BE* **no need for** |
| *There BE* no room service | | *There BE* **no evidence that** |
| *There BE* no money leave | | *There BE* **no reason to** |
| *There BE* no guide provide | | *There BE* **nothing wrong with** |
| | | *There BE* **no such thing** |
| | | *There BE* **no question of** |
| | | *There BE* **no way of** |

Looking at the bigger picture and considering both the results in Appendix A and Table 5, we see a good example of the UB model in operation in second language acquisition. Learners move from holophrase (e.g., *There BE* + a nominal subject) to low scope slot and frame system (e.g., *There BE* + *no* + NOUN + … : *there is no social programme; there is no money left*) to a fully abstracted formulaic chunk (*There is no doubt that*; *there is no need to*; *there is no point in*; *There is no need for*) and as they do, the patterns that they use more resemble those in the BNC. Let us now consider the implications for DDL.

*5. Implications for DDL*
The case study in this article looked at just one pattern that UB-based research identified as used by A1-level German learners (of English) (i.e., Römer 2019). The case study tested this in the CLC across learners of English from over 150 different first language backgrounds (see section 4). The results showed that the existential *there* + copula *be* + a nominal subject pattern is a very important one not just at A1 level but across all levels. Tracing the journey of this one pattern across the differing levels of competency, we can see that learners built on the very basic patterns of

use at A1 that largely relate to quantification and, over time, the same pattern is used in more and more complex patterns. For example, there is evidence of a move towards negated patterns that eventually take on more figurative discourse functions in expressions of stance and evaluation (e.g., *There is no doubt that*). Overall, using the CLC, the case study illustrated that even at beginner's level, the pattern was productive though in a limited manner where learners had mapped the patterns to a narrow range of meanings. Then we saw how the same pattern became a core item in the learner's repertoire, ultimately becoming more complex in both form and meaning. Such evidence could be very useful in DDL design. First of all, it tells us that working with patterns should not be left until the students are at intermediate level because, as discussed, even at A1 level, learners are sensitive to patterns. However, it also shows us that there is a need to think about how to use this information to the best effect with A1, A2 and B1-level learners. The challenge is to find a way of exploiting this kind of corpus evidence so as to seize the opportunity to create a pathway for acquisition through DDL-based materials and principled task design. Meunier (2019) notes that concordance lines are not the only possible triggers of frequency effects and form-meaning mappings in focus on form activities (i.e., language lessons which focus on grammar in an overt way). This may be a crucial point. At the core of the DDL ethos is the notion that the learner can explore and discover language patterns (see Johns and King 1991; Cobb 2005; O'Sullivan 2007; Boulton 2010; O'Keeffe 2021a), however, as discussed, meta-analysis studies show that the traditional approach of working with concordances works best for learners who are above intermediate level (Boulton and Cobb 2017). The alternative then is to think beyond concordance-line-type activities. Meunier (2019) gives ample examples of existing applications that might be suitable for repurposing. These include ones such as *Playphraseme* (see https://www.playphrase.me). This interface allows users to search for a phrase in English so as to generate very short video clips (usually about a sentence in length) from a film. The clips are then played with subtitles (also in English). Therefore, as an alternative to a screen of written concordances, a beginner-level learner works with video clips, which offer a multi-modal (i.e., not just written mode), communicatively rich context (as detailed in Meunier 2019). One of the additional modes that is provided in this format is that of sound. Therefore, the lower-level learner will get to hear the prosodic patterns as well as read the phrase in its full context

(including gaze, gesture, etc.). This experience aligns more with natural first and second language acquisition where form-frequency mapping is a multi-modal experience (usually with audio, visual, verbal, non-verbal, prosodic information) (see Meunier 2019).

Essentially, what is being argued here is that for DDL to work with lower-level learners, there is a need for greater teacher mediation because patterns and data suitable to the stage of learning need to be carefully curated and contextualized (as argued in O'Keeffe 2021a), as well as a need to provide as rich a multi-modal context as possible. This will align well with the three core factors that affect cognition processes (based on Ellis 2012b: 195):

1. *Frequency*: the more frequently something is experienced, the stronger our memory of it;
2. *Recency*: the more recently we have experienced something, the stronger our memory of it;
3. *Context*: the more times we experience 'conjunctions of features' in a context (e.g., patterns uttered in a given context and situation), the stronger the categorisation process of associating a pattern with a particular meaning in context.

As humans we are incessantly and unconsciously figuring and tallying (Ellis 2002). Even at lower levels, learners who are presented with carefully curated patterns in a context-rich format can potentially unlock connections between these patterns and their meanings. While at first, this may be in a very basic and literal way, it will ultimately accrue to more complex patterns and uses. In the context of learning a language, the quality of the input is crucial in this process.

## 6. Conclusion

As discussed in this article, DDL, in its traditional sense, is ideally driven by a student's curiosity and it leads them to find out more about forms and their meanings. This is an ideal experience that has been shown to benefit acquisition at intermediate level or higher. In light of UB insights on acquisition, this is perhaps best explained by the fact that lower-level learners have not reached the point where the form-meaning mapping process is advanced enough to cope with the cognitive demands of categorisation of concordance lines and the induction of their meaning (O'Keeffe 2021a). However, they have a capacity for form-meaning mapping (as illustrated in the case study and in Römer 2019 using a

different learner dataset) and they may simply need a more structured DDL approach. This approach needs to be informed by a theory of second language acquisition (as in this case we made use of the UB model) as well as corpus evidence (such as the results provided by Römer 2019 or in the present case study). Additionally, the nature of the language input needs structure and multimodal richness if it is going to best replicate how we acquire our first language by experiencing it in all its context.

In summary, this article makes that case for a theoretically-informed approach to DDL, not just at higher levels but across levels of proficiency. It makes a case for the UB model as a strong ally for the DDL approach. However, if we engage with UB theory, the washback is that we need to pay more attention to the mediation and curation of the 'data' we use in lessons. We need to take cognisance of factors such as the developmental stage of the learner and the CEFR offers us a structure for this. This can be aided by learner corpus evidence, as illustrated in this case study. By engaging with the UB model, there is also scope to learn how DDL can be enhanced. Especially, it can inform us as to how to make DDL more useable across the full gamut of second language learners. In conclusion, we need to focus more on the 'L' in DDL so as to deepen our understanding of the process of 'learning' in the context of patterns and frequency. This can lead to a more informed approach to differentiating DDL relative to the stage of acquisition of our learners.

*Appendix*

Appendix A. Top 20 lemmatised patterns of existential *There + BE* + a nominal subject pattern from Cambridge Learner Corpus (2019) A1 to C2 learners at performance levels (phrases per million tokens)

| | A1 | | A2 | | B1 | | B2 | | C1 | | C2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | lemma | PMW | lemma | PMW | lemma | PMW | lemma | PMW | lemma | PMW | lemma | PMW |
| 1. | there BE a lot of | 4.5 | there BE a lot of | 42.5 | there BE a lot of | 14.7 | there BE a lot of | 15.7 | there BE a lot of | 9.9 | there BE a lot of | 7.3 |
| 2. | there BE a bed , | 3.4 | there BE a new shop | 6.1 | there BE a problem with | 2.9 | there BE [number] people over | 1.3 | there BE no doubt that | 1.7 | there Be no doubt that | 3.4 |
| 3. | there BE all my friend | 0.9 | there BE not a lot | 1.5 | there BE [number] parking space | 2.8 | there BE no doubt that | 1.0 | there BE no social programme | 1.6 | there BE no need to | 1.3 |
| 4. | there BE a concert . | 0.8 | there BE a concert in | 1.2 | there BE no parking space | 2.7 | there BE a new direct | 1.0 | there BE no need to | 1.4 | there BE no point in | 1.3 |
| 5. | there BE a concert on | 0.8 | there BE many kind of | 1.2 | there BE a TV company | 2.1 | there BE no need to | 0.9 | there BE a lack of | 1.0 | there BE many people who | 1.2 |
| 6. | there BE a concert of | 0.7 | there BE lot of people | 1.2 | there BE a bed , | 2.0 | there BE a lack of | 0.9 | there BE no point in | 0.9 | there BE some people who | 1.2 |
| 7. | there BE a desk , | 0.7 | there BE lot of thing | 1.1 | there BE not enough parking | 1.0 | there BE a new shop | 0.9 | there BE not enough minibus | 0.9 | there BE a number of | 1.2 |
| 8. | there BE a bed and | 0.6 | there BE a problem with | 0.9 | there BE only four office | 0.8 | there BE a new collection | 0.7 | there BE not enough time | 0.9 | there BE no money leave | 1.1 |
| 9. | there BE my house . | 0.6 | there BE a house with | 0.8 | there BE some problem with | 0.7 | there BE more than [number] | 0.8 | there BE many people who | 0.8 | there BE a lack of | 1.0 |
| 10. | there BE my bed , | 0.6 | there BE many thing to | 0.7 | there BE [number] people over | 0.7 | there BE a number of | 0.7 | there BE nothing to do | 0.8 | there BE more and more | 0.9 |
| 11. | there BE a computer , | 0.5 | there BE not many people | 0.7 | there BE not a lot | 0.6 | there BE some thing that | 0.6 | there BE no room service | 0.7 | there BE a wide range | 0.7 |
| 12. | there BE two bed , | 0.4 | there BE a new client | 0.6 | there BE four office leave | 0.6 | there BE advantage and disadvantage | 0.6 | there BE a wide range | 0.6 | there BE no need for | 0.6 |
| 13. | there BE a concert , | 0.4 | there BE not so much | 0.6 | there BE no parking . | 0.6 | there BE a general upward | 0.6 | there BE a number of | 0.6 | there BE , however , | 0.6 |
| 14. | there BE my family and | 0.4 | there BE some problem in | 0.5 | there BE good public transport | 0.5 | there BE a wide range | 0.5 | there BE some people who | 0.6 | there BE too many people | 0.5 |
| 15. | there BE a TV , | 0.4 | there BE not so many | 0.5 | there BE my bed , | 0.5 | there BE not a lot | 0.5 | there BE more and more | 0.5 | there BE a great number | 0.5 |
| 16. | there BE a table , | 0.4 | there BE more thing to | 0.5 | there BE something wrong with | 0.5 | there BE no social programme | 0.4 | there BE no money leave | 0.5 | there BE also a lot | 0.4 |
| 17. | there BE a sofa , | 0.3 | there BE also a lot | 0.5 | there BE a desk , | 0.4 | there BE so many thing | 0.5 | there BE a variety of | 0.5 | there BE so many people | 0.4 |
| 18. | there BE a rock concert | 0.3 | there BE more and more | 0.5 | there BE a fall in | 0.4 | there BE a need for | 0.4 | there BE some thing that | 0.5 | there BE many reason for | 0.4 |
| 19. | there BE a concert at | 0.3 | there BE so many people | 0.5 | there BE no parking place | 0.4 | there BE no point in | 0.4 | there BE over [number] people | 0.5 | there BE people who do | 0.4 |
| 20. | there BE a big bed | 0.3 | there BE all kind of | 0.5 | there BE no need to | 0.3 | there BE lot of thing | 0.3 | there BE no guide provide | 0.5 | there BE a wide variety | 0.4 |

*References*

Alexopoulou, Theodora, Jeroen Geertzen, Anna Korhonen, and Detmar Meurers. 2015. Exploring big educational learner corpora for SLA research: Perspectives on relative clauses. *International Journal of Learner Corpus Research* 1(1): 96–129.

Barlow, Michael, and Suzanne Kemmer. 2000. *Usage based models of language.* Chicago: The University of Chicago Press.

Boulton, Alex. 2010. Data-Driven Learning: Taking the computer out of the equation. *Language Learning* 60(3): 534–572.

Boulton, Alex. 2012. Corpus consultation for ESP: A review of empirical research. In *Corpus-informed research and learning in ESP: Issues and applications*, edited by Alex Boulton, Shirley Carter-Thomas, and Elizabeth Rowley-Jolivet, 261–291. Amsterdam: John Benjamins.

Boulton, Alex, and Tom Cobb. 2017. Corpus use in language learning: A meta-analysis. *Language Learning* 67(2): 348–393.

Burnard, Lou, and Guy Aston. 1998. *The BNC handbook: Exploring the British National Corpus*. Edinburgh: Edinburgh University Press.

Bybee, Joan. 2008. Usage-based grammar and second language acquisition. In *Handbook of cognitive linguistics and second language acquisition*, edited by Peter Robinson, and Nick C. Ellis, 216–235. London: Routledge.

Carter, Ronald A., Michael J. McCarthy, Geraldine Mark, and Anne O'Keeffe. 2011. *English grammar today.* Cambridge: Cambridge University Press.

Chambers, Angela. 2007. Popularising corpus consultation by language learners and teachers. In *Corpora in the foreign language classroom*, edited by Encarnación Hidalgo, Luis Quereda, and Juan Santana, 3–16. Amsterdam: Rodopi.

Cobb, Tom. 1999. Applying constructivism: A test for the learner as scientist. *Educational Technology Research & Development* 47(3): 15–31.

Cobb, Tom. 2005. Constructivism, applied linguistics, and language education. *Encyclopedia of language and linguistics* Volume 3, 2nd ed., 5–88. Amsterdam: Elsevier.

Cobb, Tom, and Alex Boulton. 2015. Classroom applications of corpus analysis. In *Cambridge handbook of corpus linguistics*, edited by Douglas Biber, and Randi Reppen, 478–497. Cambridge: Cambridge University Press.

Croft, William, and Alan Cruse. 2004. *Cognitive Linguistics.* Cambridge: Cambridge University Press.

De Smet, Hendrick, and Hubert Cuyckens. 2007. Diachronic aspects of complementation: Constructions, entrenchment and the matching-problem. In *Studies in the history of the English language III: Managing chaos: strategies for identifying change in English*, edited by Christopher M. Cain, and Geoffrey Russom, 1–37. Berlin: De Gruyter Mouton.

Ellis, Nick C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24: 143–188.

Ellis, Nick C. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In *Handbook of second language acquisition*, edited by Catherine J. Doughty, and Michael H. Long, 33–68. Oxford: Blackwell.

Ellis, Nick C. 2012a. Frequency effects. In *The Routledge encyclopedia of second language acquisition*, edited by Peter Robinson, 260–265. New York: Routledge.

Ellis, Nick C. 2012b. Frequency-based accounts of second language acquisition. In *The Routledge handbook of second language acquisition*, edited by Susan M. Gass, and Alison Mackey, 193–216. London: Routledge.

Ellis, Nick C., and Fernando Ferreira-Junior. 2009. Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7: 188–221.

Ellis, Nick C., and Diane Larsen-Freeman. 2006. Language emergence: Implications for applied linguistics. Introduction to the special issue. *Applied Linguistics* 27(4): 558–589.

Ellis, Nick C., Ute Römer, and Matthew B. O'Donnell. 2016. *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar.* Malden, MA: Wiley-Blackwell.

Flowerdew, Lynne. 2015. Data-Driven Learning and language learning theories: Whither the twain shall meet. In *Multiple affordances of language corpora for Data-Driven Learning*, edited by Agnieszka Leńko-Szymańska, and Alex Boulton, 15–36. Amsterdam: John Benjamins.

Gabrielatos, Costas. 2005. Corpora and language teaching: Just a fling or wedding bells? *Teaching English as a second or foreign language* 8(4): 1–34.

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

Johansson, Stig. 2009. Some thoughts on corpora and second-language acquisition. In *Corpora and language teaching*, edited by Karin Aijmer, 33–44. Amsterdam: John Benjamins.

Johns, Tim, and Philip King. 1991. Classroom concordancing. *English Language Research Journal* 4: 17–25.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: Chicago University Press.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar. vol. 1: Theoretical prerequisites.* Sandford, CA: Sandford University Press.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar. vol. 2: Descriptive application.* Sandford, CA: Sandford University Press.

Lee, Hansol, Mark Warschauer, and Jang Ho Lee. 2019. The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics* 40(5): 721–753.

McCarthy, Michael J., Anne O'Keeffe, and Steve Walsh. 2010. *Vocabulary matrix: Understanding, learning, teaching.* Andover: Heinle Cengage Learning.

Mizumoto, Atsushi, and Kiyomi Chujo. 2015. A meta-analysis of Data-Driven Learning approach in the Japanese EFL classroom. *English Corpus Studies* 22: 1–18.

Mukherjee, Joybrato. 2006. Corpus linguistics and language pedagogy: The state of the art–and beyond. In *Corpus technology and language pedagogy*, edited by Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, 5–24. Frankfurt am Main: Peter Lang.

Meunier, Fanny. 2019. A case for constructive alignment in DDL: Rethinking outcomes, practices and assessment in (data-driven) language learning. In *Data-Driven Learning for the next generation: Corpora and DDL for pre-tertiary learners*, edited by Peter Crosthwaite, 13–30. London: Routledge.

Myles, Florence. 2015. Second language acquisition theory and learner corpus research. In *The Cambridge handbook of learner corpus*

*research*, edited by Sylviane Granger, Gaëtanelle Gilquin, and Fanny Meunier, 309–331. Cambridge: Cambridge University Press.

O'Keeffe, Anne. 2021a. Data-Driven Learning and the second language acquisition interface debate. In *Beyond the concordance: Multiple applications of language corpora for language education*, edited by Pascual Peréz-Paredes, and Geraldine Mark, 35–55. Amsterdam: John Benjamins.

O'Keeffe, Anne. 2021b. Data-Driven Learning: A call for a broader research gaze. *Language Teaching* 54(2): 259–272.

O'Sullivan, Íde. 2007. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL* 19(3): 269–286.

Perek, Florent. 2015. *Argument structure in usage-based construction grammar.* Amsterdam: John Benjamins.

Pérez-Paredes, Pascual. 2019. A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning* 35(1/2): 36–61.

Pérez-Paredes, Pascual, Geraldine Mark, and Anne O'Keeffe. 2020. *The impact of usage-based approaches on second language learning and teaching.* Cambridge education research reports. Cambridge: Cambridge University Press.

Römer, Ute. 2019. A corpus perspective on the development of verb constructions in second language learners. *International Journal of Corpus Linguistics* 24(3): 268–290.

Sharwood Smith, M. 1991. Speaking to many minds: On the relevance of different types of language information on the L2 learner. *Second Language Research* 7(2): 118–132.

Sharwood Smith, Michael. 1993. Input enhancement in instructed SLA. *Studies in Second Language Acquisition* 15(2): 165–179.

Slobin, Dan I. 1987. Thinking for speaking. *Proceedings of the thirteenth annual meeting of the Berkeley Linguistics Society*, 435–445.

Tomasello, Michael. 1992. *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tyler, Andrea E., and Lourdes Ortega. 2018. Usage-inspired L2 instruction: An emergent, researched pedagogy. In *Usage-inspired L2*

*instruction: Researched pedagogy*, edited by Andrea E. Tyler, Lourdes Ortega, Mariko Uno, and Hae I. Park, 3–26. Amsterdam: John Benjamins.

Vyatkina, Nina, and Alex Boulton (eds.). 2017. Corpora in language teaching and learning. Special issue of *Language Learning & Technology* 21(3).

Wulff, Stefanie, and Nick C. Ellis. 2018. Usage-based approaches to SLA. In *Bilingual cognition and language: The state of the science across its subfields*, edited by David T. Miller, Faith Bayram, Jason Rothman, and Ludovica Serratrice, 37–56. Amsterdam: John Benjamins.