

A CORPUS-BASED ANALYSIS OF THE EVENT-PERCEPTION CONSTRUCTIONS IN ROMANIAN

Mihaela GHEORGHE¹ 

Article history: Received 31 August 2025; Revised 2 November 2025; Accepted 7 November 2025;
Available online 12 December 2025; Available print 30 December 2025

©2025 Studia UBB Philologia. Published by Babeş-Bolyai University.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ABSTRACT. *A Corpus-Based Analysis of the Event-Perception Constructions in Romanian.* Consistent with cross-linguistic patterns, perception verbs in Romanian typically take two arguments: an Experiencer/Agent, and a Theme. The Theme argument can be expressed in three main syntactic forms: (i) as a determiner phrase (DP), (ii) as a finite or non-finite clause (CP), or (iii) as a restructured configuration. These options show different syntactic behaviours and constraints related to subject licensing, aspect, and the perception of events versus states. The paper presents a corpus-based analysis of Event-Perception Constructions (EPCs) in Romanian, focusing on the interaction between the type of perception verb (*a vedea* 'see', *a auzi* 'hear'), the semantic-aspectual class of the embedded predicate, and the choice of clausal complement (*că*-, *cum*-, or gerundive structures). Using data from the RoTenTen2021 corpus, the study examines EPCs with verbs that denote non-volitional or semi-volitional physiological and emotional events. The analysis shows that the distribution of complement types is influenced by both sensory salience (visual, auditory, or dual-modality events) and aspectual properties (atelic activities, semelfactives, punctual achievements). Gerunds are favored for processual construals, *cum*-clauses for manner or dynamic unfolding, and *că*-clauses for factual perception. Verbs with balanced perceptual profiles (e.g., *a plânge* 'cry', *a râde* 'laugh') occur freely with all complement types across modalities, while those with a strongly marked sensory channel (e.g., *a sforăi* 'snore', *a roşi* 'blush') display more restricted patterns.

Keywords: *Romanian language, perception verbs, event-perception construction, pseudo-relative clause, corpus linguistics*

¹ **Mihaela GHEORGHE** is professor of linguistics at Transilvania University in Braşov, at the Department of Theoretical and Applied Linguistics, and a senior researcher at the Institute of Linguistics „Iorgu Iordan – Alexandru Rosetti” of the Romanian Academy. Romanian syntax, mainly the syntax of relative clauses, represents the area of her research interest. Email: m.gheorghe@unitbv.ro

REZUMAT. *O analiză de corpus a construcțiilor de percepție evenimentială în limba română.* În concordanță cu modelele din alte limbi, și în limba română verbele de percepție au de obicei două argumente: Experimentator/Agent și Temă. Argumentul tematic poate fi exprimat în trei forme sintactice primare: (i) ca grup al determinantului (DP), (ii) ca propoziție subordonată (CP) sau (iii) ca structură sintactică reorganizată. Aceste opțiuni au comportamente sintactice și constrângeri diferite legate de autorizarea subiectului, de aspect și de tipul de percepție a obiectelor, a evenimentelor sau a stărilor. Articolul de față prezintă o analiză bazată pe corpus a construcțiilor de percepție a evenimentelor (EPC) în limba română, concentrându-se pe interacțiunea dintre tipul verbului de percepție (*a vedea*, *a auzi*), clasa semantico-aspectuală a predicatului din subordonată și alegerea complementului propozițional (structuri introduse prin complementizatorii *că-*, *cum-* sau construcții gerunziale). Folosind date din corpusul RoTenTen2021, studiul examinează EPC-urile cu verbe care denotă evenimente fiziologice și emoționale nonvolitive sau semivolitive. Analiza arată că distribuția tipurilor de compliniri este influențată atât de proeminența senzorială (evenimente vizuale, auditive sau cu modalitate duală), cât și de proprietățile aspectuale (activități atelice, semelfactive, realizări punctuale). Gerunziile sunt favorizate pentru construcțiile procesuale, subordonatele introduse prin *cum* pentru desfășurarea dinamică, iar subordonatele cu *că* pentru percepția factuală. Verbe cu profiluri perceptive echilibrate (de exemplu, *a plânge*, *a râde*) apar fără restricții cu toate tipurile de complementare, pe când cele cu un canal senzorial puternic marcat (de exemplu, *a sforăi*, *a roși*) prezintă limitări contextuale.

Cuvinte-cheie: *limba română, verb de percepție, construcție de percepție a evenimentelor, pseudo-relativă, lingvistică de corpus*

1. Preliminaries

1.1. *Complementation of perception verbs.* In Romanian, as in many other languages, perception predicates select two arguments. The argument receiving the theta-role of the Experiencer or of the Agent² is a [+animate] determiner phrase (DP) that occupies the subject position. The second argument, which takes the syntactic role of the direct object (DO), can be realized either as a DP (1a) or as a clausal complement (CP) (1b-e). When realized as a clause, the complement is introduced by the prototypical complementizers *că* ('that'), *să* ('to' = *SĂ_{SUBJ}*), and *dacă* ('if'), or, in specific constructions, by the wh-word *cum* ('how') (GR 2013, 88; 226–227; Nicula 2013, 316–323).

² Perception verbs can be either non-agentive: *vedea* 'see', *auzi* 'listen', *simți* 'feel', or agentive: *asculta* 'listen', *privi* 'look', *zări* 'see', *a se uita* 'gaze', etc. For a complete list, see Nicula 2012, 36.

1. a. *Ana a văzut* [_{DP} *filmul*] *ieri*.
'Ana saw the movie yesterday.'
- b. *Ana a văzut* [_{CP} *că Ion dansa cu Maria*].
'Ana saw that Ion was dancing with Maria.'
- c. *Ana n-a văzut* [_{CP} *să fi dansat Ion cu Maria*]
'Ana didn't see Ion dancing with Maria.'
- d. *Ana n-a văzut* [_{CP} *dacă Ion a dansat cu Maria*].
'Ana didn't see if Ion danced with Maria.'
- e. *Ana l-a văzut pe Ion* [_{CP} *cum dansa cu Maria*].
'Ana saw Ion dancing with Maria.'

In terms of non-finite clause structures, Romanian is distinct from other Romance languages because it only employs a gerund³ form after perception verbs (2a). Spanish (Aldama García 2016; Casalicchio 2016), Catalan (Rafel 2000) and Portuguese (Mateus et al. 2003) also accept the gerund structure (2b-d), but they also use an *infinitive* non-finite clause (3a-c), while other Romance languages⁴ like French and Italian rely solely on the infinitive (3d,e)⁵.

- | | | | | | | | |
|----|----|-------|--------|-------|-------|-------------------|------------------|
| 2. | a. | Rom. | O | văd | pe | Maria | <i>cântând</i> . |
| | | | cl.ACC | I.see | DOM | Maria | sing.GER |
| | b. | Sp. | Veo | a | María | <i>cantando</i> . | |
| | | | I.see | DOM | Maria | sing.GER | |
| | c. | Cat. | Veig | la | Maria | <i>cantant</i> . | |
| | | | I.see | DOM | Maria | sing.GER | |
| | d. | Port. | Eu | vejo | Maria | <i>cantando</i> . | |
| | | | I | see | Maria | sing.GER | |

³ The use of the infinitive within gerund contexts is also attested in Old and Pre-modern Romanian, but it is merely accidental:

(i) *Văzut-am flăcăii scuturându-și pletele (...) și fruntea lor a se încreți fără vreme (...), florile de pe capul copiilor a se vesteji și poporul căutând în beție uitarea necazurilor* (Russo, S, 228 *apud* Diaconescu 1977, 135)

'I have seen the lads shaking their hair (...) and their foreheads wrinkling without time (...), the flowers on the children's heads fading, and the people seeking in drunkenness to forget their troubles'.

(ii) *S-au auzit în tabăra vrăjmașilor multe sunete de trâmbițe a face gălceavă* (VF)

'In the enemy's camp, many sounds of trumpets were heard making noise' (*apud* Nicula 2013, 323)

⁴ The choice between infinitives and gerunds is more intricate in Romance languages. For an in-depth analysis from a generative grammar point of view, refer to Casalicchio 2016; 2024. He examines the Rhaeto-Romance dialects of northern Italy, particularly Gardenese, demonstrating that the interaction between the parameters [+/- progressive] _{v_{perc}} and [+/- Agr] gerunds accounts for the variations related to the non-finite complements of perceptive verbs.

⁵ Similarly, in Germanic languages, perception verbs select an infinitive structure (English, German, Dutch), either exclusively (German, Dutch) or alongside a structure with the present gerund or participle (English) (Niculescu 2013, 68).

3. a. Sp. Veo a María *cantar*.
 I.see DOM Maria sing.INF
 b. Cat. Veig *cantar* la Maria.
 I.see sing.INF DOM Maria
 c. Port. Eu vejo Maria (a)⁶ *cantar*.
 I see Maria (A) sing.INF
 d. Fr. Je vois Marie *chanter*.
 I see Marie sing.INF
 e. It. Vedo Maria *cantare*.
 I.see Maria sing.INF
 'I see Maria sing'

Regarding perception verbs, linguistic literature distinguishes between three types of perception: (i) non-epistemic/epistemically neutral/direct perception (4a,b), (ii) epistemic/epistemically positive/indirect perception (4c,d), and (iii) cognitive perception (4e). The specific type of perception is usually linked to particular patterns of complementation⁷, as many studies confirm (Alboiu & Hill 2013; Aldama García 2016; Avram 2003; Barwise 1981; Dik & Hengeveld 1991; Felser 1999; Guasti 1993; Nicula 2012; 2013; Usonienė 2001, among others).

4. a. O văd pe Maria *plângând*.
 cl.ACC I.see DOM Maria cry.GER
 b. O văd pe Maria *cum plânge*.
 cl.ACC I.see DOM Maria how cry.3SG
 c. O văd pe Maria *că plânge*.
 cl.ACC I.see DOM Maria that cry.3SG
 'I see Maria crying'
 d. *Văd că Maria a plâns*.
 'I see that Maria has been crying'
 e. Nu o văd pe Maria *angajându-se curând*.
 'I don't see Maria being hired anytime soon'

⁶ Alongside the bare infinitive employed as complement of a perception verb, European Portuguese also has a *Prepositional Infinitival Construction*, which consists of a DP, a P, and an infinitival verb that can be inflected for person and number (Rafel 2000, 78).

⁷ Using the semantic concept of evidentiality, Alboiu & Hill 2013, 275 argue that: "(i) a direct evidential reading arises when the perception verb selects a reduced CP complement; whereas (ii) an indirect evidential reading (inferential, evaluative) arises when the verb selects a full-fledged CP, the subject of which moves to the matrix clause (Raising-to-Object – RtoO)." The distinction largely aligns with the types of perception discussed here, in the sense that a broad semantic distinction is made between *first-hand evidentiality* (which indicates perceptual access to information) and *non-first-hand evidentiality*, encompassing both external sources (input from another person) and internal sources (inferential processes), see DTG 2019, s.v. *evidențialitate*.

In terms of their grammatical features, most perception verbs in Romanian are transitive, such as *a vedea* 'see', *a privi* 'look', *a zări* 'spot', *a auzi* 'hear', *a asculta* 'listen', and *a simți* 'feel'. The exceptions are *a se uita* 'look' (5a) and *a se holba* 'stare' (5b), which select a prepositional object, and exhibit restrictions in their complementation (5c-d).

5. a. Ion se uită la Maria *cum plânge*.
'Ion is watching Maria crying'
- b. Ion se holbează la Maria *cum plânge*.
'Ion stares at Maria as she cries'
- c. Ion_i se uită la Maria_j
Ion cl.REFL.3SG watch.3SG at Maria
pro_{i/*j} *plângând*.
cry.GER
'Ion is watching Maria as he cries'
- d. ?Ion_i se holbează la Maria_j
Ion cl.REFL.3SG stare.3SG at Maria
pro_{*i/*j} *plângând*.
cry.GER
'Ion stares at Maria as she cries'

1.2. Pseudo-relative clauses in Western Romance (PRs). There is extensive linguistic literature on a clausal argument of perception verbs that, in some Romance languages and dialects, exhibits the surface features of a relative clause: Italian (6a), French (6b), Spanish (6c), Catalan (6d), Galician (6e), and Calabrian (6f).

6. a. Ho visto Gianni *che correva*.
b. J'ai vu Jean *qui courait*.
c. He visto a Juan *que corria*.
d. He vist en Joan *que corria*.
e. Eu vin a Xoán *que corría*.
f. Vitti a Gianni *chi scappava*
have.I seen DOM John that ran.3SG.IMPERF
'I saw John running' (*apud* Rafel 2000: 68⁸)

⁸ The list of examples in Rafel 2000, 68 also includes Romanian (example (4g)): *Am văzut pe Ion că fugea*, which was glossed in the same way. I will comment on the inaccuracy of that reference in section 1.3. below.

Research in the field (Burzio 1986; Guasti 1993; Rafel 2000; Casalicchio 2016; Aldama García 2016; Graffi 2017; Cinque 2020, 244-275) has demonstrated that, despite their apparent similarity to relative constructions – mainly because of the overlapping of the complementizer *che/qui/que* with the homonymous *wh*-words – they are structurally different. These complements of perception verbs serve the discourse function of introducing a new entity within a given discourse and providing new information about that entity. The challenging aspect of understanding these constructions is that, although they are usually used to introduce an entity argument into discourse, they actually introduce an event argument. As a result, events are “disguised” as entities.

The syntactic features⁹ of PRs can be summarized as follows: (i) they have a CP structure; (ii) the initial DP is always coindexed with the subject in the subordinate clause; (iii) the initial DP can be a proper noun (a feature excluded for the antecedent of a restrictive relative clause); (iv) the verb in the matrix clause is a perception verb such as *to see*, *to hear*, *to feel*, or similar verbs that indicate a type of discovery: *to find*, *to meet*, *to discover*, *to surprise*, *to observe*, *to notice*, etc., through which the interlocutor is prompted to become aware of the presence of an entity or event; also, an interjection with a presentative meaning may appear as the predicate of the matrix clause; (v) the embedded clause is an event-type predicate; (vi) since it concerns immediate perception of an event, the temporal, modal, and aspectual features of the sentence complement depend on the corresponding features in the matrix.

For languages featuring *che/qui/que*, multiple interpretations have been proposed (see a review for (a) and (b) in Barron 2000, 12-14; Rafel 2000, 71-74; Alboiu & Hill 2013, 279): (a) the PR functions as a CP-type element with a DP raised to Spec, CP, creating a complex small clause; (b) the PR is a genuine relative clause that modifies the noun serving as the direct object of the perception verb; (c) the embedded clause is a relative clause without movement, involving a *wh*-element functioning as a secondary predicate or presentative relative clause. (Lambrecht 2000).

⁹ The theoretical framework used in the works from which I drew this synthesis is the current generative model, which divides sentence structure into three domains: CP (the functional domain of the complementizer, including complementizers, *wh*-elements, and other dislocated elements hosted by its left periphery), IP (the functional domain involving verbal inflection and pronominal clitics, responsible for, among other things, assigning the Nominative case), and vP (the lexical domain, responsible for generating the verb and arguments). Internal arguments are generated and marked casually within the lexical domain. Conversely, the external argument (the subject), although generated in the lexical domain, depends on the tense projection (TP), part of the IP domain, to be marked with the Nominative case (Chomsky 1981; see Cornilescu 2000; Stan 2005; Dragomirescu & Nicolae 2016 for Romanian). The prototypical PR can be schematically represented as in (i), adopted from Rafel 2000, 71:

(i) [_{CP} Jeani [_C qu(e) [_{IP} i [_{I'} [_{VP} pro i courait]]]]]

1.3. Are there PRs in Romanian? The presence of PRs in Romanian has received limited attention because there is no homonymy between a complementizer and a *wh*-word, thus lacking a syntactic parallel to Western Romance. However, since PRs in Romance languages correspond to gerund or infinitive structures and constructions with gerund complements of perception verbs have Romanian equivalents introduced by *cum/că*, this has led to discussions on the topic (Gheorghe 2004; 2015; SOR 2016, 468). I disagreed with Rafel (2000, 68), who grouped Romanian *că*-based constructions with PRs in Romance languages. I found the comparison forced because, in Romanian, the relative pronoun stems from Lat. *QUALEM*, not from Lat. *QUOD*, so the concept of homonymy in Romance languages doesn't apply. Nonetheless, I believe that the complement clauses of perception verbs that follow the PR model in Romance are more likely embedded clauses introduced by *cum*, especially since, in Old Romanian, the term was multifunctional, serving both as the complementizer *că* 'that' and the *wh*-word *cum* 'how' (SOR 2016, 468; GR 2013, 88). Following Alboiu & Hill 2013, I concluded that Romanian lacks PRs¹⁰ and that the only valid interpretation for the Romanian constructions headed by *cum* is that they are small clauses. The main difference between PRs in Romance languages and their Romanian equivalents is how the entity expressed by the DP is identified: in Romance PRs, the entity is cognitively inactive in discourse, introduced into the perception field of a perceiving subject through its action; in the Romanian equivalents, the entity is necessarily brought into the discourse universe and then connected with the event through a syntactic reorganization mechanism (i.e., small clauses).

2. A corpus-based analysis of EPCs in Romanian.

2.1. Methodology. In analyzing the data on EPCs in Romanian, I examined the collocations of the transitive perception verbs *a vedea* 'see' and *a auzi* 'hear' with complements that denote events: full CPs (headed by the complementizer *că*), defective CPs (headed by the complementizer *cum*), and non-finite clauses (gerunds).

I used the Romanian Web 2021 (RoTenTen2021¹¹) corpus, which I interrogated with Corpus Query Language (CQL) formulas that enable searching for complex grammatical or lexical patterns. The CQL formulas I employed are the following:

¹⁰ See Casalicchio 2016 for the same interpretation.

¹¹ The corpus is available through SketchEngine: <https://app.sketchengine.eu>.

- (i) [tag="P.*"] + [lemma="vedea/auzi/privi"] + [word="pe"] + [] + [word="că"];
- (ii) [tag="P.*"] + [lemma="vedea/auzi/privi"] + [word="pe"] + [] + [word="cum"];
- (iii) [tag="P.*"] + [lemma="holba/uita"] + [word="la"] + [] + [word="cum"];
- (iv) [tag="P.*"] + [lemma="vedea/auzi/privi"] + [word="pe"] + [] + [tag="V.*"].

The results I obtained through CQL (iii) were filtered manually to retain only instances with gerunds.

The corpus analysis did not focus on quantitative comparisons but aimed to identify as many specific usages of these parallel constructions as possible, highlighting restrictions and incompatibilities that might be overlooked when only working with constructed examples.

For some patterns, the query generated tens of thousands of examples, and I randomly sampled 200 examples for manual labeling based on the following criteria: (i) the category of the verb in the main clause (auditory perception vs. visual perception); (ii) the semantic-syntactic features of the verbs in the embedded clause according to aspect (telicity vs. atelicity) and type of event (process vs. state vs. result); (iii) the syntactic pattern (constructions with *că/cum*/gerund).

Due to time and space constraints, I have utilized only a portion of the EPC collection for the data analysis in section 2.2 below, limiting the discussion to a category of verbs that occur in the embedded clause of the perception verb or as a gerundial complement: verbs denoting non-volitional or semi-volitional physiological or emotional events. The corpus was filtered for verbs in the embedded clause, and the inventory that resulted includes: (i) psychological events: *căscă* 'yawn', *leșina* 'faint', *roși* 'blush', *sforăi* 'snore', *strănuta* 'sneeze', *tuși* 'cough', *transpira* 'sweat', *vomita* 'vomit'; (ii) emotional events: *ofta* 'sigh', *plânge* 'cry', *râde* 'laugh', *suspina* 'sob'.

2.2. The data analysis. In Romanian EPCs, the choice of the complement (*că/-cum/-gerund*) depends on the interaction between the type of perception verb and the aspectual class of the embedded predicate. The construction with *că* 'that' is the most flexible, accepting activities, achievements, and even states. *Cum* 'as' is more restrictive and is linked to process-oriented activities, especially in visual perception. The gerund works with atelic activities and is not used with stative verbs or punctual achievements.

In this section, I will present data from the corpus, and I will comment on the limits of ECP distribution in Romanian. The analysis of actual examples will confirm that the theoretical generalizations hold, but will also reveal that the semantics of the embedded verb can affect the distribution of the patterns.

2.2.1. Constructions with verbs denoting physiological events. This group includes verbs that describe bodily processes and reflexes with a clear physiological basis, such as *a căsca* ‘yawn,’ *a leșina* ‘faint,’ *a roși* ‘blush,’ *a sforăi* ‘snore,’ *a strănuta* ‘sneeze,’ *a tuși* ‘cough,’ *a transpira* ‘sweat,’ and *a vomita* ‘vomit.’ From an aspectual perspective, most of these verbs are atelic and durative, indicating ongoing activities or states (*a căsca*, *a sforăi*, *a transpira*), while *a tuși* ‘cough’ and *a strănuta* ‘sneeze’ are semelfactive in their basic reading but can also denote continuous atelic events when interpreted iteratively. The perceptual profile varies considerably across the set: some predicates are accessible to both *a vedea* ‘see’ and *a auzi* ‘hear’ (e.g., *a strănuta*, *a tuși*), while others are primarily auditory (*a sforăi*) or visual (*a roși*, *a transpira*), and a few have a strongly limited sensory channel (*a leșina*, *a vomita* - mostly visual attestations in the corpus). Regarding complementation, these verbs show the full range of options available in EPCs – gerund, *cum*-clauses, and *că*-clauses – though distribution patterns depend on both the sensory salience of the event and its aspectual profile. Gerunds usually encode a process-oriented view, *cum*-clauses emphasize manner or the unfolding of the event, and *că*-clauses present it as a factual perception. The following analyses demonstrate how these factors interact for each verb, based on corpus examples and constructed instances where relevant.

For *a căsca* ‘yawn’, only the gerund appears in the corpus, both with *a vedea* ‘see’ (7a) and *a auzi* ‘hear’ (7b), but constructions with complementizers *cum* (7c) and *că* (7d) are also perfectly possible. The variant with *că* could be omitted by speakers for reasons of euphony, but semantically and syntactically, there are no restrictions.

7.
 - a. Este contagios, respectiv caști atunci când *vezi* pe altcineva *căscând*. (libertateapentrufemei.ro)
‘It’s contagious, meaning you yawn when you see someone else yawning.’
 - b. Căscatul contagios se poate declanșa atunci când *auzim* pe cineva *căscând*. (carti-bune.ro)
‘Contagious yawning can be triggered when we hear someone yawning.’
 - c. L-au văzut *cum căsca*...
‘They saw him yawning...’
 - d. L-au văzut *că căsca*...
‘They saw him yawning...’

In case of the verbs *a leșina* ‘faint’ and *a roși* ‘blush’, the corpus rendered only examples with *a vedea* as a perception verb (8a,b) and (9a,b). The absence of contexts with *a auzi* ‘hear’ is understandable because this perception verb requires a detectable auditory input, and events like fainting or blushing do not have a clear “auditory profile”.

In terms of complementation, there is a contrast between the two verbs: *a leșina* appears only with gerunds (8a) and *cum*-clauses (8b), while *a roși* appears with gerunds (9a) and *că*-clauses (9b).

For *a leșina*, the absence of contexts with *că*-clauses is not surprising. After verbs of perception, they mainly occur when the complement introduces a finite clause representing a situation regarded as a confirmed fact, not necessarily perceived as happening in real-time. *A leșina* ‘faint’ is a punctual event, an achievement-type event, which is more naturally expressed through direct, event-based descriptions, hence the preference for *cum* (which indicates how the event unfolds from the observer’s perspective) and for the gerund (which emphasizes the ongoing process), and the oddness of an example like (8c). On the other hand, for the verb *a roși* ‘blush’, there are attested constructions with *că*, profiling a change of state outcome observable in its result (skin is red), but not with *cum*. Still, examples with *cum* can be acceptable (9c).

8.
 - a. Oare fiindcă îl văzuse pe Harry leșinând în tren și își spuse că nu era mare lucru de capul lui? (literaturapetocuri.ro)
‘Was it because he had seen Harry fainting on the train and thought he was not very bright?’
 - b. Când ies din incinta tunsătoriei le văd pe dudu cum leșină prin stația de autobuz. (misterx.ro)
‘When I leave the barber shop, I see the ladies fainting at the bus stop.’
 - c. ??Le-am văzut că leșinau...
‘I saw them fainting’
9.
 - a. Ce să spun era înduioșător s-o vezi pe regină roșind și zâmbind. (docplayer.org)
‘It was touching to see the queen blushing and smiling.’
 - b. Dacă îi văd pe unii că roșesc, la ei se reped mai aprig. (historia.ro)
‘If they see some people blushing, they rush at them more fiercely.’
 - c. Îi văd pe unii cum roșesc...
‘I see some people blushing...’

The verb *sforăi* ‘snore’ naturally pairs with the verb *auzi* ‘hear’ because it is atelic and durative, making it suitable for imperfective views, mainly as an auditory event. The corpus supports this with both *că* (10a) and gerundive (10c) complements. *Că* matches the factual auditory perception reading (*aud că sforăie*), while the gerund naturally fits with processual auditory perception. *A vedea* ‘see’ is less central here but still possible in contexts where visual cues of snoring - such as an open mouth or body movement - are noticeable. The corpus only shows the gerund (10b), which describes the event as an ongoing visual process. *Cum* is completely absent as a complementizer here, probably because snoring usually doesn't require manner-focused description, as it is mainly identified by the sound rather than a visually dynamic unfolding that needs “manner” encoding. However, examples with *cum* are not excluded (see the constructed example under 10d).

10. a. *O aude pe Ana sforăind* ca un motoraş, în fundul camerei. (natgeo.ro)
‘He hears Ana snoring like a motor at the back of the room.’
- b. *Aruncând o privire o văzu pe Maşa sforăind* în continuare. (beautyboutiquesalon.ro)
‘Glancing over, he saw Masha still snoring.’
- c. (...) *îl aud pe unul că sforăie*, pe altul că tuşeşte (...) (taifasuri.ro)
‘(...) I hear one snoring, another coughing (...)’
- d. *O aude cum sforăie*.
‘He hears her snoring.’

Sneezing is a brief, periodic physiological event that is perceivable through both modalities. It is punctual (completed), but can occur in sequences, allowing for both perfective and imperfective interpretations. The verb *vedea* ‘see’ naturally applies to the visible bodily movement associated with sneezing; *auzi* ‘hear’ applies to the distinctive sound. This dual perceptibility explains why all three complement types are attested with the two perception verbs. The corpus data shows a balanced visual–auditory profile, due to the event’s strong cues in both modalities. *A vedea + a strănuta* are the options for factual *că*, while *cum* is used for a dynamic manner; There is no gerund with *vedea* because the process-oriented visual interpretation of the event is secondary to the auditory one. *A auzi + a strănuta* have the same interpretation, i.e., *că* for factual events, and *cum* for ongoing or repeated events, but this time with sounds. Example (10d) also has an exclamative reading, due to the syntactic configuration of the matrix (a surrogate imperative). In this context, *cum* has a stronger modal value than in other similar EPCs.

11.
 - a. Când *auzea* pe cineva *strănutând*, se întorcea, zâmbind (confluente.org)
 - b. Tot de politete ține să spui "sănătate" după ce *auzi* pe cineva *că strănută*. (radiounirea.ro)
'It is also polite to say "bless you" after hearing someone sneeze.'
 - c. Când *vedeți* pe cineva *că strănută* și nu are mască, puteți suna la 112. (mediafax.ro)
'When you see someone sneezing and not wearing a mask, you can call 112.'
 - d. Câteva săptămâni parașutele le inspirăm noi, da' să-l *vezi* pe Ricci *cum strănută*! (digitalreviews.ro)
'We'll inspire the parachutes for a few weeks, but just wait until you see Ricci sneeze!'

A tuși 'cough' is a short, often repeated physiological event that can both be heard (primary) and seen (secondary, through bodily movements). Regarding aspect and (im)perfectivity, *a tuși* 'cough' is an atelic verb when repeated and punctual when isolated. *A vedea* 'see' naturally fits in contexts where the visual component is prominent (such as shoulder movement, hand to mouth, facial expression), while *a auzi* 'hear' emphasizes the distinctive sound. Concerning the complement type distribution, both *a vedea* 'see' and *a auzi* 'hear' select *că* (for factual construal of ongoing or habitual actions), as well as gerunds and *cum* for process-oriented auditory or visual depiction, emphasizing repeated or sustained coughing. There is no gerund attested for the auditory domain, but it is perfectly possible (12 e). The absence of *cum* in the visual domain suggests it is more natural with auditory than with visual perception (12f).

12.
 - a. Tot la televizor, îl *vezi* pe Arafat *tușind* în pumn. (kmkz.ro)
'Also on television, you see Arafat coughing into his fist.'
 - b. A două zi, călătorind împreună cu mașină, *l-am auzit* pe MB *cum tușea*. (dcmedical.ro)
'The next day, traveling together by car, I heard MB coughing.'
 - c. (...) îl *aud* pe unul *că sforăie*, pe altul *că tușește* (...) (taifasuri.ro)
'(...) I hear one snoring, another coughing (...)'
 - d. Să nu intrăm în magazine dacă *vedem* pe cineva *că tușește*, este febril sau transpiră. (libertatea.ro)

- ‘Let's not go into stores if we see someone coughing, feverish, or sweating.’
- e. Îl *auzeam tușind* dincolo de perete.
 ‘I could hear him coughing on the other side of the wall.’
- f. ?Îl văd cum tușește.
 ‘I see him coughing.’

A transpira ‘sweat’ is a non-auditory, gradual physiological process, making *a vedea* ‘see’ the only natural perception verb here. Auditory perception is irrelevant since there is no inherent sound associated with sweating. The complement type distribution shows that with *a vedea* ‘see’, all three complements are possible, even though the corpus only presented an example with the gerund (13a). *Că-* complements can be chosen for factual visual construal, related to an observed result (wet clothes, glistening skin) (13b), while gerund and *cum*-complements are used for process-oriented depiction, framing the observation in real time. *A transpira* ‘sweat’ is durative and atelic, so all three complements match the imperfective nature of the event.

13. a. Când *vedem* pe cineva *transpirând* putem fi siguri că îi este cald sau este tulburat (trotusaeauto.ro)
 ‘When we see someone sweating, we can be sure that they are hot or upset.’
- b. Îl *văd că transpiră* instantaneu.
 ‘I see him sweating instantly.’
- c. Îl *vedeam cum transpiră* din ce în ce mai mult.
 ‘I could see him sweating more and more.’

A vomita ‘vomit’ has both visual and potential auditory components, but the corpus shows only visual attestations (14a). *A vomita* ‘vomit’ is telic, often punctual, but can have iterative phases, so both *că* and *cum* complements are possible in Romanian (see 14b,c). Gerund supports a processual construal even for telic events by stretching the moment of observation.

- 14 a. M-*ați văzut* pe mine *vomitând* când vorbesc românește? (ebihoreanul.ro)
 ‘Have you seen me vomiting when I speak Romanian?’
- b. L-a *văzut că vomită* și s-a speriat.
 ‘He saw him vomiting and got scared.’
- c. Când l-a *văzut cum vomită*, s-a speriat.
 ‘When he saw him vomiting, he got scared.’

2.2.2. Constructions with verbs denoting emotional events. This group includes verbs whose lexical semantics denote emotional manifestations with a strong physiological component, such as *a plânge* ‘cry’, *a râde* ‘laugh’, *a suspina* ‘sob’, and *a ofta* ‘sigh’. These predicates are perceptually salient in both the visual and auditory domains to varying degrees, depending on the nature of the event: crying and laughing typically have a balanced dual profile, while sighing (*a suspina*, *a ofta*) tends to have a primary auditory component, with visual cues playing a secondary role. In terms of aspect, these verbs are atelic and readily compatible with imperfective construals, which allows them to occur with all three types of complement – gerund, *cum*-, and *că*-clauses – without strong grammatical restrictions. The corpus data show that the distribution across complement types often correlates with the sensory modality that is most salient for the event: gerunds are frequent for processual depictions, *cum* highlights manner or dynamic unfolding, and *că* encodes a factual perception. The analyses below illustrate how these tendencies play out for each verb, with examples covering both *a vedea* ‘see’ and *a auzi* ‘hear’.

A plânge ‘cry’ is a dual-modality physiological act, with clear auditory cues (sobbing, weeping) and strong visual cues (tears, facial expressions, posture). The corpus confirms this dual profile with examples of both *a vedea* ‘see’ (15a,d,f) and *a auzi* ‘hear’ (15b,c,e). All three complement types are found in both modalities: the gerund (*plângând*) describes the crying as an ongoing process; *cum*-clauses depict the manner or dynamic unfolding of the act; and *că*-clauses are used for factual descriptions, showing the crying as a confirmed state or event. Since *a plânge* ‘cry’ is atelic and compatible with imperfective construal, there are no aspectual restrictions on complement choice, and the full range of patterns is available to speakers.

15. a. O singură dată l-am văzut *plângând* (istorielocala.ro)
‘I saw him crying only once.’
- b. Ți se rupe sufletul când îl *auzi plângând*. (desprecopii.ro)
‘It breaks your heart when you hear him crying.’
- c. Mută-te în altă cameră pentru a nu-l mai *auzi* pe bebeluș
că plânge (qbebe.ro)
‘Move to another room so you can't hear the baby crying.’
- d. Nu-i nici o bucurie să-l *vezi* pe bebe *că plânge*.
(femeia.ro)
‘It's no fun to see your baby crying.’
- e. Îl *auzeam* pe tata *cum plânge*, pe mama pe fundal care
vorbea cu altcineva la telefon. (wordpress.ro)
‘I could hear my father crying, my mother in the background talking to someone else on the phone.’

- f. E teribil de dureros *să vezi* pe altul *cum plânge*, când tu nu mai ai nicio lacrimă. (wordpress.ro)
'It is terribly painful to see someone else crying when you have no tears left.'

A râde 'laugh' is also a dual-modality event, with prominent auditory cues (the sound of laughter) and equally salient visual cues (facial expressions, body movement). The corpus includes attestations for both *a vedea* 'see' (16b,d) and *a auzi* 'hear' (16a,c,e), covering all three complement types in each modality, except (16f), which is constructed, but acceptable. The gerund (*râzând*) provides a processual reading; *cum*-clauses highlight manner (e.g., *cum râde încetișor* 'how he laughs softly'); and *că*-clauses represent laughter as a factual perception. Since laughter is atelic and durative in repeated bouts, all complements are natural and semantically flexible.

16. a. Îl *aud* pe Viorel *râzând* în hohote. (ning.com)
'I hear Viorel laughing loudly.'
- b. Scriu comedie fiindcă îmi place *să-i văd* pe oameni *râzând*. (teatral.ro)
'I write comedy because I like to see people laughing.'
- c. Îmi crește inima când *aud* pe cineva *că râde* din cauza și în preajma mea. (nemira.ro)
'My heart swells when I hear someone laughing because of and around me.'
- d. Dacă te *vede* pe tine *că râzi*, va începe să râdă și el. (suntmamica.ro)
'If he sees you laughing, he will start laughing too.'
- e. Îl *auzi* pe străin *cum râde* încetișor. (editura.ro)
'You hear the stranger laughing softly.'
- f. Îl *vezi* *cum râde* încetișor.
'You see him laughing softly.'

A suspina 'sigh/soft sob' has mainly an auditory profile but is also visually noticeable through bodily cues such as chest movement and facial tension. The corpus includes both *a vedea* 'see' (17b) and *a auzi* 'hear' (17a,c) example, covering all complement types except the factive *că* and *cum* in the visual domain. The lack of *cum* in the context of *vedea* probably reflects the low visual salience of sighing unless sound is involved, though the construction is still possible. The *că*-complement characterizes the act as a factual perception, and probably contexts with *suspina* 'sob' are less frequent.

17.
 - a. Mi s-a părut că *aud* pe cineva *suspinând* sus, lângă marginea mormântului. (wordpress.ro)
'I thought I heard someone sobbing up by the edge of the grave.'
 - b. Vei sta tu nepăsător *să-i vezi* pe păcătoși *suspinând* în durerile morții. (rcrwebsite.com)
'Will you stand by indifferently and watch sinners groaning in the pains of death?'
 - c. În același timp o *auzeam* pe Monica *cum suspina* la telefon. (doxologia.ro)
'At the same time, I could hear Monica sobbing on the phone.'
 - d. O *aud* că *suspină*.
'I hear her sobbing.'
 - e. O *văd* că *suspină*.
'I hear her sobbing.'
 - f. O *văd* *cum suspină*.
'I hear her sobbing.'

A ofta 'sigh (long exhale)' shows a balanced auditory - visual profile, as the action produces both a perceivable sound and visible bodily movement. The corpus attests all three complement types, but not for both *a vedea* (18a) and *a auzi* (18b,c). However, the missing configurations are still possible; see the constructed examples (18d-f). Due to its atelic and imperfective nature, *a ofta* allows all complements without semantic restriction.

18.
 - a. Nu suport *să-i văd* pe bărbați *oftând*. (asa-si-asa.ro)
'I can't stand seeing men sighing.'
 - b. Îi *auzi* pe unii *că oftează*. (jurnalul.ro)
'You hear some people sighing.'
 - c. Zâmbetul îmi dispare atunci când îl *aud* pe Itachi *cum oftează*. (forumz.ro)
'My smile disappears when I hear Itachi sigh.'
 - d. Îi *aud* *oftând*.
'I hear them sighing.'
 - e. Îl *văd* *cum oftează*.
'I see him sighing.'
 - f. Îl *văd* *că oftează*.
'I see him sighing.'

3. Conclusions

This study provides a corpus-based analysis of Event-Perception Constructions (EPCs) in Romanian, examining how perception verb types (*a vedea* ‘see’ vs. *a auzi* ‘hear’), the semantic and aspectual features of the embedded predicate, and the choice of clausal complement (*că*-, *cum*-, or gerundive structures) interact. By analyzing a set of verbs representing non-volitional or semi-volitional physiological and emotional events, several conclusions emerge. First, the distribution of complements is strongly influenced by sensory salience. Events with a balanced dual profile (e.g., *a plânge*, *a râde*) easily combine with all complement types across modalities. Verbs with a primary auditory profile (*a sforăi*, *a ofta*) or a primary visual profile (*a roși*, *a transpira*, *a leșina*, *a vomita*) tend to have more restricted patterns, often lacking certain complements in the less salient modality. Second, aspectual class interacts with the choice of complement. Atelic, durative verbs favor gerunds for processual interpretations, whereas punctual achievements (*a leșina*) are more naturally expressed with *cum*- clauses or gerunds, which depict the unfolding of the event. Semelfactives (*a strănuta*, *a tuși*) can resemble activities when interpreted iteratively, broadening their compatibility with all complement types. Third, the different complementizers serve distinct functions: (i) *că*-clauses present the embedded event as a factual perception, with fewer restrictions on aspect but some sensitivity to modality; (ii) *cum*-clauses encode manner or dynamic unfolding, being more common in visual EPCs but also appearing in the auditory domain when the manner of sound production is salient; (iii) gerundives consistently indicate processual readings and are the most flexible across modalities. Finally, the analysis supports the idea that Romanian *cum*-clauses in EPCs should be viewed as evidential CPs with an eventive interpretation rather than as pseudo-relatives. Corpus evidence shows they pattern closely with gerundives in their event-focused sense, with an added emphasis on the manner and unfolding of the perceived event. This differs from Western Romance PRs, which have different syntactic triggers and discourse functions. Future research could expand the dataset to include quantitative frequency comparisons, explore EPCs with volitional action verbs for contrast, and study diachronic changes in the use of *că*-, *cum*-, and gerundive complements in Romanian.

WORKS CITED

- Alboiu, Gabriela, and Virginia Hill. 2013. "On Romanian Perception Verbs and Evidential Syntax." *Revue roumaine de linguistique* LVIII (3): 275–298.
- Aldama García, Nuria. 2016. *Pseudo-relatives Complement of Perception Predicates*. M.A. Diss., Universidad del País Vasco. <https://addi.ehu.es/handle/10810/21499>.
- Avram, Larisa. 2003. "An Aspectual Analysis of Gerunds." *Revue roumaine de linguistique* 48 (1–4): 203–219.
- Barron, J. 2000. "The Morphosyntactic Correlates of Finiteness." In *Proceedings of the LFG00 Conference*, edited by M. Butt and T. Holloway King, 1–19. CSLI Publications, Berkeley. <http://csli-publications.stanford.edu> (accessed 06.05.2011).
- Barwise, Jon. 1981. "Scenes and Other Situations." *The Journal of Philosophy* 78: 369–397.
- Burzio, Luigi. 1986. *Italian Syntax: A Government-Binding Approach*. Dordrecht: D. Reidel Publishing Company.
- Casalicchio, Jan, and Peter Herbeck. 2024. "Pseudorelative Clauses, Infinitives, and Gerunds with Spanish Perception Verbs: A Comparative View." *Syntax*, 1–34. <https://doi.org/10.1111/synt.12271>.
- Casalicchio, Jan. 2016. "The Use of Gerunds and Infinitives in Perceptive Constructions: The Effects of a Threefold Parametric Variation in Some Romance Varieties." In *Theoretical Approaches to Linguistic Variation*, edited by Ermenegildo Bidese, Federica Cognola, and Manuela Caterina Moroni, 53–88. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.234.03cas>.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Cinque, Guglielmo. 2020. *The Syntax of Relative Clauses: A Unified Approach*. Cambridge: Cambridge University Press.
- Cornilescu, Alexandra. 2000. "The Double Subject Construction in Romanian." In *Comparative Studies in Romanian Syntax*, edited by Virginia Motapanyane, 83–133. Amsterdam / Lausanne / New York / Oxford / Shannon / Singapore / Tokyo: Elsevier.
- Diaconescu, Ion. 1977. *Infinitivul în limba română*. București: Editura Științifică și Enciclopedică.
- Dik, Simon C., and Kees Hengeveld. 1991. "The Hierarchical Structure of the Clause and the Typology of Perception-Verb Complements." *Linguistics* 29: 231–259.
- Dragomirescu, Adina, and Alexandru Nicolae. 2016. "Originea formelor verbale «nonfinite» cu flexiune: infinitiv vs. supin." *Diacronia* 4 (1 August): art. A54, 1–16. <https://doi.org/10.17684/i4A54ro>.
- DTG. 2019. *Dicționar de termeni gramaticali*. Gabriela Pană Dindelegan (coord.), Adnana Boioc Apintei, Adina Dragomirescu, Alexandru Nicolae, and Rodica Zafiu. București: Editura Univers Enciclopedic Gold.
- Felser, Claudia. 1999. *Verbal Complement Clauses: A Minimalist Study of Direct Perception Constructions*. Amsterdam: John Benjamins.
- Gheorghe, Mihaela. 2004. *Propoziția relativă*. Pitești: Paralela 45.
- Gheorghe, Mihaela. 2015. "Notă sintactică despre construcțiile pseudo-relative în limba română." In Ionuț Pomian (coord.) and Nicolae Mocanu (ed.), *Înspre și dinspre Cluj. Contribuții lingvistice. Omagiu profesorului G.G. Neamțu la 70 de ani*, Cluj-Napoca: Scriptor Argonaut.

- GR. 2013. *The Grammar of Romanian*. Edited by Gabriela Pană-Dindelegan. Oxford: Oxford University Press.
- Graffi, Giorgio. 2017. "What Are 'Pseudo-Relatives'?" In Roberta D'Alessandro, Gabriele Iannàccaro, Diana Passino, and Anna M. Thornton (eds.), *Di tutti i colori. Studi linguistici per Maria Grossmann*, 115–132. Utrecht: Utrecht University Repository.
- Guasti, Maria Teresa. 1993. *Causative and Perception Verbs: A Comparative Study*. Torino: Rosenberg & Sellier.
- Lambrecht, Knud. 2000. "Prédication seconde et structure informationnelle: La relative de perception comme construction présentative." *Langue française* 127: 49–66.
- Mateus, Maria Helena, Ana Maria Brito, Inês Duarte, Isabel Hub Faria, et al. 2003. *Gramática da Língua Portuguesa*. Lisboa: Editorial Caminho.
- Nicula, Irina. 2012. *Modalități de exprimare a percepțiilor fizice. Verbele de percepție în limba română*. București: Editura Universității din București.
- Nicula, Irina. 2013. "The Romanian Verbs of Perception *vedea* and *auzi*: Between Direct and Indirect Perception." *Revue roumaine de linguistique* LVIII (3): 313–327.
- Niculescu, Dana. 2013. *Particularități sintactice ale limbii române din perspectivă tipologică. Gerunziul*. București: Editura Muzeului Național al Literaturii Române.
- Rafel, Joan. 2000. "From Complementizer to Preposition: Evidence from Romance." *Probus* 12 (1): 67–91.
- Stan, Camelia. 2005. *Categoria cazului*. București: Editura Universității din București.
- SOR. 2016. *Syntax of Old Romanian*. Edited by Gabriela Pană Dindelegan. Oxford: Oxford University Press.
- Usonienė, Aurelija. 2001. "On Direct/Indirect Perception with Verbs of Seeing and Seeming in English and Lithuanian." *Working Papers* 48. Lund: Lund University, Department of Linguistics: 163–181.