

# ON SHELL NOUNS, SPECIFICATIONAL CONSTRUCTIONS AND LONG INFINITIVES IN ROMANIAN

ANAÏS VAJNOVSZKI<sup>1</sup>, MARLEEN VAN PETEGHEM<sup>2</sup>

**Abstract.** This article focuses on the specificational long infinitive construction in Romanian (*Soluția este de găsit a unui serviciu* ‘The solution is finding a job’) and the shell nouns it can welcome. It examines the differences between this construction and other specificational constructions containing the long (*găsit* ‘finding’) and the short infinitive (*a găsi* ‘to find’). It shows that the construction under study is mainly used to background the internal argument of the long infinitive. It also shows that the lexeme in the noun slot is not always a typical shell noun. However, in this case, the long infinitive constituent no longer expresses a propositional content, but gives rise to several other interpretations. Our analysis is based on 200 randomized occurrences from the roTenTen corpus.

**Keywords:** shell nouns, specificational constructions, long infinitives, Romanian.

## 1. INTRODUCTION

This article deals with specificational copular constructions in Romanian as exemplified in (1). The subject of these constructions is typically headed by a shell noun, while the predicate is realized by the preposition *de* + an undetermined noun ending in *-re*, called the “long infinitive” in Romanian grammar, and the noun’s complements:

- (1) **Sarcina** limfocitelor este **de recunoaștere** a intrușilor  
task.the lymphocytes.the.GEN is of recognition GEN intruders.the.GEN  
din organism<sup>3</sup> [...] (Sketch Engine)  
from organism<sup>4</sup>  
‘The task of the lymphocytes is to recognize the intruders in the organism’

As is well known, shell nouns are abstract nouns which have the capacity to serve as “conceptual shells for complex, propositional-like pieces of information” (Schmid 2000:4), a phenomenon also known as “encapsulation” (cf. Francis 1986). In the construction under

---

<sup>1</sup> Universiteit Gent & Université Paris 3 Sorbonne-Nouvelle, anais.vajnovszki@ugent.be & anais.vajnovszki@sorbonne-nouvelle.fr.

<sup>2</sup> Universiteit Gent, marleen.vanpeteghem@ugent.be.

<sup>3</sup> For reading comfort, we have systematically indicated in bold both the shell noun and the encapsulated infinitive.

investigation this propositional content is provided by the constituent [*de* + long infinitive + complement].

The aim of this article is to examine the semantics of this specificational construction containing the long infinitive by confronting it with three other infinitive constructions: (i) the adnominal use of the long infinitive (2), and 2 specificational constructions, one with (ii) the long infinitive in a definite NP in a predicative use (3), and one with (iii) the short infinitive (4).

- (2) **Sarcina de recunoaștere** a intrușilor din organism  
 task.the of recognition GEN intruders.the.GEN from organism  
 ‘**The task of recognizing** the intruders in the organism’
- (3) **Ideea** unei naturi matematizate este **descoperirea** care  
 idea.the a.GEN nature.GEN mathematized.GEN is discovery.the.FSG which  
 încifrează sensul concret al lumii. (personal Romanian corpus of academic texts)  
 encodes sens.the concrete GEN world.the.GEN  
 ‘**The idea** of a mathematized nature is **the discovery** which encodes the concrete meaning of the world.’
- (4) **Scopul** este **de a reduce** temperatura de aprindere  
 goal.the is of to reduce temperature.the of lighting  
 a particulelor. (Sketch Engine)  
 GEN particules.the.GEN  
 ‘**The goal** is **to reduce** the temperature at which the particles catch fire.’

Our research questions are the following:

- (i) What are the inherent properties of this construction? We will answer this question within the theoretical framework of Construction Grammar by investigating the common points and differences between the specificational construction with a long infinitive constituent and the three other types of infinitive constructions.
- (ii) What kind of shell nouns are favored by this variant of the specificational construction? Can it also host other types of nouns?

Our analysis will be based on a randomized sample of 200 examples of a specificational long infinitive construction extracted from Sketch Engine. For each of these examples we have examined whether they can be paraphrased by each of the three other infinitive constructions, as well as the factors which make the paraphrase possible or impossible. The identification of those factors will shed light on the differences between the four constructions under study and, most importantly, on the specificities of the construction with the long infinitive.

This article is structured as follows: section 2 is dedicated (i) to describing the construction with the long infinitive in more detail, as well as to the differences between the long and the short infinitive and (ii) to a more in-depth overview of the shell noun problematic. Section 3 gives a description of the corpus and the methodology we have used to collect it. Section 4 outlines the results of the comparison between the

construction with the long infinitive and the other constructions mentioned above. Finally, Section 5 recaps the different types of nouns entering the construction with the long infinitive.

## 2. SCOPE OF THE STUDY

### 2.1. The long infinitive copular construction

In constructional terms, the construction under investigation has at least 4 slots: 1- an NP, which is mostly definite and is the subject of the sentence; 2- the copula *a fi* ‘to be’; 3- the preposition *de*; 4- the long infinitive with, possibly, 5- the infinitive’s complements:

- (5) [...] *dacă obiectivul este de promovare și dacă subiectul campaniei*  
 if goal.the is of promoting and if subject.the campaign.the.GEN  
*este “mass media” [...]*  
 is mass\_media  
 [1- obiectivul] [2- este] [3- de] [4- promovare]  
 ‘[...] if **the goal is promotion** and if the subject of the campaign is “mass media” [...]’  
 [1- the goal] [2- is] [3- to] [4- promote/promotion]

As mentioned above, the long infinitive is actually not a verb, but a noun which is always feminine (Diaconescu 1977, *i.a.*). It is called “long infinitive” for historical reasons because it derives from the Latin infinitive. Like the Latin infinitive, it ends in *-re*, but from a syntactic point of view it behaves like a noun in that it assigns the genitive case to the internal argument of the source verb, *i.e.* the object of transitive verbs (6a), or the subject of unaccusative verbs (7a). In contrast, the short infinitive assigns the accusative to the object of transitive verbs (6b), whereas with unaccusative verbs no corresponding construction with the short infinitive and its nominative subject is available (7b).

- (6) a. *Sarcina limfocitelor este de recunoaștere a intrușilor*  
 task.the lymphocytes.the.GEN is of recognition GEN intruders.the.GEN  
 din organism (original example)  
 from organism  
 ‘**The task** of the lymphocytes is **to recognize** the intruders in the organism’  
 b. *Sarcina limfocitelor este de a recunoaște intrușii din organism*  
 task.the lymphocytes.the.GEN is of to recognize intruders.the from organism  
 ‘**The task** of the lymphocytes is **to recognize** the intruders in the organism.’
- (7) a. *Tendința este de creștere a numărului de cazuri diagnosticate*  
 tendency.the is of growing GEN number.the.GEN of cases diagnosed.PL  
 cu această cumplită boală [...] (Sketch Engine; original example)  
 with this terrible disease  
 ‘**The tendency** is for the number of diagnosed cases from this terrible disease **to grow** [...]’

b. \***Tendința** este **de a crește** numărul de cazuri diagnosticate  
 tendency.the is of to grow number.the of cases diagnosed  
 cu această cumplită boală  
 with this terrible disease

This observation points to a first syntactic feature of the long infinitive construction: when compared to its counterpart with the short infinitive, it can welcome a wider range of verbs, since it can also host unaccusatives.

Other types of arguments of the nouns are encoded as PPs in the same way as with the source verb:

- (8) Doi, trebuie să îți alegi o dietă nu foarte restrictivă  
 two must.3SG SBJV you.DAT choose.2SG a diet NEG very restrictive.FSG  
 pentru că odată terminată o dietă restrictivă,  
 for that once finished.FSG a diet restrictive.FSG  
**tendința** e **de revenire** la vechiurile [sic] obiceiuri deci îngrășare...  
 tendency.the is of going\_back to old.the.PL habits hence fattening  
 ‘Secondly, you must choose a diet which is not very restrictive because once you finish a restrictive diet, **the tendency is to go back** to old habits and so to gain weight...’
- (9) **Tratamentul** este **de calmare** prin mixturi locale, antiinflamatoare.  
 treatment.the is of calming through ointments local.PL anti-inflammatory.PL  
 ‘**The treatment is to calm** through topical ointments, anti-inflammatory drugs.’

Furthermore, the long infinitive can also be modified by adjectives (cf. *continuuă* ‘constant’ in (10a)), which take on the feminine form through agreement with the feminine long infinitive. They are typical modifiers for a nominal constituent. In contrast, in a paraphrase with the short infinitive the adjective becomes an adverbial modifier (*continuu* ‘continuously’ in (10b) in the masculine default form), which, contrary to adjectival modifiers, is compatible with a verbal constituent:

- (10) a. [...] **tendința** este **de creștere** continuă.  
 tendency.the is of growing continuous.FSG  
 ‘[...] **the tendency** is one of constant **growth**.’
- b. **tendința** este **de a crește** continuu  
 tendency.the is of to grow continuously  
 ‘the tendency is to grow continuously’

Together with the use of the genitive case for the object, adjectival modification further confirms the nominal syntax of the long infinitive.

Finally, it has to be noted that each specificational construction has a non-copular, *i.e.* adnominal, variant. For example, an occurrence like *scopul fiind de reducere* in (4) above can be paraphrased by the NP in (11) below:

- (11) **Scopul de reducere**  
 goal.the of reducing  
 ‘**The goal of reducing**’

With respect to the distinction put forward by Grimshaw (1990), it is clear that this infinitive has an event rather than a result reading and hence requires arguments. Nevertheless, in the construction under investigation the long infinitive is never determined by an enclitic definite article. In this way, it lacks a typical nominal property. This yields a non-actualized interpretation of the nominalization, which allows the nominalization to occur without any argument.

In contrast, when the infinitive is definite, the use of the preposition *de* is not possible (12b). Without *de*, as in (12c), the construction is of the same type as example (3) above, which is one of the three competing constructions under investigation. In this case, the presence of an argument makes the construction more natural, as indicated by the ? symbol preceding example (12c).

- (12) a. **Tendința** este **de creștere** (original example from above)  
 tendency.the is of growing  
 ‘The tendency is to grow’
- b. \***Tendința** este **de creșterea**  
 tendency.the is of growing.the.FSG
- c. ?**Tendința** este **creșterea**  
 tendency.the is growing.the.FSG  
 ‘The tendency is growth’
- d. \***Tendința** este **creșterea** a  
 tendency.the is growing.the.FSG GEN  
 numărului de cazuri  
 number.the.GEN of cases  
 ‘The tendency is for the number of cases to grow’

Given that the long infinitive is less actualized than the short infinitive, which is a genuine verb, we expect a prevalence of the long infinitives in constructions without arguments, *i.e.* constructions with intransitive verbs or with transitive verbs whose arguments are omitted.

## 2.2. Shell nouns and specificational copular constructions

As we mentioned in the introduction, shell nouns are nominal lexemes which can encapsulate a piece of text. This is the reason why they frequently occur in the subject of copular specificational constructions, since the predicate of these constructions provide the propositional content. As a matter of fact, the possibility for the noun to occur in the subject of specificational constructions is taken as a criterion for shell-nounhood both in English (cf. Schmid 2000) and in French (Legallois 2008, Roze *et al.* 2014, *i.a.*)

As mentioned above, in a specificational construction the information which “fills” the shell noun can take the form of an infinitive, either long or short. However, the propositional content can also be expressed by other types of constituents. In English, these are the following, as labeled by Schmid (2018), with some examples, also taken from Schmid (2018):

- a- a *that*-clause: *The problem is that I have to leave now.*  
 b- a *to*-infinitive clause: *The idea is to have fun.*  
 c- an interrogative clause: *The question is why you have to leave.*

As for Romanian, Cuniță (2015:180), who has studied one particular kind of shell nouns, namely “attitudinal” shell nouns (cf. *aberație* ‘aberration’, *catastrofă* ‘catastrophe’), mentions subordinate clauses introduced by the complementizers *că* ‘that’ (13) or *să* (14), the latter being a clitic subjunctive marker, as well as infinitives headed by the *de* preposition (however without illustrating the latter), as possible mechanisms for shell nouns to function in context.

- (13) **Problema** e că acum nu mai pot să-l șterg,  
 problem.the is that now NEG more can.1SG SBJV-it.ACC erase.1SG  
 să-l formatez nicicum. (personal Computer-Mediated Discourse corpus)  
 SBJV-it.ACC format.1SG anyhow  
 ‘**The problem** is that now I can’t delete it anymore or format it somehow.’
- (14) **Ideea** e să scoatem poluarea din oraș nu din  
 idea.the is SBJV remove.1PL pollution.the from city NEG from  
 satele din jur. (personal Computer-Mediated Discourse corpus)  
 villages.the from around  
 ‘**The idea** is to take all the pollution out of the city, not out of the villages around it.’

However, the copular constructions with *că*- or *să*-clauses after the copula will not be analyzed nor used for comparison in this article, which will only deal with long and short infinitives (see section 3.1 for a detailed overview of the four constructions used).

Furthermore, it has to be noted that not all nouns which co-occur with long infinitives after the copula are indeed shell nouns. Sometimes the post-copular constituent can express something else than the propositional content of the noun in the pre-copular position, e.g. an action that is framed as a goal. Indeed, for (15a), which contains a long infinitive in the predicate, a paraphrase with the short infinitive (15b) is impossible, which shows that the noun in the subject slot of the construction, *i.e.* *pozele* ‘the photos’, is not a shell noun. The copular construction is then no longer specificational but predicational.

- (15) a. Pozele sunt de prezentare. (Sketch Engine)  
 photos.the are of presenting  
 ‘The photos are to present / for presenting.’  
 b. \*Pozele sunt de a prezenta.  
 photos.the are of to present

Nevertheless, certain nouns which are not typically shell nouns may be coerced as such, as is the case in (16), where the noun *puls* ‘pulse’ is used as a shell noun in a specificational construction with a *că*-clause, as a synonym of the shell noun *impresie* ‘impression’ or *opinie* ‘opinion’. Indeed, shell nouns form an open-ended category, which means that virtually any noun can become a shell noun if used the right way. Many factors can trigger the coercion of a non-prototypical shell noun into the shell noun slot of a specificational construction.

- (16) [The journalist had previous asked “What do you feel?”]  
**Pulsul meu** este că românii au  
 pulse.the mine is that Romanians.the have.3PL  
 bun-simț [...] (personal corpus of spoken Romanian)  
 good sense  
 ‘My pulse is that Romanians have common sense.’

A discussion of this types of nouns goes beyond the scope of the article. However, in section 3.4, we will discuss some examples of nouns that have been coerced into the construction under scrutiny, namely the copular construction with a long infinitive, bearing in mind the following research question: does the copular construction with a long infinitive help make (new) shell nouns?

### 3. METHODOLOGY

#### 3.1. Corpus and examples

The corpus used for this study is the roTenTen16 corpus, accessible via the online corpus creation and managing platform Sketch Engine. This is a “Romanian corpus made up of texts collected from the Internet”<sup>5</sup> and contains over 2 billion words. It has been automatically lemmatized, which allows for quick CQL searches.

In order to acquire a manageably sized corpus, we have used the “randomize” function offered by Sketch Engine when collecting our data: this function allows a user to save a randomized sample of the size of the user’s choice, “with the same number of lines generated from the same concordance [...] always produc[ing] exactly the same concordance lines”<sup>6</sup>. This allows for data representativity and reproducibility.

Another note should be made about the reliability of the examples: since many of our examples come from online productions, many contain spelling errors, such as missing letters or diacritics. For a better ease of reading, we have taken the liberty of correcting such mistakes, without ever adding any element that could change the meaning of the example.

We have gathered 200 examples of the construction under study containing the long infinitive using the following CQL request: [tag="Nc..r.\*"] [lemma="fi" & tag="V.\*"] [word="de"] [tag="N.\*" & word=".\*are|. \*ire|. \*ere" & !lemma="mirare<sup>7</sup>|părerere|parere<sup>8</sup>|culoare<sup>9</sup>"]. In what follows, we will call this construction ‘Cx0’:

<sup>5</sup> <https://www.sketchengine.eu/rotenten-romanian-corpus/>

<sup>6</sup> *Ibid.*

<sup>7</sup> The noun *mirare* ‘surprise’ was excluded from the CQL search in order to evacuate all cases of *a fi de mirare* ‘to be surprising’ (litt. ‘to be of surprise’).

<sup>8</sup> The noun *părerere* ‘surprise’ and its incorrectly spelled variant *parere* was excluded from the CQL search in order to evacuate all cases of *a fi de părerere* ‘to have the opinion’ (litt. ‘to be of opinion’).

<sup>9</sup> The noun *culoare* ‘color’ was excluded from the CQL search in order to evacuate all cases of *a fi de culoare* ‘to be colored blue / brown etc.’ (litt. ‘to be of color + blue / brown’).

- a- Cx0: the copular construction with a long infinitive not determined by an article  
Example: Sarcina limfocitelor **este de recunoaștere** a intrușilor  
 task.the lymphocytes.the.GEN is of recognition GEN intruders.the.GEN  
 din organism  
 from organism

These examples have been annotated according to whether or not they can be transformed into the following three constructions:

- b- Cx1: the adnominal construction with a long infinitive  
Example: Sarcina **de recunoaștere** a intrușilor din organism  
 task.the of recognition GEN intruders.the.GEN from organism  
 ‘The task of recognizing intruders in the organism’
- c- Cx2: the copular construction with a short infinitive  
Example: Principala sa sarcină **este de a echilibra** greutatea corpului  
 main.the his task.the is of to balance weight.the body.the.GEN  
 și nivelurile de energie  
 and levels.the of energy  
 ‘His main task is to balance the weight of the body and the energy levels’
- d- Cx3: the copular construction with a long infinitive determined by an article  
Example: Ideea unei naturi matematizate **este descoperirea**  
 idea.the a.GEN nature.GEN mathematized is discovery.the.FSG  
 care încifrează sensul concret al lumii  
 which encodes sens.the concrete GEN world.the.GEN  
 ‘The idea of a mathematized nature **is the discovery** which encodes the concrete meaning of the world’

### 3.2. Cleaning the data: eliminating irrelevant occurrences

Despite the carefully coded CQL request mentioned above, some noise needed to be manually removed from the initial dataset. Among the false positives we first find typographical inconsistencies which caused Sketch Engine’s tagger to misinterpret an element as being part of the targeted lexico-syntactic pattern. For example in (17) the extra spaces in the word *polițe* ‘policy’ caused the automatic tagger to interpret the isolated *e* as the verb *a fi* ‘to be’, whereas there is in fact no copula in the sentence.

- (17) S-a afirmat faptul ca ar fi beneficiarul unor poli t e de asigurare  
 S-a afirmat faptul că ar fi beneficiarul unor **polițe de asigurare** (without the extra spaces)  
 ‘It has been claimed that he is the beneficiary of an insurance policy’

In addition, we have decided to exclude from our dataset all occurrences where the *-re* ending is not due to the infinitival nature of the constituent, as in (18), *sărbătoare* being a noun meaning ‘feast’. Thirdly, cases of fixed syntagmas (19), such as *de vânzare* ‘for sale’, have also been excluded from our dataset.



- (18) Atmosfera în centru e **de sărbătoare**.  
 atmosphere.the in centre is of party  
 ‘The atmosphere in the city center feels like party / is a party atmosphere’
- (19) Oare binele e **de vânzare**?  
 Q good.the is of selling  
 ‘Is the good for sale?’

Finally, we would like to mention some examples which were excluded from our dataset on the basis that the *-re* form, although derived from a verb, has lost its verbal nature. To test this, we have used a syntactic test involving the paraphrase with a construction mentioned before, namely the copular construction with a subjunctive *să*-phrase. Although this particular construction is not studied here as a point of comparison (see section 3.1 for an overview of which constructions we will be using), we use it here as a test to identify shell nouns, given the affinity of this construction with shell nouns. For example, example (20a) cannot be paraphrased by (20b), whereas (1), repeated here under (21a), withstands this transformation: we have therefore excluded examples such as (20) from the dataset, but kept those that fulfill the same tests, such as (21). In other words, although the *să*-phrase construction is not taken into consideration in this study in the same way as Cx0, Cx1, Cx2 and Cx3, it is used as a test to determine which examples will be part of our final dataset.

- (20) a. Muzicologul D. Prenisnicov s-a opus adoptării  
 muzicologist.the D Prenisnicov REFL-has opposed adoption.the.DAT  
 programului de concert, considerând că **muzica**  
 programme.the.GEN of concert considering that music.the  
 este **de orientare prooccidentală** [...]  
 is of orientation pro\_occidental  
 ‘Musicologist D. Prenisnicov stood against the adoption of the concert program,  
 claiming that the music has a pro-Western orientation [...].’
- b. \*Muzica este să orienteze prooccidental  
 music.the is SBJV orients.SBJV pro-occidental
- (21) a. **Sarcina limfocitelor** este **de recunoaștere** a intrușilor  
 task.the lymphocytes.the.GEN is of recognition GEN intruders.the.GEN  
 din organism (example given in the introduction)  
 from organism  
 ‘**The task of the lymphocytes is to recognize** the intruders in the organism’
- b. **Sarcina limfocitelor** este să recunoască intrușii  
 task.the lymphocytes.the.GEN is SBJV recognize.SBJV.3PL intruders.the  
 din organism  
 from organism  
 ‘The task of the lymphocytes is to recognize the intruders in the organism’

### 3.3. Annotating the 200 examples

Since Cx0 is the starting point of our study, the syntactic manipulations that aim at identifying the features of cx0 will consist in a paraphrase of Cx0 with, respectively, Cx1, Cx2 and Cx3. In other words, in our dataset, three variables per construction were annotated,

each encoding whether a paraphrase with Cx1, Cx2 and Cx3 is possible or not. We also checked whether the resulting paraphrase could be made more acceptable by the addition of a complement (22c).

- (22) a. Deci **problema** **e de socializare** in mod real [...]
   
therefore problem.the is of socialising in mode real
   
'So **the problem is of socializing** in real life'<sup>2</sup>
- b. ?**problema** **de socializare** (reformulation of example (22) with Cx1)
   
problem.the of socializing
   
'The problem of socializing'
- c. **problema de socializare** cu oameni
   
problem.the of socializing with people
   
'The problem of socializing with people'
   
(reformulation of example (22) with Cx1, made more acceptable by the addition of the complement *cu oameni* 'with people')

### 3.4. Sorting the data

In the introduction we mentioned that shell nouns form an open category and that new lexemes can be coerced into constructions which are usually associated with shell nouns. Our corpus contains such examples of nouns whose semantics does not call for a specification somewhere else in the text, but which still integrate into Cx0, thus taking on a shell-noun-like behavior. These examples can have several readings associated with them, which are sometimes overlapping.

#### 3.4.1. *When the infinitive encodes the goal of a shell noun*

When the infinitive encodes a goal (cf. example (15) above, which we reproduce under (23) below), it is possible to paraphrase the long infinitive with a short infinitive headed by the preposition *pentru* 'for' (23b). The addition of a boulogmatic shell noun such as *goal* or *purpose* (23c) is also possible:

- (23) a. Pozele sunt de prezentare.
   
photos.the are of presenting
   
'The photos are to present / for presenting.'
- b. Pozele sunt pentru a prezenta.
   
photos.the are for to present
   
'The photos are for presenting.'
- c. Scopul pozelor este de a prezenta.
   
goal.the photos.the.GEN is of to present
   
'The goal of the pictures is to present.'

With these examples, there is no informational void to be filled somewhere else in the text via an encapsulation. Indeed, the question '*What is the N?*' (*Quel est le N?* in French, where *quel* can be translated as both *what* or *which*), which is a common test for shell-

nounhood (cf. Legallois and Gréa 2006, Legallois and Vajnovszki 2021), is most of the time not possible (23d), or will be answered by an identificational (not a specificational) construction (24b).

- (23) d. ?Care sunt pozele?  
 which are photos.the  
 ‘What / Which are the pictures?’
- (24) a. **Ultima unitate** este **de recapitulare și evaluare finală** [...]  
 last.the unit is of recapping and evaluating final.FSG  
 ‘**The last unit is to recap / for recapping and for a final evaluation**’
- b. Care este ultima unitate? Ultima unitate este ceea care îi ajută  
 which is last.the unit last.the unit is this which they.ACC helps  
 pe elevi să recapituleze materia  
 ACC students SBJV recap.SBJV.3PL material.the  
 ‘What/Which is the last unit? The last unit is the one that helps students to recap the material’

This is why we consider that, in most cases of coercion, there has been a kind of “shortcut” leading to the omission of the implied boulomaic shell noun, like in (23e) and (24c).

- (23) e. Care este scopul pozei? Scopul pozei este de a prezenta.  
 which is goal.the photo.the.GEN goal.the photo.the.GEN is of to present  
 ‘What / Which is the goal of the picture? The goal of the picture is to present.’
- (24) c. Care este scopul ultimei unități? Scopul ultimei  
 which is goal.the last.the.GEN unit.the.GEN goal.the last.the.GEN  
 unități este de a recapitula.  
 unit.the.GEN is of to recap  
 ‘What / Which is the goal of the last unit? The goal of the last unit is to recap.’

### 3.4.2. When the transient nature of the shell noun is made explicit: some cases of “defining utterances”

An important aspect of shell nouns is that they create ephemeral categories. Or, like Legallois (2008:9) puts it: “The only limit of this categorization is its duration because it only exists as long as the discourse does<sup>10</sup>” (our translation). This is why specificational constructions often welcome lexemes which do not need a specification and, hence, are at the margins of the shell noun category. Therefore, they cannot be the subject of the ‘*What is the N?*’ question, but their content can sometimes be specified. When the speaker chooses to give a specification to a non-typical shell noun, like *vacanță* ‘holidays’ (25), they can do this under the guise of a personal definition, thus expressing a “defining utterance” or “explicitation utterance” (Legallois and Vajnovszki 2021). We illustrate this possibility

<sup>10</sup> « La seule limite est la durée de cette classification qui n’existe que le temps d’un discours. » (Legallois 2008:9)

below with an example of a defining utterance from our corpus, in which the subjectivity of the definition is made explicit through the use of the prepositional phrase *pentru dumneavoastră* ‘for you’.

- (25) a. Indiferent că preferați      distractia      sau dacă **vacanța**      **pentru**  
indifferent that prefer.2PL      distraction.the      or if      holiday.the      for  
**dumneavoastră**      este      **de relaxare** [...]        
you      is      of relaxing  
‘Whether you prefer fun or whether for you **holidays are to relax**’
- b. Care este scopul vacanței      (pentru dumneavoastră)?  
which is goal.the holiday.the.GEN for      you  
Scopul vacanței      este      de a (se)      relaxa.  
goal.the      holiday.the.GEN is      of to REFL.PRON      relax  
‘What / Which is the goal of a holiday (for you)? The goal of a holiday is to relax.’
- c. Pentru dumneavoastră, vacanța înseamnă a (se)      relaxa.  
for      you      holiday.the means      to REFL.PRON relax  
‘For you, holidays mean(s) relaxing.’

However, in cases such as (25) it is difficult to determine whether it is the goal of the noun which is being specified (25b) or the content of the noun itself under the guise of a defining utterance (25c). Hence, there can be an overlap in the licensing factors that enable coercions.

### 3.4.3. When the long infinitive expresses a subtype of the head noun

Another interpretation of examples such as (25) would be to consider the long infinitive as expressing a subtype, as in the adnominal equivalent (25d), where *de relaxare* indicates a type of holiday, just like *cerneală de sublimare* ‘sublimation ink’ (26) can mean a certain type of ink (26b).

- (25) d. Ce tip de vacanță doriți?      Doresc **o vacanță de relaxare**.  
what type of holiday wish.2PL      wish.1SG a holiday of relaxing  
‘What kind of holiday do you want? I want a relaxation holiday.’
- (26) a. **Cerneala**      este      **de sublimare** fixată      termic      la  
ink.the      is      of sublimation fixed.FSG      thermally      to  
150 grade C pe material.  
150 degrees C on material  
‘The ink is for sublimation, thermally fixated onto the material with 150 degrees Celsius.’
- b. Ce tip de cerneală este asta?      Este **o cerneală de sublimare**.  
what type of ink is this      is a ink      of sublimation  
‘What kind of ink is it? It is a sublimation ink.’
- c. Care este scopul cernelii?      Scopul cernelii      este de a sublima.  
which is goal.the      ink.the.GEN      goal.the ink.the.GEN is of to sublimate  
‘What / Which is the goal of the ink? The goal of the ink is to sublimate.’

The sub-type interpretation is even more convincing for nouns such as *impresie* ‘impression’ (27) or *sentiment* ‘sentiment’ (28), which have been called “classifying terms” (Novakova *et al.* 2018 about their French equivalents). When those shell nouns are used in this way, they often co-occur with another “psychological noun<sup>11</sup>” (preceded by the preposition *de*) which specifies the subtype (*ibid.*).

- (27) a. Ca și in cazul      dicționarelor lui      Valeriu Cristea, **impresia**  
 like and in case.the      dictionaries.the.GEN      Valeriu Cristea impression.the  
 este **de regăsire**      a      unei      frumuseți originare,  
 is of finding\_again      GEN      a.GEN beauty.GEN original  
 de lustruire      a      argintăriei.  
 of polishing      GEN      silverware.the.GEN  
 ‘Just like with Valeriu Cristea’s dictionaries, **the impression** is of **finding** an original beauty again, of polishing silverware.’
- (28) a. **Sentimentul** este **de dezagregare**      sub imperiul  
 feeling.the is of disaggregation      under empire.the  
 apei,      al      lumii  
 water.the.GEN GEN      world.the.GEN  
 ‘**The feeling** is that of **disaggregation** under the force of water, of the world.’

However, for classifying terms such as *impresie* or *sentiment* the telic value has disappeared since the paraphrases with (27b) and (28b) are not acceptable. Furthermore, the questioning through ‘*What / Which is the N?*’ remains possible (27c-28c), yielding the long infinitive as an answer. Those arguments suggest that classifying nouns are not as marginal as nouns like *poză* ‘picture’ or *unitate* ‘unit’, but are closer to the prototypical shell noun.

- (27) b. \*Scopul impresiei      este de a regăsi.  
 goal.the impression.the.GEN is of to find\_again  
 ‘The goal of the impression is to find again.’  
 c. Care este **impresia**?      Este **de a regăsi**.  
 which is impression.the      is of to find\_again  
 ‘What / Which is **the impression**? It is **to find again**.’
- (28) b. \*Scopul sentimentului este de a dezagrega.  
 goal.the feeling.the.GEN is of to disaggregate  
 c. Care este **sentimentul**?      Este **de a dezagrega**.  
 which is feeling.the      is of to disaggregate  
 ‘What / Which is **the feeling**? It is **to disaggregate**.’

<sup>11</sup> In French in the original text: “Lorsqu’il est suivi d’un <SN prép.>, ce nom [*sentiment*] est classifieur (5)-(6): il sert à introduire d’autres N psychologiques (*sentiment* {*d’inquiétude, d’abandon*}), distribution qu’il ne partage pas avec émotion (\**émotion* {*d’inquiétude, d’abandon*}) : (5) Donc j’éprouve vis-à-vis de votre personne **un sentiment de méfiance**. (Dard, *Du sable dans la vaseline*, 1998)”

(6) Dans mon boulot, on a **un sentiment de ras-le-bol**. (*Libération*, 19-01-2007)”  
 (Novakova *et al.*, 2018:86).

Our dataset even contains an example in which the long infinitive acquires a different meaning when combined with a classifying noun. The verb *a respinge* is defined by the Dex (online Romanian dictionary) as ‘to move away from, to cast aside’ (our translation): this meaning is illustrated in (29) with the long infinitive *respingere* in cx0. The long infinitive retains the same meaning in the adnominal variant illustrated in (29b).

- (29) a. În cazul în care **hotărârea** este **de respingere** a înscrierii,  
 in case.the in which decision.the is of rejection GEN registration.the.GEN  
 aceasta se motivează, se comunică  
 this REFL.PRON motivates REFL.PRON communicates  
 solicitantului [...].  
 solicitant.the.DAT  
 ‘In the case in which **the decision** is that of **a rejection** of the registration, this shall be justified and communicated to the solicitant [...].’  
 b. **Hotărârea de respingere** (reformulation of example (29) with Cx1)  
 decision.the of rejection  
 ‘**The decision to reject**’

However, when preceded by the classifying noun *reacție* in Cx0 (30), the corresponding adnominal sequence (30b) takes on the meaning of ‘repulsion’ or ‘disgust’, and not of ‘rejection’<sup>12</sup>.

- (30) a. **Prima mea reacție față de o persoană care miroase a transpirație**  
 first.the mine reaction face of a person which smells PREP sweat  
 este **de respingere**  
 is of rejection  
 ‘**My first reaction when faced with a person who smells of sweat is of disgust**’  
 b. O reacție de respingere (reformulation of example (30) with Cx1)  
 a reaction of rejection  
 ‘A reaction of disgust’

This semantic but also syntactic shift from a transitive verb to an unaccusative one, is quite telling and will be further explored in the next section.

#### 3.4.4. Interim conclusion

As an interim conclusion on coercion cases, we can point out one common characteristic of all coerced nouns: the long infinitive with which they combine has somehow lost its verbal value, which results in an impossible paraphrase with Cx2 and Cx3, but not with cx1.

For example, nouns which acquire a telic value through their integrating Cx0 (31) can only enter cx1 (31b), but not Cx2 (31c) or Cx3 (31d). The only exceptions to this rather

<sup>12</sup> We would like to thank here the native speakers who confirmed this interpretation.

steadfast constructional behavior happen when a paraphrase with Cx2 and Cx3 becomes possible precisely because some kind of complement (32) to the long infinitive is already present in the original Cx0 occurrence.

- (31) a. **O pereche e de alergare montană** (amuzant considerând  
a pair is of running mountainous.FSG amusing considering  
că nu apuc să fac acest tip de sport, dar ...eh, în fine).  
that NEG get.1SG SBJV make.1SG this kind of sport but eh in end  
'One pair is for mountain running (funny considering that I never get to practice this,  
but... eh, whatever).
- b. O pereche de alergare montană (reformulation of example (31) with Cx1)  
a pair of running mountainous.FSG
- c. \*O pereche este de a alerga la munte/montan  
a pair is of to run at mountain mountainous.MSG  
(reformulation of example (31) with Cx2)
- d. \*O pereche este alergarea montană (reformulation of example (31) with Cx3)  
a pair is running.the.FSG mountainous.FSG
- (32) a. **Tratamentul este de calmare** prin mixturi locale, antiinflamatoare.  
treatment.the is of calming through ointments local.PL anti-inflammatory.PL  
'The treatment is to calm through topical ointments, anti-inflammatory drugs.'
- b. **Tratamentul de calmare** prin mixturi locale  
treatment.the of calming through ointments local.PL  
(reformulation of example (32) with Cx1)
- c. **Tratamentul este de a calma** prin mixturi locale  
treatment.the is of to calm through ointments local.PL  
(reformulation of example (32) with Cx2)
- d. **Tratamentul este calmarea** prin mixturi locale  
treatment.the is calming.the.FSG through ointments local.PL  
(reformulation of example (32) with Cx3)

The same holds true for classifying nouns (33), which can undergo transformation into cx1 (33b), but not into Cx2 (33c) or Cx3 (33d).

- (33) a. **Senzația este de durere ascuțită** sau de arsură.  
sensation.the is of hurting sharp.FSG or of burning  
'The sensation is of sharp pain or burning.'
- b. **Senzația de durere ascuțită** (reformulation of example (33) with Cx1)  
sensation.the of hurting sharp.FSG  
'The sensation of sharp pain'
- c. \*Senzația este de a durea ascuțit. (reformulation of example (33) with Cx2)  
sensation.the is of to hurt sharp.MSG
- d. \*Senzația este durerea ascuțită. (reformulation of example (34) with Cx3)  
sensation.the is hurting.the.FSG sharp. FSG

## 4. RESULTS

In this section, we present the detailed results of our corpus research, which aims at identifying the factors that block the paraphrase of the long infinitive construction with one of the other constructions or make it harder: Cx1, Cx2 and Cx3. The goal is to identify possible similarities and differences between one or more of the four constructions under scrutiny in this article. But first, we will examine the nouns which enter without any difficulty all three constructions used for comparison.

### 4.1. Which shell nouns combine with the long infinitive?

Among the most frequent shell nouns in our dataset we find well-established shell nouns, like *tendință* ‘tendency’ (39 occ.) and *efect* ‘effect’ (9 occ.). In the modal classification of shell nouns (Adler and Legallois 2018, Legallois and Vajnovszki 2021), most of these nouns fall under what these authors call alethic the modality (*acțiune* ‘action’, *activitate* ‘activity’, *fenomen* ‘phenomenon’, *rezultat* ‘result’, *trend* ‘trend’). This modal class contains shell nouns which describe reality in an objective way. Missing from our list, though, is the very frequent alethic shell noun *fact* ‘fact’, which is among the most used (alethic) shell nouns.

Our list also contains quite a lot of epistemic shell nouns, such as *risc* ‘risk’ or *tendință* ‘tendency’. These shell nouns are also descriptive, but they are based on a subjective judgment rather than on reality (Legallois and Vajnovszki 2021). Thirdly, we find some nouns which fall under relative modality, which means that they express conditions or sub-actions which make a greater end goal possible (*ibid.*). These are shell nouns like *etapă* ‘step’ or *strategie* ‘strategy’.

What is striking is that our corpus only yields a very small proportion (2 occ. overall) of boulomaic shell nouns, i.e. nouns like *idee* ‘idea’ or *obiectiv* ‘objective’. This is surprising because these two shell nouns are among the most frequent shell nouns, yet each of them only appears once in our corpus, coming behind alethic and epistemic shell nouns, which are usually rarer.

All in all, the most prototypical shell nouns of our dataset typically belong to modal classes which are descriptive (alethic and epistemic classes), not injunctive (the boulomaic class).

### 4.2. Factors that block the paraphrases

Now that we have had a look at the most frequent shell nouns in our dataset, we can examine more thoroughly the acceptability of the paraphrases with the three other constructions. We will examine if these acceptabilities coincide and we will also explain why one or more paraphrases are blocked.

#### 4.2.1. Paraphrasing with cx1: possible with all but one example

The paraphrase with Cx1 is the easiest to achieve, since almost all occurrences in our dataset allow it. This points to a greater similarity between Cx0 and Cx1 than between Cx0 and the other two constructions examined in this study. Only one exception



was found during the annotation phase, and that is with the shell noun *chestia* ‘thing’ in (34a), which cannot be paraphrased with Cx1 (34b) but which, surprisingly, can be so with Cx2 (34c).

- (34) a. [...] atunci, trebuie să notăm! cred că **toată chestia**  
 then must.3SG SBJV note.3PL think.1SG that whole thing.the  
 era de abordare a problemei...  
 was of tackling GEN problem.the.GEN  
 ‘Then, we must take note! I think **this whole thing** was of **tackling the problem**...’  
 b. \**toată chestia de a aborda problema* (reformulation of example (34) with Cx1)  
 whole thing.the of to tackle problem.the  
 c. **toată chestia** era de a aborda problema. (reformulation of example (34) with Cx2)  
 whole thing.the was of to tackle problem.the  
 ‘The whole thing was to tackle the problem.’

A possible explanation to this exclusionary behavior could be that in (34a) *chestia* is used in the same way as a “topic change marker” (cf. Schmid 2000), as in the specificational construction (35). Indeed, in this context *chestia e că* ‘the thing is’ can even occur without the conjunction *că* (36), the same way as *the thing is* or *the bottom line is* (37) are found in English without the complementizer *that* (see Tuggy 1996 and Pinson 2022). *Chestia*, just like *thing* or *bottom line* in English, seems to be on its way to becoming a discourse marker when integrating the specificational construction, and this use is also found in (34a), hence the impossible paraphrase of example (34a) with (34b) above.

- (35) **chestia** e că am greșit cam mult (Twitter example)  
 thing.the is that have.1SG mistaken quite much  
 ‘**The thing** is that I made too many mistakes’  
 (36) **Chestia** e, nu e posibil să mai mori de foame  
 thing.the is NEG is possible SBJV more die.2SG of hunger  
 și să dormi în stradă. (Sketch Engine)  
 and SBJV sleep.2SG in street  
 ‘**The thing** is, it’s not possible to starve and sleep on the street anymore.’  
 (37) But **the bottom line** is, is the country freer, stronger?  
 (example taken from Pinson 2022)

In what follows, we will examine the factors which block one or more of the paraphrases with Cx1, Cx2 and Cx3. We will start by looking at a factor which blocks the paraphrase with only Cx2, before moving on to some factors which block the paraphrase with both Cx2 and Cx3.

#### 4.2.2. Main factor which blocks the paraphrases with Cx2 and Cx2: no internal argument

Let us first have a look at the factor which blocks the paraphrase with Cx2, but not necessarily with the two other constructions. We have found only one such factor.

Among the occurrences annotated as being eligible for a paraphrase with Cx2, we find several cases where the paraphrase is possible if the long infinitive has an object, like *a producției* ‘the production’ (38), or another type of complement, like *prin mixturi locale* ‘through topical ointment’ (39), but not when the object (38b) or complement (39b) is omitted from cx2 (38c-39c).

- (38) a. **Politica urmărită** este **de uniformizare** a producției,  
 policy.the followed.FSG is of uniformising GEN production.the.GEN  
 eliminarea producătorilor particulari și îndoparea  
 eliminating.the.FSG producers.the.GEN particulars and stuffing.the.FSG  
 prostimii de către supermarketuri cu chimicale  
 stupidity.the.GEN of by supermarkets.the with chemicals  
 ‘**The policy that’s being followed (by the supermarkets)** is of **standardizing**  
 production, eliminating particulars from producers and feeding the masses with  
 chemicals’  
 b. **politica urmarită** este **de a uniformiza producția**  
 policy.the followed.FSG is of to uniformizing production.the  
 (reformulation of example (38) with Cx2)  
 c. \*politica urmărită este de a uniformiza  
 policy.the followed.FSG is of to uniformize
- (39) a. **Tratamentul** este **de calmare** prin mixturi locale, antiinflamatoare.  
 treatment.the is of calming through ointments local.PL anti-inflammatory.PL  
 ‘**The treatment is to calm** through topical ointments, anti-inflammatory drugs.’  
 b. **Tratamentul** este **de a calma** prin mixturi locale  
 treatment.the is of to calm through ointments local.PL  
 (reformulation of example (39) with Cx2)  
 c. \*Tratamentul este de a calma  
 treatment.the is of to calm

Conversely, the addition of a reflexive pronoun (*se* ‘themselves’) in (40c), in the paraphrase with Cx2, makes it more acceptable. That shows that, in Cx0, *atotsupune* ‘to submit’ (without the *se*) already had an unaccusative reading. This suggests a certain affinity of Cx0 with unaccusative verbs, while in Cx2 the unaccusative reading needs to be made explicit through the pronoun *se*.

- (40) a. **singura lor atitudine** este **de atotsupunere** față de  
 only.the their attitude is of submission face of  
 nulitățile alese în funcții importante în acest oraș.  
 nullities.the chosen.FPL in functions important.FPL in this city  
 ‘**Their only attitude is to submit** to the daft representatives in charge of this city’  
 b. ?**singura lor atitudine** este **de a atotsupune**  
 only.the their attitude is of to submit  
 (reformulation of example (40) with Cx2)

c. **singura lor atitudine** este **de a se** **atotsupune**  
 only.the their attitude is of to REFL.PRON submit  
 ‘Their only attitude is to submit themselves’

The importance of the presence or absence of an internal argument is even clearer in the following examples which contain the long infinitive *menține* ‘maintaining’, which is a transitive verb, requiring a direct object. When used without an object (41), it is impossible to paraphrase the original example, neither with Cx2 (41b), nor with Cx3 (41c):

- (41) a. **Afacerile** față de care **atitudinea** este **de menținere** și consolidare,  
 affairs.the face of which attitude.the is of maintaining and consolidating  
 gestionarea se face într-o viziune pe termen lung [...]  
 managing.the REFL.PRON makes into a vision on term long  
 ‘For business matters for which **the attitude** is **of maintaining** and of consolidating, management is done with a long-term vision’
- b. ?**Afacerile** față de care **atitudinea** este **de a menține și de a consolida**,  
 affairs.the face of which attitude.the is to of maintain and to of consolidate  
 gestionarea se face într-o viziune pe termen lung  
 managing.the REFL.PRON makes into a vision on term long  
 (reformulation of example (44) with Cx2)  
 ‘For business matters for which **the attitude** is **to maintain and to consolidate**, management is done with a long-term vision’
- c. \***Afacerile** față de care **atitudinea** este **menținerea** și  
 affairs.the face of which attitude.the is maintaining.the.FSG and  
**consolidarea**, gestionarea se face într-o viziune pe termen lung  
 consolidating.the.FSG managing.the REFL.PRON makes into a vision on term long  
 (reformulation of example (44) with Cx3)

However, when used with an object, the paraphrase with Cx2 (and, incidentally, with Cx3) is possible, as shown in (42).

- (42) a. [...] se vede că **intenția** este **de menținere** la nivel maxim  
 REFL.PRON sees that intention.the is of maintaining to level maximum  
 a sentimentelor anti-ruse în Europa și prin orice mijloace.  
 GEN feelings.the.GEN anti Russian.PL in Europe.the and through any means  
 ‘[...] one can see that **the intention** is **of maintaining** anti-Russian feelings in Europe to a maximum level and to achieve this through whichever means.’
- b. Se vede că **intenția** este **de a menține** la nivel maxim  
 REFL.PRON sees that intention.the is of to maintain to level maximum  
 sentimentele anti-ruse în Europa (reformulation of example (43) with Cx2)  
 feelings.the. anti\_Russian.PL in Europe.the  
 ‘one can see that **the intention** is **to maintain** anti-Russian feelings in Europe to a maximum level’

c. Se vede că **intenția** este **menținerea** la nivel maxim  
 REFL.PRON sees that intention.the is maintaining.the.FSG to level maximum  
 a sentimentelor anti-ruse în Europa (reformulation of example (43) with Cx3)  
 GEN feelings.the anti\_Russian.PL in Europe.the  
 ‘one can see that that **the intention** is **the keeping** of anti-Russian feelings in Europe  
 to a maximum level’

Surprisingly enough, it seems that any verb complement of the long infinitive facilitates the paraphrase with Cx2 and Cx3, as exemplified below with an original example (43) that contains a locative complement (*în mai multe locuri* ‘in several places’) post-posed to the infinitive:

- (43) a. **Procedeu** este **de injectare** în mai multe locuri astfel încât să  
 process.the is of injecting in more many places so that SBJV  
 acopere toată zona afectată.  
 covers.SBJV whole zone.the affected.FSG  
 ‘**The process** is **of injecting** in several places so that it covers the whole affected area.’
- b. **Procedeu** este **de a injecta** în mai multe locuri astfel încât să  
 process.the is of to inject in more many places so that SBJV  
 acopere toată zona afectată. (reformulation of example (45) with Cx2)  
 covers.SBJV whole zone.the affected  
 ‘**The process** is **to inject** in several places so that it covers the whole affected area.’
- c. **Procedeu** este **injectarea** în mai multe locuri astfel încât să  
 process.the is injection.the.FSG in more many places so that SBJV  
 acopere toată zona afectată. (reformulation of example (45) with Cx3)  
 covers whole zone.the affected  
 ‘**The process** is **the injection** in several places so that is covers.SBJV the whole  
 affected area.’

Another interesting phenomenon happens when we try to paraphrase with Cx2 some original examples of Cx0 containing 2 coordinated infinitives. Indeed, as we saw with example (41) above, the addition of another infinitive makes the paraphrase with Cx2 more acceptable (41b), as witnessed also in (44b) below.

- (44) a. Un astfel de formular indică dacă **ordinul** este  
 a such of form indicates if order.the is  
**de cumpărare** sau de vânzare [...]  
 of buying or of selling  
 ‘Such a form indicates whether **the order** is **for buying** or for selling [...]’
- b. ?Un astfel de formular indică dacă **ordinul** este **de a cumpăra**  
 a such of form indicates if order.the is of to buy

**sau de a vinde** (reformulation of example (47) with Cx2)

or of to sell

‘Such a form indicates whether **the order is to buy or to sell**’

As expected, when the second infinitive has an explicit object, the resulting paraphrase with Cx2 becomes even more acceptable (45b).

- (45) a. **Procedeu** este **de cumpărare**, **de negociere** cu omul  
 process.the is of buying of negotiating with person.the  
 care a stat acolo.  
 which has stayed there  
 ‘**The process is of buying, of negotiating** with the person who stayed there.’
- b. **Procedeu** este **de a cumpăra**, **de a negocia** cu omul  
 process.the is of to buy of to negotiate with man.the  
 care a stat acolo. (reformulation of example (48) with Cx2)  
 which has stayed there  
 ‘**The process is to buy, to negotiate** with the person who stayed there.’

This shows that, when compared to Cx2 and Cx3, the long infinitive copular construction (Cx0) can work more easily with infinitives used without any complement. This comes from the ability of Cx0 to integrate verbs which can omit their internal argument (section 4.2.2.) or which are intransitive (section 4.2.3.), contrary to Cx2 and Cx3. This has led us to investigate more thoroughly to what extent this construction (Cx0) favors the use of verbs which allow the verb argument to be omitted. The results of this final investigation is presented in the following section.

### 4.3. The proportion of transitive and intransitive verbs in our dataset

In line with our goal to describe the long infinitive copular construction, we wanted to determine the kinds of verbs which participate to it. Given the fact that the construction under study allows verbs appearing without any argument, we expected the proportion of intransitive verbs to make up most of the infinitives in Cx0. In order to investigate this final question, we have annotated several other aspects of the examples in our dataset, such as the presence and the kind of argument (section 4.3.1.), the proportion of intransitive verbs used without any argument (section 4.3.2.) and the overall proportion of transitive vs. intransitive verbs (section 4.3.3.).

#### 4.3.1. Overall results: a predominantly omitted argument

As expected, occurrences without a verb argument (whether the verb is transitive, and thus the patient is omitted, or whether the verb is an intransitive one) make up more than half of our dataset.

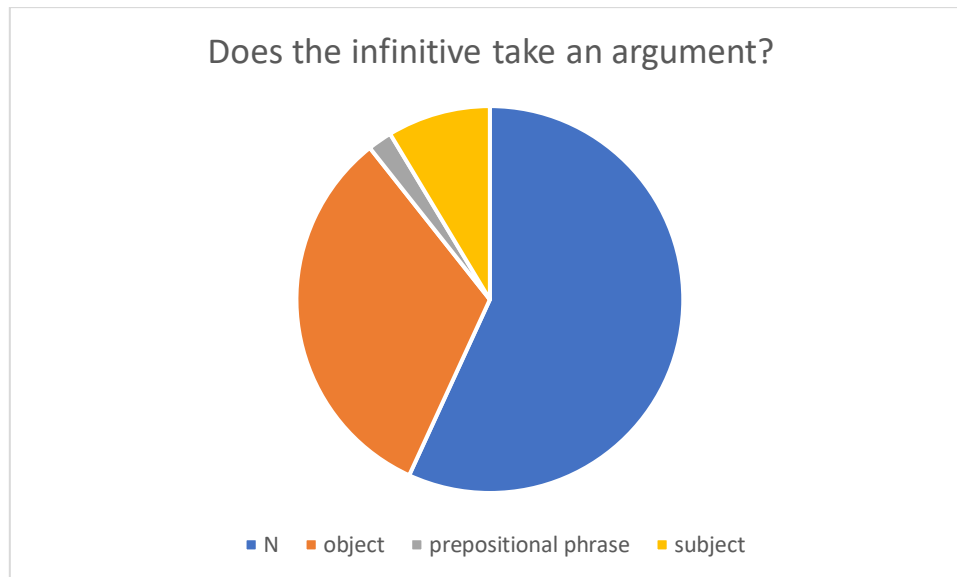


Figure 1. *The proportion of long infinitives without an argument compared to the rest*

The labels used to determine the type of argument in *Figure 1* above are the following:

- 1- “N”, when the argument is absent;
- 2- “object”, when the argument is an object, be it a genitive NP (46), an adnominal prepositional phrase (47) or even a direct object (48);
  - (46) În aceste condiții, **tendința este de defrișare a acestor suprafețe.**  
 in these conditions tendency.the is of clearing GEN these.GEN surfaces  
 In these conditions, **the tendency is to clear those surfaces.**
  - (47) **Acuzația este de deturnare de fonduri.**  
 accusation.the is of diverting of funds  
 ‘**The accusation is of embezzlement.**’
  - (48) [...] **contractul este de furnizare produse și are o valoare estimată sub 100.000 euro.**  
 contract.the is of furnishing products and has a value estimated.FSG under 100\_000  
 euro  
 ‘**This contract is a furnishing products contract and its estimated value is under 100,000 euros.**’
- 3- “prep”, when the complement is headed by a preposition other than *de* ‘of’ (49);
  - (49) **Tendențele sunt de trecere către full HD [...]**  
 tendencies.the are of passing towards full HD  
 ‘**The tendencies are of transitioning to full HD [...]**’

4- “subject” when the argument is a subject (50), encoded as a genitive NP.

- (50) **Efectele sunt de încetinire a retragerii gingiilor...**  
 effects.the are of slowing GEN retraction.the gums.the.GEN  
 ‘The effects are of slowing down the retraction of the gums...’

However, although these results seem to confirm the patient-backgrounding function of Cx0, *Figure 1* above only gives a general overview of the valency behavior of the long infinitives involved in our dataset, without taking into account the different types of verbs. Indeed, since we have seen that Cx0 offers the possibility of backgrounding the internal argument, compared to other constructions, it would be more telling to examine more closely the transitive verbs, which usually require an object.

#### 4.3.2. The proportion of transitive infinitives omitting the argument

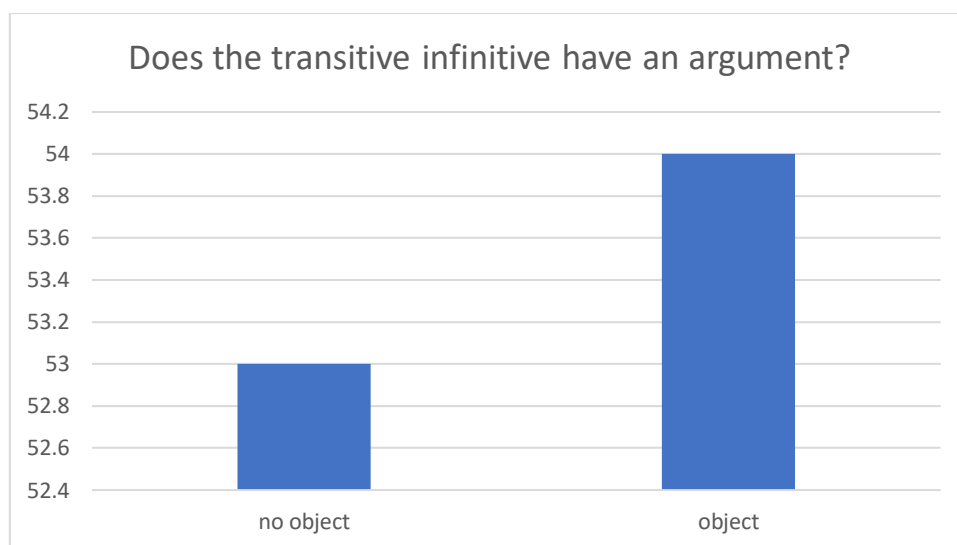


Figure 2. *The proportion of transitive infinitives without an argument (left) compared to the rest*

This proportion seems non-negligible (over a third). We give below some examples of the most frequent transitive infinitives for which the speaker harnesses the opportunity given by the construction to background the patient omitting the argument.

- (51) [...] dacă **obiectivul** este **de promovare** și dacă subiectul campaniei  
 if goal.the is of promoting and if subject.the campaign.the.GEN  
 este „mass media” [...] is mass\_media  
 ‘[...] if **the goal is promotion** and if the subject of the campaign is “mass media” [...]

- (52) Super articolul... eu acum vin de la sala de fitness... greutate  
 super article.the me now come.1SG from to room.the of fitness weights  
 și bandă de alergare... **ideea e de întreținere**, de 2-3 ori pe săptămână  
 and treadmill idea.the is of maintaining of 2 3 times on week  
 și maxim[um] 40-50 de minute... + mers la bazin de înot  
 and maximum 40 50 of minutes + going to swimming\_pool  
 și acasă cu bicicleta :) spor! numai bine!  
 and home with bicycle :) encouragement only good  
 ‘Great article... I’m just back from the gym, weights and treadmill... **The idea** is of **maintaining** [one’s body] by exercising 2-3 times a week for maximum 40-50 minutes... + going to the pool and using the indoor bike :) Good luck! Take care!’
- (53) Taninul are proprietatea de a se fixa pe colagen și elastina,  
 tannin.the has propriety.the of to REFL.PRON fixate on colagen and elastin  
**efectul fiind de conservare.**  
 effect.the being of conserving  
 ‘The tannin has the property of latching itself onto the collagen and the elastin, **the effect** being of **conservation.**’
- (54) **Procedeu** este **de cumpărare**, **de negociere** cu omul  
 process.the is of buying of negotiating with person.the  
 care a stat acolo.  
 which has stayed there  
 ‘**The process** is of **buying**, of **negotiating** with the man who has stayed there.’

Nevertheless, in some cases, the patient is heavily implied, like in (52), where it is clear that it is the body which needs to be maintained through physical exercise. In other cases, the patient can be recovered contextually, like in (53), where it is the collagen and the elastin which the tannin helps conserve, by latching onto it; and in (54), the previous sentences should make it clear which kind of patient the transactions and negotiations concern.

#### 4.3.3. *Transitive verbs without their internal argument*

Our initial expectation (see section 2.1.) was that the long infinitive would show a preference for constructions without a complement, with either intransitive verbs or with transitive verbs whose subject has been omitted. So far, we have seen that complements are indeed more often omitted than not, which partially confirms our initial expectation.

Now, in order to test this expectation with regard to the proportion of transitive verbs, we have computed the occurrences of transitive vs. intransitive verbs in our dataset. Do note however that, given the very low frequency of unergative verbs (7 occ.), we have chosen not to account for the distinction between unaccusative and unergative verbs, which have been grouped together as intransitive verbs.



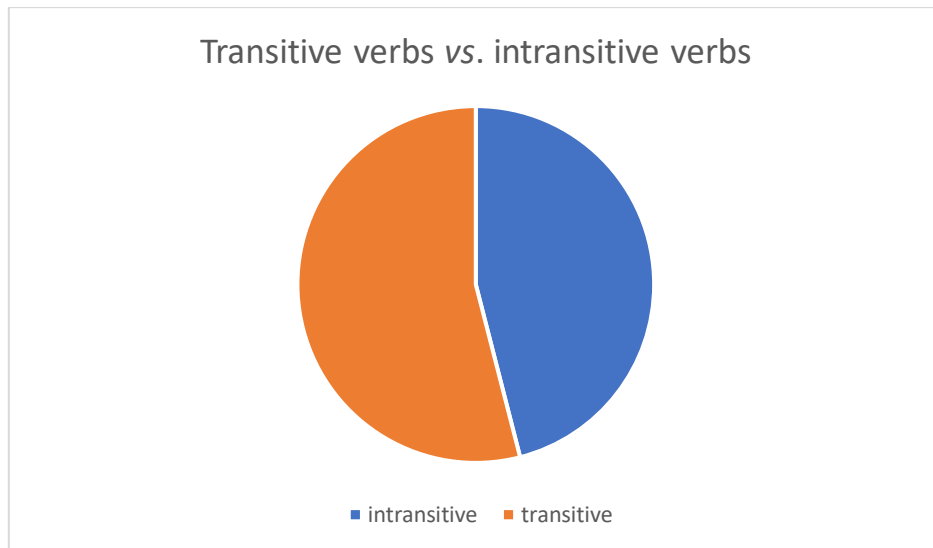


Figure 3. *The proportion of transitive vs. the 2 types of intransitive verbs (unaccusatives and unergatives) in our corpus*

The results, which seemingly contradict our initial hypothesis, actually further confirm it. Indeed, since *Cx0* seems to prefer constructions where the long infinitive is derived from a transitive verb, yet where the argument is omitted, we can conclude that one of the features of this construction is to omit the patient or any other complement. This is also confirmed by the low number of unergative verbs in our dataset: the only argument prone to omission is the external argument (55), but in all types of nominalizations external arguments are overall more omissible. Again, the majority of our data show transitive verbs, *i.e.* verbs which usually take on an argument, yet said argument is mostly absent from our data (see *Figure 1* above).

- (55) **[P]roblema e de comunicare...**  
 problem.the is of communicating  
 'The problem is a communication problem...'

All in all, these figures, together with the proportion of transitive and intransitive verbs entering *Cx0* (see *Figure 3* below), further confirm the backgrounding potential of this construction. The hypothesis according to which *Cx0* will show a preference, in the infinitive slot, for verbs without an argument is therefore borne out.

## 5. CONCLUSIONS

After a careful analysis of our dataset, we have come to several conclusions about the specificational long infinitive construction (*Cx0*). First, contrary to the constructions with the short infinitive (*Cx2*) and the definite long infinitive construction (*Cx3*), it allows in its

infinitive slot either intransitive verbs, which do not require an object, or transitive verbs used without an object. This suggests that this construction leads to a backgrounding either of the patient, in the case of transitive and unaccusative verbs, or of the agent, in the case of unergative verbs. However, since unergative verbs are quite rare in our corpus, the backgrounding mostly concerns the patient.

As for the proportion of transitive vs. intransitive verbs in the construction under study, transitive verbs are predominant. Since these occur mostly without their object, this, again, comes to support our initial hypothesis according to which the construction has a backgrounding function. In other words, the specificational long infinitive construction seems to be a form of nominalization which is used by the speaker precisely when the goal is to omit the patient of the process denoted by the verb.

Finally, regarding the semantics of the shell noun lexemes used in this construction, several observations have been made. The typical shell nouns occurring most frequently in the noun slot of the construction are associated with the following modalities: 1- alethic (nouns like *acțiune* ‘action’, *activitate* ‘activity’, *fenomen* ‘phenomenon’, *rezultat* ‘result’, *trend* ‘trend’), 2- epistemic (nouns like *risc* ‘risk’, *tendență* ‘tendency’), 3- relative (nouns like *etapă* ‘step’, *strategie* ‘strategy’) or 4- boulomaic (nouns like *idee* ‘idea’, *obiectiv* ‘goal’). However, not only typical shell nouns can enter the noun slot of the specificational long infinitive construction: there are also borderline cases of coercion of the noun. In this case, the long infinitive no longer expresses the propositional content of the noun, but triggers different semantic effects: 1- a goal reading of the infinitive, 2- a personal definition of the speaker, or 3- a subtype reading of the noun.

#### LIST OF ABBREVIATIONS

|           |                            |
|-----------|----------------------------|
| ACC       | accusative                 |
| DAT       | dative                     |
| F         | feminine                   |
| GEN       | genitive                   |
| M         | masculine                  |
| NEG       | negation                   |
| PL        | plural                     |
| PREP      | preposition                |
| REFL.PRON | reflexive pronoun          |
| SBJV      | subjunctive                |
| SG        | singular                   |
| Q         | question particle / marker |

#### REFERENCES

- Adler, S., D. Legallois, 2018, “Les noms sous-spécifiés dans le débat parlementaire: Analyse fréquentielle et catégorisation modale”, *Langue française*, 198, 2, 19–34.
- Cuniță, A., 2015, « „Cuvintele-cochilie” atitudinale și funcțiile lor în discurs », in I. Nedelcu, R. Zafiu (eds), *Variația lingvistică: probleme actuale. Actele celui de-al 14-lea Colocviu Internațional al Departamentului de Lingvistică*, București, Editura Universității din București, 177–185.
- Francis, G., 1986, “Anaphoric Nouns”, *Discourse Analysis Monographs*, 11.

- Grimshaw, J. 1990, *Argument structure*, Cambridge, Mass., The MIT Press.
- Legallois, D., 2006, “Sur quelques caractéristiques des noms sous-spécifiés”, *Scolia*, 23, 109–127.
- Legallois, D. and Gréa, Ph., 2006, “L’objectif de cet article est de... : construction spécifique et grammaire”, *Les Cahiers de Praxématique*, 46, 161–186.
- Legallois, D., 2008, “Sur quelques caractéristiques des noms sous-spécifiés”, *Scolia*, 23, 109–127.
- Legallois, D. and Vajnovszki, A., 2021, “Proposition pour une conception modale et aspectuelle des noms dits sous-spécifiés”, *Corela (Cognition, représentation, langage)* [Online], HS-34.
- Novakova, A., Sorba, J. A. Tutin, 2018, “Les noms généraux *sentiment* et *émotion*”, *Langue Française* 198, 83–98.
- Pinson, M., 2023, “The (inter)subjectification of *bottomline* phrases”, *Linguisticae Investigationes*, 45, 2, 276–295.
- Roze, C., T. Charnois, D. Legallois, S. Ferrari, M. Salles, 2014, “Identification des noms sous-spécifiés, signaux de l’organisation discursive”, *Proceedings of TALN 2014 (Volume 1: Long Papers)*, 377–388.
- Schmid, H.-J., 2000, “English Abstract Nouns As Conceptual Shells: From Corpus to Cognition”, Berlin, Mouton De Gruyter.
- Schmid, H.J., 2018, “Shell nouns in English – a personal roundup”, *Caplletra*, 64, 109–128.
- Tuggy, D.H., 1996, “The thing is that people talk that way. The question is why?”, in E. H. Casad (ed.), *Cognitive Linguistics in the Redwoods*, Berlin, Mouton de Gruyter, 713–752.
- Vergaro, C., 2018, *Illocutionary shell nouns in English*, Berne, Peter Lang.

