frontiers in Education

METHODS
published: 19 February 2019
doi: 10.3389/feduc.2019.00007

Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning

Elisa Corino and Cristina Onesti*

Dipartimento di Lingue e Letterature Straniere e Culture Moderne, Università degli Studi di Torino, Turin, Italy

When dealing with Language for Specific Purposes (LSP), teachers always have to confront with issues which are strictly linked to the specificities of the language of a given field. This is not only the case of language teachers, but it is also and particularly true for CLIL teachers in Italy, who are subject teachers sharing with language teachers some aspects of pupils' language education. This paper is grounded in the discussion of Data-driven Learning (DDL) as a scaffolding method to support the language aspects of CLIL, and the role of data-driven materials in enhancing learning in general. Corpus-based methodology in CLIL classes (the LSP learning environment par excellence) means to empower both teachers and students to develop competences in moving away from mere surface features of text to selecting and understanding meanings and structures. In doing so, they use texts with specific intentions, becoming familiar with tools such as corpora to compensate, for instance, the deficiencies of general dictionaries. Over 25 years ago Tim Johns advocated the learning-centered value of DDL, calling "every student a Sherlock Holmes." In fact, DDL good practices perfectly align with current theories and practices of SLA, namely the constructivist and learner-centered approaches to language acquisition. It underpins the mandate in contemporary communicative language instruction for the use of authentic language materials and for the development of metalinguistic knowledge and learner autonomy. The uses and benefits of corpora for language learning are widely reported in the literature, although there is still little field practice, in Italy at least. Our word wants to suggest possible good and effective practices combining CLIL needs and the DDL approach to language and content. A case study where DDL was successfully used in a vocational context will be presented. The reference corpus was based on oral and written productions by English native speakers, elicited from a picture specifically sketched for the activity (a hairdresser's salon with multiple actions and objects). Both lexical items and syntactic structures were extracted by students who were confronted with the data and had to deal with different tasks for their analysis. Results were encouraging and students who were exposed to DDL engaged in an involving activity that considerably improved their language skills in their actual working practice.

Keywords: data-driven learning, CLIL, corpus, LSP, second/foreign language teaching, vocational school

OPEN ACCESS

Edited by:
Marco Mezzadri,
University of Parma, Italy

Reviewed by:
Michele Daloiso,
Independent Researcher, Venice, Italy
Antonios Ventouris,
Aristotle University of Thessaloniki, Greece

****Correspondence:***
Elisa Corino
elisa.corino@unito.it

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *12 March 2018*
Accepted: *22 January 2019*
Published: *19 February 2019*

Citation:
Corino E and Onesti C (2019) Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning. Front. Educ. 4:7. doi: 10.3389/feduc.2019.00007

INTRODUCTION

This paper is grounded in the discussion of DDL as a scaffolding method to support the language aspects of CLIL, and the role of data-driven materials in enhancing language learning in general.

Some case studies where DDL was successfully used in CLIL classes of physics and in a vocational context will be presented. Both lexical items and syntactic structures were extracted by students who were confronted with the data and had to deal with different tasks for their analysis. Results were encouraging and students who were exposed to DDL engaged in an involving activity that considerably improved their language skills in their actual working practice.

STATE OF THE ART

Over 25 years ago Johns (1991), the "father" of Data-driven Learning (DDL), advocated the learning-centered value of DDL, calling "every student a Sherlock Holmes."

The main idea behind such image is that learners can discover grammatical patterns, word meanings or other aspects of language through searching linguistic data and investigating large amounts of authentic language.

Corpus-based language teaching has been praised as a revolution in teaching by Sinclair (2004a), although a more balanced discussion has followed (see e.g., Kaltenböck and Mehlmauer-Larcher, 2005 examining also some limits of DDL).

Since then literature on uses and benefits of corpora for language learning has rapidly grown (see among the most recent research papers Boulton and Lenko-Szymanska, 2015; Ackerley, 2017), although there is still little field practice, in Italy at least. McEnery and Xiao also argue that "while indirect uses of corpora seem to be well established, direct uses of corpora in teaching are largely confined to advanced levels such as higher education" (2010, p. 374). They also add that "corpus-based learning activities are nearly absent in general Teaching English as a Foreign Language (TEFL) classes at lower levels such as secondary education" (McEnery and Xiao, 2010, p. 374), and we might add that it is definitely absent in that teaching practices of other foreign languages.

DDL good practices align with current theories and practices of Second Language Acquisition, namely the constructivist and learner-centered approaches to language acquisition. It underpins the mandate in contemporary communicative language instruction for the use of authentic language materials and for the development in learners of metalinguistic knowledge and learner autonomy (Godwin-Jones, 2017).

The DDL approach in teaching vocabulary and grammar leads indeed to a relevant consciousness-raising of the learners, drawing the student's attention to the formal properties of the target language.

Students have to do with a "massive but controlled exposure to authentic input," so fundamental for language learning (Cobb and Boulton, 2015) and such controlled and contextualized contact fosters more language awareness, noticing and autonomy. Corpora do not only contain the answer to many possible questions: the investigation itself leads to better metalinguistic competencies. A corpus-based bottom-up approach can foster LPS competence of both content teachers and students, by offering facts of actual language usage which are hard to come by with other means (Mindt, 1997; Gavioli, 2005; Hüttner et al., 2009; Walker, 2011), especially with regard to typical choice of words (sorting them by frequency), meaning nuances and appropriate use of collocations.

Furthermore, if students have to struggle to decode and solve linguistic problems, they activate HOTS (higher order thinking skills) processes that will probably result into a longer-lasting knowledge and better language skills.

Following the path of other contemporary approaches to language teaching and learning, also DDL advocates a learning environment where the teacher is no longer the only authoritative owner of knowledge, but rather a "consultant, guide, coach, and/or facilitator" (Suan Chong, 2016). As corpora are often made up of data by native speakers, the teacher's assistance as a guide is certainly essential to give focused tips to the class and to lead students through the data discovery and interpretation. However, instead of transmitting information to the class explicitly and directly, they work as research directors and collaborators, not requiring information teachers already have, but leaving enough space for searching for a solution or a meaning. "In this framework, the teacher acts as a learning expert rather than a language expert" (Bernardini, 2004, p. 28).

Tim Johns argues that "at the heart of the approach is the use of the machine not as a surrogate teacher or tutor, but as a rather special type of informant" (1991, p. 1). Once the informant answered the question, students have to make an effort in order to "make sense of that response (possibly asking other questions in order to do so) and to integrate it with what is already known" (Johns, 1991). Corpora provide data, but do not interpret them: it is up to learners' work and responsibility to evaluate the information found.

It has to be underlined, though, that learners themselves could find Data-driven Learning more demanding and less comfortable than other traditional approaches they are used to.

Such discomfort has to be taken into account together with teachers' attitudes toward using corpus-based materials in class. Teachers are often reluctant to apply DDL: sometimes simply for lack of awareness, not being DDL present in their initial training; in other cases, because they consider it as a research activity confined to higher education.

Many authors mention the application of corpus-based materials in class as a marginal practice. Beyond the above mentioned trends, an important aspect in dealing with DDL is sufficient training which is essential not only for the learners but also for their teacher: it requires considerable investment in terms of time and practice in order to comprehend the rationale and how to use data efficiently (Meunier, 2011; Boulton, 2012).

Additionally, working with corpora in class may prove beneficial for some skills and tasks, but not for others.

We should always bear in mind that different learner profiles and individual differences coexist in a group, so data-driven reflection might not be equally appropriate for all students (Cobb and Boulton, 2015, p. 487).

In an encouraging overview, careful attention to students' learning preferences has not to be neglected: the study of Hadley and Charles (2017) stated that "Affectively, learners found semi-hard DDL to be useful but ultimately unattractive, preferring instead to focus on reading for pleasure and enjoying conversations of self-discovery with classmates." In particular, with lower proficiency learners they claim a softening of the DDL approach.

However, the case study we are going to analyze here (see section DDL in Practice) was successfully launched in a secondary school class which had a relatively low level of English and no previous experience of Data-driven Learning.

Such uncertainties cannot deny the connection Boulton sees among Data-driven Learning and already existing approaches:

"It might appeal to those who are keen to return language to a central place in their language class, and rather than expecting teachers to make the conceptual leap toward corpus linguistics, it may help to bring DDL closer to them by highlighting how DDL exploits any number of key concepts in existing approaches—including, but not limited to: authenticity, autonomy, cognitive depth, consciousness-raising, constructivism, context, critical thinking, discovery learning, heuristics, ICT, individualization, induction, learner-centeredness, learning-to-learn, lifelong learning, (meta-)cognition, motivation, noticing, sensitization and transferability" (Boulton, 2016, p. 3).

As a final remark, considering how much our everyday life is nowadays affected by Internet search, DDL could move closer to the learner by highlighting search techniques connected with their previous background: "It seems likely that many learners around the world are already Googling the Internet in ways not entirely dissimilar to DDL, a practice which may be actively encouraged by their teachers while remaining invisible in the DDL research literature" (Boulton, 2012, p. 25).

USES OF CORPORA IN LANGUAGE TEACHING AND LEARNING

Corpora are proving increasingly influential in language teaching as sources of language descriptions. Pedagogical uses of corpora include two main perspectives: indirect applications of corpora, where scholars use data to create teaching materials or reference books, and direct applications, meant as use of corpora by teachers and learners in a hands-on approach, which is the field we are analyzing herein.

The easiest way to explore corpus data is directly via concordancers: end-users may display a list of words with their immediate context—a concordance based on KWIC (keyword in context), whose visualization is able to reveal a massive amount of information about the language: idioms, collocations, fixed phrases, frequency data.

The learner is in charge of interpreting data, as already mentioned, but "the merit of the corpus is simply to enable data to be delivered in a convenient form for the investigator, whatever area of linguistics he or she is concerned with" (Leech, 1997, p. 9).

Let's now consider some concrete applications. As regards grammar, DDL approach can be effective at teaching and learning grammar due to its encouragement to be active learners (see Indra Nugraha et al., 2017) and use inductive strategies for self-discovery of regularities. Data-driven learning includes both deductive and inductive processes; however, the crucial focus has traditionally been looking at examples to induce patterns or rules.

Through four stages of teaching procedures, for example, Indra Nugraha/Miftakh/Wachyudi got very positive feedbacks by a class of midwifery major at a state university, as testified also by learners' comments like these: "I can see the example of grammar use contextually/I can see many more example sentences than in a dictionary" (Indra Nugraha et al., 2017, p. 303).

Even though some learners found that learning the grammar structures via corpus-informed activities was more difficult than learning with a traditional book and they needed help or guidance from the teacher, another experience—described by Yanto and Indra Nugraha (2017)—has underlined the pleasure of perceiving the language as not artificial but alive and up-to-date. Students found it entertaining and exciting to make grammar rule generalization on their own.

As concerns vocabulary instruction, the knowledge of a word goes obviously beyond the knowledge of its dictionary definition: it embraces also knowing the word's part of speech, spelling, morphology, variant meanings, specific uses, collocations, register.

Corpora help students to master different aspects—lexical information, patterns of textualisation, and genre-structuring features or "moves"—which are relevant to the foreign language learner who needs considerable information regarding the appropriateness and acceptability of particular linguistic choices in individual genres. And some pieces of information are not to be found either in paper or in e-dictionaries, whereas more detailed information on lexico-grammatical features—such as syntactical markedness and nuances in meaning of near-synonyms—is possible through the use of corpus linguistics.

In an online forum on corpus linguistics for EFL[1] a teacher complains about the lack of learner's dictionaries in mentioning the fact that the relative pronoun *who* can actually be used when referring to animals, the same way as *that* or *which*, as general dictionaries point out. This is exactly where a corpus can come across with new "real" language details to supplement the lexicogrammar pattern offered by the dictionaries and standard grammars.

A quick look at the occurrences of dog followed by a relative shows that all the options are possible, or better still, *who* is the most frequent among the first 20 results from the BNC.

In order to successfully search a corpus, students should develop querying skills and the ability to read the data and to classify them. Facing language in context is always fundamental: Ackerley argues that a "phrase-focused approach to teaching and learning may lead to more fluent, native-like, or expert production" (2017, p. 197). Thus, despite some starting obstacles, teachers may provide to the class a very useful resource, becoming an opportunity for lifelong learning.

Moving to the use of corpora for language for specific purposes (LSP), we are aware of many textbooks offering an

[1] https://corpling4efl.wordpress.com/tag/skell/

unrealistic idea of how people communicate in a specific field. "Exploratory corpus research may lead the ESP practitioner to new discoveries about the language used in the students' target situations, and hence to changes in the content of syllabi and materials" (Nesi, 2013), providing suggestions for the development of new tasks.

Let us consider a reflection about listing the most frequent words in a specialized field: turning to dedicated corpora may open access to a considerable exposure to authentic input which is simply not possible with normal textbooks. Besides, Leech (2011, cit. in Meunier, 2011) also "warns that frequency counts are least useful when they are based on a general corpus covering the range of the language and are more useful if they are more specific, i.e. differentiated for mode, register, text type or region."

The main problem herein is probably connected with the little sharing of specialized corpora, often not available for the public.

Obviously grammar and vocabulary dimensions cannot be considered separately: thinking of a CLIL-class, being proficient in LSP means to be able to master different linguistic aspects, among which lexicogrammatical features, patterns of textualisation, and genre-structuring features in order for the FL learner to acquire appropriateness and acceptability of particular linguistic choices in specific genres (Corino, 2014). Corpus work and DDL can thus help teachers to find "patterns of specialized phraseology, which are barely mentioned in the general bilingual and monolingual dictionaries used by their students" (Corino, 2014, p. 68).

We cannot forget that in CLIL not only the vocabulary is challenging: tasks become more cognitively demanding, new concepts and language are presented to the learners as well. Academic language learning is not only related to the understanding of content area vocabulary, it includes skills such as classifying, comparing, evaluating, synthesizing, and inferring.

Tools and Resources

Data-driven learning is based on the principle of "cutting out the middle man," a reference to learning language directly from language rather than from mediated resources such as textbooks, grammar, dictionaries, and teachers (Thomas, 2015, p. 18). Learners should learn and acquire language through direct interactions with language data, with the computer enabling the student to investigate and test hypotheses.

Nowadays there are many software products, which can help analyzing corpora. Some of the most well-known and widely used are *WordSmith Tools*, the *Compleat Lexical Tutor*, and the *SketchEngine*. Other freely available software solutions are *NoSketchEngine (and SkELL)*, *AntConc,* and *LancsBox*, just to name a few.

All of them are concordancers, namely tools or pieces of software which search a text corpus and display a list of words with their immediate context. Those words can show how language is used in an authentic environment, i.e., the KWIC concordance in which each occurrence of the chosen word is highlighted within its context.

The main added value of a corpus is his *vertical dimension*, which allows a researcher to make generalities from the recurrences (Sinclair, 2004b). The KWIC search can reveal a huge amount of information about the language such as common collocates, idioms, fixed phrases and collocations along with usage and frequency. Viewing constructions in a concordance "can be especially informative, as learners are able to see both literal and non-literal uses" of a specific word or utterance (Godwin-Jones, 2017: 12). Another important use of the KWIC is to analyze the frequency of occurrence of an expression in a corpus which can "provide guidance on how common that construction is among native speakers" (Godwin-Jones, 2017, p. 12).

Tools for corpus analysis are now user-friendlier than in the past, and even non-linguists can learn to successfully use them. It is not anymore required to get confident with the corpus query language—though it would be desirable—as user interfaces make the query process clear and easy.

Custom software certainly have more features than free tools. Nonetheless, the latter are rather comprehensive and provide users with the essential functions one may need when inquiring into language facts. And this is particularly true in a DDL environment where the inquirers are students and teachers, thus corpora and tools should be fast, efficient, as user-friendly as possible or at most designed specifically with language learners in mind.

If it is true that users today are far more familiar with computer tools, and "digital natives" regularly use search engines for language queries on the web via computers or mobile devices, it is also a fact that a computational approach to language teaching and learning has been confined to some specific languages and contexts.

The main objection teachers always raise when confronted with a DDL practice is that tools are difficult to manage and learning how to use them is time consuming. Furthermore, they are devised mainly for English, whereas other languages are often left out.

These could be valid arguments some 10 years ago, but we are now facing a new phase where tools are within teachers and students' reach and work with a wide set of languages.

Of course, there are "easy" tools and "difficult" tools, just as much as there are "ready-to-use" corpora and "customizable" corpora, which adapt to a wide range of purposes, i.e., students' age and level, or general language vs. LSP.

Among the most recent "ready-to-use" corpora with a teaching/learning purpose, SkELL (Sketch Engine for Language Learning) is certainly a good source of information and a fine resource for DDL. Using a special algorithm to select occurrences from a large multi-billion samples of text, it currently provides good KWIC examples of the word or phrase useful for language learners of English, German, Italian, Czech, and Russian.

Users just have to type the word they want to investigate and the software returns a set of occurrences, the word sketch with PoS relationships to other words, and the synonyms or semantically related words.

James Thomas has recently shared (http://www.versatile.pub/uploads/8/1/6/3/81634112/skell_trump_observation_tasks_for_students_.pdf) a worksheet devised for some guided discovery questions that learners can answer by looking at

SKELL pages. The searches are *trump up, trump card, trump and Donald Trump*.

Although this can be a good starting point for teachers who like to plan consciousness raising or language awareness activities, the heterogeneity of the corpus does not make SkELL suitable for querying specialized languages.

When dealing LSP and CLIL contexts the choice of the tool is essential, and tailor-made corpora can best meet teacher and students' needs.

One the most popular tools is AntConc. Using an accurate concordancer such as AntConc makes it nice and easy for students to extract collocates and frequency lists, which makes it useful if one wants to explore the lexicon of a dataset and define the properties and relationships of a given word. Nonetheless, the corpora generated with AntConc are lemmatized but not PoS tagged, thus excluding the possibility to choose only part of speech to be queried within the corpus. The fact that PoS tagged corpora can be uploaded into AntConc or that it can be used together with the CLAWS tagging software or other taggers is not really a workable solution for school contexts.

If the corpus is not PoS tagged, queries will generate mixed results in the case of homonyms, mixing up—for instance—verbs and nouns which will then have to be disambiguated.

To use DDL as a scaffolding methodology for CLIL and LSP, PoS tagging and advanced query functions are a significantly important added value. Adding filters to the search can in fact help in identifying verb-noun collocations, adjective-noun colligations and sorting out patterns in a more efficient and precise way. Moreover, as Shaw (2011, p. 24) states, "There is a reciprocal relationship between corpora and part of speech knowledge. To use corpora, students must have some knowledge of parts of speech, but corpus use can increase that knowledge as well."

Both the Sketch Engine and LancsBox can serve these functions in many different languages, thus answering as well the need for resources to be applied to languages other than English. The former is now "a classic," used by linguists, lexicographers and important publishers to produce their products, the latter belongs to the new generation of user-friendly tools to manage corpora.

Apart from showing the word sketch and finding the frequency of a word or phrase and its collocates, advanced tools such as Sketch Engine and LancsBox also provide for a Corpus Query Language (CQL) search allowing users to find word classes and complex grammatical patterns.

LancsBox is certainly the most user-friendly and it offers some advantages such as the search for semantic categories (place adverbs, hedges) and the GraphColl tool, which identifies collocations and displays them in a table and as a collocation graph or network. Relationships among words are visualized and can be filtered, thus helping those students who do not have a solid (meta)grammatical background and might find reading a wordsketch difficult.

As for Sketch Engine, the querying possibilities are far more developed, but teachers and students do not need such a professional level of search. Notwithstanding, the possibility to crawl the web to build up a specialized corpus with WebBootCat makes Sketch Engine a valuable resource, especially for LSP and CLIL classes.

DDL IN PRACTICE

As Corino (2014) pointed out, when dealing with LSP, teachers always have to confront with issues which are strictly linked to the specificities of the language of a given field, and the need of compensating the deficiencies of general bilingual and monolingual dictionaries. This is true for both language teachers dealing with LSP and CLIL teachers who are sharing with their colleagues some aspects of pupils' language education.

In the following paragraphs some examples of good corpus-based scaffolding practices in a CLIL physics class and an experience of DDL for LSP in a vocational school will be presented.

The examples in section DDL as Scaffolding for CLIL are the result of a planning activity carried out with CLIL teachers in methodological training courses, whereas the practices presented in section DDL as for LSP (see below for details about educational context, participants, etc.) were designed and implemented as part of an MA final dissertation in Language teaching and learning.

DDL as Scaffolding for CLIL

The main argument for using DDL in CLIL contexts is that language is the access key to content.

Snow (2010) acknowledges the language of science to be "alienating," if not downright annoying, and in fact when teachers adopt that concise and authoritative tone to explain strange-sounding phenomena which young minds could neither see nor fathom, they might transform even the mother tongue into a foreign language. The context thickens when dealing with this "alienating" LSP in a foreign language where the development of a language-aware content education is strictly required.

The Italian physics term *velocità*, for instance, is a case where disambiguation is needed and dictionaries are not conclusive in order to define the difference between its two translations in English, i.e., *speed* and *velocity*.

The Italian bilingual dictionary Ragazzini (2016) gives both options tagging them as (*fis*), but it seems to be no difference between the two English words, which are presented as synonyms. Which is not the case, though.

> Velocità f. 1 (anche fis.) speed; velocity; (velocità di variazione) rate; (ritmo) pace: (fis.) velocità angolare, angular velocity (o speed)

Turning to corpora and making students discover the difference in behavior of the two English terms also means to lead them to define the different inherent qualities of the two quantities.

In this case a corpus of 586,989 tokens was implemented with Sketch Engine (see an example in **Figure 1**) by the teacher. Students had to search for examples of KWIC starting from the

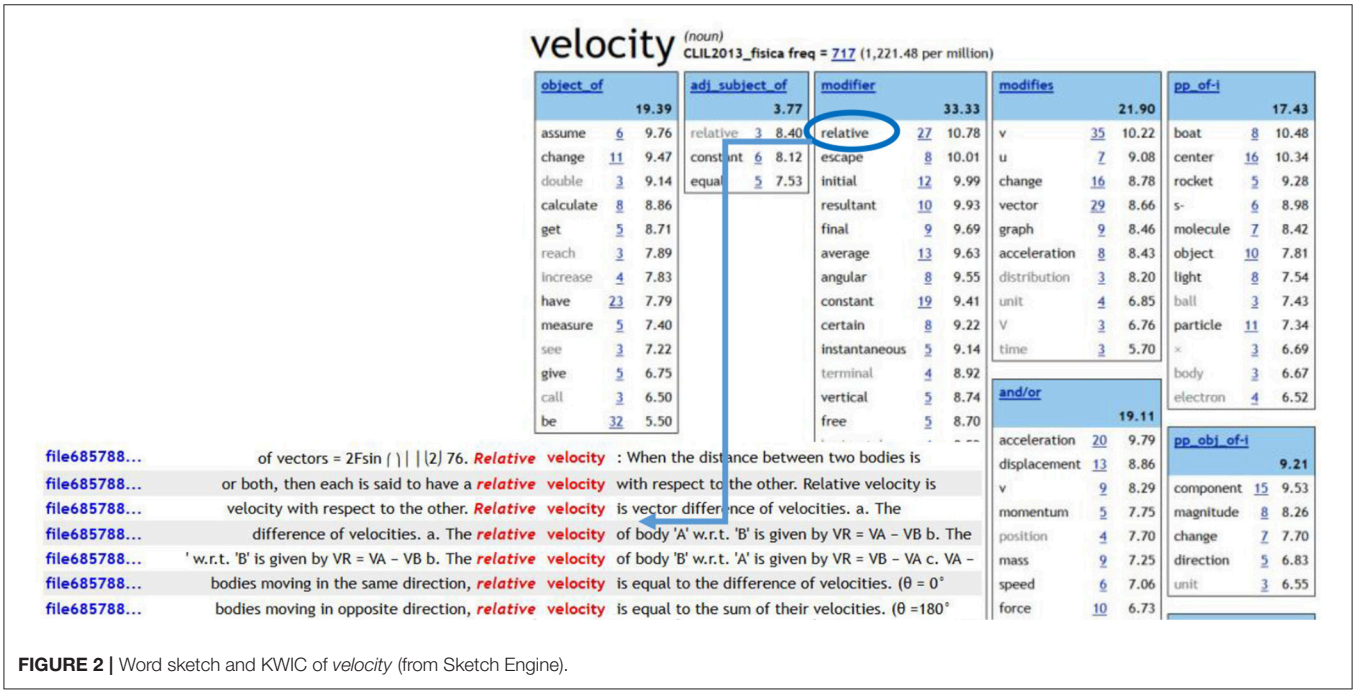
Query dog, IN/that|WP 115 (1.02 per million)

Page 1 of 6 Go Next | Last

Spoken con...	the hounds several dozen hounds off these two	dogs who	were being savaged to death and he had a heart
Written mi...	, John responded to the FUR with the story of his	dog whom	he loves. Anne responded to the QUIET STILLNESS
Written mi...	dog', and complex sentences, like 'This is the	dog that	John likes'. Understanding the second
Written bo...	in a great spurt of scarlet. A guard kicked away a	dog who	tried to run under the scaffold to lick the
Written bo...	as he choked the last struggling life out of the	dog that	he heard another scream, this time from above,
Spoken con...	to the wall. Ah ah no Who are those people back to	dogs that	in North Korea you know they don't just last till
Written bo...	any money from the cooks and porters and pet	dogs who	made up the bulk of their subscribers, quietly
Written mi...	the bundle of clothes hit the ground it was the	dog who	realised first that there was something inside
Written bo...	chin. The wool gave off a friendly smell of the	dogs who	had been lying on it. 'And here's the hot water
Written bo...	PRO Dogs awards dinner LOVE KNOWS NO FEAR PAT	Dogs who	go where others dare not MINDING MASTER LUKE
Written bo...	absence - completely out of character with the	dog that	they have lived with for 13 or so years. TIME TO
Written bo...	of us are familiar with the stories of Pavlov's	dogs who	were conditioned to salivate at the sound of a
Written bo...	kennels is unkind. I advise owners of small	dogs who	have to leave their pets in kennels to find one
Written bo...	(Dec 21-Jan 19) Your influence over another	dog who	is being a bit of a pain in the collar, will bring
Written bo...	in. My initial impression of Sadie was that of a	dog who	was totally depressed. Her eyes were dull and
Written bo...	something. I have yet to come across another	dog who	began life in the same way Sadie did which has now
Written bo...	stems from your pup living with two very gentle	dogs who	have allowed her to become number one in their
Written bo...	the nearest suitable spot then straight home;	dogs who	bounded ahead, to wait for a moment, look round,
Written bo...	; and now, as Jim himself said, he felt like a lame	dog who	had found a wonderful home. He thought it was a
Written bo...	ways, such horses are like certain breeds of	dogs who	will only accept one owner in their lifetime.

Page 1 of 6 Go Next | Last

FIGURE 1 | [lemma = "dog"][tag = "IN/that|WP"]—BNC on Sketch Engine.



velocity (noun) CLIL2013_fisica freq = 717 (1,221.48 per million)

object_of			adj_subject_of			modifier			modifies			pp_of-i		
		19.39			3.77			33.33			21.90			17.43
assume	6	9.76	relative	3	8.40	relative	27	10.78	v	35	10.22	boat	8	10.48
change	11	9.47	constant	6	8.12	escape	8	10.01	u	7	9.08	center	16	10.34
double	3	9.14	equal	5	7.53	initial	12	9.99	change	16	8.78	rocket	5	9.28
calculate	8	8.86				resultant	10	9.93	vector	29	8.66	s-	6	8.98
get	5	8.71				final	9	9.69	graph	9	8.46	molecule	7	8.42
reach	3	7.89				average	13	9.63	acceleration	8	8.43	object	10	7.81
increase	4	7.83				angular	8	9.55	distribution	3	8.20	light	8	7.54
have	23	7.79				constant	19	9.41	unit	4	6.85	ball	3	7.43
measure	5	7.40				certain	8	9.22	V	3	6.76	particle	11	7.34
see	3	7.22				instantaneous	5	9.14	time	3	5.70	×	3	6.69
give	5	6.75				terminal	4	8.92				body	3	6.67
call	3	6.50				vertical	5	8.74	and/or			electron	4	6.52
be	32	5.50				free	5	8.70			19.11			
									acceleration	20	9.79	pp_obj_of-i		
									displacement	13	8.86			9.21
									v	9	8.29	component	15	9.53
									momentum	5	7.75	magnitude	8	8.26
									position	4	7.70	change	7	7.70
									mass	9	7.25	direction	5	6.83
									speed	6	7.06	unit	3	6.55
									force	10	6.73			

file685788...	of vectors = 2Fsin ( ) | | |2| 76.	*Relative*	velocity : When the distance between two bodies is
file685788...	or both, then each is said to have a	*relative*	velocity with respect to the other. Relative velocity is
file685788...	velocity with respect to the other.	*Relative*	velocity is vector difference of velocities. a. The
file685788...	difference of velocities. a. The	*relative*	velocity of body 'A' w.r.t. 'B' is given by VR = VA – VB b. The
file685788...	' w.r.t. 'B' is given by VR = VA – VB b. The	*relative*	velocity of body 'B' w.r.t. 'A' is given by VR = VB – VA c. VA –
file685788...	bodies moving in the same direction,	*relative*	velocity is equal to the difference of velocities. (θ = 0°
file685788...	bodies moving in opposite direction,	*relative*	velocity is equal to the sum of their velocities. (θ =180°

FIGURE 2 | Word sketch and KWIC of *velocity* (from Sketch Engine).

word sketches of both *speed* and *velocity*. They could observe that *velocity* (**Figure 2**) is often modified by *resultant*, *displacement*, and *space* (terms generally associated to vector quantity), whereas *speed* (**Figure 3**) is linked through a high frequency number of occurrences to *average* (meaning scalar quantity). The Sketch diff (**Figure 4**) offered a visual summary of the uses of the two words,

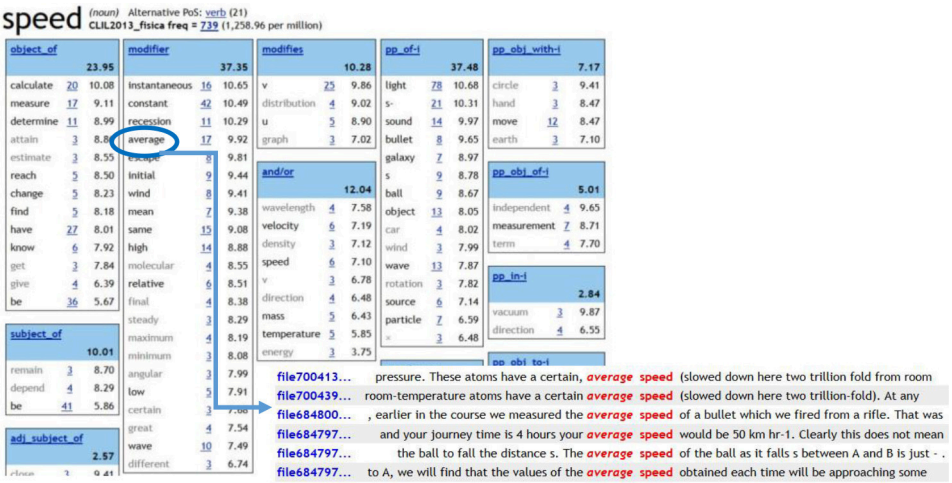


FIGURE 3 | Word sketch and KWIC of *speed* (from Sketch Engine).

of the language features—and thus content features—they have in common, and of the exclusive peculiarities of each of them. Tracing the examples confirmed the students' intuitions and gave them an insight into the definition of the specific content of the terms.

The scaffolding function of corpora partially diverge from the pure DDL definition, as given by Johns (1991, p. 3), where "the data is primary, and the teacher does not know in advance exactly what rules or patterns the learners will discover." On the contrary, and especially in school education, teachers should be well aware of the corpus composition and language content, in order to fulfill their role as guide and indirectly pilot the autonomous discovery of language and content by their students.

Full DDL-based CLIL didactic modules for physics have also been implemented. An example of worth mentioning good practice is an activity about Ideal Gas Law[2], where a corpus-based approach was used both to actively collect a LSP vocabulary and to give a warming up summary of the topics to be studied in depth throughout the unit.

After collecting and uploading a small corpus on Sketch Engine, the teacher asked students to make a word list of nouns, verbs and adjectives in order to get a handle of the lexical material they were going to deal with. The most significant items that were identified were the words *gas, temperature, volume, pressure, particle, collision, constant, proportional, universal, absolute*.

Starting from the first word on, collocations were extracted and word sketches were drawn.

Students observed that the most frequent attributes of the noun *gas* are *ideal* and *real* (**Figure 5**) and it is often associated to the expressions *temperature of …/… at temperature*; *volume of …/… at volume*; *pressure of …/… at pressure*; *state of …* etc. (**Figure 6**), and to the verbs *expand*, *compress*, *behave like,* besides occurring in the phrases *gas equation, gas law, gas state.*

FIGURE 4 | Sketch diff between *speed* and *velocity* (from Sketch Engine).[3].

As reported in Corino (2014), from the disciplinary point of view, these occurrences actually introduce through expanded contextualized examples the differences between *ideal gases* and *real gases* and the physical quantities *temperature*, *volume,* and *pressure*, which typify the state of gases.

In order to sum up their linguistic observations, students were given a table to fill in with the pieces of information gathered from KWIC and collocations, and word sketches, thus being actively involved in the bottom-up elaboration process.

[2]The Didactic Unit was experimented by professor Anna Grazia Botti.

[3]**Figures 3–5** were presented and discussed also in Corino (2014, p. 73). Picture have been reproduced with the permission of the copyright holder (University of Hildesheim). Written informed consent was obtained per email from the

	Attributes	Subj./obj. of verbs
Temperature	• Thermodynamic • High/low • Absolute • Constant • Proportional	• Increase/decrease • Rise • Keep • Measure • Depend
Volume	• Small/large • Constant • Proportional	• Increase/decrease • Occupy • Keep • Measure • Depend
Pressure	• High/low • Constant • Proportional	• Increase/decrease • Exert • Keep • Measure

Some adjectives linked to *temperature* (*thermodynamic/absolute*) are part of the definition of the Kelvin temperature scale and of the concept of absolute zero; the verbs *keep* and *constant* are part of the occurrences *provided volume / temperature / pressure is kept constant*, which express Boyle's and Gay-Lussac's laws. The presence of *proportional* in connection to the three nouns suggests a relationship between all these quantities and it is frequently connected to the adverbs *directly* and *inversely,* the numerous examples at students' disposal also offer a linguistic model for expressing direct and inverse proportionality in English.

The syntagmatic relations of the keyword *particle* give some clues on the modality of interaction between the molecules of ideal gases: it occurs with the verbs *collide* and *interact*, in particular *interact by/through/on collision,* while *collision* has its highest frequency concordances with the adjectives *elastic/inelastic.* And so on …

Starting from the 10 selected keywords this bottom-up approach has allowed students to get a sizeable portion of the LSP needed and to draw a fairly detailed mind map to scaffold further exercises such as cloze texts of reading comprehension tasks. Not to mention the content that has filtered through the language analysis.

modifier		47.16
ideal	83	12.62
natural	37	11.38
greenhouse	26	11.15
monatomic	10	10.04
real	9	9.55
theoretical	5	8.91
exhaust	5	8.85
helium	6	8.77
hot	5	8.52
inert	3	8.34
hydrogen	5	8.33
diatomic	3	8.32
coal	5	8.22
universal	3	8.21
most	3	7.84
oxygen	3	7.81
liquid	3	7.39
same	3	6.72

FIGURE 5 | Word sketch *gas*.

DDL as for LSP

A particular case of DDL applied to LSP is a recent experience carried out in a vocational context, i.e., a school for hairdressers[4].

It is an interesting and rather unusual situation, in particular for the profile of the learners, who have different ages, display low motivation, and can rely only on really basic language competence (ranging from A1 to A2).

The class selected to experience DDL was a third year hairdressing course.

Seventeen pupils participated in the project, which was developed during seven sessions of 1 or 2 h, for a total amount of 12 contact hours. Attendance was constant, with only a few absences for the whole duration of the project. The students' age ranged from 15 to 19 years old; 15 students were female and two male. Fourteen students were Italian native speakers, one student's native language was Chinese, one Moldavian, and two Rumanian.

Materials and Methods

Considering the specific field of study of the course, a drawing of a hairdressing salon (**Figure 7**) was depicted by an illustrator in order to represent the most common objects, tools, and actions in that field. The main aim was to offer native speakers enough situations and material for them to be able to describe the picture for at least 3, 4 min. The descriptions elicited would then serve as corpus data for the DDL activities.

editor of the volume G. Faaß and the Publishing Department of the University of Hildesheim.

[4]The content of this paragraph is part of the MA thesis of Buschini (2018). For a more thorough description of the activities cfr. Buschini and Corino (2019).

pp_obj_of-i		28.39
mole	9	10.09
volume	18	9.54
sample	6	9.03
behaviour	4	8.92
theory	8	8.65
emission	3	8.13
molecule	6	8.09
mass	16	8.04
pressure	7	7.79
state	6	7.74
property	3	7.22
amount	3	6.82
temperature	10	6.80
energy	17	6.23
particle	3	5.43

.[39] The *temperature* of a classical ideal gas is related to its average kinetic energy per
toms about each other. The *temperature* of a gas is a measure of the average translational
ning to the large scale, the *temperature* of a gas is something that we can determine
n J/kgK T is the absolute *temperature* of the gas , in K Now, Normal atmospheric pressure is
air was compressed, the *temperature* of the gas increased. He measured the amount of work

FIGURE 6 | Word sketch *gas*.

FIGURE 7 | Cutting edge (Maurizio Modena, 2017).

The drawing was then described by 16 English native speakers, both male and female, 13 of them being from the UK, two from the US and one from South Africa. The outcome was a set of 16 recordings lasting from 53 s to 3 min and 49 s that were then transcribed. The transcription was faithful, only signs of hesitation, false starts and repetitions were left out, and a minimal amount of punctuation was added in order to make the text clear.

Having both the oral and written version of the elicited texts makes it possible to integrate different skills in a multi-modal DDL practice, thus answering Cobb and Boulton question when they speculate on "what multi-modal or multimedia corpora can bring to the table and their impact on speaking and listening skills" (2015, p. 495).

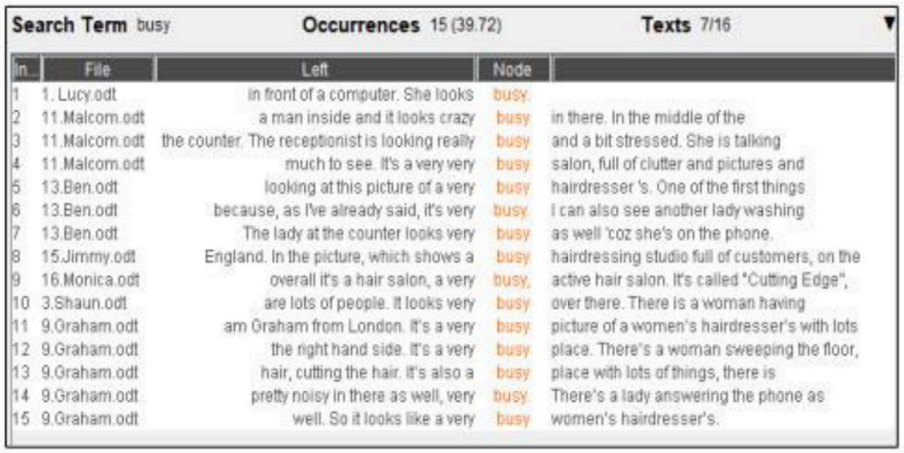
Once transcribed, the corpus was then loaded onto LancsBox, it comprises a total of 3,776 tokens and 550 lemmas.

A preliminary corpus-driven study carried out by the teacher highlighted many interesting utterances and some unpredictable ones presented too. A set of 20 words related to commonly mentioned objects and tools in the drawing was extracted using LancsBox and one recurrent grammatical structure emerged. The students had previously encountered some of the lexis which is the focus of the study but it was assumed that they had never come across the grammatical structure *have/get + object + past participle* (e.g., *have her hair cut*) which was extremely common in the native speakers' descriptions but sounds unnatural to an Italian audience.

The first part of the activity revolved around the audio files: students were divided into groups and some days before the first lesson received the oral descriptions of the picture to listen via Whatsapp, thus getting them acquainted with the principles of the flipped-classroom approach as well.

During class, learners were at first disoriented and discouraged by some aspects of authentic conversation containing random, off-topic utterances, frequent meaningless back-channeling (e.g., *yes*, *I see*), a lot of repetition. However, having the chance to listen to the recordings in their free time prior the beginning of the Data-driven learning project, helped them to start filtering the information and focus on the sentences which carried actual meaning, ignoring the utterances that were not important in order to understand the overall sense of the speech, along with interjections and false starts.

During the second preliminary session, the students were then asked to listen to the recordings and try to identify objects and actions on the picture, specifically focusing on the pronunciation of the name of the things or tools they did recognize. They were given a crossword to be filled in with the vocabulary elicited in the recordings. Completing the task was not compulsory, though. It was just a way to state a starting point and make students aware of their knowledge. The same activity will be then reproposed both in the corpus hand-on phase as guideline to explore the corpus,



Search Term busy Occurrences 15 (39.72) Texts 7/16

In...	File	Left	Node	
1	1. Lucy.odt	in front of a computer. She looks	busy	
2	11.Malcom.odt	a man inside and it looks crazy	busy	in there. In the middle of the
3	11.Malcom.odt	the counter. The receptionist is looking really	busy	and a bit stressed. She is talking
4	11.Malcom.odt	much to see. It's a very very	busy	salon, full of clutter and pictures and
5	13.Ben.odt	looking at this picture of a very	busy	hairdresser 's. One of the first things
6	13.Ben.odt	because, as I've already said, it's very	busy	I can also see another lady washing
7	13.Ben.odt	The lady at the counter looks very	busy	as well 'coz she's on the phone.
8	15.Jimmy.odt	England. In the picture, which shows a	busy	hairdressing studio full of customers, on the
9	16.Monica.odt	overall it's a hair salon, a very	busy	active hair salon. It's called "Cutting Edge",
10	3.Shaun.odt	are lots of people. It looks very	busy	over there. There is a woman having
11	9.Graham.odt	am Graham from London. It's a very	busy	picture of a women's hairdresser's with lots
12	9.Graham.odt	the right hand side. It's a very	busy	place. There's a woman sweeping the floor,
13	9.Graham.odt	hair, cutting the hair. It's also a	busy	place with lots of things, there is
14	9.Graham.odt	pretty noisy in there as well, very	busy	There's a lady answering the phone as
15	9.Graham.odt	well. So it looks like a very	busy	women's hairdresser's.

FIGURE 8 | Busy—LancsBox.

FIGURE 9 | Tray table.

and at the end of the DDL workflow to test the effectiveness of the project.

Students were eventually introduced to LancsBox and to the different features of the concordancer. In particular, the KWIC was explored in class, in order to produce basic sequences of words through the concordance lines, helping students to familiarize with the tool and show them how to analyse the text and search for specific words in context.

An initial search was carried out using the KWIC to research the word *busy* (**Figure 8**).

Once familiar with LancsBox, the students were then asked to complete the crossword previously started with the help of the audio files, the print-outs and the KWIC on LancsBox.

Once the students gained more confidence, they appreciated this hands-on experience and the general feeling was that they found concordances very useful as they were able to look for words or parts of the word somehow connected to the ones they were aiming to find and the program would return a string of words which helped the pupils to resolve the crossword in a more quickly way compared to a traditional analysis of a printed text.

An interesting discovery for the students was the use of the combination *tray table* to describe what in Italian would be called *carrello* (**Figure 9**).

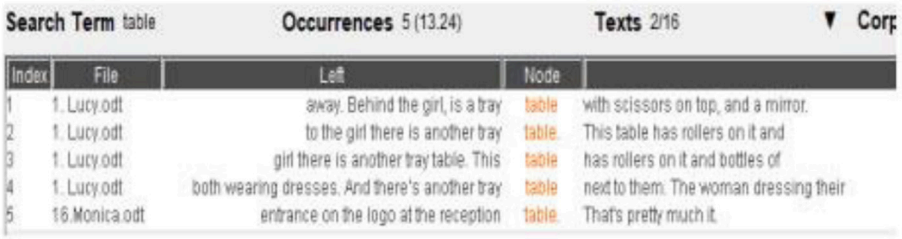
Most of the learners thought of the English word *trolley* or *table with wheels* to describe that specific object. However, those terms were not present in the corpus, whereas the object was referred to as *tray table* or *cabinet* instead. The learners discovered the answer to the definition required by crossword by searching *table* in the KWIC as shown in the figure below (**Figure 10**).

The above instance is a valid example of the fact that corpora can be used to identify potential or frequent errors in students' lexis or syntax. As Godwin-Jones (2017, p. 13) observes in fact, that "includes mismatches between the frequency of use by learners versus native speakers, including overuse, underuse, or misuse of particular words or constructions." The word *trolley* was this way identified as a mismatch and the term *tray table* used henceforth to describe the specific object. This way, students also realized that they could make guessing games thanks to the concordance software, they could, therefore, construct hypotheses for the word that they searched for and, through the KWIC, they were able to verify or discredit their hypotheses.

After the vocabulary was explored, DDL was used to introduce and inductively describe the grammar pattern *have/get + object + past participle* as well.

Students were asked to search the corpus and look for action verbs related to hairdressing. The first verb analyzed was *cut*, then the same was then done with other verbs such as *wash*, *brush*, *comb*, *straighten*, etc.

Following Johns's (1991) procedure of DDL, in the first step learners identified the structure under examination, they then classified it and in the last part the students draw a general rule describing the usage of the structure examined. So, a regular structure i.e., *have/get + object + past participle*, also



Search Term table Occurrences 5 (13.24) Texts 2/16 ▼ Corp

Index	File	Left	Node	
1	1. Lucy.odt	away. Behind the girl, is a tray	table	with scissors on top, and a mirror.
2	1. Lucy.odt	to the girl there is another tray	table.	This table has rollers on it and
3	1. Lucy.odt	girl there is another tray table. This	table	has rollers on it and bottles of
4	1. Lucy.odt	both wearing dresses. And there's another tray	table	next to them. The woman dressing their
5	16.Monica.odt	entrance on the logo at the reception	table.	That's pretty much it.

FIGURE 10 | *Table*—LancsBox.

known as causative passive, was highlighted and the students, in divided groups, examined the occurrences and tried to guess in what instances the above structure could be used. After careful consideration by the students, a rule of thumb was elicited.

The pupils came up with the idea that the morphosyntactical structure analyzed was used to describe something done by someone in someone else's interest.

RESULTS

After the DDL sessions, a test was carried out in the school hairdressing studio where the students had the chance to be in a real environment and could use real tools necessary to complete the task. They had to describe actions mimed by their classmates. The aim was to assess how much the students remembered of the lexis learned using the DDL method and whether the learners were going to use the structure *have/get +object + past participle* or not to describe the charades.

The results were surprisingly good: after more than a week, most of the learners remembered and tried to reproduce both lexical and grammatical structures and showed a great improvement in pronunciation too. Each and every group used the appropriate lexis and at least one example of the grammatical structure *have/get + object + past participle* even though it was not specifically requested in the task.

DISCUSSION

As Leech (1997) argues, by learning to interact with corpora, "students find themselves learning a great deal about language, and how to study language. They learn about the kinds of questions that can be usefully asked and answered by reference to a corpus of data."

The experiences presented in this paper suggest that DDL activities to support CLIL and to steer language learning have proven to have great benefits on students' language skills development, along with stimulating HOTS, motivation and involvement.

When students were asked about the corpus-based activities, the majority of them declared that they found it useful and relatively easy to work with the concordancer and with authentic data. All of them appreciated the new approach compared to traditional teaching methods, they found it very useful to be able to work on their own, under the guidance of the teacher and using the software to explore the language and to learn in a way different from the traditional lessons.

The overall outcomes seem to be quite positive and it looks like DDL was very well received. Furthermore, it appeared that consciousness-raising have started taking place in all the classes were the method was applied, leading to a long-term global improvement of the learning conditions, even among the students who are usually not very interested during traditional lessons. The very fact of using technological means in class stimulated the group work and promoted the participation of all students, including the ones who have learning disabilities or are not highly motivated.

These data shed some light on the reasons for encouraging more use of corpora in educational settings. It is a fact, though, that DDL in Italy—especially at high school level—is either ignored or suspiciously looked at as time consuming or too elaborate to be easily exploited with average students. On this point Hüttner et al. (2009) quote teacher education as an "interface of theory and practice," suggesting to train future teachers to work with and analyze LSP texts within an applied corpus-based/corpus-driven linguistics framework in order to prepare them to mediate insights to language and teaching practice.

ETHICS STATEMENT

Ethics approval and written informed parental consent was not required as per the authors' Institution's guidelines and national regulations because all data collected and used were the result of observations in class. No identifiable information or text was reproduced in the paper. The school "C.IA.C. (Consorzio InterAziendale Canavesano)"—Ciriè (TO)—was aware of the ongoing research and gave its approval for conducting it.

AUTHOR'S NOTE

Figures 8, **10** (Cutting Edge) were expressly designed for the teaching activity by Maurizio Modena. The figures have been used with the permission of the copyright holder.

AUTHOR CONTRIBUTIONS

EC contributed to plan the experiment, to analyze data and to write the manuscript (sections Tools and Resources, DDL in Practice, and Discussion). CO contributed to write the manuscript (sections Introduction, State of the Art, and Uses of Corpora in Language Teaching and Learning).

FUNDING

This paper was published with the funding of Dipartimento di Lingue e letterature straniere e Culture moderne–Università di Torino, *Fondo per la ricerca locale 2016*.

ACKNOWLEDGMENTS

Appreciation is expressed to Claudia Buschini and the CIAC School for permission to use their materials in this research.

REFERENCES

Ackerley, K. (2017). Effects of corpus-based instruction on phraseology in learner English. *Lang. Learn. Technol.* 21, 195–216.

Bernardini, S. (2004). "Corpora in the classroom," in *How to Use Corpora in Language Teaching*, ed. J. M. Sinclair (Amsterdam; Philadelphia, PA: John Benjamins), 15–36.

Boulton, A. (2012). What data for data-driven learning? *Eurocall Rev.* 20, 23–27.

Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *ASp* 69, 111–135. doi: 10.4000/asp.4826

Boulton, A., and Lenko-Szymanska, A. (eds.). (2015). *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam; Philadelphia, PA: John Benjamins.

Buschini, C. (2018). *Data-Driven Learning for ESP Courses: A Case Study in a Vocational School for Hairdressers*. Master's thesis, University of Turin, Turin.

Buschini, C., and Corino, E. (2019). "There's a woman having her hair cut. A case study of data driven learning in a vocational school for hairdressers," in *13th Biennial Teaching and Language Corpora (TaLC) Conference*, Cambridge.

Cobb, T., and Boulton, A. (2015). "Classroom applications of corpus analysis," in *The Cambridge Handbook of English Corpus Linguistics*, eds D. Biber, and R. Reppen (Cambridge: Cambridge University Press), 478–497.

Corino, E. (2014). "Bottom up specialized phraseology in CLIL teaching classes," in *Workshop Proceedings of the 12th Edition of the KONVENS Conference* (Hildesheim: Universitätsverlag Hildesheim), 68–76.

Gavioli, L. (2005). *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.

Godwin-Jones, R. (2017). Data-informed language learning. *Lang. Learn. Technol.* 21, 9–27.

Hadley, G., and Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Lang. Learn. Technol.* 21, 131–152.

Hüttner, J., Smit, U., and Mehlmauer-Larcher, B. (2009). ESP teacher education at the interface of theory and practice: introducing a model of mediated corpus-based genre analysis. *System* 37, 99–109. doi: 10.1016/j.system.2008.06.003

Indra Nugraha, S., Miftakh, F., and Wachyudi, K. (2017). "Teaching grammar through data-driven learning (DDL) approach. advances in social science, education and humanities research (ASSEHR)" in *Ninth International Conference on Applied Linguistics (CONAPLIN 9)* (Paris: Atlantis Press) 300–303.

Johns, T. (1991). "From printout to handout: grammar and vocabulary teaching in the context of data-driven learning," in *Classroom Concordancing. English Language Research Journal 4*, eds. T. Johns and P. King, 27–45. doi: 10.1017/CBO9781139524605.014

Kaltenböck, G., and Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: on the potential and limitations of computer corpora in language teaching. *ReCALL* 17, 65–84. doi: 10.1017/S0958344005000613

Leech, G. (1997). "Teaching and language corpora, a convergence," in *Teaching and Language Corpora (Applied Linguistics and Language Study)*, eds A. Wichmann, T. McEnery, and G. Knowles (London: Longman Publishing Group), 1–23.

Leech, G. (2011). "Frequency, corpora and language learning," in *A Taste for Corpora. In Honour of Sylviane Granger*, eds F. Meunier, S. De Cock, G. Gilquin, and M. Paquot (Amsterdam; Philadelphia, PA: Benjamins), 7–32.

McEnery, T., and Xiao, R. (2010). "What corpora can offer in language teaching and learning," in *Handbook of Research in Second Language Teaching and Learning, Vol. 2*, ed. E. Hinkel (London; New York, NY: Routledge), 364–380.

Meunier, F. (2011). Corpus linguistics and second/foreign language learning: exploring multiple paths. *Rev. Bras. Linguíst. Apl.* 11, 459–477. doi: 10.1590/S1984-63982011000200008

Mindt, D. (1997). "English corpus linguistics and the foreign-language teaching syllabus," in *Using Corpora for Language Research: Studies in Honour of Geoffrey Leech*, eds J. Thomas and M. H. Short (Harlow: Longman), 232–247.

Nesi, H. (2013). "ESP and corpus studies," in *The Handbook of English for Specific Purposes*, eds B. Paltridge, and S. Starfield (Oxford: Wiley-Blackwell), 407–426.

Ragazzini, G. (2016). *il Ragazzini 2016. Dizionario Inglese-italiano Italiano-inglese*. Bologna: Zanichelli.

Shaw, E. M. (2011). *Teaching Vocabulary Through Data-driven Learning*. MA thesis, Brigham Young University. Available onlne at: https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=4023andcontext=etd

Sinclair, J. (ed.). (2004a). *How to Use Corpora in Language Teaching*. Amsterdam; Philadelphia, PA: John Benjamins.

Sinclair, J. (ed.). (2004b). *Developing Linguistic Corpora: a Guide to Good Practice*. Available online at: https://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science* 328, 450–452. doi: 10.1126/science.1182597

Suan Chong, C. (2016). Ten *Innovations That Have Changed English Language Teaching*. Available online at: https://www.britishcouncil.org/voices-magazine/ten-innovations-have-changedenglish-language-teaching

Thomas, J. (2015). "Deriving extended collocations from full text for student analysis and synthesis," in *Multiple Affordances of Language Corpora for Data-driven Learning*, eds A. Leńko-Szymańska and A. Boulton (Amsterdam; Philadelphia, PA: John Benjamins), 85–108.

Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *Engl. Spec. Purposes* 30, 101–112. doi: 10.1016/j.esp.2010.12.003

Yanto, E. S., and Indra Nugraha, S. (2017). The implementation of corpus-aided discovery learning in english grammar pedagogy. *J. ELT Res.* 2, 66–83. doi: 10.22236/JER_Vol2Issue2pp66-83

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.