ARTIGO

Getting started with digital literacy using pedagogic corpora in young learners' English classes

Iniciando o letramento digital usando corpora nas aulas de Inglês de alunos das séries iniciais

Ana Lúcia Surerus Pitanguy MARQUES
alpitanguy@gmail.com
Universidade Federal de Minas Gerais, MG, Brasil.

Deise Prina DUTRA
deisepdutra@gmail.com
Universidade Federal de Minas Gerais, MG, Brasil.

FLUXO DA SUBMISSÃO
Submissão do trabalho: 26/04/23
Aprovação do trabalho: 11/11/23
Publicação do trabalho: 24/11/23

10.23925/2318-7115.2023v44i2e61842

Abstract
This paper addresses the provision of instruction to teachers and learners on becoming digitally literate and skilled with corpus-based tools. It proposes the compilation of specific pedagogic corpora, in this case in Geography and Science, to be used in the English young learners' classes in elementary education.The paper also illustrates the handling of user-friendly concordancer's tools of #LancsBox 6.0 to perform basic analysis of the corpora language through a more accessible technology. It concludes by summarizing the possible benefits of an alternative approach to expose learners to customized corpus-informed language in English.

Keywords: Pedagogic corpora, Digital literacy, Young learners, Concordancer tools.

Resumo

Este artigo discorre sobre a necessidade de professores e alunos se tornarem letrados digitalmente e habilitados a usar ferramentas de análise do conteúdo de corpora. Propõe-se a compilação de dados específicos, neste caso de Geografia e Ciências, para uso pedagógico em aulas de inglês do ensino fundamental I. O artigo ilustra a utilização básica das ferramentas digitais do software #LancsBox 6.0 que analisam o conteúdo através de uma tecnologia mais acessível. Conclui-se apresentando possíveis benefícios do uso de uma abordagem de ensino alternativa que exponha os jovens aprendizes ao conteúdo linguístico em inglês baseado em corpora.

Palavras-chave: *Corpora pedagógicos, Letramento digital, Jovens aprendizes, Ferramentas de concordância.*

1. Introduction

The most recent generation of young learners, currently in elementary school, is certainly one that challenges the learning / teaching boundaries of the past even further while demanding teachers' mentoring in new ways in the classroom. To this date, teachers' roles have been multi-faceted, trying to provide learners not only with the subject-matter contents of their lessons but also guiding them towards meeting learning goals. However, most of those roles are now being disputed as technology and portable devices, available to a large portion of the population, offer learners instantly the information required for their day-to-day lives. It is the turning point of the source of knowledge: the tools available can supply the present generation of learners with the right answers at the tip of their fingers.

The teachers are still invaluable as curators of the information readily available, but their roles have been changing fast as the new generation gradually takes agency of their own learning path. It is a long-held belief among educators that contemporary education should enable learners to be more engaged and committed to their own learning process and responsible for the results (Chambers, 2010). This is the window of opportunity teachers have to motivate learners to make effective use of the digital tools available to improve learning and lighten their weight as linguistic authorities they traditionally have had (Aston, 2007).

As an immediate result of the fast pace of change in the educational scenario, the sudden advance of technology has brought into the scene the need to know English to navigate and visit websites, as it is the most commonly used language on the internet. Young learners need to start learning at a much younger age and in tandem they need to acquire the ICT skills demanded by the new digital learning mode. All those involved in their literacy process should start considering that the changes have to take place urgently.

Thus, to broaden young learners' linguistic scope, we started to speculate about a way to accelerate the English learning in subjects like Geography and Science at a much younger age (Marinova-Todd *et al*, 2000), while attempting to integrate corpus-informed pedagogy in our *Fundamental* 1[1] schools. If, as a first step, second language (L2) teachers could resort to topicalized corpora, i. e., corpora in other subjects other than general English, to devise activities for cross-

[1] *Fundamental* I (Brazil) and Elementary (US) years are equivalent and will be used interchangeably.

curricular projects, they would be able to increase young learners' exposure to L2 vocabulary in their early school years. If learners could take part in the change, it would motivate and engage them in the process of discovering the intricacies of the language patterns by themselves (Schmidt, 1990; Johns, 1990; Sinclair, 1991). The corpora would offer the authentic language (McCarten, 2007) in the correct levels of proficiency and in the topics they would otherwise be only exposed to in their first language (L1). Such exposure could trigger deliberate learning, for example, with learners working with vocabulary where the primary aim of the activity is to learn intentionally and explicitly target words (Webb; Chang, 2012).

In response to the claims above, this paper addresses the demands of contemporary society for a more inclusive classroom. We focus on the advantages of using pedagogic corpora and the digital tools that facilitate its use. Section 2 discusses briefly relevant principles underlying second language acquisition which resonate with approaches now proposed for the classwork. Section 3 claims that both teachers and learners should be better oriented in the educational digital environments. It then proposes one way of improving vocabulary and language patterns learning by using pedagogic corpora and a concordancer which analyzes language to be used for class activities. In Section 4, it outlines the compilation of a pedagogic corpus with language at an appropriate level for elementary school students (Pérez-Parede, 2020) and in Section 5 describes how the contents are accessed and analyzed by a very user-friendly concordancing software - #LancsBox 6.0[2]. In Section 6, it describes how digital tools are used to examine some language combinations and patterns and hopes to motivate teachers and young learners to make meaningful use of it. In Section 7, it compares briefly the pedagogic corpus with a traditional adult corpus, in this case the BNC2014-baby. Section 8 has the authors' remarks and a conclusion showing the positive aspects of using technology with teachers and young learners.

2. Revisiting second language acquisition (SLA)

Research in the depth and breadth of SLA to date has reiterated the advantages of learners being exposed to contextualized L2 in the classroom. One widely recommended approach to vocabulary learning is the selection and use of different activities on the same topic where words

[2] < http://corpora.lancs.ac.uk/lancsbox/download.php > However, there is a new version of the software at <https://lancsbox.lancs.ac.uk/ >. Access In February 20, 2023.

are grouped semantically into lexical sets to increase the potential for the recurrence of target vocabulary exposure (Nation, 2020; Schmidt, 1990). Investigations have pointed out that noticing and discovering the relationships between the words that are presented together have a great impact on language learning (Ellis, 2012). Many of those multi-word clusters, words that are frequently together in a specific corpus, also known as formulaic language, are constantly being scrutinized by researchers. From the decade-old usage-based model (*Ibidem*, 2012) advocating the importance of frequency, context and recency[3] of formulaic language to enable learners to transform it into intake, to the more recent Formulaicity Principle[4] (O'Keeffe; Mark, 2023), it seems that language noticing, frequency of occurrence and recurrence of exposure have pivotal roles in the L2 learning literature.

Summarizing, the combination of explicit instruction (Ellis, 2002) with target language (TL) recurrence of exposure (Gabrielatos, 2005) can promote noticing (Schmidt, 1990) of words and multi-word sequences (Cortes, 2004). These can be made salient through learners' manipulation of concordance lines (Johns, 1990) in activities which can foster language retention. Concordance lines are micro contexts of specific keywords yielded by concordancer tools explained in detail in Section 6.

3. Digital literacy for teachers and learners

As of 2018 and onwards, the guidelines in the *Base Nacional Comum Curricular* (*BNCC*)[5] state that teachers need to be skilled and equipped to help young learners acquire and develop, among other competencies, the digital competency in the *Fundamental* I:

> 5. Understand, use and create digital information and communication technology, in a meaningful, reflexive and ethical way, in the various social practices (including the school ones) to communicate, access and share information, create knowledge, solve problems and take agency of own personal and collective life (BRASIL, 2018: 9)[6]

[3] Recency claims that the more recently we experience a construction, the stronger our memory of it is (O'Keeffe; Mark, 2023 adapted from Ellis, 2012).
[4] Formulaicity develops across levels of language proficiency and is a marker of an advanced learner (O'Keeffe; Mark, 2023).
[5] Brazilian National Syllabus Core.
[6] Authors' translation for: "5. Compreender, utilizar e criar tecnologias digitais de informação e comunicação de forma crítica, significativa, reflexiva e ética nas diversas práticas sociais (incluindo as escolares) para se comunicar, acessar e disseminar informações, produzir conhecimentos, resolver problemas e exercer protagonismo e autoria na vida pessoal e coletiva" (BRASIL, 2018: 9).

Therefore, it is of paramount importance to consider, first, the extent to which the teachers have access to the hardware and, second, if they have the technical skills to be able to navigate the online medium. They also need to learn to curate the appropriate information before they can implement innovations in the classroom. According to Meunier (2020), teachers themselves may still today not always see the added value of integrating them into their lessons. And recently, Crosthwaite (2022) corroborated those statements by positing that many young learners' teachers still lack both the technical and pedagogical knowledge to integrate Computer Assisted Language Learning (CALL)[7] applications into teaching practice.

The scenario described above reflects just some of the difficulties teachers have to overcome to transform their classrooms into 21st century educational environments. First, they need to learn to curate the information available online. Second, they need to learn to use the digital tools to search for language patterns, and only then they would be ready to help learners interpret the results and make effective use of the findings in the classroom. That is, when working with concordancing software and data analysis, teachers need to realize and accept the fact that their role changes fundamentally,

> as s/he is no longer the sole source of knowledge about the target language, but rather a facilitator of the learning process, helping the learners to interpret the data, and giving them advice on how best to search the corpus and analyze their search results (Chambers, 2019, p.354).

Learners also need to be digitally literate to look for information on the web. According to *BNCC,* they need to be exposed to the digital medium, learn to navigate safely and identify suitable and trustworthy sites which suit their learning goals. They also need to learn to analyze critically the information they receive or send. They need to be acquainted with the use of web search engines appropriately to understand how to find what they need and be able to understand and interpret output of its particular discursive functions in context (Hafner; Candlin, 2007). To empower learners with digital skills, Redecker insists that learning resources and activities, among other aims, should "open up learning to new, real-world contexts, which involve learners themselves in hands-on activities, scientific investigation or complex problem solving, or in other ways increase learners' active involvement in complex subject matters" (2017, p.22).

[7] The term Computer Assisted Language Learning (CALL) was coined by Hardisty and Windeatt (1989).

4. Pedagogic corpora for young learners

To change the teaching approach and empower learners with hands-on activities, this Section outlines the compilation of pedagogic corpora which are aimed to be used in language teaching in a future study and, therefore, has been designed with pedagogic purposes in mind (Pérez-Paredes, 2020; Willis, 2011). The balanced corpora, COREL-GEO+SCI[8], composed of textbooks and website oral and written texts represent the language variety in Geography and Science relevant in the specific teaching context (Friginal, 2018) of elementary school grades.

Hence, and in line with Willis's (1998) definition of pedagogic corpus as a body of texts to be used in the classroom to support teaching (texts from the learners' coursebooks) with any additional texts that the teacher may bring into the classroom (Gilquin; Granger, 2010), pedagogic corpora were compiled to suit elementary young learners' - 9 - 12 years old - needs and interests.

Small corpora would seem to be both useful as instruments of language learning in their own right, and as means of training learners to use corpora appropriately (Aston, 2007). Aston also mentions other benefits of small corpora: easier to construct, to interpret, to become familiar with and to allow language to be more fully analysable. While size is usually an issue, it should be considered hand-in-hand with the appropriateness of corpus design. In terms of suitability, however, it is often the design of a corpus as opposed to its size which is the determining factor (O'Keeffe *et al*., 2007, p.4).

Regarding corpora contents, besides the intended audience of young learners in elementary school, the authors considered their regular school syllabi, selecting subjects with an overarching reach: Geography and Science. That is, the corpora should serve the pedagogic aims of the project they had been designed for (Reppen, 2010; Xiao, 2010; Jablonkai, 2022). Topics (Table 1) were selected based on the usefulness and essentialness constructs for learners` school grades and the connection with the syllabi, so as to make the experience of working with concordance lines not only useful but also meaningful and beneficial.

[8] COREL-GEO+SCI - Corpora for Elementary levels, a conflation of two corpora: a Geography corpus and a Science corpus.

Table 1. Themes and topics for the pedagogic corpora.

30 Topics	
Animals	Climate
Habitats	Natural disasters
Life Cycles	Seasons
Food Chains	Soil
Endangered Species	Erosion
Biomes	Climate change
Ecosystems	Resources
Plants	Renewable
Lifecycles	Non-renewable
Pollination	Fossil fuels
Germination	Pollution
Photosynthesis	Recycling - Conservation
Rainforests	Humans
Bodies of Water	Communities
Landforms	The Solar System - The Earth

Source: the authors.

Texts from six different authentic sources within the A1-A2 framework of language[9] were then collected: printed workbooks A[10], printed workbooks B[11], web texts / video transcriptions and articles from assorted sources on the web.

Table 2. Subsets of corpora information.

Science: COREL-SCI	Total number of texts	Total number of tokens	Total number of types	Total number of lemmas	Mean number of words
Written textbook	334	54,821	4,032	3,429	164,13
Written web	78	22,224	3,467	3,129	284,92
Spoken web	25	13,824	2,018	1,839	552,96
TOTAL	437	90,869	9,517	8,397	207,93
Geography: COREL-GEO					
Written textbook	346	44,576	3,724	3,429	128,46
Written web	87	23,057	3,272	2,985	265,02
Spoken web	25	20,167	2,857	2,605	806,68
TOTAL	458	87,800	9,853	9,019	191,28
Overall	**895**	**178,669**	**19,370**	**17,416**	**199,40**

Source: #LancsBox.

The pedagogic corpus, COREL- GEO+SCI,[12] for this study is a conflation of two corpora with a total of 895 texts, a balanced number of 437 texts in Science (COREL- SCI) and 458 texts in Geography (COREL- GEO) separated by topic / theme (Table 2). They can also be identified by school grade if necessary. Corpus contents also make it possible to expose learners to authentic

[9] According to the CEFR – Common European Framework of Reference for Languages.
[10] 180 Days of Science and Geography, Shell Education, K to 5th grade, 2014.
[11] DK WORKBOOKS, Penguin Random House, Pre-K to 4th grade, 2016.
[12] The GEO and SCI corpora can be split in two and teachers would use them according to the syllabi.

language and actually present them with a large number of instances of a particular linguistic item to work with them all at once (Cobb, 1997).

5. The concordancer #LancsBox 6.0

To access the corpora data, it is paramount that teachers have the digital skills to work with a concordancer. Concordancers are common corpus analysis tools that search texts based on a word or phrase provided by the user and yield them in contexts called concordance lines or rank them according to their frequency in that corpora. This section explains how the concordancer was chosen and what its tools can accomplish at the user`s basic level of understanding. The literature suggests that if the digital tools are (1) hard to use or (2) perceived to be hard to use, then widespread adoption of the tools is not likely (Hendry; Sheepy, 2022). These authors mention the importance of the multidimensional construct of usability to identify and select the most appropriate concordancer to use in the classroom.

According to Hendry and Sheepy (2022), in a recent study comparing concordancing softwares, #LancsBox[13], a freely available online concordancer with a unique graphical interface, was found to be the easiest for some to use. The 6.0 version has a straightforward interface and accompanying tutorials and is very suitable as a first step to those teachers trying to get acquainted with the current technological tools. Those factors have prompted us to choose it taking into account the aims of its creators (Brezina; Gablasova, 2018). At the time, they declared that they were interested in improving learner vocabulary instruction through corpus analysis, mainly keyword and collocation analysis.

Once the software has been chosen, the first requirement for teachers and learners to output data is to know its tools. The software tutorials are short, the explanations objective, and the repetition of procedures associated with a little curiosity towards experimenting with other tools may help teachers overcome any initial barriers. The software already has many corpora embedded in its system such as American English, British English, BNC, Brown, LOB, English Literature, etc. Additionally, one can upload one's specialized corpus and use it like the example of the pedagogic corpus mentioned here and in the next Sections. By using a concordancer with

[13] <http://corpora.lancs.ac.uk/lancsbox/index.php >.

a readily-understood interface and meeting the criterium of being user-friendly, our main aim was to address the 'user-friendliness' aspect mentioned by Frankenberg-Garcia (2012) to show teachers the way to start developing their digital literacy.

6. Pedagogic corpora and the concordancer analysis tools

Hendry and Sheepy (2022, p.439) put forward the idea that "learners can use corpus analysis tools to support vocabulary acquisition (1) as a reference to identify important words to study, (2) as a reference to check for patterns in typical usage in authentic texts", language improvement and development of autonomous work. Hence, Table 3 illustrates the interface of the software #LancsBox 6.0 showing its basic tools in the black bar at the top. The tools were used, first, to extract lists of most frequent words, then, content words as well as multi-word 3-gram clusters from COREL- GEO (Corpus for Elementary Levels on Geography) and COREL- SCI (Corpus for Elementary Levels on Science) to be described in sections 6.1 to 6.4.

The next step was to have the most frequent items in the lists, and some were selected as KWICS (key words in context) to obtain the concordance lines - their micro contexts. Findings in concordance lines (KWIC tool) and collocation visualization tools (GraphColl tool) can help learners recognize and remember collocations (Hendry and Sheepy, 2022). This condensed exposure (Gabrielatos, 2005) can contribute to heightened awareness of language patterns, vocabulary expansion and retention (Granger, 1998).

Table 3. #LancsBox 6.0 interface.

Source: #LancsBox 6.0.

6.1 Words tool

First, the 'Words' tool (in the black bar in Table 3) was used to generate the lists of the most frequent words in the corpus COREL-GEO (Table 4) and COREL-SCI (Table 5).

Table 4. Most frequent words in COREL-GEO.

▼ Corpus COREL_GEO ▼ Frequency ▼ Dispersion ▼ Type

Type	▼ Frequency: 01 - Freq	Dispersion: 01_CV
the	6375.000000	0.447758
of	2569.000000	0.650346
and	2453.000000	0.677885
is	2046.000000	0.747554
a	2034.000000	0.871552
to	1859.000000	0.800018
in	1710.000000	0.752914
are	1427.000000	0.957705
it	887.000000	1.131860
that	800.000000	1.051627
you	669.000000	1.799714
on	592.000000	1.396621
they	589.000000	1.549366
can	549.000000	1.629760
water	514.000000	2.026305
for	505.000000	1.509535
people	491.000000	1.674919
from	470.000000	1.451002
this	467.000000	1.509161
or	437.000000	1.771113

Source: #LancsBox 6.0.

Table 5. Most frequent words in COREL-SCI.

▼ Corpus COREL_SCI ▼ Frequency ▼ Dispersion ▼ Type

Type	▼ Frequency: 01 - Freq	Dispersion: 01_CV
the	6157.000000	0.548700
and	2465.000000	0.658276
of	2411.000000	0.757319
a	2326.000000	0.859471
is	1954.000000	0.868201
to	1680.000000	0.780321
are	1618.000000	1.086940
in	1617.000000	0.939618
that	996.000000	1.152942
they	880.000000	1.431558
it	840.000000	1.274232
water	761.000000	2.064976
animals	713.000000	1.891993
can	680.000000	1.553853
have	628.000000	1.692628
you	567.000000	1.940654
from	544.000000	1.650789
plants	536.000000	2.124250
or	529.000000	1.704293
what	516.000000	1.547638
on	474.000000	1.580419

Source: #LancsBox 6.0.

At a glance it is possible to see that the results are somewhat similar as most frequent words are function words. The content words, though, which constitute the core vocabulary of each subject come up somewhat differently (e.g., people, plants). Therefore, it is necessary to make some changes in the header to obtain the list of most frequent nouns and verbs to get the corpora content differences. The same procedures can be used for adverbs, adjectives and other word classes of interest. First, one should left-click 'Type' at the top blue header (Table 5), changing Type to Lemma by clicking the arrow. After that, right-click on the black bar, next to the word Type, and a pop-up window will open. Add: *_v, or *_n, or *_adj or *_adv to have the most frequent words of the different word classes, one at a time.

The resulting lists (Tables 6 – 9) show nouns and verbs, KWIC options, which need to be selected by the teachers according to the relevance of the vocabulary in the syllabus and the connection with the lessons' contents. The data should be mediated by the teacher (McCarthy's, 2004), so learners can read and handle the concordance lines without difficulties to discover language patterns (Johns, 1990; Schmidt, 1990) and notice their meaning (Rutherford, 1987).

Table 6. Most frequent nouns in COREL-GEO.

▼ Corpus COREL_GEO	▼ Frequency	▼ Dispersion	▼ Lemma
Lemma	▼ Frequency: 01 - Freq	Dispersion: 01_CV	
ocean_n	546.000000	3.083121	
water_n	527.000000	2.005631	
people_n	490.000000	1.674755	
animal_n	409.000000	2.465666	
map_n	368.000000	2.172981	
earth_n	348.000000	2.360590	
river_n	309.000000	2.857697	
place_n	290.000000	2.022076	
plant_n	279.000000	2.520589	
land_n	275.000000	2.121670	
world_n	253.000000	3.312790	
planet_n	225.000000	4.992227	
city_n	218.000000	3.037777	
area_n	205.000000	2.274882	
forest_n	202.000000	3.905457	

Source: #LancsBox 6.0.

Table 7. Most frequent verbs in COREL-GEO.

▼ Corpus	COREL_GEO	▼ Frequency	▼ Dispersion	▼ Lemma
Lemma		▼ Frequency: 01 - Freq	Dispersion: 01_CV	
be_v		4139.000000	0.482737	
have_v		638.000000	1.442349	
can_v		574.000000	1.621004	
live_v		414.000000	1.680417	
do_v		365.000000	1.800258	
make_v		287.000000	2.052865	
use_v		282.000000	1.944986	
find_v		271.000000	1.932996	
call_v		264.000000	2.021651	
help_v		195.000000	2.766487	
answer_v		160.000000	1.678604	
go_v		157.000000	2.956513	
know_v		152.000000	2.669434	
read_v		151.000000	1.825758	
get_v		142.000000	2.707839	

Source: #LancsBox 6.0.

Table 8. Most frequent nouns in COREL-SCI.

▼ Corpus	COREL_SCI	▼ Frequency	▼ Dispersion	▼ Lemma
Lemma		▼ Frequency: 01 - Freq	Dispersion: 01_CV	
animal_n		883.000000	1.663656	
water_n		750.000000	2.102484	
plant_n		731.000000	1.953777	
food_n		413.000000	2.900088	
part_n		331.000000	2.501983	
energy_n		306.000000	4.698884	
earth_n		289.000000	2.690054	
rock_n		267.000000	4.894477	
thing_n		244.000000	3.167832	
sun_n		237.000000	3.585815	
air_n		232.000000	3.472406	
question_n		196.000000	1.886758	
ocean_n		194.000000	4.574690	
insect_n		193.000000	4.760802	

Source: #LancsBox 6.0.

Table 9. Most frequent verbs in COREL-SCI.

▼ Corpus	COREL_SCI	▼ Frequency	▼ Dispersion	▼ Lemma
Lemma		▼ Frequency: 01 - Freq	Dispersion: 01_CV	
be_v		3988.000000	0.587242	
have_v		852.000000	1.407591	
can_v		722.000000	1.485668	
make_v		431.000000	2.094555	
do_v		422.000000	1.802421	
live_v		404.000000	2.245170	
call_v		399.000000	1.792891	
grow_v		282.000000	2.618893	
use_v		282.000000	2.189419	
eat_v		256.000000	3.496684	
get_v		224.000000	2.428762	
help_v		220.000000	2.433215	
move_v		211.000000	3.738629	
find_v		189.000000	2.666892	

Source: #LancsBox 6.0.

6.2 KWIC tool

The concordance lines are produced using the KWIC tool, the easiest tool to handle of them all. Once teachers select the content word to be dealt with in the classroom, they left-click the KWIC tool on the header and insert it in the appropriate slot on the left next to Search. Just clicking search the lines are yielded instantly showing the KWIC in red at the center. Screenshots of two samples, 'Earth' and 'live' are below in Tables 10 and 11. The concordance lines can be handled in assorted ways in the classroom. They suit learners mainly from 3rd and 4th grades onwards, when most of them are probably already familiar with digital gadgets like tablets and phones. The teacher has at her disposal an array of ways to explore the KWIC tool for pedagogical reasons, for example, she can enlarge, print the list, cut out the lines and distribute to each learner, so they can work out a definition of the planet Earth in small groups; or individual learners can look for other words they already know and create new sentences related to planet Earth. Learners can also look for adjectives on the left side which qualify Earth, and also extract information for the description of Earth, names of planets, and so on. It will depend on the learners' grades and their syllabus. To round off the work, the teacher can show some concordance lines she had previously selected which have meaningful information for the group.

Table 10. Screenshot of KWIC Earth from COREL-GEO.

#LancsBox 6.0

KWIC | GraphColl | Wheik | Words | Ngrams | Text

Corpora | KWIC: earth ✕

Search

Search earth **Occurrences** 349 (39.78) Texts 458 **▼ Corpus** COREL_GEO **▼ Context** 10 **▼ Display**

Index	File	Left	Node	Right
1	Geo - 2r	gases that gives off great amounts of energy. Life on	Earth	depends on light and heat from the Sun. The solar
2	Geo - 2r	and asteroids. They all orbit, or travel around, the Sun.	Earth	orbits the Sun at a distance of about 93 million
3	Geo - 3r	#308646-toc> Water, wind, and other natural forces cause rocks and	earth	to wear away. These forces also move bits of rock
4	Geo - 3r	wear away. These forces also move bits of rock and	earth	to new places. This movement changes the shape of the
5	Geo - 3r	Rivers carry freshwater to people, plants, and animals all across	Earth.	They provide people with a method of transport and water
6	Geo - 3r	or more minerals. Rock makes up the outer layer of	Earth,	called the crust. The lower parts of this layer are
7	Geo - 3r	section links to articles on other celestial bodies. Mercury, Venus,	Earth,	and Mars are rocky planets. They are closer to the
8	Geo - 3r	the closest to the Sun, these planets are Mercury, Venus,	Earth,	Mars, Jupiter, Saturn, Uranus, and Neptune. The solar system is
9	Geo - 3r	gravity. This pull creates tides. As the Sun, Moon, and	Earth	move in space, they sometimes form a straight line, shown
10	Geo - 3r	that are lower than usual. At other times the Sun,	Earth,	and Moon are positioned like the corner of a square,
11	Geo - 3r	low than usual. Along the coasts of every ocean on	Earth	the water level changes on a regular basis. This movement
12	Geo - 3r	volcano erupts, hot gases and melted rock from deep within	Earth	find their way up to the surface. This material may
13	Geo - 3r	com/kids/article/water/390625> Water is the most important liquid on	Earth.	It covers almost 75 percent of Earth's surface in the
14	Geo - 3r	process known as the water cycle. Water is present on	Earth	in three states: gas, liquid, and solid. The amount of
15	Geo - 3r	constantly in these three states. The way water moves around	Earth	is called the water cycle. Water has been cycling since

Source: #LancsBox 6.0.

Table 11. Screenshot of KWIC live from COREL-GEO.

Search live **Occurrences** 315 (35.91) Texts 163/458 **▼ Corpus** COREL_GEO **▼ Conte**

Index	File	Left	Node	Right
1	Geo - 3r	rivers, and lakes. All plants and animals need water to	live.	People have many uses for water besides drinking. They use
2	Geo - 4th	is among the river's many fish. People Millions of Brazilians	live	in cities in the Amazon region. Belém, Manaus, and Santarém
3	Geo - 4th	1990s there were about 600,000 Indians in the Amazon. Most	live	in remote areas. Economy The countries of the Amazon region
4	Geo - 4th	parts of the forest threatens the plants and animals that	live	there.
5	Geo - 4th	to survive in their particular biome. Plants and animals that	live	within smaller areas of a biome also depend on each
6	Geo - 4th	as spruces and firs. Lynx, gray wolves, moose, and beavers	live	in this biome. Deciduous forests Deciduous forests are found mainly
7	Geo - 4th	are hot all year long. Prairie dogs and mule deer	live	in the temperate grasslands of North America. Giraffes, zebras, and
8	Geo - 4th	the temperate grasslands of North America. Giraffes, zebras, and lions	live	in the tropical grasslands of Africa. Deserts Deserts are the
9	Geo - 4th	Dungeness crab, and... krill krill Krill are shrimplike animals that	live	in the ocean. They belong to the group of animals
10	Geo - 4th	is the only planet we know of that humans can	live	on. Everything we need to survive exists on or near
11	Geo - 4th	Sun, the Earth would be too cold for us to	live,	and plants wouldn't be able to absorb the light from
12	Geo 3rd	very little rain each year. Very few plants or animals	live	in desert areas. Most deserts get less than 10 inches
13	Geo- 2n	many useful items and functions for not only people who	live	on the coast, but all people. The first thing that
14	Geo- 2n	action. The world's oceans are providing water for people who	live	in dry places. One of the most important ways we
15	Geo- 4th	a community of living organisms, such as plants and animals,	live	together in conjunction with nonliving aspects of their environment. Ecosystems
16	Geo- 4th	is a network of plants, animals and other organisms that	live	in an area of land. They have less water available
17	Geo_ 3r	been found in every ocean of the world. Since whales	live	in the ocean they are marine animals. They swim in
18	Geo_ 3r	oil for lamps and other things. The blue whale can	live	up to ninety years. Interesting facts: Blue whale calves can
19	Geo_ 3r	They also eat fruit as well as tree bark. Capybaras	live	in herds. The females take care of each other's babies.
20	Geo_ 3r	are reptiles and are part of the lizard family. They	live	on every continent except Antarctica. Their habitat can range from

Source: #LancsBox 6.0.

Concordance line #19 in Table 11 above has the word Capybaras, which is well known by Brazilians and could be explored in a class about animals and their behavior. The most meaningful advantage of using concordance lines from a pedagogic corpus is that the language is not only authentic but also at the appropriate level for learners. The concordance lines in Table 12 below have information about the Amazon biome and they can be worked with to prompt a discussion about the Amazon region with the 3rd and 4th graders, for example. The learners can look for the meaning of biome, very similar to Portuguese, and guess a definition. Other words scattered in the lines can also be made salient (Rutherford, 1987) with students looking for their definitions. Or else, they can speculate about the different animals which live in different biomes.

Table 12. Excerpt with the top five concordance lines about the Amazon biome.

Search live **Occurrences** 315 (35.91) Texts 163/458 **▼ Corpus** COREL_GEO

Index	File	Left	Node	
1	Geo - 3r	rivers, and lakes. All plants and animals need water to	live.	People have many uses for water besides drinking. They use
2	Geo - 4th	is among the river's many fish. People Millions of Brazilians	live	in cities in the Amazon region. Belém, Manaus, and Santarém
3	Geo - 4th	1990s there were about 600,000 Indians in the Amazon. Most	live	in remote areas. Economy The countries of the Amazon region
4	Geo - 4th	parts of the forest threatens the plants and animals that	live	there.
5	Geo - 4th	to survive in their particular biome. Plants and animals that	live	within smaller areas of a biome also depend on each

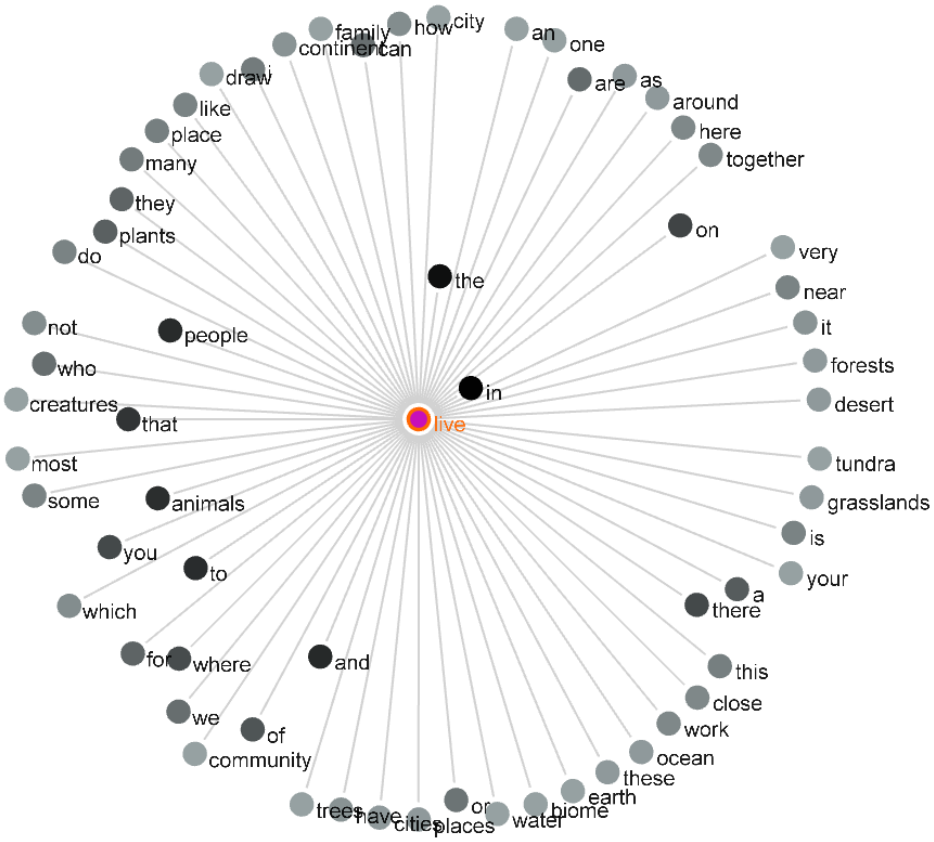
Source: #LancsBox 6.0.

6.3 GraphColl tool

A third tool, GraphColl, usually catches the attention of users in the #LancsBox. It enables them to visualize the collocates of the node chosen and the degree of mutual strength indicated by their positions in the Graph. The stronger the link, the closer to the node the collocate is in the graph.

One example is the graph of 'live' below (Table 13). The visualization helps the user to identify the closest collocates (e.g. 'animals live', 'live in forests') which are listed in Table 14 below. The table also displays the value of the selected association measure in Stats column, while Freq (coll) displays the frequency of the collocation (combination of node + collocate) and Freq (corpus) the frequency of the collocate anywhere in the corpus.

Table 13. Graphcoll of 'live' with the 64 strongest collocations in the corpus COREL-GEO.



Source: #LancsBox 6.0.

The node 'live' was selected and the Span was changed to 3<>3 (Table 14 in the header) words to each side of the node, so that we have only the strongest collocations with it. The R (right) and L (left) indicate the position in relation to the node in the concordance lines.

Table 14. The 20 strongest collocations of 'live' in the corpus COREL-GEO.

▼ Span 3 <> 3 ▼ Statistics 01 -

live

Freq: 315 - Collocates: 64

Index	Status	Position	Collocate	▼ Stat	Freq (coll.)	Freq (corpu
1	○	R	in	160.0	160	1710
2	○	R	the	122.0	122	6375
3	○	L	and	71.0	71	2453
4	○	L	people	68.0	68	491
5	○	L	to	66.0	66	1859
6	○	L	animals	64.0	64	364
7	○	L	that	56.0	56	800
8	○	R	on	40.0	40	592
9	○	R	there	36.0	36	431
10	○	L	you	36.0	36	669

Source: #LancsBox 6.0.

6.4 Ngrams tool

A fourth tool used was Ngrams to generate the word clusters (Table 15). In this example, we chose to have the program generate 3-gram sequences. Since an n-gram is a contiguous sequence of n items that come from a text or a corpus, some of them, though frequent, may not be pedagogically relevant. Even if the clusters deemed most useful for the learners are not the most frequent, teachers should choose at their discretion those more meaningful to their class. In Table 15 the first two more frequent n-grams are used for instruction 'answer the questions' and 'read the text' which are quite often used in class and may not need to be highlighted.

Table 15. Screenshot of the most frequent 3-grams multi-word clusters.

▼ Corpus COREL_GEO+SCI ▼ Frequency ▼ Dispersion ▼ Type

Type	▼ Frequency: 01 - Freq	Dispersion: 01_C
answer the questions	322.000000	1.687115
read the text	264.000000	1.853502
plants and animals	170.000000	3.863660
directions read the	164.000000	2.514339
the text and	159.000000	2.378572
part of the	142.000000	4.618029
then answer the	131.000000	2.809632
look at the	116.000000	3.197654
from the sun	108.000000	6.360234
study the photo	106.000000	2.985693
the photo then	102.000000	3.023573
live in the	100.000000	4.784212
and answer the	94.000000	3.578074
photo then answer	88.000000	3.187753
a lot of	80.000000	4.666478
text and study	78.000000	3.351886
and study the	78.000000	3.351886
text and answer	76.000000	3.677964
one of the	71.000000	5.893860
in the world	67.000000	4.914871
the united states	66.000000	5.898643
of the world	66.000000	5.452285
what is the	65.000000	5.244701
that live in	62.000000	6.306781
is called a	61.000000	6.260574
made up of	59.000000	5.589772
animals that live	58.000000	6.604447

Source: COREL-GEO + COREL-SCI.

Other n-grams, such as, 'the text and' and 'part of the' are phrase fragments that would not be relevant to be taught. However, the prepositional phrases 'in the world' and 'of the world' as well as the verb phrase 'made up of' (Table 16) would be useful for young learners. Groups of words which contain prepositions are usually those to present difficulties in the future when learners are speaking. If they are exposed to them in context, the prepositions can be internalized appropriately from start.

Table 16. 3-gram 'made up of' in the corpus COREL-GEO + COREL-SCI.

Search made up of **Occurrences** 59 (3.31) Texts 41/892 **▼ Corpus** COREL_GEO+SCI

Index	File	Left	Node	
1	Earthar	Fire? The Pacific Ring of Fire is	made up of	plate boundaries. It wraps around the Pacific
2	Geo - 3	Rock, or stone, is a hard material	made up of	one or more minerals. Rock makes up
3	Geo - 3	<https://kids.britannica.com/kids/article/sand-dune/390265> Sand is	made up of	small, loose pieces of rock, soil, minerals,
4	Geo - 3	places for dunes. Most windblown sand is	made up of	the mineral quartz. However, near volcanoes, dunes
5	Geo - 3a	.com/kids/article/the-solar-system-at-a-glance/610256> The solar system is	made up of	the Sun and everything that orbits, or
6	Geo - 4	completely of metallic elements. Because it is	made up of	metal, scientists believe the core is what
7	Geo_1s	might be an airport. A community is	made up of	places to go, things to do, and
8	Geo_3r	is not just one "bowl" It is	made up of	hundreds of similar landforms. The climate of
9	Geo_3r	natural feature of Earth's surface that is	made up of	rock, dirt or minerals. Landforms can be
10	Geo_3r	know is that our solar system is	made up of	one star, the Sun, which several planets
11	Geo_3r	completely of metallic elements. Because it is	made up of	metal, scientists believe the core is what
12	Geo_4t	Rainforest? While the Amazon Rainforest is mainly	made up of	plants and animals, there are some people
13	Geo_4t	atmosphere is all around us and it's	made up of	a layer of gases that surrounds the
14	Geo_4t	in the northern hemisphere, and is mainly	made up of	coniferous forests. The taiga covers large amounts
15	Geo_4t	small–some organisms that live there are only	made up of	a single cell. That's pretty small! The
16	Geo_5t	many things in common. Each market is	made up of	many individual shops. Shops sell a variety
17	Geo_M	place to place, but all soil is	made up of	different amounts of three types of particles:
18	Geo_M	are the layers of soil? Soil is	made up of	distinct horizontal layers. If you could take
19	Geogra	mountains that are mostly underwater. The is	made up of	islands. Read the words in the box
20	Geogra	hot places, too. A rain forest is	made up of	four layers. Each layer is home to
21	LifeSci_	backbone isn't a single bone. It is	made up of	many small bones. These bones are called
22	LifeSci_	Structures flower fruit leaf stem Plants are	made up of	several different structures. The roots of a
23	Sci_4th	the data transfer process. A transmitter is	made up of	a collection of electronic circuits. The transmitter's
24	Sci_4th	coming. Both human and animal bodies are	made up of	many sensory receptors. A sensory receptor is
25	Sci_4th	to use. Layers of Soil Soil is	made up of	many things, including humus and tiny pieces

Source: COREL - GEO+SCI.

Young learners will certainly benefit from handling the lines, distributed in small batches to each one, to identify what comes after 'made up o', for example. Is it always a noun? Can they categorize what kind of noun it is? Once they identify the noun, learners can create lists under categories such: abstract, concrete, related to people, related to things, and so on. Once done, they can create sentences using elements in the classroom or in their backpacks, or even in their homes.

7. Comparison between BNC2014-baby corpora and COREL_GEO+SCI

Many researchers admit that adult corpora are too difficult for young learners to use (Anthony, 2007) or, in the author's view, most of their contents are above the language level of young learners and also more diversified in relation to topics. This would probably yield query results that are not related to learners` questions. Table 17 below shows the difference between BCN2014-baby contents, one of the smallest adult corpus already embedded in #LancsBox 6.0, and the pedagogic corpus COREL-GEO+SCI. Even though BCN2014-baby has fewer files, the mean number of words indicate the texts are longer and probably more complex.

Table 17. Comparison between two corpora.

Comparison of corpora	Total number of files	Total number of tokens	Total number of types	Total number of lemmas	Mean number of words
COREL_GEO+SCI	892	178,384	8,873	8,187	199,9
BCN2014-baby	13	5,024,072	127,122	122,349	386,467

Source: #LancsBox 6.0.

In Table 18, the difference in the lists of most frequent words outputted from both corpora is clear. Although some words are present in both lists, such as function words (the, of, to, *etc.*), some are not the same. This is an indication of the wider scope of topics in BNC2014-baby, yielding a general frequency list which does not present any content words (nouns, verbs or adjectives) among the 23 most frequent words in the corpus. On the other hand, the COREL-GEO-SCI frequent word list carries nouns such as 'animals' and 'plants'.

Table 18. Comparison of most frequent words in both corpus.

▼ Corpus	BNC2014-baby	▼ Frequency
Type	▼ Frequency: 01 - Freq	
the	240869.000000	0.248102
and	131776.000000	0.192305
to	124487.000000	0.127398
of	114059.000000	0.416149
a	113610.000000	0.149034
in	80631.000000	0.242005
i	75495.000000	0.690811
it	62425.000000	0.474114
that	54228.000000	0.319127
you	51306.000000	0.805737
is	47136.000000	0.239216
was	44964.000000	0.499732
for	43508.000000	0.259723
on	37469.000000	0.201309
with	32413.000000	0.172421
he	31810.000000	0.915291
as	31011.000000	0.357391
but	29947.000000	0.362874
be	27633.000000	0.335723
have	26325.000000	0.359275
this	25982.000000	0.429803
at	25840.000000	0.253288
yeah	22958.000000	2.824525

▼ Corpus	COREL_GEO+SCI	▼ Freque
Type	▼ Frequency: 01 - Freq	
the	12520.000000	0.497510
of	4974.000000	0.704931
and	4910.000000	0.668675
a	4358.000000	0.864055
is	3999.000000	0.802870
to	3534.000000	0.793004
in	3322.000000	0.844992
are	3042.000000	1.027288
that	1795.000000	1.136758
it	1724.000000	1.201873
they	1469.000000	1.496569
water	1275.000000	2.097321
you	1231.000000	1.880419
can	1229.000000	1.597600
animals	1077.000000	2.258955
on	1063.000000	1.487363
from	1011.000000	1.570400
have	1000.000000	1.849618
for	974.000000	1.592980
or	964.000000	1.740909
this	889.000000	1.610406
what	871.000000	1.650278
as	813.000000	1.808480
plants	782.000000	2.445035

Source: #LancsBox 6.0.

In Table 19, the difference between the two lists of most common nouns is considerable as the COREL-GEO+SCI has a specific focus on Geography and Science lexis for elementary grades, making this corpus much more appropriate and meaningful to the young learners' goals in the classroom. The situation changes when we analyze the most common verbs in both corpora as there are many similar verbs in both lists (Table 20). Once teachers are more skilled with the tools, they can resort to comparing different corpora lists and enhance the contents of their lessons. After all, learners will probably be exposed to all of them in their general English classes.

Table 19. Screenshot of most frequent nouns in both corpora.

▼ Corpus	BNC2014-baby	▼ Frequency
Lemma	▼ Frequency: 01 - Freq	
time_n	10899.000000	0.171048
mm._n	9361.000000	3.411926
name_n	8432.000000	1.952388
year_n	8157.000000	0.453016
people_n	6877.000000	0.439037
thing_n	6770.000000	0.576718
day_n	6482.000000	0.475297
i_n	6156.000000	1.415486
way_n	5957.000000	0.241363
place_n	5035.000000	0.766334
something_n	4197.000000	0.621782
bit_n	4154.000000	0.901218
man_n	3943.000000	0.746937
life_n	3690.000000	0.363189
cos_n	3589.000000	2.935227
er_n	3388.000000	3.219952
week_n	3371.000000	0.765329
world_n	3327.000000	0.612727
erm_n	3326.000000	3.422586
work_n	3235.000000	0.549822
book_n	3173.000000	1.808641
lot_n	3045.000000	0.640578
home_n	2952.000000	0.480652
family_n	2895.000000	0.745080

▼ Corpus	COREL_GEO+SCI	
Lemma	▼ Frequency: 01 - Freq	
animal_n	1292.000000	2.063896
water_n	1277.000000	2.105728
plant_n	1010.000000	2.313096
ocean_n	740.000000	3.673924
earth_n	637.000000	2.503452
people_n	615.000000	2.188950
food_n	523.000000	3.679890
part_n	480.000000	2.724812
place_n	445.000000	2.458987
rock_n	422.000000	4.893841
thing_n	415.000000	3.101536
question_n	388.000000	1.767219
map_n	385.000000	3.152898
energy_n	385.000000	5.654122
planet_n	374.000000	5.660537
land_n	374.000000	2.614013
river_n	362.000000	3.746060
tree_n	360.000000	3.592964
sun_n	355.000000	4.233658
world_n	342.000000	4.143043
type_n	318.000000	3.175753
life_n	314.000000	3.527561
forest_n	302.000000	4.956465
area_n	301.000000	2.728337

Source: #LancsBox 6.0.

Table 20. Screenshot of most frequent verbs in both corpora.

▼ Corpus	COREL_GEO+SCI	
Lemma	Frequency: 01 -...	
be_v	8122.000000	0.532330
have_v	1489.000000	1.452832
can_v	1296.000000	1.555019
read_v	304.000000	1.751349
do_v	784.000000	1.819012
call_v	663.000000	1.900958
answer_v	248.000000	1.993854
live_v	818.000000	1.999962
use_v	564.000000	2.055708
make_v	718.000000	2.134761
find_v	460.000000	2.219331
get_v	365.000000	2.569601
help_v	415.000000	2.602819
look_v	291.000000	2.607050
know_v	275.000000	2.796530
will_v	258.000000	3.099543
take_v	227.000000	3.101444
grow_v	410.000000	3.267830
go_v	261.000000	3.284520
see_v	289.000000	3.302802
show_v	178.000000	3.305372
need_v	272.000000	3.418124
would_v	190.000000	3.454784
come_v	214.000000	3.487196

▼ Corpus	BNC2014-baby	▼ Freq
Lemma	▼ Frequency: 01 - Freq	
be_v	171302.000000	0.115794
have_v	56943.000000	0.231063
do_v	28439.000000	0.636113
say_v	21436.000000	0.679890
get_v	20109.000000	0.707861
go_v	17721.000000	0.638703
know_v	15200.000000	0.867897
think_v	13909.000000	0.798926
can_v	13081.000000	0.370849
would_v	11067.000000	0.557739
see_v	10847.000000	0.360432
make_v	10817.000000	0.240892
will_v	10680.000000	0.496453
come_v	8837.000000	0.477019
take_v	8606.000000	0.259778
could_v	7587.000000	0.448263
look_v	7430.000000	0.505924
want_v	7107.000000	0.439542
use_v	6639.000000	0.715412
give_v	5567.000000	0.242084
like_v	5142.000000	0.958033
need_v	5018.000000	0.470003
find_v	4967.000000	0.275258
tell_v	4650.000000	0.523260

Source: #LancsBox 6.0.

Remarks and conclusion

This paper has intended to raise readers' awareness of the meaningful use of pedagogic corpora and #LancsBox, a freely available concordancer, to expose young learners to authentic subject-informed English in the classrooms. It described the steps into getting started with user-friendly concordancer tools to access language from the corpora, exposing learners to one of the 21st century digital media that addresses language learning challenges.

It also showed the relevance of creating a pedagogic corpus targeted at a specific discourse community of young learners. The language output presented in Section 6 should encourage teachers to work with authentic level-appropriate subject-specific subsets of corpora that can be used in materials design (McCarthy, 2004). To succeed, Jablonkai (2022, p.474) recommends the involvement of "subject- specialist informants in the corpus building process especially for future pedagogically motivated specialized corpora" to inform teaching on a wider scale.

Until that moment in the future, this paper has been an attempt to motivate teachers to start getting acquainted with the web environment as much as they do in their daily lives using the mobile devices for different purposes. The digital tools and websites available can make

education more inclusive while targeting at much broader types of student populations, helping students with learning difficulties, etc. (Meunier, 2020). However, with content of various subjects being taught in English to young learners, the use of pedagogic corpora with the support of specific software, such as #LancsBox, can speed up both content and language learning.

In times of great cultural awareness and the need for social equity and inclusion in regular schools, English and digital literacy can be the key elements to help reduce the educational differences our youngsters experience while attending the primary school years. This empowerment in the use of digital technologies can "enhance accessibility and inclusion, differentiation and personalisation, and learners' active engagement (Meunier, 2022, p.350). A better prepared and skilled young adult will certainly stand a better chance to succeed and be prepared to seek further opportunities in the future.

References

#LancsBox - Version 6.0. **Lancaster University corpus toolbox**. Available online at #LancsBox: Lancaster University corpus toolbox. Last accessed: Jan 31 st 2023.

CHUJO, Kiyomi; ANTHONY, Lawrence; OGHIGIAN, Kathryn. **DDL for the EFL classroom - Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer.** Tokyo: Waseda University, 2009.

ASTON, Guy. Small and large corpora in language learning. In: LEWANDOWSKA-TOMASZCZYK, Barbara & MELIA, Patrick (Orgs.) **PALC '97 Proceedings of the first annual conference.** Łodz: Łodz University Press, 1997. p. 51-62.

BASE NACIONAL COMUM CURRICULAR (BNCC). **Ministério da Educação e Cultura**. Available online at http://basenacionalcomum.mec.gov.br. Last accessed: Feb 23 rd 2023.

BREZINA, Vaclav; GABLASOVA, Dana**. #LancsBox**. Lancaster: Lancaster University, 2018.

CHAMBERS, Angela. Towards the corpus revolution? Bridging the research – practice gap. **Language Teaching**, Cambridge, *v.*52, n.04, p. 460-475, 2019.

CHAMBERS, Angela. What is data-driven learning? In: O`KEEFFE, Anne & MCCARTHY, Michael (Eds.) **The Routledge handbook of Corpus Lingusitics**. London: Routledge, 2010. p. 345- 358.

COBB, Thomas. Is there any measurable learning from hands-on concordancing ? **System**, Elsevier, v.3, n.25, p. 301-315, 1997.

CORTES, Viviana. Lexical bundles in published and student disciplinary writing: examples from History and Biology. **English for Specific purposes**, v.23, n.04, p. 397 – 423, 2004.

CROSTHWAITE, Peter; STELL, Annita. It helps me get ideas on how to use my words - Primary school students' initial reactions to corpus use in a private tutoring setting. In: CROSTHWAITE, Peter (Ed.) **Data-Driven Learning for the Next Generation - Corpora and DDL for Pre-tertiary Learners.** London: Routledge, 2020. Kindle Edition, Kindle Locations: 3837-3838.

ELLIS, Nick. Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear. **Annual Review of Applied Linguistics**, Cambridge, v. 32, p. 17–44, 2012.

ELLIS, Nick. Frequency effects in language processing a review with implications for theories of implicit and explicit language acquisition. **Studies in Second Language Acquisition**, v.24, p. 143–188, 2002.

FRANKENBERG-GARCIA, Ana. Raising teachers` awareness of corpora. **Language Teaching**, v.45, n.04, p. 475-489, 2012.
FRIGINAL, Eric. **Corpus linguistics for English teachers: New tools, online resources, and classroom activities.** London: Routledge, 2018.

GABRIELATOS, Costas. Corpora and Language Teaching: Just a fling or wedding bells? **TESL – EJ**, v.8, n.04, p. 1-37, 2005.

GRANGER, Sylviane. The computer learner corpus: a versatile new source of data for SLA research. In: GRANGER, Sylviane (Org.) **Learner English on Computer.** London: Addison Wesley Longman, 1998. p. 3-18.

GILQUIN, Gaetanelle; GRANGER, Sylviane. How can data-driven learning be used in language teaching. In: O'KEEFFE, Anne & MCCARTHY, Michael (Orgs.) **The Routledge Handbook of Corpus Linguistics**. London: Routledge, 2010. p. 359-371.

HAFNER, Christoph; CANDLIN, Christopher. Corpus tools as an affordance to learning in professional legal education. **Journal of English for Academic Purposes**, v.6, n.04, p. 303–318, 2007. Available online at https://doi.org/10.1016/j.jeap.2007.09.005. Last accessed: Feb 23 rd 2023.

HENDRY, Clinton; SHEEPY, Emily. Evaluating corpus analysis tools for the classroom. In: JABLONKAI, Reka &. CSOMAY, Eniko (Orgs.) **The Routledge Handbook of Corpora and English Language Teaching and Learning.** London: Routledge, 2022. p. 437 – 459.

JABLONKAI, Reka. Building Corpora for ELT. In: JABLONKAI, Reka & CSOMAY, Eniko (Orgs.) **The Routledge Handbook of Corpora and English Language Teaching and Learning**. London: Routledge, 2022. p. 460-477.

JOHNS, Tim. Should you be persuaded: Two examples of data driven learning. In: JOHNS, Tim & KING, Philip (Orgs.) Classroom concordancing. **ELR Journal,** Birmingham v.4, p. 1 – 16, 1991.

JOHNS, Tim. From printout to handout: Grammar and vocabulary teaching in the context of data driven learning. **CALL Austria**, v.10, p. 14 – 34, 1990.

MARINOVA-TODD, Stefka; MARSHALL, Bradford; SNOW, Catherine. Three Misconceptions about Age and L2 Learning. **TESOL Quarterly**, v. 34, n.01, p. 9 – 34, 2000.

MCCARTEN, Jeanne. ***Teaching vocabulary. Lessons from the Corpus, Lessons for the Classroom.*** Cambridge: Cambridge University Press, 2007.

MCCARTHY, Michael. **Touchstone: from Corpus to Coursebook**. Cambridge: Cambridge University Press, 2004.

MEUNIER, Fanny. Revamping DDL: Affordances of Digital Technology. In: JABLONKAI, Reka & CSOMAY, Eniko (Orgs.) **The Routledge Handbook of Corpora and English Language Teaching and Learning.** London: Routledge, 2022. p. 344-360.

MEUNIER, Fanny. A case for constructive alignment in DDL - Rethinking outcomes, practices, and assessment in (data-driven) language learning. In: CROSTHWAITE, Peter (Org.) **Data-Driven Learning for the Next Generation**. New York: Routledge, 2020. Kindle Edition, p. 757-759.

NATION, Paul. What matters in vocabulary learning? **LALS**, 2020, Victoria University of Wellington, New Zealand. Webinar.

O'KEEFFE, Anne; MARK, Geraldine. Principled pattern curation to guide data-driven learning design**.** ***Applied Corpus Linguistics***. Available online at https://www.sciencedirect.com/science/article/pii/S2666799122000132. Last accessed: Jan 31 st 2023.

O'KEEFFE, Anne; MCCARTHY, Michael; CARTER, Ronald. **From corpus to classroom: language use and language teaching.** Cambridge: Cambridge University Press, 2007.

PÉREZ-PAREDE, Pascual. The pedagogic advantage of teenage corpora for secondary school learners. In: CROSTHWAITE, Peter (Org.) **Data driven learning for the next generation: Corpora and DDL for pre-tertiary learners**. London: Routledge, 2020. p. 67–87.

REDECKER, Christine. European framework for the digital competence of educators: DigCompEdu. In: PUNIE, Yves (Org.) **EUR 28775 EN.** Publications Office of the European Union, JRC107466, 2017. Available online at https://doi.org/10.2760/159770. Last accessed: Jan 31 st 2023.

REPPEN, Randi. Building a Corpus – What are the key considerations? In: O'KEEFFE, Anne & MCCARTHY, Michael (Org.) **Routledge Handbook of Corpus Linguistics**. London: Routledge, 2010. p. 31-37.

RUTHERFORD, William. **Second language grammar: Learning and teaching**. London: Longman, 1987.

SCHIMDT, Richard. The role of consciousness in second language learning**. Applied Linguistics**, v.11, no.2, p. 129–158, 1990.

SINCLAIR, John. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

XIAO, Richard. Corpus creation. In: INDURKHYA, Nitin & DAMERAU, Fred (Orgs.) **The Handbook of Natural Language Processing**. London: Taylor and Francis, 2010. p. 147–165.

WEBB, Stuart and CHANG, Anna. Vocabulary learning through assisted and unassisted repeated reading. **Canadian Modern Language Review**, v.68, n.03, p. 1–24, 2012.

WILLIS, Dave. The language syllabus: building language study into a task-based approach. In: **Classroom Matters**, v. 30, Spring 2011.< http://ihjournal.com/the-language-syllabus-building-languag-into-a-task-based-approach-by-dave-willis-2e-study > Acessed in April, 2021.

WILLIS, Jane. Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In: TOMLINSON, Brian (Org.) **Materials development in language teaching**. Cambridge: Cambridge University Press, 1998. p. 51–77.