

MLS - EDUCATIONAL RESEARCH

<http://mlsjournals.com/Educational-Research-Journal>

ISSN: 2603-5820



How to cite this article:

Abad Castelló, M. & Álvarez Baz, A. (2021). Data-driven learning and lexical combinations: a didactic proposal with near synonyms. *MLS Educational Research*, 5(2), 88-104. doi: 10.29314/mlser.v5i2.562.

DATA-DRIVEN LEARNING IN SPANISH AS A FOREIGN LANGUAGE: A CASE STUDY WITH NEAR SYNONYMS

Magdalena Abad Castelló

Instituto Cervantes de Mánchester (United Kingdom)

Malena@Abad.com · <https://orcid.org/0000-0002-1942-7421>

Antxon Álvarez Baz

Universidad de Granada (Spain)

antxon@ugr.es · <https://orcid.org/0000-0003-1955-0232>

Abstract. Language corpora have been used as a tool for language learning from the late 80s in the field of EFL (English as a foreign language) and there is a vast body of empirical research with English as target language. In the field of Spanish as a foreign language (SFL), while numerous pedagogical oriented articles can be found, empirical studies are much scarcer. A content that causes difficulties for students and that can benefit enormously from this approach is vocabulary. The aim of this case study is to show how certain activities with language corpora integrated in a teaching sequence can help students to deepen their lexical-semantic knowledge of the lexical units. This paper presents a practical application of Data-Driven Learning (DDL) in the context of Spanish as a foreign language: a didactic sequence for the acquisition of climate related near synonyms aimed at B2 students. Three different groups of British students of Spanish carried out activities with indirect and direct use of CORPES XXI. After the implementation of the teaching sequence, a questionnaire administered to the students showed satisfactory perceptions of the DDL activities, both towards printed materials and towards direct consultation of the corpus. Likewise, direct observations of the students' actions during their consultation of the corpus showed a positive attitude towards the use of corpora, although varying in degree. Results support the integration of ABD in comprehensive teaching sequences.

Key words: Corpus; Data Driven Learning (DDL); Spanish as a foreign language (SFL); teaching sequence.

APRENDIZAJE BASADO EN DATOS EN ESPAÑOL COMO LENGUA EXTRANJERA: UN ESTUDIO DE CASO CON CUASISINÓNIMOS

Resumen. Los corpus lingüísticos se vienen utilizando como herramienta para el aprendizaje de lenguas desde finales de los años 80 en el área de inglés como lengua extranjera con un extenso volumen de investigaciones empíricas. En el campo del español como lengua extranjera (ELE) abundan las propuestas didácticas mientras que los estudios empíricos son mucho más escasos. Por otro lado, un contenido que causa dificultades a los estudiantes y del que se pueden beneficiar enormemente de este enfoque es el léxico.

El objetivo de este estudio de caso es mostrar, cómo determinadas actividades con corpus lingüísticos integradas en una unidad didáctica pueden ayudar a profundizar el conocimiento léxicosemántico de las unidades léxicas. En este artículo presentamos un ejemplo de aplicación práctica del aprendizaje basado en datos (ABD) en ELE. Se trata de una propuesta didáctica dirigida a estudiantes de nivel B2 para la adquisición de cuasisinónimos relacionados con el clima. Tres grupos diferentes de estudiantes británicos realizaron actividades con uso indirecto y directo del corpus CORPES XXI. El cuestionario administrado a los estudiantes tras la implementación de la unidad didáctica mostró percepciones satisfactorias de las diferentes actividades realizadas, tanto hacia los materiales impresos como hacia la consulta directa al corpus. Asimismo, las observaciones directas de las acciones de los estudiantes durante su consulta directa al corpus mostraron una actitud positiva hacia el uso de corpus lingüísticos, aunque con diferentes grados. Estos resultados avalan la utilidad de las actividades ABD integradas en unidades didácticas completas.

Palabras clave: Corpus lingüístico, Aprendizaje basado en datos (ABD), español como lengua extranjera (ELE), unidad didáctica.

Introduction

Today there seems to be a consensus on the key role of lexical combinations in learning a foreign language (Lewis, 1993; 1997), more specifically, for the development of fluency (Wood, 2007; Thomson, 2017) and written expression (Garner, Crossley, & Kyle, 2018). Language learners are well aware of this. Also, a large majority of these learners are introduced to foreign languages with mobile applications based on sentences rather than monoverbal units. Following Nattinger and DeCarrico (1992), this preference may be due to the fact that these expressions allow lower level learners to use utterances which they cannot yet construct autonomously. Unfortunately, in subsequent stages of learning, learners often construct sentences in which they transfer the combinatorial of their native language with very unidiomatic and unnatural results (Lewis, 1993; Fenik & Dikilitas, 2014).

These pluriverbal units are difficult to teach (Nattinger & DeCarrico, 1992; Boers & Lindstromberg, 2009). Very often they are invisible to the learner, so the teacher has to help the learner to perceive these blocks and to inquire about them. To this end, the teacher must train his students to develop strategies of discovery and analysis in order to understand their meaning. However, this analysis may not be enough; the learner needs repeated encounters with these units in order to retain them in long-term memory. In this respect, linguistic corpora are an ideal tool for learning lexical combinations as they make the units visible, provide multiple encounters, and facilitate exploration and analysis of the units.

In this paper we present a proposal for the integration of DDL activities in a complete didactic unit for B2 level students, with which we want to show the benefits of this approach in the acquisition of lexical units, specifically, two groups of near synonyms related to climate. More specifically, we explore learners' perceptions towards the activities carried out with the two types of access to the corpus (printed materials and direct consultation of the corpus) and in two different environments: in the classroom and outside the classroom, as homework at home.

A first step in lexical acquisition is unit acquisition and a second step is unit retention. Craik and Lockhart (1972) with their Depth of Processing Hypothesis argue that there is a close relationship between cognitive depth and retention. Thus, actively working with a lexical unit will multiply the probability of storing that information in our long-term memory. This is what Hulstijn and Laufer (2001) called the "Involvement Load Hypothesis" of the task. That is, this load can be further optimized if the learner is involved in the task.

Recent empirical studies in SFL have explored the effectiveness of different types of explicit teaching of lexical combinations. Pérez Serrano's (2015) study showed that both explicit teaching and simple collocation highlighting are effective for collocation acquisition. Jensen (2017) tested two explicit teaching methods, contrastive analysis and translation (CAT) and form-focused teaching and both worked, concluding that "any exercise which leads students to cognitively engage with a set of previously selected collocations is likely sufficient for the learning of these items" (p. 16).

There are numerous didactic proposals for teaching lexical combinations in SFL. Higuera (2006) and Haddouch (2015) emphasize didactic sequencing. Fernández Montoro's (2015) proposal integrates lexis and culture. Chamorro (2017), like Jensen (2017), suggests contrastive analysis and translation activities. For his part, Pérez Serrano (2017) insists on the need to understand the meaning of the unit and, following de Boers (2013), suggests using linguistic motivation to deepen the knowledge of the block.

Data-driven learning (DDL) is a learner-centered approach to learning whereby, following Johns (1990), the learner is a detective investigating data drawn from linguistic corpora, and the teacher is merely a facilitator guiding the task. The most common way of presenting this data is in the form of concordance lines (see Figure 1), which have a word marked in the middle of the line, a so-called Key Word in Context (KWIC). These contexts allow the learner to observe positional and combinatorial patterns and, in this way, to grasp the different micro-meanings of words depending on their co-occurrences. In this way, DDL provides the two key steps for acquisition: grasping through the concordance lines and deep cognitive activity in the analysis and discovery tasks.

In short, it can be stated that data-driven learning brings the language learner closer to the knowledge of a native speaker. In the words of Pérez-Paredes and Zapata-Ros:

"The use of DDL activities can allow the language learner to access knowledge that, intuitively, a native speaker of the language may come to possess in at least some of the numerous registers of language use" (Pérez-Paredes & Zapata-Ros, 2018: p.7).

In English as a foreign language, the effectiveness of DDL for lexical acquisition has been proven in numerous empirical studies. Lee, Warschauer, and Lee (2019) conducted a meta-analysis of 29 empirical studies with lexis as the focus and found a significant positive impact in all cases. Also, the authors found that corpus use was most effective when matching lines were selected and when print materials were combined with direct corpus use. A final major finding was that DDL performed better in studies that explored deep knowledge than in studies that focused on precise knowledge. DDL studies on the acquisition of the deep dimension can have two foci: referential meaning or syntagmatic relations of lexical units. Among the former, Mansoor's (2014) studies on semantic prosody and Yilmaz's (2017) studies on abstract noun usage yielded positive results. Among the latter, Ackerley's (2017) studies on phraseology, Szudarski (2020) with phrasal verbs and adverbial locutions as target form and Liountou (2020), who investigated the acquisition of idiomatic expressions, showed positive effects on the acquisition of lexical units.

As for empirical research in the field of SFL, Benavides (2015) and Marcos Miguel (2020) explored grammatical aspects and Contreras Izquierdo (2019) investigated varieties of Spanish. So far, we have only found two investigations with lexical focus. In Vincze's (2015) study, students used a selection of concordance lines to correct placement errors with very positive results. More recently, Yao (2019) conducted a study, with 38 monoverbal units as target forms, in which she tested the effectiveness of the DDL approach against the use of traditional methods.

Finally, it should be noted that more and more authors insist on the need for DDL proposals that integrate DDL activities into complete didactic units with the use of lexical combinations as we propose. Following Leńko-Szymańska (2014), teachers have to develop materials suitable for the DDL approach of proven pedagogical soundness, combine them with other teaching techniques, and integrate them into their teaching context. Asención-Delaney et al. (2015) insist on this integration in teaching practice by stating that DDL "combines meaning-centered input with language-centered learning and should therefore be complemented by other learning activities that focus on the production and fluent use of new words to foster a comprehensive knowledge of vocabulary" (2015, p. 144).

Method

This case study aims to show how certain DDL activities integrated in a didactic unit can help to deepen the knowledge of specific lexical units: specifically, two groups of near synonyms. The study analyses the students' perceptions of the activities, using a questionnaire and the observations of the teacher-researcher.

The didactic unit presented was carried out with three different groups of adult learners at the Instituto Cervantes in Manchester. All three groups were students of so-called special upper level courses. These courses have no syllabus, as the syllabus is designed ad hoc for each group and course. A task-based approach is applied at the center with continuous assessment based on individual monitoring of the students' progress in completing the tasks.

The participants in the study were 23 students of whom 21 were British and their mother tongue was English, one student had German as her mother tongue and one student had Polish as a mother tongue. All students were over 46 years of age and a large majority (65.2%) were over 66 years of age. In terms of gender the representation was very even: 13 women and 10 men. The vast majority of the students (82.6%) had a high level of education: undergraduate and postgraduate degrees. All had studied other languages and half of them had reached B2 level or higher in another language.

The study materials consisted of a complete didactic unit that included two sequences with DSL activities, one of direct use and the other of indirect use of the Corpus of 21st Century Spanish or CORPES XXI. This corpus was selected mainly because of the easy navigation of the interface, which is key to introduce students to the use of this tool. Searches in this corpus are intuitive and the results appear on the same screen.

The activities were part of a complete didactic unit on climate and weather, which followed the indicators of the Instituto Cervantes Curriculum Plan (2006) (see Table 1). The target forms are two groups of near synonyms related to the theme of the unit: four nouns denoting precipitation (*shower, rain, cloudburst, and downpour*) and four adjectives qualifying temperature (*warm, heated, hot, burning, and fiery*). In selecting the DDL target forms of the teaching units, several criteria were followed: thematic, didactic (the content poses difficulties; it is motivating and meets a need), and linguistic (the semantic features of the units offer possibilities for analysis by means of a linguistic corpus).

A questionnaire was also designed (Appendix A) with 12 statements, 4-point Likert scale in which 1 expressed "strongly disagree" and 4 "strongly agree". This questionnaire was designed by the teacher-researcher, validated by four university professors and piloted in a pre-experimentation pilot study. The answers of the questionnaires were analyzed with the statistical program SPSS. At the same time, during

the implementation of the didactic unit, the teacher-researcher took note of the development of the unit and, especially, of the students' actions, reactions, and comments during the different DDL tasks.

Table 1
Talk about time

TALK ABOUT THE WEATHER	
CEFR level B2 - C1	
Approximate duration of the whole didactic unit: 10 hours in 4 sessions of two and a half hours. Approximate duration of DDL sessions: 60 minutes in session 2 and 80 minutes in session 3.	
Objectives <ul style="list-style-type: none">- Deepen in the thematic contents: climate, weather, and environment.- Develop the command of the communicative activities of the language: reading and listening comprehension, written and oral expression and interaction, as well as mediation.- Train students in the use of linguistic corpora. Indirect use (Session 2) and direct use of CORPES XXI (Session 3).- Deepen students' lexical knowledge. Specifically, aim to study the following lexical features: synonymy, connotation, and ideology, figurative literal meaning, combinatorial and register.- Know the characteristics of a textual genre: weather reports.	
PCIC General indicators Specific notions 20.4. Climate and atmospheric weather Functions 5. Socializing 2. Express opinions, attitudes, and knowledge Socio-cultural knowledge and behavior 1.14. Ecology and environment Learning procedures 1.2.2. Elaboration and integration of information <ul style="list-style-type: none">Inductive Reasoning<ul style="list-style-type: none">▪ Generalization and formulation (implicit or explicit) of rules from the observation of phenomena.▪ Inference.	
Specific linguistic contents <ul style="list-style-type: none">• Expressions to formulate hypotheses• Lexical units to talk about climate, weather, and environment.	
Contents of the activities with DDL <ul style="list-style-type: none">- Near synonyms of precipitation (<i>shower, rain, cloudburst, and downpour</i>) and their combinational.- Near synonyms for expressing high temperature (<i>warm, heated, hot, burning, and fiery</i>) and their combinational.	
Resources required <ul style="list-style-type: none">• Attached worksheets• Computers with internet connection and projector to watch the videos.• Cardboard	
Didactic unit sessions Session 1. In this session the topic is introduced. The basic vocabulary of the topic is reviewed through hypotheses about the weather in different places. Finally, the use of time as a social resource to engage strangers in conversation is explored. Session 2. In this session, we discuss precipitation and introduce students to the use of linguistic corpora with materials and direct consultation with the teacher. Students are also introduced to the analysis of lexical units and their combinational as a means of discovering meaning and usage. At the end of the class, a discussion on time and character is proposed. Session 3	

In this unit new lexical units are introduced by means of contextualized texts. Students are then introduced to the direct use of linguistic corpora to work with three of these units. The activities of analysis of the units by their context and combinational are deepened. In addition, following the written and audio-visual models, students carry out the first group task of the unit: a weather forecast.

Session 4

In this unit, we work on the topic of climate change. Lexical combinations related to the topic are studied. After receiving input on the topic through texts (jigsaw reading) and a video, the second big group task is carried out: a debate on a sustainable consumer society.

The following two sections describe the activities in sessions 2 and 3 of the sequence. It has not been considered necessary to add more data on the activities of sessions 1 and 4 as the language corpora, which are the object of the study, were not used in these sessions.

DDL Activities in Session 2

The sequence with DDL begins when the teacher presents the students with two texts taken from CORPES XXI and asks them to decide which of the two proposed titles is theirs. Afterwards, they are asked if they know what a collocation is and are asked to look for weather-related collocations in the texts (*stormy afternoon, torrential rain, thunder and lightning, lightning, electric spark, blazing sun, sweltering heat*). The teacher picks up the collocation *torrential rain* and asks the students if they know other terms related to precipitation and asks them to brainstorm. After checking prior knowledge, they are told that they are going to work with four terms that denote precipitation (shower, rain, cloudburst, and downpour). They are told that they are going to analyze the data extracted from a linguistic corpus, specifically, some concordance lines about each of the terms.

Corpus del Español del Siglo XXI
(CORPES)

The screenshot shows the 'CONCORDANCIA' section of the CORPES XXI interface. The search term 'aguacero' is entered in the search box. The results are displayed in a table with two columns of text. The interface includes navigation buttons like 'Configuración', 'Ayuda', 'Estadística', 'Modo de cita', 'Sugerencias', and 'Preguntas frecuentes'. There are also filters for 'Clase de palabra', 'Grafía original', 'Suboritos', and 'Proximidad'. The concordance lines are sorted by 'Año ascendente'.

CONCORDANCIA		Ordenar por: Año ascendente
-Cuando murió, en Nopalillo cayó tramendo	aguacero	Mirando a la ventana se quedó tiesa. Recuerdo el rojo intenso de los geranios
coqueteo menos descarado. Recuerda, aquella vez que fuimos al arroyo en medio de un	aguacero	tachonado de relámpagos y truenos de guerra. Tu tenías apenas doce años y yo te
ser mediodía las casas y los techos y las calles, están desolados, esperando el	aguacero	inclemente que seguramente arrasará por un buen rato el rescoldo de verano que
porque la fuerza avasalladora de Juan De La Cruz arrieta cada vez más bajo este	aguacero	sin fin. Hace dos horas fluye y los lirios del fondo, al lado la celba sucumbieron
estudio se oscureció del todo, quedando indefenso bajo las nubes negras de otro	aguacero	y pude claudicar sin ningún esfuerzo a las inmensas ganas de llorar amargamente
Lloré sin medir el tiempo. Nunca supe cuanto y el	aguacero	se desvaneció sin percatarme, ni sentir su leve e imperceptible despedida en forma
escucha un gran trueno y los murciélagos vuelan espantados. De seguidas un torrencial	aguacero	comienza a caer. Vemos la lluvia filtrarse por el techo y bajar en grandes goterones
El	aguacero	continua arrojando. Afuera, en un alero a las puertas de la jefatura que están
sólo con el pantalón, entra en cuadro y se para entre ellos. Todos miran caer el	aguacero	Luego, atrás, vemos pasar, desnuda a Ana María y perderse por entre los calabozos
Sasagani en la cocina - pensó Tatu, cuando las primeras gotas se convierten en un	aguacero	El señor Hinagata se puso de pie. ¡Bibebí un poco y se fue caminando muy derecho
desierto. La gente esquivaba el Congreso por causa del olor nauseabundo. De pronto, un	aguacero	y el viento del sudeste nublaron todo. Cayeron marquesinas, rodaron tachos de basura
de ronquidos y paos que tronaban a raja suelta la Marsellesa del suelo. En pleno	aguacero	, estando, legaban esos amigos de Carlos a reunirse en el añllo. Y uno se quedaba
bueno, ya lo había hecho. El sol apareció entre las nubes negando la posibilidad de	aguacero	y la ciudad fue víctima de ese resplandor cobrizo que arrastra por las aceras
mayor que los demás. Garnier pensó que se trababa de alguna romería arruinada por el	aguacero	Luis Maurat y la que entonces era su mujer los arriaban como si se tratara de
por vigorosos truenos que hacían retumbar la habitación. Luego, comenzó a caer un	aguacero	con la fuerza que solo se conocía en Mauran y en sus islas adyacentes. Todos se
-Pase usted. Durand! Veo que lo ha cogido el	aguacero	¿Una desgracia dice?
Después de ese día, comenzó a caer la desgracia sobre la isla, como si fuera un	aguacero	intermitente. Yo me acuerdo como si fuera hoy, cuando Jasmin y Petik, dos jóvenes
hizo presente con enormes goterones aliados en un comienzo. Luego, un violento	aguacero	comenzó a golpear con desesperación los tejados de las construcciones que albergaban
ocasión, en vez del acostumbrado chubasco que refrescaba el ambiente, precipitó un	aguacero	inolvidable, de siete horas de duración, que saturó todas las tierras y destruyó
campo. Llovió intensamente en todo el territorio del reino, y una vez que escapó al	aguacero	no volvió a salir el sol como era costumbre, sino que el cielo se oscureció permanentemente

Figure 1. Concordance lines of "shower."

Note: Source: Corpes XXI.

The teacher asks the students to analyze the concordance lines and take note of the adjectives and verbs that accompany the nouns. They are then to look at these collocations and answer the questions on the analysis chart on the worksheet (Appendix B).

DDL Activities in Session 3

First, the teacher asks the students to read some newspaper headlines in which the phrase "hot autumn" appears and asks them if they understand what it means. He then tells the students that, in Spanish, there are several adjectives to describe a high temperature. She shows them five texts (Appendix C) in which the adjectives are presented and asks them what information the texts provide about the adjectives and whether the meaning is clear.

Later, the teacher explains to the students that they are going to study in depth three of the adjectives (warm, burning, and heated), but this time they are going to consult directly a linguistic corpus: the CORPES XXI. First he shows them how to access the corpus and how to search for concordances. Later, he shows them how the results appear, how the number of cases appear, and the large number of screens that appear. He also encourages them to look for the nouns that appear next to the adjectives and to try to find some kind of pattern. Then ask them, in pairs, to complete the table with the number of cases and the nouns which they are combined. Finally, he asks them to try to answer the questions in the analysis table (Appendix B).

As a final activity in this sequence, students are asked to do a gap-filling exercise with the same partner (Appendix D) in which they have to complete some sentences with one of the three adjectives. Finally, there is a debriefing of the two tasks in which the teacher also answers all the questions posed.

As homework, students are asked to use CORPES XXI to look up the other two high temperature adjectives and analyze them in the same way they did in class. Then, they will apply all this new knowledge to do another gap-filling exercise but this time with the five adjectives studied (Appendix D).

Results

After the completion of the didactic unit, the questionnaire was administered to the students, which was completed by 22 students out of the initial 23. They were asked to rate their perception of the corpus activities. On the one hand, they were asked about the different tools used: contextualized texts (CT) in the printed materials and concordance lines (CL) for the two types of access to the corpus: indirect access through printed materials and direct access. The questions on indirect access differentiated between homework done in the classroom (HC) and activities done as homework (HH). At the same time, they were asked about the usefulness of the activities in relation to general comprehension and to different dimensions of the lexical unit: meaning, level of formality, and combinational.

Table 2
Descriptive statistics of perception

WORK WITH PRINTED MATERIALS		1	2	3	4	Mean	DT
C	1 This activity has helped me to understand the meaning.	1	6	15		3.64	0.58
T	2 This activity has helped me to understand the level of formality (recording).	1	10	11		3.45	0.59
	3 This activity is useful as a first approximation.	1	7	14		3.59	0.59
C	4 This activity has helped me to understand the meaning.	1	6	15		3.64	0.58
L	5 This activity has helped me to understand the level of formality (recording).		11	11		3.50	0.51
	6 This activity has helped me to understand the usual collocations of a lexical unit.		11	11		3.50	0.51
WORK WITH DIRECT ACCESS TO THE CORPUS							
C	7 This activity has helped me to understand the meaning of the lexical unit.	2	10	10		3.36	0.65
L	8 This activity has helped me to understand the level of formality (recording).	4	10	8		3.18	0.73
H	9 This activity has helped me to understand the usual collocations of a lexical unit.	4	11	7		3.14	0.71
C	10 This activity has helped me to understand the meaning of the lexical unit.	2	12	7		3.24	0.62
L	11 This activity has helped me to understand the level of formality (register).	4	12	5		3.05	0.66
H	12 This activity has helped me to understand the usual collocations of a lexical unit.	3	12	6		3.14	0.65
	TOTAL AVERAGE					3.37	0.38

As can be seen in the table above, the students perceived the DDL activities favorably with a total average of 3.28 (out of 4). The perceptions are positive for both types of tools: matching lines and contextualized texts. The statements that receive the highest scores are those referring to the corpus as a tool that helps to understand the meaning, 3.64 in the activities with printed materials, 3.36 in the direct reference activities carried out in the classroom, and 3.24 in the direct access tasks at home. Statement 3 on the usefulness of the activities as a first approximation also gets a very positive perception (3.59). The question that receives the lowest score (3.05) is the one referring to the usefulness of the concordance lines in the comprehension of the register in the homework. Likewise, the questions about the usefulness of the corpus to understand collocations receive a lower score. Finally, a very remarkable fact is the higher score of the activities carried out with printed materials compared to the activities carried out by means of direct access to the corpus. And within the activities of direct consultation, the evaluation is lower in the activities carried out at home without the teacher's guide.

Next, we proceeded to analyze whether there were significant differences according to the variables gender, age, level of education, and proficiency in another language. First of all, we analyzed the mean of the answers on the activities in printed materials, the activities of direct consultation of the corpus, and the total mean. A better evaluation was predicted for younger students, with a higher level of studies and with a command of another language. However, as shown in the following table, the differences between the averages were minimal and only the expectation of age was confirmed with very small differences, although younger students clearly preferred printed materials, while among older students the difference between the types of access was much smaller. Interestingly, students with a lower level of education showed a more positive perception towards DDL activities. As for the gender variable, females showed a more positive perception towards printed materials while males favored direct access. However, the

differences were minimal in the four variables, so an inferential statistical analysis was not performed.

Table 3

Descriptive statistics of the perception by demographic variables

	Printed material		Shortcut Classroom		Shortcut House		Total	
	Mean	DT	Mean	DT	Mean	DT	Mean	DT
GENRE								
Woman	3.62	0.36	3.23	.64	3.10	.64	3.39	0.38
Man	3.44	0.38	3.22	.68	3.20	.50	3.33	0.39
AGE								
From 46 to 65	3.76	.26	3.09	.71	3.22	.75	3.46	.40
More than 66	3.45	.40	3.28	.62	3.11	.52	3.32	.37
LEVEL OF EDUCATION								
Secondary education	3.44	.34	3.66	.57	3.33	.57	3.47	.41
Undergraduate	3.61	.40	3.03	.67	3.00	.50	3.32	.31
Postgraduate	3.51	.42	3.29	.61	3.22	.68	3.38	.46
PROFICIENCY IN ANOTHER LANGUAGE (B2+)								
No	3.53	.39	3.18	.62	3.23	.41	3.37	.36
Yes	3.37	.40	3.27	.69	3.06	.71	3.37	.41

On the other hand, during the implementation of the didactic unit, the teacher-researcher took note of the development of the unit and, especially, of the students' actions, reactions, and comments during the different DDL tasks.

The students performed well in the first activity with printed materials. The collocations of the four terms helped them to perform the analysis activity and to understand the four terms. They had more trouble differentiating between *cloudburst* and *rain* but, thanks to the context, they identified the textual genre in which the former appeared (weather reports) and the figurative sense in which the latter often appeared. In the following production activities, they used the new units appropriately, except on one occasion when they used *cloudburst* figuratively, and it did not work in that context.

The activity of direct use of a corpus was a very new experience for almost all the students. Only two of them had used a corpus before. Among the rest of the students there were very different reactions. Most of them understood immediately how to do the searches. There were pairs with a very different technological level but together they were able to do the activity in an agile way and quickly made very interesting findings. Even after the activity they undertook independent searches on different units. However, there were students who read each of the concordance lines in full and were puzzled by the incompleteness of the sentences. These students read intensively looking for each of the new terms, thus slowing down the activity. In addition, this intensive reading prevented them from focusing on the most relevant aspects such as collocations and position in the sentence of the terms, and they had problems answering the analysis questions.

The homework assignment gave them the first opportunity to work independently with the corpora. After the training session, the students reported no problems in their autonomous search. They responded very well to the gap-filling exercise. They only hesitated between the use of *heated* and *hot* with "receiving"; however, they quickly proved that both adjectives were possible in one of the sentences with a difference in degree. There was also discussion between the uses of *heated* and *fiery* in the context of a confrontation. However, several students deduced how *fiery* applied more to the environment and *heated* more to people.

Discussion and conclusions

These results show, above all, a positive perception of all the DDL activities and their usefulness in understanding the meaning of the lexical units. In the different DDL activities on the corpus data, aspects such as frequency, figurative or literal meaning, referent and connotations were explored, which helped the students to deepen their semantic knowledge of the lexical unit. For example, a semantic feature such as intensity in precipitation helps to differentiate *shower* from *cloudburst*. Although to a lesser extent according to the results, the tasks served the students to understand the combinational of the target forms and their level of formality. Through this cognitive processing, the DDL activities promote deep knowledge of the lexical units and, with it, retention of the lexical units. The use of these units in the subsequent activities of the didactic unit confirms this retention. However, another more exhaustive study with a larger number of students would be needed to statistically evaluate this improvement in the knowledge and use of lexical units.

Also, the observation of students' actions and attitudes when performing the direct corpus consultation activities shows a high involvement of all students in the task and, in many cases, a high degree of autonomy in undertaking their own searches. Following the theories of text acquisition reviewed in the introduction, engagement can also lead to retention.

Moreover, in this case study, students were introduced to the direct use of a linguistic corpus, CORPES XXI. Specifically, students were trained to search for concordance by lemma and form, to take notes on frequency, and to analyze the concordance lines of some of the resulting displays. However, the scores indicate a more favorable perception towards the activities with printed materials than towards the activities in which they used the corpus directly. Moreover, students preferred the direct use in the classroom, guided by the teacher, to the use of the corpus outside the classroom (see Table 2). Although the difference is insignificant, these results confirm the teacher-researcher's observations about the difficulties experienced by some students in the activities of direct access to the corpus in class, indicating the need for further training in corpus use.

From the implementation of the unit, several conclusions can be drawn for lexical didactics in SFL: firstly, students improved their lexical-semantic knowledge and lexical competence in general by working with different aspects of the lexical unit (combinational, connotations, literal and figurative meaning, and textual genre). Secondly, the fact that DDL activities are integrated into a didactic unit means that the focus on form and meaning translates into a creative use of lexical units and, again, this use will have an impact on the retention of the lexical unit. Finally, it should be noted that DDL does not imply an abandonment of the cognitive/constructivist principles of task-based and communicative approaches. On the contrary, it reinforces and emphasizes the central role of the learner as an active agent of learning.

This study, focused on the perceptions of a group of students and the observations of the teacher-researcher, shows how a sequence of DDL activities integrated in a didactic unit helps to deepen the lexical knowledge of SFL. The multiple contexts offered by the concordance lines, the analytical activity when interpreting these lines, and the active involvement of the student are the pillars on which the implementation of DDL in the classroom is based.

We recognize that our case study is limited by the number of students and the object of study. In the future, more research in SFL on a larger scale and with a quantitative focus would be necessary in order to observe the effects on usage as well as student perceptions. In addition, the two DDL routes, teacher-mediated access and direct access, should be further explored. We hope that this proposal will help other SFL teachers to

bring this approach into the classroom and to create materials with DDL activities integrated into teaching units.

References

- Ackerley, K. (2017). Effects of corpus-based instruction on phraseology in learner English. *Language Learning & Technology*, 21(3), 195-216. <https://doi.org/10.125/44627>
- Asención-Delaney, Y., Collentine, J. G., Collentine, K., Colmenares, J., & Plonsky, L. (2015). El potencial de la enseñanza del vocabulario basada en corpus: optimismo con precaución. *Journal of Spanish Language Teaching*, 2(2), 140-151. <https://doi.org/10.1080/23247797.2015.1105516>
- Benavides, C. (2015). Using a Corpus in a 300-Level Spanish Grammar Course. *Foreign Language Annals*, 48(2), 218-235. <https://doi.org/10.1111/flan.12136>
- Boers, F. (2013). Cognitive Linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching*, 46(2), 208-224. <https://doi.org/10.1017/S0261444811000450>
- Boers, F., & Lindstromberg, S. (2009). *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Palgrave Macmillan.
- Chamorro, D. (2012). El foco en la forma léxica: cómo enseñar vocabulario. *Mosaico. Revista para la promoción y apoyo a la enseñanza del español*, 30, 26-33.
- Chamorro, D. (2017). Dar forma al enfoque léxico: directrices para clase y tipología de actividades. En F. Herrera (Ed.), *Enseñar léxico en el aula de español: El poder de las palabras* (págs. 71-82). Difusión.
- Contreras Izquierdo, N. (2019). Innovación docente en educación superior: recursos lexicográficos en el aprendizaje basado en datos (ABD) para la enseñanza de las variedades del español. En J. Gázquez Linares, M. Molero Jurado, A. Barragán Martín, M. Simón Márquez, Á. Martos Martínez, J. Soraino Sánchez, & N. F. Oropesa Ruiz (Edits.), *Innovación docente e investigación en Arte y Humanidades* (págs. 521-536). Dykinson.
- Corpus del español del siglo XXI. Manual de consulta en línea (v0.93 beta)*. (s.f.). <https://webfml.rae.es/CORPES/view/inicioExterno.view>
- Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.
- Fenik, S., & Dikilitas, K. (October de 2014). Integrating Corpora into Collocation-based Vocabulary Learning. *Humanising Language Teaching*, 5. <http://www.hltmag.co.uk/oct14/idea01.htm>
- Fernández Montoro, D. (2015). *Enseñar cultura a través del léxico: una combinación para favorecer el aprendizaje en el aula de ELE*. [Tesis doctoral sin publicar]. Universidad de Granada. <http://digibug.ugr.es/handle/10481/40373>
- Garner, J., Crossley, S., & Kyle, K. (2018). Beginning and intermediate L2 writer's use of N-grams: an association measures study. *International Review of Applied Linguistics in Language Teaching*, 58(1), 51-74. <https://doi.org/10.1515/iral-2017-0089>.
- Haddouch, B. (2015). *Las unidades fraseológicas en la enseñanza del español (caso del alumnado marroquí del Instituto Cervantes de Tetuán)*. [Tesis doctoral sin publicar]. Universidad Mohammed V-Agdal de Rabat.

- <https://www.mecd.gob.es/educacion/mc/redele/biblioteca-virtual/numerosanteriores/2015/memorias-master/balsam-haddouch.html>
- Higueras, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. Arco Libros.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539-558.
- Jensen, E. (2017). No Silver Bullet: L2 Collocation Instruction in an Advanced Spanish Classroom. *L2 Journal*, 9(3), 1-21.
- Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of datadriven. *CALL Austria*, 10, 14-34.
- Lee, H., Warschauer, M., & Lee, J. H. (2019). The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis. *Applied Linguistics*, 40(5), 721-753. <https://doi.org/10.1093/applin/amy012>
- Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260-278. <https://doi.org/10.1017/S095834401400010X>
- Lewis, M. (1993). *The Lexical Approach*. Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Language Teaching Publications.
- Liontou, T. (2020). The effect of data-driven learning activities on young EFL learners' processing of English idioms. En P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation* (págs. 208-227). Routledge.
- Mansoor, N. (2014). Teaching Semantic Prosody of English Verbs through the DDL Approach and its Effect on Learners' Vocabulary Choice Appropriateness in a Persian EFL Context. *Advances in Language and Literary Studies*, 5(2), 149-161. <https://doi.org/10.7575/aiac.all.v.5n.2p.149>
- Marcos Miguel, N. (2020). Exploring tasks-as-process in Spanish L2 classrooms: Can corpus-based tasks facilitate language exploration, language use, and engagement? *International Journal of Applied Linguistics* (Special issue article). <https://doi.org/10.1111/ijal.12314>
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Pérez Serrano, M. (2015). *Un enfoque léxico a prueba: efectos de la instrucción en el aprendizaje de las colocaciones léxicas* (Vol. 10). [Tesis doctoral]. Universidad de Salamanca. <http://infoling.org/search/tesis/ID/162#.WwBEJEgvy7A>
- Pérez Serrano, M. (2017). *La enseñanza-aprendizaje del vocabulario en ELE desde los enfoques léxicos*. Arco Libros.
- Pérez-Paredes, P., & Zapata-Ros, M. (2018). Patrones de Pensamiento Computacional y corpus lingüísticos: el aprendizaje de lenguas con datos lingüísticos. *RED. El aprendizaje en la Sociedad del Conocimiento*. <http://eprints.rclis.org/32209/>
- Szudarski, P. (2020). Effects of data driven learning on enhancing the phraseological knowledge of secondary school learners of L2 English. En P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation* (págs. 133-149). Routledge.
- Thomson, H. (2017). Building Speaking Fluency with Multiword Expressions. *TESL CANADA JOURNAL*, 34(3), 26-53. <http://dx.doi.org/10.18806/tesl.v34i3.1272>
- Vincze, O. (2015). *Learning multiword expressions from corpora and dictionaries*. [Tesis doctoral]. Universade da Coruña.
- Wood, D. (November de 2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT Journal*, 29(2), 209-230. <https://doi.org/10.37546/JALTJJ29.2>

- Yao, G. (2019). Vocabulary learning through data-driven learning in the context of Spanish as a foreign language. *Research in Corpus Linguistics*, 7, 18-46.
- Yilmaz, M. (2017). The Effect of Data-driven Learning on EFL Students' Acquisition of Lexico-grammatical Patterns in EFL Writing. *Eurasian Journal of Applied Linguistics*, 3(2), 75-88. <https://doi.org/10.32601/EJAL.460966>

Receipt date: 12/05/2020
Revision Date: 02/25/2021
Acceptance date: 07/10/2021

Appendices

Appendix A

Questionnaire on activities with corpus in the didactic unit

We would be grateful if you could answer this questionnaire.

We guarantee the confidentiality of the data and thank you for your invaluable cooperation.

First name _____ Last name(s) _____

1 Working with units in printed material.

The lexical units have been presented in the materials with several activities. Read each statement and put a cross next to the relevant number.

(4) strongly agree, (3) agree, (2) disagree, (1) strongly disagree.

a. The lexical unit has been presented within a short text.

1 This activity has helped me to understand meaning.	1 2 3 4
2 This activity has helped me to understand the level of formality (register).	1 2 3 4
3 This activity is useful as a first approximation.	1 2 3 4

b. The lexical unit has been presented in seven or more sentences.

4 This activity has helped me to understand meaning.	1 2 3 4
5 This activity has helped me to understand the level of formality (register).	1 2 3 4
6 This activity has helped me to understand the usual collocations of a lexical unit.	1 2 3 4

2 Work with direct access to the Corpus

a. A direct querying activity of multiple concordance lines from a corpus has been done in class.

3 This activity has helped me to understand the meaning of the lexical unit.	1 2 3 4
4 This activity has helped me to understand the level of formality (register).	1 2 3 4
5 This activity has helped me to understand the usual collocations of a lexical unit.	1 2 3 4

b. A direct consultation activity of multiple concordance lines from a corpus has been done outside of class as homework.

6 This activity has helped me to understand the meaning of the lexical unit.	1 2 3 4
7 This activity has helped me to understand the level of formality (register).	1 2 3 4
8 This activity has helped me to understand the usual collocations of a lexical unit.	1 2 3 4

Thank you very much!

Appendix B

Analysis tables

¿Qué adjetivos se usan para cada palabra?

aguacero	
chaparrón	
chubasco	
palo de agua	

¿Qué verbos se utilizan para cada palabra?

aguacero	
chaparrón	
chubasco	
palo de agua	

- ¿Qué palabras connotan más intensidad?
- ¿Qué término se utiliza de manera figurativa?
- ¿Qué término es más técnico?
- ¿Qué palabra se asocia con el verano?
- ¿Qué términos son más "latinos"?

	Número de casos	Sustantivos con los que se combina
caliente	10.274 casos en 4.230 documentos	<i>cama</i>
cálido		
acalorado		

- ¿Qué término es el más frecuente?
- ¿Qué términos se utilizan con frecuencia de manera figurada?
- ¿Qué término es más técnico?
- ¿Qué palabra tiene una connotación más positiva?

	Número de casos	Sustantivos con los que se combina
caldeado/a		
caluroso/a		

- ¿Qué término es el más frecuente?
- ¿Se utilizan de manera figurada?
- ¿Qué connotaciones tienen?
- ¿Se pueden aplicar a personas?

Appendix C

Contextualized texts

Me dijeron que en el monasterio ya tienen la habitación **caldeada** y oliendo a rosas, la cunita con paños hervidos y blancos, las toallas, Federico, las tetas de las vacas preparadas para ser ordeñadas. Todo listo para recibir a la niña a punto de nacer.

Afuera hay cuatro duchas con cortinas y tres WC con puerta. También hay una pequeña pileta con agua **caliente** y jabón donde pueden meter los pies. Alrededor de la pileta hay dos bancas de madera.



¿Ocho abdominales? Un gusano estaba en mejores condiciones que yo, y hasta tenía mejores perspectivas con las flexiones de brazos. Me sentía **acalorado**, agitado, me dolía y me picaba todo, regresé a la casa con el peor de los humores.

Pierre insistió en llevarme en el **cabriolet**, pero hacía tiempo que no andaba por las calles y rechacé su ofrecimiento. La tarde era hermosa y **cálida** y el crepúsculo me tomó paseando junto al río.

No, nunca, afirmó rotundo Tomás, mientras se protegía con una mano a modo de pantalla, del sol que le alumbraba los ojos. Ya no queda sombra, dijo a continuación. Con los cuerpos casi pegados contra los nichos, el sol brillaba en sus cabezas y les bañaba los hombros y la parte superior del torso. Va a ser un día muy **caluroso**, aseveró Alfredo (...).

Source: CORPES XXI

Appendix D

Gap filling activities

Activity 1 (done during class)

Complete the following sentences with one of the three adjectives studied.

warm heated fiery

- a. The congressional president resigned from his post during a _____ session in which the majority of legislators were debating his replacement.
- b. The _____ front will leave us mid-afternoon high or medium clouds. Temperatures will continue to rise.
- c. After the game we were hungry and bought some delicious _____ dogs from a stall.
- d. Feeling _____, he rolled up his sleeves and took off his shoes.
- e. Since this morning there is a very unpleasant _____ wind.
- f. The living room was very cozy: it was painted with _____ colors and furnished with a pine wood coffee table and chairs and a rustic style sideboard.

Activity 2 (done as homework)

With all that you have found out about these five adjectives, you are going to complete the following sentences. In some cases there may be more than one correct answer.

warm hot heated burning fiery

1. A vending machine for _____ drinks had been installed in the school.
2. The press conference yesterday became the scene of a _____ confrontation between the president and the journalist.
3. The musician was moved by the audience's _____ reception at his tribute concert.
4. Low-cost _____ water production systems are needed.
5. Some employees had been laid off and the atmosphere was very _____.
6. The Solymar Hotel welcomes you in a functional and _____ environment, both for your business meetings and congresses, or working days.
7. My teacher told me that if he noticed me _____, he could take off my jacket.
8. For the coolest summer and the _____ winter: Air conditioning products, Heating, Boilers, Fireplaces, Stoves.
9. An enviable natural setting accompanied by a mild climate, mild in winter, and not very _____ in summer.
10. The heating had been turned on, but the room was not yet _____.

(Activity taken and adapted from Chus Fernandez, University of Salford)