The Routledge Handbook of Corpora and English Language Teaching and Learning

Reka R. Jablonkai, Eniko Csomay

DDL for Younger Learners

Publication details
Peter Crosthwaite
Published online on: 29 Jul 2022

How to cite :- Peter Crosthwaite. 29 Jul 2022, DDL for Younger Learners from: The Routledge Handbook of Corpora and English Language Teaching and Learning Routledge
Accessed on: 13 Dec 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
25
DDL FOR YOUNGER LEARNERS

Peter Crosthwaite

25.1 Introduction
Advances in the availability, functionality, and user-friendliness of language corpora have resulted in a concomitant rise in the number of Data-Driven Learning (DDL) studies focusing on young learners (YLs) in recent years, with generally positive qualitative and even quantitative results now finally reported in the literature. However, there are still a number of barriers to the implementation and integration of DDL in the primary and secondary classrooms, including the lack of “pedagogic processing” of existing (mainly adult-focused) materials, a lack of software for specific use with YLs, and a lack of corpus literacy reported for pre-/in-service teacher trainees of YLs that often leads to a reluctance to adopt DDL pedagogy. This chapter discusses the affordances of direct, hands-on corpus use by pre-tertiary (language) learners for the purposes of data-driven learning, while also outlining the challenges involved in the successful implementation of DDL with younger learners. I summarise the successes (and occasional failures) of such studies to date, while reflecting on strategies to increase the adoption of DDL for YLs in future research.

25.2 Review of current state of research
Data-driven learning typically involves direct learner engagement with language corpus data, either through the use of teacher-printed concordance materials or learner-led hands-on corpus consultation, involving concordancing software or (increasingly) the use of online multimodal corpus-based applications that allow for students to learn and internalise statistical and contextual information about language in use in the process of working as “language detectives”, with “every student a Sherlock Holmes” (Johns, 1997, p. 101). DDL, as a pedagogical approach, is proposed to promote autonomous and constructivist learning in its adoptees, in that learners “learn best when they discover or can be led to discover themselves” (Cobb, 1999, p. 15). DDL provides its adoptees increased opportunities for data-enhanced, learner-centred focus-on-form (Bernardini, 2001; Long, 1991), allowing for inductive learning where the learner is responsible for the detective work, as well as allowing language teachers to aid learners’ deductive reasoning through guided support and scaffolding (Flowerdew, 2009).
Doubtlessly, other chapters in this volume cover the general affordances of DDL in greater detail (Fiona & Karlsen, this volume; Meunier, this volume). Hence, the discussion now moves on to DDL for YLs (defined in the present chapter as children aged between 6 (the first grade of primary school in Australia and also in numerous other contexts) and 16 years old (the final year of secondary school in Australia and some other contexts too). While DDL studies on learners younger than six are certainly possible, it is by first grade that many children will have learned the basics of reading in their own languages, and so will have attained a key prerequisite for consulting concordance data.

Despite a large increase in the number of DDL studies in the last 10 years as corpora increase in size and accessibility and as corpus applications increase in processing power, user-friendliness, and data visualisation, there still remains a dearth of DDL research that deals with its affordances for YLs. Three recent meta-analyses of the DDL literature (Boulton & Cobb, 2017; Lee et al., 2019; Pérez-Paredes, 2019a) have pointed to a severe lack of empirical data for DDL at the secondary years of education. At the time of writing, just over 5% of published studies explicitly state that they were conducted in high-school settings (Boulton, 2019), while the number of DDL studies on primary-age data “can probably be counted on one hand” (Crosthwaite & Stell, 2019, p. 150) – although I hope this is not still the case by the time this chapter goes to print. I expand on possible reasons for the lack of DDL studies with YLs in the following section.

25.3 Core issues and topics

This chapter now focuses on two main issues regarding the use of corpora and DDL with YLs. The first issue is whether YLs’ ability or aptitude to engage with DDL may be different from that of older learners, while the second is related to more adult concerns – namely, whether and how DDL may be integrated by teachers into the YL classroom.

25.3.1 Younger learners and corpus consultation

For many parents who lived through the pre-internet era, the connected world they live in today is now very different from the one that they were raised in. Today, many children are learning (about) languages through internet-connected tablet devices and interactive games, are searching for information online through search engines or wikis, are viewing instructional videos online, and are communicating with others in real time via video-conferencing and chat software. While a welcome and fascinating development, we need to be careful not to assume that YLs are already equipped with the technical knowledge and ICT skills required to adopt DDL-like learning practices, as one might expect if one erroneously considers YLs as “digital natives” (Prensky, 2001). While younger children may indeed have acquired a range of ICT skills even at the primary levels of education, there may still be a gap between their technical abilities and the kind of information literacy, inductive noticing, and deductive reasoning abilities required for successful engagement with corpus data as would normally be required for DDL to be effective. For example, when exploring the affordances of the web-as-corpus with YLs, such learners have been shown to adopt an “in and out” approach to consulting online search engine data (Thompson, 2013), and it is only after a process of explicitly training YLs in deductive reasoning and query refinement that YLs can turn basic search engine consultation into a more linguistically-oriented language learning opportunity (Gatto, 2019). While adult learners also often require explicit training in these processes for DDL to be effective (Boulton, 2009), children lack the acquired contextual knowledge
DDL for younger learners

from which to effectively apply deductive reasoning and may lack the cognitive development required for fast induction to take place (Piaget, 1970). YLs are also developing in their L1 literacy, leaving sudden exposure to a multitude of authentic (i.e. unmodified) L1 corpus data potentially problematic (let alone L2 data), and corpus output may require significant modification to be comprehensible (see sections below). Therefore, there is still a major role to play for teachers of YLs in effectively preparing their students for DDL, and there should be no expectation that YLs will simply “pick it up themselves” if left to their own devices.

That said, if YLs are able to receive initial guidance in the basics of corpus literacy, it does not take them very long to develop autonomous and rather idiosyncratic corpus usage habits. In one of the very few studies involving primary age learners, Crosthwaite and Stell (2019) introduced the online SketchEngine for Language Learning (SKeLL, Baisa & Suchomel, 2014) corpus platform with two 5-year-old boys within a private home tutoring setting. The tutor forwarded copies of the students’ writing to the researcher who provided feedback on aspects of their writing that would be amenable for corpus consultation, including certain query strings the students could use. For 10 minutes before their regular writing instruction, the students consulted SKeLL as the tutor discussed with them and made notes about their experiences. Overall, both students reacted positively towards corpora as a tool that could help them resolve lexical issues in their writing and both reported higher general self-efficacy as a result of their new-found language problem-solving skills after just five short sessions with corpora. In particular, the children took the opportunity to progress through the DDL training and revisions at their own pace, eschewed the query syntax provided by the “expert”, and reported frequently accessing the corpus platform long after its introduction – something less commonly reported in studies involving adult learners. That YLs may be better at DDL than their teachers think was also a finding in a study by Kim (2019) within the Korean primary EFL context. Following the introduction of paper-based concordance materials, three teachers and eighteen 6th grade students were interviewed about their experiences. Interestingly, while the teachers were convinced that a substantial level of guidance and scaffolding would be required for their students to adopt DDL practices, the students themselves reported enjoying great progress discovering the target rules exemplified through the concordances through peer discussion with their classmates. Additionally, Liontou (2019) experimented with DDL for the teaching of idioms to intermediate level EFL learners in Greece, involving just a single hour of training in how to use the BYU-COCA corpus platform (Davies, 2009) combined with printed extension activities involving corpus data. The results of pre- and post-tests on learners’ knowledge of idioms were significantly higher for students who access these materials than those who learned only through consulting school textbooks, an achievement made more impressive as a result of the very limited training time provided.

25.3.2 Issues with the integration of corpora into the YL classroom

As mentioned, the use of corpora in the primary/secondary classroom is still incredibly limited. While teachers of YLs are very familiar with the terms Computer-Assisted Language Learning (CALL) and Information and Communication Technology (ICT) (Voogt et al., 2013), they are, of course, quite unfamiliar with the term DDL. A key aspect of CALL theory and research over the past 20 years has been the integration and normalisation of technology into classroom settings (e.g. Bax, 2003; Chambers et al., 2004). Numerous CALL technologies including smartphones, PowerPoint, e-whiteboards have now become normalised within many teaching and learning contexts, yet the integration and normalisation of corpus
technology into pre-tertiary (and even, admittedly, tertiary) classroom practice have failed to materialise, despite numerous recently reported innovations and an increase in empirical support. For such integration to occur, Eng (2005) suggests a three-phase sequence from an ‘emerging’ phase of infrastructure development, an ‘application’ phase within existing CALL/ICT related teaching and learning processes, and an ‘infusion’ phase where teachers (and their students) are able to use the technology innovatively for a range of pedagogical purposes. In my introduction to the first edited volume on DDL for YLs (Crosthwaite, 2019), I argued that while the ICT infrastructure generally exists (at least in most rich Western contexts, an issue to which I return later), corpus infrastructure is not yet appropriate for general use with YLs (see also Braun, 2007). Regarding application, the current state of DDL research in pre-tertiary contexts is mostly restricted to teacher training contexts, with an incredibly small number of studies providing empirical data of corpus use with actual students. Even when introducing DDL into teacher training contexts, the results are not always positive (see below). This situation precludes there is still a long way to go before the final infusion stage outlined by Eng (2005), where teachers and students are able to use the technology in creative and unforeseen ways. As Lee (2011) suggests, “the appropriate and effective use of corpora in the classroom is partly a technical issue, but primarily a pedagogical one” (p. 159, emphasis mine).

To help overcome this situation, Whyte and Schmid (2018) suggest three current critical issues be addressed for the successful integration of CALL technology in younger (English) learner classrooms, namely (1) classroom interaction and digital interactivity, (2) design and implementation of teaching and learning tasks, and (3) challenges of orchestrating complex technology-mediated interaction with YLs. It must be said here that Whyte and Schmid are discussing CALL generally, so we must now fill in the blanks regarding the application of these recommendations for DDL.

Regarding the first issue mentioned above, teachers must learn how to apply the technical competence gained post-corpus training into appropriate didactic and dialogic opportunities for meaningful interaction between themselves, the corpus technology, and their students (Beauchamp & Kennewell, 2010; Glover et al., 2007) if such learners are to appropriately develop the type of constructivist, take-charge approach to learning espoused by DDL within their current learning practices. However, corpus literacy among teacher trainees is still sorely lacking (Chen et al., 2019). Many teachers of YLs also lack the technological, pedagogical, and content knowledge (TPACK, Koehler & Mishra, 2009) to integrate CALL applications (including, but not restricted to corpora) into teaching practice (Taghizadeh & Yourdshahi, 2020), and more focus needs to be given to developing teachers’ use of technology that can support their own pedagogical work and core teaching performance (McKenney & Visscher, 2019).

With specific reference to teacher training for DDL, Schaeffer-Lacroix (2019) conducted a series of training sessions in DDL-focused lesson planning with trainee secondary teachers of L2 German in France. Despite 15 hours of training on technology-enhanced learning (including a number of sessions on DDL), most teachers found it difficult to design appropriate DDL tasks for younger secondary school learners, failed to incorporate all but the simplest concordance-reading activities into their lesson planning, or insisting on presenting paper-based concordances only for their learners due to doubts about the abilities of their learners to engage with corpora directly. Latif (2021) reported on DDL training for 19 secondary school teachers in the Gulf EFL context, which resulted in initially very strong support for the implementation of corpora within their teaching practice. However, when interviewed again two years later, the same teachers reported that they had not been able to successfully implement
DDL for younger learners
corpus use into their classroom practice to the extent they had planned to following their training. Similar concerns have also been reported in a study by Leńko-Szymańska (2017) on the experiences of pre-service teachers who were trained in DDL lesson planning over the course of a semester, with the trainees mastering only basic corpus consultation competence post-training, and lacked the pedagogic skills to put what they had learned about corpora into classroom practice. However, the learners in Schaeffler-Lacroix’s (2019) and Leńko-Szymańska’s (2017) studies were fortunate to at least have the opportunity to take part in extended DDL training. Generally, in-service pre-tertiary teachers appear to have very little time for extra professional development activities (Bingimlas, 2009), and my previous attempts to organise a series of workshops on how corpora can be used in EAL lesson planning and materials development in Australia were met positively during initial correspondence. However, despite invites going out to over 30 staff in Brisbane Catholic schools, only three were able to attend a single four-hour session during the busy summer period. Complications related to COVID-19 have also made further engagement more difficult.

Despite such complications, small-scale opportunities for professional development have proven effective, with Tyne (2012) working together with two secondary school teachers of L2 Spanish in France to identify specific instances of where corpora could be used to support learning outcomes in current pedagogical sequences. The teachers reported being better able to carry out their lesson and activity planning following corpus literacy training, leading Tyne to suggest that DDL techniques – with a little help – can be aligned to suit the everyday practices of “ordinary” teachers (p. 136).

Regarding the second area mentioned by Whyte and Schmid (2018) above, learner-centeredness, authentic language use, and reflective learning practices are crucial elements in the design of successful CALL teaching and learning tasks for YLs, and it is not difficult here to see the parallels between these elements and what has been proposed to occur during learners’ engagement with corpora during DDL. However, regarding DDL instruction, there is still a lack of constructive alignment (Biggs, 1996) between CALL curricula, teaching methods, and assessment tasks, with learners often not playing an active nor reflective role in their own learning (as required for constructivist learning to occur); with learners’ output measured mainly through cloze tasks rather than text construction/revision; and with DDL mainly limited to out-of-class experiments rather than embedded into the curriculum (Boulton, 2010; Meunier, 2019).

In the secondary classroom, Wicher (2019) suggests that the current didactics of language learning involve the well-known presentation-practice-production (PPP) models of instruction including textbooks, form-focused exercises, and pair/group communication, alongside Task-based Language Teaching (TBLT) approaches (e.g. Shintani, 2011) that combine authentic tasks with focus-on-form. Wicher believes DDL can cater for both paradigms, with concordances used at the presentation stage, learner interaction with/modification of concordance results at the practice stage, and teacher/peer discussion of concordance findings at the production stage under a PPP model. Likewise, TBLT offers “the most fertile terrain for DDL” (Wicher, 2019, p. 39) through opportunities for input enhancement at the pre-task stage and consciousness-raising focus-on-form at post-task stages. For either to succeed, and given the complexities of teaching and learning at the secondary level, Wicher suggests an increased need for flexible design including internal differentiation (see also Coffey, 2018) of input enhancement – including modifying concordances – for learners of different technical/linguistic abilities. This should help to reduce any negative consequences for weaker learners as they try DDL for the first time, and while this may result in a trade-off regarding the “authenticity” of corpus data, i.e. unmodified data obtained from a corpus containing naturally
occurring language samples, and its comprehensibility, Wicher suggests that such “corpus worship” (Gabrielatos, 2005) needs to give way to internal differentiation if DDL is to be adopted by both YLs and their teachers.

As for learner-centeredness at the primary school level, traditional approaches to input enhancement and exposure to target forms in the primary L1/L2 learner classroom include teachers’ read-aloud stories (Lin, 2014) and home-book reading (Roberts, 2008), along with multimodal approaches such as songs and physical gestures (Total Physical Response, Asher, 2009), and explicit teaching and rote memorisation. There are also an increasing number of apps and online games that are used both in and out of class in many contexts. Certainly, for primary-age learners, having access to multimodal forms of data for DDL is very important, given that such learners progress in their literacy development from pictures with word captions, to picture books, then on to chapter books. In particular, while all language learners benefit from repeated exposure to target language forms in context, younger L2 learners stand to benefit more from multimodal corpus consultation as they lack either the fast-mapping ability young L1 learners have to learn new words from a single encounter (Clark, 1993) or the socio-pragmatic competence utilised for language acquisition by older learners (Goto-Butler, 2019). While there are corpora of children’s chapter books available for use with traditional concordancers (Montag et al., 2015) as well as corpora containing graded texts that are specifically written for learners at primary/secondary grades (e.g. the Weebit corpus, Vajjala & Meurers, 2012), there are still few concordancers designed specifically for use for primary-age students. One useful exception is the work of Eri Hirata (2016, 2019), who developed a multimodal corpus tool (or MmCT) for use with primary-age L2 English learners in the Japanese context. The MmCT boasts a MovieConc feature where learners can access video, audio, and subtitle text together on different sides of the same screen, and where the video/audio can be played at slower speeds to aid comprehension. Learners can search for target words or larger expressions and be taken directly to relevant movie clips containing these expressions. Teacher trainees reported preference towards using this multimodal resource as compared to the use of the text-only BYU-COCA online platform (Davies, 2009) in the classroom with their primary-age learners (Hirata, 2019).

Regarding the third item discussed by Whyte and Schmid (2018), there is a paradox for YLs where engagement with CALL technology is difficult due to limited linguistic resources, but where interaction with said technology is also required to gain the relevant linguistic resources. Whyte and Schmid discuss this in relation to L2 learning, although I suggest this can apply equally to primary-age L1 learners using corpora originally designed for adults. While there have been numerous reports in the literature of adult learners (or teachers) finding corpus consultation difficult, for YLs the idea of pedagogic processing, or, in other words, the simplifying of corpus materials for use with YLs, can greatly increase the likelihood of success with YLs’ initial forays into corpus consultation (Flowerdew, 2009; Pérez-Parades, 2010; Wicher, 2019). Kim (2019) reported that for younger L2 learners, it was also advantageous to have access to L1 translations or glosses for difficult/complex vocabulary, although such corpora are relatively rare. In addition, corpora designed specifically for pedagogic use with YLs – or pedagogic corpora – can have a significant advantage over commonly used online corpora such as BYU-COCA or the BAWE within SketchEngine (Alsop & Nesi, 2009) in the secondary classroom (Pérez-Parades, 2019b). While the latter was primarily designed for research purposes and L1 linguistic representativeness, pedagogic corpora are specifically designed to accommodate specific L1/L2 learners, levels, and needs. A useful example is the SACODEYL pedagogic corpus (Braun, 2007; Pérez-Parades, 2019) that is composed of data taken from video recordings of teenagers’ oral interactions on everyday topics across multiple
DDL for younger learners

L1s. This data allows both vertical and horizontal reading to allow learners to see more clearly how discourse is structured, while the data is segmented and annotated according to teachable discourse functions, e.g. the use of conditionals, which makes searching the corpus for particular functions much easier for teachers and learners. In addition, having the video recordings available for listening practice and other important multimodal aspects of discourse such as, gesture, provides an additional entry point into the data that may be useful for secondary learners who find reading concordances difficult.

Despite the lack of pedagogic corpora currently available, Boulton (2019) suggests there may actually be a great deal of DDL going on in pre-tertiary classrooms although this may not be in a form where students sit at a computer using a typical concordancer. Rather, Meunier (2019) suggests expanding the scope of tools and tasks commonly used in DDL activities by leveraging tools that promote constructivist learning through data consultation but that are different in form and function from traditional concordancers. By generating appropriate accompanying tasks that require the same kind of noticing and pattern-sleuthing processes as found under a more typical DDL approach (see Kennedy & Miceli, 2010 for an overview of such processes at the tertiary level), one may greatly expand their repertoire of available DDL tools to include more multimodal resources, and – crucially – tools that YLs in particular can both easily use and find appealing. Meunier uses the websites PlayPhrase.me (https://www.playphrase.me/), a web resource where students can search for phrases used in a database of film clips and see the phrases and accompanying clips together), and LyricsTraining (http://fr.lyricstraining), which presents gap-fill questions or a karaoke mode based on a database of song lyrics organised by proficiency level as potential resources that teachers looking to adopt DDL practice can try with YLs, despite these resources being very different in feel and function to traditional concordancers.

25.4 Current contributions and research

Outside of the Crosthwaite (2019) dedicated edited volume on DDL for YLs, a number of studies featuring corpus use with younger learners have been published over the last few years, and this section provides an overview of this research.

For the acquisition of collocational patterns, Boontam and Phoocharoensil (2018) used paper-based corpus materials with young fourth grade Thai EFL students at pre-intermediate levels of L2 proficiency, as they learned about the English prepositions *during*, *among*, and *between* through DDL. The learners read teacher-printed concordance materials containing the target items, extracting from the Graded Readers Corpus within the Compleat Lexical Tutor (http://www.lextutor.ca/, Cobb, 2005). The results of pre-/post-tests on the target items found significant improvement in learners’ knowledge of these constructions at the post-test stage, measured via gap fill, sentence building, and grammaticality judgement tasks. The students themselves reported high satisfaction with the DDL approach; in particular, in gaining the ability to learn about language by themselves, while only six of thirty learners reported difficulty interpreting the selected concordances. Also covering EFL learners’ development of L2 English prepositions, Özbay and Olgun (2017) compared learners’ acquisition of these forms using traditional textbook-based instruction versus printed concordances on groups of high-school students in Turkey across intermediate and advanced L2 proficiency levels. The researchers found significant differences between experimental and control groups’ knowledge of L2 preposition collocations following 15 weeks of instruction. Finally, Saeedakhtar et al. (2020) explored the use of DDL for the learning of verb–noun collocations with lower-intermediate level secondary school learners in the Iranian context. Comparing
hands-on direct corpus use with hands-off DDL involving printed concordances, participants using both approaches reported learning gains over a control group not using DDL in immediate post-tests, although learners in the hands-on condition outperformed learners in the hands-off condition in delayed post-tests.

For vocabulary, Soruç and Tekin (2017) explored the development of secondary school students’ L2 English lexical knowledge using DDL in the Ugandan context, comparing a paper dictionary-only control group with an experimental group who received training in understanding pre-selected concordance output before consulting a range of online corpora (the authors do not say which, unfortunately). Using a standardised pre-/post-/delayed post-test procedure, the authors report significant improvement in the experimental group over the improvement noted for the control group, as well as positive qualitative reports from the learners themselves. Tekin and Soruç (2016) also looked at L2 English vocabulary development in a Turkish high school setting using the British National Corpus (although they do not say which website was used to access the corpus). Learners received training in using the various functions of the platform together with printed concordances in order to investigate target items from the same standardised vocabulary test reported in Soruç and Tekin (2017). Semi-structured interviews with participants revealed five common themes arising from the DDL treatment, including “innovative”, “autonomous”, “easy and fun”, and “practical” as positive aspects of corpus use, with “complex” (Soruç & Tekin, 2017, p. 1276) representing learners’ negative opinions of the treatment, with the latter related to concordances being cut-off or being too difficult to decode. Overall learner satisfaction with the DDL treatment was over 75%. Also in Turkey, Yılmaz and Soruç (2017) trained twenty 14–16-year-old EFL students on how to use BYU-COCA over eight hours in a private language school setting. Again, performance on pre-/post-test vocabulary tests was significantly higher for those using corpora than a control group who did not. Also focusing on DDL for vocabulary is a study by Karras (2016) conducted in international secondary schools in the Vietnamese EFL context. A longitudinal experimental/control group training procedure on vocabulary acquisition was performed over an eight-week period, with the experimental group adopting a DDL approach through the Compleat Lexical Tutor software (Cobb, 2005), and with the control group using an online dictionary. The study is notable for its distinct treatment of EFL students at international schools as either non-local ‘third culture kids’ (TCKs) from a range of countries or ‘educational cross-cultural kids’, who comprise local children attending an international school. This is an important distinction for DDL in that Karras (2016) claims that TCK learners, in particular, are atypical of most secondary EFL learners, as there may be “a tripartite relationship between the TCK experience, attending an international school, and their linguistic and academic prowess” (p. 182). This tripartite relationship may result in increased motivation to try new approaches when learning second languages, including the kind of autonomous, constructivist learning practices associated with direct DDL.

In terms of improving grammar, Moon and Oh (2018) explored the use of DDL with Korean middle school EFL learners to “unlearn” (p. 48) the overgeneration of the copula before thematic verbs (e.g. he is dance very well), a common error in Korean L2 English. Almost 200 students were divided into either a traditional textbook-based instructional group or an experimental DDL group, who received printed concordances from a graded reader corpus composed of articles written for Time Magazine for Kids (www.timeforkids.com) as well as from a learner corpus generated from the learners’ own writings, before consulting these corpora hands-on using Wordsmith Tools (Scott, 2008). Learners taking the DDL treatment produced significantly fewer redundant copula post-training than the control group. In particular, the authors stress the importance of exposing learners to negative evidence of
DDL for younger learners

erroneous forms via consulting a learner corpus generated from their own writings, as seen in Crosthwaite (2017). Additionally, in a study focusing on the Greek EFL context, Michali and Patsala (2020) used a combination of printed concordance handouts and instructional materials on querying the British National Corpus and COCA to aid English grammar teaching with 36 students at junior high school. After just two teaching hours, the young learners reported very positive opinions about the use of corpora for learning about both grammar and vocabulary, although reported that the examples taken from the two online corpora were above their level of proficiency.

25.5 Future directions of research

One of the major issues still facing those who wish to increase DDL adoption in YL contexts is simply a lack of empirical data on its effectiveness in the classroom. Whenever writing grant applications involving corpus consultation for YLs, I always seem to receive feedback along the lines of “but there is no evidence that it works”, despite the wealth of evidence now available for how older learners can and do benefit from DDL. Of the few DDL studies that have been performed in YL contexts, and aside from the empirical studies reported in the section above, most studies deal with aspects of teacher training for YL or YL teachers’ reactions to corpus use, leaving data from students themselves sorely lacking. This situation is likely because teacher training contexts are as far as most researchers have been able to penetrate into pre-tertiary school systems to conduct DDL research, due to the numerous ethical and bureaucratic complexities involved in actually getting access to young children from whom to collect data. In addition, of the studies already cited in this chapter involving YL teacher training (e.g. Kim, 2019; Schaeffer-Lacroix, 2019), despite initial positivity during such training, many teachers remain sceptical of how YLs will react to or engage with corpus use, and given the vast range of other workload demands placed on primary and secondary teachers, many are still unwilling to take the risk of even partial classroom implementation during actual teaching periods. These rational fears (Boulton, 2009) are potentially preventing the kind of large-scale funded projects required to empirically test the affordances of DDL with YLs from getting off the ground, and it may take more significant developments involving DDL with older learners in mainstream education before the pendulum can eventually swing to increased opportunities for YLs.

Another area for future research is that while many DDL studies have been conducted in technologically-rich Western or East Asian countries, the benefits of DDL are yet to be realised in regions such as the Middle East, S.E. Asia, and Africa. As a result, little is currently known about teacher trainees’ or educators’ acceptance of or preparation for DDL in these regions. This situation may be reflective of a digital divide (Lozano & Izquierdo, 2019) between relatively ICT-poor countries and their richer neighbours. Reporting on teachers of YLs’ attempts to implement CALL technology into the classroom in the Middle East, Taghizadeh & Yourdshahi (2020) claim that a large number of teachers receive little to no CALL training within initial teacher education. In addition, limited ICT infrastructure and a lack of support from schools were major barriers to teachers’ willingness to undertake appropriate professional development to remedy their technological or pedagogical shortcomings. The above-mentioned work by Soruç and Tekin (2017) is the first DDL study that I am aware of that has tested the affordances of DDL in the African context. Crosthwaite et al. (2021) and Crosthwaite (2020) used Zoom workshops to introduce DDL to trainee L2 English teachers at a teacher training facility in Indonesia. Generally, positive perceptions of DDL were reported for secondary school trainees, while primary school trainees were more negative in their
views. More studies (aside from those already mentioned in this chapter) from other contexts including South America, Africa, and S.E. Asia are of course welcome. Additionally, it would also be beneficial for the field as a whole if more DDL studies involving YLs were conducted where the target language is other than English. While Di Vito (2019) investigated the acquisition of L2 French within the Italian YL context, and Schaeffer–Lacroix’s (2019) looked at teachers of L2 German in the French context, there is a significant need for DDL studies featuring YLs for languages other than English. This is highlighted as a key factor in Vyatkina’s (2020) explanation for the underutilisation of corpora in mainstream education, where she calls for a broadening of contexts for DDL beyond EFL/ESL, broadening access to DDL resources, increased scaffolding of DDL for teachers, and an increase in open DDL resources and corpora for general use.

Further reading


Crosthwaite, P. (2019). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge. This is the first edited volume of DDL studies specifically targeting research involving YLs. The book is organised into three sections, with the first dealing with overcoming challenges for DDL with YLs, the second on applying new DDL methods with YLs, while the third section covers a small number of empirical studies involving YLs corpus use at the primary and secondary levels of education.

Sealey, A., & Thompson, P. (2004). ‘What do you call the dull words?’ primary school children using corpus-based approaches to learn about language. *English in Education*, 38(1), 80–91. This is one of the very few studies dealing with DDL for primary school education, exploring the potential for teaching Key Stage 2 pupils in the UK context using corpus materials. The study includes valuable examples of the teaching activities used, as well as extracts of discussions between the researcher and two groups of pupils as they attempt to use corpora to learn the distinction between open and closed word classes.

References


DDL for younger learners


Butler, Y. G. (2019). Teaching vocabulary to young second-or foreign-language learners: What can we learn from the research. *Language Teaching for Young Learners, 1*(1), 4–33. https://doi.org/10.1075/ltyl.00003.but


Fukuoka Jo Gakuen University Bulletin. Faculty of International Career Development, 2, 19–32.

on the integration of DDL in teacher education. In P. Crosthwaite (Ed.), Data driven learning for 
the next generation: Corpora and DDL for pre-tertiary learners (pp. 88–105). Routledge. https://doi. 
org/10.4324/9780429425899-6

program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), Teaching and language 
corpora (pp. 100–115). Longman.

international school students in Vietnam. ReCALL, 28(2), 166–186.

learners to a corpus as a reference resource. Language Learning & Technology, 14(1), 28–44. http://dx.doi.org/10125/44201

Kim, H. (2019). The perception of teachers and learners towards an exploratory corpus-based 
grammar instruction in a Korean EFL primary school context. Primary English Education, 25(1), 
123–152.

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? 
Contemporary Issues in Technology and Teacher Education, 9(1), 60–70.


Lee, S. (2011). Challenges of using corpora in language teaching and learning: Implications for sec-


classrooms. English Language Teaching, 7(7), 68–81. https://doi.org/10.5539/elt.v7n7p68

Liontou, T. (2019). The effect of data-driven learning activities on young EFL learners’ processing of 
English idioms. In P. Crosthwaite (Ed.), Data driven learning for the next generation: Corpora and DDL 
for pre-tertiary learners (pp. 208–227). Routledge. https://doi.org/10.4324/9780429425899-12

R. Ginsberg, & C. Kramsch (Eds.), Foreign language research in cross-cultural perspectives (pp. 39–52). 
Benjamins. http://dx.doi.org/10.1075/sibil.2.07lon


Greek EFL learners’ attitudes towards corpus-based teaching, INTED2020 Proceedings, 6871–6881. 
https://doi.org/10.21125/inted.2020.1827


Meunier, F. (2019). A case for constructive alignment in DDL: Rethinking outcomes, practices and 
assessment in (data-driven) language learning. In P. Crosthwaite (Ed.), Data driven learning for the 
next generation: Corpora and DDL for pre-tertiary learners (pp. 13–31). Routledge. https://doi.org/ 
10.4324/9780429425899-2

%2F0956797615594361

Moon, S., & Oh, S. Y. (2018). Unlearning overgenerated be through data-driven learning in the sec-
ondary EFL classroom. ReCALL, 30(1), 48–67. https://doi.org/10.1017/S095834417000246

Özbay, A. S., & Olgun, O. (2017). The application of DDL for teaching preposition collocations to 

DDL for younger learners


