ARTIGO

Somewhere in between grammar and lexis: the role of verb types in learner production of caused-motion constructions

Entre a gramática e o léxico: o papel dos verbos na produção de construções de movimento causado por aprendizes

Rodrigo Garcia Rosa
rodrigo.rosa@usp.br
Universidade de São Paulo, São Paulo, SP, Brasil.

Abstract

This article deals with the selection, extraction and analysis of learner written production of English caused-motion constructions. The data extraction targeted four groups of learners with regard to their L1s (Brazilian Portuguese, Spanish, Italian and French) and their levels of English proficiency (from A1 to C2 in the CEFR). The data were extracted from EFCamDAT corpus, a morpho-syntactically tagged corpus of learner written production. The analysis targets the lexical variability in the production of learners across levels of proficiency and with different L1s so as to verify the extent to which we can claim learners have access to schematic caused motions or whether they appeal to specific verb classes, namely, instantiation and modification verbs. The results, especially the quantitative ones, indicate a strong reliance on verbal over constructional knowledge in the production of English caused-motion structures. The study also sheds light on developmental factors involving the relationship between grammar and lexis.

Keywords: Caused-motions Constructions; Modification and Instantiation verbs; Proficiency; Learner language; Corpora.

Resumo

Este artigo trata da seleção, extração e análise da produção escrita de aprendizes de construções de movimento causado em inglês. A extração de dados teve como foco quatro grupos de aprendizes, considerando suas línguas maternas (português brasileiro, espanhol, italiano e francês) e seus níveis de proficiência em inglês (de A1 a C2, segundo o QECR). Os dados foram extraídos do corpus EFCamDAT, um corpus de produção escrita de aprendizes com anotação morfossintática. A análise examina a variabilidade lexical na produção dos aprendizes, considerando diferentes níveis de proficiência e L1s, a fim de verificar até que ponto é possível afirmar que eles têm acesso a padrões esquemáticos de movimento causado ou se recorrem a classes verbais específicas, nomeadamente verbos de instanciação e de modificação. Os resultados, especialmente os quantitativos, indicam uma forte dependência do conhecimento verbal em

The Especialist

10.23925/2318-7115.2025v46i1e68552

OPEN ACCESS

FLUXO DA SUBMISSÃO:
Submissão do trabalho: 03/10/2024
Aprovação do trabalho: 07/08/2025
Publicação do trabalho: 26/09/2025

AVALIADO POR:
Deise Prina Dutra (UFMG)
Maria José Finatto (UFRS)

EDITADO POR:
Luciana Kool Modesto-Sarra (PUC-SP)

COMO CITAR:
ROSA, R. G. Entre a gramática e o léxico: o papel dos verbos na produção de construções de movimento causado por aprendizes. The Especialist, [S. l.], v. 46, n. 1, p. 910–937, 2025. DOI: 10.23925/2318-7115.2025v46i1e68552.

detrimento do conhecimento construcional na produção de estruturas de movimento causado em inglês. O estudo também lança luz sobre fatores de desenvolvimento que envolvem a relação entre gramática e léxico.

Palavras-chave: *construções de movimento causado; verbos de modificação e instanciação; proficiência; língua de aprendizes; corpora.*

1. Introduction

The idea that foreign language learners have access to L2 constructions is one of the theses defended by Cognitive Construction Grammar (Goldberg, 1995, 2006, 2013, 2018) and its application to L2 acquisition (Baicchi, 2015; Ellis, 2013; Gries & Wulff, 2005). Such a perspective is also the one used in this article in the treatment of the expressions below, which are all instances of caused-motion constructions, structures formally represented as [subj [v obj obl]] and functionally portraying a scene in which X CAUSES Y TO MOVE Z.

(1) John put the roses in the vase.

(2) She brought the problem to my attention.

(3) They laughed me out of their office.

(4) The boys drank themselves into stupor last night.

This study investigates the production of the type of constructions exemplified above by foreign language learners in order to help shed some light on the following research question, especially relevant to constructional studies: *do foreign language learners have L2 constructions as abstract as the kind of caused motions represented in the sentences above?* This is especially relevant because learners' capacity to produce sentences like (1) and (2) could be said to stem from their lexical knowledge of the argument structures of *put* and *bring*, since these verbs are two-participant events that require both a direct object and an oblique directional/locative argument. Nevertheless, the same cannot be said about (3) and (4) in which the main verbal predicates, respectively *laugh* and *drink,* do not require a third oblique argument (*out of their office* and *into stupor*). Therefore, if production of (3) and (4) is attested, it does not seem to be plausible to claim that such production is the result of lexical knowledge. To put it differently, constructions of the type exemplified in (3) and (4) could only be the result of more abstract representations in learners' cognition and not the result of specific lexicalization patterns.

This paper seeks to investigate this matter with the aid of corpus linguistic tools by looking at the production of learners of English in the EFCamDAT corpus (Education First - Cambridge Open Language Database). The investigation is based on the analysis of the production of learners with four L1 backgrounds, namely, Brazilian Portuguese, Spanish, Italian and French and four levels of proficiency in the CEFR (A2 to C1).

2. Cognitive and corpus linguistics

Cognitive linguistics, in general, and Cognitive Construction Grammar (CCxG), in particular, consider speakers' knowledge of language to be an inventory of symbolic structures that are conditioned to the communicative function of language within particular speech communities. Such a characteristic makes CCxG a theory of language that is deep-rooted in the usage of conventional constructions, that is, linguistic symbols that belong to the mental grammar of speakers, both natives and nonnatives. In addition to that, the *usage-based model* (Bybee, 2013; Langacker, 2013) acknowledges the importance of studies centered on the frequency of language expressions since frequency of use may be seen as systematic evidence of linguistic conventionalization. Thus, it can also be used as a diagnosis of what is more or less stable and prototypical in the cognition of speakers.

Bybee (2006, 2013) claims that the frequency of constructions has great impact on the modeling of grammar, and this can be shown by the capacity of speakers to recognize what is more or less conventional in their language. Likewise, Evans and Green (2006) draw attention to the relationship between frequency of use and the entrenchment of constructions.

> [...] the central claim of Cognitive Grammar, with respect to the usage-based thesis, is that usage affects grammatical representation in the mind. Furthermore, frequency of use correlates with entrenchment. Two main types of frequency effects have been described in the literature: token frequency and type frequency. Each of these gives rise to the entrenchment of different kinds of linguistic units. While token frequency gives rise to the entrenchment of instances, type frequency gives rise to the entrenchment of more abstract schemas. (Evans; Green, 2006, p.118)

To the authors, both kinds of frequency effects, *type* and *token* frequency, have a direct connection with the entrenchment of language patterns in the mind. In this way, checking frequency patterns might seem to be an effective way to probe into cognitive representations and strategies. Along the same lines, other researchers working at the interface of cognitive and corpus linguistics have identified strong correlations between the mental representation of linguistic patterns and their frequency in textual data. Schmid (2010), for instance, systematizes this correlation by postulating a principle which he names *From-Corpus-to-Cognition Principle.* The principle states that "frequency in text instantiates entrenchment in the cognitive system" (Schmid, 2000, p. 39). Inspired by Halliday's (1993) observations about the relationship between frequency in texts and probabilities in the system, Schmid (2000) postulates the principle, but warns that frequency in texts can only be taken as evidence of cognitive entrenchment if data retrieval and analysis are conducted in accord with standard practices and clear criteria. These criteria, according to Bybee (2006, 2013), Evans and Green (2006), Schmid (2000) and others, have been the focus of attention to those working in a corpus linguistics paradigm, a relatively modern methodological approach that takes language probabilities, as opposed to possibilities, with systematic and scientific rigor.

An exhaustive account of corpus linguistics as an area of investigation, its main methods and research applications goes beyond the objectives of this article, but a few aspects must be covered to contextualize our data selection and extraction. The first refers to the view of language shared among corpus linguists, that is, what they consider language to be for. To these researchers, languages can only be conceived as possible within and through human interaction in discourse (Teubert, 2009; Sinclair, 1991; Lindquist, 2009; McEnery & Hardie, 2012); that is, to these corpus linguists the idea that language cannot be detached from its contexts of use and the speakers that use it is a rather indisputable fact. This view of language places corpus linguistics within a broader area of functional studies, given the similarities shared between functionalist theories and corpus methods. McEnery and Hardie (2012, p. 168) summarize such compatibility by stating that

> Language is not seen as an abstract, isolated system, but one that is *used* to communicate meaning, and which is shaped by the ways it is used, by the contexts in which it occurs and by the structure of human cognition [...] [t]he emphasis on language *in use* makes functionalism compatible with corpus linguistics in a way that formalist linguistics is not.

Given the importance that *use* has in functionalist and corpus studies, it seems to be more than reasonable to state that observations about the structure of language and how it is used could only be made if the analysis relies on naturally occurring data. This empirical perspective makes it possible for corpus linguists to describe language from a *probabilistic* perspective. Also, the method is capable of presenting the analyst with patterns that speakers are unlikely to conjure up through introspection alone. Fillmore (1992, p. 35) draws attention to this fact by stating that "every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way". These words summarize two determining factors in research done from a corpus linguistics perspective, that is, the *objectivity* of the data being used (as opposed to subjective introspection) and a commitment to the *replicability* of the analyses, since "corpus data can easily be verified by other researchers" (Svartvik, 1992, p. 8).

Although corpus linguists are not the only researchers working from empirical perspectives that favor the use of samples of naturally occurring language in their analyses, only corpus linguistics is theoretically devoted to the creation and discussion of systematic methods of large data selection, compilation and observation. These data, or databases, can be used by researchers both in *quantitative* studies, usually centered on the conventionalization of linguistic structures, and in *qualitative* studies, which can help develop and/or (re)shape existing theories of language (McEnery & Hardie, 2012). Whether the corpus study is quantitative, qualitative or both quantitative and qualitative, though, is directly connected to the research questions and objectives. Other aspects related to the research aims, which may affect the accuracy of the analysis and the results, concern the corpus *representativeness*, *size* and *nature*. The discussion of these aspects will be particularized to the learner corpus used in this study.

3. Data extraction: EFCamDAT Corpus[1]

Education First - Cambridge Open Language Database (EFCamDAT) is an online linguistic database with 83,543,480 word tokens available. It contains records of written assignments of 174,743 EFL learners, whose proficiency levels range from A1 to C2 in the CEFR (Common European

[1] corpus.mml.cam.ac.uk.

Framework of Reference). The data are composed of essays written by learners as part of their course on *Englishtown*, an online English course owned by EF Education First. The entire course on *Englishtown* is formed of 16 proficiency levels, which are paired up with the CEFR levels as follows.

Table 1. EFCamDAT proficiency levels and the CEFR bands.

Englishtown levels	1-3	4-6	7-9	10-12	13-15	16
CEFR bands	A1	A2	B1	B2	C1	C2

Source: available in https://ef-lab.mmll.cam.ac.uk/EFCAMDAT.html

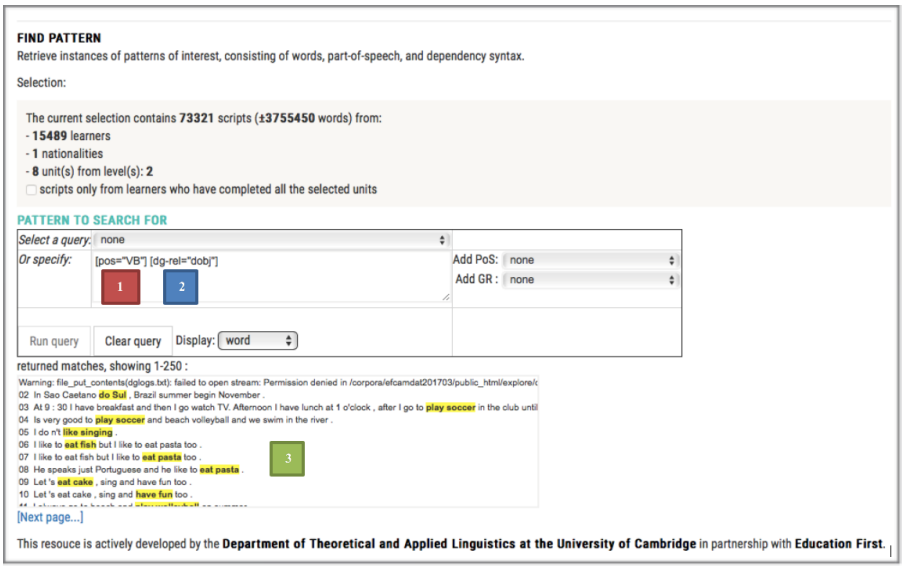
The 16 levels on *Englishtown* are used in the corpus as the criteria for proficiency segmentation. Each level of proficiency contains 8 units of work on a range of receptive and productive tasks and the written essays constituting the data on EFCamDAT cover an array of topics. The corpus data contain a total of 1,180,310 scripts[2] and 7,126,752 sentences produced by learners with a wide range of L1 backgrounds. There are 198 nationalities represented in the corpus and among these, Brazilian learners comprise the largest group, featuring 40,4% of the scripts and 31,078,406 number of words. This corresponds to 37,20% of the word tokens in the corpus. The data are annotated with part of speech tags (PoS), for which The Penn Treebank Tagset was used, and also contains some grammatical dependencies parsed with SyntaxNet Parser.

The script selection makes it possible for the user to select specific *teaching levels*, *script topics*, the learner *nationalities by continent* as well as the *countries* the researcher would like to restrict his search to. Once you complete the selection, the corpus provides the possibility of combining PoS searches with grammatical dependency relations. This combination is of extreme relevance to those interested in specific grammatical dependency relations, since the corpus has not been parsed for all sorts of grammatical constructions, such as caused motions, for instance.

Below is an example of a simple search that combines the PoS *verb base* ([pos="VB"]) and a grammatical relation of *direct object* ([dg-rel="dobj"]). The search results are shown in the form of concordance lines in which the searched items are highlighted in yellow (Fig.01, number 3).

[2] A reviewer questioned the use of the word *script* in place of the more generally accepted term *text*. As Fig.1 shows, the term *script* is what is used in the EFCamDat corpus interface. So, we adhere to such a use.

Figure 1. Example of a verb-object search on EFCamDAT.



Source: available in https://ef-lab.mmll.cam.ac.uk/EFCAMDAT.html

All things considered, the data on EFCamDAT follow the standard criteria for corpus data compilation, as discussed by Granger (1998, 2008), and allows the researcher to carry out investigations on a number of lexical and grammatical constructions with a variety of L1 backgrounds. In order to summarize the main corpus characteristics, Table 2 presents the main features of EFCamDAT in light of the criteria defined in Granger (1998).

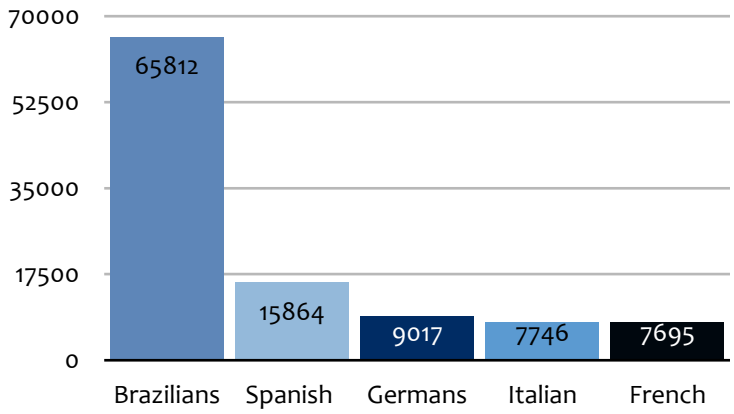
Table 2. EFCamDAT main features.

Language	**EFCamDAT**	**Learner**	**EFCamDAT**
Medium	Written	Age	Not specified
Genre	Essays	Gender	Not specified
Topic	Varied	Mother tongue	Controlled by nationality
Technicality	Not technical	Region	Specified
Task setting	Untimed/unassisted	Other foreign languages	Not specified
		Level	Teaching level
		Learning context	EFL learners
		Practical experience	Not specified

Source: author's data

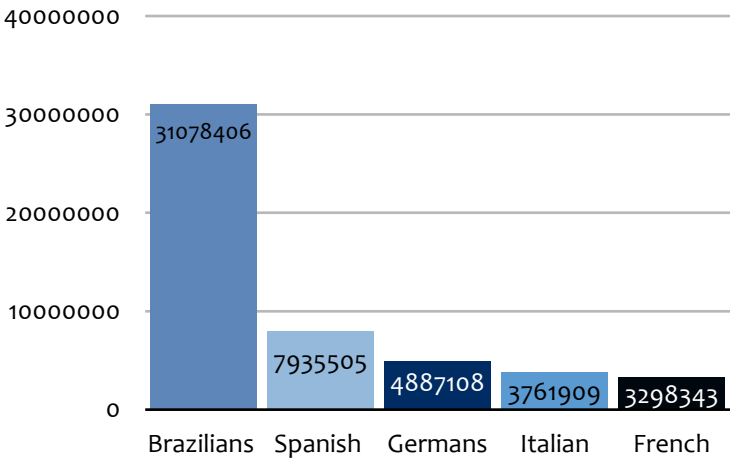
Considering that EFCamDAT does not contain specific information on the L1s of learners, this variable had to be tackled inferentially by the selection of the countries where these languages are spoken, that is, the 7.935.505 word-tokens of Spanish, for instance, reflect data extracted from learners of 20 different nationalities: Spain, Mexico, Costa Rica, El Salvador, Puerto Rico, Guatemala, Dominican Republic, Panama, Honduras, Cuba, Nicaragua, Colombia, Peru, Argentina, Venezuela, Chile, Ecuador, Paraguay, Bolivia and Uruguay. Below are two charts with information about the number of learners selected and the number of word-tokens per L1.

Figure 2. Number of learners per language.



Source: author's data

Figure 3. Number of word tokens per language.



Source: author's data

Having described the corpus as well as our data selection, the next section presents the criteria adopted for data extraction of learner caused-motion constructions from EFCamDAT.

3.1 Search syntax

As previously discussed, caused-motion constructions are linguistically characterized as follows:

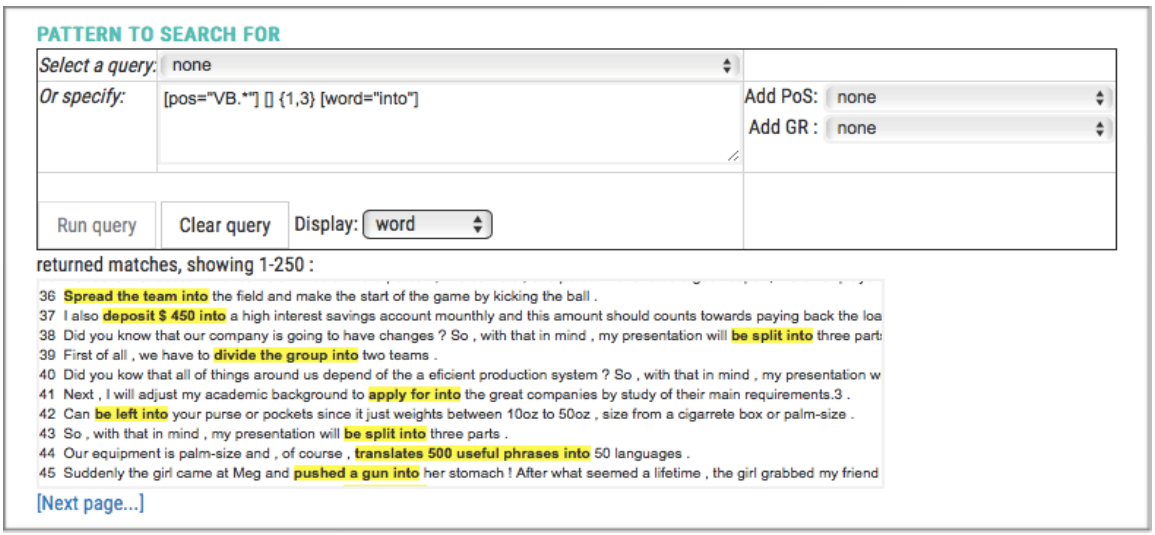
(5) a. Form: [subj [v obj obl]]
b. Function: X CAUSES Y TO MOVE Z

In this syntactic pattern, the verbal gap must be filled by non-static verbs, and the oblique argument takes the form of a PP that indicates the DIRECTION towards which the dislocated THEME will be caused to move. The non-adjacent relationship between the non-static verb and the directional PP, though, poses a problem for the search syntax in corpora that are not semantically tagged, given that the sheer syntactic sequence [subj [v obj obl]] can license a number of regular transitive constructions with normal adjunction, be it nominal (6) or verbal (7).

(6) I'm not crazy. I [$_{VP}$ saw [$_{NP}$ a man [$_{PP}$ with a knife]]]. (COCA/Movie/2016)
(7) I like to cook dinner with him and [$_{VP}$ [$_{V}$ watch [$_{NP}$ TV]] [$_{PP}$ at home]].
(COCA/Magazine/2006)

Therefore, the first challenge that data extraction imposed on our analysis relates with EFCamDAT search tools which, as we have shown, provide two main search possibilities. That is, one that specifies the PoS of the searched items and/or a search by dependency relations. Thus, in view of the impossibility of looking for semantically tagged caused-motion relations, we had to look for patterns that could potentially provide us with the constructions under examination. Below is our final search syntax and some concordance lines (Fig.4).

Figure 4. Search syntax for caused motions on EFCamDAT.



Source: available in https://ef-lab.mmll.cam.ac.uk/EFCAMDAT.html

Fig.4 spells out the search syntax used, that is, [pos="VB.*"] [] {1,3} [word="into"]. In it, we opted for the lemmatized form of the verb ([pos="VB.*"]) so as to capture all morphological instances of the verbs occurring in the pattern. The lemmatized verb is followed by [] {1,3}, which allows for the corpus to bring any sequence containing from 1 to 3 word-forms occurring between the lemmatized verb and the preposition/particle. A first search was tried with the syntactic tag [dg-rel="dobj"], a dependency tag for direct objects. Although the results with this dependency relation did reveal some instances of caused motions, simply allowing for the corpus to consider any element ranging from 1 to 3 elements covered both the instances captured by the syntactic tag [dg-rel="dobj"] and also some other instances left aside. In light of this comparison, we decided for the gap {1,3}. At last, in place of using specific prepositions/particles in the search, the first tested search specified the introducer of the last argument with a PoS tag provided in the corpus, [pos="IN"] (= *prepositions*). After eyeballing the resulting concordance lines and drawing a comparison with the final search, we opted to define each of the prepositional items, since this search generated more accurate data than the general one with the tag for prepositions. Separate searches were conducted for each of the following prepositions/particles: *off, out, into, onto, across, through, inside, outside, down the, up the.* With *down* and *up*, the definite article *the* was included as a means to enforce a prepositional use, rather than an adverbial one, and also to avoid

phrasal verbs. This final search resulted in the following figures for the pattern [lemmatized verb + {1,3} + preposition/particle]:

Table 3. Results for the search [pos="VB.*"] [] {1,3} [word="prep"] by proficiency levels.

	Portuguese	**Spanish**	**German**	**Italians**	**French**
A1	5637	1216	592	473	352
A2	5592	1322	967	747	566
B1	5427	1573	1059	774	531
B2	3767	1201	1703	933	806
C1	914	342	660	344	182
C2	156	46	119	84	0
Total	21493	5654	5100	3271	2437

Source: author's data

The figures above represent the absolute numbers of sequences the corpus search brought by different L1s and CEFR levels. As was expected, given the different sizes of data and learners, Portuguese presented almost four times more sequences than the second largest set of data, that is, Spanish, and almost ten times more than French, the language with the least amount of data and the lowest number of learners.

A sample of some resulting concordance lines by Portuguese and Spanish-speaking learners is provided below.

Portuguese

(8) a. It's very easy and fast, I *don't need go out* of my home. (BR/A1)[3]

b. There's *an Italian restaurant across* the street. (BR/A1)

c. I do the shopping, *watch cinema and eat out* once a week with my girlfriend. (BR/A2)

d. Now, I can respond quickly to professional emails and easily *access our database through* application. (BR/B1)

[3] The codes at the end of learner samples refer to the learner profile classification we have attributed to the data. BR refers to *Brazilian* and *A1* refers to the CEFR level. The codes used for the other languages are SP, ITA, FR and GER, respectively, for Spanish, Italian, French and German.

Spanish

(9) a. I don't usually *surf the internet through* the night. (SP4/A2)

b. Fortunately we *did manage to pay off* our mortgage. (SP10/B2)

c. I'm planning to *create a fund into* which the economically better-off students can pay in order to finance educational trips. (SP13/C1)

d. Apparently, thief *intented*[4] *to get into* the real state according with the neighbors who heard the sound of someone breaking a window (SP15/C1)

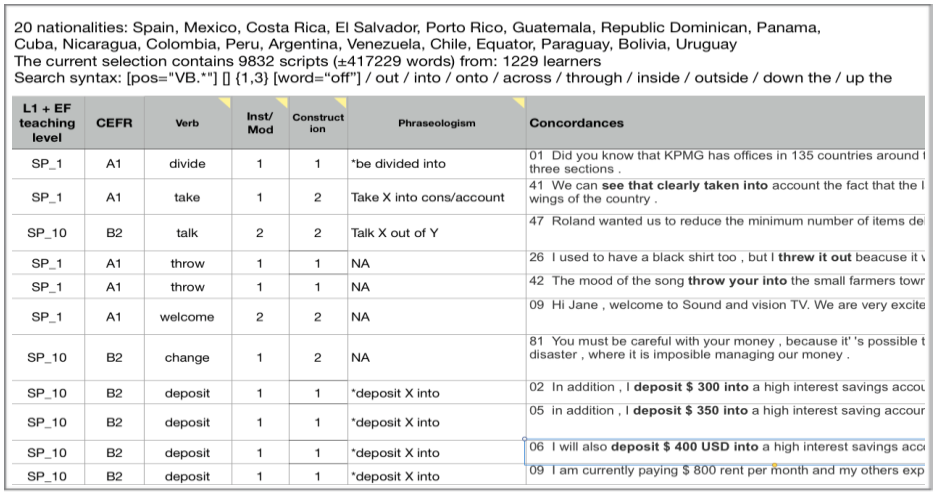
The data generated by the search syntax and exemplified above illustrate the first challenge of looking for caused-motion relations in the corpus. All the sequences in italics are formal instantiations of the sequence [pos="VB.*"] [] {1,3} [word="prep"], but none of them represent the structure we were looking for. Some of them (8a, for example) did bring cases of movement, but not of the caused-motion kind. Therefore, after computing all the sequences brought by the search syntax, the next step was to classify the resulting concordance lines and separate them into real instances of caused-motion constructions and random patterns. In other words, the "raw" data had to be semantically annotated so we could have a clear quantitative picture of the production of caused-motion constructions by the selected groups of learners.

3.2 The semantic annotation

After extracting the data from EFCamDAT, the next step was to find, among all the given concordances containing the sequence [subj [v obj obl]], those whose meanings instantiated the semantic reading of caused motion, that is, X CAUSES Y TO MOVE Z. Another important aspect of the analysis, which was also meant to help decide on the constructional semantic reading, was to isolate the verbs of each sentence. The criteria adopted for the analysis of occurrences were defined in different columns, as Fig.5 presents.

[4] All the learner data here analyzed will be kept in their original forms. Therefore, no grammatical inaccuracies, misspelled words, etc. will be corrected.

Figure 5. Spreadsheet with the data.



20 nationalities: Spain, Mexico, Costa Rica, El Salvador, Porto Rico, Guatemala, Republic Dominican, Panama, Cuba, Nicaragua, Colombia, Peru, Argentina, Venezuela, Chile, Equator, Paraguay, Bolivia, Uruguay
The current selection contains 9832 scripts (±417229 words) from: 1229 learners
Search syntax: [pos="VB.*"] [] {1,3} [word="off"] / out / into / onto / across / through / inside / outside / down the / up the

L1 + EF teaching level	CEFR	Verb	Inst/ Mod	Construct ion	Phraseologism	Concordances
SP_1	A1	divide	1	1	*be divided into	01 Did you know that KPMG has offices in 135 countries around t three sections .
SP_1	A1	take	1	2	Take X into cons/account	41 We can **see that clearly taken into** account the fact that the l wings of the country .
SP_10	B2	talk	2	2	Talk X out of Y	47 Roland wanted us to reduce the minimum number of items de
SP_1	A1	throw	1	1	NA	26 I used to have a black shirt too , but I **threw it out** beacuse it v
SP_1	A1	throw	1	1	NA	42 The mood of the song **throw your into** the small farmers towr
SP_1	A1	welcome	2	2	NA	09 Hi Jane , welcome to Sound and vision TV. We are very excite
SP_10	B2	change	1	2	NA	81 You must be careful with your money , because it' 's possible t disaster , where it is imposible managing our money .
SP_10	B2	deposit	1	1	*deposit X into	02 In addition , I **deposit $ 300 into** a high interest savings accou
SP_10	B2	deposit	1	1	*deposit X into	05 in addition , I **deposit $ 350 into** a high interest saving accour
SP_10	B2	deposit	1	1	*deposit X into	06 I will also **deposit $ 400 USD into** a high interest savings acc
SP_10	B2	deposit	1	1	*deposit X into	09 I am currently paying $ 800 rent per month and my others exp

Source: author's data

It is relevant to mention that the third column isolates the verbs used in the sentences and it is followed by other columns relevant to the analysis of the data, namely *verb categorization* (*instantiating* vs. *modifying* verbs), *constructional reading* (*literal* vs. *figurative*) and, lastly, whether or not the sentences produced represented some attested *phraseologisms*[5]. These points will be discussed in more detail in the analysis.

After the data tabulation, an extensive eyeball inspection was carried out with all the 37956 concordances with the aim to separate the data into two subgroups, that is, those that represented real caused-motion instances and those representing any other random pattern[6]. For matters related to how that was coded in the spreadsheet, we adopted the tag NA (= not applicable) for the sentences that did not contain caused-motion occurrences.

[5] This column, though, was not used as a definitional criterion in the analysis of the data. This column was added a posteriori as a result of the observation of the behavior of some types of caused motions, especially figurative ones with verbs of modification.

[6] Ideally, the annotation of caused-motion constructions should be done computationally. However, to our knowledge, no existing framework has reached an ideal level of accuracy for such cases. To make up for the subjectivity of the semantic annotation, we have conducted an informal double-checking test by selecting a random sample of 50 concordance lines, spread along the five annotated languages, and asked three PhD candidates in linguistics (one native speaker of English and two other proficient speakers) to annotate them segmenting the data into caused motions and random patterns. The test resulted in 95% of convergence between their answers and our annotation of the data.

The first segmentation into *caused motions* and *random patterns* reduced the entire data (37956 occurrences) to 5807 occurrences of caused motions distributed amongst five groups of learners. This corresponds to 15,29% of all the dataset. The absolute values and the corresponding percentages of caused motions and random patterns are provided below.

Table 4. Semantic annotation of caused motions and random patterns.

	Spanish	Portuguese	French	Italian	German
Caused motion	552 (9,76%)	3165 (14,72%)	402 (16,49%)	559 (17,08%)	1129 (22,13%)
Random pattern	5102 (90,23%)	18328 (85,27%)	2035 (83,50%)	2712 (82,91%)	3971 (77,86%)
Total	5654	21493	2437	3271	5100

Source: author's data

A first look at the absolute values already presents a significant reduction in the number of occurrences for caused motions across the board, but the figures also hint at the productivity of the X CAUSES Y TO MOVE Z reading for the pattern [pos="VB.*"] [] {1,3} [word="prep"] in our group of languages, since the contrast with our control group (German), in spite of being significant, is not huge. A more visual distribution in percentages is presented below.

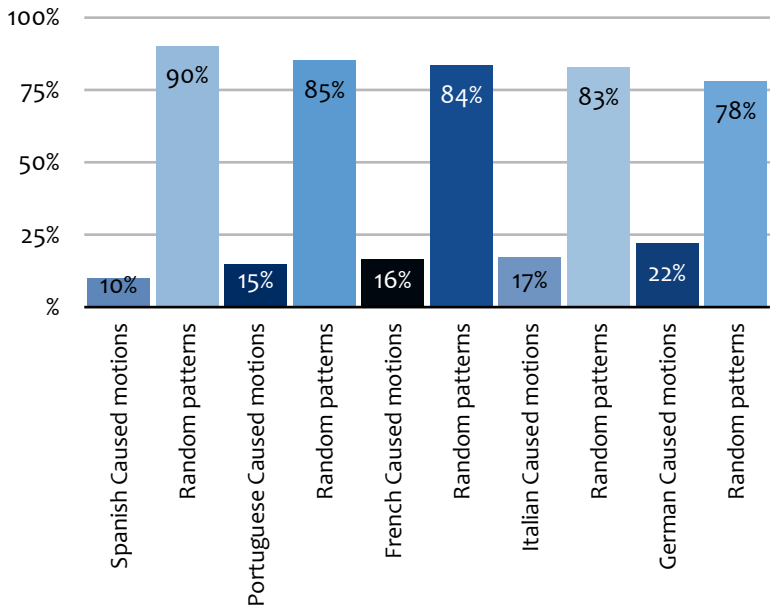


Figure 6. Caused motions vs. random patterns.

Source: author's data

Source: author's data

It is important to point out that the observed distribution above seems to be stable across levels of proficiency. The absolute values of caused motions per level of proficiency are provided in the table below.

Table 5. Caused motions out of random patterns per levels of proficiency.

CEFR levels	Brazilians	Spanish	Germans	Italians	French
A1	92 / 5637	10 / 1216	67 / 592	25 / 433	12 / 352
A2	1050 / 5593	72 / 1302	173 / 967	105 / 747	88 / 566
B1	1014 / 5427	215 / 1573	207 / 1059	168 / 774	98 / 531
B2	777 / 3764	191 / 1181	460 / 1703	114 / 913	178 / 802
C1	211 / 914	58 / 336	187 / 660	68 / 320	24 / 182
C2	18 / 156	05 / 46	20 / 119	10 / 84	0

Source: author's data

The table above clearly shows that there is relatively mild development in the use of the structure as the levels of proficiency increase. However, an observation of the absolute values of the proficiency levels raises an issue that should not be ignored. The low levels of occurrence at A1 and C2 might compromise our analytical capacity to generalize about these groups of learners (e.g. Spanish A1). The statistical significance of these groups of learners and levels, vis-à-vis other languages and levels, might take the analyses to draw inaccurate conclusions. Given the small size of this and other A1 and C2 datasets, their removal from our main database seemed to be the best option in order not to jeopardize our results.

The following section presents the analysis of the data as for the variable we subjected the data to, that is, verb type categorization: verbs of instantiation and modification.

4. Results and discussion

As discussed previously, caused-motion constructions are, among other things, characterized by the types of events that can be integrated with the constructional scheme, that

is, whether the constructions are fused with *instantiating* and *modifying verbs* (Cabrera & Zubizarreta, 2004; Zubizarreta & Oh, 2007). However, the existence of caused-motion structures in speakers' construction is better attested when we are confronted with expressions like Goldberg's classic '*she sneezed the foam off the cappuccino*' in which the directional argument (the oblique PP '*off the cappuccino*') is not predicted by the argument structure of the verb. In other words, since *modifying verbs* such as *talk*, *sneeze*, *laugh*, etc. do not subcategorize for complements (neither direct nor indirect objects), the directional PP is said to be provided by the construction. Expressions with *instantiating* verbs like *put*, *send*, *throw*, on the other hand, do not seem to reinforce the existence of the construction, since the argument structure of this verb class mirrors the number of arguments predicted in the schematic caused motion.

In light of this linguistic constraint, in real communication, learners would have to deal with two types of caused-motion constructions, which are exemplified below:

Caused motions with instantiating verbs:

(10) If the winners *get the ball into the box* for him in the right areas, he will score goals (COCA/Blog/2012)

(11) *Get your child into the habit* of keeping his hands away from his cold sore (COCA/Magazine/2000)

Caused motions with modifying verbs:

(12) Eventually he *laughed me right out of the office*. (COCA/Spoken/2014)

(13) They branded him a cowardly bureaucrat and *laughed the project out of existence*. (COCA/TV/1998)

Therefore, taking into account these constructional aspects of caused motions, the extracted learner data were analyzed in terms of types of verbs, *instantiation* or *modification,* so as to determine whether learners relied on their lexical knowledge of specific ditransitive verbs or whether they displayed knowledge of schematic structures while producing caused motions. This analysis is meant to give us the conditions to check the validity of the general research question presented in the introduction, that is, *do foreign language learners have L2 constructions as abstract*

as the kind of caused motions represented in the sentences above? But it is also able to shed some light on developmental factors by showing the extent to which learners' production is affected more by levels of proficiency than by L1 backgrounds.

To map out the relationship between the production of caused motions and specific verb classes and types, we first looked at the lexical variability in the verbal slot of learners' concordance lines.

4.1 Lexical variability: constructions and verbs

A first look at the grammatical properties of caused-motion constructions produced by learners aimed at probing into the lexical variability in the verbal slot. Considering the number of caused motions as absolute values, learners of the five groups under examination used a variety of verbs, as is shown in Table 6.

Table 6. Variability of verbs by groups of learners.

Portuguese (80 verbs)	add, allow, arrest, bring, carry, cast, change, commit, convert, crash, cut, dislocate, direct, distribute, divide, download, drive, drop, eat, engage, enter, force, get, head, help, hit, hurtle, implant, include, incorporate, insert, integrate, introduce, invite, jump, keep, kick, launch, lead, leave, let, loan, lock, make, move, place, play, plug, promote, publish, pull, push, put, receive, rewrite, save, scare, send, separate, serve, shoot, sign, slam, soak, split, spread, squeeze, store, take, talk, throw, toss, transform, translate, turn, upload, use, vote, walk, welcome.
Spanish (55 verbs)	divide, take, talk, throw, welcome, change, deposit, drag, get, guide, let, make, push, put, turn, type, bring, incorporate, load, pull, split, translate, lead, separate, set, transform, convert, export, process, vote, force, introduce, move, download, toss, enter, walk, disperse, help, involve, roll, add, hit, keep, kick, place, score, send, pour, rub, cross, express, include, launch, merger*.
Italian (66 verbs)	bring, launch, put, split, take, deposit, divide, help, keep, point, pull, push, type, add, allow, enter, get, immerse, make, plunge, transform, translate, compound, convert, cut, incorporate, lead, throw, turn, elect, include, organize, vote, haul, lock, move, set, tuck, change, squeeze, thrust, assist, breathe, guide, load, talk, integrate, send, sting, crash, draw, drive, fill, hit, kick, pour, shoot, download, introduce, ravish, conduct, drop, find, follow, search, sweep.
French (57 verbs)	bring, disclose, take, chase, deposit, divide, get, grab, integrate, push, put, save, stick, transfer, turn, watch, adapt, annoy, dive, log, make, slip, split, translate, call, involve, publish, transform, vote, welcome, find, move, project, charge, thrust, convert, talk, introduce, send, walk, decelerate, drive, develop, enter, fill, hit, let, lose, place, pull, share, show, throw, download, drop, mix, see.
German (118 verbs)	bring, disclose, divide, guide, introduce, invest, lead, let, plug, show, take, chase, convert, deposit, dip, drag, fasten, get, grab, help, involve, keep, loan, make, place, press, pump, push, put, save, send, spend, splash, split, transform, transport, turn, accompany, acquire, add, change, cut, load, pack, pull, spread, translate, type, bury, carry, draw, drive, incorporate, invite, leave, project, scare, separate, sink, slap, stab, stuff, integrate, elect, fill, find, fly, follow, remodel, transform, vote, accept, combine, laugh, lure, manoueuver*, move, ride, throw, trigger, tug, welcome, call, crack, implement, prevent, shock, thrust, elevate, order, set, squeeze, stress, talk, walk, compress, enter,

	sell, force, fling, have, kick, shoot, stretch, tuck, break, choose, create, flick, hit, line, play, rescue, roll, sort, write, download, soak.

Source: author's data

A simple eyeball inspection of the verbs in Table 6 shows that there is a clear difference in the range and variability of verbs used amongst the four target groups. Germans, our control group, used twice as many verbs as the Spanish-speaking learners, the group with the lowest lexical variability among the Romance language speakers. However, the range and number of verbs used is especially significant when compared proportionally with the number of caused motions used by different L1 speakers and at different levels of proficiency. Table 7 presents this ratio.

Table 7. Ratio of verbs and caused-motion constructions.

CEFR levels	**Portuguese**	**Spanish**	**Italian**	**French**	**German**
A2	0.043 (46/1050)	0.236 (17/72)	0.076 (8/105)	0.136 (12/88)	0.160 (27/168)
B1	0.052 (55/1014)	0.120 (26/215)	0.184 (31/168)	0.204 (20/98)	0.183 (38/207)
B2	0.052 (41/777)	0.125 (24/191)	0.159 (29/182)	0.134 (24/178)	0.132 (62/469)
C1	0.199 (42/211)	0.327 (19/58)	0.338 (23/68)	0.583 (14/24)	0.283 (53/187)

Source: author's data

Ellis and Ferreira-Junior's (2009) study sought to determine learners' level of reliance on the argument structure of verbs when learners produce specific argument structure constructions. The study showed that learners tend to be conservative in that they either opt for verbs that are semantically compatible with constructions (eg. *give* in the ditransitive construction) or they make use of semantically generic verbs (eg. *go* in the locative construction). In view of this, the ratio between constructions and the verbal variability in our data may be an inferential cue to the level of entrenchment of caused motions, since the more lexically varied learners' production is, the less likely it is that they are relying on the argument structure of specific verbs when producing caused motions. In other words, in scenarios where the relationship between constructions and verbs is closer to a one-to-one correspondence, the use of caused motions is less likely to be due to chance or to the mastery of specific verb semantics.

Such an observation seems to be plausible if we analyze the data in Table 7 vertically so as to verify developmental factors. Brazilian A2 learners, for example, display a ratio of 0.043 showing that there are 4 verbs to each 100 caused motions (4:100). This is a somewhat stable scenario for the other languages, which also show a varied increase towards C1. Nevertheless, at first glance, Spanish A2 level learner production seems to refute the analysis, given the high level of proportion between verbs and constructions at this stage (>0.2). Having said that, a closer look at the data as a whole suggests that the high ratio of Spanish A2 is likely to reflect a lack of instances of caused motions in the data in absolute terms. To put it differently, the fewer occurrences of caused motions we have, the higher the proportion between verbs and constructions will naturally be. This would also account for Spanish B1 0.120 featuring half the ratio of A2 (0.236), since Spanish B1 has three times as many occurrences as A2; thus, this dataset is, contrastively, more susceptible to statistical dispersion. All the same, the low absolute numbers in Spanish A2 are in line with the expectation one would have for low-level learners; in other words, when produced, caused motions seem to be marginal and likely to be a by-product of the lexical knowledge of certain verbs, as Ellis and Ferreira-Junior's (2009) findings pointed out.

In order to determine whether learner production of caused motions is the result of constructional knowledge or lexical knowledge of specific verbs, we looked specifically at the verbs in their production in light of the segmentation between *verbs of instantiation* and *verbs of modification*. The following section deals with such issues.

4.2 Instantiating and modifying verbs in learner data

Verb and construction ratios can shed some light on the matter of learner production of caused motions, but the type of verb used (*instantiating* or *modifying*) can be an insightful linguistic heuristic to determine whether or not caused motions are productive in learner language. This is due to the fact that, with verbs of modification, the constructional status of caused motions is evident, given that the directional PP is not predicted by the argument structure of the verb, hence being provided by the semantics of the construction (Goldberg, 1995, 2003, 2006, 2013). The sentences below, occurrences of caused motions produced by learners, exemplify both types of verbs.

(14) a. My career may be described with a headline: *turning ideas into highly effective operations*. (BR/B2)

b. You'll cut the lemons and *squeeze them into the cup* (BR/B1)

(15) a. Its not easy to buy green product because they are more expensive that others, but we should sacrifice some luxuries to buy and *incorporate them into our everyday lives* in order to produce a positive impact on the environment (SP/C1)

b. Any way, we have to follow rules and laws to *keep us out of problems*. (SP/B1)

(16) a. Don't *take the document out of the office* for respect the confidential information (ITA/A2)

b. …and *plunged the knife into his chest*, then he fell to the ground and she looked horrified. (ITA/C1)

(17) a. *Do not bring out a PC of the office*. - Do not have a lunch in a meeting room. (FR/A2)

b. The father took fingerprints of Tom few days before the murder then he *called him into the room*. (FR/B2)

(18) a. Each monthly payment would be like *putting money into a high interest account*. (GER/B2)

b. Other stories are depicting mermaid *squeezing the life out of drowning men* while these tried to rescue them. (GER/C2)

(14) to (18) represent the five groups of learners under scrutiny and they contain the two types of verbs discussed. After being isolated in one column (cf. Fig.5), the verbs were categorized as verbs of *instantiation* or *modification*. For such categorization, we used a test proposed by Goldberg (1995, pp. 43-44) to determine verbal meaning as well as the number and type of participant roles. The idea is to insert the verb into the gerundial structure "*No _____-ing occurred*" so as to interpret the roles implicitly thought to be involved in the frame. In (14a), for example, the simple "No *turn*ing occurred" could only have a "two-participant [change of direction] interpretation". However, with this interpretation, the THEME argument of (14a), that is, *ideas* would have to be left out of the structure in that *turn* would not select for this argument (eg. *…turn into effective operations*). The verb would, ultimately, have to be classified as a *modifying* verb, since it does not reflect the argument structure of caused motions. That is certainly not the

case of (14a) in which *turn* does mean *cause to become*. For these cases, Goldberg (1995) claims that complements must be included so as to accurately portray the intended scene. Thus, in this case, the gerundial structure would be *"No turning of ideas into operations occurred"* and the "three-participant interpretation" could be inferred. Below is the test applied to the verbs of sentences (14) to (18).

- (14a) - No *turning of something into something* occurred (three-participant [cause to become] interpretation)
- (14b) - No *squeezing* occurred (two-participant interpretation)
- (15a) - No *incorporating of something into something* occurred (three-participant interpretation)
- (15b) - No keeping occurred (two-participant [stay in position] interpretation)
- (16a) - No *taking something out of something* occurred (three-participant [cause to move] interpretation)
- (16b) - No *plunging* occurred (one-participant interpretation)
- (17a) - No *bringing something out of something* occurred (three-participant [cause to move] interpretation)
- (17b) - No *calling* occurred (one-participant interpretation)
- (18a) - No *putting something into something* occurred (three-participant [cause to move] interpretation)

The verbs in (a) were all classified as *instantiating verbs*, since their conceptual structures predict the realization of all the participant roles present in the sentences. In other words, the semantic requirements of these verbs are a perfect match for the constructional scheme X CAUSES Y TO MOVE Z. However, the same cannot be said about the verbs in (b), which were, in our analysis, classified as *modifying verbs*. *Squeeze*, *keep*, *plunge* and *call* do not contain any sort of motion feature in their conceptual structure. '*Squeezing something into the cup*' cannot be analyzed as a more specific type of '*squeezing*' like '*taking something out of something*' can be a more specific

kind of '*taking*'. In '*You'll cut the lemons and squeeze them into the cup*', then, the PP '*into the cup*' is provided by the semantics of the construction that licenses the expression and with which the verb *squeeze* is satisfactorily fused.

All the 5807 verb tokens, represented by 184 verb types, were isolated and analyzed as for their numbers of participant roles. A second verification for participant roles was carried out for predicates about which the gerundial test was not accurate. The second verification of verbal valency was done on *FrameNet* (by checking the *core* and the *peripheral* FEs (frame elements)).

Frame element	Core type	Verb: divide
Agent	Core	"**We** can DIVIDE the chores between us ," said Dr Maingay with enthusiasm
Cause	Core	**the social alienation** that often DIVIDES cancer patients from
Parts	Core	Rats were DIVIDED **into a control group** (group A)
Place	Peripheral	DIVIDE richer **from poorer peasants**
Recipients	Extra-thematic	"We can DIVIDE the chores **between us** ," said Dr Maingay with enthusiasm

Below is an example of one of the verbs analyzed.

Figure 7. FrameNet entry for *divide.*

Source: available in https://framenet.icsi.berkeley.edu/fndrupal/luIndex

Therefore, the criterion to define items as *instantiating* or *modifying* verbs was whether or not verbs specified the same number and types of participant roles as the caused-motion constructions. Verbs with one-participant interpretation like *talk*, *walk*, *laugh* and two-participant interpretation like *kick*, *scare*, *let* and *type* were classified as *modifying* verbs. Verbs of three-participant interpretation like *bring*, *place*, *get*, *send*, etc. were all classified as *instantiating* verbs.

On the role of verb types on learners' performance, the analysis sought to verify how representative both types of verbs were in each group of language and at each proficiency level. Instantiating verbs were intuitively expected to be more frequent than modifying verbs across the board, but, conversely, the use of modifying verbs was expected to increase as the levels of proficiency advanced, from A2 to C1; this is associated with the idea that the more modifying verbs are used, the more entrenched caused motions seem to be in learners' construction. Table 8 presents the distribution of both types of verbs.

Table 8. Instantiating and modifying verbs per language and level of proficiency.

CEFR levels		Portuguese	Spanish	Italian	French	German
A2	Instantiating	97%	91%	99%	98%	88%
	Modifying	3%	9%	1%	2%	12%
B1	Instantiating	94%	95%	94%	91%	74%
	Modifying	6%	5%	6%	9%	26%
B2	Instantiating	91%	95%	90%	91%	44%
	Modifying	9%	5%	10%	9%	56%
C1	Instantiating	73%	64%	75%	71%	66%
	Modifying	27%	36%	25%	29%	34%

Source: author's data

An observation of the data from a horizontal perspective shows that German (the typologically related L1) outperforms speakers of Romance languages (the target group of learners) in the use of modifying verbs, with B2 German caused motions with verbs of modification being ten times as frequent as the Spanish B2 group, for example. However, when the level of proficiency reaches C1, the figures stabilize across L1s with Spanish L1 speakers mildly outperforming the control group of German learners. This seems to suggest that, despite the typological differences between English (the target language) and the group of Romance languages, the variable *level of proficiency* seems to be more determining than L1 background in the use of caused motions. The vertical inspection of the figures endorses such an analysis since it shows a rather steady increase in the number of caused motions with modifying verbs with all learners reaching similar percentages in the use of caused motions with modifying verbs.

Overall, the descriptive statistical data discussed thus far seem to endorse that *learner (L1 Romance/L2 English) production of caused motions is affected by the role of the verb*, since *instantiating* verbs are demonstrably more frequently used across languages and proficiency levels and *modifying* verbs seem to cluster around the end of the proficiency scale with figures reaching a solid 36% (Spanish C1) and 29% (French C1).

Conclusions

In this article, we have analyzed the uses of caused-motion constructions by learners of four different Romance languages, namely Brazilian Portuguese, Spanish, Italian and French. The proposed analysis aimed at showing learner production of English caused motions and, to that end, we have looked at the relationship between verbs and constructions, firstly by looking at the verbal lexical variability in the verb-construction ratio and then by probing into the segmentation between instantiating and modifying verbs. The idea of using different types of verbs meant to provide an inferential answer to the question: *do foreign language learners have L2 constructions as abstract as the caused-motions construction?* But it also provided information on the interplay between constructional complexity and proficiency levels since the analyzed data showed that the use of caused motions with verbs of modification both increased as the proficiency levels rose before stabilizing at C1 irrespective of the L1 background. This suggests that learners seem to rely a lot more on the role of verbs than on that of schematic constructions as the study by Ellis and Ferreira-Junior (2009) had already showed.

Although this study is not centered on teaching applications, its findings may contribute to this area in that they provide interesting aspects to be empirically tested in the classroom environment and applied in foreign language classes. Since learners seem to rely on the lexical knowledge of instantiating verbs, these could be targeted by the teacher as a way to develop more schematic knowledge of caused-motion constructions, especially those involving modifying verbs like *talk, walk, laugh*, etc. Also, learners' processing of such constructions must be investigated in future studies in order to determine whether the low figures witnessed in the corpus analysis will also be present in the processing data. This can also be used to inform the teaching of these constructions.

Informações complementares:

a) Declaração de contribuição das autoras e dos autores:

O planejamento e a escrita foram feitos pelo autor.

b) Disponibilidade de dados de pesquisa e outros materiais:

Os dados públicos que apoiam as conclusões deste estudo - Somewhere in between grammar and lexis: the role of verb types in learner production of caused-motion constructions - estão, respectivamente, disponíveis em corpus.mml.cam.ac.uk e https://www.english-corpora.org/coca/ .

c) Declaração de conflito de interesse:

Declaro não haver conflitos de interesse.

d) Avaliação por pares:

✓ **Avaliador 1:** Deise Prina Dutra (correções obrigatórias)

The text presents an interesting issue (caused-motion construction in a learner corpus) and the study design is well done. I suggest a few revisions before it is published.
1. Section 2: "researchers working at the interface between cognitive and corpus linguistics have drawn strong correlations between the mental representation of language patterns and text frequency."
2. Section 3 - corpus description: Why does the author refer to "scripts" and not "texts" ? For example, in the sentence " The corpus data contain a total of 1,180,310 scripts and 7,126,752 sentences produced by learners with a wide range of L1 backgrounds. " Could "scripts" be called "texts"?
If you choose to use "script", it is important to explain why. This is not
3. Use of language "all things considered" appear twice in the text. As a low frequency expression in academic texts (see BNC and COCA), it would be better to substitute, at least one of the occurrences, by a more frequent expression in academic genres.
4. Insert commas after "analysis" and "reading": " Another important aspect of the analysis, which was also meant to help decide on the constructional semantic reading, was to isolate the verbs of each sentence.
The comment below Chart 3 is hard to understand since there is no reference to the CEFR levels in the chart and the paragraph starts with this sentence "The chart above clearly shows that there is relatively mild development in the use of the structure as the levels of proficiency increase. " The paragraph needs to be rewritten and/or the chart has to be changed.
5. Check spelling: Paragraph below Table 2 - "por example" rather than "for example"

✓ **Avaliador 2:** Maria José Finatto (aceitar)

O trabalho é excelente e merece ser publicado. A revisão da literatura foi bastante pertinente tendo em vista o objetivo do trabalho. Recomendo que os autores possam considerar, incluir, ainda, alguma referência, em Linguística de Corpus, sobre o tema dos lexical bundles, dado que o título instaura algo entre o léxico e gramática. Aspecto a ser retomado no fechamento. Essa seria a única restrição para, salvo melhor juízo, tornar o trabalho 100% pronto para publicação..

Referências

BAICCHI, A. **Construction learning as a complex adaptive system**: psycholinguistic evidence from L2 learners of English. Berlin: Springer, 2015.

BYBEE, J. L. From usage to grammar: the mind's response to repetition. **Language**, v. 82, p. 711-733, 2006.

BYBEE, J. L. Usage-based theory and exemplar representations of constructions. In: TROUSDALE, T.; HOFFMANN, G. (Eds.). **The Oxford handbook of construction grammar**. New York: Oxford University Press, 2013. p. 49-69.

CABRERA, M.; ZUBIZARRETA, M. L. Constructional properties versus lexical specific transfer: overgeneralized causatives in L2 English and L2 Spanish. **BUCLD 28 Proceedings Supplement**, 2004.

ELLIS, N. C. Second language acquisition. In: TROUSDALE, T.; HOFFMANN, G. (Eds.). **The Oxford handbook of construction grammar**. New York: Oxford University Press, 2013. p. 365-378.

ELLIS, N.; FERREIRA-JUNIOR, F. Constructions and their acquisition: islands and the distinctiveness of their occupancy. **Annual Review of Cognitive Linguistics**, v. 7, p. 187-220, 2009.

EVANS, V.; GREEN, M. **Cognitive linguistics**: an introduction. Edinburgh: Edinburgh University Press, 2006.

FILLMORE, C. J. "Corpus linguistics" vs. "computer-aided armchair linguistics". In: **Directions in corpus linguistics**: proceedings from a 1991 Nobel Symposium on Corpus Linguistics. Berlin: Mouton de Gruyter, 1992. p. 35-66.

GOLDBERG, A. E. **A construction grammar approach to argument structure**. Chicago: The University of Chicago Press, 1995.

GOLDBERG, A. E. **Constructions at work**: the nature of generalization in language. Oxford: Oxford University Press, 2006.

GOLDBERG, A. E. Constructionist approaches. In: TROUSDALE, T.; HOFFMANN, G. (Eds.). **The Oxford handbook of construction grammar**. New York: Oxford University Press, 2013. p. 15-32.

GOLDBERG, A. E.; CASENHISER, D.; SETHURAMAN, N. A lexically based proposal of argument structure meaning. **Chicago Linguistic Society**, 2003.

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

GRANGER, S. Learner corpora. In: LÜDELING, A.; KYTÖ, M. (Eds.). **Corpus linguistics**: an international handbook. v. 1. Berlin: Walter de Gruyter, 2008. p. 259-275.

GRIES, S. Th.; WULFF, S. Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. **Annual Review of Cognitive Linguistics**, n. 3, p. 182-200, 2005.

HALLIDAY, M. A. K. Quantitative studies and probabilities in grammar. In: HOEY, M. (Ed.). **Data, description, discourse**: papers on the English language in honour of John Sinclair. London: HarperCollins, 1993. p. 1-25.

LANGACKER, R. W. **Essentials of cognitive grammar**. New York: Oxford University Press, 2013.

LINDQUIST, H. **Corpus linguistics and the description of English**. Edinburgh: Edinburgh University Press, 2009.

MCENERY, T.; HARDIE, A. **Corpus linguistics**. Cambridge: Cambridge University Press, 2012.

SCHMID, H. **English abstract nouns as conceptual shells**: from corpus to cognition. Berlin: Mouton de Gruyter, 2000.

SCHMID, H. Does frequency in text instantiate entrenchment in the cognitive system? In: GLYNN, D.; FISHER, K. (Eds.). **Quantitative methods in cognitive semantics**: corpus-driven approaches. Berlin: De Gruyter Mouton, 2010. p. 101-135.

SINCLAIR, J. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.

SVARTVIK, J. **Directions in corpus linguistics**: proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991. Berlin: Mouton de Gruyter, 1992. (Trends in Linguistics Studies and Monographs, 65).

TEUBERT, W. Corpus linguistics: an alternative. **Semen**, n. 27, p. 130-152, 2009.

ZUBIZARRETA, M. L.; OH, E. **On the syntactic composition of manner and motion**. Massachusetts: The MIT Press, 2007.