SUBJUNCTIVE AND INFINITIVE CLAUSES IN ROMANIAN. THE VIEW FROM COROLA

MARIA AURELIA COTFAS[1]

Abstract. The paper looks at the distribution of finite and non-finite subject clauses in Romanian and how these are represented in the CoRoLa corpus. More particularly, we aim to investigate how the corpus can be effectively and efficiently used to either establish or enrich contexts of distribution or otherwise to confirm previous analyses on the distribution and properties of our analyzed structures (triggers, presence or absence of complementizers, type and interpretation of embedded subject, properties of the embedded predicate, control vs raising, etc.)

Keywords: infinitive, subjunctive, subject, (non-)agreeing, corpus data, query.

1. INTRODUCTION & SCOPE OF ANALYSIS

The paper looks at the distribution of finite (subjunctive) and non-finite (infinitive) subject clauses in Romanian (object clauses are excluded, for space considerations). Of course, a lot has been written on the topic. Our purpose here is not to give an overview of these or to propose a novel approach, but to look at these contexts through the lens of the CoRoLa corpus and to see what it can reveal about subject clauses in Romanian and how the corpus can be utilized as an efficient tool in the study of particular linguistic contexts. More precisely, our goal is to investigate how efficient and easy-to-use search queries can be built, such that they can lead to relevant examples meant to shed more light on the typical distribution and properties of subject clauses.

Corpora can be used both to establish and systematize contexts of distribution, as well as to confirm – and gather more evidence to back up – already known classifications or theories. Thus, our purpose is twofold: on the one hand, starting from search algorithms as clearly defined as possible, we will attempt a classification of main clause triggers; on the other hand, starting from previous analyses, we mean to analyze the compatibility of main predicates with different clause sizes and the restrictions they impose on the complement and verify whether or to what extent previous assumptions are confirmed.

The structure of the paper is as follows: Section 2 gives a short overview of the complementation system in Romanian, a daunting task, given the complexity of the topic and the many variables at work (finiteness, internal struture (i.e., number and type of complemenizers, type of available subjects, temporal reference, etc.)). Section 3, the bulk of the paper, looks at how finite (subjunctive) and non-finite (infinitive) subject clauses are reflected in the CoRoLa corpus (sections 3.2. and 3.1., respectively), focusing on issues

[1] University of Bucharest, maura.cotfas@gmail.com; maura.cotfas@lls.unibuc.ro

RRL, **LXIV**, *3*, p. 293–306, Bucureşti, 2019

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

concering the queries for establishing their distribution (be + adjective, be + past participle of transitive verbs), an attempt at a classification of the typical adjectives and verbs in the main clause, observations concerning agreeing vs non-agreeing variants in the be + past participle distribution (i.e., whether the participle agrees or not in number and gender with the embedded subject), findings concerning clause size, (preference for) presence vs absence of complementizers, type of embedded subject, etc. Section 4 draws the conclusion.

2. A BRIEF OVERVIEW OF ROMANIAN COMPLEMENTATION

As already stated, this is a very sketchy presentation, given the amount of complexity underlying Romanian complementation, for both finite and non-finite clauses. As far as non-finite clauses are concerned, Romanian has two main types of clauses, each of which presents interesting sub-classifications.

Even though, diachronically, the infinitive has been mostly replaced by the subjunctive (the *conjunctiv*), especially as complement to verbs, it is still a productive construction in Modern Romanian, particularly in the formal register, in temporal adjuncts, in argumental subject position and/or as complement to various adjectives. In object position, it mostly survived as a bare infinitive (see below) in the complement position to dynamic or circumstantial modal verbs (*a putea* 'can', *a şti* 'know (how)'). Infinitive clauses can be classified in three types, according to their make-up or internal structure. Thus, we distinguish between a) the bare infinitive, b) the *a*-infinitive and c) the *de*-infinitive, all of which will be discussed below.

The second type of non-finite clause in Romanian is the supine. For space considerations, we do not include any data on supines in our analysis. Supine clauses have a typical make-up: they display a (prepositional) complementizer *de*, followed by the (invariable) past participle form of the verb. They, too, can appear as subjects and objects and fall into two sub-types: Restructuring and Non-Restructuring[2].

Let us now briefly come back to infinitive clauses. The three types are illustrated below. As already stated, the bare infinitive in (1) has a very limited distribution as complement to ability modals. The example also illustrates the four conjugations:

(1) Pot cînt*a*$_{\text{ConjI}}$ /prind*e*$_{\text{ConjII}}$ fluturi/ tăc*ea*$_{\text{ConjIII}}$/cit*i*$_{\text{ConjIV}}$
'I can sing /catch butterflies /shut up /read

The long infinitive or the *a*-infinitive displays an extra infinitive particle *a*, followed by the base form of the verb, as already shown in (1). The *a*-infinitive can occupy both subject and object positions, as shown in (2a) and (2b). The status of *a* as a complementizer or modal particle (similar to subjunctive *să*) has been widely discussed in the literature.

[2] For more details on the properties of the two types of supine clauses, see Cornilescu & Cosma (2014)

(i) E important /bine de mers acolo. (subject, after evaluative predicates (be+adj))
is importand/good DE go-part there

(ii) Am terminat de citit. (object, after aspectual (terminative) verbs)
I have finished DE read-part

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

Given that the structure of the Romanian infinitive allows an extra, complementizer-like element, i.e., *de*, we take *a* as a modal particle, on a par with *să*.

(2) a. A citi e o binecuvântare./b. Am început a citi.
'To read is a blessing.' / 'I have started to read.'

The third type of infinitive, the *de*-infinitive**,** appears mainly as complement to nouns and adjectives, but it can also be selected by some verbs and predicates.

(3) a. nevoia de a citi / b. fericit de a citi/c.Este bine de a citi./d. A încetat de a cânta.
the need DE A read / happy DE A read /is good DE A read/ has ceased DE A to sing

In this paper we focus on subject infinitives (like (2) and (3) above).

Turning now to finite complementation, indicative clauses are, as is expected, the first type that comes to mind. Their inflectional paradigm (i.e., person and number agreement) varies according to each of the four conjugations illustrated above. They may be subjects as well as objects, mainly after *verba dicendi* and extensional or weakly-intensional verbs, i.e., those which commit the speaker to the truth of the embedded proposition, thereby accounting for mood choice. The indicative is visible not only on inflection, but on the complementizer also: there is a specific indicative complementizer *că*, which, unlike *ca*, the typical subjunctive complementizer (see below), can never be deleted. However, just like *ca*, but unlike the modal particle *să*, the indicative *că* does not form a cluster with the embedded verb. Namely, full-fledged nouns (such as the embedded subject *Maria* in (4b)) can appear between *că* and the indicative verb.

(4) a. Cred *(că) Maria e foarte deşteaptă. b. Au spus că Maria a plecat devreme.
'I think that Maria is very smart.' 'They said that Maria left early.'

As for the subjunctive, unlike its West Romance sisters, Romanian marks it analytically via a specific mood particle *să*, analyzed as deriving from the Latin *si* (DEX 1998: 949). Except for the particle, subjunctive inflection is poor, being visible only on the 3rd person singular and plural, cf. (5):

(5) (el, ea/ei/ele) pleac***ă*** / să *pleacă/să plec***e**
(he/she/they) leave-ind SĂ leave-ind/leave-**subj**

Also, as already hinted at above, there is a subjunctive complementizer *ca*, different from the indicative *că* and further proof that in Romanian mood selection is visible higher than in other Romance, i.e., it is marked on the introductory element.

In modern (standard) Romanian, *ca* and *să* may but need not co-occur: the complementizer may not be lexicalized, but when the two do co-occur, they are never adjacent. Namely, there must always be lexical material in between:

(6) Vreau *să* plec***e*** toată lumea mâine / ***ca*** mâine ***să*** plec***e*** toată lumea.
want SBJV leave all people tomorrow / that tomorrow SBJV leave all people

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

'I want everyone to leave tomorrow'

Simplifying matters to some extent and taking into account only lexical(ized) material, we can thus divide the class of Romanian subjunctives into two (sub-)types: *să*-subjunctives and *ca*-subjunctives. The former are those complementizer-less clauses where *să* forms a cluster with the verb, such that no full-fledged element can intervene. The subject is always post-verbal (7), whereas the latter are clauses introduced by *ca*, inevitably followed by the *să* + verb cluster, but always with overt material (adverbs, topicalized objects, the embedded subject itself) in between, as already shown above and below in (8).

(7) Vreau să vină Ion /* să Ion vină.
want SBJV come Ion / * SBJV Ion come

(8) Vreau **ca** Ion **să** plece mâine /**ca** mâine **să** plece.
want that Ion SBJV leave tomorrow/ that tomorrow SBJV leave-3.sg/pl

From the finite class, we will focus on subjunctive clauses and not on indicatives. This is because, for our present purposes, the subjunctive/infinitive competition is more interesting than the subjunctive/indicative one. As East Romance and part of the 'Balkan Sprachbund', Romanian is a 'subjunctive language', unlike its Romance sisters: the subjunctive has diachronically come to replace the infinitive (especially) in complement positions, restraining the distribution of the latter, which in Modern Romanian survives in legal, official texts and with a limited number of predicates. The Romanian subjunctive is therefore the counterpart of the infinitive in 'infinitive languages', i.e. those that use infinitives extensively after main predicates (e.g., West Romance, Germanic, etc.). For more details on infinitive vs subjunctive languages, see Cotfas (2012).

3. ROMANIAN SUBJECT CLAUSES IN THE COROLA CORPUS

3.1. Infinitival Subject Clauses

In our analysis of subject infinitives in the CoRoLa corpus, we started from a few basic facts. Firstly, in point of make-up, what we were to look for were either *a*-infinitives or *de-a*-infinitives – as only these last two can also appear as subjects. Secondly, as regards the type of embedded subject, infinitives are known to either appear subjectless – in which case we assume there to be a null subject, interpreted either generically or as anaphoric to (or 'controlled' by) the main clause subject – or with their own subject. Thirdly, in what concerns typical triggers, we focused on predicates featuring the copula or auxiliary use of *a fi* 'to be' and investigated two main types of distribution, one involving copula + adjective, the other one involving passive be + the past participle form of transitive verbs (i.e., passivized transitive triggers). Relevant examples are given below, in (9):

(9) a. Este important [(de) a spune adevărul] (be + adj; null subject)
is important (DE) A say the truth
b. Este esențial [(de) a veni toți la aceeași oră] (be + adj; overt subject)
is essential (DE) A come all at same hour

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

c. Era stabilit [(de) a transfera circulația în suburbii] (passive trans.V; null subject)
was planned DE A transfer traffic-the in suburbia
d. Era planificat [(de) a sosi fiecare neînsoțit] (passive trans. V; overt subject)
was planned (DE) A arrive each unaccompanied

An interesting 'complication' for the be + past participle distribution is that sometimes the embedded subject appears in the main clause (the subject is 'shared' by the two predicates) and agrees in phi-features (number and gender) with the main clause predicate. Thus, (10a) below contrasts with (10b) in that there is no agreement in (10a) (the main clause verb appears with default 3rd person agreement and the subject remains embedded), whereas in (10b) the plural noun subject of the embedded predicate agrees in number and gender with the main clause (passivized) verb (see the plural marking on the auxiliary and the feminine plural morphology on the past participle. For simplicity of exposure, let us call these the agreeing (10b) vs non-agreeing variants (10a):

(10) a. **Este/A fost** estimat [a avea loc mai multe demonstrații].
is /has been estimated to take place several demonstrations
b. Demonstrațiile **sunt** estimat***e*** a avea loc mâine.
Demonstrations-the are estimated to take place tomorrow

Now let us define and establish our corpus research questions. First off, as far as the be + adjective distribution is concerned, what we aim to get from the corpus is a) whether a classification into classes of adjectives (according to their semantics) is possible or at least feasible and b) whether a correlation can be established between the adjectival class and the (preference for the) presence vs absence of the complementizer "de".

Secondly, with respect to the be + past participle distribution, a similar attempt at a classification of typical verb classes is desirable (though perhaps a bit more strenuous to achieve), alongside findings concerning a) the compatibility or preference of a given verbal class with/for the agreeing or non-agreeing variant and b) the correlation between the (preference for) presence vs absence of "de" and the (non-)agreeing variants (i.e., can we get sufficient and convincing evidence that agreeing variants, as instances of raising/backward agree, are smaller in size than their non-agreeing counterparts and are therefore compatible with *a*- rather than *de*-infinitives?). Last but not least, there is also the issue of register: considering the relevant data and the fact that the corpus provides extensive contextual evidence, what is there to be said (or added) about the typical occurrence contexts of our infinitives? Let us tackle these questions in turns.

For the first investigated distribution context, the search algorithm we used is illustrated in (11). We aimed to find as many examples as possible of this type and sift through the data in order to establish a classification of those adjectives which, when in combination with the copula, function as typical main clause triggers for subject infinitives.

(11) [drukola/m=ctag:v3 & drukola/m=person:third & drukola/m=tense:present]
[drukola/m=ctag:r][drukola/m=pos:particle & drukola/m=type:infinitive]

A thorough analysis of the examples yielded by the search query in (11) allowed us to come up with a three-fold classification of adjectives into modal, emotive-evaluative and

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

temporal – with the provision that the last class is more restrained and may sometimes feature adverbs instead of adjectives.

The class of 'modal' adjectives consists of items which express modal notions such as (lack of) obligation, necessity, interdiction, (lack of) permission: necesar 'necessary', (im)posibil '(im)possible', (im)probabil '(im)probable/ (un)likely', permis 'permitted', interzis 'forbidden', obligatoriu 'obligatory', a.s.o.. The great majority of these examples comes from the judicial jargon (i.e., OUG, Eur-Lex), so they are confined to the written, formal literary style. In (12), a tentative ranking is offered:

(12) necesar (1,417 occurences) > posibil (cca 319) > permis (95) > obligatoriu (56)) > imposibil (36) > interzis (33) > probabil (16) > improbabil (2)

The corpus data shows that some of these modal adjectives frequently co-occur with dative experiencer arguments in the main clause (particularly *imposibil* 'impossible') and *permis/interzis* 'permitted/forbidden'). This Experiencer nominal, usually doubled by a pronominal, is the one that controls the reference of the embedded null subject (13). Moreover, *a fi* 'be' may alternate with other copula-like verbs (14).

(13) ...suferea că-*i* este imposibil [a-i schimba destinul..]
suffered because to him was impossible A change her destiny

(14) a. se dovedeşte imposibil a se respecta dispoziţiile ...
proves impossible A se-pass respect dispositions the …
b. devine necesar a se face aplicarea prevederilor
becomes necessary A se-pass make applicable the decisions

The second adjectival class is probably the bulkiest, as under the "evaluative" label we include emotive and manner adjectives (*important, indispensabil, util, firesc, vital, minunat, extraordinar, nemaipomenit, bine, normal, esenţial, dificil, potrivit, suficient (= de ajuns), deprimant, adecvat, impropriu, omeneşte, corect, nimerit, binevenit, descurajant, periculos, cumplit* a.o. 'important, useful, natural, vital, wonderful, extraordinary, good, normal, essential, correct, welcome, dangerous, etc.); *uşor, greu, dificil, lesne, pripit, solicitant, epuizant, obositor*, a.o. 'easy, difficult/hard, quick, tiring, exhausting'.

These examples are no longer restricted to the law jargon, but the register remains formal, nonetheless, as they mostly come from literary texts or from newspaper excerpts. In point of frequency, (15) gives an overall picture, with occurrences in between brackets:

(15) i. bine (216) > important (45) > util & suficient (a bit over 30/each) >
corect (22) > esenţial, destul (cca 5 each) > firesc, periculos (2) > indispensabil, periculos (1),
ii. dificil (40) > greu (37) > uşor (22) > lesne ("e lesne" 11) > pripit (1)

As with the first class, some of these adjectives may also co-occur with dative experiencer nominals (Bene-/Male-Ficiary) in the main clause, particularly manner adjectives. This Experiencer controls the reference of the embedded null subject, as shown in (16), where the be-copula has been replaced by another copula-like verb, *a veni* (come):

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

(16) […] *îmi* vine /e(ste) greu/uşor [a scrie]
to me-Dat comes/ is difficult/easy [to write]

The search query in (11) has also yielded examples of the type in (17), where instead of an adjective we actually have an adverb (*devreme* 'early')[3]. Other such temporal adverbs may be *târziu*, 'late' or *curând* 'soon', for example (CoRoLa has revealed no example with the last one), or even the adjective *tardiv* 'belated'. Although adverbs seem predominant, adjectives are not excluded. The overall scarcity of this class shows that the infinitive in this distribution is obsolete and likely to disappear. This goes in hand with the fact that the very few examples of this type are confined to the literary register.

(17) Este (prea) devreme a face urări
is too early to make well-wishes

Overall, what we notice from a careful consideration of the be + adjective distribution is that modal adjectives are the most frequent (though not the class with the most numerous members). This could be accounted for by the fact that these examples come from legalese, a field where infinitives have survived at ease.

Quite often, it seems that our infinitives appear with null subjects, interpreted either generically (notice the embedded (*se*-)passives) or retrieved from the larger context, sometimes via a main clause nominal in the dative (particularly with modal and manner adjectives). Also, even though 'be' is predominant, it may on occasion alternate with other copula-like verbs, though specific compatibilities are probably due to idiosyncrasies.

As to the compatibility of be + adjective with a *de-a*-infinitive, we have modified the search query to accommodate the presence of the complementizer:

(18) [drukola/m=ctag:v3 & drukola/m=msd:vmip3s & drukola/m=number:singular & drukola/m=person:third &drukola/m=type:main][drukola/m=ctag:r] [drukola/m=case:accusative & drukola/m=ctag:s & drukola/m=wordformation:compound][drukola/m=pos:verb

Many of the hits obtained were, unfortunately, not relevant, because they represented cases of lexical selection and hence had to be discarded (i.e., cases where the adjective itself selects the preposition 'de', such as *departe de a, gata de a, demn de a*, etc.). As an alternative, we conducted a search with individual adjectives (+ *de-a*-Infinitives), which turned extremely few results (cf. a scrutiny of the first 16 pages of CoRoLa). As it turns out, therefore, *de*-infinitives are much less productive in the be + adjective distribution, probably because our adjectives are highly compatible with the *de*-supine, which creates a competition of sorts. While less frequent, however, *de*-infinitives are not

[3] This is probably due to the fact that the query in (11) does not specify specific inflection (in terms of number and gender) for the adjective. That is, our search is set to look for examples where what we take for an adjective appears in the default (singular, masculine) form. Thus, there is no way for the system to differentiate between a masculine, singular adjective like *crud* or *tardiv* and an adverb like *curând* or *devreme*. Had we set the query to search only for inflected forms of the adjective, we would probably not have come across examples featuring adverbs. But looking for examples with inflected adjectives was not deemed necessary for the purposes at hand.

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

impossible, and appear with generic null embedded subject, as shown in (19), all from CoRoLa (necesar '*necessary*', again, is the most frequent).

(19) a. este util de a pregăti … / b. Este necesar de a încuraja eforturile..
is useful DE A prepare is necessary DE A encourage efforts-the..
c. Este important de a accentua că … /d.nu e ușor de a rezista în fața …
is important DE A emphasize that.. / not is easy DE A resist before …

Let us now finally turn our attention to the second type of distribution, the one involving passivized transitive main predicates. We will focus here less on the classification of the predicates into semantic classes and more on the distribution and behaviour of the agreeing vs non-agreeing variants illustrated above, under (10). Syntactically, the non-agreeing variant corresponds to a control infinitival whose (null) subject is either controlled by the main subject or otherwise understood as arbitrary (i.e., generic), while the agreeing variant corresponds to a raising (or backward agree) analysis whereby the two predicates share the same subject, such that the main predicate agrees in number and gender with the embedded subjects (which more often than not appears 'raised' in the main clause).

Traditional analyses of control vs raising assume different clause sizes of the complement, such that control infinitivals are more robust than raising ones, which need to be transparent or 'shrunken' domains, so that agreement (or raising of the embedded subject) is facilitated. Our expectation therefore is that the non-agreeing variant would be more compatible with a *de-a*-Infinitive than agreeing variants.

In order to investigate the compatibility of both the agreeing and the non-agreeing variants with an *a*- and a *de*-infinitive, we needed to create carefully constructed search queries, meant to reflect the two types of distribution. That is, we have 4 possibilities:

a) non-agreeing + a-Inf (este/era/a $\text{fost}_{\textbf{sg}}$ + $\text{past participle}_{\text{sg}}$ + **a-Inf**)
b) non-agreeing + de-Inf (este/era/a $\text{fost}_{\textbf{sg}}$ + $\text{past participle}_{\text{sg}}$ + **de-a-Inf**)
c) agreeing + a-Inf (sunt/erau/au $\text{fost}_{\textbf{pl}}$ + $\text{past participle}_{\text{pl}}$ + **a-Inf**)
d) agreeing + de-a-Inf (sunt/erau/au $\text{fost}_{\textbf{pl}}$ + $\text{past participle}_{\text{pl}}$ + **de-a-Inf**)

In (20), we have the query used for a) above.

(20) [drukola/m=ctag:va3s & drukola/m=**number:singular** &drukola/m=**person:third** & drukola/m=type:auxiliary & drukola/m=verbform:indicative]
[drukola/m=ctag:vpsm & drukola/m=**gender:masculine** & drukola/m=**number:singular** & drukola/m=type:main
& drukola/m=verbform:participle][drukola/m=ctag:qn & drukola/m=pos:**particle** & drukola/m=type:**infinitive**]

What is immediately observable from the examples yielded by the system following the query above is that it is actually difficult to tease apart non-agreeing from agreeing variants, since the query provides examples of the type in (21a, b) below, as well as of the type in (22a, b). Even though similar on the surface, the difference between (21) and (22) is subtle, as the former represent instances of non-agreeing infinitives, whereas the latter are examples of agreeing variants where the agreement just happens to be with a singular and masculine subject, given in italics ('the movie' in (22a), 'the fish' in (22b)).

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

(21) a. (Pe mese) era stabilit a *se* aşeza.....
(on tables) was settled to be laid...
b. (pt elevi) este recomandat a *se* înființa……
(for students) is recommended to be set up

(22) a. *Filmul* este aşteptat a avea premieră ...
movie-the is expected to have premiere ...
b. *Peştele* este considerat a fi şi un aliment care...
fish-the is considered to be too a product which....

On closer inspection, though, the above examples can be teased apart if attention is paid to other relevant details, such as the presence of embedded passive morphology in the non-agreeing variants (*se*-passives rather than *fi*-passives – see the italicized items in (21a, b) above, both with embedded transitive verbs (*lay, set up*)). The presence of the passive both on the main and on the embedded infinitival predicate shows that we are dealing here with disjoint (null) subjects, which is also in accordance with the type of selecting predicates (*decide, recommend*), i.e., futurate verbs of the mandative class, which, according to traditional analyses of control, take tensed complements (i.e., infinitives which are future-oriented). Conversely, in (22) we have an overt subject which furthermore appears displaced in the main clause; there is no passive in the infinitive clause, which actually features unaccusative predicates. Moreover, the matrix predicates are no longer futurate verbs, but believe-type verbs (non-factive verbs of propositional attitude).[4] Overall, therefore, the formula in (20) may yield either variant, but there are other means to settle whether we are dealing with a raising/agreeing variant where agreement is with a masculine and sigular shared subject, as it would be the default case for non-agreeing variants.

In (23), we give the query meant to test the compatibility of non-agreeing variants with a *de*-Infinitive (the possibility listed in b) above). This yielded few results, mostly with futurate-verbs as the one in (24):

(23) [drukola/m=ctag:va3s & drukola/m=**number:singular** &drukola/m=**person:third** & drukola/m=type:auxiliary] [drukola/m=ctag:vpsm & drukola/m=**gender:masculine** & drukola/m=**number:singular** & drukola/m=verbform:participle][drukola/m=case:accusative & drukola/m=**type:preposition**][drukola/m=ctag:qn & drukola/m=pos:**particle** & drukola/m=type:**infinitive**]

(24) Era preconizat de a avea cca. 10000 de tranzistoare
was anticipated DE A have cca 10,000 transistors

For the possibility in c), the search algorithm is given in (25):

[4] As far as the class of the selecting predicate is concerned, let us notice that, in accordance with Abusch (2004) and Wurmbrand (2014) for English (among others), as well as with Cornilescu (2017), Cornilescu and Cotfas (2017) – for Romanian – the raising vs control distinction (i.e., what we have here termed, for infinitives in particular, non-agreeing and agreeing infinitives, should no longer reside on the class of selecting verb (e.g., believe-verbs for the latter, future-oriented verbs for the former) and hence on the temporal/modal specification of the embedded clause (untensed/realis for raising, tensed/irrealis for control), since raising/agreeing variants are clearly allowed by futurate verbs also (see, for example (10b)). This is confirmed by the corpus data – but we do not discuss it here.

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

(25) [drukola/m=ctag:va3p & drukola/m=**number:plural** & drukola/m=person:third & drukola/m=type:auxiliary & drukola/m=verbform:indicative][drukola/m=ctag:vppf & drukola/m=**number:plural** & drukola/m=verbform:participle] [drukola/m=ctag:qn & drukola/m=pos:**particle** &drukola/m=type:**infinitive**]

Again, the system provides examples which are similar on the surface structure, but have different underlying syntax, as it happens, for example, in English when it comes to teasing apart cases of Direct Object control (with 3-argument verbs) from Accusative + Infinitive/raising to object constructions (with bi-argumental predicates)[5]. This is illustrated for Romanian with examples from the corpus: in (26a), an instance of object control (with 'oblige'), the subject is base-generated in the main clause and is then passivized, resulting in participle agreement; conversely, (26b) is an instance of raising/agreement with a nominal base-generated in the infinitival clause, but which happens to surface in the main clause. In such cases, therefore, the examples have to be weighed attentively and one has to use her syntactic knowledge of argument structure, next to contextual hints regarding the class of matrix trigger or the type of embedded predicate. One other reliable test would be to consider whether the subject can remain embedded, in which case a raising analysis is at stake (understood as long distance agree)[6]. Constraining the query to only look for examples where the subject noun remains *in-situ* could be a way out, but since case is not morphologically marked in Romanian and the specifications/tags for nouns do not include case clues, this will not yield the desired results. Be that as it may, the query can provide a valuable pool of examples wherefrom the relevant ones can be sifted away.

(26) a. aceştia **sunt obligaţi a** pune la dispoziția medicului ….
these ones are obliged-masc.pl to lay at the disposal of the doctor….
b. Veteranii **sunt consideraţi a** fi nişte bătrâni asistaţi
veterans-the are considered-masc.pl to be some old men assisted

A very interesting and valuable find for our present purposes is that, as far as the compatibility of the raising/agreeing option with a *de-a*-Infinitive is concerned, the corpus data is quite consistent and points to a clear incompatibility. A search with a choice selection of individual predicates was undergone (*consider, expect, conceive, estimate, plan*), in the agreeing variant and with a *de-a* infinitive (in parallel with an *a*-Infinitive). The results are listed below, under (27):

(27) a. **sunt** considera**te a** ... – 1,598 matches /a'. **sunt** considera**te de a** … **- 0 matches**
b. **sunt** estima**te a** … – 28 matches / b'. **sunt** estima**te de a** … **- 0 matches**
c. este concepu**tă a** ... – 9 matches / c'. este concepu**tă de a** ... – **0 matches**

[5] (i) $John_i$ obliged $them_j$ [PRO_j to leave] > $They_j$ are obliged (by $John_i$) [PRO_j to ….]
(ii) They consider [the veterans to be too old] > The $veterans_i$ are considered [t_i to be too old]

[6] For example, this would show that while the pronominal subject cannot be moved in the infinitival clause in (26a), this is possible with "the veterans" in (26b):
(i) a. ?? sunt obligaţi a pune (aceştia) la dispoziția medicilor (aceştia)...
are obliged to lay (these ones) at disposal-the of doctors (these ones)
b. sunt consideraţi a fi asistaţi veteranii
are considered to be assisted veterans-the

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

d. este aşteptat**ă a** … – 4 matches / d'. este aşteptat**ă de a** ... – **0 matches**
e. **erau** preconiza**te a** ... – 2 matches / e'. **erau** preconiza**te de a** … – **0 matches**

We thus have a confirmation that agreeing infinitives favour smaller-sized complements. This is a welcome result, given that, syntactically, raising infinitives have always been taken as less opaque/more transparent domains, unlike control configurations (which, in Romanian, seem to allow *de*-infinitives, cf. (24) above).

3.2. Subjunctive Clauses

Taking into account that the subjunctive has come to replace the infinitive in many contexts of coccurrence in Modern Romanian and that, expectedly, the two are very much alike distributionally, this section will capitalize on many of the assumptions and observations made above for infintives.

In our quest for data involving subjunctive subject clauses (both *ca*- and *să*-subjunctives), we started from the same set of distributional contexts: copula + adjective and passivized transitive verbs. For the first type of distribution, we queried the corpus for both *să*- and *ca*-subjunctives, cf. (28a, b). In the former, we tried to fine-tune the query so as to make sure that *ca* is seen by the system as the subjunctive introductory element (since it is followed by *să*, with intervening material in between) (i.e., *ca* + noun + *să')*. This way, we managed to sift out irrelevant examples where *ca* was used as preposition. The total number of matches was significantly less than the hits following (28a), which is to be expected, given that *să* is the actual subjunctive marker and taking into account the (recent?) gradual obsolescence of *ca*, even in contexts with displaced material in the pre-verbal (*să*+verb cluster) domain.

(28) a. [drukola/m=ctag:v3 & drukola/m=number:**singular** & drukola/m=person:third &drukola/m=tense:present & drukola/m=verbform:indicative][drukola/m=ctag:r & drukola/m=pos:adverb & drukola/m=type:**general**][drukola/m=ctag:qs & drukola/m=**pos:particle** & drukola/m=type:**subjunctive**]
b. [drukola/m=ctag:v3 & drukola/m=number:**singular** & drukola/m=person:third & drukola/m=verbform:indicative][drukola/m=ctag:r & drukola/m=**degree:positive** & drukola/m=type:general][drukola/m=ctag:rc & drukola/m=type:portmanteau][drukola/m=number:singular & drukola/m=pos:**noun** & drukola/m=type:**common**][]{2}[drukola/m=ctag:qs & drukola/m=pos:**particle** & drukola/m=type:**subjunctive**]

The classification of adjectival classes discussed above is maintained, with much more examples overall and a more robust class 3 (the one featuring temporal adverbs/adjectives; there were 23 matches for *este prea devreme + să 'it is too early + să-subjunctive'*). Modal adjectives are again the most frequent, but the examples are no longer limited to the judicial register, as expected with subjunctives. Besides the overwhelming ubiquity of *să*-subjunctives (for reasons discussed above), we could not establish a clear-cut correlation between one adjectival class and (lack of) preference for *ca*-subjunctives.

Let us now finally turn to the second type of distribution, namely the one involving passivized transitive main verbs. As with infinitives, we have here four possibilities to

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

investigate: the non-agreeing variant with both a *să*- and a *ca*-subjunctive, and the agreeing variant with the two types of subjunctive complements. In (29a) we list the query for non-agreeing + *să* and in (29b) the one with *ca* (+ noun + *să*)

(29) a. [drukola/m=ctag:va3s & drukola/m=number:**singular** & drukola/m=**person:third**&drukola/m=type:auxiliary& drukola/m=verbform:indicative] drukola/m=ctag:vpsm & drukola/m=**gender:masculine** & drukola/m=**number:singular** & drukola/m=verbform:participle][drukola/m=ctag:qs & drukola/m=pos:**particle** & drukola/m=type:**subjunctive**]
b. [drukola/m=ctag:va3s & drukola/m=**number:singular** & drukola/m=**person:third** & drukola/m=type:auxiliary & drukola/m=verbform:indicative][drukola/m=ctag:vpsm & drukola/m=**gender:masculine** & drukola/m=verbform:participle] [drukola/m=ctag:rc & drukola/m=type:portmanteau][drukola/m=pos:**noun** & drukola/m=type:**common**][drukola/m=ctag:qs & drukola/m=pos:**particle** & drukola/m=type:**subjunctive**]

The staggering difference seen above with the be+ adjective distribution holds for the non-agreeing participle distribution as well: there were a lot more hits following (29a) vs a meagre result of for (29b) (just 2 pages worth of examples). Just as in the case of infinitives in this distribution, the system generates examples that can equally be agreeing or non-agreeing instances, with the same verb: (30a, a') are instances of agreement with a masculine singular shared subject ('the movie', 'the tournament') vs (30b), instances of free-subjunctives, i.e., independent domains both temporally and in point of subject reference[7].

(30) a. Inițial *filmul* era planificat să apară în 2009, ..
initially movie-the was planned SBJV appear in 2009
a'. *turneul* era planificat să se desfăşoare doar în Japonia
tournament-the was scheduled SBJV SE unfold only in Japan
b. …în Comisiile …, unde era planificat să se dezbată şi Legea Minelor
in committees …where was planned SBJV SE debate also law-the mines-of
b'. [După lansarea **ei** (USS Hawaii)] .. era planifica**t** să fie converti**tă** în ...
after launch hers (USS Hawaii) was planned SBJV be converted-fem into..

What about agreeing variants? According to CoRoLa, *să*-subjunctives are widely compatible with a matrix agreeing participle – not at all a surprise, given that Romanian has hyper-raising even from indicative clauses. A careful scrutiny of the examples yielded by the query in (31) below shows that raising is available not just with believe-verbs, but also with futurate matrix predicates – as expected (see fn. 4 above).

(31) [drukola/m=ctag:va3p & drukola/m=**number:plural** & drukola/m=person:third & drukola/m=type:auxiliary & drukola/m=verbform:indicative][drukola/m=ctag:vppf

[7] Notice that *se* in (30a') is inherent (unaccusative), whereas in (30b) it is passive. This is not a trivial matter.

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

& drukola/m=**number:plural** & drukola/m=verbform:participle]
[drukola/m=ctag:qs & drukola/m=pos:**particle** & drukola/m=type:**subjunctive**]

As for *ca*-subjunctives, most of the examples on the pattern [plural copula + plural participle + *ca*] yield irrelevant results, where *ca* is a preposition, not a complementizer, cf. *cincizeci de miliarde sunt aşteptate ca cifră de export* 'fifty billion are expected **as** export figures'; *schimbările de preț sunt estimate ca fiind* .. 'price changes are estimated **as** being'

Moreover – and once more just like with the agreeing options and a *de*-infinitive – the corpus data shows that neither are agreeing subjunctives compatible with the complementizer. This might reflect a more general tendency towards a dwindling use of *ca*, but it is telling for an analysis of these contexts as instances where the embedded subject enters a syntactic relation with the matrix predicate. A comparative search of the corpus with individual passivized agreeing verbs followed by either *să* or *ca* reveals the same picture as with infinitives and *a*- vs *de*-infinitives. The results are summarized in (32) below:

(32) a. **sunt** estima**te să** – 13 matches / a'. **sunt** estima**te ca** .. – 4 matches, **0 relevant**
b. **sunt** aşteptați **să** ... – 150 matches / b'. **sunt** aşteptați **ca** ... – **0 matches**
c. **erau** concepu**te să** – 5 matches/ c'. **erau** concepu**te ca** – 10 matches, **0 relevant**
d. **sunt** concepu**te să** – 84 matches /d'. **sunt** concepu**te ca** – 53 matches, **0 relevant**
e. **erau** planifica**te să** ... – 5 matches / e'. **erau** planifica**te ca** .... – **0 matches**

4. CONCLUSIONS

The present paper represents an attempt to look at infinitive and subjunctive (subject) clauses as they are represented in the CoRoLa corpus. Our aim was to suggest ways in which effective search queries can be built and used so that valuable input can be obtained, on the basis of which classifications can be done or refined and analyses confirmed, all the while gathering an ever-increasing pool of data.

We investigated the distribution of the two types clauses by focusing on two different triggers: copula + adjective and passivized transitive verbs. In the latter distribution, Romanian clauses are interesting in that they allow two patterns with different underlying syntax: an agreeing variant where the embedded subject is shared between the two predicates and agrees in number and person with the (passivized) main verb (and which, syntactically, corresponds to raising) and a non-agreeing variant where the infinitive has its own subject (null or non-Nominative) and where there is no agreement on the matrix predicate, corresponding to (non-obligatory) control instances (arbitrary, implicit)).

For the first type of distribution, the corpus data has helped establish a typology (and frequency) of adjectives into modal, emotive-evaluative and temporal, each with its own particularities. For the second, the corpus turned out to be a useful tool not only in confirming typical classes of triggers (i.e., propositional attitude verbs (believe-verbs) and futurate verbs), but also in confirming the preference for smaller-sized complement with raising/agreeing variants and in revealing that the latter are available with futurate verbs as well. Also, for cases of ambiguity (i.e., when the query cannot differentiate a non-agreeing variant from an agreeing one with a masculine singular shared subject), the examples provide other contextual disambiguating clues, such as preference for the subject to surface

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)

in the matrix and for stative embedded predicates – for agreeing variants, vs the pervasive presence of embedded *se*-passives and eventive predicates – with non-agreeing instances.

REFERENCES

Abusch, D., 2004, "On the Temporal Composition of Infinitives", in: J. Gueron, J. Lecarme (eds), *The Syntax of Time*, MIT Press, 27–53.

Cornilescu, A, M.A. Cotfas, 2017, "On clause-size, tense and case in Romanian infinitive constructions", conference paper AICED 2019, Bucharest, June 2017.

Cornilescu, A., 2017, "Remarks on Romanian Clausal Passives", paper at the Cambridge workshop on Voice, May 22-24, 2017.

Cornilescu, A., R. Cosma, 2014, "On the Functional Structure of the Romanian de-supine", in: R. Cosma, S. Engelberg, S. Schlotthauer, S. Stănescu, G. Zifonun (eds), *Komplexe Argumentstrukturen. Kontrastive Untersuchungen zum Deutschen, Rumänischen und Englischen,* Berlin/München/Boston, de Gruyter [Konvergenz und Divergenz 3], 283–335.

CoRoLa = Corpus computațional de referință pentru limba română, corola.racai.ro

Cotfas, M.A., 2012, *On the Syntax of the Romanian Subjunctive: Control and Obviation*, doctoral dissertation, University of Bucharest.

DEX= *Dicționarul explicativ al limbii române*, Academia Română, Institutul de lingvistică „Iorgu Iordan", ediția a II-a, București, Univers Enciclopedic.

Wurmbrand, S., 2014, "Tense and Aspect in English Infinitives", *Linguistic Inquiry*, 45, 3, 403–447.

BDD-A30405 © 2019 Editura Academiei
Provided by Diacronia.ro for IP 216.73.216.169 (2026-11-07 03:06:21 UTC)