DOI: http://dx.doi.org/10.5007/2175-8026.2023.e94720

THE PROCESSING OF ENGLISH CAUSED-MOTION CONSTRUCTIONS BY EFL LEARNERS

Rodrigo Garcia Rosa[1*]
[1]Universidade de São Paulo, São Paulo, SP, Brasil

Abstract

This article investigates the processing of English caused-motion constructions by Brazilian EFL learners. Drawing on corpus data on the production of such constructions by Brazilian learners, this study experimentally tests the four following constructional domains of caused motions discussed in Rosa (2020): (1) literal caused motions with instantiating verbs (e.g., *put the toys into the box*); (2) figurative caused motions with instantiating verbs (e.g., *get yourself in trouble*); (3) literal caused motions with modifying verbs (e.g., *they laughed him out of the office*); and (4) figurative caused motions with modifying verbs (e.g., *she talked me into stupor*). An acceptability judgment task with 120 EFL Brazilian learners at two levels of proficiency (B2 and C1) was devised so as to investigate the development of caused motions by testing two aspects and their effect on the comprehension of the structure: (1) level of proficiency, and (2) the degree of linguistic complexity of the structures in question. The results of the experimental study showed that learners' proficiency level does affect their comprehension of the structures. However, the linguistic complexity of the structures seems to play an even greater role in that learners exhibit a descending level of comprehension of caused motions that is proportional with the ascending level of semantic and syntactic complexity of the structures.

Keywords: Language processing, EFL learners, caused-motion constructions, acceptability judgement task.

[*] Professor of English at the department of Modern Languages of the Universidade de São Paulo. Email: rodrigo.rosa@usp.br ORCID: 0000-0002-1538-1542.

1. Introduction

1.1 The integration of corpus studies and experimental research

Corpus linguistics integrates a family of language theories and methods that sees usage events as the parameters against which theories should be tested and developed. In view of this, corpus-based studies adopt an empirical perspective, which, by and large, is opposed to the idea of using introspection in the judgment of linguistic expressions as starting points of observation (McEnery & Hardie, 2012). This general refutation to subjective introspection is justified in that corpus studies are a lot more centered on what is *probable* and a lot less on what is *possible* in language structure (Berber Sardinha, 2004).

Nevertheless, within the broader area of corpus studies, research is commonly classified into two subareas depending on how corpora data are viewed; that is, studies using corpora as the only reliable source for the analysis and explanation of language behavior are commonly referred to as *corpus-driven*. For corpus-driven studies, the adoption of independent linguistic theories is not the usual practice, as this might jeopardize the reality of the facts given the usual adoption of pre-corpus concepts and principles. Such concepts and principles are said to affect, influence and, ultimately, bias the analysis. In a different group are studies that view corpus linguistics as an empirical methodology in which researchers can find the empirical support for theoretical explanations and generalizations about the structure of languages, language varieties, language change, etc. These are usually categorized as *corpus-based* studies. For corpus-based linguistics, corpora analyses can greatly assist the language analyst to investigate and understand facts about language while providing them with quantitative data that serve as the empirical basis upon which theories can be refined, reformulated or even refuted. For many of the linguists who subscribe to this perspective in corpus studies, corpus linguistics can be defined as a methodology and not as a theory. Lindquist (2009, p. 1) claims that "corpus linguistics is thus a methodology, comprising a large number of related methods which can be used by scholars of many different theoretical leanings".

According to McEnery and Hardie (2012), most recently the dialogue between corpus studies and other linguistic approaches, initiated by those categorized under the corpus-based umbrella term, gained traction and favored the collaboration between corpus-based studies and certain areas of psycholinguistics, especially those of an experimental nature. Such a dialogue became feasible as both language approaches are based on very similar linguistic-philosophical principles, that is, corpus linguistics and experimental studies both rely on empirically quantified data aiming at language generalizations and explanations. In practical terms, the main differences between the approaches in question boil down to the types of research conducted in each area (Gilquin & Gries, 2009). While many of the studies in corpus linguistics are concerned with the occurrence of linguistic expressions in specific genres of discourse,

terminological analyses, lexicographic and translation studies, experimental research is mainly concerned with areas related to language learning, L1 and L2 acquisition and language processing. Having said that, for some corpus-based linguists (Gilquin & Gries, 2009; Gries & Wulff, 2005; McEnery & Hardie, 2012), both approaches can complement each other in interesting ways, generating more comprehensive and reliable explanations for an array of linguistic phenomena. This is possible, according to these studies, because the main limitations of one approach can be complementarily tackled with the aid of the other.

It is well known that electronic corpora tools and analyses have provided researchers with unmatched quantitative approaches and techniques, thus enriching the investigation of all sorts linguistic phenomena. However, certain aspects of language processing, such as cognitive processes underlying online language production, for example, will hardly ever be captured in corpora, even if exhaustive verifications of millions of texts are done with that purpose.

Two other constraints of corpus studies, especially those concerning learner corpora are highlighted by Gilquin and Gries (2009) and are of special significance to our study. These limitations are:

i. not always can corpus studies reveal which language constructions are more or less easily processed by speakers (native or nonnative), since language production entails language processing, but the opposite is not necessarily true. That is, it is not plausible to claim that only because X is processed, X will necessarily be produced, especially in L2 settings;

ii. not always are language corpora the best sources of data, especially if one is interested in studying language patterns with low discursive salience (low frequency of occurrence).

In order to illustrate, but also reinforce the second point raised above, we evoke a maxim in experimental studies when these are compared to observational data. Nevertheless, in place of declaring it ourselves, let a corpus linguist do that.

> […] unless the corpus represents the whole population, **the absence of evidence is not the evidence of absence**. In other words, if an expression does not appear in a corpus, this doesn't mean that this expression is non-existent. (Brezina, 2018)

Especially for those concerned with the acquisition and the teaching of foreign languages, the points raised above are of great relevance. The first one deals with a *sine qua non* condition for language production, since learners will only be able to produce constructions if they are first able to process them. On this issue, Ionin and Zyzik (2014, p. 37) state that

> "[t]he fact that a learner has not produced a particular expression, or has made a production error, does not necessarily reflect a lack of linguistic knowledge. Other factors, such as avoidance, phonological complexity, or difficulty with retrieval from memory (to name but a few) may be responsible. Conversely, the production of certain frequent formulaic expressions may reflect rote memorization rather than linguistic knowledge. For these reasons (among others), SLA researchers may use other data collection tools to study learners' linguistic knowledge." (our highlight)

As for the second point, analyzing constructions with low discursive salience in learner corpora might not bring the expected results. As discussed in Rosa (2020), L2 learners are especially sensitive to aspects of the target language in ways that L1 acquirers are not. Furthermore, Ellis (2013) mentions the importance of *input frequency*, *discursive salience* and its relation to *perception* as key ingredients for the entrenchment of L2 constructions. That said, the points in (i) and (ii) above are intertwined in that constructions with low discursive salience, the case of caused-motions constructions, will have low frequency. This factor in turn may compromise learners' perception of the item when it occurs, thus making the learning of these constructions harder.

The aspects outlined thus far, as well as the general low frequency of caused motions analyzed in the learner language (Rosa, 2020), are discussed in more depth below.

1.2 Learner production of caused-motion constructions: from observation to experimentation

Rosa (2020) dealt with the observation of learner data in the EFCamDAT[1] corpus by investigating the use of four constructional domains of caused motions (see examples below) by learners of four L1 Romance languages, namely Brazilian Portuguese, Spanish, Italian and French.

(1) Caused motions with *instantiating* verbs denoting *literal* movement.
e.g., If the wingers *get the ball into the box* for him in the right areas, he will score goals (COCA/Blog/2012)

(2) Caused motions with *instantiating* verbs denoting *figurative* movement.
e.g., *Get your child into the habit* of keeping his hands away from his cold sore (COCA Magazine/2000)

(3) Caused motions with *modifying* verbs denoting *literal* movement.
e.g., Eventually he *laughed me right out of the office*. (COCA/Spoken/2014)

(4) Caused motions with *modifying* verbs denoting *figurative* movement.
e.g., They branded him a cowardly bureaucrat and *laughed the project out of existence*. (COCA/TV/1998)

The analysis aimed at showing developmental factors in the use of caused motion events as well as cross-linguistic comparisons amongst the four

targeted groups of learners. From a developmental perspective, the data showed that learners' use of the constructions is affected by the types of verbs used (*instantiation* vs. *modification*) and constructional reading (*literal* vs. *figurative*), but they tend to get more salient towards the end of the proficiency continuum, as was expected, both generally and by the hypotheses of the study. From a cross-linguistic perspective, the analysis showed that the four L1s behaved quite similarly, since no Romance language stood out from the remaining ones as for the two variables analyzed. Although the analysis was quite straightforward vis-à-vis its objectives and results, one of the four constructional domains discussed (example 4), composed of figurative caused motions containing verbs of modification, did not behave as expected in spite of the theoretically viable predictions for the domain.

The analysis of the learner corpus aimed at probing into learners' *production* of caused-motion constructions as for general frequencies of occurrence (thus investigating productivity), as well as from a more qualitative point of view; that is, we also looked at the types of verbs occurring in the verbal slots of constructions (*instantiation* vs. *modification*) and the constructional reading of learners' production (*literal* vs. *figurative*) for the directional PPs (e.g., 'into the box' in (1)). The data analyzed showed that:

i. *Instantiating* verbs are generally more frequently used than *modifying* verbs;

ii. *Modifying* verbs are used more frequently in more advanced levels;

iii. *Literal* caused motions outnumber *figurative* caused motions across levels of proficiency;

iv. *Figurative* caused motions are more salient towards the end of the proficiency scale.

Although the results above did confirm our two hypotheses for learner production, that is, that there was an ascending level of difficulty through the domains, the analysis correlating both variables, which generated four constructional domains (repeated below), did not confirm our unified hypotheses that "learner production of caused motions is affected by the different types of constructions (1, 2, 3 and 4) and this effect follows an ascending level of constructional complexity from (1) to (4)".

(1) Caused motions with *instantiating* verbs denoting *literal* movement.
e.g., If the wingers *get the ball into the box* for him in the right areas, he will score goals (COCA/Blog/2012)

(2) Caused motions with *instantiating* verbs denoting *figurative* movement.
e.g., *Get your child into the habit* of keeping his hands away from his cold sore (COCA Magazine/2000)

(3) Caused motions with *modifying* verbs denoting *literal* movement.
e.g., Eventually he *laughed me right out of the office*. (COCA/Spoken/2014)

(4) Caused motions with *modifying* verbs denoting *figurative* movement.
e.g., They branded him a cowardly bureaucrat and *laughed the project out of existence*. (COCA/TV/1998)

According to the data analyzed, learners' performance in constructional domain 4 was not as low as expected, in comparison with the other domains (63% of all caused motions at Italian C1 level were of the fourth type). A qualitative look at the data containing figurative caused motions with verbs of modification suggested a possible explanation for the quantitative discrepancy with the expected outcome: the role of conventional phrases - memorized phraseologisms and/or fixed expressions lifted from the tasks learners were exposed to (Ionin & Zyzik, 2014).

All things considered, the results for our four constructional domains, especially domain 4, called for an experimental intervention given that a correlation between learners' descending performance and the ascending constructional complexity from domains (1) to (4) can be attested from the perspective of language comprehension. Therefore, in this part, an acceptability judgment task was devised to test the following hypothesis:

> Learner comprehension of caused motions is affected by the different types of constructions ((1), (2), (3) and (4)) and this effect follows an ascending level of constructional complexity from (1) to (4). Thus, B2 and C1[2] level learners are expected to have a descending level of recognition that is proportional with the ascending level of constructional complexity of each domain.

To empirically test our hypothesis, experimental intervention was needed and an acceptability judgement task which involved 120 Brazilian EFL learners distributed in four groups (2 groups of 30 B2-level learners and 2 groups of 30 C1-level learners) was devised. Below we detail the experimental design.

2. Experimental design: aims and objectives

2.1 The task

In spite of being traditionally referred to as *grammaticality judgement tasks*, the term is somewhat misleading even in formal linguistics, claim Schütze and Sprouse (2014), since *grammaticality* is believed to be an internal property of language knowledge, which judgement tasks are meant to skillfully capture and take snapshots of. In other words, one cannot reply to something that is internal and unconscious, which is what grammar is thought to be like in formal linguistics. The most appropriate term to use would be *acceptability judgment task*, inasmuch as the fact that reaction of this nature has to do with *agreeing*,

endorsing or, as the specialized literature calls it, *accepting* if something is *natural*, *possible*, *correct*, *appropriate*, etc. However, *acceptability* is itself a *percept* much like *brightness*, *loudness, temperature* and *pain* are (Schütze & Sprouse, 2014, p. 28). So, the linguist's job is to indirectly assess such a percept by asking subjects to judge how acceptable language patterns are. So, acceptability judgement tasks are, in fact, a *reported perception of acceptability* (Sprouse & Almeida, 2013), thus a deductive method for observing speakers' knowledge of language. In other words, they aim to measure what speakers think of what they tacitly know *about* and *of* language.

Schütze and Sprouse (2014) list some of the practical advantages of using acceptability judgment tasks. These are:

i. They provide evidence for phenomena that rarely occur in spontaneous speech;

ii. They do not contain normal speech production errors;

iii. The researcher does not need some highly technological apparatus for the task.

However, one of the main concerns of the researcher working from an experimental paradigm must be the careful design of the experiment in order to minimize the effect of natural extraneous variables that might interfere in the result. That said, things related to the *type* of test, the *instructions* given, the *materials* used, the *experimental items* and the *sample size*, are of extreme relevance since these will guarantee the validity of the results obtained. Below we present each of these for the experiment we conducted.

2.2 Participants

The corpus analysis was meant to check learners' production of the target constructions, but also served to define the profile of the experiment participants.

In view of the fact that no significant cross-linguistic differences were felt among the four targeted groups of Romance languages (Brazilian Portuguese, Spanish, Italian and French), we did not feel compelled to apply the experimental task to all four language groups. For that reason, our experiment was restricted to the group of learners most widely represented in our corpus analysis, Brazilian EFL learners. We believe the results on comprehension from this group could be extended to how speakers of other Romance languages would react to the stimuli for two main reasons: 1) performance in the use of caused motions by speakers of Romance languages among the four constructional domains followed, by and large, very similar routes of descending performance; but 2) Brazilian Portuguese speakers were the only group who behaved in accord with the expected outcome for constructional domain 4. Thus, whatever the experimental results for Brazilian

learners are, it seems to be plausible to expect, based on the observational data analyzed, that the remaining speakers would perform similarly to or better than our experimental group.

As for participants' proficiency levels, we decided to restrict our experimental groups to the highest levels of analysis in our corpus investigation: B2 and C1 levels. This way, the experiment could shed some light on learners' comprehension of the constructions; but the comparison across such levels could also reveal features of development and the extent to which proficiency affects the comprehension of the constructions under investigation. Since we aimed at carrying out the experiment on an online platform for surveys and experiments[3], we established 70 learners per language and proficiency level so as to anticipate and minimize data that, in the end, could not be used in the analysis. We managed to collect the responses of 162 participants (81 of each level) and, after eliminating incomplete responses and some outliers[4], the experimental group came down to a total of 120 participants, 2 groups of 30 B2 and 2 groups of 30 C1 level participants.

The participant recruitment was done with students from a language institute and, for the selection of participants' proficiency levels, we used the leveling criterion of the institute, which uses the CEFR. For B2 learners, we recruited participants doing an upper-intermediate course and for the C1 level, advanced-level learners were invited to take part in the experiment[5].

In the next section, we present the experimental items and discuss their composition.

2.3 Experimental items

The items were created and divided into two types vis-à-vis our hypothesis: type of verb (*instantiation* or *modification*) and interpretation (*literal* or *figurative* motion). No syntactically ungrammatical/unacceptable sentence containing the target language was included, since the aim of the experiment was to show how close to "natural" participants would rate the items and the extent to which these would be affected by the grammatical properties of these sentences.

The instantiation verbs used in the sentences were extracted from Hampe's (2010) corpus study on the most frequent verbs used in caused-motion constructions in the ICE corpus. The modification verbs used in the stimuli were also validated on the COCA corpus, but most of them were taken from publications on caused-motion constructions (Goldberg, 1995; Goldberg and Jackendoff, 2003; Cabrera & Zubizarreta, 2004; Rappaport Hovav & Levin, 1995).

In the item composition process, sentences were controlled for: 1) number of words (each item has 15 words); 2) number of syllables (Mean= 20.3); and 3) frequency of the verb on COCA (among the first 500 most frequent verbal lemmas). The critical elements in the sentences (i.e., Verb + Object + Oblique Argument) were also controlled for and the vast majority is composed of 7 words, ranging from 6 to 10 words. The acceptability of the target items and the different interpretations (*literal* or *figurative*) of the sentences were validated with 5 native

speakers of English and the critical elements were all checked for occurrence in the COCA corpus.

The experiment is composed of 48 target items (12 sentences per domain) and 30 distractor sentences. According to Keating & Jegerski (2015), a rule of thumb on the number of distractors is to include 75% of fillers in relation to the target stimuli. In this case, 30 sentences account for 75%. Fillers with off-target language were included to remove bias and to stop participants from discovering the main focus of the experiment. The fillers show 4 types of constructions: 1) Relative Clauses; 2) Passives; 3) Caused motions with unaccusative verbs (=unacceptable); and 4) Intransitive motion constructions. 70% of the fillers (21 sentences) are unacceptable. These were included in order to stop learners from generalizing that all sentences were acceptable.

Below we present the target items divided into the four constructional domains.

Domain A: caused motions with instantiation verbs denoting literal movement

1. You need to *put the mixture directly into the milk* while it is still hot.

2. Sam wants to *bring more young people into our discussion group* before it gets full.

3. We have to *get the wine out of the fridge* before it gets too cold.

4. She will *take those silver rings out of her nose* because her dad hates them.

5. They plan to *return all the policemen to the streets* to reduce the crime rates.

6. You have to *send this pack out of the country* using a special delivery service.

7. I can't *carry these kinds of metal objects into the airport* without declaring them first.

8. The coach will *lead the winning team onto the field* to have the celebration party.

9. You should not *throw the sliced vegetables into the pan* after the soup is hot.

10. You need to *drop the keys in the blue box* before you leave the room.

11. You need to *pull these plants out of the ground* before they destroy the vegetables.

12. They will *push the chairs out of the way* for us to use this space.

Domain B: caused motions with instantiation verbs denoting figurative movement

1. We need to *put this political crisis to an end* before it is too late.

2. You should always *bring the issues to my attention* before you go and make decisions.

3. The results of this investigation can *get the president in trouble* in the near future.

4. This government can *take the country into another economic depression* in less than a year.

5. Peter and Liz want to *turn their plans and ideas into reality* with our help.

6. This project will *send all the team into despair* because of its size and complexity.

7. We really need to *carry the new plans into effect* in time for the holidays.

8. Education can always *lead people out of the darkness of ignorance* no matter the age.

9. These actions can *throw the population into confusion* because they aren't clear or well defined.

10. We should *drop the problems out of our consciousness* if we want to solve them.

11. The policeman could *pull the girl out of danger* before the animal could reach her.

12. These bad decisions can *push many small businesses into financial trouble* in the near future.

Domain C: caused motions with modifying verbs denoting literal movement

13. They will *laugh me and my team out of the office* if we present this.

14. This new policy may *run many good professionals out of the country* in a month.

15. You should *work the butter and the milk into the eggs* by using a fork.

16. You must *play the ball off your right foot* to reduce the number of mistakes.

17. They only need to *speak some words into the microphone* and the spectators go crazy.

18. The magician says he can *talk the hat off your head* with his magic tricks.

19. You must *breathe lots of air into the lungs of the patient* to save him.

20. Patients *cry themselves into the emergency room* when they are in a lot of pain.

21. This circus artist can *sneeze spaghetti out of his nose* in front of the audience.

22. One parent can't *drive the kids out of the country* without the other parent's permission.

23. I had to *scream the kids back into their homes* because of the heavy rain.

24. She was able to *walk all the horses into the stables* before it started raining.

Domain D: caused motions with modifying verbs denoting figurative movement

25. My kids always *laugh themselves into exhaustion* when they see puppets and clowns on TV.

26. These decisions can *run the healthcare system into the ground* in one month or two.

27. Scientists can *work themselves into severe mental breakdowns* if they do not get enough rest.

28. After football matches some guys *drink themselves into madness* on the streets around the stadium.

29. You can *speak your problems into existence* when you talk about them all the time.

30. The doctors need to *talk some sense into her mind* before she makes more mistakes.

31. This method teaches you to *breathe yourself out of fear and anxiety* in 8 minutes.

32. At night some kids *cry themselves into exhaustion* before they finally stop and fall asleep.

33. I literally *sneeze myself into a terrible headache* when I eat anything with black pepper.

34. Paul always *drives his parents to desperation* when he behaves the way he did yesterday.

35. You cannot *scream yourself out of trouble* if you have a difficult problem to solve.

36. Doctors say people can *walk themselves out of their bad mood* if they walk regularly.

2.4 Procedure

The experimental items above were divided into two tasks of 24 target items each, with 6 items per domain (task A and task B), plus the number of distractors (=18). The division into two tasks was meant to reduce the total number of items per experiment, but also served as a way of relexicalizing the items, thus isolating effects of vocabulary knowledge on learners' responses. Table 1 shows the distribution.

Table 1: Distribution of participants per task

	B2	**C1**	**Total**
Test A	30	30	60
Test B	30	30	60
Total			120

Source: the author

The test was conducted online with the platform Qualtrics. First, participants were provided with some information about the test and the context of its application. Then, a page to collect their authorization to use the data appeared. If participants did not agree to have their responses used in the research, the system would automatically direct them to another page thanking them for participating. Next, language proficiency level was checked (upper-intermediate or advanced) before the first question appeared.

Participants were asked to choose numbers (1-5) on a Likert scale going from "Not Natural (1)" to " Natural (5)". An additional option "I don't know" was included and these answers were disregarded in the analysis. The task

design was based on Suzuki et al. (2016). The experimental items were randomly presented (target items of the four domains and distractors). Participants were not allowed to go back to previous responses, skip sentences or choose more than one option per item. All of these presentation features were automatically programmed on Qualtrics.

The test was untimed and 50% of the participants did it in an unsupervised manner[6], but instructions demanded they did not use any material for consultation during the execution of the task and they were asked not to take longer than 20 minutes to perform it[7]. The next section presents the results and the discussion.

3. Results

The four constructional domains (A-D) were isolated and participants' ratings were computed into a spreadsheet (from 0 (="I don't know") to 5 (="Natural)). The conventional parametric statistics treatment advises that, in Likert scale tasks, the analyst must calculate the *mean* of each participant, as well as its standard deviation. These two values are meant to be computed so as to obtain participant's z-score for each response. The z-score represents a standardized response of participants as each one of them is expressed in standard deviation units from the participants' mean (Schütze & Sprouse, 2013, p. 43). The z-score transformation is a standardization process that removes possible distortions within the data. However, parametric tests are only necessary for continuous data, not categorical data. In our case, each of the constructional domains was grouped and seen as an "island" for comparison between B2 and C1 levels. This way, the data is categorical, rather than continuous, and does not require a parametric treatment. Thus, our aim was basically to obtain a mean of each participant's response, from which we calculated another mean. This latter value would represent the level's response to the sum of experimental items contained in that particular domain. For example, in Task_ B_CMIL (literal caused motions with instantiating verbs), the mean of B2's response was 4.19 and C1 4.4. This shows, as expected, albeit timidly, the improved performance of C1-level learners over B2 participants.

In the next sections we compare both levels of proficiency by constructional domain before crossing the variables and levels.

4. Discussion

4.1 Literal caused motions with instantiating verbs

This constructional domain (A) was expected, according to our central hypothesis, to be the least complex group for learners to process. As the experimental items below demonstrate, the argument structure of verbs reflects the construction's argument structure. Also, the high frequency of the verbs selected for these items reinforces the thesis that these are familiar lexical

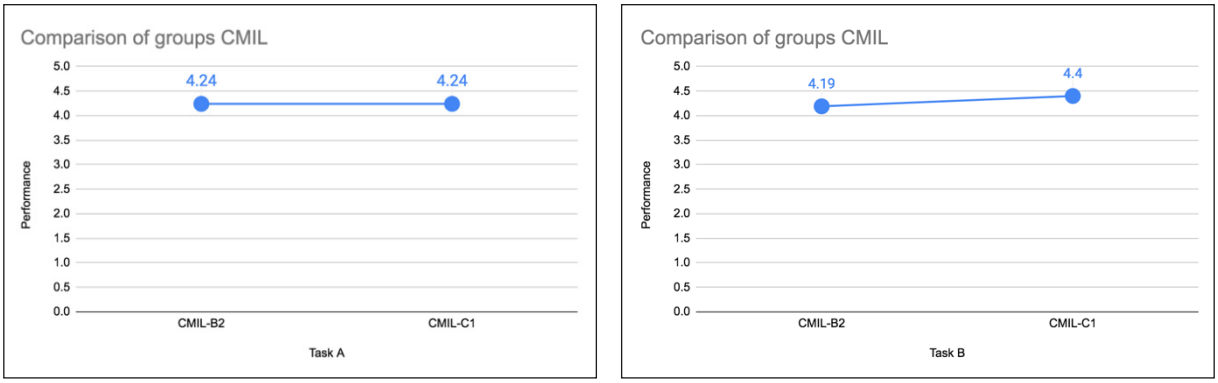
constructions, especially at B2 and C1 levels. The items are repeated below for convenience.

Domain A: CMIL

- You need to *put the mixture directly into the milk* while it is still hot.
- Sam wants to *bring more young people into our discussion group* before it gets full.
- We have to *get the wine out of the fridge* before it gets too cold.
- She will *take those silver rings out of her nose* because her dad hates them.
- They plan to *return all the policemen to the streets* to reduce the crime rates.
- You have to *send this pack out of the country* using a special delivery service.
- I can't *carry these kinds of metal objects into the airport* without declaring them first.
- The coach will *lead the winning team onto the field* to have the celebration party.
- You should not *throw the sliced vegetables into the pan* after the soup is hot.
- You need to *drop the keys in the blue box* before you leave the room.
- You need to *pull these plants out of the ground* before they destroy the vegetables.
- They will *push the chairs out of the way* for us to use this space.

The results for constructional domain A are presented in the Figures below.

Figure 1: Comparison of domain A at levels B2 and C1.



Source: the author

Since Tasks A and B were composed of items alternately taken from the sum of items in domain A, the difference between responses may be the result of different levels of familiarity with the grammar of specific verbs composing the task in question. In the left-hand graph above (Task A), no difference at all was observed between B2 and C1, whereas in the graph on the right-hand side (Task B), C1 is shown outperforming B2 as expected.

4.2 Figurative caused motions with instantiating verbs

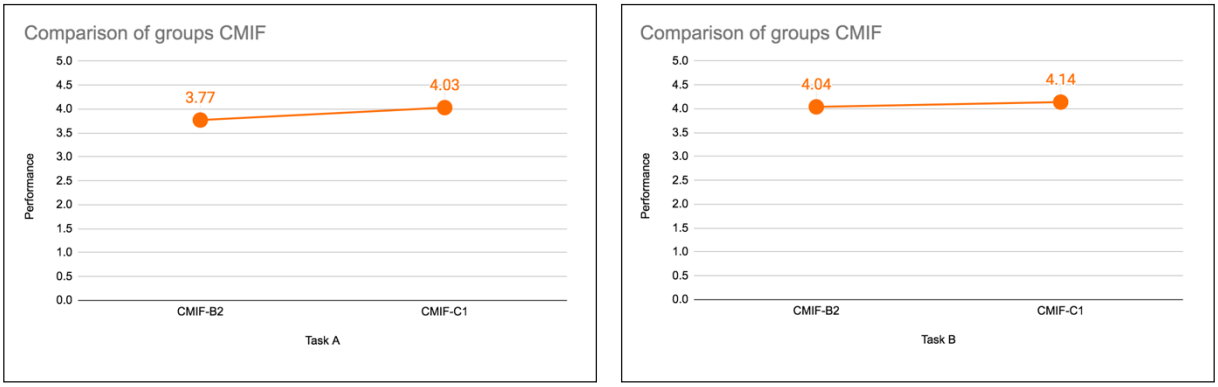
Differently from the results above, the second group involves the *figurative reading* variable, which adds an extra layer of complexity for learners to handle. The oblique argument is construed as a figurative space from or to which the events cause the THEME to dislocate. In the items below these arguments are respectively coded in *to an end, to my attention, in trouble, into another economic depression, into reality, into despair, into effect, out of the darkness of ignorance, into confusion, out of our consciousness* and *into financial trouble*.

Domain B: CMIF

- We need to *put this political crisis to an end* before it is too late.
- You should always *bring the issues to my attention* before you go and make decisions.
- The results of this investigation can *get the president in trouble* in the near future.
- This government can *take the country into another economic depression* in less than a year.
- Peter and Liz want to *turn their plans and ideas into reality* with our help.
- This project will *send all the team into despair* because of its size and complexity.
- We really need to *carry the new plans into effect* in time for the holidays.
- Education can always *lead people out of the darkness of ignorance* no matter the age.
- These actions can *throw the population into confusion* because they aren't clear or well defined.
- We should *drop the problems out of our consciousness* if we want to solve them.
- The policeman could *pull the girl out of danger* before the animal could reach her.
- These bad decisions can *push many small businesses into financial trouble* in the near future.

The observation of learners' responses to domain B (Figure 2) is in line with the expectation for B2 performance in comparison with C1 learners. In both tasks, C1 level learners outperform B2 by more than 0.10. It is important to state that 0.10 does not seem to be a significant difference at first glance, but one must consider the scale of granularity adopted. In the computation of the data, we decided to maintain the mean on the 0-5 scale and make the difference more visual with the aid of line graphs.

Figure 2: Comparison of domain B at levels B2 and C1.



Source: the author

Studies of L1 acquisition and usage suggest that metaphorically extended caused motions are lexically licensed by low-level constructions (fixed phraseologisms) as a result of processes of conventionalization between the verb and the PP argument (Boas, 2013; Hampe, 2010). The second experimental item of domain B is an example of this. *Bring X to one's attention* has a 3.57 MI-score on COCA, which means that the words *bring* and *attention* maintain a mutual level of attraction that is statistically significant. In other words, they form a phraseologism. One could claim that the inability to process phraseologisms entails lack of lexical knowledge and, thus, cannot be used as evidence for the absence of constructional knowledge. That is certainly true for "decoding idioms" (Fillmore, Kay & O'Connor, 1988) - idioms like *kick the bucket*, whose opaque meanings cannot be decoded from their constitutive parts. *Bring X to one's attention*, though, is an "encoding idiom" - one which can be inferentially understood by learners on the basis of its constituent parts. Thus, no specific lexical knowledge would be necessary for an encoding idiom like *bring X to one's attention* to be interpreted by learners. Moreover, knowledge of the phraseologism would certainly yield better and faster results in a timed lexical recognition task, for instance, but in untimed acceptability judgement tasks which contain encoding idioms like this, using the schematic knowledge of X CAUSES Y TO MOVE Z might lead to similar results.

All in all, domain B shows that the figurative variable does affect learners' acceptability judgment of caused motions and this effect is also differently felt by B2 and C1 levels, with C1 outperforming B2, as expected.

4.3 Literal caused motions with modifying verbs

This constructional domain isolates the effect that the type of verb has on learners' interpretation of the items. This variable is believed to put a strain on the learner in different ways from that produced by the need for figurative interpretation. In this particular case, learners must solely rely on the constructional meaning to apprehend both the argument structure of the expressions (some of the verbs are prototypically monoargumental: *laugh, work, talk, breathe, cry, sneeze, scream*) and also interpret the semantics of modification

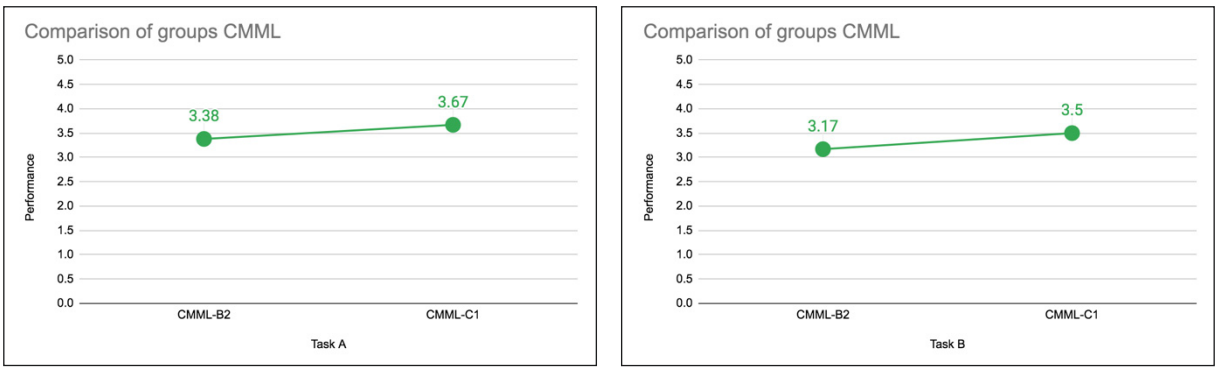
that these verbs convey. In other words, in order to interpret the sentences, learners must be able to apprehend that 'They will ***laugh*** *me and my team* ***out*** *of the office* if we present this' means 'They will ***make*** *me and my team* ***leave*** *the office* ***by laughing at us*** if we present this'.

Domain C: CMML

- They will *laugh me and my team out of the office* if we present this.
- This new policy may *run many good professionals out of the country* in a month.
- You should *work the butter and the milk into the eggs* by using a fork.
- You must *play the ball off your right foot* to reduce the number of mistakes.
- They only need to *speak some words into the microphone* and the spectators go crazy.
- The magician says he can *talk the hat off your head* with his magic tricks.
- You must *breathe lots of air into the lungs of the patient* to save him.
- Patients *cry themselves into the emergency room* when they are in a lot of pain.
- This circus artist can *sneeze spaghetti out of his nose* in front of the audience.
- One parent can't *drive the kids out of the country* without the other parent's permission.
- I had to *scream the kids back into their homes* because of the heavy rain.
- She was able to *walk all the horses into the stables* before it started raining.

As expected, learners' responses to this domain showed effects of proficiency with C1 outperforming B2 in both tasks A and B. We will contrast values across domains to isolate the effect of types of verbs and interpretation in section 4.5, but it is imperative to quickly contrast here the values in domain B (CMIF) to this domain. The mean in the previous domains varied within the range of 4.0, whereas domain C, with the marked variable (i.e., verbs of *modification*), reduced the scale of variation to around 3.0. This reinforces the thesis that both B2 and C1 level learners, albeit to different degrees, have a strong reliance on verb meaning for the interpretation of sentences. As the graphs below demonstrate, within the range of 3.0, the difference in response between B2 and C1 is, respectively, 0.29 and 0.33 (M=0.31).

Figure 3: Comparison of domain C at levels B2 and C1.



Source: the author.

4.4 Figurative caused motions with modifying verbs

The last constructional domain, thought to be the most complex and the most laborious domain in processing terms, also shows variation within the range of 3.0 with C1 level learners performing better than B2, as expected. In this constructional group, both marked variables, type of verb (*modification*) and interpretation (*figurative*) characterize the items. With these, learners are faced with the interpretation task of assigning caused-motion meaning to sentences composed of verbs which do not reflect the argument structure of X CAUSES Y TO MOVE Z, like instantiation verbs do. Learners must also be able to apply the modification reading to such expressions (*'**breathe** yourself **out** of fear and anxiety' = '**move away** from fear and anxiety **by breathing**'*), but differently from domain C, in CMMFs the PP argument is also a figuratively construed space from or to which the constructional causal event dislocates the THEME. The experimental items are repeated below.

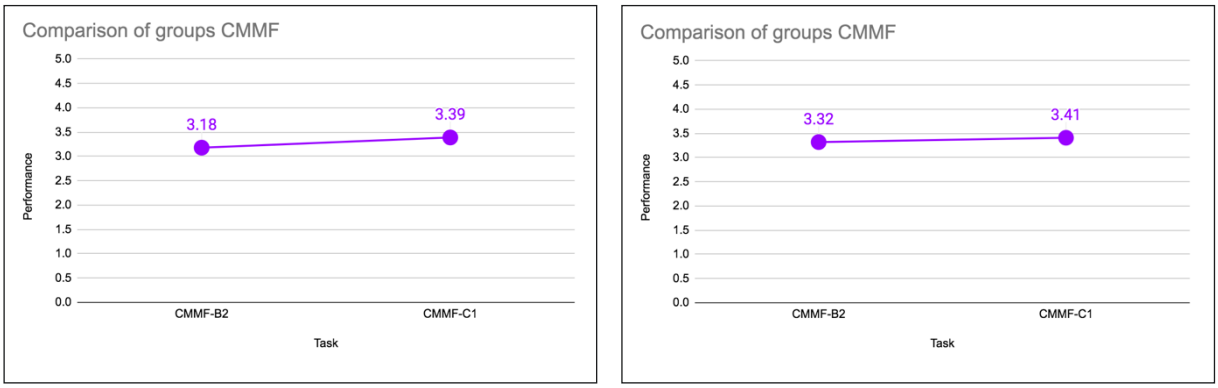
Domain D: CMMF

- My kids always *laugh themselves into exhaustion* when they see puppets and clowns on TV.
- These decisions can *run the healthcare system into the ground* in one month or two.
- Scientists can *work themselves into severe mental breakdowns* if they do not get enough rest.
- After football matches some guys *drink themselves into madness* on the streets around the stadium.
- You can *speak your problems into existence* when you talk about them all the time.
- The doctors need to *talk some sense into her mind* before she makes more mistakes.
- This method teaches you to *breathe yourself out of fear and anxiety* in 8 minutes.
- At night some kids *cry themselves into exhaustion* before they finally stop and fall asleep.
- I literally *sneeze myself into a terrible headache* when I eat anything with black pepper.
- Paul always *drives his parents to desperation* when he behaves the way he did yesterday.
- You cannot *scream yourself out of trouble* if you have a difficult problem to solve.
- Doctors say people can *walk themselves out of their bad mood* if they walk regularly.

The incremental layers of complexity that characterize this domain seem to have had a perceived effect on learners' responses, differently from the

results obtained in the observational data discussed in Rosa (2020). The lowest performance of this group in relation to the previous constructional domains confirms our experimental hypothesis of a descending level of performance at levels B2 and C1, but also shed some light on the unexpected results seen in the learner corpus data. There, the analysis of some concordance lines suggested the 'good' performance of learners was due to lexical knowledge interference in the production of CMMFs. By contrast, here learners could not resort to memorized chunks of language for the interpretation of the expressions and had to rely on a schematic X CAUSES Y TO MOVE Z for the interpretation, since no experimental item above contains statistically significant phraseologisms[8].

Figure 4: Comparison of domain D at levels B2 and C1



Source: the author.

In Section 4.5, proficiency levels B2 and C1 across domains will be consolidated.

4.5 Proficiency levels across constructional domains

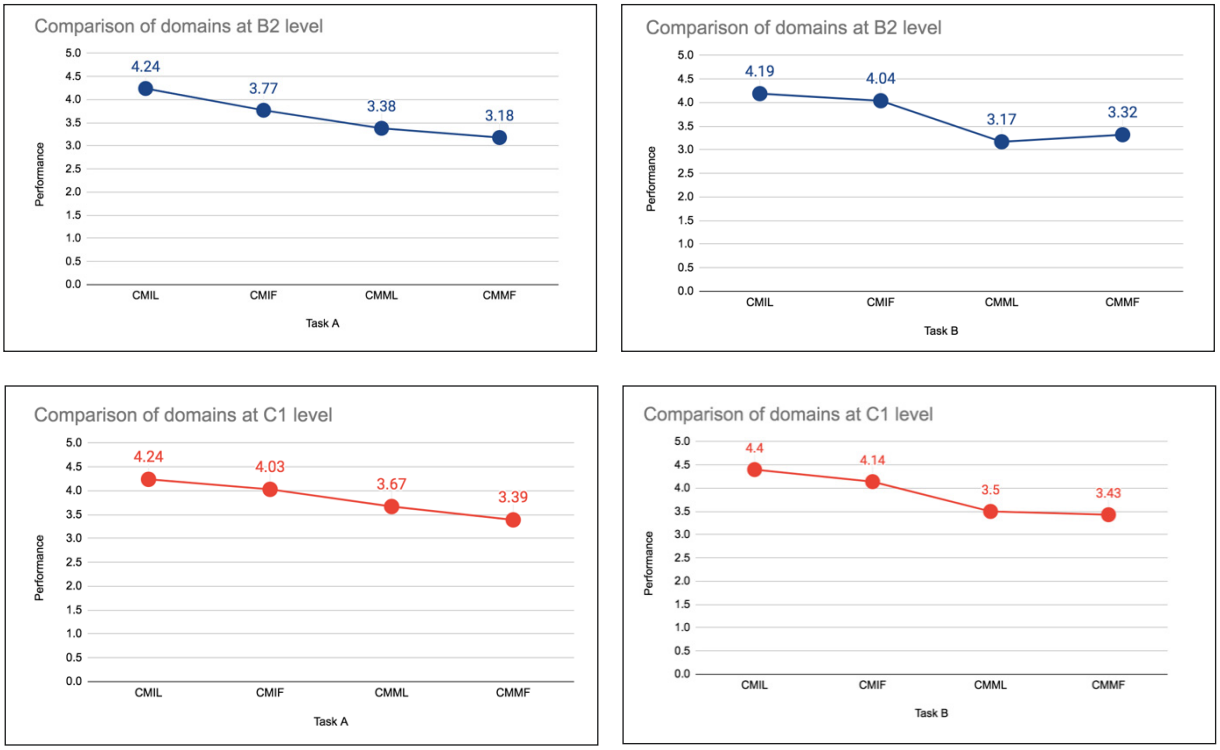
The data analyzed thus far seem to validate our experimental hypothesis (repeated below).

> Learner comprehension of caused motions is affected by the different types of constructions (A, B, C and D) and this effect follows an ascending level of constructional complexity from A to D. Thus, B2 and C1 level learners are expected to have a descending level of recognition that is proportional with the ascending level of constructional complexity of each domain.

The consolidated graphs below show that both proficiency levels did have a gradual impoverished performance as they advanced through the given constructional domains. That is, the level of linguistic complexity of structures did affect learners' comprehension of caused-motion constructions, but differently from the analyzed corpus data, the elicited data confirmed that variables *type of*

verb and *construction interpretation* affect this performance with a difference in rating ranging between 0.85 to 1.06 from domains A to D.

Figure 5: Comparison across domains at levels B2 and C1



Source: the author.

The difference in performance between domains A to D in both tasks A and B are presented in Table 2. For this purpose, let delta (Δ) be the difference between the absolute values of CMIL and CMMF.

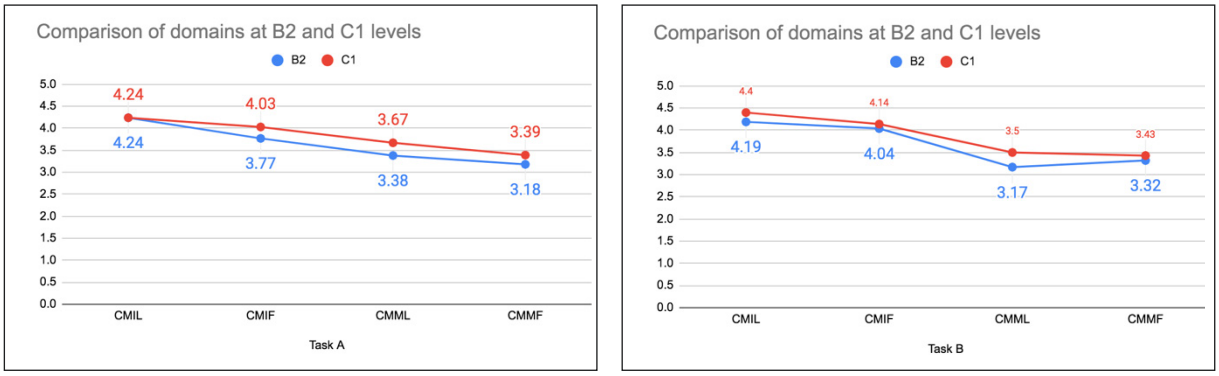
Table 2: B2 and C1 performance from CMIL to CMMF

	B2	**C1**
Test A	Δ = 1.06 (CMIL 4.24 > CMMF 3.18)	Δ = 0.85 (CMIL 4.24 > CMMF 3.39)
Test B	Δ = 0.87 (CMIL 4.19 > CMMF 3.32)	Δ = 0.97 (CMIL 4.4 > CMMF 3.43)

Source: the author

A comparison in performance across all the domains with both B2 and C1 levels is presented below and, as expected, both proficiency levels descend in their comprehension, but to a lesser degree at C1.

Figure 6: Comparison of levels B2 and C1 across domains

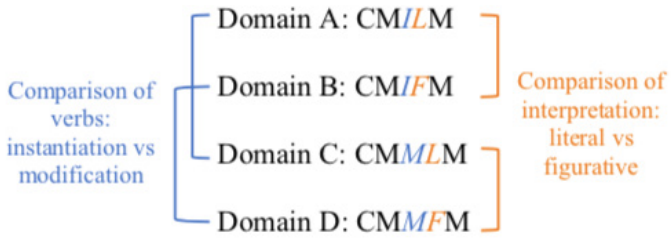


Source: the author

4.6 Proficiency levels across constructional domains

As was done with the observational data in Rosa (2020), we have contrasted domains so as to isolate the effect of *types of verbs* and *constructional interpretation* as follows:

Figure 7: Comparison of domains per types of variables

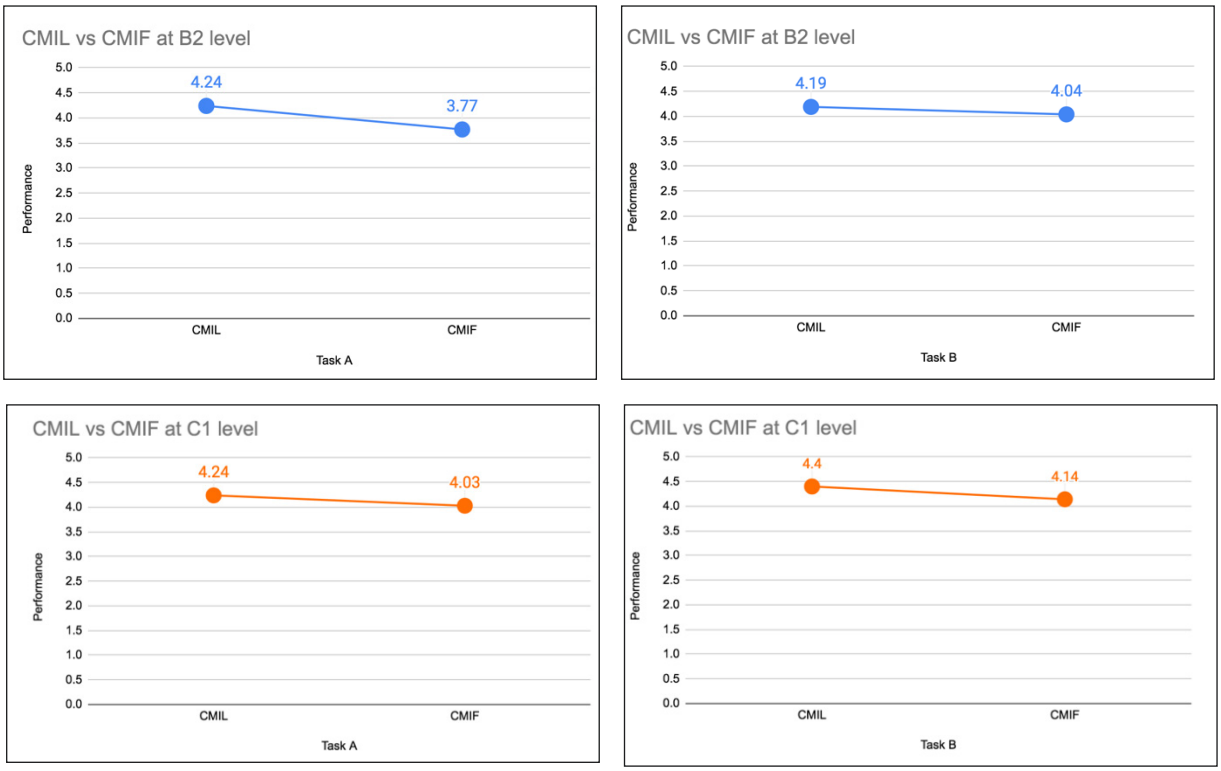


Source: the author

4.6.1 Constructional interpretation across domains

Let us first contrast the effect that the marked *figurative reading* had on the interpretation of the items in domains A and B, that is, constructions with instantiating verbs.

Figure 8: CMIL (domain A) vs. CMIF (domain B) at levels B2 and C1

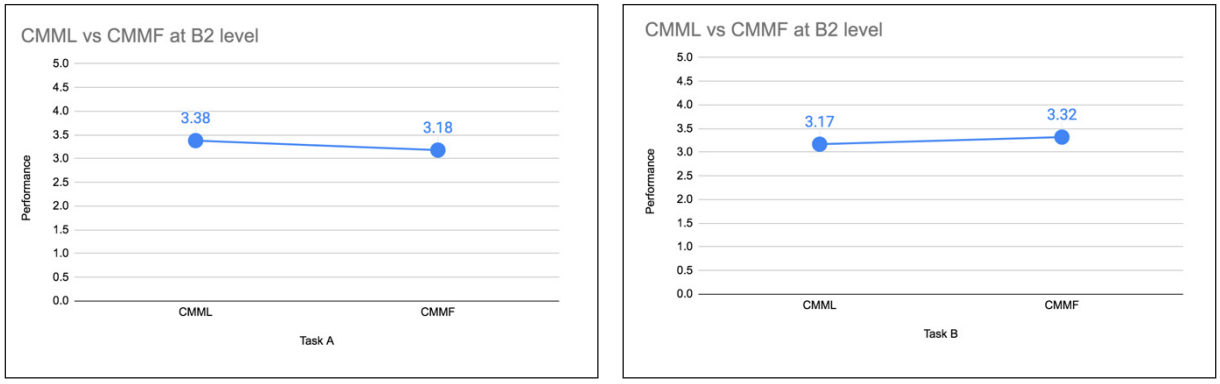


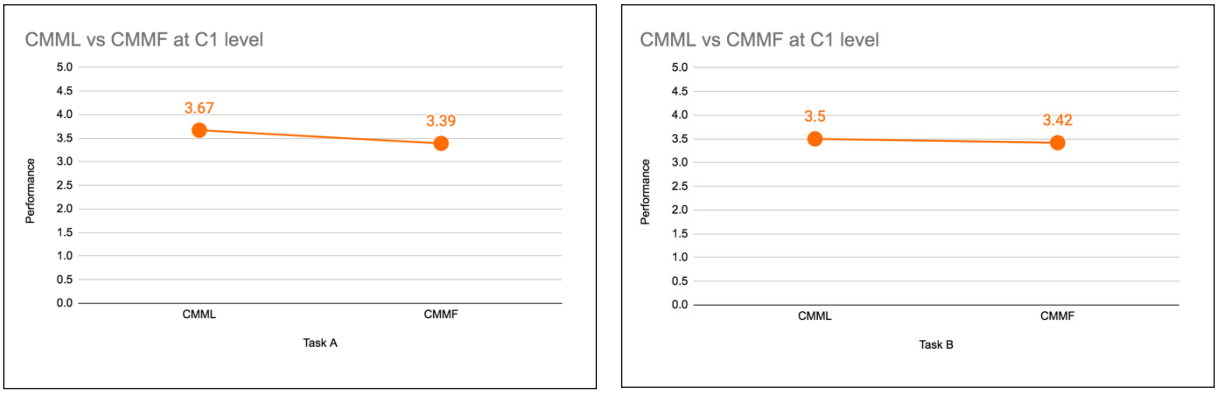
Source: the author

In the four given scenarios, the presence of the variable *figurative meaning* had an effect on the processing of expressions at both levels of proficiency. Although both levels descended in performance, C1 showed a smaller difference (M=0.23) compared to B2 (M=0.31), in accord with the anticipated outcome.

Below are the graphs isolating the *figurative* variable for the items in domains C and D, that is, in contexts with modifying verbs.

Figure 9: CMML (domain C) vs. CMMF (domain D) at levels B2 and C1





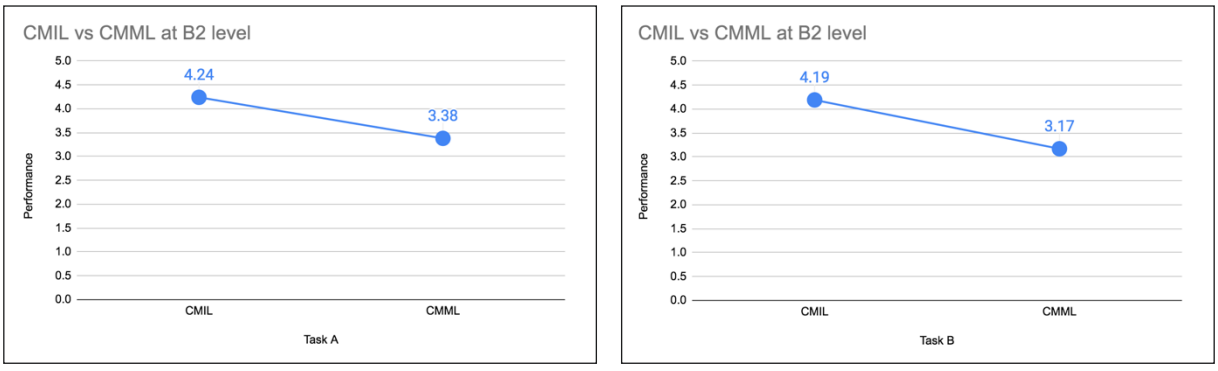
Source: the author

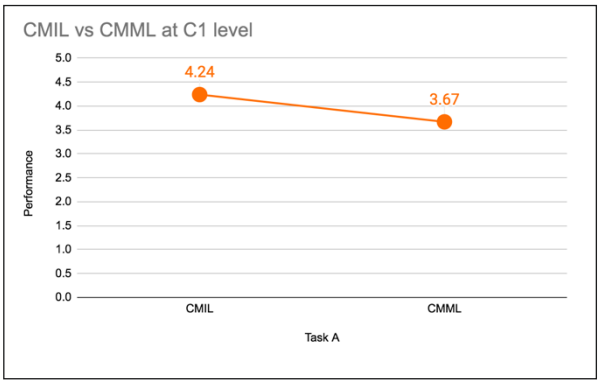
Participants of both levels of proficiency in task A performed according to the predictions made in our hypothesis, but B2 level learners in task B showed an expected improvement of 0.15 in CMMFs, as opposed to C1 learners who performed the same task and showed a decrease of 0.07. The data in task B of B2 may have been skewed by learners who either took a guess while responding to some of the stimuli or who were helped by extralinguistic factors, such as the consultation of materials or the Internet while performing the task[9]. In any case, while it is true that this comparison of domains slightly deviates from the expected outcome, it is important to foreground that the other groups all perform according to expectation. Also, the varying levels in this contrast occur within range of 3.0 and not 4.0 as the previous contrast between CMIL and CMIF. This shows, at least in 3 out of 4 groups, that the presence of modifying verbs in association with figurative readings does affect data interpretation. This contrast is dealt with in the following section.

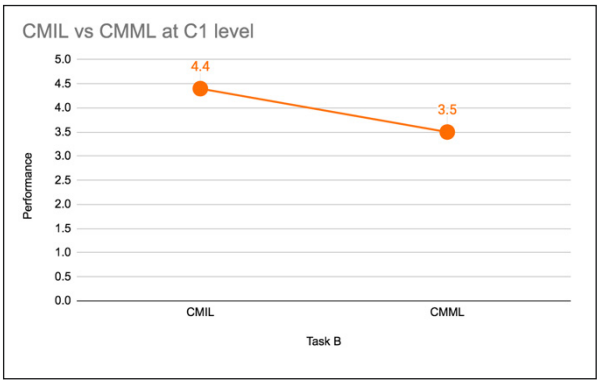
4.6.2 Types of verbs across domains

This contrast of domains aims at isolating the effect of the marked modifying verbs in the contexts of literal and figurative interpretation. Given that this variable demands that learners operate both with the argument structure of schematic constructions and with the modification that specific verbs assign to X CAUSES Y TO MOVE Z, this variable is expected to result in lower performances when compared to the degree that figurative reading affects the responses to the stimuli. Below are the graphs contrasting domains A and C.

Figure 10: CMIL (domain A) vs. CMML (domain C) at levels B2 and C1





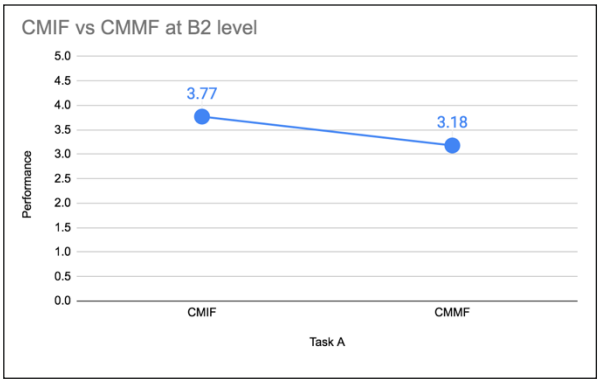


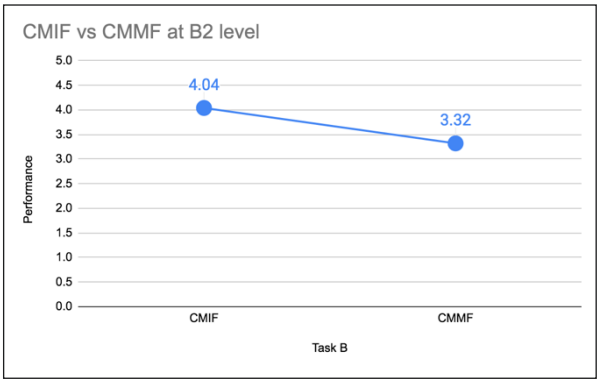
Source: the author

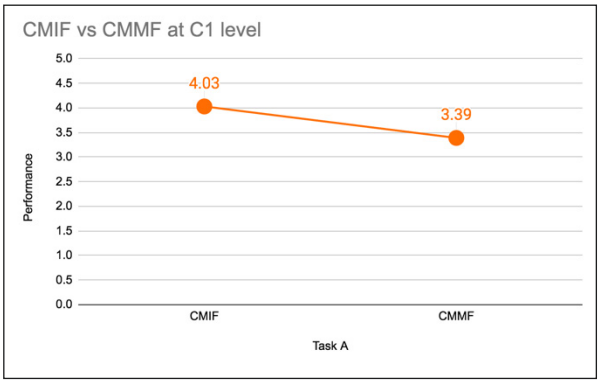
The data in the graphs above exhibit the expected results in all four constructional groups, thus foregrounding the effect that modifying verbs have on the interpretation of caused motions. B2 groups showed a decreased performance with CMMLs in both tasks (A and B) with a mean of almost 1.0 (M=0.94). To a lesser extent, C1 level learners also had their interpretation of CMMLs lowered (M=0.73), but as predicted, the type-of-verb variable was less detrimental to C1's interpretation than it was to B2.

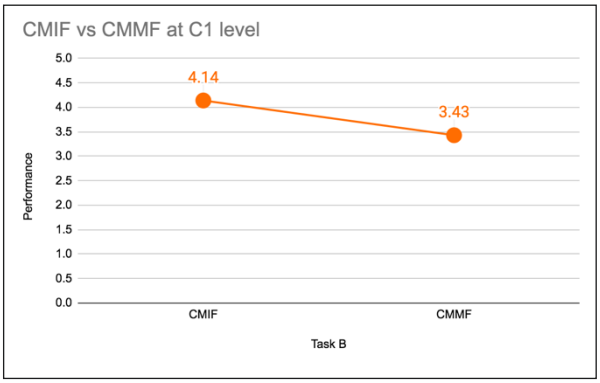
The same observed distribution, with CMMFs being lower than CMIFs were verified in the contrast between domains B and D. However, although C1 performs slightly better than B2, the difference is smaller than in the previous group: B2 (M=0.65) and C1 (M=0.67). Just as for CMMF when contrasted with CMML, the data below surprises vis-à-vis the predictions for domain D.

Figure 11: CMIF (domain B) vs. CMMF (domain D) at levels B2 and C1









Source: the author

The experimental data analyzed here seem to confirm the hypothesis that the interaction between *figurative* readings and *modifying* verbs in caused-motion constructions does pose an interpretation challenge to learners in more significant ways than domains A, B or C do.

5. Conclusions

This article dealt with the comprehension of caused-motion constructions with 4 groups of Brazilian EFL learners at two levels of proficiency: B2 and C1. We have applied an acceptability judgment task to 120 learners (60 B2-level learners and 60 C1-level learners) in two tasks containing the four constructional domains below:

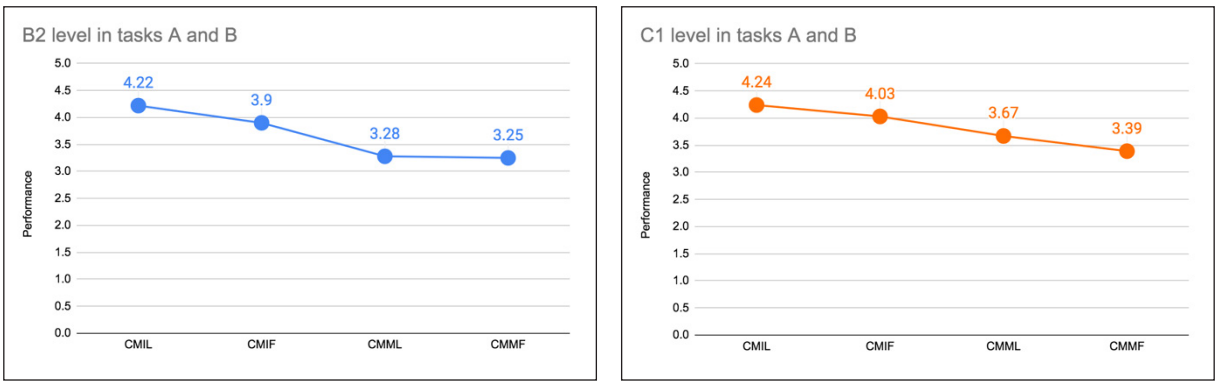
- Domain A: Caused motions with *instantiating* verbs denoting *literal* movement.
- Domain B: Caused motions with *instantiating* verbs denoting *figurative* movement.
- Domain C: Caused motions with *modifying* verbs denoting *literal* movement.
- Domain D: Caused motions with *modifying* verbs denoting *figurative* movement.

The data were compiled and described in light of the four domains above with the guidance of our central hypothesis, that is:

> Learner comprehension of caused motions is affected by the different types of constructions (A, B, C and D) and this effect follows an ascending level of constructional complexity from A to D. Thus, B2 and C1 level learners are expected to have a descending level of recognition that is proportional with the ascending level of constructional complexity of each domain.

As the analyses and discussions carried out throughout section 4 showed, our hypothesis was confirmed in that all learners demonstrated a decreased level of comprehension from domains A to D. The data also showed that proficiency level does affect the degree of interpretation of caused motions in these four domains in that C1 outperformed B2 learners across the constructional domains analyzed. We end this discussion with a consolidation of all the data. The responses of both tasks A and B were put together so as to visualize whether the expected descending line of performance could be observed.

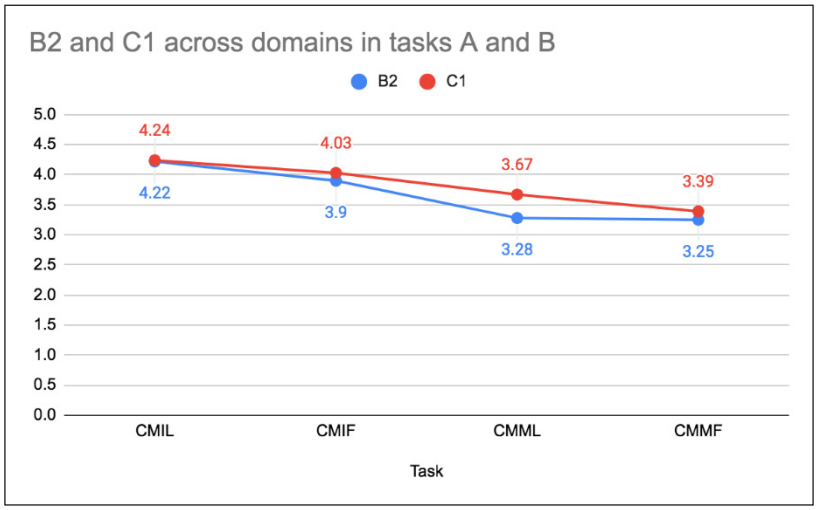
Figure 12: Performance of levels B2 and C1 with tasks A and B



Source: the author

The graphs above represent the responses of B2 and C1 learners to the 48 experimental stimuli distributed across the four domains. As the lines show, the performance expected from learners in the hypothesis was met and, comparatively, C1 outperforms B2 even when the differences between groups of learners is neutralized (Graph 13).

Figure 13: Comparison of B2 and C1 performances with combined tasks A and B



Source: the author

Notes

1. EF-Cambridge Open Language Database (https://corpus.mml.cam.ac.uk/efcamdat2).
2. In the learner corpus analysis, the other CEFR levels of proficiency (A1, A2 and B1) did not display relevant production of caused motions. Therefore, in the experimental part, we narrowed the analysis down to B2 and C1 levels.
3. www.qualtrics.com
4. We have considered an outlier any response that systematically chose 1 or 5 across all the questions in our survey.

5. In order for participants to be prototypical representatives of the CEFR levels B2 and C1, we have selected only participants who were respectively doing an upper-intermediate 2 level (out of three stages) and an advanced 2 level (out of three stages). Our intention with this was to avoid having either participants entering the targeted CEFR level (which could characterize real knowledge of the previous level) or exiting the targeted CEFR (which could characterize knowledge of the following level).
6. This researcher supervised the execution of the task with 50% of participants. With the other 50%, instructions were given face-to-face, but the test was sent by email.
7. Qualtrics' estimated time to perform this task was 18 minutes. We included 2 extra minutes in the instructions given that this was a survey with learners.
8. The only likely candidate for the status of phraseologism 'talk some sense into' has a 0.61 MI-score on COCA, which is considerably far from the widely accepted 3.0 for statistical significance.
9. That could be the case of those learners who did the task in an unsupervised mode.

References

Berber Sardinha, T. (2004). *Linguística de Corpus*. Barueri: SP: Manole.

Brezina, V. (2018). *Statistics in Corpus Linguistics - A Practical Guide*. Cambridge: Cambridge University Press.

Boas, H. C. (2013) Cognitive Construction Grammar. In T. Hoffmann, G. Trousdale (Ed.), *The Oxford Handbook of Construction Grammar* (Chapter 13, pp. 233 – 252). New York: Oxford University Press.

Cabrera, M., & Zubizarreta, M. L. (2004) Constructional properties versus lexical specific transfer: overgeneralized causatives in L2 English and L2 Spanish. BUCLD, 28 *Proceedings Supllement*.

Ellis, N. C. (2013). Second language acquisition. In T. Hoffmann, G. Trousdale (Ed.), *The Oxford Handbook of Construction Grammar* (Chapter 20, pp. 365 – 378). New York: Oxford University Press.

Fillmore, C., P. Kay, & M. O'Connor. (1988). Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language* 64, pp. 501–538.

Gilquin, G. & Gries, S. (2009). Corpora and experimental methods: a state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1), pp. 1-26.

Goldberg, A. E. (1995). A *Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Goldberg, A. E. & Jackendoff, R. (2003). The resultative as a family of constructions. *Language* 80, pp. 532–68.

Gries, S. Th. & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3, p. 182-200.

Keating, D & Jegerski, J. (2005) Experimental Designs in Sentence Processing Research: A Methodological Review and User's Guide. *Studies in Second Language Acquisition*, v37 n1, pp. 1-32.

Levin, B. and Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.

Hampe, B. (2010). Metaphor, constructional ambiguity and the causative resultatives. In Handl, S. & Schmid, J. (Ed.) *Windows to the mind: Metaphor, metonymy and conceptual blending*. (pp. 185–215). Berlin & New York: Mouton de Gruyter.

Ionin, T., & Zyzik, E. (2014) Judgment and Interpretation Tasks in Second Language Research. *Annual Review of Applied Linguistics*, 34, pp. 37-64. doi:10.1017/S0267190514000026

Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.

Rosa, R. G. (2020). Construções de movimento causado em um corpus de aprendizes de inglês: da observação à experimentação. [PhD dissertation, University of São Paulo].

Schütze, C. & Sprouse, J. (2014) Judgement Data. *Research Methods in Linguistics*. In Podesva, R. J. & Sharma, D. (Ed.) (Chapter 3, pp. 27-50). Cambridge: Cambridge University Press.

Sprouse, J. & Almeida, D. (2013). The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*, 28, pp. 222-228. doi: 10.1080/01690965.2012.703782.

Suzuki, K., Shioda, K., & Kikuchi, N., Maetsu, M. and Hirakawa, M. (2016). Cross-linguistic Effects in L2 Acquisition of Causative Constructions. BUCLD 40 Online Proceedings Supplement.

Recebido em: 29/06/2023
Aceito em: 18/08/2023